Investigating novel vaccine candidates for *Clostridium difficile*

Catherine Louise Hall

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy
of the
University of London

March 2020

Department of Infection Biology

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by Medical Research Council

Research group affiliation(s): Dr Lisa Dawson
Professor Brendan Wren

Declaration

I Catherine Hall, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

[redacted]

26th September 2019

Abstract

Clostridium difficile infection (CDI) is the most common cause of nosocomial diarrhoea worldwide. Current treatment options have varied levels of success and there is no licensed vaccine. Most vaccines investigated to date use the *C. difficile* toxins, as these induce a protective immune response. However, this is only against the symptoms of CDI, therefore there is a need for antigens capable of preventing colonisation and transmission.

The results of a *C. difficile* specific pan-protein array, screened against blood samples from CDI patients and healthy controls, was analysed and a shortlist of proteins identified where the IgG antibody response was significantly higher in the control group. Three of these were expressed and purified in *E. coli,* and tested against patient samples in an ELISA. For two of the three proteins, there was a higher IgG response in the healthy control group (albeit not significant) confirming the array results.

Inactivation of three immunogenic proteins from the array in *C. difficile* R20291 revealed the putative permease CDR20291_0342, was not involved in efflux of the antimicrobials tested. Inactivation of the putative pilin protein, CDR20291_3343, did not influence surface motility, but motility was almost abolished in a flagella mutant. The putative cobalt transporter CDR20291_0330, was proposed to contribute to a putative vitamin B12 synthesis pathway, which is required for ethanolamine utilisation. Neither R20291Δ*0330* or R20291 utilised ethanolamine in the conditions tested but it was found that inactivation of one component of the two-component regulatory system in *C. difficile* 630Δ*erm* abolished ethanolamine utilisation.

Finally, mechanisms of synthesising *C. difficile* glycoconjugates using bioconjugation were investigated, using a carrier protein and portion of the R20291 flagella glycan. A short acceptor peptide was designed and found to be glycosylated at all sites with a glycan from *Campylobacter jejuni* but it was not possible to detect glycosylation with the flagella glycan from *C. difficile*.

Word count: 300

Acknowledgements

I would firstly like to express my deepest thanks to my supervisors Dr Lisa Dawson and Professor Brendan Wren. Thank you for your constant support and unwavering encouragement over the past 4 years and for sharing your expertise and enthusiasm for science. Thank you for providing me with so many wonderful opportunities throughout the PhD and for your valuable comments on drafts of this thesis.

I would also like to extend this to members of the Wren group, both past and present, who have been wonderful colleagues to spend 4 years with. A special thanks to my 210 lab buddies; Alex, Ian, Allie and Mark- you got the brunt of all my stories and questions and were there to share the highs and lows of Popmaster, it would not have been as enjoyable without you. Thank you to the glycoengineering group aka “The Sugababes”; Jon, Vanessa, Jen, Emily, Sherif, Marta, Tim and Lizzie- you have all been so generous with your time, knowledge and advice. I am also very grateful to those who contributed their time and expertise to this project outside of LSHTM, including Dr Hilary Browne and Dr Kevin Vervier at the Wellcome Trust Sanger Institute, all members of the MRC programme grant and those patients who contributed to the study and enabled this research to take place.

Finally, I would like to thank my amazing friends- Alex, Cat, Sarah and Laura for being a never-ending source of encouragement, laughs and kindness and for always being there to brighten my day. Thank you to my family, to John and Grace but most of all I would like to thank my parents. I am eternally grateful for your constant love, faith and support, without you this would have been impossible.

Table of Contents

Contents

List of Figures

List of Tables

List of Abbreviations

AMP. Antimicrobial peptide

AMR. Antimicrobial resistance

BCAA. Branched chain amino acid

BHI. Brain heart infusion

BHIS. Brain heart infusion supplemented

BSA. Bovine serum albumin

CDH. *Clostridium difficile* healthy

CDI. *Clostridium difficile* infection

CDU. *Clostridium difficile* diarrhoea unknown cause

CFU. Colony forming unit

DDM. N-dodecyl-β-D-maltoside

DTT. Dithiothreitol

EIA. Enzyme immunoassay

ECF. Energy coupling factor

EDTA. Ethylenediaminetetraacetic acid

ELISA. Enzyme linked immunosorbent assay

FMT. Faecal microbiota therapy

GDH. Glutamate dehydrogenase

GT1. Glycosyltransferase 1

GT2. Glycosyltransferase 2

GT3. Glycosyltransferase 3

HA. Haemagluttinin

HA1. Homology arm 1

HA2. Homology arm 2

HIS. Histidine

HMW. High molecular weight

HRP. Horseradish peroxidase

IPTG. Isopropyl β-D-1-thiogalactopyranoside

IVTT. *In vitro* transcription translation

LB. Lysogeny broth

LC MS/MS. Liquid chromatography–mass spectrometry

LLO. Lipid-linked oligosaccharide

LMW. Low molecular weight

MIC. Minimum inhibitory concentration

µG. Microgram

MOPS. 3-(*N*-morpholino) propanesulfonic acid

ML. Millilitre

MLST. Multi locus sequence typing

MS. Mass spectrometry

NAAT. Nucleic acid amplification test

NMR. Nucleic magnetic resonance

ORF. Open reading frame

OST. Oligosaccharyltransferase

PBS. Phosphate buffered saline

PCR. Polymerase chain reaction

PHE. Public Health England

PMC. Pseudomembranous colitis

PPI. Proton pump inhibitors

RPM. Revolutions per minute

RT. Room temperature

SBA. Soy bean agglutinin

SDS. Sodium dodecyl sulfate

SEC. Size exclusion chromatography

SLP. Surface layer protein

SNP. Single nucleotide polymorphism

SOC. Super Optimal broth with Catabolite repression

SOE. Splicing by overlap extension

T2SS. Type 2 secretion system

T4P. Type IV pili

TEV. Tobacco etch virus

TMB. 3,3',5,5'-tetramethylbenzidine

TMD. Transmembrane domain

UCI. University of California, Irvine

UoL. University of Liverpool

UndPP. Undecaprenol pyrophosphate

UTR. Untranslated region

WTSI. Wellcome Trust Sanger Institute

W/V. Weight per volume

1. Introduction

1.1. An introduction to *Clostridium difficile*

In 1974, Tedesco and colleagues found 21% of patients receiving clindamycin experienced diarrhoea, with almost half of these also presenting with inflammatory lesions along the colonic mucosa, known as pseudomembranous colitis (PMC) [1]. Three years later, the causative agent *Clostridium difficile* was implicated [2-4]. First identified in infants in 1935, the strict anaerobe *C. difficile* is a Gram-positive spore forming organism from the Firmicutes phylum of bacteria [5]. More recently, the *Clostridium* genus was restricted to *Clostridium butyricum* and its related species. As *C. difficile* is phylogenetically distant from this species, it was reassigned into the *Clostridioides* genus [6]. It is found in 4-15% of healthy adults but can cause serious disease when established as a dominant member of the gut microbiome, following disturbance to the protective gut microbiome [7]. The symptoms of *C. difficile* infection (CDI) are primarily associated with the action of Toxins A and B, which the pathogen releases into the gut, resulting in breakdown of the epithelial cell barrier [8-10]. CDI is a major cause of nosocomial diarrhoea but can have more severe outcomes, including PMC, toxic megacolon and fatalities [11, 12].

1.1.1. *C. difficile* and the intestinal microbiota

The microbiota of the intestinal tract is one of the most influential microbiomes in the human body, ranging from roles in intestinal illness such as inflammatory bowel disease through to development of the immune system [13-16]. Microbiome composition displays high inter-person variability, influenced by diet, environment and health status, but is relatively stable within individuals throughout adulthood [17-19]. Difficulties arise following community disturbance and loss in diversity, a process accelerated with advanced age, particularly in institutionalised settings [20, 21] and as a result of exposure to broad-spectrum antibiotics [22-24].

A robust gut microbiome is crucial for protection against CDI. The microbiome can complete with *C. difficile* for space, nutrients and access to the epithelial cell lining and produces antimicrobial

peptides (AMPs), which may be toxic to *C. difficile* [25]. Specific species within this microbiome are particularly important in providing protection from CDI. Louie *et al.* found include *Bacteroides* and *Prevotella* species are important in the maintenance of a healthy microbiome as patients with lower counts of these were more likely to experience recurrence of CDI [26].

Additionally, microbiome-mediated bile salt metabolism is also important, as certain bile salts can inhibit outgrowth of *C. difficile*. Primary bile salts such as glyco- and tauro- cholate and chenodeoxycholate are released into the gut by the gall bladder, for the digestion of dietary fats. They are then metabolised by the gut microbiome to release taurine and glycine, leaving behind the unconjugated cholate and chenodeoxycholate. Specific members of the gut microbiome can then further process these into secondary bile salts, deoxycholate and lithocholate [27]. Bile salts promote germination of *C. difficile* spores but interestingly secondary bile salts such as deoxycholate prevent their outgrowth [28, 29]. Therefore, secondary bile salts and the bacteria harbouring the enzymes to produce them are an important means of preventing *C. difficile* colonisation (although colonisation doesn't automatically equate to CDI (1.1.2.1)). Buffie *et al.* identified *Clostridium scindens* as a member of the gut microbiome that harbours the 7α-dehydroxylating enzymes required for the production of secondary bile salts. Administration of this species to mice resulted in resolution of CDI [30]. Furthermore, Mullish *et al.* recently reported bile salt hydrolases, which mediate deconjugation of glycine and taurine from primary bile salts, are found in lower proportions in those with CDI [31].

The increased understanding of the relationship between bile salts and particular species of the microbiome and *C. difficile* is now being investigated for its suitability as a treatment for CDI, with delivery of the whole microbiome via faecal transplants already in practice (1.4.4, 1.4.5).

1.1.1.1. *C. difficile* metabolism

Defined minimal media is a valuable tool for bacterial study as it enables precise manipulation of the growth environment. Early work by Karasawa *et al.* identified the growth requirements for *C.*

difficile, including a combination of amino acids, salts, trace salts, vitamins and glucose as a carbon source [32]. This study also found, for the strains of *C. difficile* they investigated, proline, cysteine, isoleucine, leucine, valine and tryptophan were all essential for growth [32]. Many of these amino acids can be used in Stickland fermentation, a primary energy source for *C. difficile* involving the oxidation and reduction of amino acids [33]. Minimal media has been used to elucidate the nutritional relationship between toxin production and the metabolism of substrates in the host gut, such as glucose, sorbitol and ethanolamine [34][35, 36]. Theriot *et al*. used minimal media to demonstrate the range of carbohydrates *C. difficile* can utilise as a carbon source, including fructose, mannitol and sorbitol, which all increase in abundance following antibiotic ablation of the gut microbiome(i.e. a CDI susceptible state), in the mouse model [37]. Finally, in a recent study, Collins *et al.* used minimal media to support their finding that dietary trehalose in humans is associated with the emergence of hypervirulent strains of *C. difficile* [38].

1.1.2. *C. difficile* infection

1.1.2.1. Outcomes of infection and asymptomatic carriage

C. difficile's presence in the gut can result in a range of outcomes, from asymptomatic colonisation through to severe disease and fatalities. Although definitions vary, asymptomatic colonisation typically refers to patients who are positive for *C. difficile* (i.e. by faecal culture) but are negative for the detection of free Toxin A and B in stool and do not display clinical signs of CDI [7]. Current estimates indicate 4-15% of adults are asymptomatically colonised [7]. Rates of colonisation in infants under two years old tend to be higher than in adults, reaching over 50% in some studies [39-41]. However, reports of actual CDI in this cohort are rare, which has been suggested to be a result of immaturity of the infant mucosa, rendering it resistant to Toxins A and B [39, 42-44].

Colonisation with *C. difficile* is associated with increased risk of CDI [45]. In a recent review, Crobach *et al*. found the risk of developing CDI was increased in those colonised with toxigenic C. difficile [7]. Importantly, patients colonised with *C. difficile* are also a source of nosocomial transmission. High

levels of spores have been identified on the skin of these individuals which are then shed into the environment [46-50]. Lawley *et al.* found treating immunocompetent mice colonised with *C. difficile* with clindamycin resulted in disruption of the microbiome, *C. difficile* outgrowth and anincrease in sporulation. This increase in the presence of spores in the mouse faeces was termed a "super shedder" state and resulted in increased transmission to other animals [51]. As many hospitalised patients are on antibiotics, this study offers a plausible role for asymptomatically colonised patients in the transmission of CDI.

Although CDI can result in self-limiting diarrhoea, it can also develop into a severe infection leading to PMC and toxic megacolon [12, 52]. Furthermore, there are reports of systemic and extra-intestinal outcomes of CDI, including multiple organ dysfunction syndrome, liver abscesses and sepsis [53-55]. Of the total 13, 286 people diagnosed with CDI in England in 2017/18, 15.2% died within 30 days of onset [56]. Recurrence of infection is another well-established aspect of CDI and can be differentiated into relapse, i.e. infection with the same strain or re-infection, with a new isolate [57, 58]. Estimates put recurrence of CDI at 15-30% following primary infection but following two recurrent episodes, 45-65% will experience another case of CDI [59-61]. Multiple risk factors have been associated with recurrence of CDI, which often overlap with those for the acquisition of primary CDI, including antibiotic exposure and advanced age [62, 63].

1.1.2.2. Risk factors

Antibiotic use is a major risk factor for CDI, as this depletes the microbiome, resulting in loss of colonisation resistance. Although the majority of antibiotics can disturb the microbiome to some extent, clindamycin, cephalosporins and fluoroquinolones are particularly associated with development of CDI [64, 65]. The predisposition of those receiving clindamycin treatment to CDI has been known since the 1970s [1, 3, 66] but the importance of fluoroquinolones only came to light in the early 2000s, when a hypervirulent ribotype 027 strain emerged in Canada which was highly fluoroquinolone resistant [67-69]. Restriction of fluoroquinolone use resulted in a substantial

decrease in fluoroquinolone resistant infections [70]. Proton pump inhibitors (PPI) are used to reduce gastric acid production and have been implicated in increasing susceptibility to CDI, by reducing stomach acid. However, this association is controversial, as spores are the major route of *C. difficile* transmission and these are unaffected by stomach acid[71-73]. A systematic review and meta-analysis conducted by Kwok *et al.* found a probable association between PPI use and CDI, but a similar study in the same year found low quality evidence for this [71, 74].

Another predominant risk factor is advanced age (>65 years old). In the UK, the highest rates of CDI are found in people over 85 years of age [75-78]. Elderly people are more likely to experience the common risk factors of CDI, including hospitalisation, presence of co-morbidities and antibiotic use [79-83]. Furthermore, the diversity of the microbiome decreases with age, which could reduce colonisation resistance [17, 20, 21, 84, 85]. Waning of the humoral immune response with age is often mentioned as a risk of CDI due to the importance of antibodies in protection from infection, but there is a lack of evidence to fully support this (1.4.1) [86-88]. Kyne *et al.* found patients with higher IgG responses to Toxin A had reduced recurrence of infection, but IgG titres were not affected by age [89, 90]. Those with co-morbidities are also at an increased risk of CDI, both due to the fact they are likely immunocompromised and often require stays in hospital, which increases exposure to *C. difficile* [80, 83]. Illnesses, such as inflammatory bowel disease and surgery of the gastrointestinal tract are particularly associated with the development of CDI [76, 91-93].

1.1.2.3. Diagnosis

PHE recommends screening for CDI in patients with diarrhoea which cannot be attributed to an underlying condition or therapy [94]. Diagnostics for *C. difficile* target either the organism or Toxins A and B. Toxigenic culture utilises selective agar to isolate *C. difficile* from a faecal sample before testing for toxin using an established method such as PCR or cell culture. This method is sensitive and specific for *C. difficile* and allows for detection of toxin producing strains [95, 96] but is also time consuming. Therefore, enzyme immunoassay (EIA) based detection of the *C. difficile* metabolic

enzyme glutamate dehydrogenase (GDH), directly on patient samples is a more frequently utilised test for the detection of *C. difficile* [97]. Cytotoxin assays detect the toxins by exposing *in vitro* cell lines to faecal samples and monitoring cell morphology for cytopathic effects [95, 97, 98]. However, these assays are also time consuming and require cell culture facilities and therefore have been replaced by the more rapid, nucleic acid amplification tests (NAATs) and enzyme immunoassays (EIAs) [99]. NAATs are a rapid PCR-based test which usually amplify conserved regions of the genes encoding Toxin A (*tcdA*) and B (*tcdB*) but use of these alone without confirmation of actual toxin production runs the risk of over diagnosing CDI [100, 101]. EIAs allow detection of Toxins A and B in faecal samples and sensitivity of these has improved in recent years, but it is still very variable [95, 99]. Crobach *et al.* reviewed the sensitivity and specificity for the different available EIAs and NAATs for toxin detection and GDH testing and compared these to use of toxigenic tissue culture. The sensitivity was lowest and most variable for the Toxin EIA, while the sensitivity and specificity of the Toxin NAAT and GDH EIA were comparable and both >80% [101]. Detection of the organism cannot differentiate between toxigenic and non-toxigenic strains, therefore both organism and toxin-specific tests are required to confirm CDI [101]. PHE recommends initial testing with a GDH EIA or Toxin NAAT and positive results followed up with a Toxin EIA to confirm CDI [94].

1.2. Epidemiology of infection

1.2.1. Typing and phylogeny

C. difficile is an extremely diverse species with many strains that can be grouped into distinct clades. *C. difficile* harbours a large, mosaic genome which enables it to occupy different niches. The environmental challenges of occupying these niches both in and out of the gut are believed to have been the driving force behind the highly diverse *C. difficile* strains seen today. There are a number of mechanisms behind this diversity, particularly the presence of transposable elements, which are particularly prevalent within the *C. difficile* genome (~11%) [102]. Additionally, there are regions where mutations are more common, for example within the PaLoc locus which encodes the genes

required for toxin production. Toxinotyping is a method of typing strains based on mutations within this region [103].

A number of techniques are utilised to differentiate strains based on different factors. Multi-locus typing (MLST) is a well-established typing system throughout bacteria and is based on sequencing conserved house-keeping genes [104]. These have been identified for *C. difficile* and used to assign a sequence type to different strains [104, 105]. An alternative to this is toxinotyping, a typing method based on the genes encoding Toxins A and B, *tcdA* and *tcdB*, respectively [103, 106]. One of the most heavily relied upon systems is ribotyping, which groups strains based on the pattern of amplified intergenic regions of the 16S and 23S ribosomal RNA [107, 108].

Stabler *et al.* used MLST of 385 diverse strains to assess evolutionary relatedness of these (Figure 1.1). Five clades were identified, each including clinically relevant strains in humans [109]. Clade 1 contains strain 630 (ribotype 012) which was isolated from a patient with PMC in 1982 and was the first *C. difficile* genome to be fully sequenced and annotated, and therefore is a routine lab strain [110, 111]. Clade 2 contains ribotype 027 strains, including the epidemic BI/NAP1/027 strains which are highly virulent and fluoroquinolone resistant [68, 112, 113]. Clades 3 and 4 harbour ribotypes 023 and 017, respectively. Ribotype 017 strains are notable as they are Toxin A-B+ but still cause infection in humans, while ribotype 023 strains have increased in recent years and are capable of causing severe infections [114, 115]. Finally, strains from clade 5 are very distinct from the other clades and include the ribotype 078 strains which are also associated with hyper-virulence and are frequently isolated from animals [116, 117].

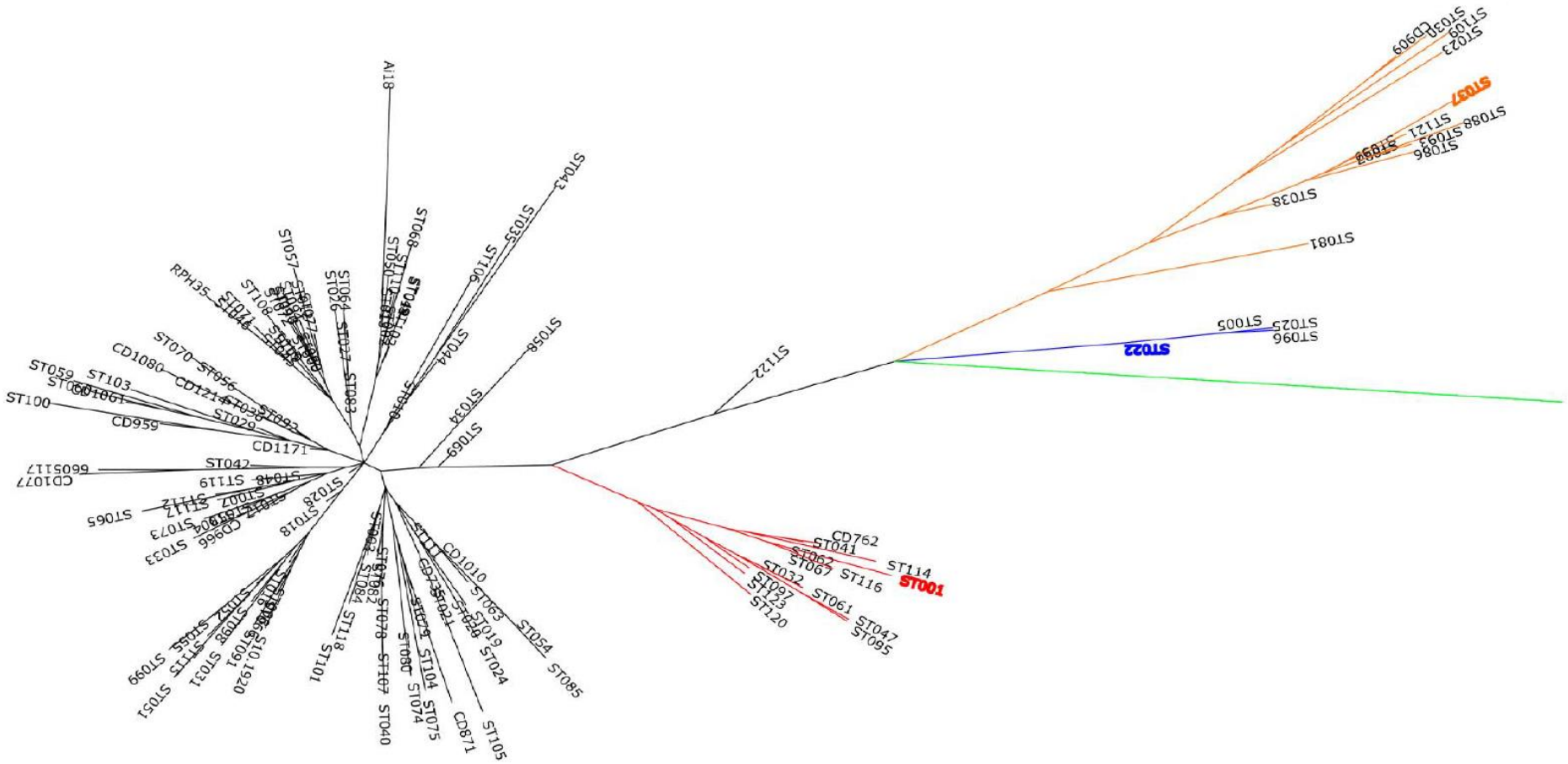


Figure 1.1. Evolutionary relatedness of *C. difficile* strains. Clustering of strains by multi-locus sequence typing resulted in 5 distinct clades which all include strains capable of causing human infection. Clade 1 (black), clade 2 (red), clade 3 (blue), clade 4 (orange), clade 5 (green). Image taken from Stabler *et al.* [109].

1.2.2. Changing pattern of CDI

Historically, *C. difficile* was considered a serious yet manageable infection, even following the rising incidence of PMC in the 1970s, mirroring the beginning of widespread antibiotic use. However, in the early 2000s, reports began of increased incidence and severity of CDI in Quebec, Canada, which quickly spread to cover North America and arrived in the UK with the highly publicised Stoke Mandeville epidemic . The causative agent, an emerging hypervirulent BI/NAP1/027 strain was isolated and identified as a ribotype 027 strain [69, 112, 118-121]. R20291 is the hypervirulent 027 strain isolated from the Stoke Mandeville epidemic [113]. Multiple features of this strain were associated with its increased severity of infection and spread, including increased toxin production, increased spore germination and a single nucleotide polymorphism (SNP) in DNA gyrase, conferring high-level fluoroquinolone resistance [68, 122-125]. This is of particular importance as the use of fluoroquinolones was widespread at the time. He *et al.* identified two distinct lineages (FQR1 and FQR2) of the epidemic BI/NAP1/027 strain which acquired the same SNP in DNA gyrase independently [68]. Increased toxin production was suggested to be a result of modifications within TcdC, the negative regulator of Toxin A and B expression, but there are conflicting reports over the influence of this [122, 126, 127]. Furthermore, comparing pre and post epidemic strains of BI/NAP1/027 found there was no difference in the sequence of the Pathogenicity Locus (PaLoc) encoding the toxins and their regulators and accessory genes [68].

In 2007/8, over 55,000 cases of CDI were reported in the UK, with over 8000 deaths, 5 times higher than those caused by MRSA [67]. More recently, these figures have reduced but CDI remains a persistent burden, particularly in healthcare associated settings, with over 12,000 cases reported in the UK between April 2018 and March 2019 [128]. Although both hospital and community onset cases of CDI have reduced, including those caused by ribotype 027 strains, this reduction is much less rapid for community onset cases, which now account for 66% of annual diagnoses [128].

The increase in prominence of community infections changed the view of *C. difficile* as a predominantly nosocomial pathogen, although definitions of community infection can vary and do not always exclude recent hospital contact [83, 129-131]. Those with community acquired CDI tend to be younger and harbour fewer risk factors than nosocomial patients [131]. In terms of the source of infection, environmental contamination with *C. difficile* spores has been reported numerous times [132-134]. Furthermore, *C. difficile* has also been detected in a number of animals, including household pets and farm animals, but whether this a source of transmission to humans is still under investigation [135-138]. One study isolated strains from cats and dogs being kept as pets, which are known to also infection humans, including those from ribotype 078 and 027. However, no simultaneous cases of CDI were reported between the pet and their owner [139].Ribotype 078 strains are particularly common in animals [117, 140]. Dingle *et al.* recently demonstrated 76.5% of sequenced ribotype 078 strains, which are frequently isolated from animals, carried the *tetM* gene, which confers tetracycline resistance, an antibiotic frequently used in agriculture [141]. Ribotype 078 strains began to emerge in 2005 and cause similar levels of severe diarrhoea and mortality as hypervirulent ribotype 027 strains [116, 142, 143]. However, compared to these strains, 078 infections are more often found in younger people and as a cause of community infections [116].

The changing epidemiology described above primarily relates to Europe and North America, where formal surveillance programmes were introduced following the ribotype 027 epidemic of the 2000s. Such surveillance programmes are not universal, particular in low resource areas where other infections can pose a greater burden, which hampers understanding of the global epidemiology of CDI. Furthermore, this can result in *C. difficile* being labelled a low prevalence pathogen when in fact this is due to incomplete testing, typing and reporting [144]. A recent meta-analysis noted a lack of incidence data from Latin America, Eastern Mediterranean, South-East Asia and Africa [145]. Perhaps unsurprisingly, another study found guidelines for management and infection control of CDI were either lacking or completely absent in these same regions [146]. In Asia, including China and South Korea, it has been documented that 027 strains did not pose the same burden as they did in

the West but instead, clade 4 ribotype 017 strains, which are typically Toxin A negative and Toxin B positive, were prominent [147]. A recent phylogenetic analysis of these strains found these in fact emerged in North America and reached Asia via Europe [148]. CDI has been identified in a number of African countries but these tend to be small studies with limited associated data such as sequence type or ribotype of the strains [149]. One study in South Africa found 50% of the 32 strains recovered from patients were ribotype 017 [150].

1.3. Virulence factors of *C. difficile*

1.3.1. Toxins

The major virulence determinants of *C. difficile* are Toxins A and B, which are the primary mediators of the symptoms of CDI [53]. These toxins are composed of an enzymatic glucosyltransferase domain (including a catalytic and cysteine protease domain), translocation and receptor binding domain, which all have a functional role in toxin activity [151]. Upon secretion from *C. difficile*, the receptor binding domains of Toxins A and B bind intestinal epithelial cells, resulting in host-mediated endocytosis [152, 153]. Potential receptors have been identified for both Toxins A and B, namely a glycoprotein (gp96) on the surface of colonocytes for Toxin A and the polio virus-like 3 receptor on colonic epithelial cells for Toxin B [154, 155]. Once internalised, the pH drop in the acidified endosome initiates a conformational change of the toxin resulting in pore formation within the endosome membrane and release of the glucosyltransferase domain into the cytosol [156, 157]. This domain can then go on to glucosylate Rho and Rac GTPases, resulting in disruption of the actin cytoskeleton and breakdown of the epithelial barrier [158, 159]. Toxins A and B are encoded by *tcdA* and *tcdB*, respectively, within the 19.6 Kb PaLoc [111, 160]. Three other genes are also present within this locus; *tcdR, tcdC* and *tcdE.* TcdE is a holin-like protein thought to mediate release of the toxins into the gut lumen [161]. TcdR is a positively regulator of toxin production and TcdC was initially believed to be an anti-sigma factor that repressed toxin expression, but this has since been debated [126, 162-164].

Some strains such as those from ribotype 027 also harbour the ADP-ribosylating binary toxin, which is encoded separately to the PaLoc [118, 165]. This toxin is also internalised and acts on the actin mesh under the cell membrane, resulting in protrusions from the cell that increase adherence of *C. difficile* [166-168]. The role of binary toxin in infection is not as pronounced as for Toxins A and B but some studies have found an association between its presence and an increased severity of infection or occurrence of relapse [169, 170]. However, in a hamster model of infection, Toxin A- Toxin B- Binary+ strains were able to colonise but did not cause symptoms of CDI [170].

The relative contributions of the *C. difficile* toxins in disease remains a source of debate. Lyras *et al.* found inactivation of Toxin A resulted in death of hamsters in a similar manner to the wild-type infection, whereas inactivation of Toxin B resulted in a significant increase in survival. Therefore, the authors concluded only Toxin B was essential for infection [10]. However, two following studies by Kuehne *et al.* reported both Toxin A only and B only strains can cause fulminant disease in the hamster model [8, 9]. Continuing on from this, Carter *et al.* used three different animal models to evaluate the influence of the toxins and found virulence was only attenuated upon inactivation of Toxin B [53]. From a human infection perspective, Toxin A-B+ strains have been isolated frequently and can cause CDI, but Toxin A+B- infections are very rare [114, 171, 172]. Finally, there are circulating Toxin A and B negative strains which can colonise the host but are not believed to cause infection in humans, although a toxin negative strain did result in mild caecal pathology in a hamster model of infection [173, 174].

1.3.2. Sporulation and Germination

As a strict anaerobe, survival of vegetative cells outside of the gut or indeed any anaerobic environment would not be possible, therefore *C. difficile* produces highly resistant spores which can tolerate aerobic conditions, high temperatures and environmental desiccation [64, 132, 175-177]. Spores are the major route of *C. difficile* transmission, and only germinate once in the gut, a process influenced by bile salt composition (section 1.1.1) [28-30, 178]. Recent work has identified the spore

proteins CspC and CspA as the receptors detecting bile salt germinants [179, 180]. It has also been suggested that spores may be a source of the high recurrence of CDI, persisting within the gut until conditions are suitable for germination and outgrowth [178, 181]. The difficulty in spore eradication contributes to the rapid spread of CDI within healthcare settings [182, 183]. Furthermore, as discussed above, people asymptomatically colonised with *C. difficile* shed spores into their environment, a process which is enhanced in mice upon antibiotic treatment, which also contributed to transmission of *C. difficile* [46-48, 51]. Deakin *et al.* demonstrated that the master regulator Spo0A is essential for sporulation, and its inactivation still enables acute infection in mice but relapse of infection is lost, as is host to host transmission [178].

1.3.3. Motility

1.3.3.1. Flagella

Flagella are whip-like structures that extend from the cell surface and propel bacteria towards nutrient sources or away from potential stressors. They are frequently associated with virulence, particularly for adherence and colonisation in the host. *C. difficile* flagella facilitate swimming through semi-solid agar *in vitro* which is lost upon inactivation of the flagella, through mutation of the major structural subunit (FliC) or flagella cap (FliD) [184-187]. Other phenotypes attributed to the *C. difficile* flagella differ between strains. For example, inactivation of the flagella in 630 increased adherence to Caco2 cells *in vitro* but the opposite was found for strain R20291 [186, 188]. Dapa *et al*. found removal of the flagella resulted in reduced biofilm formation in R20291 [189].

Post-translational modification of the flagella also influences its function. Glycan modification of FliC for strains 630 (Type A) and R20291 (Type B) has been characterised in detail and revealed two distinct structures [184, 187, 190] (Figure 1.2). This variation in glycan structure is reflected in the glycosylation gene content of the flagella locus from different *C. difficile* strains [185]. Glycosylation of the flagella is essential for motility in both strains. Furthermore, modification or loss of these glycans increases cell aggregation, which is not seen with removal of the entire flagella [184, 187].

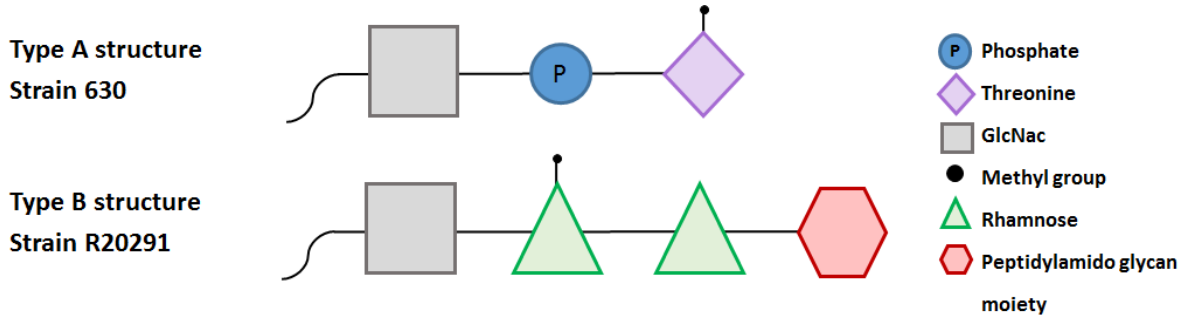


Figure 1.2. Flagella glycosylation in *C. difficile* 630 and R20291. Schematic of the solved glycan structures of the Type A and Type B modifications on the flagellum subunit, FliC. This is modified from an image by Dr Alexandra Faulds-Pain [184, 187].

Faulds-Pain *et al.* found wild-type 630Δ*erm* colonised a mouse model to the same extent as 630Δ*erm* with no flagella or harbouring a truncated flagella glycan. However, *C. difficile* was detected in the faeces for fewer days post-inoculation for the glycosylation mutant, suggesting this does play a role in colonisation [184]. Regarding the flagella directly, Dingle *et al.* found 630Δ*erm* lacking a flagella was more virulent in a hamster model and suggested increased toxin production could be behind this [186].

The link between the flagella operon and toxin production was found to be the sigma factor SigD, encoded with the early stage flagella genes. SigD positively regulates expression of late stage flagella genes, including *fliC* and the glycosyltransferases, as well as *tcdR*, the positive regulator of Toxin A and B [191]. Consequently in a *sigD* mutant strain, expression of the flagella biosynthesis genes and Toxin genes *tcdA* and *tcdB* is reduced [191, 192]. Expression of these genes is further influenced by the regulation of *sigD*, via both cyclic-diguanylate-guanosine monophosphate (c-di-GMP) and a genetic switch controlling phase variation [192-194]. c-di-GMP is a secondary messenger with a whole array of regulatory targets, but at high level is primarily associated with promoting entry into the sessile lifestyle, including a reduction in motility and increase in pili formation and cell aggregation [195, 196]. Cellular levels of c-di-GMP are tightly controlled by the action of two sets of enzymes which respond to signals to either synthesise (diguanylate cyclases) or degrade (phosphodiesterases) this secondary messenger [197]. In 630, there are 37 of these enzymes, 31 of

which are conserved in R20291 [198]. C-di-GMP controls expression by binding to a riboswitch in the 5' untranslated region of its target gene. The binding of c-di-GMP to a class I riboswitch results in premature translational termination and consequently no production of downstream gene products [199]. Class II riboswitches undergo a conformational change when bound by c-di-GMP which enables expression of the downstream genes [198, 200, 201]. Binding of c-di-GMP to the class I riboswitch within the 5' untranslated region of the early stage flagella genes results in translational termination and consequently SigD is not produced [202, 203]. Furthermore, orientation of a genetic switch downstream of this riboswitch but before start of the flagella genes can also controlSigD, which in turn affects the production of Toxins A and B and flagella biosynthesis via phase variation [193, 194].

1.3.3.2. Pilli

Pili are filamentous structures which extend from bacterial cells and can mediate adhesion and exchange of genetic material. Type 4 pili (T4P) are a sub-type of these, found in Gram-positive organisms and are associated with gliding and twitching motility, adhesion and biofilm formation [204]. *C. difficile* is predicted to encode a major (*CDR20291_3340-3350*) and minor (CDR20291_3153-3158) T4P locus which show a low level of divergence between strains and are thought to be core components of the *C. difficile* genome [205, 206]. Additional pilins encoded elsewhere on the chromosome display greater strain variation [204, 207]. Predicted T4P structures have been identified on the surface of *C. difficile in vitro* and pili like structures have been identified in hamster models [208-210] (Figure 1.3). Although *C. difficile* is also decorated with flagella structures on the surface, these are usually thicker than pili and surface appendages could still be identified in a flagella mutant, suggesting pili are also present [195]. T4P are composed of several pilin units each with a specific role for formation of the full pilus structure. In *C. difficile*, PilA1 is the predicted major structural unit with no visible pili in a *pilA1* mutant in R20291 [211].

As with the flagella, pili expression is also regulated by c-di-GMP, but this is via a type II riboswitch, meaning expression is upregulated in the presence of c-di-GMP. This corresponds with the role of c-di-GMP, mediating the switch to the sessile lifestyle [196, 197]. In *C. difficile*, cell aggregation increases following expression of a diguanylate cyclase for c-di-GMP synthesis which is lost when PilA1 is disrupted [195]. T4P have also been demonstrated to have a role in motility on solid surfaces and initial biofilm formation [195, 212]. Recently, McKee *et al.* reported that T4P negative mutants were less able to adhere to epithelial cells *in vitro* and were cleared more quickly compared to the wild-type in a mouse model of infection [213].

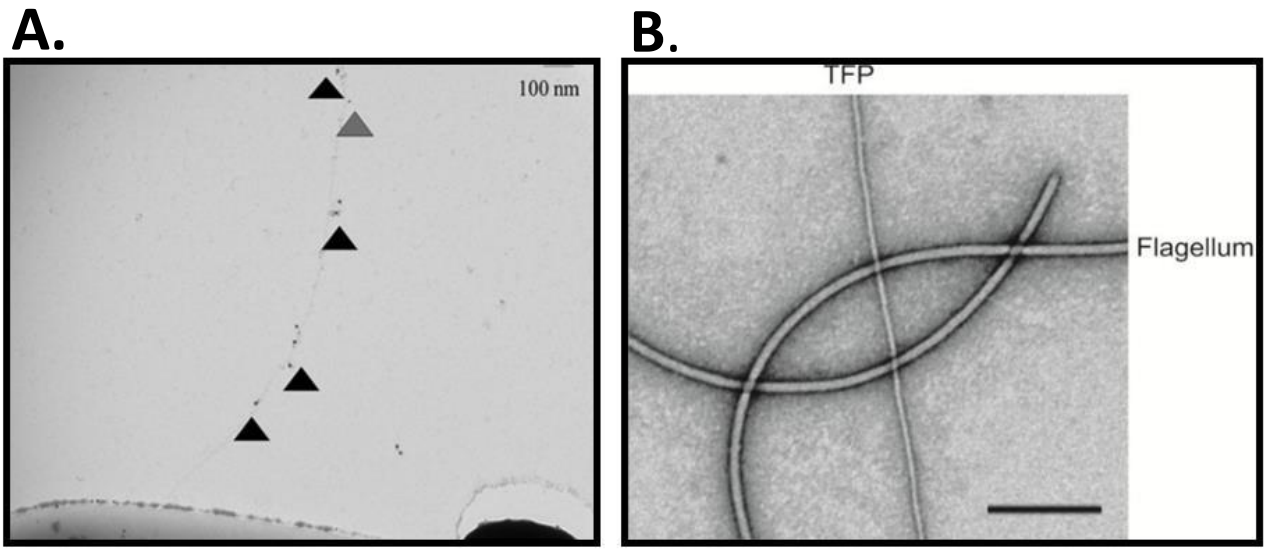


Figure 1.3. Visualisation of the Type 4 pili on the surface of *C. difficile*. (A) Immunogold staining of the PilJ and PilA units on the surface of *C. difficile* (This research was originally published in the Journal of Biological Chemistry. Piepenbrink KH, Maldarelli GA, de la Peña CF, Mulvey GL, Snyder GA, De Masi L, von Rosenvinge EC, Günther S, Armstrong GD, Donnenberg MS and Sundberg EJ. Structure of *Clostridium difficile* PilJ Exhibits Unprecedented Divergence from Known Type IV Pilins. J Biol Chem. 2014; 289:4334-4345. © the American Society for Biochemistry and Molecular Biology or © the Piepenbrink KH, Maldarelli GA, de la Peña CF, Mulvey GL, Snyder GA, De Masi L, von Rosenvinge EC, Günther S, Armstrong GD, Donnenberg MS and Sundberg EJ)[209] (B) Comparison of flagella and pili structures on the surface of *C. difficile* [208].

1.3.4. The *C. difficile* cell surface

1.3.4.1. Proteins

C. difficile harbours a dynamic cell surface. As a Gram-positive organism, it is encased within a single lipid membrane surrounded by a thick layer of peptidoglycan. Outside of this is an additional surface

layer (S-layer), a paracrystalline array that covers cell [214]. This is composed of the SlpA protein, which is expressed then post-translationally cleaved into a high molecular weight (HMW) and low molecular weight (LMW) component, by the cysteine protease Cwp84 [215-217]. The HMW protein interacts with the cell membrane and is highly conserved between strains whereas the LMW protein is exposed and variable, leading to suggestions it is involved in immune invasion [218, 219]. After much difficulty, in 2017 Kirk *et al*. isolated a strain of *C. difficile* that harboured point mutations within SlpA and consequently did not produce an S-layer [220]. This had profound effects on the bacteria including defects in sporulation and toxin production and increased sensitivity to the innate immune response. Furthermore, although this strain could colonise the hamster gut, it was not virulent [220].

In addition to the S-layer, there are a number of other surface proteins decorating the cell that can contribute to adherence. This includes collagen (CbpA) and fibrinogen binding proteins (Fbp68 and FbpA), sorted proteins (CD3392), lipoproteins and cell wall proteins [221-226]. CwpV is the largest cell wall protein and its expression is phase variable [227, 228]. Although its role in virulence and host colonisation is not well understood, CwpV does contribute to cell-aggregation and resistance from bacteriophage infection [228, 229].

1.3.4.2. Surface polysaccharides and glycosylation

Both glycosylated proteins and surface polysaccharides have been identified in *C. difficile,* although for some their purpose within *C. difficile* is poorly understood and therefore their contributions to virulence are often putative. As described above, flagella glycosylation in strains 630 and R20291 has been characterised in detail, resulting in resolution of these two distinct glycan structures and *in vitro* demonstration of their essentiality for flagella-mediated motility (1.3.3.1) [184, 187, 190]. Glycosylation of the LMW component of the S-layer is also restricted to certain strains. Characterisation of the S-layer from a ribotype 005 strain identified a complex glycan composed of three domains; a core peptide–linked tetrasaccharide, a repeating pentasaccharide and a rhamnose-

rich non-reducing end. Loss of or modification to this glycan via interrupting genes within its synthesis locus is associated with alterations in sporulation and adherence [230].

C. difficile encodes three surface polysaccharides; PS-I, PS-II and PS-III [231-233]. The latter is a lipid-bound glycosyl phosphate polymer of the lipoteichoic acid (LTA) family and is sometimes referred to as a LTA or as having LTA domains. To our knowledge, both PS-II and PS-III have been found in all strains examined so far whereas PS-I has only been found in a small number of strains so far and doesn't appear to be unique to a particular ribotype [234]. The structures for all three polymers have been solved and investigated for their immunogenicity and as vaccine candidates (1.6.3, Table 1.2) but their role in virulence is largely unknown [234-238]. Willing *et al*. identified a set of genes downstream of the large S-layer biogenesis locus which they named the anionic polymer (AP) locus. This was predicted to encode the synthesis of the hexaglycosyl phosphate repeating polymer PS-II, based on the presence of glycosyltransferases and genes related to the synthesis of surface polymers. It was not possible to inactivate genes within this locus using insertional inactivation, suggesting its essentiality. However, using anti-sense RNA knockdown experiments targeting three of the genes within this locus, disruption of S-layer assembly was demonstrated, suggesting a role for this polymer in correct assembly of this para-crystalline array [218].

1.3.5. Mechanisms for persistence within the gut

Although the mechanisms enabling *C. difficile*'s persistence within the gut are not fully characterised, a number of processes have been identified which are likely to contribute, including sporulation (1.3.2). *C. difficile* is one of few bacteria to produce the phenolic compound *para*-cresol, a result of tyrosine fermentation [239], which is toxic to some members of the gut microbiome- particularly Gram-negative bacteria [240]. As *C. difficile* itself is comparably tolerant to *p*-cresol [240-242], its production has been suggested to confer a competitive advantage over other gut bacteria and facilitate maintenance of gut dysbiosis, therefore inhibiting the recovery of microbiota-mediated colonisation resistance [10]. Indeed, Passmore *et al.* found that in a relapse mouse model of CDI, the

C. difficile colony forming units (CFU) in the mouse stool was significantly reduced in a mutant which could not produce *p*-cresol compared to the wild-type, suggesting production of this compound facilitates relapse of infection [240].

Furthermore, like many bacterial species, *C. difficile* can form biofilms- compact bacterial communities encased in a protective polymeric matrix. These have been repeatedly identified *in vitro* and mat-like structures were identified in a hamster model of infection [189, 243-246]. Biofilms have been suggested as a source of *C. difficile* relapse by maintaining a pool of viable cells that are protected from environmental stressors and only released under suitable conditions. Dapa *et al.* found within these biofilms, *C. difficile* is less susceptible to vancomycin, one of the major antibiotic therapies for CDI [189].

1.3.6. Antimicrobial resistance

Antibiotics remain the first line treatment for CDI, and metronidazole, vancomycin and fidaxomycin are all recommended in the UK (1.4.3) [75]. With the exception of fluoroquinolone resistance in epidemic ribotype 027 strains, antibiotic resistance against first line antibiotic treatments in *C. difficile* remains relatively low, and antimicrobial resistance (AMR) has not been identified as a cause of treatment failure [247]. Fidaxomycin has only relatively recently been introduced and as such AMR is very rare [248, 249]. Metronidazole resistance has increased in recent years but a study of 39 sites across 29 European countries found average metronidazole resistance rates were 0.2% over 3 years [249]. This does display geographic variation, as a study in China found metronidazole resistance rates of 15.6% [250]. The European study described above found vancomycin resistance was also low, and only found in 0.1% of isolates [249] and the mechanism of resistance was not investigated.

In order to survive and proliferate within the gut environment, *C. difficile* also encodes resistance mechanisms to antimicrobial peptides, which are secreted by both host cells and the indigenous microbiota. *C. difficile* can expel AMPs from the cell via the ABC transporters CprABC and ClnRAB.

Additionally, interaction with AMPs is reduced by the *dlt* operon, which encodes for the addition of D-alanine esterases to teichoic acids within the cell membrane, conferring a positive charge on the cell [251-253].

1.4. Prevention and treatment of *C. difficile* infection

1.4.1. The host response to *C. difficile*

The host defences play a key role in preventing CDI, and can be broadly divided into three arms; the microbiome (1.1.1), innate immunity and adaptive immunity. Breakdown of the intestinal epithelial cell (IEC) lining by Toxins A and B enables influx of commensal and pathogenic bacteria into the lamina propria. In response to this, a pro-inflammatory cytokine and chemokine cascade is initiated, to recruit immune cells such as macrophages and mast cells to the site of infection, in an effort to restore intestinal homeostasis. Studies have demonstrated increased release of IL-23, IL-8, IL-1β and IFN-γ in response to CDI, which can recruit immune cells such as neutrophils and dendritic cells [254-258]. The latter is particularly important in initiation of adaptive immunity and T-cell responses [256, 259]. Furthermore, surface layer proteins (SLPs) and flagellin proteins of *C. difficile* are also recognised via pathogen recognition receptors, TLR4 and TLR5, respectively [260-262].

This innate response acts to clear the invading pathogen, and is required for the initiation of the adaptive immune response but overwhelming inflammation can in fact exacerbate host tissue damage [263, 264]. Furthermore, leucocytosis, a sign of the systemic inflammatory response, is associated with increased severity of CDI [265]. Buonomo *et al.* demonstrated a reduction in CDI-associated mortality in two different mouse models which both lacked IL-23 signalling [266]. Overall, it appears that the strength and type of inflammation can differentiate between a protective or pathogenic innate immune response.

Although there is a T-cell component to the adaptive immune response to *C. difficile*, this is predominantly B-cell driven, especially to Toxins A and B. The toxins were believed to be

immunologically distinct with no cross protection between them but in one study, mice immunised with Toxin A only also developed anti-Toxin B antibodies [267, 268]. Understanding of the humoral immune response to *C. difficile* is largely based on comparisons between the anti-Toxin A and B antibody titres of healthy controls and those with CDI, under the assumption that higher antibody responses in healthy participants relates to protection from infection. These studies have demonstrated an important role for humoral immunity in protection from CDI, with a number of studies finding higher anti-Toxin A and anti-Toxin B antibodies in healthy control groups compared to those with CDI, or those who have a single infection compared to those who experience recurrences [89, 264, 269-271]. However, this is not a consistent phenomenon and other studies reported no difference in antibody response between healthy and patient groups [272, 273]. Warny *et al.* found patients on immunosuppressants mounted a weak antibody response to *C. difficile* but only suffered from mild CDI. Although not demonstrated in the study, the authors suggest a link back to the inflammatory response in CDI and how the suppression of the immune system dampened this response, limiting the host tissue damage associated with severe CDI [274].

Regarding antibody responses to non-toxin antigens, Pechine *et al.* found a significantly higher antibody response to the surface proteins Cwp66 and Cwp84 and the flagella proteins FliC and FliD in a healthy control group compared to patients with CDI [273]. As a gut pathogen, a robust mucosal response, mediated primarily by secretory IgA, is presumably important for protection from infection. There is evidence for a mucosal antibody response to *C. difficile* but these studies are limited, in part due to the challenges in obtaining appropriate samples to measure this response [274-276]. Oberli *et al.* (2011) screened the PS-II surface polysaccharide against stool supernatants from 10 hospitalised patients with and without CDI. Anti-PS-II antibodies were detected in 6 patient samples, three of which had CDI or were borderline for infection [235].

1.4.2. Healthcare Practice

Since mandatory reporting began in in 2007/08, there has been a 77.9% reduction in cases of CDI. This has largely been attributed to changes in healthcare practices, which focused on reducing antimicrobial prescribing and improved infection control [277]. Furthermore, these were enforced by fining NHS Trusts for cases of CDI [278].

A number of infection control practices were implemented including hand washing with soap, improved *C. difficile*-specific decontamination of rooms and equipment, patient isolation and protective clothing for healthcare workers but it appears that targeting antibiotic use has had the greatest effect on the reduction of hospital-acquired CDI [277, 279-281]. A study of US hospitals by Kazakova *et al.* found an over 20% reduction in the administration of fluoroquinolones and 3rd and 4th generation cephalosporins resulted in an 8% and 13% reduction in cases of hospital-acquired CDI, respectively. [282]. Similar findings have been reported for the UK [280, 283]. Dingle *et al.* reported that it was fluoroquinolone restriction specifically that resulted in the decline of CDI, via the substantial reduction of fluoroquinolone-resistant cases that were a major problem at the time [70].

1.4.3. Antibiotics

Antibiotics are the first line treatment for CDI. Current UK guidelines recommend metronidazole for mild and moderate cases of CDI, while vancomycin and fidaxomycin are reserved for severe infections. There is a lack of definition as to what constitutes severe infection but current markers include raised blood creatine levels and white cell counts, a temperature over 38.5 ^{o}C and severe colitis [75]. Although vancomycin is superior to metronidazole for treating severe CDI [57], both antibiotics are broad-spectrum, meaning their use can result in collateral damage to the gut microbiome, which delays the restoration of colonisation resistance [24, 284]. This has been suggested as a contributing factor to the high rates of relapse following antibiotic treatment (1.1.2.1) [285].

Fidaxomycin is narrow-spectrum and recommended for treating recurrent infections or those at high risk of recurrence [75, 286-289]. Louie *et al.* compared treatment of CDI patients with vancomycin and fidaxomycin and found perturbation of the microbiome following treatment in the vancomycin group only, where recurrence of infection was also higher (23% compared to 11% for fidaxomycin [26]. This supports the notion that preservation of the microbiome is important in the prevention of recurrent infection. However, recurrence in 11% of cases is still concerningly high and similar levels of recurrence post-fidaxomycin treatment have been reported elsewhere, demonstrating the challenges in antibiotic management of CDI [287]. Fidaxomycin is also more expensive than vancomycin, costing £1350 for a 10-day course as opposed to £189 to £378 for vancomycin [290]. Interestingly, a modelling study conducted from the perspective of NHS Scotland found the total cost of the two treatments was actually quite similar when factoring in the reduced need for hospitalisation in the fidaxomycin treatment group. They determined that fidaxomycin was cost-effective for treating severe and first recurrence CDI [290].

1.4.4. Faecal microbiota therapy

For those experiencing recurrent CDI, faecal microbiota therapy (FMT) offers an effective alternative to antibiotics. Here, faecal extracts from healthy donors are administered to patients with chronic CDI to repopulate the gut and reinstate colonisation resistance [289, 291, 292]. van Nood *et al*. reported 81% of patients with recurrent CDI were free from infection 10 weeks following administration of vancomycin with bowel lavage and nasoduodenal administration of FMT, as opposed to 23% in the vancomycin plus bowel lavage only group [293]. Additional studies have also demonstrated the superiority of FMT over antibiotics alone, even fidaxomycin, for the treatment of recurrent CDI [294-296]. The British Society of Gastrology recommends FMT for patients with two recurrent episodes of *C. difficile*, or a single recurrent case but at risk of further recurrence [297].

Logistically, delivery of FMT is more complex than for antibiotics. It requires careful sourcing and screening of faecal samples to prevent transmission of pathogenic material to the patient. Infections

arising from FMT have so far been rare but there have been cases of norovirus and *Escherichia coli* bacteraemia [298-300]. Isolation, screening and administration of fresh faecal matter in a timely manner is challenging, but this can be overcome by sample freezing, which does not influence treatment efficacy [301]. Finally, a number of routes of administration have been trialled, including using nasogastric and nasointestinal tubes or colonoscopy [294, 295]. These have all proved effective delivery systems but there are concerns over conducting such invasive procedures in patients who are already unwell. Youngster *et al.* and Hirsch *et al.* both used orally ingested capsules for delivery of FMT and reported clinical cure rates of 90% and 89% for CDI patients, respectively. However, in some cases repeat administration of the FMT capsules was required and both studies had very small sample sizes- maximum 20 patients [302, 303].

1.4.5. CDI therapies in development

Following on from FMT, there has been interest in the identification of a "minimal microbiota" that can be given to the patient to restore colonisation resistance. Lawley *et al.* administered six different gut bacteria to mice and found restoration of a healthy microbiome and clearance of CDI [304]. As discussed (1.1.1), Buffie *et al.* administered *C. scindens* to mice which resolved CDI [30], but this approach should be pursued with caution, due to the possible link between bile salts and gastrointestinal cancers [30, 305]. Alternatively, there has been promising progress in the use of non-toxigenic *C. difficile* strains to prevent CDI, which have now entered clinical trials [306-308]. A Phase II trial administered non-toxigenic *C. difficile* to patients who had already completed metronidazole or vancomycin treatment for primary CDI. The rate of relapse in the group colonised with non-toxigenic *C. difficile* was 2% as opposed to 31% in the colonised group [307]. One possible drawback of this approach is the opportunity for transfer of the PaLoc locus from toxigenic strains in the gut, resulting in the treatment strains becoming toxigenic and therefore capable of causing infection [309].

Bezlotoxumab is a humanised monoclonal antibody raised agaisnt Toxin B [310]. Monoclonal antibodies can be delivered to the patient for immediate protection agaisnt CDI and doesn't rely on the immune status of the host, an important note considering how many CDI patients are unwell or immuncompromised. Bezlotoxumab has been evaluated in Phase II and III clinical trials and has been licensed for use in Europe [310-313]. It is specifically targeted for the recurrence of infection. A study in Finland found of those receiving bezlotoxumab, 73% did not experience recurrent infection within the three months following treatment, and found this was as effective as using FMT for recurrent infections [314]. Using probiotics to protect against CDI is not currently recommended in the UK, due to the lack of evidence of their efficacy and need for large scale evaluations [75]. The Cochrane Library recently published a systematic review of the use of probiotics for prevention of CDI. Meta-analysis of 31 randomised control trials found moderate evidence that probiotics are effective at preventing *C. difficile*-associated disease [315]. Finally, Nale *et al.* found bacteriophage therapy delays the onset of symptoms of CDI in a hamster model of infection [316].

1.5. Vaccine Development

Immunisation is one of the most successful public health interventions ever implemented and mass immunisation programmes are credited in saving approximately 2.5 million lives, annually [317]. Immunisation can provide long-term or even lifelong protection from infection, reduce carriage and transmission and in some cases even result in total eradication of the pathogen, as was the case for smallpox in 1980, and the animal pathogen rinderpest in 2011 (https://www.who.int/csr/disease/smallpox/en/) [318]. The basis of vaccination involves presenting the immune system with all or part of the target pathogen, to enable generation of a protective immune response without development of infection. There are many different vaccine designs available, which can be grouped into live attenuated, where the whole organism is presented but not capable of causing fulminant infection, or inactivated which can be the whole cell or components of it (Table 1.1).

Type of vaccine	Description	Licensed example
Live attenuated	Modified version of the whole organism which can no longer cause full infection	Measles, mumps and rubella
Whole cell	Killed whole organism so no longer infectious	Influenza
Toxoid	Inactivated or de-toxified toxins normally released by the pathogen during infection	Tetanus and diphtheria
Recombinant	Expression of an immunogenic antigen from the target pathogen in another organism that is safer to handle and generates high yields of the desired antigen	Human papilloma virus
Glycoconjugate	Immunogenic polysaccharide conjugated to a protein carrier to enable development of immunological memory	*Haemophilus influenza* Type B and meningitis C

Table 1.1. Different types of vaccination used for human infections. The information in this table was obtained from "The Green Book", the UK guide on immunisation, printed by the Government and updated annually (https://www.gov.uk/government/collections/immunisation-against-infectious-disease-the-green-book#the-green-book).

1.5.1. Glycoconjugate vaccines- use and production

Bacterial cell surfaces are decorated with a host of glycans, many of which are immunogenic and therefore attractive vaccine candidates. However, polysaccharides alone induce a T-cell independent response, which does not result in immunological memory and is particularly weak in young children, the elderly and the immunocompromised [319-322].

This can be overcome by covalent attachment of the polysaccharide to a carrier protein [323]. The resulting glycoprotein can be processed by antigen presenting cells and presented within the major histocompatibility complex II, enabling T-cell recognition. This induces a robust T-cell mediated response, including antibody isotope switching and the generation of immune memory [324-326]. The carrier proteins used to present the glycans are typically well-characterised immunogenic proteins, such as the inactivated diphtheria or tetanus toxin [327]. A more recent approach uses an immunogenic carrier protein from the same species as the glycan, to boost the organism-specific immune response, or from a different species, to produce a multipathogen vaccine [238, 328-330].

The development of glycoconjugate vaccines have been an invaluable contribution to the immunisation field and has resulted in the licensing and wide-spread use of vaccines against *Haemophilus influenza* Type b and meningitis C, among others [331]. These vaccines are synthesised using chemical conjugation, whereby the individual protein and glycan components are produced and isolated separately before being covalently joined (Figure 1.4) [332]. Both before and after conjugation, all components are subject to a series of purification steps, to remove any contaminating material. This technique is vital for the production of the currently available glycoconjugate vaccines but does carry certain limitations. These include cost and yield loss resulting from the multiple rounds of purification, batch to batch variation and working with pathogenic bacteria to isolate immunogenic material [332, 333]. As a result of this, alternative means of synthesising these vaccines is of great interest, particularly bioconjugation, whereby bacterias' innate ability to synthesis and conjugate complex glycans is re-appropriated for vaccine synthesis.

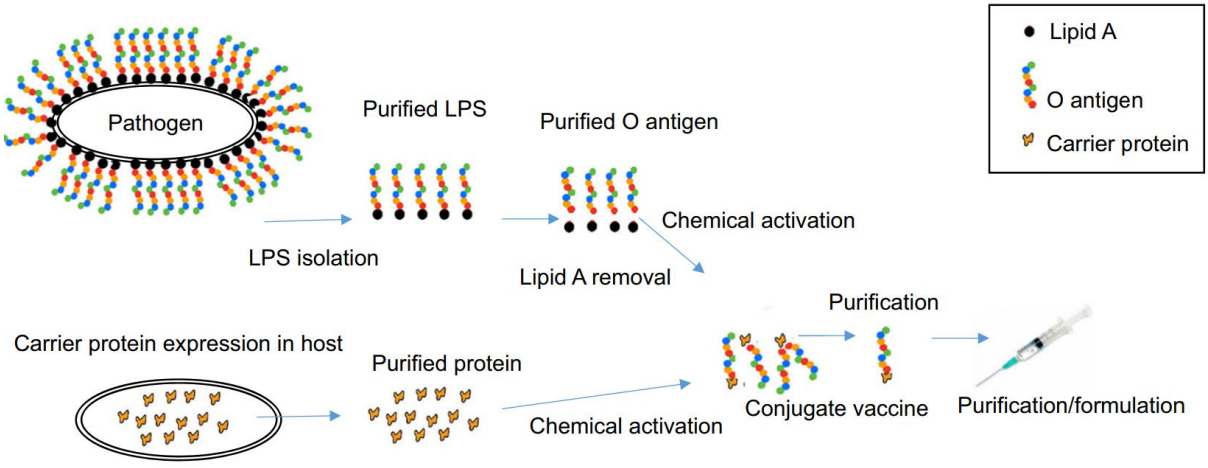


Figure 1.4. Chemical conjugation of glycoconjugate vaccines. The glycan and protein components of the vaccine are first produced separately, either recombinantly or in the native host. These are then isolated, purified to remove contaminating material then conjugated together before a final round of purification. Image from Kay *et al.* [332] (no changes were made to the original image (http://creativecommons.org/licenses/by/4.0/)).

1.5.1.1. Bacterial Glycosylation

Bacterial glycosylation is a relatively recent discovery, with glycoproteins originally believed to be restricted to higher organisms. Investigations into prokaryotes has identified glycosylation

machinery and glycoproteins in a diverse range of bacterial species, and this is a continuing area of research [334, 335]. Glycoproteins can serve a multitude of functions for the cell, including conferring motility, colonisation of the host and immune evasion [184, 187, 336-339]. Glycans have been identified on the FliC protein of the flagella and S-layer proteins in *C. difficile*, but this is not universal to all strains (1.3.3.1, 1.3.4.2) [184, 185, 187, 230].

Glycosylation is one of the most predominant post-translational modifications of bacterial proteins, and bacteria employ a number of mechanisms to achieve this. Firstly, glycoproteins can be *N*-linked or *O*-linked, primarily characterised by the attachment of the glycan to an asparagine or serine/threonine residue within the acceptor protein, respectively. Furthermore, glycan synthesis can either be *en bloc*, where the whole polysaccharide is fully synthesised on a lipid linker within the inner membrane, prior to conjugation, or can be built via the addition of monosaccharides directly onto the carrier protein, sequentially [335, 340-344].

The glycosyltransferase class of enzymes are central to these processes, and catalyse both the assembly of the glycan and its transfer to the target protein. In relation to glycan synthesis, the individual specificity of the glycosyltransferase dictates which monosaccharides are incorporated into the glycan and the glycosidic linkages by which they are joined [345, 346]. Oligosaccharyltransferases (OSTs) are a type of glycosyltransferase which mediate attachment of the glycan to the acceptor protein. These also hold their own specificities including the type of glycan they can transfer and the glycosylation sequence they recognise in the acceptor protein [347-351].

Examples of two distinct glycosylation systems can be found in *Campylobacter jejuni. O*-linked glycans are often associated with flagella and are an important component of flagella assembly and function [336, 352, 353]. Indeed, the FlaA flagellin subunit from *C. jejuni* is *O*-glycosylated and the glycan is assembled sequentially, directly onto serine and threonine residues within FlaA before it is trafficked to the surface of the cell [335, 354-356] (Figure 1.5b). The mechanism of glycosylation for

the flagella has not yet been characterised in *C. difficile* but is proposed to follow a similar method to the one described here.

The *N*-linked glycosylation pathway in *C. jejuni* was the first example of a bacterial *N*-linked glycosylation system [357, 358] (Figure 1.5a). Synthesis of the *C. jejuni* heptasaccharide utilises the *en bloc* mechanism of glycan assembly, whereby the sugar is built onto the undecaprenol pyrophosphate (UndPP) lipid linker on the cytoplasmic face of the inner membrane, to produce lipid-linked oligosaccharide (LLO). This LLO can then be trafficked to the periplasm by the flippase enzyme, PglK. Once in the periplasm, the CjPglB OST transfers the glycan onto the asparagine residue within the target acceptor protein [335, 342, 359-361]. CjPglB recognises a specific sequon within the acceptor protein that includes the target asparagine residue, D-X-N-Y-S/T, where X and Y can be any amino acid except proline [348]. *N*-linked glycosylation systems harbouring enzymes orthologous to PglB have now been identified in a number of bacteria, including *Helicobacter, Sulfurovum* and *Deffibacter* [362, 363]. A similar system has not been identified in *C. difficile* although there are genes within the anionic polymer locus for synthesis of the surface polysaccharides that have putative functions similar to those within this system, including a putative flippase [218].

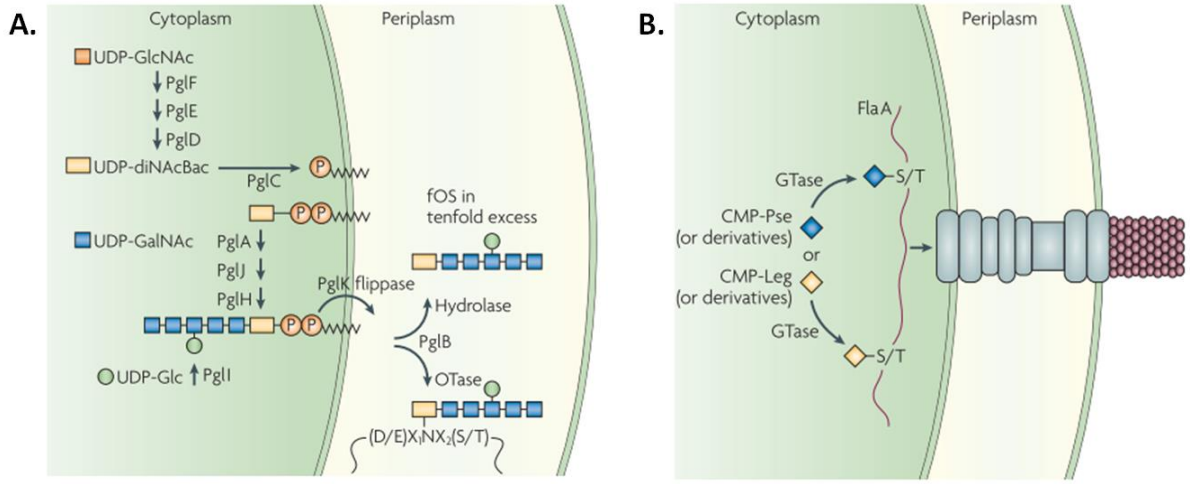


Figure 1.5. *Campylobacter jejuni* *N*-linked and *O*-linked glycosylation pathways. (A) The *N*-linked glycosylation locus from *C. jejuni.* Here, the glycan is assembled onto the undecaprenol pyrophosphate (UndPP) lipid linker at the cytoplasmic face of the inner membrane. The UndPP-bound glycan is then transferred to the periplasm by the flippase PglK, before it is picked up by the oligosaccharyltransferase PglB and conjugated to a specific asparagine residue within the acceptor protein. (B) The *O*-linked glycosylation system from *C. jejuni* involves sequential addition of the individual monosaccharides onto the FlaA subunits of the flagella, before it is trafficked out of the cell. Image from Nothaft and Szymanski [335]. Reprinted by permission from Copyright Clearance Centre: Springer Nature, Nature Reviews Microbiology, Protein glycosylation in bacteria: sweeter than ever, Nothaft H and Szymanski CM, 2010.

1.5.1.2. Bioconjugation

Bioconjugation harnesses the naturally occurring bacterial glycosylation pathways described above and repurposes them for glycoconjugate vaccine synthesis. The idea behind this is to build a "bacterial vaccine factory" which can overcome some of the shortfalls of chemical conjugation already discussed (1.5.1). Bioconjugation is often conducted in *E. coli* as the host organism, which offers a number of advantages over working with the target pathogen include, safety and ease of culture and the number of existing tools available for genetic manipulation.

The development of bioconjugation and in particular the use of CjPglB is a result of a series of pioneering studies conducted in the mid-2000s. In 2002, Wacker *et al.* reconstituted the entire *C. jejuni pgl* locus in *E. coli* and demonstrated glycosylation of the acceptor protein AcrA [341] which was followed by further in depth characterisation of the biosynthesis locus [342]. Feldman *et al.* demonstrated CjPglB mediated transfer of O-antigens from *E. coli* and *Pseudomonas aeruginosa*,

which are distinct from its native heptasaccharide substrate. This demonstrated the promiscuity of CjPglB, a fundamental component of biocojugation. Further studies then defined the requirements of CjPglB, including an accessible D-X-N-Y-S/T recognition sequon within the acceptor, which contains the asparagine residue to be glycosylated and the need for an acetoamide group at the reducing end of the glycan substrates [347, 348, 364].

Based on this understanding, it has become possible to build a diverse range of glycoconjugates which in its simplest form, can be achieved with plasmid mediated expression of three components; the carrier protein, the glycan synthesis locus and the OST, which conjugates the two together. How this work in practice when co-expressed in the same host is depicted in Figure 1.6. The glycan is synthesised onto the undecaprenol pyrophosphate lipid linker within the inner membrane then transferred into the periplasm by a dedicated flippase enzyme. CjPglB can then recognise the lipid-linked glycan and transfer it to the provided acceptor protein, which harbours the recognised glycosylation sequon. This results in a synthesised glycoconjugate which can be purified from the cell [332, 333, 365].

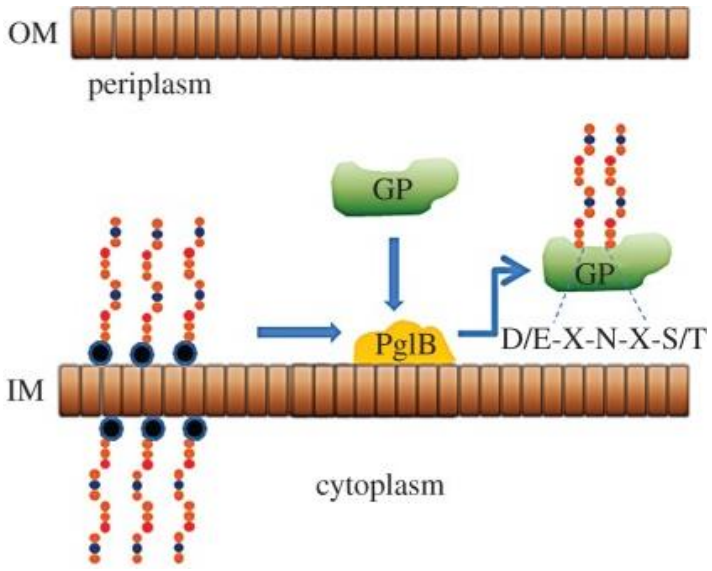


Figure 1.6. Production of glycoconjugates using *C. jejuni* PglB-mediated bioconjugation. The glycan biosynthesis locus is expressed in the host cell to enable its synthesis on the undecaprenol lipid linker within the inner membrane. This is then transferred into the periplasm by a specific flippase enzyme. Once in the periplasm it can be recognised by the oligosaccharyltransferase PglB from *C. jejuni* which transfers the glycan from the lipid linker onto an asparagine residue in the D-X-N-Y-S/T glycosylation sequence, within the carrier protein. GP, glycoprotein, IM, inner membrane, OM, outer membrane. Image from Cuccui *et al.* [365].

Obviously within this, the requirements of CjPglB have to be adhered to, including the D-X-N-S/T recognition sequon and presentation of the polysaccharide on a lipid linked donor. Regarding the acceptor protein, some may contain natural glycosylation sites, such as *C. jejuni* AcrA, but others have to be modified first, either through mutagenesis to introduce internal glycosylation sites or through the addition of a series of glycosylation sequons at the N- and/or C-terminus of the protein, known as a glycotags [366]. Furthermore, as a lipid-linked glycan is required, the glycosylation locus transferred into the host cell, must encode the requisite components to assemble this, including glycosyltransferases to build the glycan onto UndPP and a flippase to transfer it to the periplasm.

There are now multiple examples of CjPglB-based glycoconjugate synthesis, including for *Francisella tularensis, Streptococcus pneumoniae, Burkholderia pseudomallei,* and *Staphylococcus aureus* [365, 367-369]. Importantly, this has also evolved from academic research into real world applications, with bioconjugation-synthesised vaccines against *S. pneumoniae* (NCT03303976), extra-intestinal *E. coli, Shigella flexnerii* and *Shigella dysentariae* now in Phase I clinical trials [370-372].

1.6. Vaccine development in *C. difficile*

All current *C. difficile* management approaches are unified by their requirement for further interaction with the gut microbiota and lack of long-term protection. The ideal model of CDI prevention would circumvent these issues and may be achievable through immunisation. This would not necessarily be a vaccine administered universally, but instead would target those harbouring risk factors that increase their susceptibility to infection, similar to the method already applied for the flu and pneumonia vaccines [373-375]. People with increased risk for acquiring CDI include the elderly, those on certain antibiotics and people in, or with planned stays in hospital [376, 377]. Additionally, this could be used in those who have already suffered from an episode of CDI to prevent recurrence of infection [378]. Increased interest in a *C. difficile* vaccine has also emerged from the AMR field, where vaccination has been targeted as a means of reducing AMR. This currently focuses on those pathogens where AMR is considered the greatest burden, which includes *C. difficile* [317].

1.6.1. Analysis of the humoral immune response

Identification and analysis of antigenic proteins, for example during vaccine candidate discovery, can be laborious but is necessary before assessment in animal models of infection. Enzyme-linked immunosorbent assays enable measurement of the humoral immune response to antigens of interest by binding the protein to the bottom of a 96-well plate then exposing it to the test sample which is detected with a labelled secondary antibody. Although widely applied and very useful, ELISAs are laborious and therefore unsuitable for mass screening for multiple antigens and samples. In recent years the same principle has been scaled up to a protein array format, whereby proteins of interest are bound to a glass slide then probed with the same antibody-containing patient sample simultaneously [379, 380]. This platform enables whole proteome screening of organisms of interest, with patient samples (usually sera or plasma), offering a rapid means of profiling the humoral immune response to these proteins and from there identifying potential vaccine or diagnostic candidates. Furthermore, combining these arrays with a cell free transcription translation system can enable rapid and high-throughput antigen production, which would otherwise be a rate limiting step [380-383]. These arrays have now been used for a multitude of infectious diseases, including but not limited to; *N. meningitidis, Burkholderia pseudomallei, Brucella melitensis, human papilloma virus* and *Salmonella Typhi* [384-389].

1.6.2. Infection models for *C. difficile*

Animal models have enabled study of *C. difficile* in the host, including identification of the factors involved in colonisation and pathogenesis and as a means of evaluating novel antimicrobials or vaccine candidates prior to human trials.

Hamster models have been in use since the 1970s and are a favourable infection model as the course of CDI within them shares similarity with that seen in humans, including the development of diarrhoea, intestinal injury, pseudomembranous colitis and death [390]. They have enabled elucidation of a variety of aspects of CDI, including the effects of different strains on the host,

treatment of CDI and the role of the toxins in disease, although this remains controversial (1.3.1) [8, 10, 173, 210, 391, 392]. Prior to inoculation with *C. difficile* spores, clindamycin is administered to induce gut dysbiosis and provide an environment suitable for *C. difficile* outgrowth and the development of CDI. The same process is less effective in mice, believed to be a result of a more robust gut microbiota. To overcome this, Chen *et al.* used an antibiotic cocktail that induces sufficient gut dysbiosis that following treatment with clindamycin, mice developed CDI [393]. As mouse models are less sensitive to *C. difficile,* it is possible to use them to study other aspects of infection, such as colonisation and relapse [240, 394-396]. Furthermore, these models can enable *in vivo* assessment of vaccine candidates, including analysis of antibody functionality towards a specific target, such as neutralising activity.

1.6.3. Current vaccine development for *C. difficile*

Using immunisation to manage CDI has been of interest since the late 1980s, when investigations into the protective effects of Toxins A and B first began [397]. Since then, there has been substantial research in both animal models and humans, dedicated to characterising the immunogenicity of these antigens and into their use within a *C. difficile* vaccine, using whole inactivated toxins or specific immunogenic domains (1.3.1) [267, 276, 398-406]. Leuzzi *et al.* demonstrated immunisation with both toxins was required to provide optimal protection from infection [267]. Both anti-*C. difficile* vaccines in active clinical trials are based on the use of both Toxins A and B.

In 2017, the *C. difficile* vaccine development field suffered a major setback when Sanofi Pasteur's Phase III clinical trial of Cdiffense, based on formalin inactivated Toxins A and B, was terminated prematurely (NCT01887912) [407]. This was due to interim results indicating it was unlikely the trial would meet its primary objective of preventing primary CDI. Consequently, Sanofi Pasteur have now halted their *C. difficile* vaccine development programme (https://www.sanofi.com/en/media-room/press-releases/2017/2017-12-01-22-00-00). Pfizer are currently undertaking Phase III testing of their detoxified Toxins A and B vaccine, but as mutation of the glucosyltransferase domains did

not sufficiently reduce toxicity, these are also chemically detoxified (NCT03090191). Previous Phase II trials of this vaccine found it was safe and immunogenic within their target population of 65-85 year olds (NCT02561195) [408, 409]. A chimeric vaccine consisting of a single fusion protein combining the binding domains of Toxins A and B is being developed by Valneva and has completed a Phase II trial (NCT02316470) which demonstrated this vaccine was safe and immunogenic [410]. However, concerns have been raised over how removal of other immunogenic domains of the toxin would influence the overall immune response [375].

There has also been interest in alternative mechanisms of toxin delivery, including the use of DNA vaccines and expression of Toxins A and B from *Vibrio cholera*, *Salmonella typhimurium* and *Lactococcus lactis* [411-415]. Permpoonpattana *et al*. demonstrated the production of specific serum IgG and faecal IgA antibodies to the C-terminal repeats of Toxin A, when presented on the surface of *Bacillus subtilis* spores and orally administered to mice. Furthermore, 6/8 hamsters orally immunised with these recombinant spores survived challenge with *C. difficile* strain 630 [268]. One drawback of this approach is the risk of these spores germinating within the gut, leading to release of genetically modified organisms. This offers a novel delivery system with the opportunity of induction of both the systemic and mucosal antibody responses and is being prepared for Phase I clinical trials. Only one study has investigated the binary toxin as a potential vaccine candidate. Secore *et al.* (2017), found that administering the binary toxin with Toxins A and B increased survival in immunised hamsters when challenged with a NAP1 strain of *C. difficile*, when compared to use of Toxins A and B alone [416].

Despite significant effort into the production and testing of toxin-based immunisation, there remains no commercially available *C. difficile* vaccine, and termination of the Cdiffense trial so late in development highlights the challenges faced. Additionally, a frequently raised concern over the toxin approach is the lack of protection from colonisation, meaning shedding and transmission of infection can continue [267, 403]. Understandably, there are now increased calls for use of a combination of

antigens that can protect against symptomatic infection, colonisation and possibly even against spores.

Investigations into antigens that could offer this has been almost exclusively restricted to pre-clinical development, using antigens from a variety of sources, including spores and the cell surface (Table 1.2). Immunisation of mice with S-layer protein resulted in a reduced faecal bacterial count [417] and serum antibodies were raised in hamsters sub-cutaneously immunised with the cell wall protein Cwp84 [395]. All three surface polymers have been investigated and found to induce a PS-specific antibody response in a mouse model of infection, when chemically conjugated to CRM_{197}, a non-toxic version of the diptheria toxin and a widely employed carrier protein within glycoconjugate vaccines [235, 418-420]. Interestingly, PS-II conjugated to the Toxin B glucosyltransferase domain induced a PS-specific antibody response while also retaining immunogenicity of the toxin [238]. This is promising for the combination of toxin and non-toxin antigens in future vaccine formulations. Only PS-III-CRM_{197} has been evaluated in challenge experiments, where colonisation was significantly reduced in immunised mice compared to controls [420]. Matrivax are currently testing a combined toxin and PS-II vaccine in Phase I clinical trials (https://www.matrivax.com/c-difficile-vaccine).

Two of the most promising antigens emerging from pre-clinical development are the *C. difficile* exosporium proteins, CdeC and CdeM. These conferred protection in mouse and hamster models of infection (higher dose in the hamster model) and were delivered via the intra-peritoneal route with an aluminium adjuvant. 10/10 mice and 10/10 hamsters survived challenge with *C. difficile* following immunisation with CdeC, compared to 9/10 of mice and 8/10 hamsters immunised with CdeM [421].

It is notable that the functionality of the antibodies detected to the cell surface antigens above was not assessed and therefore their role in an opsonophagocytic or neutralisation response is unknown. The above immunisation was delivered parentally. Although *C. difficile* is a gut pathogen and therefore the gut immune response is very important, the studies described above demonstrating protection from CDI using toxin fragments was achieved using parental administration. Furthermore,

attempts to provide protection and generate antibody responses using oral vaccination of non-toxin antigens have had limited success [422-424]. 50% of hamsters survived challenge with *C. difficile* spores following oral immunisation with pectin beads loaded with FliC, compared to 17% of hamsters which were immunised with beads only. However, these hamsters did not demonstrate increased serum anti-FliC antibody responses and no FliC specific antibodies could be detected in their faeces [422].

	Target	Formulation	Model, dosage, antibiotic for dysbiosis	Route	Measure of immunogencitiy	Outcome	Ref.
Surface poly-saccharides	PS-I	PS-I-CRM_{197} + FA	Mouse (C57BL/6), 3	N/S	Antibody titre	Anti-PS-I serum IgG response	[419]
	PS-I and PS-II	PS-I and PS-II free polysaccharide	Pregnant sow, 2	IM	Antibody titre	Anti-PS-I and anti-PS-II IgM response	[234]
	PS-II	PS-II-CRM_{197}	Mouse (C57BL/6), 3	SC	Antibody titre	Serum anti-PS-II IgG response	[235]
		PS-II-CRM_{197} + MF59	Mouse (BALB/c), N/S	N/S	Antibody titre	Serum anti-PSII IgG response.	[236]
		PS-II chemically conjugated to binding domain of Tox A or glucosyltransferase domain of Tox B, both + MF59	Mouse (BALB/c), 3	IP	Antibody titre	Serum PS-II-Tox B IgG response comparable to using PS-II-CRM_{197} (lower for PS-II-Tox A). Antibody response lower in PS-II-Tox A glycoconjugate compared to Tox A alone. Toxin neutralising antibodies detected for both.	[238]
	PS-III	LTA of PS-III-ExoA or HSA, Rabbit FA, mouse SA	Mouse (BALB/c) and rabbit, 3	Mouse-IP Rabbit-SC	Antibody titre	Serum anti-PS-III IgG response	[237]
	PS-III	Portion of PS-III-CRM_{197} alone or + FA or Alum	Mouse (C57BL/6), 3, clindamycin	SC	Antibody titre and CFU for *C. difficile* colonisation	Serum anti-PS-III IgG response and significant reduction in *C. difficile* colonisation upon challenge.	[420]
Cell surface	FliC	Recombinant FliC in pectin beads or alone	Hamster, 3, clindamycin	O, IP, IR	Antibody titre, faecal cell count, survival	50% hamsters survived when orally immunised with FliC loaded pectin beads as opposed to 17% with beads only. Using FliC only, 0%, 17% and 33% survived challenge when immunised by the O, IR and IP route, respectively. No surviving hamster had detectable *C. difficile* in faeces but only the increased serum anti-FliC response was for IP-administered FliC only.	[422]

	Target	Formulation	Model, dosage, antibiotic for gut dysbiosis	Route	Measure of immunogencitiy	Outcome	Ref.
Cell Surface	FliC	Recombinant FliC alone or with recombinant receptor binding domains of Tox A and B. All + Alum	Hamster, 3, clindamycin Mouse (C57BL/6), 3, antibiotic cocktail* + clindamycin	IP	Antibody titre, stool spore count and survival	Serum anti-FliC antibodies. Increased anti-Tox A antibodies when administered with FliC. All mice immunised with FliC +/- Tox A and B survived challenge and had significantly lower stool spore count compared to controls. 5/14 hamsters immunised with low dose FliC died following challenge, 8/14 died in high dose.	[425]
	FliD	Recombinant FliD adsorbed to *Bacillus subtilis* spores alone or also expressing CotB- human IL-2	Mouse (BALB/c), oral 9 and IN 8	O and IN	Antibody titre	No anti-FliD antibodies detected following O immunisation. IN immunised mice had significantly higher serum anti-FliD IgG compared to controls but no difference in IgA in gastrointestinal tract. Serum IgG highest in spore expressing FliD and CotB-IL-2.	[424]
		Recombinant FliD and flagella preparation + CT	Mouse (C3H), 3, amoxicillin + clavulanic acid	IR	CFU for *C. difficile* colonisation	6 days post-challenge: colonisation higher in the immunised compared to controls, by day 30 colonisation in the immunised group was significantly lower than controls	[426]
	FliD + Cwp84	Recombinant FliD and recombinant Cwp84 +CT	Mouse (C3H), 3, amoxicillin + clavulanic acid	IR	CFU for *C. difficile* colonisation	6 days post-challenge: colonisation higher in the immunised group compared to the control group. By day 30 colonisation in the immunised group was significantly lower than controls.	[426]
	Cwp84	Recombinant Cwp84 + FA (SC) or CT (R)	Hamster, 3, clindamycin	SC, IR, IG	Antibody titres and CFU for *C. difficile* colonisation	Serum antibodies to Cwp84 highest via SC route, no significant difference in antibody response when immunised by R or IG route compared to pre-immunisation.	[395]
	GroEL	Recombinant GroEL + CT	Hamster, 3, clindamycin Mouse (C3H), 4, cefoxitin	Hamster-IR, Mouse-IN	Antibody titres, colonisation in mouse, survival in hamster	No difference in survival between control and immunised for hamster. Colonisation significantly reduced in immunised mice 10 days post-challenge, also had significantly higher serum anti-GroEL antibody response	[396]

	Target	Formulation	Model, dosage, antibiotic for gut dysbiosis	Route	Measure of immunogencitiy	Outcome	Ref.
Cell Surface	SLP	Recombinant SlpA + CT or recombinant *C. difficile* FliC	Mouse (C57BL/6), 3, antibiotic cocktail	IR	Faecal bacterial count	Faecal bacterial count significantly reduced 10 days post-challenge for immunised compared to controls, no difference between adjuvants	[417]
		Purified SLP containing equimolar amounts of LMW and HMW proteins + Alum, Ribi or CT	Hamster,3, clindamycin Mouse (BALB/c), 3	IP or IN	Antibody titres, intestinal damage and survival	Mice: immunised via IP with SLP+CT or Ribi developed serum anti-SLP IgG, more modest response for serum IgA. Mucosal IgA no different from controls. Hamsters: immunised via IP route with SLP+Alum, developed serum anti-SLP IgG but all died on challenge. Immunised via IP with SLP+Ribi- 2/5 survived challenge or SLP+Ribi+CT- 2/3 survived- all had erosion of epithelium.	[427]
	Cell wall extract	Cell wall extract presumed to include Cwp66 and S-layer protein + CT	Mouse (C3H)	IR	CFU for *C. difficile* colonisation	6 days post-challenge: colonisation higher in the immunised group compared to the control group. By day 30 colonisation in the immunised group was significantly lower than controls.	[426]
	Membrane fraction	Membrane fraction from a strain of non-toxigenic *C. difficile*	Mouse (C3H), 3, cefoxitin Hamster, 3, clindamycin	Mouse-O, IR, IN and SC Hamster-IR	Antibody titres and colonisation by CFU count in mice, survival for hamsters	Anti-membrane fraction IgG antibodies detected in sera for all routes of immunisation except O, with highest via SC. Overall lower intestinal IgA response, highest via IR. Hamster survival 7 days post challenge was 2/40 for controls and 14/40 for IR immunised.	[423]
Spores	CdeC, CdeM, BclAI, SleC, CotA	Recombinant CdeC, CdeM, BclAI, SleC, CotA, all + Alum	Mouse (C57BL/6), 3, antibiotic cocktail Hamster, 3, clindamycin	IP	Antibody titres, colonisation and survival	Serum IgG response for all proteins. Best protection 15 days post-challenge for CdeC (100% survival) and CdeM (90% survival).	[421]

Table 1.2. Pre-clinical development of non-toxin antigens for inclusion within a *C. difficile* vaccine. LTA, lipoteichoic acid, CFU, colony forming units, SLP, surface layer protein, HMW and LMW are the high and low molecular weight products of the cleaved S-layer protein, SlpA. Carrier proteins; CRM_{197}, detoxified diphtheria toxin, regions of *C. difficile* toxins A and B and HSA, human serum albumin. Adjuvants; MF59, squalene-based, CT, non-toxic derivative of the cholera toxin, Alum, aluminium hydroxide, SA, Sigma adjuvant, FA, Freund's adjuvant and Ribi, oil in water emulsion of bacterial cell wall components. Route of administration; IP, intraperitoneal, IR, intrarectal, SC, subcutaneous, O, oral, IG, intragastric, IM intramuscular, N/S not stated. *Antibiotic cocktail consists of kanamycin, gentamycin, colistin, metronidazole and vancomycin.

