Characterising novel biomarkers for use in point of need sensors:

A *Clostridium difficile* case study

Beth Lawry

Submitted for the degree of Doctor of Philosophy

Newcastle University

Faculty of Medical Science

Institute of Cellular Medicine

2016

Thank you to Mum, Dad and the rest of the Lawry clan (yes Grace, that includes you) for all of their support/listening to me whinge. Also, a special thank you to my rock, Tim, for keeping me sane….ish

Thesis Abstract

Clostridium difficile infection (CDI) is the leading cause of nosocomial diarrhoea in the developed world and community-acquired infection is increasing in incidence. The current 'gold standard' for diagnosing CDI is stool culture, followed by a cell cytotoxicity neutralisation assay. However, these methods are labour intensive and time consuming, so are seldom used. More rapid tests for diagnosing CDI, for example, toxin enzyme immunoassays, are less sensitive and/or specific. Accurate and early detection, especially at the appropriate point of need (PoN), is crucial for the successful treatment of CDI and minimising transmission of the disease between patients.

In this study, a novel bioinformatics approach was used to identify protein sequence diagnostic biomarkers that are unique for bacteria within a particular group of interest (GOI). This approach incorporated a previously designed system, IDRIS, to mine all fully annotated genomes within the NCBI RefSeq database, together with a bioinformatics workflow that was designed and applied within this project. The workflow uses several bioinformatics tools to identify biomarkers that are a minimum of 15 amino acids, are unique to a GOI and are surface accessible within whole cells. Biomarkers from surface accessible proteins were selected as they would provide targets that facilitate PoN detection of whole cells, by minimising the need for time consuming pre-treatment of samples.

The most promising target identified was a 16 amino acid biomarker within the surface layer protein, SlpA. Monoclonal antibodies (mAbs) were produced against this unique biomarker and were shown to bind specifically to whole *C. difficile* cells, including strains from each of the 13 surface layer type clades, which represent the entire *C. difficile* species. The selected mAb did not bind to species closely related to *C. difficile,* highlighting the molecules promise for PoN CDI detection. The sensitive and specific binding of the developed mAb highlights the value of the bioinformatics approach, which could be used to provide similar results for other bacteria, including other pathogens.

Table of contents

List of Figures

List of tables

List of equations

Abbreviations

Ab	Antibody
Ap	Ampicillin
A.U.	Arbitrary Units
BLAST	Basic Local Alignment Search Tool
B. subtilis	*Bacillus subtilis*
BSA	Bovine serum albumin
C. difficile	*Clostridium difficile*
CV	Coefficient of variation
^{o}C	Degrees Celsius
DB	Database
ECL	Enhanced Chemiluminescence
ELISA	Enzyme-linked immunosorbent assay
FBS	Foetal bovine serum
g	Gram
x *g*	Acceleration due to gravity
gb	Gigabyte
g/l	Gram per litre
h	Hour
HCl	Hydrochloric acid
HMM	Hidden Markov model
HMW	High molecular weight
HRP	Horseradish peroxidase
HT	Haemorrhagic toxin
ID	Identification
IF	Immunofluorescence
IgG	Immunoglobulin G

IPTG	Isopropyl-β-D-thiogalactopyranoside
IR	Infra-red
ITC	Isothermal titration calorimetry
kb	Kilobase
kDa	Kilodalton
K_d	Equilibrium dissociation constant
l	Litre
LMW	Low molecular weight
LT	Lethal toxin
M	Molar
mA	Milliamp
mAbs	Monoclonal antibodies
mg	Milligram
min	Minutes
ml	Millilitre
mm	Millimeter
mM	Millimole
MST	Microscale thermophoresis
nl	Nanolitre
OD	Optical density
O/N	Overnight
OPD	o-Phenylenediamine
PAGE	Polyacrylamide-gel-electrophoresis
PDB	Protein Data Bank
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline with 0.05% Tween 20
PCR	Polymerase chain reaction
PFA	Paraformaldehyde

PoN	Point of need
R	Ribotype
RAM	Random-access memory
r.p.m	Revolutions per minute
RT	Room temperature
s	Second
SDS	Sodium dodecylsulphate
SLP	Surface layer protein
SlpA	Surface layer protein A
SLT	Surface layer type
SLTS	Surface layer type strains
SPI	Signal peptidase I
SPR	Surface plasmon resonance
TAE	Tris-acetate-EDTA buffer
TEMED	N,N,N',N'-tetramethyl-ethylenediamine
Tris	Tris(hydroxymethyl)aminoethane
µg	Microgram
µm	Micrometer
v/v	Volume for volume
w/v	Weight for volume
>	Greater than
<	Less than
≥	Greater than or equal t

Chapter 1 General introduction

1.1 Biosensors and point of need testing

Biosensors are devices that combine biological recognition molecules with physical or chemical transducers. The detection element can be qualitative, for example a colour change, or can be quantitative, for instance a digital readout (Newman and Setford, 2006).

In 1962, Leland C. Clark designed the original biosensor, when he expanded his innovative oxygen electrode to create the first glucose meter (Clark and Lyons, 1962). The basic design of this glucose sensor has been the foundation for several different enzyme sensors and in 1970, Anton Clemens produced a quantitative sensor that could be used by the patient to self-monitor blood glucose levels (Tonyushkina and Nichols, 2009). This advancement in technology, along with the discovery of insulin, has been said to be the most important in controlling diabetes mellitus (Clarke and Foster, 2012).

Another highly successful biosensor is the pregnancy test, which utilises a sandwich based, enzyme-linked immunosorbent assay (ELISA) and a lateral flow strip. A sandwich ELISA detects antigens between two layers of antibodies. One antibody layer is bound to a solid surface and the secondary antibodies are fluid and conjugated to a detection molecule, such as gold nanoparticles (Zhou *et al.*, 2015). The antigen binds to the stationary antibodies and the fluid, secondary antibodies bind and sandwich the antigen, producing a positive signal, such as a detectable colour (Figure 1-1 A). In pregnancy tests this assay is performed using lateral flow strips, porous strips that substances can travel through (Figure 1-1 B). Antibodies bound to the strip are used to detect the hormone, human chorionic gonadotropin, which is produced by fertilised ova (Peters, 2008). Conjugated secondary antibodies then bind to the hormone, producing a positive coloured line. Both the blood glucose monitor and the pregnancy test are examples of for rapid point of need, testing at or near the site of patient care, (PoN) sensors.

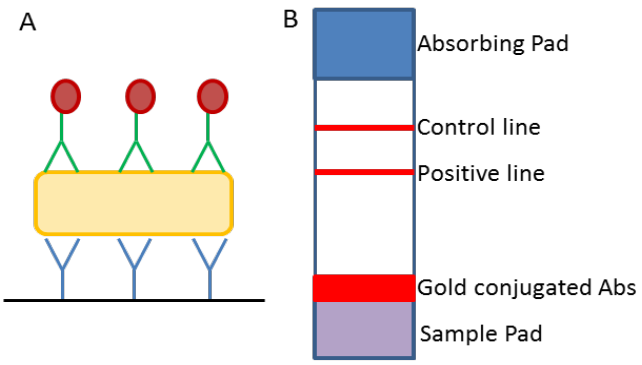


Figure 1-1. Diagram of a sandwich enzyme-linked immunosorbent assay (A) showing the stationary antibodies capturing, and the gold conjugated Abs sandwiching, the antigen. Diagram (B) is a sandwich ELISA in use in a lateral flow strip. The conjugated antibodies produce a coloured line when bound to the antigen.

PoN biosensors that are rapid, specific and sensitive are essential for swift diagnosis and the best possible patient outcome (Byrne *et al.*, 2009). They are a vital element of an affordable, worldwide healthcare system (Sharma *et al.*, 2015). It is also crucial for PoN sensors to be simple to use and provide an easily understandable output (Brandt, 2012). Typically, rapid biosensors are used for medical diagnostics, however, they are also used for environmental monitoring (Ramnani *et al.*, 2015), veterinary medicine (Mettler *et al.*, 2005) and the food industry, for example in early detection of salmonella (Taban *et al.*, 2013).

Phenotypic methods are the most common means of identifying bacteria in clinical laboratories (Janda and Abbott, 2002), however, there are many other techniques that can be used. These bacterial identification techniques include, Southern blot hybridisation of specific DNA fragments (Nilsson *et al.*, 2000), PCR amplification of genes or regions of interest (Harper-Owen *et al.*, 1999) and microarrays of specific DNA sequences (Mitterer *et al.*, 2004). Traditional bacterial typing technologies are often difficult to standardise, lack sufficient ease of use and are not usually suitable for PoN (Jakupciak and Colwell, 2009). Fast and user-friendly PoN sensors would have a huge impact on the early diagnosis and treatment of pathogen related infections (Veigas *et al.*, 2015) and in turn could reduce the transmission rate of infection (Goldenberg *et al.*, 2012). The quick response of accurate treatment could also lower the use of incorrect antimicrobials and therefore, reduce the spread of antimicrobial-resistant bacteria (Ramirez-Rosales and Cantu-Llanos, 2012). The majority of available PoN sensors are not specifically for bacteria and include urine analysis, blood-glucose testing and blood-gas analysis (Shearer *et al.*, 2009; Remes-Troche, 2012).

However, there are some bacterial PoN tests, including tests for Group A Streptococcus (Li *et al.*, 2011) and bacterial vaginosis (Bradshaw *et al.*, 2005).

Two of the major pathogens in the developed world are methicillin-resistant *Staphylococcus aureus* (MRSA) (Barra-Carrasco *et al.*, 2013) and *Clostridium difficile* (*C. difficile*), which cause a high percentage of infections and economic burden (Leekha *et al.*, 2013). Currently there is no single PoN sensor for either of these pathogens that is rapid, sensitive, easy to use and specific.

In order to create a PoN sensor that meets all of the necessary criteria, a biomarker that can be used to identify and distinguish the bacteria must be characterised. For PoN bacterial diagnosis, it is important to minimise the pre-treatment of samples and, therefore, reduce the time taken to perform the test. Ideally the biomarker should be on the surface of, or secreted by the bacteria, so the cells do not need to be lysed prior to testing. The biomarker also needs to be present at a high enough concentration to be detected in complex samples such as serum or faeces (Giljohann and Mirkin, 2009). Genomic and proteomic data can be used not only to determine biomarkers that are unique to bacteria, but when analysed with subcellular localisation bioinformatics tools the data can also be used to differentiate between the biomarkers that are surface exposed and those that are not (Tiwari and Srivastava, 2014).

The subject of this thesis is the exploration and proof of concept of a novel 'top-down' bioinformatics-laboratory approach, to characterise bacterial biomarkers for use in PoN sensors. Although this case study uses *C. difficile*, which will be discussed in detail within this introduction, the bioinformatics approach and downstream laboratory analysis can be utilised for biomarker discovery for any organism of interest. *C. difficile* was selected for the case study as it is a pathogenic organism that caused a worldwide epidemic in 2000 (Carroll and Bartlett, 2011). Also, *C. difficile* associated outcomes would benefit from PoN diagnosis however, there currently is no PoN test used (Goldenberg *et al.*, 2014).

1.2 Biomarkers

A definition of biomarker from the U.S. National Institute of Health is "a biological characteristic, objectively measured (i.e. with acceptable accuracy and reproducibility) and

used as an indicator for a physiological or pathological process, or of the activity of a medicine." (De Gruttola *et al.*, 2001; Group., 2001). The term biomarker therefore covers a broad variety of possibilities. The term 'bacterial biomarker' narrows the scope to a biological characteristic of a bacterium. A bacterial biomarker can take several forms, firstly, the presence of a bacterial species can indicate a particular condition, such as sepsis (Mohan and Harikrishna, 2015). Secondly, the presence of host biomarkers such as cytokines or procalcitonin, which can be measured to identify the presence of a bacterial infection (Holub *et al.*, 2013). Thirdly, the presence of a bacterial gene, or part of a gene, established with sequence-based identification (Jarvinen *et al.*, 2009). Finally, protein biomarkers can be used to identify specific bacteria using whole-cell matrix-assisted laser desorption ionization–time of flight mass spectrometry (Fagerquist, 2013). However, this diagnostic method is not suitable for PoN testing, since the required equipment is typically large and expensive (Suarez *et al.*, 2013). Any biomarker detection that requires culture of the microorganism are often discounted for use in PoN sensors due to the time-consuming nature of the method (Emerson *et al.*, 2008).

Accuracy is fundamental to the use of biomarkers for either screening or diagnostic purposes and is outlined as the ability to correctly classify subjects into clinically relevant groups, for example those infected, or not, with *C. difficile* (Soreide, 2009). The diagnostic accuracy of a PoN test, when incorporating the detection of a biomarker, is determined by two aspects. The first is the accuracy of the biomarker detection molecule and second, the exclusive presence of the biomarker within the organism to be detected (Giljohann and Mirkin, 2009). The discovery of biomarkers has traditionally been carried out using a knowledge driven approach by reference to the literature and user expertise and experience (Bravo *et al.*, 2014). Large-scale screening studies and taxonomic analysis can also reveal bacterial traits that may act as suitable biomarkers for a particular species or strain (Szafranski *et al.*, 2015). However, with the advent of high-throughput genomic scale studies, the systematic discovery of biomarkers now becomes possible through the comparative analysis of 'omics' type datasets (Segata *et al.*, 2011).

1.3 Computational determination of biomarkers

Traditional methods of biomarker identification analysed particular genes or proteins of interest using techniques such as Northern blotting, 2D-PAGE electrophoresis or microarray data (Jemal *et al.*, 2002; Leirdal *et al.*, 2004; Jarvinen *et al.*, 2009). However, more recently, bioinformatics has emerged as a major tool in biomarker identification (Li *et al.*, 2002; Zhang *et al.*, 2014).

The amount of genomic and proteomic data that is accessible has grown exponentially over the past three decades as shown by Figure 1-2 taken from The European Nucleotide Archive statistics page[1] which is part of the European Bioinformatics Institute (EMBL-EBI, 2015). With the necessary computational power and skills, this huge amount of data can be mined to search for possible biomarkers that are unique to a particular bacterial species, or that differentiate strains within species. The available sequence data allows biomarkers that are fragments of proteins to be identified, unlike traditional methods that usually searched for entire proteins (Rabilloud and Triboulet, 2013). Furthermore, this data and computational skill enables a novel 'top-down' approach, were the sequence databases are searched prior to laboratory work. Traditional phenotypic approaches usually use a 'bottom-up' approach in which laboratory work identifies a target that is then searched for within bioinformatics databases, to determine specificity to the species in question (Song *et al.*, 2010).

[1] http://www.ebi.ac.uk/ena/about/statistics

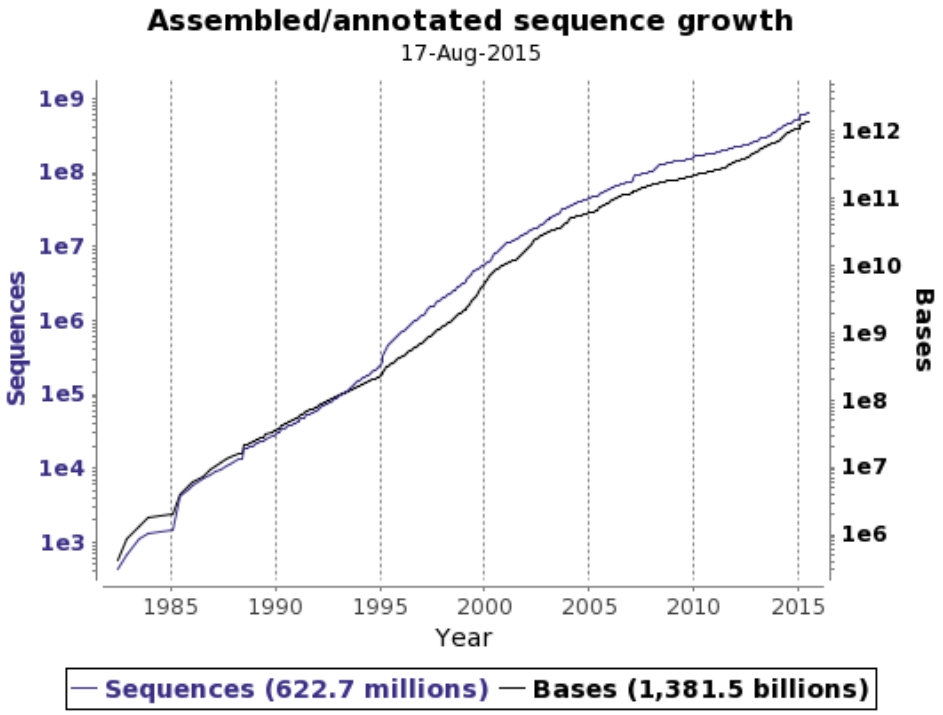


Figure 1-2. Increase in sequence data taken from The European Nucleotide Archive (ENA) , which is part of the European Bioinformatics Institute http://www.ebi.ac.uk/ena/about/statistics and shows the exponential rise, since 1982, of annotated DNA sequences (blue) and the number of DNA bases (black) contained within ENA.

Computational pipelines are often utilised for comparative biomarker discovery, in various ways and for a multitude of functions. One of the most common use of pipelines involves mass spectrometry proteomic analysis of samples, followed by a computational search of proteomics databases to identify the protein (Higdon and Kolker, 2007; Song *et al.*, 2010; Zhang *et al.*, 2014). This pipeline has more recently been used to determine proteins that are potential biomarkers for the differentiation of disease and health (Zhang *et al.*, 2014).

Genomic data has also been analysed using computational pipelines, including transcriptomics data from microarrays and more recently next generation sequencing, to determine biomarkers from gene expression data (Allison *et al.*, 2006; Prassas *et al.*, 2012; Simon and Roychowdhury, 2013). The most straightforward genome pipeline analyses the putative protein products predicted to be encoded by a bacterial genome sequence. A comparison is carried out between all the proteins from one genome with those in another to determine the genes, encoding proteins, that are similar between strains (termed orthologues) and those that might be unique to a given strain, and thus potential biomarkers (Zhao *et al.*, 2015). Bioinformatics tools have also been used to model graphs of

protein similarity data, to enable easy visualisation of protein similarity networks (Enright and Ouzounis, 2001).

1.4 *Clostridium difficile* overview

Clostridium difficile (*C. difficile*) is a Gram-positive, spore forming, anaerobic bacillus, which causes *C. difficile* infection (CDI) (Hernández-Rocha *et al.*, 2013). In developed countries, CDI is the foremost source of nosocomial diarrhoea (Evans and Safdar, 2015) and, in many areas, *C. difficile* has a greater mortality and morbidity rate than methicillin-resistant *Staphylococcus aureus* (MRSA) (Tam Dang *et al.*, 2012; Howerton *et al.*, 2013). The economic burden of CDI is huge, in the United States alone it has been estimated at over $3 billion per year (Howerton *et al.*, 2013). Currently, there is not a single CDI diagnostic test that is used as the worldwide standard as there is variance in the accuracy, cost and operational speed of the differing tests (Hansen *et al.*, 2010). These tests will be described in detail within section 1.14 an accurate and rapid, PoN diagnostic system would reduce the rates of morbidity, mortality and costs associated with CDI (Barbut *et al.*, 2014).

1.5 *Clostridium difficile* background

In 1935, Hall and O'Toole first described *C. difficile*, naming it *Bacillus difficilis* due to their difficulty in isolating the organism and its slow growth in culture. The bacillus was isolated from the stools of healthy neonates where it resided within the normal intestinal flora of these infants (Hall and O'Toole, 1935). The bacterium was later renamed *Clostridium difficile* and in 1978 was shown to be a pathogen when Bartlett and colleagues demonstrated that the bacteria produced the cytotoxin found in the stools of pseudomembranous colitis (PMC) patients (Bartlett *et al.*, 1978a; Bartlett *et al.*, 1978b; Larson *et al.*, 1978). *C. difficile* can affect both humans and animals, and colonisation can be asymptomatic or cause CDI (Bartlett and Gerding, 2008). Toxigenic *C. difficile* strains, those producing toxins, cause CDI primarily after the administration of antibiotics (Bartlett and Gerding, 2008). The infection is often nosocomially acquired, however, community-acquired CDI is on the increase (Gupta and Khanna, 2014).

1.6 Symptoms of *Clostridium difficile* infection

The symptoms of CDI can range from mild self-limiting diarrhoea, cramping or fever through to life threatening PMC (Loo *et al.*, 2005). CDI can infrequently progress to fulminant CDI (FCDI), at a rate of around 1-3 % of total CDI (McMaster-Baxter and Musher, 2007). FCDI includes, developing colonic perforation and toxic megacolon, an acute form of colon distension, which has a high mortality rate of 38 % to 80 % (Foglia *et al.*, 2012). The symptoms of FCDI include severe abdominal pain, dehydration, hypotension, oliguria or anuria, and marked leucocytosis (McMaster-Baxter and Musher, 2007).

1.7 *C. difficile* toxins and pathogenicity locus

C. difficile has two toxins that are recognised as the main virulence factors for CDI, toxin A (TcdA) and toxin B (TcdB). These toxins are both large, TcdA has a molecular weight of 308kDa and TcdB is 270kDa, homologous to each other and both cytotoxic and enterotoxic (Rupnik *et al.*, 2009; O'Donoghue and Kyne, 2011; Calderaro *et al.*, 2013). The *tcdA* and *tcdB* genes are encoded in a 19.6-kb pathogenicity locus (PaLoc) which also contains 3 other genes *tcdR*, *tcdE* and *tcdC* (Figure 1-3) (O'Connor *et al.*, 2009). Non-toxigenic strains of *C. difficile*, that do not produce either TcdA or TcdB, lack the PaLoc and do not induce symptoms of CDI (Kachrimanidou and Malisiovas, 2011).

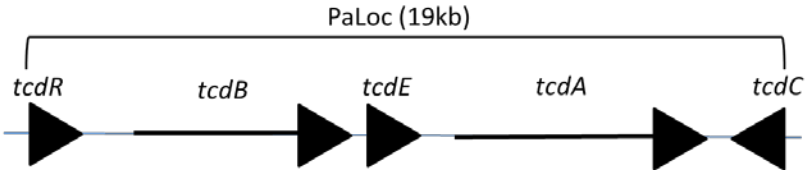


Figure 1-3. The *C. difficile* pathogenicity locus . PaLoc contains the genes encoding toxins A and B and the *tcdC* gene, a putative repressor of toxin A and B production.

TcdA and TcdB are monoglucosyltransferases that catalyse the addition of glucose to the host Rho proteins (Davies *et al.*, 2011). This modification inactivates the Rho-type GTPases, which can no longer polymerise actin filaments, causing cell disruption, collapse of the cytoskeleton and cell death (Carroll and Bartlett, 2011; Yang *et al.*, 2015; Zhang *et al.*, 2015).

There is some controversy over which toxin is largely responsible for virulence in *C. difficile* (King *et al.*, 2015). Early studies reported that TcdA was the essential virulence toxin (Rupnik *et al.*, 2009) whereas, other studies indicated that only TcdB was required for virulence, showing 100-1000 fold more toxicity, to cells *in vitro*, than TcdA (Lyras *et al.*, 2009). However, Kuehne and co-workers showed that both TcdA, TcdB cause *C. difficile* virulence (Kuehne *et al.*, 2010; Kuehne *et al.*, 2014). The variety in results may be caused by differences in animal models and mutant strains used. The contradictory findings do show that neither of the toxins can be ignored in the pathogenicity of *C. difficile*.

C. difficile can also produce a third toxin, an actin-ADP-ribosylating toxin, called *C. difficile* transferase (CDT) , also known as a binary toxin (Schwan *et al.*, 2009). CDT consists of two unlinked polypeptides, CDTa and CDTb that are encoded by the CdtLoc, found in approximately 6 %-28 % strains (Carroll and Bartlett, 2011; Goldenberg and French, 2011). CDT causes microtubule-based protrusions from intestinal epithelial cells that facilitate *C. difficile* colonisation (Schwan *et al.*, 2009). Some *C. difficile* strains that only produce CDT display pathogenicity, even without the main virulence factors TcdA and TcdB, however, these strains are uncommon (Eckert *et al.*, 2014).

1.8 Transmission of *Clostridium difficile* and route of infection

C. difficile is spread via the faecal-oral route, either through vegetative cells or resilient spores (Cohen *et al.*, 2010). *C. difficile*, in either form, is excreted in carriers' faeces, transmitted by person-person or person-object contact, and re-enters the gut via ingestion. The key reservoirs for *C. difficile* are colonised people, either asymptomatic or CDI patients, and inaminate objects, on which spores can survive for months (Gerding *et al.*, 1995; Fordtran, 2006).

The gut is usually protected from intestinal pathogens, such as *C. difficile*, by the host microbiota, a population of commensal bacteria that inhabit the gastrointestinal mucosal surfaces (Guarner, 2006; Dethlefsen *et al.*, 2008; Ubeda *et al.*, 2010). The microbiota provides both physical and chemical barriers, which confer pathogenic resistance via the occupation of mucosal attachment sites, production of antimicrobial substances and consumption of nutrient sources (Servin, 2004; Sekirov *et al.*, 2010; Ubeda *et al.*, 2010). It

has been shown by Dethlefsen and colleagues that more than 5,600 bacterial taxa exist within this gut microbiota (Dethlefsen *et al.*, 2008).

Figure 1-4 provides an excellent overview of the route of CDI and was kindly provided by Dr Susan Poutanen[2], taken from work performed with Simor (Poutanen and Simor, 2004). This figure will be used throughout this section to describe transmission of CDI. Once ingested, the majority of vegetative cells are destroyed by the stomach acid however, the hardy *C. difficile* spores can survive this acidic environment (Poutanen and Simor, 2004). If the gut microbiota is disrupted, usually via the use of broad spectrum antibiotics, it can allow the ingested *C. difficile* spores to germinate into vegetative cells (Figure 1-4). These, and any pre-existing vegetative *C. difficile* cells, then have the space to proliferate (Chang *et al.*, 2008). When toxigenic strains of *C. difficile* colonise the colon, toxins A and or B are produced, causing damage to the intestinal mucosa (Voth and Ballard, 2005). The toxins then inactivate host rho-proteins, causing disruption of the cytoskeleton and cell apoptosis (Rea *et al.*, 2012). The resulting disruption of the gut epithelial cells leads to watery diarrhoea, pseudomembrane formation and colitis (Poutanen and Simor, 2004). These symptoms are exacerbated by toxin-mediated recruitment of a proinflammatory host immune response (Prevention, 2008).

[2] Departments of Laboratory Medicine and Pathobiology and Medicine, University of Toronto

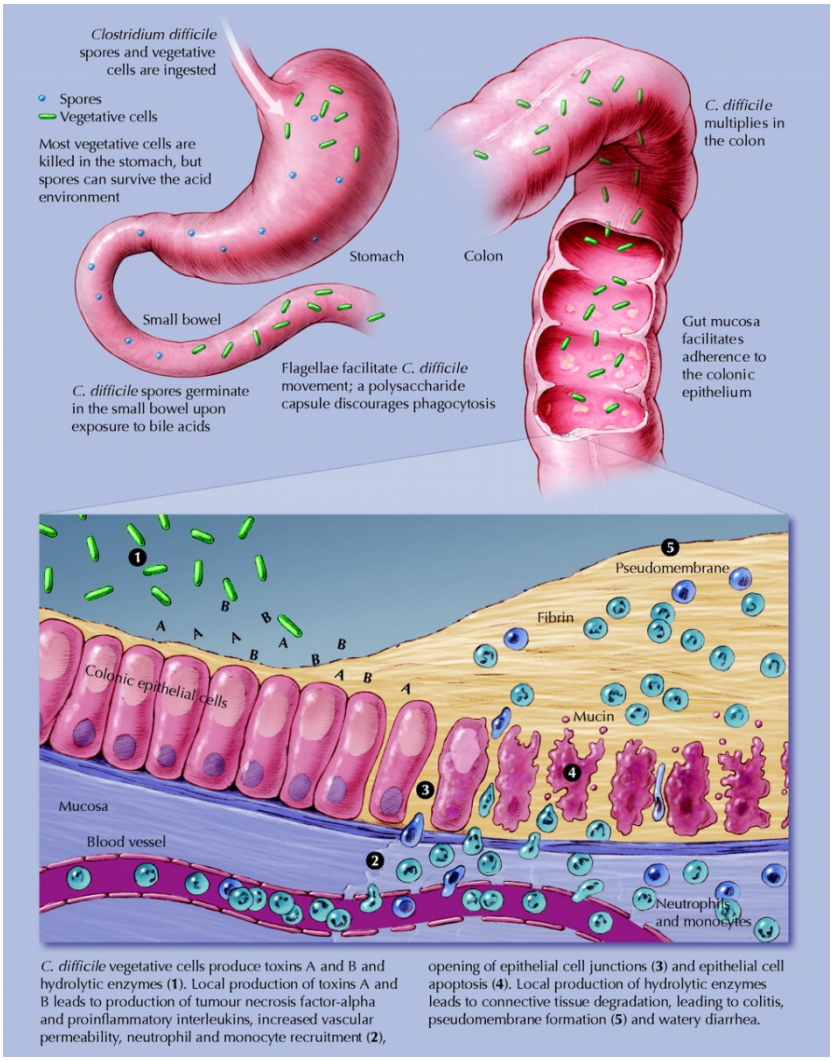


Figure 1-4. How *C. difficile* causes CDI. *C. difficile* is ingested in the form of vegetative cells or spores. The stomach acid kills the vegetative cells however the spores survive, germinating into vegetative cells which can produce toxins. The toxins cause damage to the gut epithelium and induce a host proinflammatory response, leading to CDI (Poutanen and Simor, 2004).

1.9 Asymptomatic carriage of *Clostridium difficile*

Asymptomatic carriage of *C. difficile* is relatively common, the organism is present in the stool of around 5 %-7 % of healthy adults (Foglia *et al.*, 2012; Galdys *et al.*, 2014) ranging up to 30 % of patients in long-term care facilities (Riggs *et al.*, 2007). It has been reported that colonisation of *C. difficile* occurs in 70 % of infants up to the age of three, when the rate of carriage then declines to the values found among healthy adults (Bryant and McDonald, 2009). This colonisation of infants, occasionally with toxigenic strains, rarely leads to CDI, the reasons for which are not understood. However, several postulations have been made

including, the neutralising effect of maternal antibodies and the lack of mature receptors for *C. difficile* toxins (Eglow *et al.*, 1992; Gerding *et al.*, 2012). There is some evidence that screening for asymptomatic *C. difficile* carriers, upon admission to hospital, could reduce the number of CDI cases (Lanzas and Dubberke, 2014).

1.10 Causes and risk factors of *Clostridium difficile* infection

The major risk factors for CDI are antibiotic exposure, hospitalisation and advanced age, usually over 65 years (Taslim, 2009; Carroll and Bartlett, 2011; He *et al.*, 2013). CDI is predominantly regarded as a nosocomial disease, affecting patients often with underlying disorders (Fordtran, 2006; Tattevin *et al.*, 2013) with more than 300,000 nosocomial cases per year in the United States (Rupnik *et al.*, 2009). However, there has been a worldwide increase in community-acquired CDI, even within reasonably healthy people (Prevention, 2008; Gupta and Khanna, 2014).

The risk of CDI in people aged 65 or over has been shown to be 10 times greater than in the younger population (Pépin *et al.*, 2004). In the U.S. in 2008, 93 % of deaths related to CDI were in patients older than 65 years of age. Also in this year, *C. difficile* was reported as the eighteenth leading cause of death in this age group (Miniño *et al.*, 2010). The increased risk of CDI in the elderly may be due to age-related changes in gut microbiota and immune senescence (Simor *et al.*, 2013). Unlike other enteric infections, *C. difficile* has an elevated rate of relapse (Eyre *et al.*, 2012). Recurrence of CDI is a large problem, with up to 30 % of patients developing a second episode (Howerton *et al.*, 2013) and 60 % of these patients being affected by further recurrences (Kachrimanidou and Malisiovas, 2011). Being aged 65 and over is a major risk for recurrent CDI (Buffie *et al.*, 2012). The relapse rate is likely to be due to the high prevalence of spore contamination within primary-care facilities and decreased gut microbiota from previous treatment.

Initially, clindamycin was considered as the main antibiotic risk factor for CDI however, fluoroquinolones and broad-spectrum β-lactam agents have recently been validated as principal triggers (Tattevin *et al.*, 2013). The rise in fluoroquinolone risk, is parallel to the increase in fluoroquinolone-resistant strains (Rupnik *et al.*, 2009). Kachrimanidou and

Malisiovas assert that 85 % of CDI patients have received antibiotics within the 28 days prior to disease onset (Kachrimanidou and Malisiovas, 2011).

Although it has not been fully studied, there is evidence that CDI has been previously initiated by transmission from asymptomatic *C. difficile* carriers (Clabots *et al.*, 1992). These carriers may have the potential to contribute significantly to the disease transmission in long-term care facilities (Riggs *et al.*, 2007). Although it is not currently recommended, there have been studies that suggest screening for CDI would lower the rate of infection (Cohen *et al.*, 2010; Leekha *et al.*, 2013; Lanzas and Dubberke, 2014). The main reason for the lack of screening is due to the traditional, accurate *C. difficile* detection methods, such as toxigenic culture, being time-consuming, costly and labour-intensive. Therefore, if a rapid, accurate and cost-effective technique was available, screening for asymptomatic carriers could occur and the associated risk of such carriers transmitting *C. difficile* could be decreased (Lanzas and Dubberke, 2014).

1.11 *Clostridium difficile* spores

C. difficile sporulates during unfavourable conditions, producing metabolically dormant, endospores (Carroll and Bartlett, 2011). These spores, unlike the vegetative *C. difficile*, can tolerate oxygen and survive in the environment for prolonged lengths of time (Buffie *et al.*, 2012). Spores are shed in the faeces of CDI patients at an estimated rate of between 1×10^4 and 1×10^7 spores per gram of stool (Dawson *et al.*, 2011). These spores are also extremely difficult to eradicate (Alfa *et al.*, 2010) as they have resistance to heat, bleach and radiation (Babakhani *et al.*, 2012). The great persistence and high number of *C. difficile* spores, means that they are prevalent on hospital surfaces and re-infection is a problem (Eyre *et al.*, 2012). Spore contaminated objects within the nosocomial setting range from thermometers and tube feeding equipment, to toilets and carpets (Fordtran, 2006). This pervasiveness of spores provides an environmental, transmission reservoir for CDI (Lawley *et al.*, 2012).

1.12 Hypervirulent *Clostridium difficile* strains

The large increase in incidence of CDI between 2000 and 2010 has been ascribed to an epidemic strain designated restriction endonuclease type B1, North American pulsed-field type 1, PCR ribotype 027 (BI/NAP1/027) (Buckley *et al.*, 2011; Carroll and Bartlett, 2011). This fluoroquinolone-resistant *C. difficile* variant has caused outbreaks in the United States, Canada and across Europe, and is associated with an increased mortality rate of 15 % and recurrence rate of 47 % (Cartman *et al.*, 2010). In 2007/08, at the height of the *C. difficile* epidemic, BI/NAP1/027 caused more than 40 % of CDI cases in the UK (Brazier *et al.*, 2008). He and associates report that fluoroquinolone resistance was the key factor in the rapid emergence of BI/NAP1/027 during the early 2000s (He *et al.*, 2013). The BI/NAP1/027 strain has been classed as 'hypervirulent' and as such, has an increased virulence compared with usual strains due to its amplified toxin production (Heeg *et al.*, 2012) - producing toxins A and B, 16 and 23 times greater, respectively, than non-epidemic strains (Vaishnavi, 2011).

The BI/NAP1/027 strain contains alterations to the *tcdC* gene within the PaLoc region. The *tcdC* gene (Figure 1-3), which is thought to be a repressor of toxin A and B production, has an 18-bp deletion (Bartlett *et al.*, 1978b; Carroll and Bartlett, 2011) and contains a nonsense mutation at position 117 (Cartman *et al.*, 2010). This genetic shift results in the production of a truncated protein which is less effective at minimising toxin A/B production (Miller *et al.*, 2002; Carter *et al.*, 2011). This protein mutation is hypothesised to be a contributor to the increased levels of the toxin production, associated with the hypervirulent BI/NAP1/027 variant (Warny *et al.*, 2005; Carter *et al.*, 2011).

There is some evidence that *C. difficile* BI/NAP1/027 hypersporulates, producing a greater number of spores compared with non-epidemic strains (Wilcox and Fawley, 2000; Merrigan *et al.*, 2010). Although Heeg and colleagues have shown that sporulation rates within variant strains differ, hypersporulation could play a factor in the high transmission rate of epidemic BI/NAP1/027 strains (Heeg *et al.*, 2012). Hypervirulent strains are also linked with the production of the binary toxin CDT, which is thought to aid pathogen attachment and colonisation (Kelly *et al.*, 1994; Schwan *et al.*, 2014).

1.13 *Clostridium difficile* related mortality rates

Since 1978, CDI has accounted for 20-25 % of all cases of antibiotic-associated diarrhoea (AAD) (Bartlett, 2007) and at least 90 % of pseudomembranous colitis incidences (Bartlett *et al.*, 1978b; Kelly *et al.*, 1994). Between 0.5 % and 1.5 % of all hospital admissions develop CDI (O'Brien *et al.*, 2007), within which the mortality rate is 1.5-6 % (Miller *et al.*, 2002; Loo *et al.*, 2005) and 6-30 % suffer from pseudomembranous colitis (McMaster-Baxter and Musher, 2007). In many areas, *C. difficile* has a greater mortality and morbidity rate than methicillin-resistant *Staphylococcus aureus* (MRSA) (Tam Dang *et al.*, 2012; Leekha *et al.*, 2013).

Between 2000 and 2005, incidences of CDI, in the United States (US) and parts of Europe, tripled and the associated morbidity and mortality rates increased (Gerding *et al.*, 2008; Rupnik *et al.*, 2009). The rising rates in CDI were attributed to the hypervirulent strain BI/NAP1/027 (Loo *et al.*, 2005; McDonald *et al.*, 2005) and it was estimated that in the U.S. up to 20,000 patients die annually from CDI (Rupnik *et al.*, 2009).

A total of 14,165 CDI cases were reported within the NHS between April 2014 and March 2015, which was a 6 % increase on the 2013/14 rates, however, when compared to the epidemic of 2007/08, it remains a huge reduction in cases (Gerver *et al.*, 2015). The most recent mortality rates associated with CDI in England and Wales are for 2012, when 1646 deaths were attributed to the infection, equating to 15.3 deaths per million population (National Statistics, 2013). Again, this is a large decrease compared to 2007 when 4056 deaths were attributed to CDI (Hudson, 2010). Infection rates are not down to the pre-epidemic numbers, however, the introduction of antimicrobial stewardship, infection control barriers and education of health-care workers and patients has caused a significant decrease in CDI cases and deaths (Ghose, 2013).

1.14 Current diagnosis techniques for *Clostridium difficile* infection

There are several methods for diagnosing CDI, however, no single technique has been implemented universally (Hansen *et al.*, 2010). The two key factors for any diagnosis are sensitivity and specificity. Sensitivity is the accuracy in which the method can detect those with CDI and does not produce false negatives. Specificity is the accuracy in which the

method determines those that are 'disease-free' and does not produce false positives (Parikh *et al.*, 2008). Since non-toxigenic strains do not cause CDI, *C. difficile* can be present within samples without infection (Kachrimanidou and Malisiovas, 2011). Therefore, to diagnose CDI, determining the presence of *C. difficile* is not enough, either toxin A or B needs to be identified too (Crobach *et al.*, 2009).

Initially CDI should be predicted through clinical diagnosis and then confirmed with laboratory methods, suspected CDI patients should have a history of antibiotic use coupled with diarrhoea with a characteristic foul odour (Bartlett and Gerding, 2008). The majority of laboratory diagnostic methods are performed using stool as the test sample, however, the range of tests vary widely.

1.14.1 Enzyme immunoassays

Toxin enzyme immunoassays (EIAs), the CDI diagnostic method most routinely used in American laboratories, detects toxins A, B or both (Calderaro *et al.*, 2013). Detection is performed using antibodies against the specific toxins either via solid-phase, well type, or membrane assays (Barra-Carrasco *et al.*, 2013). These assays are fairly rapid and inexpensive however, they have relatively low sensitivity and specificity (Johnson *et al.*, 2001; Lalande *et al.*, 2011). Bartlett and Gerding state that EIAs should not be used as a stand-alone test for CDI, as the EIAs for toxin A only or both toxins can miss ~40 % of diagnoses (Bartlett and Gerding, 2008) , compared with stool culture (Delmee *et al.*, 2005) and the cell cytotoxicity assay (Shanholtzer *et al.*, 1992). A great deal of research is currently being performed to find methods that can amplify the toxin signal, therefore improving the sensitivity of toxin tests (Song *et al.*, 2015).

There are also EIAs that detect the presence of glutamate dehydrogenase (GDH), a metabolic enzyme produced by *C. difficile* and also many other bacteria (Commichau *et al.*, 2008; Harper *et al.*, 2010; Burnham and Carroll, 2013). Typically, bacterial GDHs are cytoplasmic or membrane-associated proteins however, GDH can also be found in the extracellular culture medium of *C. difficile* (Girinathan *et al.*, 2014). GDH EIAs tend to have high sensitivity as the enzyme is produced in large quantities and therefore few false negatives are produced (Tenover *et al.*, 2011). However, GDH EIAs do not detect the

C. difficile toxins and are unable to determine the *C. difficile* strains that are actually causing infection, therefore, the test is poorly specific (Goldenberg *et al.*, 2010). Other bacteria also possess GDH, and so these EIAs can display cross-reactivity, for example with *C. sordellii* (Lyerly *et al.*, 1988; Burnham and Carroll, 2013). Although GDH EIAs are a good screen for CDI they are not suitable for use as a standalone test (Brecher *et al.*, 2013)

1.14.2 Nucleic acid amplification techniques

In 2010, the US Food and Drug Administration (FDA) approved the first commercial real-time *C. difficile* polymerase chain reaction (PCR) assay (Kvach *et al.*, 2010). The PCR diagnostic method amplifies regions in either the *tcdA* or *tcdB* genes. Some commercially available kits also amplify the *tcdC* gene deletion to detect the hypervirulent strain BI/NAP1/027 (Barra-Carrasco *et al.*, 2013). PCR comes under the category of nucleic acid amplification techniques (NAATs), which provide diagnosis for *C. difficile* that is sensitive and specific. It has been stated that NAATs are superior to all other methods, with the exception of toxigenic culture (Carroll and Bartlett, 2011). However, because NAATs are highly sensitive at detecting toxigenic *C. difficile* and not the toxins themselves, studies have reported that it has caused positive results for patients who are perhaps carriers of toxigenic *C. difficile* but do not have clinical CDI (Planche *et al.*, 2013; Koo *et al.*, 2014). Another disadvantage of NAATs are the associated cost as they require expensive equipment and technical skills to perform the technique (Kufelnicka and Kirn, 2011; Buchan and Ledeboer, 2014).

1.14.3 Stool culture

Stool culture (SC) is a highly sensitive method for detecting *C. difficile,* however, this technique is not specific as SC cannot differentiate between asymptomatic carriers and CDI sufferers, as it does not detect toxins (Bartlett and Gerding, 2008). For the method, stool samples are subjected to alcohol shock, inoculated onto cycloserine cefoxitin fructose agar plates, and grown anaerobically. Colonies are determined as *C. difficile* by morphology, Gram stain and latex agglutination. This determines if the sample is positive or negative for CDI (Peterson *et al.*, 2007; Eastwood *et al.*, 2009).

1.14.4 Cell cytotoxicity neutralization assay

The cell cytotoxicity neutralization assay (CCNA) is used directly with stool specimens. The supernatants from faecal samples are anaerobically incubated with cultured cells, such as Vero cell lines (Kachrimanidou and Malisiovas, 2011), for at least 24 hours. If the cell monolayer shows a typical cytopathic effect, via cell rounding, then a toxin antiserum is added. If the cell rounding is reversed then a positive result for toxigenic *C. difficile* is recorded (Hernández-Rocha *et al.*, 2013). CCNA is not generally used in clinical settings as it is time consuming and requires technical equipment for the upkeep of cell lines (Reller *et al.*, 2010). CCNA has often been used as the reference standard, when assessing other *C. difficile* diagnostic methods, yet the CCNA technique is not standardised, with Laboratories using different cell lines and techniques (Calderaro *et al.*, 2013). These differences lead to varying results for the sensitivity and specificity of the technique. Overall, CCNA is regarded as relatively specific, however, Cohen and associates believe that it is too insensitive to be used as a reference standard (Cohen *et al.*, 2010).

1.14.5 Toxigenic culture

Toxigenic culture (TC) is the gold standard for diagnosing *C. difficile,* in the sense that it is the most sensitive and specific method and is often used as a comparison for the success rates of other CDI determining techniques (Peterson *et al.*, 2007). During toxigenic culture, faecal samples are treated in the same manner as SC. Positive *C. difficile* colonies are then tested for toxin production using the CCNA method described above (Hernández-Rocha *et al.*, 2013). This method is laborious and slow, requiring 72-96 hours, thus it is not practical for clinical settings and therefore seldom used (Lalande *et al.*, 2011).

1.14.6 Overview of the advantages and disadvantages of all CDI diagnostic tests

There have been many studies on the sensitivity and specificity of CDI diagnosis and variations occur between studies, such as differences in the ‘gold standard’ reference, for which all the other methods are tested against, and dilutions/storage times of samples. These variations mean that the results of such studies are inconsistent, although there is a general trend for each diagnostic. Table 1-1 provides an overview of the ‘pros and cons’ of different CDI diagnostic tests (O'Brien *et al.*, 2007; Hansen *et al.*, 2010; Shetty *et al.*, 2011;

Fraser and Swiencicki, 2013; Issarachaikull *et al.*, 2015). The table demonstrates the lack of reasonably priced, rapid, sensitive and specific diagnostic systems.

Table 1-1. A comparison of the sensitivity, specificity, cost and completion time, for differing CDI diagnostic methods , providing an overview of the related advantages and disadvantages.

Method	Sensitivity (%)	Specificity (%)	Cost ($)	Time (hours)	Advantages	Disadvantages
Toxin Enzyme Immunoassays	Low	High	5-15	< 2	Rapid and simple to use	Low sensitivity
Glutamate Dehydrogenase	Medium	Med/high	5-15	< 1	Rapid and can be used as initial screening tool	Requires further toxigenic testing
Nucleic acid amplification tests (PCR)	High	High	40-50	1-4	Relatively sensitive and specific	Requires expensive specialist equipment and skills
Cell Cytotoxicity Neutralisation Assay	High	High	20-30	24-48	High specificity	Labour intensive, slow, requires specialist equipment
Stool Culture	High	Med/high	20	> 48	Sensitive	Slow and not specific for toxigenic strains
Toxigenic Culture	High	High	10-30	≥ 96	Highly sensitive	Labour intensive, very slow, requires specialist equipment

1.14.7 Current recommendations for diagnosing CDI

Current UK and European guidelines for CDI diagnosis do not recommended toxin EIAs for use as a stand-alone test, due to their lack of sensitivity (Crobach *et al.*, 2009; Wilcox, 2013). However, toxin EIAs or GDH EIA are recommended as part of a two test algorithm, which begins with a test for the presence of *C. difficile* using either NAATs or GDH EIA, followed by toxin EIA (Crobach *et al.*, 2009; Wilcox, 2013). If the initial test is negative then the second test is not performed (Bartlett, 2007).

1.15 Current treatment for *Clostridium difficile* infection

Accurate and early detection is crucial for the successful treatment of CDI and for minimising transmission of *C. difficile* from patient to patient (Barbut *et al.*, 2014). Treatment for CDI is dependent on the severity of the infection and whether it is the initial infection or relapse. Metronidazole is used in initial episodes of mild-to-moderate CDI and vancomycin is usually administered for the first episode of severe CDI (Cohen *et al.*, 2010). Vancomycin is not used for moderate CDI as there is concern that its use may select for

vancomycin-resistant enterococci (Al-Nassir *et al.*, 2008). This antibiotic treatment directly suppresses *C. difficile*, presumably allowing recovery of the indigenous microbiota, thus restoring pathogenic resistance (Chang *et al.*, 2008). Howerton and associates state that the failure rate of CDI treatment, with first-line antibiotics, can reach 38 % (Howerton *et al.*, 2013).

The narrow-spectrum antibiotic, fidaxomicin, is more effective against *C. difficile in vitro* than vancomycin (Louie *et al.*, 2011). Furthermore, this antibiotic has no activity against Gram-negative bacteria and therefore causes limited damage to the host's microbiota (Karas *et al.*, 2010; Ghose, 2013). *In vivo* trials of fidaxomicin have shown similar results to vancomycin, and lower rates of recurrence (Crook *et al.*, 2012). In 2012, fidaxomicin was approved for the treatment of CDI in Europe (Wilcox, 2013), however, mainly due to the cost, fidaxomicin is not universally used (Bartsch *et al.*, 2013).

An emerging area of CDI treatment is with the use of monoclonal antibodies against either toxin A or B. A phase II study, which administered the monoclonal antibodies with antibiotics, saw a reduced CDI recurrence rate, down to 7 % compared with 25 % in the patients which received only antibiotics (Lowy *et al.*, 2010). Unfortunately, due to the huge cost of clinical testing and getting the necessary approvals, no monoclonal antibodies for the treatment of CDI have yet come to market.

Faecal microbiota transplantation (FMT) is the transplantation of faeces from a healthy donor to CDI patients either by nasogastric tube, colonoscopy or enema and is another treatment used to reduce recurring CDI (Rohlke and Stollman, 2012; Rodriguez *et al.*, 2015). The faeces convey the healthy microbiota to the new host, re-establishing the host's ability to defend itself from *C. difficile* colonisation. FMT has been shown to be successful in preventing CDI recurrence in 87 % of patients (Cammarota *et al.*, 2014) and in a study comparing FMT with vancomycin, recurrent CDI was resolved using FMT 13 out of 16 patients compared with only 4 out of 13 for vancomycin (Drekonja *et al.*, 2015).

1.16 The *Clostridium difficile* S-Layer

Although toxins A and B have been shown to be major virulence factors in CDI, other features influencing *C. difficile* colonisation of the host, may also play a key role (Vedantam *et al.*, 2012). The *C. difficile* surface layer (S-Layer), is involved in pathogen-host interactions and has been linked as a potential virulence factor (Spigaglia *et al.*, 2013).

The *C. difficile* S-layer is a paracrystalline structure that completely coats the outer surface of vegetative cells (Reynolds *et al.*, 2011). Unlike other bacterial S-layers, which are made of one protein, the *C. difficile* S-layer consists of two proteins, the high molecular weight (HMW) and the low molecular weight (LMW) surface layer proteins (SLPs) (Fagan and Fairweather, 2011). These SLPs are post-translationally cleaved from a single precursor protein, SlpA (Dang *et al.*, 2012). The precursor protein is encoded by the *slpA* gene which, after translocation to the cell surface (Calabi *et al.*, 2001), is post-translationally cleaved by the cysteine protease Cwp84 to form the individual LMW and HMW SLPs (Kirby *et al.*, 2009; Dang *et al.*, 2012). The site of cleavage has been shown to be highly conserved across *C. difficile* strains (Calabi *et al.*, 2001). Post-cleavage, the two SLPs associate into heterodimers and form a large percentage of the S-layer (Drudy *et al.*, 2004). A diagram of the *C. difficile* cell surface, with the LMW and HMW SLPs forming the S-layer is provided in Figure 1-5. The cytoplasmic membrane is surrounded by a thick peptidoglycan cell wall and the entire cell is encapsulated with the S-layer which is mainly made from the mature LMW and HMW SLPs that are cleaved from SlpA. The S-layer also contains some other cell wall proteins such as the adhesion protein Cwp66 (Dingle *et al.*, 2013).

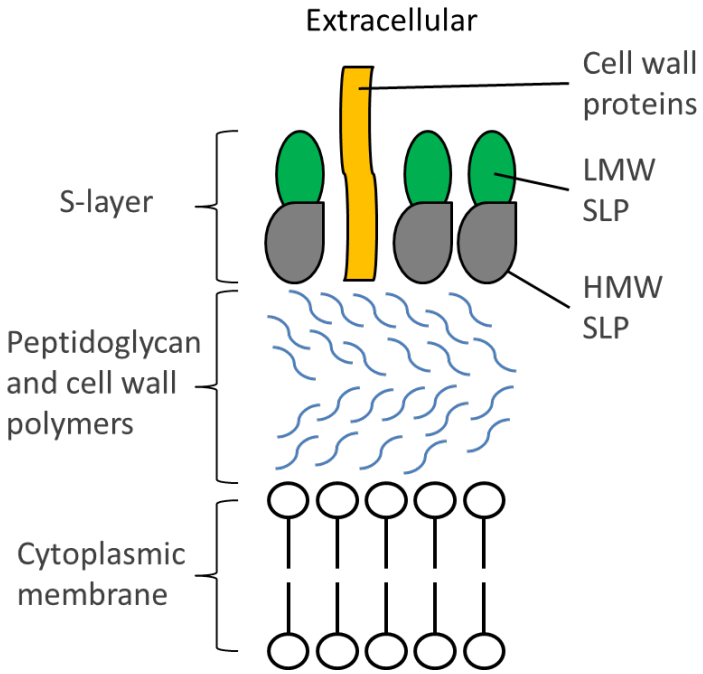


Figure 1-5. Structure of the Gram-positive *C. difficile* cell surface . A lipid bilayer forms the cytoplasmic membrane which is surrounded by a thick peptidoglycan cell wall. The entire cell is encapsulated with the S-layer which is made mainly from the mature LMW and HMW SLPs that are cleaved from SlpA. The S-Layer also contains some other cell wall proteins.

The *slpA* and *cwp66* genes are located within the 36.6-kb cell wall protein (*cwp*) gene cluster, which also encodes the ATPase SecA2 to which the secretion of both SlpA and Cwp66 are dependent on (Fagan and Fairweather, 2011; Dingle *et al.*, 2013). SlpA has 28 paralogs, including Cwp66, each of which are predicted to be surface associated (Sebaihia *et al.*, 2006).

The HMW SLP is formed from the more conserved C-terminal end of SlpA, whereas, the highly variable N-terminal forms the LMW SLP after cleavage (Cerquetti *et al.*, 2000). Both SLPs vary in sizes between variant *C. difficile* strains, with LMW SLP ranging between 36-45kDa and HMW SLP 48–56 kDa (Calabi *et al.*, 2001; Karjalainen *et al.*, 2002).

The SLPs are involved in the mechanism of gut colonisation and in the process of adhesion to the intestinal mucosa (Cerquetti *et al.*, 2002). Calabi and colleagues showed strong and specific binding of purified SLPs to human epithelial cells and gastrointestinal tissues, with recombinant HMW SLP binding with a much higher affinity than the LMW SLP. When anti-

HMW antibodies were introduced the binding of the SLPs to human cells was reduced (Calabi *et al.*, 2002).

Fagan and Fairweather state that "SlpA is the most highly expressed protein in *C. difficile*, accounting for 10-15 % of total cellular protein" and it has been shown that SlpA is also highly expressed in all stages of growth (Savariau-Lacomme, 2003; Fagan and Fairweather, 2011).

1.17 Aims and Objectives

The main aim of this project was to determine the effectiveness of a novel bioinformatics workflow in identifying bacterial biomarkers that are unique to organisms within a group of interest. *C. difficile* was used as a case study within this project with the aim of identifying a unique surface associated biomarker that could be used in a point of need (PoN) sensor to detect the bacteria.

The initial objective was to develop a workflow that could differentiate surface associated biomarkers. The biomarker could be part of, or a whole protein that was unique to *C. difficile* and found within no other sequenced organisms. The biomarker must also be common to all strains of *C. difficile*.

The second objective was to validate the biomarker for potential use in a PoN sensor using laboratory methods. This objective included establishing the presence of the biomarker within multiple *C. difficile* strains, whilst confirming the absence of the biomarker in closely related species.

Chapter 2 Laboratory methods

2.1 Bacterial storage and growth

All *C. difficile* strains were cultured in brain heart infusion broth (BHI) supplemented with 0.5 % yeast extract and 0.1 % L-cysteine (Sigma), or cooked meat broth (Sigma) and on plates made from *C. difficile* agar base (Sigma) with *C. difficile* C.D.M.N. – selective supplement (Oxoid), 1 vial to every 500ml *C. difficile* agar, and 7 % sheep's blood. *C. sordellii* and *P. anaerobius* were cultured in BHI or on soy-tryptone agar (15 g/l tryptone, 5 g/l soytone, 5 g/l NaCl, 15 g/l agar) supplemented with 7 % sheep's blood. *C. hiranonis* was cultured in PY + X medium (Table 2-1 and Table 2-2). All incubations were anaerobic, using anaerobic jars and Anaerocult A gas packs (Merck), at 37°C. The broths were incubated for 12-16 h and the agar plates were incubated for 48 h.

Table 2-1. PY + X medium for growth of *C. hiranonis*

Ingredients	Quantity
Tryptone	5 g/l
Peptone (pepsin digested)	5 g/l
Yeast	10 g/l
Salt solution (see Table 2)	40 ml/l
L-Cysteine	0.5 g/l
D-Glucose	5 g/l

Table 2-2.Salt solution used in the PY + X medium for *C. hiranonis* culture

Ingredients	Quantity
$CaCl_2$ x 2 H_2O	0.25 g/l
$MgSO_4$ x 7 H_2O	0.5 g/l
K_2HPO_4	1 g/l
KH_2PO_4	1 g/l
$NaHCO_3$	10 g/l
NaCl	2 g/l

For short term storage, cultures were anaerobically kept on agar plates at room temperature. For long term storage, liquid cultures were grown to exponential phase (OD_{600} 0.4-0.6) and mixed with sterile glycerol to a final concentration of 25 % and stored at -80°C.

2.2 Bacterial strains

Table 2-3. List of bacterial strains used throughout this project

Species	Strain	Type
C. difficile	630	Reference Strain
C. difficile	R002*	Ribotype
C. difficile	R005*	Ribotype
C. difficile	118497G*	Clinical isolate
C. difficile	128703G*	Clinical isolate
C. difficile	994535*	Clinical isolate
C. difficile	Ox160	SLTS
C. difficile	Ox575	SLTS
C. difficile	Ox858	SLTS
C. difficile	Ox1121	SLTS
C. difficile	Ox1145	SLTS
C. difficile	Ox1192c	SLTS
C. difficile	Ox1342	SLTS
C. difficile	Ox1396	SLTS
C. difficile	Ox1424	SLTS
C. difficile	Ox1437a	SLTS
C. difficile	Ox1523	SLTS
C. difficile	Ox1533	SLTS
C. difficile	Ox1896	SLTS
C. difficile	Ox2404	SLTS
B. subtilis	BSB1	Lab strain
E. coli	Rosetta pLMW1-262-1	Lab strain
C. sordellii	ATCC 9714	Lab strain
C. hiranonis	DSM-13275	Lab strain
P. anaerobius	ATCC 27337	Lab strain
S. aureus	ATCC 29213	Lab strain

C. difficile strains marked with an asterisk were kindly donated by Dr Kerry Hill from the Freeman Hospital, Newcastle upon Tyne. The *C. difficile* Oxford strains (Ox), are the 14 S-Layer type strains, which were kindly donated by Dr Kate Dingle, Nuffield Department of Medicine, Oxford. The *E. coli* Rosetta pLMW1-262-1 was kindly donated by Professor Neil Fairweather, Imperial College London, and was used to produce the recombinant LMW SLP.

2.3 DNA Extraction

Cultures of bacteria were grown as described in section 2.1.1, 50 ml O/N and DNA extraction was performed with a DNeasy Blood and Tissue kit (Qiagen), following the protocol for Gram-positive bacteria, with the volume of elution buffer decreased to 30 µl. The DNA was quantified using a NanoDrop[3] spectrophotometer.

2.4 Polymerase chain reaction (PCR)

Clone Manager Version 9[4] was used to model PCR primers. The 'PCR clone wizard' tool was followed, with differing inputs depending on the target. The presence of primers, in all *C. difficile* genomes, was checked using Jalview in 'clustalx' view. The extracted DNA was amplified using Phusion polymerase (NEB), the conditions shown in Table 2-4 and the designed primers:

Forward *SlpA* - ATGAGTATAGCTCCAGTTGC

Reverse *SlpA* – ATCTTCATCACCATCTCCTGC

Table 2-4. PCR conditions used with Phusion polymerase

Temperature	Time	Stage	Cycle number
98 °C	30 s	Initial denaturation	1
98 °C	10 s	Denaturation	30
x °C = T_m (lowest) – 5 °C	20 s	Anneal	30
72 °C	30 s/kb of product	Extension	30
72 °C	5 min	Final Extension	1
4 °C	hold		

2.5 Agarose Gel Electrophoresis

The PCR products were separated on a 0.8 % w/v agarose/TAE buffer gel with 0.5 µl/ 10 ml agarose Safe View (NBS biologicals), at 90 v for 1 h, with a 1kb DNA ladder (Promega), to determine the product's size.

[3] A micro-sample spectrophotometer that can be used to measure DNA, RNA and protein. From Thermo Scientific. http://www.nanodrop.com/Default.aspx

[4] Scientific & Educational Software copyright 1994-2007 http://www.scied.com/pr_cmbas.htm

2.6 DNA Sequencing

The PCR product was purified using a Qiagen QIAquick PCR Purification Kit and quantified using the Nano-Drop spectrophotometer with ND1000 V3.7.1 software. Samples with values over 40 ng/µl were sent for sequencing by Geneius Laboratory[5] with the primers:
Forward *SlpA* - ATGAGTATAGCTCCAGTTGC

Reverse *SlpA* – ATCTTCATCACCATCTCCTGC

2.7 Low pH glycine, surface protein extraction

Cells from O/N, 50 ml liquid cultures were harvested via centrifugation at 2700 x *g* for 15 min 4^{o}C and washed in 10 ml 0.1 M PBS. The cell pellet was resuspended in 500 µl, 0.2 M glycine pH 2.2 and incubated at RT for 30 min. Finally, the solution was centrifuged at 2700 x *g* for 15 min and the supernatant was removed and neutralised with 2 M Tris, pH 7.

2.8 Determination of protein concentration

An indication of protein concentration was gained using a Nano-Drop spectrophotometer with ND1000 V3.7.1 software. To determine protein concentrations more accurately, a bicinchoninic acid (BCA) assay kit (Thermo scientific) was used, as per the manufacturer's instructions, using BSA as the standard protein. If required, samples were diluted to get the absorbance values within the range of the standard curve, an example of a standard curve is displayed in the Appendices Figure 9-1.

2.9 Sodium Dodecyl Sulphate-polyacrylamide gel electrophoresis

Proteins were separated according to their molecular weights by discontinuous sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Briefly, samples were mixed 2:1 with loading buffer (1.8 ml 10 % SDS (w/v), 1.8 ml 50 % glycerol (v/v), 0.5 ml 1 M Tris/HCl pH 6.8, 0.1 % bromophenol blue (w/v) and 0.5 ml β-mercaptoethanol), boiled for 10 min and pulse spun in a centrifuge prior to being loaded onto a polymerised acrylamide gel

[5] http://www.geneiuslabs.co.uk/

at 10 % (v/v). Samples were either whole cells or 5 µg protein sample. The gels were used in a BioRad gel tank system and a voltage of 80-150 V was applied for 1-2 h, with a lower voltage being applied as the samples migrate through the 'stacking gel'. The recipe for the stacking and separating gels are shown in Table 2-4. Gels were then stained, with constant agitation, in coomassie staining solution (1 g/L coomassie brilliant blue R250, 50 % methanol (v/v), 40 % ddH_2O (v/v) and 10 % acetic acid (v/v)) for 1 h and destained in destaining solution (30 % methanol (v/v), 60 % ddH_2O (v/v) and 10 % acetic acid (v/v)) until background was cleared.

Table 2-5. SDS-PAGE recipe for both the stacking and resolving part of two gels.

Component	Stacking gel	Resolving gel (10 %)
ddH_2O	2.45 ml	3.1 ml
Buffer (I & II)[1]	1.25 ml(II)	2.5 ml (I)
Acrylamide solution[2]	0.75 ml	3.3 ml
10% SDS	50 µl	100 µl
2% TEMED	250 µl	500 µl
1.4% APS	250 µl	500 µl

[1] Buffer I, 1.5 M Tris/HCl, pH 8.8; Buffer II, 0.5 M Tris/HCl, pH 6.8.
[2] Acrylamide

2.10 Scanning of SDS-PAGE gels

Gels were placed inside a plastic wallet and scanned using an Epson Perfection V350 scanner and professional software package.

2.11 Western blot

Proteins were separated by SDS-PAGE and transferred on to pre-activated Amersham hybond PVDF blotting membrane (GE Healthcare) using a BioRad wet-blot system. The membrane was activated upon submersion in methanol for 15 s and equilibration in transfer buffer (20 mM Tris, 192 mM glycine, 0.1 % SDS (w/v), 20 % methanol (v/v)). Transfer occurred on ice with transfer buffer, at 350 mA for 1 h. The PVDF membrane was incubated overnight at 4^oC with blocking buffer (0.1 M PBS, 5 % non-fat milk powder (w/v) and 0.05 % tween-20 (v/v)). The membrane was then incubated with the primary antibody diluted in blocking buffer at RT for 1.5 - 2 h. The membrane was then washed with blocking buffer 4 times for 5 min each. The membrane was then incubated at RT for 1 h with peroxidase-conjugated goat anti-mouse IgG (F_c specific) (Sigma), diluted 1:120,000 in blocking buffer.

The membrane was washed for 15 min with PBS with 0.05 % tween-20 (v/v) and then another 4 x 5 min with PBS only. For visualisation of antibody binding, the membrane was treated with the Enhanced Chemiluminescence (ECL) kit from Amersham, GE Healthcare, as per manufacturer's instructions. The bands were viewed using an ImageQuant LAS4000mini biomolecular imager (GE Healthcare) and processed with the accompanying software and ImageJ.

2.12 Production of HMW SLP monoclonal antibodies

The biomarker recognition molecules that were selected for this project were monoclonal antibodies (mAbs). AbMART, a monoclonal antibody manufacturing service, was employed to produce mAbs against the *C. difficile* biomarker (Table 4-1). To produce the mAbs, AbMART used synthetic linear peptides, 10 amino acids in length, as the antigen. To retain some level of confidentiality and as a confirmatory exercise of the bioinformatics work performed within this project, the entire *C. difficile* 630 HMW SLP sequence, taken from NCBI YP_001089306.1 (Appendices, section 9.3) was sent to AbMART, rather than the specific biomarker sequence. AbMART utilised bioinformatics techniques to analyse the HMW SLP sequence and provide possible epitope regions that were predicted to be exposed on the surface of the protein. The *C. difficile* biomarker was returned as a number of potential epitope sites, supporting the prediction obtained within this research that the biomarker is surface exposed. AbMART was then instructed to focus on the region containing the biomarker, in order to produce the mAbs.

2.13 Hybridoma culture and production of antibody

All cell culture reagents were purchased from Sigma. Hybridoma cells were grown at 37^{o}C in a 5% CO_2 incubator, in basic media (RPMI 1640, 10 % foetal calf serum (FCS) (v/v), 100 mg/l penicillin-streptomycin and 2 mM glutamine (w/v)), initially with the addition of 10 % Condimed (v/v) to aid recovery. When cells were fully confluent they were grown in media with reducing FCS content until they were able to grow efficiently in serum free media (10 %, 5 %, 2 %, 0.5 %, serum free). Large scale growth of the cells was performed in specialist media, EX-CELL® 610-HSF Serum-Free Medium for Hybridoma Cells, Low-protein, with L-glutamine, to which 100 mg/l penicillin-streptomycin was added. Cells were left in

this media for 8-10 days until many of the cells were dead, the supernatant was then harvested at 13000 rpm for 5 mins.

2.14 Antibody purification via affinity chromatography

Antibody purification was performed with an ÄKTA start and a HiTrap Protein G HP, 1 ml column (GE Healthcare Life Sciences) and all solutions were filtered and degassed. The column was flushed ≥ 10 ml of wash buffer (20 mM phosphate pH 7.2) and a minimum of 260 ml cell culture supernatant was added to the column at a 4:5 ratio with 100 mM Na phosphate, pH 7.2. The column was washed with 20 ml wash buffer and eluted with 10 ml, 100 mM glycine/HCl, pH 2.0 in 1 ml fractions. UV absorbance chromatograms for the purification were collected using a chart recorder and OD280 was of the purified antibody was determined via NanoDrop spectrophotometer.

2.15 Isotyping of mAbs

Rapid Mouse Isotyping Kit - Gold Series (Cambridge Bioscience) was used to type the mAbs according to the manufacturer's instructions. Briefly, 40 µl cell culture supernatant was added to lateral flow strip 1 in order to indicate the Ig type. If IgG was indicated then the strip 2 was utilised to determine the IgG type and if the light chains were Kappa or Lambda again with 40 µl cell culture supernatant. Both strips were read within 15 min of adding the supernatant.

2.16 Whole cell Enzyme-linked immunosorbent assay

Overnight bacterial cultures were grown to an OD_{600} 0.6 and 8 ml of culture was centrifuged at 2700 x g for 10 min and washed three times with 0.1 M PBS. The cells were resuspended in 6ml 50mM carbonate/bicarbonate buffer, pH 9.6 and 150 µl was added to the wells of a Nunc Maxisorp 96 well, flat-bottomed plate (Affymetrix eBioscience) and left O/N at 4^{o}C. Cells were checked for adherence to the ELISA plate using an inverted microscope, however quantitative analysis of cell adhesion could not be performed. The cell solution was removed and washed 3 x with 0.1 M PBS, 0.05 % tween pH 7.4 (PBS-T), the plate was then blocked in blocking buffer (PBS-T with 5 % non-fat milk), for either 2 h at room temperature or O/N at 4^{o}C. The blocking buffer was removed and the plate was washed twice with PBS-T.

The mAbs were diluted to differing concentrations (10 µg/ml – 0.1 ng/ml) in blocking buffer and 100 µl was added to the wells and incubated O/N at 4°C. The plate was washed 4 x PBS-T before the secondary antibody, peroxidase-conjugated goat anti-mouse IgG (F_c specific) (Sigma), was diluted 1:20,000 in blocking buffer and added to the wells. The plate was incubated for 2 h at RT before removing and washing the plate 4 x PBS-T. One Ophenylenediamine (OPD) tablet was added to 50 ml substrate buffer (0.05 M citric acid, 0.05 M Na_2HPO_4, pH 5) and then 50 µl of 30 % hydrogen peroxide (v/v)was added. To each of the wells, 150 µl of the OPD/substrate solution was added, developed for 30 min and the absorbance read at 450 nm with a plate reading spectrophotometer. Alternatively, the OPD reaction was stopped with 100 µl stop solution (2 M H_2SO_4) and the absorbance read at 490 nm.

2.17 Overproduction of recombinant *C. difficile* 630 LMW-SLP

E. coli Rosetta containing pET 28a pLMW1-262-1, was kindly donated by Neil Fairweather, Imperial College London and was grown aerobically, O/N at 37°C in either LB agar or LB broth, both with 50 mg/l kanamycin and 12.5 mg/l chloramphenicol (kan/chl).

For large scale expression, 20 ml LB broth, 50 mg/l kanamycin and 12.5 mg/l chloramphenicol was inoculated with *E. coli* Rosetta pLMW1-262-1 and grown O/N at 37°C with constant agitation at 200 x rpm. This 20 ml culture was used to inoculate 1 l LB broth kan/chl and grown at 30°C until OD_{600} 0.6. Protein expression was induced with the addition of IPTG to a final concentration of 1 mM and the culture was grown for 5 h at 30°C.

To determine solubility of protein 1 ml aliquots were removed pre-induction and hourly upon induction. The cells were pelleted, the supernatant was removed and the cells were resuspended in 1ml 25 mM Tris/HCl, 200 mM NaCl, pH 7.5. DNase (Sigma) and protease inhibitor cocktail (Sigma) was added to both the resuspended cells and the supernatant before the cells were lysed following the sonication method. Both the lysed cells and supernatant were analysed via SDS-PAGE to determine the location of the majority of the expressed protein.

2.18 Immobilised metal affinity chromatography (IMAC)

IMAC was used to purify recombinant *C. difficile* 630, LMW SLP. The recombinant LMW SLP protein was labelled with a 6 x His-tag and therefore, purification was performed with a gravity column of affinity resin with bound bivalent nickel. The *E. coli* Rosetta containing pET 28a pLMW1-262-1 cell supernatant was harvested at 10000 x *g* for 20 min at 4^{o}C. The nickel column was washed with equilibration buffer (25 mM Tris/HCl, 200 mM NaCl, 5 mM imidazole, pH 7.5) and the cell supernatant was loaded on to the column at a 1:1 dilution with equilibration buffer. The column was then washed with wash buffer (25 mM Tris/HCl, 200 mM NaCl, 20 mM imidazole pH 7.5) and elution was performed stepwise with additions of imidazole (50 – 300 mM imidazole) in elution buffer (25 mM Tris/HCl, 200 mM NaCl, x mM imidazole pH 7.5). The protein containing fraction was dialysed into 10 mM HEPES pH 7.5, 150 mM NaCl.

2.19 Sonication of bacterial cells

DNase and protease inhibitor cocktail (both Sigma) were added to the cell solution pre-lysis and sonication was performed on ice. Cells were disrupted with a digital sonifier (Branson) using 10 % increments in sonication power (10 % - 40 %) for 30 s on and 1 min off cycles.

2.20 Fluorochrome labelling of bacteria

C. difficile strains and the three closely related species were labelled with the following method. Bacteria were grown to exponential phase OD_{600} 0.5, 3 ml was washed with 0.1 M PBS and fixed in 4 % paraformaldehyde for 20 min at RT. The cells were washed 3 x PBS and to quench the remaining PFA, the pellet was resuspended in 1ml of 20 mM NH_4Cl_2 for 15 min. The cells were pelleted and again washed in 3 x PBS. The pellet was resuspended in 0.5 ml PBS, 2 % BSA and 5 µg/ml mAb at RT for 1 h. Cells were washed 3 x PBS and resuspended in 0.5 ml PBS, 2 % BSA and 1 drop Alexa Fluor 594, goat anti-mouse IgG (life technologies) and incubated for 30 min at RT in the dark. Cells were washed 3 x PBS and resuspended in 100 µl PBS and for IF microscopy or 500 µl for flow cytometry. *C. difficile* controls were produced following the same protocol however either the primary mAb or secondary antibody incubation step was removed.

2.21 Immunofluorescent microscopy

Glass slides were prepared with 1 % (w/v) agarose in sterile water and 500 µl molten agarose was pipetted onto a Hendley-Essex multispot microscope slide with a plain glass microscope slide (VWR) placed on top to spread out the agarose, which was allowed to solidify for 2 min. The top slide was slid off to reveal a smooth layer of agarose and 1 µl of fixed immunofluorescent cells added to each spot and covered with a cover slip. Microscopy was performed using a Nikon M200 inverted microscope and Nikon 100x 1.30 oil objective coupled to a Photometrics CoolSNAP HQ CCD camera. The system was controlled and images acquired with MetaMorph v7.7.80 software, aided using the Nikon Ultimate Focus plugin. Phase contrast images were obtained using an exposure time of 100 ms, and 1000 ms for fluorescence images with ET-mCherry filter (Chroma 49008 ET560/40x).

2.22 Flow Cytometry

Fluorochrome labelled cells were produced as above and analysed using a Fortessa X20 flow cytometer with a Yellow/Green laser 561 nm, band pass filter 610/20. A laser beam with an optimal wavelength of 590 nm is required for excitation of Alexa Fluor 594, which upon excitation, emits maximum fluorescence at 617 nm. The operational software package that was used with the Fortessa X20 flow cytometer was BD FACSDiva. For each sample 100,000 to 1,000,000 events were recorded, and a dot plot display of forward scatter (FSC) *versus* side scatter (SSC) was used to gate the negative sample of just fixed cells with no fluorescence.

FlowJo software version 7.6 was used to analyse the flow cytometry data. Briefly, the data was viewed using a histogram plot with the 561_610/20 fluorescence on the x-axis. The event count was normalised when overlaying results to display the % of max on the y-axis. To achieve the % of max, the fluorescence value which has the largest cell count, for each sample, is changed to 100 % and therefore, samples containing different numbers of cells, when normalised, all have a peak at 100 % of max (Treister and Roederer, 2015). This normalisation allows the fluorescence intensity and distribution of cell percentage across this intensity, for each sample, to be easily compared.

The median and percentage robust coefficient of variance (% rCV) were calculated using the FlowJo analysis software.

The median is the relative intensity value below which 50 % of the events are found; i.e., it is the 50th percentile. In general, the median is a more robust estimator of the central tendency of a population than the mean.

The %rCV = 100 * 1/2(Intensity[at 84.13 percentile] – Intensity [at 15.87 percentile]) / Median. The robust CV is not as skewed by outlying values as the CV.

The flow cytometry results were subjected to statistical analysis, see section 2.1.28.

2.23 Immunogold labelling and transmission electron microscopy

C. difficile and *C. sordellii* cultures were grown O/N to OD600 0.5 and 1.5ml was removed, washed in 0.1 M PBS and fixed in 1 ml 4 % paraformaldehyde (PFA) at RT. The cells were washed 3 x PBS and to quench the remaining PFA, and the pellet was resuspended in 1ml of 20 mM NH_4Cl_2 for 15 min. The cells were pelleted and again washed 3 x PBS. The pellet from each tube was resuspended in 200 µl H_2O and immediately 5 µl of cell suspension was added to on 200 mesh, carbon-coated, plasma etched grid for 5 min before excess liquid was removed with filter paper, by touching edge for 10 s. Grids were then rinse in 10 µl PBS 3 x 5 min before blocking with 10 µl normal goat serum 1:10 in PBS, 1 % BSA for 30 min. Rinse again with 10 µl PBS 3 x 5 min before incubating with 10 µl *C. difficile* mAb, diluted 1:25 in PBS, 1 % BSA for 1 h, RT. Grids were rinsed in 10 µl PBS 3 x 5 min before adding 10 µl anti-mouse IgG, 9-11 nm gold conjugated (Sigma), diluted 1:20 in PBS, 1 % BSA and incubating for 1 h at RT. Rinse grids 4 x 5 min with PBS and then 5 x 5 min with H_2O, allow grids to air dry and negatively stain with 1.5 % phosphotungstic acid. The grids were viewed using a Philips CM100 TEM with Compustage and high resolution digital image capture, at varying magnifications.

2.24 Dot Blot

Dots were marked on to Amersham hybond PVDF blotting membrane (GE Healthcare) in pencil, to mark where to add the analyte. The membrane was activated upon submersion in methanol for 15 s and equilibration in transfer buffer (20 mM Tris, 192 mM glycine, 0.1 %

SDS, 20 % methanol). Serial dilutions of either whole cells or low-pH glycine extracted SLPs were added in 6 – 10 µl dots and left to dry for a minimum of 2 h. The serial dilutions were prepared in PBS and were prepared as the following undiluted, 1/2, 1/4, 1/100 and 1/1000 depending on the experiment. The primary antibody that was being tested or purified SLPs were bound directly to the membrane and used as positive controls. The membrane was blocked in with blocking buffer (0.1 M PBS, 5 % non-fat milk powder and 0.05 % tween-20) for 1 h. The membrane was then incubated at RT for 45 min with the *C. difficile* mAbs, at 2.5 – 10 µg/ml, diluted in blocking buffer. The membrane was then washed with 3 x PBS-T for 5 min each. The membrane was then incubated at RT for 30 min with peroxidase-conjugated goat anti-mouse IgG (F_c specific) (Sigma), diluted 1:120,000 in blocking buffer. The membrane was washed for 15 min with PBS-T and then twice for 5 min with PBS only. For visualisation of antibody binding, the membrane was treated with the Enhanced Chemiluminescence (ECL) kit from Amersham, GE Healthcare, as per manufacturer's instructions. The positive signals were viewed using an ImageQuant LAS4000mini biomolecular imager (GE Healthcare) and processed with the accompanying software and ImageJ.

2.25 Size exclusion chromatography to purify the SLP-complex

The SLPs from *C. difficile* 630 were extracted using the low-pH glycine technique (Calabi *et al.*, 2001) The LMW/HMW SLP complex was purified from the SLPs using a Superdex200 HiLoad 16/600 size exclusion column (GE Healthcare) and an Äkta Prime$^+$ system (GE Healthcare). All solutions were filtered (0.45 µm) and degassed. The column was washed with 1.5 x volume H_2O prior to equilibration with 1.5 x volume 10 mM HEPES pH 7.5, 150 mM NaCl. The flow rate used was 1 ml/min with fractions of 1 ml or 2 ml collected and analysed by SDS-PAGE for protein content and purity and then BCA assay for concentration.

2.26 Microscale thermophoresis (MST)

MST was used to analyse the interactions of the mAbs and the purified *C. difficile* 630 LMW/HMW SLP complex. The protocol was carried out as per manufacturer's instructions and as described in (Jerabek-Willemsen *et al.*, 2011) using a Monolith NT.115TM series MST machine.

The purified SLP complex was fluorescently tagged, as per manufacturer's (Nanotemper) instructions, using the amine reactive dye (NT-647 N-hydroxysuccinimide (NHS)). Before beginning the experiment, the concentration of fluorescent protein, LED power and capillary coating (standard, hydrophobic, hydrophilic, and premium) was optimised to provide the required 200-1500 fluorescent units. The LED power used was 90 % and the capillary coating was standard. The unlabelled mAb was serially diluted 1:1, 16 times, in 0.1 M PBS with 0.05 % Tween, to a total of 10 µl, before 10 µl of fluorescent dye was added, taking into account the dilution effect of combining the labelled and unlabelled proteins. An initial 'Cap scan' was completed to measure the fluorescence of each sample and determine any anomalous samples. Experiments were performed using a Monolith NT.115TM series, MST machine and the resulting temperature jump and subsequent thermophoresis data was used to trace unlabelled ligand concentration against normalised fluorescence trace (FNorm). The dissociation constant (K_d) was estimated using the Nanotemper Analysis software.

2.27 Lateral flow strip production

Lateral flow strip, pore size approx. 15um (Sartorius) was added to a support strip of card. A control line, 1 mg/ml rabbit anti-mouse IgG (Sigma), diluted in 50mM carbonate/bicarbonate buffer, pH 9.6, was drawn on to the membrane using a 20 µl pipette tip. The HMW/LMW SLP complex was also made to 1 mg/ml in carbonate buffer and the 'test' line was drawn below the control. The membrane was left to dry for a minimum of 2 h. The strip was then blocked for 15 min at RT in 1 % non-fat milk powder (w/v) and 1 % empigen (v/v) (Sigma) with gentle agitation. A strip of absorbent pad was added to the top of the strip, at the end where the control line is and a section was left at the bottom with no membrane as an area to load the sample.

2.28 Gold conjugation of IgG

The antibody was diluted to 15 µg/ml in 5mM TES buffer, 1 % SDS (w/v), 5 mM EDTA and 10 mM Tris-HCl, pH 7.4. A 1 ml aliquot of 20nm (BBI solutions) was centrifuged for 20 min at 13,000 rpm. The supernatant was removed and the gold pellet was resuspended in 1 ml of the pre-prepared antibody solution and incubated at RT for 2 h before adding 100 µl 200 mg/ml BSA in dH20, vortexed and incubated for a 1.5 h. The sample was then centrifuged

for 20 min at 13,000 rpm. The supernatant was removed and the pellet resuspended in X ml PBS, pH 7.4 with 0.05 % Tween (v/v) and 2 % BSA (w/v) until and OD_{280} of around 5 was achieved.

2.29 Statistical analysis

Statistical analysis of results was performed using the, IBM SPSS version 21. Several tests were employed depending on the data. Means were compared using a One-way ANOVA with Bonferroni and Tukey post hoc tests. The flow cytometry data was analysed using a Mann-Whitney test and a Wilcoxon Signed Ranks test. A statistical significance value of $P = < 0.05$ was used throughout the analysis.