1.7. Aims

This research aimed to use the output of a novel pan-proteome array to identify reactive *C. difficile* antigens in both *C. difficile* patients and healthy age-matched controls. The results of this array were used to identify novel immunogenic vaccine candidates and characterise these to determine their function within *C. difficile*. Additionally, this study aimed to develop novel bioconjugation technology to produce a *C. difficile*-specific glycoconjugate, as proof of principle for *C. difficile* vaccine synthesis using this system.

1. Identify potential *C. difficile* vaccine candidates using data from a high throughput pan-proteome array.

- Analyse data resulting from a high-throughput *C. difficile* pan-proteome array designed to detect antibody response to *C. difficile* proteins with samples from infected patients compared to matched controls.
- Identify potential vaccine candidates, with a higher response in healthy controls compared to CDI cases.

2. Characterise novel vaccine candidates for *C. difficile in vitro* and validate their immunogenicity.

- Make clean gene inactivation mutants in these strains using allele exchange mutagenesis and characterise the phenotype of these mutants *in vitro* based on their putative function.
- Clone, express and purify these vaccine candidates for validation of their immunogenicity with patient samples using indirect ELISA assays.

3. Design and construct a protein carrier for bioconjugation with the *C. difficile* flagella glycan.

- Design a carrier protein suitable for bioconjugation and optimise *in vitro* glycosylation of the carrier protein using the oligosaccharyltransferase PglB and the heptasaccharide from *C. jejuni.*

4. Employ bioconjugation to synthesise the flagella glycan from the hypervirulent *C. difficile* strain R20291 on a suitable acceptor protein.

- Design, test and optimise an *E. coli* system to facilitate the production of the *C. difficile* flagella glycan and its conjugation onto a carrier protein.
- Use a potential vaccine candidate identified on the pan-protein array as an acceptor protein for bioconjugation with the flagella glycan.

2. Materials and Methods

2.1. Materials

All oligonucleotides were ordered from Integrated DNA Technologies (IDT) and enzymes from New England Biolabs (NEB), unless otherwise stated (Appendix A). Plasmid construction is detailed in Appendix B.

2.2. Growth conditions

Strains used in this work are listed in Table 2.1. All strains were stored in 20% glycerol (Sigma Aldrich) at -80°C.

2.2.1. *C. difficile* growth

Clostridium difficile strains were routinely cultured on brain heart infusion supplemented (BHIS) agar; brain heart infusion (BHI) agar (Oxoid) supplemented with 0.5% w/v yeast (Sigma Aldrich or Bacto) and 0.1% w/v L- cysteine (Sigma Aldrich), or in liquid cultures of BHIS broth; BHI broth supplemented with 0.5% w/v yeast and 0.1% w/v L- cysteine. All work with live *C. difficile* was performed at 37°C in an anaerobic atmosphere (10% CO_2, 10% H_2, 80% N_2) in a Whitley MG500 workstation and liquid cultures incubated shaking at 65 rpm. When required, the following was added; 0.1% w/v sodium taurocholate (Sigma Aldrich) for germination of *C. difficile* spores, 250 µg/ml D-cycloserine and 8 µg/ml cefoxitin (ThermoFisher Scientific) for selection of *C. difficile* and 15 µg/ml thiamphenicol (Sigma Aldrich) for selection of *C. difficile* harbouring plasmids with a thiamphenicol resistance cassette. With the exception of BHIS agar used to reactivate strains from glycerol, all media and reagents were pre-equilibrated for a minimum of 4 hours in the anaerobic workstation before use.

Strain	Details	Source
Clostridium difficile		
630Δ*erm*	Erythromycin sensitive derivative of PCR-ribotype 012 isolated in Zurich, Switzerland in 1982	[111, 428]
630Δ*erm*_*1910*::CT	630Δ*erm* with ClosTron insertion with *CD1910*, encoding EutV of the ethanolamine utilisation locus	Alexandra Faulds-Pain
630Δ*erm*_*1911*::CT	630Δ*erm* with ClosTron insertion with *CD1911*, encoding EutW of the ethanolamine utilisation locus	Alexandra Faulds-Pain
R20291	Fully sequenced PCR-ribotype 027 hypervirulent strain isolated from the Stoke Mandeville hospital epidemic	[113]
R20291Δ*0330*	*CDR20291_0330* gene deletion mutant in R20291, a putative cobalt binding protein	This study
R20291Δ*3343*	*CDR20291_3343* gene deletion mutant in R20291, a putative pilin protein	This study
R20291Δ*0342*	*CDR20291_0342* gene deletion mutant in R20291, a putative lipid transporter	This study
R20291_*fliC*::CT	R20291 with ClosTron insertion within *fliC*	[187]
R20291_*fliC*::CTΔ*3343*	R20291 with ClosTron insertion within *fliC* and clean deletion of *CDR20291_3343*	This study
E. coli		
Top10	Commercially available strain, electro-competent. F− mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 nupG recA1 araD139Δ(ara-leu)7697 galE15 galK16 rpsL(Strr) endA1 λ	Invitrogen
DH5α	Commercially available strain, chemically competent.	NEB
DH10β	Commercially available strain, chemically competent.	NEB
BL21 (DE3)	Protein expression strain: F-ompT hsdSB(rB - mB -) gal dcm(DE3)	Invitrogen
CA434	Kanamycin resistant conjugation donor strain. *E. coli* HB101 [F− mcrB mrr hsdS20(rB− mB−) recA13 leuB6 ara-14 proA2 lacY1galK2 xyl-5 mtl-1 rpsL20(Smr) glnV44 λ−] containing plasmid R702	[429]
CLM24	*E. coli* K12 strain W3110 with inactivated ligase (*waaL*)	[430]
CLM24*pglB*	CLM24 with *pglB* integrated into *cedA* on the chromosome	Jon Cuccui
O13	*E. coli* isolate harbouring the O13 antigen, from Public Health England Culture collection	Public Health England Culture Collection

Table 2.1 List of strains used in this study.

2.2.1.1. *C. difficile* minimal media

The *C. difficile* minimal media recipe used was based on that described by Karasawa *et al.* [32] and simplified by Cartman *et al.* [431] (Table 2.2). tThis was used for motility assays performed in minimal media agar. Growth kinetics in ethanolamine minimal media were performed as described by Nawrocki *et al.* [36] (Table 2.3) which uses a reduced concentration of cas-amino acids and glucose compared to the Cartman recipe. Finally, the valine utilisation experiments were performed according to Cartman *et al*., [431] but cas-amino acids were replaced with individual amino acids as described in the Karasawa method (Table 2.4) [32]. All components were dissolved in d.H_2O before sterilising with 0.22 µM syringe filter units (Merck Millipore). Iron sulphate ($FeSO_4$) solution was prepared anaerobically and added to pre-equilibrated liquid media. For solid minimal media, agar was added to the required percentage and $FeSO_4$ added prior to pre-equilibrating due to the difficulties in preparing plates in the anaerobic work station.

Component	Stock Concentration (mg/ml)	Final Concentration (mg/ml)
Amino acids		
Cas-amino acids	50	10
L-tryptophan	2.5	0.5
L- cysteine	2.5	0.5
Salts		
Na_2HPO_4	50	5
$NaHCO_3$	50	5
KH_2PO_4	9	0.9
NaCl	9	0.9
Trace Salts		
$(NH_4)_2SO_4$	2	0.04
$CaCl_2 \cdot 2H_2O$	1.3	0.026
$MgCl_2 \cdot 6H_2O$	1	0.02
$MnCl_2 \cdot 4H_2O$	0.5	0.01
$CoCl_2 \cdot 6H_2O$	0.05	0.001
Glucose	200	10
$FeSO4 \cdot 7H_2O$	0.4	0.004
Vitamins		
D-biotin	0.1	0.001
Calcium-D-pantothenate	0.1	0.001
Pyridoxine	0.1	0.001

Table 2.2. Minimal media recipe. Based on Cartman *et al.* [431]

	Stock Concentration (mg/ml)	Final Concentration (mg/ml)
Amino acids		
Casamino acids	12.5	1.25
L-tryptophan	0.625	0.0625
L- cysteine	0.625	0.0625
Salts		
Na_2HPO_4	50	5
$NaHCO_3$	50	5
KH_2PO_4	9	0.9
NaCl	9	0.9
Trace Salts		
$(NH_4)_2SO_4$	2	0.04
$CaCl_2{\cdot}2H_2O$	1.3	0.026
$MgCl_2{\cdot}6H_2O$	1	0.02
$MnCl_2{\cdot}4H_2O$	0.5	0.01
$CoCl_2{\cdot}6H_2O$	0.05	0.001
Glucose	9	0.9
Ethanolamine	14.6	1.46
$FeSO4{\cdot}7H_2O$	0.4	0.004
Vitamins		
D-biotin	0.1	0.001
Calcium-D-pantothenate	0.1	0.001
Pyridoxine	0.1	0.001

Table 2.3. Minimal media recipe for ethanolamine utilisation. Based on recipe by Nawrocki *et al.*[36].

Amino acid	Final concentration (mg/ml)
Valine	0.3
Proline	0.6
Leucine	0.4
Isoleucine	0.3
Tryptophan	0.1
Cysteine	0.5
Histidine	0.1
Arginine	0.2
Methionine	0.2
Threonine	0.2
Glycine	0.1

Table 2.4. Amino acid composition for valine utilisation in minimal media. Recipe from Karasawa *et al.*[32].

2.2.2. *E. coli* growth

Escherichia coli strains were cultured in lysogeny broth (LB) or on LB agar (Merck). LB was purchased (Merck), or prepared using 1% w/v tryptone (Sigma Aldrich), 0.5% w/v yeast and 0.5% w/v sodium chloride (Sigma Aldrich). Unless otherwise stated, all cultures were incubated at 37°C in an aerobic atmosphere, and liquid cultures shaken at 180 rpm. When required, LB was supplemented with antibiotics, as described in Table 2.5.

Antibiotic	Final concentration for *E. coli* (μg/ml)	Final concentration for *C. difficile* (μg/ml)
Thiamphenicol	N/A	15
Chloramphenicol	25	N/A
Kanamycin	50	N/A
Ampicillin	25	N/A
Spectinomycin	250	N/A

Table 2.5. List of antibiotics used in *C. difficile* and *E. coli*. N/A- not used for this bacteria.

2.3. Bioinformatics

The GENtle software was routinely used for viewing DNA sequences (http://gentle.magnusmanske.de/), designing plasmids and primers and analysing DNA sequencing results. The assembled genome sequences of *C. difficile* strains 630 and R20291 were downloaded as

EMBL genome files from the Wellcome Trust Sanger website (https://www.sanger.ac.uk/resources/downloads/bacteria/clostridium-difficile.html) and viewed using Artemis (https://www.sanger.ac.uk/science/tools/artemis) [432]. DNA and protein sequences of interest were extracted from these genome files and used for further analysis, including BLAST searches for prediction of gene function and for the identification of related genes (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome).

For putative protein analysis, the Phyre^2 Server (Protein Homology/analogY Recognition Engine V 2.0) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id) [433] and pFam 32.0 database (https://pfam.xfam.org/) [434] were used for predicting protein function and structure, and identification of conserved domains. SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP/) [435] was used for detection of possible signal peptides, TMHMM Server v 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) [436] for identification of transmembrane domains and Protein Molecular Weight for the estimation of protein size from the amino acid sequence (https://www.bioinformatics.org/sms/prot_mw.html). Peptide2 (https://www.peptide2.com/N_peptide_hydrophobicity_hydrophilicity.php) was used for hydrophobicity predictions based on amino acid sequence.

2.4. DNA manipulation

2.4.1. DNA isolation

DNA was isolated from *C. difficile* by Chelex preparation. 1.5 ml of overnight liquid culture was centrifuged at 14, 500 x *g* for 2 mins, the supernatant discarded and pellet resuspended in 300 µl 5% w/v Chelex (Sigma Aldrich). Suspensions were vortexed thoroughly before boiling then chilling on ice for 10 mins each. Following a final centrifugation step at 14, 500 x *g* for 1 min, 200 µl of the supernatant was stored at -20^{o}C. DNA concentration and purity was measured using a NanoDrop1000 spectrophotometer (NanoDrop Technologies, Inc). For DNA purification, the

QIAquick PCR Purification Kit (Qiagen) or the Monarch PCR & DNA Cleanup Kit (NEB) were used. For isolation of plasmid DNA from *E. coli*, the QIAprep Spin Miniprep Kit (Qiagen) or Monarch Plasmid Miniprep Kit (NEB) were used. All kits were used according to manufacturer's instructions and DNA products stored at -20°C.

2.4.2. Polymerase chain reaction

All oligonucleotides used are listed in Appendix A. Prior to ordering, these were checked for melting temperature and secondary structures using the NEB Tm calculator (https://tmcalculator.neb.com/) and NetPrimer (http://www.premierbiosoft.com/netprimer/), respectively. PCR reactions were performed using 1 unit of high-fidelity Phusion DNA polymerase (NEB), 1x Phusion HF buffer (20 mM Tris-HCl, 100 mM KCl, 1 mM DTT, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 200 µg/ml BSA, 50% Glycerol, 1X stabilizers) (NEB), 0.2 µM dNTPs (NEB), 0.2 µM of each primer and approx. 10 ng of template DNA. Annealing temperature and extension times were tailored to the specific reaction but the standard protocol followed; 98°C for 30 secs then 35 cycles of the following, 98°C for 10 secs, melting temperature (Tm) of the primers for 30 secs and 72°C for 2 mins then a final 4 min step at 72°C. These were run in a DNA Engine Tetrad 2 thermal cycler (Bio-Rad Laboratories Inc.).

2.4.3. Agarose gel electrophoresis

DNA samples were run on 1% w/v agarose gels containing a 1:10,000 dilution of GelRed Nucleic Acid Gel Stain (Biotium), in 1x TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA). Between 20 and 50 µl of DNA were mixed with 1x DNA Gel Loading Dye (NEB) before running on the gel alongside a 1 Kb DNA Hyper Ladder (Bioline). Routine conditions for resolving the gel were 100-120 volts for 30 to 60 mins. Gels were visualised using a Gene Genius Bio Imaging System (Syngene) and DNA purified from the extracted agarose band using the QIAquick Gel Extraction Kit (Qiagen) or Monarch DNA Gel Extraction Kit (NEB), according to manufacturer's instructions.

2.4.4. Restriction digestion

A list of plasmids used in this study can be found in Appendix A. Digestion reactions to be used for cloning contained 1x CutSmart reaction buffer (10 mM Bis-Tris-Propane-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA (NEB), 50-100 ng of insert or plasmid DNA, 20 units of each restriction enzyme and d.H_2O to a final 50 µl volume. Reactions were incubated at 37°C for 3 hours for plasmid DNA or 1 hour for insert DNA. 5' phosphate groups were removed from the digested plasmid to prevent self-ligation by adding 5 units Antarctic phosphatase (NEB) and 1x Antarctic phosphatase buffer (50 mM Bis-Tris-Propane-HCl, 1 mM $MgCl_2$, 0.1 mM $Zn_6`Cl_2$) (NEB), directly to the digestion reaction, incubating at 37°C for 15 mins then inactivating the enzyme at 65°C for 5 mins. Dephosphorylated digested plasmid was resolved on an agarose gel, extracted and purified (2.4.3) while the insert was purified directly after digestion (2.4.1). For test digests to confirm presence of an insert, reactions included 1x CutSmart reaction buffer, 500 ng/ml BSA, 5 µl plasmid DNA and 10 units of each restriction enzyme, made up to a total 20 µl reaction volume with d.H_2O.

To remove any residual backbone when amplifying large fragments from plasmid DNA, 20 units DpnI and 1x CutSmart reaction buffer was added to 50 µl PCR reactions immediately after amplification. These were incubated for 1 hour at 37°C then the enzyme inactivated with incubation at 80°C for 20 mins.

2.4.5. DNA ligation

Ligation reactions contained insert and plasmid DNA at a 3:1 concentration ratio, 1 unit of T4 DNA ligase buffer (10 mM Tris-HCl pH 7.4, 50 mM KCl, 1 mM DTT, 0.1 mM EDTA and 50% glycerol) (Promega) and 1x T4 DNA ligase (Promega). Reactions were either incubated for 1 hour at 16°C or on ice overnight then dialysed on 0.45 µM membrane filters (Millipore) in sterile d.H_2O for 30 minutes, before using immediately for *E. coli* transformation (2.6.2) or freezing at -20°C.

2.4.6. Gibson assembly

In addition to restriction/ligation cloning, plasmids were also constructed using Gibson assembly [437]. This involves the amplification of insert and plasmid DNA to include overlapping regions which can anneal following incubation with exonuclease, an enzyme which creates single-stranded 3' overhangs. DNA polymerase then completes each annealed fragment and DNA ligase seals any nicks in the assemblies.

Plasmids were constructed using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB), according to manufacturer's instructions. Briefly, the NEBuilder online tool (https://nebuilder.neb.com/) was used to design oligonucleotides for amplification of the plasmid and insert fragments with 5' and 3' overhangs. Following PCR amplification of each fragment (2.4.2) and DpnI treatment of the plasmid (2.4.4), these were resolved on an agarose gel and purified (2.4.3). 50-100 ng of the purified plasmid fragment was incubated at 50^{o}C for 15-30 mins with 2x excess of the insert fragment, 1x NEBuilder DNA Assembly HiFi DNA Master Mix and d.H_2O to a 20 µl total reaction volume. Reactions were then chilled on ice before either immediate transformation into competent cells (2.6.2) or storage at -20^{o}C.

2.4.7. DNA sequencing

DNA and plasmid samples were sequenced using the Eurofins or Source Sanger sequencing services, according to company requirements.

2.4.8. Screening of transformants and colony PCR

Colonies were screened for the desired construct using test digestion or colony PCR. For the former, single colonies to be screened were grown overnight in LB supplemented with appropriate antibiotic before plasmid isolation (2.4.1) and digestion with restriction endonucleases flanking the insert site (2.4.4). These were then visualised on a 1% w/v agarose gel (2.4.3) to determine if the released product equated to the size of the desired insert. Colony PCR was used to screen the transformants directly from the plate by resuspending each colony in 50 µl 1x sterile phosphate buffered saline

(PBS). 2 µl of resuspension was used in a 50 µl PCR reaction with d.H_2O volume adjusted accordingly (2.4.2). The initial 98°C denaturation step was extended to 8 mins to aid cell lysis and release of DNA. When possible, oligonucleotides were used that annealed to the regions of the plasmid flanking the insert to ensure that amplification was not from residual insert, carried over from the ligation reaction. Test digestion and colony PCR results indicating presence of the insert were confirmed by sequencing (2.4.7).

2.5. Manipulation of *E. coli*

2.5.1. Preparation of electrocompetent *E. coli*

An overnight culture of the *E. coli* strain to be made electro-competent was diluted 1:100 into 100 ml LB, supplemented with antibiotic as needed and incubated shaking at 180 rpm until $OD_{595nm\ of}$ 0.5 and 1 was reached. Cells were chilled on ice for 15 mins before harvesting at 14, 500 x g, 4°C for 15 mins. Supernatant was discarded and cells re-suspended in 100 ml pre-chilled d.H_2O and centrifuged as described previously. Pelleted cells were re-suspended in 8 ml pre-chilled 10% glycerol and centrifuged. Supernatant was discarded and cells re-suspended in 500 µl 10% glycerol and stored at -80°C in 55 µl aliquots.

2.5.2. Transformation of competent *E. coli*

An aliquot of competent cells were thawed on ice and mixed with approx. 1-5 ng plasmid DNA. Electro-competent cells were electroporated at 2.5 KV in a Gene Pulser Xcell (Bio-Rad Laboratories Inc.) before recovery in 500 µl super optimal broth (SOC) (0.5% Yeast Extract, 2% Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM Glucose) for 1 hour at 37°C, shaking incubation. Chemically competent NE5 alpha cells (NEB) were chilled on ice for 30 mins, incubated at 42°C for 30 secs, chilled on ice for another 2 mins then incubated with 950 µl SOC buffer for 60 mins at 37°C, shaking incubation. Following recovery, both cell types were transferred to 5 ml LB or 100 µl was plated neat or 1:10 dilution on LB agar plates. Both broth and plates were supplemented with

the antibiotic corresponding to the resistance cassette on the plasmid and incubated overnight at 37°C.

2.6. Manipulation of *C. difficile*

2.6.1. Transformation of *C. difficile*

As transformation of *C. difficile* cells requires conjugation with a competent donor strain of *E. coli*, the kanamycin resistant *E. coli* CA434 strain was used. CA434 cells were grown overnight in 5 ml LB broth supplemented with 50 µg/ml kanamycin and 25 µg/ml chloramphenicol to select for the strain and plasmid, respectively. Cells were prepared for mating by pelleting 1.5 ml of overnight culture for 2 mins at 14,500 x *g*, discarding the supernatant, then gently re-suspending cells in 500 µl 1x PBS before centrifugation for 1 min at 14, 500 x *g*. Supernatant was again discarded and cells transferred to the anaerobic workstation and re-suspended in 200 µl overnight *C. difficile* culture before plating in 20 µl aliquots on BHI agar. When needed, a heat shock protocol was used to try and increase conjugation efficiency [438]. Here, pelleted *E. coli* cells were resuspended in 200 µl overnight *C. difficile* culture immediately after the first centrifugation step, without washing the cells with PBS, and incubated at 52^{o}C for 5 minutes then 37^{o}C for 2 minutes before spot plating as described. Following 24 hour conjugation, growth was re-suspended in sterile 1x PBS and spread plated in 100 µl aliquots on BHIS agar supplemented with D-cycloserine and cefoxitin, for isolation of *C. difficile* only and thiamphenicol to select for the plasmid. Following 48-72 hours incubation, colonies were restreaked to confirm successful transformation.

2.6.2. *C. difficile* mutagenesis

Genes of interest were deleted from the *C. difficile* chromosome using allele exchange mutagenesis [439]. This approach does not require a negative selection marker and instead relies on homologous recombination to remove the gene from the chromosome. This results in the generation of in-frame deletion mutants without polar effects to downstream genes, as can occur with insertional mutagenesis methods, such as ClosTron [440, 441].

2.6.2.1. Plasmid construction for allele exchange mutagenesis

Allele exchange cassettes were built using splicing by overlap extension (SOE) PCR [442] then cloned into pMTL82151 (http://www.clostron.com/) for conjugation into *C. difficile*. Each cassette consisted of the two ~1200 bp DNA regions (homology arms) immediately flanking the gene for deletion, which enable the homologous recombination event to occur. Homology arms 1 and 2 (HA1 and HA2) were amplified from *C. difficile* R20291 genomic DNA (gDNA) to include 5' (HA1) and 3' (HA2) restriction endonuclease sites for cloning into pMTL82151 and 3' (HA1) and 5' (HA2) complimentary overhangs, comprised of the first and last 21 bp of the gene for deletion. These complimentary regions permit joining of the two constructs in a second round of amplification, using the forward primer of HA1 and reverse primer of HA2. Allele exchange constructs were inserted into pMTL82151 using restriction-ligation cloning (2.4.4, 2.4.5), with correct plasmid assembly confirmed using test digestion (2.4.4) and DNA sequencing (2.4.7).

2.6.2.2. Conjugation

Each pMTL82151 plasmid carrying a different allele exchange cassette was transformed into electro-competent CA434 cells (2.6.2) and conjugated with *C. difficile* R20291 (2.7.1). Following resuspension and spread plating of the conjugations, plates were incubated for 72 hours, before re-streaking the largest colonies 3-4 times on BHIS agar, supplemented with D-cycloserine, cefoxitin and thiamphenicol. Larger colonies were preferentially restreaked, as these were most likely to have undergone the single cross over event and therefore harbour the plasmid on the chromosome (Figure 2.1). This is due to the use of the "pseudo-suicide" plasmid, pMTL82151, which replicates poorly in *C. difficile* so better growth is obtained when the plasmid has integrated and the thiamphenicol resistance cassette is therefore replicated with the chromosome.

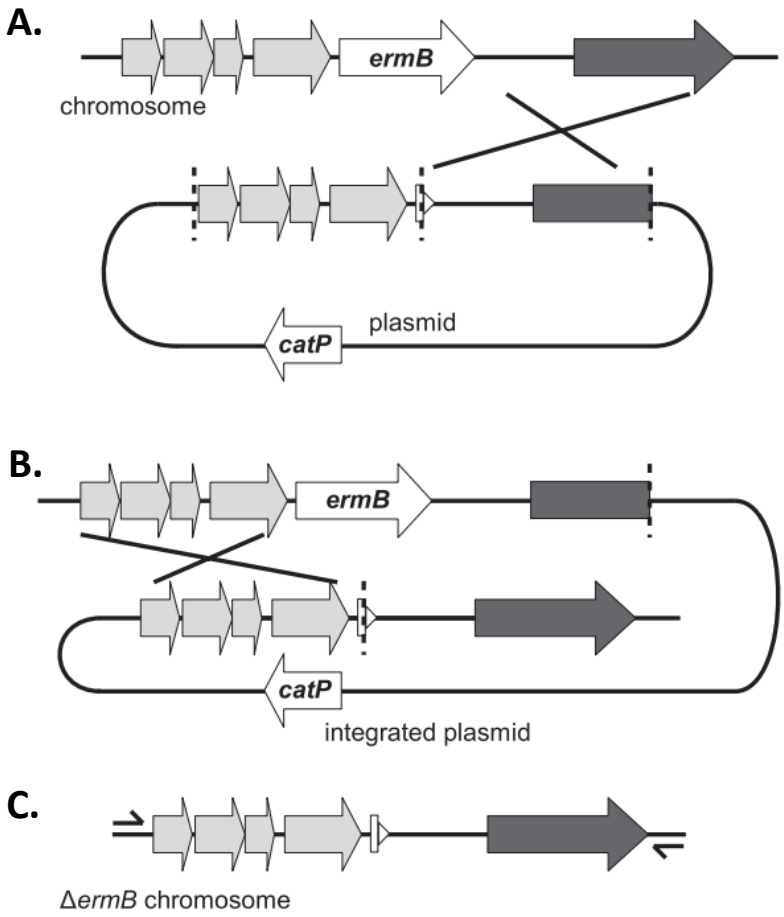


Figure 2.1. Diagram of allele exchange mutagenesis. Example of deletion of *ermB* from the chromosome using allele exchange mutagenesis. (A) Single cross event resulting from one of the regions of homology (dark grey), leading to incorporation of the whole plasmid onto the chromosome. (B) Second cross over event with the second regions of homology (light grey), resulting in deletion of *ermB* from the chromosome (C) If the second recombination event occurs at the same region of homology as the first, the plasmid is lost and the strain reverts to wild-type sequence. Image taken from Faulds-Pain *et al*. [439].

2.6.2.3. Isolation and screening of gene deletion mutants

Predicted single cross over colonies were passaged 4-5 times on non-selective BHIS agar to remove the selection pressure to maintain the plasmid and enable the second recombination event to occur (Figure 2.1). Following passage, colonies were patch plated on non-selective BHIS agar and BHIS agar containing thiamphenicol to identify those which were had lost the plasmid. Colonies demonstrating thiamphenicol sensitivity were screened by PCR to differentiate between wild-type revertants (those which have lost the plasmid without complete integration) and double crossovers mutants, using

primers flanking the gene for deletion. Primers annealing upstream of the 5' of HA1 and downstream of the 3' of HA2 were used to amplify this region for sequencing, to ensure the gene had successfully been deleted from the chromosome and results were not due to interference from residual plasmid DNA.

2.7. Phenotypic assays

2.7.1. Growth kinetics of *C. difficile*

To determine the growth rate of *C. difficile*, three colonies were used to inoculate a primary culture of 5 ml BHIS or BHI broth. Following overnight incubation, this was used to inoculate 25 cm^3 vented tissue culture flasks containing 15 ml BHI or BHIS broth to a starting OD_{595nm} 0.05. OD_{595nm} readings were taken every hour for the first 8 hours then a final reading at 24 hours using a CO 8000 Cell Density Meter spectrophotometer (Biochrom WPA). Cultures were set up in duplicate and three independent replicates were performed.

2.7.1.1. Ethanolamine Utilisation

An overnight 5 ml BHIS culture of *C. difficile* was diluted 1:100 into 10 ml fresh BHIS broth and shaking incubated to OD_{595nm} 0.6. This was back-diluted in ethanolamine minimal media (Table 2.3) containing either 0.9 mg/ml glucose (Sigma Aldrich), 1.46 mg/ml ethanolamine (Sigma Aldrich) or water (no carbon control) to OD_{595nm} 0.1. Dilutions were used to inoculate 15 ml or 20 ml minimal media to a starting OD_{595nm} 0.01. Cultures were incubated shaking for 72 hours, with OD_{595nm} taken at; 16, 20, 24, 40, 44, 48, 64, 68 and 72 hours. Cultures were set up in duplicate and assay performed in three independent replicates.

To investigate the role of cobalt in ethanolamine utilisation, an overnight 10 ml minimal broth, containing 0.9 mg/ml glucose and no cobalt was inoculated using three colonies of *C. difficile*. This was diluted 1:100 into a 10 ml minimal broth containing 0.9 mg/ml glucose and no cobalt, grown to OD_{595nm} 0.6 then back-diluted to OD_{595nm} 0.1 in media matching the final growth media, specifically ethanolamine minimal media +/- cobalt and either 0.9 mg/ml glucose (Sigma Aldrich), 1.46 mg/ml

ethanolamine (Sigma Aldrich) or water (no carbon control). This was used to inoculate a single minimal media flask per condition to OD_{595nm} 0.01, then readings taken as above. The assay performed in two independent replicates.

2.7.1.2. Valine Utilisation

Three colonises from a plate culture were used to inoculate 10 ml minimal media containing all 11 amino acids plus glucose and incubated overnight (Table 2.4). This overnight was used to inoculate fresh 15 ml minimal media to a starting OD 0.05. Minimal media was either plus or minus glucose, and contained either all 11 amino acids (Table 2.4), 6 amino acids (proline, leucine, cysteine, valine, isoleucine, tryptophan) or 5 amino acids (proline, leucine, cysteine, isoleucine, tryptophan). 200 µl of culture was removed every hour for 8 hours then at a final 24 hours for OD_{595} readings taken using a plate reader. This was performed in two independent replicates.

2.7.2. Colony forming units

In order to determine the number of colony forming units (CFUs) of *C. difficile*, liquid cultures were serially diluted 1:10 until the 1:100,000 dilution, in 1x PBS. Each sample was vortexed between dilutions. 10 µl of neat *C. difficile* and each dilution were plated in triplicate on BHIS plates (without L-cysteine) containing 0.1% w/v sodium taurocholate to ensure germination of spores. For spore counts, 1 ml of neat *C. difficile* culture was transferred to 1.5 ml microcentrifuge tubes and heated at 65^{o}C for 20 mins, before serial dilution and plating. Plates were incubated anaerobically overnight at 37^{o}C and number of colonies counted by eye to the lowest dilution where colonies could still be differentiated. Colony counts were multiplied by 100 to obtain the number of colonies per ml then multiplied by the relevant dilution factor to determine the final CFU. Vegetative cell count was calculated by subtracting the spore count from the total cell count. This was performed in three independent replicates.

2.7.3. Antimicrobial susceptibility

The sensitivity of *C. difficile* to different antibiotics and antimicrobial peptides (AMPs) was measured using the broth dilution method [148, 443] under anaerobic conditions. 24-well flat-bottom plates contained each antibiotic concentration in duplicate (doubling dilutions were prepared from 32 to 0.5 µg/ml for lincomycin and 512 to 0.5 µg/ml for bacitracin and erythromycin), a growth control (no antibiotic) and broth sterility control (no inoculant). Antibiotic stocks were prepared at either 10 mg/ml or 1 mg/ml then diluted to the desired concentration in BHI broth, before aliquotting 990 µl into each well. An overnight *C. difficile* liquid culture was diluted 1:100 in 10 ml BHI broth and incubated shaking until OD_{595nm} 0.3. 10 µl of culture was added to each well and plates were incubated shaking for 16 hours. If *C. difficile* growth had sedimented to the bottom of the plate well, this was resuspended before OD_{595nm} readings were taken using an ELx800 Absorbance microplate reader (Biotek). The minimum inhibitory concentration (MIC) determined in the first assay was confirmed with two more replicates at three concentrations; the MIC, and 2x and 0.5x this concentration (no 2x concentration was used for bacitracin and erythromycin as there was no growth at the highest concentration tested, 512 µg/ml).

2.7.4. Colony morphology

To visualise colony morphology, overnight cultures in BHIS broth were diluted 1:100 into fresh BHIS and left to grow until OD_{595nm} reached between 0.3 and 0.4. These were then diluted to between 1:10,000 and 1:100,000 in sterile 1x PBS and 100 µl of each spread plated on 1.8% BHIS agar supplemented with thiamphenicol, 1% glucose, and 0 or 25 ng/ml anyhydrotetracycline (Atc), for induction of the Ptet promoter. Following 5 days incubation, the colonies were photographed using a Canon 600D SLR.

2.7.5. Swimming motility

Swimming motility of *C. difficile* was assessed using minimal media with 0.3% agar and thiamphenicol. When required for induction of the Ptet promoter, anhydrotetracycline was added at

25, 50, 100 or 250 ng/ml. *C. difficile* colonies growing for 24-48 hours were picked using a sterile toothpick and the centre of the swimming plate stab inoculated to approximately halfway through the agar. Plates were incubated statically for 5 days then two perpendicular measurements of the diameter taken (Figure 2.1a), which were averaged. Plates were also photographed using a Canon 600D SLR. Each plate was set up in duplicate and the assay performed in three independent replicates.

2.7.6. Surface motility

Surface motility of *C. difficile* was assessed using 1.8% BHIS agar. Higher agar percentages increase surface hardness which enables the swarming movement to occur. An overnight broth culture of *C. difficile* was diluted 1:100 into fresh 10 ml BHIS and 5 µl spotted in quadruplicate onto BHIS agar plates supplemented with thiamphenicol, 1% glucose and 0 or 25 ng/ml Atc, once OD_{595nm} reached 0.3-0.4. Plates were incubated statically and 2 measurements taken for each spot every day for 5 days- at the widest diameter of the colony and perpendicular to this measurement, which were then averaged (Figure 2.1b). Plates were also photographed on day 5 using a Canon 600D SLR. The assay was performed in triplicate.

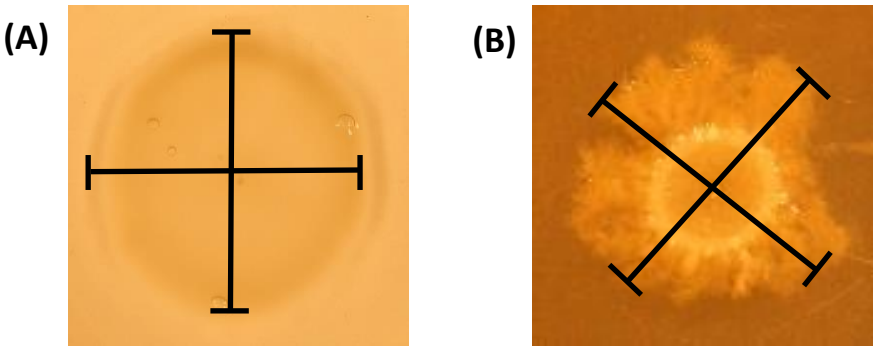


Figure 2.2. Motility measurements. Schematic of measuring diameter of bacteria in motility assays. (A) Swimming motility diameter, taking two perpendicular readings (B) swarming motility measurement, taking the diameter at the widest point and the diameter perpendicular to this.

2.7.7. Biolog

Biolog Phenotypic Microarrays assays enable high-throughput analysis of bacterial growth through comparing growth in many different conditions, simultaneously. Biolog assays were performed at the Wellcome Trust Sanger Institute by Dr Hilary Brown and Dr Kevin Vervier. In brief, these were conducted using an AN MicroPlateTM (Biolog) that was exposed to oxygen for 30 mins before transfer to the anaerobic cabinet. A plate culture of *C. difficile* was swabbed from the agar surface and resuspended in pre-reduced AN Inoculating Fluid (Biolog) then transferred to the AN Microplate, 100 µl per well. Plates were sealed in an oxygen impermeable bag then transferred to the OmniLog incubator for 24 hours, which takes automated OD readings of cell growth.

2.8. Protein techniques

2.8.1. SDS-PAGE

15 µl of protein sample was mixed with 1x NuPAGE LDS sample buffer (ThermoFisher Scientific) to a total 20 µl volume and heated at 70^{o}C for 5 mins. Standard conditions used a pre-cast 10% or 4-12% NuPAGE Bis-Tris gel (ThermoFisher Scientific) with PageRuler Prestained Protein Ladder (ThermoFisher Scientific) or PageRuler Plus Prestained Protein Ladder (ThermoFisher Scientific) in 1x 3-(*N*-morpholino)propanesulfonic acid (MOPS) SDS Running Buffer (50 mM MOPS, 50 mM Tris Base, 0.1% sodium dodecyl sulfate (SDS), 1 mM EDTA, pH 7.7) (ThermoFisher Scientific) for 50-60 mins at 180 V. For resolution of smaller proteins, 1x 2-(*N*-morpholino)propanesulfonic acid (MES) SDS Running Buffer (50 mM MES, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.3) was used. For separation and visualisation of proteins under 15 KDa, the same protocol was followed, but samples were resolved on 16% Tricine Protein Gels (ThermoFisher Scientific) in Tricine SDS Running Buffer (ThermoFisher Scientific) with SeeBlue Pre-Stained Protein Standard ladder (ThermoFisher Scientific). Gels were visualised using Western blotting (2.8.3) or by staining with Coomassie (2.8.2).

2.8.2. Coomassie Staining

Gels were incubated rocking with Coomassie Brilliant Blue R-250 stain (10% acetic acid, 10% ethanol, 1 mg/ml Coomassie) at room temperature (RT) for 20-60 mins, then destained overnight in 10% ethanol, 10% acetic acid. Gels were visualised on an Azure c600 Gel Imaging System (Azure Biosystems).

2.8.3. Fluorescent Western Blotting

The iBlot 2 Dry Blotting System (ThermoFisher Scientific) was used for transfer of samples from the SDS-PAGE gel to a nitrocellulose membrane (ThermoFisher Scientific). Transfer stacks were assembled and loaded onto the iBlot according to manufacturer's instructions and transferred for 7 mins (1 min at 20 V, 4 mins at 23 V, 2 mins at 25 V). Following transfer, membranes were washed for 5 mins in 1x PBS + 0.1% Tween20 (Sigma Aldrich) then incubated with blocking buffer (1x PBS + 0.1% Tween20 and 5% skimmed milk) for 1 hour. When blotting with the soy bean agglutinin (SBA) lectin (Vector Laboratories), milk was not included as this lectin binds galactose. All incubation and wash steps were conducted at RT on a rocker. The blocking buffer was replaced with 10 ml fresh blocking buffer containing primary antibody and incubated for 1 hour. A list of antibodies used in this work are listed in Table 2.6. Membranes were washed 3 x 5 min with 1x PBS + 0.1% Tween20 before incubation with the secondary antibody at the appropriate dilution, in the dark. The membrane was again washed 3 x in 1x PBS + 0.1% Tween20 before visualisation on an Odyssey near-infrared imager (LI-COR Biosciences). Images were scanned at either 680 nm or 800 nm wavelengths, depending on the IRdye conjugated to the secondary antibody.

Antibody	Dilution	Suppllier
Mouse anti-Histag	1:5000	AbCam
Rabbit anti-Histag	1:5000	AbCam
Goat anti-mouse IRDye conjugated 600CW	1:5000	LI COR
Goat anti-mouse IRDye conjugated 800CW	1:5000	LI COR
Goat anti-rabbit conjugated to HRP	1:20,000	Sigma Aldrich
Goat anti-human IgG	1:10,000	Sigma Aldrich

Table 2.6. List of antibodies used in this study.

2.8.4. Indirect Enzyme Linked Immunosorbent Assay (ELISA)

Indirect ELISAs were performed in Nunc Maxisorp or Medisorp 96-well flat-bottom plates (ThermoFisher Scientific). Protein samples were serially diluted to the concentrations required in sodium carbonate buffer (Na_2CO_3, pH 9.6) or 1x PBS, and 50 µl added to each well before static incubation overnight at 4°C. Wells were washed 4 x 2 mins with 1x PBS + 0.05% Tween20 before 2 hour shaking incubation at 500 rpm with 300 µl blocking buffer (1x PBS + 0.05% Tween20 + 10% milk (5% milk for human sera samples). Wells were washed once for 2 mins with 1x PBS + 0.05% Tween20 then incubated shaking for 2 hours with 50 µl of human sera sample diluted 1:00 in 1x PBS + 0.05% Tween20 + 1% milk or primary antibody. (Table 2.6). Wells were washed again 4 x 2 mins before incubation with 50 µl horse radish peroxidase (HRP)-conjugated secondary antibody in 1x PBS + 0.05% Tween20 + 1% milk for 90 mins, shaking incubation. A final 4 x 2 mins wash steps were performed with 1x PBS + 0.05% Tween20 before 100 µl of the 3,3',5,5'-tetramethylbenzidine (TMB) substrate (ThermoFisher Scientific) was added and the plates incubated in the dark for 2-15 minutes, until colour development had saturated. At this point, the reaction was stopped with 1 M H_2SO_4 and plates read at a 450 nm wavelength.

2.8.5. Protein expression in *E. coli*

For protein expression tests in *E. coli*, 5 ml LB supplemented with appropriate antibiotic was inoculated with a single *E. coli* colony or directly from a glycerol stock and incubated overnight. This was diluted 1:100 into fresh 10 ml LB supplemented with antibiotic to maintain the plasmid and grown to OD_{595nm} 0.4-0.8, before induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (ThermoFisher Scientific) or 0.02-1% L-arabinose (Sigma Aldrich). L-arabinose can be metabolised by *E. coli* so was added again before leaving the culture shaking overnight at 30°C or 37°C. Following overnight growth, 200 µl of culture was removed to analyse the whole cell lysate and the remaining culture centrifuged at 14, 500 x *g* for 5 mins at 4°C. The supernatant was discarded and pellet either immediately extracted or stored at -20°C. For large scale expression using 1 to 2 litres of LB, the same

process was followed but the initial overnight culture was increased as necessary and cells pelleted at 7459 x *g* for 30 mins at 4°C.

2.8.6. Protein extraction from *E. coli*

2.8.6.1. Ribolyser

Pelleted bacteria were resuspended in 1-2 ml binding buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10-25 mM imidazole) and loaded into Lysing Matrix B- 2 ml Bead Beating Tubes (MP Biomedicals) before ribolysing in a FastPrep-24 Classic Instrument (MP Biomedicals) for 40 secs. Ribolysed extracts were centrifuged a 14,500 x *g* for 2 mins and the supernatant stored at 4°C, to reduce protein degredation.

2.8.6.2. Cell homogeniser

Bacterial pellets obtained from the 1-2 litre cultures were resuspended in binding buffer to a maximum 50 ml volume which was diluted further in binding buffer for viscous resuspensions, before passing through the cell homogeniser, which uses high pressure to break the cells open. The homogeniser was first washed with 50 ml 20% ethanol, 50 ml 70% ethanol and 50 ml d.H_2O before 20 ml binding buffer was added. Once the last 10 ml of binding buffer was in the chamber of the homogeniser, the resuspended pellet was added, a maximum of 50 ml at a time, keeping the lysate on ice between runs. Cells were homogenised in 10 ml aliquots and the whole slurry was passed through three times or until the lysate started to become less opaque and viscous. The cell lysate was then centrifuged at 30, 966 x *g* for 1 hour before filtering the supernatant with a 0.22 µM filter and storing at 4°C. cOmplete™ mini EDTA-free Protease Inhibitor Cocktail tablets (Merck) (1 tablet per 10 ml lysate) were added to cell lysates to reduce proteolytic activity and protein degradation.

2.8.7. Protein purification

2.8.7.1. NiNTA agarose purification of His-tagged recombinant proteins

Proteins harbouring a hexa-histidine tag (His-tag) were purified using nickel affinity chromatography. This was done in three different ways, depending on the volume of sample and intended use of the

purified protein. For clarified cell lysates under 1 ml, the supernatant was incubated with 100 µl Ni-NTA agarose (Qiagen) in a microcentrifuge tube, rotating for 20-40 mins at 4°C. The agarose-protein complex was harvested at 14,500 x *g* for 1 minute, resuspended in 1 ml binding buffer and rotated for 10 mins at 4°C, a total of three times. After the last wash and centrifugation step, the cells were centrifuged again to remove any residual wash buffer and supernatant discarded. 100 µl of elution buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM imidazole) was added to the pelleted resin and the tube inverted for 2 mins at RT. Following a final centrifugation step at 14,500 x *g* for 1 minute, 100 µl of the supernatant containing the eluted protein was stored at 4°C.

For separation of Histagged and non-Histagged protein following incubation with tobacco etch virus (TEV) protease (Sigma Aldrich and NEB), a gravitational flow column was used. Here, protein samples were incubated with 500 µl to 1 ml Ni-NTA resin (depending on the volume of sample and amount of protein predicted) for 1 hour at 4°C, rotating. The column was then stood upright to allow the resin to settle to the bottom and the flow through released. The resin was then washed three times using 20 column volumes (volume of resin) of binding buffer each time. The sample was eluted using 1 column volume of elution buffer.

2.8.7.2. NiNTA agarose purification using HPLC

Clarified and filtered cell lysate was passed over a 1 ml HisTrap HP column (GE Healthcare Life Sciences) to isolate His-tagged protein, using an AKTApurifier system (GE Healthcare Life Sciences). Applying a flow rate of 1 ml/min for the entire procedure, the column was washed with d.H_2O then pre-equilibrated with binding buffer before loading of the sample. The column was washed again with 7 column volumes of binding buffer, before elution in 1 ml fractions using a gradient of 25 mM to 300 mM imidazole over 30 column volumes. The column was washed with d.H_2O then 20% ethanol before storage. Protein from fractions corresponding to a peak on the UV_{280nm} trace of the nickel affinity purification were run on an SDS-PAGE gel (section 2.8.1) to detect fractions harbouring the protein of interest.

2.8.7.3. Gel filtration

Proteins requiring additional purification were passed over a gel filtration column (GE Healthcare Life Sciences) using the AKTApurifier, for size exclusion chromatography (SEC). The column was washed with 1 column volume of d.H_2O then SEC buffer (50 mM Tris-HCl pH 8.0 and 300 mM NaCl) before the protein sample was eluted in 1 ml fractions over 1.5 column volumes. The column was washed with 1 column volume of d.H_2O and 20% ethanol before storage. As with His-purification, protein from fractions corresponding to a peak on the UV_{280nm} trace were run on an SDS-PAGE gel (2.9.1) to detect the protein of interest.

2.8.8. Cleavage with the tobacco etch virus (TEV) protease

When required, proteins were expressed with a TEV cleavage site between the His-tag and protein coding sequence, enabling removal of the Histag following nickel affinity purification. Eluted fractions confirmed to harbour the protein of interest by SDS-PAGE were pooled and exchanged into TEV buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM DTT) using Vivaspin protein concentrator spin columns (GE Healthcare Life Sciences). These were incubated with 1 µg of TEV protease (NEB or Sigma Aldrich) for every 100 µg of target protein in a Slide-A-Lyser Dialysis Cassette (ThermoFisher Scientific), stirring overnight in 3 l of TEV buffer at 4^oC. Following incubation, samples were purified using nickel affinity chromatography section 2.8.7.2) to remove the His-tagged TEV protease and isolate the cleaved protein of interest in the flow through fraction.

2.8.9. Protein identification

To confirm the identity of purified proteins, samples were sent to King's College London- Centre for Excellence for Mass Spectrometry for liquid chromatography-mass spectrometry (LC-MS/MS) analysis. Samples were run on an SDS-PAGE gel (section 2.9.1), the band corresponding to the protein of interest excised and stored in MilliQ d.H_2O before processing. Results were viewed using Scaffold Proteome Software (http://www.proteomesoftware.com/products/scaffold/).

2.8.10. Protein quantification

The bicinchoninic acid assay (BCA) assay (ThermoFisher Scientific) was used for quantification of protein samples in a 96-well plate format, using known concentrations of BSA to generate a standard curve, according to manufacturer's instructions. Results were read at a 595 $_{nm}$ wavelength on a plate reader and data plotted in Microsoft Excel to generate a standard curve, from which protein concentration could be calculated by rearranging the formula $y=mx+c$, where y= absorbance of protein of interest, m=gradient, c=y-intercept, and x=concentration of the protein of interest.

2.9. Bioconjugation

The expression, extraction, purification and visualisation of proteins within the bioconjugation work was conducted as described in 2.8.

2.9.1. *In vitro* glycosylation

For *in vitro* glycosylation experiments, 1 L cultures of the strains carrying the acceptor protein, oligosaccharyltransferase (OST) and lipid-linked glycan donor were grown separately overnight at 37^{o}C and induced with 1 mM IPTG or 0.2% L-arabinose, as required (2.8.4). Cells were harvested at 7459 x *g* for 30 mins at 4^{o}C then washed three times in S30 buffer (10 mM Tris acetate, 14 mM magnesium acetate, 60 mM potassium acetate, pH 8.2) at a total 50 ml volume, centrifuging for 15 mins at 2880 x *g* between each wash. Following the final wash step, pellets were resuspended in 20 ml S30 buffer and lysed in the cell homogeniser (2.8.6.2). Lysates were clarified at 2880 x *g* for 15 mins then stored at -80^{o}C in 1 ml aliquots or used immediately in the reaction set up. When concentration of the glycan bound to the inner membrane was required, the supernatant isolated following clarification of the cell lysate was centrifuged at 100, 000 x *g* for 1 hour at 4^{o}C and the pellet resuspended in 1 ml S30 buffer containing 1 mM n-dodecyl-β-D-maltoside (DDM). Each reaction contained 20 µl of the PglB OST, 1 mM DDM, 1% $MnCl_2$ and a range of volumes of the acceptor protein and lipid-linked glycan, made up to a total 600 µl volume with S30 buffer. Reactions were incubated overnight at 100 rpm, 30^{o}C then purified as described in section 2.8.7.1.

2.10. Statistical Analysis

For all statistical tests, $p<0.05$ was considered significant linear regression was performed in Stata and data transformed to Log10 prior to analysis. Students *t*-tests were performed in Microsoft Excel and Wilcoxon Rank tests in SPSS.

3. Identification of novel immunogenic *C. difficile* proteins and expression and purification in *E. coli*

3.1. Introduction

Significant research has been dedicated to identifying candidates for inclusion within an anti-*C. difficile* vaccine. This has mostly centred on the use of Toxins A and B to provide protection from the symptoms of *C. difficile* infection (CDI). However, despite numerous toxin-based approaches entering clinical trials, there remains a lack of licensed vaccine against *C. difficile* [375]. The focus on the toxins as key targets for immunisation partly stems from their induction of the humoral immune response. Higher systemic antibody responses towards Toxin A and/or Toxin B have been associated with protection from infection, reduced duration of infection, prevention of mortality and reduced chance of recurrence [86, 269-271, 274, 444, 445]. However, these associations are not consistent, with some reports describing no difference in antibody response or higher antibody responses in those who developed infection or relapsed compared to healthy controls or asymptomatic carriers of *C. difficile* [272-275, 446].

Beyond the toxins, there has been some interest in alternative antigens for inclusion in a *C. difficile* vaccine, but none of these have reached later stage clinical trials. Higher sera IgG was found towards the cell surface proteins Cwp66 and Cwp84 and the flagella components FliC and FliD in healthy controls compared to patients with *C. difficile* associated-disease [273]. In addition, the major surface layer protein SlpA has been demonstrated to be immunogenic in mice [447] and higher anti-SLP IgG titres were reported for healthy controls compared to those with CDI [272]. However, the exposed, low molecular weight portion of SlpA is variable between strains, which could limit vaccine coverage [448]. Overall, understanding of *C. difficile* antigens other than the toxins remains limited. This in turn restricts development of vaccines which target both the symptoms of CDI and colonisation by the pathogen, an approach which has received increased interest in light of the challenges associated with

toxin-only vaccines [449]. Antigen discovery can arise from a number of avenues, including antigenic components from other bacterial species, analysis of patient antibody responses during natural infection and rational hypotheses, such as the use of conserved cell surface proteins, which are exposed to the immune system. Once identified, typical *in vitro* characterisation of these as immunogenic proteins usually utilises immunoblots or enzyme-linked immunosorbent assays (ELISAs). Although different protocols are available, the basic principle of ELISAs is antibody-mediated detection and quantification of a specific ligand, including antigens and other antibodies. These have been widely used to investigate antibody responses to *C. difficile*, particularly in relation to the toxins and are a well-established and valuable technique in quantifying humoral immune responses [86, 450]. However, both upscaling production of the antigens to be screened and testing these within a 96-well plate format are labour intensive and unsuitable for the screening of multiple antigens and/or human samples and can also suffer from inter-assay variability.

Just as DNA microarrays have been influential in the understanding of bacterial genomes and expression profiles, particularly before the advent of widely accessible and affordable sequencing technologies, protein microarrays offer a high-throughput means of quantifying antibody responses to a large number of antigens [380]. Here hundreds or thousands of antigens are expressed and printed onto glass slides, which can be probed with the sample of choice, usually human sera or plasma, simultaneously. Reactivity to the antigens is measured as with plate-type ELISAs, i.e. antigen-bound sera or plasma antibodies are probed with anti-human antibodies harbouring a reporter enabling their fluorometric or colourmetric detection. This large-scale characterisation of the humoral response supports understanding of infection progression as well as vaccine and diagnostic discovery and has been used for investigation of a number of pathogens, including but not limited to; *Leptospira interrogans*, *Francisella tularensis, Salmonella typhi*, *Coxiella burnetii* and *Plasmodium falciparum* [384, 386, 451-453].

In this chapter it was aimed to use the results of a human *C. difficile*-specific protein array from an ongoing Medical Research Council programme grant of my supervisors, to identify novel vaccine candidates and purify these candidates to validate their immunogenicity for potential use in an anti-*C. difficile* vaccine.

3.2. Design, probing and analysis of the *C. difficile* protein array

The *C. difficile* protein array was funded by a Medical Research Council Programme Grant entitled "Interactions between *C. difficile*, the intestinal microbiota and the host response in hospitalised patients". This was awarded to a consortium of institutions, lead by London School of Hygiene and Tropical Medicine (LSHTM), comprising the University of Liverpool (UoL), Wellcome Trust Sanger Institute (WTSI) and the University of California, Irvine (UCI). The project was established to investigate the factors relating to progression and outcome of CDI, through analysis of a large cohort of longitudinal patient samples with accompanying clinical metadata. Analysis covered a range of approaches including profiling of the humoral immune response using a pan-protein array and sequencing of stool samples for analysis of the gut microbiota. The results from analysis of the humoral immune response were used for the identification of novel vaccine candidates in this study, as described below. Section 3.2 describes the work conducted by the consortium, both prior to and during my PhD, with those who completed the work named throughout.

3.2.1. Sample collection

Patient recruitment and sample collection was conducted between 2013 and 2015. Patient recruitment and sample collection, storage and processing were conducted by UoL and the Royal Liverpool University Hospital. This was led by Professor Sir Munir Pirmohamed and involved a team of researchers, medics and NHS staff, including but not limited to: Dr Paul Roberts, Dr Neil French, Dr Fabio Miyajima and Ms Alejandra Doce-Carracedo. Patients were recruited from hospitals across Merseyside and Cheshire, including; Royal Liverpool and Broadgreen Hospitals, Aintree University

Hospitals, Wirral (Arrowe Park) Hospital, St. Helens & Knowsley (Whiston) Hospital and Countess of Chester Hospital.

For recruitment into the study, participants had to be over 18 years of age and capable of providing written consent to confirm their willingness to participate. Patients/participants were recruited if they met the criteria for one of the three study groups; *C. difficile* infection (CDI), diarrhoea of unknown cause (CDU) and healthy controls (CDH) (Table 3.1). The patients were recruited in line with the ethics sought by UoL and awarded by Liverpool Research Ethics Committee (REC Reference 08/H1005/32) with additional approval granted by LSHTM for working with the samples at LSHTM (Appendix C). Diagnosis of *C. difficile* was conducted on patient stool samples by the NHS microbiology laboratory using NHS-approved methods, including an ELISA-based detection of the *C. difficile* Toxins A and B. During the study an additional test was adopted by participating hospitals for CDI detection- the glutamate dehydrogenase assay (GDH). Additional diagnostics available for some but not all samples included Toxin B PCR and microbiological culture. CDI patients were differentiated into mild and severe infection based on faecal white cell count.

Study Group	Definition	Description
CDI	Patients with *C. difficile* infection	Positive for *C. difficile* by Toxin A/B ELISA, as diagnosed by the NHS lab. Later into the study, GDH testing was also adopted at the recruiting hospitals and microbiological culture was performed.
CDH	Healthy controls	Healthy people who were not hospitalised, had no history of CDI within the last 12 months and Toxin A/B and culture negative. Where possible these were age and sex-matched to the CDI group.
CDU	Patients with diarrhoea of unknown origin	Patients with diarrhoea but negative for *C. difficile* by Toxin A/B ELISA. These were either GDH positive or negative.

Table 3.1. Description of participant study groups for the *C. difficile* protein array. ELISA, enzyme-linked immunosorbent assay, GDH, glutamate dehydrogenase assay, CDI, *C. difficile* infection.

Upon recruitment, stool, urine and blood baseline samples were taken. The blood was fractionated into plasma and sera. Where possible, a maximum of three follow-up samples were taken longitudinally at 2, 4 and 6 weeks post-baseline. Time between initial NHS lab diagnosis and baseline sample collection ranged from approximately 1 to 7 days, in which time the patients may have already begun antibiotic treatment for *C. difficile* with metronidazole, vancomycin and/or fidaxomycin, as per NHS guidelines (Figure 3.1) [454]. Metadata was obtained throughout the study for recruited participants, including; previous history of CDI and/or previous antibiotic usage, severity of CDI and the PCR ribotype of infection, relapse or continuous colonisation, date, duration and type of antibiotics, proton pump inhibitors and nutritional supplements. The NHS diagnostic results for each patient were also extracted retrospectively, including some retrospective culturing form the lab stool sample.

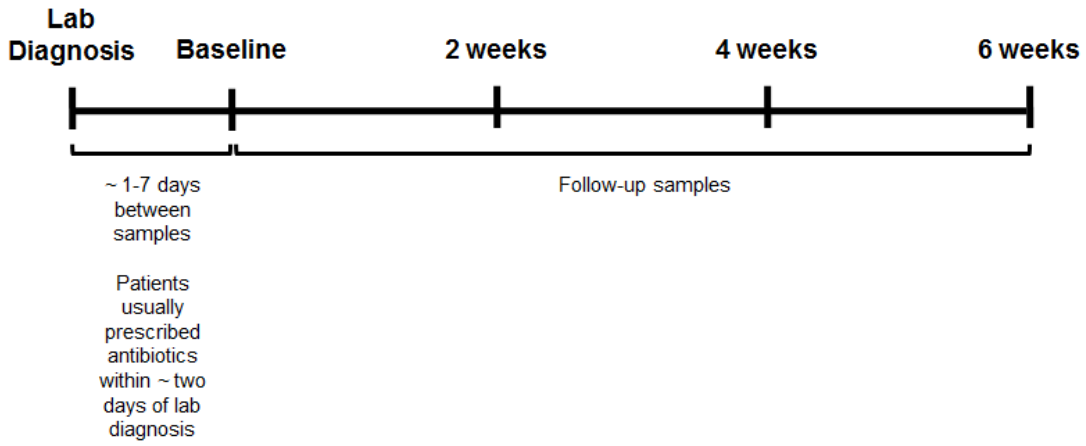


Figure 3.1. Timeline for diagnosis and collection of participant samples. Baseline represents samples taken on recruitment into the study, with up to three additional samples taken in follow-up. Antibiotic prescription describes those administered for treatment of CDI.

3.2.2. Probing of the array

The proteome of the hypervirulent *C. difficile* strain R20291 was used as the base for selection of proteins to be screened on the pan-protein array, with additional proteins included from three other lineages of *C. difficile* namely strains 630 (ribotype 012), CF5 (ribotype 017) and M120 (ribotype 078). Putative protein function as annotated in the *C. difficile* genome and pBLAST analysis enabled selection

of proteins for screening. This consisted of a number of different potential antigens, including surface-associated proteins, virulence factors and strain specific unique proteins, including transposases. Analysis and selection of the proteins was performed by Dr Richard Stabler and Dr Lisa Dawson at LSHTM.

The Felgner group at UCI have established a pan-protein array platform for the screening of bacterial, viral and parasitic proteomes against human samples, in the pursuit of vaccine and diagnostic discovery [380, 385, 455]. Production and probing of the *C. difficile* pan-protein array was undertaken by the consortium partners in the Felgner group at UCI Irvine, as described previously [380, 456]. Briefly, the coding sequence of the target proteins and the pXT7 plasmid for protein expression were amplified by PCR with complimentary overhangs to one another. Amplicons were transformed together into competent *E. coli* DH5α cells and incubated overnight in LB supplemented with 50 µg/ml of kanamycin for plasmid selection. From this, the ligated plasmids were isolated and a proportion of these confirmed by sequencing to encode the correct insert, with a total cloning efficiency rate of ~93%.

Protein expression was achieved using an *in vitro* transcription/translation (IVTT) cell free system, using the RTS 100, *E. Coli* HY Kit (Roche) [457]. Synthesised plasmids encoding the construct for expression with ribosome binding site and T7 promoter and terminator, were incubated with all the components required for transcription and translation, including T7 polymerase for production of mRNA and an *E. coli* lysate providing the ribosomal machinery required for translation [380]. Instead of purifying the protein, the IVTT reactions were centrifuged with 0.2% Tween20, then the supernatants printed directly onto nitrocellulose-coated glass FAST slides (Whatman) using an Omni Grid 100 microarray printer (Genomic Solutions). To confirm protein expression, the coding sequences for each protein also harboured a 5' 10XHistag (HIS) and 3' haemagluttinin tag (HA), enabling detection of full protein. In total, 3165 proteins were cloned, then successfully printed and screened on the array, by UCI. Arrays were printed throughout the study, as required.

3.2.3. Optimisation of the pan-protein array and pilot study

Plasma samples were selected as preliminary work found no difference in antibody response between plasma and sera samples. It was also attempted to use the stool samples to probe the array, to assess the level of secretory IgA present within the gut. However, unacceptable levels of background were produced in these assays which made it impossible to determine *C. difficile*-specific signal, therefore this was reserved for later use following optimisation.

A pilot study was performed in 2015 using 189 plasma samples from the CDI (split into severe CDP and mild CDN) and CDU group, and a third control group collected by UCI, called GCRC. This is a standard healthy control group utilised in their arrays, constituting students and staff recruited from the UCI campus. This protein array was probed separately with IgG and IgA, using anti-human antibodies conjugated to the Cy5 fluorophore. As the IgA response was low across the sample set, the IgG response was analysed. This analysis was conducted by UCI following probing of the array, who then provided a short list of the top 70 reactive antigens which were significantly different in signal in at least one of the between group comparisons (CDI vs CDU, CDP vs CDU, CDI vs GCRC) (Appendix D). The healthy control group from California were not age and sex matched to the CDI patients and little was known about these samples. As these were likely to be younger people these samples were deemed unsuitable as *C. difficile* primarily affects those over 65 years of age and therefore this is the target group for investigation. The CDU group was therefore used as a comparator. This group had diarrhoea but were negative for the *C. difficile* toxin and therefore assumed to be *C. difficile* negative. At this stage, the results were used to look for immunogenic proteins in the CDI group, to confirm these antigens are recognised by the human immune response. The proteins within the shortlist fell into many different functional categories but the majority came from one of four groups; enzymes, transport, cell surface proteins and hypothetical/uncharacterised proteins. From the shortlist of immunogenic proteins, three were selected for further characterisation *in vitro*. Proteins were preferentially selected that were predicted to be surface exposed and present in multiple clinically relevant strains as identified by pBLAST. Furthermore, proteins with a putative function that suggests

an important role in survival that would be valuable to characterise were also prioritised. This resulted in the selection of three proteins for characterisation *in vitro*, CDR20291_0330, a putative cobalt binding protein, CDR20291_3343 a putative pili protein and CDR20291_0342 a putative ABC transporter permease, as described in Chapter 4, 5 and 6, respectively.

3.2.4. Screening of CDI patients against matched healthy controls

The protein arrays were probed in 2016 with 279 baseline plasma samples, from the CDI (128), CDH (71) and CDU (80) groups. The CDH group are matched controls from the Liverpool cohort of samples. The printed slides were incubated for 30 mins with protein array-blocking buffer (GVS Life Sciences, Sanford, USA) then incubated at 4^{o}C overnight with 1:100 plasma, diluted in blocking buffer. Prior to probing, plasma samples were pre-absorbed with *E. coli* lysate to quench any anti- *E. coli* antibodies, which can mask organism-specific antibody responses [380]. Slides were washed, then incubated with anti-human IgG or IgA simultaneously, before washing again and visualising the results using an ArrayCAM imaging system, which enables two colour scanning of the arrays. IgG and IgA signals were differentiated by the Quantum dot conjugated to the anti-human antibodies, which have different fluorescent spectral profiles [456, 458].

3.2.5. Analysis of array results

Processing and analysis of the raw data obtained following probing of the array was conducted by Dr Li Liang at UCI and Dr Lisa Dawson and Professor Taane Clarke at LSHTM, between 2016 and 2017. Firstly, background signal was removed by subtracting the mean of the no DNA control spots from the raw data values for each protein. The no DNA controls represent plasmid-free *E. coli* lysate from the IVTT system. From this, the mean IgG reactivity for each protein was calculated for the different patient groups and any statistically significant differences between the CDI and CDH groups was determined using a Student's *t*-test and Wilcoxon Rank test, where $p<0.05$ was considered significant. The IgA signal was consistently much lower than the IgG response and it was difficult to extract

meaningful data, therefore the IgG response was used for identification of potential vaccine candidates.

3.3. Identification of novel immunogenic *C. difficile* proteins

In order to identify immunogenic antigens that could serve as potential vaccine candidates, those proteins where the IgG reactivity signal was higher in CDH over the CDI group were investigated. It was hypothesised that a higher response in the CDH group suggests antibody development against the target protein has a protective effect against infection.

3.3.1. Identification of immunogenic proteins for recombinant expression in *E. coli*

A table of the normalised intensity values for each antigen on the array was compiled, including the total number of samples reacting to each antigen (out of a total 279), where a minimum signal intensity of 0.5 was produced in both groups (Table 3.2). To identify potential vaccine candidates, these were sorted by the largest difference in signal intensity between the CDH and CDI group, where the CDH signal was significantly higher than in the CDI group, as determined by *t*-test or Wilcoxon rank test. From this antigen shortlist, 6 targets were selected for further investigation (in bold, Table 3.2), taking into consideration conservation across different *C. difficile* strains from the 5 clades; (630 (ribotype 012), R20291 (ribotype 027), M120 (ribotype 078), M68 (ribotype 017) and CD305 (ribotype 023)), as determined by pBLAST and sequence analysis. Additional factors include the mean signal in the two groups and putative annotation of the protein, selectively choosing those most likely to be surface exposed and a range of different putative functions to screen a variety of different candidates. Signal to purified Toxin A was significantly higher in the CDH compared to the CDI cases, but was excluded due to volume of existing research [86, 274, 445]. Two proteins were not conserved through all strains but were included for analysis, especially considering any eventual *C. difficile* vaccine would unlikely to be based on a single antigen, meaning the lack of the protein in all strains wouldn't necessarily limit vaccine coverage. *CDR20291_2697* was present in M68 but contained multiple stop codons and therefore was unlikely to produce a functional protein. However, the IgG response to this

protein was one of the highest for the CDH group and a high proportion of samples recognised this protein (228/279) (Table 3.2). *CDR20291_0342* was found to not be present in the ribotype 023 CD305 strain but as before, this wouldn't be used in a vaccine alone and was one of the immunogenic proteins chosen for *in vitro* characterisation based on the pilot study. Of the three proteins selected for further investigation from the pilot study, both CDR20291_0330 and CDR20291_0342 demonstrated a significantly higher IgG response in the CDH group in this analysis, confirming their potential as vaccine candidates. The response to CDR20291_3343 was higher in the CDH group although this was not significant.

Three additional proteins were also included for use as controls. These covered three different reactivity profiles- higher in the CDI group compared to the CDH group (CDR20291_3064) (negative control), high in both groups (CDR20291_2503) (positive control) and low in both groups (CDR20291_2536) (negative control). Therefore in total, 10 candidates were taken forward for expression, purification and testing in the ELISAs.

Name	Putative Annotation	No. of samples with signal (279)	IgG signal		CDH minus CDI	Wilcoxon rank test	T-test	Outcome
			Mean CDH	Mean CDI				
CDCF5_2692	Cell surface protein	135	4.51	1.24	3.27	0.01	0.00	Not conserved
CDM120_orf03606	Putative class B sortase	210	6.02	2.76	3.26	0.01	0.04	Not conserved
CDCF5_0223	Flagellin subunit	213	6.67	3.51	3.16	0.00	0.02	Not conserved
CDR20291_2697	**Hypothetical membrane protein**	**228**	**6.79**	**4.02**	**2.77**	**0.01**	**0.20**	**CANDIDATE**
CDM120_0411	Signaling protein	278	24.95	22.69	2.26	0.00	0.42	Not conserved
CD0416	Exported protein	207	3.29	1.15	2.14	0.00	0.01	Not conserved
CD0386	Collagen-binding surface protein	243	6.85	4.76	2.09	0.00	0.11	Not conserved
CDR20291_2235	Membrane protein (HP)	213	10.07	8.16	1.91	0.01	0.63	High response in CDI
CD1105	DNA primase	270	7.82	6.30	1.52	0.00	0.13	Not conserved
CDR20291_2226	Hypothetical membrane protein	179	4.49	3.01	1.49	0.00	0.41	Unknown function
CDR20291_2640	**Accessory gene regulator**	**220**	**3.07**	**1.64**	**1.43**	**0.00**	**0.12**	**CANDIDATE**
CDM120_2021	Type I-B CRISPR-associated endonuclease Cas1	190	2.40	1.06	1.33	0.03	0.41	Not conserved
CDR20291_2491	CdtA (adp-ribosyltransferase	192	5.25	4.01	1.24	0.00	0.41	Not conserved
CDR20291_0330	**Cobalt transport protein**	**156**	**2.07**	**0.86**	**1.21**	**0.15**	**0.03**	**CANDIDATE**

Name	Putative Annotation	No. of samples with signal (279)	IgG signal		CDH minus CDIN	Wilcoxon rank test	T-test	Outcome
			Mean CDH	Mean CDIN				
CD0430	Membrane protein	238	2.46	1.32	1.14	0.00	0.00	Not conserved
CDR20291_2389	Competence protein	175	2.05	0.92	1.13	0.02	0.15	Unlikely to be surface exposed
CDR20291_0342	**ABC transporter, permease protein**	**212**	**2.55**	**1.44**	**1.11**	**0.01**	**0.04**	**CANDIDATE**
CDR20291_3155	**Type IV pilin**	**179**	**2.12**	**1.02**	**1.10**	**0.02**	**0.13**	**CANDIDATE**
CD0453B	Hypothetical protein	247	2.49	1.46	1.03	0.00	0.00	Not conserved
CDR20291_0781	Hypothetical membrane protein	188	2.03	1.01	1.03	0.00	0.03	Unknown function
CDR20291_0966	Transglycosylase	169	1.94	0.92	1.01	0.01	0.01	Unlikely to be surface exposed
CDM120_0286	Hypothetical protein	198	2.69	1.68	1.01	0.00	0.11	Not conserved
CDR20291_1324	Uncharacterized protein	272	4.54	3.54	0.99	0.00	0.06	Not conserved
CDR20291_0002	Mechanosensitive ion channel protein	176	1.96	1.00	0.96	0.00	0.24	Low signal in both groups
CDR20291_2253	**Exported protein**	**167**	**2.59**	**1.68**	**0.91**	**0.03**	**0.33**	**CANDIDATE**
CDR20291_3343	**Pilin protein**	**101**	**0.28**	**0.19**	**0.09**	**0.1**	**0.3**	**CANDIDATE**

Table 3.2. Top vaccine candidates from the protein array, based on difference in IgG reactivity. The IgG response of the 279 samples to the screened proteins was compared to identify the top 25 antigens with the largest difference in signal between the CDH and CDI sample groups, where the signal was significantly higher ($p<0.05$) in the CDH group, as determined by *t*-test or Wilcoxon rank test. From this list, six proteins were selected for purification and further screening with patient samples (bold). This was based on those with interesting putative functions that were likely to be surface exposed and conserved across multiple *C. difficile* strains. Proteins were considered conserved when they were identified in 4/5 strains investigated using pBLAST; 630, R20291, CD305, M68 and M120. The seventh bolded antigen is CDR20291_3343 which was selected for characterisation based on the result of a pilot study. CDH, healthy controls, CDI, patients with *C. difficile* infection.

3.4. Selection of regions for recombinant expression in *E. coli*

Cell surface proteins often encode transmembrane domains (TMDs) for their anchoring into the cell membrane. As these are hydrophobic, when expressed recombinantly it can make isolation of soluble protein challenging. It was therefore decided to use a combination of bioinformatics approaches, namely Phyre[2] and TMHMM Server v2.0, to identify predicted TMDs within each array candidate. Additionally, Phyre[2] and SignalP were used for identification of signal sequences, which would not be present in the mature protein *in vivo*. Few signal peptides were identified during protein analysis, however many harboured putative *N*-terminal TMDs. It is possible that these TMDs are in fact signal peptides as both harbour a number of hydrophobic residues.

In some instances, it was not possible to exclude all TMDs. For example, CDR20291_0330 is a small protein with two putative TMDs. Selecting the region between these domains would result in expression of a very small peptide which was not practical, therefore the region between these domains was expressed with the second TMD. CDR20291_2640 is TMD rich and there were no predicted exposed regions of suitable length for expression therefore the whole protein was included. Finally, as CDR20291_0342 is a large protein, two regions were selected for expression which were predicted to be extracellular, S1 and S3. Modified protein sequences were codon optimised for expression in *E. coli* and ordered for synthesis as G-block gene fragments (IDT) for downstream cloning (Figure 3.2).

CDR20291_3155

LINLINKKRKGFTLVEMIVVVTILGVISSIALVKYSKVQESAKLNADYTNAANIVTAASM
AINDDEKTIDSLSVETLKEKGYLNTVPVPQSTSGKFELVINDSGTDISVNINSKQFYPK

CDR20291_0342

MKDNLKIALRYIISYKARSLAIALSIILATSLIVGIGTLSRSAQQAEVDKLKRETGSDHV
YFKDINKNQLEYIKSRKDIKNLGITSHYGYTDPNERLAINLEYANKNYLTNQSKLIKGHL
PKASNEVVVEKWILNSLGLKPEINQNITFKLYQKEKPETFKVVGILEDRPIEKNKGTCEI
FLNLNESKLDKFSYAYIEFNNGIDINTSIDNIVKNTMLDENSVGKNKMLIESTRENGTLD
NSSKYTAITMSLFSGIVIYSIYVISIYQRIQEYGILRAIGSTNFKIFKFMFYELFILALI
AIPIGICTGIGGAQIFNRTAGNIQFEGNVTVTPFVIPDKIILLSIGSIILTILIISFFTY
LKIRRISPIDSIRKTFGTDKNINKVNSLISKLTLNISVTKSISAKNIFRNKKGFIIIILS
MSLGGIMIIKENYKYSFSDIQNKNGQEETYMNADFILTNFFLKLNQSNKADSFKDIKGLN
DNQIDKIKNINGIDKVKTASILDTRIEVDKLNGLDYYNIINSTPYYKDFPLFIKNKNTGK
YTMKQKLRGYNDEMINSLEKYLVSGSINLERMKKENLAILYVPQVSKTNKYKLSYTPGTE
TPAVDIKVGDTIKVKYPKGEIDQDLYIKLKDNYEYLEYEFKVGAIVSYPFADNYLYSGDQ
GIDVITSNDYLKKLTGVDKYNVVYVDINKNANVKKINTLLGEIGSESPGTTTANMLMEKE
NFDKMTARALTYAYGIVAVMFIISVFNIINNISYNLTSRTSEFGMLRAIGISERGFKNMI
LYEGILYGILSSVITIVVGLIIQFRMYYTYNFVSYGLGFSIDYKIYILVVLANIIVGILA
TYIPLRKINKISIVEAINITE

CDR20291_0330

MSAKTKTKTNILLLLLVVALIIFPLLVNSGAEYGGADGQAESEITKINPDYEPWFSSPYE
PPSGEIESLLFSAQAALGAGVIGYILGVQKGKRS

CDR20291_2697

MFEKKKKELFPWKISALILVGVFVLSSGVYIGTRLKNNGVQTSKVVSASANMDEKSKEAF
GLNKNCEVWLQSEGENETNYKRTLTMAGIVPESLLDKTKSEIVSYFKKEYPTKKIKSMDK
NEIILVESKYSDNKDVQAMNATIKGKYTIENDNGNIALYKYDDSGSKTLVEKTKIRVDSL
PKTVQDEIKKGVVMDTEEEAYSRLEDFAS

CDR20291_2503

MLTPIEIENKEFKKGLRGYRDEEVDEFLDIIKEDYESLCRENTALKEKLGLYQDQVNKYE
NIEETLKATLITAQSAAEDTCSAANKKAKIIVEEADLKARQIIEQANNRVIEIRKEYDAL
VKEFKIFRNKFKSLLEDEIRSVDEIFYDVDEKCVNAFEGTTVYNYSNDEAATTLE

CDR20291_2640

MFKRYAEKMTSVLICNNMIDNNESKVYSYGFEILIAFIVNITTMLFIGFLFGKFTYVLFF
LMCYCPIRQFSGGYHADNYFRCLLTFIFIILSTILIIENINIDLFKNIIMIIASVSWVGI
CVLCPIEHRSNPISDREKFVYKKTAIFISTVVLLITILSLSLSIFVDYFTYSAFAMFWIF
VMLVLGKLKAKV

CDR20291_3343

LRKWNKFKSERGAALVLVLIVVALLSIVGLIFSNQIANRIKSTKTTNEGIQAKYLAETCV
ENSIDKAYEKLYDELEKMDNEFKSENQEKSISRSKLRNISDEDFNNQDEKNIEAERLGYM
NNINFYLNKASSDLEKASMELKKLYDLDMLDYRDIEYVDANIISHRDSILEICKNYTSGD
ISKINEYILKEDIDSTTLIEAKLVNNDILLKMFLEENKIENEHLNSAFSHTYKALDNISL
AMQNMIEYRHTFHIDEPKVEVSNGIPDSQQYYELIQNPIINSMEYIWNSKWDTLENLLEI
LPNQTQGFNSLRVHLRNNVRKFEKLSDNISSGKKNTAKNFLKYKELLYEISDQCNQLKSM
SYEKIPVKYDNMALITTFDYIQNELLAEIKCRLKELKPQEIDKTEGITIKIPFYKADYDM
TKEGWPKLKENGSGAELSLMVTGDKDGIKEVEVTDGKKNIIGLGVEENSNSKYKVDAIVN
FNLNIDTNVVGNYDIKDKILINHDISSYKKVN

CDR20291_2253

MLKKIVIISCIVVVLIILSKIVDDNIKEDASIPNVNKETLEYFRKNYKEDIITCAEEDLN
NDGKKDLVVIYKKSNNSNEMVVVVSDKNSHYITKPIPAPIENQTITFKNIDDKAPIEVIV
SGSKNGNVGYAIYRVEGKKFVDLFGEDMDKCC

CDR20291_2536

MKKKIVILGAFILGIFISTYIKSLDPNRVYITLGQTESIESEVENTKKEIKNLKETLKNN
KKTLNEYKEAKKNGNESIQTLMQKELEEVKELSGYSTVTGPGITITMRDSERELKDGQNP
NDLIIHDIDILRVLNDLKKAGARAISINGERVLATSKIKCSGATITVNDTTYGQPFIIKA
IGDIDLLSAAINSPESYAKSLKDIYGINIKVQENNGNKVNLFAK

CDR20291_3064

MDFISSILQGLLSILFNFTGDFGISIVLITLIVKLVLMPLSLKQRFSVKKQQELAEKMEH
IKEKYKNDTKELEKQIQVSSVESMKSMLGCLIVLVQMPVVYALYHVCLDMTKEFTTIIIP
WIANLGTFDNLFILPCIYTLIMLASNLVNYIPYLKINSQVKFNKQMAVVTTITSLILTVR

Figure 3.2. Analysis of the *C. difficile* protein array candidates for recombinant expression in *E. coli*. Each array candidate was assessed for transmembrane domains (TMDs) (underlined) and signal peptides (bold) using a combination of Phyre[2], TMHMM server v2.0 and SignalP. Both signal peptides and TMDs were removed where possible as these are hydrophobic and can hinder purification of soluble protein. Regions selected for expression are shaded in grey. Two regions of CDR20291_0342 were proposed to be surface exposed and were selected for expression; S1 and S3.

3.5. Expression of recombinant *C. difficile* proteins in *E. coli*

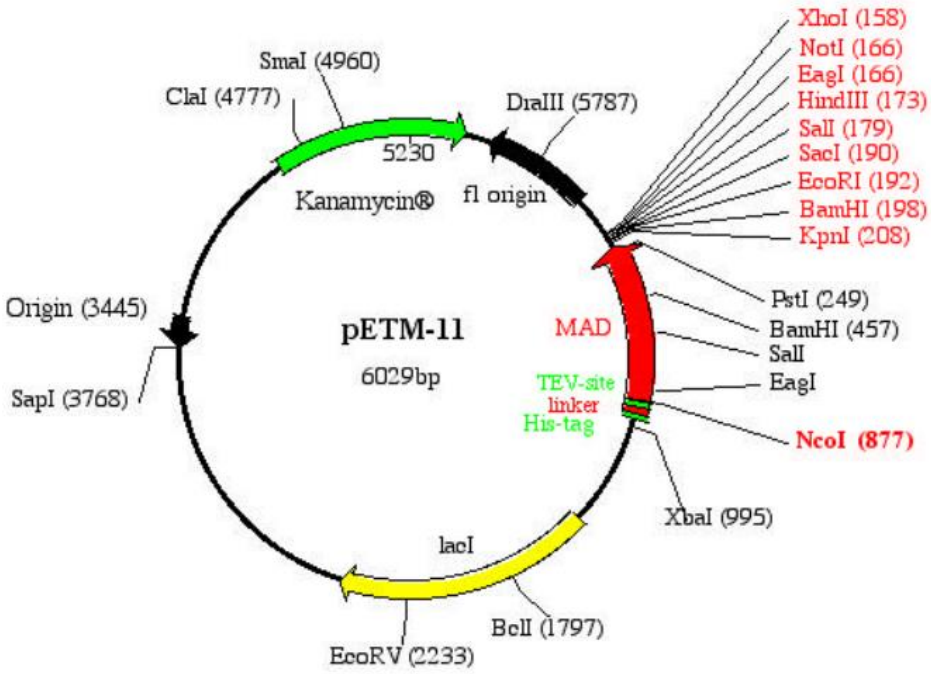
The array candidates selected for screening were recombinantly expressed in *E. coli* (Table 3.3). Large scale protein expression and purification is routine procedure and optimised for *E. coli*. Using this host also enabled expression from the pETM11 plasmid (Figure 3.3) [459], where protein expressionis IPTG inducible. The candidate protein is cloned downstream of an IPTG inducible promoter, and transcriptionally fused to an N-terminal 6XHistag for nickel affinity purification of the desired protein, which is separated from the protein coding sequence by a tobacco etch virus (TEV) protease cleavage site. This enables cleavage of the 6XHistag following translation and purification, which provides an additional round of purification and removes the chance of background reactivity from the 6XHistag.

Name	Putative function	IgG response on protein array	Size with 6XHistag (KDa)	Size without 6XHistag (KDa)
R20291_3155	Pilin protein from the minor pili locus, PilA2	Higher in CDH	12.37	9.23
R20291_0342 S1	Permease component of an ABC transporter (S1 region)	Higher in CDH	26.39	23.25
R20291_0342 S3	Permease component of an ABC transporter (S3 region)	Higher in CDH	36.37	33.23
R20291_0330	Cobalt binding protein	Higher in CDH	10.83	7.69
R20291_3343	Minor pilin, PilK, encoded within the major pili locus	Higher in CDH	58.74	55.6
R20291_2253	Exported protein	Higher in CDH	20.14	17
R20291_2697	Membrane protein	Higher in CDH	22.03	18.89
R20291_2640	Accessory regulator	Higher in CDH	25.3	22.16
R20291_2503	Cell-division initiation protein	High in CDH and CDI	20.3	17.16
R20291_2536	Exported protein	Low in CDH and CDI	27.73	24.59
R20291_3064	Membrane protein insertase	Higher in CDI	6.11	2.97

Table 3.3. Protein array candidates selected for recombinant expression in *E. coli*. Putative function of the *C. difficile* proteins selected for recombinant expression in *E. coli*, along with the predicted size before and after cleavage of the 6XHistag. IgG response is also included to describe the different candidates- higher in CDH (vaccine candidate), High in CDH and CDI (positive control), higher in CDI and low in both CDI and CDH (negative controls).

3.5.1. Cloning and test expression of recombinant proteins

Codon optimised G-block fragments encoding the recombinant *C. difficile* protein sequences were inserted immediately downstream of the TEV cleavage site in pETM11 (Figure 3.3) using either restriction digest (section 2.4.4), for 0342 S1, 0342 S3, 2253, 3155 and 0330 or Gibson assembly (section 2.4.6), for 2697, 3343 and 2640. It was not possible to successfully clone 2536, 2503 and 3064 into pETM11, therefore these were not taken further. All plasmids were confirmed to be correct by DNA sequencing.




```
   T7 promoter  -->          lac operator    XbaI
GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAAT
CTTTAATTATGCTGAGTGATATCCCCTTAACACTCGCCTATTGTTAAGGGGAGATCTTTA

          rbs                          His.Tag
AATTTTGATTTAACTTTAAGAAGGAGATATACCATGAAACATCACCATCACCATCACCCC
TTAAAACTAAATTGAAATTCTTCCTCTATATGGTACTTTGTAGTGGTAGTGGTAGTGGGG
                                  METLysHisHisHisHisHisHisPro

                          TEV site      NcoI
ATGAGCGATTACGACATCCCCACTACTGAGAATCTTTATTTTCAG GGCGCCATGGCGGCG
TACTCGCTAATGCTGTAGGGGTGATGACTCTTAGAAATAAAAGTC CCGCGGTACCGCCGC
MetSerAspTyrAspIleProThrThrGluAsnLeuTyrPheGln|GlyAlaMETAlaAla
```

Figure 3.3. Plasmid map of pETM11, used for recombinant expression of *C. difficile protein* array candidates in *E. coli*. Plasmid map and sequence of the pETM11 plasmid for protein expression. This harbours an N-terminal 6XHistag followed by a TEV protease cleavage site, enabling removal of the 6XHistag following protein purification. Image from Dümmler *et al.* [459] and the EMBL Protein Expression and Purification Facility (https://www.embl.de/pepcore/pepcore_services/strains_vectors/vectors/pdf/pETM-11.pdf).

The pETM11 plasmids harbouring the different recombinant proteins were transformed into the *E. coli* protein expression strain, BL21. As an initial test to confirm expression of the proteins, 10 ml cultures of the BL21 strain carrying the different pETM11 plasmids were induced overnight with 1 mM IPTG and cells harvested the next day. Extracted protein was purified using nickel affinity resin then probed with anti-His antibody in an immunoblot (Figure 3.4).

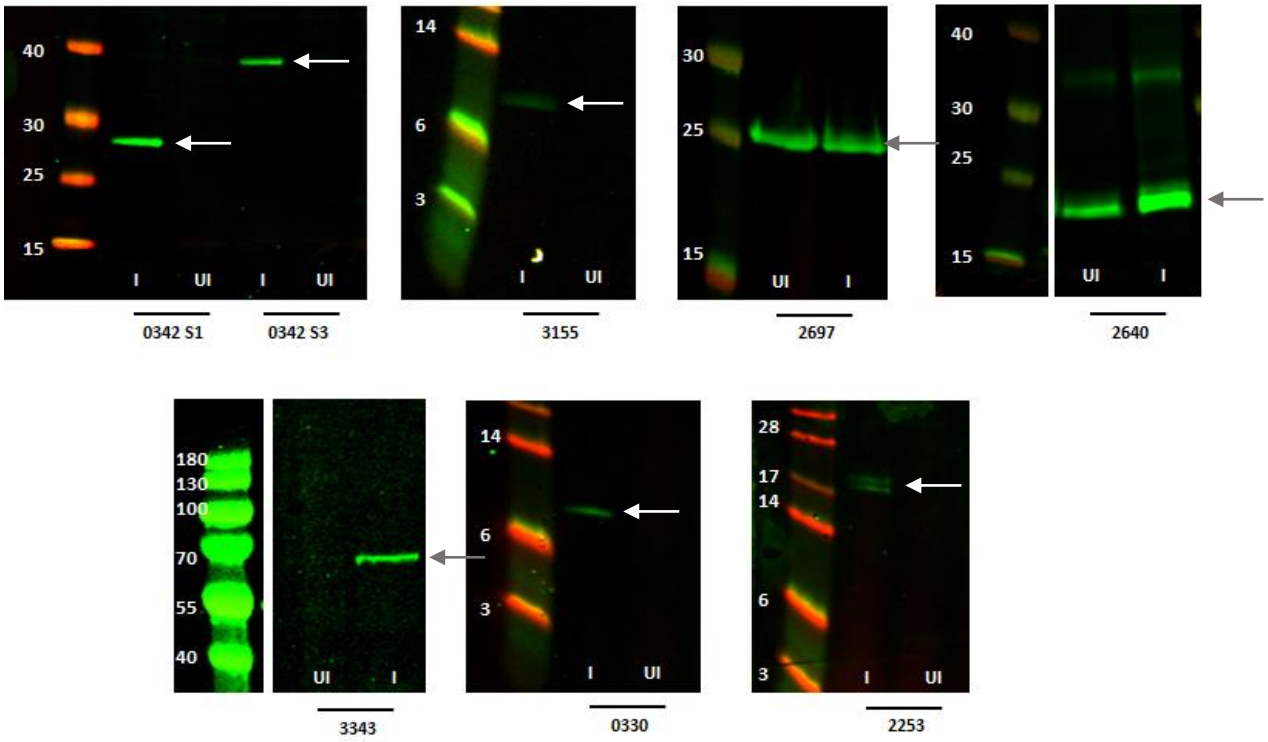


Figure 3.4. Recombinant expression of *C. difficile* protein array candidates in *E. coli*. Each pETM11 plasmid harbouring a different construct for expression was induced overnight in 10 ml LB with 1 mM IPTG. The cells were disrupted and protein extracted and purified by nickel affinity chromatography. Purified protein samples were run on an SDS-PAGE gel then detected on an anti-his immunoblot (green). UI, uninduced, I, induced. Arrows indicate the band corresponding to the desired protein.