Chapter 3 Bioinformatic approaches to bacterial biomarker identification and selection

3.1 Introduction

In the context of this project a biomarker was defined as a putative epitope that was unique to a group of bacterial strains, known as the group of interest (GOI), and had the potential to act as a diagnostic target. A novel bioinformatics approach was utilised that identifies unique biomarkers using existing genome sequence data. The aim of this bioinformatics approach was to identify proteins, or protein fragments that could be used as epitopes for the development of diagnostic reagents such as antibodies or aptamers however, it could also be used to differentiate strains within species or identify particular traits of interest. The approach builds on previous work from the AptamemsID project in which cloud computing workflows were developed for the identification of peptide signatures as putative biomarkers (Flanagan *et al.*, 2014). Identification of these peptide signatures required exploitation of all fully annotated, bacterial genomes contained within RefSeq (Pruitt *et al.*, 2002).

There are more than 1400 species of microorganisms that cause disease within humans (Taylor *et al.*, 2001; Ecker *et al.*, 2008) and 2764 organisms in NCBI RefSeq with fully sequenced genomes, each on average encoding approximately 4000 proteins (Brocchieri and Karlin, 2005). To determine which protein or fragment is unique to the GOI, each of these proteins must be compared to all of the other proteins from all other genomes. This all-against-all comparison is an extremely computationally demanding task and to achieve the results in a reasonable time frame requires a parallel computing approach or the use of cloud computing. Although the bioinformatics approach could be used to determine biomarkers for any organism with a fully annotated genome within RefSeq, the focus of this project was to perform the analysis using *C. difficile* as a case study. Only biomarkers that were predicted to be surface associated were selected in order to aid the rapidity of any related PoN detection sensor. The surface association of the target would provide accessibility to the sensor and would negate the need for time-consuming pre-treatment of cells.

3.1.1 The IDRIS system

The bioinformatics workflow used within this project utilised a system, from here on termed the IDRIS system, which was developed by Keith Flanagan[6] and colleagues as part of the AptaMemsID and i-sense projects. This system performed a comparative analysis of the predicted proteomes of all complete genome sequences within the NCBI RefSeq database. Initially the IDRIS system utilised RefSeq release 47 (2011/05) and then automatically updated the system with each new fully annotated genome added to RefSeq. Two cloud computing workflows, ApID1 and ApID2 (Flanagan *et al.*, 2014), are utilised within the IDRIS system and were created using the computing platform Microbase[7]. A diagram of the entire IDRIS system is provided in Figure 3-2. The initial workflow, ApID1 splits the translated amino acid sequences from these RefSeq genomes in to 15 amino acid strings, termed tokens. The 'tokenisation' of the amino acid sequences is performed by a sliding window that moves across the annotated genomes, one amino acid at a time, depositing each of these tokens into the TokenDB (Figure 3-1).

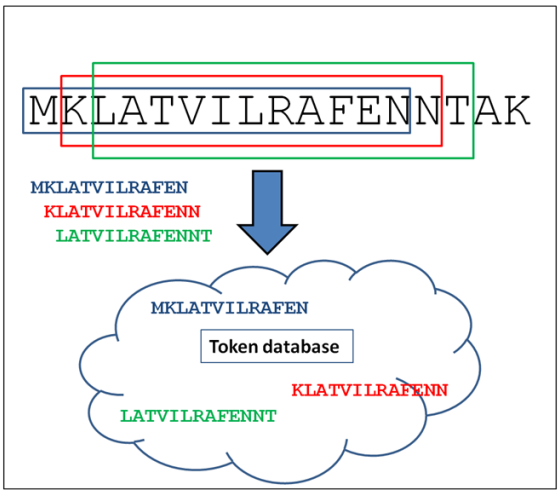


Figure 3-1. 'Tokenisation' of the amino acid sequences from the NCBI RefSeq database . A 15 amino acid sliding window moves step wise, one amino acid at a time, across the translated sequences, forming tokens that are then deposited into the TokenDB.

Additionally, ApID1 uses the annotated protein sequences and the subcellular localisation tools, SignalP, LipoP, PSORTb and TMHMM (which are described further in section 4.1.2) to

[6] ICOS, School of Computing Science, Newcastle University

[7] http://www.microbasecloud.com/

predict the subcellular localisation of each protein. The results from these tools are deposited into the subcellular localisation database (Figure 3-2).

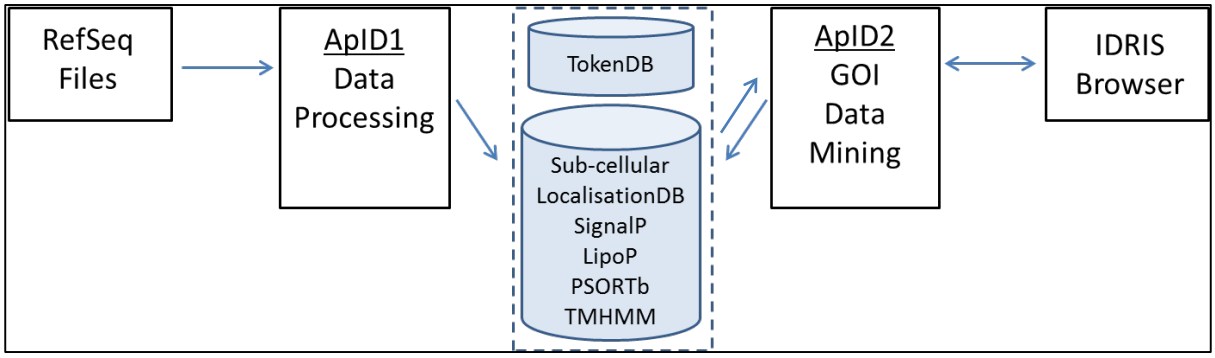


Figure 3-2. The IDRIS system . All complete bacterial genomes within the NCBI RefSeq database are used as the input for the IDRIS system. The ApID1 data processing workflow takes the amino acid sequences and splits them into 15 amino acid tokens, depositing them in to the TokenDB. ApID1 also predicts the subcellular localisation of the translated proteins, using the subcellular localisation tools, SignalP, LipoP, PSORTb and TMHMM. The results from these tools are placed into the subcellular localisation database. A GOI is created within the ApID2 data mining workflow, and the workflow searches the TokenDB for tokens that are common to all bacterial strains within this GOI. The common tokens are then searched for similarity against the remaining tokens within the TokenDB. Only tokens that are unique to the GOI but are found within all strains of the GOI are provided by ApID2 as results. When the tokens that are unique to the GOI are identified, ApID2 then mines the subcellular localisation database for the protein data that is associated to the relevant tokens.

Within the second workflow, termed ApID2, a bacterial GOI is created and the TokenDB is mined for tokens that are common to all of the strains within this GOI. ApID2 then searches the TokenDB for these common tokens, which are also unique to the GOI. That is, the system checks that the token is not located within the predicted proteins from any other bacteria, based on the genome sequence, except those that are within the GOI. The TokenDB uses an exact sequence matching approach to determine whether any two tokens are the same or different. Consequently, a token annotated by the system as unique, may still exist in other organisms with amino but with one or more amino acid changes. In the worst case, a token listed as unique to a particular GOI may only differ by a single amino acid to another token in an organism outside of that GOI. As a result, the manual analysis, within the bioinformatics approach, was necessary to determine the tokens that could be used as a *C. difficile* biomarker.

An overview of how ApID2 performs this search has been provided in Figure 3-3, with *C. difficile* as the GOI. When the tokens that are unique to the GOI are identified, ApID2 then mines the subcellular localisation databases for the data for the token associated proteins (Figure 3-2).

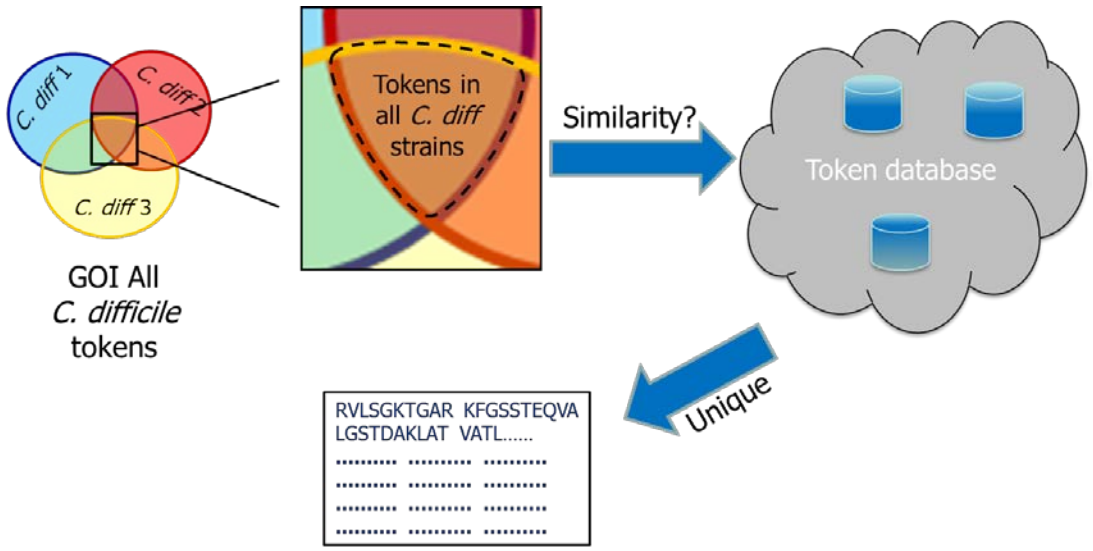


Figure 3-3. An example of ApID2 searching the TokenDB for a specified group of interest . The set of tokens common to all organisms defined in the GOI (in this case *C. difficile*) is found. The token database is then searched for each of these 'common' tokens to determine whether they occur in any other organism. If a given token is found in an organism outside those defined in the GOI, then it is filtered from the result list since it is not unique to the GOI. The tokens that are present in all members of the GOI, and do not have an exact match outside the GOI are classified as 'GOI-unique' and returned to the user for further analysis.

The 2015 version of the IDRIS system integrates a Web-based browser – termed the IDRIS browser. The browser provides a simple graphical user interface, which can be used to query the IDRIS system to find peptide tokens, or complete proteins, that are unique to a given group of bacterial genomes of interest and to view the cellular location prediction for that protein.

In this study, the original 2012 IDRIS system was used by Keith Flanagan to predict unique tokens whose parent proteins could be surface associated. The updated 2015 IDRIS system, including Web-browser, was then used within the results of this project to update the results and ensure that any unique biomarkers were still valid.

The 2012 IDRIS system contained 888,314,351 distinct tokens from around 1400 organisms, three of which were *C. difficile* strains. Whereas, the updated 2015 system held 1,350,122,042 distinct tokens, from 2764 organisms, nine of which were *C. difficile* strains. This huge amount of additional data in the 2015 IDRIS system, highlights the need for the validation of the initially identified tokens. The results of the original *C. difficile* search, performed by Keith Flanagan, presented 9793 protein IDs that contained at least one unique token. These proteins were used in the downstream subcellular localisation workflow.

3.1.2 Predicting protein localisation within the cell

Computational tools to predict the protein structure and function of sequences play a major part in microbiological research, providing the ability to rapidly identify cell-surface-exposed proteins and thus enable the discovery of diagnostic biomarkers (Gardy and Brinkman, 2006). Owing to the importance of this analysis, many protein subcellular localisation tools have been developed over the past 25 years (Restrepo-Montoya *et al.*, 2009).

The subcellular localisation prediction within this project used a number of tools (SignalP, LipoP, PSORTb and TMHMM), which were automatically performed with ApID2 and manually checked. There were 9793 proteins that were predicted using the IDRIS system to contain tokens unique to *C. difficile* and the subcellular localisation results of each protein were provided in a spreadsheet. The results were filtered via the following options, SignalP *D- score* ≥ 0.45, LipoP *SPI*, PSORTb final localisation - cell wall and TMHMM ≥ 1 transmembrane region count. Once the proteins had been analysed by the localisation tools and were filtered via the above scores, the 9793 proteins containing possible tokens were reduced to 28 that were positive for all four localisation tools (Appendices Table 9-1). Therefore, these 28 proteins were classed as exposed on the surface of *C. difficile*. These 28 proteins were manually checked using the web interfaces for each tool and more information of each tool is given below.

3.1.2.1 SignalP

A signal peptide is an N-terminal amino acid sequence that controls entry into the secretory pathway and determines protein localisation either within organelles integrated into the cell membrane, or secreted (Nielsen *et al.*, 1997). Signal peptides are usually 15-40 amino acids in length and are cleaved during translocation across the membrane (Nielsen, 1998). SignalP (Petersen, 2011) is a tool that is used to predict the presence of a signal peptide region within proteins. The proteins identified as containing unique tokens were analysed using SignalP version 4.1, with the Gram-positive organism group selected. The *C- score* is the raw cleavage score and produces a 'high' value to mark the first amino acid after the signal peptide has been cleaved - the first amino acid of the mature protein. The *S- score* is highest at amino acid positions that SignalP predicts are part of the signal peptide and low at the

amino acid positions forecast to be within the mature protein. The *Y- score* is an average of the *C- score* and the slope of the *S- score*, providing a better cleavage site prediction than the *C- score* alone. The *D- score* is a weighted average of the mean *S-* and maximum *Y- scores* and was used to discriminate between signal peptides and non-signal peptides. The presence of a signal peptide was determined by a *D- score* ≥ 0.45, and it was the *D- score* alone that was used as the deciding factor for continuing investigation with proteins or disregarding them.

3.1.2.2 LipoP

LipoP discriminates between lipoprotein signal peptides, other signal peptides and N-terminal membrane helices by analysing the protein sequence for particular sites. LipoP (Juncker *et al.*, 2003) was designed for use with protein sequences from Gram-negative bacteria. However, it has also been shown to perform well with sequences from Gram-positive bacteria, identifying 92.9% of lipoproteins in a Gram-positive test set (Rahman, 2008). The cleavage site for signal peptidase type I *SpI* indicates a signal peptide is present, whereas signal peptidase type II *SpII* indicates a lipoprotein. Bacterial lipoproteins are a set of membrane proteins that have many different functions (Kovacs-Simon *et al.*, 2011) but are anchored to the cell membrane (Juncker *et al.*, 2003). Therefore, LipoP (version 1.0) was used to determine the proteins that had a cleavage site for *SpI.*

3.1.2.3 PSORTb

PSORTb version 3.0.2 (Yu *et al.*, 2010), incorporates several algorithms to predict the transmembrane helices and the cellular position of proteins. For Gram-positive bacteria these localisation predictions can be cytoplasmic, cytoplasmic membrane, cell wall or extracellular. A final prediction of cell wall localisation was selected for this project.

3.1.2.4 TMHMM

TMHMM (Transmembrane hidden Markov model) is a membrane protein topology prediction method based on a hidden Markov model (Krogh *et al.*, 2001). TMHMM predicts the number of transmembrane helices within a protein, providing evidence of its location.

3.1.2.5 Phobius

Although not part of the automated IDRIS system, Phobius was executed manually as a verification step of the localisation tools. Phobius is a combined transmembrane topology and signal peptide predictor (Kall *et al.*, 2004) and was used to verify the presence of a signal peptide and the predicted non-cytoplasmic localisation of the protein.

3.1.3 Surface Layer type clades

A project led by Dr Kate Dingle, Nuffield Department of Clinical Medicine, used large-scale, whole genome sequencing to determine the diversity and evolution of the *SlpA* gene within 57 *C. difficile* strains. Using the *SlpA* sequences, the researchers were able to group strains within the *C. difficile* species into 13 surface layer type (SLT) clades (Dingle *et al.*, 2013). Currently, all sequenced *C. difficile* strains fall in to these 13 clades which therefore, represent the known SlpA diversity within the *C. difficile* species.

3.1.4 The Pathosystems Resource Integration Centre

The Pathosystems Resource Integration Centre (PATRIC) provides bioinformatics data on bacterial pathogens (Wattam *et al.*, 2014), including phylogenetic information about *C. difficile* and closely related organisms. The phylogenetic trees displayed on the PATRIC website[8] are constructed with an automated pipeline that utilises the amino acid sequences from each genome. Genomes can be submitted to PATRIC via NCBI GenBank, RefSeq, or collaborator request. The PATRIC automated pipeline identifies homologous proteins in a two round process. In the first round, the annotated proteins of a single genome from each distinct species is selected and are searched against each other using the alignment tool BLAT[9] and the top hits are clustered with MCL[10], defining the initial seed sets for the homology groups. In the second round, the seed sets are aligned using MUSCLE[11] and

[8] https://www.patricbrc.org/portal/portal/patric/Home

[9] http://genome.ucsc.edu/FAQ/FAQblat.html

[10] http://www.micans.org/mcl/

[11] http://www.ebi.ac.uk/Tools/msa/muscle/

hidden Markov models (HMMs) are built with hmmbuild[12]. All genomes are searched with hmmsearch[13], the results of which are used to define the homology groups that are aligned with MUSCLE and poorly aligned groups are removed with Gblocks[14]. The main tree is estimated using FastTree15, with random samples of 50% of the homology groups from the MUSCLE alignment. 100 of these 50% homology group trees are made using FastTree and the support values on the main tree indicate the number of times a particular branch was observed in the support trees. PATRIC was used within this project to determine the species that are most similar to *C. difficile*.

3.2 Bioinformatics methods

3.2.1 The bioinformatics approach used to predict unique, surface associated *C. difficile* biomarkers

The manual section of the bioinformatics approach (Figure 3-4) was performed within this project and details are to follow. The initial two steps of the approach were automated within the IDRIS system, which was developed and performed by Dr Keith Flanagan[16]. More detail about the IDRIS system can be found in Chapter 5. The bioinformatics approach was followed in order to predict biomarkers that were unique to *C. difficile*, exposed on the surface of the bacteria and expressed in all stages of organism growth.

[12] http://www.csb.yale.edu/userguides/seq/hmmer/docs/node19.html

[13] http://www.csb.yale.edu/userguides/seq/hmmer/docs/node26.html

[14] http://molevol.cmima.csic.es/castresana/Gblocks.html

[15] http://www.microbesonline.org/fasttree/

[16] ICOS, School of Computing Science, Newcastle University

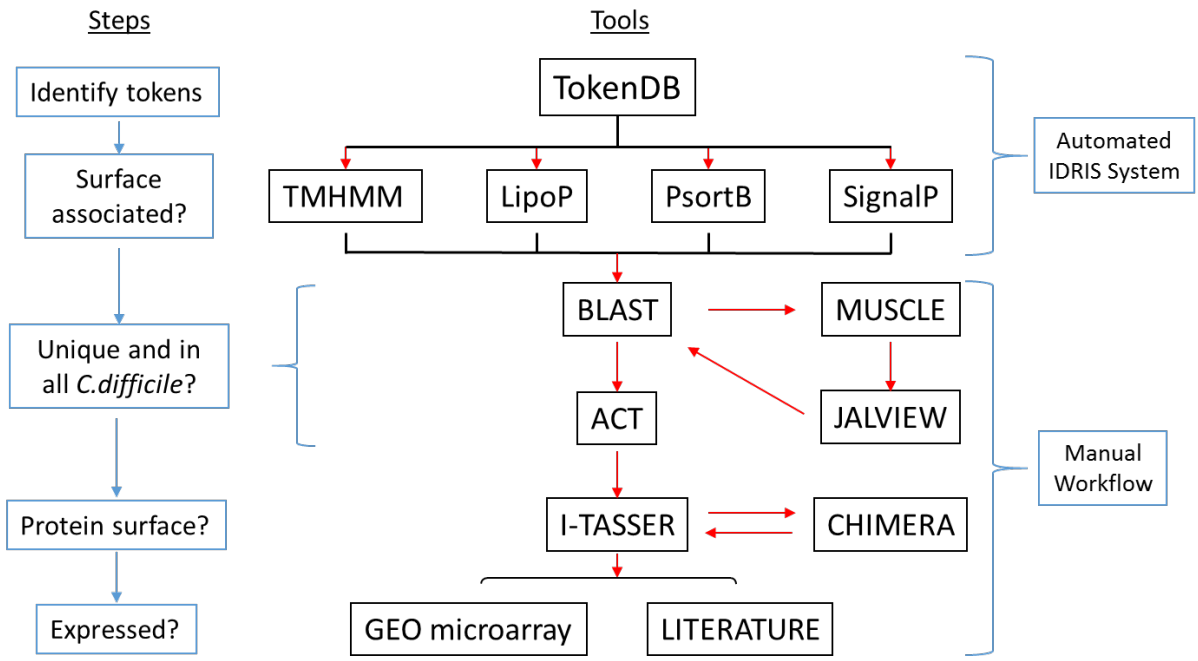


Figure 3-4. The bioinformatics approach consists of an automated part and a manual part. The automated section is performed by ApID2 within the IDRIS system and uses TokenDB to identify bacterial tokens that are unique to a specified organism group of interest. The protein sequences containing the identified tokens are then analysed by a number of subcellular localisation prediction tools again using ApID2. For the work presented here, only the proteins indicated to be surface-associated are of interest. A manual workflow containing bioinformatics analysis steps are then performed using a variety of tools to determine a biomarker within the protein that is unique to and yet common to all strains within the chosen group. The biomarker position within the protein is then identified before searching for evidence of the expression of the protein.

3.2.2 Manual subcellular localisation of proteins

Although the subcellular protein localisation tools were performed within the automated IDRIS system manual confirmation of some of the IDRIS system results was also performed. In addition to the tools used within the approach, Phobius, a combined transmembrane topology and signal peptide predictor was also used to analyse the protein sequences. The tools were applied via their relevant web pages to the individual protein sequences in FASTA format.

The options selected for each tool are provided within Table 3-1. When the option was given, Gram-positive was selected and also output with graphics was chosen. The results were filtered via the following options, SignalP *D- score* ≥ 0.45, LipoP *SPI*, PSORTb final localisation - cell wall and TMHMM ≥ 1 transmembrane region count.

Table 3-1. The five tools used to predict the subcellular localisation of proteins , performed manually. The table contains the options chosen when available for the input sequence and the output.

Tool	Gram	Output format	Organism
SignalP v 4.1	Gram-positive	Standard	
LipoP 1.0 server		Standard	
PSORTb v. 3.0.2	Gram-positive	Normal	Bacteria
TMHMM v2.0		Extensive with graphics	
Phobius		Long with graphics	

3.2.3 Identifying unique protein regions using BLAST, MUSCLE and Jalview

The identification of the unique biomarkers, within the 'unique token' containing proteins provided by the IDRIS system, was performed manually. The IDRIS system uses an exact sequence matching approach to determine whether any two tokens are the same or different. Consequently, a token annotated by the system as unique, may not be *wholly* unique. In the worst case, a token listed as unique to a particular GOI may only differ by a single amino acid to another token in an organism outside of that GOI. As a result, manual analysis was necessary to determine the biomarkers that were at least 15 amino acids in length and had a minimum of four amino acids difference from the closest related sequence.

To locate the biomarkers that met the minimum four amino acid threshold, the individual protein sequences were searched for similarity to other proteins within UniProt – release 15.0 using BLAST (Uniprot-Consortium., 2012). The sequences of the proteins with the most similarity were downloaded in their FASTA format, for SlpA the most similar 22 proteins were used. The sequences were then aligned using EBI MUSCLE version 3.5 (Edgar, 2004) and the resulting sequence alignment was viewed using Jalview version 2.6.1 (Waterhouse *et al.*, 2009). Within Jalview, the amino acids coloured via 'percentage identity' with the regions with >80 % similarity coloured in dark purple and the regions with no similarity are coloured in white. Sequences of 15 amino acids or more, which were identical in all of the *C. difficile* proteins and not present in any other bacterium, were manually identified and selected for downstream analysis.

Jalview was also used to predict the hydrophobicity of the amino acids within any conserved sequences. The colour selection used was 'hydrophobicity*',* which uses colours hydrophobic and hydrophilic amino acids, red and blue respectively. Unique regions with hydrophilic sections of 5 continuous amino acids or more were retained for further analysis.

The regions that were predicted to be unique and contain hydrophilic sections, were entered into NCBI BLASTP 2.2.26 (Altschul *et al.*, 1990) and searched for similarity against all organisms within NCBI's non-redundant DB. Only sequences retaining the minimum five amino acid difference from other sequences were continued as promising biomarkers. BLASTP searches were often performed with the *C. difficile* species excluded from the search. This exclusion enabled the closest related sequence that was not located within *C. difficile* to be identified.

The similarity searches performed with NCBI BLASTP were also used to produce phylogenetic trees of the SlpA protein by selecting the top 15 hits and the 'distance tree of results' on the BLASTP results page. The following phylogenetic tree options were used to construct the tree:- 'Fast Minimum Evolution' was selected for '*Tree method*', '0.85' for '*Max Seq Difference*', 'Grishin (protein)' for '*Distance*' and finally a 'slanted cladogram' was chosen for '*Layout'*.

3.2.4 Identify biomarkers that are conserved across *C. difficile*

The sequences for all complete *C. difficile* genomes (Table 3-2) were downloaded from NCBI in FASTA format and searched with BLASTN for similarities, to create comparison files. The genome files were entered into the Artemis Comparison Tool (ACT) (Carver *et al.*, 2005) with one designated as the reference sequence and one as the subject sequence. The comparison files were also loaded into the ACT, which was then used to display the comparisons of the *C. difficile* genomes. ACT provides a six-frame translation of the DNA sequence, and displays pairwise comparisons between two or more genome sequences, which can be used to identify regions of similarity. However, since there were nine *C. difficile* strains, the ACT viewing window was not large enough to display the translated amino acid sequences. Instead, the biomarker was searched within the amino acid sequences and highlighted within the DNA sequences of the comparisons, as this allowed

space for the visualisation of conservation. The red bands represent the forward matches and blue bands the reverse matches between the genomes, with the intensity of the colour being proportional to the percent identity of the match. White areas are displayed where there is no BLAST similarity present and therefore indicate variable regions. A searched sequence that is matched turns yellow and if the sequence is conserved throughout the strains, all will display the same yellow sequence (Carver *et al.*, 2005).

Table 3-2. The *C. difficile* strains contained within the NCBI database, which have fully sequenced and annotated genomes.

C. difficile strain	NCBI Taxon ID
630	272563
2007855	699033
BI1	699034
CD196	645462
R20291	645463
CF5	699036
M120	699035
M68	699037
ATCC 42355	499175

3.2.5 Identifying the biomarker position within the protein

The following method was used to identify the biomarkers that were exposed on the protein surface. Initially a search of the protein data bank, RSCB PDB[17] was performed to identify if it contains any structures for the proteins. For this project, the protein structures were not solved and therefore predictions of protein structures had to be made by sending the protein sequence to I-TASSER protein structure and function predictor (Zhang, 2008). I-TASSER uses LOMETS (Wu and Zhang, 2007) and SPICKER (Zhang and Skolnick, 2004) to construct five full-length atomic models of the sequence.

These I-TASSER models were then observed using the UCSF Chimera software package that was downloaded onto a local machine. Chimera was developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311) (Pettersen *et al.*, 2004). The view was selected to show the protein surface and the biomarker was coloured red on the resulting images.

[17] www.rcsb.org

3.2.6 Determining expression of the biomarker containing protein

A thorough literature search is one of the most effective ways to determine if a protein of interest is in fact expressed. The protein literature search was performed using NCBI Pubmed[18] and Science Direct[19], with protein expression at different stages of growth being a key feature of the positive proteins.

Another method to evidence protein expression was using microarray data downloaded from the Gene Expression Omnibus (GEO) (Edgar *et al.*, 2002). The interpreted dataset was for *C. difficile* 630, GSE25474 – 'Comparison of the expression profiles of 630E strain after 4h and 10 h of growth' and GSE25475 - 'Comparison of the expression profiles of 630E strain and a sigH mutant after 10 h of growth' (Saujet *et al.*, 2011). These gene expression profiles were made under laboratory conditions, with *C. difficile* 630 grown anaerobically in brain heart infusion broth for different growth periods. The expression values were downloaded from GEO and entered into an Excel spreadsheet. The values were ordered from high to low, with a total of 15744 unique probe results for each microarray chip. Therefore, position 1 in the spreadsheet is the gene with the highest expression value and position 15744 in the spreadsheet is the gene with the lowest expression value. The positions of the genes of interest, within the ordered results, were noted and their percentage position within the total gene number was calculated. For example, if the gene of interest was in position 221 then (221/15744)*100 would provide a percentage of 1.4% showing that the gene is expressed within the top 1.4 % of all genes under those particular experimental conditions.

3.2.7 Identifying protein domains

The *C. difficile* SlpA sequence was submitted to InterProScan (Jones *et al.*, 2014) version 5, via the Web page, using the default settings. The sequence was analysed by InterProScan by comparing it to the protein signatures and predictive models contained within the InterPro Consortium (Mitchell *et al.*, 2015).

[18] http://www.ncbi.nlm.nih.gov/pubmed

[19] http://www.sciencedirect.com/

3.2.8 Identifying species that are closely related to *C. difficile*

The SlpA sequence was analysed using BLASTP and a 'distance tree of results' was produced on the BLASTP results page using the top 15 hits. The following phylogenetic tree options were used to construct the tree:- 'Fast Minimum Evolution' was selected for '*Tree method*', '0.85' for '*Max Seq Difference*', 'Grishin (protein)' for '*Distance*' and finally a 'slanted cladogram' was chosen for '*Layout'*.

3.2.9 Average Nucleotide Identity calculator

The average nucleotide identity (ANI) calculator[20] (Rodriguez-R and Konstantinidis, 2014) estimates the average nucleotide identity between two genomes to determine the likelihood that the sequences are from the same species or not. If the ANI value is above 95 % then the sequences are predicted to be from the same species. The identity is predicted using both one-way ANI and two-way ANI between two genomic datasets, as calculated by Goris and colleagues, 2007. The FASTA sequences of bacterial genomes were downloaded from NCBI and uploaded via the Web page to the ANI calculator, two sequences at a time.

3.2.10 Aligning bacterial genomes to view regions of conservation, insertion and deletion

The complete *C. difficile* genomes of strains 630 and R20291 were used as reference genomes for the alignment of contigs present in *C. sordellii* draft genomes. The sequences were aligned with CONTIGuator 2 (Goris *et al.*, 2007) by uploading the *C. difficile* FASTA sequence into the 'Reference file' input and the *C. sordellii* genomes into the 'Contig file' input. The criteria for the 'Contigs profiling' was kept at the default settings.

[20] http://enve-omics.ce.gatech.edu/ani/

3.3 Bioinformatics Results

The bioinformatics approach (Figure 3.4) was developed within this project and used to determine biomarkers that were predicted to be accessible from the surface of whole *C. difficile* cells. The results of each step will be described in detail within this section.

3.3.1 Predicting protein localisation within the cell

Once the 9793, token-containing, *C. difficile* proteins were identified within the TokenDB, the next ApID2 automated step was the downstream analysis of the location of these proteins within the cell 'Surface associated?' step (Figure 3.4). Only proteins that were surface-associated were deemed as positive and manually checked using the web interface for the localisation tools. Throughout this section, the surface layer protein SlpA (NCBI Accession YP_001089306) was used as an example of a positive prediction, as this protein was determined to contain the most promising biomarker. In addition to the positive SlpA result, an example of a negative result for each tool is also provided.

3.3.1.1 SignalP

For SlpA (Figure 3-5 A), all signal peptide scores predict the cleavage site to be between position 24 and 25. Therefore, the signal peptide was 24 amino acids in length and position 25 was the first amino acid of the mature protein. Figure 3-5 also shows an example of a negative SignalP result from the analysis of the protein sodium:amino acid symporter, GenBank accession YP_003214646 (Figure 3-5 B). All scores for this protein are low and therefore none of the amino acids were predicted to be involved with a signal peptide.

The SlpA *D- score* was 0.783, indicating the presence of a signal peptide. The *D- score* for sodium:amino acid symporter was 0.146 therefore, no signal peptide was predicted and the protein was disregarded from further analysis.

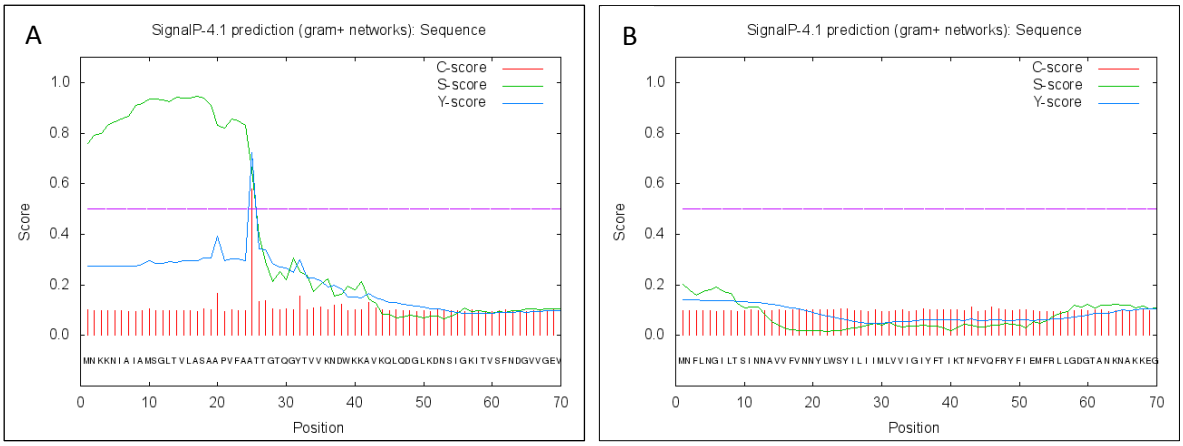


Figure 3-5. SignalP analysis of *C. difficile* 630 SlpA YP_001089306 (positive example, A) and Sodium:amino acid symporter YP_003214646 (negative example, B). The scores for SlpA determine that a signal peptide exists, which is cleaved between amino acids 24 and 25. The *C- score* (red) is the raw cleavage score and produces a high value to mark the first amino acid after the signal peptide has been cleaved, and therefore indicates the first amino acid of the mature protein. The *S- score* (green) is highest at amino acid positions that are predicted to be part of the signal peptide and low at the amino acid positions forecast to be within the mature protein. The *Y- score* (blue) is an average of the *C- score* and the slope of the *S- score*, providing a better cleavage site prediction than the *C- score* alone. In contrast, all scores for 'Sodium:amino acid symporter' are low, indicating that no signal peptide has been predicted.

3.3.1.2 LipoP

Figure 3-6 (A) shows the LipoP prediction of the *SpI* site for the *C. difficile* SlpA protein. The proteins that were deemed negative either had no putative cleavage sites and were predicted to be cytoplasmic proteins, or they were classed as having a lipoprotein signal peptidase *SPII* such as the *C. difficile* stage II sporulation protein D (NCBI accession

YP_003213187.1) (Figure 3-6 B). All proteins with *SpI* were progressed to the next localisation tool.

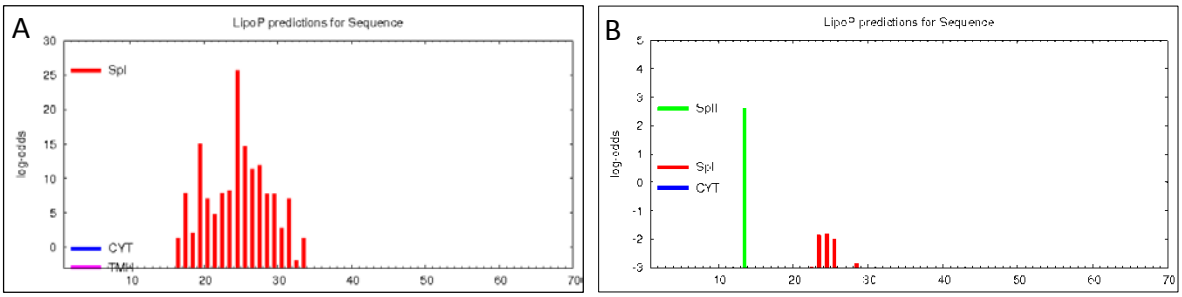


Figure 3-6. LipoP predictions of *C. difficile* 630 SlpA (A) and stage II sporulation protein D, YP_003213187.1 (B). SlpA is predicted to have a signal peptide *SpI* cleavage site between amino acid 24 and 25, indicated by the red line. The stage II sporulation protein D, is predicted to have a *SpII* cleavage site.

3.3.1.3 PSORTb

PSORTb predicted that SlpA is non-cytoplasmic and is cell wall associated, predicting the surface localisation of the protein (Figure 3-7, top). In contrast, the tool predicted the ABC transporter permease, NCBI Accession YP_001088460, to be within the cytoplasmic membrane, with 8 membrane helices calculated (Figure 3-7, bottom). Only proteins that were predicted to be cell wall associated were used for further analysis.

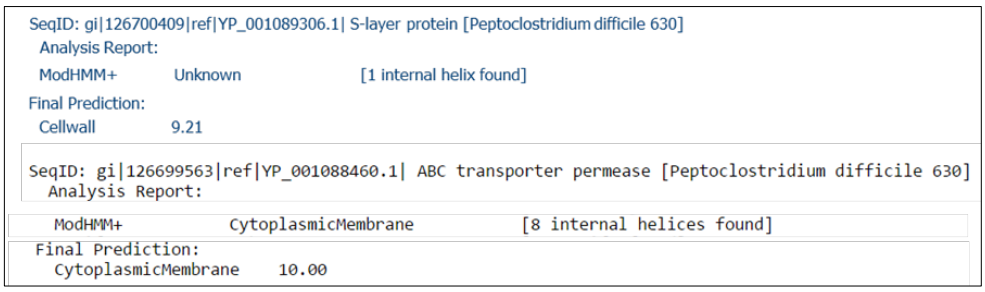


Figure 3-7. PSORTb analysis of *C. difficile* 630 SlpA (top blue) and ABC transporter permease (bottom black). SlpA is predicted to be localised to the cell wall and the ABC transporter permease to be located within the cytoplasmic membrane.

3.3.1.4 TMHMM

The TMHMM Server v. 2.0 result for SlpA predicts that there is a single transmembrane domain at the N-terminus (Figure 3-8 A), which corresponds to the signal peptide that remains within the membrane once it has been cleaved. The TMHMM result for the 50S ribosomal protein L15, (YP_003213150) indicated the absence of any transmembrane helices (Figure 3-8 B). A minimum of one transmembrane region was deemed to be a positive result for the purposes of this project. Proteins with zero transmembrane regions were disregarded.

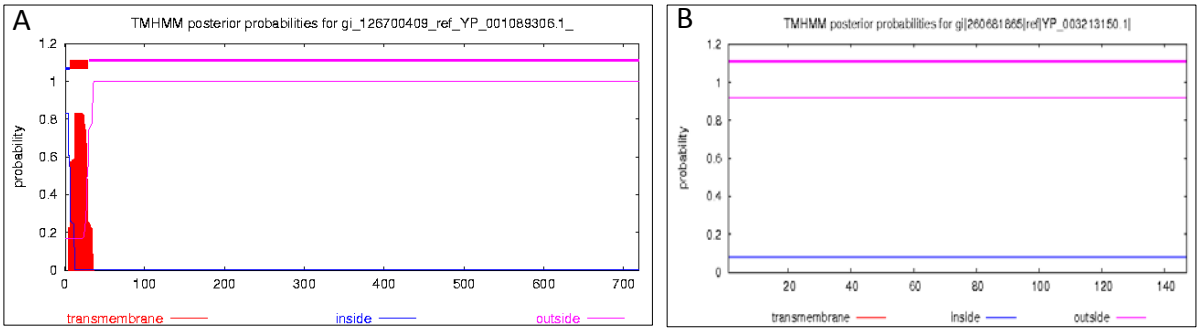


Figure 3-8. TMHMM predictions of *C. difficile* 630 SlpA (A) and ribosomal protein L15, YP_003213150 (B). SlpA is predicted to have one initial transmembrane region, indicated by the red signal at the beginning of the sequence. Most of the sequence is predicted to be outside of the membrane, indicated by the pink line. The result for YP_003213150, 50S ribosomal protein L15, predicts zero transmembrane regions on the protein.

3.3.1.5 Phobius

Phobius was used as a verification step for the 28 proteins identified as surface associated by the IDRIS system. The Phobius results for SlpA (Figure 3-9), verified the results from the surface prediction pipeline by confirming a signal peptide, with the remainder of the protein being non-cytoplasmic.

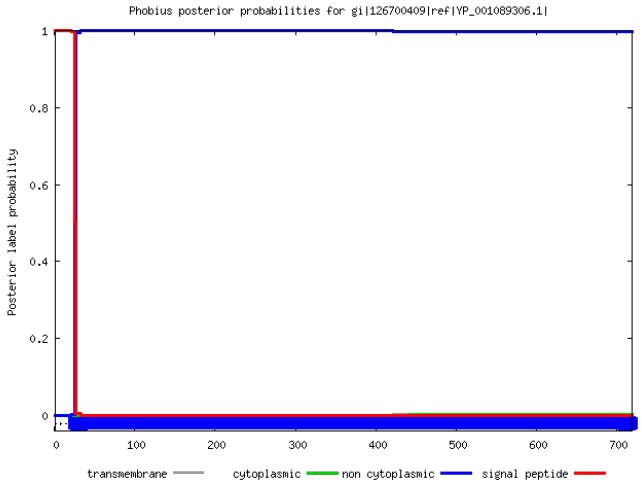


Figure 3-9. Phobius results for SlpA . A signal peptide (red line) is predicted at the N-terminal and the remainder of the protein is predicted to be non-cytoplasmic (blue line). The x-axis displays the number of amino acids and the y-axis is the probability that those amino acids are the predicted localisation. The first few amino acids have a signal peptide with a probability of one, which is the highest score, then reduces to zero indicating the end of the signal peptide. The non-cytoplasmic signal is also predicted at a probability of one for the remainder of the sequence.

3.3.2 Identifying unique biomarkers with a minimum of 15 amino acids

3.3.2.1 Basic Local Alignment Search Tool

To determine which of the 28 proteins contained a token that was suitable for sensitive detection, the Basic Local Alignment Search Tool (BLAST) was used to identify tokens that differed from the closest relative by a minimum of four amino acids. A BLAST search for the SlpA protein from *C. difficile* SlpA was carried against the Uniprot database (28/06/12) (Figure 3-10). The results with the greatest similarity to the SlpA protein were all from *C. difficile*, with the green colour indicating a high percentage identity of matched sequences to the subject.

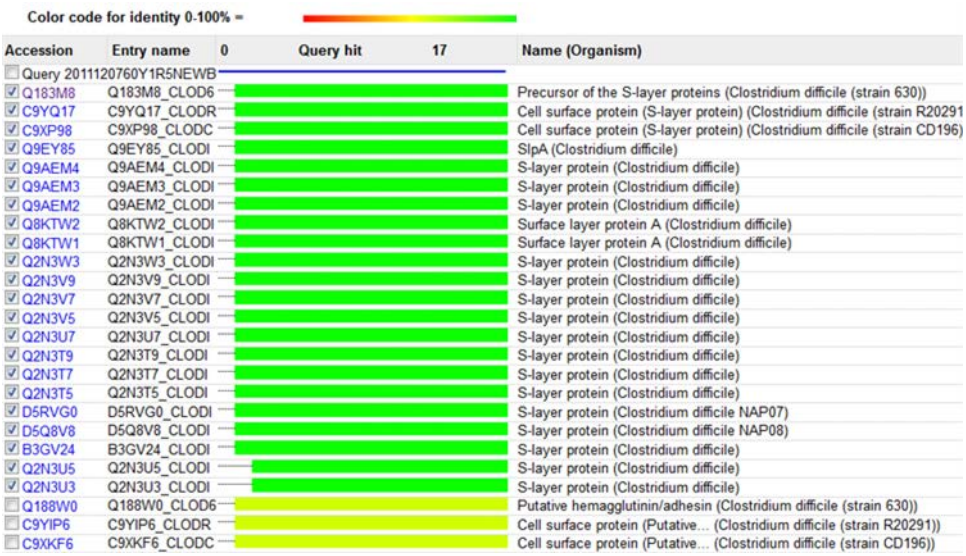


Figure 3-10. Uniprot BLAST of *C. difficile* 630 SlpA . The first column on the left is the protein accession ID, the second is the entry name, which includes the accession ID and associated species ID. The Query hit provides a visual means of recognising the identity of the subject and query sequence, where green indicates very high identity and red indicates low identity. The far right column provides the name of the protein and the organism in which it is located. All of the top hits (dark green) are within *C. difficile* strains. The related protein sequence was downloaded 28/07/12.

3.3.2.2 Viewing the alignments of proteins

The SlpA alignment (Figure 3-11) clearly shows a possible biomarker, outlined by the black box, which was 16 amino acids in length and conserved across all 22 *C. difficile* proteins. Using the colouring options within Jalview, the amino acids were coloured via their hydrophobicity, with blue depicting hydrophilic and red representing hydrophobic, 10 out of 16 amino acids within the predicted biomarker were coloured blue/purple.

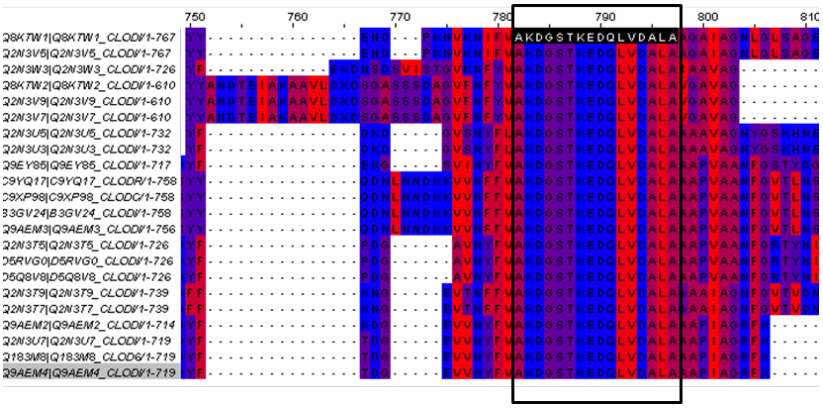


Figure 3-11. The SlpA protein alignment of the top 22 hits from the Uniprot BLAST against *C. difficile* 630 SlpA. The sequences were viewed using Jalview and the amino acids coloured according to hydrophobicity (blue is hydrophilic and red is hydrophobic). The conserved biomarker, AKDGSTKEDQLVDALA, is highlighted by the black rectangle.

3.3.2.3 Searching for the SlpA biomarker within all fully sequenced *C. difficile* genomes

The initial 2012 IDRIS system contained only three *C. difficile* strains 630, R20291 and CD196, which increased to nine as the complete genome sequences were added to NCBI RefSeq. To provide the evidence of the SlpA biomarker being common to all nine fully sequenced *C. difficile* strains, the genome sequences were compared using the Artemis Comparison Tool (ACT) (Carver *et al.*, 2005). The genomes were searched for the SlpA biomarker amino acid sequence, and the corresponding DNA sequences were highlighted in yellow. All nine *C. difficile* strains contained the SlpA biomarker (Figure 3-12). The SlpA biomarker is seen to be conserved across all the strains, with the sequence surrounding the SlpA biomarker displaying some insertions and deletions between strains.

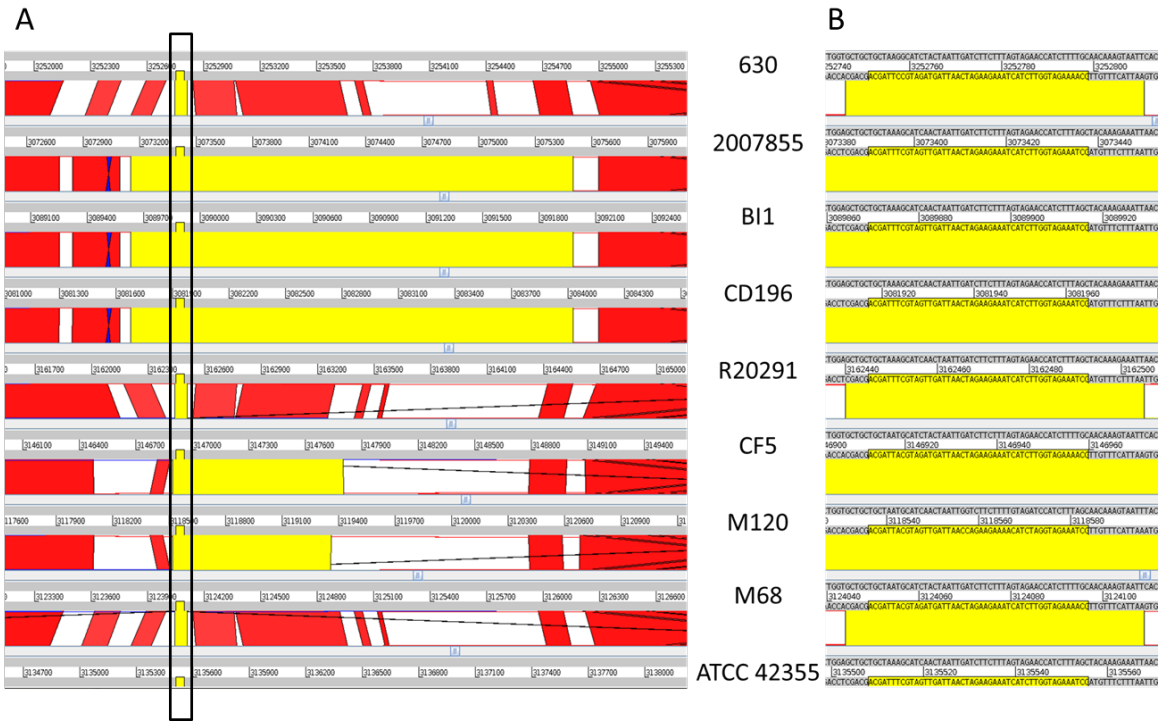
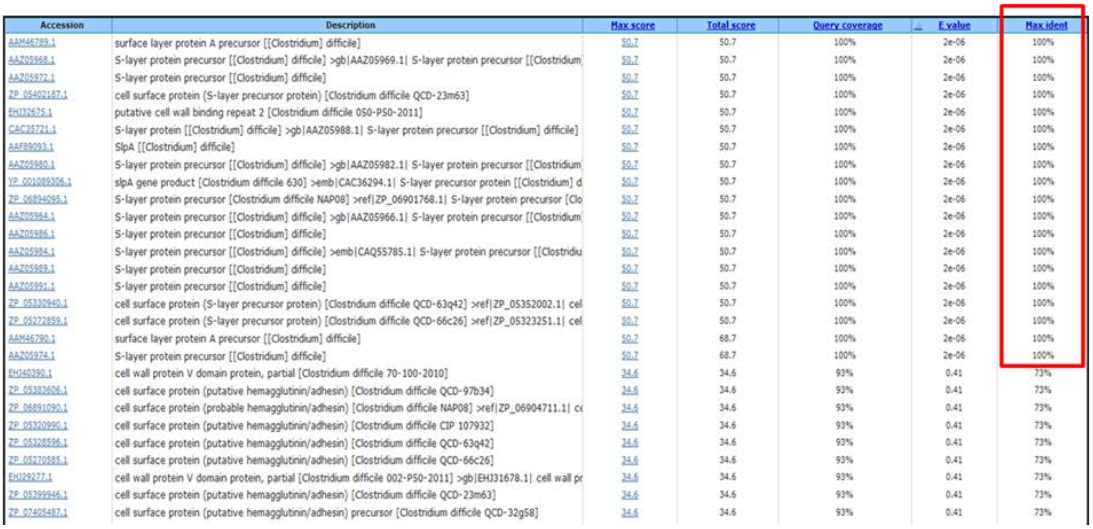


Figure 3-12. Screenshots taken from the Artemis Comparison Tool showing comparisons of nine *C. difficile* genomes (A), and a zoomed view of the region containing the biomarker (B). Nine different sequence tracks are shown (grey with coordinate bars), one for each strain present in the comparison. Between these tracks, a series of coloured bands represent forward-matched regions (red) and inversions (blue) of each pairwise comparison. The yellow bands indicate the currently selected pairwise match. In addition, a smaller yellow highlighted area can be seen protruding into the grey sequence tracks and surrounded by the black box indicating the SlpA biomarker (A). It is this biomarker that is in the highlighted region included within the zoomed-in view (B). The highlighted sequence within the zoomed view is the DNA sequence that codes for the SlpA biomarker. ACT was searched by inputting the amino acid sequence of the biomarker, which was subsequently reverse-translated and displayed on the corresponding DNA sequence.

3.3.2.4 Determining the uniqueness of the SlpA biomarker

The SlpA biomarker 'AKDGSTKEDQLVDALA' was searched for similarity using NCBI BLASTP, which searches the protein database using a protein sequence. The search was performed on 28/6/2012 and showed the biomarker located in its entirety (marked in red) within *C. difficile* only, with the top hits all being *C. difficile* proteins (Figure 3-13).



Accession	Description	Max score	Total score	Query coverage	E value	Max ident
AAM46789.1	surface layer protein A precursor [[Clostridium] difficile]	50.7	50.7	100%	2e-06	100%
AAZ05968.1	S-layer protein precursor [[Clostridium] difficile] >gb|AAZ05969.1| S-layer protein precursor [[Clostridium	50.7	50.7	100%	2e-06	100%
AAZ05972.1	S-layer protein precursor [[Clostridium] difficile]	50.7	50.7	100%	2e-06	100%
ZP_05402187.1	cell surface protein (S-layer precursor protein) [Clostridium difficile QCD-23m63]	50.7	50.7	100%	2e-06	100%
EHJ32675.1	putative cell wall binding repeat 2 [Clostridium difficile 050-P50-2011]	50.7	50.7	100%	2e-06	100%
CAC38721.1	S-layer protein [[Clostridium] difficile] >gb|AAZ05988.1| S-layer protein precursor [[Clostridium] difficile]	50.7	50.7	100%	2e-06	100%
AAF89093.1	SlpA [[Clostridium] difficile]	50.7	50.7	100%	2e-06	100%
AAZ05980.1	S-layer protein precursor [[Clostridium] difficile] >gb|AAZ05982.1| S-layer protein precursor [[Clostridium	50.7	50.7	100%	2e-06	100%
YP_001089306.1	slpA gene product [Clostridium difficile 630] >emb|CAC36294.1| S-layer precursor protein [[Clostridium] d	50.7	50.7	100%	2e-06	100%
ZP_06894085.1	S-layer protein precursor [Clostridium difficile NAP08] >ref|ZP_06901768.1| S-layer protein precursor [Clo	50.7	50.7	100%	2e-06	100%
AAZ05964.1	S-layer protein precursor [[Clostridium] difficile] >gb|AAZ05966.1| S-layer protein precursor [[Clostridium	50.7	50.7	100%	2e-06	100%
AAZ05986.1	S-layer protein precursor [[Clostridium] difficile]	50.7	50.7	100%	2e-06	100%
AAZ05984.1	S-layer protein precursor [[Clostridium] difficile] >emb|CAQ55785.1| S-layer protein precursor [[Clostridiu	50.7	50.7	100%	2e-06	100%
AAZ05989.1	S-layer protein precursor [[Clostridium] difficile]	50.7	50.7	100%	2e-06	100%
AAZ05991.1	S-layer protein precursor [[Clostridium] difficile]	50.7	50.7	100%	2e-06	100%
ZP_05330940.1	cell surface protein (S-layer precursor protein) [Clostridium difficile QCD-63q42] >ref|ZP_05352002.1| cel	50.7	50.7	100%	2e-06	100%
ZP_05272859.1	cell surface protein (S-layer precursor protein) [Clostridium difficile QCD-66c26] >ref|ZP_05323251.1| cel	50.7	50.7	100%	2e-06	100%
AAM46790.1	surface layer protein A precursor [[Clostridium] difficile]	50.7	68.7	100%	2e-06	100%
AAZ05974.1	S-layer protein precursor [[Clostridium] difficile]	50.7	68.7	100%	2e-06	100%
EHJ40290.1	cell wall protein V domain protein, partial [Clostridium difficile 70-100-2010]	34.6	34.6	93%	0.41	73%
ZP_05383606.1	cell surface protein (putative hemagglutinin/adhesin) [Clostridium difficile QCD-97b34]	34.6	34.6	93%	0.41	73%
ZP_06891090.1	cell surface protein (probable hemagglutinin/adhesin) [Clostridium difficile NAP08] >ref|ZP_06904711.1| c	34.6	34.6	93%	0.41	73%
ZP_05320990.1	cell surface protein (putative hemagglutinin/adhesin) [Clostridium difficile CIP 107932]	34.6	34.6	93%	0.41	73%
ZP_05328596.1	cell surface protein (putative hemagglutinin/adhesin) [Clostridium difficile QCD-63q42]	34.6	34.6	93%	0.41	73%
ZP_05270585.1	cell surface protein (putative hemagglutinin/adhesin) [Clostridium difficile QCD-66c26]	34.6	34.6	93%	0.41	73%
EHJ29277.1	cell wall protein V domain protein, partial [Clostridium difficile 002-P50-2011] >gb|EHJ31678.1| cell wall pr	34.6	34.6	93%	0.41	73%
ZP_05399946.1	cell surface protein (putative hemagglutinin/adhesin) [Clostridium difficile QCD-23m63]	34.6	34.6	93%	0.41	73%
ZP_07405487.1	cell surface protein (putative hemagglutinin/adhesin) precursor [Clostridium difficile QCD-32g58]	34.6	34.6	93%	0.41	73%

Figure 3-13. BLASTP search for similarity of the SlpA biomarker AKDGSTKEDQLVDALA to other protein sequences within the NCBI database (28/06/12). The biomarker was 100 % conserved within *C. difficile* strains.

A minimum difference of four amino acids between any possible biomarker and the closest related sequence, located in an organism not located within the GOI, was selected (the 'uniqueness threshold').The closest similarity to the SlpA biomarker, which was much further down the hits shown within Figure 3-13 and was not located within *C. difficile*, was found in the two-component LuxR family transcriptional regulator from the Gram-negative species *Novosphingobium nitrogenifigens* (*N. nitrogenifigens*) (Figure 3-14). There were five differences between the two amino acid sequences and therefore the SlpA biomarker was still classed as unique to *C. difficile*.

```
>ref|ZP_08209781.1|  two component LuxR family transcriptional regulator [Novosphingobium
nitrogenifigens DSM 19370]
 gb|EGD58158.1|  two component LuxR family transcriptional regulator [Novosphingobium
nitrogenifigens DSM 19370]
Length=208

 Score = 30.8 bits (65),  Expect = 6.3
 Identities = 11/15 (73%), Positives = 12/15 (80%), Gaps = 2/15 (13%)

Query  2    KDGSTKEDQLVDALA  16
            KDGS  ED+L DALA
Sbjct  103  KDGS--EDELLDALA  115
```

Figure 3-14. BLASTP results (28/06/12) of the protein with the closest similarity to the SlpA biomarker that was not from a *C. difficile* strain. The protein that was most similar was the two-component LuxR family transcriptional regulator from *Novosphingobium nitrogenifigens*.

The entire protein sequence of the two-component LuxR family transcriptional regulator was analysed using SignalP and was shown to be lacking a signal peptide, and therefore was not predicted to be transported out of the cytoplasm (Figure 3-15).

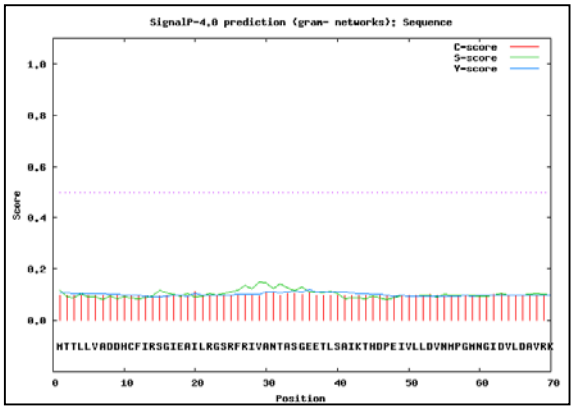


Figure 3-15. SignalP results of the two-component LuxR family transcriptional regulator protein sequence . The results show that there is no signal peptide predicted.

3.3.3 Structural prediction of SlpA and location of the biomarker

3.3.3.1 The SlpA domains

InterProScan (Jones *et al.*, 2014) was utilised to analyse the SlpA protein sequence. The InterPro Consortium employs several databases, including PROSITE (Sigrist *et al.*, 2002), HAMAP (Pedruzzi *et al.*, 2015) and SUPERFAMILY (Gough *et al.*, 2001), to gain the required protein signatures and predictive models. Figure 3-16 integrates the subcellular localisation results of an InterProScan version 5 analysis of *C. difficile* 630 SlpA, and the literature which states the LMW and HMW SLP cleavage position (Calabi *et al.*, 2001). The domains are mapped on to the SlpA sequence, with the biomarker also highlighted at its amino acid

position, which is predicted to be within a putative cell wall binding domain.

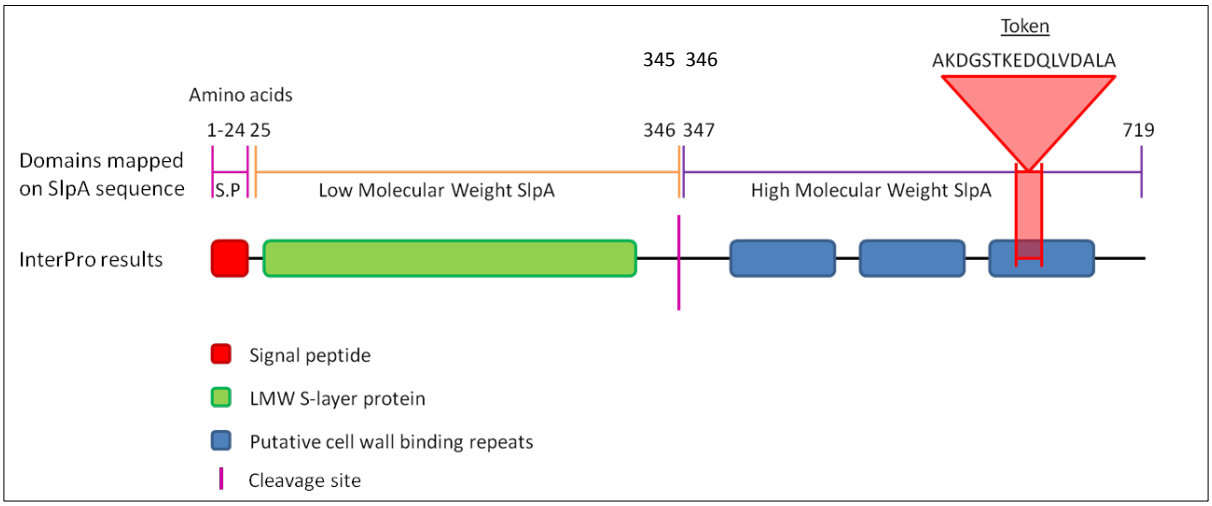


Figure 3-16. Combining the results from the surface localisation tools and an InterProScan of the sequence to show the domains and biomarker location on *C. difficile* 630 SlpA . The cleavage site of SlpA into the mature LMW AND HMW SLPs was taken from the literature.

3.3.3.2 Predicting the structure of the HMW SLP and the surface location of the biomarker

The hydrophobicity view within Jalview (Figure 3-18) was used to predict the largely hydrophilic biomarker location on the surface of the protein. However, further evidence of the location was required to show that the biomarker is accessible to potential PoN sensors. A search of the protein data bank, RSCB PDB[21] for SlpA structures revealed that the protein structure for only LMW SLP had been solved, PDB ID: 3CVZ (Fagan *et al.*, 2009). Neither the HMW SLP, which contains the biomarker, nor the whole complex was structurally solved and subsequently, tools to predict protein structure were used. The HMW portion of SlpA was sent to the I-TASSER protein structure and function predictor (Zhang, 2008), which constructs five full-length atomic models of the sequence. Each model has a confidence score (C-score) that estimates the quality of the prediction and usually ranges from -5 to 2. These models were then observed using the viewing platform UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311) (Pettersen *et al.*, 2004). The view was

[21] www.rcsb.org

selected to show the surface of the protein and the biomarker was coloured red on the resulting images. All five models predicted the biomarker to be surface exposed (Figure 3-17) and therefore if would likely be accessible to detection molecules. Model A had the greatest C-score with -2.68 and is therefore indicated to be the most likely structure of the HMW SLP. The C-scores of the remaining four models were as follows, -2.78 (B), -3.01 (C), -3.10 (D) and -4.55 (E).

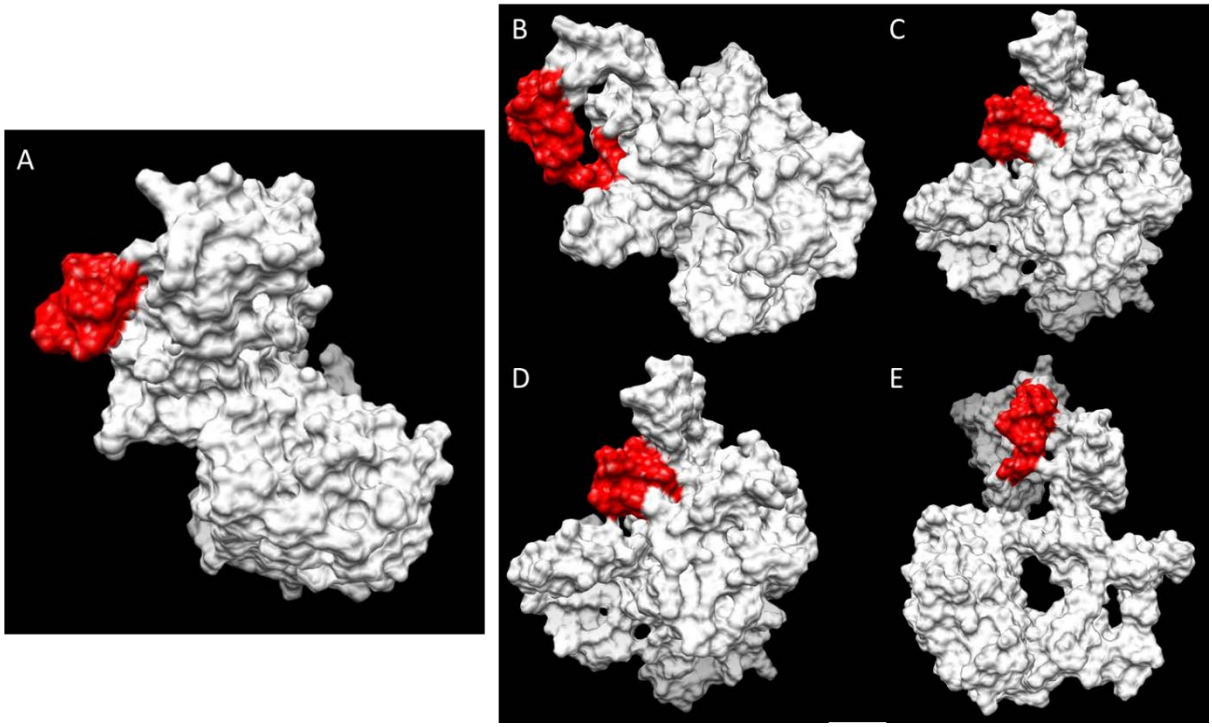


Figure 3-17. Chimera images of the five HMW SLP structures predicted using I-TASSER . The PDB files were loaded, the surface of the protein was displayed in white, and the biomarker sequence highlighted in red. All five models displayed the biomarker as being surface-accessible. Model A had the greatest confidence score of -2.68, closely followed by B with -2.78. The confidence scores for the remaining three models were -3.01 (C), -3.10 (D) and -4.55 (E).

3.3.3.3 Expression of SlpA

As discussed in Chapter 1 section 1.16, SlpA is post-translationally cleaved into the LMW and HMW SLPs which non-covalently bind to form the S-Layer, a paracrystalline layer that encapsulates the *C. difficile* cell. SlpA is described in the literature as one of the most highly expressed *C. difficile* proteins, within all stages of growth (Savariau-Lacomme, 2003; Fagan and Fairweather, 2011).

Microarray data for *C. difficile* was downloaded from the Gene Expression Omnibus (GEO) (Barrett *et al.*, 2011) and used to determine expression of the *slpA* gene. The dataset used

was GSE25474 – 'Comparison of the expression profiles of 630E strain after 4 h and 10 h of growth' (Saujet *et al.*, 2011) and GSE25475 - 'Comparison of the expression profiles of 630E strain and a SigH mutant after 10 h of growth', with all strains grown anaerobically in brain heart infusion (BHI) medium (Saujet *et al.*, 2011). SigH is an alternative sigma factor, which is a key element in the initiation of *C. difficile* sporulation. There were six microarrays in total, three different conditions with two dyes for each: cyanine dye Cy3 and Cy5. The dyes fluorescently label the *C. difficile* cDNA, which hybridizes to the DNA on the microarray and provides a quantifiable fluorescent signal. It is the fluorescence intensity that is measured to determine the expression of the particular gene. The results for the three conditions were very similar for both dyes. The *slpA* gene, named CD2793 in the data, was expressed in the top 6.4 % of genes for the 4 h growth, 1.2 % for 10 h growth and 2.4 % for 10 h growth of the SigH mutant. These results show that under differing conditions of growth stage and mutation, *slpA* remains one of the most highly expressed genes within *C. difficile*.

3.3.4 Surface Layer type clades

To enable additional analysis of the possible SlpA biomarker, Kate Dingle kindly sent sequences of the surface-layer type strains (SLTS) from each of the 13 groups. The *SlpA* gene from each SLTS and *C. difficile* 630 was translated into an amino acid sequence using Artemis genome browser, aligned with MUSCLE, with the resulting alignment viewed with Jalview. The amino acids were coloured via their percentage identity with dark purple defining >80 % identity across the sequences. The biomarker is denoted by the black box and is shown to be 100 % conserved across all 14 *C. difficile* strains (Figure 3-18).

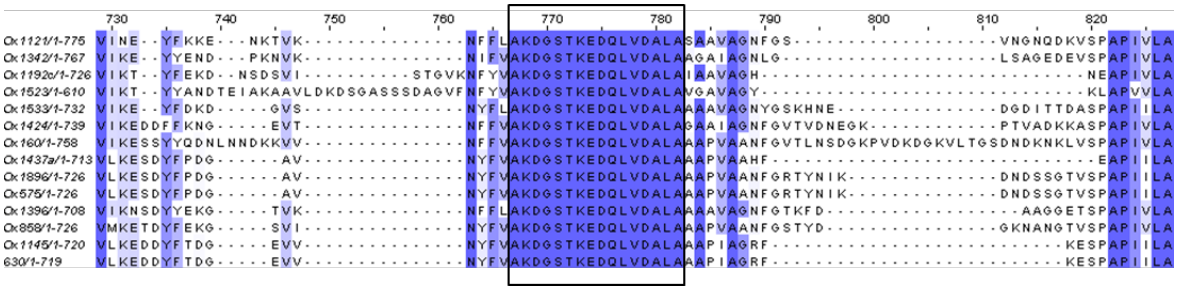


Figure 3-18. Jalview image of the alignments of SlpA sequences from the 13 SLTS and *C. difficile* 630 . Amino acids are coloured according to percentage identity, with the dark purple representing >80 % identity, purple >60 % and light purple >40 % sequence identity. The biomarker is denoted by the black box and displays 100 % sequence identity across all 14 *C. difficile* strains, SLTS Ox1121, Ox1342, Ox1192c, Ox1523, Ox1533, Ox1424, Ox160, Ox1437a, Ox1896, Ox575, Ox1396, Ox858, Ox1145 and the reference strain 630.

The sequences within Jalview were also explored to identify which of the 13 SLT clades *C. difficile* 630 belongs too. With only 11 out of the 719 amino acids being different, *C. difficile* 630 was most similar to Ox1145, SLT 7, displaying 98.5% identity (Figure 3-19).

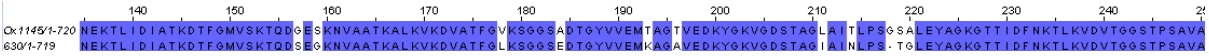


Figure 3-19. Jalview image of the 11 differences out of 719 amino acids in the SlpA sequences between Ox1145 (SLT clade 7) and *C. difficile* 630 . In the case of these two sequences the dark purple represents amino acids with 100 % identity and the white represents amino acids with no similarity.

3.3.5 Identifying species that are closely related to *C. difficile*

To be able to perform appropriate laboratory tests for the specificity of detection molecules produced against the SlpA biomarker, it was important to identify species that were closely related to *C. difficile*. These species could then be tested to ensure they do not provide false positives, the lack of which is vital for a PoN sensor. The closely related species are more likely to have similar protein sequences and perhaps comparable epitopes to which detection molecules may be more likely to non-specifically bind to.

The BLASTP results identified *Clostridium hiranonis* and *Peptostreptococcus anaerobius* as having protein sequences that were the most related to SlpA. Therefore, a 'distance tree of results' was produced on the BLASTP results page, using the top 15 hits. (Figure 3-20). Although, the *C. hiranonis* and *P. anaerobius* proteins were annotated as hypothetical, the protein search still highlights these strains as important to investigate further in future work.

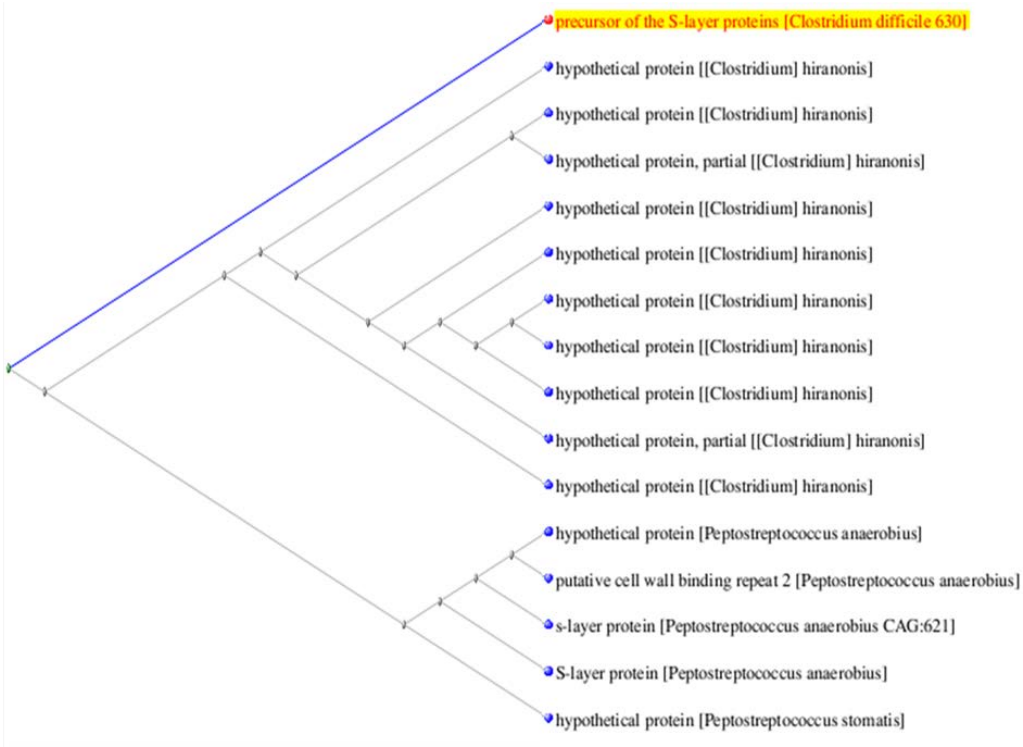


Figure 3-20. Slanted cladogram of BLASTP results from a *C. difficile* 630 SlpA similarity search, excluding *C. difficile* . The search was performed on 27/2/13 and the results were viewed with the 'Fast Minimum Evolution' for tree method, '0.85' for max seq difference and 'Grishin (protein)' for distance.

In addition to investigating the similarity of SlpA to other proteins, the similarity of full genome sequences to *C. difficile* was also explored using PATRIC. A screenshot of the phylogeny tree for Clostridium was taken on 28/2/2013, and displays the organisms that were closely related to *C. difficile* (Figure 3-21). The organisms with the ticks in the adjacent boxes were selected as the closely related species that required investigation within the laboratory work. The three were selected because *C. sordellii* was the species most closely related to *C. difficile* in PATRIC and both *C. hiranonis* and *P. anaerobius* were some of the top hits in the BLASTP search of the SlpA protein (Figure 3-20). Although *C. bartlettii* DSM 16795 was also shown to be a closely related species within PATRIC, this species was not identified within the BLASTP search of SlpA and therefore was not selected. However, if the time and resources are available then this species should also be included in future work.

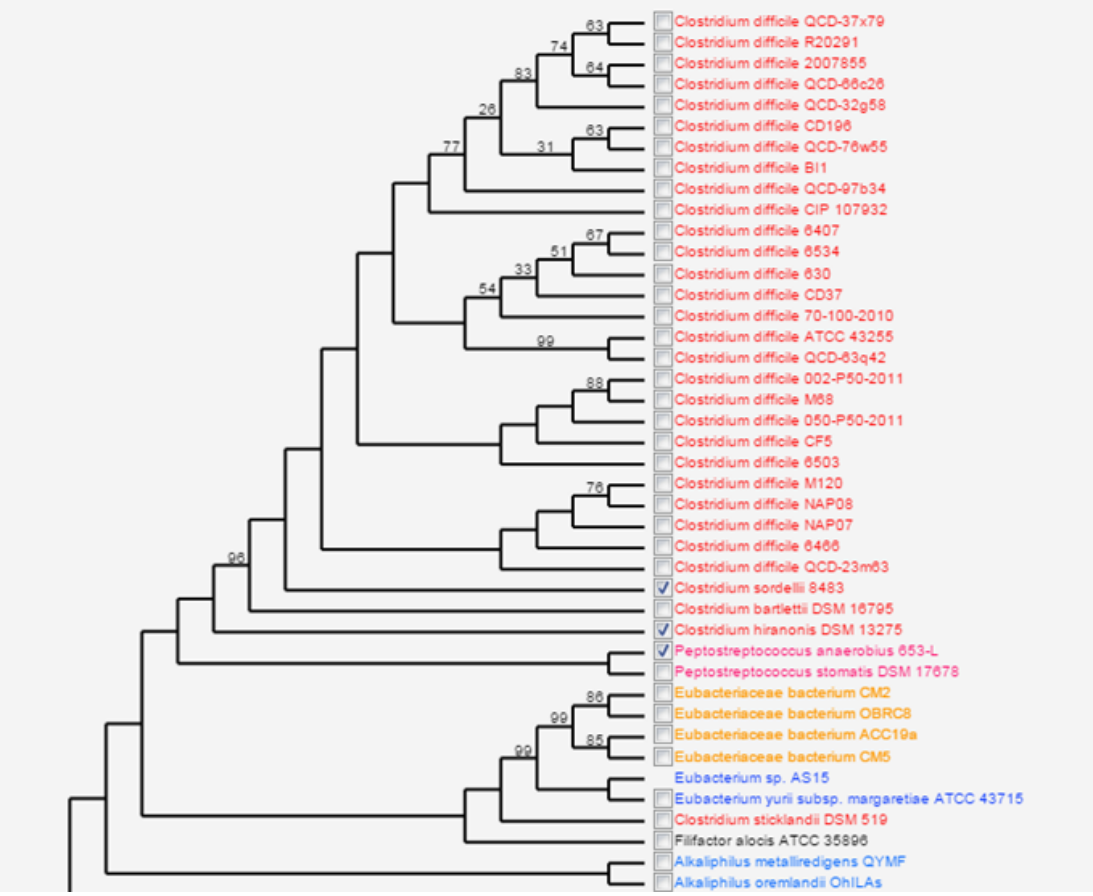


Figure 3-21. Phylogenetic tree of *C. difficile* and closely related species based on the amino acid sequence files from their genomes. Taken from PATRIC, The Pathosystems Resource Integration Centre (28/2/13).

3.3.5.1 Updating the search for SlpA sequence similarity

Since the initial bioinformatics work performed in July 2012, many more fully annotated genomes and sequence files were added to the NCBI and Uniprot databases over the duration of the project.. BLASTP searches were performed frequently throughout the project to determine whether the SlpA biomarker remained unique to *C. difficile*. A recent similarity search of the SlpA biomarker sequence was performed on 13/11/15, again using BLASTP and with *C. difficile* excluded from the search. When compared to the top 'non *C. difficile'* hit in 2012, which was the *N. nitrogenifigens* two-component LuxR family transcriptional regulator, the more recent search had 25 proteins with higher sequence identity to the biomarker across 14 different organisms (Figure 3-22). There were two identical hits displayed for *N. nitrogenifigens*, the two-component LuxR family transcriptional regulator and the DNA-binding response regulator, which are highlighted with a red box (Figure 3-20).

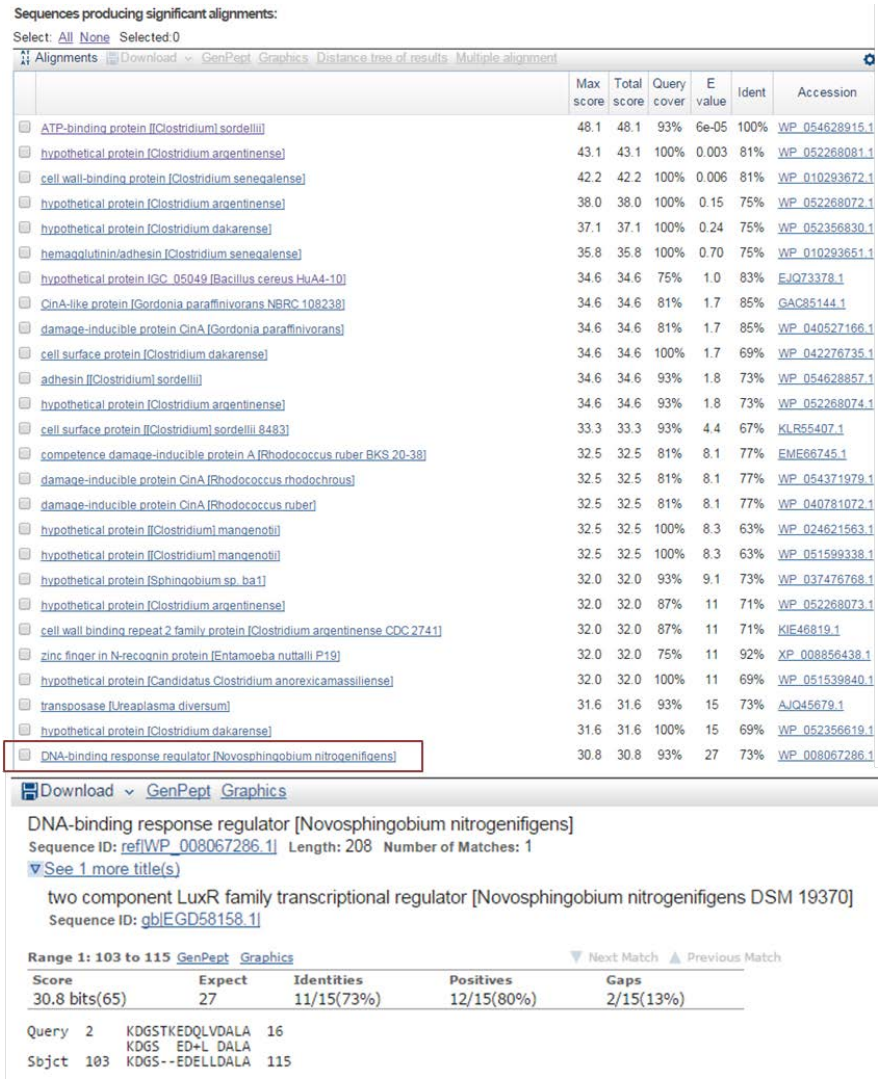


Figure 3-22. BLASTP search results for the SlpA biomarker with *C. difficile* excluded from the search (13/11/15) . The result for *N. nitrogenifigens*, highlighted with a red box, has been shown in more detail below the main results table.