When induced with 1 mM IPTG, a His-reactive band was detected at approximately the correct size for 0342 S1, 0342 S3, 3155, 3343, 2640, 2697 and 0330. Although for some proteins the level of expression appeared to be low (CDR20291_0330 and CDR20291_3155), it was deemed sufficient to be taken forward into upscale and further purification. A double band was produced for 2253 and therefore was excluded from further study. Higher levels of expression were achieved for CDR20291_2697 and CDR20291_2640, which was presumably the reason for detection of protein in the uninduced control, due to the promoter not being completely off.

3.5.2. Purification of the *C. difficile* recombinant proteins from *E. coli*

In order to obtain sufficient protein for screening with the serum samples, eight litres of *E. coli* culture was prepared for expression of each *C. difficile* protein; 0342 S1, 0342 S3, 0330, 3155, 3343, 2697 and 2640. Cultures were induced overnight at 37°C with 1 mM IPTG then cells harvested and lysed using a

cell homogeniser. The supernatant of the lysate was obtained by centrifugation then passed over a 1 ml HisTrap HP column on an AKTA purifier system for isolation of the target protein via the N-terminal 6XHistag. The purified protein was eluted from the column in 1 ml fractions and an aliquot of each corresponding to the AKTA peak were run on an SDS-PAGE gel to identify which held the protein of interest. Even though 0342 S3, 3343 and 2640 were detected in the test cultures, and despite repeated attempts with the large scale cultures at different temperatures and in difference flask sizes, very little or no protein was detected following purification of these in the larger scale set up and would not have been sufficient for downstream processing, therefore these were also not taken forward. This left four purified proteins to take forward; 0342 S1, 0330, 3155 and 2697.

The fractions containing the eluted protein were pooled for each candidate then incubated overnight with TEV protease to remove the 6XHistag. Following incubation, the samples were run through an additional nickel affinity purification step but in this instance, any protein still harbouring the 6XHistag and the 6XHis-tagged TEV protease bind to the column, allowing the isolation of cleaved protein only in the flow-through. This also served as an additional purification step, as any histidine rich proteins pulled out in the first nickel affinity purification, also bound to the column (Figure 3.5). We thank the Issaacson group at Kings College London for their kind gift of purified TEV protease.

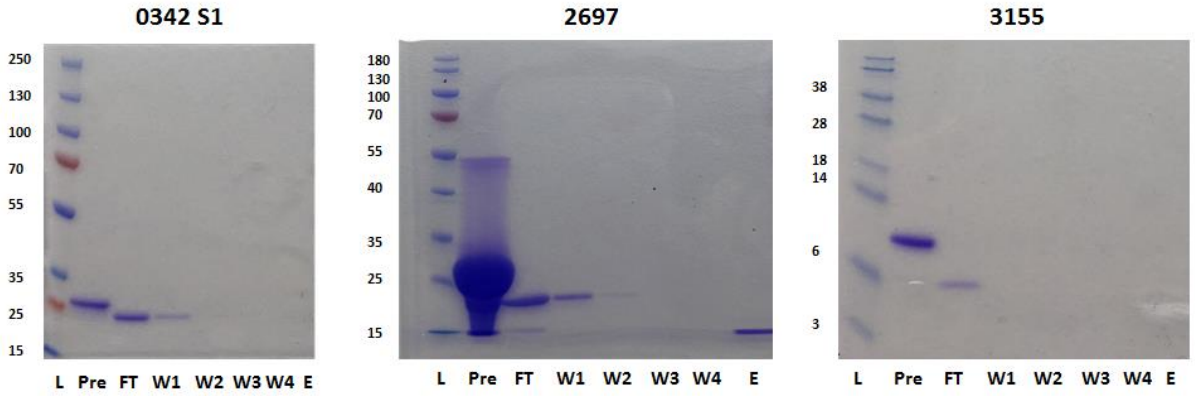


Figure 3.5. Cleavage of the 6XHistag from the nickel affinity purified proteins using TEV protease. Proteins purified using nickel affinity chromatography were incubated overnight with tobacco etch virus (TEV) protease for removal of the N-terminal 6XHistag. The protease cleaves at a recognition site upstream of the start of the recombinant protein, which can then be isolated in the flow-through fraction of a nickel affinity purification, where the 6XHistagged TEV protease remains bound to the column. Each protein isolated in the flow-through of the nickel affinity purification was run on an SDS-PAGE gel to ensure the correct protein had been isolated and visualise the size shift resulting from removal of the 6XHistag. The predicted size of the protein pre and post incubation with TEV, respectively, is 26.4 KDa to 23.3 KDa for 0342 S1, 22 to 18.9 KDa for 2697 and 12.4 to 9.2 KDa for 3155. Lanes represent the different stages of purification; Pre- pre-TEV cleavage and the rest of the different stages of nickel affinity purification; FT- flow-through, W1- wash 1, W2- wash 2, W3- wash 3, W$- wash 4, E-elution.

The concentration of 0330 was low even before incubation with the TEV protease, and very little was left following the second nickel affinity purification, therefore this was not taken forward. As a final step, to ensure only the protein of interest would be tested on our ELISAs and minimise the risk of *E. coli* contamination, the three 6XHistag cleaved proteins were purified by gel filtration, which separates protein based on size. Again the fractions eluted from the column were run on an SDS-PAGE gel to confirm the protein location and presence (Figure 3.6). To ensure the correct protein had been purified, an aliquot 3155 and 0342 S1 was sent to Kings College London Centre of Excellence for Mass Spectrometry for liquid chromatography mass spectrometry (LC MS/MS). 2697 was analysed using LC MS/MS by Dr Sherif Abouelhadid at Imperial College London. Analysis for all three revealed the correct protein had been purified. This provided further confirmation that the correct protein had been purified for 3155, which was running at approx. 6 KDa on the SDS-PAGE gel as opposed to the predicted 9.2 KDa size of the protein.

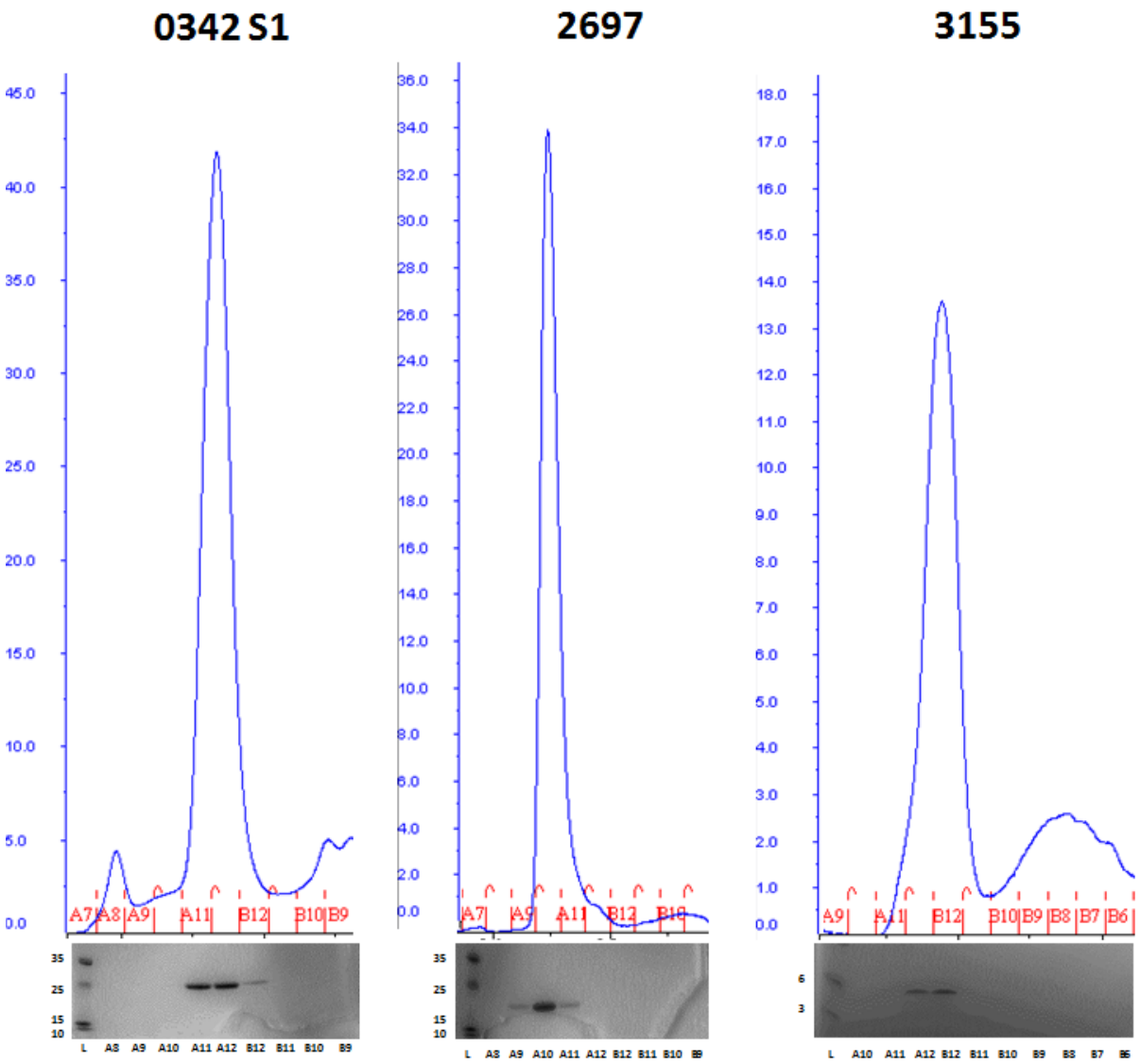


Figure 3.6. Purification of the *C. difficile* recombinant array candidates using gel filtration. Flow-through samples isolated form the nickel affinity purification following TEV mediated cleavage of the 6XHistag were passed through a gel filtration column as an additional purification step. This separates protein based on size and was successful for all three proteins, as demonstrated in the corresponding SDS-PAGE gels.

3.6. Screening of recombinant *C. difficile* proteins against CDI patient sera

Three of the potential vaccine candidates identified on the protein array (2697, 0342 S1 and 3155) were purified to a sufficient concentration for screening against serum samples from patients with CDI and controls. To do so, an indirect ELISA was utilised. In this method, the protein antigens were bound to the bottom of a 96-well plate before probing with human sera (using a protocol from [460] and modified by Dr Tim Scott). Serum-derived antibodies bound to the protein were then detected using

a HRP-conjugated anti-human IgG antibody which enables colourmetric quantification of antibody binding when incubated with the 3,3',5,5'-tetramethylbenzidine (TMB) substrate.

3.6.1. Optimisation of ELISAs for protein screening

The initial optimisation steps for the ELISAs, to determine the optimal plate type and coating buffer for adhering the protein to the plate, were performed using each protein still harbouring their 6XHistag, to enable detection with an anti-His antibody rather than the limited human serum samples. Each protein was serially diluted in coating buffer, either carbonate/bicarbonate or PBS, and used to coat a Nunc Maxisorp flat-bottom or Nunc Medisorp flat-bottom plate. These plates are designed for use in immunoassays, with the Medisorp plate recommended for more hydrophobic proteins. Plates were incubated with the protein overnight then probed with rabbit anti-His antibody followed by HRP-conjugated goat anti-rabbit IgG, with thorough washing steps between incubations. The TMB substrate which reacts with the HRP-conjugate was added to each well and the reaction stopped with 1 M H_2SO_4 when the two highest concentrations of protein appeared to be the same colour by eye, suggesting the reaction had reached saturation point. OD_{450} readings were taken and the absorbance of each protein in the different conditions compared. All proteins were successfully bound to the plate, as demonstrated by the increase in absorbance reading with increasing protein concentration (Figure 3.7). There was no difference between the Maxisorp and Medisorp plates, however, carbonate-bicarbonate buffer did appear to confer improved protein coating over PBS (Figure 3.7). From this, Medisorp plates with carbonate-bicarbonate coating buffer were selected for the screening conditions.

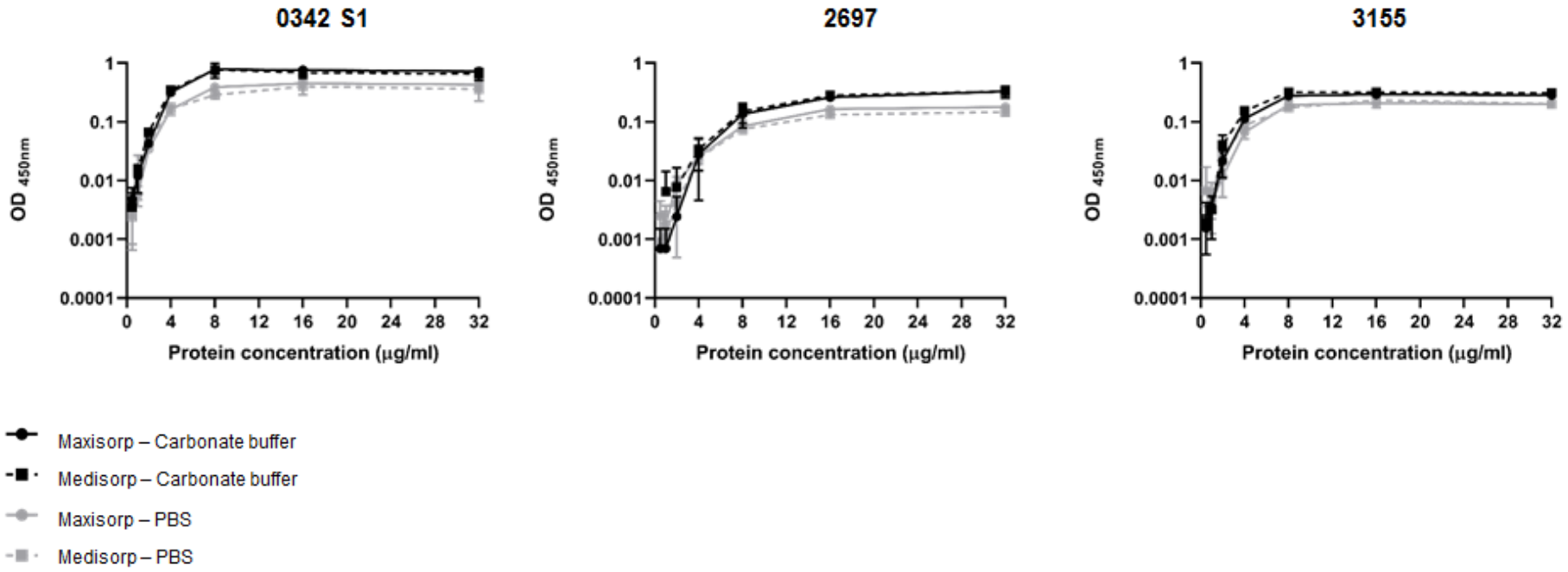


Figure 3.7. Optimisation of ELISAs using 6XHistagged protein. Plate type and coating buffer for use in the ELISAs were determined prior to testing with human sera using each of the purified proteins still harbouring their 6XHistag. Each protein was serially diluted (32 µg/ml to 0.5 µg/ml) in carbonate-bicarbonate buffer or phosphate buffered saline and bound to a Nunc Maxisorp or Medisorp plate overnight. Proteins were identified using anti-his primary antibody followed by a secondary antibody conjugated to horseradish peroxidase which was detected using TMB substrate. Reactions were stopped with 1 M H_2SO_4 following appropriate colour development and results read using a plate reader at OD 450_{nm}. Each protein concentration was tested in duplicate and the assay performed twice. Error bars denote standard deviation from the mean.

3.6.1. Testing patient samples

A total of 39 serum samples were provided by UoL for screening against the three purified proteins. These samples are in the pipeline for screening on the protein array by UCI (includes longitudinal sampling of CDI patients). The samples provided by UoL consisted of 10 samples from the CDI group, nine from the CDH group (Table 3.1) and 20 from a diarrhoea CDU group. Due to time and protein sample limitations 10 samples from the CDU group were run for 2697 and 0342 S1. There was insufficient protein available to run these against 3155. As it wasn't possible to clone the desired control, purified human IgG was used as a positive control for secondary antibody binding and buffer only as a negative control.

To optimise the assay for use with human serum samples, one CDI and one CDH sample were screened against serial dilutions of the three purified proteins, from 1.5 µg/ml to 50 µg/ml. A broader dilution range was used compared to the anti-his ELISAs as the reactivity of the serum with the purified protein was unknown. A buffer only control was used to determine background signal for each serum sample and purified human IgG was used as a positive control for the secondary antibody.

A signal was generated for all three proteins with both CDI and CDH serum samples, resulting in saturation at the highest protein concentrations (Figure 3.8). Signal strength varied greatly between both proteins and samples, therefore three concentrations of protein were selected for the full scale sample testing; 12.5 µg/ml, 6.25 µg/ml and 3.125 µg/ml. These concentrations translated to a signal between baseline and saturation in the test assay.

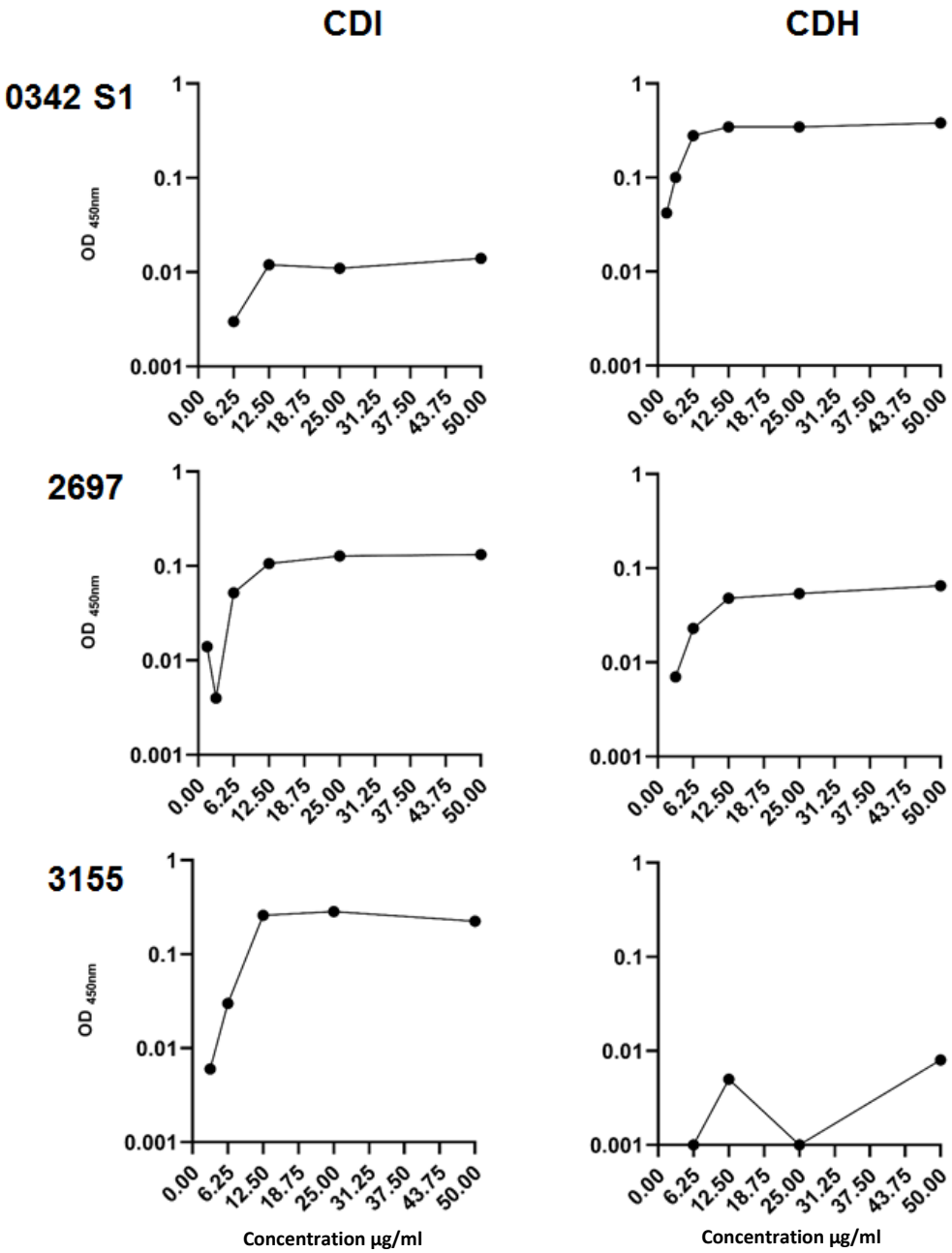


Figure 3.8. Optimisation of ELISAs using human plasma samples. To determine the concentration range of each protein for screening human plasma samples, a range of protein concentrations were first probed with one CDI and one CDH plasma sample. Plasma was diluted 1:100 before incubating with the serially diluted protein (50 µg/ml to 1.5 µg/ml). Human antibodies in the plasma were identified using anti-human IgG secondary antibody which was conjugated to horseradish peroxidase, enabling colourimetric detection upon addition of the TMB substrate. Reactions were stopped with 1M H_2SO_4 following appropriate colour development and results read using a plate reader at OD 450_{nm}. This assay was performed once, as a screen to ensure sera responses could be detected.

Sample testing was conducted for the 10 CDI and 9 matched CDH samples, as described above, and the difference between the IgG signal in response to the different antigens was plotted for both groups and assessed using both a Students *t*-test and Wilcoxon rank test (Figure 3.9a). Any significant differences were set at a cut off $p < 0.05$. As seen in the optimisation steps (Figure 3.8), there were variations in the signal between the three protein antigens tested. The IgG response varied between protein antigen and sera samples (Figure 3.9a), with a similarly varied response pattern also seen in the results from the protein array when all patients' screened are plotted individually (Figure 3.9b). However, unlike in the array results where all three proteins were associated with a significantly higher IgG response in the CDH group (Figure 3.9b), there was no significant difference in IgG response between the CDH and CDI groups for any protein, although the mean IgG signal against 0342 S1 and 2697 did look to be higher by ELISA in the CDH compared to the CDI group (Figure 3.9a).

Proteins 2697 and 0342 S1 were also screened against 10 serum samples from the CDU group. These samples were included as an additional group of patients who were in hospital and negative for the *C. difficile* toxin by ELISA, but still had diarrhoea of an unknown cause. Further investigation of these samples by the UoL using additional *C. difficile* diagnostics found some samples within this group were GDH and *C. difficile* culture positive, suggesting these patients had CDI. Therefore, the available metadata was used to categorise our CDU samples into two groups i) GDH positive culture positive (GDH +) which were most likely CDI positive patients and ii) GDH negative culture negative (GDH-), which potentially have diarrhoea from an unrelated source (non-CDI) (Figure 3.9a). There was no significant difference in response for either antigen when comparing the response in the GDH positive and negative groups.

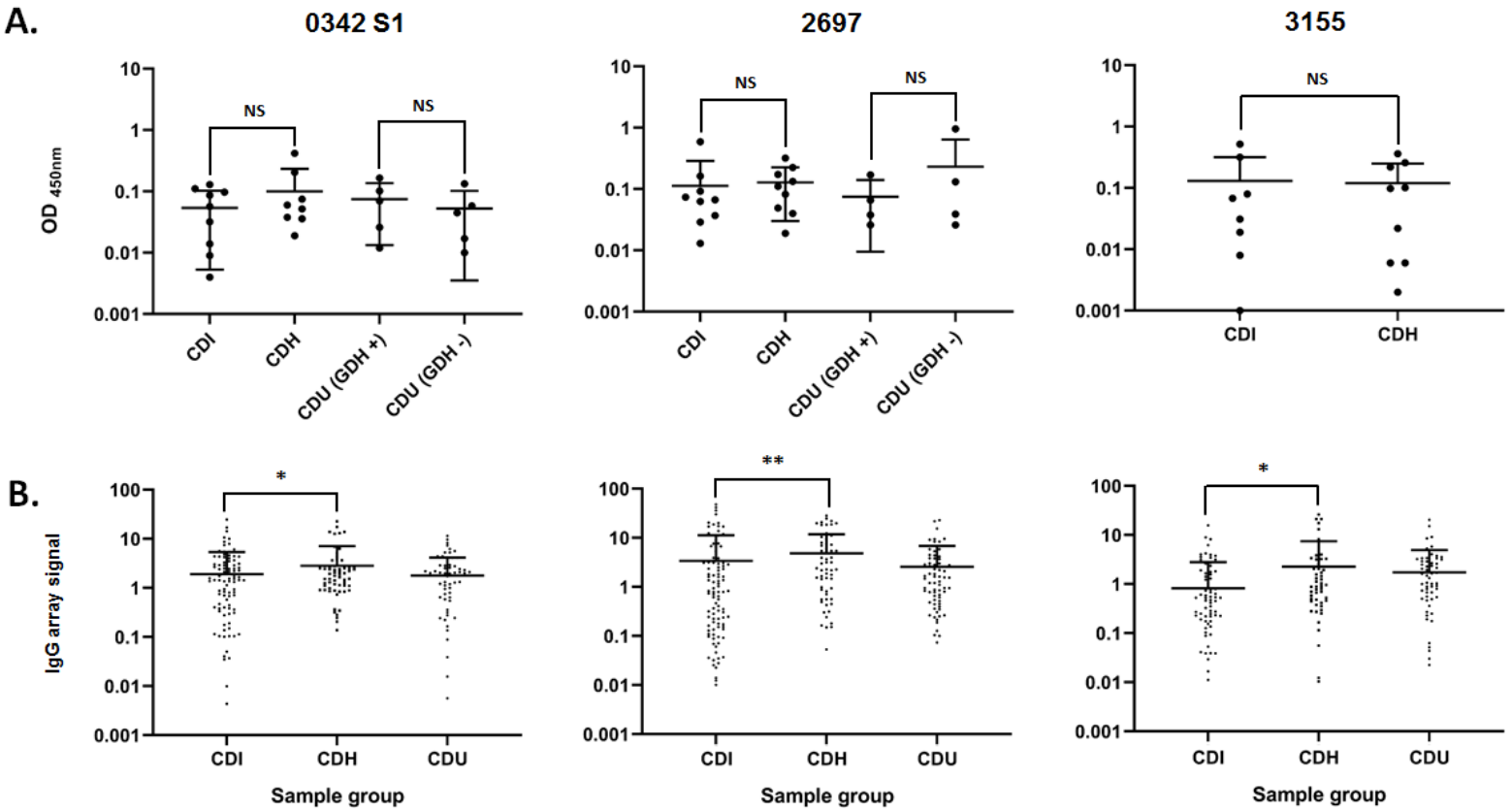


Figure 3.9. Screening of human blood samples to quantify IgG reactivity towards three *C. difficile* proteins. Three *C. difficile* proteins were screened against human samples in either an ELISA or protein array format. (A) Three recombinant proteins from *C. difficile* were screened against 29 sera samples; 10 from patients with CDI (CDI), nine from healthy controls (CDH) and 10 from people with diarrhoea of unknown cause (CDU) which were further split depending on if the glutamate dehydrogenase (GDH) test was positive or negative. This was performed in an indirect ELISA using a horseradish peroxidase conjugated secondary antibody which recognises human IgG and enables chemiluminescent detection of IgG binding. (B) Results of the three *C. difficile* proteins on the protein array when screened against 279 human plasma samples; 128 CDIN, 71 CDH and 80 CDU. Statistical significance was determined by *t*-test or Wilcoxon rank test where $p<0.05$ was significant, *$p<0.05$, **$p<0.005$, NS- non-significant. Each protein was tested in duplicate at three concentrations and the assay performed once. Error bars denote standard deviation from the mean.

Additional data for each patient in the CDI group is detailed in Table 3.4, including culture result, ribotype of current infection, previous cases of CDI, severity of infection, relapse and antibiotic treatment prior to and during the study. The response to each protein on the ELISA was conditionally formatted, with highest responses in red and lowest in green and compared against this patient data to investigate whether there appear to be an association between sample response and patient outcome.

From this, there did not appear to be a clear association between response to the samples in the ELISA and the outcome of *C. difficile* infection. Patients CDI 04 and CDI 05 were both continuously colonised with *C. difficile* throughout the study period and demonstrated a low IgG response to all three proteins. This suggests that they had not mounted a strong immune response to these antigens. However, there were also instances of patients clearing the infection whilst demonstrating no response to 0342 S1 or 2697 (CDI 09) and on the reverse, patient CDI 01 demonstrated a comparably high IgG response to all three proteins yet relapse of CDI was reported 2 weeks post-diagnosis. Therefore, from this sample set it was not possible to conclude as to whether 0342 S1, 3155 or 2697 would provide protection from infection if included as a vaccine component.

ELISA IgG response of serum samples from the CDIN group										
0342 S1	0.087	0.009	0.111	0.004	0.032	0.097	0.014	0.129	0.000	0.056667
2697	0.590	0.074	0.092	0.029	0.013	0.067	0.163	0.037	0.000	0.063333
3155	0.068	0.517	0.008	0.001	0.080	0.031		0.317		0.018833
Clinical data for participants in the CDIN group										
Sample No.	CDI 01	CDI 02	CDI 03	CDI04	CDI 05	CDI 06	CDI 07	CDI 08	CDI 09	CDI 10
Culture result	POS	POS	POS	POS	POS	POS	POS	POS	POS	POS
GDH result	N/A	N/A	N/A	N/A	POS	N/A	N/A	N/A	N/A	POS
Ribotype	049	002	078	017	002	002	N/A	027	014	N/A
Previous CDI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Severity of CDI	MILD	SEVERE	MODERATE	SEVERE	MODERATE	SEVERE	MILD	SEVERE	MODERATE	SEVERE
Relapse	Diarrhoea -2 weeks	Continued colonisation	NO	Continued colonisation	Continued colonisation	Continued colonisation	NO	NO	NO	NO
Previous antibiotics	CIP	-	-	-	BEN, AMX	GEN, AMC, TZP, MEM	SXT, GEN, FID, TZP	CEF, TZP	AMC	AMC, GEN, TZP
CDI treatment	MTZ, VANC	FID	NONE	FID	MTZ, FID	FID	FID	VANC, MTZ	MTZ	MTZ
Antibiotics during study	-	-	-	-	FLU, BEN	MEM	-	-	-	-
PPIs or H2Bs	YES	YES	NO	YES	NO	YES	YES	NO	YES	YES

Table 3.4. IgG reactivity of three *C. difficile* proteins screened against sera samples from patients with CDI, with matching clinical data. CDIN, blood samples from patients with CDI, POS-positive, NEG-negative, CIP-ciprofloxacin, BEN-benzlycillin, AMX-amoxicillin, GEN-gentamicin, AMC-co-amoxiclav, TZP-tazocillin, MEM-meropenem, SXT-co-trimoxazole, FID-fidaxomycin, CEF-ceftriaxone, MEZ, metronidazole, VANC-vancomycin, FLU-flucloxcillin, N/A- not available, PPI-proton pump inhibitor, H2B- histamine-2 receptor blocker.

3.7. Discussion

This chapter describes the identification of novel, immunogenic *C. difficile* antigens that hold potential as vaccine candidates for protection against CDI. These were identified based on the results of a *C. difficile*-specific protein array, a high-throughput platform for rapid screening of multiple antigens against human samples. Array-based characterisation of the humoral immune response during CDI is one component of an ongoing programme grant investigating progression of CDI, including analysis of host biomarkers and the intestinal microbiota.

Protein arrays are yet to be fully exploited for the investigation of *C. difficile*. Negm *et al.* demonstrated their utility for screening purified Toxin A and B and three ribotype specific SLPs from *C. difficile*, with sera samples from patients with CDI, cystic fibrosis patients negative for symptomatic CDI and healthy controls [272]. Here, higher anti-SLP IgG responses were identified in the healthy controls compared to those with CDI, but there was no difference in the two groups when comparing the anti-toxin response [272]. An additional study found a higher anti-Toxin A and B IgA response and a higher level of neutralising IgA towards Toxin A in those with cystic fibrosis compared to those with CDI and healthy controls [461].

The studies described above screened samples against array slides printed with purified protein. Although this is possible for a small number of antigens, it becomes impractical to prepare purified antigenic material for whole proteome analysis. The use of IVTT for the arrays described in this chapter enabled transcription and translation of the target protein in a one tube, cell-free system which is directly printed on the slides for probing [457]. Consequently, over 3000 *C. difficile* proteins were screened with plasma samples from patients with CDI, healthy controls (CDH) and those with diarrhoea of an unknown aetiology (CDU).

From the primary pilot screen three candidates were prioritised to follow-up with for phenotypic characterisation as these were immunogenic proteins in the top 100 reactive antigens. Subsequently,

in the second array of 279 samples, 25 potential vaccine candidates were shortlisted, which displayed a significantly higher IgG response in the CDH over the CDI group. Two of these were CDR20291_0330 and CDR20291_0342, those proteins chosen in the pilot screen. The third protein, CDR20291_3343, was associated with a higher response in the CDH group but this was not significant. However, this antigen was still investigated as pilins are a popular target for vaccines in many species [462-465]. A higher antibody response in healthy people suggests a protective effect from infection. This has been demonstrated for other protein in multiple studies which showed higher anti-toxin and anti-cell surface protein antibodies, such as FliC and Cwp84, in those without CDI or asymptomatic carriers of *C. difficile*, compared to those with CDI [86, 270, 273, 461]. However, these associations are not consistent across all studies. The relative contribution of antibodies to Toxin A versus Toxin B, the antibody isotype (IgG and IgA) and the overall role of anti-toxin antibodies in acquisition of infection, severity and duration of CDI and relapse, varies between reports [269, 271, 272, 275, 444-446]. Indeed, in our antigen shortlist, purified toxin A was present (but not pursued due to the wealth of information already available), yet purified toxin B and domains of toxins A and B produced using IVTT were not. One possible reason for the inconsistencies across the existing literature is the large variation in sample cohorts between reporting studies, including; geographical location, antibiotic administration, average age of the patient, co-morbidities, control groups selected for comparison and the overall number of samples. Further investigations should aim to stratify the existing data to facilitate meaningful comparisons in conjunction with the design and performance of additional research.

In relation to the study groups investigated here, the IgG response between the CDI and CDH group were compared as the third group, the diarrhoea, *C, difficile* toxin negative CDU group has not been fully characterised. This group consists of hospitalised patients with diarrhoea of an unknown cause and negative for *C. difficile* Toxins A and B. Further analysis of these has found a subset of patients to be positive for *C. difficile* by GDH test and culture, therefore the lack of toxin detection could either

indicate a false-negative result by Toxin A/B ELISA or asymptomatic colonisation [466, 467]. Stratification of this group is ongoing, but in order to provide a valuable contribution towards analysis of the humoral immune response, would ideally include:

CDI positive. Patients with toxigenic *C. difficile* causing symptomatic infection that was initially missed due to lack of sensitivity in toxin detection.

Asymptomatic carriage. Patients negative for the toxins but still colonised with *C. difficile*. Ideally these would be differentiated into toxigenic and non-toxigenic colonising strains.

CDI negative. Toxin, GDH and culture negative. This presents another control group who are hospitalised but free from both *C. difficile* infection and colonisation.

Those who are asymptomatically colonised with toxigenic *C. difficile* are an interesting control group, as it suggests a host response is preventing the development of CDI. Previously, higher anti-toxin antibodies have been found in those asymptomatically colonised with *C. difficile* compared to CDI patients [86]. The existing sample cohort described here lacks a specified asymptomatic colonisation group but this could be derived from further analysis of the diarrhoea-positive CDU patients, including samples that are stool culture positive for *C. difficile* but free toxin negative. However, for these it would be very important to confirm the absence of free *C. difficile* toxin, ideally in combination with a diagnosis of the actual cause of diarrhoea, to confidently exclude symptomatic *C. difficile* infection. Following categorisation of these patients, associated longitudinal samples could be analysed to aid differentiation between transient and persistent colonisation, and how this relates to humoral immunity [7, 52]. Within this, it is obviously important to consider other factors that could prevent the switch from asymptomatic colonisation to CDI, such as the presence of certain bacterial taxa within the gut microbiota and antibiotic exposure [7, 468, 469].

Ideally, an anti-*C. difficile* vaccine would prevent both the symptoms of CDI and colonisation of the pathogen. Comparison of the three groups described above plus the healthy controls could enable

further dissection of humoral immunity in these different states, with an aim to identify antibody responses specific to each. For example, by taking those patients who are in hospital with diarrhoea but negative for symptomatic CDI and comparing between those who are positive or negative for *C. difficile* colonisation, it may be possible to identify a specific anti-colonisation antibody signature.

Screening of a larger proportion of the total sample cohort could aid in answering some of the questions raised above, especially when including the analysis of longitudinal samples and the host microbiota. This work is ongoing, with additional plasma samples being screened against a smaller number of antigens on a specifically designed down selected array than tested previously, focusing on proteins which were identified by the IgG response in the whole proteome screen as potential diagnostic or vaccine candidates.

High sera IgG responses to *C. difficile* in healthy people has been reported previously and was the basis for selection of potential vaccine candidates in this study. It has been postulated that the high rate of circulating, systemic antibodies to *C. difficile* in healthy people is a result of prior exposure, for example in the environment or during colonisation in infancy [272, 470, 471]. This enables the development of immune memory which is activated upon exposure to or transient colonisation with *C. difficile* [472, 473].

However, this does raise the question as to why serum antibodies are so important for protection from a gut pathogen, as the secretory IgA response within mucosal immunity would presumably be more relevant. Robust mucosal immunity has been demonstrated to provide protection from numerous enteric pathogens, such as *Salmonella Typhi* [474]. Indeed, Warny *et al*. found higher faecal anti-Toxin A IgA in patients with mild CDI over those with more severe or recurrent infections [274] and comparisons of colonic biopsies between CDI patients and healthy controls found higher numbers of IgA-secreting cells in the latter group [475]. Furthermore, Hong *et al.* demonstrated protection from *C. difficile* colonisation in hamsters orally immunised with *Bacillus subtilis* spores expressing a portion of Toxin A, with a high secretory IgA response to Toxin A detected 14 days after completion of

immunisation [276]. As for the systemic response, its development and role is not fully understood. Suggestions include toxin-mediated disruption of the epithelial cells lining the gut enables circulating antibodies to encounter *C. difficile*, either by *C. difficile* antigens entering the mucosal tissue or systemic antibodies moving into the gut lumen [476, 477].

The main limitation of protein arrays is the lack of information on antibody functionality. Antibody binding does not necessarily equate to a protective immune response and it is the activity of that antibody when bound to the antigen that is important. This is therefore an important area for future work, focusing on antibody binding to bacteria cells and neutralisation and opsonophagocytic assays to determine the response of the antibodies to the antigen of interest. It is also important to note that *C. difficile* is a complex infection and acquisition of CDI is influenced by a number of factors in addition to the antibody response, including immune capacity of the patient, previous exposure to *C. difficile*, state of the gut microbiome and co-morbidities. This can be addressed in part using patient metadata in conjunction with the antibody response results. Another drawback to using this technology in *C. difficile* is that the *E. coli* background can mask immunogenicity of the *C. difficile* proteins as the systemic antibody response is not high enough to overcome it. Blood-borne pathogens have a much higher systemic antibody response meaning this background is less of a problem [384]. This was the reason behind the use of purified protein for the ELISAs in this study. Regions of the protein selected for expression were based on likelihood of exposure to the immune system, meaning signal peptides and transmembrane domains were removed. As these are regions of hydrophobicity, their removal also aids purification of soluble protein. The regions selection for expression are therefore unlikely to be identical to those printed for the array, which were not purified and instead spotted directly on the slides. It is therefore possible that removal of the hydrophobic domains may have influenced the immune recognition of the proteins, not necessarily in terms of reactivity of the domains themselves, but how their removal may alter folding of the protein, and consequently the loss of certain conformational epitopes.The proteins were screened with polyclonal sera which is very likely to

recognise numerous epitopes, both conformational and linear, but the measured response could still be reduced if particularly immunodominant epitopes were lost. Indeed, if the whole protein were required for future investigations, a denaturing agent such as urea could be used to solubilise the protein, enabling purification [478]. It is also noteworthy that one limitation of protein arrays, even when the whole protein is successfully expressed, is the difficultly in ensuring the protein is folded correctly when printed on the array, in order to ensure conformational epitopes are presented as they would be *in vivo* [479]. In addition to protein folding, the success in cloning and expression of the protein may also affect the antibody response as protein concentration is not quantified before spotting on the array when using the IVTT system. Prior to testing, the ELISAs were optimised using both anti-His antibodies and human sera samples. When detecting bound anti-IgG antibodies using the TMB substrate, following incubation it was necessary to stop all reactions at the same point in order to enable comparison across the different serum samples. However, due to the natural variation in reactivity between protein antigens and serum samples, both found here and reported previously, some samples reached saturation before others had a detectable response [273, 445]. In the future, and with unlimited protein and sera samples, checkerboard assays could be performed to enable detection of the optimal combination of protein concentration and serum dilution. Here all combinations of protein and serum concentrations are tested until the correct ratio is determined [480]. Final testing could then use this information to include a larger range of protein concentrations and serum dilutions to better determine antibody responses in a varied sample population.

Although this study provided a valuable starting point in the identification of novel vaccine candidates using a protein array, only a small sample set was investigated in the ELISAs, which limited conclusions to be drawn over the immunogenicity of the antigens of interest. The IgG response to 2697, 3155 and 0342 S1 was significantly higher in healthy controls compared to those with CDI when assessed by protein array, whereas although 0342 and 2697 signals were higher in the CDH group by ELISA, these differences weren't significant. This may be a result of the small sample sizes available for statistical

analysis compared to the 279 samples from the protein array which would likely have resulted in a higher power analysis. Furthermore, as noted, there was substantial variation in samples responses and as these were randomly selected, it may be that those tested were skewed towards a particular response. Regarding antigen choice, if this were to be repeated it would be useful to also include those proteins where a high response was identified in all patient groups. This would be a means of validating the array and identifying immunoreactive candidates, before then undergoing further analysis of any differences in response between patient groups groups. The down selected protein arrays that are currently underway will probe 250 IVTT expressed antigens, including 2697, 0342 S1 and 3155 both purified and expressed using IVTT, against a number of additional samples including those tested here by ELISA. This will facilitate comparison of purified versus IVTT expressed protein, as well as between the protein arrays and ELISAs for measuring IgG reactivity. Previously, Negm *et al.* found significant correlation between ELISA and protein array quantification of anti-toxin antibodies, although this association was weaker for Toxin B compared to Toxin A [272].

4. Functional characterisation of CDR20291_0330, a putative cobalt transport protein

4.1. Introduction

The *C. difficile* protein array identified a number of immunogenic proteins, including CDR20291_0330. In the pilot study this protein was found to be immunogenic then in a follow up, using healthy matched controls to the CDI patients, a higher serum IgG response to the protein was detected in healthy controls compared to patients with CDI. This response is associated with potential vaccine candidates, as a higher response in the control group suggests a protective effect of harbouring antibodies to these proteins.

CDR20291_0330 is a putative cobalt binding protein. Cobalt, along with many other micronutrients, is essential for bacterial survival, although it is usually only required in minute concentrations by the cell. Cobalt is primarily required for synthesis of the different cobalamin compounds through its incorporation into the corrin ring [481, 482]. Vitamin B_{12} (cyanocobalamin) is one of the most notable members of this groups and is important in many cellular processes, as are its derivatives (hydroxycobalamin, methylcobalamin and adenosylcobalamin), particularly as enzyme co-factors [483]. For clarity, from here on the different vitamers will all be referred to as B_{12}.

Some bacteria, along with archaea, are unique in their ability to synthesise B_{12}, with both aerobic and anaerobic biosynthesis pathways identified in a broad range of organisms [484, 485]. Additionally, dedicated transporters can be utilised for direct uptake of B_{12} from the environment [486]. Cobalt and other micronutrients are often scarce within the environments that bacteria reside, meaning high affinity uptake transporters are employed to scavenge for necessary substrates. This includes energy coupling factor (ECF) transporters which mediate the uptake of micronutrients via hydrolysis of ATP [487]. They are similar to, but functionally distinct from ABC transporters and are composed of three

universal units; an A unit with ATPase activity, a T unit forming the transmembrane channel and an S unit for substrate binding [488, 489].

Without *in vitro* characterisation, identification of transporters and more importantly, their targeted substrate is often based on a combination of factors, including; phylogenetic analysis, associated regulatory elements and the putative or actual function of genes with which they are co-localised [490]. Using this system, Rodionov *et al.* identified the cobalt ECF transporter, CbiMNQO (Figure 4.1) [490]. This transporter encodes two substrate binding components (CbiM and CbiN), a permease, (CbiQ) and an ATPase, (CbiO). It was confirmed to mediate high affinity uptake of cobalt when CbiMNQO transporters from *Rhodobacter capsulatus* and *Salmonella enterica* serovar *Typhimurium* were reconstituted *in vitro* [490, 491].

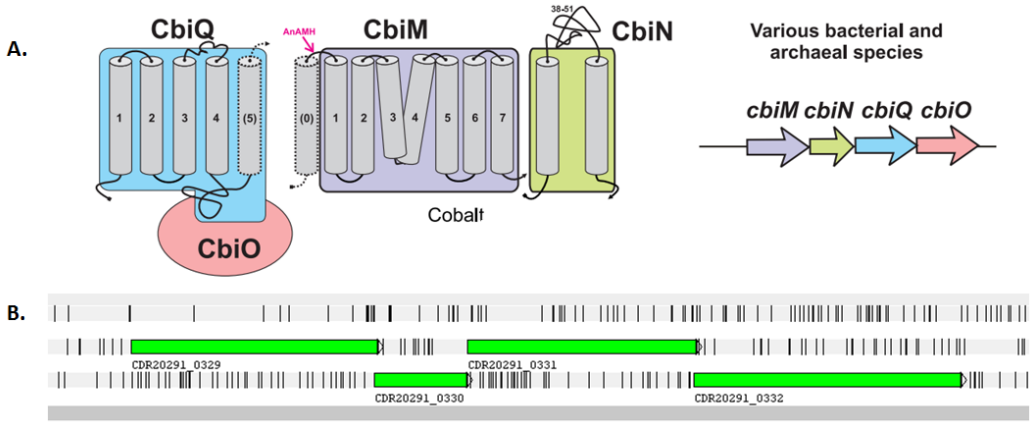


Figure 4.1. Organisation of the CbiMNQO transporter. (A) Schematic of the suggested organisation of the CbiMNQO cobalt transporter in bactera, including the ATPase (CbiO), transmembrane components (CbiQ) and substrate binding domains (CbiM and CbiN). Image taken from Rodionov *et al.* [490]. (B) Genomic organisation of the putative CbiMNQO transporter in *C. difficile* R20291 including; CbiM (*CDR20291_0329*), CbiN (*CDR20291_0330*), CbiQ (*CDR20291_0331*) and CbiO (*CDR20291_0332*).

CbiMNQO and other genes encoding cobalt/B_{12} transporters or B_{12} synthesis pathways are often encoded downstream of a B_{12} riboswitch. These regulatory elements are encoded within the 5' untranslated region of their associated gene and can be directly bound by B_{12} leading to structural changes within the RNA, resulting in transcriptional termination or translational sequestration [492, 493]. Three riboswitches have been annotated within the *C. difficile* 630 genome, upstream of the

cbiMNQO operon, a putative B_{12} transporter (BtuFCD) and within the putative B_{12} biosynthesis pathway [111, 492]. Micronutrient uptake transporters such as CbiMNQO are encoded with a riboswitch to prevent build-up of their substrate within the cell, whichcan be toxic [494].

CDR20291_0330 is the putative cobalt binding protein (CbiN) from CbiMNQO and we sought to characterise the role of this protein within the cell, specifically in terms of the requirements for cobalt and consequently B_{12} in *C. difficile*.

4.2. Bioinformatics

When searching the amino acid sequence of CDR20291_0330 against other strains of *C. difficile*, this protein was found to be highly conserved and present in strains from different clades (Table 4.1).

Strain	Ribotype	Percentage identity (%)
R20291	027	100
CD196	Historic 027	100
630	012	96.8
M120	078	96.8
M68	017	96.8
CD305	023	96.8

Table 4.1. Conservation of CDR20291_0330 amino acid sequence across *C. difficile* strains. The conservation of the CDR20291_0330 amino acid sequence across a number of *C. difficile* strains from diverse ribotypes was assessed using pBLAST searches with R20291 as the comparator sequence.

Phyre^2 analysis identified CDR20291_0330 as a putative membrane protein with a signal peptide (1 to 21 amino acids) and single transmembrane domain. However, CbiN has previously been identified as harbouring two transmembrane domains flanking an extracytoplasmic loop [490]. This organisation was supported when the amino acid sequence was analysed using the transmembrane domain prediction software, TMHMM Server v. 2.0., which identified two transmembrane domains. As with the Phyre^2 analysis, SignalP 5.0 also identified a signal peptide, belonging to the Sec pathway. For annotation of the CDR20291_0330 sequence with this domains, see Chapter 3, Figure 3.2.

In addition to the presence of a B_{12} riboswitch, the annotation of CbiMNQO as a cobalt transporter is supported by the relatively high degree of sequence similarity between CDR20291_0330 and CbiN from *S. typhimurium* (92% query cover and 60% percentage identity) and *R. capsulatus* (query cover 90% and percentage identity 49%) (Figure 4.2). As discussed above, both of these transporters have

been demonstrated to be highly specific for cobalt uptake, and inactivation of CbiN abolished transport [491].

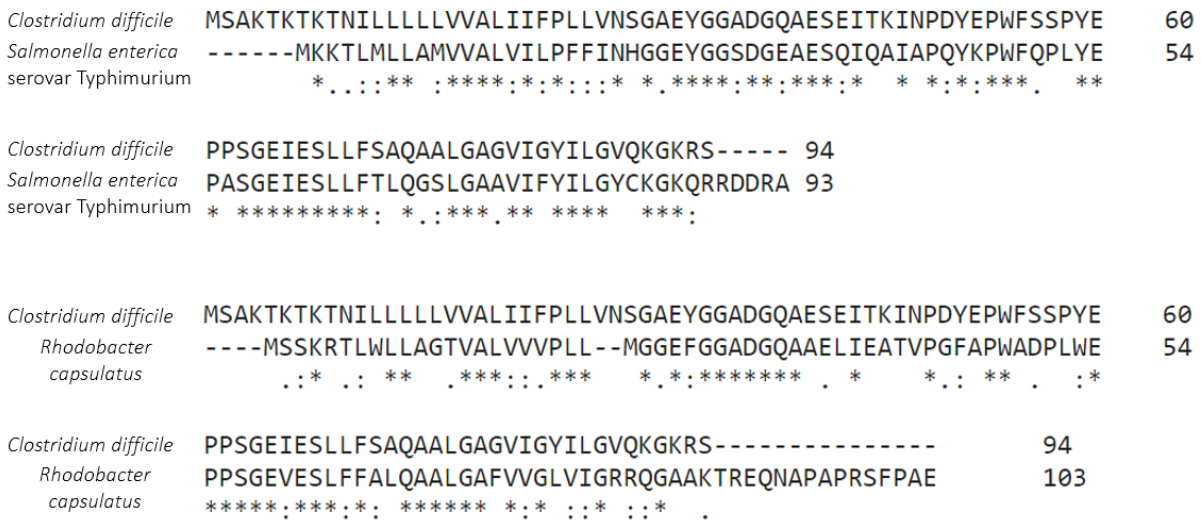


Figure 4.2. Alignment of amino acid sequences of CbiN. The amino acid sequence of CDR20291_0330 from *C. difficile* was aligned with the CbiN sequence from *Salmonella typhimurium* (92% query cover and 60% percentage identity) and *Rhodobacter capsulatus* (query cover 90% and percentage identity 49%) using pBLAST.

4.2.1. Cobalt and B_{12} uptake in *C. difficile*

As discussed, prokaryotes encode many transport mechanisms to ensure they maintain sufficient intracellular levels of B_{12} and cobalt, but no putative cobalt or B_{12} transporters from *C. difficile* have been characterised *in vitro*. Genes were investigated that have been annotated as cobalt or B_{12} transport proteins in *C. difficile,* in addition to CbiMNQO, using pBLAST to analyse both the putative transporter and any associated genes (Table 4.2). Of note are two putative CorA transporters, which are typically associated with magnesium uptake, but have also been found to transport zinc and cobalt other bacteria [495]. Neither of these putative transporters were encoded with B_{12} riboswitches or genes relying on B_{12}/cobalt to function.

Gene (630/R20291)	Putative function	Co-localisation	B_{12} riboswitch?	Comments
CD0313/CDR20291_0317	Cadmium translocating ATPase. Heavy metal domain, implicated in inorganic ion transport including cobalt, cadmium, lead, and zinc	Encoded with CD0312, putative metalloregulator from the SmtB/ArsR family	No	
CD0591/CDR20291_0516	Heavy metal translocating ATPase	Encoded with the hypothetical protein CD0592	No	
CD00101-102/CDR20291_99-100	Putative ATPase and permease components of an ECF transporter	None	No	No S unit or riboswitch could be identified
CD0430-CD0431/not found in R20291	CD430 and CD431 are putative cbiQ and cbiO components of an ECF transporter	Encoded with a putative S unit (CD0429) and two predicted transcriptional regulators (CD0427 and CD0428)	No	No CbiN component identified, low homology of putative S unit with CbiM, could not be found in R20291
CD1044/CDR20291_900	HlyC/CorC family of transporters for metal ion uptake, including cobalt	None	No	
CD1831/CDR20291_1725	Magnesium transporter within CorA family. CorA transporters are also associated with zinc and cobalt transport	None	No	
CD2122/ CDR20291_2029	Magnesium transporter within CorA family. CorA transporters are also associated with zinc and cobalt transport	None	No	

Table 4.2. Putative cobalt transporters in *C. difficile*. List of putative cobalt transporters in *C. difficile*, analysed using pBLAST. The presence of B_{12} riboswitches and B_{12}-related genes was also investigated. The previously identified CbiMNQO transporter (CDR20291_0329-0332) is not included.

In relation to B_{12} uptake, the only probable B_{12} transporter is BtuFCD (*CD2997-2999/CDR20291_2832-2835*). This is a well-established B_{12} transporter in other organisms but has not been investigated in *C. difficile* [486, 496, 497]. In the *C. difficile* genome it is encoded downstream of a B_{12} riboswitch [111, 492]. Interestingly, both 630 and R20291 encode this transporter yet in R20291, there is a transposase insertion towards toward the 3' of BtuC (*CD2998*/*CDR20291_2833)* which encodes the permease component (Figure 4.3). Therefore, whether BtuCDE is actually functional in R20291 is unknown.

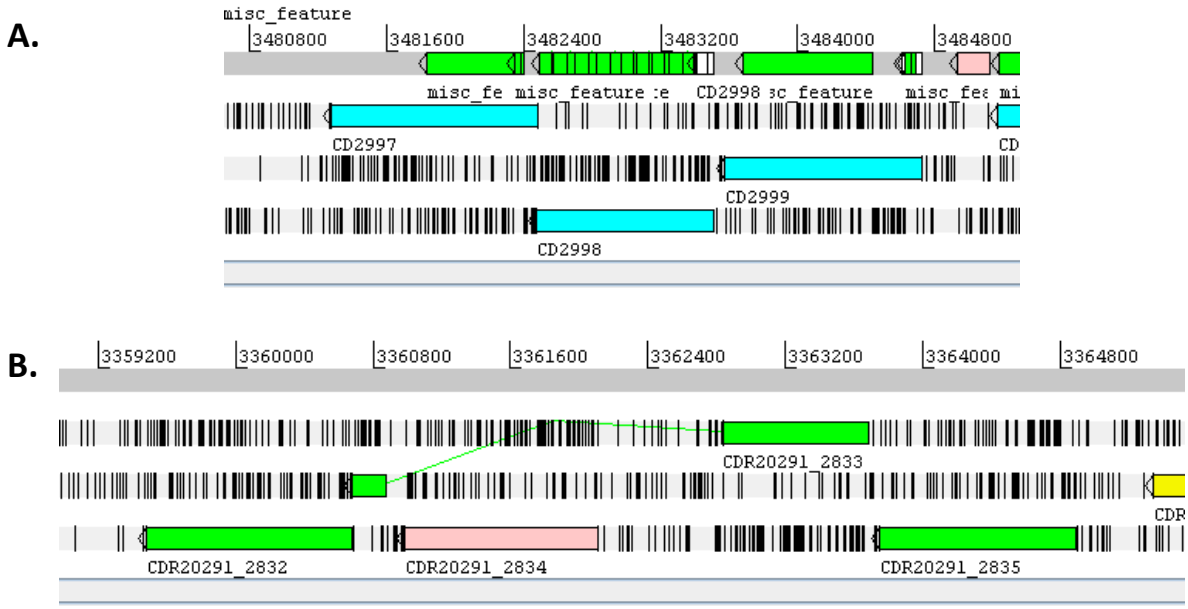


Figure 4.3. Comparison of the putative BtuCDE B_{12} transporter from *C. difficile* 630 and R20291. The putative BtuFCD transporter for uptake of vitamin B_{12} was identified in the genome of *C. difficile* strains 630 and R20291. There is a transposase insertion (*CDR20291_2834*) within the BtuC gene in R20291 (*CDR20291_2833*).

4.3. Construction of a *CDR20291_0330* gene deletion mutant

The major function for cobalt within bacteria is for B_{12} synthesis. *C. difficile* encodes the anaerobic pathway for B_{12} synthesis, suggesting it is capable of *de novo* synthesis. Therefore, it was hypothesised that inactivation of the CbiMNQO transporter, the only putative cobalt transporter identified in *C. difficile* to date, would impact on B_{12} synthesis in *C. difficile*. The requirements for cobalt and B_{12} in *C. difficile* are unknown but interestingly, screening of a transposon library in *C. difficile* found

inactivation of *cobT*, a gene within the anaerobic synthesis pathway resulted in loss of colonisation in mice, suggesting this is an essential pathway for *C. difficile* survival in the gut [498].

The loss of CbiN was hypothesised to result in loss of function of the transporter as as previous work in *Salmonella typhimurium* demonstrated loss of CbiN results in abolishment of transporter activity *in vitro* [491].

To achieve this, *CDR20291_0330* was deleted from the *C. difficile* chromosome using allele exchange mutagenesis, a technique to generate in-frame deletion mutants in *C. difficile*, which reduce the risk of polar effects resulting from insertional inactivation [439]. This utilises two homology arms (>1000 bp) flanking the gene for deletion, which enable the homologous recombination events to occur (Figure 4.4). These were introduced into R20291 using pMTL82151, a replication defective "pseudo-suicide" plasmid, which is only stable once integrated on the chromosome. Integration results in improved growth on thiamphenicol so larger colonies were selected for passage on non-selective agar to facilitate loss of the plasmid. This resulted in reversion to wild-type or a second cross over event leading to a gene deletion mutant which were differentiated by PCR.

CDR20291_0330 is only 286 bp, therefore for screening purposes, primers flanking the gene rather than the entire homology region were used, so the size shift would be easily visible by gel electrophoresis. As can be seen in Figure 4.4, clones were isolated where their corresponding amplicon was smaller compared to that from wild-type R20291. These potential deletion mutants were confirmed by amplification and sequencing of the entire homology region.

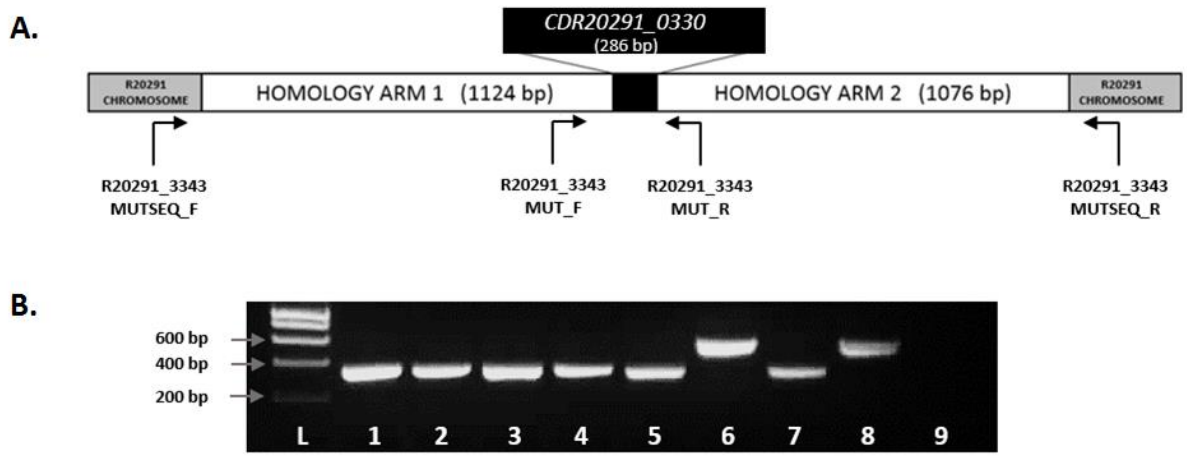


Figure 4.4. Allele exchange mutagenesis of *CDR20291_0330* in R20291. (A) Schematic of the allele exchange construct for generation of the *CDR20291_0330* gene deletion mutant. This is annotated with the primers used for construction of the cassette and for screening to detect gene deletion mutants. (B) PCR screen of potential *CDR20291_0330* gene deletion mutants using primers R20291_0330_MUT_f and R20291_0330_MUT_r. 1 to 7- individual clones from R20291, 8- wild-type R20291 gDNA, 9-dH_2O. The predicted size of amplicons with and without *CDR20291_0330* are 610 bp and 367 bp, respectively.

4.3.1. Growth kinetics in rich media

To determine whether R20291Δ*0330* had any growth impairments, OD_{595} readings were taken for R20291 wild-type and R20291Δ*0330* when growing in rich media, BHI (Figure 4.5). There was no difference in growth rate between the two strains.

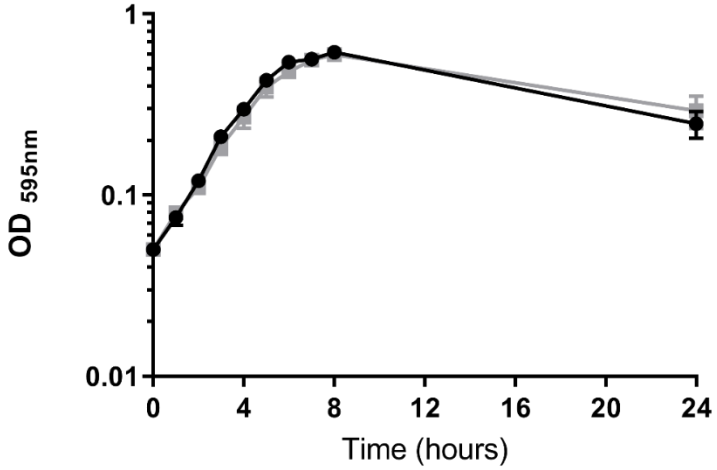


Figure 4.5. Growth kinetics of R20291 and R20291Δ*0330* in BHIS. R20291 and R20291Δ*0330* were grown overnight in BHI then diluted into fresh BHI media to a starting OD_{595} 0.05. OD readings were taken every hour for the first 8 hours then at 24 hours. Black- R20291 and grey- R20291Δ*0330*. Each culture was set up in duplicate and three independent replicates were performed. Error bars represent standard deviation.

4.3.2. Growth kinetics in minimal media

It may be that in rich media, B_{12}-dependent mechanisms are non-essential or that cobalt and B_{12} are freely available and the cell is capable of their uptake. To try and dissect this, the exact growth conditions were specified using *C. difficile* minimal media. Originally outlined by Karasawa *et al*. [32] and further modified to remove surplus ingredients by Cartman *et al.* [499], this media comprises the minimal components required for *C. difficile* growth and enables manipulation of the growth environment. Cobalt is provided in the form of cobalt chloride, but B_{12} was removed by Cartman *et al.*, suggesting it is non-essential in this setting. However, the effect of removing both B_{12} and cobalt has not been examined. We hypothesised that in the wild-type, absence of B_{12} in the media would result in no growth, but this would be compensated by providing cobalt to enable *C. difficile* to synthesise its own B_{12}. Furthermore, we hypothesised that in the mutant, the addition of cobalt would not enable growth due to the loss of function of the CbiMNQO transporter. Strains were inoculated

directly from plate cultures into minimal media containing either cobalt and B_{12}, cobalt or B_{12} or neither. These were grown for 24 hours then diluted into fresh minimal media matching the overnight conditions, and growth rate measured every hour at OD_{595}. Neither strain exhibited any differences in growth kinetics when comparing between growth media (Figure 4.6).

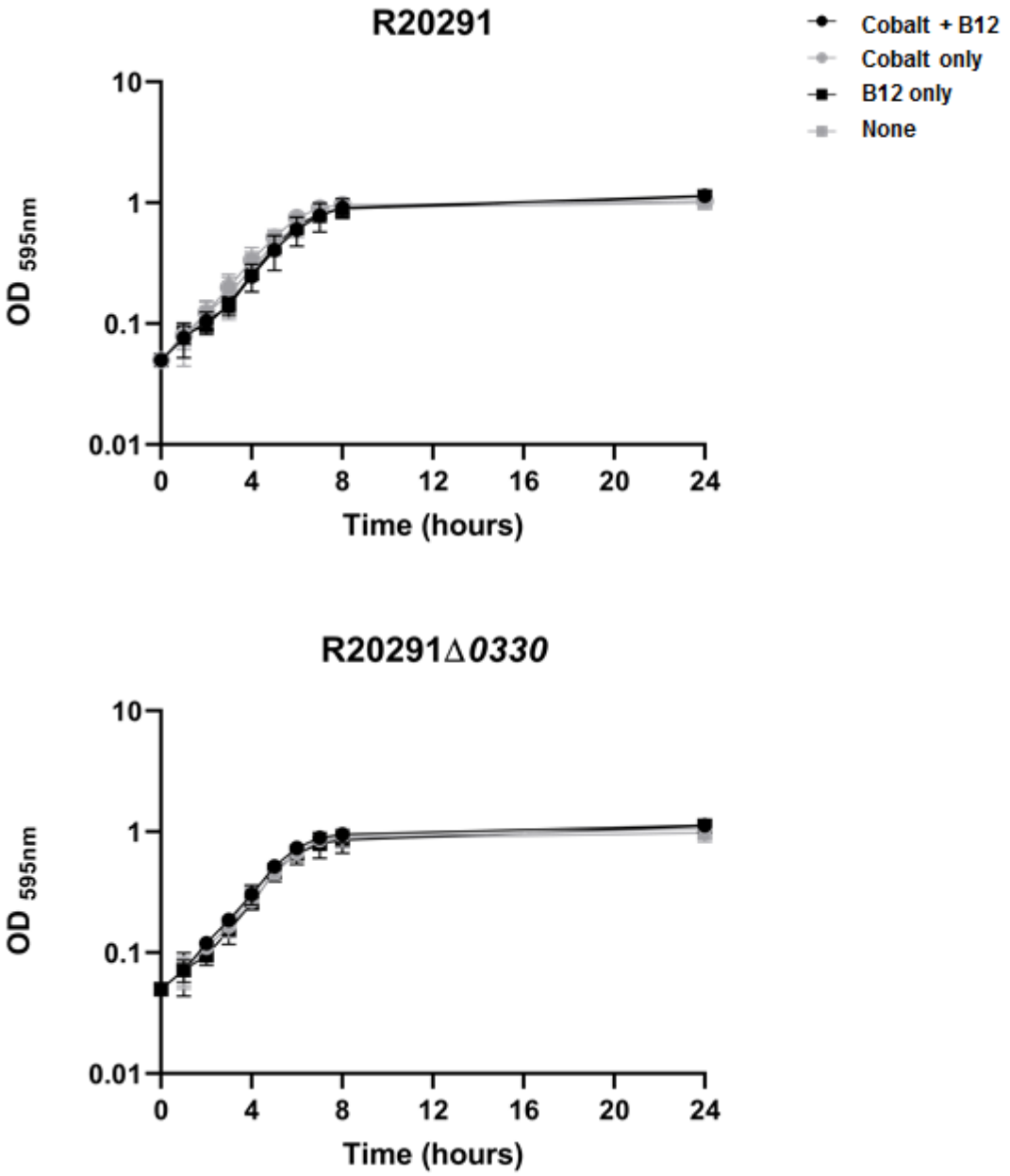


Figure 4.6. Growth kinetics of R20291 and R20291Δ*0330* in minimal media. R20291 and R20291Δ*0330* were grown for 24 hours in minimal media containing either cobalt and B_{12}, cobalt or B_{12} or neither then diluted into fresh minimal media matching the overnight conditions to a starting OD_{595} 0.05. OD readings were taken every hour for the first 8 hours then at 24 hours. Each culture was set up in duplicate and three independent replicates were performed. Error bars represent standard deviation.

4.4. Cobalt and B_{12} requirements in *C. difficile*

In both rich and minimal media, where cobalt and B_{12} availability was controlled, no growth impairments were detected in either strain. Therefore, we aimed to identify conditions where B_{12} was essential, for example for a particular metabolic pathway. Once established, this could be used to assess cobalt and B_{12} dependency and any impact of inactivating *CDR20291_0330*. Previous work has identified genes in *C. difficile* that are predicted to encode for B_{12}-dependent enzymes [483, 500], including MetH and EutBC. These genes in 630 and R20291 were assessed for co-localisation with a B_{12} riboswitch or B_{12}/cobalt transporter and presence of a B_{12}-independent isomer of the enzyme, which bacteria can utilise during B_{12} deficiency and are often coupled to B_{12} riboswitches [483, 492]. Genes involved in the synthesis of B_{12} were excluded.

The putative enzyme required for the final stage of methionine synthesis, MetH (*CD3596/CDR20291_3434*) is not co-localised with a B_{12} transporter or riboswitch. However, *C. difficile* also encodes the B_{12}-independent isomer of this enzyme- MetE (*CD0130/CDR20291_0129*), which we hypothesised would compensate for the lack of B_{12} if investigated in growth assays, resulting in no observable phenotype. The EutBC (*CD1913-1914/CDR20291_1834-1835*) enzyme is also not found with a B_{12} riboswitch or uptake system but no B_{12}-independent isomer has been identified in *C. difficile* or other species [483]. The lack of a B_{12} independent isomer as well information on EutBC's requirement for B12 in other species made this an interesting candidate for further work. EutBC encodes ethanolamine ammonia lyase, the enzyme responsible for breakdown of ethanolamine which can be used as a carbon and nitrogen source [501]. Ethanolamine utilisation was therefore selected for further work using investigations in minimal media growth assays, which can be manipulated to generate the desired conditions [32, 431].

4.4.1. Ethanolamine Utilisation

Phosphotidylethanolamine is the principal bacterial phospholipid and a key component of the cell membrane. Many gut bacteria are capable of metabolising phosphotidylethanolamine, which releases

ethanolamine into the intestinal environment. Consequently, a number of enteric bacteria use ethanolamine as a carbon source, including *Salmonella* and *Enterococcus* [502].

C. difficile harbours a 20 gene locus predicted to encode the ethanolamine utilisation pathway [36, 111, 502] (Figure 4.7), which is upregulated in the presence of ethanolamine [36]. *C. difficile* 630 can use ethanolamine as a carbon source *in vitro*, although current evidence is less conclusive for R20291. In *Enterococcus,* the *eut* operon is positively regulated following detection of ethanolamine by a two- [503, 504] EutW, which have also been identified within the *C. difficile eut* operon- CD1910 and CD1911, respectively [36].

The central enzyme for ethanolamine breakdown is the ethanolamine ammonia lyase (EutBC), which is B_{12}-dependent. The *eut* locus is therefore often found downstream of a B_{12} riboswitch, but no putative riboswitch could be identified in *C. difficile* in this study or others [111]. The locus encodes a number of other genes, which includes those that form the bacterial microcompartment, within which ethanolamine is utilised [505].

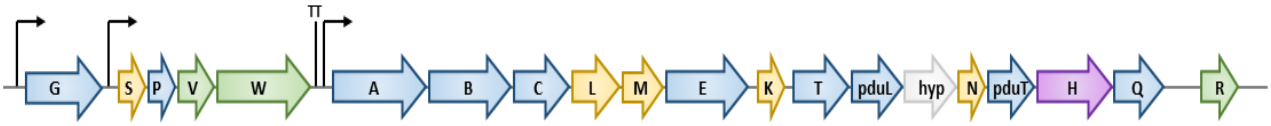


Figure 4.7. **The ethanolamine utilisation (*eut*) locus in *C. difficile***. Schematic of the putative ethanolamine utilisation locus in *C. difficile* based on annotation and diagram by Nawrocki *et al*. [36] (2018 Society for Applied Microbiology and John Wiley & Sons Ltd). There are three predicted promoters, upstream of *eutG*, *eutS*, and *eutA* and a transcriptional terminator following *eutW*. Genes encoded within the locus vary in their predicted function and include; metabolic enzymes (blue), regulatory proteins (green), bacterial microcompartment units (yellow), ethanolamine transporter (purple) and a protein of unknown function (white).

Due to the requirement for cobalt in B_{12} synthesis and the association between B_{12} and ethanolamine utilisation, this was selected as a suitable assay for screening cobalt/B_{12} dependent conditions in *C. difficile*. We hypothesised that absence of B_{12} would result in *C. difficile* being unable to utilise ethanolamine. Consequently, it was predicted that ethanolamine utilisation would be abolished for R20291Δ*0330* if B_{12} was not provided, as this strain is hypothesised to not import cobalt into the cell,

preventing B_{12} synthesis. To begin with, growth kinetics were performed using R20291 and R20291Δ*0330* with either glucose, ethanolamine or water as a no carbon control. Although no difference in OD between ethanolamine and water was identified by Nawrocki *et al.* when investigating R20291, the number of colony forming units (CFUs) was higher in ethanolamine, therefore both were measured here [36]. Cultures were grown overnight in BHIS broth, then diluted 1:10 into fresh BHIS broth and grown until OD_{595} 0.6. These were then diluted to OD_{595} 0.1 in minimal media matching the final growth conditions and diluted 1:10 into the test flasks to obtain a starting OD_{595} 0.01.

Unlike Nawrocki *et al.* who observed differences in growth between glucose and ethanolamine within the first 12 hours [36], our study observed limited growth and therefore the total growth time was extended up to 72 hours, taking readings at 16, 20, 24, 40, 44, 48, 64, 68 and 72 hours. This enabled determination as to whether the time delay was due to *C. difficile* adapting to the change between rich and minimal media and/or using other preferred carbon sources before switching to the ethanolamine-dependent pathway. CFUs were taken at 16, 44 and 72 hours.

Both R20291 and R20291Δ*0330* had grown in glucose to a comparable level after 16 hours, but there was no increased growth in ethanolamine compared to the water only control at any time point (Figure 5.8). After 72 hours, the vegetative cell count was significantly higher in ethanolamine compared to the no carbon control for R20291, with the same comparison not significant in R20291Δ*0330*. However, in both strains the total number of CFUs appeared to be either steady or decreasing in ethanolamine and no carbon across the three time points, suggesting this may be a result of a faster reduction in viable *C. difficile* in the no carbon compared to ethanolamine, rather than growth in ethanolamine. Due to the lack of growth of the wild-type R20291 strain in ethanolamine, it was not possible to answer our hypothesis.

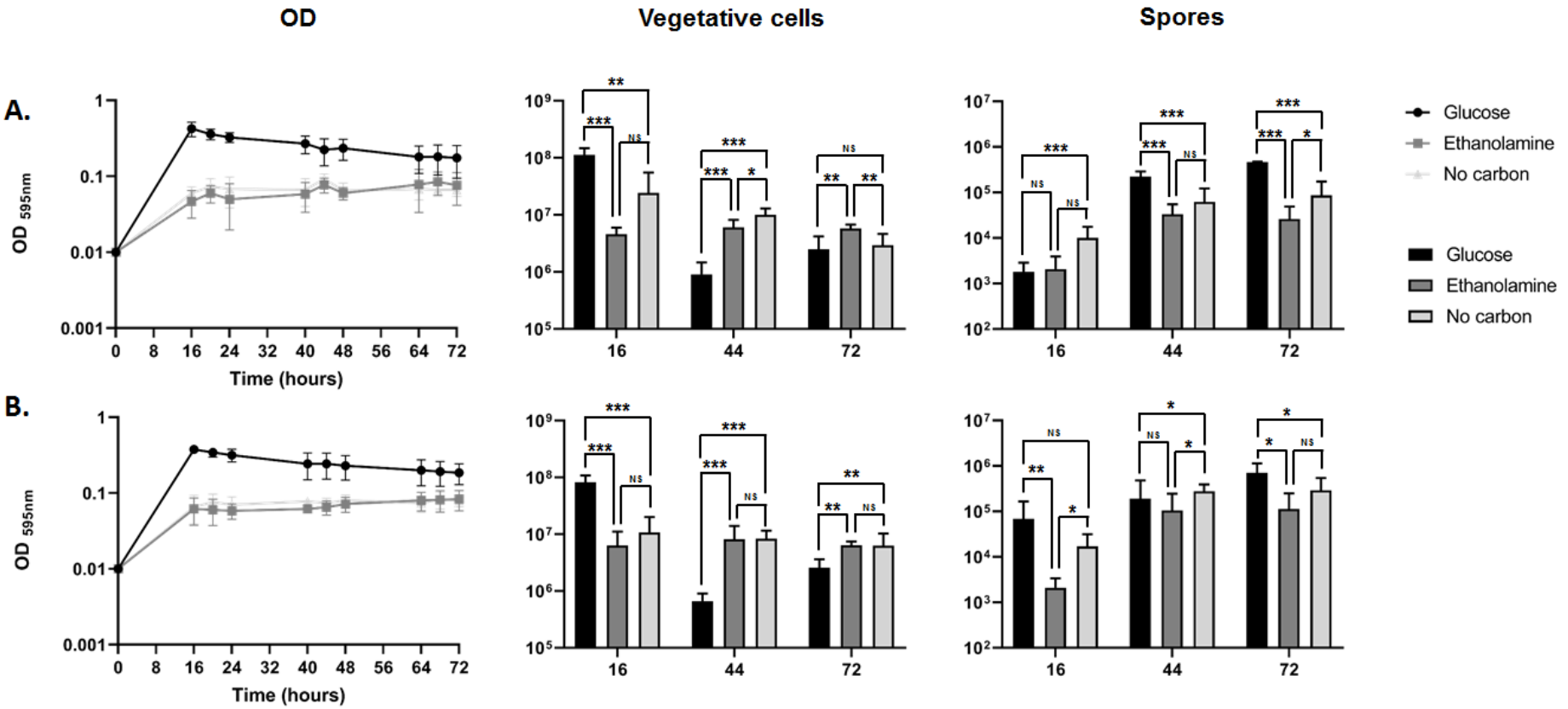


Figure 4.8. Growth kinetics of R20291 and R20291Δ*0330* in ethanolamine. Strains were grown to OD 0.6 in rich media before inoculation into minimal media containing glucose, ethanolamine or water (no carbon control). OD_{595} readings were taken at 16, 20, 24, 40, 44, 48, 64, 68 and 72 hours. Colony forming units for the total cell and spore counts were also taken at 16, 44 and 72 hours. Spore counts were incubated at 65°C for 20 mins before plating to remove vegetative cells then vegetative cell count was calculated by subtracting the spore count from the total cell count. (A) R20291 and (B) R20291Δ*0330*. Differences in CFU counts between strains was analysed for significance using a linear regression. * $p>0.05$, $p>0.005$, $p>0.0005$, NS- not significant. Each culture was set up in duplicate and three independent replicates were performed. Error bars represent standard deviation.

The same assay was repeated using 630Δ*erm*, 630Δ*erm_1910*::CT and 630Δ*erm_1911*::CT. The latter strains harbour ClosTron insertions within EutV and EutW, respectively, which form the ethanolamine sensor (these strains were constructed by Dr Alexandra Faulds-Pain). Firstly, the growth of all strains was compared in BHI and found to be comparable, demonstrating these mutations do not impair growth in rich media (Figure 4.9).

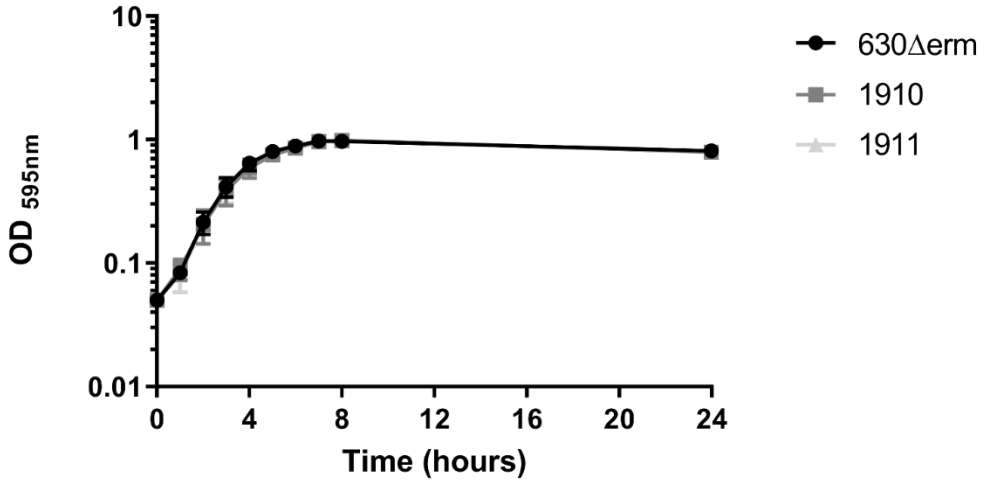


Figure 4.9. Growth kinetics of 630Δ*erm*, 630Δ*erm_1910*::CT and 630Δ*erm_1911*::CT in BHIS. All strains were grown overnight in BHI then diluted into fresh BHI media to a starting OD_{595} 0.05. OD readings were taken every hour for the first 8 hours then at 24 hours. Each culture was set up in duplicate and three independent replicates were performed. Error bars represent standard deviation.

In the minimal media (Figure 4.10), all strains grew in glucose after 16 hours. There was no significant difference in total cell counts between strains in the glucose or no carbon conditions at any time point (Figure 4.11). At 16, 20 and 24 hours, growth in ethanolamine was comparable to that in no carbon for all strains. Following this, the OD and vegetative cell count slowly decreased until reaching a plateau, in glucose and the no carbon control.

Interestingly, by 44 hours, the spore counts in ethanolamine were significantly higher than the no carbon control for 630Δ*erm* and 630Δ*erm_1910*::CT. By 72 hours, both the vegetative cell and spore count were significantly higher in ethanolamine compared to no carbon and this difference was also visible by OD, suggesting both of these strains were utilising ethanolamine. There was no significant

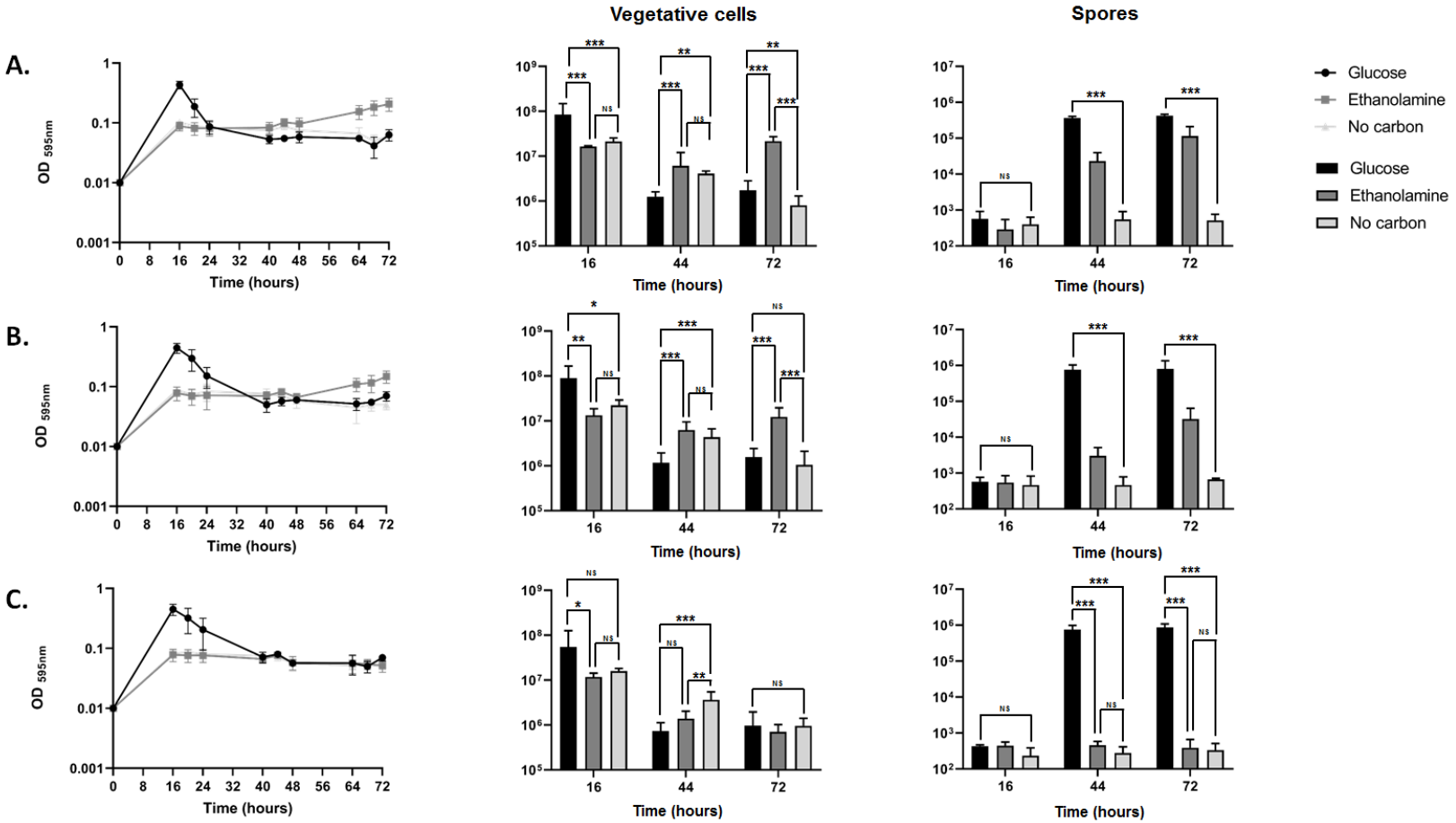


Figure 4.10. Growth kinetics of 630Δ*erm*, 630Δ*erm_1910*::CT and 630Δ*erm_1911*::CT in ethanolamine. Strains were grown to OD 0.6 in rich media before inoculation into minimal media containing glucose, ethanolamine or water (no carbon control). OD_{595} readings were taken at 16, 20, 24, 40, 44, 48, 64, 68 and 72 hours. Colony forming units for the total cell and spore counts were also taken at 16, 44 and 72 hours. Spore counts were incubated at 65°C for 20 mins before plating to remove vegetative cells then vegetative cell count was calculated by subtracting the spore count from the total cell count. (A) 630Δ*erm*, (B) 630Δ*erm_1910*::CT and (C) 630Δ*erm_1911*::CT. Differences in CFU counts between strains was analysed for significance using a linear regression. * $p>0.05$, $p>0.005$, $p>0.0005$, NS- not significant. Each culture was set up in duplicate and three independent replicates were performed. Error bars represent standard deviation.

increase in OD, vegetative cell or spore count for 630Δ*erm_1911*::CT in ethanolamine compared to no carbon. At 44 and 72 hours, growth in ethanolamine was significantly higher in 630Δ*erm* and 630Δ*erm-_1910*::CT compared to 630Δ*erm_1911*::CT (Figure 4.11). Collectively, this demonstrates that 630Δ*erm* can use ethanolamine as an energy source for growth and inactivation of 630Δ*erm_1911*::CT only, abolishes this phenotype.

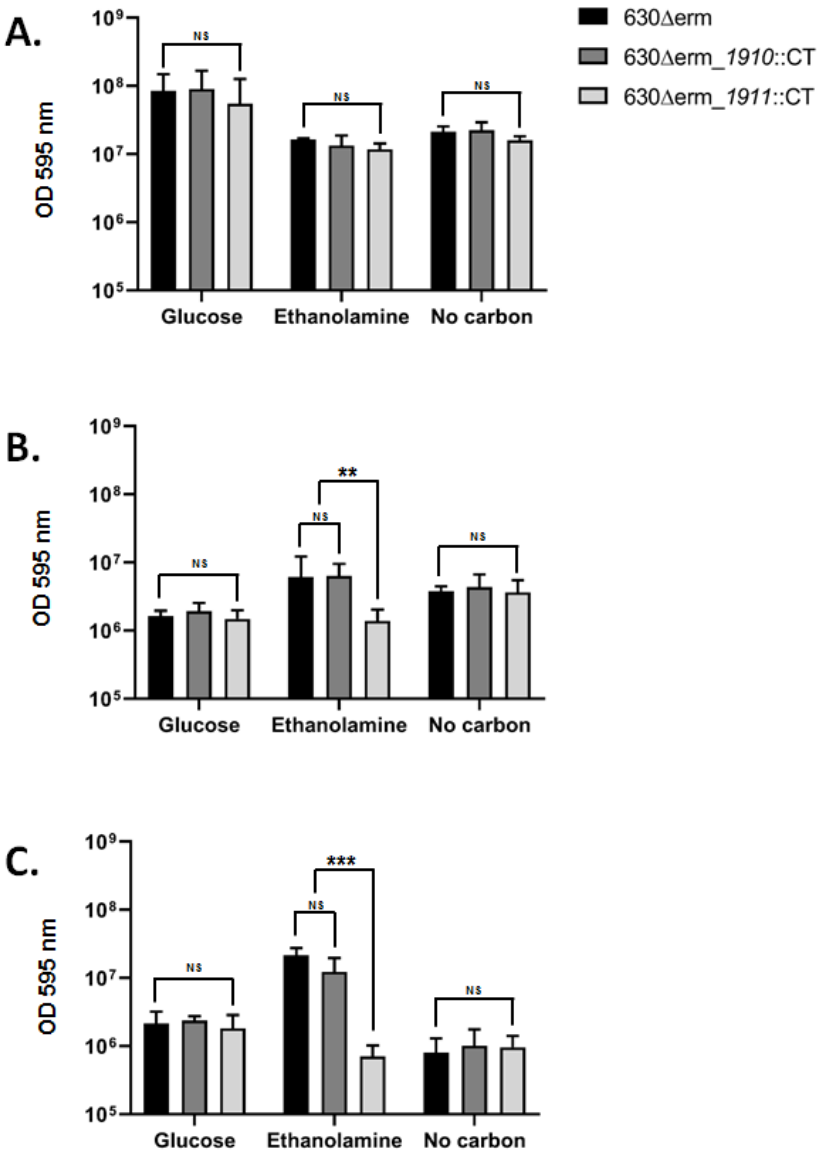


Figure 4.11. Colony forming units of 630Δ*erm*, 630Δ*erm_1910*::CT and 630Δ*erm_1911*::CT when grown in glucose, ethanolamine and no carbon. Total cell counts of 630Δ*erm*, 630Δ*erm_1910*::CT and 630Δ*erm_1911*::CT when grown in glucose, ethanolamine or water (no carbon) at (A) 16 hours, (B) 44 hours, (C) 72 hours. Differences in CFU counts between strains was analysed for significance using a linear regression. * $p>0.05$, $p>0.005$, $p>0.0005$, NS- not significant. Each culture was set up in duplicate and three independent replicates were performed. Error bars represent standard deviation.

4.4.1.1. Cobalt requirements in ethanolamine utilisation

To evaluate the B_{12} requirements of ethanolamine utilisation, the 630Δ*erm* growth kinetics were repeated, with and without cobalt. As cobalt is only required at trace level by the cell, it was important to reduce the chance of carryover of cobalt already present within the cell or rich media. Therefore, the assay was performed as before, but for the steps using rich media, this was replaced with minimal media supplemented with glucose but cobalt was excluded.

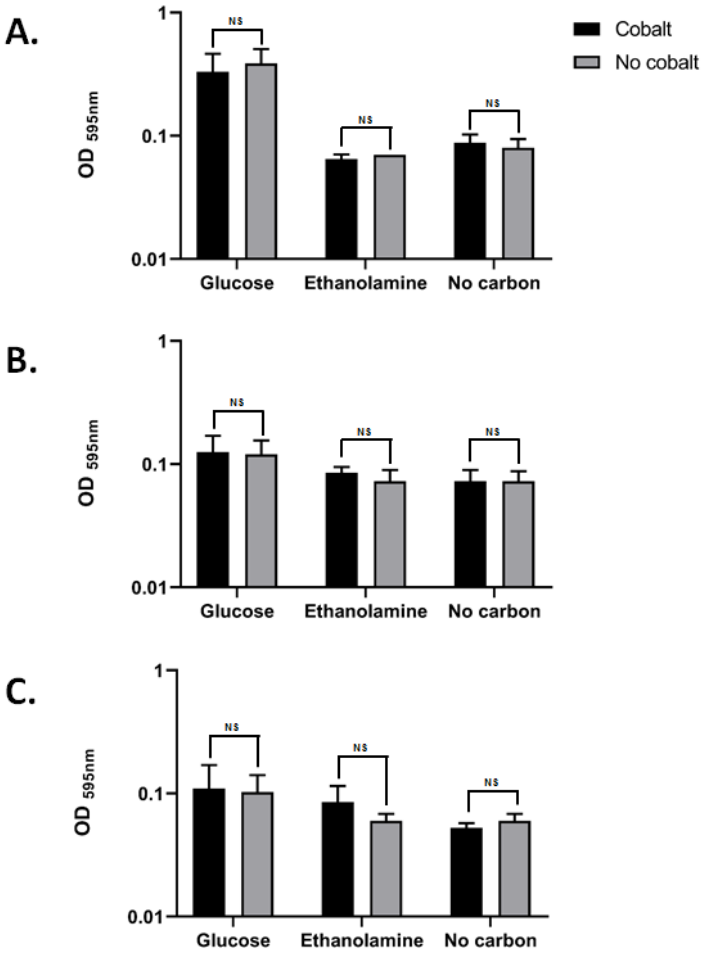


Figure 4.12. Growth kinetics of 630Δ*erm* in glucose, ethanolamine or water, with and without cobalt. 630Δ*erm* grown overnight in minimal media with glucose but no cobalt was subbed into fresh media and grown to OD 0.6 before inoculation into test minimal media containing glucose, ethanolamine or water (no carbon control), with and without cobalt. OD_{595} readings were taken at 16, 20, 24, 40, 44, 48, 64, 68 and 72 hours. (A) 16 hours (B) 44 hours and (C) 72 hours. Differences in CFU counts between strains was analysed for significance using a Student's t-test, NS- not significant. Each culture was set up in duplicate and two independent replicates were performed. Error bars represent standard deviation.

We hypothesised, that in the absence of cobalt and B_{12}, *C. difficile* would not be able to utilise ethanolamine. However, there was no difference in growth in glucose between the plus and minus cobalt cultures. Interestingly, even by 72 hours, there was no growth in ethanolamine for strains growing with or without cobalt. This suggests that minimal carryover from the rich media in the previous experiments may have provided an essential factor that supported ethanolamine utilisation. Overall, due to the lack of growth in ethanolamine in both conditions, it was not possible to determine the cobalt and consequently B_{12} requirements.

4.5. Discussion

This chapter aimed to investigate the role of CDR20291_0330, a putative cobalt transport protein (CbiN) [483], from the CbiMNQO ECF transporter, which mediates high affinity uptake of cobalt [490, 491]. In pursuit of this, *CDR20291_0330* was deleted from the chromosome using allele exchange mutagenesis, which enables removal of the target with minimal disruption to neighbouring genes [439]. It was also aimed to investigate cobalt and B_{12} requirements in *C. difficile*, an understudied area especially considering the importance of B_{12} in many cellular processes. *C. difficile* encodes a complete anaerobic B_{12} biosynthesis pathway, for which cobalt is an essential component [111, 481, 482]. Transcriptomics data suggests genes within the *cbiMNQO* operon are involved in the central metabolism pathways in *C. difficile* as they are downregulated in a Spo0A mutant (*cbiM* and *cbiO*) and under predicted negative regulation by CodY (*cbiM*), two major regulators of *C. difficile* [506, 507].

In both rich and minimal media, no growth defect was observed in R20291 wild-type or R20291Δ*0330* when cobalt and/or B_{12} was removed, suggesting in these conditions it is non-essential. However, *C. difficile* encodes putative mechanisms for both B_{12} uptake and *de novo* synthesis, the latter of which constitutes a large locus that is likely to pose a metabolic burden so maintenance of this suggests B_{12} is required by the bacteria [111, 483, 484]. To further examine this, the expression profiles of these transporters and synthesis enzymes could be monitored to determine when they are required and under what conditions. For those genes downstream of a B_{12} riboswitch, including the *cbiMNQO* operon and *cbiP* within the B_{12} synthesis pathway [111], these are likely to be regulated at the translational level, meaning transcriptomics would be less informative. Instead, a translational fusion could be employed, such as a chromosomal SNAP-tag or LOV domain, which will be translated with the gene product and provide a detectable marker [508-510]. Upon identification of expression conditions, spent growth media from *C. difficile* which lacked exogenous $B_{12,}$ but was cultured with and without cobalt could be assessed for the presence of *C. difficile*-synthesised B_{12}, using high performance liquid chromatography or by using the growth of *Lactobacillus leichmannii* in the spent

culture media which is auxotrophic for B_{12} [511]. When testing R20291Δ*0330* in these conditions, it would be necessary to consider the activity of other putative cobalt transporters (Table 5.2), but this would only be relevant once a cobalt/B_{12} dependent condition had been defined. Alternatively, to investigate the activity of CDR20291_0330 more directly, ligand binding assays could be performed, using fluorescence spectrometry to measure binding affinity of the purified protein with cobalt and other such as nickel or iron [512].

With an aim to provide a growth environment where cobalt/B_{12} is essential, ethanolamine utilisation was assayed. Ethanolamine breakdown is mediated by the B_{12}-dependent ethanolamine ammonia lyase (EutBC) which is encoded within the putative ethanolamine utilisation operon (*eut)* in *C. difficile*. There are only two publications previously dedicated to ethanolamine utilisation in *C. difficile*. The first characterises the ethanolamine microcompartment within which it is hypothesised it is metabolised whereas the second focuses on growth of *C. difficile* in ethanolamine, the genes involved, and how this impacts virulence [36, 505]. Therefore, the latter publication was used as a starting point for our own growth assays.

Nawrocki *et al.* demonstrated a biphasic growth curve for strain 630Δ*erm* in minimal media, which they attributed to the differences in carbon utilisation- amino acid metabolism in the first instance followed by an additional growth phase relying on glucose or ethanolamine after ~ 4 and 10 hours, respectively [36]. No substantial growth was obtained in any condition within the standard 8 hour time period (data not shown) therefore the growth kinetics were extended with a final reading at 72 hours. There was a ten-fold increase in OD (0.01 to 0.1) in the first 16 hours of growth in the no carbon control, which is likely a combination of carry-over from rich media and amino acid metabolism. By 16 hours, the OD for growth in glucose was comparable to the maximum absorbency readings obtained in the previous publication.

Regarding ethanolamine, growth was not detected by OD in 630Δ*erm* until the 64-hour time point. Although this lag time is substantial, diauxic growth is not unusual in bacteria and can be related to

preferential metabolism of more available energy sources [513, 514]. Therefore, *C. difficile* may be depleting the different substrates available before switching to the ethanolamine-dependent lifestyle. This response is reflected by *in vivo* transcriptomic data, where increased expression of the *eut* operon, encoding ethanolamine utilisation, is not reported until 38 hours post infection, whereas glucose utilisation is believed to occur at ~14 hours [35]. The differences between our data and the previous study are surprising, as the assay set up was based on their published methods. It may be that the cas-amino acid composition used here was richer and took longer to deplete, which delayed the switch to ethanolamine utilisation. Alternatively, it may be an outcome of differences in ethanolamine, as our study used liquid ethanolamine compared to ethanolamine hydrochloride in the previous study, which may be easier to metabolise.

Growth of R20291 was not detected in ethanolamine by OD or CFU, therefore it was not possible to determine the impact of deleting *CDR20291_0330* in this assay. Nawrocki *et al.* found the increased expression of the *eut* operon from basal level in the presence of ethanolamine was more pronounced in 630 compared to R20291 [36]. Therefore in this study, the *eutVW* genes encoding a two-component response regulator for detection of ethanolamine were compared for 630 and R20291, to see whether sequence differences could be impairing detection of ethanolamine, but the sequences were highly conserved across the two strains (data not shown).

No ethanolamine utilisation was observed when 630Δ*erm* was sub-cultured in minimal as opposed to rich media before inoculation of the test flasks, which prevented analysis of cobalt requirements in this process. This was interesting, as it suggests lack of a component required for ethanolamine utilisation that is carried over in sufficient amounts when inoculating from rich media. This could be further investigated by comparing the minimal media at inoculation, 44 and 72 hours between the rich and minimal inoculants using nuclear magnetic resonance, to compare the metabolite profile, including B_{12}. B_{12} could also be spiked into the media to try and promote ethanolamine utilisation.

Two distinct mechanisms of regulating ethanolamine utilisation have been identified. Enterobacteriaceae, such as *Salmonella Typhimurium*, encode EutR, a constitutively expressed positive transcriptional regulator. This upregulates expression of the entire *eut* operon by binding a single promoter, but only in the presence of two ligands- ethanolamine and B_{12} [515]. Firmicutes (like *C. difficile*) lack the EutR regulator and instead encode the EutVW two-component regulatory system, as demonstrated in *Enterococcus faecalis* and *Listeria monocytogenes*, which encode very similar *eut* operons to *C. difficile* [503, 504, 516]. In the presence of ethanolamine, EutW autophosphorylates then goes onto phosphorylate the response regulator EutV. EutV binds RNA to interrupt transcriptional terminators and enable expression of the operon (Figure 4.13) [503, 504]. Unlike EutR, B_{12} does not directly interact with EutVW, but does maintain regulation over the *eut* operon, via a riboswitch. Although B_{12} riboswitches usually result in the termination of expression upon B_{12} binding, presence of B_{12} actually facilitates expression of the *eut* operon. This is through interaction with a non-coding RNA, EutX, which binds and sequesters EutV, preventing induction of the operon. EutX harbours a B_{12} binding domain which when bound, results in transcriptional termination of EutX, producing a shorter fragment that can no longer bind EutV, allowing induction of the ethanolamine utilisation pathway (Figure 4.13) [504, 517, 518].

As no B_{12} riboswitch has been identified within the *C. difficile eut* operon, it is possible B_{12} is not required for ethanolamine utilisation in *C. difficile*. This seems unlikely, as no B_{12}-independent isomer has been identified for ethanolamine breakdown in *C. difficile* or other species, the locus still encodes EutT, which is required for processing of B_{12} into the active co-factor and very early research using Clostridial (species not specified) cell free extracts found a cobinamide co-enzyme was necessary for ethanolamine breakdown [111, 483, 492, 519].

Furthermore, the authors of a recent review into ethanolamine note how the riboswitch within the EutX RNA is dissimilar to other B_{12} riboswitches and therefore may not be identified using standard riboswitch identification techniques [504]. This also acts *in trans*, meaning it could be located

elsewhere on the genome [501]. Future work should look at identifying an ortholog of EutX in *C. difficile* in addition to monitoring expression of the *eut* operon in the presence of ethanolamine, plus and minus B_{12}.

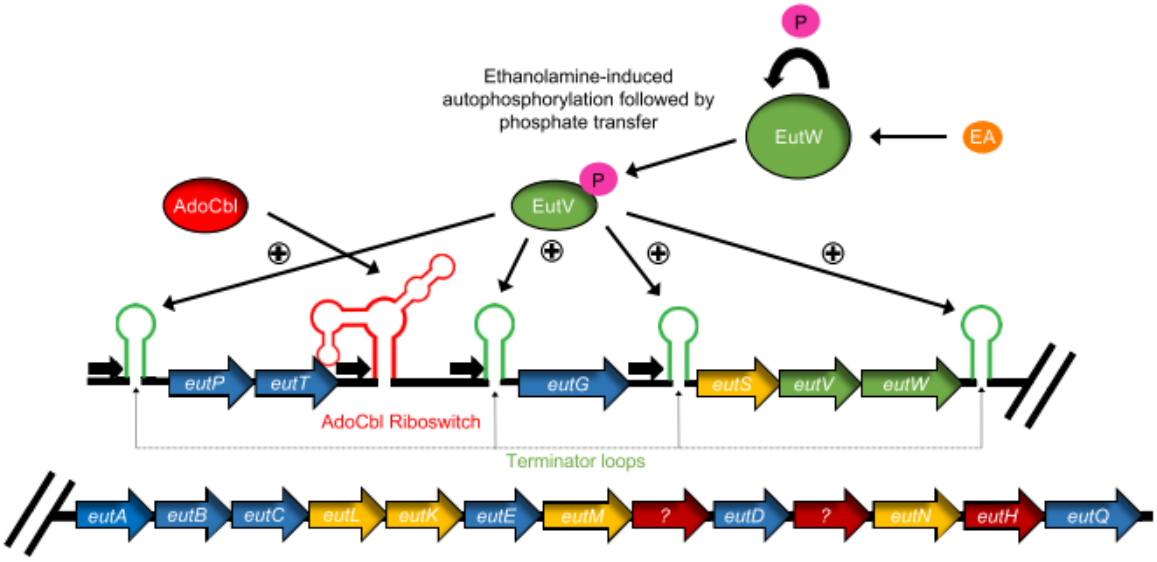


Figure 4.13. Regulation of ethanolamine utilisation in *Enterococcus faecalis*. Regulation of the *eut* operon for ethanolamine utilisation in *E. faecalis* which requires both ethanolamine and B_{12}. The histidine kinase sensor, EutW, autophosphorylates in the presence of ethanolamine and phosphorylates the response regulator, EutV. EutV has an RNA-binding domain which enables binding and disruption of transcriptional terminators, resulting in induction of the operon. However, in the absence of B_{12}, EutV is sequestered by the small RNA, EutX (not shown), and is only released upon B_{12} binding, which interrupts the EutV binding site. Image taken from Kaval and Garsin [501].

Separate to the requirement for B_{12}, inactivation of EutW (CD1911), but not EutV (CD1910) abolishes ethanolamine utilisation in *C. difficile*, suggesting EutV is dispensable in this process. El Papa and Perego noted that both EutV and EutW had RNA binding activity *in vitro*. RNA binding is the method of anti-termination used to enable expression of the *eut* operon [503]. In *E. faecalis,* inactivation of EutV abolished ethanolamine utilisation, although it is noteworthy that this is an insertional mutant and therefore may have influenced the activity of downstream genes [503]. The mutants used in our study were also generated using insertional inactivation and although not confirmed using complementation, the results suggests there were no polar effects on downstream genes. As the mutation in EutW, which is encoded downstream of EutV, displayed no growth with ethanolamine, this would also be expected in the EutV mutant, if its inactivation influenced expression of EutW. However, this is not the case, as the EutV mutant shows no difference from wild-type in ethanolamine

utilisation. Regarding EutW, Nawrocki *et al.*, demonstrated two transcriptional units in the presence of ethanolamine in *C. difficile*, one of which ends after *eutW* [36]. Therefore, the genes downstream of this are under the control of a separate promoter and unlikely to be susceptible to polar effects.

Alternatively, it may be that an orphan response regulator encoded elsewhere on the genome is able to compensate for the loss of EutV. Orphan response regulators are encoded without a histidine kinase partner and can regulate genes elsewhere on the genome. *C. difficile* encodes 51 response regulators and 5 of these are considered orphan kinases (a 6th has been identified as regulating lantibiotic resistance) [252]. To investigate whether these could be involved in ethanolamine regulation, the sequence of EutV could be compared to the sequences of these different orphan response regulators to see if they may also be involved.

Limitations to this study which should be addressed with future work include the lack of complementation for CDR20291Δ*0330,* 630Δ*erm_eutV*::CT and in particular, 630Δ*erm_eutW*::CT, as this is the mutant where a different phenotype from the wild-type was observed. For the reasons previously discussed, polar effects on neighbouring genes were not predicted to have influenced the results described here, but it is still important to reintroduce the gene into the cell to confirm that the altered phenotype is a result of loss of the target gene, only. Another limitation regarding ethanolamine utilisation is the difficulty in recreating the *in vivo* gut environment within the laboratory. This includes the challenges of culturing *C. difficile* in minimal media, in replicating the dynamic environment of the gut within the laboratory and the possibility of carryover of compounds from starting cultures. There is media available designed to mimic the gut niche which could be an improvement on standard minimal media. A recent study investigating ethanolamine utilisation in uropathogenic *E. coli* did so using an artificial urine medium supplemented with ethanolamine [520]. However, here it is important to note that urine is usually a sterile environment, whereas the microbiome in the gut is constantly using and releasing many different metabolites.

There has been significant development in the use of substrate binding proteins as vaccine candidates. Antibodies raised against the methionine binding protein MetQ, from *Neisseria gonorrhoeae* are bactericidal and prevent gonococcal binding to epithelial cells *in vitro* [521] and immunisation of mice with recombinant SBP2 from *Moraxella catarrhalis* increased clearance of infection from the lung [522]. Immunising mice with PiuA and PiaA from *Streptococcus pneumoniae* increased survival upon challenge [523]. Interestingly, this was not a result of impairing the activity of these essential iron uptake components, but was in fact due to the elicited opsonophagocytic response, demonstrating how essentiality in the host is not necessarily a requirement for a good vaccine target [524]. PiuA has been included within PnuBioVax, an anti-*S. pneumoniae* vaccine which was recently demonstrated to be safe and immunogenic in Phase I clinical trials (NCT02572635) [525]. Also in clinical trials is the manganese binding protein, MntC from *Staphylococcus aureus*, which is currently in Phase II (NCT02388165) [526]. In relation to use of CDR20291_0330 in a vaccine, additional studies are required to confirm this proteins suitability for inclusion within a vaccine. Production of antibodies against this protein would enable probing of intact *C. difficile* cells to ensure the protein is expressed on the surface, or the different fractions of the *C. difficile* cell could be probed.

5. Characterisation of CDR20291_3343, a putative pilin protein

5.1. Introduction

Motility is a key process for many bacterial species. It can facilitate movement towards nutrient sources or away from noxious agents and environmental stressors, interactions between cells to aggregate and form biofilms and is an important virulence factor, particularly for colonisation of the host [184, 189, 527, 528].

Bacteria employ a number of different mechanisms for motility. Although use and definition of the different motility terminology can vary, swimming and swarming are typically considered a product of the flagella, a whip-like surface structure that propels bacteria through their environment, via a motor rotating the flagella filament [529]. The two mechanisms are differentiated depending on if the motility is through a medium (swimming) or on the surface (swarming) [530]. Flagella-independent motility is often attributed to the action of the type IV pili (T4P), hair-like appendages on the cell surface [204]. Rather than forceful propulsion of the bacteria through or on a medium, pili drag the bacteria across a surface by continuous assembly and disassembly of the polymerised filament, in a grappling hook mechanism known as twitching motility [531-533]. Gliding motility is also used to describe T4P-mediated motility, for example in *Clostridium perfringens* [534], but is also linked to non-T4P actions such as focal adhesions or sliding due to bacterial division [530, 531].

C. difficile motility is primarily dictated by the action of the peritrichous flagella decorating the cell surface [184, 187]. These are essential for movement through semi-solid agar, as movement is abolished upon inactivation of the flagella filament gene *fliC*, or removal of the whole flagella glycan (some mutants of this are still motile) [184, 185, 187]. *C. difficile* also encodes the hypothetical major (CDR20291_3340-3350) and minor (CDR20291_3153-3158) T4P locus which are thought to be core components of its genome [111, 205, 206]. These loci are predicted to encode major and minor pilins that form the pilus filament and the accessory components required for pilus assembly or retraction. Using microscopy, predicted T4P structures have been identified on the surface of wild-type *C. difficile*

in vitro [209, 211] which have been differentiated from flagella based on size and use of a flagella mutant as well as through use of immunogold staining to PilA1 and PilJ [195, 209, 211]. Furthermore, pili like structures have been identified by microscopy in hamster infection models [210].

CDR20291_3343 is encoded within the predicted major T4P locus of *C. difficile*, and has been previously identified as a putative minor pilin protein (Figure 3.1a) [204, 207, 534]. At a predicted 57.7 KDa, CDR20291_3343 is larger than most other pilins but still possesses the conserved N-terminal hydrophobic domain and predicted pre-pilin peptidase cleavage site (Figure 3.1b) [207].

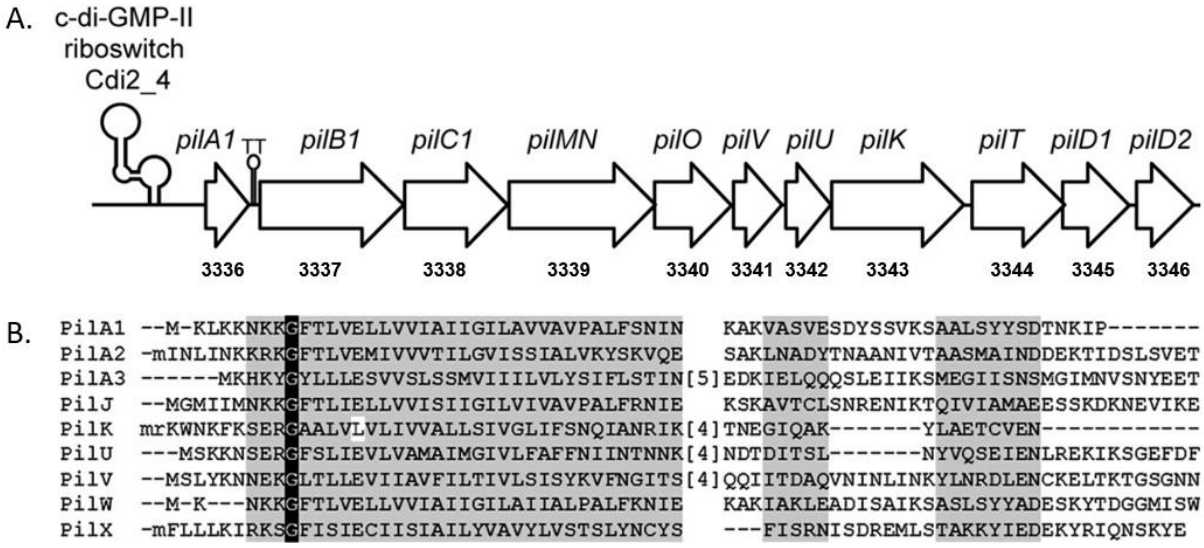


Figure 5.1. The *C. difficile* type 4 pili locus. (A) Schematic of the major type 4 pili locus in *C. difficile* strain R20291 [195]. The locus is preceded by a c-di-GMP dependent type II riboswitch (Cdi2_4) which controls its expression. Putative functions have been assigned to each gene within the locus including the major structural pilin proteins; *pilA1*, the ATPase, *pilB1* and the predicted tip of the pilus, *pilK*. (B) Comparison of the N-terminal amino acid sequences from different *C. difficile* R20291 pilin proteins [207]. Conserved regions are shaded in grey, including the N-terminal hydrophobic region. The glycine residue in black is the pre-pilin peptidase cleavage site. At position +5 after this site, PilK lacks the conserved glutamic acid residue, predicted to be required for interaction with neighbouring pilins via salt bond formation. Images from a) Bordeleau *et al.* and b) adapted from Maldarelli *et al.* [195, 207] (by permission of Oxford University Press).

Pilus assembly in *C. difficile* has not been characterised and T4P assembly is less well understood compared to Gram-negative systems. Based on the amino acid similarity to other T4P locus, it is predicted that the *C. difficile* putative T4P assembly follow the ATPase mediated route, wherein each pilin protein is actively secreted from the cell, in a tip to base manner, to form the polymerised

filament [535-537]. Pilins are suspected to interact via salt bridge formation between a conserved glutamic acid residue at +5 after the prepilin peptidase cleavage site and an N-terminal phenylalanine on a neighbouring pilin (Figure 5.1) [207]. Interestingly, this glutamic acid residue is absent in CDR20291_3343, replaced instead with a hydrophobic leucine residue, which presumably abolishes this interaction. This has resulted in CDR20291_3343 being assigned into the GspK family of proteins. GspK from the Enterotoxigenic *E. coli* type 2 secretion system (T2SS), which share amino acid similarity with T4P, is also large, lacks the conserved glutamic acid residue and forms the tip of the pseudopilin [207, 538, 539]. CDR20291_3343 has therefore been annotated as PilK, hypothesised to form the tip of the pilus. [204, 207].

As a member of the major pili locus (Figure 5.1), *pilK* expression is upregulated by the secondary messenger cyclic diguanosine monophosphate (c-di-GMP), via a positive transcriptional type II riboswitch, Cdi2_4. Riboswitches are RNA elements within the 5' untranslated region (UTR) of their target genes which, when bound by their effector, can promote or inhibit expression at the transcriptional or translational level. Cyclic-di-GMP is a central regulator of a number of key bacterial processes and in elevated concentrations is associated with promoting cell entry into the sessile lifestyle [196]. As such, elevated levels of c-di-GMP are associated with repression of the flagella and a reduction in related motility, whereas biofilm formation and pili production are upregulated [198]. Expression of *pilK* in strain 630 is higher in the presence of elevated levels of c-di-GMP and *CDR20291-_3343* expression is higher in plate and biofilm cultures compared to planktonic growth as well as in biofilm cultures [195, 212].

Previous characterisation of the role of the *C. difficile* pili has identified a number of associated phenotypes, including mediating surface motility, auto-aggregation and biofilm formation and binding to human cells lines and colonisation [195, 208, 212, 213]. These investigations have primarily been conducted in strains 630 and R20291, using wild-type strains as well as those with inactivation of PilA1, the major structural subunit forming the polymerised filament and PilB1, the ATPase driving pilus

formation. Neither of these mutant strains produce T4P, as shown by transmission electron microscopy [195].

CDR20291_3343 was found to be immunogenic in patients with CDI on a pilot study using a *C. difficile* specific pan-protein array (Chapter 3). A number of bacteria encode immunogenic pili, some of which have since been investigated as vaccine candidates [462, 463]. Therefore, this chapter aimed to further investigate the role of pili in *C. difficile,* specifically in relation to the role of CDR20291_3343 within pili-attributed motility, by comparing individual contributions of flagella and pili within *C. difficile* motility.

5.2. Bioinformatics

The amino acid sequence of CDR20291_3343 was highly conserved across a diverse selection of *C. difficile* strains (Table 5.1). The comparison tool pBLAST revealed no putative conserved domains within the amino acid sequence of this gene, but as already discussed, CDR20291_3343 harbours the conserved *N*-terminus associated with pilin proteins [207].

Strain	Ribotype	Percentage Identity (%)
R20291	027	100
CD196	Historic 027	99.8
630	012	99
M120	078	94
M68	017	97.6
CD305	023	98

Table 5.1. Conservation of CDR20291_3343 amino acid sequence across *C. difficile* strains. The conservation of the CDR20291_3343 amino acid sequence across a number of *C. difficile* strains from diverse ribotypes was compared using pBLAST searches with R20291 as the comparator sequence.

5.3. Construction of a *CDR20291_3343* gene deletion mutant in R20291 and R20291_*fliC*::CT

Previous work characterising the *C. difficile* pili-associated phenotype has been conducted in strains still harbouring their flagella [208]. Although increasing c-di-GMP levels required for pili expression should result in suppression of the flagella, there may be some residual activity, and removing the flagella altogether would aid in precise identification and be a more robust means of investigating pili and pili-associated phenotypes. To further dissect the contributions of each of these surface appendages on motility and other functions within the cell, in frame deletion mutants were constructed in this study, which lack flagella, pili or both (Table 5.2). It was hypothesised that loss of PilK could result in one of two phenotypes: either that PilK is required for pili length regulation and that a knockout would result in a hyper-extended pili without PilK present to act as a pilin cap [540]; or that PilK is essential for pilus assembly and knockout results in the loss of the pilus altogether. It was found that in a *pilK* mutant, the major structural pilin PilA1 can be detected in the cell but is no longer secreted as it is in the wild-type, suggesting that the pilus isn't assembled [541] (Neil Fairweather, personal communication). Therefore, throughout this study the PilK mutant has been treated as a pili deficient strain.

[540][541]

Strain	High level c-di-GMP		Low level c-di-GMP	
	Flagella	**Pili**	**Flagella**	**Pili**
R20291	x	✓	✓	x
R20291Δ*3343*	x	x	✓	x
R20291_*fliC*::CT	x	✓	x	x
R20291Δ*3343*_*fliC*::CT	x	x	x	x

Table 5.2. Predicted phenotype of flagella and pili inactivated strains, in the presence and absence of c-di-GMP. The influence of high or low-level c-di-GMP on expression of flagella and pili in four strains; wild-type R20291, R20291Δ*3343* where no pili are expected to be assembled, R20291_*fliC*::CT where no flagella are produced and R20291Δ*3343*_*fliC*::CT, lacking both pili and flagella.

To generate the desired strains, an allele exchange cassette was built targeting *CDR20291_3343* (Figure 5.2), to delete this gene from R20291 wild-type and R20291_*fliC*::CT. The latter strain does not produce flagella due to a ClosTron insertion within the flagellin gene, *fliC* [187]. The generation of the allele exchange mutants was performed as described in 4.3.2, including construction of the 3870 bp allele exchange cassette. Following transformation of R20291 and passage on selective and non-selective media, the homology region from thiamphenicol sensitive colonies was amplified to screen for the loss of the 1539 bp region of *CDR20291_3343* from both R20291 and R20291_*fliC*::CT (Figure 5.2). It was not possible to amplify the entire homology region (2144 bp), in the potential mutants or wild-type control gDNA from R20291, therefore two products were generated instead, using one primer outside of the homology region (R20291_3343_MUTSEQ_f or R20291_3343_MUTSEQ_r) and a reverse primer adjacent to the deletion of *CDR20291_3343* (R20291_3343_MUT_f and R20291_3343_MUT_r). Potential double crossovers and therefore gene deletion mutants were identified for both backgrounds and the PCR products were sequenced to confirm *CDR20291_3343* had been successfully deleted from the chromosome and no single nucleotide polymorphisms had been accrued in the amplified region in the process.

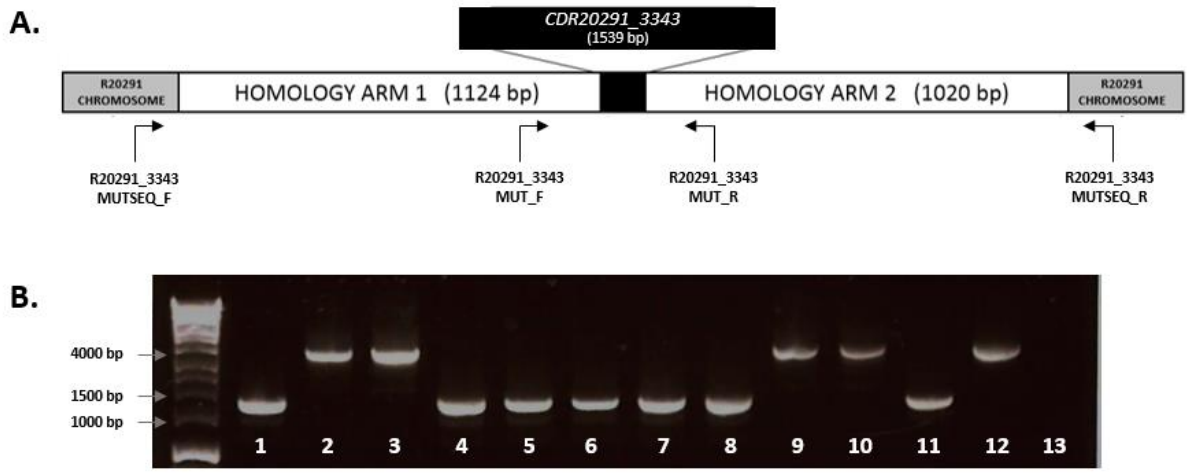


Figure 5.2. Allele exchange mutagenesis of *CDR20291_3343* in R20291 and R20291_*fliC*::CT. (A) Schematic of the allele exchange construct for generation of the *CDR20291_3343* gene deletion mutants. This is annotated with the primers used for screening to detect gene deletion mutants. (B) PCR screen of potential *CDR20291_3343* gene deletion mutants using primers R20291_3343_MUTSEQ_f and R20291_3343_MUT_r. 1 to 6- individual clones from R20291, 7-11- individual clones from R20291_*fliC*::CT, 12- wild-type R20291 gDNA, 13-dH_2O control. The predicted size of amplicons with and without *CDR20291_3343* are 3683 bp and 2184 bp, respectively. Clones from lanes 2, 3, 9 and 10 clones are predicted wild-type revertants with the remaining representing gene deletion mutants, which were confirmed with DNA sequencing.

5.4. Selection of an expression system for c-di-GMP

Pili expression is upregulated in the presence of c-di-GMP, therefore it is necessary to elevate the cellular levels of this secondary messenger for study. This was achieved using the *dccA* gene, which encodes a cyclic diguanylase, the enzyme responsible for synthesis of c-di-GMP. Although 15 genes encoding cyclic diguanylases have been predicted in *C. difficile,* previous work has demonstrated that expression of the *dccA* gene is sufficient to achieve high intracellular levels of c-di-GMP [198, 542]. Two plasmids have previously been constructed which encode *dccA* under control of either a constitutive P_{cwp2} promoter (pECC12) or an anhydrotetracycline (Atc) inducible promoter, P_{tet} (pECC17). These were a kind gift from the Fairweather Laboratory [221]. Initially, the pECC12 plasmid for constitutive expression of c-di-GMP was preferentially used over the anhydrotetracycline (Atc) inducible promoter, as motility assays take a number of days and the ability to maintain induction for that length of time, was unknown.

5.4.1. Swimming motility with constitutive expression of c-di-GMP

In the presence of elevated levels of c-di-GMP, the flagella filament gene, *fliC*, is repressed, due to interaction of c-di-GMP with the type I riboswitch upstream of *fliC* [198]. As the flagella are the driving force behind swimming motility, reduced expression of *fliC* results in loss of this phenotype [198]. To confirm expression of constitutive *dccA* resulted in the anticipated c-di-GMP induced effects on motility, assays were performed to monitor the swimming capabilities of different strains as a marker of flagella expression. Colonies taken from agar plates were inoculated into 0.3% minimal media agar using sterile cocktail sticks and incubated for 5 days (Figure 5.3).

As previously reported, the flagella-negative strains, which harbour a ClosTron insertion within *fliC* did not move beyond the inoculation (stab) point, whereas R20291 and R20291Δ*3343* demonstrated a halo swimming pattern. Surprisingly, constitutive c-di-GMP expression in R20291 and R20291Δ*3343* did not result in a reduction in swimming motility- in fact, both strains appeared to swim more when c-di-GMP was present. This experiment was terminated early due to this finding, as it was not in keeping with a reduction in expression of *fliC* that should be observed in the presence of c-di-GMP [198]. The swimming observed may have been due to condensation on the plates, or due to an unusual phenotype of the constitutively expressed c-di-GMP, which leads to artificially high levels of the messenger. Overall, it is clear these strains are still swimming in the presence of c-di-GMP, suggesting pECC12 is unsuitable for use.

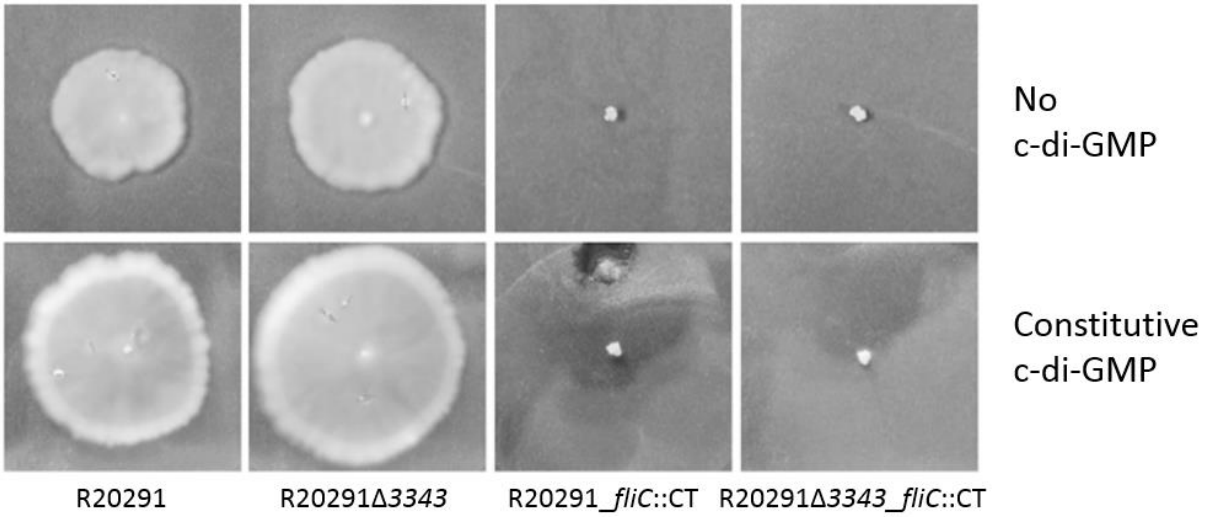


Figure 5.3. Swimming motility with and without constitutive expression of c-di-GMP. Individual colonies were taken from plate cultures and stab inoculated into 0.3% minimal media agar plates then incubated at 37°C, anaerobically. Strains either contained no plasmid or pECC12, for constitutive expression of the *dccA* gene, leading to high levels of c-di-GMP within the cell. Each culture was set up in duplicate and the assay performed once.

5.4.2. Inducible control of c-di-GMP mediated motility

To replace constitutive expression of c-di-GMP, the pECC17 plasmid was utilised, which encodes the *dccA* gene under control of the Atc-inducible promoter, P_{tet}. Swimming motility was assessed in R20291 carrying pECC17 or the empty vector, pASF085, in 0.3% minimal media agar plates (Figure 5.4) with a range of Atc concentrations to test induction of c-di-GMP at; 0, 25, 50, 100 or 250 ng/ml and 15 µg/ml thiamphenicol (for plasmid maintenance). It was hypothesised that increasing the level of induction would result in a reduction in swimming motility in the wild-type strains due to an increase in the cellular level of c-di-GMP. Plates were incubated for 5 days, and the diameter measured at 5 days to determine whether sufficient Atc would still be present to induce c-di-GMP expression, as measured by a reduction in swimming, compared to the empty vector control (pASF085).

When carrying the empty vector (pASF085), there was no significant difference in swimming motility when comparing +/- Atc, until the Atc concentration reached 250 ng/ml (Figure 5.4). Here, swimming was significantly lower compared to in the absence of Atc (p<0.05), suggesting this concentration is toxic to the cell, or has a deleterious effect on growth. Swimming motility was significantly reduced

in the presence of c-di-GMP, even in the lowest concentrations of Atc induction (25 ng/ml) (p<0.05), demonstrating flagella repression in these conditions. As hypothesised, swimming was significantly reduced in the presence of c-di-GMP (p<0.05) (Figure 5.4), and was almost completely abolished in concentrations of 50 ng/ml Atc. Further experiments were undertaken in a concentration of 25 ng/ml to minimise the deleterious effect that Atc can have on growth at high concentrations (Figure 5.4).

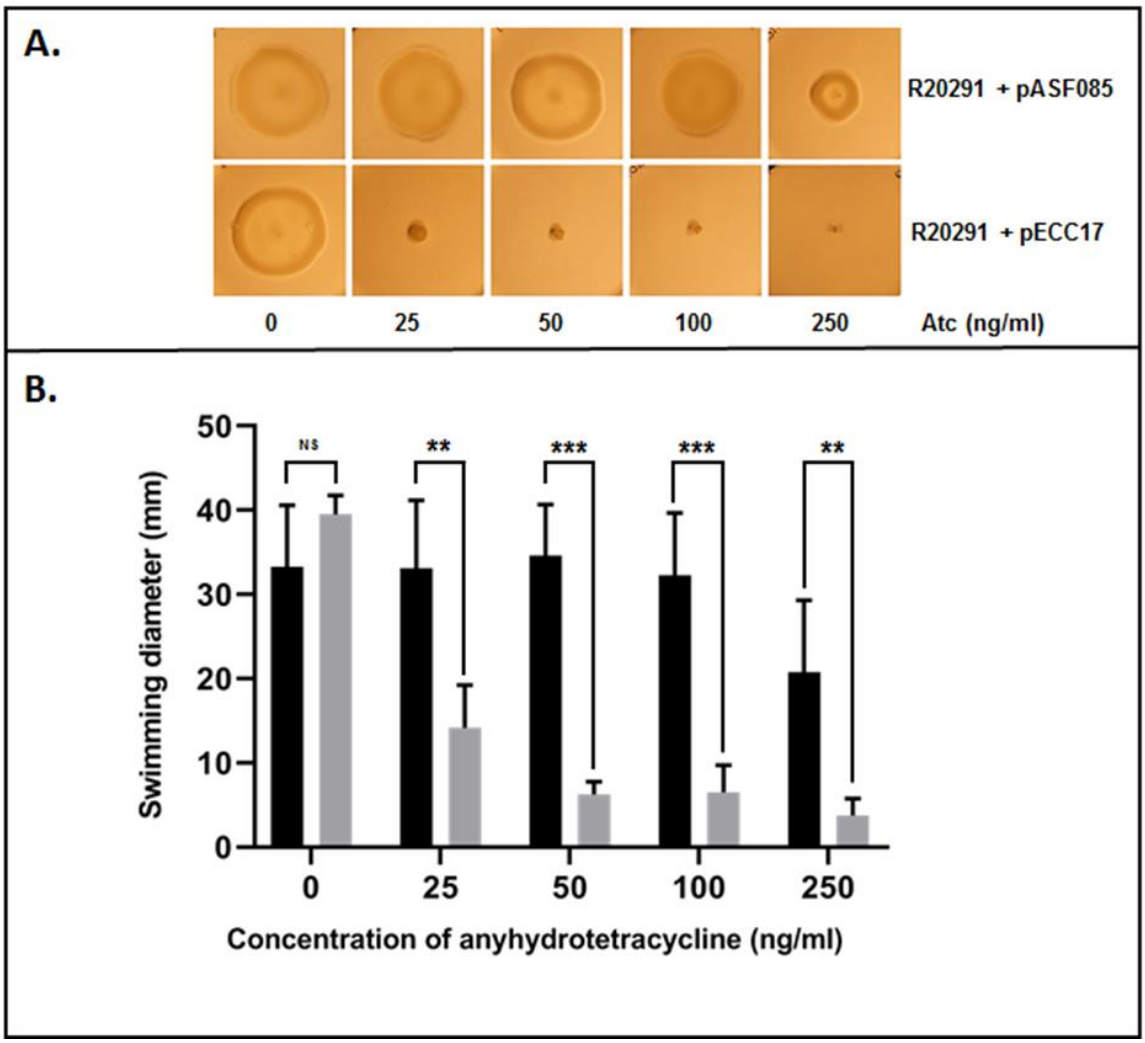


Figure 5.4. Swimming motility with increasing concentrations of c-di-GMP. Individual colonies of *C. difficile* were used to inoculate the centre of 0.3% minimal media agar plates, supplemented with 0, 25, 50, 100 or 250 ng/ml Atc and 15 µg/ml thiamphenicol. R20291 either carried the empty vector pASF085 or pECC17, for inducible expression of c-di-GMP. Plates were photographed after 5 days incubation and two perpendicular measurements were taken of the diameter. A linear regression was performed to assess any differences in swimming motility when c-di-GMP is induced, for each concentration of Atc compared to the empty vector control. (A) Photo of swimming motility of *C. difficile* R20291 after 5 days incubation, (B) Average diameter of each motility halo after 5 days growth. NS not significant, * p<0.01, ** p<0.001, *** p<0.0001. Each culture was set up in duplicate with three independent replicates performed. Error bars represent standard deviation.

5.4.3. Growth kinetics with Atc and c-di-GMP

The swimming defect of R20291 in the presence of 250 ng/ml Atc, even without induction of c-di-GMP, warranted growth assessment of R20291 and R20291_*fliC*::CT in the presence of 25 ng/ml Atc, to ensure this concentration of Atc does not have deleterious effects on the cell. This was performed using R20291 and R20291_*fliC*::CT harbouring either pASF085 or pECC17, to also test if c-di-GMP expression affected the growth rate (Figure 5.5). Strains were grown in BHIS with either 0 or 25 ng/ml Atc and 15 µg/ml thiamphenicol and OD_{595} readings taken every hour for the first 8 hours then after 24 hours of growth.

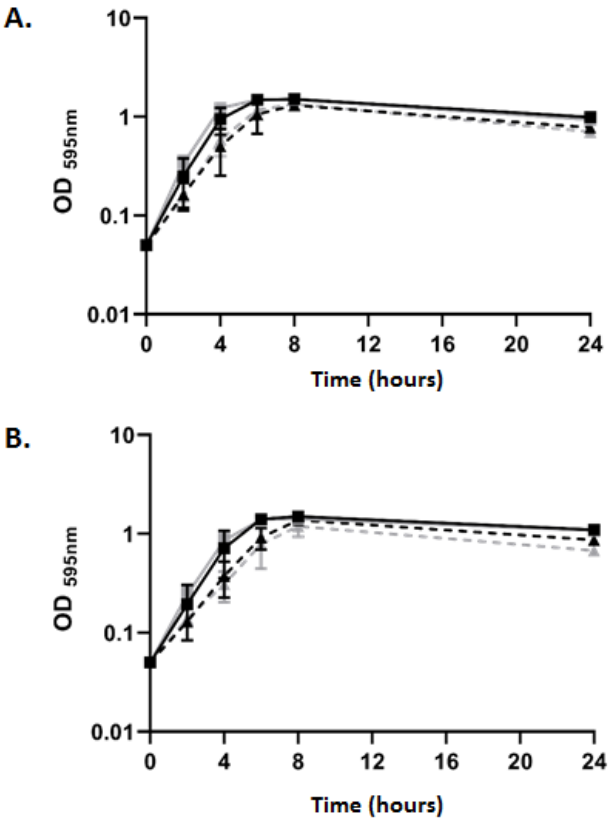


Figure 5.5. Growth kinetics with anhydrotetracycline inducible c-di-GMP. R20291 and R20291_*fliC*::CT harbouring either pASF085 or pECC17 were grown overnight in BHIS then diluted to a starting OD_{595} 0.05 in fresh BHIS supplemented with 0 or 25 ng/ml Atc and 15 µg/ml thiamphenicol. OD_{595} readings were taken every hour for the first 8 hours of growth then at 24 hours. (A) 0 ng/ml Atc and (B) 25 ng/ml Atc. Black solid line- R20291 + pASF085, grey solid line- R20291 + pECC17, black dashed line- R20291_*fliC*::CT + pASF085, R20291_*fliC*::CT + pECC17. Each culture was set up in duplicate with three independent replicates performed. Error bars represent standard deviation.

There was no difference when comparing the same strain in each of the four growth conditions. When comparing R20291 to R20291_*fliC*::CT, the latter demonstrated a reduced growth rate during exponential phase with and without Atc, but at 8 hours there was no difference in OD. The aflagellate strains appeared to clump more in the growth flasks, which may have interfered with the accuracy of cell density readings.

5.5. Swimming motility

Once the conditions for induction of c-di-GMP were established, swimming motility was assessed in all strains with and without induction of c-di-GMP. As swimming motility is a flagella-driven action, we hypothesised that only the absence of the flagella would lead to a reduction in swimming motility and there would be no difference between the PilK positive and negative strains (Figure 5.6). As previously described, a colony of each strain carrying either pASF085 or pECC17 was inoculated into 0.3% minimal media agar +/- 25 ng/ml Atc and containing 15 µg/ml thiamphenicol and incubated for 5 days.

As expected, strains without flagella did not disperse beyond the inoculation point and swimming was significantly reduced in flagella-positive strains in the presence of c-di-GMP (Figure 5.6) There was no significant difference in swimming motility between R20291 and R20291Δ*3343*, suggesting pili do not influence this phenotype (Figure 5.6), and that this motility is driven by expression of the flagella locus mediated via the type I riboswitch in the presence of c-di-GMP.

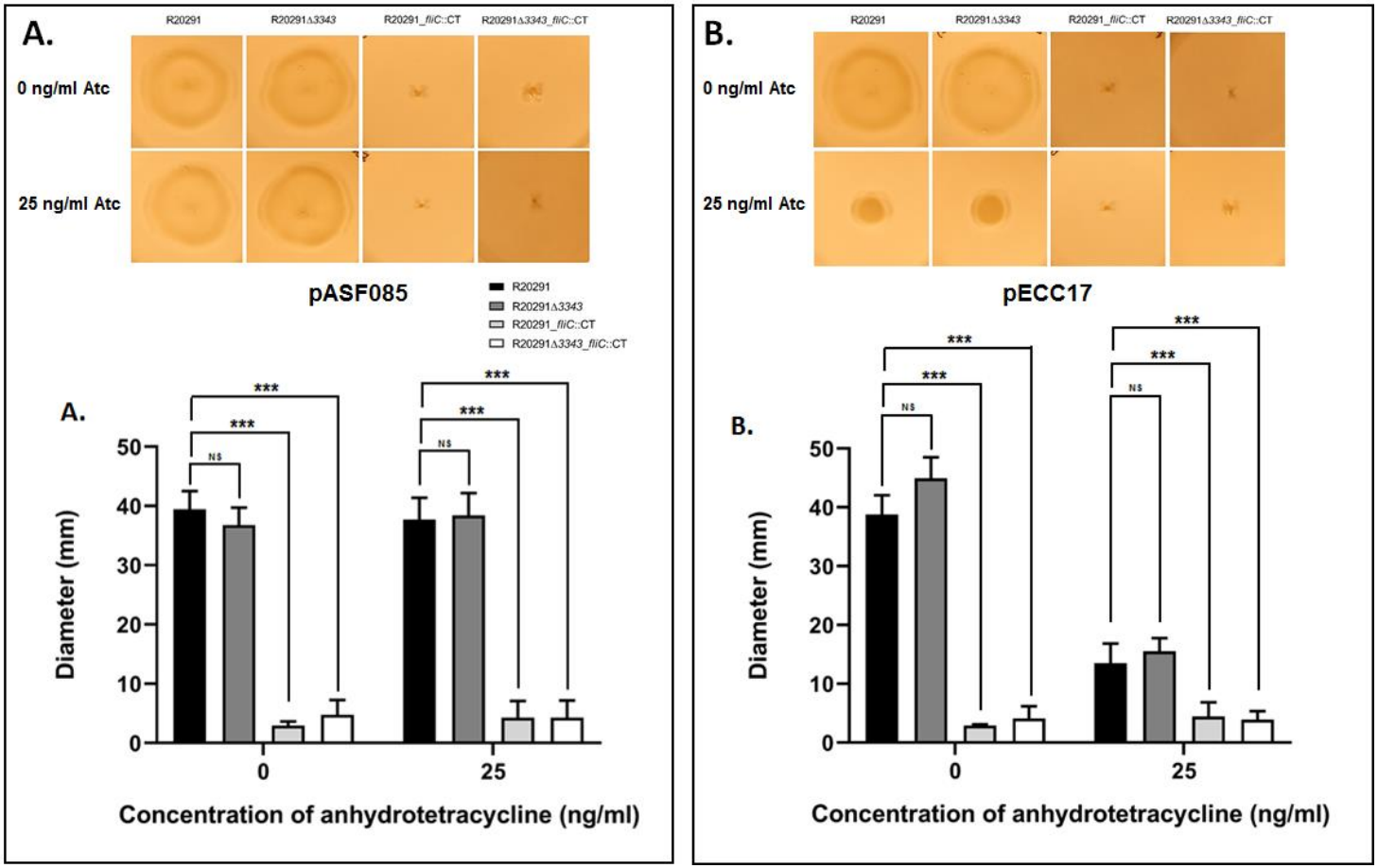


Figure 5.6. Swimming motility in the presence of inducible c-di-GMP. Individual *C. difficile* colonies were inoculated into the centre of 0.3% minimal media agar plates, supplemented with 0 or 25 ng/ml Atc for induction of c-di-GMP and 15 µg/ml thiamphenicol. Each strain carried either the empty vector pASF085 or pECC17, for Atc-mediated induction of c-di-GMP. Plates were photographed after 5 days incubation and two perpendicular measurements were taken of the diameter. A linear regression analysis was performed to assess any differences in swimming motility compared to the wild-type R20291. (A) Swimming motility with the empty vector pASF085, (B) swimming motility with pECC17, carrying Atc-inducible *dccA* for expression of c-di-GMP. NS not significant, * p<0.01, ** p<0.001, *** p<0.0001. Each culture was set up in duplicate with three independent replicates performed. Error bars represent standard deviation.

5.6. Colony morphology

Pili and flagella are both surface appendages and can direct movement of the cell, therefore it was interesting to see whether inactivation of either of these influenced colony morphologies. We hypothesised that in the presence of c-di-GMP, when pili expression is induced, cells would display a pili-mediated colony morphology which would be absent in the presumed pili-negative strain, R20291Δ*3343*. Strains carrying pASF085 or pECC17 were grown in BHIS supplemented with 15 µg/ml thiamphenicol to OD_{595} 0.3 - 0.4, then diluted in PBS to ensure single colonies when plated on 1.8 % BHIS agar + 1 % glucose +/- 25 ng/ml Atc and 15 µg/ml thiamphenicol and incubated for 5 days (Figure 5.7). 1.8% agar was selected as pili activity tends to increase with a harder surface [208]. The aflagellate strains displayed a typical *C. difficile* morphology in all conditions. In the presence of high-level c-di-GMP only, R20291 and R20291Δ*3343* displayed numerous fronds growing outwards from the centre colony (Figure 5.7). However, as this occurs in strains with and without PilK, it suggests this is not a pili-mediated phenotype, which may indicate that FliC plays a role in this extended colony morphology due the absence of this phenotype in the *fliC* mutants.

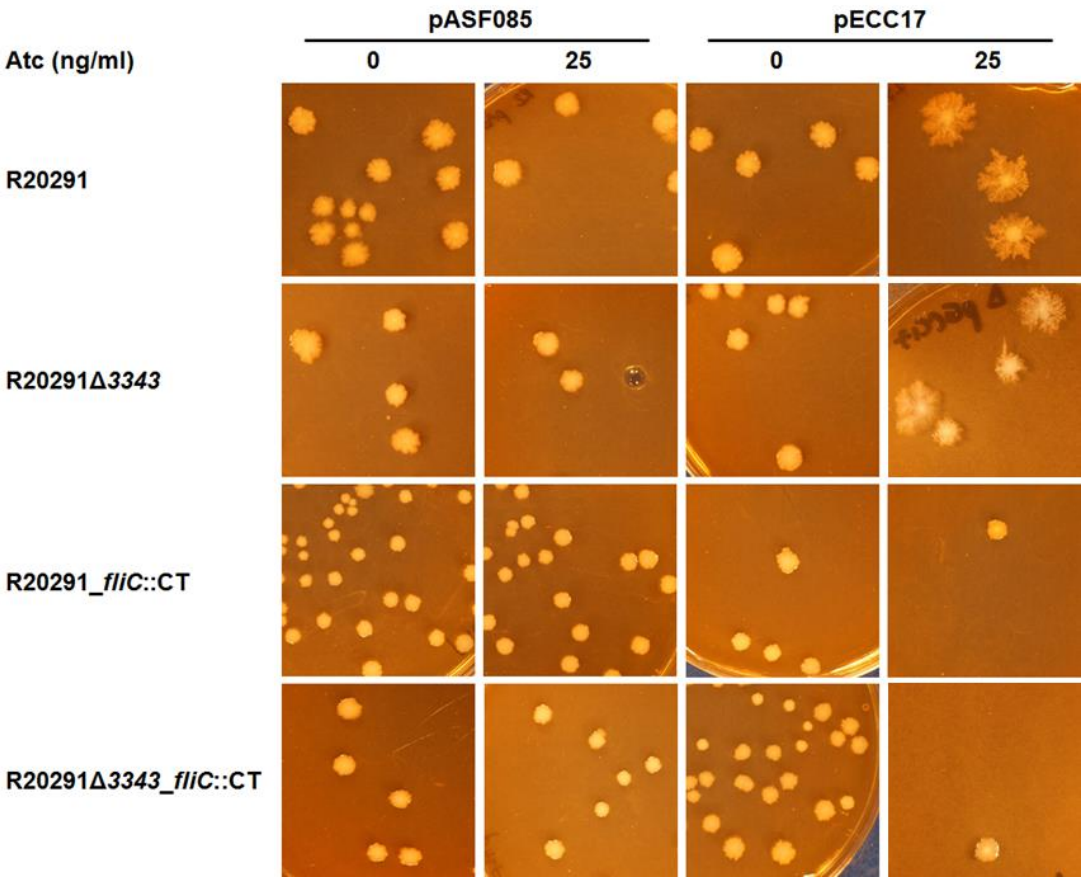


Figure 5.7. Colony morphology of *C. difficile*. Strains of *C. difficile* harbouring pili and flagella (R20291), either pili or flagella (R20291_*fliC*::CT or R20291Δ3343, respectively), or neither (R20291Δ*3343_fliC*::CT), were serially diluted then spread onto 1.8% BHIS agar plates supplemented with 15 µg/ml thiamphenicol and incubated for 5 days in the presence or absence of Atc for induction of c-di-GMP (pECC17) or with an empty vector (pASF085). Each culture was set up in duplicate with three independent replicates performed.

5.7. Surface motility

Both swarming and twitching motility enable bacteria to move across solid surfaces. While the former is attributed to the action of the flagella, twitching motility is associated with the pili, powered by their continuous assembly and disassembly. This has been demonstrated in many species, including *C. difficile* [208]. Surface hardness of the agar is believed to enable twitching motility, therefore 1.8% BHIS agar was used, based on previous work [208]. The assay was set up as outlined for colony morphology, but instead of plating dilutions, 5 µl of neat culture was spotted directly onto the surface of 1.8% agar with 15 µg/ml thiamphenicol, in quadruplicate. Two perpendicular measurements of the diameter of each growth spot were taken daily, for 5 days when plates were also photographed. It

was hypothesised that as pili are associated with surface motility, under c-di-GMP expression the bacteria would move outward from the centre spot of inoculation but this phenotype would be lost in the R20291Δ*3343* mutant strain.

Figure 5.8 presents the image of each growth spot and the total diameter reached after 5 days incubation. When carrying the empty vector pASF085, there was no significant difference between strains in the presence or absence of Atc. Upon c-di-GMP induction, the conditions under which flagella are down-regulated and pili are expressed, the motility radius for all strains increased in size, significantly (p<0.001). The flagella-positive strains dispersed significantly more than the aflagellate strains, and R20291 reached a larger total diameter than R20291Δ*3343*, although both displayed the frond phenotype as seen in the colony morphology assays (Figure 5.8).

The aflagellate strains did disperse beyond the initial inoculation site more under c-di-GMP induction, but this is a much smaller, more uniform halo compared to the flagella positive strains. In addition, a similar pattern is seen in both R20291_*fliC*::CT and R20291Δ*3343*_*fliC*::CT, suggesting this may be a general cellular effect of c-di-GMP, rather than a specific effect mediated by *fliC* or *pilK*. As was observed with the colony morphologies (Figure 3.7), the flagella positive strains produced much larger and more irregular swarms under c-di-GMP induction compared to uniduced strains and induced strains without flagella (Figure 5.8).

Overall, in the presence of c-di-GMP, surface motility was reduced in a PilK inactivated strain compared to the wild-type, but only in strains still harbouring an active flagella. Inactivation of the flagella showed the most substantial influence on this motility.

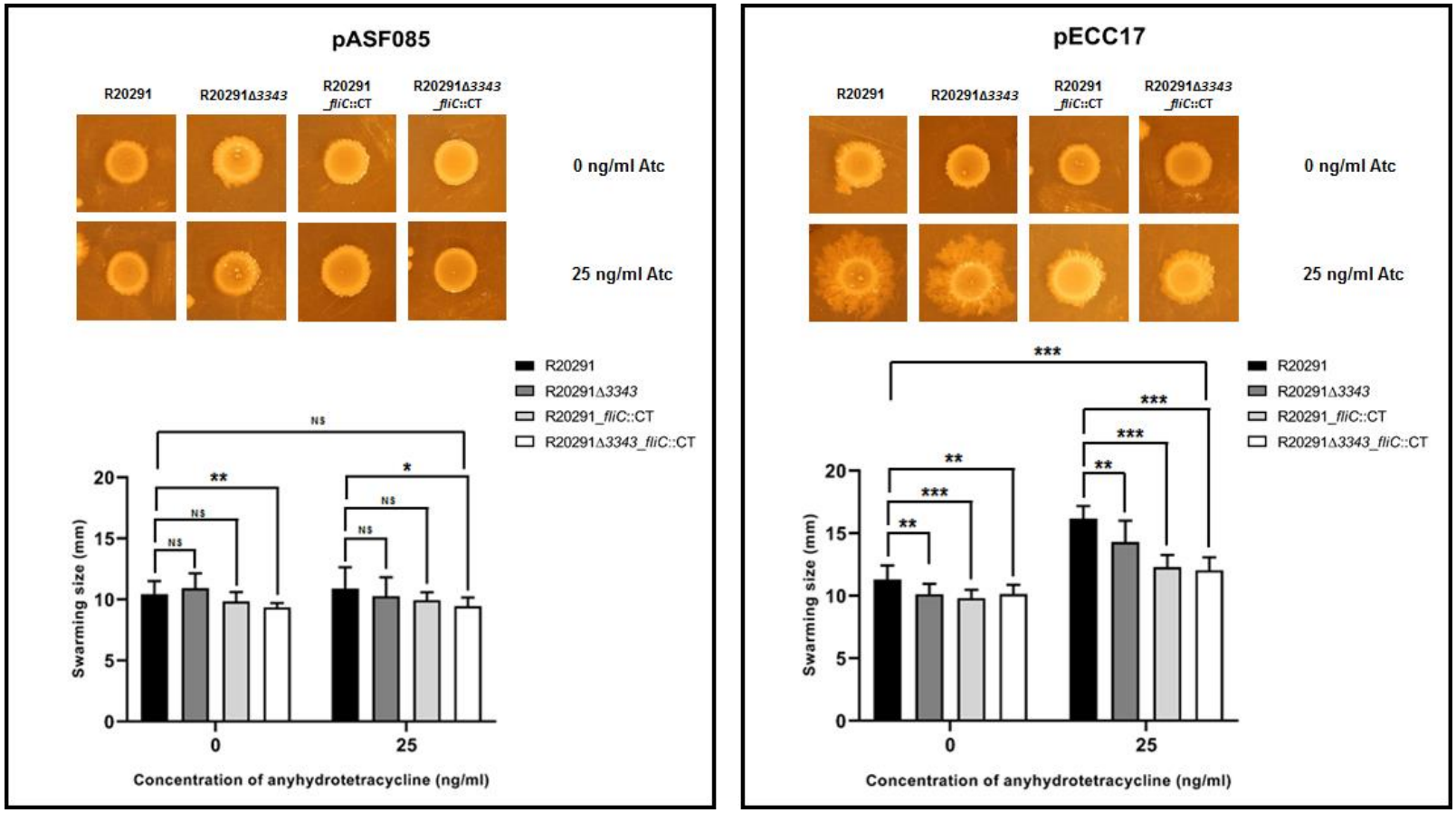


Figure 5.8. Surface motility of *C. difficile*. Strains of *C. difficile* harbouring pili and flagella (R20291), either pili or flagella (R20291_*fliC*::CT or R20291Δ*3343*, respectively), or neither (R20291Δ*3343*_*fliC*::CT), were spotted onto 1.8% BHIS agar plates with 15 µg/ml thiamphenicol and incubated for 5 days in the presence or absence of Atc for induction of c-di-GMP (pECC17) or with an empty vector (pASF085). Two measurements of the diameter of each inoculum spot were taken daily and plates photographed after 5 days. NS not significant, * p<0.01, ** p<0.001, *** p<0.0001. Each culture was set up in duplicate with three independent replicates performed. Error bars represent standard deviation.

5.8. Discussion

This chapter investigated the role of CDR20291_3343, a putative pilin protein predicted to form the tip of the pilus, PilK. CDR20291_3343 was investigated in relation to its role within the *C. difficile* T4P, assessed using motility assays in both flagella-positive and flagella-negative strains. To aid this, allele exchange mutagenesis was successfully applied to generate clean *CDR20291_3343* deletion mutants in a R20291 wild-type and R20291_*fliC*::CT backbone.

We found that in surface motility and colony morphology assays, under c-di-GMP induction, where flagella are downregulated and pili expression is induced, both R20291 and the *pilK* mutant, R20291Δ*3343*, produced frond like filaments, extending from the initial inoculation site (Figure 5.7 and 5.8), a phenotype previously attributed to the action of the pili [208]. The diameter of the total growth site, including the fronds, was significantly larger compared to the same strains growing without c-di-GMP induction. This phenotype was previously observed in R20291 and attributed to pili-mediated motility due to the observation that induction of c-di-GMP leads to downregulation of flagella expression and induction of pili expression [198, 208]. However, the total motility on the surface of the agar for the flagella-negative strains did not increase significantly upon c-di-GMP induction, and they remained significantly smaller than the flagella-positive strains and the frond phenotype was absent (Figure 3.6). From these results, three potential mechanisms driving the surface motility phenotype were considered; i) the action of the pili, ii) the action of the flagella or iii) pili and flagella independent effect, of induction by c-di-GMP.

If surface motility and frond formation in the presence of c-di-GMP were due solely to the action of the pili, then loss of surface motility would be expected in the *pilK* mutant (R20291Δ*3343* and R20291Δ*3343_fliC*::CT), as the pilus should not be assembled in this mutant [541]. However, both R20291 and R20291Δ*3343* were able to move across the agar surface. These results indicate the phenotype seen here is likely to be independent of the pili. There was a small but significant reduction in motility in R20291Δ*3343* compared to wild-type R20291, however, it was still capable of movement

and produced the distinctive frond display. Although this suggests that the pili could still be functional in the absence of PilK, it may also be influenced by the variability within surface motility assays. It has previously been reported that the conditions for testing can influence this phenotype, as has been described for organisms such as *Pseudomonas aeruginosa*, where motility has been thoroughly investigated. The composition and hardness of the agar surface can influence results which therefore can provide room for error, particularly when measuring millimetre differences in motility [530, 543].

It may be that PilK inactivation does not abolish pili formation in *C. difficile*, as this has not been confirmed using microscopy. This was attempted in this study by providing samples of each strain with and without c-di-GMP induction to Tony Fearns in the Gutierrez laboratory at the Francis Crick Institute for scanning electron microscopy, but no data had been obtained at the time of writing. Even so, the lack of secretion of the major pili structural subunit, PilA1, in a PilK mutant is highly suggestive of lack of pilus assembly [541]. This was demonstrated in strain 630 but the pili loci are highly conserved between 630 and R20291, albeit differentially responsive to c-di-GMP [208]. Furthermore, evidence that this is a pili-independent phenotype is provided in a recent thesis by Couchman, who could detect no difference in swarming motility between R20291 wild-type and a ClosTron PilB1 mutant [541]. PilB1 is the ATPase driving assembly of the pili and in the mutant, pili cannot be detected on the surface of the cell via microscopy [208].

Interestingly, both our results and those obtained by Couchman contradict the only published data on "pili-mediated" motility in *C. difficile*. Here, Purcell *et al*. also utilised ClosTron mutagenesis to inactivate PilB1 in R20291, but found in their strain, the surface motility phenotype was abolished, including in the presence of c-di-GMP [208]. Furthermore, where the authors demonstrate substantial surface motility in R20291 without c-di-GMP induction, we see very little expansion from the inoculation site without c-di-GMP induction (Figure 5.8). Couchman speculated that the insertional inactivation used to remove PilB1 activity in the published work may have inactivated other genes in the operon, as complementation of PilB1 was not demonstrated. Therefore, they speculate that

inactivation of PilB1 has disrupted genes required for functioning of the secondary pili gene cluster, which may be the true mediator of motility, but this has not been tested [541].

Alternatively, is this simply a result of high-level expression of a prolific secondary messenger with multiple cellular targets? Overexpression of c-di-GMP can influence cellular morphology, with published microscopy identifying elongation and curling of cells [221]. However, if this were purely an effect of c-di-GMP on the cell, independent of flagella or pili, one may hypothesise this would be observed in flagella and non-flagella strains, alike.

This study is the first to test surface motility of *C. difficile* flagella-negative strains. Aflagellate strains were found to be significantly less motile than strains harbouring their flagella and did not produce the filamentous fronds seen in the flagellate strains. If this phenotype were truly a result of the action of the flagella filament, then it would be present in low-level c-di-GMP (i.e. the empty vector with and without Atc and c-di-GMP but without induction), when the flagella are expressed and active. However, it is only found upon induction of c-di-GMP, conditions which downregulate the flagella, as demonstrated previously [198] and in the swimming assays in this study. Future work should determine *fliC* expression in the surface motility assay without and without c-di-GMP to confirm it is not being expressed.

If not a direct result of the flagella filament, could this be related to regulation? The flagella operon in *C. difficile* does not just encode for expression and assembly of the flagella, it also can also mediate regulation of a number of cellular processes, most notably toxin expression. Toxin expression is positively regulated by the alternative sigma factor, SigD, which is encoded with the early stage flagella genes. SigD regulates expression of the late stage flagella genes, in addition to a number of other genes throughout the *C. difficile* genome [191]. Pili formation was demonstrated in a SigD mutant strain, but as the motility phenotype seen here is not believed to be pili-related, its activity cannot be ruled out [195]. However, a previous ClosTron insertion with FliC did not inactivate SigD, as toxin expression was unaffected, therefore it is likely SigD activity is not disrupted in our FliC mutant [544].

Alternatively, is this a result specific to the inactivation of FliC? As mentioned, previous inactivation of FliC did not inactivate SigD, but did result in differential expression of 310 genes *in vivo* and 258 genes, *in vitro*, although only 36 of these were shared across the two groups. [544]. Further investigation of surface motility in *C. difficile* would also have to consider how PilB1 contributes to this, as surface motility was lost in a PilB1 ClosTron mutant [208].

Interestingly, during preparation of this discussion a publication was deposited on the pre-peer review site, BioRxiv, which identified a putative phosphorelay system, named the colony morphology regulators (*cmr*), involved in regulation of *C. difficile* motility [545]. This is composed of a histidine kinase (*cmrS*) and two response regulators (*cmrT* and *cmrS*), positively regulated by a c-di-GMP riboswitch and under additional control of an additional genetic switch which mediates phase variation. The authors related two colony morphologies to this system- rough and smooth. In the presence of c-di-GMP all colonies were of the rough phenotype, but in the absence of c-di-GMP, the genetic switch determined between the rough (ON) and smooth (OFF) morphologies. These are believed to be adaptations to the environment, as rough colonies were associated with the fronding phenotype in surface motility assays (although without induction of c-di-GMP) and the smooth colonies were isolated from the swimming agar [545]. This supports our findings of a difference in behaviour of our R20291 and R20291Δ*3343* strains in swimming and surface motility assays, when both are under c-di-GMP induction.

The *cmr*-mediated spreading of bacteria in the surface motility assays was suggested to be a result of bacterial chaining, as identified using scanning electron and light microscopy, although the exact targets of the response regulators is unknown and there may be multiple [545]. Importantly this appears to be a flagella and T4P independent phenotype, as rough colonies were produced upon c-di-GMP induction in a SigD and PilB1 mutant. Furthermore, transcript levels of T4P and flagellum genes were not altered when the two response regulators CmrT and CmrR were expressed *in trans* [545].

This is interesting as the authors demonstrate surface motility of R20291 without c-di-GMP, which is lost in the PilB1 mutant. It therefore brings into question the exact contribution of pili and the *cmr* response regulator in this motility. Furthermore, in our results, we see no surface motility unless c-di-GMP is induced and as mentioned there is no different between wild-type and the pili-negative strain. Finally, how do flagella relate to this? Our surface motility phenotype is completely absent in FliC negative strains, yet flagella are not believed to be regulated by *cmr*. The current study did not assess surface motility in the SigD mutant. It would interesting to further investigate the activity of *cmr* in the FliC mutant to see if rough and smooth colonies can be isolated, and the effect of over expression of the two response regulators on colony morphology and surface motility phenotypes.

Irrespective of the mechanism(s) behind surface motility, it appears this is not a reliable or definitive measure of pili activity. A number of other *C. difficile* phenotypes have been associated with the pili, including aggregation, biofilm formation, adhesion to host cells, and *in vivo* colonisation [195, 208, 212, 213]. Assessment of these mechanisms may provide further insight into the relationship between flagella and pili and any associated regulation in these phenotypes. The recent finding relating to a role for them in *in vivo* colonisation was interesting and is one of the strongest publications to date to link the *C. difficile* pili to an active role in virulence and colonisation of the host, rather than exclusively *in vitro* phenotypes [213].

The lack of a reliable method of measuring motility was a limitation to this study as was the use of c-di-GMP as a means of studying flagella and pili expression and their associated motility phenotypes. Although this is the standard means of such assessments of motility and is one of the regulatory mechanisms for these surface appendages *in vivo*, it is difficult to tease apart specific influence of the flagella and pili on motility when c-di-GMP has so many targets within the cell [202, 203]. Additionally, despite some preliminary evidence from other groups and ongoing work with a collaborator [541], it has not yet been possible to confirm CDR20291Δ*3343* is a pili-negative strain. It would be interesting to repeat these assays using a PilA1 mutant, which is known to not produce pili [195] and see if the

surface motility phenotype is the same as seen here. Finally, as with the CDR20291Δ*0330* work, these phenotypes have not been complemented. As there is a lack of resistance markers available for plasmid maintenance in *C. difficile* and a plasmid is already in use for expression of c-di-GMP, there are two options for construction of these complements. R20291_*3343* could be reintroduced onto the genome using allele exchange mutagenesis and watermarked by changing the stop codon so the complement could be differentiated from the wild-type by sequencing. This would not be applicable for R20291_*fliC*::CT as this is a ClosTron mutant so instead the *fliC* gene could be cloned into the same plasmid as the *dccA* gene for c-di-GMP expression, under its native promoter.

The whole pilus or different pilin components have been demonstrated to induce an immune response in mice for *Neisseria meningitidis* and *P. aeruginosa*, and an anti-pili vaccine is licensed for the animal pathogen *Moraxella bovis* [207]. As pili tend to be assembled with the hydrophobic N-terminal buried within the central helix, the exposed C-terminal domain is predicted to be the immunogenic portion [546].

In the pan-protein array described in Chapter 3, CDR20291_3343 was found to be immunogenic in a pilot study. The follow up array found a higher response to the protein in healthy controls compared to those with CDI, although this was not significantly different. Individual pilins from *C. difficile* (PilK not tested) induced an immune response in the mouse model of CDI [207], and in a follow up study, three pilin proteins, PilA, PilJ and PilW were taken forward for challenge experiments to determine protective efficacy. However, in these experiments, only weak antibody responses were raised to these pili and no protection was demonstrated. As the authors were unable to replicate their own results in relation to antibody generation against the pilins, they suggest these results may be a result of the mouse model differing between the first and second studies. Furthermore, even the dmLT adjuvant didn't induce an immune response in these experiments, which had been used successfully in previous mouse studies [546]. Therefore, the true immunogenic properties and vaccine potential of these pilins warrants further investigation. Additionally, reviews on the design of anti-*C. difficile*

vaccines have noted the likely need for a multi subunit approach, covering many surface antigens in addition to the toxin, rather than requiring on individual component [449].

6. Characterisation of CDR20291_0342, a putative permease protein

6.1. Introduction

ATP-binding cassette (ABC) transporters are ubiquitous integral membrane proteins in prokaryotes. Through hydrolysis of ATP, they transport a variety of substrates across the lipid bilayer, both in and out of the cell. Their preferred substrates are diverse, as is their function, which includes but is not limited to; nutrient acquisition, antimicrobial resistance, export of virulence factors and trafficking of cell surface components [547]. The number of ABC transporters a bacterial genome encodes appears to correlate with both genome size and the environmental niche within which the bacteria resides [548].

CDR20291_0342 has been annotated as the putative permease component of an ABC transporter [111], but information regarding this protein in the literature is limited. Forgetta *et al.* found *CDR20291_0342* was one of the genes identified with a high SNP prevalence, which differentiated hypervirulent 027 isolates from other *C. difficile* strains, although the impact of this on the activity or putative function of the transporter was not determined [549]. Transcriptomic analysis of *C. difficile* identified that *CDR20291_0342* was significantly downregulated during germination [506].

This chapter aimed to characterise the role of CDR20291_0342 and its associated ABC transporter within *C. difficile*, using allele exchange mutagenesis to inactive the gene and a series of phenotyping assays to assess the impact of this gene's inactivation.

6.2. Bioinformatic analysis of CDR20291_0342

CDR20291_0342 is a predicted 212 KDa protein, conserved across a diverse range of *C. difficile* strains (Table 6.1). Amino acid sequence similarity analysis across a diverse selection of *C. difficile* strains was 86%, slightly reduced compared to the level of conservation reported for CD20291_3343 and CDR20291_0330, where amino acid similarity was over 90% for all strains. This is likely a result of the high SNP prevalence within this gene, as discussed above [549]. Following on from this, the amino acid sequence of CDR20291_0342 from strains 630 and R20291 were aligned which showed the high SNP prevalence did introduce non-synonymous mutations, but none resulted in a stop codon and therefore these SNPs do not affect expression of the full protein (Appendix E). The CD305 genome was unavailable for the first analysis but when investigated at a later date, it was found CDR20291_0342 and the two genes it is encoded with were absent from this genome, and from another ribotype 023 strain.

Strain	***Ribotype***	***Percentage identity (%)***
R20291	027	100
CD196	Historic 027	100
630	012	86
M120	078	86
M68	017	86
CD305	023	-

Table 6.1 Conservation of CDR20291_0342 amino acid sequence across *C. difficile* strains. The conservation of the CDR20291_0342 amino acid sequence across a number of *C. difficile* strains from diverse ribotypes was compared using pBLAST searches with the R20291 amino acid sequence as the comparator sequence. CDR20291_0342 and the two genes it is encoded with are absent from the ribotype 023 strain CD305.

Phyre[2] analysis identified 8 putative transmembrane domains within the protein, and it is encoded in the genome alongside a putative ATPase (*CDR20291_0341*), supporting its annotation as the permease component of an ABC transporter (Figure 6.1a and b). A third, hypothetical protein is

encoded alongside these (CDR20291_0340), but analysis of the protein provided no indication as to its function or the potential role of the transporter.

Analysing the amino acid sequence using pFam and Phyre^2 corroborated the annotation of *CDR20291_0342* as a permease. 21% of the whole protein was modelled with 99.2% confidence against the N-terminus of the LolE protein from *Acinetobacter baumannii*, with high amino acid similarity also found when modelled against MacB. LolE is the transmembrane component of the lipoprotein releasing system LolCDE, which traffics lipoproteins to the outer membrane in *E. coli* [550]. Bioinformatic analysis can help identify putative lipoproteins by screening for the consensus lipobox sequence within 40 residues of the N-terminus of the protein [551]. The amino acid sequence of theCDR20291_0340 hypothetical protein neighbouring CDR20291_0342 was analysed using The Database of Bacterial Lipoproteins (https://www.mrc-lmb.cam.ac.uk/genomes/cgi/dolop/newlipo.cgi), but was not a predicted lipoprotein.

The MacB transporter has been identified as a means of resistance to macrolides and antimicrobial peptides in a number of species, although it can export a range of compounds from the cell, including virulence factors [552, 553]. In addition to the MacB and LolE domains, pFam also highlighted a SalY domain, related to antimicrobial peptide (AMP) transport (Figure 6.1 c). Taken together, these results indicate the transporter may be involved in mediating efflux from the cell, for either lipoproteins, antibiotics or antimicrobial peptides.

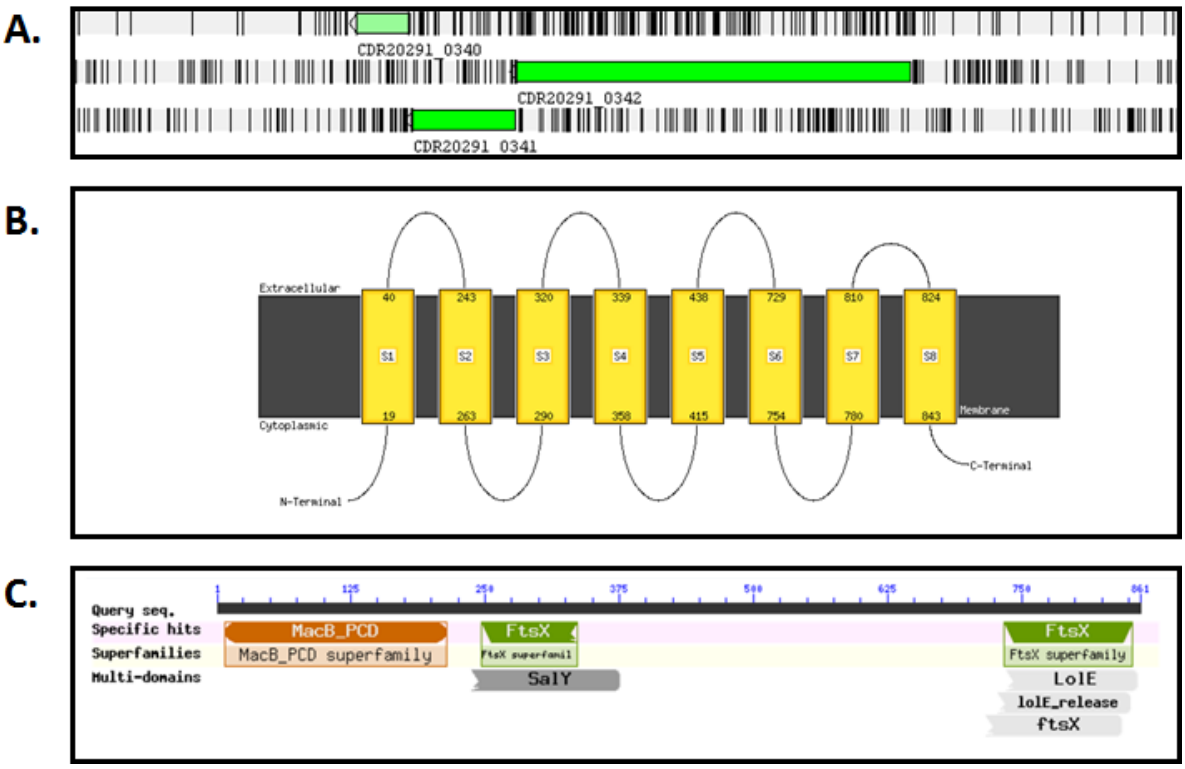


Figure 6.1. Bioinformatic analysis of CDR20291_0342. (A) The position of the putative permease *CDR20291_0342* in the R20291 genome, encoded with the other components of a predicted ABC transporter, including the putative ATPase, *CDR20291_0341* and a hypothetical protein, *CDR20291_0340*. (B) A schematic of the organisation of CDR20291_0342 within the cell membrane, produced from Phyre analysis of the CDR20291_0342 amino acid sequence, revealing 8 putative transmembrane domains. (C) Analysis of CDR20291_0342 amino acid sequence using the pFam database, revealing regions of homology with domains of efflux transporters LolE and MacB.

6.3. Construction of a CDR20291_0342 gene deletion mutant

Deletion of *CDR20291_0342* from the chromosome could aid understanding as removal of the permease component is hypothesised to inactivate the entire transporter. Allele exchange mutagenesis was used to delete *CDR20291_0342* from the chromosome, as described in 4.3.2 and 5.4.1 (Appendix B). Potential double cross over mutants were screened using primers flanking both the entire homology region (MUTSEQ_F and R) and the gene only (SEQ_F and R) (Figure 6.2). Four colonies generated amplicons matching wild-type R20291, suggesting wild-type revertants. Two amplicons were ~1.2 Kb smaller than the wild-type amplicon, indicating a gene deletion mutant, which was later confirmed with sequencing across the deletion.

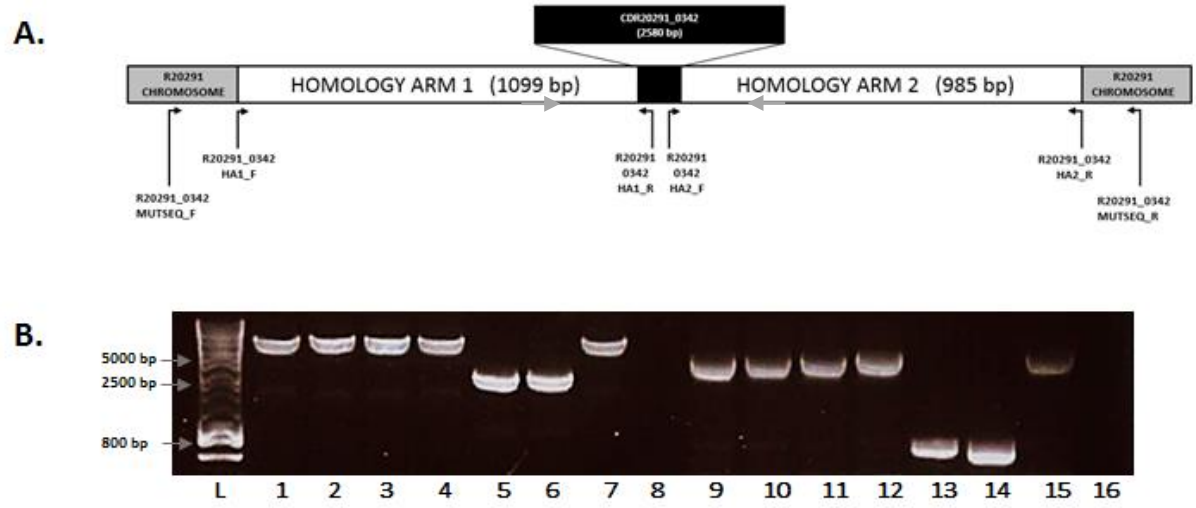


Figure 6.2. Allele exchange mutagenesis of *CDR20291_0342* in R20291. (A) Schematic of the allele exchange construct for generation of the *CDR20291_0342* gene deletion mutant. This is annotated with the primers used for screening to detect gene deletion mutants. (B) PCR screen of potential *CDR20291_0342* gene deletion mutants. Lanes 1 to 6- individual clones from R20291, 7- R20291 wild-type, 8- water only, all screened using primers R20291_0342_MUTSEQ_f and R20291_0342_MUTSEQ_r. Lanes 9-16- individual clones from R20291, 15- R20291 wild-type, 16- water only, all screened using primers R20291_0342_SEQ_f and R20291_0342_SEQ_r (grey arrows). The predicted size of amplicons with and without *CDR20291_0342* are 4800 bp and 2262 bp, respectively for the MUTSEQ primers and 3288 bp and 750 bp, respectively, for the SEQ primers. Clones from lanes 5 and 6 and 13 and 14 were suspected gene deletion mutants, which were confirmed by sequencing.