The top three BLASTP results against the SlpA biomarker are shown in more detail in Figure 3-23. The proteins that contain sequences with the greatest similarity to the biomarker are the ATP binding protein and S-Layer protein from *C. sordellii* 8483 (Figure 3-23 A), which displays 100 % identity with 93 % query cover, that is 15 out of the 16 amino acids. There are several other clostridia that also contain protein sequences with high similarity to the SlpA biomarker: *C. argentinense* CDC 2741 (Figure 3-23 B) also had two hits, which was second only to *C. sordellii* 8483. The protein sequence with the third greatest similarity to the biomarker was within *C. senegalense* JC122 (Figure 3-23 C). As shown in Figure 3-23 there was only one amino acid difference between the biomarker and both *C. sordellii* proteins, ATP binding protein and S-Layer protein, three amino acid differences shown for

the *C. argentinense* hypothetical protein and *C. senegalense* cell wall-binding protein. These variances are below the minimum four amino acid difference that was selected as the biomarker '*uniqueness*' threshold. Therefore, any detection molecule chosen to bind to the SlpA biomarker may not have the necessary specificity and would require further laboratory analysis to confirm the binding specificity. However, it has been reported in several studies that antibodies and aptamers have extremely high specificities and even a single amino acid difference within the epitope is enough to diminish antibody/antigen binding (Zegers *et al.*, 1995; Zheng *et al.*, 1995; Negri *et al.*, 2012).

Figure 3-23. Top three BLASTP results for the SlpA biomarker, with *C. difficile* excluded from the search (13/11/15). The query is the SlpA biomarker sequence and the subject is the protein sequence that shows high similarity. The top hit is located within *C. sordellii* 8483 (A), the protein with the second greatest similarity is within *C. argentinense* CDC 2741 (B) and the third hit is from *C. senegalense* JC122 (C).

The subcellular localisation checks (the second step in the bioinformatics approach Figure 3-4) were also performed, using all five of the proteins containing the highly similar sequence to the SlpA biomarker. All five proteins were predicted using SignalP to have a signal peptide and were classified as being cell wall associated. The location indicates that the proteins may interfere with the detection molecules, designed against the biomarker,

and therefore will require further testing before the specificity of the molecules can be categorised.

3.3.5.2 Updating the search for closely related species

Due to the changes displayed by the BLASTP search results of the SlpA biomarker performed on 13/11/15 (Figure 3-22), an update of the BLASTP search with the entire SlpA protein was also performed on 13/11/2015. The results were displayed as a slanted cladogram, with the same options selected as described in section 3.3.5. This analysis revealed that *C. sordellii* 8483 S-layer protein was the closest related protein to *C. difficile* SlpA with both *C. hiranonis* and *P. anaerobius* also featuring (Figure 3-24). The *C. sordellii* 8483 genome was added to the NCBI database on 09/06/2015 and therefore wasn't included in the earlier bioinformatics work.

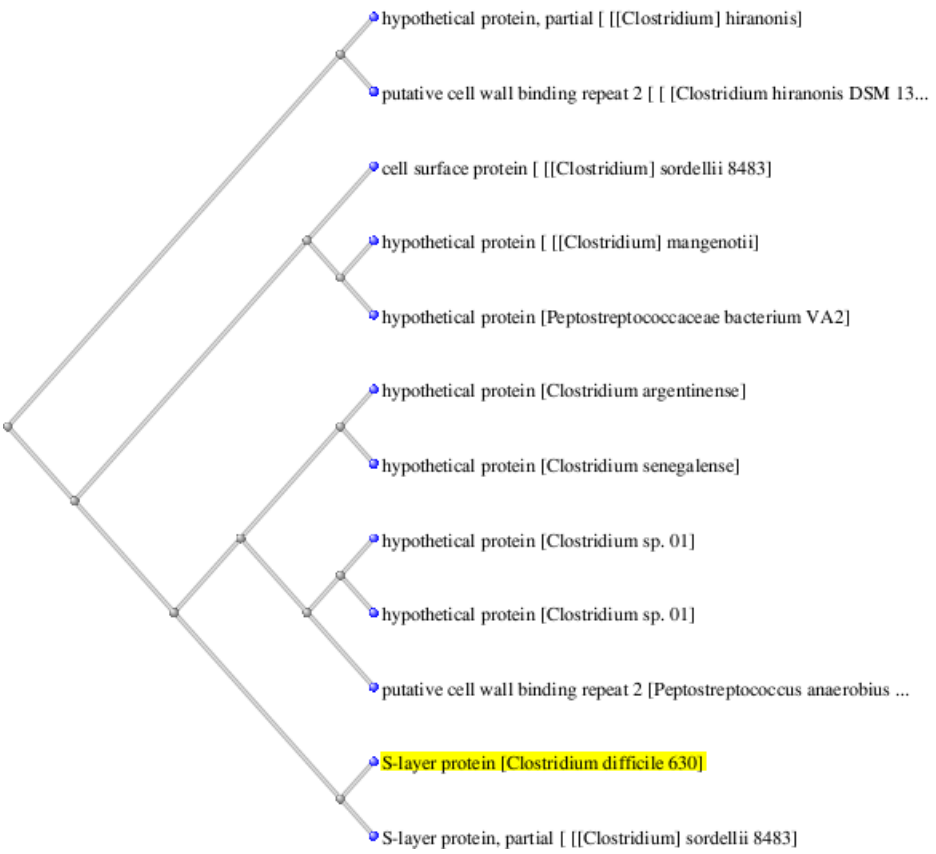


Figure 3-24. Slanted phylogenetic tree of BLASTP results from a *C. difficile* 630 SlpA similarity search , excluding *C. difficile*. Search performed 13/11/15.

3.3.6 Analysis of the *C. sordellii* 8483 genome

The position of the *C. sordellii* 8483 genome in the PATRIC clostridia phylogeny was noted in 2013. However, the sequence of this strain was not added to NCBI until 2015. Upon further investigation, the insertion date of the strain, displayed within PATRIC, was stated as December 2014, which is over a year later than the screenshot with *C. sordellii* 8483 was taken (Figure 3-21). Correspondence with the PATRIC authors confirmed that the sequence was submitted directly by a research group, before the official 'insertion date', to determine where *C. sordellii* 8483 would fit in the clostridia phylogeny tree. The high identity of the SlpA biomarker between *C. difficile* and this *C. sordellii* strain is not replicated in any other *C. sordellii* strains with sequences submitted to NCBI. For example, Figure 3-25 shows the top BLASTP hits for the biomarker search against *C. sordellii* ATCC 9714 (A) and *C. sordellii* VPI 9048 (B), exhibiting poor similarity (search performed 18/11/15).

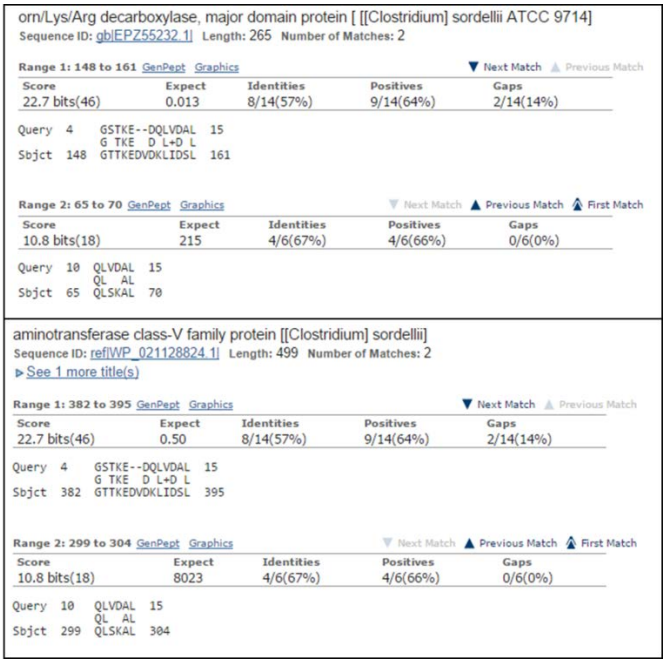


Figure 3-25. BLASTP search showing similarity to the SlpA biomarker within *C. sordellii* ATCC 9714 (A) and *C. sordellii* VPI 9048 (B) sequences, performed 18/11/15.

A comparative genomics study was performed with *C. sordellii* 8483, VPI 9048, ATCC 9714 and *C. difficile* 630 and R20291. The initial step looked at the genome sizes and number of associated proteins located within each of these strains, data taken from the NCBI database. The genome size and number of proteins for *C. sordellii* 8483 is much larger than those seen in other *C. sordellii* and *C. difficile* strains (Table 3-3).

Table 3-3. The genome size (bp) and the number of proteins found within *C. difficile* 630 and R20291, and *C. sordellii* ATCC 9714, VPI 9048 and 8483 . All strains have their associated Accession number. The figures for *C. sordellii* 8483 have been highlighted in yellow.

Strain	Genome Size (bp)	Protein number
C. difficile 630 Accession: NC_009089.1	4,290,252	3767
C. difficile R20291 Accession: FN545816.1	4,191,339	3508
C. sordellii ATCC 9714 Accession: PRJEB7738	3,534,945	3435
C. sordellii VPI9048 Accession: AQGJ01	3,571,992	3586
C. sordellii 8483 Accession: AJXR01	7,810,226	7850

The FASTA sequences for the five *C. difficile* and *C. sordellii* strains were compared in a pairwise analysis using the average nucleotide identity (ANI) calculator[22] (Rodriguez-R and Konstantinidis, 2014). The identity is predicted using both one-way ANI and two-way ANI between two genomic datasets, as calculated by (Goris *et al.*, 2007). Furthermore, the ANI calculator supports both complete and draft genomes, estimating the average nucleotide identity between two genomes to determine the likelihood that the sequences are from the same species or not. If the ANI value is above 95 % then the sequences are predicted to be from the same species.

An example of the results produced by the ANI calculator is provided in Figure 3-26, which was gained from the identity predictions between *C. difficile* 630 and *C. sordellii* 8483. The two genomes generated a similarity of 99.36 %, which classes them as the same species.

2222 http://enve-omics.ce.gatech.edu/ani/

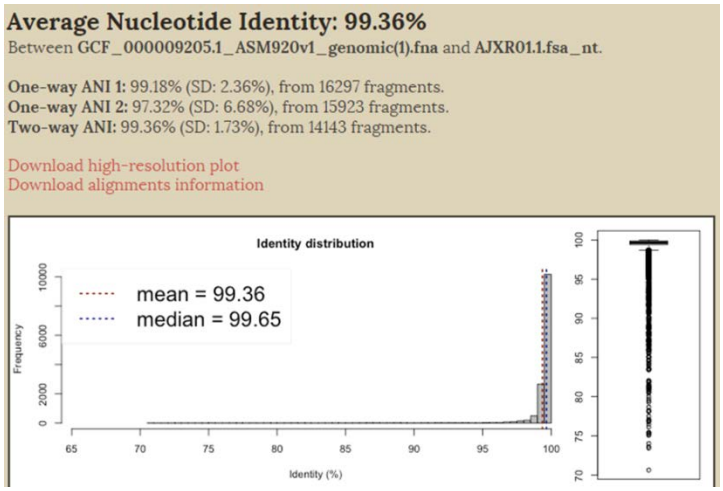


Figure 3-26. ANI calculator results of *C. difficile* 630 compared to *C. sordellii* 8483 . The histogram displays the distribution of the sequence fragments displaying identity between the two strains. The mean value of the histogram is at identity 99.36 % indicating that the two strains are highly similar and are in fact the same species.

The ANI results for all five of the tested *C. sordellii* and *C. difficile* strains are shown in Table 3-4. *C. sordellii* 8483, highlighted in yellow, was classed as the same species as both *C. sordellii* and *C. difficile*. In contrast to this, the *C. sordellii* strains, ATCC 9714 and VPI 9048 were only similar to each other, as were the *C. difficile* strains, 630 and R20291.

Table 3-4. The ANI percentages of comparisons made between *C. difficile* 630, R20291, and *C. sordellii* ATCC 9714, VPI 9048 and 8483 . All percentages above 95 % indicate the strains are the same species. The results for *C. sordellii* 8483 have been highlighted in yellow.

	C. difficile 630	*C. difficile* R20291	*C. sordellii* 8483	*C. sordellii* VPI 9048	*C. sordellii* ATCC 9714
C. difficile 630	100				
C. difficile R20291	98.96	100			
C. sordellii 8483	99.36	98.92	100		
C. sordellii VPI 9048	78.31	78.25	98.38	100	
C. sordellii ATCC 9714	78.27	78.01	98.76	98.43	100

The complete *C. difficile* genomes of strains 630 and R20291 were used as reference genomes for the alignment of the contigs present in the three *C. sordellii* draft genomes. The sequences were aligned with CONTIGuator 2 (Galardini *et al.*, 2011) and an overview of the results is provided in Table 3-5. The results show that the two *C. difficile* sequences

displayed excellent alignment, with zero unmapped contigs. Furthermore, as expected due to the ANI scores, both *C. sordellii* VPI 8048 and ATCC 9714 did not align well to *C. difficile* 630. In contrast, half of the *C. sordellii* 8483 contigs align to *C. difficile* 630.

Table 3-5. Overview of the CONTIGuator 2 results for the alignments of *C. sordellii* ATCC9714, VPI 9048 and 8483, and *C. difficile* R20291 , all aligned against the reference genome, *C. difficile* 630. Also displayed are the percentage of unmapped contigs.

	C. difficile R20291	*C. sordellii* ATCC 9714	*C. sordellii* VPI8048	*C. sordellii* 8483
CONTIGuator results against *C.difficile* 630	Very similar, very good alignment	Not similar, no alignment	Not similar, very poor alignment	Some similarity, alignment shown for half
Unmapped contigs (%)	0	-	95	57

When the total length of the mapped contigs was displayed in a pairwise comparison view, it was shown that 4,143,197 bp from the *C. sordellii* 8483 genome mapped to the reference *C. difficile* strain, which is 4,290,252 bp in length. This contig mapping, displays that almost the entire reference genome aligns with, and is therefore highly similar to, parts of the *C. sordellii* 8483 DNA sequence.

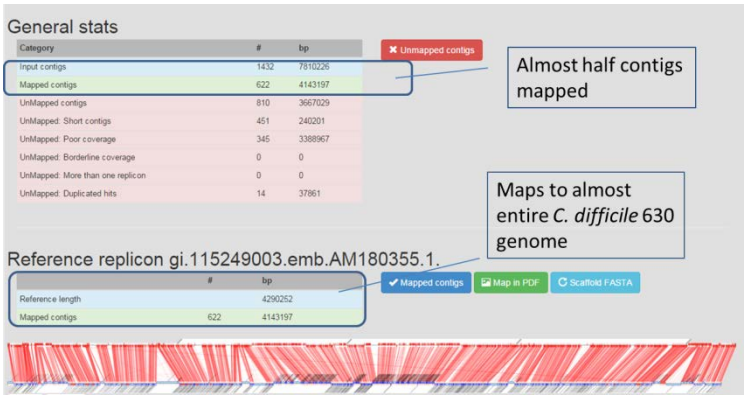


Figure 3-27. CONTIGuator results of *C. sordellii* 8483 aligned against *C. difficile* 630 . There are 622 mapped contigs that correlate to 4,143,197 bp aligned with the *C. difficile* genome.

To investigate the alignment further, an area of the CONTIGuator result was enhanced and several contigs were selected for downstream analysis (Figure 3-28). The blue sections represent insertions and indicate where the contigs do not map to the reference genome, and the red sections are successfully aligned contigs.

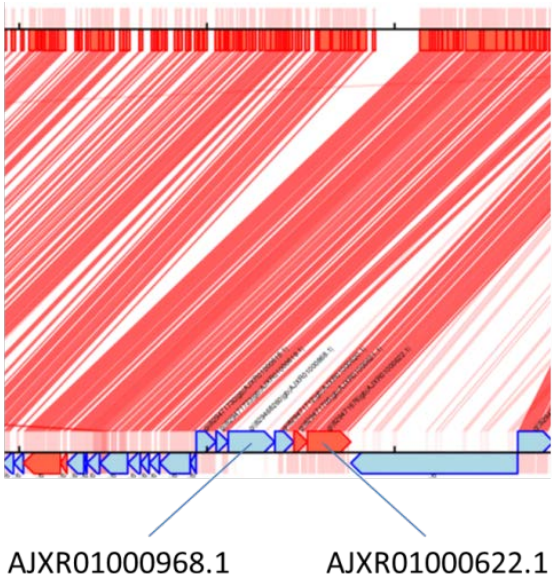


Figure 3-28. A magnified image of a section the *C. sordellii* 8483 (bottom strand) alignment to *C. difficile* 630 (top strand). The blue arrows display contigs that did not map to the reference genome and the red arrows show the contigs that have mapped to the *C. difficile* 630 genome. One of each of the mapped and unmapped contigs have been labelled on the image.

To clarify the similarity of the aligned contigs (red), the sequences for a number of these contigs were searched using NCBI nucleotide BLAST (BLASTN) and were highly similar to only *C. difficile* strains, and displaying very little similarity to the *C. sordellii* strains (search performed 20/11/15). An example of this similarity is shown in the BLASTN results of contig AJXR01000622.1, which had 99 % coverage and 99 % identity in *C. difficile* strains and only 29 % coverage and 82 % identity to *C. sordellii* (Figure 3-29).

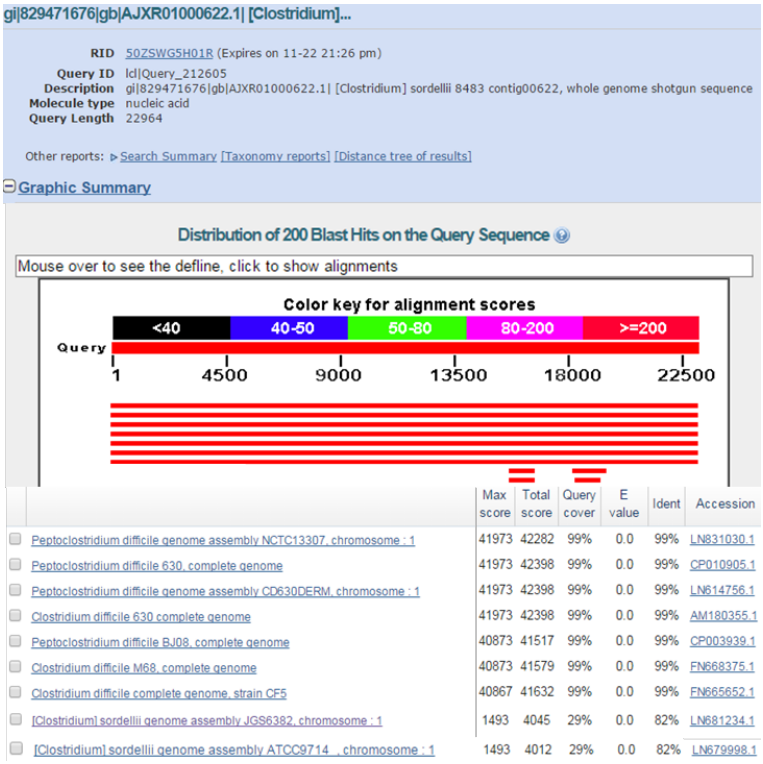


Figure 3-29. BLASTN results of the *C. sordellii* 8483 contig, AJXR01000622.1, which mapped to *C. difficile* 630 using CONTIGuator 2 . The length of the red lines on the image, display the relative length of the sequence and the colour displays the high alignment score. The bacterial strains with the largest query cover and percentage identity with the contig are all *C. difficile* strains (search performed 20/11/15).

Several contigs that were shown to be insertions and not match *C. difficile* (blue contigs), were also analysed using BLASTN (search performed 20/11/15). The results displayed similarity to only *C. sordellii* strains and not to *C. difficile*. The strain that showed the greatest similarity was *C. sordellii* ATCC 9714, which was the top hit for all unmapped contigs that were searched. An example of the BLASTN results is shown for the *C. sordellii* 8483 contig, AJXR01000968.1 (Figure 3-30). The sequence input for the *C. sordellii* 8483 genome contains close to all of the *C. difficile* 630 DNA sequence and also almost all the *C. sordellii* ATCC 9714 DNA sequence.

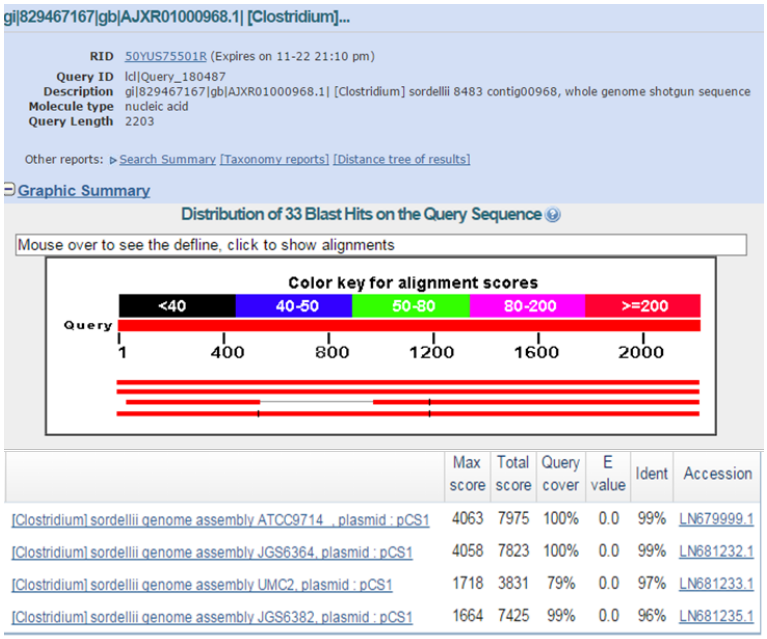


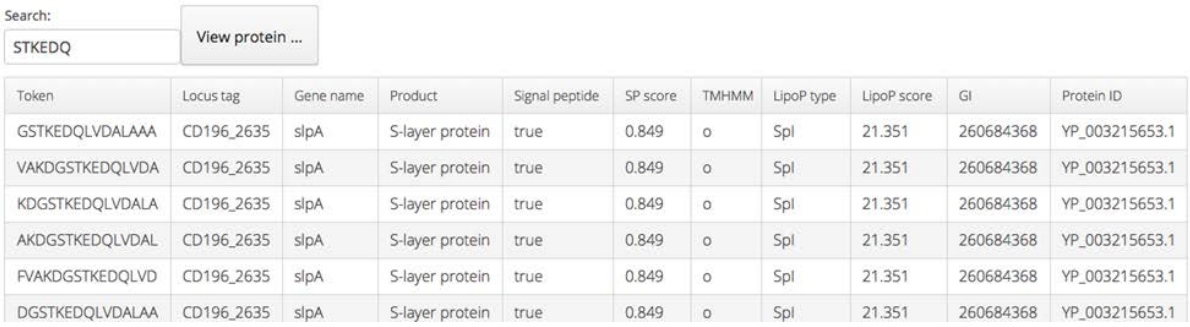
Figure 3-30. BLASTN results of the *C. sordellii* 8483 contig, AJXR01000968.1, which did not map to *C. difficile* 630 using CONTIGuator 2 . The length of the lines on the image indicates the relative length of the sequence, while the colour displays the alignment score (where red indicates the highest scores). All sequence similarity hits are seen within the *C. sordellii* species (search performed 20/11/15).

This evidence indicates that the *C. sordellii* 8483 sequence which was submitted to NCBI was not a pure *C. sordellii* strain, but also contained a contaminant from a *C. difficile* strain. In addition, the data suggests that the two strains within the sample are *C. sordellii* ATCC 9714 and *C. difficile* 630. Consequently, the similarity data of *C. sordellii* to the SlpA biomarker can be disregarded. However, additional *C. sordellii* strains will require lab-based testing to ensure there is no non-specific binding of the detection molecules. Also, it would be useful to obtain the *C. sordellii* 8483 sample from the researchers who submitted the sequence and perform in-house next generation sequencing on a single colony taken this sample.

3.3.7 Verifying the validity of the SlpA biomarker with the 2015 IDRIS system

The 2015 IDRIS system contained 461,807,691 more distinct tokens, from around 1300 more organisms, than the 2012 system. Using the *C. difficile* strains as the GOI the 2012 system presented 9793 protein IDs that contained at least one unique token whereas, the 2015 IDRIS system provided 3204 unique token-containing proteins. The 2015 IDRIS system contains nine *C. difficile* strains and therefore a decrease in the number of 'unique tokens' was expected. A search of the 2015 token list, identified the SlpA biomarker as still present

within the 'unique token' list (Figure 3-31). Therefore, the increase in sequence data within IDRIS did not alter the most promising result.


Search:
STKEDQ
View protein ...

Token	Locus tag	Gene name	Product	Signal peptide	SP score	TMHMM	LipoP type	LipoP score	GI	Protein ID
GSTKEDQLVDALAAA	CD196_2635	slpA	S-layer protein	true	0.849	o	SpI	21.351	260684368	YP_003215653.1
VAKDGSTKEDQLVDA	CD196_2635	slpA	S-layer protein	true	0.849	o	SpI	21.351	260684368	YP_003215653.1
KDGSTKEDQLVDALA	CD196_2635	slpA	S-layer protein	true	0.849	o	SpI	21.351	260684368	YP_003215653.1
AKDGSTKEDQLVDAL	CD196_2635	slpA	S-layer protein	true	0.849	o	SpI	21.351	260684368	YP_003215653.1
FVAKDGSTKEDQLVD	CD196_2635	slpA	S-layer protein	true	0.849	o	SpI	21.351	260684368	YP_003215653.1
DGSTKEDQLVDALAA	CD196_2635	slpA	S-layer protein	true	0.849	o	SpI	21.351	260684368	YP_003215653.1

Figure 3-31. The IDRIS Web browser displaying the results of the 2015 IDRIS search for unique tokens within all *C. difficile* strains within the system as the GOI. A search of the results for the SlpA biomarker 'AKDGSTKEDQLVDALA' identified it as present, spanning several tokens. A screenshot of this search was provided and shows the token sequence, locus tag, gene name, protein name. The results table also displays if a signal peptide is present within the protein sequence and the results of several cell-surface localisation tools, SignalP, TMHMM and LipoP. Finally, the accession numbers of the proteins are provided in the final two columns.

3.4 Discussion

A search of the literature confirmed the novelty of the bioinformatics approach used in this work. There have been a number of previous studies that have used a "sliding window"-based approach to break down genomic or proteomic data into tokens. However, this has not occurred for the entire RefSeq dataset, and nor have they been used for the identification of protein fragments that are unique to an organism. In these previous studies, "tokens" were used in smaller data sets for the identification of insertion sites in transposon libraries (DeJesus and Ioerger, 2013) or the identification of prophages (Akhter *et al.*, 2012).

Although *C. difficile* was used in this project as a case study, the bioinformatics approach could be exploited to identify unique regions within any organism contained within the database. Furthermore, the IDRIS system not only distinguishes unique bacterial biomarkers for whole species, it could also be used to ascertain sequences that distinguish strains within species. An example of the IDRIS system providing sequences for inter-species differentiation was shown for MRSA strains within the *S. aureus* species (Flanagan *et al.*, 2014).

3.4.1 Using the IDRIS system with *C. difficile* species as the GOI

The original 2012 IDRIS system contained three *C. difficile* strains which were used as the GOI, providing 9793 *C. difficile* protein IDs that contained unique tokens. With the addition of almost double the total number of organisms in the 2015 TokenDB (from 1400 to 2764), and an increase from three to nine *C. difficile* strains within the GOI, there was a decrease in the number of unique token-containing proteins. This decrease was roughly a third of the results gained from the 2012 system, from 9739 to 3204 unique token containing proteins. A decrease in results between the two IDRIS system searches was not surprising. The larger the number of organisms within the database, the greater the chance that: (a) a new species within the GOI has subtle variations in its sequences that reduce the number of exact token matches among the rest of the GOI members, and, (b) that a closely-related strain from outside the GOI contains an exact match to a token that is common to the members of the GOI. This large increase in tokens from the newly deposited strains therefore decreases the likelihood that a token common to the GOI is also unique. Consequently, with the ever increasing rate of bacterial sequence submissions to NCBI, it is likely that fewer unique tokens will be found for GOIs. Conversely, the more sequences that are searched, the better the chance there is of identifying truly unique tokens. Therefore, it is predicted that the number of tokens will eventually reach a plateau no matter how many more strains are added. Figure 3-32 provides some evidence of this, showing a graph that was produced from the unique token results from the GOI of *Chlamydia trachomatis*. This species was selected simply because IDRIS contains sequence data from a high number of strains from this species. As the number of strains within the GOI increases, the number of tokens that are common, yet unique to this GOI, deceases. After an initial swift decrease, the line begins to plateau, showing relatively little difference in the number of tokens between 40-73 strains within the GOI.

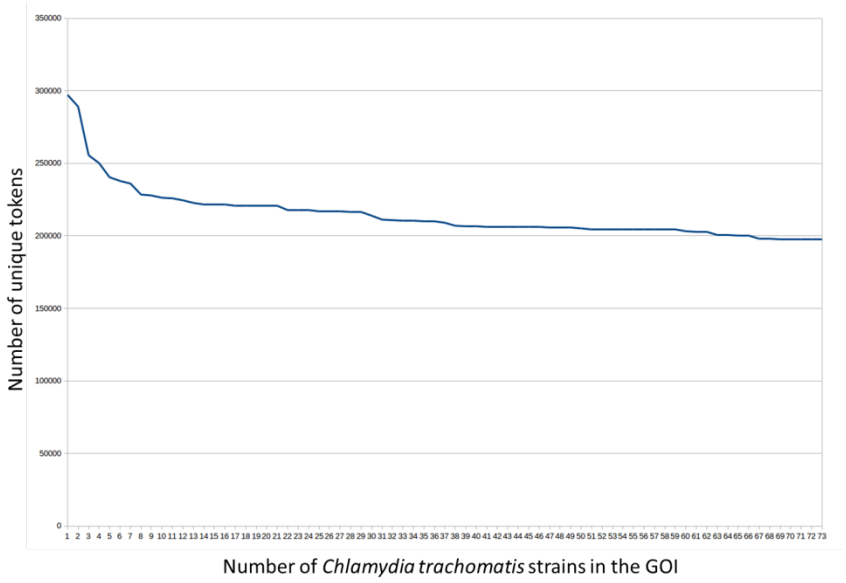


Figure 3-32. The resulting number of 'unique tokens' when using GOI's with differing numbers of *Chlamydia trachomatis* strains , ranging from 1-73. As the number of strains within the GOI increases, the number of 'unique tokens' decreases, initially quite rapidly and then reaching a relative plateau.

3.4.2 Subcellular localisation of the 9739 *C. difficile* proteins

The results of the token containing proteins, from the 2012 IDRIS system, were progressed to the second step of the bioinformatics approach - the subcellular localisation of the proteins (Figure 4-4). Although this was an automated step using ApID2, within the IDRIS system, the positive results of proteins predicted to be located on the surface of the cell were manually validated with the subcellular localisation tools. The displayed results from the individual subcellular localisation tools all analysed the *C. difficile* SlpA precursor sequence, as this was the most promising protein. The SignalP results for SlpA identified a signal peptide at amino acid positions 1-24, which correlated with the literature (Calabi *et al.*, 2001; Emerson and Fairweather, 2009). This signal peptide indicates that the protein is destined for the secretory pathway and therefore is not a cytoplasmic protein (Nielsen *et al.*, 1997). The protein sequences that were disregarded did not contain a predicted signal peptide. Further analysis of the signal peptide was performed using LipoP, which predicted *SPI* to cleave the SlpA signal peptide and therefore, SlpA is not predicted to be a lipoprotein. In prokaryotes, the *SPI* cleavage motif is most commonly associated with alanine (Payne *et al.*, 2012) and the SlpA cleavage site is no different as it's predicted to be between two alanine's (Appendices, section 9.3). Proteins were disregarded if they were predicted to

contain signal peptidase II (SPII), an aspartic acid protease that cleaves signal peptides to form lipoproteins (Dalbey *et al.*, 2012). Proteins were also disregarded if they were predicted to be a cytoplasmic protein (CYT) (Petersen, 2011).

Analysis of SlpA with both PSORTb and TMHMM, predicted that it is a cell wall protein with only one transmembrane helix at the N-terminal. The transmembrane helix and signal peptide prediction tools both search for hydrophobic regions within sequences, often causing the same sequence region to be predicted as both a signal peptide and a transmembrane helix (Kall *et al.*, 2004). Since the signal peptide is cleaved within the membrane, it is unsurprising that the signal peptide is often categorised as a transmembrane region, which was the case with the *C. difficile* SlpA. Any proteins that were classed as cytoplasmic or cell membrane proteins were disregarded from further analysis. From the initial 9793 proteins that contained at least one token that was unique to *C. difficile*, there were 28 that were positive with the criteria as defined in the results section. All four subcellular localisation tools and Phobius classed SlpA as cell wall associated. Therefore, the protein is likely to be accessible for any detection molecules within a PoN sensor, without the need of pre-treatment of cells.

3.4.3 Identifying a unique biomarker within the 28 *C. difficile* proteins

The TokenDB identified a token as unique if the entire 15 amino acid sequence was not located within another organism. However, since the TokenDB performs exact sequence matching, it is possible that the 'uniqueness' of a given token might only be due to a single amino acid mismatch. Consequently, in the worst case, a 'unique token' could share up to 14 amino acids with a token in other species.

Since there are only 20 amino acids (Bettelheim *et al.*, 2015), it was highly unlikely that a single token would be completely unique to a group of organisms. As the threshold of similarity increases from zero to 14, the likelihood of finding a unique token increases. In order to gain the maximum number of tokens, it was decided to set the similarity threshold of the TokenDB to 14, for example 14 out of the 15 amino acids could be located in other bacteria. For future work, the TokenDB would require testing at a similarity threshold decreasing from 14 amino acids, until the optimum number was identified. There would

need to be a compromise between a large enough number of tokens within the chosen location of a protein, and the threshold being low enough to ensure that the tokens contained sections of wholly unique sequences of amino acids.

Due to this similarity threshold, it was imperative that the 28 proteins predicted as being cell wall associated were analysed manually for unique biomarkers, with a minimum length of 15 amino acids. Many proteins did not contain unique biomarkers that were 15 amino acids or longer. Often there were predicted biomarkers that contained less than 4 unique amino acids, which was not deemed adequate for a specific detection molecule. Biomarkers were classed as viable candidates if BLASTP analysis determined that the most similar sequence to the biomarker, not located within *C. difficile,* was different by a minimum of four amino acids. This number was selected to try and ensure that any biomarker recognition molecules would not non-specifically bind to similar sequences. Due to the high specificity of antibodies and aptamers, even a residue difference of one is likely to remove the specificity of the binding molecules (Zegers *et al.*, 1995; Zheng *et al.*, 1995; Negri *et al.*, 2012). For future work, epitopes can be produced with differing numbers of amino acid changes, using site-directed mutagenesis. The minimum number of changes that removes binding of the recognition molecules can then be determined and used as the threshold for biomarker similarity.

Out of the 28 identified, the protein with the most promising biomarker was SlpA, within which, a 16 amino acid biomarker was identified using Uniprot BLAST, Jalview and BLASTP. This biomarker also contained hydrophilic regions, which indicated that this section of the sequence would likely be located on the external surface of the protein (Lodish *et al.*, 2000). As a result, these hydrophilic sections were predicted to be more accessible to detection molecules, such as antibodies. For a bacterial PoN sensor, it is vital that all strains of the species of interest can be detected and consequently, all strains must contain the specific recognition region, which will reduce the potential of obtaining false negatives. ACT revealed the conservation of the SlpA biomarker in all nine of the fully sequenced *C. difficile* strains. Furthermore, Jalview was used to identify the biomarker within sequences from all of the 13 SLT clades. Since these represent all known *C. difficile* strains, the presence of the SlpA biomarker demonstrates that the sequence occurs across the *C. difficile* species. The

SlpA sequence for *C. difficile* 630 places this strain into SLT clade 4, which is represented in the group of sequences provided by Dr Kate Dingle, by SLTS Ox1145. This similarity was established with the Jalview results, as well as being revealed in the research performed by Kate Dingle and colleagues (Dingle *et al.*, 2013).

The SlpA biomarker BLASTP search performed in 2012 revealed the non *C. difficile* strain with the closest similarity to be from *N. nitrogenifigens*. In contrast, when the same search was performed in November 2015, there were several organisms that contained sequences with greater similarity to the SlpA biomarker than *N. nitrogenifigens*. This increase in organisms with similar sequences to the SlpA biomarker is not a surprise since, as shown in Figure 1-2, the number of submitted sequences is rising exponentially. Within the three years that the IDRIS system has been running, the number of genomes and related protein sequences that it contains has doubled. Many of the similar sequences, identified by the 2015 BLASTP search, were located in other clostridia species that were sequenced a few months preceding the search. The most similar hit to the SlpA biomarker with the 2015 BLASTP search was within *C. sordellii* 8483, which is discussed in detail within in section 4.3.7 . Here, the discussion focuses on two other species that also displayed high similarity to the biomarker, *C. argentinense* and *C. senegalense*.

Except for *C. sordellii*, *C. argentinense* displayed the greatest similarity to the SlpA biomarker with a difference of only three amino acids. *C. argentinense* is the name for the species that was once Group IV *C. botulinum* and was originally isolated from soil in Argentina in 1970 (Gimenez and Ciccarelli, 1970) and has not been associated with botulism in humans, animals or birds (Peck, 2014). Non-toxigenic samples have been noted to be isolated from blood and wounds (Johnson, 2005) but not from faecal samples. Therefore, although there is a similarity within these bacteria to the SlpA biomarker, any diagnostic sensor that is focussed on faecal samples may not come into contact with this species, thus false positives may not be an issue. However, non-specific binding of any recognition molecules with this species should be determined.

C. senegalense also displayed high similarity to the SlpA biomarker within the 2015 BLASTP search, with three amino acids different. *C. senegalense* was originally isolated from the faeces of a healthy Senegalese male (Mishra *et al.*, 2012; Ahmed *et al.*, 2014). Since the

bacteria were isolated from a faecal sample and similarity to the SlpA biomarker has been shown, there may be an issue with a *C. difficile* PoN sensor made to target this biomarker. However, there was no evidence to suggest that *C. senegalense* has been isolated from faecal samples in more developed countries and there was very little literature on this strain. Therefore, *C. senegalense* may not be common enough to cause any false positives when detecting *C. difficile*. However, this issue would need to be investigated in future work. If any of the *C. argentinense* or *C. senegalense* strains could be obtained for further work, these species should be analysed for binding to the detection molecules produced against the SlpA biomarker. If it is deemed necessary, it could be possible to use a second *C. difficile* biomarker, as identified using the IDRIS system, to differentiate those samples that truly contain *C. difficile*.

3.4.4 SlpA is expressed

Data from microarray experiments was used to provide an idea of the expression of SlpA protein, rather than conclusive evidence. The data shows that the *slpA* gene is one of the most highly expressed genes in *C. difficile* 630. The high *slpA* expression was shown at two different time points, 4 h and 10 h growth and also within a mutant strain lacking SigH, an alternative sigma factor, which is a key element in the initiation of sporulation (Saujet *et al.*, 2011). The main evidence for SlpA protein expression however, was provided with a thorough literature search. The SlpA protein forms the *C. difficile* S-layer that completely coats the outer surface of vegetative cells (Reynolds *et al.*, 2011). Therefore, a high rate of protein synthesis is required to maintain the intact S-layer (Sleytr and Messner, 1983). There is a very high rate of SlpA synthesis in uninhibited cells (Dang *et al.*, 2012) and Fagan and Fairweather (2011) state that it is the most highly expressed protein in *C. difficile*, with expression occurring in all stages of growth (Savariau-Lacomme, 2003; Fagan and Fairweather, 2011). Researchers have shown that SlpA is vital for cell growth (Dembek *et al.*, 2015) and it has not been possible to create SlpA mutants (Dang *et al.*, 2010; Fagan and Fairweather, 2011). SlpA being an essential protein would explain why SlpA is highly expressed in all stages of growth and in mutants. The literature fairly conclusively shows that SlpA is expressed, providing the evidence required to continue with SlpA as a potential biomarker. Additionally, since SlpA is one of the most expressed proteins and is on the

outside of the cell, the literature also highlights the suitability of the protein to provide a biomarker for a PoN sensor. The evidence suggests that any biomarker within the SlpA protein will be produced in abundance, presenting many epitopes for detection molecules to bind to. If definitive proof of the expression of SlpA was required, reverse-transcription polymerase chain reaction (RT-PCR) could be utilised. This technique includes the isolation of *C. difficile* RNA, cDNA is then produced with reverse transcriptase and provides the sample for traditional PCR amplification of the *slpA* region (Calabi *et al.*, 2001; Bustin, 2002). Although it is less sensitive than RT-PCR another possibility to determine SlpA expression is to use Northern blotting, which uses hybridising probes complementary to that of the specific SlpA mRNA (Trayhurn, 1996).

3.4.5 The biomarker is located on the surface of SlpA

The SlpA protein is post-translationally cleaved into the LMW and HMS SLP (Dang *et al.*, 2012). The SlpA biomarker is located within the HMW SLP and therefore, to determine the location of the biomarker, the HMW SLP structure was required. I-TASSER predicted the structure and the biomarker was forecast to be completely accessible on the surface of the protein in all five of the models. Since the model structures are predictions, it is difficult to say with complete certainty that the biomarker is located on the surface of the HMW SLP. However, since all five models predicted the biomarker surface location, including model A with the highest C-score, it is likely that the biomarker is exposed on the surface of the protein. This indication of surface exposure, along with the hydrophobicity predictions shown with Jalview, was enough evidence to proceed with this biomarker as the target sequence for detecting *C. difficile*. The laboratory work was used to determine if this biomarker is indeed accessible to detection molecules such as antibodies, and the specificity of these to *C. difficile*.