6.4. Antimicrobial susceptibility

As the bioinformatic analysis identified domains relating to antimicrobial efflux, presumably for the purposes of resistance, it was hypothesised that CDR20291_0342 and its associated transporter could be an antimicrobial efflux pump. To investigate this, the sensitivity of R20291 and R20291Δ*0342* against the macrolide erythromycin, the lincosamide lincomycin and the cyclic peptide bacitracin were all investigated. These were chosen as the amino acid sequence CDR20291_0342 harboured amino acid similarity to the antimicrobial efflux pump MacB, which can export macrolides from the cell in other species [552]. Macrolide efflux pumps can often also transport lincosamides and therefore lincomycin was investigated. Finally, as a cyclic peptide bacitracin is sometimes classed as an antimicrobial peptide and was therefore investigated as CDR20291_0342 harbours a SalY domain which is annotated in pFam as being associated with antimicrobial peptide resistance (Figure 6.1). The

broth dilution method was used to determine the MIC of both strains against the three compounds, but no difference in MIC was found for any of the antimicrobials (Table 6.2) in either strain.

Strain	Bacitracin	Erythromycin	Lincomycin
R20291	>512 µg/ml	> 512 µg/ml	20 µg/ml
R20291Δ*0342*	>512 µg/ml	> 512 µg/ml	20 µg/ml

Table 6.2. Antimicrobial susceptibility of R20291Δ*0342*. R20291 and R20291Δ*0342* were tested for susceptibility to erythromycin, bacitracin and lincomycin using the broth dilution method. Strains were incubated in 24-well tissue culture plates containing 1 ml BHI broth supplemented with bacitracin and erythromycin at 512, 256 and 128 µg/ml and lincomycin at 40, 20 and 10 µg/ml. Pre-equilibrated plates were inoculated with a day culture of each strain to a starting OD_{595} 0.003 and incubated for 16 hours before the OD_{595} was taken for each well using a plate reader. Each concentration was tested in duplicate and three independent replicates were performed.

6.5. Growth kinetics

To determine whether loss of *CDR20291_0342* influenced growth, both the mutant and wild-type were grown in BHI media and OD_{595} taken every hour for 8 hours then a final reading at 24 hours. There was no difference in growth rate between the two strains (Figure 6.3).

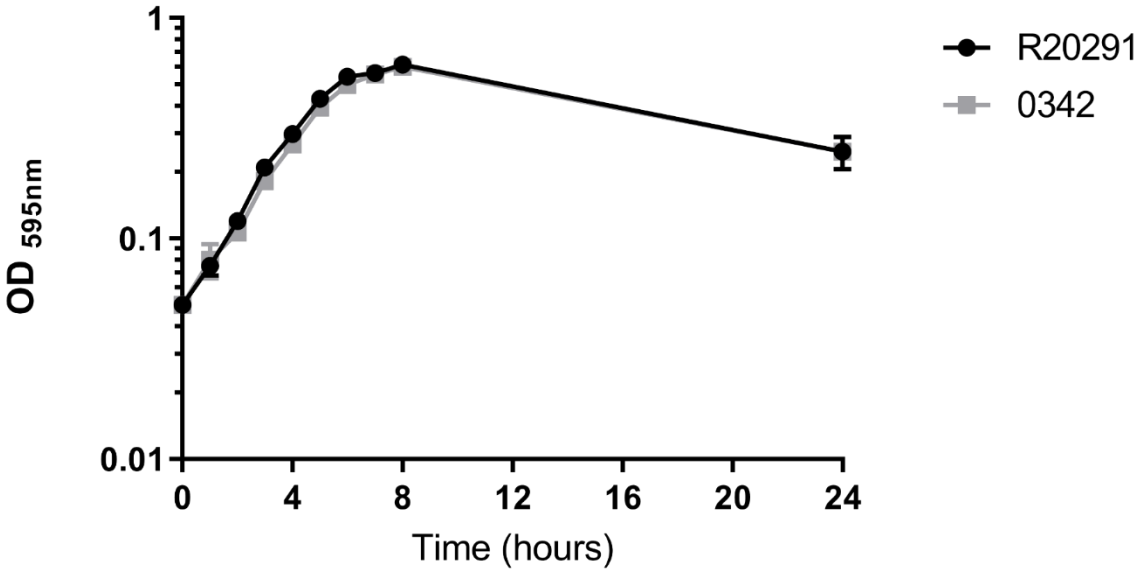


Figure 6.3. Growth rate of R20291Δ*0342* in rich media. R20291 and R20291Δ*0342* were grown overnight in BHI broth then diluted to a starting OD_{595} 0.05 in fresh media. Absorbance readings were taken every hour for the first 8 hours followed by a final 24 hour reading. Cultures were set up in duplicate and three independent replicates were performed. Error bars represent standard deviation.

6.6. Carbon utilisation

6.6.1. Phenotypic microarrays

Where the role of a particular protein within the cell is not easily decipherable from bioinformatics analysis alone, large-scale screening methodologies can be of use, enabling rapid identification or exclusion of potential functions. Biolog Phenotypic Microarrays assays are an established methodology for high-throughput screening of a number of different growth conditions in a plate format, a much simpler set up than traditional growth kinetics [554], and have recently been optimised for the study of anaerobes.

The Biolog assays were performed and analysed at The Wellcome Trust Sanger Institute by Dr Hilary Browne and Dr Kevin Vervier, respectively. Strains R20291 and R20291Δ*0342* were grown with 95 different carbon sources, broadly grouped into acids, amino acids and carbohydrates. Two other, unrelated strains were screened simultaneously- R20291Δ*0330* and R20291_*2436*::CT. Plates were inoculated with *C. difficile* overnight culture and incubated in sealed anaerobic bags within an OminoLog for automated cell density readings. A minimum of seven replicates were performed per strain and analysis by Dr Vervier allocated a growth/no growth outcome for each strain in every tested growth condition, based on OD readings. Tables 6.3, 6.4 and 6.5 detail growth outcomes in amino acids, acids and carbohydrates, respectively. Any carbon source where no growth was recorded for any strain was excluded.

Although two versions of the wild-type R20291 strain were tested, these performed poorly in the assay, with strains failing to grow in the majority of tested carbon sources, including well established sugars such as glucose. Therefore, to try and extract any potential phenotypes for R20291Δ*0342*, all strains were compared to one another. It was noted that R20291Δ*0342* did not grow with L-valine, aspartic acid plus L-valine or alpha-ketovaleric acid. Both R20291_*2436*::CT and R20291Δ*0330* grew with L-valine and these strains plus one of the wildtype R20291 isolates grew with L-valine and aspartic

acid and alpha-ketovaleric acid. As valine was central to all these compounds, growth of R20291 and R20291Δ*0342* in L-valine was taken forward for further investigation.

	Amino acids					
Carbon Source	**R20291 (1)**	**R20291 (2)**	**R20291_*2436*::CT**	**R20291Δ*0330***	**R20291Δ*0342***	**No. strains**
Alaninamide	-	-	+	-	-	**1**
L-Alanine	-	+	+	+	-	**3**
L-Alanyl-L-glutamine	-	-	+	-	-	**1**
L-Alanyl-L-Histidine	-	-	+	-	-	**1**
L-Alanyl-L-Threonine	-	-	+	+	-	**2**
Glycyl-L-Aspartic Acid	-	-	+	-	-	**1**
Glycyl-L-Glutamine	-	+	+	-	-	**2**
Glycyl-L-Proline	-	-	+	-	-	**1**
L-Phenylalanine	-	-	+	-	+	**2**
L-Serine	-	-	+	+	-	**2**
L-Threonine	-	-	+	-	-	**1**
L-Valine	-	-	+	+	-	**2**
L-Valine plus L-aspartic acid	+	-	+	+	-	**3**
Deoxyadenosine	-	-	+	-	-	**1**
Inosine	-	-	+	-	-	**1**
Thymidine	-	-	+	-	-	**1**
Uridine	-	-	+	-	-	**1**
Thymidine monophosphate	-	-	+	-	-	**1**
Uridine-Monophosphate	-	-	+	-	-	**1**
No. growth conditions (n=19)	**1**	**2**	**19**	**5**	**1**	

Table 6.3. Amino acids utilisation in R20291 and R20291Δ*0342*. Growth of R20291 and R20291Δ*0342* compared to each other and two other strains; R20291Δ*0330* and R20291_*2436*::CT in the presence of different amino acids as carbon sources. This was performed using a Biolog Phenotypic Assay on an AN microplate. Green and plus indicates growth and red and minus indicates no growth. The last column indicates the total number of strains growing in each condition and the last row displays the number of growth conditions each strain grew in. Each strain was tested in seven independent assays.

	Acids					
Carbon source	**R20291 (1)**	**R20291 (2)**	**R20291_*2436*::CT**	**R20291Δ*0330***	**R20291Δ*0342***	**No. strains**
Acetic Acid	-	-	+	-	-	**1**
Glyoxylic Acid	-	-	+	-	-	**1**
Itaconic Acid	-	-	-	+	-	**1**
alpha-Ketobutyric Acid	-	-	+	+	-	**2**
alpha-Ketovaleric Acid	-	+	+	+	-	**3**
DL-Lactic acid	-	-	-	+	-	**1**
L-Lactic Acid	-	+	+	+	-	**3**
Pyruvic Acid	-	-	+	-	+	**2**
Pyruvic Acid Methyl Ester	-	-	+	-	-	**1**
Succinic Acid	-	-	+	-	-	**1**
Succinic Acid Mono-Methyl Ester	-	-	+	-	+	**2**
m-Tartaric Acid	-	-	-	+	-	**1**
No. growth conditions (n=12)	**0**	**2**	**9**	**6**	**2**	

Table 6.4. Acid utilisation in R20291 and R20291Δ*0342*. Growth of R20291 and R20291Δ*0342* compared to each other and two other strains; R20291Δ*0330* and R20291_*2436*::CT in the presence of different acids as carbon sources. This was performed using a Biolog Phenotypic Assay on an AN microplate. Green and plus indicates growth and red and minus indicates no growth. The last column indicates the total number of strains growing in each condition and the last row displays the number of growth conditions each strain grew in. Each strain was tested in seven independent assays.

	Carbohydrates					
	R20291 (1)	**R20291 (2)**	**R20291_*2436*::CT**	**R20291Δ*0330***	**R20291Δ*0342***	**No. strains**
D-Cellobiose	-	-	+	-	-	**1**
beta-Cyclodextrin	-	-	-	+	-	**1**
Dextrin	-	-	+	-	-	**1**
Dulcitol	-	+	+	-	-	**2**
D-Gluconic Acid	-	-	+	-	-	**1**
alpha-D-Glucose	-	-	+	-	-	**1**
D-Glucose6-Phosphate	-	+	+	+	-	**3**
Maltotriose	-	-	+	-	-	**1**
D-Mannose	-	-	+	-	-	**1**
3-Methyl-D-Glucose	-	-	+	+	+	**3**
alpha-Methyl-D-Galactoside	-	-	+	-	-	**1**
beta-Methyl-D-Glucoside	-	-	+	-	-	**1**
D-Trehalose	-	-	+	-	-	**1**
Turanose	-	-	+	-	-	**1**
No. growth conditions (n=14)	**0**	**2**	**13**	**3**	**1**	

Table 6.5. Carbohydrate utilisation in R20291 and R20291Δ*0342*. Growth of R20291 and R20291Δ*0342* compared to each other and two other strains; R20291Δ*0330* and R20291_*2436*::CT in the presence of different amino acids as carbon sources. This was performed using a Biolog Phenotypic Assay on an AN microplate. Green and plus indicates growth and red and minus indicates no growth. The last column indicates the total number of strains growing in each condition and the last row displays the number of growth conditions each strain grew in. Each strain was tested in seven independent assays.

6.6.2. Valine utilisation

To further investigate the importance of valine in the growth of R20291 and any differences in R20291Δ*0342*, a series of growth kinetics were conducted using the defined minimal media as described in 2.7.1.2. Both strains were grown with and without glucose with one of three amino acid preparations. In the strains tested, Karasawa *et al.* identified 11 amino acids required for optimal growth of *C. difficile* and 6 amino acids essential for growth; valine, proline, tryptophan, leucine, isoleucine and cysteine. Both of these amino acid preparations were tested in addition to the basic 6 but with valine removed to confirm this was essential in R20291. Furthermore, it was hypothesised that if R20291 were able to grow in the 6 amino acid composition without glucose, i.e. using the amino acids as carbon sources only, that this growth would be impaired in the R20291Δ*0342* strain if valine utilisation were not possible.

Strains were inoculated directly from plate cultures into individual defined minimal media overnights, which matched the final growth conditions and incubated for 24 hours. Only the growth obtained in 11 amino acids with glucose was sufficient for inoculation of the final growth assays, therefore this was used as the inoculant for all conditions. It is likely the drastic change in environment from a rich plate culture to very minimal nutrient broth did not facilitate growth in the other conditions in the overnight set up. OD_{595} readings were taken using a plate reader every hour for 8 hours then a final reading at 24 hours by removing 100 µl of culture for measurement. The results obtained from growth of both strains in all three amino acid set ups are detailed in Figure 6.4.

Unsurprisingly, the maximum OD_{595} achieved decreased as the number of amino acids was reduced, and both strains exhibited poor growth without glucose. Interestingly, both strains appeared to be capable of growing without valine, suggesting this is not essential for growth of R20291. Finally, there was no difference in growth between the strains in any of the conditions, suggesting inactivation of *CDR20291_0342* does not impact on growth in the conditions tested.

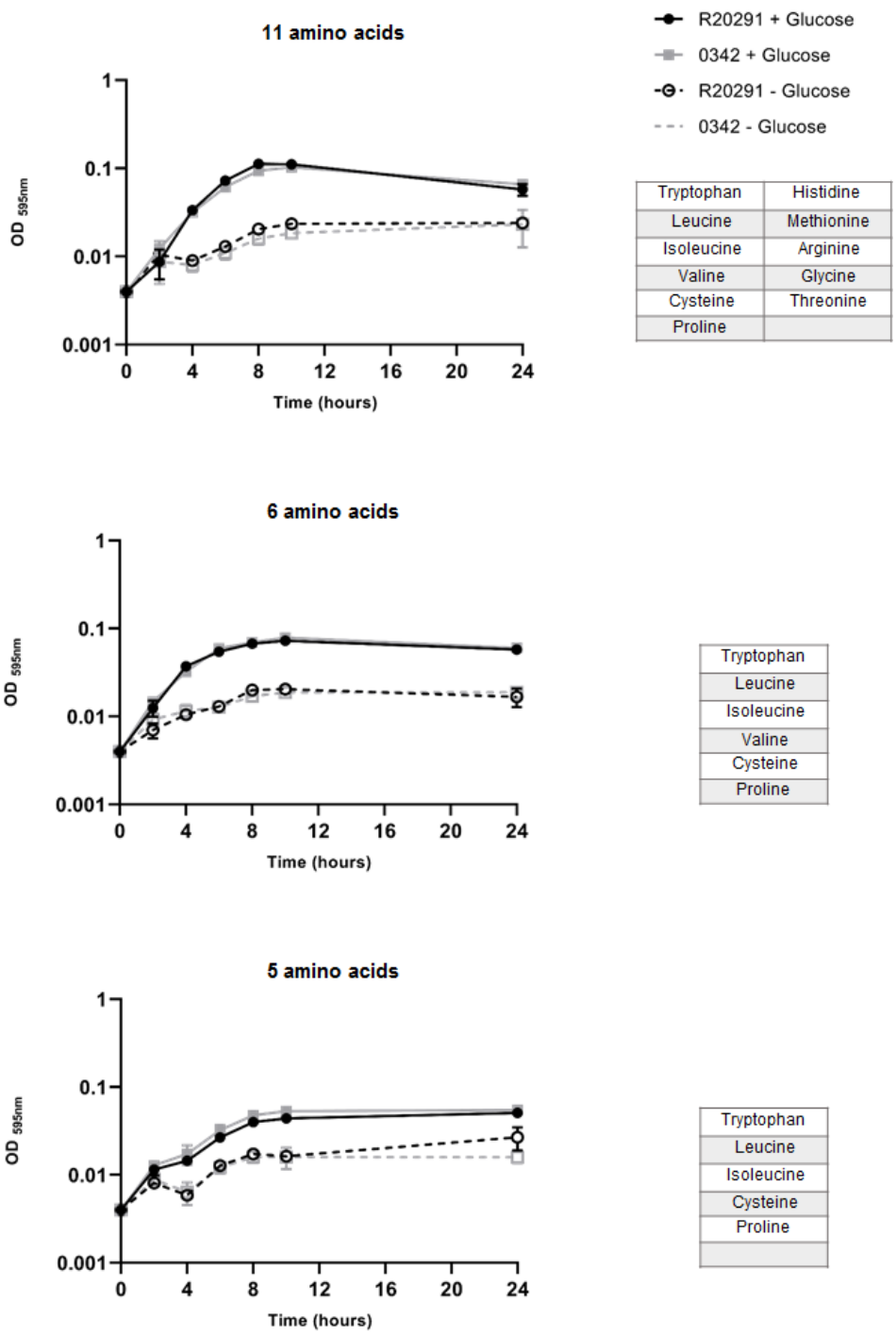


Figure 6.4. Growth kinetics of R20291 and R20291Δ*0342* with differing amino acid compositions. Three growth conditions were tested, with and without glucose, using 11, 6 or 5 amino acids, as described in the table next to each growth curve. Strains were incubated for 24 hours in defined media supplemented with the 11 amino acid composition and glucose, which was then used to inoculate the final growth conditions. 100 µl of culture was removed every hour for 8 hours and at 24 hours to take an OD_{595} readings using a plate reader. Cultures were set up in quadruplicate and two independent replicates were performed. Error bars represent standard deviation.

6.7. Discussion

This study aimed to characterise the role of the putative permease protein, CDR20291_0342, within *C. difficile*. This was identified in Chapter 3 as a potential vaccine candidate, as the protein array results revealed antibody responses to the protein were higher in healthy people compared to those with CDI. To investigate its function, the gene was successfully deleted from the chromosome of R20291 using allele exchange mutagenesis. As this removed the permease component within a putative ABC transporter, it was hypothesised the activity of the transporter would also be abolished.

The N-terminal portion of CDR20291_0342 was found to share high similarity with the N-terminal domains of LolE and MacB, components of export ABC transporters. The LolCDE transporter from *E. coli* is involved in sorting of lipoproteins to the outer membrane [550]. Lipoproteins can have a variety of potential roles in virulence, including mediating attachment to host cells. Indeed, the *C. difficile* lipoprotein CD0873 was found to be important for binding to Caco-2 cells *in vitro* [223, 555]. Lipoproteins harbour an N-terminal motif known as the lipobox [551]. No lipobox was identified for CDR20291_0340, therefore this line of enquiry was not pursued.

MacB transporters have a number of functions, but were first identified as efflux transporters of macrolide antibiotics [553]. They have since been shown to pump out a variety of antimicrobials, including colisitin and bacitracin [556, 557]. Antimicrobial resistance (AMR) rates within *C. difficile* can vary depending on geographical location and ribotype prevalence [249]. For example, hypervirulent strains within the 027 ribotype harbour a mutation in DNA gyrase, which confers high-level fluoroquinolone resistance [123]. Strain 630 is highly resistance to erythromycin via the presence of two *ermB* genes on the Tn*5398* transposon inserted on the chromosome [558-560]. These encode for methylases, which modify the macrolide binding site within the 23S rRNA, conferring resistance [561]. Interestingly, there are some high-level erythromycin resistance strains of *C. difficile*, such as R20291, which do not carry the *ermB* gene, and for which the origin of resistance has not been identified. This

suggests there are unidentified mechanisms of resistance mediating this phenotype [205]. As R20291_0342 shared similarity with the MacB transporter, the mutant and wild-type strains were investigated for sensitivity to the macrolide erythromycin and lincosamide lincomycin, however, no difference was found in MIC for either antibiotic. It is not uncommon for efflux pumps to increase the MIC to a particular antimicrobial rather than be the critical switch between susceptibility and resistance [562], and as erythromycin was resistant to the highest concentration tested (512 µg/ml) any smaller changes above that would not have been identified.

No difference was found in susceptibility of R20291 and R20291Δ*0342* to the common antimicrobial peptide bacitracin. AMPs are abundant within the gut and can interrupt the integrity of the cell membrane, therefore bacteria require defence mechanisms for survival [563]. ABC transporters are a very common mechanism of AMP resistance in Gram-positive species and *C. difficile* encodes the *cpr* and *cln* operons for resistance to nisin and galidermin, and cathelicidin, respectively [253, 564]. However, it is unlikely this is an exhaustive list of AMP resistance mechanisms within *C. difficile*, as inactivation of the *cpr* operon did not completely abolish resistance to its associated AMPs. R20291Δ*0342* is a putative efflux transporter and has potential links to efflux of antimicrobials. Although there was no difference in bacitracin sensitivity between the wild-type and mutant strains, some AMP-resistance mechanisms require pre-conditioning with their related AMP before their activity can be detected *in vitro*, as was the case for the Cpr transporter [564]. This would be an interestingly line of investigation, where gene expression could also be monitored and the repertoire of AMPs screened expanded to include other AMPs common to the gut environment.

Despite identification of conserved domains within CDR20291_0342, it may be that the orthologue is a product of similar physical functions i.e. efflux, and unrelated to the purpose of the transporter on which the protein was modelled. For example, Phyre^2 modelled CDR20291_0342 against LolE with 99.2% confidence, but this was only based on 21% of the protein. In these instances, where clues to

protein function are limited, high-throughput screening assays can be a valuable source of potential candidates for transport.

Biolog Phenotypic Microarrays have been employed for many bacterial and fungal species and have enabled elucidation of nutritional and chemical phenomes [565-567]. Conditions required for different cellular pathways can also be screened, as was demonstrated for *C. difficile* toxin production [568]. Disappointingly, our R20291 strain performed poorly in this assay and didn't grow in the majority of conditions tested. Previous work using the Biology Phenotypic Arrays to assess the growth of R20291 found this strain in fact had an expanded nutrient utilisation profile compared to other strains of *C. difficile*. This and other studies identified numerous conditions where R20291 was able to grow, which was not identified within our assays [38, 565]. The reason behind this poor growth is unknown. The other strains (mutants) also appeared to be growing weakly in these biological runs, with the exception of R20291_*2436*::CT which grew well in most conditions. On a technical note, it is my experience and others that the organism does not grow well in a 96-well plate format even in rich media, so this may have also influenced the Biolog results.

Despite this, there were some compounds worth following up on, including L-valine, L-valine plus aspartic acid and alpha-keto valeric acid which were unable to support growth of R20291Δ*0342*. As these all had the common component of L-valine, the growth of R20291 and R20291Δ*0342* with and without valine was assessed to determine whether the Biolog results were a true phenotype. Previous studies have found *C. difficile* is auxotrophic for valine, along with five other amino acids (proline, cysteine, leucine, isoleucine and tryptophan) [32]. However, until now, this hasn't been assessed for R20291. *C. difficile* strain R20291 was able to grow in minimal media with the 5 amino acids and the biggest impact on growth resulted from resulting from removal of glucose, rather than L-valine. Furthermore, no difference between the mutant and wild-type strain was found in any of the conditions tested. This indicates that the results from the Biolog were most likely due to inter-plate variation, rather than an observable phenotype.

Valine is a branched chain amino acid (BCAA) that is an important donor for Stickland fermentation, the major source of energy production within *C. difficile*. Both the valine synthesis pathway and BCAA transporter, BrnQ are regulated by CodY, a global regulator within *C. difficile*, which can bind valine directly to suppress transcription of the CodY regulon [507]. Considering the synthesis pathways available for valine in *C. difficile* and the multiple BCAAs that can be used in Stickland fermentation, it is slightly surprising that valine is an essential requirement in minimal media. Our results suggest valine is in fact not essential for growth of R20291. It may be that carry over from the overnight conditions provided adequate valine to support growth, although this is likely to only sustain initial growth (i.e. 0-2 hours) and be insufficient for the full growth curve obtained. Even so, future work could look at removal of any residual media with washing of the pellet or further dilution of the primary cultures before inoculation of the test media. Pellet washing was not attempted in these experiments, as when trialled previously with minimal media experiments, growth was very slow and the bacteria did not demonstrate the typical lag, exponential and stationary growth phases, even in conditions that could support normal growth.

As with the other proteins undergoing phenotypic characterisation, no complements were constructed for CDR20291_0342, although here there was less of a requirement as the phenotype of the mutant did not differ from wild-type in the conditions tested. The difficulties in working with minimal media and carryover were similar to those faced with the CDR20291_0330 growth experiments. The contribution of media carryover is important to identify in these experiments to see whether valine is truly unnecessary for growth or if the bacteria were relying on the small concentration likely to still be in the media. It could also be interesting to extend the growth time to see if there are any downstream differences in growth rate, as was seen for 630 when grown in ethanolamine. If it were possible to obtain gut filtrate, the growth and antimicrobial sensitivity assays could be repeated in these conditions to determine whether an environment more similar to the *in vivo* conditions highlights any differences. In relation to further characterisation of CDR20291_0342, as the role of this protein and its associated transporter within *C. difficile* remains unknown, future

experiments should focus on large scale screening methodologies. For example, transcriptomics comparing the mutant to wild-type R20291 could determine whether gene expression is altered when *CDR20291_0342* is inactivated. Any patterns of differential expression obtained could then inform functions related to CDR20291_0342. Alternatively, adaptive transcriptomics could be used to monitor the gene expression of *CDR20291_0342* under different conditions and/or over a period of time. For example, adaptive transcriptomics in a mouse model was used to investigate the virulence of an epidemic ribotype 027 strain [35].

CDR20291_0342 was highlighted as a potential vaccine candidate as a higher sera IgG response in healthy controls compared to CDI patients was found in both the pan-protein array and using the purified protein in an ELISA (albeit not significantly higher for the ELISA). CDR20291_0342 was conserved across a number of clinically relevant strains except the ribotype 023 CD305. Although this could reduce vaccine coverage, it is unlikely that *C. difficile* vaccines of the future will be single component vaccines and instead will constitute multiple immunogenic components include anti-colonisation factors and Toxins A and B. From a basic accessibility perspective, ABC transporters are encoded within the cell membrane, often with large extra cellular portions or exposed substrate binding components, which are therefore available for recognition by the immune system. Furthermore, a number of ABC transporters have been identified as immunogenic and assessed in vaccine trials [547]. This includes vaccines against *Streptococcus pneumoniae*, *Mycobacterium tuberculosis*, *Moraxella catarrhalis* and *Yersinia pestis* [547, 569, 570]. In relation to *C. difficile*, a study of immunoreactive cell surface proteins included an ABC transporter that was recognised by human sera [571], but no ABC transporter antigens have been evaluated in clinical trials, suggesting this is a potentially exciting and novel area.

7. Expressing the *C. difficile* flagella glycan in *E. coli*

7.1. Introduction

C. difficile vaccine development to date has primarily relied on the immunogenic properties of Toxins A and B, with all vaccines currently in clinical trials based on these components [375]. However, these trials have not led to the production of a licenced vaccine, therefore alternative approaches to *C. difficile* vaccine design is a current imperative. As described in Chapter 1 (section 1.6.3), a number of *C. difficile* antigens have been investigated as vaccine candidates, including spore coat proteins CdeM and CdeC, cell surface proteins SlpA and Cwp84 and the surface polysaccharides PS-I, PS-II and PS-III [235, 419, 421, 425]. The latter have been tested as glycoconjugates, a very effective vaccine design whereby polysaccharides are conjugated to proteins to stimulate a T-cell dependent immune response. leading to immunological memory, a process not possible with polysaccharide alone [324]. *C. difficile* PS-II has been chemically conjugated to the diphtheria toxin CRM_{197}, the enterotoxigenic *E. coli* enterotoxin B subunit, and portions of Toxins A and B from *C. difficile*. Antibodies specific to PS-II are raised in mice, but the protective efficacy of this construct is yet to be demonstrated in challenge studies [234, 235, 238].

The glycoconjugates described above were synthesised chemically, a multistep process requiring production and purification of the protein and glycan components separately, before they can be coupled together and then purified [572]. In more recent years, the development of bioconjugation has provided an alternative and in some ways, superior means of synthesising glycoconjugates, by exploiting the innate capacity of bacteria to build and conjugate complex glycans [341, 365, 367, 368]. Bacterial glycosylation systems can differ in a number of ways, including by which residues within the acceptor protein the glycan is attached to, either asparagine (*N*-linked) or serine/threonine (*O*-linked). Furthermore, these differ by where the glycan is synthesised, which can be directly on to the acceptor protein or first built on a lipid linker (e.g. UndPP) within the inner membrane before transfer to the acceptor protein by an oligosaccharyltransferase (OST) [335]. Transfer of bacterial glycosylation

systems into *E. coli* cells can result in an *in vivo* "vaccine factory", as demonstrated at the basic research level, but also advanced to such a stage that vaccines prepared using this method are now in clinical trials [332, 370-372]. Bioconjugation offers a number of advantages over chemical conjugation, including ease and safety of working with *E. coli* cells over the native organism, improved isolation of fully synthesised, intact conjugates and fewer rounds of purification, meaning reduced cost of manufacture [333].

Although the term bioconjugation covers a number of methodologies, arguably the best characterised is the use of the OST PglB from the *Campylobacter jejuni N*-linked glycosylation pathway, in *E. coli*. The *N*-linked glycan from *C. jejuni* is first synthesised onto the UndPP lipid linker, on the cytoplasmic leaflet of the inner membrane. This is then translocated by a flippase enzyme to the periplasm. Here, it is recognised by the OST PglB, which attaches the glycan to the D-X-N-Y-S/T recognition sequence within a given acceptor protein [341, 348]. PglB can be expressed in *E. coli* and used to transfer a diverse range of glycan substrates to acceptor proteins [430]. In its simplest form, this can be achieved with co-expression of the acceptor protein, OST PglB and biosynthesis locus of the desired glycan in a suitable host cell, such as *E. coli* (Figure 7.1).

Use of bioconjugation to build a *C. difficile* specific glycoconjugate offers many advantages over chemical conjugation [332, 333]. For example, a multi-component vaccine could be designed using an immunogenic surface antigen, such as one or more of those identified by the protein array, fused to domains of Toxins A and B conjugated to a *C. difficile* specific glycan. Previous work has demonstrated that recombinant Toxin B fragments chemically conjugated to *C. difficile* PS-II result in a Toxin B and PS-II specific antibody response in a mouse model [238].

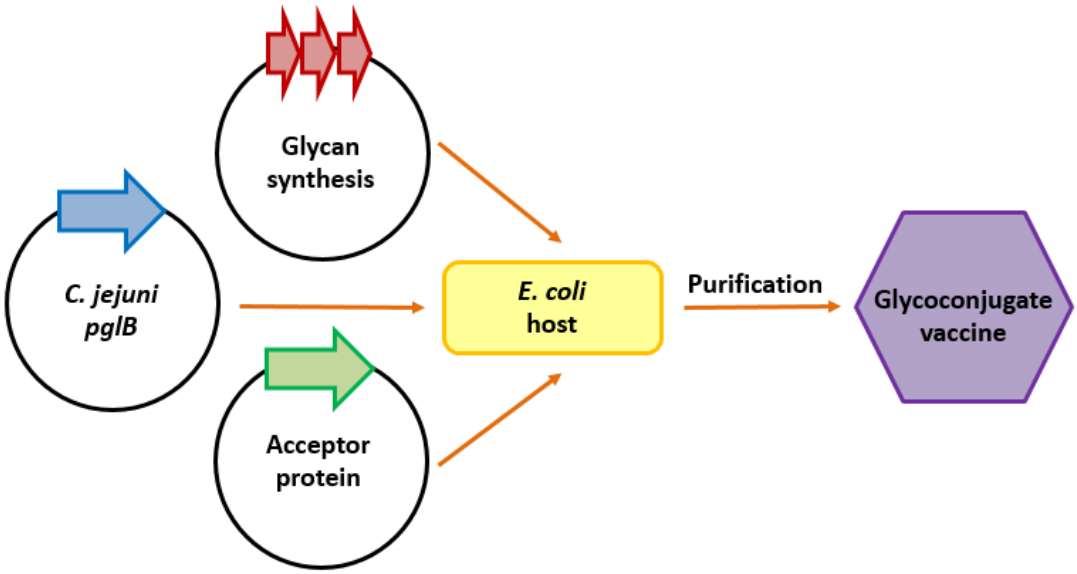


Figure 7.1. Assembly of glycosylation machinery in *E. coli*. The OST PglB from *C. jejuni* can be expressed in *E. coli* and used to attach a broad range of glycans onto a specific acceptor protein. This can be achieved by plasmid-mediated expression of each of the three required components; PglB, the acceptor protein and the machinery for bulding the glycan, in a suitable host cell such as *E. coli.*

The aim of this work was to develop the use of bioconjugation, namely the *C. jejuni* OST PglB in *E. coli*, to build a *C. difficile*-specific glycoconjugate.This would provide a novel mechanism for the synthesis of glycoconjugate vaccines for *C. difficile*. Furthermore, this would represent the first example of the functional transfer of *C. difficile* glycosylation enzymes into *E. coli* by combining *O*-linked and *N*-linked glycosylation systems to build glycan structures, which represents an important contribution to the bioconjugation toolbox.

7.2. Using bioconjugation to express the *C. difficile* flagella glycan

7.2.1. Glycan selection

Interest in the *C. difficile* glycome has increased in recent years and consequently so has understanding of the different sugars within and on the cell surface (Introduction, section 1.3.4.2). This review was used to inform glycan selection for inclusion within the *C. difficile*-specific glycoconjugate in this study. It was essential that the native locus encoding glycan synthesis had been identified in *C. difficile* and ideally, characterised, in order to inform the role of enzyme(s) required for glycan synthesis. This then informed which genes were required for transfer into *E. coli*. Structural information about the *C. difficile* glycan was also essential to know what to screen for.

The glycans decorating the *C. difficile* flagella have been resolved by mass spectrometry (MS) and nucleic magnetic resonance (NMR) and their glycosylation loci characterised in detail for ribotype 012 (strain 630) and ribotype 027 (strain R20291) [184, 185, 187, 190]. A flagella glycan was therefore deemed the most appropriate substrate for this proof of concept study. The flagella glycans are *O*-linked, meaning they are attached to serine and threonine residues within FliC. The glycan structures have been resolved by NMR, and are shown in Figure 7.2. The flagella from strain 630 is decorated with a *N*-acetyl-glucosamine (GlcNac) joined to a methylated threonine via a phosphodiester bond [184]. The flagella glycan from ribotype 027 is more intricate, with a GlcNac starting sugar followed by a methylated di-rhamnose and a terminal sulfonated peptidylamido glycan moiety [190]. As proof of principle, the first three sugars of the 027 strain glycan (Rha-2-α-Rha3OMe-3-β-GlcNac-) were selected for transfer into *E. coli* and glycosylation of an acceptor protein. These precursor sugars (GlcNac and rhamnose) should all be readily available in *E. coli* and the glycan harbours an acetoamide group at the C2 position of the reducing end sugar (GlcNac), a structural requirement of PglB [364]. It was decided that to begin with the whole glycan wouldn't be transferred due to the complexity of the terminal peptidylamido glycan moiety. Furthermore, the

addition of individual enzymes rather than the whole flagella glycosylation locus from *C. difficile* enables greater control and insight during cloning, optimisation of engineering and trouble shooting.

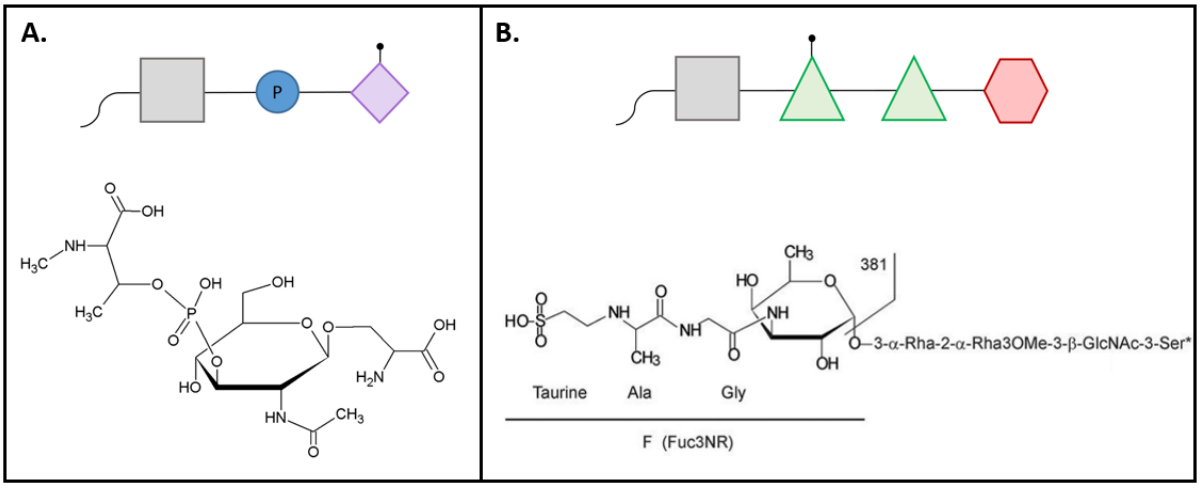


Figure 7.2. The structure of *C. difficile* flagella glycans. The flagella glycan structure as resolved by nuclear magnetic resonance (NMR) imaging for (A) strain 630, from Faulds-Pain *et al.* [184] and (B) strain R20291, from Bouche *et al.* [190]. Above each glycan structure is a simplified schematic of the glycan organisation, modified from an image by Dr Alexandra Faulds-Pain; grey box- GlcNac, green triangle- rhamnose, red hexagon- peptidylamido glycan moiety, blue circle- phosphate group, purple diamond- threonine, black dot- methyl group.

The flagella glycosylation locus begins with *fliC*, encoding the structural unit of the flagella filament, followed immediately downstream by the enzymes required for glycosylation [113, 184, 185] (Figure 7.3). Glycosyltransferase 1 (GT1) adds the starting GlcNac sugar directly onto serine and threonine residues within FliC, followed by addition and methylation of the di-rhamnose by glycosyltransferase 2 (GT2), encoded by *CDR20291_0242*. This enzyme is thought to be responsible for both glycosylation and methylation of the rhamnose sugars and harbours putative active sites for both functions present within its structure [187]. Only GT2 was transferred into *E. coli* for expression of the glycan, as described in the next section.

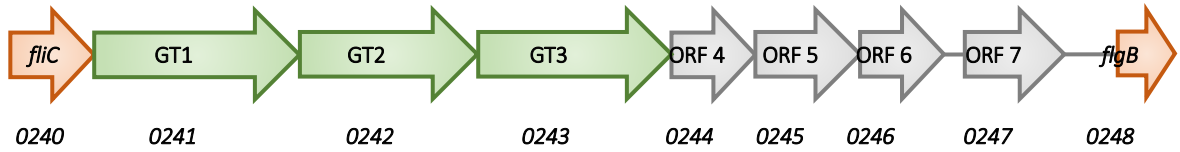


Figure 7.3. The flagella glycosylation locus from R20291. The organisation of the flagella glycosylation locus including; the flagella filament gene *fliC*, the glycosyltransferases GT1 and GT2, responsible for addition of the GlcNac and di-rhamnose onto FliC, respectively, an additional glycosyltransferase (GT3) and two open reading frames (ORF 5 and ORF 6) required for synthesis of the full glycan structure and *flgB*, which encodes the putative flagella basal body rod gene. Below is the numbering of the genes in *C. difficile* R20291 [113, 184, 187].

7.2.2. Adaptation of *C. jejuni N*-linked glycosylation

The production of glycosylated protein using PglB involves the following three stages:

1. Synthesis of the glycan in the cytoplasm on an undecaprenol pyrophosphate (UndPP) linker within the inner membrane
2. Transfer of the synthesised glycan into the periplasm by a membrane bound flippase
3. Periplasmic transfer of the glycan onto an acceptor protein by PglB

Glycoconjugates previously synthesised using PglB and its associated *N*-linked glycosylation pathway tend to include glycans originating from organisms with complementary pathways of synthesis, facilitating their transfer into *E. coli*. For example, *Francisella tularensis* already encodes enzymes for synthesis of the glycan onto UndPP and for its transfer into the periplasm, enabling successful recombinant glycoconjugate vaccine synthesis [365, 573]. *C. difficile* is a Gram-positive organism, which lacks a periplasm, and therefore encodes neither of the enzymes described above (those mediating synthesis of the glycan onto UndPP and its transfer into the periplasm). Instead it is proposed to follow other mechanisms of flagella glycosylation and synthesise the glycan in the cytoplasm directly onto serine/threonine residues (*O*-linked) within FliC.

In this proof of concept study, elements of the *E. coli* O-antigen locus were hijacked for synthesis of the *C. difficile* flagella glycan. The *E. coli* strain CLM24 does not build its native O-antigen (O16) (Figure 7.4) due to insertional inactivation of *wbbL* (a gene necessary for synthesis) with an IS5 element [574].

However, as the other genes in the loci are still present, they could be reappropriated. The first step requires the synthesis of the glycan onto UndPP in the inner membrane. As the O16 antigen and flagella glycan share the same starting sugar, GlcNac, the enzyme responsible for its placement on UndPP in *E. coli*, WecA, was selected as the starting enzyme for *C. difficile* flagella glycan synthesis. This replaces the native *C. difficile* enzyme, CDR20291_0241 (GT1), which adds GlcNac directly onto the flagellin subunit, FliC. Once UndPP-linked GlcNac was available, GT2 from *C. difficile* could be expressed for addition of the di-rhamnose. In relation to the transfer of the flagella glycan into the periplasm, *E. coli* encodes its own flippase, Wzx. The promiscuity of this flippase is under debate, but as both the *E. coli* O16 antigen and *C. difficile* flagella glycan begin with a GlcNac linked to L-rhamnose, it was proposed this level of similarity should be permissive for transport.

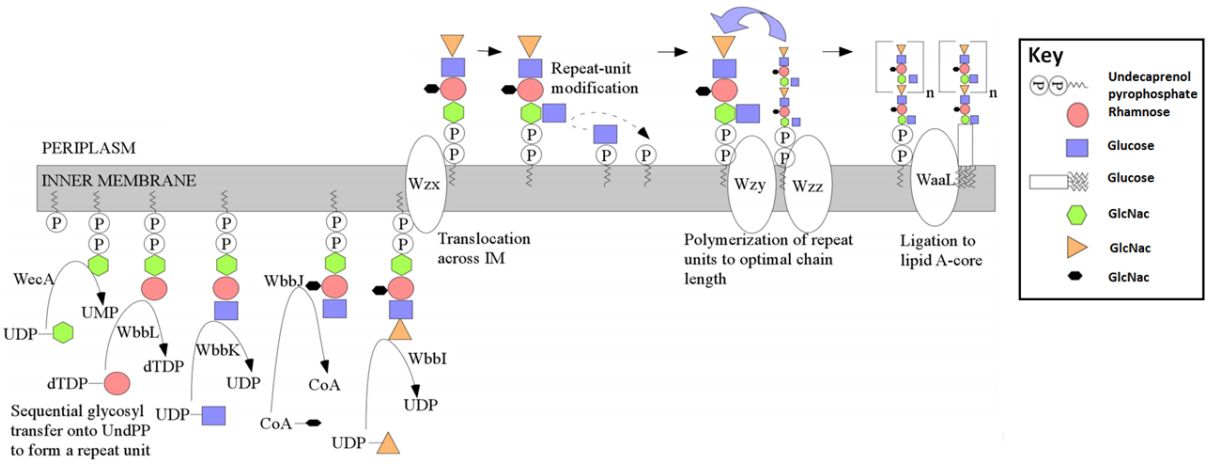


Figure 7.4. Biosynthesis pathway of the *E. coli* O16 antigen. Pathway for synthesis of the *E. coli* O16 antigen. This pathway is harboured by K-12 strains of *E. coli* (standard laboratory strains) but these do not produce their O-antigen due to a mutation in *wbbL*. Glycosyltransferases sequentially add individual sugars to UndPP within the inner membrane. The glycan is then transferred into the periplasm (Wzx), polymerised (Wzy), then ligated to Lipid A on the surface of the cell (WaaL). Image from Hong *et al.* [575].

Finally, the *C. difficile* flagella glycan is *O*-linked, meaning it is built directly onto serine and threonine residues within FliC, whereas PglB is an *N*-linked system and attaches glycans to asparagine residues within a D-X-N-Y-S/T recognition sequon. The *C. difficile* flagella glycan contains a GlcNac starting sugar which should be permissible for PglB, therefore the glycan should be transferred as long as a suitable

acceptable protein with a D-X-N-Y-S/T recognition site is provided (section 7.2.3). The concept of different enzymes working cooperatively to synthesise a glycan structure has been reported previously, as when the *C. jejuni* heptasaccahride was expressed in a WecA-positive *E. coli*, structures were identified beginning with both bacillosamine, the native starting sugar, and GlcNac, due to the activity of WecA [576]. Figure 7.5 demonstrates the proposed bioconjugation pathway.

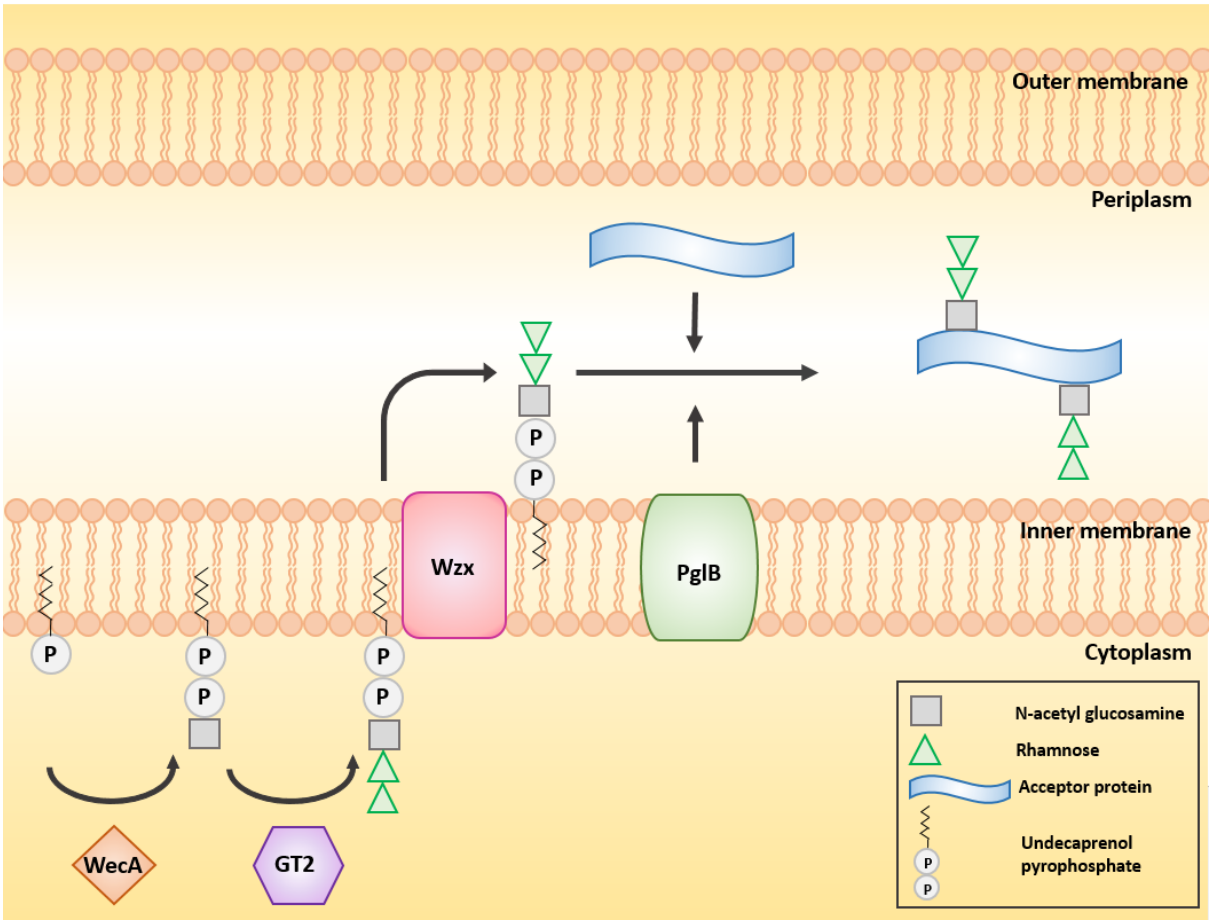


Figure 7.5. Proposed pathway for the expression of the *C. difficile* flagella glycan in *E. coli*. The *E. coli* enzyme, WecA, adds GlcNac onto UndPP then the *C. difficile* GT2 (CDR20291_0242) builds the methylated di-rhamnose structure onto this. The *E. coli* flippase transfers the glycan into the periplasm where it can be recognised by *C. jejuni* PglB and conjugated to the D-X-N-Y-S/T recognition sequon within the acceptor protein.

7.2.3. Acceptor protein design

Once the desired glycan for synthesis was identified, it was necessary to select a suitable acceptor protein to receive the flagella glycan. The estimated size of the glycan when fully synthesised (Rha-2-α-Rha3OMe-3-β-GlcNac-) is approximately 0.4 KDa, which would result in a minor change to the protein molecular weight and therefore difficult to resolve by SDS-PAGE. Gel-based assessment of changes to protein molecular weight is a useful indicator of whether glycosylation has occurred, providing a quick and easy screening method during optimisation. This is especially important considering the lack of antibodies and lectins to this structure and the more time consuming and expensive process of MS, using which isn't feasible for assessing many different strains and growth conditions.

To account for the anticipated minor change to protein mass upon glycosylation, a short reporter protein with multiple glycosylation sequons was deemed the ideal candidate for glycosylation. Minor size changes are easier to distinguish on smaller proteins and increasing the number of potential glycosylation sites, which should increase the number of glycans occupying the protein and therefore increase the total shift molecular weight (i.e. if all 7 potential sites were occupied with the 0.4 KDa glycan this would equate to a total 2.8 KDa size difference). At this stage the immunogenic properties of the protein were not a priority, as the aim of the work was to demonstrate functional transfer of *C. difficile* GT2 into *E. coli* and synthesis of the Rha-2-α-Rha3OMe-3-β-GlcNac- structure.

The *C. jejuni* periplasmic lipoprotein, AcrA, is a well-studied and routinely utilised carrier protein in bioconjugation, with multiple glycosylation sequons [341, 430, 577]. At ~40 KDa, its size made it unsuitable for this work, as very small changes to protein mass by glycosylation with the flagella glycan would not be visible by SDS-PAGE. Therefore a truncated ~13 KDa fragment of the protein harbouring a single native glycosylation site was designed for use in this study. This fragment was derived from a publication identifying the smallest domain of AcrA that could be glycosylated by PglB, using different lengths of the coiled coil domain still flanked by the lipoyl region [348]. In addition to the native

glycosylation site, the authors engineered two more glycosylation sequons into the coiled coil domain, by changing one residue per sequon at F115D and T145D. All three sites were demonstrated to be glycosylated with the *C. jejuni* heptasaccharide by PglB. Therefore the region containing all these and the flanking lipoyl domain was selected as the acceptor. A fourth native glycosylation sequon also present in the sequence was excluded as its distance from other sequons meant accommodating it would result in a significant increase in protein size (Figure 7.6a). To further improve the opportunities for glycosylation, a "glycotag" was added to the C-terminus consisting of 4 glycosylation sites, including the optimal sequon DQNAT, separated by glycine-arginine-glycine linkers [578]. Glycotags are commonly employed in bioconjugation to improve glycosylation of the protein, particularly in the absence of native or easily engineered internal glycosylation sequons [576]. Finally, an N-terminal signal sequence from the *E. coli* DsbA protein was included for localisation to the periplasm where PglB functions and a C-terminal 6xHistag for purification (Figure 7.6b and c) [579, 580].

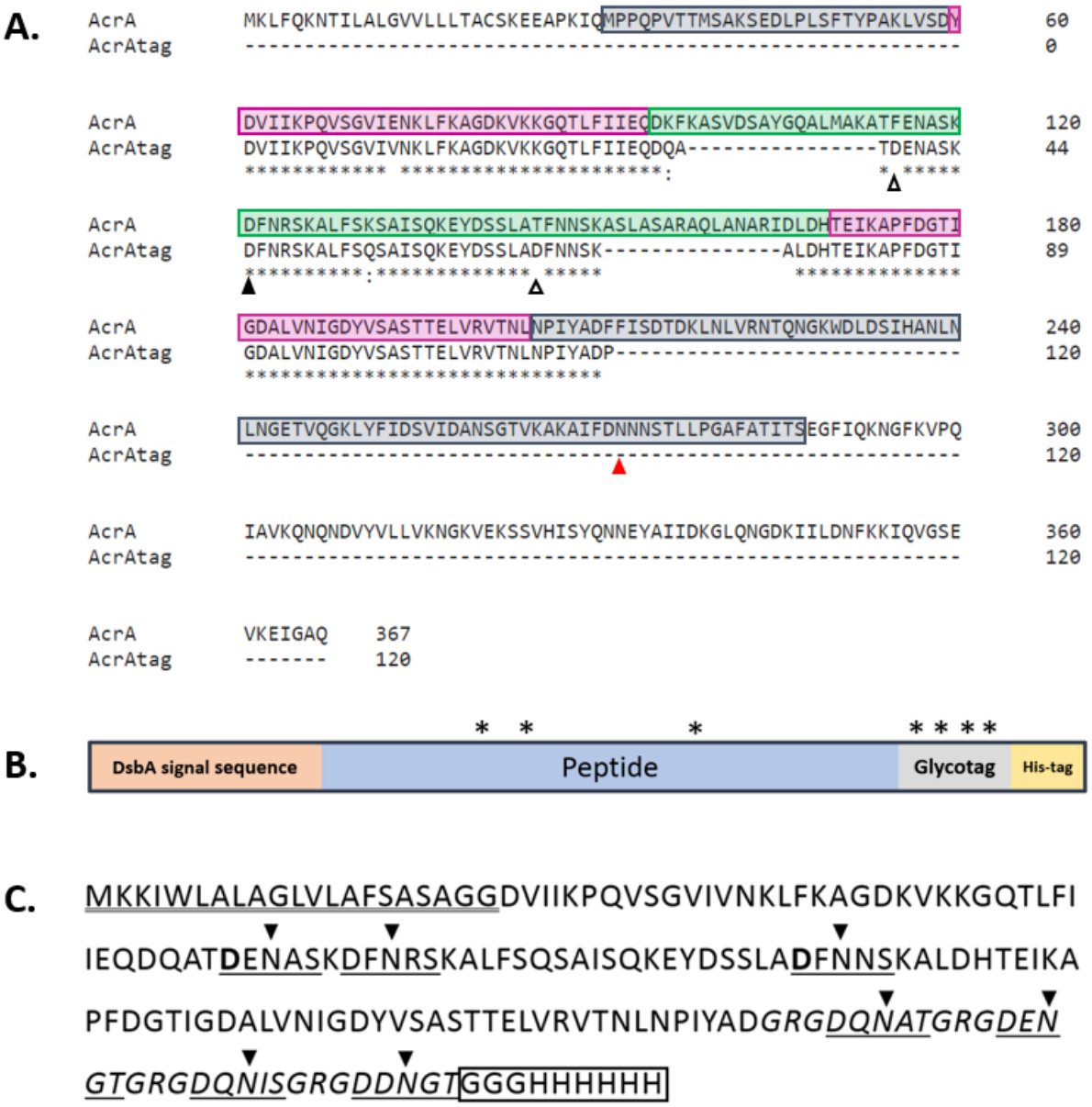


Figure 7.6. Design and sequence of the acceptor protein, AcrAtag. A short acceptor protein was designed for use in the glycosylation experiments, comprising a fragment of the *C. jejuni* periplasmic lipoprotein, AcrA with a four sequon glycotag. (A) Sequence alignment of full length AcrA and AcrAtag. The coloured regions represent the different protein domains; β barrel sandwich hybrid (grey), lipoyl domain (pink) and coiled coil domain (green) [348]. Arrows denote *N*-glycosylation sequons, either native (filled black), native but not used for AcrAtag (filled red) or engineered (empty). : defines residues mutated to reduce proteolytic cleavage. (B) Schematic of AcrAtag with N-terminal DsbA signal sequence for localisation to the periplasm, where PglB functions, the peptide sequence containing one native and two engineered glycosylation sequons (asterisks), the glycotag comprising four glycosylation sequons and a C-terminal 6xHis tag for nickel affinity purification. (C) Full sequence of AcrAtag; DsbA signal sequence (double underlined), *N*-glycosylation sequon (single underlined), glycosylated asparagine residue (arrow), mutated residue to generate another glycosylation sequon (bold), glycotag (italics), 6xHis tag (box).

7.3. Glycosylation trials using a multi-vector system

7.3.1. System design

Initial glycosylation trials utilised a multi plasmid system for expression of the different components of the glycosylation pathway, including: the acceptor protein AcrAtag, the glycosyltransferase GT2, and the OST, PglB (Figure 7.1). This approach has previously been used to build glycoconjugates for *C. jejuni*, *F. tularensis*, *Staphylococcus aureus* and *Burkholderia pseudomallei* [365, 368, 369, 576]. *E. coli* strain CLM24 was selected as the expression host and is derived from the K-12 strain W3110, but lacks a functional WaaL ligase. WaaL traffics glycans from the periplasmic UndPP to Lipid A on the cell surface. Deletion of *waaL* results in accumulation of UndPP-linked glycan in the periplasm [430]. The CLM24 strain used also harboured transposon-integrated *pglB* on the chromosome (CLM24 *cedA*::*pglB*, referred to here as CLM24*pglB*, kindly provided by Dr Jon Cuccui, LSHTM, London, UK), under IPTG inducible expression. Chromosomal integration of *pglB* removes the need for a third plasmid in the system which reduces the burden of plasmid incompatibility when expressing multiple constructs within one cell and enables a wider variety of plasmids to be trialled for expression of GT2 and AcrAtag.

The pEXT series of plasmids were selected for expression of GT2 and AcrAtag (Figure 7.7). pEXT20, 21 and 22 all express from a P_{tac} promoter, but differ in origin of replication and resistance cassette, meaning they can be stably maintained within the same cell [581]. They range in copy number from high (pEXT20) to low (pEXT21 and pEXT22), so plasmid selection can influence the concentration of the desired construct within the cell. As the ideal ratio of glycan to protein was not yet known, AcrAtag and GT2 were cloned into both pEXT20 and pEXT21, offering maximum flexibility during optimisation of protein expression and glycosylation. pEXT22 was not taken forward as both this plasmid and the CLM24*pglB* strain are kanamycin resistant.

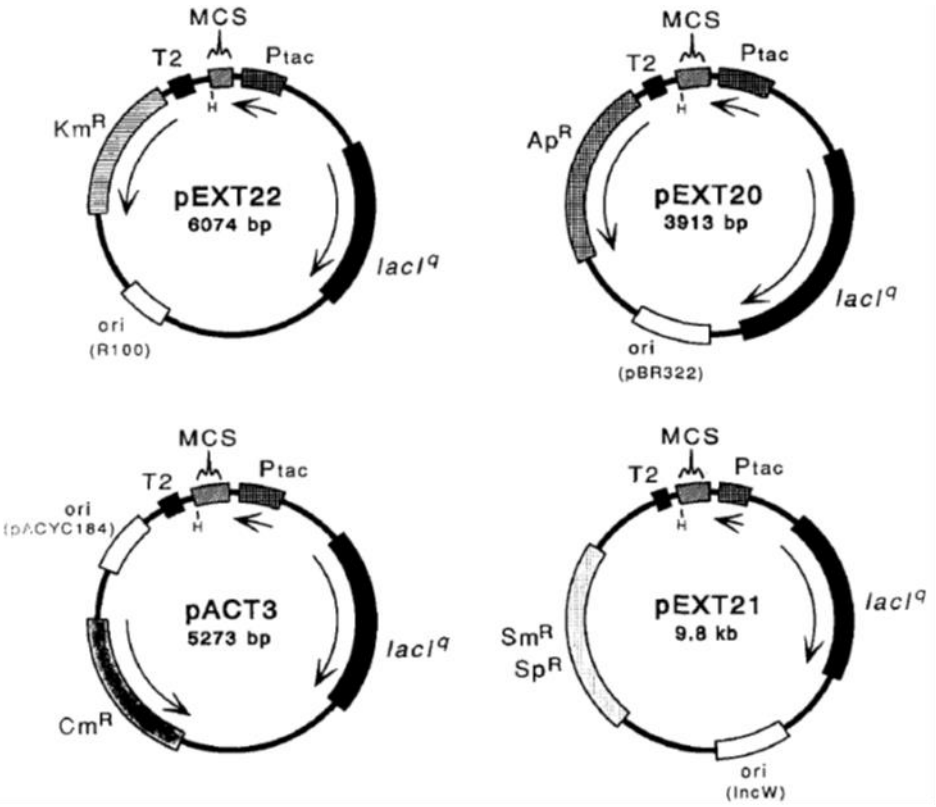


Figure 7.7. Maps of the pEXT series of plasmids. Plasmids pEXT20 and pEXT21 were used for expression of the glycosyltransferase GT2 and acceptor protein AcrAtag. Expression is driven from a Ptac promoter in all plasmids but they differ by origin of replication and antibiotic resistance cassette, meaning they can co-express within the same cell. Image taken from Dykxhoorn *et al.* [581].

7.3.2. Confirming glycosylation of IPTG-inducible AcrAtag

The gene encoding GT2, *CDR20291_0242*, was amplified from *C. difficile* genomic DNA and ligated into pEXT20 and pEXT21. The acceptor protein, AcrAtag, was synthesised as a G-block gene fragment (section 7.2.3) and ligated into pEXT20 and pEXT21. Correct construction of all plasmids was confirmed with test digestion and sequencing.

The AcrAtag protein harbours a C-terminal 6xHistag, which was used to confirm protein expression. GT2 was not constructed with a 6xHistag due to concerns of the tag interfering with enzyme function. CLM24*pglB* cells harbouring the pEXT20_AcrAtag construct were induced overnight at 37^{o}C with 1 mM IPTG, harvested cells lysed in a ribolyser then the soluble protein fraction purified using nickel affinity

resin and visualised on an anti-His immunoblot (Figure 7.8). The predicted molecular weight of AcrAtag with a 6xHistag, but without the DsbA signal sequence (which is cleaved once the protein is localised to the periplasm) is ~17.27 KDa. The anti-His immunoblot detected a protein in the induced strain carrying pEXT20_AcrAtag, which was within the expected size range and absent from both the uninduced control and the induction of empty pEXT20, confirming expression of AcrAtag.

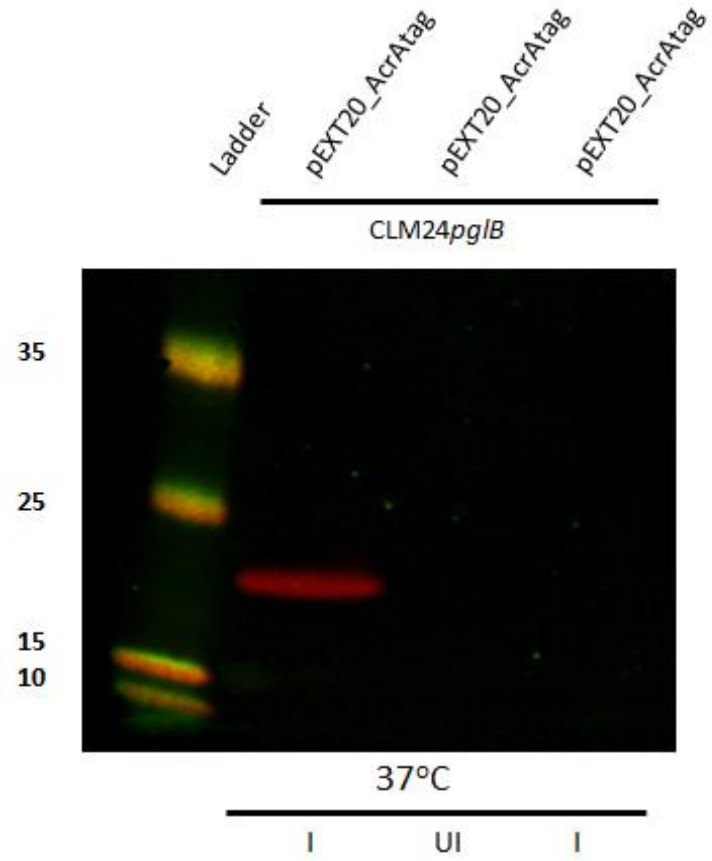


Figure 7.8. Expression of the acceptor protein, AcrAtag. CLM24*pglB cells* harbouring pEXT20_AcrAtag or empty pEXT20 only were incubated overnight at 37°C with 1 mM IPTG and the nickel affinity purified soluble fraction was probed for the ~17.27 KDa AcrAtag on an anti-His immunoblot (red). I, induced, UI, uninduced.

The functional transfer of the *pgl* locus encoding the *C. jejuni* heptasaccharide synthesis machinery, into *E. coli* and glycosylation of a chosen acceptor protein has been demonstrated previously [341]. This can be utilised as a useful indicator as to whether the protein of interest can be glycosylated and how many sequons are occupied, which in turn indicates the maximum size change to be expected when testing the *C. difficile* glycan. It also confirms the protein is being localised to the periplasm

where glycosylation occurs and is a permissible substrate for PglB. The *pgl* glycosylation locus encoded on pACYC (pACYC*pgl*Δ*pglB*) contains the complete machinery required for synthesis of the *C. jejuni* heptasaccharide (*pglB* is chromosomally integrated), including glycosyltransferases to build the glycan onto UndPP and a flippase for transport across the inner membrane.

To confirm AcrAtag could be glycosylated, CLM24*pglB* or CLM24 cells carrying pEXT20_AcrAtag and the constitutively expressed pACYC*pgl*Δ*pglB* was induced overnight at 37^{o}C or 30^{o}C with 1 mM IPTG. Cells were harvested and lysed and proteins purified by nickel affinity chromatography then resolved by SDS-PAGE and probed by Western blot with anti-His antibody and SBA lectin. The 6XHistagged AcrAtag and galactose residues within the *C. jejuni* heptasaccharide were detected by anti-His antibody and the soy bean agglutinin (SBA) lectin, respectively (Figure 7.9 panel 1 and 2). It was hypothesised that when expressing AcrAtag in the presence of PglB and the *C. jejuni* heptsaccharide, a ladder of bands would be produced relating to occupation of each glycosylation site with glycan. Indeed, when expressed together, a laddering pattern of bands is produced, corresponding to unglycosylated material at the bottom, followed by an increasing number of glycosylation sites occupied by the ~1.2 KDa heptasaccharide, up to a maximum of 6, one fewer than the total number available. Only unglycosylated protein is present when the same constructs are expressed in CLM24 cells without *pglB* on the chromosome and glycosylation appears to be more pronounced at 30^{o}C over 37^{o}C (Figure 7.9).

The uninduced control only relates to AcrAtag, as the *C. jejuni* glycosylation locus is under constitutive expression. Here, the banding pattern of glycosylation can still be detected, which is attributed to leaky expression of the acceptor protein, AcrAtag, as expression of the *C. jejuni* glycan locus alone does not result in any banding pattern (Figure 7.9, panel 3, lane 5). In the uninduced control, no unglycosylated protein was detectable by the anti-His antibody, suggesting less protein is favourable for glycosylation and results in a more efficient reaction, as the higher bands of glycosylation are comparable to induced conditions. This could also be due to the higher sensitivity of the SBA lectin

compared to the anti-His antibody (Figure 7.9). On a repeat expression test of the *C. jejuni* glycan and AcrAtag (Figure 7.9, panel 4), a banding pattern was produced which corresponds to all seven glycosylation sites being occupied, confirming all are available for occupancy and AcrAtag is an appropriate acceptor protein for use with PglB.

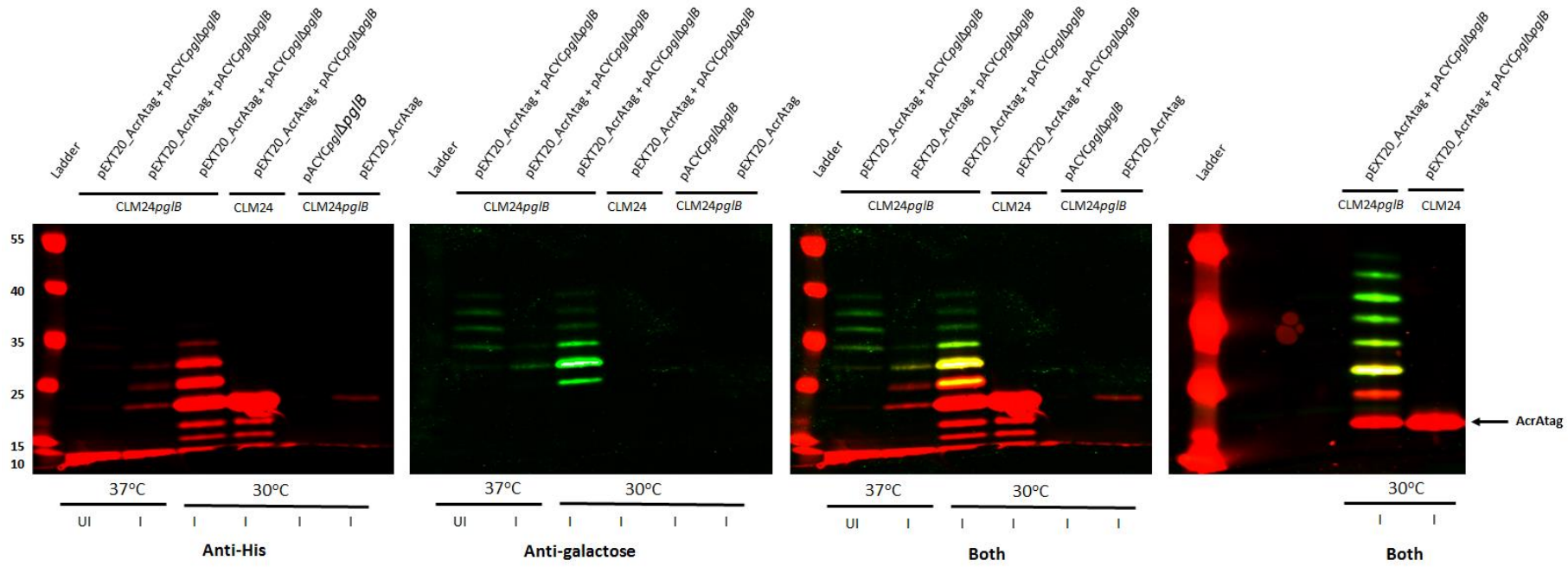


Figure 7.9. Glycosylation of IPTG-inducible AcrAtag with the *C. jejuni* heptasaccharide. To confirm AcrAtag could be glycosylated, the constitutive pACYC*pgl*Δ*pglB* construct, encoding the *C. jejuni* heptasaccharide locus was co-expressed with pEXT20_AcrAtag in CLM24*pglB and* CLM24 (no *pglB*) cells. These were incubated overnight at 37°C and 30°C, then induced with 1 mM IPTG: I, induced, UI, uninduced. AcrAtag was extracted and nickel affinity purified then detected by immunoblot using anti-His antibody to AcrAtag and soy bean agglutinin (SBA) lectin to the *C. jejuni* glycan. The first three panels present the immunoblot images with anti-His only (red), anti-glycan (anti-galactose) only (green) then the two overlaid. The last panel displays a repeat expression test.

7.3.3. Screening for glycosylation with GT2

After confirming AcrAtag is a functional acceptor protein for glycosylation by PglB, AcrAtag was co-expressed with GT2 from *C. difficle*, from the pEXT21, to screen for glycosylation with the truncated *C. difficile* flagella glycan. CLM24*pglB* or CLM24 cells expressing AcrAtag (pEXT20) and the *C. difficile* GT2 (pEXT21) were induced overnight with 1 mM IPTG at 37°C or 30°C then processed as described above. Purified protein was probed with anti-His antibody only, there is no specific antibody against the *C. difficle* glycan (Rha-2-α-Rha3OMe-3-β-GlcNac). As can be seen in Figure 7.10, there is no discernible difference between the molecular weight of AcrAtag when expressed with GT2 in the presence or absence of PglB, suggesting glycosylation had not occurred. Indeed it is possible that the size shift is too slight to be visible on a gel (maximum 2.8 KDa), as even though the *C. jejuni* heptasaccharide occupied all potential sites, this was to varying degrees and the same phenotype may not be achieved across different glycans.

In case this was an issue with the ratio of acceptor protein to glycan, the glycosylation trial was repeated, but with AcrAtag expressing from pEXT21 and GT2 from pEXT20. Although less AcrAtag protein would be expected from expression in pEXT21, due to a much lower copy number, no protein could be detected by anti-His immunoblot, indicating lack of AcrAtag expression. As this is the same backbone (pEXT21) expressing GT2 in the first experiments, it suggests GT2 was also not being expressed in the system, for an unknown reason (data not shown).

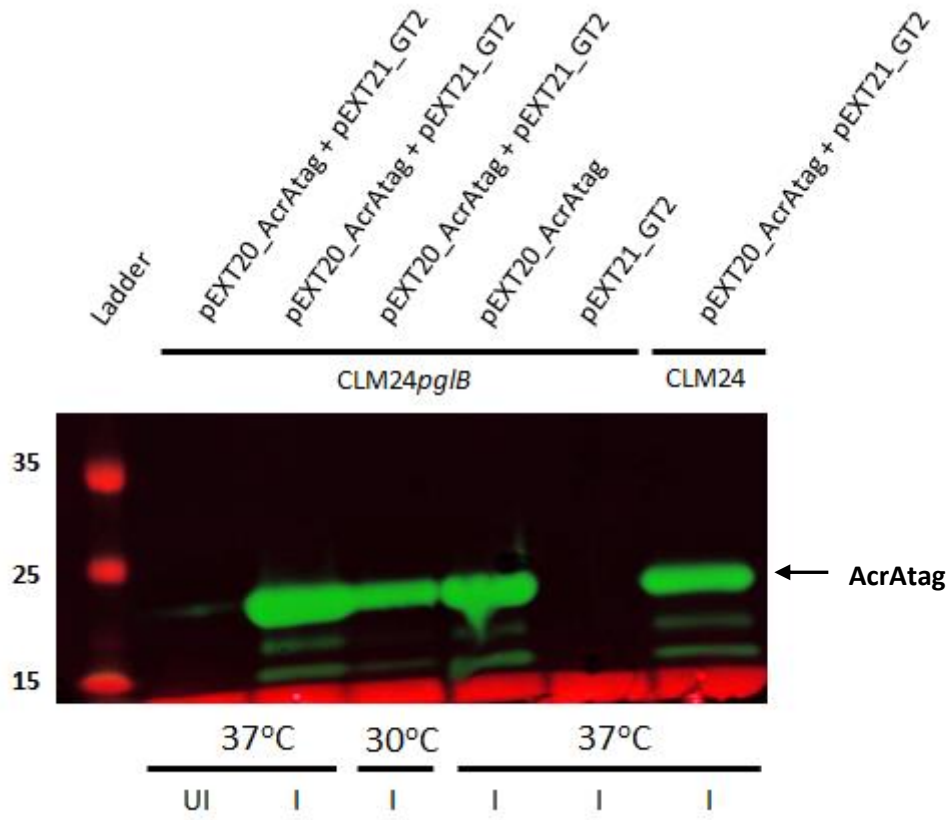


Figure 7.10. Expression of AcrAtag with *C. difficile* GT2. pEXT21_GT2, encoding the second glycosyltransferase from the R20291 flagella glycosylation locus was co-expressed with pEXT20_AcrAtag in CLM24*pglB* and CLM24 cells (no *pglB*). These were incubated overnight at 37°C and 30°C and induced with 1 mM IPTG; I, induced, UI, uninduced. AcrAtag was extracted and nickel affinity purified then detected by immunoblot using anti-His antibody to AcrAtag.

7.4. Glycosylation trials using a single plasmid system

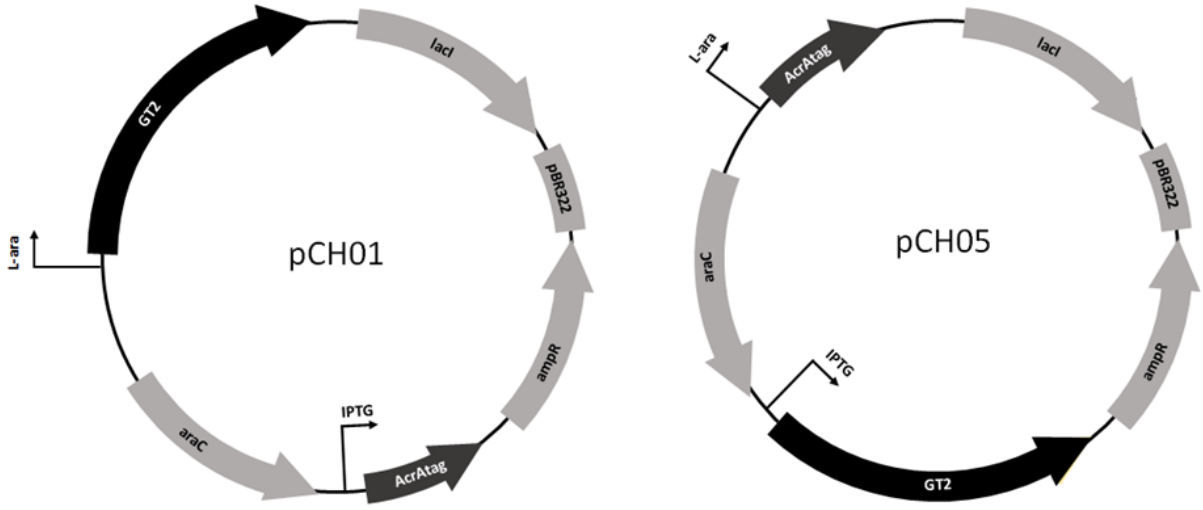
7.4.1. System design

To overcome the difficulties in working with pEXT21, and the lack of protein/glycan expression an alternative system was adopted where AcrAtag and GT2 could be expressed from the same plasmid. This is advantageous as the acceptor protein and glycosyltransferase would be under the induction of different inducible promoters (IPTG and L-arabinose), meaning their expression could be independently controlled. Furthermore, expression from the L-arabinose promoter is easier to titrate, meaning the expression of its downstream gene(s) can be more tightly controlled. The dual expression plasmids, namely the pCH series, were constructed by Gibson assembly (Methods, section 2.4.6) using pEXT20 and pEC415, which share the same pBR322 origin of replication and ampicillin resistance

cassette [582]. Dr Ian Passmore designed the system of incorporating pEC415 and pEXT20 to build a dual expression plasmid, which I then adapted for incorporation of my constructs.

All plasmids were constructed using Gibson assembly, amplifying the plasmid and insert fragments with complimentary overhangs to permit joining of the two constructs. This was firstly used to build pEXT20 and pEC415 harbouring either AcrAtag or GT2 which were then used in synthesis of the dual expression plasmids. The backbone of the pCH series of plasmids was a 3485 bp fragment of the backbone of pEXT20 (3913 bp), containing the pBR322 origin of replication, the *lacI* repressor gene, the ampicillin resistance cassette, the IPTG-inducible P_{tac} promoter for expression of AcrAtag or GT2 but without *tet* gene, as tetracycline resistance was not required. In place of the *tet* gene, a 4176 bp fragment of pEC415 was inserted, containing the L-arabinose inducible P_{Bad} promoter, again controlling AcrAtag or GT2 and the *araC* gene, for activation of expression. Plasmids were built with L-arabinose inducible GT2 and IPTG inducible AcrAtag (pCH01) or in the opposite arrangement (pCH05) (Figure 7.11).

In addition to building these plasmids with the standard GT2 construct used previously, a codon optimised version of GT2 was designed using the IDT Codon Optimisation tool and ordered for synthesis as a G-block. Codon usage differs between species and therefore some may be less well translated when expressed in the non-native strain, depending on the pool of tRNA available. By codon optimising the sequence for *E. coli*, protein expression can be improved. Multiple attempts were made to insert codon optimised GT2 into pEXT20 or pEC415 and many clones were obtained following ligation and transformation of Top10 cells. However, all clones harboured multiple SNPs or deletions within the coding region of GT2, suggesting codon optimised GT2 may be toxic to the cell, so this was not pursued.



	pCH01	pCH05
L-arabinose	GT2	AcrAtag
IPTG	AcrAtag	GT2

Figure 7.11. Plasmid maps of pCH01 and pCH05. Schematic of the plasmid design for dual expression of GT2 (*CDR20291_0242*) and AcrAtag. The pBR322 origin of replication, ampicillin resistance cassette, *lacI* gene and IPTG inducible promoter all originate from pEXT20, whereas the L-arabinose inducible promoter and *araC* gene originate from pEC415. The table describes the inducer for each construct in plasmids pCH01 and pCH05.

7.4.2. Confirming glycosylation of L-arabinose-inducible AcrAtag

Before testing of the dual plasmids was undertaken, it was first necessary to confirm expression and glycosylation of the L-arabinose inducible AcrAtag. pEC415_AcrAtag, which was constructed as an intermediate plasmid during generation of pCH05, was transformed into CLM24*pglB* and CLM24 cells harbouring pACYC*pgl*Δ*pglB* (Figure 7.12) then induced overnight at 37°C or 30° with 1 mM IPTG for induction of PglB and a range of L-arabinose concentrations (0.05, 0.1, 0.2, 0.4, 1%), to determine whether varying the amount of acceptor protein available could improve efficiency of the glycosylation reaction.

As seen with pEXT20_AcrAtag, a ladder of eight bands corresponding to occupation of all seven glycosylation sites was identified (Figure 7.12). Interestingly, glycosylation was more pronounced with

an increase in AcrAtag and at 37°C over 30°C, the latter being the opposite finding compared to using pEXT20_AcrAtag (Figure 7.9). It is noteworthy that in this test, more protein appears to be available for glycosylation at 37°C, which could explain the improved glycosylation (Figure 7.12). In summary, AcrAtag expressing from a L-arabinose inducible promoter can be glycosylated and remains a suitable acceptor protein.

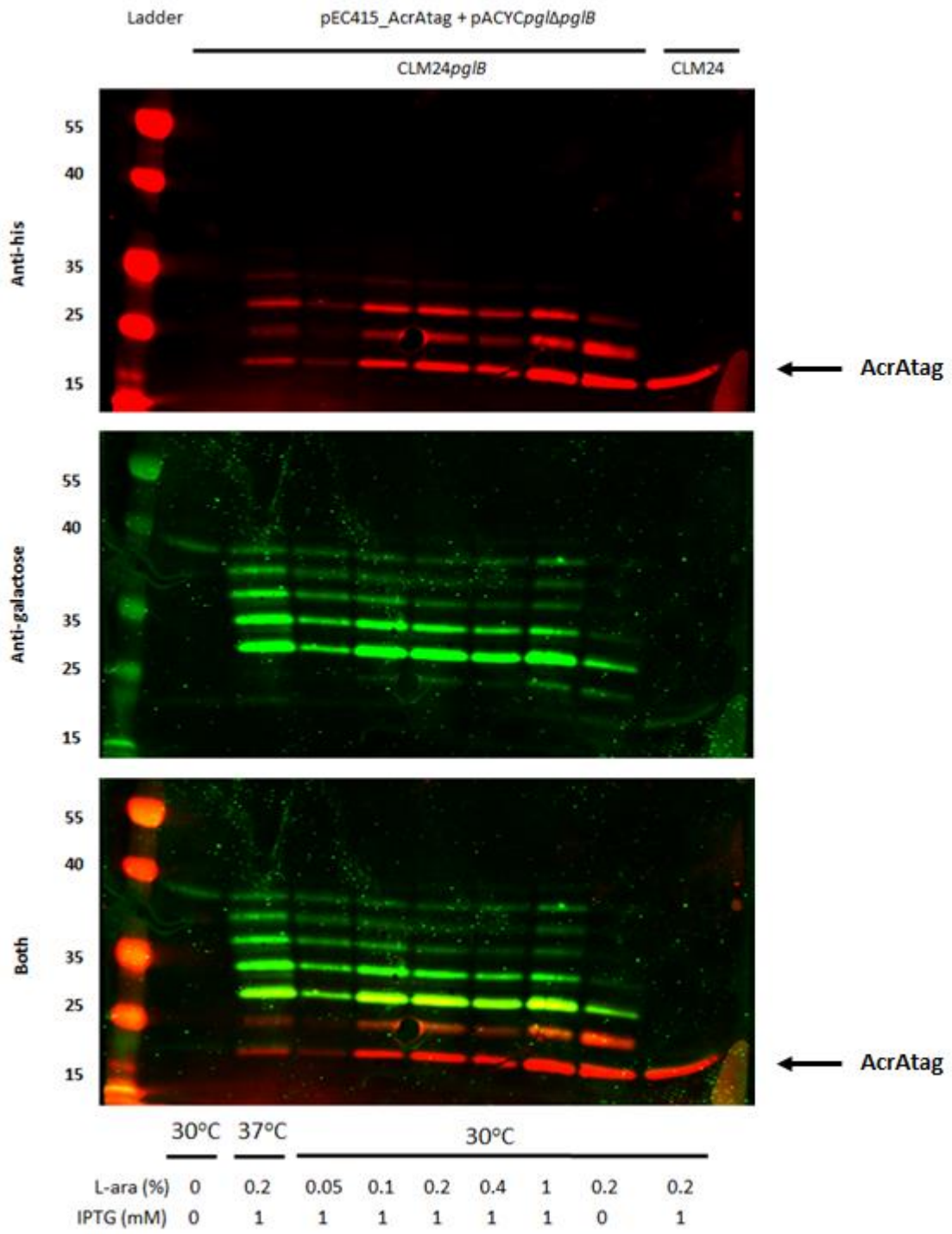


Figure 7.12. Glycosylation of L-arabinose-inducible AcrAtag with the *C. jejuni* heptasaccharide. To confirm AcrAtag could be glycosylated when expressed from an L-arabinose inducible promoter, the constitutive pACYC*pgl*Δ*pglB* construct, encoding the *C. jejuni* heptasaccharide locus was co-expressed with pEC415_AcrAtag in CLM24*pglB* and CLM24 cells. These were incubated overnight at 37°C and 30°C and induced with 1 mM IPTG and a range of L-arabinose concentrations. AcrAtag was extracted and nickel affinity purified then detected by immunoblot using anti-His antibody to AcrAtag and SBA lectin to the glycan. The three panels present the immunoblot images with anti-His only (red), anti-glycan only (green) then the two overlaid. When both AcrAtag and the *C. jejuni* heptasaccharide are expressed together in the presence of PglB, 7 bands are produced in addition to AcrAtag alone (~17.27 KDa).

7.4.3. Screening for glycosylation with dual expression pCH01 and pCH05 plasmids

Following detection of AcrAtag glycosylated with the *C. jejuni* heptasaccharide using pEC415_AcrAtag, glycosylation was tested for using pCH01 and pCH05. Either pCH01 or pCH05 was transformed into CLM24*pglB* and CLM24 cells, which were induced overnight with 1 mM IPTG for PglB and GT2/AcrAtag induction and a range of L-arabinose concentrations (0.05, 0.1, 0.2, 0.4, 1%) for GT2/AcrAtag induction, at 37^{o}C or 30^{o}C. Cells were lysed and the soluble fraction nickel affinity purified then visualised on an anti-His immunoblot. This revealed no obvious shift in molecular weight of AcrAtag under any condition, suggesting glycosylation had not occurred (Figure 7.13).

Additional His-reactive bands were detected above the band corresponding to AcrAtag alone, which can be indicative of glycosylation as seen with the *C. jejuni* heptasaccharide Figures 7.9 and 7.12). The shift in size of these bands is larger than the predicted size of the glycan, but this could be a result of polymerisation of the glycan by the native *E. coli* O-antigen polymerase, Wzy. However, after increasing the intensity of the green channel, the same banding pattern was identified in the PglB-negative control, suggesting this is not a result of glycosylation and instead is likely oligomerisation of AcrAtag.

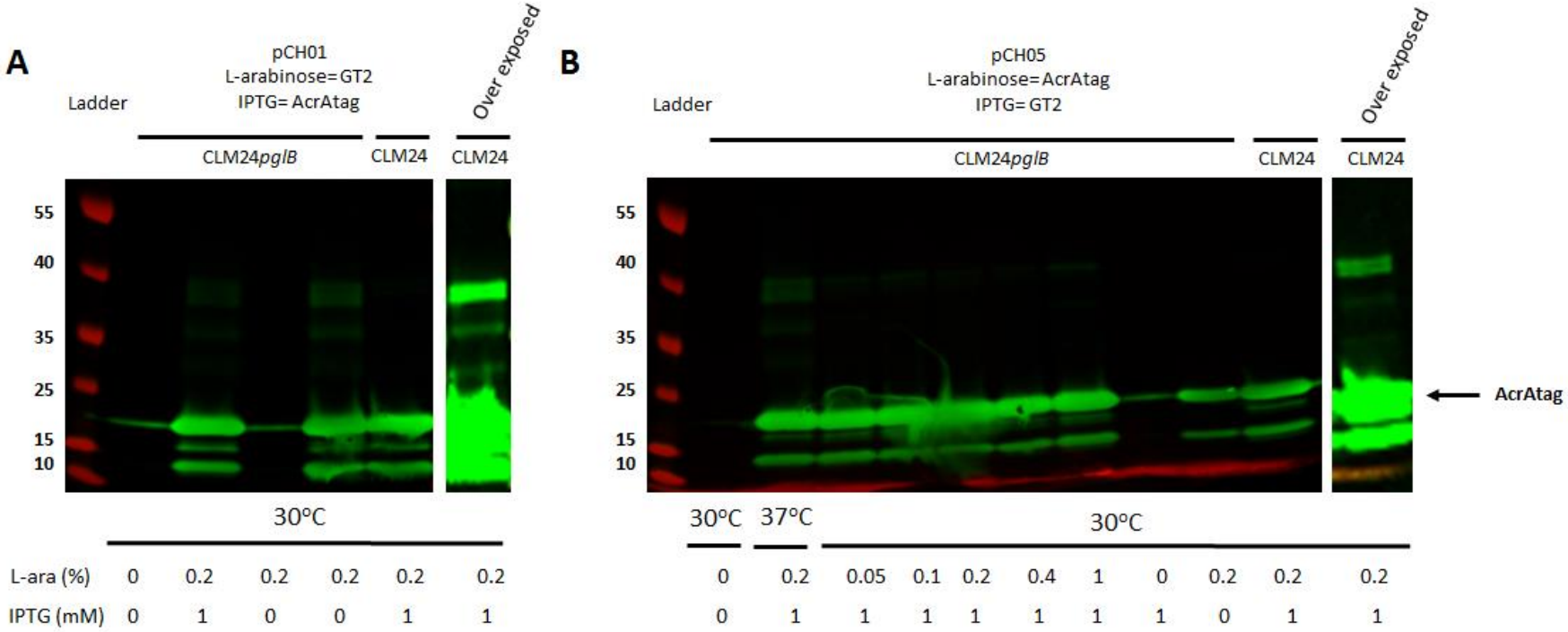


Figure 7.13. Expression of the pCH01 and pCH05 dual expression plasmids in the presence of PglB. To screen for glycosylation of AcrAtag when expressing from pCH01 or pCH05, plasmids were transformed into CLM24*pglB* or CLM24 cells and induced overnight at 37°C or 30°C with 1 mM IPTG and a range of L-arabinose concentrations. AcrAtag was nickel affinity purified then detected by anti-His immunoblot (green). (A) Nickel affinity purified protein isolated after induction of pCH01 and (B) Nickel affinity purified protein isolated after induction of pCH05.

To confirm the absence of detectable glycosylation so far was not the result of lack of GT2 expression, the GT2 construct was built with an additional C-terminal 6XHistag. This was inserted into pEC415 using Gibson assembly, transformed into CLM24*pglB* and CLM24 cells and induced overnight with 0.05, 0.1, 0.2, 0.4 or 1% L-arabinose at 37°C and 30°C, with or without induction of PglB (1 mM IPTG) (Figure 7.14). Purified protein was visualised using an anti-His immunoblot which revealed a protein approximately 72 KDa in size, corresponding to the estimated size of GT2-His, confirming it is expressing in the presence and absence of PglB. Expression appears to be improved at 37°C compared with 30°C, although higher levels of the glycosyltransferase doesn't necessarily equate to improved glycosylation. Expression is also increased in the absence of PglB, an unsurprising result considering the burden of PglB on the cell.

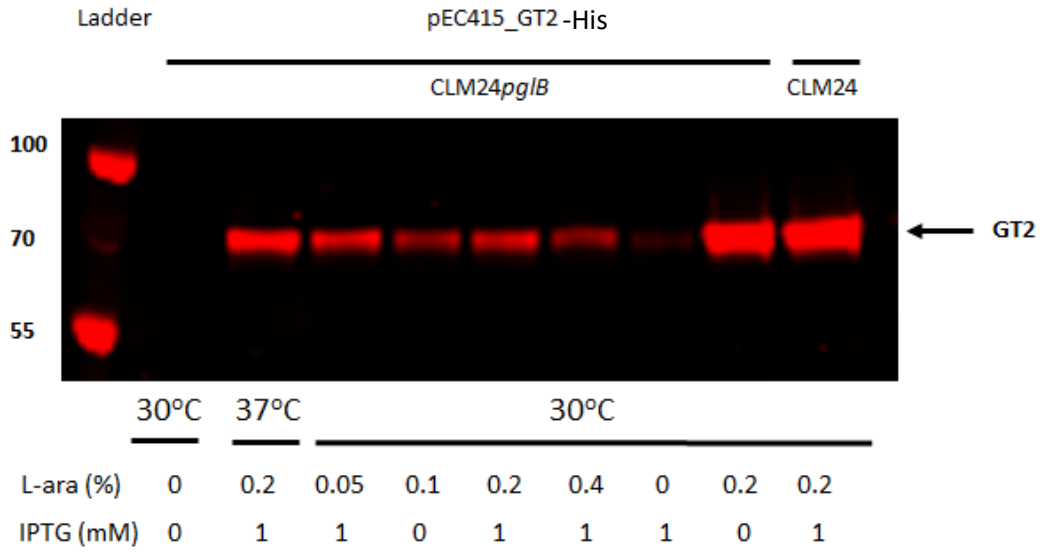
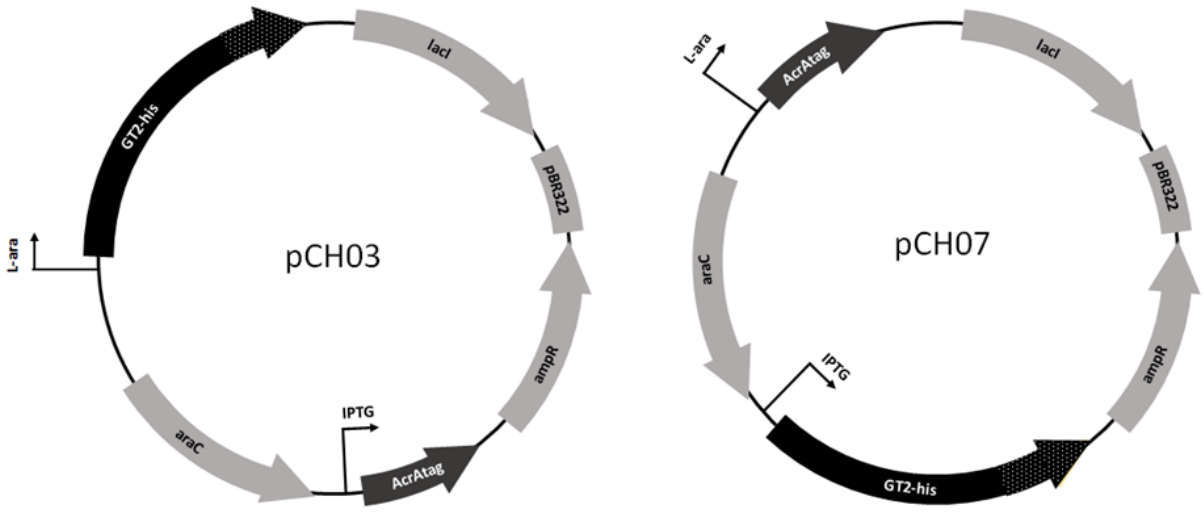


Figure 7.14. Confirming the expression of GT2. pEC415_GT2-His harbouring a C-terminal 6XHistag was expressed overnight at 37°C and 30 °C with a range of L-arabinose concentrations. Protein was extracted and the soluble fraction nickel affinity purified before probing on an anti-His immunoblot.

GT2 in the dual expression plasmids was replaced with the 6XHis-tagged construct to produce pCH03 and pCH07, allowing the monitoring of GT2 expression throughout the different screening conditions. (Figure 7.15). As before, these were synthesised using Gibson assembly, amplifying GT2-His and the backbone of pCH01 and pCH05 without the GT2 construct, with complementary overhangs to enable ligation.



	pCH03	pCH07
L-arabinose	GT2	AcrAtag
IPTG	AcrAtag	GT2

Figure 7.15. Plasmid maps of pCH03 and pCH07. Schematic of the plasmid design for dual expression of GT2-His (*CDR20291_0242*) and AcrAtag. The pBR322 origin of replication, ampicillin resistance cassette, *lacI* gene and IPTG inducible promoter all originate from pEXT20, whereas the L-arabinose inducible promoter and *araC* gene originate from pEC415. The table describes the inducer for each construct in plasmids pCH03 and pCH07.

Testing conditions mirrored those used for pCH01 and pCH05 (Figure 7.13), using both CLM24*pglB* and CLM24 strains with induction at 37°C and 30°C and a titration of L-arabinose for AcrAtag induction. In order to improve purification of AcrAtag and only isolate protein available for glycosylation, periplasmic extractions were performed prior to nickel affinity purification. Here, periplasmic protein is released using a sucrose buffer with lysozyme while the inner membrane is left intact, meaning the cytoplasmic protein is not released. Following removal of the periplasmic fraction, the cytoplasmic protein was also extracted using freeze thaw cycles, which should contain GT2-His.

No glycosylation was visible when GT2-His and AcrAtag were expressed from pCH03 or pCH07 (Figure 7.16). Again, the additional bands above AcrAtag only were present in both CLM24*pglB* and CLM24 lanes. There was some contamination of cytoplasmic protein, as demonstrated by the presence of an anti-His reactive band at ~ 70 KDa, the size of GT2. When the cytoplasmic fractions were probed on an anti-His immunoblot, for pCH07, GT2-His expression was confirmed in all test conditions,

suggesting expression of the glycosyltransferase was not the reason for lack of detectable glycosylation. Cytoplasmic fractions were not tested for pCH03 as GT2-His was clearly visible in the periplasmic fractions, likely due to contamination during processing.

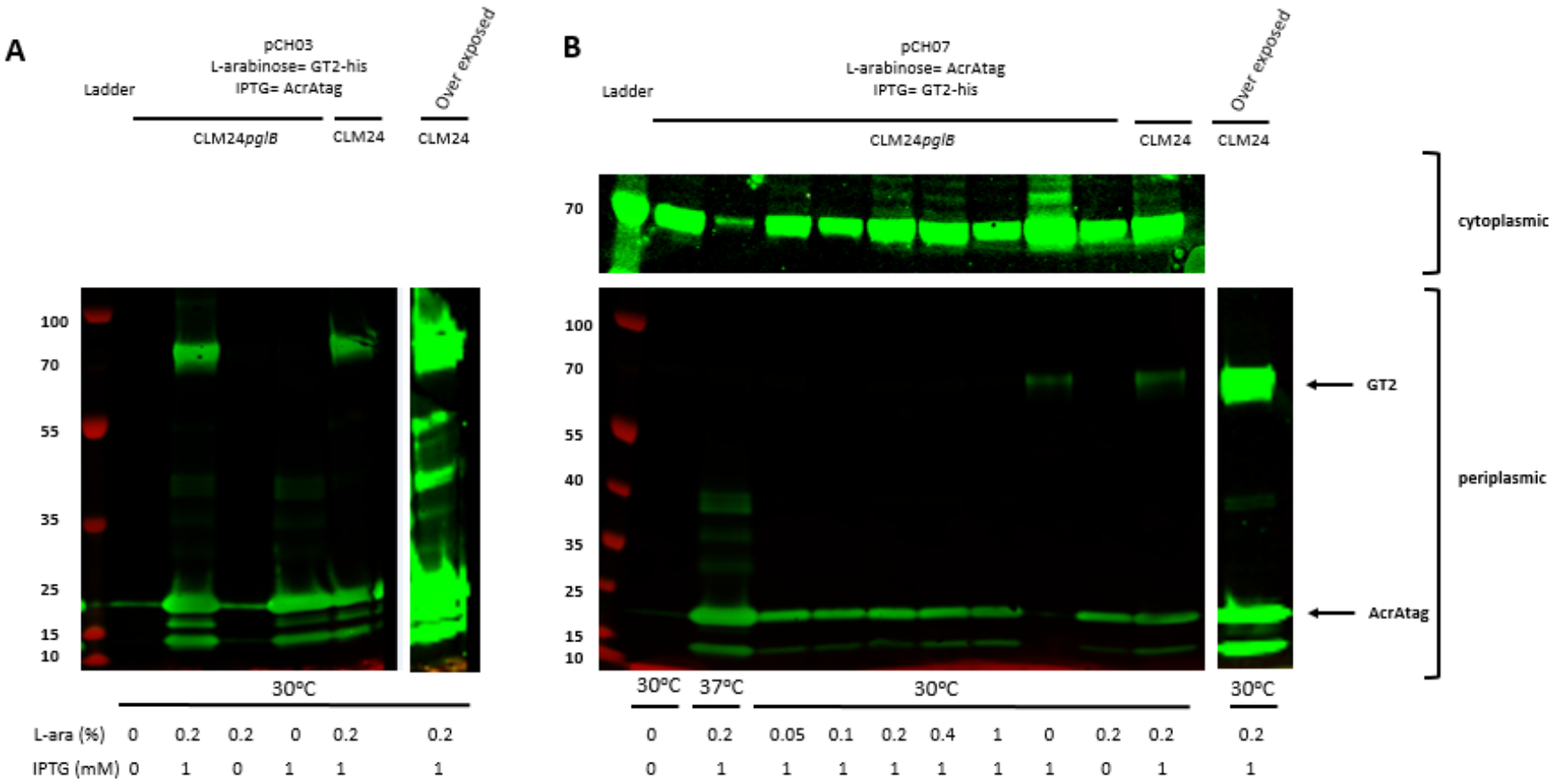


Figure 7.16. Expression of the pCH03 and pCH07 dual expression plasmids in the presence of PglB. To screen for glycosylation of AcrAtag when expressing from pCH03 or pCH07, plasmids were transformed into CLM24*pglB* or CLM24 cells and induced overnight at 37°C and 30°C with 1 mM IPTG and a range of L-arabinose concentrations. AcrAtag was isolated in periplasmic extractions and GT2-His in a cytoplasmic extraction (pCH07 only), then both nickel affinity purified and detected by anti-His immunoblot (green). (A) Nickel affinity purified protein isolated after induction of pCH03 and (B) Nickel affinity purified protein isolated after induction of pCH07.

7.5. Glycosylation trials using an *in vitro* glycosylation system

A recently published study demonstrated a method for *in vitro* protein glycosylation using a "one pot" system [583]. The system uses individual culture lysates harvested from an *E. coli* expressing the acceptor protein, *E. coli* expressing the glycan (which is built onto the lipid linker in the membrane) and *E. coli* expressing the OST (i.e. PglB). These are incubated together overnight before nickel affinity purifying the protein from the reaction and screening for glycosylation. By preparing the components for glycosylation separately, the ratio of glycan, acceptor protein and PglB can be more easily manipulated. This system also offered the possibility to overcome one of the major potential barriers of PglB-mediated production of the *C. difficile* glycan, the transfer of the glycan from the cytoplasm to the periplasm, as the promiscuity of the *E. coli* Wzx flippase is still not defined. It is assumed that PglB requires glycan built on UndPP but that this is not restricted by which leaflet of the inner membrane this is presented on, therefore removing the need for transfer to the periplasm. We therefore hypothesised that *in vitro* glycosylation could avoid the step of transferring the glycan into the periplasm, enabling glycosylation of AcrAtag with the truncated *C. difficile* flagella glycan.

7.5.1. Testing GT2 using *in vitro* glycosylation

To confirm correct set up of the *in vitro* assay and glycosylation of AcrAtag in this system, both this protein and the Exotoxin A (ExoA) acceptor protein from *Pseudomonas aeruginosa,* containing nine glycosylation sequons, were incubated with lysate harbouring UndPP-linked *C. jejuni* heptasaccharide. Individual cell lysates containing PglB, *C. heptasaccharide* built onto UndPP and ExoA were provided by Dr Elizabeth Atkins. A 1 L culture of the acceptor protein (pEXT20_AcrAtag) in CLM24 cells was induced overnight with 1 mM IPTG at 37^{o}C then washed with S30 buffer before lysing on the cell homogeniser. Lysates of the acceptor protein, *C. jejuni* heptsaccharide and PglB were incubated in S30 buffer overnight at 30^{o}C, then the protein nickel affinity purified before probing on an anti-His, SBA anti-galactose immunoblot (Figure 7.17). A single, ~90 KDa band was detected when ExoA and *C. jejuni*

heptasaccharide were incubated without PglB, with a strong anti-galactose signal produced when PglB was added into the reaction, demonstrating that glycosylation had occurred.

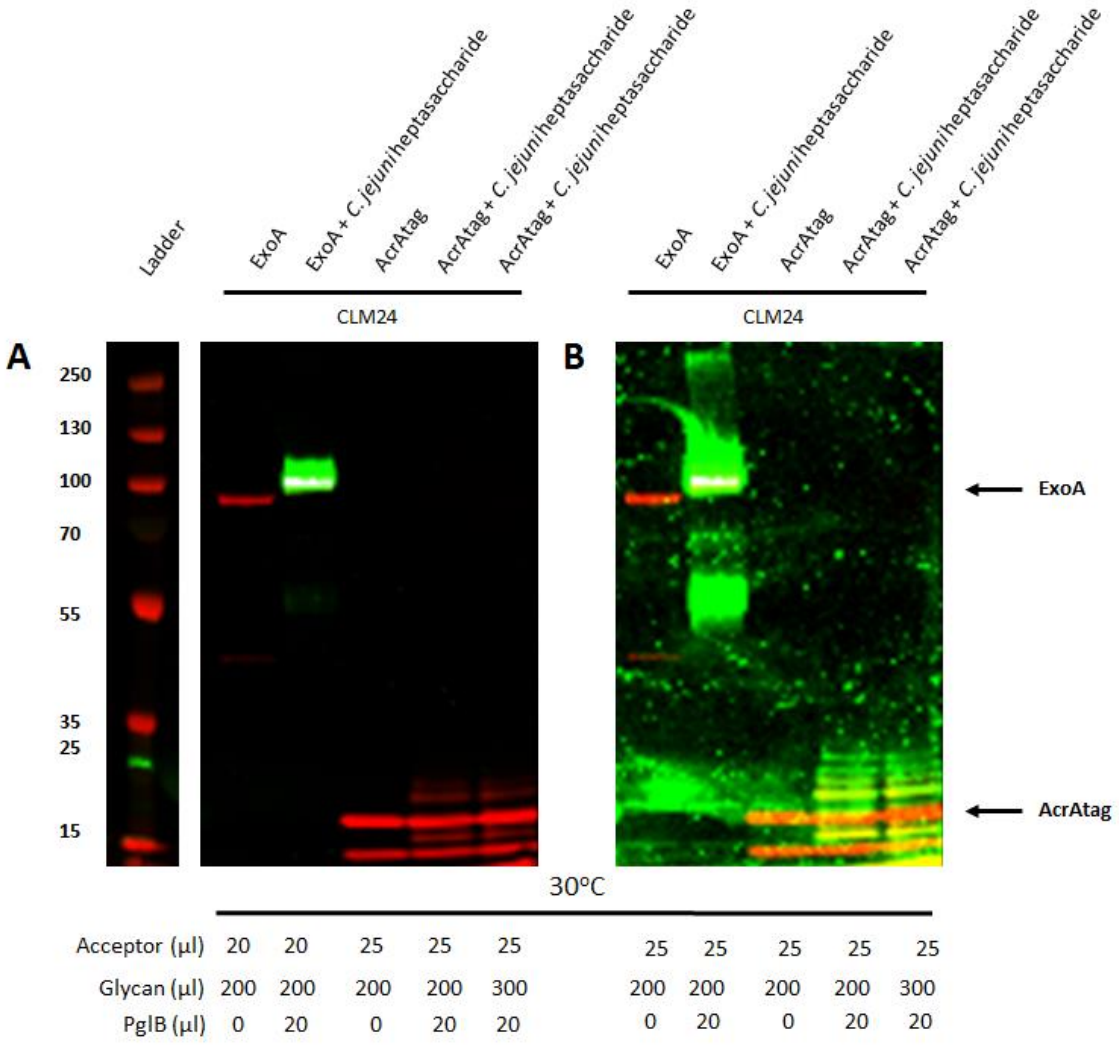


Figure 7.17. *In vitro* glycosylation of ExoA and AcrAtag with the *C. jejuni* heptasaccharide. Lysates of the *C. jejuni* heptasaccharide, PglB, and acceptor proteins ExoA or AcrAtag were incubated together overnight at 30°C, then the nickel affinity purified protein probed on an anti-His (red), anti-galactose (green) immunoblot. (A) standard exposure, (B) increased exposure for the green channel.

A single band at ~ 20 KDa was detected representing AcrAtag only, which increased to multiple bands when PglB was added. Glycosylation had occurred, demonstrated by the increased size of AcrAtag, but compared to *in vivo,* only glycosylation of four of the available seven available sites was detected

in vitro (Figure 7.17). This may be a result of insufficient glycan availability, as in the *in vivo* system the heptasaccharide is constitutively expressed.

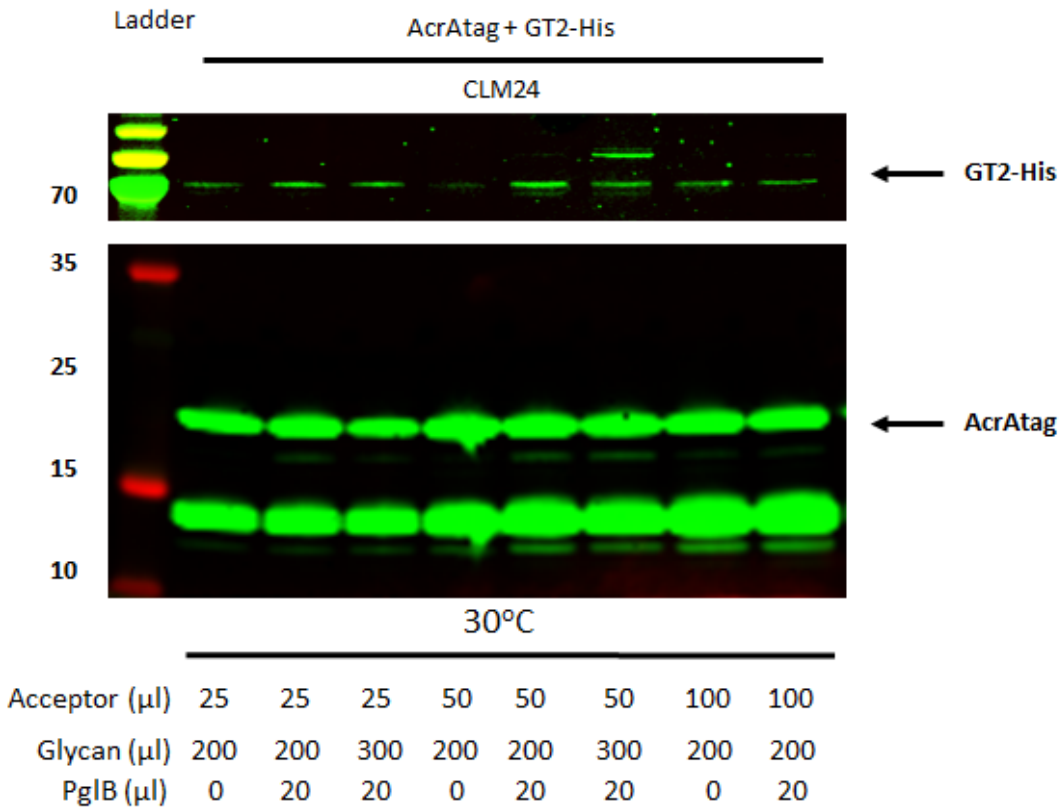


Figure 7.18. *In vitro* glycosylation trials of AcrAtag with the *C. difficile* GT2. Lysates of CLM24 harbouring GT2-His, PglB, and AcrAtag were incubated together overnight at 30°C, then the nickel affinity purified protein probed on an anti-His (green) immunoblot.

Once the conditions for glycosylation had been established, the *C. jejuni* heptasaccharide was replaced with GT2-His. Between 200 and 300 µl of GT2-His donor was incubated with either 25, 50 or 100 µl of AcrAtag. As before, reactions were incubated overnight at 30°C, then the protein was nickel affinity purified and probed on an anti-His immunoblot (Figure 7.18). There was no difference in electrophoretic mobility of AcrAtag under any condition, suggesting glycosylation had not occurred. Only a faint band was detectable corresponding to GT2-His.

Previous investigations with AcrAtag have detected additional His-reactive bands below the size of whole AcrAtag, attributed to proteolytic degradation of the protein. An additional band was also

detected here, more pronounced than previous assays which is assumedly also a result of proteolytic cleavage which have been enhanced in the upscaled set up and further incubation steps. Inclusion of a protease inhibitor in further experiments may limit this effect, although this would first have to be tested within the control assay to ensure protease inhibitor did not interfere with the i*n vitro* glycosylation reaction using the *C. jejuni* glycan.

7.6. The *E. coli* O13 antigen

The lack of glycosylation detected so far necessitated the investigation of alternative mechanisms of expressing the truncated flagella glycan. One such option was the use of a glycan from a different species that shared structural similarity with the glycan of interest. Depending on similarity, this could then be expressed directly or further modified to achieve the desired structure. Furthermore, use of a complementary glycan with a native synthesis pathway, more compatible with PglB could also increase the chance of glycosylation.

In the search for a suitable glycan, it was important to identify a structure with the correct linkage between the monosaccharides, as this stereochemistry can dictate antibody binding and reactivity. A number of L-rhamnose rich glycans were identified, such as from the Group A Streptococcus surface polysaccharide, but this lacked the required linkage [584]. Interestingly, the O-antigen of *Shigella flexneri* 2a harbours a GlcNac starting sugar followed by an L-rhamnose, attached via a β1-3 linkage, identical to the starting sugars of the R20291 flagella glycan (Figure 7.19). Although the next L-rhamnose is attached by a second β1-3 linkage, this is attached to the subsequent L-rhamnose by a β1-2 linkage, also found in the R20291 glycan. Modification of the O-antigen synthesis locus, may enable production of the desired structure, including removal of the glucose moiety on the first L-rhamnose, but first it was necessary to determine whether the acceptor protein (AcrAtag) could be glycosylated with the whole *E. coli* O13-antigen.

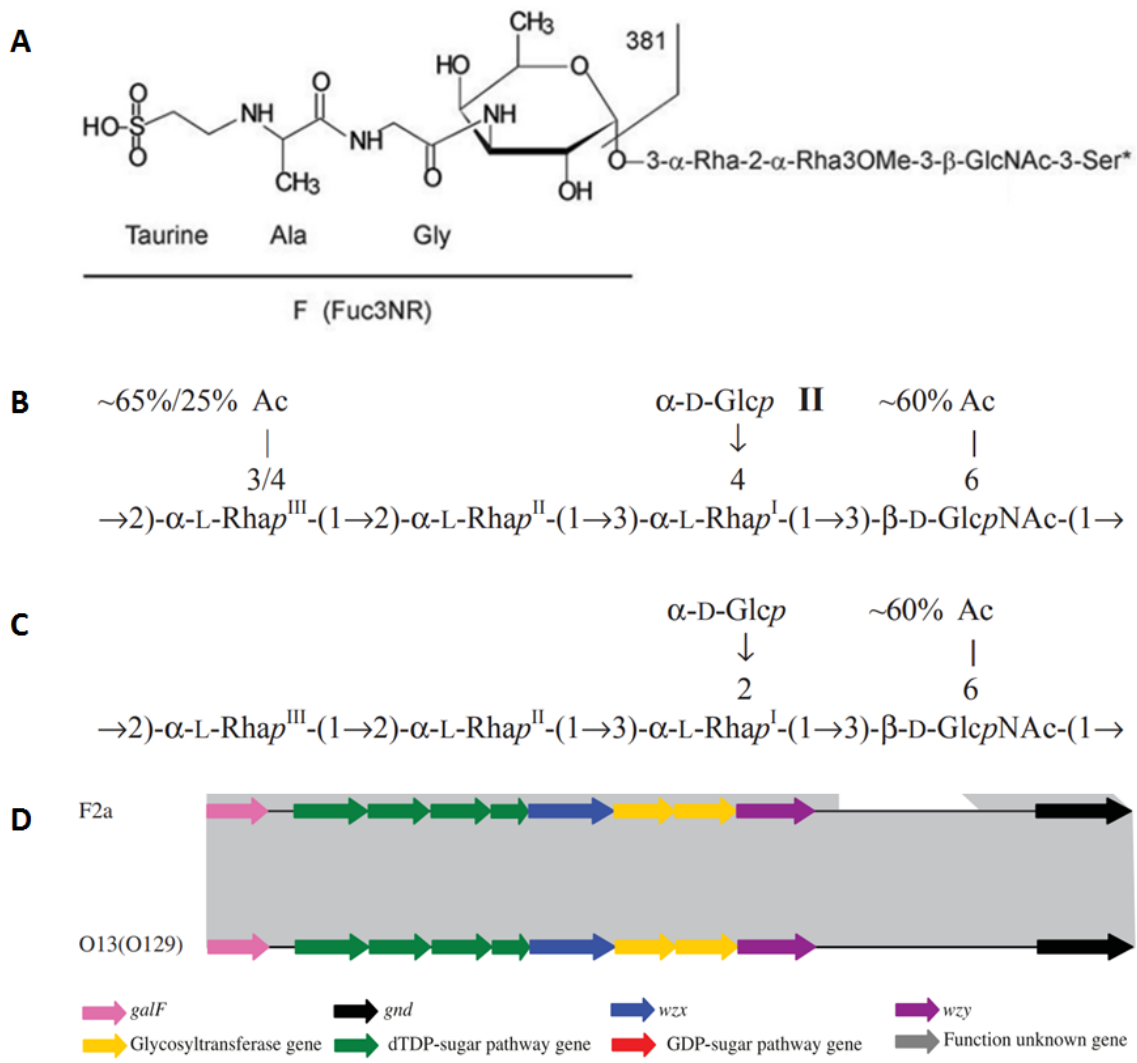


Figure 7.19. Glycan structures that share similarity with the truncated *C. difficile* flagella glycan. (A) NMR resolved structure of the structure of full length flagella glycan from *C. difficile* R20291, adapted from Bouche *et al.* [190], (B) organisation of the O-antigen from *Shigella flexnerii* serotype 2a, adapted from Perepelov *et al.* [585], (C) organisation of the *E. coli* O13-antigen [585] and (D) comparison of the loci encoding the *S. flexnerii* 2a O-antigen (F2a) and *E. coli* O13-antigen (O13), adapted from Liu *et al.* [586] (by permission of Oxford University Press).

As a serious cause of diarrheal disease, *S. flexnerii* is a Category 2, Schedule 5 pathogen, requiring increased security measures in the laboratory. Fortunately, the *E. coli* O13 antigen harbours an identical O-antigen, in structure and synthesis locus and antibodies cross react against both O-antigens [587]. As this strain does not require the same security precautions, it was selected for glycosylation trials with AcrAtag.

7.6.1. Glycosylation with the *E. coli* O13 antigen *in vivo*

E. coli strain O13 was acquired from the Public Health England Culture Collection and transformed with pEXT21_*pglB* or pGVXN115, encoding a copy of PglB harbouring two amino acid alterations [588], rendering it inactive. Both pEXT20_AcrAtag and pEC415_AcrAtag (IPTG induced or L-arabinose induced, respectively) were trialled as the acceptor for glycosylation, for IPTG or L-arabinose induction of AcrAtag, respectively. Strains were induced overnight at 37°C or 30°C with 1 mM IPTG and 0.2% L-arabinose, or a titration of L-arabinose for pEC415_AcrAtag. Cells were harvested and lysed, protein purified using nickel affinity chromatography then resolved using SDS-PAGE and probed on an anti-His immunoblot (Figure 7.20). Using pEXT20_AcrAtag, it was not possible to detect glycosylation in any condition. At ~ 1.05 KDa, a single copy of the O13 antigen is much larger than the truncated flagella glycan (~0.4 KDa), meaning unglycoyslated and glycosylated species would be more easily distinguished by gel electrophoresis. Furthermore, if all seven glycosylation sequons were occupied, this could result in a total 7.399 KDa shift, in addition to potential polymerisation by the *E. coli* Wzy polymerase.

In a second trial, using L-arabinose induced AcrAtag, a ladder of bands was produced above the unglycosylated AcrAtag when induced at 37°C with pEXT21_*pglB*. This was absent in the CLM24 trial and therefore indicates glycosylation had occurred. Furthermore, the 15 bands detectable above the unglycosylated AcrAtag, exceed the seven glycosylation sites available, suggesting polymerisation of the glycan. In relation to the trials performed at 30°C, the same laddering effect is absent, and again the suspected oligomerisation of AcrAtag was demonstrated in the presence and absence of PglB. There is a single band immediately above the unglycosylated AcrAtag, which potentially could be evidence of glycosylation as this is absent from the PglB-negative lane.

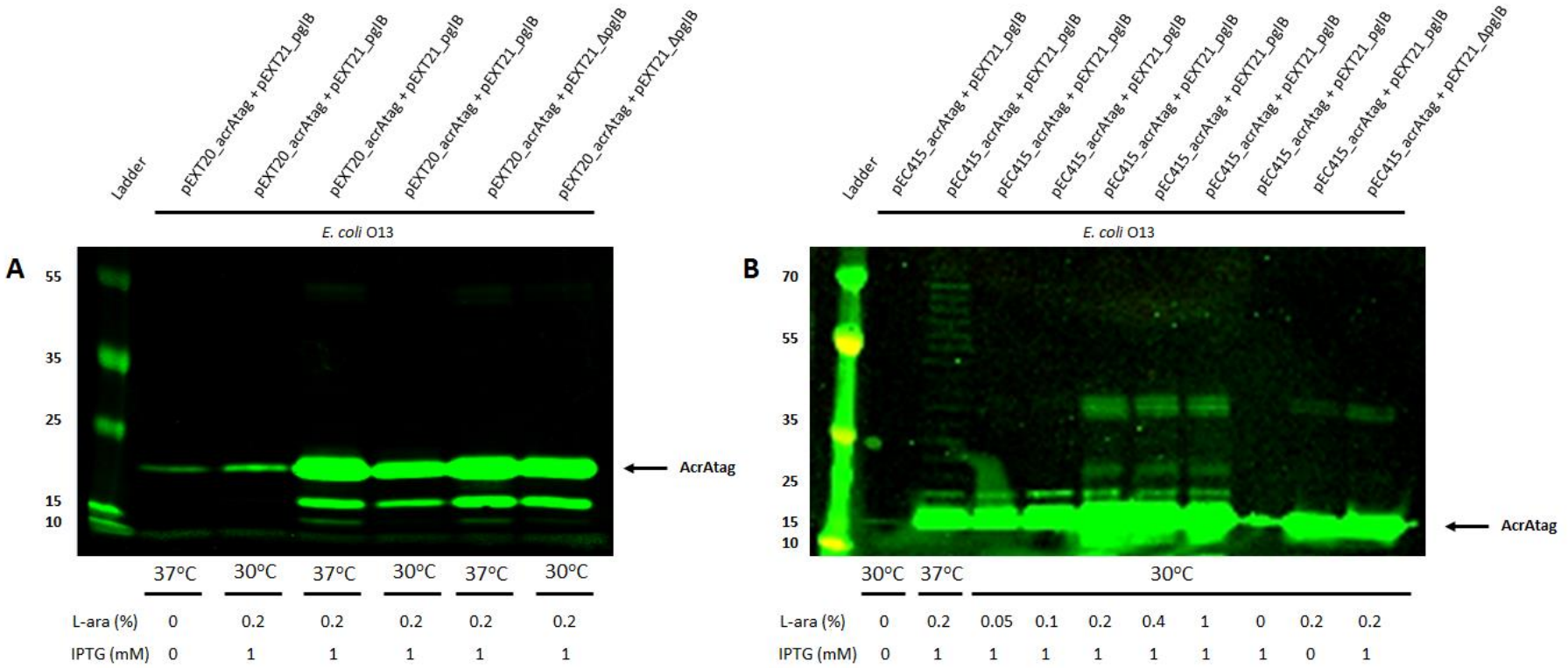


Figure 7.20. Glycosylation of AcrAtag with the *E. coli* O13-antigen. AcrAtag was expressed in a strain of *E. coli* O13 with PglB or inactivated PglB at 37°C and 30°C then the soluble protein fraction nickel affinity purified. Protein samples were probed on an anti-his immunoblot (green). (A) Expression of IPTG inducible AcrAtag with PglB in *E. coli* O13 and (B) expression of L- arabinose inducible AcrAtag with PglB in *E. coli* O13.

As demonstrated with the *C. jejuni* glycosylation trials, anti-glycan antibodies and lectin offer an additional means of confirming glycosylation and are often more sensitive to detecting glycosylated protein than anti-his antibody alone. An antibody specific for the *S. flexnerii* O-antigen, and consequently the *E. coli* O13 antigen was purchased, but unfortunately no reactivity was detected. This is likely due to low antibody concentrations as the antibody purchased was not optimised for use in immunoblotting. In summary, incubation of AcrAtag with PglB and the *E. coli* O13 antigen results in glycosylation and polymerisation.

Although not possible within the time constraints of this work, the next steps would be to investigate mutagenesis of the O13-antigen synthesis loci from *E. coli* to generate a modified glycan to fully match the truncated *C. difficile* flagella glycan structure.

7.7. Discussion

The work in this chapter aimed to utilise the recombinant PglB glycosylation pathway in *E. coli* to engineer a *C. difficile* specific glycoconjugate, the first example of bioconjugation with a *C. difficile* glycosyltransferase. Despite multiple attempts using a number of approaches, it was not possible to detect glycosylation of our targeted acceptor protein with the truncated *C. difficile* flagella glycan, via a recombinant *C. difficile* glycosyltransferase (GT2). This may be due to issues with detection of such a small glycan (only 0.4 KDa) or the lack of transfer of the glycan onto the AcrAtag carrier protein. To account for this, AcrAtag from the glycosylation trials using the dual expression system and *in vitro* glycosylation are being analysed by NMR, which is a much more sensitive means of glycan detection.

The aims of this work were not only novel in relation to *C. difficile,* but also within the field of bioconjugation, in terms of building an *O*-linked glycan from a Gram-positive organism using an *N*-linked system in a Gram-negative bacteria. These aims feed into ongoing research developing the bioconjugation "toolbox", identifying novel enzymes and pathways and using unrelated enzymes together to expand the bioconjugation repertoire. For example, Hug *et al*. demonstrated the synthesis and conjugation of a human Lewis antigen in *E. coli*, using enzymes from *Haemophilus influenzae* and *Helicobacter pylori* [589]. Despite these successes, this remains a challenging research area, requiring optimisation and modification of glycosylation pathways and identification of new glycoengineering tools. Glycoengineering in *E. coli* is a technology in its infancy and all new attempts to couple and express different protein/glycan combinations add to the knowledge bank. This study designed, expressed and glycosylated a novel acceptor protein, AcrAtag, for use in glycosylation trials. Many common acceptors are between 40 and 70 KDa. Therefore, AcrAtag represented an alternative short reporter peptide that could be glycosylated at all seven sites and enable detection of modification with smaller glycans.

In relation to production of the flagella glycan using bioconjugation, if it is confirmed by NMR that AcrAtag is not glycosylated, there are three possible stages which may have prevented synthesis.

Firstly, is the glycan being synthesised onto UndPP within the inner membrane? As *C. difficile* doesn't encode the enzyme for this step, WecA from *E. coli* was used which transfers GlcNac to UndPP [575]. As GlcNac is also the starting sugar for the flagella glycan it was hypothesised that *C. difficile* GT2 would then build the rhamnose residues onto this GlcNac. However, this requires GT2 to interact with a membrane-associated GlcNac residue, differing from its proposed native function of building the L-rhamnose residues onto GlcNac conjugated to the flagella filament [335]. Glycan synthesis can require cooperative action of glycosyltransferases, meaning GT2 function may have been impaired without its native accompanying glycosyltransferases, or in the presence of the unfamiliar WecA. For example, the streptococcal virulence-associated protein, Fap1, is only decorated with its *O*-linked glycan upon interaction of two of the required glycosyltransferases [590]. In the absence of suitable methods to isolate UndPP-linked glycan still attached to the inner membrane, it is difficult to determine whether the glycan is being synthesised. However, no glycosylation was detected in the *in vitro* system either. This system is hypothesised to only require glycan built onto UndP, irrespective of its position in the cytoplasm or periplasm. Therefore, if glycosylation is not occurring, this suggests the glycan is not being built. It it also possible that glycan synthesis was inefficient and the levels of UndPP-linked glycan were too low to be available as a substrate for PglB. It was possible to glycosylate both AcrAtag and ExoA with the *C. jejuni* heptasaccharide, therefore it is possible that currently, the assay requires plenty of optimised acceptor protein and glycan for high-level glycosylation.

The next step is whether the glycan is being translocated across the membrane. This is normally performed by a dedicated flippase enzyme, but as this is not required by the native pathway of glycan synthesis in *C. difficile*, it was necessary to reappropriate the *E. coli* O16-antigen flippase, Wzx, present in the strains of *E.coli* used in this study (CLM24). As the flagella glycan and O16 antigen share the same two starting sugars, albeit with different linkages, it was proposed that this would be a sufficient level of similarity for translocation. However, the promiscuity of Wzx in *E. coli* and other species is subject to debate. Early research demonstrated full complementation of Wzx_{O16} with a variety of *E. coli* Wzx flippases, providing the native *O*-antigen shared the same starting sugar (GlcNac) as O16

[591]. However, further work found this transfer from the cytoplasm to the periplasm was only possible when Wzx was overexpressed from a plasmid. This was compensating for inefficient translocation compared to the native Wzx, as confirmed through chromosomal integration of the flippases [575]. The translocation rate was also improved if the target antigen starting sugars were shared between Wzx homologs.

From a methodological stand point, overexpression of Wzx could be exploited to improve translocation and accordingly glycosylation, although the consequences of overexpression of a large membrane bound transporter on a cell already expressing an OST, glycosyltransferase and acceptor protein would have to be considered. Furthermore, understanding of the activity and specificity of Wzx is still limited and does not explain the activity of all Wzx homologs. For example, diverse antigens from *Pseudomonas aeruginosa* and *Yersinia pseudotuberculosis* all share the same Wzx [592, 593] and recent work building a human-like glycan (Man3GlcNAc2) in *E. coli* O16 did so without the provision of a specific flippase, suggesting this relied on Wzx_{O16} [594]. A non-Wzx based mechanism of translocation could also be explored, such as the ABC transporter PglK from *C. jejuni*. This was demonstrated to almost fully complement the activity of two different *E. coli* Wzx flippases and transfer their corresponding glycans, suggesting promiscuity [359].

Glycosylation of AcrAtagwith the flagella glycan was also attempted using *in vitro* glycosylation but no glycosylation was detected. As discussed, it was hypothesised that this system could overcome the need for translocation of the glycan to the periplasm if PglB only requires UndPP-bound glycan, regardless of membrane orientation. However, the PglB-glycan interaction and the *in vitro* glycosylation method is not yet well enough understood to confirm this, meaning the translocation step could be essential for PglB recognition in the *in vitro* system. A way of evaluating this would be to set up an *in vitro* glycosylation reaction whereby ExoA is incubated with either *C. jejuni* heptasaccharide expressed from a wild-type *pgl* locus, or from a strain where the native flippase, PglK,

is inactivated so UndPP-linked glycan would be locked in the cytoplasmic leaflet of the inner membrane.

Use of an *E. coli* O13-antigen, which shares structural similarity with the truncated flagella glycan offered a promising alternative approach of producing a *C. difficile*- specific glycoconjugate. Glycosylation of AcrAtag with the full O13-antigen structure was demonstrated, which provided a strong starting position for further modification of the locus to obtain the desired composition. The O-antigen locus of *E. coli* O13 has been identified and annotated, but the function of the individual glycosyltransferases have not yet been characterised. The number of sugars and linkages exceed the number of identified glycosyltransferases (two), suggesting these enzymes may perform multiple roles. Use of non-native glycans has been discussed previously, particularly using the pre-existing pool of diverse *E. coli* O-antigens. A combination of PglB with different *E. coli* O-antigens have been used to build glycoconjugates specific for the human blood group B antigen and the Vi antigen from *Salmonella enterica serovar Typhi* [595, 596]. The latter included manipulation of the encoding synthesis loci to obtain a structure that cross reacted with antibodies raised against the native glycan. On the other hand, Nishiuichi *et al.* demonstrated cross reactivity between the O-antigens of *E. coli* O157 and *Citrobacter freundii* and suggested use of the latter in production of an anti-O157 glycoconjugate vaccine, which would avoid working directly with the pathogenic strain [597].

The *C. difficile* flagella glycan was selected for use as a well characterised glycan for the purposes of proof of principle experiments rather than as a strong immunogenic candidate. The immunogenic properties of the glycan are unknown and flagella glycans can be associated with immune evasion, masking immunogenic epitopes on the flagella filament [339]. The exact immunogenic properties remain to be determined but it may be that the whole or a shorter fragment of the glycan is immunogenic, indeed L-rhamnose is typically immunogenic as it is not found in human glycans, which also reduces the chances of off-target effects [598].

The surface polysaccharide PS-II has already received attention for inclusion with an anti-*C. difficile* vaccine, and antibodies have been raised to the structure *in vivo* [234-237]. This anionic polymer has been found in all *C. difficile* strains investigated to date, and the structure has been solved [231, 233]. However, the enzymes responsible for each step of synthesis remain rather elusive, partially due to the detrimental effects on the cell when genes within the synthesis loci were deleted or knocked down, as is the case for PS-II [218]. Therefore, although interesting, this was an unsuitable target for study, based on current knowledge. Based on partial annotation of the putative synthesis locus, genes encoding homologs of UndPP and for translocation across the membrane have been identified [218], suggesting similarity with the *N*-linked glycosylation pathway using PglB. If it were possible to transfer this locus into *E. coli* for use with PglB, one stalling point could be the presence of a mannose at the reducing end of the glycan, a non-permissible substrate for PglB. However, specificity of this interaction is subject to debate, and the utilisation of modified PglB by structure guided directed evolution or use of PglB orthologs from different species may be able to overcome this in the future [332].

The absence of a screening method for the glycan, such as a specific antibody was a limitation of this study. Although protein samples are currently under analysis by NMR, this is not a suitable method for screening for glycosylation for multiple different candidates from many different conditions. Regarding synthesis of the truncated flagella glycan from *C. difficile*, although it was possible to confirm *C. difficile* GT2 was being expressed in *E. coli* by anti-His Western blot, it was not possible to confirm whether the enzyme was active in the conditions for testing, either in the *E. coli* cell or the cell-free system. A means of answering this would be to use the native components from *C. difficile* within the cell-free system. As a panel of flagella mutants are already available for R20291, the lysate of a *C. difficile* strain expressing GT1 and GT2 but not FliC could be incubated in the cell-free system with recombinant FliC. If the native components are able to glycosylate the flagella protein then this could act as confirmation that the *C. difficile* glycosyltransferases are active.

There remains no licensed vaccine for *C. difficile* and the current toxin-based formulations have not yielded any success. A vaccine offers an effective control strategy for CDI, especially for at risk groups and glycoconjugate vaccines in particular have been highly successful in the control of a number of very problematic infections with over a billion doses used each year [332]. Despite the challenges described here, the rapid advances being made in the relatively new field of bioconjugation means future progress may enable the development of *C. difficile*- specific bioconjugation. It may be that identification of new technologies or enzymes allow this, such as the use of *O*-linking rather than *N*-linking OSTs [599].

8. Discussion

C. difficile can cause severe infections that are fatal for those at risk. Patterns of CDI have seen distinct changes over the past 40 years, including the emergence of hypervirulence and the associated epidemic in the mid-2000s through to the increase in community infections still seen today [69, 112, 129, 131]. As a prominent nosocomial infection, CDI poses a significant financial burden on healthcare systems, and although this is difficult to quantify, one study reported CDI cost over $4.8 billion to US acute care facilities in 2008 [600]. Current treatment options are hampered by the high rates of recurrent infections and their inability to provide long-term protection from CDI [61, 62]. There are significant benefits to be gained from an effective *C. difficile* vaccine, particularly one that is able to prevent both infection and colonisation simultaneously. The colonisation aspect is very important as those who are asymptomatically colonised with *C. difficile* have been identified as a major source of *C. difficile* transmission, therefore targeting these people could result in a steep decrease in transmission [47, 51].

Vaccine development is complex and challenging. A prime example of this is the lack of a currently marketed *C. difficile* vaccine, despite a number of high-profile clinical trials [408-410]. Furthermore, no colonisation specific antigens are yet to reach Phase II or III clinical trials. This study aimed to use a large-scale protein array to identify immunogenic *C. difficile* proteins, that could be useful anti-colonisation factors for a vaccine, then characterise a subset of these to further understand their role in the bacteria. Additionally, this study aimed to explore the development of glycoconjugate vaccines for *C. difficile,* which have received limited attention. Specifically, this focused on developing bioconjugation technology for the synthesis of a *C. difficile* specific glycoconjugate.

8.1. Identifying novel immunogenic candidates from *C. difficile*

A *C. difficile* specific pan-protein array was the starting point for identification of novel immunogenic protein candidates in *C. difficile*. A protein array of this scale is the first of its kind for *C. difficile* and enables proteome level analysis of the humoral immune response using patients with CDI and healthy

controls. In Chapter 3, the results from this array were utilised to identify protein candidates where a higher response was recorded in the healthy control group compared to those with CDI, which suggested a response to these proteins provided protection from infection. These antigens constituted a range of mainly surface proteins that were conserved across important clinical strains of *C. difficile.*

Non-toxin antigens investigated so far tend to be well characterised surface proteins of *C. difficile* such as flagella or surface layer proteins [395, 417, 424, 425]. We identified seven potential vaccine candidates on the pan-protein array where a higher IgG response was found in the healthy control group compared to the CDI patients, suggesting that prior exposure had resulted in circulating protective antibodies. These proteins have not yet been investigated in *C. difficile* either in terms of their immunogenicity or function within the pathogen, which demonstrates the utility of the pan-protein array in the identification of novel candidates. The reactivity of these antigens was confirmed using a specifically designed ELISA to minimise the cross reactivity to *E. coli,* which is also present in the human gut therefore there is likely to be substantial reactivity to this bacteria, especially considering the array proteins are printed within *E. coli* lysate [380]. Although samples are exposed to *E. coli* lysate prior to screening on the array, this is unlikely to remove all background. Therefore, by purifying these proteins it made it easier to detect any specific antibody responses without the interference of *E. coli* background. Of the three array proteins purified and analysed by ELISA, the mean signal intensity was higher in the CDH group for proteins CDR20291_0342 and CDR20291_2697 compared to the CDI, albeit not significantly so. This is most likely due to the small sample size tested in the ELISA. As discussed in Chapter 3, the inclusion of these purified proteins in a down selected array will help further analyse their immunogenicity.

Furthermore, previous studies have used ELISAs to analyse antibody content in sera from those convalescing from CDI [275, 470]. This could be explored further here, as the advantages with this study are that longitudinal samples will also be screened in this array, using these results in

combination with the available metadata, will enable comparison of antibody responses in recovered patients compared to those with continuous colonisation or those with recurrent CDI infections. Analysis of mucosal humoral immunity to CDI is limited [270, 274, 276, 427], which may in part be due to the ease in isolating blood samples as opposed to biopsies or intestinal lavage [399, 475, 476]. Faecal samples are a non-invasive means of analysing this response, but as seen with the protein array screening, cross-reactivity and high background of these samples can be problematic. Furthermore, concentration of antibody in the stool can be influenced by the size of sample and number of bowel movements [601]. ELISA based screening of stool samples has been previously achieved, using purified toxin antigens [274]. Using purified proteins like those produced in this study could be a means of analysing these stool samples, either in an ELISA or by printing the purified protein on an array, with reduced background due to the purification process removing most if not all of the *E. coli* contaminants.

The antigens selected for investigation were found in different clinically relevant strains of *C. difficile*, which is important for vaccine coverage. Due to time constraints it wasn't possible to take this further, but future work should focus on the safety and suitability of these potential vaccine candidates as well as their efficacy *in vivo*. Firstly, antibiotic use is a major risk factor for CDI, due to their devastating effects on the gut microbiome, so if CDI can be prevented when patients have been treated with antibiotics for an unrelated condition, this would improve clinical outcomes, prevent CDI and therefore prevent further antibiotic use. It would need to be determined whether there is any cross reactivity between these potential vaccine candidates and the healthy microbiota, including other *Clostridial* species, which could it could lead to unintended effects on the microbiome. To answer this, work in ongoing with the WTSI, to look for orthologs of the protein antigens discussed here within their database of sequenced strains of the gut microbiome. Furthermore, only one strain from each clade was available for analysis of protein conservation throughout *C. difficile*. Although the standard reference strains were used, there is still variation seen within clades and therefore future work should look at improving coverage of the different *C. difficile* strains investigated.

The second consideration is the interaction between the protein antigens and the antibodies that recognise them. Direct antibody/antigen interactions can be assessed using an immunofluorescence test [602]. However, it is not sufficient for the antigen to simply bind the protein, this must either prevent growth or result in an opsonophagocytic response, to clear the infection from the host [603]. The genes encoding the three proteins selected for *in vitro* characterisation, *CDR20291_0330*, *CDR20291_3343* and *CDR20291_0342* were all successfully deleted from the R20291 genome, suggesting these are not essential for survival of *C. difficile*, in the *in vitro* conditions tested. Instead, an opsonophagocytosis assay can be used to monitor antibody activity, an assay used routinely for evaluation of responses to pneumococcal vaccines, among others [604-606]. They could also be tested in a mouse model to determine whether they clear CDI or reduce the infective titre.

Finally, and most importantly for any vaccine candidate, is analysis of the antigens of interest *in vivo*, which was not possible in this study due to time, funds and absence of testing facilities. Use of a hamster or mouse model offer their own individual benefits and can elucidate immune response to these antigens. A standard approach to begin with would test each antigen individually, starting with intra-peritoneal administration with an adjuvant, of which many are available including Alum or inactivated cholera toxin [395, 396, 420, 422, 425-427]. Following initial inoculation, two booster doses would be administered before antibiotic-induced gut dysbiosis and challenge with CDI. CDI results in acute infection in hamster models, therefore animal survival can be taken as a measure of protection. As hamsters are very sensitive to the *C. difficile* toxins, inactivated copies of Toxins A and B which have been previously utilised as vaccine candidates these should be co-administered with the antigens of interest. Alternatively, a number of mouse models are available that can enable the exploration of other measures of protection such as colonisation. This method of assessing *in vivo* immunogenicity and protection for antigens of interest has been widely used for the study of *C. difficile* antigens [238, 267, 390, 396]. Following on from this, for promising candidates, oral routes of administration could be investigated, as this is the natural route of infection for CDI. Different delivery

systems have been investigated such as pectin beads and expression on the surface of *B. subtilis* spores, the latter of which was particularly successful [268, 276, 422].

8.2. *In vitro c*haracterisation of novel candidates

Three immunogenic array candidates were selected for further characterisation of their role within the cell; CDR20291_3343, the putative tip of the T4P, CDR20291_0330, a putative cobalt transport protein and CDR20291_0342, a putative ABC transporter. Characterisation and understanding of the role of a vaccine or drug target is an important aspect of its development. Other non-toxin *C. difficile* antigens investigated in previous studies have been functionally characterised, including in *C. difficile* FliC and SlpA [273, 447].

Deletion of *pilK* (R20291Δ*3343*) was deemed to result in a pili-negative strain, as previous work found PilA1 (the major pilin subunit) was not secreted in this mutant, suggesting lack of pilus assembly (Neil Fairweather, *personal communication*) [541]. In order to confirm this, work is ongoing with a collaborator at the Francis Crick Institute, using scanning electron microscopy to identify the pili on the surface of *C. difficile*. In future assays, ideally a *pilA1* mutant would also be used as no pili are observed on the surface of *C. difficile* by microscopy in a *pilA1* mutant [195].

In this study, a panel of pili and/or flagella negative strains were constructed in R20291 to assess *C. difficile* motility in the presence and absence of c-di-GMP. In high c-di-GMP, *fliC* expression is repressed whereas pili expression is induced, via the action of c-di-GMP on riboswitches upstream of these loci [198]. This study found the *pilK* mutant was still capable of surface motility and in fact it was inactivation of the flagella that almost abolished this motility. A previous study found inactivation of the putative pili ATPase, *pilB1,* abolishes surface motility on agar, but this was conducted in a flagellate strain [208]. This suggests a role for FliC in surface motility, even though *fliC* is down-regulated in the presence of high-level c-di-GMP [198].

One issue in the assessment of pili phenotypes in *C. difficile* is the need for c-di-GMP to induce expression [195, 203]. This secondary messenger has multiple targets and therefore can have an

overwhelming effect on the cell, and the levels of c-di-GMP can influence the observed phenotypes [203]. Therefore, future work could investigate introducing an inducible promoter onto the chromosome before *fliC* and the pili locus to improve control and measure the effects of each in a more defined manner, without the need for c-di-GMP. Alternatively, *pilK* could be overexpressed in the the *flic/pilK* mutant strain constructed in this study. These approaches could also be expanded to include analysis of other pili and flagella related phenotypes, such as cell aggregation and biofilm formation [189, 195, 198, 212].

The putative transporters CDR20291_0330 and CDR20291_0342 were also investigated for their role in *C. difficile*. In the conditions tested, it was not possible to determine a cobalt or B_{12} related phenotype *in vitro* for R20291Δ*0330*. In many bacterial species ethanolamine is used as a carbon source and is regulated by a B_{12} dependent riboswitch [501, 502]. Although we were unable to identify a B_{12} riboswitch upstream of or within the ethanolamine utilisation operon, to our knowledge, no B_{12} independent ethanolamine utilisation pathways have been identified in bacteria [504]. Herein, studies with ethanolamine found regulation of this pathway may differ from that previously reported for another enteric Gram-positive pathogen, *E. faecalis*, as only the histidine kinase (EutW, *CD1911*) part of the two-component regulatory system was required for ethanolamine utilisation [503]. It was also found that growth in ethanolamine could not be detected until 64 hours into the growth kinetics, as opposed to 10 hours in a previous study of *C. difficile* [36].

Regarding characterisation of CDR20291_0342, a Biolog assay found this strain did not grow on L-valine, L-valine and aspartic acid and alpha-ketovaleric acid, but the wild-type did. To further investigate this, R20291 and R20291Δ*0342* were grown in minimal media including six amino acids previously documented to be essential for *C. difficile* growth (valine, isoleucine, leucine, proline, cysteine and tryptophan) [32]. There was no difference in growth between the strains, but both strains still grew upon valine removal, suggesting this is not essential. [32]. Although the overnight cultures

used to inoculate the growth assays were also in minimal media, these contained valine to support the sufficient growth required to use this as an inoculant, therefore there is a chance for carry over.

It would be prudent to investigate these mutants in a mouse model of CDI, to determine whether inactivation of these transporters impairs colonisation. Furthermore, this would also be beneficial from a nutritional requirement stand point, as it is very challenging to recreate the niche environment of the gut *in vitro,* particularly as culturing *C. difficile* in minimal media can be difficult. This would be particularly interesting for the CDR20291_0330 mutant, as previous work found a mutant of *cobT*, a component of the B_{12} biosynthesis pathway, was unable to colonise mice, suggesting it is essential for survival within the host [498]. B_{12} is also believed to be a very important vitamin for the gut microbiome, and 80% of sequenced species within this community utilise corrinoids (compound group including B_{12}) [607, 608].

8.3. Bioconjugation of a *C. difficile*-specific glycoconjugate

Chapter 7 describes the investigation of bioconjugation technology as a means of producing a *C. difficile* specific glycoconjugate, using a truncated form of the flagella glycan from strain R20291. This involved a novel approach of combining enzymes from Gram-positive and Gram-negative bacteria to create a hybrid system to synthesise a native *O*-linked glycan using an *N*-linked glycosylation system.

A short, novel acceptor protein, AcrAtag, was designed and shown to be glycosylated at all seven sites with the positive control, C*. jejuni* heptasaccharide *in vitro.* However, despite multiple attempts using a range of techniques, it was not possible to identify glycosylation of this acceptor protein with the desired trisaccharide from the *C. difficile* flagella glycan. Indeed, each trisaccharide would only result in a 0.4 KDa change in size to the protein, and therefore successful glycosylation of all seven sites would result in a 2.8 KDa size shift which may not be visible by SDS-PAGE and subsequent immunoblot. Furthermore, there was still a distinct band of unglycosylated AcrAtag in the glycosylation experiments with the *C. jejuni* heptsaccharide. Therefore, if only a small percentage of the total protein is being glycosylated, this would also make it difficult to resolve this difference. To account for this, work is

ongoing with a collaborator at Leiden University, to screen for the glycan addition using the much more sensitive method of nuclear magnetic resonance (NMR) imaging.

Different techniques were applied to try and identify potential barriers to glycosylation such as transfer of the glycan from the cytoplasm to the periplasm using the *E. coli* flippase enzyme Wzx, the specificity of which is debatable [575, 591]. This transfer was not an issue for the *C. jejuni* glycan as it encodes its own specific flippase. The *in vitro* glycosylation approach was used as this is hypothesised to not require the translocation step and instead just needs glycan presented on a lipid-linked glycan (i.e. bound to UndPP) [583]. Again, glycosylation was not detected by SDS-PAGE and immunoblot, but herein we still have the issue of minimal size change, so this will also be screened by NMR as above. If glycosylation has not occurred, it is possible the glycan is not being synthesised. *In vitro* glycosylation is a relatively crude method so an alternative way to study glycan synthesis and glycosyltransferase interaction would be to purify the individual components required for glycosylation (PglB and each glycosyltransferase) then incubate these together *in vitro* [609, 610]. This would however require substantial optimisation to purify functional enzymes and identify appropriate buffers for their activity.

The R20291 flagella glycan was selected for reconstitution in *E. coli* as it has been characterised in detail, meaning this understanding could inform selection of the enzymes to be used for bioconjugation [187]. These were proof of principle experiments to determine whether this hybrid glycosylation system could be developed. Glycosylation of bacterial flagella has been linked to immune evasion, by masking important immunogenic epitopes [339], suggesting this might not be useful as a vaccine candidate. However, Valiente *et al.* found no difference between TLR5 recognition of the flagella protein in the absence and presence of the glycan, suggesting it does not mask this epitope [187, 611]. Alternatively, Bouche *et al.* suggested immune evasion could be achieved by the terminal peptidylamido sugar moiety which is taurine-like, an amino sulfonic acid associated with immune evasion [190]. The terminal moiety was omitted in the proof of principal study. R20291 harbouring an

absent or truncated flagella glycan has not been studied in animal models, therefore although it is possible the flagella glycan is involved in immune evasion in *C. difficile,* this remains hypothetical [187, 190]. Antibodies to the flagella filament proteins FliC and FliD were identified in 15/17 patients with CDI [612] while another study found antibody responses to FliC and FliD were higher in a control group compared to those with *C. difficile* associated disease [273].

It may be that a portion of this glycan is immunogenic. For example, removal of the peptidylamido sugar terminus, would expose the GlcNac di-Rhamnose targeted in this study. Rhamnose is often immunogenic as it is a non-human glycan [598]. Mutants in the R20291 glycosylation pathway, leading to truncated glycans have already been constructed [187], but not tested in an animal model, therefore this is a promising next step. Use of a mouse model could aid understanding of virulence and colonisation of R20291 strains with truncated or absent glycans. Furthermore, Martin *et al.* used glycan arrays to screen sera from mice immunised with the full repeating unit from surface polysaccharide PS-I against different portions of this repeat unit, to identify the epitopes within this that are immunogenic and recognised *in vivo* [419]. A similar study could be conducted for the flagella glycan to determine, which portions if any, are recognised by the immune system. Regarding glycan arrays, this approach could be expanded for analysis of the *C. difficile* glycome, screening both full and truncated glycans and polysaccharides to identify immunogenic structures and epitopes. Combining this with the existing bank of patient samples already collected for the protein array could produce informative results.

Although in this study it has not been possible to conclusively determine whether transfer of the glycan was successful, this does not mean it is an unachievable approach, especially considering this study was the first attempt at using bioconjugation with *C. difficile*. Bioconjugation is a large and rapidly advancing discipline, meaning new glycoengineering enzymes and techniques are continually being explored and developed. It would be interesting to use the results of a *C. difficile* glycan array to determine which glycans or polysaccharides would be useful for incorporation into a glycoconjugate

vaccine. Although OST (the enzyme that transfers the glycan to the acceptor protein) substrate specificity can be a stalling point in glycan transfer, this is the subject of ongoing investigation. Recent research has expanded the repertoire of glycans it is possible to transfer, through modification of existing OSTs to reduce specificity as well as identification of new OSTs with different specificities, such as those from deep sea vent bacteria [349, 350, 362, 613].

8.4. Future work and final conclusions

In addition to those described above, there are a number of areas for future work. For the protein array, the next steps should address those points raised in Chapter 3, including assessment of antibody functionality and characterisation of patient groups. As discussed, there are many factors influencing susceptibility to CDI other than antibody response. Over the course of this study, additional metadata has become available which is currently being processed, and will hopefully aid in understanding of co-morbidities and immune status of the patient. As anti-Toxin antibodies are correlated with exposure to *C. difficile*, anti-toxin response in each patient could be used as a marker of prior exposure to the pathogen. Regarding quality of the microbiome and its role in protection from CDI, the WTSI has sequenced the gut microbiome for a subset of patients using their faecal samples, so using this data can hopefully provide a picture of microbiome quality of the patient, which can then be linked back to the antibody response of the patient and their experience of CDI. Furthermore, as discussed, protein arrays use recombinantly expressed protein which may not fold to match the structure presented *in vivo*. As conformational epitopes are an important part of antigen recognition, flow cytometry could be used to monitor antibody binding to whole bacterial cells [614, 615].

Regarding those proteins taken forward for phenotypic characterisation in *C. difficile*, there are a number of avenues to pursue. Firstly, it would be prudent to whole genome sequence the three gene deletion mutants to ensure no other single nucleotide polymorphisms have been accrued during the mutagenesis process which may be influencing the phenotypes seen. For CDR20291_0330, it would be interesting to study cobalt and B_{12} uptake in *C. difficile*, which very little is known about. Flow

cytometry has previously been used to study the activity of bacterial multidrug efflux transporters in *E. coli, Francisella tularensis* and *Burkholderia pseudomallei*, [616, 617] so a similar approach could be applied for *C. difficile*, using labelled cobalt and B_{12}. As discussed in Chapter 4, the environment of the gut is difficult to replicate *in vitro* and is subject to a whole host of other factors, particularly metabolism of nutrients by the microbiome. It would therefore be interesting to look at the metagenomics and metabolomics of ethanolamine utilisation by the gut microbiome, but data related to this is currently limited [618]. One recent study in weaned rats found administration of ethanolamine resulted in a shift in microbiome metabolism towards lipid and sugar biosynthesis and metabolism and altered the composition of the predominant species composing the microbiome [619]. For CDR20291_0342, inactivation of the gene encoding this protein did not appear to impact valine utilisation, therefore CDR20291_0342 and its associated transporter may have an alternative target. It would be interesting to assess the protein composition of the cell membrane in the mutant compared to the wild-type to determine whether another transporter is compensating for the loss of CDR20291_0342. This could also give an indication as to the function of CDR20291_0342. Bacterial transporters have a wide array of functions, meaning there are a lot of possible phenotypes that could be screened for. For example, Edwards *et al.* identified two permease proteins in *C. difficile* which inhibit sporulation [620]. There are a number of areas for future work regarding CDR20291_3343 and the role of the pili and flagella in surface motility, particularly as both the flagella and c-di-GMP are within complex regulatory networks. This could be further investigated by looking at RNA seq data with and without c-di-GMP induction, ideally using colonies from the surface motility plates, comparing those within the original spot of inoculant to those on the outer edge. This could aid in identifying which proteins are contributing to the surface motility phenotypes in the different conditions reported here. For example, it may be that flagella proteins are required for function of the pili but these are not expressed in the *fliC* mutant. To follow up on this, it would be interesting to investigating *C. difficile* strain M120, which lacks the F3 region of the flagella operon, including the regulator SigD [205, 621].

The longer-term objective of this work, would be to take the promising candidates from the pan-protein array and fuse these to appropriate glycotag for bioconjugation with the desired glycan. This could use the binding protein CDR20291_0330, as a recent study demonstrated glycosylation of the streptococcal binding protein PiuA with the capsule 4 polysaccharide using bioconjugation resulted in protection in mice that was comparable to the commercially available vaccine [367]. Another alternative formula would be use of the binding domain of Toxin B, which can retain immunogenicity when fused [238, 578]. Further validation of the potential bioconjugation techniques evaluated here, provide an exciting concept for future vaccine designs.

This study has used a combination of approaches to both expand our understanding of *C. difficile* biology and contribute to the ongoing development of a *C. difficile* vaccine. The results of a large-scale screen of the *C. difficile* proteome were used to identify immunogenic proteins for potential inclusion within a *C. difficile* vaccine. Characterisation of these *in vitro* has led to further understanding of both motility and nutritional metabolism of *C. difficile*. Finally, the use of bioconjugation was explored for the development of *C. difficile* glycoconjugate vaccines.

References

1. Tedesco, F.J., R.W. Barton, and D.H. Alpers, *Clindamycin-associated colitis. A prospective study.* Ann Intern Med, 1974. **81**(4): p. 429-33.
2. Larson, H.E., et al., *Clostridium difficile and the aetiology of pseudomembranous colitis.* Lancet, 1978. **1**(8073): p. 1063-6.
3. Bartlett, J.G., et al., *Clindamycin-associated colitis due to a toxin-producing species of Clostridium in hamsters.* J Infect Dis, 1977. **136**(5): p. 701-5.
4. Bartlett, J.G., et al., *Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia.* N Engl J Med, 1978. **298**(10): p. 531-4.
5. Hall, I.C.O.T., E., *Intestinal flora in new-born infants: with a description of a new pathogenic anaerobe, Bacillus difficilis.* Am J Dis Child, 1935. **49**: p. 390-402.
6. Lawson, P.A., et al., *Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prevot 1938.* Anaerobe, 2016. **40**: p. 95-9.
7. Crobach, M.J.T., et al., *Understanding Clostridium difficile Colonization.* Clin Microbiol Rev, 2018. **31**(2).
8. Kuehne, S.A., et al., *The role of toxin A and toxin B in Clostridium difficile infection.* Nature, 2010. **467**(7316): p. 711-3.
9. Kuehne, S.A., et al., *Importance of toxin A, toxin B, and CDT in virulence of an epidemic Clostridium difficile strain.* J Infect Dis, 2014. **209**(1): p. 83-6.
10. Lyras, D., et al., *Toxin B is essential for virulence of Clostridium difficile.* Nature, 2009. **458**(7242): p. 1176-9.
11. Mora Pinzon, M.C., et al., *Outcomes of Community and Healthcare-onset Clostridium difficile Infections.* Clin Infect Dis, 2018. **68**(8): p. 1343-1350.
12. Sayedy, L., D. Kothari, and R.J. Richards, *Toxic megacolon associated Clostridium difficile colitis.* World journal of gastrointestinal endoscopy, 2010. **2**(8): p. 293-297.
13. Pascal, V., et al., *A microbial signature for Crohn's disease.* Gut, 2017. **66**(5): p. 813.
14. Ni, J., et al., *A role for bacterial urease in gut dysbiosis and Crohn's disease.* Sci Transl Med, 2017. **9**(416).
15. Franzosa, E.A., et al., *Gut microbiome structure and metabolic activity in inflammatory bowel disease.* Nature Microbiology, 2019. **4**(2): p. 293-305.
16. Lazar, V., et al., *Aspects of Gut Microbiota and Immune System Interactions in Infectious Diseases, Immunopathology, and Cancer.* Frontiers in Immunology, 2018. **9**(1830).
17. Claesson, M.J., et al., *Gut microbiota composition correlates with diet and health in the elderly.* Nature, 2012. **488**(7410): p. 178-84.
18. Costello, E.K., et al., *Bacterial community variation in human body habitats across space and time.* Science, 2009. **326**(5960): p. 1694-7.
19. Wu, G.D., et al., *Linking long-term dietary patterns with gut microbial enterotypes.* Science, 2011. **334**(6052): p. 105-8.
20. Claesson, M.J., et al., *Composition, variability, and temporal stability of the intestinal microbiota of the elderly.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(Suppl 1): p. 4586-4591.
21. Zwielehner, J., et al., *Combined PCR-DGGE fingerprinting and quantitative-PCR indicates shifts in fecal population sizes and diversity of Bacteroides, bifidobacteria and Clostridium cluster IV in institutionalized elderly.* Exp Gerontol, 2009. **44**(6-7): p. 440-6.
22. Dethlefsen, L., et al., *The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing.* PLoS Biol, 2008. **6**(11): p. e280.
23. Buffie, C.G., et al., *Profound Alterations of Intestinal Microbiota following a Single Dose of Clindamycin Results in Sustained Susceptibility to Clostridium difficile-Induced Colitis.* Infection and Immunity, 2012. **80**(1): p. 62-73.

24. Lewis, B.B., et al., *Loss of Microbiota-Mediated Colonization Resistance to Clostridium difficile Infection With Oral Vancomycin Compared With Metronidazole.* J Infect Dis, 2015. **212**(10): p. 1656-65.
25. Buonomo, E.L. and W.A. Petri, Jr., *The microbiota and immune response during Clostridium difficile infection.* Anaerobe, 2016. (41): p. 79-84.
26. Louie, T.J., et al., *Fidaxomicin preserves the intestinal microbiome during and after treatment of Clostridium difficile infection (CDI) and reduces both toxin reexpression and recurrence of CDI.* Clin Infect Dis, 2012. **55** (Suppl 2): p. S132-42.
27. Bäumler, A.J. and V. Sperandio, *Interactions between the microbiota and pathogenic bacteria in the gut.* Nature, 2016. **535**(7610): p. 85-93.
28. Wilson, K.H., *Efficiency of various bile salt preparations for stimulation of Clostridium difficile spore germination.* J Clin Microbiol, 1983. **18**(4): p. 1017-9.
29. Sorg, J.A. and A.L. Sonenshein, *Bile salts and glycine as cogerminants for Clostridium difficile spores.* J Bacteriol, 2008. **190**(7): p. 2505-12.
30. Buffie, C.G., et al., *Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile.* Nature, 2015. **517**(7533): p. 205-8.
31. Mullish, B.H., et al., *Microbial bile salt hydrolases mediate the efficacy of faecal microbiota transplant in the treatment of recurrent Clostridioides difficile infection.* Gut, 2019. **68**(100: p. 1791-1800.
32. Karasawa, T., et al., *A defined growth medium for Clostridium difficile.* Microbiology, 1995. **141** (Pt 2): p. 371-5.
33. Bouillaut, L., W.T. Self, and A.L. Sonenshein, *Proline-dependent regulation of Clostridium difficile Stickland metabolism.* J Bacteriol, 2013. **195**(4): p. 844-54.
34. Karlsson, S., et al., *Toxins, butyric acid, and other short-chain fatty acids are coordinately expressed and down-regulated by cysteine in Clostridium difficile.* Infect Immun, 2000. **68**(10): p. 5881-8.
35. Janoir, C., et al., *Adaptive strategies and pathogenesis of Clostridium difficile from in vivo transcriptomics.* Infect Immun, 2013. **81**(10): p. 3757-69.
36. Nawrocki, K.L., et al., *Ethanolamine is a valuable nutrient source that impacts Clostridium difficile pathogenesis.* Environ Microbiol, 2018. **20**(4): p. 1419-1435.
37. Theriot, C.M., et al., *Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection.* Nat Commun, 2014. **5**: p. 3114.
38. Collins, J., et al., *Dietary trehalose enhances virulence of epidemic Clostridium difficile.* Nature, 2018. **553**(7688): p. 291-294.
39. Collignon, A., et al., *Heterogeneity of Clostridium difficile isolates from infants.* Eur J Pediatr, 1993. **152**(4): p. 319-22.
40. Enoch, D.A., et al., *Clostridium difficile in children: colonisation and disease.* J Infect, 2011. **63**(2): p. 105-13.
41. Rousseau, C., et al., *Prevalence and diversity of Clostridium difficile strains in infants.* J Med Microbiol, 2011. **60**(Pt 8): p. 1112-8.
42. Bolton, R.P., et al., *Asymptomatic neonatal colonisation by Clostridium difficile.* Arch Dis Child, 1984. **59**(5): p. 466-72.
43. Eglow, R., et al., *Diminished Clostridium difficile toxin A sensitivity in newborn rabbit ileum is associated with decreased toxin A receptor.* J Clin Invest, 1992. **90**(3): p. 822-9.
44. Nicholson, M.R., I.P. Thomsen, and K.M. Edwards, *Controversies Surrounding Clostridium difficile Infection in Infants and Young Children.* Children (Basel, Switzerland), 2014. **1**(1): p. 40-47.
45. Zacharioudakis, I.M., et al., *Colonization with toxinogenic C. difficile upon hospital admission, and risk of infection: a systematic review and meta-analysis.* Am J Gastroenterol, 2015. **110**(3): p. 381-90.

46. Blixt, T., et al., *Asymptomatic Carriers Contribute to Nosocomial Clostridium difficile Infection: A Cohort Study of 4508 Patients.* Gastroenterology, 2017. **152**(5): p. 1031-1041.
47. Lanzas, C., et al., *Epidemiological model for Clostridium difficile transmission in healthcare settings.* Infect Control Hosp Epidemiol, 2011. **32**(6): p. 553-61.
48. Riggs, M.M., et al., *Asymptomatic Carriers Are a Potential Source for Transmission of Epidemic and Nonepidemic Clostridium difficile Strains among Long-Term Care Facility Residents.* Clinical Infectious Diseases, 2007. **45**(8): p. 992-998.
49. Curry, S.R., et al., *Use of multilocus variable number of tandem repeats analysis genotyping to determine the role of asymptomatic carriers in Clostridium difficile transmission.* Clin Infect Dis, 2013. **57**(8): p. 1094-102.
50. Eyre, D.W., et al., *Asymptomatic Clostridium difficile colonisation and onward transmission.* PLoS One, 2013. **8**(11): p. e78445.
51. Lawley, T.D., et al., *Antibiotic treatment of clostridium difficile carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts.* Infection and immunity, 2009. **77**(9): p. 3661-3669.
52. Smits, W.K., et al., *Clostridium difficile infection.* Nature reviews. Disease primers, 2016. **2**: p. 16020-16020.
53. Carter, G.P., et al., *Defining the Roles of TcdA and TcdB in Localized Gastrointestinal Disease, Systemic Organ Damage, and the Host Response during Clostridium difficile Infections.* mBio, 2015. **6**(3): p. e00551-15.
54. Dobson, G., C. Hickey, and J. Trinder, *Clostridium difficile colitis causing toxic megacolon, severe sepsis and multiple organ dysfunction syndrome.* Intensive Care Med, 2003. **29**(6): p. 1030.
55. Roy, M., K. Dahal, and A.K. Roy, *Invading beyond bounds: extraintestinal Clostridium difficile infection leading to pancreatic and liver abscesses.* BMJ Case Rep, 2017. **28**.
56. Public Health England. *30-day all-cause fatality subsequent to MRSA, MSSA and Gram-negative bacteraemia and C. difficile infections, 2017/18.* 2018. (https://www.gov.uk/government/statistics/mrsa-mssa-and-e-coli-bacteraemia-and-c-difficile-infection-30-day-all-cause-fatality)
57. Zar, F.A., et al., *A comparison of vancomycin and metronidazole for the treatment of Clostridium difficile-associated diarrhea, stratified by disease severity.* Clin Infect Dis, 2007. **45**(3): p. 302-7.
58. Musher, D.M., et al., *Relatively poor outcome after treatment of Clostridium difficile colitis with metronidazole.* Clin Infect Dis, 2005. **40**(11): p. 1586-90.
59. Song, J.H. and Y.S. Kim, *Recurrent Clostridium difficile Infection: Risk Factors, Treatment, and Prevention.* Gut Liver, 2019. **13**(1): p. 16-24.
60. Doh, Y.S., et al., *Long-Term Clinical Outcome of Clostridium difficile Infection in Hospitalized Patients: A Single Center Study.* Intest Res, 2014. **12**(4): p. 299-305.
61. Barbut, F., et al., *Epidemiology of recurrences or reinfections of Clostridium difficile-associated diarrhea.* Journal of clinical microbiology, 2000. **38**(6): p. 2386-2388.
62. Dharbhamulla, N., et al., *Risk Factors Associated With Recurrent Clostridium difficile Infection.* Journal of clinical medicine research, 2019. **11**(1): p. 1-6.
63. Deshpande, A., et al., *Risk factors for recurrent Clostridium difficile infection: a systematic review and meta-analysis.* Infect Control Hosp Epidemiol, 2015. **36**(4): p. 452-60.
64. Rupnik, M., M.H. Wilcox, and D.N. Gerding, *Clostridium difficile infection: new developments in epidemiology and pathogenesis.* Nat Rev Microbiol, 2009. **7**(7): p. 526-36.
65. Slimings, C. and T.V. Riley, *Antibiotics and hospital-acquired Clostridium difficile infection: update of systematic review and meta-analysis.* J Antimicrob Chemother, 2014. **69**(4): p. 881-91.
66. Johnson, S., et al., *Epidemics of diarrhea caused by a clindamycin-resistant strain of Clostridium difficile in four hospitals.* N Engl J Med, 1999. **341**(22): p. 1645-51.

67. Cartman, S.T., et al., *The emergence of 'hypervirulence' in Clostridium difficile.* Int J Med Microbiol, 2010. **300**(6): p. 387-95.
68. He, M., et al., *Emergence and global spread of epidemic healthcare-associated Clostridium difficile.* Nat Genet, 2013. **45**(1): p. 109-13.
69. Loo, V.G., et al., *A predominantly clonal multi-institutional outbreak of Clostridium difficile-associated diarrhea with high morbidity and mortality.* N Engl J Med, 2005. **353**(23): p. 2442-9.
70. Dingle, K.E., et al., *Effects of control interventions on Clostridium difficile infection in England: an observational study.* Lancet Infect Dis, 2017. **17**(4): p. 411-421.
71. Tleyjeh, I.M., et al., *Association between proton pump inhibitor therapy and clostridium difficile infection: a contemporary systematic review and meta-analysis.* PLoS One, 2012. **7**(12): p. e50836.
72. McDonald, E.G., et al., *Continuous Proton Pump Inhibitor Therapy and the Associated Risk of Recurrent Clostridium difficile Infection.* JAMA Intern Med, 2015. **175**(5): p. 784-91.
73. Novack, L., et al., *Acid suppression therapy does not predispose to Clostridium difficile infection: the case of the potential bias.* PLoS One, 2014. **9**(10): p. e110790.
74. Kwok, C.S., et al., *Risk of Clostridium difficile infection with acid suppressing drugs and antibiotics: meta-analysis.* Am J Gastroenterol, 2012. **107**(7): p. 1011-9.
75. Public Health England, *Updated guidance on the management and treatment of Clostridium difficile infection*. 2019 (https://www.gov.uk/government/publications/clostridium-difficile-infection-guidance-on-management-and-treatment).
76. Furuya-Kanamori, L., et al., *Comorbidities, Exposure to Medications, and the Risk of Community-Acquired Clostridium difficile Infection: a systematic review and meta-analysis.* Infect Control Hosp Epidemiol, 2015. **36**(2): p. 132-41.
77. Vardakas, K.Z., et al., *Risk factors for development of Clostridium difficile infection due to BI/NAP1/027 strain: a meta-analysis.* International Journal of Infectious Diseases, 2012. **16**(11): p. e768-e773.
78. Loo, V.G., et al., *Host and pathogen factors for Clostridium difficile infection and colonization.* N Engl J Med, 2011. **365**(18): p. 1693-703.
79. Gao, L., et al., *Medication usage change in older people (65+) in England over 20 years: findings from CFAS I and CFAS II.* Age Ageing, 2018. **47**(2): p. 220-225.
80. Olsen, M.A., et al., *Increasing Age Has Limited Impact on Risk of Clostridium difficile Infection in an Elderly Population.* Open forum infectious diseases, 2018. **5**(7): p. 160.
81. Jump, R.L., *Clostridium difficile infection in older adults.* Aging health, 2013. **9**(4): p. 403-414.
82. Dharmarajan, T., et al., *Co-morbidity, not age predicts adverse outcome in clostridium difficile colitis.* World J Gastroenterol, 2000. **6**(2): p. 198-201.
83. Wilcox, M.H., et al., *A case–control study of community-associated Clostridium difficile infection.* Journal of Antimicrobial Chemotherapy, 2008. **62**(2): p. 388-396.
84. Woodmansey, E.J., *Intestinal bacteria and ageing.* J Appl Microbiol, 2007. **102**(5): p. 1178-86.
85. Rea, M.C., et al., *Clostridium difficile carriage in elderly subjects and associated changes in the intestinal microbiota.* J Clin Microbiol, 2012. **50**(3): p. 867-75.
86. Kyne, L., et al., *Asymptomatic carriage of Clostridium difficile and serum levels of IgG antibody against toxin A.* N Engl J Med, 2000. **342**(6): p. 390-7.
87. Gupta, S.B., et al., *Antibodies to Toxin B Are Protective Against Clostridium difficile Infection Recurrence.* Clin Infect Dis, 2016. **63**(6): p. 730-734.
88. McGlauchlen, K.S. and L.A. Vogel, *Ineffective humoral immunity in the elderly.* Microbes Infect, 2003. **5**(13): p. 1279-84.
89. Kyne, L., et al., *Association between antibody response to toxin A and protection against recurrent Clostridium difficile diarrhoea.* Lancet, 2001. **357**(9251): p. 189-93.

90. Shin, J.H., K.P. High, and C.A. Warren, *Older Is Not Wiser, Immunologically Speaking: Effect of Aging on Host Response to Clostridium difficile Infections.* J Gerontol A Biol Sci Med Sci, 2016. **71** (7): p. 916-922.
91. Adejumo, A.C., O. Akanbi, and L. Pani, *Among inpatients, ischemic bowel disease predisposes to Clostridium difficile infection with concomitant higher mortality and worse outcomes.* Eur J Gastroenterol Hepatol, 2019. **31**(1): p. 109-115.
92. Ticinesi, A., et al., *Multimorbidity in elderly hospitalised patients and risk of Clostridium difficile infection: a retrospective study with the Cumulative Illness Rating Scale (CIRS).* BMJ Open, 2015. **5**(10): p. e009316.
93. Leffler, D.A. and J.T. Lamont, *Clostridium difficile infection.* N Engl J Med, 2015. **372**(16): p. 1539-48.
94. Public Health England., *Updated guidance on the diagnosis and reporting of Clostridium Difficile.* 2019. (https://www.gov.uk/government/publications/updated-guidance-on-the-diagnosis-and-reporting-of-clostridium-difficile).
95. Carey-Ann, B.D. and K.C. Carroll, *Diagnosis of Clostridium difficile Infection: an Ongoing Conundrum for Clinicians and for Clinical Laboratories.* Clinical Microbiology Reviews, 2013. **26**(3): p. 604.
96. Bayardelle, P., *Importance of culture for detection of Clostridium difficile toxin from stool samples to report true incidence and mortality related to C. difficile in hospitals.* Clin Infect Dis, 2009. **49**(7): p. 1134-5.
97. Eastwood, K., et al., *Comparison of Nine Commercially Available Clostridium difficile Toxin Detection Assays, a Real-Time PCR Assay for C. difficile tcdB, and a Glutamate Dehydrogenase Detection Assay to Cytotoxin Testing and Cytotoxigenic Culture Methods.* Journal of Clinical Microbiology, 2009. **47**(10): p. 3211.
98. Planche, T.D., et al., *Differences in outcome according to Clostridium difficile testing method: a prospective multicentre diagnostic validation study of C. difficile infection.* Lancet Infect Dis, 2013. **13**(11): p. 936-45.
99. Guery, B., T. Galperine, and F. Barbut, *Clostridioides difficile: diagnosis and treatments.* BMJ, 2019. **366**: p. 4609.
100. Polage, C.R., et al., *Overdiagnosis of Clostridium difficile Infection in the Molecular Test Era.* JAMA internal medicine, 2015. **175**(11): p. 1792-1801.
101. Crobach, M.J., et al., *European Society of Clinical Microbiology and Infectious Diseases: update of the diagnostic guidance document for Clostridium difficile infection.* Clin Microbiol Infect, 2016. **22** (Suppl 4): p. S63-81.
102. Knight, D.R., et al., *Diversity and Evolution in the Genome of Clostridium difficile.* Clin Microbiol Rev, 2015. **28**(3): p. 721-41.
103. Rupnik, M., et al., *A novel toxinotyping scheme and correlation of toxinotypes with serogroups of Clostridium difficile isolates.* J Clin Microbiol, 1998. **36**(8): p. 2240-7.
104. Griffiths, D., et al., *Multilocus Sequence Typing of Clostridium difficile.* Journal of Clinical Microbiology, 2010. **48**(3): p. 770.
105. Lemee, L., et al., *Multilocus sequence analysis and comparative evolution of virulence-associated genes and housekeeping genes of Clostridium difficile.* Microbiology, 2005. **151**(Pt 10): p. 3171-80.
106. Rupnik, M., et al., *Comparison of toxinotyping and PCR ribotyping of Clostridium difficile strains and description of novel toxinotypes.* Microbiology, 2001. **147**(Pt 2): p. 439-47.
107. Gurtler, V., *Typing of Clostridium difficile strains by PCR-amplification of variable length 16S-23S rDNA spacer regions.* J Gen Microbiol, 1993. **139**(12): p. 3089-97.
108. Wilcox, M.H., et al., *Changing epidemiology of Clostridium difficile infection following the introduction of a national ribotyping-based surveillance scheme in England.* Clin Infect Dis, 2012. **55**(8): p. 1056-63.