3.4.6 Summary of why the SlpA biomarker was selected to differentiate *C. difficile* from other organisms

Unlike other token containing proteins that did not contain biomarkers over the 15 amino acid threshold, SlpA did. The 16 amino acid SlpA biomarker 'AKDGSTKEDQLVDALA' was determined to be the most promising biomarker for differentiating whole *C. difficile* cells from other bacteria. Selection of this biomarker was due to several reasons, firstly, the

protein in which the biomarker resides, SlpA, was predicted by the five subcellular localisation tools to be surface associated. Secondly, the biomarker sequence is found within all sequenced *C. difficile* strains, including all SLTS that represent the entire *C. difficile* species. Also, this biomarker, within the 2012 searches, was classed as 'unique' to *C. difficile* as it was not entirely located within other species and was different to the closest related species by five amino acids. Finally, the biomarker was predicted to be on the surface of SlpA, which is expressed in all stages of growth and completely encapsulates the cell, therefore, a great number of biomarker epitopes are predicted for recognition molecules to bind to. The combination of all of these results provides information that the SlpA biomarker was the most promising.

3.4.7 Analysis of *C. sordellii* 8483 and the issues that can arise with poor sequence data

As shown with the analysis of the *C. sordellii* 8483 strain, bioinformatics analyses can only be as good as the sequences that are provided. If the initial biomarker search had been performed post-addition of the *C. sordellii* 8483 sequence to NCBI, then the SlpA biomarker would not have been identified and yet it was one of the most promising sequences provided with the bioinformatics approach. When a database such as the TokenDB is reliant on the sequences contained within NCBI, it is vital that the sequences are as accurate and true to the species of which it is labelled. One of the main problems is that a huge amount of sequence data is submitted and as a result, full reviews of all data are not possible and therefore, accuracy is not always achieved. An issue that arose when Keith Flanagan was creating the TokenDB was the poorly annotated genomes within RefSeq. A quote taken from the NCBI RefSeq page states that:

'RefSeq provides a comprehensive, integrated, non-redundant, well-annotated set of sequences'[23]

However, many of the sequences were not well-annotated and were not suitable for use within the TokenDB. The different databases published by the NCBI such as Assembly,

[23] http://www.ncbi.nlm.nih.gov/refseq/about/

RefSeq and Nucleotide are not always simple to navigate and information is not always well linked. With the ever-increasing amount of data being submitted to NCBI and other databases, it is important to have well established guidelines that researchers can follow when adding data. It is up to the researchers to ensure that sequence data is from a pure sample and is accurate to the species it is labelled with. Anomalies such as the huge difference in genome size seen with *C. sordellii* 8483 compared to other *C. sordellii* strains, should be noted and investigated by the researchers in order to prevent fundamental errors such as those seen with this submission.

3.5 Summary

The IDRIS system and the manual steps that were developed within this project were combined into a bioinformatics approach. This approach was used to successfully identify a biomarker that is common to all members of the *C. difficile* species, and yet found within no other organism. The biomarker was a 16 amino acid sequence within the precursor protein SlpA, which was further determined to be within the HMW SLP section of this protein. The sequence is predicted to be on the surface of the protein, which is in turn on the surface of the bacteria and therefore is forecast to be surface accessible. However, laboratory experiments will have to be performed to determine if the location of the biomarker is indeed accessible to antibodies. If the SlpA biomarker is proven unsuccessful, the *C. difficile* token-containing proteins, will be re-visited to identify alternative biomarkers. Although this bioinformatics approach has been used to identify *C. difficile* biomarkers, the IDRIS system and downstream analysis can be used to search for biomarkers within any species that has been fully sequenced and is annotated. These biomarkers could be used to identify a species as a whole or used to differentiate different strains within a species.

Chapter 4 The development of monoclonal antibodies for the specific detection of *Clostridium difficile*

4.1 Introduction

4.1.1 Bacterial identification

Traditionally, bacterial identification and taxonomy was based on morphological features, using bacterial culture and colony categorisation. However, this phenotypic identification is time-consuming and prone to human error (Peng *et al.*, 2010). In the second half of the 21st century, great advances were made in the technology available for bacterial taxonomy. In 1977, Carl Woese utilised 16S ribosomal ribonucleic acid (rRNA) sequences for phylogenetic taxonomy, resulting in the creation of the third domain, archaea (Woese and Fox, 1977). In the same year, Frederick Sanger developed DNA sequencing, a step that would revolutionise bacterial taxonomy (Sanger *et al.*, 1977). With this ability to sequence genes, bacteria could be grouped based on their genetic make up. As accessibility to this new DNA sequencing technique increased, so too did the amount of data being produced. Subsequently, in 1979, the nucleic acid database, GenBank, was established to store genomic data (Choudhuri, 2014). Another major technological advancement in bacterial taxonomy was the invention of polymerase chain reaction (PCR) by Kary Mullis in 1983, which enabled genes and specific bacterial identifiers, to be amplified and detected relatively quickly (Saiki *et al.*, 1988; Shafique, 2012). These new methods were the first major steps in molecular diagnostics and meant that bacteria could be identified and grouped in the space of hours as opposed to days, as seen with the use of traditional phenotypic taxonomy. These molecular biology techniques also decreased the need for human interpretation, an aspect which often caused errors in traditional phenotypic diagnostics (Tang *et al.*, 1998).

Rapid and accurate identification of bacterial pathogens is an important goal for clinical microbiology laboratories and front line clinicians (Tang *et al.*, 1998). However, with traditional techniques of time-consuming culture this goal is impossible to acheive. In contrast, recent methods of genotypic bacterial identification, such as PCR, has enabled more rapid diagnostics. It is not currently practical to use these techniques at the PoN, as it

requires specialist skills and laboratory equipment (Kufelnicka and Kirn, 2011; Buchan and Ledeboer, 2014).

4.1.2 *C. difficile* strain typing

Over the years, *C. difficile* strains have been categorised using a variety of techniques, leading to the same strain having a number of titles. Serotyping by slide agglutination or ELISA was the traditional method of grouping *C. difficile*, splitting strains into ten groups, each denoted by a letter (Karjalainen *et al.*, 2002). As molecular techniques advanced, new typing methods have emerged. Pulse field gel electrophoresis (PFGE), the use of alternating directions of electrical fields to resolve DNA fragments, is currently the standard *C. difficile* typing method used in Canada and the USA, providing strains with the nomenclature North American Pulsotype (NAP) and a number e.g. NAP1 (Janezic and Rupnik, 2010). Whereas in Europe, the standard method used to type *C. difficile* strains is PCR ribotyping, a form of genomic fingerprinting. Every *C. difficile* strain contains multiple copies of rRNA operons, each with a differing length of intergenic spacer between the 16S and 23S ribosomal subunits (Figure 4-1A). Ribotyping utilises these differences in intergenic spacer lengths, using PCR and primers within the 16S and 23S subunits to amplify these regions (Bidet, 1999). The PCR products are separated on an agarose gel, producing a band barcode that is unique to that particular *C. difficile* strain (Figure 4-1B). Strains that have been categorised using PCR ribotyping are titled R and a number, for example R002 and R050, which are two of the strains used in the initial laboratory work within this project.

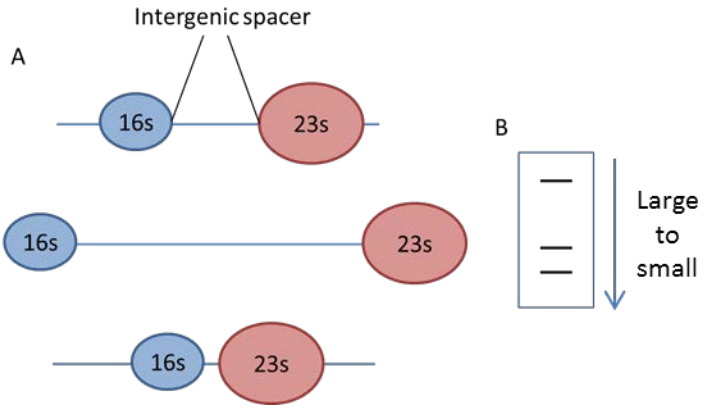


Figure 4-1. A diagram representing the principle of PCR ribotyping. The different sized intergenic spacers, within one strain, are amplified (A) and separated on an agarose gel. The bands on the gel produce a barcode that is specific to that *C. difficile* strain (B)

4.1.3 *C. difficile* biomarkers

Current CDI biomarkers include the *C. difficile* toxin genes identified using PCR, the toxins themselves detected by EIAs and the genomic fingerprints produced by PFGE or ribotyping. Other possible biomarkers are produced by the host in response to the infection, such as, fecal lactoferrin, calprotectin and cytokines, however, these are not specific to CDI (Burnham and Carroll, 2013). Genotyping of SlpA has also been shown as a possible biomarker for use in both identification and typing of *C. difficile* (Karjalainen *et al.*, 2002; Dingle *et al.*, 2013). However, the only current CDI biomarker that is rapid enough for PoN detection is toxin EIAs and as discussed in Chapter 1. Section 1.14, this method does not have the sensitivity required for a stand-alone test.

4.1.4 *C. difficile* genome sequencing

In 2006 the first *C. difficile* genome, *C. difficile* 630, was fully sequenced (Sebaihia *et al.*, 2006), having being isolated from a patient with severe pseudomembranous colitis (Wüst *et al.*, 1982). In the literature, *C. difficile* 630 is used as the reference strain (Kurka *et al.*, 2014) and during the bioinformatics section of this work the SlpA precursor protein from *C. difficile* 630, was used for the detailed analysis. For these reasons *C. difficile* 630 was chosen as the primary strain for the laboratory work within this project.

4.1.5 Antibodies

Antibodies are immunoglobulin (Ig) molecules secreted by the immune cells, B-lymphocytes. Monoclonal antibodies are made from identical B-lymphocytes, from the same parent clone, and are monovalent, that is, they bind to the same epitope. Monoclonal antibodies have been generated in the laboratory since 1975 when César Milstein and Georges J. F. Köhler invented the technique (Liu, 2014).

Antibody binding occurs between the antigen binding fragment (F_{ab}) of the antibody (**Figure 4-2**) and the epitope of the antigen. This binding is via transient, non-covalent interactions, the strength of which depends on the spatial complementarity of the antibody and antigen (Janeway *et al.*, 2001)

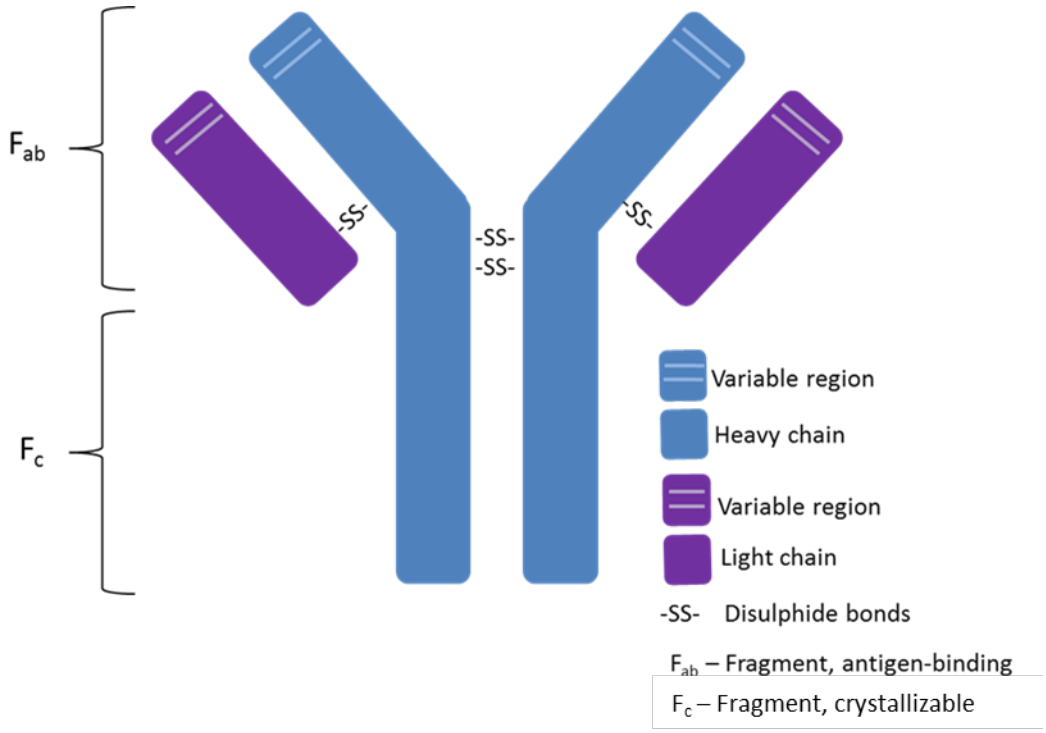


Figure 4-2. Molecular structure of Immunoglobulin G (IgG) antibody. The heavy and light chains are linked with disulphide bonds, as are the two heavy chains. The fragment antigen-binding (F_{ab}) fragment is the region where the antibody binds to antigens. The variable region contains the epitope binding sites, whereas the remainder of the light and heavy chains are conserved sequence. The fragment crystallizable (F_c) of an IgG is composed of the constant regions of the heavy chains.

The regions of the F_{ab} that bind antigens are the variable regions of the light and heavy chains, VL and VH respectively. Each of the VL and VH has three hypervariable loops named complementary-determining regions (CDRs). It is these CDRs that determine the antigen specificity of the antibody (Martin, 2010). Other than the highly variable CDRs, Abs have uniform protein structures, including the fragment crystallisable (F_c) region (Janeway *et al.*, 2001). Antigens can have epitopes that are conformational, where the bound amino acids have been brought together via protein folding, or linear, where the protein sequence has no tertiary structure and the antibody binds to a single, continuous sequence (Janeway *et al.*, 2001).

Antigen/Ab binding occurs via non-covalent interactions between the side chains of amino acids. These non-covalent interactions include hydrogen bonds, van der Waals forces, hydrophobic interactions and ionic bonds (Elgert, 2009). Non-covalent interactions are generally weak, however, the simultaneous combination of multiple interactions, form a

tight bond between antibody and antigen. This tight bond is reversible, enabling the complete separation of the two molecules without alteration (Mak and Saunders, 2001).

The weakest of the non-covalent interactions are the van der Waals forces, which include all intermolecular forces between electrically neutral atoms. Van der Waals interactions are caused by the fluctuating electron density of nearby atoms (Mak and Saunders, 2001). The varying distribution of electrons around an atom can form transient dipoles, which in turn can induce a dipole in a neighbouring atom. If close enough together, these dipoles cause attraction/repulsion between the atoms (Craig *et al.*, 2010).

Hydrogen bonds are stronger than van der Waals forces and are formed when a hydrogen atom is bound to a highly electronegative atom, such as oxygen, causing the hydrogen atom to be δ^+. The hydrogen is then attracted to other nearby electronegative atoms that have a lone pair of electrons, forming a bond (Lodish *et al.*, 2000). Amino acids with polar, uncharged side chains, for example serine (S) or glutamine (Q), have the ability to either donate or accept hydrogen atoms, enabling the formation of hydrogen bonds (Lieberman *et al.*, 2007).

Hydrophobic interactions can have similar strengths to hydrogen bonds and are formed by the preferential clustering between hydrophobic amino acid regions, when in aqueous solution (Troy and Beringer, 2006; Craig *et al.*, 2010). It is estimated that up to fifty percent of the interactions between the amino acids of the antibody and the epitope are caused by hydrophobic interactions (Elgert, 2009).

Finally, the strongest of the non-covalent interactions are the ionic bonds, such as salt bridges. A salt bridge is an interaction between a negatively charged carboxylate ion and a positively charged amino group (Snape *et al.*, 2014). The amino acids with charged side chains are able to form salt bridges and include lysine (K) and aspartic acid (D) (Schwarz *et al.*, 2004).

The contribution that each non-covalent interaction has to binding, is dependent on the particular Ab, with all interactions working together to confer antibody specificity. Even when Abs are raised against the exact same epitope, the affinity and specificity of these Abs can vary widely (Janeway *et al.*, 2001).

4.1.6 Monoclonal antibody production

For over 40 years, researchers have utilised animals to produce custom designed Abs that bind to antigens with high specificity. To produce Abs, animals such as mice, rabbits or goats are immunised with antigens, such as whole organisms, proteins or linear peptides. B-lymphocytes are presented with these 'foreign' antigens and begin producing Abs to bind to them. B-lymphocytes are removed from the spleen of the animal and fused with immortal myeloma cells. Myeloma cells do not produce any other immunoglobulins and lack the hypoxanthine-guanine-phosphoribosyltransferase gene, which is vital for the synthesis of nucleotides (Liu *et al.*, 2015a). The fused B-lymphocytes and myeloma cells are known as hybridomas. The hybridomas are grown in selective media so that only fused cells, which have the immortality from the myeloma cells and the nucleotide synthesising properties of the B-lymphocytes, survive. For monoclonal antibodies, antibodies made from single clones, the hybridomas are individually cultured in different wells, allowing antibodies from a single clone to be produced (Leenaars and Hendriksen, 2005).

For this project, synthetic, linear peptides were used as the antigen. There is not a definitive optimal length for a linear antigen, as there are a number of contributing factors. One of those factors is the current technology used for peptide synthesis, which results in decreasing yields as peptide length increases (Saper, 2009). Additionally, self-interaction can be an issue with larger peptide sequences, as they can form unintended conformational epitopes for antibodies to bind to (Hancock and O'Reilly, 2005). If antibodies bind to these conformational epitopes, they may not bind to the intended linear epitopes and therefore would not bind to the sequence of interest. In contrast, peptides that are too short are unlikely to cause an immunogenic reaction and are therefore, unsuitable for antibody production. Hancock and O'Reilly state that linear antigens of ten to twenty amino acids in length are optimal (Hancock and O'Reilly, 2005). Hansen and colleagues mention four to ten amino acids as optimal (Hansen *et al.*, 2013) and Berglund and colleagues affirm that linear epitopes usually range from six to nine amino acids (Berglund *et al.*, 2008).

4.1.7 Species and strains used during the laboratory work

Throughout the laboratory work, a number of different *C. difficile* strains were used, there is a full list of which in Table 2-3. It is important to note that the *C. difficile* reference strain,

C. difficile 630, was used throughout the work as a positive control. There were three species, closely related to *C. difficile*, which were used to determine any non-specific binding of the Abs. These species were ascertained with the phylogenetic analysis within section 3.3.5. *C. hiranonis* DSM-13275 and *P. anaerobius* VPI 4330 contained proteins with similarity to the *C. difficile* SlpA protein so were chosen as controls. *C. sordellii* 8483 was predicted to be a closely related species within PATRIC, however, this strain could not be obtained and therefore *C. sordellii* ATCC 9714 was used. Upon addition of the *C. sordellii* 8483 genome to NCBI in 2015, which was post laboratory work, analysis of this genome was performed in section 3.3.6. This bioinformatics investigation confirmed that the sequence is not accurate for this organism and contains *C. difficile* reads. Consequently, it is recognised that the PARTIC classification of *C. sordellii*, being the most closely related species to *C. difficile*, is not accurate. This species does, however, remain a very useful control as it is in the same genus and shows some similarity to the biomarker. Throughout this project, all three of these strains, *C. hiranonis* DSM-13275, *P. anaerobius* VPI 4330 and *C. sordellii* ATCC 9714 are collectively regarded as 'the closely related strains'. The same strains were used as controls throughout the project, so for ease of reading, they will be referred to by species name only, within the bulk of the text.

4.2 Results

4.2.1 Determining the presence of the biomarker within three *C. difficile* strains

The unique *C. difficile* biomarker predicted with the bioinformatics workflow is shown in Table 4-1. PCR amplification of the SlpA gene from three *C. difficile* provided products of the expected length, which was 1468 for *C. difficile* 630 and 1469 for *C. difficile* R002 and R050 (Figure 4-3A). The sequences of the PCR products were translated into amino acid sequences using ExPASY translate (Gasteiger *et al.*, 2003) and the biomarker was shown to be present in all three of the sequenced *C. difficile* strains (Figure 4-3b).

Table 4-1. The sequence predicted using the TokenDB and bioinformatics workflow to be a unique biomarker for *C. difficile*.

C. difficile HMW SLP biomarker sequence (16 amino acids)
AKDGSTKEDQLVDALA

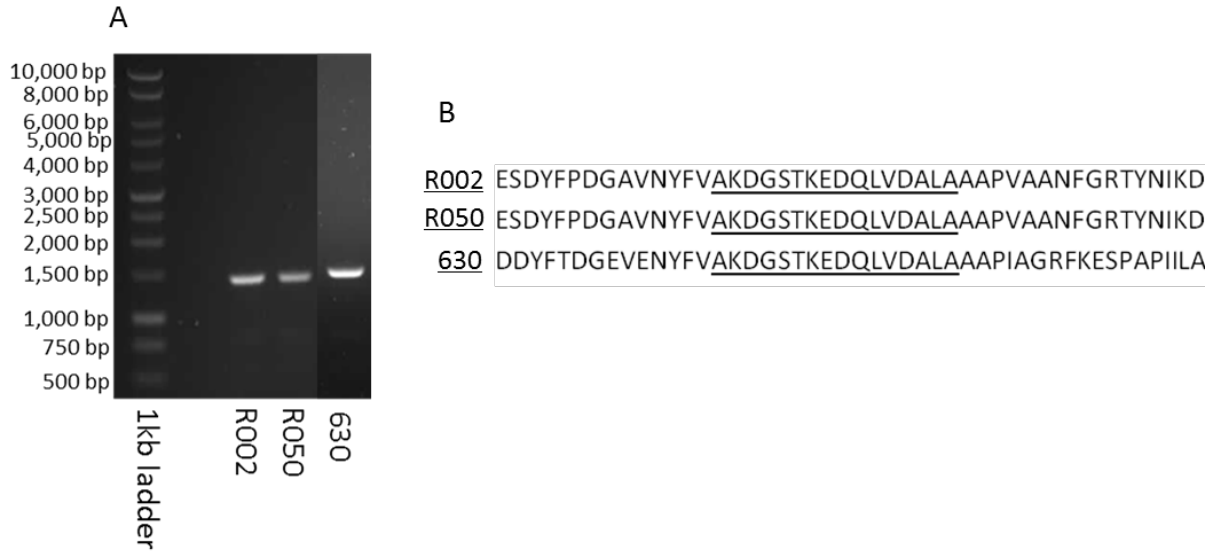


Figure 4-3. PCR products from the *SlpA* amplification, separated by agarose gel electrophoresis (A) and the sequence data of the purified PCR products, with the biomarker underlined (B).

4.2.2 Analysis of the surface layer proteins of *C. difficile*

Low-pH glycine extracted SLPs were separated on an SDS-PAGE gel (Figure 4-4). *C. difficile* 630 and R002 displayed two distinct protein bands, with a relatively high concentration of protein at sizes correlating to the HMW and LMW SLP within the literature. *C. difficile* R050 on the other hand, revealed a single band, containing a relatively high protein concentration, which was smaller in size than observed in the other *C. difficile* strains however, it remained within the HMW SLP size range (see discussion). The R050 LMW SLP was at the same size as the LMW SLP in R002 and 630 however, in comparison it had a smaller amount of protein.

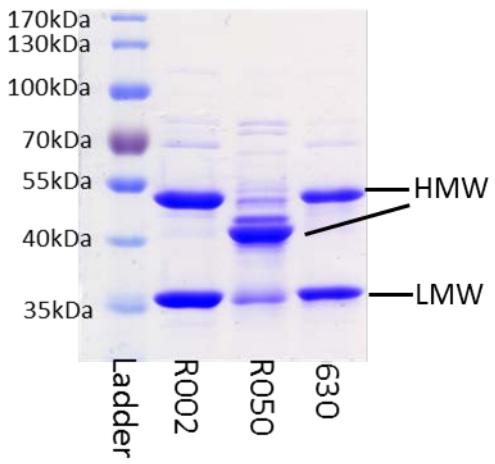


Figure 4-4. SDS-PAGE gel of the extracted surface layer proteins from *C. difficile* R002, R050 and 630. The HMW SLP varies in size and along with the LMW SLP is labelled. There are several other surface proteins present within the samples.

4.2.3 Production of HMW SLP monoclonal antibodies

AbMART synthesised several epitope sequences and used these linear peptides to produce mouse mAbs, via the hybridoma technique, discussed in section 4.1.6. The eight mAbs that were provided by AbMART are listed, alongside their ten amino acid epitope sequence (Table 4-2).

Table 4-2. The *C. difficile* antibodies produced by AbMART and their corresponding epitope sequences. The region contained within the C. difficile biomarker is highlighted in yellow.

Antibody ID	Epitope sequence (biomarker highlighted)
Ab370	KDGSTKEDQL
Ab521	STKEDQLVDA
Ab498	STKEDQLVDA
Ab491	FVAKDGSTKE
Ab652	KEDQLVDALA
Ab493	VDALAAAPIA
AbB31	STKEDQLVDA
AbB84	STKEDQLVDA

4.2.4 Determining mAb binding to the surface layer proteins of *C. difficile*

Initial experiments were performed to screen the eight mAbs for successful binding to the surface layer proteins of *C. difficile*. Figure 4-5 shows examples of two mAbs, Ab521 (A) and Ab652 (B) successfully bound to both the extracted SLPS (S) and residual pellet (P) of *C. difficile* 630. The dark dots indicate binding, which is not seen for the negative control, *C. sordellii* ATCC 9714. The positive control of the primary antibody (Positive Ab…) produced a clear mark. There were four mAbs which revealed positive results, Ab521, Ab652, Ab491 and Ab493, the latter two are displayed within the Appendices Figure 9-2.

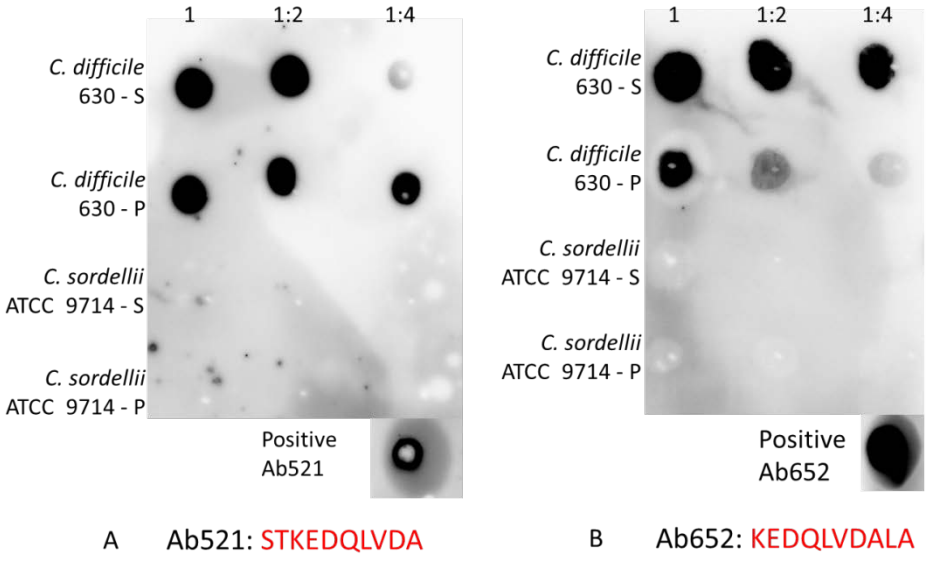


Figure 4-5. Dot blots with Ab521 (A) and Ab652 (B) of low pH glycine extracted surface proteins (S) and the residual pellet, post extraction (P) for *C. difficile* 630 and *C. sordellii* ATCC 9714. The primary antibody was directly blotted on to the membrane as a positive control. The biomarker regions of the antibody epitopes are displayed in red.

Within the dot blots, the remaining four mAbs produced negative binding with two examples, Ab370 (A) and Ab498 (B), shown in Figure 4-6. The results for AbB31 and AbB84 are displayed in the Appendices Figure 9-3. All four of these mAbs do not bind to either the *C. difficile* or *C. sordellii* surface layer proteins or residual pellet and were excluded from further analysis.

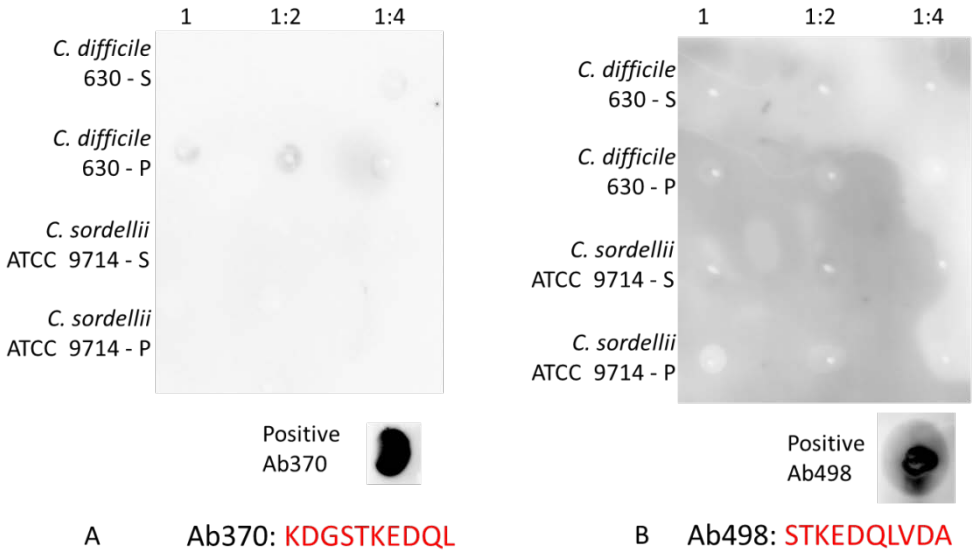


Figure 4-6. Dot blots with Ab370 (A), Ab498 (B) against low pH glycine extracted surface proteins (S) and the residual pellet, post extraction (P) for *C. difficile* 630 and *C. sordellii* ATCC 9714. The primary antibody was directly blotted on to the membrane as a positive control. The biomarker regions of the antibody epitopes are displayed in red.

4.2.5 Binding of the mAbs to proteins within *C. difficile* and the closely related strains

Western blots were performed with the extracted SLPs (S) and the residual pellet (P), of *C. difficile* 630 and the closely related species *C. sordellii* and *P. anaerobius* (Figure 4-7). Concomitant SDS-PAGE gels of the samples were run to show the loading of the proteins and if the low pH glycine method extracts any proteins from the closely related species.

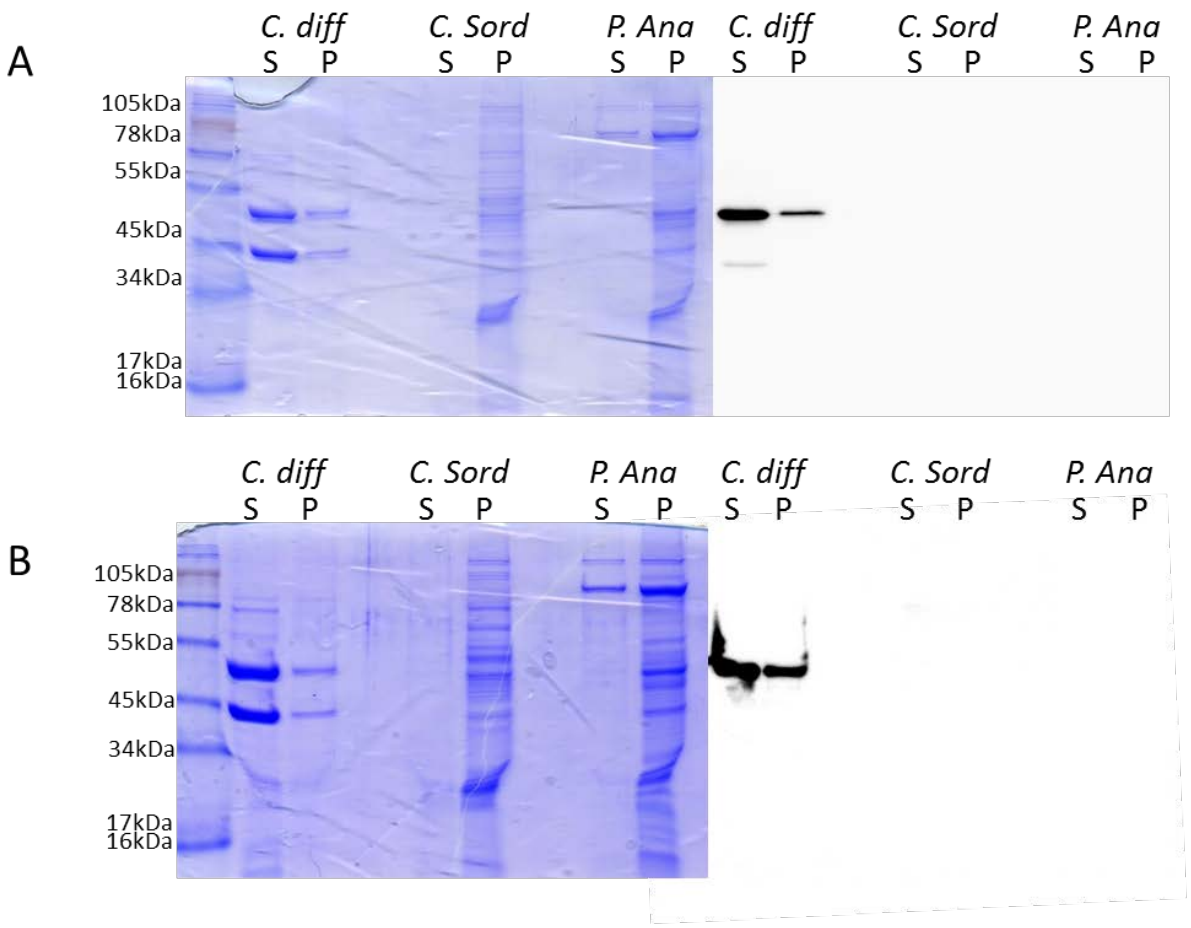


Figure 4-7. Western blot (right) and concomitant SDS-PAGE gel (left) with Ab521 (A) and Ab652 (B) against the low pH glycine extracted surface proteins (S) and residual pellet (P) of *C. difficile* 630 and the closely related strains *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330.

All four of the mAbs, Ab491, Ab493, Ab521 and Ab652 showed binding to proteins in both the extracted surface layer proteins and pellet of *C. difficile.* The mAbs bound strongly to a protein around 50 kDa in size, there was no signal displayed for any of the four mAbs with the closely related species. Two of the Western blots for Ab521 and Ab652, are shown in Figure 4-7, (A) and (B) respectively, and the blots for Ab491 and Ab493 are displayed in the Appendices Figure 9-4.

4.2.6 Binding of the four mAbs to proteins within commensal bacteria that are also found within faecal samples

Western blots were used to determine binding of the mAbs to denatured proteins from two *C. difficile* strains, 630 and R002, and the Gram-positive commensals, *S. aureus* ATCC 29213

and *B. subtilis* BSB1. Similarly to the previous Western blots, both the surface extracted proteins and the residual pellets were used.

All four mAbs, Ab491, Ab493, Ab521 and Ab652 produced positive signals for both *C. difficile* strains and not against the commensal strains. The Western blots for two of these mAbs, Ab521 and Ab652 are shown in Figure 4-8 with the remaining two displayed in the Appendices Figure 9-5.

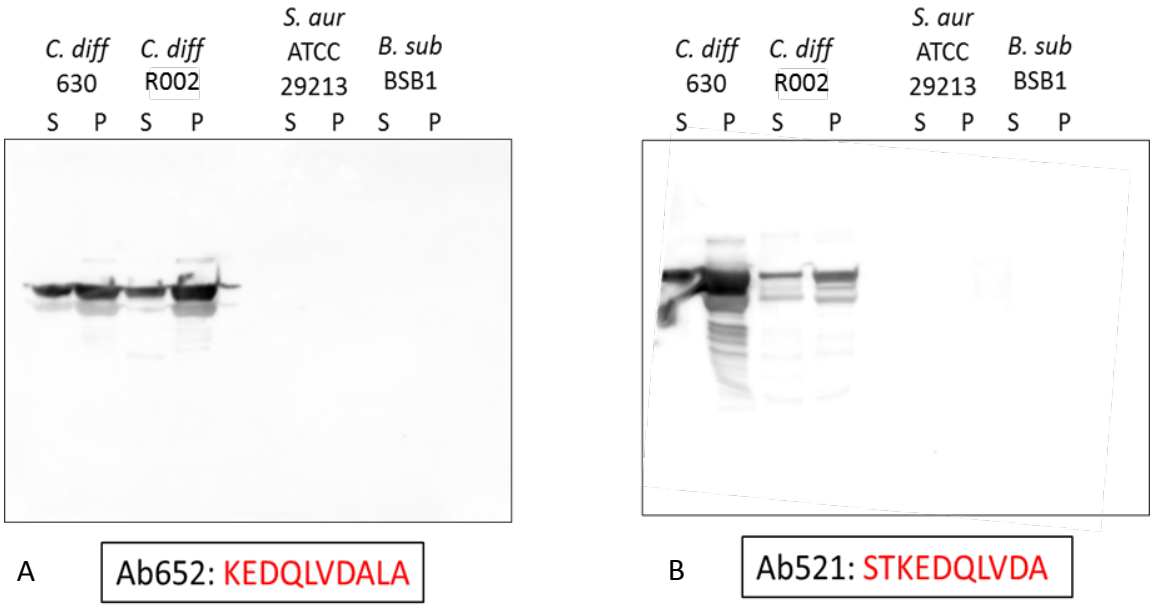


Figure 4-8. Western blots with Ab652 (A) and Ab521 (B) against the low pH glycine surface layer proteins (S) and remaining pellet (P) of *C. difficile* 630 and R002, *S. aureus* ATCC 29213 and *B. subtilis* BSB1.

4.3 Discussion

4.3.1 The *C. difficile* biomarker is present in the genome

The gene that contains the bioinformatics predicted unique biomarker, *SlpA*, is 2157 bp in size. Although, the size of this gene varies within other *C. difficile* strains, for example CD196 is 2274 bp and R20291 is 2277 bp in length. A smaller region of the gene, 1468 bp in size, which contained the biomarker, was selected for amplification as it enabled the consistent use of the same primers for all three strains. The genomes for R002 and R050 have not been fully sequenced however the *SlpA* gene was within the NCBI database and consequently, the primers were checked against the sequence prior to use. The product from the *SlpA* gene

was successfully amplified in all three *C. difficile* strains, 630, R002 and R050 (Figure 4-3). The PCR products when run on the agarose gel were all around the expected size. The lack of a fully anotated genome sequence within NCBI meant that the bioinformatics workflow did not include R002 and R050, therefore, the biomarker had not previously been shown to be present in these strains. The *SlpA* region was sequenced, translated in to an amino acid sequence and the biomarker was shown to be present in all three of the *C. difficile* strains. Although there were not a large number of *C. difficile* strains sequenced, it was important to show the presence of the biomarker within the *C. difficile* genome of some strains used within this project. This was especially important in strains R002 and R050 that were previously unknown to contain the sequence. The bioinformatics work showed that the biomarker was present in all currently sequenced *C. difficile* strains. Determining the presence of the biomarker was the first step in the laboratory workflow.

4.3.2 Expression of HMW SLP

Upon SDS-PAGE separation of the surface proteins of *C. difficile* 630, R002 and R050, the resulting image displayed several protein bands of differing sizes (Figure 4-4). It is stated throughout the literature that the HMW and LMW SLPs are the predominant surface layer proteins, with sizes that differ between strains, the LMW SLP is often around 37 kDa and the HMW SLP 50 kDa (O'Brien *et al.*, 2005; de la Riva *et al.*, 2011). However, these protein sizes can range from 30-43 kDa for the LMW SLP and 42-56 kDa for the HMW SLP (Cerquetti *et al.*, 2000; Calabi *et al.*, 2001; Fagan *et al.*, 2009). The samples for *C. difficile* R002 and 630 clearly displayed two prominent bands, both at sizes corresponding to the same samples within the literature (Calabi and Fairweather, 2002; Eidhin *et al.*, 2006). R050 on the other hand, displayed one dominant band at around 42 kDa, which is at the lower size of the HMW SLP range, and several lesser bands including one at the LMW SLP size of R002 and 630, 37 kDa. An SDS-PAGE gel of R050 could not be found in the literature, however, there are ribotypes that display a single, dominant band within their SLPs. Ribotype 167 has been shown to have a single dominant band at the HMW SLP and a very faint band at around the size of LMW SLP although, experiments to prove that it is the LMW SLP have not been performed (Calabi and Fairweather, 2002). There are also ribotypes R005 and R054 that

have a dominant HMW SLP however, these strains do not have bands around the LMW SLP (Eidhin *et al.*, 2006).

The SDS-PAGE performed during this research shows that not only are the LMW and HMW SLPs present within the surface layer, but there are also other, less prevalent proteins. These proteins are from the cell wall protein (CWP) family, which is formed by the 28 SLP paralogs that have been identified in *C. difficile* (Sebaihia *et al.*, 2006). CWP proteins can be visible in surface protein extracts on SDS-PAGE gels, and include the adhesion protein Cwp66, and the protease Cwp84 which are named after their size, 66 kDa and 84 kDa respectively (Waligora *et al.*, 2001; ChapetonMontes *et al.*, 2011). Cwp84 is responsible for cleaving the precursor SlpA protein in to the mature LMW and HMW SLPs (Péchiné *et al.*, 2005; Pantaléon *et al.*, 2015). The Cwps range in size from Cwp 25 which is 33.8 kDa up to CwpV, which is the largest of the cell wall proteins and is 167 kDa in size (Fagan *et al.*, 2011). The SDS-PAGE results corroborate the microarray data detailed in section 3.3.3.3 and the literature, defining that HMW SLP is expressed in *C. difficile* (Savariau-Lacomme, 2003; Fagan and Fairweather, 2011).

4.3.3 Production of mouse mAbs

Eight mouse mAbs were produced against the *C. difficile* biomarker and closely surrounding sequence, using epitopes of ten amino acids in length. From the eight mAbs there were two, Ab491 and Ab493 that had epitopes which were not fully conserved within the biomarker (Table 4-2). Whilst it was realised that the specificity of these peptides may not be retained, these mAbs were kept and used for an exploratory exercise. Further analysis of the two mAbs was required as it was still possible that they could bind to the epitope region conserved within the biomarker and, if not, they served to inform about the design criteria for future applications of this technique. It was hypothesised that these mAbs would not bind as successfully and specifically as the mAbs that had epitopes that were fully conserved within the biomarker.