109. Stabler, R.A., et al., *Macro and micro diversity of Clostridium difficile isolates from diverse sources and geographical locations.* PLoS One, 2012. **7**(3): p. e31559.
110. Wust, J., et al., *Investigation of an outbreak of antibiotic-associated colitis by various typing methods.* J Clin Microbiol, 1982. **16**(6): p. 1096-101.
111. Sebaihia, M., et al., *The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome.* Nat Genet, 2006. **38**(7): p. 779-86.
112. Kuijper, E.J., B. Coignard, and P. Tull, *Emergence of Clostridium difficile-associated disease in North America and Europe.* Clin Microbiol Infect, 2006. **12** (Suppl 6): p. 2-18.
113. Stabler, R.A., et al., *Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium.* Genome Biol, 2009. **10**(9): p. 102.
114. van den Berg, R.J., et al., *Characterization of toxin A-negative, toxin B-positive Clostridium difficile isolates from outbreaks in different countries by amplified fragment length polymorphism and PCR ribotyping.* J Clin Microbiol, 2004. **42**(3): p. 1035-41.
115. Wren, M.W., et al., *Detection of Clostridium difficile infection: a suggested laboratory diagnostic algorithm.* Br J Biomed Sci, 2009. **66**(4): p. 175-9.
116. Goorhuis, A., et al., *Emergence of Clostridium difficile infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078.* Clin Infect Dis, 2008. **47**(9): p. 1162-70.
117. Knetsch, C.W., et al., *Whole genome sequencing reveals potential spread of Clostridium difficile between humans and farm animals in the Netherlands, 2002 to 2011.* Euro Surveill, 2014. **19**(45): p. 20954.
118. McDonald, L.C., et al., *An epidemic, toxin gene-variant strain of Clostridium difficile.* N Engl J Med, 2005. **353**(23): p. 2433-41.
119. Pepin, J., et al., *Increasing risk of relapse after treatment of Clostridium difficile colitis in Quebec, Canada.* Clin Infect Dis, 2005. **40**(11): p. 1591-7.
120. Smith, A., *Outbreak of Clostridium difficile infection in an English hospital linked to hypertoxin-producing strains in Canada and the US.* Euro Surveill, 2005. **10**(6): p. e050630.2.
121. Labbé, A.-C., et al., *Clostridium difficile infections in a Canadian tertiary care hospital before and during a regional epidemic associated with the BI/NAP1/027 strain.* Antimicrobial agents and chemotherapy, 2008. **52**(9): p. 3180-3187.
122. Warny, M., 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**(9491): p. 1079-84.
123. Drudy, D., et al., *gyrA mutations in fluoroquinolone-resistant Clostridium difficile PCR-027.* Emerging infectious diseases, 2007. **13**(3): p. 504-505.
124. Burns, D.A., J.T. Heap, and N.P. Minton, *The diverse sporulation characteristics of Clostridium difficile clinical isolates are not associated with type.* Anaerobe, 2010. **16**(6): p. 618-22.
125. Moore, P., et al., *Germination efficiency of clinical Clostridium difficile spores and correlation with ribotype, disease severity and therapy failure.* J Med Microbiol, 2013. **62**(Pt 9): p. 1405-13.
126. Dupuy, B., et al., *Clostridium difficile toxin synthesis is negatively regulated by TcdC.* J Med Microbiol, 2008. **57**(Pt 6): p. 685-9.
127. Smits, W.K., *Hype or hypervirulence: a reflection on problematic C. difficile strains.* Virulence, 2013. **4**(7): p. 592-596.
128. Public Health England., *Annual epidemiological commentary: Gram-negative, MRSA and MSSA bacteraemia and C. difficile infection data, up to and including financial year April 2018 to March 2019.* 2019 (https://www.gov.uk/government/statistics/mrsa-mssa-and-e-coli-bacteraemia-and-c-difficile-infection-annual-epidemiological-commentary).
129. Chitnis, A.S., et al., *Epidemiology of Community-Associated Clostridium difficile Infection, 2009 Through 2011.* JAMA Internal Medicine, 2013. **173**(14): p. 1359-1367.

130. Khanna, S. and D.S. Pardi, *The growing incidence and severity of Clostridium difficile infection in inpatient and outpatient settings.* Expert Rev Gastroenterol Hepatol, 2010. **4**(4): p. 409-16.
131. Khanna, S., et al., *The epidemiology of community-acquired Clostridium difficile infection: a population-based study.* The American journal of gastroenterology, 2012. **107**(1): p. 89-95.
132. Jump, R.L.P., M.J. Pultz, and C.J. Donskey, *Vegetative Clostridium difficile survives in room air on moist surfaces and in gastric contents with reduced acidity: a potential mechanism to explain the association between proton pump inhibitors and C. difficile-associated diarrhea?* Antimicrobial agents and chemotherapy, 2007. **51**(8): p. 2883-2887.
133. al Saif, N. and J.S. Brazier, *The distribution of Clostridium difficile in the environment of South Wales.* J Med Microbiol, 1996. **45**(2): p. 133-7.
134. Kim, K.H., et al., *Isolation of Clostridium difficile from the environment and contacts of patients with antibiotic-associated colitis.* J Infect Dis, 1981. **143**(1): p. 42-50.
135. Rupnik, M., *Is Clostridium difficile-associated infection a potentially zoonotic and foodborne disease?* Clin Microbiol Infect, 2007. **13**(5): p. 457-9.
136. Pelaez, T., et al., *Characterization of swine isolates of Clostridium difficile in Spain: a potential source of epidemic multidrug resistant strains?* Anaerobe, 2013. **22**: p. 45-9.
137. Riley, T.V., et al., *Gastrointestinal carriage of Clostridium difficile in cats and dogs attending veterinary clinics.* Epidemiol Infect, 1991. **107**(3): p. 659-65.
138. Borriello, S.P., et al., *Household pets as a potential reservoir for Clostridium difficile infection.* Journal of Clinical Pathology, 1983. **36**(1): p. 84.
139. Rabold, D., et al., *The zoonotic potential of Clostridium difficile from small companion animals and their owners.* PLoS One, 2018. **13**(2): p. e0193411.
140. Keel, K., et al., *Prevalence of PCR Ribotypes among Clostridium difficile Isolates from Pigs, Calves, and Other Species.* Journal of Clinical Microbiology, 2007. **45**(6): p. 1963.
141. Dingle, K.E., et al., *A Role for Tetracycline Selection in Recent Evolution of Agriculture-Associated Clostridium difficile PCR Ribotype 078.* MBio, 2019. **10**(2).
142. Burns, K., et al., *Infection due to C. difficile ribotype 078: first report of cases in the Republic of Ireland.* J Hosp Infect, 2010. **75**(4): p. 287-91.
143. Jhung, M.A., et al., *Toxinotype V Clostridium difficile in humans and food animals.* Emerg Infect Dis, 2008. **14**(7): p. 1039-45.
144. Martin, J.S., T.M. Monaghan, and M.H. Wilcox, *Clostridium difficile infection: epidemiology, diagnosis and understanding transmission.* Nat Rev Gastroenterol Hepatol, 2016. **13**(4): p. 206-16.
145. Balsells, E., et al., *Global burden of Clostridium difficile infections: a systematic review and meta-analysis.* J Glob Health, 2019. **9**(1): p. 010407.
146. Balsells, E., et al., *Infection prevention and control of Clostridium difficile: a global review of guidelines, strategies, and recommendations.* J Glob Health, 2016. **6**(2): p. 020410.
147. Burke, K.E. and J.T. Lamont, *Clostridium difficile infection: a worldwide disease.* Gut Liver, 2014. **8**(1): p. 1-6.
148. Cairns, M.D., et al., *Comparative Genome Analysis and Global Phylogeny of the Toxin Variant Clostridium difficile PCR Ribotype 017 Reveals the Evolution of Two Independent Sublineages.* Journal of Clinical Microbiology, 2017. **55**(3): p. 865-876.
149. Roldan, G.A., A.X. Cui, and N.R. Pollock, *Assessing the Burden of Clostridium difficile Infection in Low- and Middle-Income Countries.* J Clin Microbiol, 2018. **56**(3).
150. Rajabally, N., et al., *A comparison of Clostridium difficile diagnostic methods for identification of local strains in a South African centre.* J Med Microbiol, 2016. **65**(4): p. 320-327.
151. Pruitt, R.N., et al., *Structural organization of the functional domains of Clostridium difficile toxins A and B.* Proc Natl Acad Sci U S A, 2010. **107**(30): p. 13467-72.
152. Frisch, C., et al., *The complete receptor-binding domain of Clostridium difficile toxin A is required for endocytosis.* Biochem Biophys Res Commun, 2003. **300**(3): p. 706-11.

153. Papatheodorou, P., et al., *Clostridial glucosylating toxins enter cells via clathrin-mediated endocytosis.* PLoS One, 2010. **5**(5): p. e10673.
154. LaFrance, M.E., et al., *Identification of an epithelial cell receptor responsible for Clostridium difficile TcdB-induced cytotoxicity.* Proc Natl Acad Sci U S A, 2015. **112**(22): p. 7073-8.
155. Na, X., et al., *gp96 is a human colonocyte plasma membrane binding protein for Clostridium difficile toxin A.* Infect Immun, 2008. **76**(7): p. 2862-71.
156. Barth, H., et al., *Low pH-induced formation of ion channels by clostridium difficile toxin B in target cells.* J Biol Chem, 2001. **276**(14): p. 10670-6.
157. Qa'Dan, M., L.M. Spyres, and J.D. Ballard, *pH-induced conformational changes in Clostridium difficile toxin B.* Infect Immun, 2000. **68**(5): p. 2470-4.
158. Just, I., et al., *Glucosylation of Rho proteins by Clostridium difficile toxin B.* Nature, 1995. **375**(6531): p. 500-3.
159. Just, I., et al., *The enterotoxin from Clostridium difficile (ToxA) monoglucosylates the Rho proteins.* J Biol Chem, 1995. **270**(23): p. 13932-6.
160. Braun, V., et al., *Definition of the single integration site of the pathogenicity locus in Clostridium difficile.* Gene, 1996. **181**(1-2): p. 29-38.
161. Govind, R. and B. Dupuy, *Secretion of Clostridium difficile toxins A and B requires the holin-like protein TcdE.* PLoS Pathog, 2012. **8**(6): p. e1002727.
162. Moncrief, J.S., L.A. Barroso, and T.D. Wilkins, *Positive regulation of Clostridium difficile toxins.* Infect Immun, 1997. **65**(3): p. 1105-8.
163. Matamouros, S., P. England, and B. Dupuy, *Clostridium difficile toxin expression is inhibited by the novel regulator TcdC.* Mol Microbiol, 2007. **64**(5): p. 1274-88.
164. Carter, G.P., et al., *The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of Clostridium difficile.* PLoS Pathog, 2011. **7**(10): p. e1002317.
165. Carter, G.P., et al., *Binary toxin production in Clostridium difficile is regulated by CdtR, a LytTR family response regulator.* J Bacteriol, 2007. **189**(20): p. 7290-301.
166. Popoff, M.R., et al., *Actin-specific ADP-ribosyltransferase produced by a Clostridium difficile strain.* Infect Immun, 1988. **56**(9): p. 2299-306.
167. Schwan, C., et al., *Clostridium difficile toxin CDT hijacks microtubule organization and reroutes vesicle traffic to increase pathogen adherence.* Proc Natl Acad Sci U S A, 2014. **111**(6): p. 2313-8.
168. Papatheodorou, P., et al., *Lipolysis-stimulated lipoprotein receptor (LSR) is the host receptor for the binary toxin Clostridium difficile transferase (CDT).* Proc Natl Acad Sci U S A, 2011. **108**(39): p. 16422-7.
169. Best, E.L., J. Freeman, and M.H. Wilcox, *Models for the study of Clostridium difficile infection.* Gut microbes, 2012. **3**(2): p. 145-167.
170. Geric, B., et al., *Binary toxin-producing, large clostridial toxin-negative Clostridium difficile strains are enterotoxic but do not cause disease in hamsters.* J Infect Dis, 2006. **193**(8): p. 1143-50.
171. Sawabe, E., et al., *Molecular analysis of Clostridium difficile at a university teaching hospital in Japan: a shift in the predominant type over a five-year period.* Eur J Clin Microbiol Infect Dis, 2007. **26**(10): p. 695-703.
172. Drudy, D., S. Fanning, and L. Kyne, *Toxin A-negative, toxin B-positive Clostridium difficile.* Int J Infect Dis, 2007. **11**(1): p. 5-10.
173. Buckley, A.M., et al., *Susceptibility of hamsters to Clostridium difficile isolates of differing toxinotype.* PloS one, 2013. **8**(5): p. e64121-e64121.
174. Natarajan, M., et al., *A clinical and epidemiological review of non-toxigenic Clostridium difficile.* Anaerobe, 2013. **22**: p. 1-5.
175. Pickering, D.S., et al., *Investigating the transient and persistent effects of heat on Clostridium difficile spores.* J Med Microbiol, 2019. **68** (10): p. 1445-1454.

176. Dawson, L.F., et al., *Hypervirulent Clostridium difficile PCR-ribotypes exhibit resistance to widely used disinfectants.* PLoS One, 2011. **6**(10): p. e25754.
177. Lawley, T.D., et al., *Use of purified Clostridium difficile spores to facilitate evaluation of health care disinfection regimens.* Appl Environ Microbiol, 2010. **76**(20): p. 6895-900.
178. Deakin, L.J., et al., *The Clostridium difficile spo0A gene is a persistence and transmission factor.* Infect Immun, 2012. **80**(8): p. 2704-11.
179. Francis, M.B., et al., *Bile acid recognition by the Clostridium difficile germinant receptor, CspC, is important for establishing infection.* PLoS Pathog, 2013. **9**(5): p. e1003356.
180. Shrestha, R., A.M. Cochran, and J.A. Sorg, *The requirement for co-germinants during Clostridium difficile spore germination is influenced by mutations in yabG and cspA.* PLoS Pathog, 2019. **15**(4): p. e1007681.
181. Marsh, J.W., et al., *Association of Relapse of Clostridium difficile Disease with BI/NAP1/027.* Journal of Clinical Microbiology, 2012. **50**(12): p. 4078.
182. Landelle, C., et al., *Contamination of healthcare workers' hands with Clostridium difficile spores after caring for patients with C. difficile infection.* Infect Control Hosp Epidemiol, 2014. **35**(1): p. 10-5.
183. Sjoberg, M., et al., *Transmission of Clostridium difficile spores in isolation room environments and through hospital beds.* Apmis, 2014. **122**(9): p. 800-3.
184. Faulds-Pain, A., et al., *The post-translational modification of the Clostridium difficile flagellin affects motility, cell surface properties and virulence.* Molecular Microbiology, 2014. **94**(2): p. 272-289.
185. Twine, S.M., et al., *Motility and flagellar glycosylation in Clostridium difficile.* J Bacteriol, 2009. **191**(22): p. 7050-62.
186. Dingle, T.C., G.L. Mulvey, and G.D. Armstrong, *Mutagenic analysis of the Clostridium difficile flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters.* Infect Immun, 2011. **79**(10): p. 4061-7.
187. Valiente, E., et al., *Role of Glycosyltransferases Modifying Type B Flagellin of Emerging Hypervirulent Clostridium difficile Lineages and Their Impact on Motility and Biofilm Formation.* J Biol Chem, 2016. **291**(49): p. 25450-25461.
188. Baban, S.T., et al., *The role of flagella in Clostridium difficile pathogenesis: comparison between a non-epidemic and an epidemic strain.* PLoS One, 2013. **8**(9): p. e73026.
189. Dapa, T. and M. Unnikrishnan, *Biofilm formation by Clostridium difficile.* Gut Microbes, 2013. **4**(5): p. 397-402.
190. Bouche, L., et al., *The Type B Flagellin of Hypervirulent Clostridium difficile Is Modified with Novel Sulfonated Peptidylamido-glycans.* J Biol Chem, 2016. **291**(49): p. 25439-25449.
191. El Meouche, I., et al., *Characterization of the SigD Regulon of C. difficile and Its Positive Control of Toxin Production through the Regulation of tcdR.* PLOS ONE, 2013. **8**(12): p. e83748.
192. McKee, R.W., et al., *The second messenger cyclic Di-GMP regulates Clostridium difficile toxin production by controlling expression of sigD.* J Bacteriol, 2013. **195**(22): p. 5174-85.
193. Anjuwon-Foster, B.R. and R. Tamayo, *Phase variation of Clostridium difficile virulence factors.* Gut Microbes, 2017: **9**(1): p. 76-83.
194. Anjuwon-Foster, B.R. and R. Tamayo, *A genetic switch controls the production of flagella and toxins in Clostridium difficile.* PLoS Genet, 2017. **13**(3): p. e1006701.
195. Bordeleau, E., et al., *Cyclic di-GMP riboswitch-regulated type IV pili contribute to aggregation of Clostridium difficile.* J Bacteriol, 2015. **197**(5): p. 819-32.
196. Purcell, E.B. and R. Tamayo, *Cyclic diguanylate signaling in Gram-positive bacteria.* FEMS Microbiol Rev, 2016. **40**(5): p. 753-773.
197. Hengge, R., *Principles of c-di-GMP signalling in bacteria.* Nat Rev Microbiol, 2009. **7**(4): p. 263-73.

198. Purcell, E.B., et al., *Cyclic Diguanylate Inversely Regulates Motility and Aggregation in Clostridium difficile.* Journal of Bacteriology, 2012. **194**(13): p. 3307-3316.
199. Sudarsan, N., et al., *Riboswitches in eubacteria sense the second messenger cyclic di-GMP.* Science, 2008. **321**(5887): p. 411-3.
200. Lee, E.R., et al., *An allosteric self-splicing ribozyme triggered by a bacterial second messenger.* Science, 2010. **329**(5993): p. 845-848.
201. Tamayo, R., *Cyclic diguanylate riboswitches control bacterial pathogenesis mechanisms.* PLoS Pathog, 2019. **15**(2): p. e1007529.
202. Soutourina, O.A., et al., *Genome-wide identification of regulatory RNAs in the human pathogen Clostridium difficile.* PLoS Genet, 2013. **9**(5): p. e1003493.
203. McKee, R.W., C.K. Harvest, and R. Tamayo, *Cyclic Diguanylate Regulates Virulence Factor Genes via Multiple Riboswitches in Clostridium difficile.* mSphere, 2018. **3**(5).
204. Melville, S. and L. Craig, *Type IV Pili in Gram-Positive Bacteria.* Microbiology and Molecular Biology Reviews : MMBR, 2013. **77**(3): p. 323-341.
205. Stabler, R.A., et al., *Comparative phylogenomics of Clostridium difficile reveals clade specificity and microevolution of hypervirulent strains.* J Bacteriol, 2006. **188**(20): p. 7297-305.
206. Janvilisri, T., et al., *Microarray Identification of Clostridium difficile Core Components and Divergent Regions Associated with Host Origin.* Journal of Bacteriology, 2009. **191**(12): p. 3881-3891.
207. Maldarelli, G.A., et al., *Identification, Immunogenicity and Crossreactivity of Type IV Pilin and Pilin-like Proteins from Clostridium difficile.* Pathogens and disease, 2014. **71**(3): p. 302-314.
208. Purcell, E.B., et al., *Regulation of Type IV Pili Contributes to Surface Behaviors of Historical and Epidemic Strains of Clostridium difficile.* J Bacteriol, 2015. **198**(3): p. 565-77.
209. Piepenbrink, K.H., et al., *Structure of Clostridium difficile PilJ Exhibits Unprecedented Divergence from Known Type IV Pilins.* The Journal of Biological Chemistry, 2014. **289**(7): p. 4334-4345.
210. Goulding, D., et al., *Distinctive profiles of infection and pathology in hamsters infected with Clostridium difficile strains 630 and B1.* Infect Immun, 2009. **77**(12): p. 5478-85.
211. Piepenbrink, K.H., et al., *Structural and evolutionary analyses show unique stabilization strategies in the type IV pili of Clostridium difficile.* Structure, 2015. **23**(2): p. 385-96.
212. Maldarelli, G.A., et al., *Type IV pili promote early biofilm formation by Clostridium difficile.* Pathog Dis, 2016. **74**(6).
213. McKee, R.W., et al., *Type IV Pili Promote Clostridium difficile Adherence and Persistence in a Mouse Model of Infection.* Infect Immun, 2018. **86**(5): p. e00943-17.
214. Fagan, R.P. and N.F. Fairweather, *Biogenesis and functions of bacterial S-layers.* Nat Rev Microbiol, 2014. **12**(3): p. 211-22.
215. Fagan, R.P., et al., *Structural insights into the molecular organization of the S-layer from Clostridium difficile.* Mol Microbiol, 2009. **71**(5): p. 1308-22.
216. Calabi, E., et al., *Molecular characterization of the surface layer proteins from Clostridium difficile.* Mol Microbiol, 2001. **40**(5): p. 1187-99.
217. Kirby, J.M., et al., *Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of Clostridium difficile.* J Biol Chem, 2009. **284**(50): p. 34666-73.
218. Willing, S.E., et al., *Clostridium difficile surface proteins are anchored to the cell wall using CWB2 motifs that recognise the anionic polymer PSII.* Mol Microbiol, 2015. **96**(3): p. 596-608.
219. Dingle, K.E., et al., *Recombinational switching of the Clostridium difficile S-layer and a novel glycosylation gene cluster revealed by large-scale whole-genome sequencing.* J Infect Dis, 2013. **207**(4): p. 675-86.
220. Kirk, J.A., et al., *New class of precision antimicrobials redefines role of Clostridium difficile S-layer in virulence and viability.* Science translational medicine, 2017. **9**(406): p. 6813.

221. Peltier, J., et al., *Cyclic diGMP regulates production of sortase substrates of Clostridium difficile and their surface exposure through ZmpI protease-mediated cleavage.* J Biol Chem, 2015. **290**(40): p. 24453-69.
222. Donahue, E.H., et al., *Clostridium difficile has a single sortase, SrtB, that can be inhibited by small-molecule inhibitors.* BMC Microbiol, 2014. **14**: p. 219.
223. Kovacs-Simon, A., et al., *Lipoprotein CD0873 is a novel adhesin of Clostridium difficile.* J Infect Dis, 2014. **210**(2): p. 274-84.
224. Hennequin, C., et al., *Identification and characterization of a fibronectin-binding protein from Clostridium difficile.* Microbiology, 2003. **149**(Pt 10): p. 2779-87.
225. Barketi-Klai, A., et al., *Role of fibronectin-binding protein A in Clostridium difficile intestinal colonization.* J Med Microbiol, 2011. **60**(Pt 8): p. 1155-61.
226. Tulli, L., et al., *CbpA: a novel surface exposed adhesin of Clostridium difficile targeting human collagen.* Cellular Microbiology, 2013. **15**(10): p. 1674-1687.
227. Emerson, J.E., et al., *A novel genetic switch controls phase variable expression of CwpV, a Clostridium difficile cell wall protein.* Mol Microbiol, 2009. **74**(3): p. 541-56.
228. Reynolds, C.B., et al., *The Clostridium difficile cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function.* PLoS Pathog, 2011. **7**(4): p. e1002024.
229. Sekulovic, O., et al., *The Clostridium difficile cell wall protein CwpV confers phase-variable phage resistance.* Mol Microbiol, 2015. **98**(2): p. 329-42.
230. Richards, E., et al., *The S-layer protein of a Clostridium difficile SLCT-11 strain displays a complex glycan required for normal cell growth and morphology.* J Biol Chem, 2018. **293**(47): p. 18123-18137.
231. Ganeshapillai, J., et al., *Clostridium difficile cell-surface polysaccharides composed of pentaglycosyl and hexaglycosyl phosphate repeating units.* Carbohydr Res, 2008. **343**(4): p. 703-10.
232. Danieli, E., et al., *First synthesis of C. difficile PS-II cell wall polysaccharide repeating unit.* Org Lett, 2011. **13**(3): p. 378-81.
233. Reid, C.W., et al., *Structural characterization of surface glycans from Clostridium difficile.* Carbohydr Res, 2012. **354**: p. 65-73.
234. Bertolo, L., et al., *Clostridium difficile carbohydrates: glucan in spores, PSII common antigen in cells, immunogenicity of PSII in swine and synthesis of a dual C. difficile-ETEC conjugate vaccine.* Carbohydr Res, 2012. **354**: p. 79-86.
235. Oberli, M.A., et al., *A possible oligosaccharide-conjugate vaccine candidate for Clostridium difficile is antigenic and immunogenic.* Chem Biol, 2011. **18**(5): p. 580-8.
236. Adamo, R., et al., *Phosphorylation of the synthetic hexasaccharide repeating unit is essential for the induction of antibodies to Clostridium difficile PSII cell wall polysaccharide.* ACS Chem Biol, 2012. **7**(8): p. 1420-8.
237. Cox, A.D., et al., *Investigating the candidacy of a lipoteichoic acid-based glycoconjugate as a vaccine to combat Clostridium difficile infection.* Glycoconj J, 2013. **30**(9): p. 843-55.
238. Romano, M.R., et al., *Recombinant Clostridium difficile toxin fragments as carrier protein for PSII surface polysaccharide preserve their neutralizing activity.* Toxins (Basel), 2014. **6**(4): p. 1385-96.
239. Elsden, S.R., M.G. Hilton, and J.M. Waller, *The end products of the metabolism of aromatic amino acids by Clostridia.* Arch Microbiol, 1976. **107**(3): p. 283-8.
240. Passmore, I.J., et al., *Para-cresol production by Clostridium difficile affects microbial diversity and membrane integrity of Gram-negative bacteria.* PLoS Pathog, 2018. **14**(9): p. e1007191.
241. Dawson, L.F., R.A. Stabler, and B.W. Wren, *Assessing the role of p-cresol tolerance in Clostridium difficile.* J Med Microbiol, 2008. **57**(Pt 6): p. 745-9.
242. Hafiz, S. and C.L. Oakley, *Clostridium difficile: isolation and characteristics.* J Med Microbiol, 1976. **9**(2): p. 129-36.

243. Dawson, L.F., et al., *Characterisation of Clostridium difficile biofilm formation, a role for Spo0A.* PLoS One, 2012. **7**(12): p. e50527.
244. Poquet, I., et al., *Clostridium difficile Biofilm: Remodeling Metabolism and Cell Surface to Build a Sparse and Heterogeneously Aggregated Architecture.* Front Microbiol, 2018. **9**: p. 2084.
245. Pantaleon, V., et al., *Clostridium difficile forms variable biofilms on abiotic surface.* Anaerobe, 2018. **53**: p. 34-37.
246. Soavelomandroso, A.P., et al., *Biofilm Structures in a Mono-Associated Mouse Model of Clostridium difficile Infection.* Frontiers in Microbiology, 2017. **8**(2086).
247. Baines, S.D. and M.H. Wilcox, *Antimicrobial Resistance and Reduced Susceptibility in Clostridium difficile: Potential Consequences for Induction, Treatment, and Recurrence of C. difficile Infection.* Antibiotics (Basel, Switzerland), 2015. **4**(3): p. 267-298.
248. Goldstein, E.J., et al., *Comparative susceptibilities to fidaxomicin (OPT-80) of isolates collected at baseline, recurrence, and failure from patients in two phase III trials of fidaxomicin against Clostridium difficile infection.* Antimicrob Agents Chemother, 2011. **55**(11): p. 5194-9.
249. Freeman, J., et al., *The ClosER study: results from a three-year pan-European longitudinal surveillance of antibiotic resistance among prevalent Clostridium difficile ribotypes, 2011-2014.* Clin Microbiol Infect, 2018. **24**(7): p. 724-731.
250. Jin, D., et al., *Molecular Epidemiology of Clostridium difficile Infection in Hospitalized Patients in Eastern China.* J Clin Microbiol, 2017. **55**(3): p. 801-810.
251. McBride, S.M. and A.L. Sonenshein, *The dlt operon confers resistance to cationic antimicrobial peptides in Clostridium difficile.* Microbiology, 2011. **157**(Pt 5): p. 1457-65.
252. Suarez, J.M., A.N. Edwards, and S.M. McBride, *The Clostridium difficile cpr locus is regulated by a noncontiguous two-component system in response to type A and B lantibiotics.* J Bacteriol, 2013. **195**(11): p. 2621-31.
253. Woods, E.C., et al., *The C. difficile clnRAB operon initiates adaptations to the host environment in response to LL-37.* PLoS Pathog, 2018. **14**(8): p. e1007153.
254. Jafari, N.V., et al., *Clostridium difficile-mediated effects on human intestinal epithelia: Modelling host-pathogen interactions in a vertical diffusion chamber.* Anaerobe, 2016. **37**: p. 96-102.
255. McDermott, A.J., et al., *Interleukin-23 (IL-23), independent of IL-17 and IL-22, drives neutrophil recruitment and innate inflammation during Clostridium difficile colitis in mice.* Immunology, 2016. **147**(1): p. 114-24.
256. Jafari, N.V., et al., *Clostridium difficile modulates host innate immunity via toxin-independent and dependent mechanism(s).* PLoS One, 2013. **8**(7): p. e69846.
257. Mahida, Y.R., et al., *Effect of Clostridium difficile toxin A on human intestinal epithelial cells: induction of interleukin 8 production and apoptosis after cell detachment.* Gut, 1996. **38**(3): p. 337.
258. Ng, J., et al., *Clostridium difficile toxin-induced inflammation and intestinal injury are mediated by the inflammasome.* Gastroenterology, 2010. **139**(2): p. 542-52.
259. Jafari, N.V., et al., *Host immunity to Clostridium difficile PCR ribotype 017 strains.* Infect Immun, 2014. **82**(12): p. 4989-96.
260. Ryan, A., et al., *A role for TLR4 in Clostridium difficile infection and the recognition of surface layer proteins.* PLoS Pathog, 2011. **7**(6): p. e1002076.
261. Batah, J., et al., *Clostridium difficile flagella predominantly activate TLR5-linked NF-kappaB pathway in epithelial cells.* Anaerobe, 2016. **38**: p. 116-24.
262. Batah, J., et al., *Clostridium difficile flagella induce a pro-inflammatory response in intestinal epithelium of mice in cooperation with toxins.* Sci Rep, 2017. **7**(1): p. 3256.
263. Kelly, C.P., et al., *Neutrophil recruitment in Clostridium difficile toxin A enteritis in the rabbit.* J Clin Invest, 1994. **93**(3): p. 1257-65.

264. Solomon, K., et al., *Mortality in patients with Clostridium difficile infection correlates with host pro-inflammatory and humoral immune responses.* J Med Microbiol, 2013. **62**(Pt 9): p. 1453-60.
265. Bulusu, M., et al., *Leukocytosis as a harbinger and surrogate marker of Clostridium difficile infection in hospitalized patients with diarrhea.* Am J Gastroenterol, 2000. **95**(11): p. 3137-41.
266. Buonomo, E.L., et al., *Role of interleukin 23 signaling in Clostridium difficile colitis.* J Infect Dis, 2013. **208**(6): p. 917-20.
267. Leuzzi, R., et al., *Protective efficacy induced by recombinant Clostridium difficile toxin fragments.* Infect Immun, 2013. **81**(8): p. 2851-60.
268. Permpoonpattana, P., et al., *Immunization with Bacillus Spores Expressing Toxin A Peptide Repeats Protects against Infection with Clostridium difficile Strains Producing Toxins A and B.* Infection and Immunity, 2011. **79**(6): p. 2295-2302.
269. Leav, B.A., et al., *Serum anti-toxin B antibody correlates with protection from recurrent Clostridium difficile infection (CDI).* Vaccine, 2010. **28**(4): p. 965-9.
270. Islam, J., et al., *The role of the humoral immune response to Clostridium difficile toxins A and B in susceptibility to C. difficile infection: a case-control study.* Anaerobe, 2014. **27**: p. 82-6.
271. Wullt, M., et al., *IgG antibody response to toxins A and B in patients with Clostridium difficile infection.* Clin Vaccine Immunol, 2012. **19**(9): p. 1552-4.
272. Negm, O.H., et al., *Profiling Humoral Immune Responses to Clostridium difficile-Specific Antigens by Protein Microarray Analysis.* Clinical and vaccine immunology : CVI, 2015. **22**(9): p. 1033-1039.
273. Pechine, S., et al., *Immunological properties of surface proteins of Clostridium difficile.* J Med Microbiol, 2005. **54**(Pt 2): p. 193-6.
274. Warny, M., et al., *Human antibody response to Clostridium difficile toxin A in relation to clinical course of infection.* Infection and immunity, 1994. **62**(2): p. 384-389.
275. Johnson, S., D.N. Gerding, and E.N. Janoff, *Systemic and mucosal antibody responses to toxin A in patients infected with Clostridium difficile.* J Infect Dis, 1992. **166**(6): p. 1287-94.
276. Hong, H.A., et al., *Mucosal antibodies to the C terminus of toxin A prevent colonization of Clostridium difficile.* Infection and Immunity, 2017. **85**(4): p. e01060-16.
277. Public Health England., *Clostridium difficile infection: How to deal with the problem.* 2008. (https://www.gov.uk/government/publications/clostridium-difficile-infection-how-to-deal-with-the-problem).
278. Walker, A.S., et al., *Fairness of financial penalties to improve control of Clostridium difficile.* Bmj, 2008. **337**: p. 2097.
279. Baur, D., et al., *Effect of antibiotic stewardship on the incidence of infection and colonisation with antibiotic-resistant bacteria and Clostridium difficile infection: a systematic review and meta-analysis.* Lancet Infect Dis, 2017. **17**(9): p. 990-1001.
280. Valiquette, L., et al., *Impact of a reduction in the use of high-risk antibiotics on the course of an epidemic of Clostridium difficile-associated disease caused by the hypervirulent NAP1/027 strain.* Clin Infect Dis, 2007. **45** (Suppl 2): p. 112-21.
281. Donskey, C.J., *Preventing Transmission of Clostridium difficile: Is the Answer Blowing in the Wind?* Clinical Infectious Diseases, 2010. **50**(11): p. 1458-1461.
282. Kazakova, S.V., et al., *Association between Antibiotic Use and Hospital-Onset Clostridioides difficile Infection in U.S. Acute Care Hospitals, 2006-2012: an Ecologic Analysis.* Clin Infect Dis, 2019. **70**(1): p. 11-18.
283. Pereira, J.B., et al., *Association between Clostridium difficile infection and antimicrobial usage in a large group of English hospitals.* British journal of clinical pharmacology, 2014. **77**(5): p. 896-903.
284. Jakobsson, H.E., et al., *Short-Term Antibiotic Treatment Has Differing Long-Term Impacts on the Human Throat and Gut Microbiome.* PLOS ONE, 2010. **5**(3): p. e9836.

285. Chang, J.Y., et al., *Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea.* J Infect Dis, 2008. **197**(3): p. 435-8.
286. Tannock, G.W., et al., *A new macrocyclic antibiotic, fidaxomicin (OPT-80), causes less alteration to the bowel microbiota of Clostridium difficile-infected patients than does vancomycin.* Microbiology, 2010. **156**(Pt 11): p. 3354-9.
287. Cornely, O.A., et al., *Treatment of first recurrence of Clostridium difficile infection: fidaxomicin versus vancomycin.* Clin Infect Dis, 2012. **55** (Suppl 2): p. S154-61.
288. Louie, T.J., et al., *Fidaxomicin versus vancomycin for Clostridium difficile infection.* N Engl J Med, 2011. **364**(5): p. 422-31.
289. Drekonja, D., et al., *Fecal Microbiota Transplantation for Clostridium difficile Infection: A Systematic Review.* Ann Intern Med, 2015. **162**(9): p. 630-8.
290. Nathwani, D., et al., *Cost-effectiveness analysis of fidaxomicin versus vancomycin in Clostridium difficile infection.* The Journal of antimicrobial chemotherapy, 2014. **69**(11): p. 2901-2912.
291. Allegretti, J.R., et al., *The evolution of the use of faecal microbiota transplantation and emerging therapeutic indications.* Lancet, 2019. **394**(10196): p. 420-431.
292. Gough, E., H. Shaikh, and A.R. Manges, *Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent Clostridium difficile infection.* Clin Infect Dis, 2011. **53**(10): p. 994-1002.
293. van Nood, E., et al., *Duodenal Infusion of Donor Feces for Recurrent Clostridium difficile.* New England Journal of Medicine, 2013. **368**(5): p. 407-415.
294. Kelly, C.R., et al., *Effect of Fecal Microbiota Transplantation on Recurrence in Multiply Recurrent Clostridium difficile Infection: A Randomized Trial.* Ann Intern Med, 2016. **165**(9): p. 609-616.
295. Hvas, C.L., et al., *Fecal Microbiota Transplantation Is Superior to Fidaxomicin for Treatment of Recurrent Clostridium difficile Infection.* Gastroenterology, 2019. **156**(5): p. 1324-1332.
296. Cammarota, G., et al., *Randomised clinical trial: faecal microbiota transplantation by colonoscopy vs. vancomycin for the treatment of recurrent Clostridium difficile infection.* Aliment Pharmacol Ther, 2015. **41**(9): p. 835-43.
297. Mullish, B.H., et al., *The use of faecal microbiota transplant as treatment for recurrent or refractory Clostridium difficile infection and other potential indications: joint British Society of Gastroenterology (BSG) and Healthcare Infection Society (HIS) guidelines.* Gut, 2018. **67**(11): p. 1920-1941.
298. Schwartz, M., M. Gluck, and S. Koon, *Norovirus gastroenteritis after fecal microbiota transplantation for treatment of Clostridium difficile infection despite asymptomatic donors and lack of sick contacts.* Am J Gastroenterol, 2013. **108**(8): p. 1367.
299. Kociolek, L.K. and D.N. Gerding, *Breakthroughs in the treatment and prevention of Clostridium difficile infection.* Nat Rev Gastroenterol Hepatol, 2016. **13**(3): p. 150-60.
300. Quera, R., et al., *Bacteremia as an adverse event of fecal microbiota transplantation in a patient with Crohn's disease and recurrent Clostridium difficile infection.* J Crohns Colitis, 2014. **8**(3): p. 252-3.
301. Lee, C.H., et al., *Frozen vs Fresh Fecal Microbiota Transplantation and Clinical Resolution of Diarrhea in Patients With Recurrent Clostridium difficile Infection: A Randomized Clinical Trial.* JAMA, 2016. **315**(2): p. 142-9.
302. Hirsch, B.E., et al., *Effectiveness of fecal-derived microbiota transfer using orally administered capsules for recurrent Clostridium difficile infection.* BMC Infect Dis, 2015. **15**: p. 191.
303. Youngster, I., et al., *Oral, capsulized, frozen fecal microbiota transplantation for relapsing Clostridium difficile infection.* JAMA, 2014. **312**(17): p. 1772-8.

304. Lawley, T.D., et al., *Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice.* PLoS Pathog, 2012. **8**(10): p. e1002995.
305. Bernstein, H., et al., *Bile acids as carcinogens in human gastrointestinal cancers.* Mutat Res, 2005. **589**(1): p. 47-65.
306. Gerding, D.N., et al., *Administration of spores of nontoxigenic Clostridium difficile strain M3 for prevention of recurrent C. difficile infection: a randomized clinical trial.* JAMA, 2015. **313**(17): p. 1719-27.
307. Gerding, D.N., S.P. Sambol, and S. Johnson, *Non-toxigenic Clostridioides (Formerly Clostridium) difficile for Prevention of C. difficile Infection: From Bench to Bedside Back to Bench and Back to Bedside.* Frontiers in microbiology, 2018. **9**: p. 1700.
308. Villano, S.A., et al., *Evaluation of an oral suspension of VP20621, spores of nontoxigenic Clostridium difficile strain M3, in healthy subjects.* Antimicrob Agents Chemother, 2012. **56**(10): p. 5224-9.
309. Brouwer, M.S., et al., *Horizontal gene transfer converts non-toxigenic Clostridium difficile strains into toxin producers.* Nat Commun, 2013. **4**: p. 2601.
310. Wilcox, M., M.B. Dorr, and A. Pedley, *Bezlotoxumab and Recurrent Clostridium difficile Infection.* N Engl J Med, 2017. **376**(16): p. 1594-6.
311. Lowy, I., et al., *Treatment with monoclonal antibodies against Clostridium difficile toxins.* N Engl J Med, 2010. **362**(3): p. 197-205.
312. Wilcox, M.H., et al., *Bezlotoxumab for Prevention of Recurrent Clostridium difficile Infection.* N Engl J Med, 2017. **376**(4): p. 305-317.
313. National Institue for Healthcare and Excellence, *Preventing recurrence of Clostridium difficile infection: bezlotoxumab.* 2017. (https://www.nice.org.uk/advice/es13/chapter/Key-points).
314. Oksi, J., et al., *Real-world efficacy of bezlotoxumab for prevention of recurrent Clostridium difficile infection: a retrospective study of 46 patients in five university hospitals in Finland.* Eur J Clin Microbiol Infect Dis, 2019. **38**(10): p. 1947-1952.
315. Goldenberg, J.Z., et al., *Probiotics for the prevention of Clostridium difficile-associated diarrhea in adults and children.* Cochrane Database Syst Rev, 2017. **12**: p. 006095.
316. Nale, J.Y., et al., *Bacteriophage Combinations Significantly Reduce Clostridium difficile Growth In Vitro and Proliferation In Vivo.* Antimicrobial agents and chemotherapy, 2015. **60**(2): p. 968-981.
317. Hall, C.L. and S. Bunn, *Antimicrobial Resistance and Immunisation*. 2018, Parliamentary Office of Science and Technology: London, UK. (https://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-0581)
318. World Organisation for Animal Health., *No more deaths from rinderpest. OIE's recognition pathway paved way for global declaration of eradication by FAO member countries in June.* 2011. (https://www.oie.int/for-the-media/press-releases/detail/article/no-more-deaths-from-rinderpest/).
319. Blanchard-Rohner, G. and A.J. Pollard, *Long-term protection after immunization with protein-polysaccharide conjugate vaccines in infancy.* Expert Rev Vaccines, 2011. **10**(5): p. 673-84.
320. Stein, K.E., *Thymus-independent and thymus-dependent responses to polysaccharide antigens.* J Infect Dis, 1992. **165** (Suppl 1): p. S49-52.
321. Astronomo, R.D. and D.R. Burton, *Carbohydrate vaccines: developing sweet solutions to sticky situations?* Nature reviews. Drug discovery, 2010. **9**(4): p. 308-324.
322. Weksler, M.E., *Changes in the B-cell repertoire with age.* Vaccine, 2000. **18**(16): p. 1624-8.
323. Avery, O.T. and W.F. Goebel, *Chemo-immunological syudies on conjugated carbohydrate proteins: V. The immunological specificity of an antigen prepared by combining Type III penumococcus with foreign protein.* J Exp Med, 1931. **54**(3): p. 437-47.

324. Avci, F.Y., et al., *A mechanism for glycoconjugate vaccine activation of the adaptive immune system and its implications for vaccine design.* Nature Medicine, 2011. **17**: p. 1602.
325. De Gregorio, E. and R. Rappuoli, *From empiricism to rational design: a personal perspective of the evolution of vaccine development.* Nat Rev Immunol, 2014. **14**(7): p. 505-14.
326. Sun, X., et al., *Polysaccharide structure dictates mechanism of adaptive immune response to glycoconjugate vaccines.* Proceedings of the National Academy of Sciences, 2019. **116**(1): p. 193.
327. Micoli, F., R. Adamo, and P. Costantino, *Protein Carriers for Glycoconjugate Vaccines: History, Selection Criteria, Characterization and New Trends.* Molecules, 2018. **23**(6).
328. Michon, F., et al., *Multivalent pneumococcal capsular polysaccharide conjugate vaccines employing genetically detoxified pneumolysin as a carrier protein.* Vaccine, 1998. **16**(18): p. 1732-41.
329. Pozzi, C., et al., *Opsonic and protective properties of antibodies raised to conjugate vaccines targeting six Staphylococcus aureus antigens.* PLoS One, 2012. **7**(10): p. e46648.
330. Laird, R.M., et al., *Evaluation of a conjugate vaccine platform against enterotoxigenic Escherichia coli (ETEC), Campylobacter jejuni and Shigella.* Vaccine, 2018. **36**(45): p. 6695-6702.
331. Pollard, A.J., K.P. Perrett, and P.C. Beverley, *Maintaining protection against invasive bacteria with protein-polysaccharide conjugate vaccines.* Nat Rev Immunol, 2009. **9**(3): p. 213-20.
332. Kay, E., J. Cuccui, and B.W. Wren, *Recent advances in the production of recombinant glycoconjugate vaccines.* NPJ Vaccines, 2019. **4**: p. 16.
333. Terra, V.S., et al., *Recent developments in bacterial protein glycan coupling technology and glycoconjugate vaccine design.* J Med Microbiol, 2012. **61**(Pt 7): p. 919-26.
334. Messner, P., *Prokaryotic glycoproteins: unexplored but important.* Journal of bacteriology, 2004. **186**(9): p. 2517-2519.
335. Nothaft, H. and C.M. Szymanski, *Protein glycosylation in bacteria: sweeter than ever.* Nat Rev Microbiol, 2010. **8**(11): p. 765-78.
336. Szymanski, C.M., D.H. Burr, and P. Guerry, *Campylobacter protein glycosylation affects host cell interactions.* Infection and immunity, 2002. **70**(4): p. 2242-2244.
337. Abouelhadid, S., et al., *Quantitative Analyses Reveal Novel Roles for N-Glycosylation in a Major Enteric Bacterial Pathogen.* MBio, 2019. **10**(2): p. e00297-19.
338. Cuccui, J., et al., *The N-linking glycosylation system from Actinobacillus pleuropneumoniae is required for adhesion and has potential use in glycoengineering.* Open Biol, 2017. **7**(1): p. 160212.
339. Hanuszkiewicz, A., et al., *Identification of the flagellin glycosylation system in Burkholderia cenocepacia and the contribution of glycosylated flagellin to evasion of human innate immune responses.* J Biol Chem, 2014. **289**(27): p. 19231-44.
340. Faridmoayer, A., et al., *Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation.* Journal of bacteriology, 2007. **189**(22): p. 8088-8098.
341. Wacker, M., et al., *N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli.* Science, 2002. **298**(5599): p. 1790-3.
342. Linton, D., et al., *Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway.* Mol Microbiol, 2005. **55**(6): p. 1695-703.
343. Grass, S., et al., *The Haemophilus influenzae HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis.* Mol Microbiol, 2003. **48**(3): p. 737-51.
344. Gross, J., et al., *The Haemophilus influenzae HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification.* J Biol Chem, 2008. **283**(38): p. 26010-5.
345. Breton, C., et al., *Structures and mechanisms of glycosyltransferases.* Glycobiology, 2005. **16**(2): p. 29-37.

346. Schmid, J., et al., *Bacterial Glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products.* Frontiers in microbiology, 2016. **7**: p. 182-182.
347. Kowarik, M., et al., *N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase.* Science, 2006. **314**(5802): p. 1148-50.
348. Kowarik, M., et al., *Definition of the bacterial N-glycosylation site consensus sequence.* The EMBO Journal, 2006. **25**(9): p. 1957-1966.
349. Napiorkowska, M., et al., *Structure of bacterial oligosaccharyltransferase PglB bound to a reactive LLO and an inhibitory peptide.* Sci Rep, 2018. **8**(1): p. 16297.
350. Ielmini, M.V. and M.F. Feldman, *Desulfovibrio desulfuricans PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements.* Glycobiology, 2011. **21**(6): p. 734-42.
351. Ollis, A.A., et al., *Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences.* Sci Rep, 2015. **5**: p. 15237.
352. Goon, S., et al., *Pseudaminic acid, the major modification on Campylobacter flagellin, is synthesized via the Cj1293 gene.* Molecular Microbiology, 2003. **50**(2): p. 659-671.
353. Ewing, C.P., E. Andreishcheva, and P. Guerry, *Functional Characterization of Flagellin Glycosylation in Campylobacter jejuni 81-176.* Journal of Bacteriology, 2009. **191**(22): p. 7086.
354. Schirm, M., et al., *Identification of unusual bacterial glycosylation by tandem mass spectrometry analyses of intact proteins.* Anal Chem, 2005. **77**(23): p. 7774-82.
355. Schoenhofen, I.C., et al., *The CMP-legionaminic acid pathway in Campylobacter: biosynthesis involving novel GDP-linked precursors.* Glycobiology, 2009. **19**(7): p. 715-25.
356. Thibault, P., et al., *Identification of the carbohydrate moieties and glycosylation motifs in Campylobacter jejuni flagellin.* J Biol Chem, 2001. **276**(37): p. 34862-70.
357. Szymanski, C.M., et al., *Evidence for a system of general protein glycosylation in Campylobacter jejuni.* Mol Microbiol, 1999. **32**(5): p. 1022-30.
358. Young, N.M., et al., *Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni.* J Biol Chem, 2002. **277**(45): p. 42530-9.
359. Alaimo, C., et al., *Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides.* Embo j, 2006. **25**(5): p. 967-76.
360. Kelly, J., et al., *Biosynthesis of the N-linked glycan in Campylobacter jejuni and addition onto protein through block transfer.* J Bacteriol, 2006. **188**(7): p. 2427-34.
361. Perez, C., et al., *Structure and mechanism of an active lipid-linked oligosaccharide flippase.* Nature, 2015. **524**(7566): p. 433-8.
362. Mills, D.C., et al., *Functional analysis of N-linking oligosaccharyl transferase enzymes encoded by deep-sea vent proteobacteria.* Glycobiology, 2015. **26**(4): p.398-409.
363. Jervis, A.J., et al., *Characterization of N-Linked Protein Glycosylation in Helicobacter pullorum.* Journal of Bacteriology, 2010. **192**(19): p. 5228.
364. Wacker, M., et al., *Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems.* Proc Natl Acad Sci U S A, 2006. **103**(18): p. 7088-93.
365. Cuccui, J., et al., *Exploitation of bacterial N-linked glycosylation to develop a novel recombinant glycoconjugate vaccine against Francisella tularensis.* Open Biol, 2013. **3**(5): p. 130002.
366. Fisher, A.C., et al., *Production of secretory and extracellular N-linked glycoproteins in Escherichia coli.* Appl Environ Microbiol, 2011. **77**(3): p. 871-81.
367. Reglinski, M., et al., *A recombinant conjugated pneumococcal vaccine that protects against murine infections with a similar efficacy to Prevnar-13.* npj Vaccines, 2018. **3**(1): p. 53.

368. Garcia-Quintanilla, F., et al., *Production of a recombinant vaccine candidate against Burkholderia pseudomallei exploiting the bacterial N-glycosylation machinery.* Front Microbiol, 2014. **5**: p. 381.
369. Wacker, M., et al., *Prevention of Staphylococcus aureus infections by glycoprotein vaccines synthesized in Escherichia coli.* J Infect Dis, 2014. **209**(10): p. 1551-61.
370. Huttner, A., et al., *Safety, immunogenicity, and preliminary clinical efficacy of a vaccine against extraintestinal pathogenic Escherichia coli in women with a history of recurrent urinary tract infection: a randomised, single-blind, placebo-controlled phase 1b trial.* Lancet Infect Dis, 2017. **17**(5): p. 528-537.
371. Riddle, M.S., et al., *Safety and Immunogenicity of a Candidate Bioconjugate Vaccine against Shigella flexneri 2a Administered to Healthy Adults: a Single-Blind, Randomized Phase I Study.* Clin Vaccine Immunol, 2016. **23**(12): p. 908-917.
372. Hatz, C.F., et al., *Safety and immunogenicity of a candidate bioconjugate vaccine against Shigella dysenteriae type 1 administered to healthy adults: A single blind, partially randomized Phase I study.* Vaccine, 2015. **33**(36): p. 4594-601.
373. Berical, A.C., et al., *Pneumococcal Vaccination Strategies. An Update and Perspective.* Annals of the American Thoracic Society, 2016. **13**(6): p. 933-944.
374. Baguelin, M., et al., *Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study.* PLoS medicine, 2013. **10**(10): p. e1001527-e1001527.
375. Riley, T.V., D. Lyras, and G.R. Douce, *Status of vaccine research and development for Clostridium difficile.* Vaccine, 2019. **37**(50): p. 7300-7306.
376. Baggs, J., et al., *Identification of population at risk for future Clostridium difficile infection following hospital discharge to be targeted for vaccine trials.* Vaccine, 2015. **33**(46): p. 6241-9.
377. van Kleef, E., et al., *The projected effectiveness of Clostridium difficile vaccination as part of an integrated infection control strategy.* Vaccine, 2016. **34**(46): p. 5562-5570.
378. Lee, B.Y., et al., *The potential value of Clostridium difficile vaccine: an economic computer simulation model.* Vaccine, 2010. **28**(32): p. 5245-53.
379. MacBeath, G. and S.L. Schreiber, *Printing proteins as microarrays for high-throughput function determination.* Science, 2000. **289**(5485): p. 1760-3.
380. Davies, D.H., et al., *Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery.* Proc Natl Acad Sci U S A, 2005. **102**(3): p. 547-52.
381. Syafrizayanti, et al., *Personalised proteome analysis by means of protein microarrays made from individual patient samples.* Sci Rep, 2017. **7**: p. 39756.
382. He, M. and M.J. Taussig, *Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method).* Nucleic Acids Res, 2001. **29**(15): p. e73-3.
383. Carlson, E.D., et al., *Cell-free protein synthesis: applications come of age.* Biotechnol Adv, 2012. **30**(5): p. 1185-94.
384. Dent, A.E., et al., *Plasmodium falciparum Protein Microarray Antibody Profiles Correlate With Protection From Symptomatic Malaria in Kenya.* J Infect Dis, 2015. **212**(9): p. 1429-38.
385. Felgner, P.L., et al., *A Burkholderia pseudomallei protein microarray reveals serodiagnostic and cross-reactive antigens.* Proc Natl Acad Sci U S A, 2009. **106**(32): p. 13499-504.
386. Liang, L., et al., *Immune profiling with a Salmonella Typhi antigen microarray identifies new diagnostic biomarkers of human typhoid.* Sci Rep, 2013. **3**: p. 1043.
387. Liang, L., et al., *Systems biology approach predicts antibody signature associated with Brucella melitensis infection in humans.* J Proteome Res, 2011. **10**(10): p. 4813-24.
388. Steller, S., et al., *Bacterial protein microarrays for identification of new potential diagnostic markers for Neisseria meningitidis infections.* Proteomics, 2005. **5**(8): p. 2048-55.

389. Luevano, M., et al., *High-throughput profiling of the humoral immune responses against thirteen human papillomavirus types by proteome microarrays.* Virology, 2010. **405**(1): p. 31-40.
390. Spencer, J., et al., *Vaccination against Clostridium difficile using toxin fragments.* Gut Microbes, 2014. **5**(2): p. 225-232.
391. Buckley, A.M., et al., *Infection of hamsters with the UK Clostridium difficile ribotype 027 outbreak strain R20291.* Journal of medical microbiology, 2011. **60**(Pt 8): p. 1174-1180.
392. Browne, R.A., et al., *The protective effect of vancomycin on clindamycin-induced colitis in hamsters.* Johns Hopkins Med J, 1977. **141**(4): p. 183-92.
393. Chen, X., et al., *A mouse model of Clostridium difficile-associated disease.* Gastroenterology, 2008. **135**(6): p. 1984-92.
394. Seekatz, A.M., et al., *Fecal Microbiota Transplantation Eliminates Clostridium difficile in a Murine Model of Relapsing Disease.* Infect Immun, 2015. **83**(10): p. 3838-46.
395. Pechine, S., et al., *Immunization of hamsters against Clostridium difficile infection using the Cwp84 protease as an antigen.* FEMS Immunol Med Microbiol, 2011. **63**(1): p. 73-81.
396. Pechine, S., et al., *Immunization using GroEL decreases Clostridium difficile intestinal colonization.* PLoS One, 2013. **8**(11): p. e81112.
397. Kim, P.H., J.P. Iaconis, and R.D. Rolfe, *Immunization of adult hamsters against Clostridium difficile-associated ileocecitis and transfer of protection to infant hamsters.* Infect Immun, 1987. **55**(12): p. 2984-92.
398. Torres, J.F., et al., *Evaluation of formalin-inactivated Clostridium difficile vaccines administered by parenteral and mucosal routes of immunization in hamsters.* Infect Immun, 1995. **63**(12): p. 4619-27.
399. Ward, S.J., et al., *Local and systemic neutralizing antibody responses induced by intranasal immunization with the nontoxic binding domain of toxin A from Clostridium difficile.* Infect Immun, 1999. **67**(10): p. 5124-32.
400. Aboudola, S., et al., *Clostridium difficile vaccine and serum immunoglobulin G antibody response to toxin A.* Infect Immun, 2003. **71**(3): p. 1608-10.
401. Kotloff, K.L., et al., *Safety and immunogenicity of increasing doses of a Clostridium difficile toxoid vaccine administered to healthy adults.* Infect Immun, 2001. **69**(2): p. 988-95.
402. Sougioultzis, S., et al., *Clostridium difficile toxoid vaccine in recurrent C. difficile-associated diarrhea.* Gastroenterology, 2005. **128**(3): p. 764-70.
403. Siddiqui, F., et al., *Vaccination with parenteral toxoid B protects hamsters against lethal challenge with toxin A-negative, toxin B-positive clostridium difficile but does not prevent colonization.* J Infect Dis, 2012. **205**(1): p. 128-33.
404. Wang, H., et al., *A chimeric toxin vaccine protects against primary and recurrent Clostridium difficile infection.* Infect Immun, 2012. **80**(8): p. 2678-88.
405. Zhang, B.Z., et al., *A DNA vaccine targeting TcdA and TcdB induces protective immunity against Clostridium difficile.* BMC Infect Dis, 2016. **16**(1): p. 596.
406. Winter, K., et al., *Vaccination against Clostridium difficile by Use of an Attenuated Salmonella enterica Serovar Typhimurium Vector (YS1646) Protects Mice from Lethal Challenge.* Infect Immun, 2019. **87**(8): p. e00089-19.
407. de Bruyn, G., et al., *Defining the optimal formulation and schedule of a candidate toxoid vaccine against Clostridium difficile infection: A randomized Phase 2 clinical trial.* Vaccine, 2016. **34**(19): p. 2170-8.
408. Sheldon, E., et al., *A phase 1, placebo-controlled, randomized study of the safety, tolerability, and immunogenicity of a Clostridium difficile vaccine administered with or without aluminum hydroxide in healthy adults.* Vaccine, 2016. **34**(18): p. 2082-91.
409. Kitchin, N., et al., *A Phase 2 Study Evaluating the Safety, Tolerability, and Immunogenicity of Two 3-Dose Regimens of a Clostridium difficile Vaccine in Healthy US Adults Aged 65 to 85 Years.* Clin Infect Dis, 2019. **70**(10): p. 1-10.

410. Bezay, N., et al., *Safety, immunogenicity and dose response of VLA84, a new vaccine candidate against Clostridium difficile, in healthy volunteers.* Vaccine, 2016. **34**(23): p. 2585-92.
411. Gardiner, D.F., et al., *A DNA vaccine targeting the receptor-binding domain of Clostridium difficile toxin A.* Vaccine, 2009. **27**(27): p. 3598-604.
412. Jin, K., et al., *Protective antibody responses against Clostridium difficile elicited by a DNA vaccine expressing the enzymatic domain of toxin B.* Hum Vaccin Immunother, 2013. **9**(1): p. 63-73.
413. Ryan, E.T., et al., *Protective immunity against Clostridium difficile toxin A induced by oral immunization with a live, attenuated Vibrio cholerae vector strain.* Infect Immun, 1997. **65**(7): p. 2941-9.
414. Ward, S.J., et al., *Immunogenicity of a Salmonella typhimurium aroA aroD vaccine expressing a nontoxic domain of Clostridium difficile toxin A.* Infect Immun, 1999. **67**(5): p. 2145-52.
415. Yang, X.Q., et al., *The protective effect of recombinant Lactococcus lactis oral vaccine on a Clostridium difficile-infected animal model.* BMC Gastroenterol, 2013. **13**: p. 117.
416. Secore, S., et al., *Development of a Novel Vaccine Containing Binary Toxin for the Prevention of Clostridium difficile Disease with Enhanced Efficacy against NAP1 Strains.* PLoS One, 2017. **12**(1): p. e0170640.
417. Bruxelle, J.F., et al., *Clostridium difficile flagellin FliC: Evaluation as adjuvant and use in a mucosal vaccine against Clostridium difficile.* PLoS One, 2017. **12**(11): p. e0187212.
418. Pichichero, M.E., *Protein carriers of conjugate vaccines: characteristics, development, and clinical trials.* Human vaccines & immunotherapeutics, 2013. **9**(12): p. 2505-2523.
419. Martin, C.E., et al., *Immunological evaluation of a synthetic Clostridium difficile oligosaccharide conjugate vaccine candidate and identification of a minimal epitope.* J Am Chem Soc, 2013. **135**(26): p. 9713-22.
420. Broecker, F., et al., *Synthetic Lipoteichoic Acid Glycans Are Potential Vaccine Candidates to Protect from Clostridium difficile Infections.* Cell Chem Biol, 2016. **23**(8): p. 1014-1022.
421. Ghose, C., et al., *Immunogenicity and protective efficacy of Clostridium difficile spore proteins.* Anaerobe, 2016. **37**: p. 85-95.
422. Bruxelle, J.F., et al., *Protection against Clostridium difficile infection in a hamster model by oral vaccination using flagellin FliC-loaded pectin beads.* Vaccine, 2018. **36**(40): p. 6017-6021.
423. Senoh, M., et al., *Development of vaccine for Clostridium difficile infection using membrane fraction of nontoxigenic Clostridium difficile.* Microb Pathog, 2018. **123**: p. 42-46.
424. Potocki, W., et al., *The combination of recombinant and non-recombinant Bacillus subtilis spore display technology for presentation of antigen and adjuvant on single spore.* Microb Cell Fact, 2017. **16**(1): p. 151.
425. Ghose, C., et al., *Immunogenicity and protective efficacy of recombinant Clostridium difficile flagellar protein FliC.* Emerg Microbes Infect, 2016. **5**: p. e8.
426. Pechine, S., et al., *Diminished intestinal colonization by Clostridium difficile and immune response in mice after mucosal immunization with surface proteins of Clostridium difficile.* Vaccine, 2007. **25**(20): p. 3946-54.
427. Ni Eidhin, D.B., et al., *Active immunization of hamsters against Clostridium difficile infection using surface-layer protein.* FEMS Immunol Med Microbiol, 2008. **52**(2): p. 207-18.
428. Hussain, H.A., A.P. Roberts, and P. Mullany, *Generation of an erythromycin-sensitive derivative of Clostridium difficile strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites.* J Med Microbiol, 2005. **54**(Pt 2): p. 137-41.
429. Purdy, D., et al., *Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier.* Mol Microbiol, 2002. **46**(2): p. 439-52.

430. Feldman, M.F., et al., *Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli.* Proc Natl Acad Sci U S A, 2005. **102**(8): p. 3016-21.
431. Cartman, S.T. and N.P. Minton, *A mariner-Based Transposon System for In Vivo Random Mutagenesis of Clostridium difficile.* Applied and Environmental Microbiology, 2010. **76**(4): p. 1103-1109.
432. Carver, T., et al., *Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data.* Bioinformatics (Oxford, England), 2012. **28**(4): p. 464-469.
433. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server.* Nat Protoc, 2009. **4**(3): p. 363-71.
434. Bateman, A., et al., *The Pfam protein families database.* Nucleic Acids Res, 2004. **32**(Database issue): p. 138-41.
435. Petersen, T.N., et al., *SignalP 4.0: discriminating signal peptides from transmembrane regions.* Nat Methods, 2011. **8**(10): p. 785-6.
436. Moller, S., M.D. Croning, and R. Apweiler, *Evaluation of methods for the prediction of membrane spanning regions.* Bioinformatics, 2001. **17**(7): p. 646-53.
437. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.
438. Kirk, J.A. and R.P. Fagan, *Heat shock increases conjugation efficiency in Clostridium difficile.* Anaerobe, 2016. **42**: p. 1-5.
439. Faulds-Pain, A. and B.W. Wren, *Improved bacterial mutagenesis by high-frequency allele exchange, demonstrated in Clostridium difficile and Streptococcus suis.* Appl Environ Microbiol, 2013. **79**(15): p. 4768-71.
440. McAllister, K.N., et al., *Using CRISPR-Cas9-mediated genome editing to generate C. difficile mutants defective in selenoproteins synthesis.* Sci Rep, 2017. **7**(1): p. 14672.
441. Kuehne, S.A. and N.P. Minton, *ClosTron-mediated engineering of Clostridium.* Bioengineered, 2012. **3**(4): p. 247-54.
442. Horton, R.M., et al., *Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension.* Gene, 1989. **77**(1): p. 61-8.
443. Andrews, J.M., *Determination of minimum inhibitory concentrations.* J Antimicrob Chemother, 2001. **48** (Suppl 1): p. 5-16.
444. Solomon, K., *The host immune response to Clostridium difficile infection.* Therapeutic Advances in Infectious Disease, 2013. **1**(1): p. 19-35.
445. Mulligan, M.E., et al., *Elevated levels of serum immunoglobulins in asymptomatic carriers of Clostridium difficile.* Clin Infect Dis, 1993. **16** (Suppl 4): p. S239-44.
446. Sanchez-Hurtado, K., et al., *Systemic antibody response to Clostridium difficile in colonized patients with and without symptoms and matched controls.* J Med Microbiol, 2008. **57**(Pt 6): p. 717-24.
447. Bruxelle, J.F., et al., *Immunogenic properties of the surface layer precursor of Clostridium difficile and vaccination assays in animal models.* Anaerobe, 2016. **37**: p. 78-84.
448. Calabi, E., et al., *Binding of Clostridium difficile Surface Layer Proteins to Gastrointestinal Tissues.* Infection and Immunity, 2002. **70**(10): p. 5770-5778.
449. Péchiné, S., et al., *Targeting Clostridium difficile Surface Components to Develop Immunotherapeutic Strategies Against Clostridium difficile Infection.* Frontiers in microbiology, 2018. **9**: p. 1009-1009.
450. Laughon, B.E., et al., *Enzyme immunoassays for detection of Clostridium difficile toxins A and B in fecal specimens.* J Infect Dis, 1984. **149**(5): p. 781-8.
451. Lessa-Aquino, C., et al., *Distinct antibody responses of patients with mild and severe leptospirosis determined by whole proteome microarray analysis.* PLoS Negl Trop Dis, 2017. **11**(1): p. e0005349.

452. Nakajima, R., et al., *Towards Development of Improved Serodiagnostics for Tularemia by Use of Francisella tularensis Proteome Microarrays.* J Clin Microbiol, 2016. **54**(7): p. 1755-1765.
453. Vigil, A., et al., *Profiling the humoral immune response of acute and chronic Q fever by protein microarray.* Mol Cell Proteomics, 2011. **10**(10): p. M110.006304.
454. Public Health England., *Clostridioides difficile infection: guidance on management and treatment.* 2019. (https://www.gov.uk/government/publications/clostridium-difficile-infection-guidance-on-management-and-treatment).
455. Gaze, S., et al., *An immunomics approach to schistosome antigen discovery: antibody signatures of naturally resistant and chronically infected individuals from endemic areas.* PLoS Pathog, 2014. **10**(3): p. e1004033.
456. Taghavian, O., et al., *Antibody Profiling by Proteome Microarray with Multiplex Isotype Detection Reveals Overlap between Human and Aotus nancymaae Controlled Malaria Infections.* Proteomics, 2018. **18**(13): p. e1870115.
457. Liang, L., et al., *Identification of potential serodiagnostic and subunit vaccine antigens by antibody profiling of toxoplasmosis cases in Turkey.* Mol Cell Proteomics, 2011. **10**(7): p. M110.006916.
458. Jain, A., et al., *Evaluation of quantum dot immunofluorescence and a digital CMOS imaging system as an alternative to conventional organic fluorescence dyes and laser scanning for quantifying protein microarrays.* Proteomics, 2016. **16**(8): p. 1271-1279.
459. Dümmler, A., A.-M. Lawrence, and A. de Marco, *Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors.* Microbial Cell Factories, 2005. **4**: p. 34-34.
460. Ihssen, J., et al., *Structural insights from random mutagenesis of Campylobacter jejuni oligosaccharyltransferase PglB.* BMC biotechnology, 2012. **12**: p. 67-67.
461. Monaghan, T.M., et al., *High prevalence of subclass-specific binding and neutralizing antibodies against Clostridium difficile toxins in adult cystic fibrosis sera: possible mode of immunoprotection against symptomatic C. difficile infection.* Clinical and experimental gastroenterology, 2017. **10**: p. 169-175.
462. Loh, J.M.S., et al., *Mucosal vaccination with pili from Group A Streptococcus expressed on Lactococcus lactis generates protective immune responses.* Scientific Reports, 2017. **7**(1): p. 7174.
463. Horzempa, J., et al., *Immunization with a Pseudomonas aeruginosa Pilin Provides O-Antigen-Specific Protection.* Clinical and Vaccine Immunology, 2008. **15**(4): p. 590.
464. Boslego, J.W., et al., *Efficacy trial of a parenteral gonococcal pilus vaccine in men.* Vaccine, 1991. **9**(3): p. 154-62.
465. Lepper, A.W., et al., *A Moraxella bovis pili vaccine produced by recombinant DNA technology for the prevention of infectious bovine keratoconjunctivitis.* Vet Microbiol, 1993. **36**(1-2): p. 175-83.
466. Humphries, R.M., D.Z. Uslan, and Z. Rubin, *Performance of Clostridium difficile toxin enzyme immunoassay and nucleic acid amplification tests stratified by patient disease severity.* J Clin Microbiol, 2013. **51**(3): p. 869-73.
467. Shim, J.K., et al., *Primary symptomless colonisation by Clostridium difficile and decreased risk of subsequent diarrhoea.* Lancet, 1998. **351**(9103): p. 633-6.
468. Reeves, A.E., et al., *The interplay between microbiome dynamics and pathogen dynamics in a murine model of Clostridium difficile Infection.* Gut Microbes, 2011. **2**(3): p. 145-58.
469. Zhang, L., et al., *Insight into alteration of gut microbiota in Clostridium difficile infection and asymptomatic C. difficile colonization.* Anaerobe, 2015. **34**: p. 1-7.
470. Viscidi, R., et al., *Serum antibody response to toxins A and B of Clostridium difficile.* J Infect Dis, 1983. **148**(1): p. 93-100.
471. Rousseau, C., et al., *Clostridium difficile carriage in healthy infants in the community: a potential reservoir for pathogenic strains.* Clin Infect Dis, 2012. **55**(9): p. 1209-15.

472. Pechine, S. and A. Collignon, *Immune responses induced by Clostridium difficile.* Anaerobe, 2016.
473. Kelly, C.P. and L. Kyne, *The host immune response to Clostridium difficile.* J Med Microbiol, 2011. **60**(Pt 8): p. 1070-9.
474. Wijburg, O.L., et al., *Innate secretory antibodies protect against natural Salmonella typhimurium infection.* J Exp Med, 2006. **203**(1): p. 21-6.
475. Johal, S.S., et al., *Colonic IgA producing cells and macrophages are reduced in recurrent and non-recurrent Clostridium difficile associated diarrhoea.* J Clin Pathol, 2004. **57**(9): p. 973-9.
476. Johnson, S., et al., *Selective neutralization of a bacterial enterotoxin by serum immunoglobulin A in response to mucosal disease.* Infect Immun, 1995. **63**(8): p. 3166-73.
477. Giannasca, P.J., et al., *Serum antitoxin antibodies mediate systemic and mucosal protection from Clostridium difficile disease in hamsters.* Infect Immun, 1999. **67**(2): p. 527-38.
478. Wingfield, P.T., *Overview of the purification of recombinant proteins.* Current protocols in protein science, 2015. **80**: p. 6.1.1-6.1.35.
479. Vigil, A., D.H. Davies, and P.L. Felgner, *Defining the humoral immune response to infectious agents using high-density protein microarrays.* Future microbiology, 2010. **5**(2): p. 241-251.
480. Näslund, K., et al., *Development and evaluation of an indirect enzyme-linked immunosorbent assay for serological detection of Schmallenberg virus antibodies in ruminants using whole virus antigen.* Acta veterinaria Scandinavica, 2014. **56**(1): p. 71-71.
481. Moore, S.J. and M.J. Warren, *The anaerobic biosynthesis of vitamin B12.* Biochem Soc Trans, 2012. **40**(3): p. 581-6.
482. Fang, H., J. Kang, and D. Zhang, *Microbial production of vitamin B12: a review and future perspectives.* Microb Cell Fact, 2017. **16**(1): p. 15.
483. Rodionov, D.A., et al., *Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes.* J Biol Chem, 2003. **278**(42): p. 41148-59.
484. Roth, J.R., et al., *Characterization of the cobalamin (vitamin B12) biosynthetic genes of Salmonella typhimurium.* Journal of Bacteriology, 1993. **175**(11): p. 3303-3316.
485. Warren, M.J., et al., *The biosynthesis of adenosylcobalamin (vitamin B12).* Nat Prod Rep, 2002. **19**(4): p. 390-412.
486. Reynolds, P.R., G.P. Mottur, and C. Bradbeer, *Transport of vitamin B12 in Escherichia coli. Some observations on the roles of the gene products of BtuC and TonB.* J Biol Chem, 1980. **255**(9): p. 4313-9.
487. Slotboom, D.J., *Structural and mechanistic insights into prokaryotic energy-coupling factor transporters.* Nat Rev Microbiol, 2014. **12**(2): p. 79-87.
488. Rodionov, D.A., et al., *A novel class of modular transporters for vitamins in prokaryotes.* Journal of bacteriology, 2009. **191**(1): p. 42-51.
489. Hebbeln, P., et al., *Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module.* Proc Natl Acad Sci U S A, 2007. **104**(8): p. 2909-14.
490. Rodionov, D.A., et al., *Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters.* J Bacteriol, 2006. **188**(1): p. 317-27.
491. Siche, S., et al., *A bipartite S unit of an ECF-type cobalt transporter.* Res Microbiol, 2010. **161**(10): p. 824-9.
492. Vitreschak, A.G., et al., *Regulation of the vitamin B(12) metabolism and transport in bacteria by a conserved RNA structural element.* RNA, 2003. **9**(9): p. 1084-1097.
493. Nahvi, A., J.E. Barrick, and R.R. Breaker, *Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes.* Nucleic Acids Res, 2004. **32**(1): p. 143-50.
494. Ranquet, C., et al., *Cobalt stress in Escherichia coli. The effect on the iron-sulfur proteins.* J Biol Chem, 2007. **282**(42): p. 30442-51.
495. Guskov, A. and S. Eshaghi, *The mechanisms of Mg2+ and Co2+ transport by the CorA family of divalent cation transporters.* Curr Top Membr, 2012. **69**: p. 393-414.

496. DeVeaux, L.C. and R.J. Kadner, *Transport of vitamin B12 in Escherichia coli: cloning of the btuCD region.* J Bacteriol, 1985. **162**(3): p. 888-96.
497. Korkhov, V.M., S.A. Mireku, and K.P. Locher, *Structure of AMP-PNP-bound vitamin B12 transporter BtuCD-F.* Nature, 2012. **490**(7420): p. 367-72.
498. Fernandes, N., R. Chaudri, and R. Fagan. *Assessing the fitness of Clostridium difficile mutants using Transposon Directed Insertion Site Sequencing (TraDIS)*. in *Microbiology Society Annual Conference 2017*. 2017. Edinburgh: Microbiology Society.
499. Cartman, S.T., et al., *Precise manipulation of the Clostridium difficile chromosome reveals a lack of association between the tcdC genotype and toxin production.* Appl Environ Microbiol, 2012. **78**(13): p. 4683-90.
500. Rodionov, D.A. and M.S. Gelfand, *Identification of a bacterial regulatory system for ribonucleotide reductases by phylogenetic profiling.* Trends Genet, 2005. **21**(7): p. 385-9.
501. Kaval, K.G. and D.A. Garsin, *Ethanolamine Utilization in Bacteria.* mBio, 2018. **9**(1): p. e00066-18.
502. Garsin, D.A., *Ethanolamine Utilization in Bacterial Pathogens: Roles and Regulation.* Nature reviews. Microbiology, 2010. **8**(4): p. 290-295.
503. Del Papa, M.F. and M. Perego, *Ethanolamine Activates a Sensor Histidine Kinase Regulating Its Utilization in Enterococcus faecalis.* Journal of Bacteriology, 2008. **190**(21): p. 7147.
504. Fox, K.A., et al., *Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in Enterococcus faecalis.* Proc Natl Acad Sci U S A, 2009. **106**(11): p. 4435-40.
505. Pitts, A.C., et al., *Structural insight into the Clostridium difficile ethanolamine utilisation microcompartment.* PLoS One, 2012. **7**(10): p. e48360.
506. Dembek, M., et al., *Transcriptional analysis of temporal gene expression in germinating Clostridium difficile 630 endospores.* PLoS One, 2013. **8**(5): p. e64011.
507. Dineen, S.S., S.M. McBride, and A.L. Sonenshein, *Integration of metabolism and virulence by Clostridium difficile CodY.* J Bacteriol, 2010. **192**(20): p. 5350-62.
508. Cassona, C.P., et al., *A Fluorescent Reporter for Single Cell Analysis of Gene Expression in Clostridium difficile.* Methods Mol Biol, 2016. **1476**: p. 69-90.
509. Pereira, F.C., et al., *The spore differentiation pathway in the enteric pathogen Clostridium difficile.* PLoS Genet, 2013. **9**(10): p. e1003782.
510. Buckley, A.M., et al., *Lighting Up Clostridium Difficile: Reporting Gene Expression Using Fluorescent Lov Domains.* Scientific Reports, 2016. **6**: p. 23463.
511. Hoffmann, C.E., E.L. Stokstad, et al., *The microbiological assay of vitamin B12 with Lactobacillus leichmannii.* J Biol Chem, 1949. **181**(2): p. 635-44.
512. Muller, A., et al., *Conservation of structure and mechanism in primary and secondary transporters exemplified by SiaP, a sialic acid binding virulence factor from Haemophilus influenzae.* J Biol Chem, 2006. **281**(31): p. 22212-22.
513. Chu, D. and D.J. Barnes, *The lag-phase during diauxic growth is a trade-off between fast adaptation and high growth rate.* Sci Rep, 2016. **6**: p. 25191.
514. Solopova, A., et al., *Bet-hedging during bacterial diauxic shift.* Proc Natl Acad Sci U S A, 2014. **111**(20): p. 7427-32.
515. Blackwell, C.M. and J.M. Turner, *Microbial metabolism of amino alcohols. Formation of coenzyme B12-dependent ethanolamine ammonia-lyase and its concerted induction in Escherichia coli.* Biochem J, 1978. **176**(3): p. 751-7.
516. Tsoy, O., D. Ravcheev, and A. Mushegian, *Comparative Genomics of Ethanolamine Utilization.* Journal of Bacteriology, 2009. **191**(23): p. 7157.
517. Mellin, J.R., et al., *Riboswitches. Sequestration of a two-component response regulator by a riboswitch-regulated noncoding RNA.* Science, 2014. **345**(6199): p. 940-3.
518. DebRoy, S., et al., *Riboswitches. A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator.* Science, 2014. **345**(6199): p. 937-40.

519. Bradbeer, C., *The clostridial fermentations of choline and ethanolamine. II. Requirement for a cobamide coenzyme by an ethanolamine deaminase.* J Biol Chem, 1965. **240**(12): p. 4675-81.
520. Dadswell, K., et al., *Bacterial Microcompartment-Mediated Ethanolamine Metabolism in Escherichia coli Urinary Tract Infection.* Infect Immun, 2019. **87**(8).
521. Semchenko, E.A., C.J. Day, and K.L. Seib, *MetQ of Neisseria gonorrhoeae Is a Surface-Expressed Antigen That Elicits Bactericidal and Functional Blocking Antibodies.* Infect Immun, 2017. **85**(2): p. e00898-16.
522. Otsuka, T., et al., *Substrate binding protein SBP2 of a putative ABC transporter as a novel vaccine antigen of Moraxella catarrhalis.* Infect Immun, 2014. **82**(8): p. 3503-12.
523. Brown, J.S., et al., *Immunization with components of two iron uptake ABC transporters protects mice against systemic Streptococcus pneumoniae infection.* Infect Immun, 2001. **69**(11): p. 6702-6.
524. Jomaa, M., et al., *Antibodies to the iron uptake ABC transporter lipoproteins PiaA and PiuA promote opsonophagocytosis of Streptococcus pneumoniae.* Infect Immun, 2005. **73**(10): p. 6852-9.
525. Hill, S., et al., *Immunogenicity and mechanisms of action of PnuBioVax, a multi-antigen serotype-independent prophylactic vaccine against infection with Streptococcus pneumoniae.* Vaccine, 2018. **36**(29): p. 4255-4264.
526. Begier, E., et al., *SA4Ag, a 4-antigen Staphylococcus aureus vaccine, rapidly induces high levels of bacteria-killing antibodies.* Vaccine, 2017. **35**(8): p. 1132-1139.
527. Haiko, J. and B. Westerlund-Wikström, *The role of the bacterial flagellum in adhesion and virulence.* Biology, 2013. **2**(4): p. 1242-1267.
528. Duan, Q., et al., *Flagella and bacterial pathogenicity.* J Basic Microbiol, 2013. **53**(1): p. 1-8.
529. Terashima, H., S. Kojima, and M. Homma, *Flagellar motility in bacteria structure and function of flagellar motor.* Int Rev Cell Mol Biol, 2008. **270**: p. 39-85.
530. Kearns, D.B., *A field guide to bacterial swarming motility.* Nature reviews. Microbiology, 2010. **8**(9): p. 634-644.
531. Mattick, J.S., *Type IV pili and twitching motility.* Annu Rev Microbiol, 2002. **56**: p. 289-314.
532. Merz, A.J., M. So, and M.P. Sheetz, *Pilus retraction powers bacterial twitching motility.* Nature, 2000. **407**(6800): p. 98-102.
533. Skerker, J.M. and H.C. Berg, *Direct observation of extension and retraction of type IV pili.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6901-4.
534. Varga, J.J., et al., *Type IV pili-dependent gliding motility in the Gram-positive pathogen Clostridium perfringens and other Clostridia.* Mol Microbiol, 2006. **62**(3): p. 680-94.
535. Sauvonnet, N., et al., *Pilus formation and protein secretion by the same machinery in Escherichia coli.* The EMBO journal, 2000. **19**(10): p. 2221-2228.
536. Kline, K.A., et al., *A tale of two pili: assembly and function of pili in bacteria.* Trends in microbiology, 2010. **18**(5): p. 224-232.
537. Xia, J., et al., *Type IV pilus biogenesis genes and their roles in biofilm formation in the biological control agent Lysobacter enzymogenes OH11.* Appl Microbiol Biotechnol, 2018. **102**(2): p. 833-846.
538. Craig, L., et al., *Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions.* Mol Cell, 2006. **23**(5): p. 651-62.
539. Korotkov, K.V. and W.G. Hol, *Structure of the GspK-GspI-GspJ complex from the enterotoxigenic Escherichia coli type 2 secretion system.* Nat Struct Mol Biol, 2008. **15**(5): p. 462-8.
540. Korotkov, K.V., M. Sandkvist, and W.G. Hol, *The type II secretion system: biogenesis, molecular architecture and mechanism.* Nat Rev Microbiol, 2012. **10**(5): p. 336-51.
541. Couchman, E., *Investigating the Type IV Pili of Clostridium difficile and Clostridium sordellii.* 2016. PhD Thesis, Imperial College, London.

542. Bordeleau, E., et al., *c-di-GMP turn-over in Clostridium difficile is controlled by a plethora of diguanylate cyclases and phosphodiesterases.* PLoS Genet, 2011. **7**(3): p. e1002039.
543. Morales-Soto, N., et al., *Preparation, Imaging, and Quantification of Bacterial Surface Motility Assays.* Journal of Visualized Experiments : JoVE, 2015(98): p. 52338.
544. Barketi-Klai, A., et al., *The flagellin FliC of Clostridium difficile is responsible for pleiotropic gene regulation during in vivo infection.* PLoS One, 2014. **9**(5): p. e96876.
545. Garrett, E.M., et al., *Phase variation of a signal transduction system controls Clostridioides difficile colony morphology, motility, and virulence.* PLoS Biol, 2019. **17**(10): p. e3000379.
546. Maldarelli, G.A., et al., *Pilin Vaccination Stimulates Weak Antibody Responses and Provides No Protection in a C57Bl/6 Murine Model of Acute Clostridium difficile Infection.* J Vaccines Vaccin, 2016. **7**(3): p. 321.
547. Garmory, H.S. and R.W. Titball, *ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies.* Infection and immunity, 2004. **72**(12): p. 6757-6763.
548. Harland, D.N., et al., *An association between ATP binding cassette systems, genome sizes and lifestyles of bacteria.* Res Microbiol, 2005. **156**(3): p. 434-42.
549. Forgetta, V., et al., *Fourteen-Genome Comparison Identifies DNA Markers for Severe-Disease-Associated Strains of Clostridium difficile.* Journal of Clinical Microbiology, 2011. **49**(6): p. 2230-2238.
550. Yakushi, T., et al., *A new ABC transporter mediating the detachment of lipid-modified proteins from membranes.* Nat Cell Biol, 2000. **2**(4): p. 212-8.
551. Babu, M.M., et al., *A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins.* Journal of bacteriology, 2006. **188**(8): p. 2761-2773.
552. Greene, N.P., et al., *Antibiotic Resistance Mediated by the MacB ABC Transporter Family: A Structural and Functional Perspective.* Front Microbiol, 2018. **9**: p. 950.
553. Kobayashi, N., K. Nishino, and A. Yamaguchi, *Novel macrolide-specific ABC-type efflux transporter in Escherichia coli.* J Bacteriol, 2001. **183**(19): p. 5639-44.
554. Shea, A., et al., *Biolog phenotype microarrays.* Methods Mol Biol, 2012. **881**: p. 331-73.
555. Kovacs-Simon, A., R.W. Titball, and S.L. Michell, *Lipoproteins of Bacterial Pathogens.* Infection and Immunity, 2011. **79**(2): p. 548-561.
556. Crow, A., et al., *Structure and mechanotransmission mechanism of the MacB ABC transporter superfamily.* Proc Natl Acad Sci U S A, 2017. **114**(47): p. 12572-12577.
557. Rouquette-Loughlin, C.E., J.T. Balthazar, and W.M. Shafer, *Characterization of the MacA-MacB efflux system in Neisseria gonorrhoeae.* J Antimicrob Chemother, 2005. **56**(5): p. 856-60.
558. Mullany, P., M. Wilks, and S. Tabaqchali, *Transfer of macrolide-lincosamide-streptogramin B (MLS) resistance in Clostridium difficile is linked to a gene homologous with toxin A and is mediated by a conjugative transposon, Tn5398.* J Antimicrob Chemother, 1995. **35**(2): p. 305-15.
559. Farrow, K.A., D. Lyras, and J.I. Rood, *Genomic analysis of the erythromycin resistance element Tn5398 from Clostridium difficile.* Microbiology, 2001. **147**(Pt 10): p. 2717-28.
560. Spigaglia, P. and P. Mastrantonio, *Analysis of macrolide-lincosamide-streptogramin B (MLS(B)) resistance determinant in strains of Clostridium difficile.* Microb Drug Resist, 2002. **8**(1): p. 45-53.
561. Arthur, M., A. Brisson-Noel, and P. Courvalin, *Origin and evolution of genes specifying resistance to macrolide, lincosamide and streptogramin antibiotics: data and hypotheses.* J Antimicrob Chemother, 1987. **20**(6): p. 783-802.
562. Webber, M.A. and L.J. Piddock, *The importance of efflux pumps in bacterial antibiotic resistance.* J Antimicrob Chemother, 2003. **51**(1): p. 9-11.
563. Nawrocki, K.L., E.K. Crispell, and S.M. McBride, *Antimicrobial Peptide Resistance Mechanisms of Gram-Positive Bacteria.* Antibiotics, 2014. **3**(4): p. 461-492.

564. McBride, S.M. and A.L. Sonenshein, *Identification of a genetic locus responsible for antimicrobial peptide resistance in Clostridium difficile.* Infect Immun, 2011. **79**(1): p. 167-76.
565. Scaria, J., et al., *Comparative nutritional and chemical phenome of Clostridium difficile isolates determined using phenotype microarrays.* Int J Infect Dis, 2014. **27**: p. 20-5.
566. Blumenstein, K., et al., *Phenotype MicroArrays as a complementary tool to next generation sequencing for characterization of tree endophytes.* Front Microbiol, 2015. **6**: p. 1033.
567. Greetham, D., *Phenotype microarray technology and its application in industrial biotechnology.* Biotechnol Lett, 2014. **36**(6): p. 1153-60.
568. Lei, X.-H. and B.R. Bochner, *Using phenotype microarrays to determine culture conditions that induce or repress toxin production by Clostridium difficile and other microorganisms.* PloS one, 2013. **8**(2): p. e56545-e56545.
569. Murphy, T.F., et al., *ATP-Binding Cassette (ABC) Transporters of the Human Respiratory Tract Pathogen, Moraxella catarrhalis: Role in Virulence.* PloS one, 2016. **11**(7): p. e0158689-e0158689.
570. Tanabe, M., et al., *The ABC Transporter Protein OppA Provides Protection against Experimental Yersinia pestis Infection.* Infection and Immunity, 2006. **74**(6): p. 3687.
571. Wright, A., et al., *Immunoreactive cell wall proteins of Clostridium difficile identified by human sera.* J Med Microbiol, 2008. **57**(Pt 6): p. 750-6.
572. Berti, F. and R. Adamo, *Antimicrobial glycoconjugate vaccines: an overview of classic and modern approaches for protein modification.* Chem Soc Rev, 2018. **47**(24): p. 9015-9025.
573. Prior, J.L., et al., *Characterization of the O antigen gene cluster and structural analysis of the O antigen of Francisella tularensis subsp. tularensis.* J Med Microbiol, 2003. **52**(Pt 10): p. 845-51.
574. Liu, D. and P.R. Reeves, *Escherichia coli K12 regains its O antigen.* Microbiology, 1994. **140**(1): p. 49-57.
575. Hong, Y. and P.R. Reeves, *Diversity of O-Antigen Repeat Unit Structures Can Account for the Substantial Sequence Variation of Wzx Translocases.* Journal of Bacteriology, 2014. **196**(9): p. 1713.
576. Nothaft, H., et al., *Engineering the Campylobacter jejuni N-glycan to create an effective chicken vaccine.* Scientific Reports, 2016. **6**: p. 26511.
577. Nita-Lazar, M., et al., *The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation.* Glycobiology, 2004. **15**(4): p. 361-367.
578. Chen, M.M., K.J. Glover, and B. Imperiali, *From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni.* Biochemistry, 2007. **46**(18): p. 5579-85.
579. Schierle, C.F., et al., *The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the Escherichia coli periplasm via the signal recognition particle pathway.* J Bacteriol, 2003. **185**(19): p. 5706-13.
580. Hochuli, E., et al., *Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent.* Bio/Technology, 1988. **6**(11): p. 1321-1325.
581. Dykxhoorn, D.M., R. St Pierre, and T. Linn, *A set of compatible tac promoter expression vectors.* Gene, 1996. **177**(1-2): p. 133-6.
582. Schulz, H., H. Hennecke, and L. Thony-Meyer, *Prototype of a heme chaperone essential for cytochrome c maturation.* Science, 1998. **281**(5380): p. 1197-200.
583. Jaroentomeechai, T., et al., *Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery.* Nat Commun, 2018. **9**(1): p. 2686.
584. Kabanova, A., et al., *Evaluation of a Group A Streptococcus synthetic oligosaccharide as vaccine candidate.* Vaccine, 2010. **29**(1): p. 104-14.

585. Perepelov, A.V., et al., *Shigella flexneri O-antigens revisited: final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity.* FEMS Immunol Med Microbiol, 2012. **66**(2): p. 201-10.
586. Liu, B., et al., *Structure and genetics of Shigella O antigens.* FEMS Microbiol Rev, 2008. **32**(4): p. 627-53.
587. Morona, R., et al., *Characterization of the rfc region of Shigella flexneri.* J Bacteriol, 1994. **176**(3): p. 733-47.
588. Ihssen, J., et al., *Production of glycoprotein vaccines in Escherichia coli.* Microb Cell Fact, 2010. **9**: p. 61.
589. Hug, I., et al., *Exploiting bacterial glycosylation machineries for the synthesis of a Lewis antigen-containing glycoprotein.* J Biol Chem, 2011. **286**(43): p. 37887-94.
590. Bu, S., et al., *Interaction between two putative glycosyltransferases is required for glycosylation of a serine-rich streptococcal adhesin.* J Bacteriol, 2008. **190**(4): p. 1256-66.
591. Marolda, C.L., J. Vicarioli, and M.A. Valvano, *Wzx proteins involved in biosynthesis of O antigen function in association with the first sugar of the O-specific lipopolysaccharide subunit.* Microbiology, 2004. **150**(Pt 12): p. 4095-105.
592. Kenyon, J.J., M.M. Cunneen, and P.R. Reeves, *Genetics and evolution of Yersinia pseudotuberculosis O-specific polysaccharides: a novel pattern of O-antigen diversity.* FEMS microbiology reviews, 2017. **41**(2): p. 200-217.
593. Islam, S.T. and J.S. Lam, *Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway.* Can J Microbiol, 2014. **60**(11): p. 697-716.
594. Valderrama-Rincon, J.D., et al., *An engineered eukaryotic protein glycosylation pathway in Escherichia coli.* Nat Chem Biol, 2012. **8**(5): p. 434-6.
595. Shang, W., et al., *Production of human blood group B antigen epitope conjugated protein in Escherichia coli and utilization of the adsorption blood group B antibody.* Microb Cell Fact, 2016. **15**(1): p. 138.
596. Wetter, M., et al., *Engineering, conjugation, and immunogenicity assessment of Escherichia coli O121 O antigen for its potential use as a typhoid vaccine component.* Glycoconj J, 2013. **30**(5): p. 511-22.
597. Nishiuchi, Y., et al., *Structure and serologic properties of O-specific polysaccharide from Citrobacter freundii possessing cross-reactivity with Escherichia coli O157:H7.* FEMS Immunol Med Microbiol, 2000. **28**(2): p. 163-71.
598. Mistou, M.-Y., I.C. Sutcliffe, and N.M. van Sorge, *Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria.* FEMS microbiology reviews, 2016. **40**(4): p. 464-479.
599. Pan, C., et al., *Biosynthesis of Conjugate Vaccines Using an O-Linked Glycosylation System.* MBio, 2016. **7**(2): p. e00443-16.
600. Dubberke, E.R. and M.A. Olsen, *Burden of Clostridium difficile on the healthcare system.* Clin Infect Dis, 2012. **55** (Suppl 2): p. S88-92.
601. Guy, B., *Evaluation of events occurring at mucosal surfaces: techniques used to collect and analyze mucosal secretions and cells.* Clinical and diagnostic laboratory immunology, 2002. **9**(4): p. 753-762.
602. Illingworth, J.J., et al., *Functional Comparison of Blood-Stage Plasmodium falciparum Malaria Vaccine Candidate Antigens.* Frontiers in immunology, 2019. **10**: p. 1254-1254.
603. Lee, H., et al., *The 7-valent pneumococcal conjugate vaccine elicits cross-functional opsonophagocytic killing responses to Streptococcus pneumoniae serotype 6D in children.* BMC Infect Dis, 2013. **13**: p. 474.
604. Ramachandran, G., et al., *Opsonophagocytic Assay To Evaluate Immunogenicity of Nontyphoidal Salmonella Vaccines.* Clinical and Vaccine Immunology, 2016. **23**(6): p. 520.
605. Romero-Steiner, S., et al., *Use of Opsonophagocytosis for Serological Evaluation of Pneumococcal Vaccines.* Clinical and Vaccine Immunology, 2006. **13**(2): p. 165.

606. Song, J.Y., et al., *Pneumococcal vaccine and opsonic pneumococcal antibody.* J Infect Chemother, 2013. **19**(3): p. 412-25.
607. Degnan, P.H., et al., *Human gut microbes use multiple transporters to distinguish vitamin B12 analogs and compete in the gut.* Cell Host Microbe, 2014. **15**(1): p. 47-57.
608. Degnan, P.H., M.E. Taga, and A.L. Goodman, *Vitamin B12 as a modulator of gut microbial ecology.* Cell Metab, 2014. **20**(5): p. 769-778.
609. Mahour, R., et al., *Establishment of a five-enzyme cell-free cascade for the synthesis of uridine diphosphate N-acetylglucosamine.* J Biotechnol, 2018. **283**: p. 120-129.
610. Rexer, T.F.T., et al., *One pot synthesis of GDP-mannose by a multi-enzyme cascade for enzymatic assembly of lipid-linked oligosaccharides.* Biotechnol Bioeng, 2018. **115**(1): p. 192-205.
611. Hayashi, F., et al., *The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5.* Nature, 2001. **410**(6832): p. 1099-103.
612. Pechine, S., C. Janoir, and A. Collignon, *Variability of Clostridium difficile surface proteins and specific serum antibody response in patients with Clostridium difficile-associated disease.* J Clin Microbiol, 2005. **43**(10): p. 5018-25.
613. Ihssen, J., et al., *Increased efficiency of Campylobacter jejuni N-oligosaccharyltransferase PglB by structure-guided engineering.* Open biology, 2015. **5**(4): p. 140227-140227.
614. Moor, K., et al., *Analysis of bacterial-surface-specific antibodies in body fluids using bacterial flow cytometry.* Nat Protoc, 2016. **11**(8): p. 1531-53.
615. Alvarez-Barrientos, A., et al., *Applications of flow cytometry to clinical microbiology.* Clinical microbiology reviews, 2000. **13**(2): p. 167-195.
616. Haynes, M.K., et al., *High-Throughput Flow Cytometry Screening of Multidrug Efflux Systems.* Methods Mol Biol, 2018. **1700**: p. 293-318.
617. Jindal, S., et al., *Involvement of multiple influx and efflux transporters in the accumulation of cationic fluorescent dyes by Escherichia coli.* BMC Microbiol, 2019. **19**(1): p. 195.
618. Ravcheev, D.A., et al., *Comparative Genomic Analysis Reveals Novel Microcompartment-Associated Metabolic Pathways in the Human Gut Microbiome.* Frontiers in Genetics, 2019. **10**(636).
619. Zhou, J., et al., *Ethanolamine enhances intestinal functions by altering gut microbiome and mucosal anti-stress capacity in weaned rats.* Br J Nutr, 2018. **120**(3): p. 241-249.
620. Edwards, A.N., K.L. Nawrocki, and S.M. McBride, *Conserved oligopeptide permeases modulate sporulation initiation in Clostridium difficile.* Infection and immunity, 2014. **82**(10): p. 4276-4291.
621. Stevenson, E., N.P. Minton, and S.A. Kuehne, *The role of flagella in Clostridium difficile pathogenicity.* Trends Microbiol, 2015. **23**(5): p. 275-82.

Appendices

A.

Table A1. Plasmids used in this study

Plasmid	Description	Source
pETM11	Protein expression vector with T7 promoter and kanamycin resistance cassette and N-terminal 6XHistag with TEV cleavage site.	[1]
pETM11_0342 S1	pETM11 with residues 41-242 from CDR20291_0342	This study
pETM11_0342 S3	pETM11 with residues 439-728 from CDR20291_0342	This study
pETM11_3343	pETM11 with residues 34-479 from CDR20291_3343	This study
pETM11_0330	pETM11 with residues 22-95 from CDR20291_0330	This study
pETM11_3155	pETM11 with residues 34-56 from CDR20291_3155	This study
pETM11_2253	pETM11 with residues 1-132 from CDR20291_2253	This study
pETM11_2697	pETM11 with residues 43-209 from CDR20291_2697	This study
pETM11_2640	pETM11 with residues 1-191 from CDR20291_2640	This study
pMTL82151	Replication-defective *C. difficile* vector with chloramphenicol resistance cassette and multiple cloning site (MCS).	[2]
pMTL82151_0342	pMTL82151 vector harbouring allele exchange mutagenesis cassette for *CDR20291_0330* between BamHI and HindIII	This study
pMTL82151_3343	pMTL82151 vector harbouring allele exchange mutagenesis cassette for *CDR20291_3343* between EcoRI and BamHI	This study
pMTL82151_0330	pMTL82151 vector harbouring allele exchange mutagenesis cassette for *CDR20291_0342* between EcoRI and XmaI	This study
pEXT20	High copy number protein expression vector with ampicillin resistance cassette and P_{tac} promoter.	[3]
pEXT21	Low copy number protein expression vector with spectinomycin resistance cassette and P_{tac} promoter.	[3]
pEXT21_pglB	pEXT21 carrying full length PglB	[4]
pGVXN115	pEXT21 carrying inactivated PglB	GlycoVaxyn
pACYC*pgl*Δ*pglB*	pACYC carrying *pgl* glycosylation locus from *Campylobacter jejuni* with inactivated PglB	[5]
pEXT20_AcrAtag	pEXT20 harbouring the acceptor protein AcrAtag	This study
pEXT20_GT2	pEXT20 harbouring *CDR20291_0242* (GT2)	This study
pEXT21_AcrAtag	pEXT21 harbouring the acceptor protein AcrAtag	This study
pEXT21_GT2	pEXT21 harbouring *CDR20291_0242* (GT2)	This study
pEXT20_GT2-His	pEXT20 harbouring *CDR20291_0242* with C-terminal 6XHistag (GT2-His)	This study
pEC415	Ampicillin resistant with P_{BAD} promoter and *araC* gene for L-arabinose induction	[6]
pEC415_AcrAtag	pEC415 harbouring the acceptor protein AcrAtag	This study
pEC415_GT2	pEC415 harbouring *CDR20291_0242* (GT2)	This study
pEC415_GT2-His	pEC415 harbouring *CDR20291_0242* with C-terminal 6XHistag (GT2-His)	This study

pCH plasmids	Backbone of pEXT20 with P_{BAD} promoter and *araC* gene from pEC415, enabling dual expression of constructs under IPTG and L-arabinose induction	This study
pCH01	AcrAtag under P_{tac} promoter and GT2 under P_{BAD} promoter	This study
pCH03	AcrAtag under P_{tac} promoter and GT2-his under P_{BAD} promoter	This study
pCH05	GT2 under P_{tac} promoter and AcrAtag under P_{BAD} promoter	This study
pCH07	GT2-His under P_{tac} promoter and AcrAtag-His under P_{BAD} promoter	This study
pECC12	Chloramphenicol resistant, constitutive expression of the diguanylate cyclase *dccA*	[7]
pECC17	Chloramphenicol resistant, inducible expression of the diguanylate cyclase *dccA*, under control of P_{tac} promoter	[7]
pASF085	Empty plasmid version of pECC17	Neil Fairweather

Table A2. Oligonucleotides used in this study.

Primer Name	Sequence
Construction of plasmids for expression of recombinant *C. difficile* proteins in *E. coli*	
PAC_R20291_3155_F	TACGGGGTCTCCCATGTACAGCAAAGTACAGGAGTC
PAC_R20291_3155_R	TCGAAGAATTCTTACTTGGGGTAAAACTGCTTTG
PAC_R20291_0342S1_F	TACGGGGTCTCCCATGCGCAGCGCTCAACAAGCCGA
PAC_R20291_0342S1_R	TCGAAGAATTCTTAGGAGTTGTCCAACGTACCGT
PAC_R20291_0342S3_F	TACGGGGTCTCCCATGGACATCCAGAATAAGAACGG
PAC_R20291_0342S3_R	TCGAAGAATTCTTAACGAGCGGTCATTTTATCAA
PAC_R20291_0330_F	TACGGGGTCTCCCATGATCTTCCCCCTGTTAGTAAA
PAC_R20291_0330_R	TCGAAGAATTCTTAACTGCGCTTCCCCTTTTGAA
PAC_R20291_2253_F	TACGGGGTCTCCCATGTTGAAAAAAATCGTCAT
PAC_R20291_2253_R	TCGAAGAATTCTTAGCAGCATTTATCCATGTCTT
PAC_R20291_2697_GibF	GGTGCTCGAGTTACGACGCAAAGTCCTC
PAC_R20291_2697_GibR	GGGCGCCATGTCTAAAGTAGTATCTGCGTC
pETM11_2697_f	CTACTTTAGACATGGCGCCCTGAAAATAAAG
pETM11_2697_r	TGCGTCGTAACTCGAGCACCACCACCAC
PAC_R20291_2640_GibF	GCTCGAGTTAAACTTTAGCCTTTAATTTGCCTAAAACC
PAC_R20291_2640_GibR	GGGCGCCATGTTCAAGCGCTACGCCGAAAAAATG
pETM11_2640_f	AGCGCTTGAACATGGCGCCCTGAAAATAAAG
pETM11_2640_r	GGCTAAAGTTTAACTCGAGCACCACCAC
PAC_R20291_3343_GibF	GCTCGAGTTAGTTTACTTTCTTGTAGGAGCTAATG
PAC_R20291_3343_GibR	GGGCGCCATGAACCAAATCGCGAATCGC
pETM11_3343_f	CGATTTGGTTCATGGCGCCCTGAAAATAAAG
pETM11_3343_r	GAAAGTAAACTAACTCGAGCACCACCAC
T7_promoter	TAATACGACTCACTATAGGG
T7_terminator	GCTAGTTATTGCTCAGCGG
Construction of plasmids for allele exchange mutagenesis	
R20291_3343_HA1_F	TTGACGGAATTCGGTCTTACTTTATTAGAAGT
R20291_3343_HA1_R	TTAGTTTACTTTTTTGTATGATTTATTCCACTTCCTCAA
R20291_3343_HA2_F	TTGAGGAAGTGGAATAAATCATACAAAAAAGTAAACTAA
R20291_3343_HA2_R	CCGATTGGATCCCAGTACACCTACTAAGTGCC
R20291_0330_HA1_F	TATGGCGGATCCAGAAATTTTCACAGAAAAGG
R20291_0330_HA1_R	TTAACTTCTCTTTCCTTTTTGCGTCTTTGTTTTAGCACTCAT
R20291_0330_HA2_F	ATGAGTGCTAAAACAAAGACGCAAAAAGGAAAGAGAAGTTAA
R20291_0330_HA2_R	ACGGTAAAGCTTATTCCATTAACAGCTACTAG
R20291_0342_HA1_F	TTGACGGAATTCCGTCACTATTTACAAGTATA
R20291_0342_HA1_R	TGTAATGTTTATAGCTTCAACTATTTTTAAATTATCTTTCAT
R20291_0342_HA2_F	ATGAAAGATAATTTAAAAATAGTTGAAGCTATAAACATTACA
R20291_0342_HA2_R	CCGATTCCCGGGATCCACTTCAAAGTTGGCAA
R20291_0342_SEQ_F	ATATCCTAAACTTAATTGTA
R20291_0342_SEQ_R	GAATTAGGTTATACTGTTGA
R20291_3343_MUT_F	GAGAAGAGATAATCTATAAT

R20291_3343_MUT_R	CCTATTAGTATTACATCTGG
R20291_0330_MUT_F	GAGTGATGGCTTCAGCAATT
R20291_0330_MUT_R	ACACAAACAACTAGAATACT
M13_f	TGTAAAACGACGGCCAGT
M13_r	CAGGAAACAGCTATGACC
Construction of plasmids for bioconjugation	
acrAtag_f	TGATTGAATTCGCATGAAAAAAATCTGGTTAGC
acrAtag_r	TGATTAAGCTTTTAGTGGTGGTGAT
R20291_0242_F	TTGACGGAATTCATGAATACACCAATAGTGGT
R20291_0242_R	ACGGTAGAGCTCTTATTCATATCTCTTCCCCT
pEXTseq_f	GGTATGGCTGTGCAGGTCGT
pEXTseq_r	GCTTAATTTGATGCCTGGCA
pEC415_0242_f	GATATGAATAAGAATTCTTGAAGACGAAAGG
pEC415_0242_r	GTGTATTCATTATGTTATTCCTCCTTATTTAAAATG
0242_pEC415_f	GAATAACATAATGAATACACCAATAGTGGTTA
0242_pEC415_r	TCAAGAATTCTTATTCATATCTCTTCCCCTC
pEC415_acrAtag_f	CCACCACTAAGAATTCTTGAAGACGAAAGG
pEC415_acrAtag_r	AGATTTTTTTCATTATGTTATTCCTCCTTATTTAAAATG
acrAtag_pEC415_f	GAATAACATAATGAAAAAAATCTGGTTAGCTT
acrAtag_pEC415_r	TCAAGAATTCTTAGTGGTGGTGATGGTG
pEC415_0242tag_f	GAACGGGAGTCCAAAAATTGAGCTCG
pEC415_0242tag_r	GTAAATGCATTTCTGTGTGAGCAAAAAC
pEXT20_0242tag_f	TCACACAGAAATGCATTTACGTTGACAC
pEXT20_0242tag_r	CAATTTTTGGACTCCCGTTCTGGATAA
pEC415_0242his_f	CCACCACTAAGAATTCTTGAAGACGAAAGG
pEC415_0242his_r	GTGTATTCATTATGTTATTCCTCCTTATTTAAAATG
0242his_pEC415_f	GAATAACATAATGAATACACCAATAGTGGTTAATG
0242his_pEC415_r	TCAAGAATTCTTAGTGGTGGTGATGGTG
0242hisgib_f	AACAGAATTCATGAATACACCAATAGTGGTTAATGAAAAAT
0242hisgib_r	GCCCGCCCCCTTCATATCTCTTCCCCTCAATAAAATAAAG
pEXT20hisgib_f	GAGATATGAAGGGGGCGGGCATCACCAT
pEXT20hisgib_r	GTGTATTCATGAATTCTGTTTCCTGTGTGAAATTGTTATCCG
pEC415screen_f	GTCGAGCTAGTAAAAGCATT
pEC415screen_r	ATAGGCGTATCACGAGG

B.

Plasmid Construction

i. Plasmid construction for expression of *C. difficile* recombinant proteins in *E. coli*

The DNA sequence encoding each *C. difficile* protein selected for recombinant expression in *E. coli* was firstly codon optimised for *E. coli* using the IDT Codon Optimisation Tool (https://www.idtdna.com/CodonOpt), then ordered for synthesis as gBlock gene fragments. Regions of each protein selected for expression were cloned into the pETM11 plasmid using two different strategies.

pETM11_0342S1, pETM11_0342S3, pETM11_0330 and pETM11_3155 were built using restriction/ligation cloning. pETM11 encodes an N-terminal his-tag followed by a TEV cleavage site and NcoI recognition site for insertion of the coding sequence. However, inserting the coding sequence into this site would result in a frame shift, meaning the correct protein would not be expressed. To overcome this, the sequence encoding for the region of the protein to be expressed was amplified from the gBlocks using primers PAC_R20291_name of protein, which contained a 5' restriction sites, BsaI on the forward primer and XhoI on the reverse. BsaI is a type II endonuclease which cuts downstream of its recognition site so the cutting site was designed to leave an overhang complimentary to pETM11 following digestion with NcoI but which would result in loss of the final guanidine nucleotide of the NcoI recognition sequence, meaning the coding sequence would be in frame. pETM11 was digested with NcoI and XhoI then ligated with the respective insert. Following ligation, transformation into electro-competent Top10 cells and plated on LB agar supplemented with kanamycin (to select for pETM11). Plasmids were checked by colony PCR using the T7 promoter and terminator primers flanking the insertion site and test digest with XbaI and XhoI, then confirmed with DNA sequencing, again using the T7 primers.

pETM11_2640, pETM11_2697 and pETM11_3343 were built using Gibson assembly, as described in Methods, section 2.4.6. The backbone of pETM11 was amplified to include complimentary overhangs to 2640, 2697 or 3343, using primers pETM11_name of protein. Protein sequences were amplified to include complimentary overhangs to pETM11, using primers PAC_R20291_name of protein_Gib. Amplified sequons were then joined as described in Methods section 2.4.6 and screened for insertion of the insert using PCR with the T7 primers followed by DNA sequencing with the T7 primers.

ii. Plasmid construction for allele exchange mutagenesis

pMTL82151_3343, pMTL82151_0330 and pMTL82151_0342 were constructed using restriction ligation. SOE PCR was used to generate DNA constructs comprising two ~1200 bp DNA regions homologous to those flanking the gene for deletion, with 42 bp of the original gene remaining and restriction sites at the 5' and 3'. Construction of the allele exchange cassette Initially required amplification of the two homology arms using primers HA1_F and HA1_R or HA2_F and HA2_R (each of these is preceded by R20291_name of construct) with 21 bp of the beginning or the end of the gene for deletion incorporated, respectively. HA1_R and HA2_F primers contain overhangs complimentary to the other primer, so amplification of both products using HA1_F and HA2_R generates one construct with both regions of homology and 42 bp of the gene for deletion. These joined constructs were amplified with restriction digest sites for cloning into pMTL82151. pMTL82151 was digested with restriction enzymes matching those for digestion of the insert which were; EcoRI and BamHI for 3343, EcoRI and XmaI for 0342 and BamHI and XmaI for 0330. Following ligation, transformation and selection on Lb chloramphenicol plates and plasmid isolation all plasmids were screened by PCR using with MUTSEQ_f and MUTSEQ_r, which cover the whole construct, including homology arms, or SEQ_F or SEQ_R which flank the gene for deletion (each of these primers is preceded with R20291_name of construct). Potential gene deletion mutants were confirmed by DNA sequencing with M13_F and M13_R. Additional information about the screening for each construct is provided in the Chapters 4, 5 and 6.

iii. Plasmid construction for bioconjugation

Glycosyltransferase 2.

For expression of the *C. difficile* flagella glycosyltransferase 2 (GT2) in *E. coli*, the gene encoding this protein, *CDR20291_0242*, was amplified from *C. difficile* gDNA using primers R20291_0242_F and R These primers added EcoRI and SacI restriction sites to the 5' and 3' of the GT2 sequence, respectively so these could be cloned into the pEXT20 and pEXT21 vectors which share the same multiple cloning site (MCS), using restriction/ligation cloning. Plasmids were screened for the GT2 insert using test digestion with EcoRI and SacI and sequencing with primers pEXTseq_F and pEXTseq_R.

Acceptor protein.

For expression of the acceptor protein AcrAtag, this was first ordered for synthesis as a gBlock then amplified with primers acrAtag_F and acrAtag_R These primers added EcoRI and HindIII restriction sites to the 5' and 3' of the GT2 sequence, respectively so these could be cloned into the pEXT20 and pEXT21 vectors which share the same multiple cloning site (MCS), using restriction/ligation cloning. Plasmids were screened for the GT2 insert using DNA sequencing with primers pEXTseq_F and pEXTseq_R.

Dual expression plasmids.

For construction of one plasmid expressing both the acceptor protein acrAtag and GT2 (pCH01, pCH03, pCH05 and pCH07), Gibson assembly was utilised for all plasmids (Methods, section 2.4.6). Codon optimised *R20291_0242* was ordered for gBlock synthesis with a C-terminal 6XHistag. 0242, 0242His and AcrAtag were inserted into pEXT20 and pEC415 before synthesis of the dual expression plasmid (AcrAtag and GT2 were already inserted into pEXT20, as described above). For insertion of 0242 into pEC415, the insert was amplified with 0242_pEC415_f and 0242_pEC415_r and the plasmid amplified with pEC415_0242_f and pEC415_0242_r. For insertion of 0242his into pEC415, the insert was amplified with 0242his_pEC415_f and 0242his_pEC415_r and the plasmid amplified

with pEC415_0242his_f and pEC415_0242his_r. For insertion of 0242his into pEXT20, the insert was amplified with 0242hisgib_f and 0242hisgib_r and the plasmid amplified with pEXT20hisgib_f and pEXT20hisgib_r. For insertion of AcrAtag into pEC415, the insert was amplified with acrAtag_pEC415_f and acrAtag_pEC415_r and the plasmid amplified with pEC415_acrAtag_f and pEC415_acrAtag_r. All constructed plasmids were screened using DNA sequencing with primers pEXTseq_f and r for pEXT and pEC415screen_f and r for pEC415. The same primers were used for synthesis of pCH01, pCH03, pCH05 and pCH07. For pCH01, pEXT20_acrAtag was amplified with pEXT20_0242tag_f and r and pEC415_GT2 was amplified with pEC415_0242his_f. For pCH03, pEXT20_acrAtag was amplified with pEXT20_0242tag_f and r and pEC415_GT2-His was amplified with pEC415_0242his_f. For pCH05, pEXT20_GT2 was amplified with pEXT20_0242tag_f and r and pEC415_AcrAtag was amplified with pEC415_0242his_f. For pCH07, pEXT20_GT2-His was amplified with pEXT20_0242tag_f and r and pEC415_AcrAtag was amplified with pEC415_0242his_f. All plasmids were synthesised using Gibson assembly (Methods, 2.4.6).

C.

Ethics approval awarded to University of Liverpool for collection of patient samples, and ethical approval letter to work on the samples at LSHTM.

NHS

National Research Ethics Service

Liverpool (Adult) Research Ethics Committee
Bishop Goss Complex
Victoria Building
Rose Place
Liverpool
L3 3AN

Telephone: 0151 330 2077
Facsimile: 0151 330 2075

23 May 2008

Professor Munir Pirmohamed
Professor of Clinical Pharmacology & Honorary Consultant Physician
University of Liverpool
Sherrington Building
Ashton Street
Liverpool
L69 3GE

Dear Professor Pirmohamed

Full title of study: **Clostridium difficile-associated toxin disease: development of a tool to predict individual susceptibility based on environmental and genetic factors**

REC reference number: **08/H1005/32**

Thank you for your letter of 23 May 2008, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Ethical review of research sites

The Committee has designated this study as exempt from site-specific assessment (SSA. There is no requirement for [other] Local Research Ethics Committees to be informed or for site-specific assessment to be carried out at each site.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission at NHS sites ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Dr Lisa Dawson

LSHTM

16 May 2019

Dear Dr Lisa Dawson,

Study Title: Protein candidates for vaccines

LSHTM ethics ref: 17248

Thank you for your application for the above research, which has now been considered by the Observational Committee.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Investigator CV	Bio Sketch Dawson 20180525	04/04/2019	1
Local Approval	STK Trust Appr_11.12.13	08/04/2019	1
Protocol / Proposal	1_PIS Case	08/04/2019	1
Protocol / Proposal	2_PIS Control 1	08/04/2019	1
Protocol / Proposal	3_PIS Control 2	08/04/2019	1
Protocol / Proposal	4_PIS Control 3	08/04/2019	1
Protocol / Proposal	9_Consultee PIL	08/04/2019	1
Consent form	1_PIS Case	08/04/2019	1
Consent form	2_PIS Control 1	08/04/2019	1
Consent form	3_PIS Control 2	08/04/2019	1
Consent form	9_Consultee PIL	08/04/2019	1
Consent form	4_PIS Control 3	08/04/2019	1
Other	view	08/04/2019	Lisa certificate
Protocol / Proposal	STK Trust Appr_11.12.13	08/04/2019	1
Protocol / Proposal	WNN Trust Approval_08.05.14	08/04/2019	1
Protocol / Proposal	WTH Trust Appr_04.11.09	08/04/2019	1

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the Committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.

At the end of the study, the CI or delegate must notify the committee using an End of Study form.

D.

Table D1. Results of the *C. difficile* specific pan-protein array pilot study. IgG responses of patients with severe (CDP) or mild CDI (CDN) CDI, patients with diarrhoea but presumed negative for diarrhoea (CDA, also referred to as CDU), healthy controls (GCRC).

ID	Avg CDA plasma baseline	Avg CDP plasma baseline	Avg CDN + CDP plasma baseline	Avg CDN + CDP all	P CDA vs CDN+CDP all	P CDA vs CDN + CDP plasma baseline	P CDA vs CDP plasma baseline	P GCRC vs CDN + CDP all
CDR20291_0227	6144.904	7236.992	6444.779	8808.842	0.134563	0.8774	0.65397	1.37 e-09
CDR20291_0253	2929.519	2061.185	2496.648	3776.027	0.628835	0.810764	0.63782	0.000107
CDR20291_0297-s5	7514.212	5416.108	5630.54	7427.788	0.976856	0.551462	0.506568	6.83 e-06
CDR20291_0330	458.9038	1382.569	2772.714	3655.602	4.31 e-06	0.016092	0.271572	0.22 e-10
CDR20291_0337	2076.827	2530.531	2399.714	3292.967	0.265747	0.836679	0.850785	3.75 e-06
CDR20291_0342	659.9808	2764.108	3076.801	3434.668	3.53 e-07	0.014683	0.024184	0.953175
CDR20291_0351	1633.673	2923.377	2898.214	3508.464	0.065406	0.413163	0.580447	5.51 e-06
CD0386-s3	5907.519	2197.108	4876.366	6174.393	0.913644	0.694244	0.15155	0.003545
CDM120_0411	6452.135	6202.415	6078.844	6051.099	0.697008	0.737936	0.8369999	0.003055
CDR20291_0424	4223.288	1355.338	2288.17	3073.255	0.326864	0.12365	0.021706	4.25 e-08
CDR20291_0516	2917.673	1994.031	2409.388	3376.512	0.694831	0.687362	0.520863	1.8 e-05
CDR20291_0561	3170.212	1868.492	2552.279	3181.422	0.991648	0.597553	0.270667	0.000121
CDR20291_0575	3173.596	3245.031	3541.214	4594.399	0.539788	0.915216	0.817912	1.82 e-06
CDR20291_0582-s5	3259.058	1317.415	2338.018	3520.512	0.830152	0.49249	0.157111	0.007659
CDR20291_0584-s2	3080.673	1362.454	2117.083	3148.56	0.952295	0.427461	0.17608	0.001936
CDR20291_0584-s6	4192.904	1865.608	2741.583	3513.89	0.468682	0.159364	0.33205	0.045754
CDR20291_0671	7803.442	5364.492	5843.779	7404.087	0.880914	0.49509	0.423467	1.95 e-05
CDR20291_0971	2889.75	152.9923	1573.257	3117.86	0.844591	0.301737	0.22826	0.635679
CDR20291_1023	2831.981	5391.492	4488.931	4886.165	0.116289	0.3752	0.357941	0.00016
CDR20291_1276	4511.135	2272.8	2184.54	3101.213	0.586.213	0.381.1789	0.429434	5.75 e-05
CDR20291_1282	1863.058	1239.377	2103.714	3128.021	0.263601	0.260274	0.608231	0.007803
CDR20291_1360	771.9038	2335.954	2186.779	3610.375	0.01167	0.318176	0.518304	0.020121
CDR20291_1379	3532.904	2166.762	2574.844	3345.399	0.836152	0.349673	0.252066	1.31 e-07
CDR20291_1383	3075.212	1249.108	1966.757	3479.704	0.750299	0.401788	0.187119	1.11 e-09
CDR20291_1391	7328.596	4229.838	4721.105	6269.56	0.758529	0.47506	0.444531	3.05 e-07
CDR20291_1409	3447.365	3263.069	3800.974	4776.315	0.145891	0.729741	0.877919	0.12296
CDR20291_1480	3543.288	1820.723	2121.192	3241.225	0.80299	0.27311	0.215	5.21 e-08
CDR20291_1491	7783.135	4243.877	5230.214	7022.201	0.850555	0.546081	0.434268	0.01379
CDR20291_1529	3914.442	1434.223	2123.931	3673.027	0.853136	0.191493	0.080964	1.61 e-06
CDR20291_1569	43011.06	40223.95	41127.87	42080.46	0.636248	0.408221	0.273375	2.00 e-13
CDR20291_1606	4439.288	2406.838	2895.561	3543.28	0.397861	0.10936	0.109474	1.23 e-11
CDR20291_1795-s2	1838.904	1016.915	1817.474	3141.896	0.087277	0.98245	0.370239	0.00483
CDR20291_1795-s6	5702.058	2109.646	2619.279	4068.674	0.398846	0.134849	0.088068	4.17 e-05
CDR20291_1802	3245.058	1508.492	2361.388	3369.626	0.934291	0.5739	0.281065	2.43 e-10
CDR20291_1825	3255.288	4164.377	3484.431	4587.728	0.175933	0.880246	0.71088	0.000144
CDR20291_1911	3982.058	1915.454	2622.279	3232.003	0.274367	0.077412	0.015258	0.015406
CDR20291_1931	3499.827	3099.8	2953.757	3539.728	0.980218	0.788819	0.88748	6.62 e-07
CDR20291_1985	6583.904	4359.915	1964.322	3656.74	0.14877	0.032321	0.016669	0.000684
CDR20291_2034	4106.596	2666.569	3491.388	4562.985	0.765295	0.715413	0.383649	0.002388
CDR20291_2174-s2	3335.212	1987.3	3050.779	4360.165	0.401495	0.833523	0.32286	1.21 e-05
CDR20291_2178	4605.904	2574.069	3280.061	3378.979	0.452697	0.283085	0.140074	1.14 e-10
CDR20291_2226	7194.519	2036.3	2512.301	4198.506	0.169586	0.046612	0.026392	0.212682
CD2293	2598.904	4675.685	3889.214	4586.692	0.067664	0.346958	0.313845	0.001408
CDR20291_2298	3888.981	1306.954	2407.54	3471.901	0.855278	0.531547	0.273618	0.002115
CDR20291_2346	3103.058	2668.762	3238.061	4003.345	0.471989	0.922956	0.796855	1.36 e-10
CDR20291_2345	4741.288	2634.377	2728.974	3419.087	0.429258	0.251236	0.258885	2.97 e-05
CDR20291_2359	4090.288	1968.8	2709.931	3439.069	0.542547	0.235962	0.086836	0.005127
CDR20291_2503	20811.13	19762.15	20058.26	20183.22	0.600999	0.571591	0.468663	1.1 e-05
CDR20291_2559	5183.904	7958.492	6889.17	7913.967	0.178063	0.477314	0.375692	0.000797
CDR20291_2653	3186.212	1511.108	1746.822	3330.105	0.920003	0.321351	0.265558	0.000113
CDR20291_2660	2751.288	1562.415	2300.474	3185.973	0.6830392	0.704769	0.33489	7.44 e-09
CDR20291_2671	8171.135	1630.492	4392.627	5547.075	1.571289	0.431691	0.170442	0.000565
CDR20291_2686-s2	5702.827	2572.069	3650.996	5261.842	0.888845	0.53332	0.343185	5.83 e-08
CDR20291_2687	4635.519	1652.992	2284.431	3294.105	0.364527	0.130176	0.066246	0.023405
CDR20291_2700	4204.058	2040.185	2499.692	3081.961	0.295762	0.139928	0.085553	00000166

CDR20291_2744	1472.981	2868.992	3654.344	4075.393	0.010575	0.198603	0.57161	7.69 e-07
CDR20291_2805	3266.904	1532.992	2705.496	4153.47	0.610615	0.758221	0.331671	1.85 e-11
CDR20291_2873	4931.135	2154.146	2755.888	3221.369	0.179694	0.104482	0.048627	0.427101
CDR20291_2903	2773.981	2059.915	2341.344	3068.434	0.782354	0.711357	0.59073	3.05 e-09
CDR20291_2971	3595.442	2070.262	2644.409	3504.548	0.937897	0.451552	0.274288	0.000286
CDR20291_3061	3625.058	1781.685	3191.127	4477.345	0.599553	0.806212	0.276782	2.85 e-05
CDR20291_3081	3296.365	2960.223	2221.627	3288.285	0.997177	0.644664	0.890525	5.31 e-05
CDR20291_3279	2315.365	3092.262	2644.148	4125.728	0.079993	0.782977	0.6503	6.96 e-09
CDR20291_3280	5422.673	2042.608	2205.496	3188.111	0.187998	0.069511	0.065385	2.4 e-05
CDR20291_3328	4986.288	3789.3	3462.409	4095.003	0.6715	0.516457	0.634058	5.61 e-09
CDR20291_3334	4317.288	1884.223	2604.17	3440.213	0.375513	0.119303	0.032547	0.00015
CDR20291_3335	3464.519	1915.877	2444.235	3220.632	0.850218	0.457725	0.298	1.16 e-06
CDR20291_3343	1416.75	1385.108	2500.801	3287.446	0.006311	0.195467	0.964582	0.010197
CDR20291_3468	2586.596	2496.415	2554.083	3881.039	0.315473	0.793494	0.80625	0.000415
CDR20291_3498	4548.442	2225.877	2958.996	3973.919	0.723494	0.351677	0.192918	1.55 e-08
CDR20291_3504	3814.365	2317.415	2595.974	3219.644	0.637533	0.373492	0.305057	1.39 e-05

E.

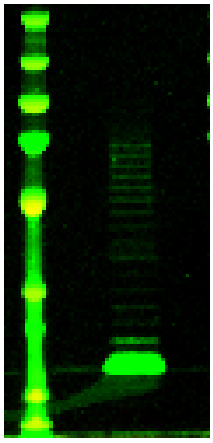
Alignment of CDR20291_0342 amino acid sequence from *C. difficile* strains 630 and R20291. Amino acid sequence of CDR20291_0342 from R20291 was aligned against CDR20291_0342 from strain 630 using pBLAST.

Score	Expect	Method	Identities	Positives	Gaps
1352 bits(3499)	0.0	Compositional matrix adjust.	685/861(80%)	768/861(89%)	3/861(0%)

```
Query  1    MKDNLKIALRYIISYKARSLAIALSIILATSLIVGIGTLSRSAQQAEVDKLKRETGSDHV  60
            MKDNLKI LRYI SYKARSLAI +SIILATSL+VGIGTLSRSAQQA+VD+LKRE G+DHV
Sbjct  1    MKDNLKIVLRYIKSYKARSLAIVVSIILATSLVVGIGTLSRSAQQADVDRLKRELGTDHV  60

Query  61   YFKDINKNQLEYIKSRKDIKNLGITSHYGYTDPNERLAINLEYANKNYLTNQSKLIKGHL  120
            YFK+I KNQL+ IKS KDIKNLGITS+Y  TD NE+L IN+ YANKNYLT+ SKLIKG
Sbjct  61   YFKNIKKNQLKSIKSSKDIKNLGITSYYASTDVNEKLPINILYANKNYLTSHSKLIKGRF  120

Query  121  PKASNEVVVEKWILNSLGLKPEINQNITFKLYQKEKPETFKVVGILEDRPIEKNKGTCEI  180
            PKASNEVVVEKWILNS+GLKPEINQN+TFKLYQKE PETFKVVGILEDR IEK  G CEI
Sbjct  121  PKASNEVVVEKWILNSMGLKPEINQNLTFKLYQKENPETFKVVGILEDRFIEKKMGQCEI  180

Query  181  FLNLNESKLDKFSYAYIEFNNGIDINTSIDNIVKNTMLDENSVGKNKMLIESTRENGTLD  240
            F NLNESKL  F+  YIEFN G DINT+IDNI+KN MLDE S+G+N MLIEST +NGTLD
Sbjct  181  FSNLNESKLSDFT-TYIEFNEGSDINTNIDNIIKNAMLDEKSIGRNSMLIESTMKNGTLD  239

Query  241  NSSKYTAITMSLFSGIVIYSIYVISIYQRIQEYGILRAIGSTNFKIFKFMFYELFILALI  300
            NSSKY AI MSLF+GIVIYSIYVISIYQR+QEYG+LRAIGSTNF+IFK M YELFILALI
Sbjct  240  NSSKYVAIAMSLFAGIVIYSIYVISIYQRVQEYGVLRAIGSTNFRIFKLMLYELFILALI  299

Query  301  AIPIGICTGIGGAQIFNRTAGNIQFEGNVTVTPFVIPDKIILLSIGSIILTILIISFFTY  360
            A+PIGIC G+ GAQIFNR+ GNIQFEGN+ VTPFVIPDKIILLSI   IL ILIISFFTY
Sbjct  300  AMPIGICIGMVGAQIFNRSVGNIQFEGNINVTPFVIPDKIILLSIACTILIILIISFFTY  359

Query  361  LKIRRISPIDSIRKTFGTDKNINKVNSLISKLTLNISVTKSISAKNIFRNKKGFIIIILS  420
            LKIRRISPID+IRKTFG DKN+ KVN LISKLTLN+SVTKSISAKNIFRNKKGF+II+LS
Sbjct  360  LKIRRISPIDAIRKTFGADKNVTKVNYLISKLTLNMSVTKSISAKNIFRNKKGFMIIVLS  419

Query  421  MSLGGIMIIKENYKYSFSDIQNKNGQEETYMNADFILTNFFLKLNQSNKADSFKDIKGLN  480
            MS+GGI++IK+NYKYSF D+QNK+ QE+TYMNADFIL+N +L  N++    SFKDIKG +
Sbjct  420  MSIGGIIVIKDNYKYSFLDVQNKSVQEQTYMNADFILSNVYL--NETGNNTSFKDIKGFS  477

Query  481  DNQIDKIKNINGIDKVKTASILDTRIEVDKLNGLDYYNIINSTPYYKDFPLFIKNKNTGK  540
            D+QIDKIKNI+GIDKVKTASIL+TRIEV+K+N LDYY  INS PYYK  PLFIKNK TGK
Sbjct  478  DSQIDKIKNIDGIDKVKTASILNTRIEVNKINQLDYYEAINSKPYYKLNPLFIKNKTTGK  537

Query  541  YTMKQKLRGYNDEMINSLEKYLVSGSINLERMKKENLAILYVPQVSKTNKYKLSYTPGTE  600
            YT+KQKL+GYNDEMI SLEKYLVSG+INL++MKKENLAI+YVP++SKT KY  ++TPG
Sbjct  538  YTLKQKLKGYNDEMIKSLEKYLVSGNINLKKMKKENLAIVYVPEISKTQKYTKTFTPGEG  597

Query  601  TPAVDIKVGDTIKVKYPKGEIDQDLYIKLKDNYEYLEYEFKVGAIVSYPFADNYLYSGDQ  660
            TPAVDIKVGDTIKVKYPKG +D  LY  LKDNYEYLEYEFKVGAI+SY FADN +YSGD
Sbjct  598  TPAVDIKVGDTIKVKYPKGGMDTGLYESLKDNYEYLEYEFKVGAIISYTFADNGMYSGDD  657

Query  661  GIDVITSNDYLKKLTGVDKYNVVYVDINKNANVKKINTLLGEIGSESPGTTTANMLMEKE  720
            G+DVITS++YLKKLTGVDKYN++Y D+N+NAN KKIN LLG+IGSE PGTTT +M++EKE
Sbjct  658  GVDVITSSEYLKKLTGVDKYNLIYADLNENANKKKINNLLGKIGSEVPGTTTVDMMVEKE  717

Query  721  NFDKMTARALTYAYGIVAVMFIISVFNIINNISYNLTSRTSEFGMLRAIGISERGFKNMI  780
            N DKMT+R + Y+YGIVA+MFIISV NIINNISYNLTSRTSEFGMLRAIGISER FKNMI
Sbjct  718  NSDKMTSRTMIYSYGIVAIMFIISVLNIINNISYNLTSRTSEFGMLRAIGISEREFKNMI  777

Query  781  LYEGILYGILSSVITIVVGLIIQFRMYYTYNFVSYGLGFSIDYKIYILVVLANIIVGILA  840
            LYEGILYG+LSS+ITIV GLIIQ +MYY   F+S+GLGFSIDYKIYILVV+ANIIVGILA
Sbjct  778  LYEGILYGVLSSIITIVSGLIIQLKMYYMQGFISHGLGFSIDYKIYILVVVANIIVGILA  837

Query  841  TYIPLRKINKISIVEAINITE  861
            TYIP RKINKISIVEAINITE
Sbjct  838  TYIPSRKINKISIVEAINITE  858
```

F.

Repeat glycosylation test of transfer of the *E. coli* O13-antigen onto AcrAtag. *E. coli* O13 cells harbouring pEC415_AcrAtag and pEXT21_pglB were induced at 37°C overnight with 1 mM IPTG and 0.2% L-arabinose. Cells were lysed and protein was nickel affinity purified before resolving on an SDS-PAGE gel then probing on an anti-His immunoblot. (1) Uninduced, (2) Induced with L-arabinose and IPTG (3) Induced (inactivated copy of PglB).



L 1 2 3

G.

Copyright

All images used in this thesis which derive from other works have been cited and permissions granted as required by the source material. Details of copyright relating to each figure can be found below.

Figure 1.1

From: Stabler, R.A., et al., Macro and micro diversity of Clostridium difficile isolates from diverse sources and geographical locations. PLoS One, 2012. **7**(3): p. e31559.

PLOS applies the Creative Commons Attribution (CC BY) license to articles and other works we publish. If you submit your paper for publication by PLOS, you agree to have the CC BY license applied to your work. Under this Open Access license, you as the author agree that anyone can reuse your article in whole or part for any purpose, for free, even for commercial purposes. Anyone may copy, distribute, or reuse the content as long as the author and original source are properly cited. This facilitates freedom in re-use and also ensures that PLOS content can be mined without barriers for the needs of research.

Figure 1.2

From: Faulds-Pain, A., et al., The post-translational modification of the Clostridium difficile flagellin affects motility, cell surface properties and virulence. Molecular Microbiology, 2014. **94**(2): p. 272-289.

This article is available under the terms of the Creative Commons Attribution License (CC BY) (which may be updated from time to time) and permits use, distribution and reproduction in any medium, provided that the Contribution is properly cited. Permission is not required for this type of reuse.

Figure 1.3a

From: Piepenbrink, K.H., et al., Structure of Clostridium difficile PilJ Exhibits Unprecedented Divergence from Known Type IV Pilins. The Journal of Biological Chemistry, 2014. 289(7): p. 4334-4345.

Parties other than the authors seeking to reuse JBC content for non commercial purposes are welcome to copy, distribute, transmit and adapt the work at no cost and without permission as long as they attribute the work to the original source using the citation above. Examples of noncommercial use include:

- Reproducing a figure for educational purposes, such as schoolwork or lecture presentations.
- Appending a reprinted article to a Ph.D. dissertation.

Figure 1.3b

From: Purcell, E.B., et al., Regulation of Type IV Pili Contributes to Surface Behaviors of Historical and Epidemic Strains of Clostridium difficile. J Bacteriol, 2015. 198(3): p. 565-77.

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.

Figure 1.4

From: Kay, E., J. Cuccui, and B.W. Wren, Recent advances in the production of recombinant glycoconjugate vaccines. NPJ Vaccines, 2019. **4**: p. 16.

Figure 1.5

From: Nothaft, H. and C.M. Szymanski, Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol, 2010. 8(11): p. 765-78

Portions	Figure 1a and 4a
Requestor Location	Ms. Catherine Hall Room 180 London School of Hygiene and Tropical Me London, WC1E7HT United Kingdom Attn: Ms. Catherine Hall

Figure 1.6

From: Cuccui, J., et al., Exploitation of bacterial N-linked glycosylation to develop a novel recombinant glycoconjugate vaccine against Francisella tularensis. Open Biol, 2013. 3(5): p. 130002.

Figure 2.1

From: Faulds-Pain, A. and B.W. Wren, Improved bacterial mutagenesis by high-frequency allele exchange, demonstrated in Clostridium difficile and Streptococcus suis. Appl Environ Microbiol, 2013. 79(15): p. 4768-71.

Figure 3.3

From: Dümmler, A., A.-M. Lawrence, and A. de Marco, Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. Microbial Cell Factories, 2005. 4: p. 34-34.

Figure 4.1

From: Rodionov, D.A., et al., Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. J Bacteriol, 2006. 188(1): p. 317-27.

Figure 4.7

From: Nawrocki, K.L., et al., Ethanolamine is a valuable nutrient source that impacts Clostridium difficile pathogenesis. Environ Microbiol, 2018. 20(4): p. 1419-1435.

your thesis/dissertation commercially. You should also duplicate the copyright notice that appears in the Wiley publication in your use of the Material. Permission is granted solely for use in conjunction with the thesis, and the material may not be posted online separately. Any third-party material is expressly excluded from this permission. If any material appears within the article with credit to another source, authorisation from that source must be obtained.

Figure 4.13

From: Kaval, K.G. and D.A. Garsin, Ethanolamine Utilization in Bacteria. mBio, 2018. **9**(1): p. e00066-18.

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. You are not required to obtain permission to reuse this article.

Figure 5.1a

From: Bordeleau, E., et al., Cyclic di-GMP riboswitch-regulated type IV pili contribute to aggregation of Clostridium difficile. J Bacteriol, 2015. 197(5): p. 819-32.

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.

Figure 5.1b

From: Maldarelli, G.A., et al., Identification, Immunogenicity and Crossreactivity of Type IV Pilin and Pilin-like Proteins from Clostridium difficile. Pathogens and disease, 2014. 71(3): p. 302-314.

Jan 13, 2020

This Agreement between Ms. Catherine Hall ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4746991504076
License date	Jan 13, 2020
Licensed content publisher	Oxford University Press
Licensed content publication	Pathogens and Disease
Licensed content title	Identification, immunogenicity, and cross-reactivity of type IV pilin and pilin-like proteins from *Clostridium difficile*
Licensed content author	Maldarelli, Grace A.; De Masi, Leon
Licensed content date	Aug 1, 2014
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Investigating novel vaccine candidates for Clostridium difficile
Publisher of your work	London School of Hygiene and Tropical Medicine
Expected publication date	Feb 2020
Permissions cost	0.00 USD

Value added tax	0.00 USD
Total	0.00 USD
Title	Investigating novel vaccine candidates for Clostridium difficile
Institution name	London School of Hygiene and Tropical Medicine
Expected presentation date	Feb 2020
Portions	Figure 2
Requestor Location	Ms. Catherine Hall Room 180 London School of Hygiene and Tropical Me London, WC1E7HT United Kingdom Attn: Ms. Catherine Hall
Publisher Tax ID	GB125506730
Total	0.00 USD

Figure 7.2a

From: Faulds-Pain, A., et al., The post-translational modification of the Clostridium difficile flagellin affects motility, cell surface properties and virulence. Molecular Microbiology, 2014. 94(2): p. 272-289.

Figure 7.2b

From: Bouche, L., et al., The Type B Flagellin of Hypervirulent Clostridium difficile Is Modified with Novel Sulfonated Peptidylamido-glycans. J Biol Chem, 2016. 291(49): p. 25439-25449.

Figure 7.4

From: Hong, Y. and P.R. Reeves, Diversity of O-Antigen Repeat Unit Structures Can Account for the Substantial Sequence Variation of Wzx Translocases. Journal of Bacteriology, 2014. 196(9): p. 1713.

Figure 7.7

From: Dykxhoorn, D.M., R. St Pierre, and T. Linn, *A set of compatible tac promoter expression vectors.* Gene, 1996. **177**(1-2): p. 133-6.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Mar 10, 2020

This Agreement between Ms. Catherine Hall ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4785321318029
License date	Mar 10, 2020
Licensed Content Publisher	Elsevier
Licensed Content Publication	Gene
Licensed Content Title	A set of compatible tac promoter expression vectors
Licensed Content Author	Derek M. Dykxhoorn,Rebecca St. Pierre,Thomas Linn
Licensed Content Date	Jan 1, 1996
Licensed Content Volume	177
Licensed Content Issue	1-2
Licensed Content Pages	4
Start Page	133
End Page	136
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	Investigating novel vaccine candidates for Clostridium difficile
Institution name	London School of Hygiene and Tropical Medicine
Expected presentation date	Mar 2020
Portions	Fig 1
Requestor Location	Ms. Catherine Hall Room 180 London School of Hygiene and Tropical Me London, WC1E7HT United Kingdom Attn: Ms. Catherine Hall
Publisher Tax ID	GB 494 6272 12
Total	0.00 GBP
Terms and Conditions	

Figure 7.19a

From: Bouche, L., et al., The Type B Flagellin of Hypervirulent Clostridium difficile Is Modified with Novel Sulfonated Peptidylamido-glycans. J Biol Chem, 2016. **291**(49): p. 25439-25449.

Figure 7.19b and c

From: Perepelov, A.V., et al., Shigella flexneri O-antigens revisited: final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity. FEMS Immunol Med Microbiol, 2012. 66(2): p. 201-10.

Permissions cost

0.00 USD

Value added tax

0.00 USD

Total

0.00 USD

Title

Investigating novel vaccine candidates for Clostridium difficile

Institution name

London School of Hygiene and Tropical Medicine

Expected presentation date

Feb 2020

Portions

Table 1

Requestor Location

Ms. Catherine Hall

Room 180

London School of Hygiene and Tropical Me

Figure 7.19d

From: Liu, B., et al., Structure and genetics of Shigella O antigens. FEMS Microbiol Rev, 2008. 32(4): p. 627-53.

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS
Jan 13, 2020

This Agreement between Ms. Catherine Hall ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4747010943603
License date	Jan 13, 2020
Licensed content publisher	Oxford University Press
Licensed content publication	FEMS Microbiology Reviews
Licensed content title	Structure and genetics of *Shigella* O antigens
Licensed content author	Liu, Bin; Knirel, Yuriy A.

Licensed content date	Jul 1, 2008
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Investigating novel vaccine candidates for Clostridium difficile
Publisher of your work	London School of Hygiene and Tropical Medicine
Expected publication date	Feb 2020
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	Investigating novel vaccine candidates for Clostridium difficile
Institution name	London School of Hygiene and Tropical Medicine
Expected presentation date	Feb 2020
Portions	Figure 4
Requestor Location	Ms. Catherine Hall Room 180 London School of Hygiene and Tropical Me London, WC1E7HT United Kingdom Attn: Ms. Catherine Hall
Publisher Tax ID	GB125506730
Total	0.00 USD
Terms and Conditions	

Appendix references

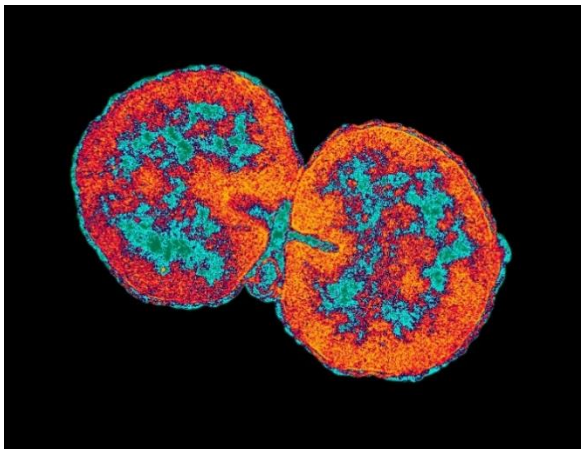
1. Dümmler, A., A.-M. Lawrence, and A. de Marco, *Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors.* Microbial Cell Factories, 2005. **4**: p. 34-34.
2. Heap, J.T., et al., *A modular system for Clostridium shuttle plasmids.* J Microbiol Methods, 2009. **78**(1): p. 79-85.
3. Dykxhoorn, D.M., R. St Pierre, and T. Linn, *A set of compatible tac promoter expression vectors.* Gene, 1996. **177**(1-2): p. 133-6.
4. Ihssen, J., et al., *Production of glycoprotein vaccines in Escherichia coli.* Microb Cell Fact, 2010. **9**: p. 61.
5. Linton, D., et al., *Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway.* Mol Microbiol, 2005. **55**(6): p. 1695-703.
6. Schulz, H., H. Hennecke, and L. Thony-Meyer, *Prototype of a heme chaperone essential for cytochrome c maturation.* Science, 1998. **281**(5380): p. 1197-200.
7. Peltier, J., et al., *Cyclic diGMP regulates production of sortase substrates of Clostridium difficile and their surface exposure through ZmpI protease-mediated cleavage.* J Biol Chem, 2015. **290**(40): p. 24453-69.

HOUSES OF PARLIAMENT
PARLIAMENTARY OFFICE OF SCIENCE & TECHNOLOGY

POSTNOTE

Number 581 July 2018

Antimicrobial Resistance and Immunisation



Antimicrobial resistance (AMR) is an urgent global health threat that, if left unchecked, could account for an estimated 10 million deaths per year by 2050.[1] Immunisation is one potential way of reducing AMR. This POSTnote describes the role for immunisation in tackling AMR, both globally and in the UK, the use of existing vaccines and how vaccine development aligns with public health priorities.

Overview

- Antimicrobial resistance (AMR) has reached a point where some infections may become untreatable.
- Immunisation is one strategy to tackle AMR, by decreasing rates of infection and thereby antibiotic use and preventing the development of resistant infections.
- The World Health Organization has developed a list of pathogens where AMR is of most concern and new antibiotics are needed; there is no equivalent for vaccines.
- Quantifying the impact of immunisation on AMR and incorporating this into calculating the cost-effectiveness of vaccines is still an area of ongoing research.
- Using immunisation to tackle AMR depends on wider use and increased uptake of existing vaccines, and increasing the development of new ones.

Background

Antimicrobial drugs kill or inhibit the growth of microbes (for example, bacteria, viruses and fungi). Microbes adapt by developing or acquiring traits that make them resistant to these drugs. This process is accelerated by widespread (including inappropriate) use of antimicrobials in humans and animals.[2,3] When resistance occurs in microbes capable of causing disease (pathogens), treatment options become limited. There are ~700,000 deaths globally every year from drug-resistant infections.[4] AMR poses a significant health and economic burden, and is a priority for the UN General Assembly.[5] Plans to tackle AMR recognise the need for multiple approaches, including immunisation and improved sanitation, but the focus has tended to be on more careful use of existing drugs and the development of new ones.[4-8] In 2014, the UK Government commissioned the (O'Neill) Review on AMR, to assess the global burden of resistance and make recommendations. It concluded that vaccines have been overlooked as a tool to reduce AMR and should be an investment focus.[9]

Immunisation as an AMR intervention

Immunisation confers protection from infection by introducing a non-harmful form or component of the pathogen in a vaccine. The body develops an immune response (such as antibodies) without disease. For many vaccines, high uptake in a population generates herd immunity. Non-immunised people are indirectly protected by being surrounded by immunised people who do not transmit the infection.[10] Some bacteria are naturally carried in or on the body, and often beneficially, but can be transmitted to and lead to disease in susceptible people.[11] Immunisation can prevent carriage, thereby reducing transmission to non-immunised people.[12] Mass immunisation programmes save ~2.5 million lives a year, globally.[13]

Immunisation can reduce the AMR burden through two main mechanisms. Firstly, it prevents infections (including resistant ones), disease and deaths, and negates the need for expensive, more complex drugs to treat resistant infections.[14,15] Secondly, preventing infection avoids the need for treatment, so antimicrobial use is reduced, in humans and animals.[16-18] One study estimated increasing uptake of the pneumococcal conjugate vaccine (PCV) could reduce antibiotic (antimicrobials for bacteria) use for pneumococcal pneumonia in children aged under 5 years by 47%.[19] Antibiotic use is linked to the development of AMR; as use decreases, so too the pressure for resistance to develop in the pathogen.[2,20,21]

Vaccines offer long-term protection from infection in contrast to antibiotics, and many vaccines still effective today were

introduced decades ago.[22] However, if a pathogen undergoes high rates of mutation, the vaccine will need to be changed, such as for influenza.[23] Furthermore, if a vaccine only covers some strains of a pathogen, infections caused by other strains can occur or increase.[24,25]

Priority Infections

The World Health Organization (WHO) and the US Centers for Disease Control and Prevention (CDC) have priority lists of bacterial pathogens of most concern, due to risk of resistance (Box 1).[26,27] The O'Neill Review identified key contexts where immunisation could reduce AMR. This includes vaccinating against: infections acquired in hospital or the community; viral infections, for which antibiotics are ineffective but may be prescribed for symptoms of, or to prevent (secondary) bacterial infections.[9]

Box 1. Antibiotic Resistant Priority Pathogens

There is no specific list of the most important resistant pathogens for the UK but lists have been developed by CDC and WHO, focusing on bacteria. Many pathogens are on both lists and categorised by threat level; the most urgent/critical threats are described here. In 2013, CDC identified the most urgent threats for the US as *Clostridium difficile*, carbapenem-resistant *Enterobacteriaceae* and *Neisseria gonorrhoeae*.[27] In 2017, WHO identified pathogens for which research and development of new antibiotics was most needed (*Mycobacterium tuberculosis* has a dedicated programme). Critical threats include:[26]

- *Acinetobacter baumannii* that are resistant to carbapenem drugs
- *Pseudomonas aeruginosa* that are carbapenem-resistant
- *Enterobacteriaceae* that are carbapenem-resistant and/or resistant to 3rd generation cephalosporin drugs.
- Carbapenem resistance is an issue as this is a "last resort" class of antibiotic, used for treatment when other options have failed.[28] There are no licensed vaccines for any of the pathogens listed above, although some are currently in development.[29]

The Global Context

The burden of disease and AMR varies considerably across the world. Many serious infections disproportionally affect low- and middle-income countries (LMICs).[30-32] HIV, tuberculosis and malaria are the "big three" infections, so-called because of the health burden they pose. Resistance is an issue for treatment of all three.[33-36] Vaccine development for each is ongoing; there is no HIV vaccine and the major vaccine in development for malaria and the only available vaccine for tuberculosis provide a sub-optimal level of protection.[29,37,38]

There are a number of infections where new vaccines or increased uptake of existing ones could reduce mortality and/or antibiotic use, such as those caused by group A streptococci, pneumococci, influenza and respiratory syncytial virus (RSV).[19,39-43] In 2016, there was an outbreak of typhoid fever in Pakistan, resistant to multiple antibiotics.[44] A new typhoid vaccine has been recommended, with higher efficacy than previous vaccines.[45-47]

The UK Context

Many infections that present a global problem are also an issue in the UK. Vaccines could be a useful tool in managing infections where antibiotic treatment is undermined by resistance, such as gonorrhoea (Box 2).

Cases of gonorrhoea have increased since 2008 and many antibiotics are no longer recommended for routine use due to the emergence of resistance. A recent case of gonorrhoea in the UK was resistant to both current primary recommended treatments.[51,48-50]

Box 2. Gonorrhoea

Gonorrhoea is a common bacterial sexually transmitted infection and can result in infertility if left untreated. Antibiotic treatment options are limited by high levels of AMR.[51,52] There is no vaccine currently available.[29] Data from New Zealand show that a vaccine against meningitis B had an estimated 31% efficacy against gonorrhoea. The bacteria causing meningitis B and gonorrhoea are related. This has led to suggestions of optimising the next generation of this vaccine to prevent gonorrhoea.[53] In 2015, the UK became the first country to offer routine meningitis B immunisation, using a vaccine which shares a component with the vaccine used in New Zealand.[54]

Healthcare-associated infections (HCAIs)

HCAIs incur significant costs to the NHS and can result in severe outcomes, such as blood-stream infections (BSIs).[55-57] *Clostridium difficile* and methicillin-resistant *Staphylococcus aureus* (MRSA) infections remain a burden in healthcare settings, but have decreased substantially in recent years (due to better infection prevention and control, and reduced prescribing). Between 2007/08 and 2017/18, there was a 76.1% decrease in *C. difficile* infections and 81% decrease in MRSA BSIs.[58] Another cause of HCAIs are the "Gram-negative" class of bacteria, where resistance is significant.[55,59] Most Gram-negative BSIs are caused by *Escherichia coli*, *Klebsiella* and *Pseudomonas aeruginosa*.[55] Almost half of *E. coli*-BSIs are resistant to co-amoxiclav, a routinely used antibiotic, and resistance to carbapenem antibiotics (Box 1) is low but increasing annually.[55] The Government has a target to reduce the incidence of Gram-negative BSIs by 50% by 2021, through improved infection prevention and control.[55]

There are no licensed vaccines for any of these infections.[29] It is suggested that these vaccines could be targeted to people with an increased risk of infection, such as those with planned surgical procedures.[9,60,61] Vaccines are in development for MRSA and *C. difficile* and three *E. coli* vaccines are in early trials, but there are no candidates in trials for *P. aeruginosa* or *Klebsiella*.[29] Although Gram-negative bacteria can cause serious infections at certain sites in the body, many normally live in the gut of healthy people.[62-64] It is unclear how immunisation would impact gut bacteria and consequently overall health.[11,65]

Viral infections

Lack of rapid diagnostics for use at the time of consultation means many antibiotics are prescribed on the basis of symptoms alone, which can be caused by multiple pathogens, for example for respiratory tract infections (RTIs).[66-68] This can lead to unnecessary antibiotic use for RTIs caused by viruses that can resolve without antibiotic treatment, but which are prescribed due to the risk of bacterial infection, particularly in children and the elderly.[69-73] Vaccinating against common viral causes of RTIs, such as influenza virus, can reduce antibiotic use and acquisition of secondary bacterial infections.[17,74,75] RSV is the most

common cause of acute lower respiratory tract infections in children under 5 years of age globally and is an issue for adults with other conditions and/or lowered immune systems.[76,77] There is no licensed vaccine but there are many vaccine candidates in different stages of clinical trials.[78-80]

The influenza vaccine is offered to all over 65s, at-risk groups (including pregnant women), and more recently to children, to provide protection from infection and to decrease transmission from children to vulnerable older individuals.[81] Vaccine uptake in the UK varies between groups but is high (>70%) for over 65s.[82] Variations in the circulating strains causing infections each season can lead to mismatches, where the available vaccine does not fully target the current strain. There is ongoing research in developing a vaccine that covers all strains, to reduce influenza infections[83-86] and with further benefits of reduced antibiotic use and secondary ear infections.[87-89] Population-wide influenza vaccination in Ontario was associated with a 64% decrease in influenza-associated respiratory antibiotic prescriptions.[17] Research by Public Health England (PHE) is assessing any changes in antibiotic prescribing and secondary lower respiratory tract infections in the context of the influenza immunisation programme in England.[90]

Decisions about Vaccine Use

There are two main approaches to reducing AMR using immunisation: maximising the utility of existing vaccines by ensuring good uptake and offering them more widely as appropriate, and making decisions about which vaccines should be developed and introduced and when and where to use them most efficiently.[9] Decisions about choices for national immunisation programmes vary between countries due to vaccine availability and affordability, disease risk, vaccine efficacy and different frameworks for advice.[91,92]

Measuring the Impact of Vaccines on AMR

Vaccines may have several direct and indirect effects on AMR:[93]

- preventing disease and deaths
- reducing the progression to and severity of disease
- reducing transmission of infection between people
- reducing antibiotic use and pressure for resistance
- reducing GP visits and hospital stays.

Monitoring and quantifying the effects of immunisation on AMR and assigning economic benefit to these outcomes is challenging and requires surveillance data and complex analysis (Box 3).[7,93,94] The committee that advises the UK Government on vaccination considers AMR in its decision-making process insofar as it is able, based on the evidence (Box 4).[95]

Surveillance and Data Collection

Understanding disease burden and the proportion that is resistant is important in informing design and use of AMR interventions, including vaccines.[4,96-98] The UK has separate surveillance systems to monitor disease and AMR.[55,100] Some lower income countries, in which AMR has a disproportionate impact, do not have the capacity for detailed surveillance, which limits understanding.[101,102] Programmes to increase surveillance in these regions include The Fleming Fund and the WHO Global AMR Surveillance System.[103,104] Monitoring antibiotic use is one way to quantify the impact of immunisation on AMR but this is complex.[105-108] The Wellcome Trust has commissioned a review of studies on immunisation and antibiotic use. It is also encouraging data collection about antibiotic use during clinical trials and post-vaccine roll out studies.[109]

Box 3. Mathematical Modelling and Cost-Effectiveness Analysis

Mathematical modelling of the impact of a vaccine on a target population and cost-effectiveness evaluations are an important part of decision-making about vaccines in the UK. This is a statutory requirement before a vaccine can be recommended to Ministers.[92] Modelling of immunisation and AMR is a useful tool to increase understanding and predict impact but its utility in informing public health strategies is restricted by the complexity in modelling this relationship.[93] For example, although antibiotic use has accelerated development of AMR, this can vary depending on the drug, pathogen and host setting.[94] Furthermore, antibiotic use can also pressure the bacteria carried in the body of healthy people to develop resistance.[99] Existing studies aiming to model AMR in the context of immunisation are restricted to pneumococcal and *S. aureus* infections.[93]

Box 4. Scientific Advice to the UK Government on Immunisation

The Government amends the national immunisation schedule on the advice of the Joint Committee on Vaccination and Immunisation (JCVI). Decision making is informed by a range of evidence including disease burden, vaccine availability, safety and efficacy and cost-effectiveness. The AMR burden for an infection is considered where possible. This is limited by the complex challenges in quantifying the impact of vaccination on the AMR burden and assigning economic benefit to this impact. Researchers are developing models to predict these complex effects. The JCVI has identified two infections for which vaccines are unavailable but could be beneficial and cost-effective, RSV and Group B *Streptococcus* (GBS).[95,110] These could also be beneficial from an AMR standpoint as both infections lead to potentially avoidable antibiotic use. Antibiotics are ineffective against RSV but may still be prescribed, including for secondary bacterial infections.[41] For GBS, there are global differences in screening practices and antibiotic use.[111-113]

Utility of Existing Vaccines

The O'Neill Review recommended wider use of existing vaccines, such as those for pneumococcal infections.[9] After the pneumococcal conjugate vaccine (PCV) was introduced in the UK in 2006, the incidence of disease across all age groups caused by types ("serotypes") contained in the vaccine fell by 97% by 2016, and infections resistant to a certain class of antibiotics (macrolides) also decreased (Box 5).[25,114] Reduced prevalence of drug-resistant infections and antibiotic use through immunisation depends on achieving good vaccine uptake. There are major global disparities in uptake, as a consequence of supply and affordability, weaknesses in health systems and public attitudes.[115,116]

Box 5. Pneumococcal Conjugate Vaccine (PCV)
Streptococcus pneumoniae can cause severe disease, including meningitis, septicaemia and pneumonia, with under 5 year-olds and the elderly most at risk.[117] There are >100 identified serotypes that vary in their prevalence, drug-resistance and ability to cause disease [118,119] PCV7, a vaccine covering 7 serotypes, has been linked to reduced antibiotic use and resistant infections in some regions.[120-122] A reduction in vaccine-targeted serotypes was followed by an increase in disease caused by other serotypes, known as serotype replacement.[25] New vaccines are required to cover serotypes arising through this process, for example PCV13 is now used in many countries, including the UK.[25,117]

The Global Context
The effectiveness of national immunisation programmes relies on high vaccine uptake in the target populations.[10,123] However, many countries do not have the infrastructure or means to achieve this. For example, the WHO estimates that global coverage of PCV for children under 5 is 42%.[124] Gavi is a public-private partnership that aims to increase vaccine coverage in the world's poorest countries.[125] Gavi's portfolio contains 11 vaccines,[126] and they estimate that their provision of *Haemophilus influenzae* type b, pneumococcal and meningococcal vaccines between 2001 and 2030, could mean that 500 million doses of antibiotics would not be used. AMR is a recent addition to Gavi's vaccine investment decision criteria.[127]

The UK Context
The UK has one of the most comprehensive immunisation schedules and high uptake.[128] There are still inequalities in uptake in certain socioeconomic and ethnic groups, but PHE and NHS England are working to reduce them.[129,130] Public perceptions of vaccination also influence uptake and can lead to increases in vaccine-preventable infections.[131,132] Survey data from 2016 showed that 94% of parents in England had confidence in the immunisation programme.[133] This high level of confidence is not reflected worldwide, including some high income European countries.[134]

Developing New Vaccines

Vaccine development requires advanced technologies, is expensive, time-consuming and subject to high attrition, so companies require markets to make research and development (R&D) commercially viable.[9,135-136] There are a number of vaccines in development for pathogens on the WHO and CDC lists (Box 1), but some vaccines may be less attractive as they offer lower returns, such as those targeting diseases that mainly affect LMICs.[29,137,138] A key question is how to stimulate and prioritise the development of new vaccines in the context of AMR. The O'Neill review highlighted the need to support early research and maintain a viable market.[9] However, the majority of funding for new AMR products (excluding direct industry investment) targets development of new antibiotics.[139] Several government AMR initiatives are ongoing, including vaccine development (Box 6). Other proposed methods to encourage development include product development partnerships, market entry rewards and tax credits. [4,40,140]

Vaccine Development and Public Health Priorities

Where collections of symptoms could be caused by several different pathogens (such as RTIs), vaccinating against many of these (such as influenza and RSV) could have a greater impact on antibiotic use, by preventing infection and reducing the need for medical intervention.[77,40,141] It has been suggested that future research could focus on developing vaccines that preferentially target resistant pathogens or are against HCAIs.[7,40,61,142] As HCAIs are more likely to affect certain groups, vaccinating the whole population may not be appropriate.[9,143] However:

- unwell or elderly people may not be able to develop a protective immune response after vaccination[9,143]
- identifying the target population risks missing people[61]
- for those entering hospitals in an emergency, there may not be time to be vaccinated and generate immunity, although treatments (monoclonal antibodies) that confer immediate, short-term immunity are being developed.[145]

Public Health Priorities
Guidance on public health priorities can inform research and help manufacturers understand potential markets.[146] For example, target product profiles that outline what is expected from a vaccine can bring stakeholders together and coordinate development.[147,148] The Wellcome Trust has commissioned work to assess vaccine development for all pathogens on the WHO priority pathogen list, including:[109]

- R&D pipeline - past failures, existing efforts and need
- market analysis - sizing calculations and target population
- payers - who would pay or support access to vaccines
- barriers - commercial, clinical trials and delivery
- benefits for each pathogen and any cross-protection.

Vaccine Development
Various initiatives can support the pre-clinical development of promising vaccine candidates; CARB-X is a public-private partnership funding development of antimicrobial products for priority pathogens (Box 1).[149] The UK government funded BactiVac Network aims to accelerate development of anti-bacterial vaccines for LMICs and covers UK needs following investment from the Industry Strategy Challenge Fund.[150] The planned UK National Vaccine Development and Manufacturing Centre aims to support later stages of development.[151] Medicines regulators such as the European Medicines Agency are co-operating with other bodies to discuss vaccine regulation in the context of AMR.[152]

Box 6. Government AMR Strategies and Immunisation
The Department of Health and Social Care (DHSC) leads the 5-Year AMR Strategy which includes increasing uptake of immunisation and investment in new vaccines.[153] The £50 million Global AMR Innovation Fund - set up in response to the O'Neill review - funds R&D in underinvested areas.[154] This focuses on LMICs and includes the development of alternatives to antibiotics, such as vaccines, with investment in CARB-X.[155] Innovate UK works with GAMRIF and conducts other AMR and vaccine work.[151] The Department for International Development contributes to the UK AMR Strategy and is the biggest contributor to Gavi.[156] The AMR Funders Forum includes research councils, government and charities and coordinates UK AMR research. This includes vaccine development and work to identify which interventions will have the most impact in different settings.[157]

POST is an office of both Houses of Parliament, charged with providing independent and balanced analysis of policy issues that have a basis in science and technology. POST is grateful to Catherine Hall for researching this briefing, to the Medical Research Council for funding her parliamentary fellowship, and to all contributors and reviewers. For further information on this subject, please contact the co-author, Dr Sarah Bunn.

Endnotes (fully referenced version available online)

1 Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations, The Review on Antimicrobial Resistance, Chaired by Jim O'Neill and published by the Wellcome Trust and HM Government, 2014
2 Goossens H et al, Lancet, 365, 579-587, 2005
3 Verweij PE et al, Lancet Infect Dis, 9, 789-795, 2009
4 Tackling Drug-Resistant Infections Globally, The Review on Antimicrobial Resistance, Chaired by Jim O'Neill and published by the Wellcome Trust and HM Government, 2016
5 Political Declaration of the high-level meeting of the General Assembly on antimicrobial resistance, United Nations General Assembly, 2016
6 Global Action Plan on Antimicrobial Resistance, World Health Organisation, 2015
7 Lipsitch M & Siber GR, MBio, 7, 2016
8 https://www.nature.com/news/vaccines-promoted-as-key-to-stamping-out-drug-resistant-microbes-1.22324
9 Vaccines and alternative approaches: reducing our dependence on antimicrobials, The Review on Antimicrobial Resistance, Chaired by Jim O'Neill and published by the Wellcome Trust and HM Government, 2016
10 Heymann DL & Aylward RB, Curr Top Microbiol Immunol, 304, 2006
11 The Microbiome and Human Health, POSTnote 574, Parliamentary Office of Science and Technology, 2018
12 Trotter CL et al, Vaccine, 26, 4434-4445, 2008
13 Global Vaccine Action Plan 2011-2010, World Health Organisation
14 Antimicrobial Resistance, World Health Organisation, 2018
15 Marks SM et al, Emerg Infect Dis, 20, 812-821, 2014
16 Grijalva CG, Lancet Infect Dis, 14, 175-177, 2014
17 Kwong JC et al, Clin Infect Dis, 49, 750-756, 2009
18 Markestad A & Grave K, Fish Vaccinol, 90, 365-369, 1997
19 Laxminarayan R et al, Lancet, 387, 168-175, 2016
20 Costelloe C et al, BMJ, 340, 2010
21 Livermore DM et al, J Antimicrob Chemother, 68, 2667-2674, 2013
22 Kennedy DA & Read AF, Proc Biol Sci, 284, 2017
23 https://www.cdc.gov/flu/about/viruses/change.htm
24 Martcheva M et al, J R Soc Interface, 5, 3-13, 2008
25 Ladhani SN et al, Lancet Infect Dis, 18, 441-451, 2018
26 Global Priority List of Antibiotic-resistant bacteria to guide research, discovery and development of new antibiotics. World Health Organisation, 2017
27 https://www.cdc.gov/drugresistance/biggest_threats.html
28 Meletis MD, Ther Adv Infect Dis, 3, 15-21, 2016
29 Antimicrobial Resistance Benchmark, Access to Medicine Foundation, 2018
30 Vitoria M et al, Am J Clin Pathol, 131, 844-848, 2009
31 Preventing disease through healthy environments: a global assessment of the burden of disease from environmental risks, World Health Organisation, 2016
32 Lozano R et al, Lancet, 380, 2095-2128, 2012
33 Goldberg DE et al, Cell, 148, 1271-1283, 2012
34 HIV Drug Resistance Report 2017, World Health Organisation, 2017
35 Global Tuberculosis Report 2017, World Health Organisation, 2017
36 Global report on antimalarial efficacy and drug resistance: 2000-2010, World Health Organisation, 2010
37 Mangtani P et al, Clin Infect Dis, 58, 470-480, 2014
38 Agnandji ST et al, N Engl J Med, 367, 2284-2295, 2012
39 Zühlke LJ et al, Curr Treat Options Cardiovasc Med, 19, 15, 2017
40 The Value of Vaccines in the Avoidance of Antimicrobial Resistance, Centre on Global Health Security, Chatham House, 2017
41 van Houten CB et al, Pediatr Infect Dis, 2018
42 Global Action Plan for Influenza Vaccine Global Progress Report, World Health Organisation, 2014
43 Shi T et al, Lancet, 390, 946-958, 2018
44 Klemm EJ et al, MBio, 9, 2018
45 Summary of the October 2017 meeting of the Strategic Advisory Group on Experts on Immunisation, World Health Orgnaization, 2017
46 Jin C et al, Lancet, 390, 2472-2480, 2017
47 Typhoid vaccines: WHO position paper, March 2018 – Recommendations, World Health Organisation, 2018
48 MacFadden DR et al, Lancet Infect Dis, 18, 599, 2018
49 UK case of *Neisseria gonorrhoeae* with high-level resistance to azithromycin and resistance to ceftriaxone acquired abroad, Public Health England, 2018
50 Unemo M & Shafer WM, Clin Microbiol Rev, 27, 587-613, 2014
51 Surveillance of antimicrobial resistance in Neisseria gonorrhoeae in England and Wales. Key findings from the Gonococcal Resistance to Antimicrobials Surveillance Programme (GRASP), Public Health England, 2017
52 Unemo M & Shafer WM, Clin Microbiol Rev, 27, 587-613, 2014
53 Petousis-Harris H et al, Lancet, 390, 1603-1610, 2017
54 https://www.nhs.uk/conditions/vaccinations/meningitis-b-vaccine/
55 English Surveillance Programme for Antimicrobial Utilisation and Resistance, Public Health England, 2017
56 Preventing healthcare associated Gram-negative bloodstream infections: an improvement resource, Public Health England, 2017
57 Plowman R et al, J Hosp Infect, 47, 198-209, 2001
58 Annual epidemiological commentary: Gram-negative bacteraemia, MRSA bacteraemia, MSSA bacteraemia and C.difficile infections, up to and including financial year April 2017 to March 2018, Public Health England, 12 July 2018
59 Blair JMA et al, Nat Rev Microbiol, 13, 42-51, 2015
60 Van Kleef et al, Vaccine, 34, 5562-5570, 2016
61 McIntosh EDG, Ther Adv Vaccines, 6, 19-27, 2018
62 Jansen KU & Anderson AS, Hum Vaccin Immunother, 22, 1-17, 2018
63 Moradigaravand D et al, MBio, 8, 2017
64 Marcel JP et al, Clin Microbiol Infect, 14, 895-907, 2008
65 Hays MP et al, BMC Res Notes, 11, 401, 2016
66 Pouwels KB et al, J Antimicrob Chemother, 73, 19-26, 2018
67 Dolk FCK et al, J Antimicrob Chemother, 73, 19-26, 2018
68 Rapid Diagnostics: Stopping Unnecessary Use of Antibiotics, The Review on Antimicrobial Resistance, Chaired by Jim O'Neill and published by the Wellcome Trust and HM Government, 2015
69 Respiratory tract infections (self-limiting): prescribing antibiotics, National Institute for health and Care Excellence, 2008
70 Jain S et al, N Engl J Med, 372, 835-845, 2015
71 Pavia AT, Clin Infect Dis, 52, 284-289, 2011
72 Bruning AHL et al, Clin Infect Dis, 65, 1026-1032, 2017
73 Infection Control Precautions to Minimise Transmission of Acute Respiratory Infections in Healthcare Settings, Public Health England, 2016
74 Norhayati MN et al, Cochrane Database Syst Rev, 2017
75 Ampofo K et al, Pediatrics, 122, 229-237, 2008
76 Nair H et al, Lancet, 375, 1545-1555, 2010
77 Lee N et al, Clin Infect Dis, 57, 1069-1077, 2013
78 Simoes EAF et al, Infect Dis Ther, 7, 87-120, 2018
79 Regulating Clinical Trials, POSTnote 561, Parliamentary Office of Science and Technology, 2017
80 https://www.path.org/publications/files/CVIA_RSV_snapshot_fs.pdf
81 The national flu immunisation programme 2018/19 letter, Public Health England, 2018
82 Seasonal influenza vaccine uptake in GP patients: winter season 2017 to 2018, Public Health England, 2018
83 Zimmerman RK et al, Clin Infect Dis, 63, 1564-1573, 2016
84 Zhou F et al, Curr Opin Immunol, 29, 88-95, 2018
85 Erbleding EJ et al, J Infect Dis, 2018
86 The Lancet Infectious Diseases, Lancet Infect Dis, 18, 475, 2018
87 Kash JC & Taubenberger JK, Am J Pathol, 185, 1528-1536, 2015
88 Ozgur SK et al, Pediatr Infect Dis J, 25, 401-404, 2006
89 Marchisio P et al, Pediatr Infect Dis J, 28, 855-859, 2009
90 Personal Communication: Dr Richard Pebody, Public Health England
91 Burchett HE et al, Health Policy Plan, 27, 2012
92 Joint Committee on Vaccination and Immunisation Code of Practice, 2013
93 Atkins KE et al, Lancet Infect Dis, 2017
94 Holmes AH et al, Lancet, 387, 176-187, 2016
95 Personal Communication: Professor Andrew Pollard, Chair Joint Committee on Vaccination and Immunisation
96 Hyde TB et al, Vaccine, 31, 94-98, 2013
97 Global Framework for Immunisation Monitoring and Surveillance, World Health Organization, 2007
98 The Role of Vaccination in Reducing Antimicrobial Resistance, Vaccines Europe, 2013
99 Tedijanto C et al, bioRxiv, 288704, 2018
100 Surveillance of Infectious Diseases, POSTnote 462, Parliamentary Office of Science and Technology, 2014
101 Laxminarayan R & Heymann DL, BMJ, 344, 2012
102 Seale AC, Wellcome Open Res, 2, 2017
103 http://www.flemingfund.org/
104 http://www.who.int/glass/en/
105 Personal Communication, Professor Marc Lipsitch
106 Fireman B et al, Pediatr Infect Dis J, 22, 10-16, 2003
107 Lau WCY et al, Vaccine, 33, 5072-5079, 2015
108 Ben-Shimol S et al, Clin Infect Dis, 63, 611-618, 2016
109 Personal Communication: Dr Elizabeth Klemm and Dr Zoe Seager, The Wellcome Trust
110 Cromer D et al, Lancet Public Health, 2, 367-374, 2017
111 Hughes RG, Brocklehurst P, Steer PJ, Heath P, Stenson BM on behalf of the Royal College of Obstetricians and Gynaecologists, BJOG, 124, e280-e305, 2017
112 Martinez de Tejada B, Int J Environ Res Public Health, 11, 7993-8009, 2014

113 Le Doare K et al, Clin Infect Dis, 65, 143-151, 2017
114 Henderson KL et al, J Antimicrob Chemother, 65, 369-70, 2010
115 Access to Vaccines Index 2017, Access to Medicine Foundation, 2017
116 Larson HJ et al, Vaccine, 32, 2150-2159, 2014
117 Pneumococcal: the green book, chapter 25, Public Health England, 2018
118 Weiser JN et al, Nat Rev Microbiol, 16, 355-367, 2018
119 Song JH et al, vaccine, 30, 2728-2737, 2012
120 Palmu AA et al, Lancet Infect Dis, 14, 205-212, 2014
121 Kyaw MH, N Engl J Med, 354, 1455-1463, 2006
122 Von Gottberg A et al, N Engl J Med, 371, 1889-1899, 2014
123 NHS Public Health Functions Agreement 2015-2016 Core Service Specification National immunisation programme, Department of Health, 2014
124 Immunisation Coverage, World Health Organisation, 2018
125 https://www.gavi.org/about/gavis-partnership-model/
126 https://www.gavi.org/support/nvs/
127 https://www.gavi.org/about/strategy/vaccine-investment-strategy/
128 Childhood Vaccination Coverage Statistics, National Statistics, 2017
129 UK Trends in Infectious Diseases, POSTnote 545, Parliamentary Office of Science and Technology, 2017
130 Immunisations: reducing differences in uptake in under 19s, National Institute of health and Care Excellence, 2009 (updated 2017)
131 Dubé E et al, Humm Vaccin Immunother, 9, 1763-1773, 2013
132 Omer SB et al, N Engl J Med, 360, 1984-1988, 2015
133 Parental attitudes to childhood immunisation, Public Health England, 2016
134 Larson HJ et al, EBioMedicine, 12, 295-301, 2016
135 Rappuoli R et al, Lancet, 378, 360-368, 2011
136 Pronker ES et al, PLoS One, 8, 2013
137 Oyston P & Robinson K, J Med Microbiol, 61, 889-894, 2012
138 Making Markets for Vaccines: Ideas to Action, Center for Global Development, 2005
139 Tackling Antimicrobial Resistance, Ensuring Sustainable R&D, The Organisation for Economic Co-operation and Development (OECD), 2017
140 Incentives for New Drugs to Tackle Anti-Microbial Resistance, Office of Health Economics, 2017
141 https://www.gov.uk/government/publications/respiratory-infections-laboratory-reports-2018
142 Joice R & Lipsitch M, PLoS One, 8, 2013
143 Knisely JM et al, Clin Infect Dis, 63, 657-662, 2016
144 Dorrington MG & Bowdish DM, Front Immunol, 28, 171, 2013
145 Sparrow E et al, Bull World Health Organisation, 95, 235-237, 2017
146 Report of an expert panel discussion on: The future of vaccines: The next decade, Association of the British Pharmaceutical Industry, 2018
147 Lee BY & Burke DS, Vaccine, 28, 2806-2809, 2010
148 Follow-up Report for the German GUARD Initiative, Breaking through the Wall, A call for concerted action an antibiotics research and development, 2017
149 https://carb-x.org/
150 https://www.birmingham.ac.uk/research/activity/immunology-immunotherapy/research/bactivac/index.aspx
151 Personal Communication: Phil Packer, Innovate UK
152 Personal Communication, Dr Marco Cavaleri. European Medicines Agency
153 UK Five Year Antimicrobial Resistance Strategy 2013 to 2018, Department of Health and Department for Environment, Food and Rural Affairs, 2013
154 https://www.gov.uk/government/news/expert-advisory-board-to-support-the-global-amr-innovation-fund
155 https://www.gov.uk/government/news/30-million-of-funding-to-tackle-antimicrobial-resistance
156 https://www.gavi.org/funding/donor-contributions-pledges/
157 https://mrc.ukri.org/research/initiatives/antimicrobial-resistance/antimicrobial-resistance-funders-forum/