4.3.4 Binding of mAbs to the native SLPs of *C. difficile*

Dot blots were used to ascertain if the mAbs bound to the *C. difficile* surface layer proteins. However, dot blots do not convey information about the size of the biomolecule to which

the mAbs bind to, nor about the specificity of antibody binding to individual proteins. Therefore this technique was used as an indicator for the successful antibody recognition of *C. difficile* proteins.

The results from the dot blots displayed four mAbs that successfully recognised and bound to *C. difficile* proteins and not to *C. sordellii*. Within these four mAbs there was Ab491 and Ab493 which have epitopes that are not fully conserved within the biomarker and Ab521 and Ab652 which do (Figure 4-5). The remaining four mAbs, Ab370, Ab498, AbB31 and AbB84 did not display clear binding to *C. difficile* and were therefore disregarded from further experiments (Figure 4-6). The unsuccessful dot blots demonstrated the need for positive controls, which showed that the technique was successfully working and the lack of binding was not due to error within the experiment. Perhaps the reason for the lack of binding was due to the dot blots being performed on proteins in their native state whereas, the mAbs were produced against linear epitopes. Consequently, the mAbs may not fit the conformational structure of the protein thus binding cannot occur. For a PoN sensor to be as rapid and as easy to use as possible, it is important that the mAbs bind to the native *C. difficile* proteins. This native binding would perhaps enable detection without the need for pre-treatment of samples.

The four positive mAbs showed binding to both the surface extracted proteins and the residual bacterial pellet. Since HMW SLP is one of the most abundant *C. difficile* proteins, it is likely that some of this protein is retained by the cell pellet, post low-pH glycine extraction, providing the mAb its epitope.

4.3.5 Binding of mAbs to the HMW SLP and not to species related to *C. difficile*

Within this chapter, there were two closely related species that were used for the laboratory analysis of the mAbs. These species were *P. anaerobius*, chosen due to its similarity to the SlpA protein, and *C. sordellii* which shows some similarity to the biomarker.

Western blots were used to determine mAb binding to denatured proteins in *C. difficile* and closely related strains. SDS-PAGE gels were run concomitantly to the Westerns so that the size of the protein to which the mAb binds to, could be established. Unlike dot blots, Western blots can be used with a molecular marker to ascertain the size of the antigen. The

SDS-PAGE gels also provided information about the loading of the samples, showing that the low pH glycine, surface protein extraction, provided proteins in the supernatant for *C. difficile* but not for *C. sordellii* and only a small amount for *P. anaerobius*. This lack of protein is evidence for using the cell pellet as well as the extracted proteins, allowing the mAbs to be tested for non-specific binding to the majority of proteins within the cell. It is unknown if *C. sordellii* has an S-layer however, Couchman and co-workers predict that *C. sordellii* does (Couchman *et al.*, 2015). Since there is no clear evidence that *C. sordellii* has an S-layer, there is also no extraction method or evidence that the low pH glycine method would provide proteins. *P. anaerobius* on the other hand, has a defined S-layer and is likely to be the proteins that were successfully extracted with the low pH glycine method (Kotiranta *et al.*, 1995; Bradshaw *et al.*, 2014).

All four Western blots, the two shown within this chapter and the two within the Appendices, demonstrate that the four mAbs bind to *C. difficile* at a protein size corresponding to the HMW SLP, around 50 kDa (Figure 4-7). As well as the *C. difficile* binding, all four of the mAbs show no binding to the closely related strains.

When analysed using the ExPASy compute pI/Mw tool (Gasteiger *et al.*, 2003) the amino acid sequence for the HMW SLP of *C. difficile* 630 was calculated to be 39.45 kDa. This weight is considerably lower than what the protein migrates to on the SDS-PAGE gel, at around 50 kDa. This difference in size has been noted in the literature and other studies suggesting that it is due to aberrant migration of the protein within the gels (Calabi *et al.*, 2001). Qazi and colleagues showed that although the HMW SLP runs at a larger size on SDS-PAGE gels, the actual mass, determined using mass spectrometry, was the same as predicted with ExPASy (Qazi *et al.*, 2009).

4.3.6 Specificity of mAbs to denatured *C. difficile* proteins

The four mAbs successfully bound to the two *C. difficile* strains, 630 and R002 at a band corresponding to the HMW SLP and to some smaller proteins that may have been produced from degradation of the HMW SLP (Figure 4-8). Additionally, all four mAbs did not bind to *S. aureus* or *B. subtilis*, which are in the same phylum as Clostridium. *S. aureus* was chosen for testing as it is a common commensal bacterium and also a facultative human pathogen that can be found in human faeces (Bhalla *et al.*, 2007; Ibarra *et al.*, 2013; Kim *et al.*, 2015).

Although *B. subtilis* was traditionally viewed as a soil microbe it has also been shown to reside in the gastrointestinal tract (Hong *et al.*, 2009) and was therefore also tested for mAb binding. None of the four mAbs bound to either *S. aureus* or *B. subtilis* and although many more bacterial species would need to be tested for non-specific binding, it is promising start. For use in PoN sensors, it is vital that the mAbs detect *C. difficile* only as non-specific binding would cause false positives which, could lead to misdiagnosis and mistreatment. Ideally, groups of bacteria that would be tested for non-specific binding with the *C. difficile* mAbs would be those that reside in the gastrointestinal tract, those often found in faeces and those associated with diarrhoea. However, due to time constraints, this project focussed on the three strains closely related to *C. difficile*. The closely related strains have the most similar sequences and proteins to *C. difficile*. Therefore, if the antibodies do not bind to these strains, then they are less likely to bind to more distantly related species, with less similar sequences and proteins.

Since the four mAbs, Ab491, Ab493, Ab521 and Ab652 all bound to the HMW SLP and did not bind to any of the four non *C. difficile* species, they were progressed to the next stage of binding analysis. Furthermore, the intellectual property (IP) rights and cell lines for these promising mAbs were purchased from AbMART, enabling production of the mAbs to be performed in house.

4.4 Summary

The *C. difficile* biomarker was shown to be present within three, tested, *C. difficile* genomes, verifying what was seen in the predictions from the bioinformatics work. Eight mAbs were produced spanning the biomarker and from these eight, four were found to bind to the native *C. difficile* proteins within dot blots and the denatured *C. difficile* HMW SLP. The four mAbs did not bind to native *C. sordellii* proteins nor did they bind to denatured proteins from the closely related strains, *C. sordellii* and *P. anaerobius*. Additionally, these Abs did not bind to the common bacteria *S. aureus* or *B. subtilis*. The four putatively specific mAbs were, Ab491, Ab493, Ab521 and Ab652 and the IP rights and cell lines that produce these mAbs were purchased from AbMART. The remaining four mAbs were disregarded since their potential for use in a PoN sensor to detect *C. difficile* is low.

Chapter 5 Characterisation of the relative binding efficiency of Ab491, Ab493, Ab521 and Ab652

5.1 Introduction

5.1.1 The binding affinities of monoclonal antibodies

As described in Chapter 4.1.5, antibodies bind to antigens through transient, non-covalent interactions. The strength of this binding is termed the binding affinity and the tighter the complementary fit of the antibody/antigen complex, the stronger the interaction (Elgert, 2009). The antibody binding affinity is determined by the equilibrium dissociation constant (K_d) and provides a measure for the proportion of unbound antibody and antigen when at equilibrium (Laguna *et al.*, 2015). The K_d is expressed in molar units of concentration (M) of the free antibody and antigen. The lower the K_d, the fewer dissociated antibody and antigen molecules there are and thus, more antibody/antigen complexes are present, indicating a stronger binding affinity of the antibody (Craig *et al.*, 2010).

A. $Ab + Ag \rightleftarrows AbAg$

B. $Kd = \frac{[Ab][Ag]}{[AbAg]}$

Equation 5-1. The reversible reaction of the bound antibody/antigen (AbAg) complex forming from the individual antibody (Ab) and antigen (Ag) molecules (A). The calculation of the K_d, the ratio of the concentration of unbound antibody and Ag, and the concentration of the AbAg complex (B).

Several methods can be used to determine the K_d of mAb interactions. One such method is isothermal titration calorimetry (ITC), a label-free method to measure the dissipated or absorbed heat of a reaction (Jerabek-Willemsen *et al.*, 2011). Using ITC, both enthalpic and entropic binders can be differentiated, with the K_d and number of binding sites per molecule being obtained (Liang, 2008). However, in order to get a quantifiable reaction, ITC requires a high concentration of sample at a minimum of 1 mg/ml. Furthermore, interactions with a small enthalpy change cannot be measured with ITC (Jerabek-Willemsen *et al.*, 2011). This high-sample consumption and inability to measure small changes are limitations of ITC. Another method that can be used to determine molecular interactions is surface plasmon resonance (SPR). SPR is a label-free technique that measures changes in

binding-induced refractive index. Unfortunately, this method requires the immobilisation of one of the sample components to the surface of the SPR chip, which is a time consuming process (Sundberg *et al.*, 2007).

5.1.2 Microscale thermophoresis

Thermophoresis is the directed movement of molecules induced by temperature gradients (Wienken *et al.*, 2010). Microscale thermophoresis (MST) uses this thermophoretic principle to characterise the formation of small molecule complexes in solution. During MST analysis, one of the samples is labelled with a fluorophore and the movement of this molecule is measured using a charge-coupled device camera (Zillner *et al.*, 2012). An overview of how the molecular movement is measured is provided in Figure 5-1, which was kindly provided by NanoTemper[24]. The 'initial state' of the baseline fluorescence is measured so that homogenous distribution of the labelled molecules can be monitored. An infra-red (IR) laser is then applied to a focused heating area, causing a temperature jump (T-jump) and the molecules typically migrate to a cooler temperature. A steady-state is reached when the movement away from the T-jump is counterbalanced by mass diffusion – the spread of molecules in a solution (Seidel *et al.*, 2013). In constant buffer conditions, the thermophoresis of the antibody complex with the fluorescently labelled antigen differs compared to the antigen alone. This change in thermophoresis is caused by the alteration in size, charge and solvation entropy of the molecules (Jerabek-Willemsena *et al.*, 2014). When the IR laser is switched off, an inverse T-jump occurs along with back diffusion of the fluorescent molecules (Entzian and Schubert, 2015).

[24] http://www.nanotemper-technologies.com/

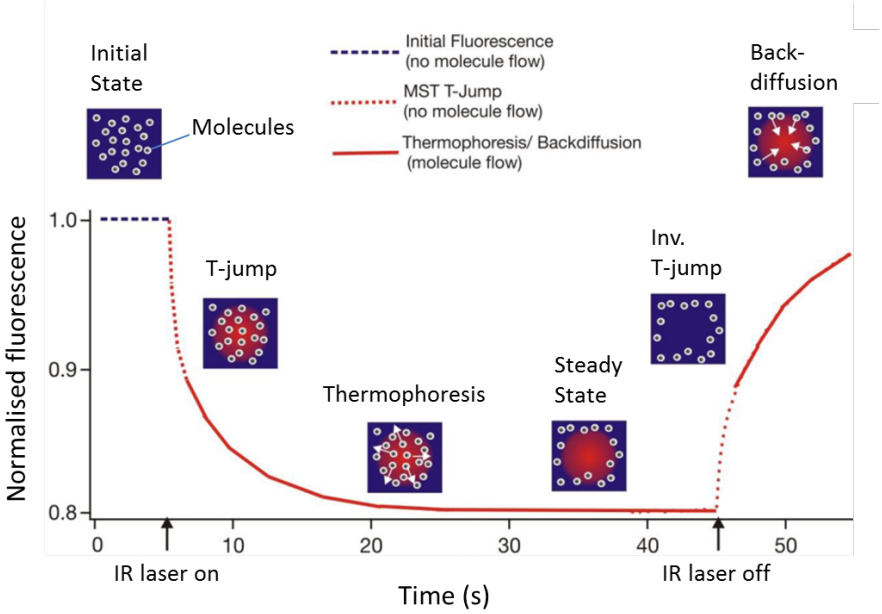


Figure 5-1. MST measurement of a fluorescently labelled molecule interacting with another molecule, provided by NanoTemper. The initial state of the baseline fluorescence is measured prior to the IR laser being turned on. The IR laser is applied and a focussed area is heated, causing a T-jump from which molecules generally diffuse. A steady state is achieved when the movement of molecules away from the T-jump is counterbalanced by mass diffusion. The IR laser is then turned off causing an inverse T-jump and the molecules diffuse back across the cooler temperature.

MST has several advantages over more traditional methods for determining binding interactions. Firstly, unlike SPR, MST is immobilisation-free, with a simple protocol that reduces the preparation time of samples. Secondly, MST is performed in micrometre-sized capillaries. Unlike ITC, only a small concentration of sample is required, often in the nanomolar range (Wienken *et al.*, 2010; Liu *et al.*, 2015b). Thirdly, MST measures molecule movement optically and is entirely contact-free, minimising possible contamination of the sample. Finally, MST is rapid, with measurements and associated K_d calculated within approximately 15 minutes (Wienken *et al.*, 2010; Entzian and Schubert, 2015). In consideration of the rapidity of testing and minimal sample consumption, MST was used to determine the K_d of the mAbs within this project.

5.1.3 ELISA

The ELISA technique was invented in 1971 by Eva Engvall and Peter Perlmann (Engvall and Perlmann, 1971). An ELISA measures the concentration of an analyte, in its native form, via the binding of an antibody. If equal concentrations of analyte are used, then the relative rate of antibody binding can be quantified (Ozdemir *et al.*, 2013). ELISAs are often used for

the analysis of antibody binding. With regards to *C. difficile*, the majority of Abs analysed bind to the toxins (Fraser and Swiencicki, 2013), however, ELISAs have been used with other *C. difficile* antigens. Kandalaft and colleagues used ELISA to determine llama antibody binding to the SLPs of whole *C. difficile* cells. Rather than for use in diagnostics, these llama Abs were shown to inhibit *C. difficile* motility and were hypothesised to be possibly of therapeutic value (Kandalaft *et al.*, 2015). Research by Merrigan and co-workers also provided evidence that SLP Abs could be used therapeutically, by showing that these Abs reduced *C. difficile* attachment to the host (Merrigan *et al.*, 2013).

There are several types of ELISA – direct, indirect, sandwich and competitive – with each having a slightly different format (Gan and Patel, 2013). Throughout this project the indirect ELISA method was performed. The antigen was bound directly onto the surface of the ELISA plate, with the primary antibody of interest then being then bound to the antigen and then finally an enzyme-conjugated secondary antibody bound to the primary Ab.

5.1.4 Isotyping and purification of mAbs

Abs can have multiple variants of both the light and heavy chains. The phenotypic deviations in the constant regions of the heavy chains define the Ig isotype, causing differing functions and antigen responses (Abbas and Lichtman, 2004). As also observed in humans, there are also five mouse antibody isotypes: IgA, IgD, IgE, IgG, and IgM. IgG also has five subclasses: IgG1, IgG2a, IgG2b, IgG2c, IgG3 (Michaelsen *et al.*, 2004; Johnson, 2013). Abs are often purified using an affinity chromatography method that utilises proteins that bind to specific Igs. These proteins are crosslinked to agarose beads and used within a flow through column. When the cell culture media, in which the Abs are produced, is washed through the column, the Abs form a complex with the proteins attached to the agarose beads (Sugiura *et al.*, 2000). The non-covalent interactions binding the antibody and protein are then disrupted with extreme pH, eluting the mAbs from the agarose beads (Janeway *et al.*, 2001). There are two main proteins that are used in immunoprecipitation, Protein A and Protein G. Protein A is a cell wall protein from *Staphylococcus aureus* (Graille *et al.*, 2000) and Protein G, a cell wall protein isolated from Group G streptococci (Sjöbring *et al.*, 1991). Both proteins bind mammalian IgGs via their fragment crystallisable (F_c) regions however, protein G binds at greater avidity than protein A (Akerstrom *et al.*, 1985).

5.1.5 The *C. difficile* SLT clades

As described in section 3.3.4 all currently sequenced *C. difficile* strains can be grouped into 13 SLT clades. Dr Kate Dingle not only provided the sequences of the 13 SLTS for the bioinformatics work but also kindly donated the corresponding strains for use within the laboratory experiments. A 14th SLTS was also sent, Ox2404, which resides in SLT clade 2, along with Ox858. Although the SlpA sequences of the two strains were the same, there were differences in other SLPs. Ox2404 lacks Cwp10 and Ox858 lacks the 5' Cwp84, both of which are cell wall proteins and therefore may influence binding of the mAbs. Furthermore, Cwp84 is the protease that cleaves SlpA into the mature LMW and HMW SLPs (Pantaléon *et al.*, 2015) and consequently any alterations to this protein may directly affect the structure of the HMW SLP and thus mAb binding. For these reasons, all 14 SLTS were used throughout the cell mAb binding analysis.

5.2 Results

5.2.1 Relative binding of the four mAbs

Direct ELISAs were performed with whole cells of *C. difficile* R002 and R050, clinical isolates 118497G, 128703G, 994535 and the reference strain 630. The closely related strains, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330 were also used as negative controls. The cells were incubated in duplicate with Ab491, Ab493, Ab521 or Ab6521 and the protocol was followed as in methods section 2.16. The ELISA results show that the two closely related species, *P. anaerobius* and *C. sordellii* provided very low absorbance readings for Ab491, Ab521 and Ab652 (Figure 5-2). The low absorbance readings indicate that there was little binding to these species. When incubated with *P. anaerobius*, Ab521 had the lowest absorbance reading with Ab652 and Ab491 displaying slightly higher values. Ab493 on the other hand, bound to *P. anaerobius* at 0.13, a rate that was double that observed with the other three mAbs, Ab652 being the highest at 0.50. This measurement was greater than the absorbance of Ab493 against four out of the six *C. difficile* strains, 630 – 0.09, R050 – 0.08, 118497G – 0.09, and 994535 – 0.02.

Upon incubation with *C. sordellii*, Ab521 and Ab652 gave the highest absorbance values of the four mAbs, however, there is a clear increase in absorbance when incubated with *C. difficile*. Ab521 produces an absorbance value of 0.02 with *C. sordellii* and 0.26-0.64 when incubated with *C. difficile* and Ab652 0.05 with *C. sordellii* and 0.23-0.46 with *C. difficile*. The difference in absorbance between the *C. difficile* and closely related strains was, depending on the strain, four to eight times greater for Ab652 and eight to twenty-one times greater for Ab521.

Within the *C. difficile* strains, Ab521 and Ab652 had the greatest absorbance values. Ab521 gave the highest *C. difficile* absorbance values, with the exception of strain 630 where Ab652 was greater. For five of the six *C. difficile* strains, R002, R050, 630, 118497 and 128703G, Ab491 and Ab493 displayed absorbance measurements that were less than half of those seen with Ab521.

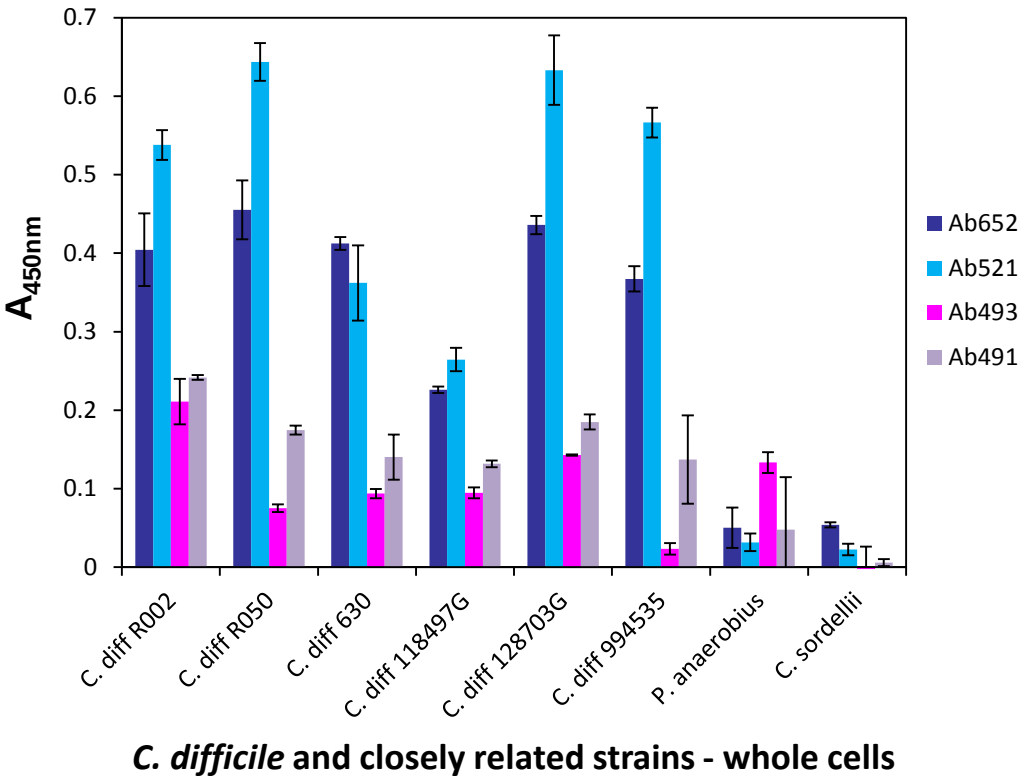


Figure 5-2. Absorbance readings from an ELISA of antibodies Ab491, Ab493, Ab521 and Ab652 binding to whole cells of *C. difficile* (*C. diff*) R002, R050. 630, 118497G, 128703G and 994535, and closely related species *P. anaerobius* VPI 4330 and *C. sordellii* ATCC 9714. The error bars show the standard deviation of the two replicates for each sample.

These ELISA results show that Ab521 and Ab652 bind with better efficiency and more specificity to *C. difficile* than Ab491 and Ab493. Due to these ELISA results, Ab491 and Ab493 were disregarded and the remainder of the project was performed focussing on the two more promising mAbs, Ab521 and Ab652.

5.2.2 Production, purification and isotyping of the mAbs

The isotypes of the mAbs were determined, IgG2b for Ab521 and IgG1 for Ab652. The mAbs were purified from the cell culture supernatant, around 3 ml of each mAb was purified at concentrations of 2.45 mg/ml for Ab521 and 0.27 mg/ml for Ab652. An example of the chromatograph produced with the ÄKTA and Unicorn start software is provided in Figure 5-3 A and an SDS-PAGE gel of Ab521 and Ab652 is shown in Figure 5-3 B. The gel displays protein bands for both mAbs at around 55 kDa and 25 kDa, which approximately equate to the heavy and light chains of the antibodies, respectively.

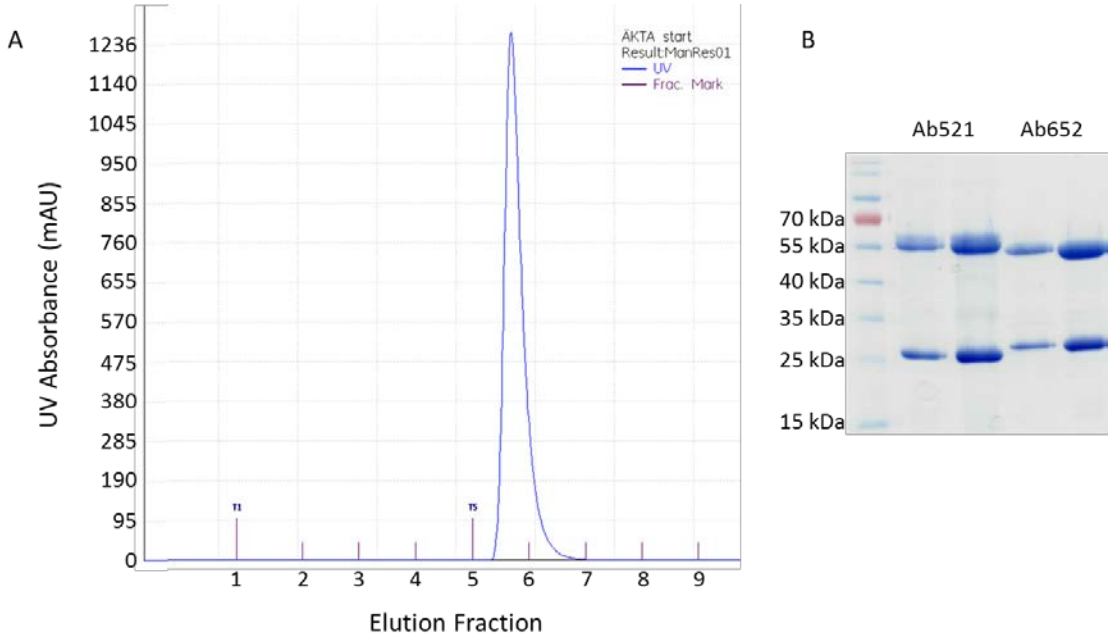


Figure 5-3. Chromatograph from the purification of Ab521 from cell culture media using a protein G agarose column on the ÄKTA Start chromatography system (A). The low pH elution buffer is added at elution fraction 1, eluting the antibody from the protein G column, as represented by the peak in the UV absorbance (blue line). The elution fractions that are predicted to contain antibody were separated on an SDS-PAGE gel to determine the presence and purity of antibody (B). Both Ab521 and Ab652 were separated on the reducing gel and show protein bands at around 55 kDa and 25 kDa.

5.2.3 Production of recombinant *C. difficile* 630 LMW SLP

E. coli Rosetta pLMW1-262-1 which was provided by Professor Neil Fairweather[25], was cultured and the recombinant protein was determined to be in the soluble fraction. Large quantities of bacteria were grown and the media was purified using immobilised metal affinity chromatography with a nickel column Figure 5-4. This plasmid produces a stable, truncated LMW SLP that is 262 amino acids in length and has a molecular weight of 28 kDa. As seen with the native HMW and LMW SLPs, the recombinant LMW SLP runs aberrantly on SDS-PAGE gels with an apparent molecular weight of 33 kDa.

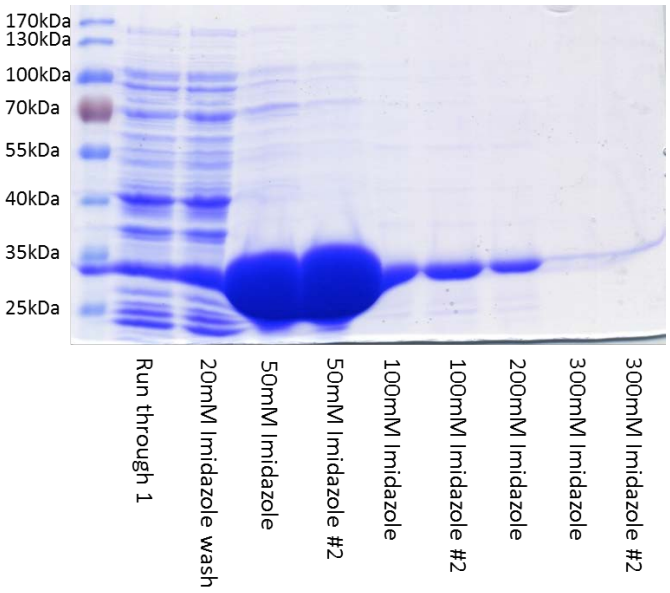


Figure 5-4. SDS-PAGE gel (10%) of the fractions from the nickel column purification of the recombinant *C. difficile* 630 LMW SLP. All fractions were in 25 mM Tris/HCl, 200 mM NaCl, pH 7.5 buffer and the stated imidazole concentration. The LMW SLP eluted from the column with 50 mM imidazole and separated on the gel at around between 25 and 35 kDa.

5.2.4 Specificity of the mAbs to the HMW SLP from the 14 SLTS

As shown in Figure 5-5, Ab521 (5 µg/ml) bound specifically to the HMW SLP of all eight SLTS, and did not bind to either naturally occurring or the recombinant LMW SLP. There were however, some double bands within these Western blots, corresponding approximately to the molecular weight of the HMW SLP (Figure 5-5). The SLTS Ox1145 had a clear double band whereas, Ox575 and Ox2404 displayed fainter double bands. The extra Ox1145 band could also be seen on the SDS-PAGE gel, therefore, the band contained enough protein to

[25] Department of Life Sciences, Imperial College London

be detected by the coomassie stain. The molecular weight of the additional band was marginally lower than that of the HMW SLP and therefore could be indicative of degradation of this protein.

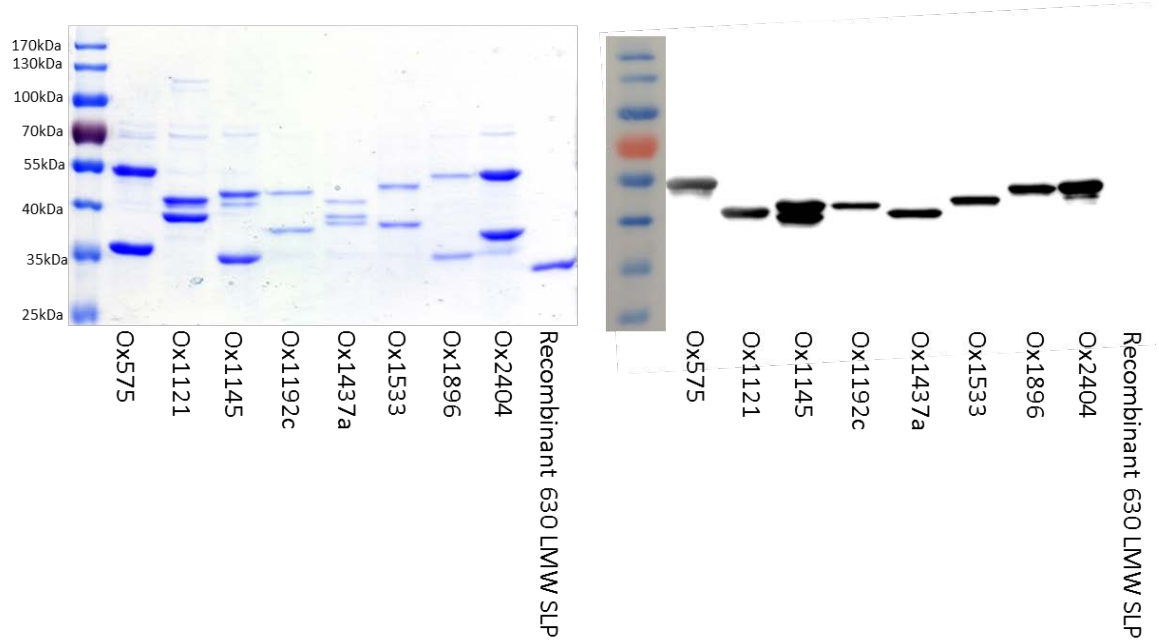


Figure 5-5. SDS-PAGE gel 10 % and corresponding Western blot using Ab521, 5 µg/ml. The samples separated on the gel were the low pH extracted surface layer proteins of eight *C. difficile* S-Layer type strains and the recombinant *C. difficile* 630 LMW SLP. The Ab521 binds specifically to the HMW SLP from all SLTS.

Ab652 (5 µg/ml) also bound specifically to the HMW SLP within the extracted surface layer proteins, and did not bind to any of the LMW SLPs (Figure 5-6). As seen for Ab521, there were also some double bands visible for strains Ox1145, Ox575 and Ox2404 within the Ab652 Western blot.

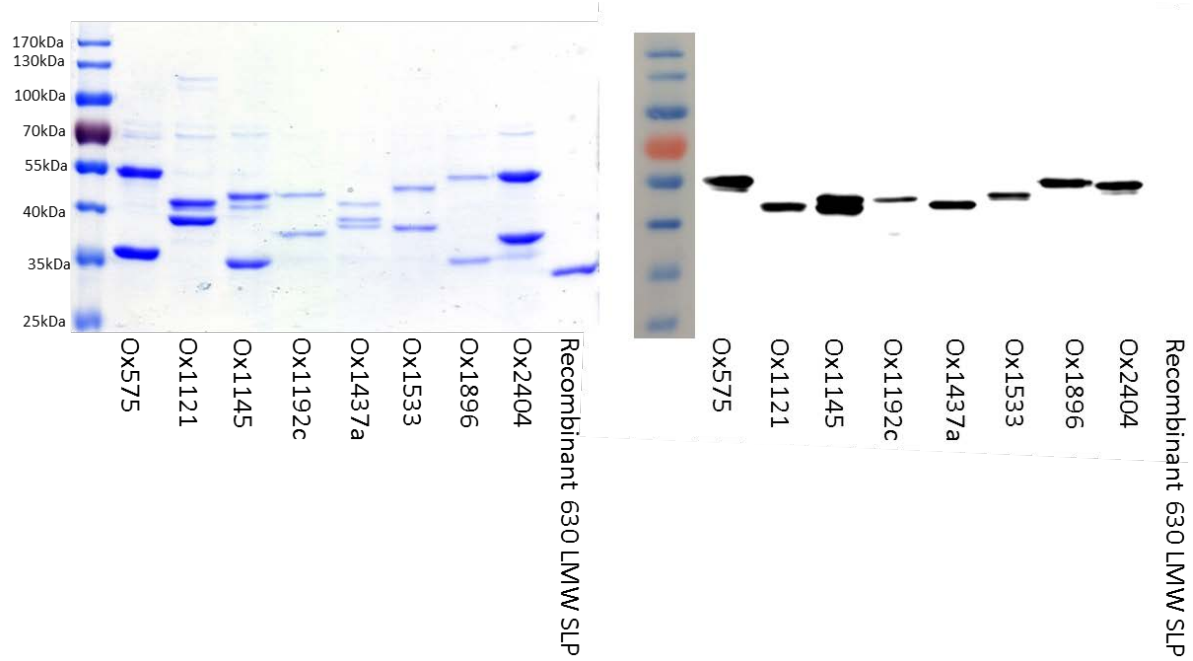


Figure 5-6. SDS-PAGE gel 10 % and corresponding Western blot using Ab652, 5 µg/ml. The samples separated on the gel were the low pH extracted surface layer proteins of eight *C. difficile* S-Layer type strains and the recombinant *C. difficile* 630 LMW SLP. The Ab521 binds specifically to the HMW SLP from all SLTS.

Rather than test the mAbs against the SLPs of the remaining six SLTS, the whole cell lysates for all fourteen SLTS were tested for mAb binding. These results would not only determine binding to the HMW SLP in all fourteen SLTS, but they would also provide evidence that the mAbs do not non-specifically bind to proteins present within whole cell lysates.

5.2.5 Specificity of Ab521 and Ab652 to proteins within whole cell lysates

It was important to test the mAbs against as many proteins as possible from the closely related strains. For this reason, whole cell lysates were used as the samples within Western blots. *C. hiranonis* DSM-13275, which displays reasonably high similarity to the SlpA protein and to the biomarker, was included as one of the closely related strains. Also used were *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330, along with whole cell lysates from all fourteen SLTS. The Western blots were used to test for binding with both Ab521 and Ab652. As shown in Figure 5-7, Ab521 at 5 µg/ml binds to the HMW SLP within all fourteen SLTS and not to any proteins within the closely related strains. Furthermore, the positive binding occurs in one clear band, demonstrating the specificity to the HMW SLP and that the previous Western blot (**Figure 5-5**) may have shown some degradation of the HMW SLP.

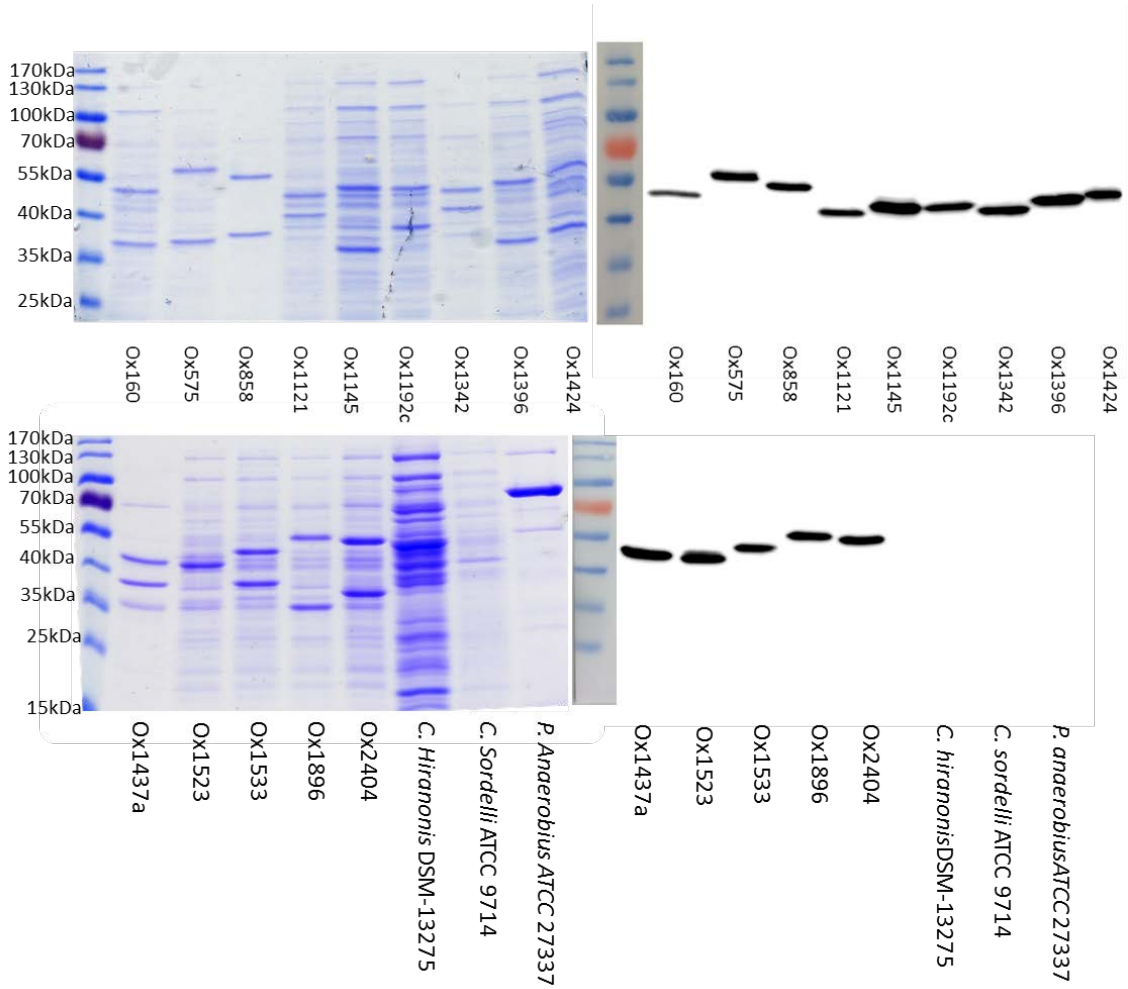


Figure 5-7. Western blots and the corresponding SDS-PAGE gels using Ab521 at 5 µg/ml against the whole cell extracts of all 14 *C. difficile* SLTS and the closely related strains, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. Ab521 binds to the HMW SLP in all *C. difficile* SLTS and does not bind to proteins in the whole cell lysates of the closely related strains.

As seen with Ab521, Ab652 also binds specifically to the fourteen SLTS, with one clear band, and does not show binding to the closely related strains (Figure 5-8). The results for the Western blot against Ox160 – Ox1424 was performed with 5 µg/ml Ab652, which showed clear binding but not as strong as seen with Ab521, although binding was not quantified. Therefore, to verify that non-specific binding to the closely related strains was not occurring, the Western blot with these strains was performed with Ab652 at a higher concentration of 10 µg/ml.

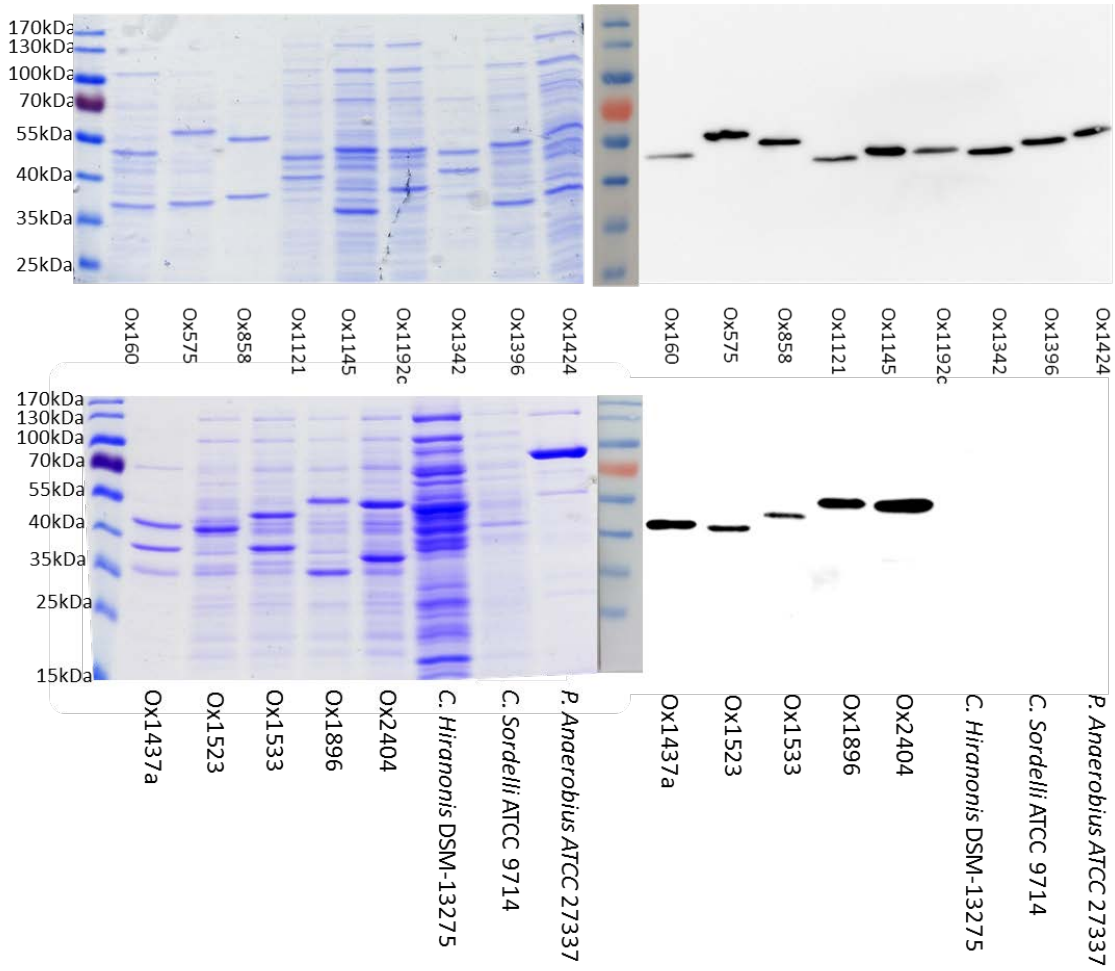


Figure 5-8. Western blots and the corresponding SDS-PAGE gels using Ab652 at 5 µg/ml (top) and 10 µg/ml (bottom) against the whole cell extracts of all 14 *C. difficile* SLTS and the closely related strains, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. Ab652 binds to the HMW SLP in all *C. difficile* SLTS and does not bind to proteins in the whole cell lysates of the closely related strains.

5.2.6 MST analysis of Ab521 and Ab652 binding to the HMW/LMW complex

The normalised fluorescence (F_{Norm}) is a ratio of the 'steady state' fluorescence after thermodiffusion (F_1) divided by the 'initial state' (Figure 5-1) of fluorescence before laser heating (F_0) (Equation 5-2). If there is binding of the unlabelled and labelled proteins then the F_{Norm} creates a binding curve, which displays the unbound molecules and an increase in F_{Norm} for the bound molecules.

$$F_{Norm} = \frac{F_1}{F_0}$$

Equation 5-2. The equation used to determine normalised fluorescence (F_{Norm}) from microscale thermophoresis data. F_{Norm} is the ratio of the fluorescence after thermodiffusion (F_1) divided by the initial fluorescence before laser heating (F_0).

The binding curve created with the MST measurements of Ab521 interacting with the *C. difficile* 630 HMW/LMW SLP complex (black diamonds Figure 5-9), shows the typical sigmoidal curve that is created when binding between molecules occurs (Jerabek-Willemsena *et al.*, 2014). The unbound HMW/LMW complex at the lower Ab521 concentrations provides an 'unbound' reading of around F_{Norm} [‰] 808 units. As the antibody concentration increases so too did the amount of complex bound with Ab521, reaching a maximum fluorescence value that plateaus and provided a 'bound' reading of around F_{Norm} [‰] 832 units. A K_d of 5.21±0.917 nM was calculated from the sigmoidal binding curve using the NanoTemper analysis software. .

The blue triangular markers display the thermophoretic movement of the control, the HMW/LMW SLP complex with only PBS with 0.05 % TWEEN and no antibody (Figure 5-9). Within all 16 control measurements, the fluorescence readings all remain within the 'unbound' value range. These readings show that the binding curve produced by the interaction of the HMW/LMW complex and Ab521 is not an artifact of usual thermophoretic movement of the complex alone.

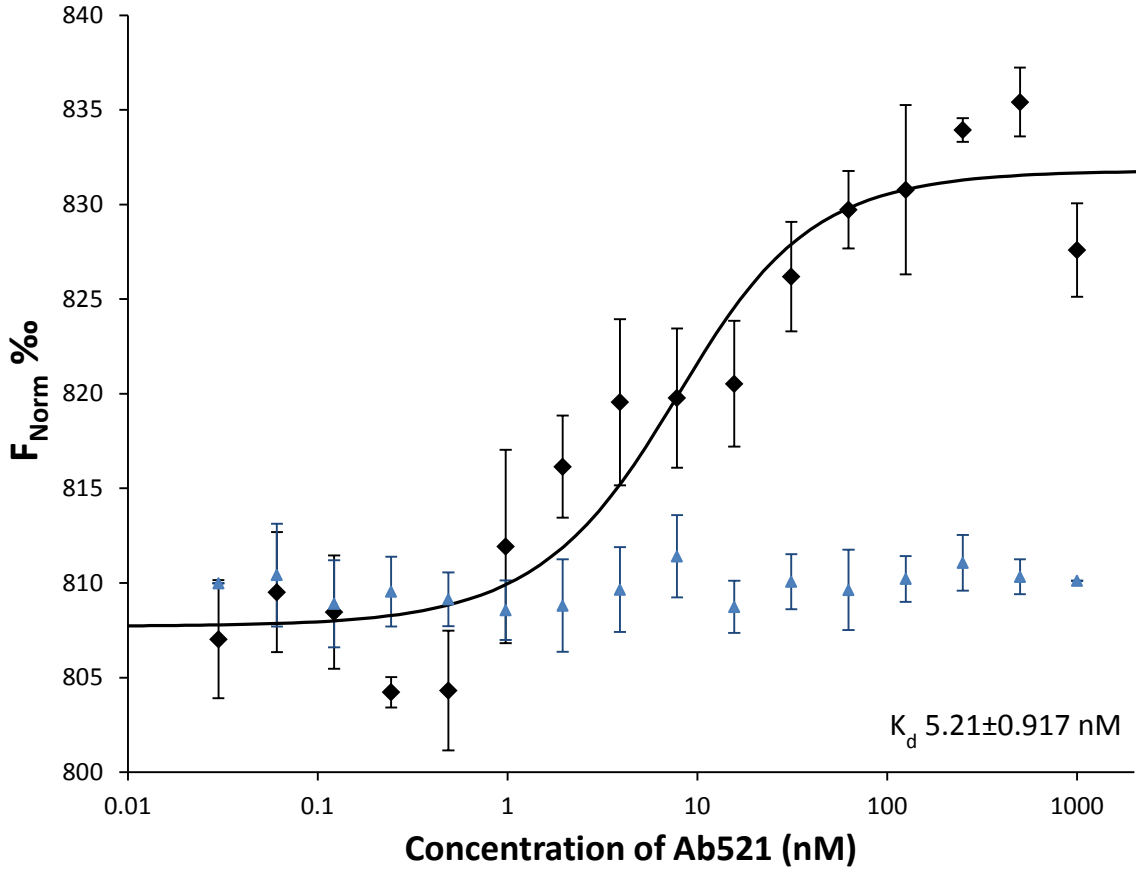


Figure 5-9. Binding curve produced by microscale thermophoresis analysis of the fluorescently labelled HMW/LMW complex interacting with either PBS with 0.05 % TWEEN (Blue triangles) or increasing concentrations of Ab521 (black diamonds). FNORM is the normalised fluorescence

The MST testing of the HMW/LMW SLP complex and Ab652 yielded no significant binding results and a K_d was not attained. Perhaps further optimisation of the experimental conditions could enable clear interactions between the SLP complex and Ab652. However, due to time constraints, it was not possible to perform these optimisation steps within this project.

5.3 Discussion

5.3.1 Ab521 and Ab652 are the most promising of the four mAbs

The absorbance values from the whole cell ELISA (Figure 5-2), are directly proportional to the amount of primary antibody bound to the whole cell antigens. For PoN sensors it is important that the antibodies bind to the organism of interest, with high affinity, allowing rapid and clear detection (Ahmed *et al.*, 2014). Weak binding may not clearly identify

bacteria from complex samples, such as faecal samples tested for CDI. Thus, the results show that due to their stronger binding, Ab521 and Ab652 are the most promising of the four mAbs for *C. difficile* detection. Furthermore, both of these mAbs show very little binding to the closely related species, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330 – up to 21 times less – when compared to the *C. difficile* strains. Thus, Ab521 and Ab652 would be highly unlikely to cause false positives by interacting with these two organisms (Owen, 2003). In contrast, Ab493 bound to *P. anaerobius* with greater efficiency than to four out of the six tested *C. difficile* strains. Therefore, accurate diagnosis would not be possible, as this binding would produce false positives, at least to this species, within a diagnostic sensor setting. The low rate of Ab493 binding to the *C. difficile* strains also limits the use of this molecule within a PoN sensor. Ab493 would be unlikely to detect *C. difficile* strains with similarity to R050, 630, 118497G and 994535, causing any sensor to have poor sensitivity. Although slightly higher than Ab493, poor binding to the *C. difficile* strains was also displayed with Ab491. This lack of binding is evidence that neither Ab491 nor Ab493 would be useful for use as detection molecules within PoN sensors.

One possible reason for the difference in mAb binding could be due to the 10 amino acid epitope against which they were produced. Ab521 and Ab652 displayed elevated binding, and both have epitopes that are 100 % conserved within the predicted *C. difficile* biomarker. Whereas, the epitopes for Ab491 and Ab493 are only partially found within the unique biomarker (Table 5-1).

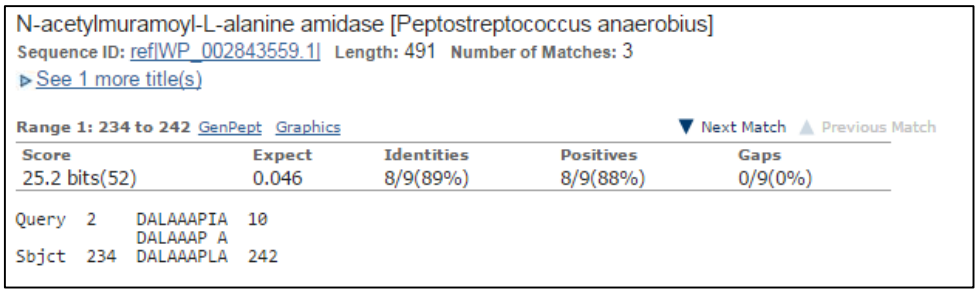
Table 5-1. The antibody ID and corresponding epitope to which the mAb was produced. The region found within the predicted C. difficile biomarker is highlighted.

Antibody ID	Epitope sequence (biomarker highlighted)
Ab491	FVAKDGSTKE
Ab493	VDALAAAPIA
Ab521	STKEDQLVDA
Ab652	KEDQLVDALA

Only the regions that are fully conserved within the biomarker are predicted to be unique and surface exposed. Therefore, regions of the Ab491 and Ab493 epitopes that are not located within the biomarker may not be fully surface accessible, causing the lower rate of antibody binding. The Jalview hydrophobicity view of the sequences, within the

bioinformatics results (Figure 3-11) show that the ten amino acid epitope for Ab493 has only two amino acids that are hydrophilic: D and P. Consequently, the epitope may not be completely exposed on the surface of the HMW SLP and could be the reason for the low antibody binding to the *C. difficile* strains. The epitope for Ab491 on the other hand has seven amino acids that are hydrophilic, which could explain the stronger binding to *C. difficile* than Ab493. Although the number of hydrophilic residues could indicate surface exposure of the epitope, Ab521 has seven hydrophilic residues and Ab652 has less, yet both mAbs have much greater binding efficiency than Ab493. Therefore, the hydrophilic nature of the amino acids is not the only factor in antibody binding to its antigen. The lower binding efficiency of Ab491 could perhaps be due to the inability of Ab491 to bind completely to the native protein.

The regions outside of the biomarker may not be unique to *C. difficile* and could cause cross-reactivity of the mAbs with other organisms. A BLASTP search of the Ab493 epitope identified that eight from the ten amino acid sequence are located within a surface protein of *P. anaerobius* (Figure 5-10). Therefore, it is possible that Ab493 recognised this sequence within *P. anaerobius*, causing the non-specific binding. The BLASTP results for Ab491, Ab521 and Ab652 epitopes did not identify sequences with a high percentage identity within *P. anaerobius,* data shown in the Appendices section 9.8. This lack of sequence identity was predicted, as there was little binding of the three mAbs to *P. anaerobius*.


N-acetylmuramoyl-L-alanine amidase [Peptostreptococcus anaerobius]
Sequence ID: ref|WP_002843559.1| Length: 491 Number of Matches: 3
See 1 more title(s)

Range 1: 234 to 242 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
25.2 bits(52)	0.046	8/9(89%)	8/9(88%)	0/9(0%)

```
Query  2    DALAAAPIA  10
            DALAAAP A
Sbjct  234  DALAAAPLA  242
```

Figure 5-10. Most similar BLASTP result of the Ab493 epitope 'VDALAAAPIA' searched within *P. anaerobius*. Eight out of the ten amino acids are matched to a surface protein within the organism.

5.3.2 Purification of Ab521 and Ab652

The cell lines for two of the mAbs, Ab521 and Ab652, were grown in large quantity and cell death was permitted to produce antibody. Upon isotyping, Ab521 was found to be IgG2b

and Ab652 was IgG1. Since Protein A has extremely poor reactivity with mouse IgG1 (Kronvall *et al.*, 1970), Protein G was chosen as the affinity chromatography protein. For both Ab521 and Ab652, 260 ml of culture supernatant was purified, producing 3 ml at 2.45 mg/ml and 0.27 mg/ml respectively. Therefore, the minimum starting concentration of the mAbs within the hybridoma supernatant was 28 µg/ml for Ab521 and 3 µg/ml for Ab652, which are within the expected range of 1-60 µg/ml (Pandey, 2010).

5.3.3 Specific binding of the mAbs to HMW SLP in all 14 SLTS

For the initial Western blots within this chapter, the extracted SLPs of the SLTS were used as the protein samples, with both mAbs displaying binding to more than one band on some of the strains. The same protein samples were used in the Western blots for both mAbs, therefore, it seems likely that the extra bands were due to HMW SLP degradation rather than non-specificity of the mAbs. The Western blots performed with whole cell lysates were immediately loaded onto an SDS-PAGE gel, to reduce the likelihood of any degradation. Within these Western blots, both Ab521 and Ab652 bound to a single band at the corresponding HMW SLP size for all fourteen SLTS. This binding showed specificity to the HMW SLP and consequently established that the previous double bands were due to degradation.

The SDS-PAGE gels for the extracted *C. difficile* SLPs and the whole cell lysates all displayed the HMW and LMW SLP, apart from Ox1523, which only had a single predominant band consistent with the HMW SLP. As described in section 6.3.2, *C. difficile* strains, such as ribotype 167, have been shown in the literature to contain a single dominant band (Calabi and Fairweather, 2002). To determine their similarity, the R167 SlpA sequence was taken from NCBI and aligned with the SlpA sequence for Ox1523. Only 2 out of 610 amino acids were different so, R167 belongs to SLT clade 11 along with Ox1523, with this similarity displaying that the single dominant band was not an anomaly. Whilst Kate Dingle and colleagues were creating the SLT clades they determined that R050 also belongs within SLT clade 11 (Dingle *et al.*, 2013). As seen with Ox1523 and R167, R050 also displayed a single dominant band, shown in the SDS-PAGE gel within Figure 4-4.

Using the sequences from the full *C. difficile* genome, the biomarker was shown to be unique to the HMW SLP and not present within any other *C. difficile* proteins. As predicted the mAbs produced against this biomarker bound solely to the HMW SLP. The mAbs did not bind to the 28 SlpA paralogs that are present within *C. difficile* (Sebaihia *et al.*, 2006) showing specificity to the HMW SLP even when compared to closely related *C. difficile* proteins. These proteins were denatured and so the Western blots do not provide evidence for HMW SLP specificity with native *C. difficile* proteins.

For the whole cell Western blots with SLTS Ox160 to Ox1424, both mAbs were used at 5 µg/ml. The signal for Ab652 appears weaker compared to Ab521 however, the signals were not quantified because more accurately quantifiable tests will be used to determine relative binding affinities in chapter 8. When the Ab652 concentration was increased to 10 µg/ml for the whole cell Western blot of Ox1437a to Ox2404 and the control strains, the signal seemed to be similar to that seen in the same Western blot for Ab521 used at a concentration of 5 µg/ml. This concentration difference indicates that Ab521 may have a better binding efficiency than Ab652, however, further experimental work was performed in chapter 8 to determine if this was correct.

All of the *C. difficile* strains, which have currently had their genome sequenced, fall within one of the 13 SLT clades, therefore, the SLTS represent the entire *C. difficile* species. Both Ab521 and Ab652 bind to all of the SLTS, indicating that the mAbs would bind to the denatured HMW SLP of any *C. difficile* strain. This detection of all strains within the species is pivotal in preventing false negatives within a PoN sensor. The next step was to determine whether Ab521 and Ab652 could detect strains from all of the native SLTS. The lack of binding to the closely related strains, as shown by the mAbs, is also vital, as it will help prevent false positives in a PoN sensor. There were three closely related strains tested for mAb binding, *P. anaerobius*, *C. sordellii* and *C. hiranonis*. Therefore, further specificity testing of Ab521 and Ab652, with more bacterial species is required. To investigate the mAb specificity fully, testing of *C. difficile* and other species was performed under native conditions.

5.3.4 Ab521 displays a high binding affinity with the HMW/LMW SLP complex

Attempts were made to produce recombinant HMW SLP, as this could be used directly within MST, however, these attempts were unsuccessful. Upon discussion with Dr Paula Salgado[26], an expert on the *C. difficile* S-Layer, it was recognised that the HMW SLP is unstable and recombinant production achieved only a minimal amount of the protein. In contrast, when the HMW SLP forms the native HMW/LMW SLP complex, it is stable and can be isolated directly from *C. difficile* cells, in a quantity sufficient for MST analysis (Fagan *et al.*, 2009). Therefore, the HMW/LMW complex was used for mAb binding analysis, with the protein complex being fluorescently labelled and the mAb titrated against it.

The MST analysis of Ab521 interacting with the HMW/LMW complex provided a binding curve, from which the binding affinity was determined to be 5.21±0.917 nM. In general terms, antibodies with a K_d less than 1 µM have high affinity (Laguna *et al.*, 2015), therefore, Ab521 is likely to have a very strong affinity to the HMW/LMW SLP complex. Optimisation of the experiment may provide improved 'unbound' measurements that show less fluctuation than seen within these results. Further optimisation may too provide readings that fit closer to the binding curve and therefore, provide a K_d with increased accuracy. Optimisation could include, the buffer used, agents used to prevent non-specific binding and the cappillaries used.

Unfortunately, the MST yielded no significant results with Ab652. One of the issues with using Ab652 is the low concentration in comparison to Ab521, 0.21 mg/ml and 2.41mg/ml respectively. Further antibody production may provide a higher concentration or the Ab652 sample could be concentrated using a centrifuge concentrator or vacuum concentrating system, such as a SpeedVac. If optimisation of the MST analysis is not successful, then more traditional methods could be used to determine the K_d of the mAbs. Since the HMW SLP cannot be produced in large quantities, the sample availability is relatively limited and therefore, SPR would be performed as opposed to ITC.

Another possible explanation for the lack of good quality MST data could be due to the presence of the fluorophore label of the HMW SLP potentially inhibiting the Ab521 binding

[26] Institute for Cell and Molecular Biosciences, Newcastle University

site. The protein was labelled with the fluorophore via amine coupling. Another NanoTemper kit is available that couples the fluorophore via cysteine residues, however, the HMW/LMW SLP complex does not contain cysteine and consequently this kit could not be used.

Since the MST technique was not successfully optimised for interactions between the mAbs and HMW/LMW SLP complex, the analysis of the mAb interaction with the recombinant LMW SLP also could not be performed. This interaction would be a vital control for any binding analysis of the mAbs and would be performed with the optimised MST method or using SPR for future work. The LMW SLP produced was a truncated version of the protein, missing the latter 59 residues. As a result of this truncation a stable protein is formed that could be used for downstream analysis. A consequence of the missing amino acids is that when used as a negative control for mAb binding, the full LMW SLP is not tested. Therefore, there is a possibility that the mAb could be binding to these residues when analysed for binding to the HMW/LMW SLP complex.

5.4 Summary

Using whole-cell ELISAs, the four promising mAbs were reduced to two, Ab521 and Ab652. These two mAbs displayed greater binding to the *C. difficile* strains and demonstrated greater specificity when compared to Ab491 and Ab493. For these reasons, Ab491 and Ab493 were disregarded from further suitability testing.

The successful antibodies, Ab521 and Ab652 displayed specific binding to the HMW SLP within representatives from the entire *C. difficile* species, and not to the closely related strains. Although not fully determined, it was indicated that Ab521 had a very strong binding affinity. The two mAbs displayed binding to both the denatured and native *C. difficile* proteins and therefore showed potential for use as detection molecules. The suitability of these molecules for PoN sensors will be investigated further in the next chapter.

Chapter 6 Specific whole cell binding of Ab521 and Ab652 to the *C. difficile* SLTS

6.1 Introduction

6.1.1 Whole cell detection

Rapid identification of bacteria can be critical to the clinical outcome of patients (Ahmed *et al.*, 2014). PoN diagnosis of CDI would reduce severe outcomes and complications by enabling quicker, more accurate treatment, and reducing transmission by implementing more prompt isolation measures (Barbut *et al.*, 2014). However, testing for CDI generally does not occur at PoN. Clinical decisions regarding patient treatment and isolation, often have to be made before the results are received from more laborious testing (Goldenberg *et al.*, 2014). This treatment of CDI, without full clinical results, often leads to inappropriate treatment, with one study showing that 74 % of people on empirical therapy, subsequently had negative CDI test results (Saade *et al.*, 2013). The turnaround time for CDI testing varies greatly, depending on the setting. In some hospitals, results are provided within eight hours (Culbreath *et al.*, 2012), whereas others take 17-32.3 h (Grein *et al.*, 2014) and long-term care facilities can take up to four days (Guerrero *et al.*, 2011). Several studies have shown that rapid diagnosis, of either positive or negative CDI cases, results in fewer days in isolation, fewer hospital days, and a reduction in associated costs for both result groups (Barbut *et al.*, 2014; Sewell *et al.*, 2014). PoN sensors that detect whole bacteria, without the need for pre-treatment of samples, would enable this much needed rapid diagnosis to occur (Ahmed *et al.*, 2014).

Non-toxigenic *C. difficile* – strains that do not produce toxins - can reside in the colon without causing infection. Therefore, in order to diagnose CDI, it is not enough to detect the organism alone and the *C. difficile* toxins must also be detected (Kachrimanidou and Malisiovas, 2011). Current UK recommendations for CDI diagnosis state that both the organism and the *C. difficile* toxins should be detected, using a two-step algorithm of either a GDH EIA or NAAT, followed by toxin EIA (Robotham and Wilcox, 2012). Since NAAT requires specialist equipment and skills (Freifeld, 2012), in its current form it cannot be

performed PoN easily, whereas simplified EIAs could be. Cepheid GeneXpert®[27] is a NAAT that was successfully trialled at PoN diagnostic, however, not only is specialist equipment needed but the turnaround time was longer than EIAs, around 1.85 h (Goldenberg *et al.*, 2014). Many diagnostics companies have their own versions of toxin or GDH EIA (Barbut *et al.*, 2014; Davies *et al.*, 2015), however, Alere has the European and US patent for the dual use of both GDH and toxin detection together in a single EIA (Boone *et al.*, 2013). Due to this monopoly on dual use tests, diagnostic companies are interested in identifying a biomarker, other than GDH, that rapidly detects the *C. difficile* organism. A successful whole cell biomarker, and associated detection molecules, would allow companies to manufacture a double EIA of their own.

Chapter 6 describes the specificity of *C. difficile* Ab521 and Ab652 when binding to the native, surface extracted proteins of *C. difficile* using the dot blot technique Figure 4-5. The chapter also describes how the mAbs specifically bound to the HMW-SLP within the whole cell extracts of all 14 SLTS, albeit in its denatured form. These initial, specific binding results indicate that the *C. difficile* target is a promising biomarker. However, to explore the potential of this biomarker for use in a rapid PoN sensor, it is important that the mAbs bind to whole *C. difficile* cells. Although initial data indicated that Ab521 and Ab652 do indeed bind to the native HMW SLP within whole *C. difficile* strains (Figure 5-2) the strains tested were not the SLTS and therefore did not represent the *C. difficile* species as defined by the Dingle study. If these mAbs bind to the whole cells of all 14 SLTS then it would indicate that they could bind to any *C. difficile* strain, without the need for time consuming pre-treatment of cells.

6.1.2 Immunofluorescence microscopy and flow cytometry

Immunofluorescence (IF) is based on pioneering work performed by Coons and colleagues in the 1940s/50s (Coons *et al.*, 1941; Coons and Kaplan, 1950) and utilises fluorescently labelled antibodies to detect molecules, such as a specific bacterial surface antigen (Abbas and Lichtman, 2004). There are two main types of IF firstly, direct IF, which uses a single

[27] http://www.cepheid.com/us/cepheid-solutions/systems/genexpert-systems/genexpert-iv

fluorescently labelled antibody that binds directly to the antigen, providing a fluorescent signal where the binding occurs (Storch, 2000). Secondly indirect IF, which utilise an unlabelled, primary antibody that binds to its antigen before a fluorescently labelled, secondary antibody binds to the F_c region of the primary (Odell and Cook, 2013). For several reasons, such as cost, amount of signal and flexibility, indirect IF is the most predominantly used method (Robinson *et al.*, 2009). Within this project, indirect IF was utilised and measured using two techniques, fluorescence microscopy and flow cytometry. Both of these techniques enabled single-cell analysis, with flow cytometry also providing information at population level and microscopy delivering the visualisation of specific binding (Godfrey *et al.*, 2005).

Within the literature, the majority of the flow cytometry work with *C. difficile* uses the toxins rather than the organism itself (Lanis *et al.*, 2010; Monaghan *et al.*, 2013). One of the projects that used flow cytometry with whole *C. difficile* cells, applied the method to determine adherence of *C. difficile* to gut epithelial cells (Drudy *et al.*, 2001). Using whole *C. difficile* cells with IF microscopy is slightly more abundant in the literature, with several papers utilising indirect IF. Fagan and colleagues labelled the LMW SLP of *C. difficile* cells to determine the importance of the secretory proteins SecA1 or SecA2 (Fagan and Fairweather, 2011). Also, Kovacs-Simon and co-workers used indirect IF to determine the adherence properties of the *C. difficile* lipoprotein CD0873 (Kovacs-Simon *et al.*, 2014). The literature search also identified a paper from 1982 that had trialled direct IF microscopy as a method to detect *C. difficile* in faecal samples (Wilson *et al.*, 1982).

6.1.3 Transmission Electron Microscopy

Transmission electron microscopy (TEM) employs electron beams to convey an image of an object at much greater magnification than light microscopy, which is achieved by the considerably smaller wavelength of electron beams compared to visible light (Serdyuk *et al.*, 2007; Leadley, 2010). Electrons scatter when passing through a specimen and the denser the specimen, the more the electrons scatter, producing a darker image (Leadley, 2010). Immunogold labelling of samples prior to TEM utilises gold conjugated Abs to define specific localisation of antigens (Kaur *et al.*, 2002). When imaged, the gold particles appear as dark

circular, uniform dots, as the electrons cannot pass through the gold and is conveyed to the imaging screen as darkness (Hoppert and Holzenburg, 1998).

6.2 Results

6.2.1 Qualitative whole cell binding of Ab521 and Ab652 to the SLTS

Both Ab521 (A) and Ab652 (B) (Figure 6-1) produced a signal, indicating binding to the whole cells from all 14 *C. difficile* SLTS, *C. difficile* 630 and also the positive controls. Neither antibody produced a signal against the three closely related species, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The Ab521 signal for the undiluted cells of SLTS Ox160, was much weaker than the more dilute samples. The Ab521 signal for all but Ox160 and Ox858 was positive across all of the cell dilutions whereas, Ab652 had strong positive signals for the undiluted cells and when cells were diluted 1 in 100. However, the signal was much weaker for the 1 in 1000 dilution of all *C. difficile* strains. This difference in binding between the two mAbs indicates that Ab521 has a stronger binding affinity than Ab652.

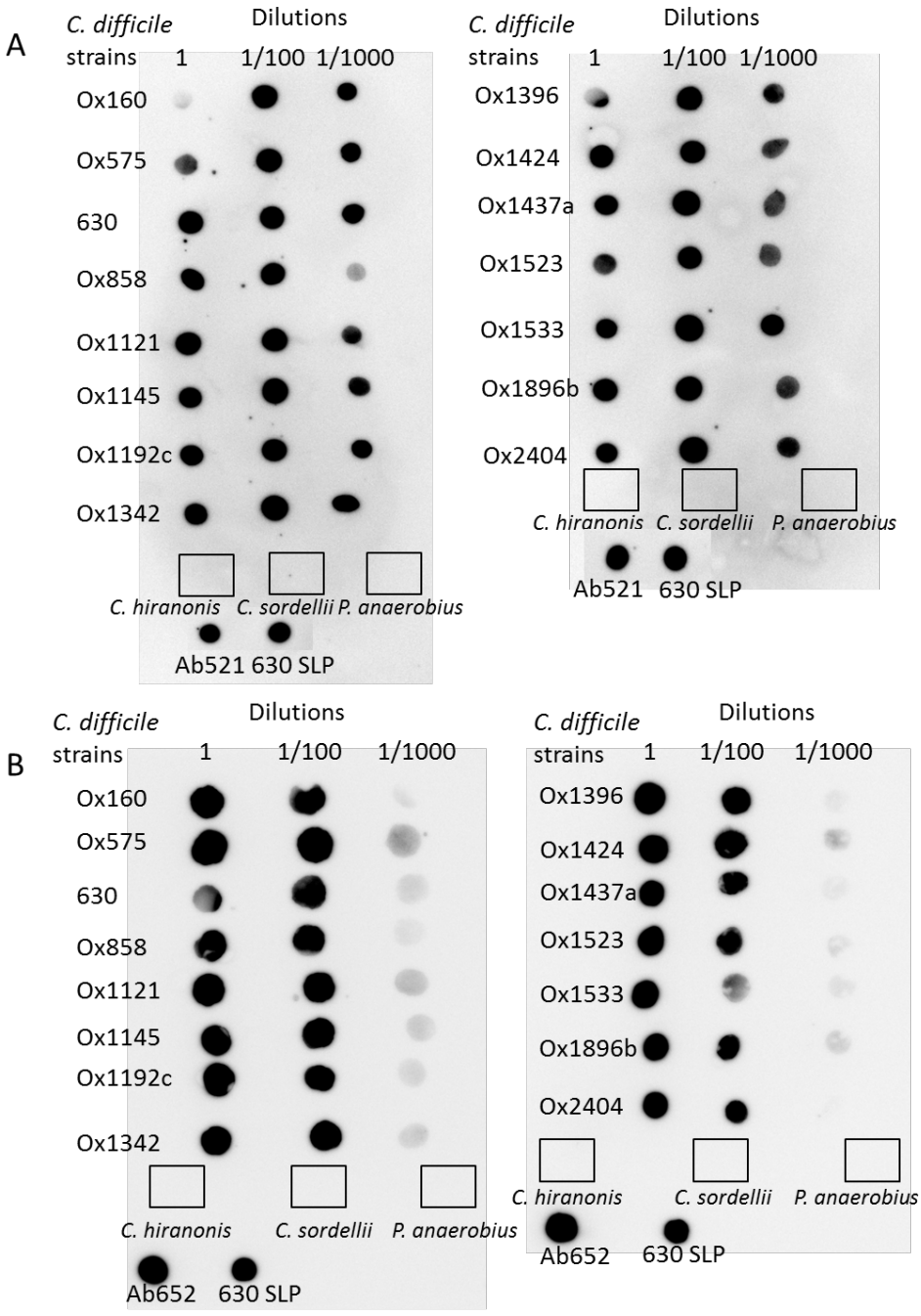


Figure 6-1. Whole cell dot blots of all 14 *C. difficile* SLTS, *C. difficile* 630, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The cells were bound directly to the membrane at differing dilutions, the closely related strains were all applied undiluted. Positive controls were used of the primary antibody and low pH glycine extracted surface layer proteins from *C. difficile* 630, bound directly to the membrane. The blots were probed with 5 µg/ml Ab521 (A) and Ab652 (B) followed by HRP conjugated anti-mouse IgG.

6.2.2 Semi-quantitative whole cell binding of mAbs to the *C. difficile* species

Statistical analysis of the results was performed using the, IBM SPSS version 21, and a one-way ANOVA with Bonferroni and Tukey post hoc tests. A statistical significance value of P= < 0.05 was used throughout the analysis. Unless stated, the P value provided is from both the Bonferroni and Tukey tests.

When tested using whole cell ELISA, Ab521 at concentrations 5 µg/ml to 100 ng/ml displayed a significant increase in absorbance values between the *C. difficile* strains, the 14 SLTS (Ox…) and strain 630, when compared to the closely related species, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330 P = < 0.05 (Figure 6-2). The lowest of the absorbance values for any of the *C. difficile* strains was 1.54 whereas the highest for the closely related species was significantly less at 0.12. The lower concentrations, 100 ng/ml and 1 µg/ml showed almost no binding to the closely related species. There was some binding of Ab521 at 5 µg/ml to the closely related species however, even the species with the greatest absorbance, *P. anaerobius*, showed binding that was significantly lower than the *C. difficile* strains. Therefore, Ab521 showed specificity to *C. difficile* when tested against *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The Ab521 dilutions provided very similar readings for each *C. difficile* strains, with 5-0.1 µg/ml showing very little difference in absorbance. The binding displayed across the *C. difficile* strains was quite uniform, with a difference of around 25 % between the strain with the maximum absorbance, Ox160, and the *C. difficile* strain with the minimum, Ox1396. On account of this clear binding to all *C. difficile* strains, Ab521 demonstrated its potential for sensitive detection of whole *C. difficile* cells.

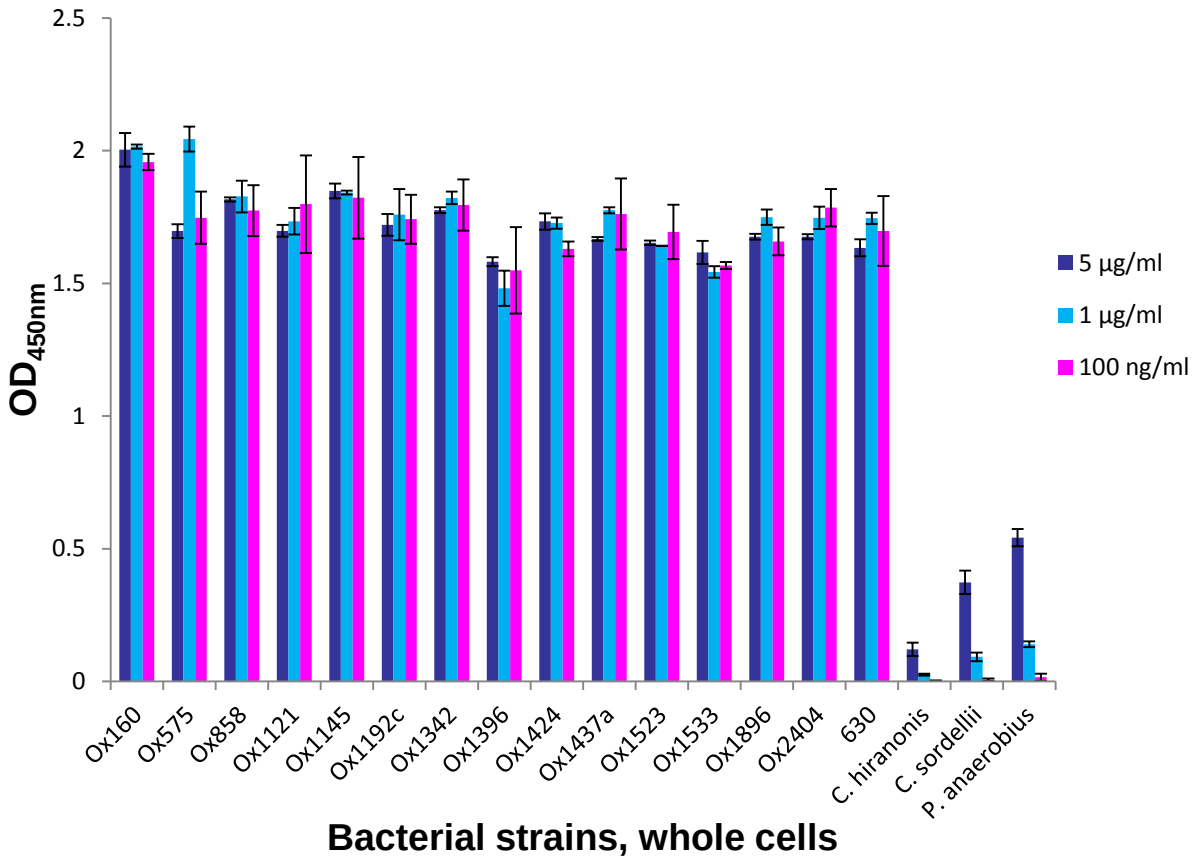


Figure 6-2. Absorbance readings at 450 nm for the Ab521 ELISA at concentrations 5 µg/ml to 100 ng/ml, with whole cells from all 14 *C. difficile* SLTS, *C. difficile* 630 and the closely related species, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The error bars show the standard deviation seen between the two replicates of each sample.

Since Ab521 at 5-0.1 µg/ml bound to *C. difficile* cells at an apparently similar saturating load, further dilutions of the antibody were tested, down to 0.1 ng/ml. Ab521 produced a significant difference in absorbance, P= < 0.05, between *C. difficile* and the closely related strains at concentrations as low as 1 ng/ml (

Figure 6-3). At the lowest concentration of 0.1 ng/ml, the absorbance values were not significantly different between the control species and *C. difficile* strains, P= > 0.05. Therefore, this Ab521 concentration would not be suitable to detect *C. difficile* cells with the required sensitivity.

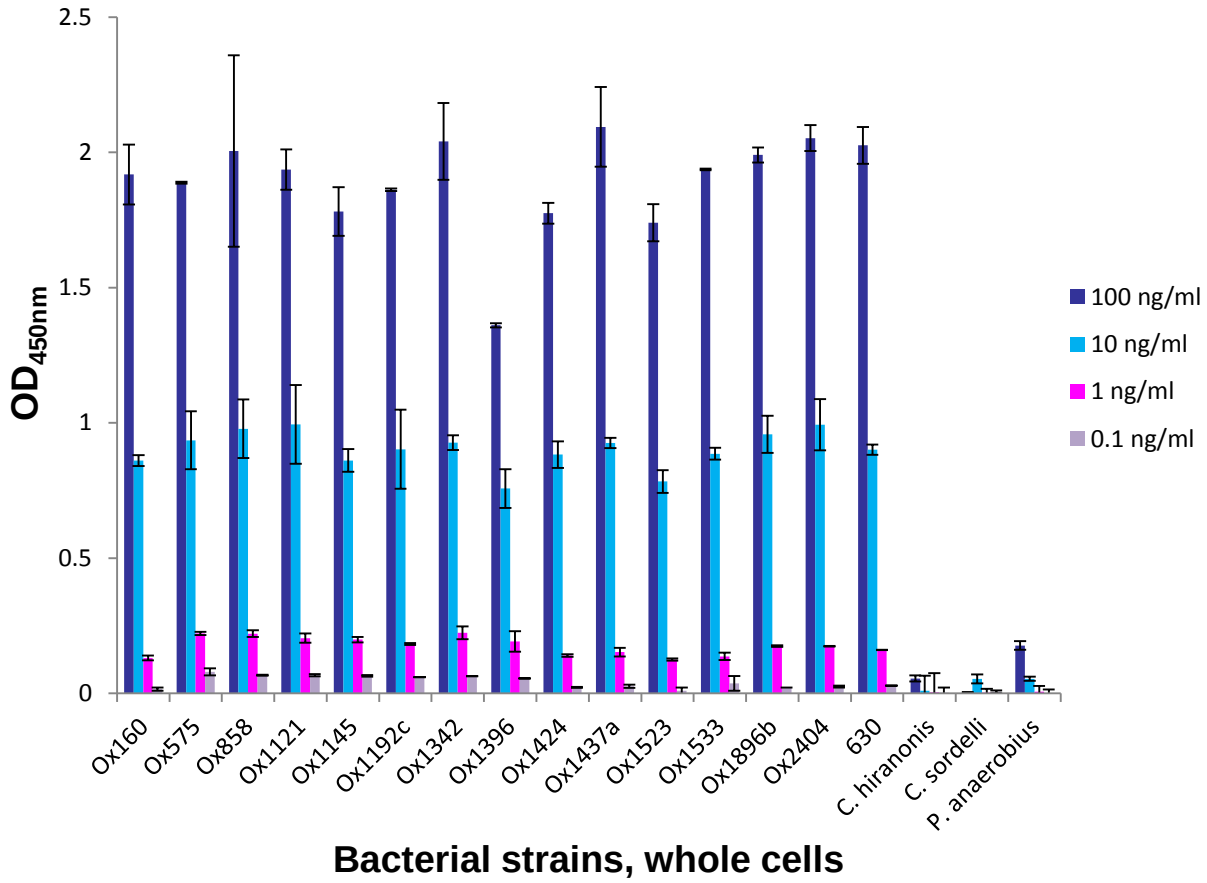


Figure 6-3. Absorbance readings at 450 nm for the Ab521 ELISA at concentrations 100-0.1 ng/ml, with whole cells from all 14 *C. difficile* SLTS, *C. difficile* 630 and the closely related species, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The error bars show the standard deviation seen between the two replicates of each sample.

The absorbance results of each Ab521 concentration from the whole cell ELISAs, from 5 µg/ml to 0.1 ng/ml, were plotted on a graph for each bacterial strain (Figure 6-4). At 0.1 ng/ml the absorbance values are very similar for all strains, displaying why P= > 0.05 at this antibody concentration. The three closely related strains, *C. hiranonis*, *C. sordellii* and *P. anaerobius* are grouped together at the lower absorbance values, across the antibody dilutions. Conversely, the *C. difficile* strains are grouped together with increasing absorbance values as the antibody concentration increases. The graph shows an apparent antibody saturating load at concentrations of 100 ng/ml to 5 µg/ml, however, an increase in absorbance is still seen with Ox1396 between these concentrations. The Ab521 concentration at which the absorbance difference between the closely related species and the *C. difficile* strains is at its greatest is 100 ng/ml. Although there remains a clear differentiation between these groups at the lower concentrations of 1-10 ng/ml.

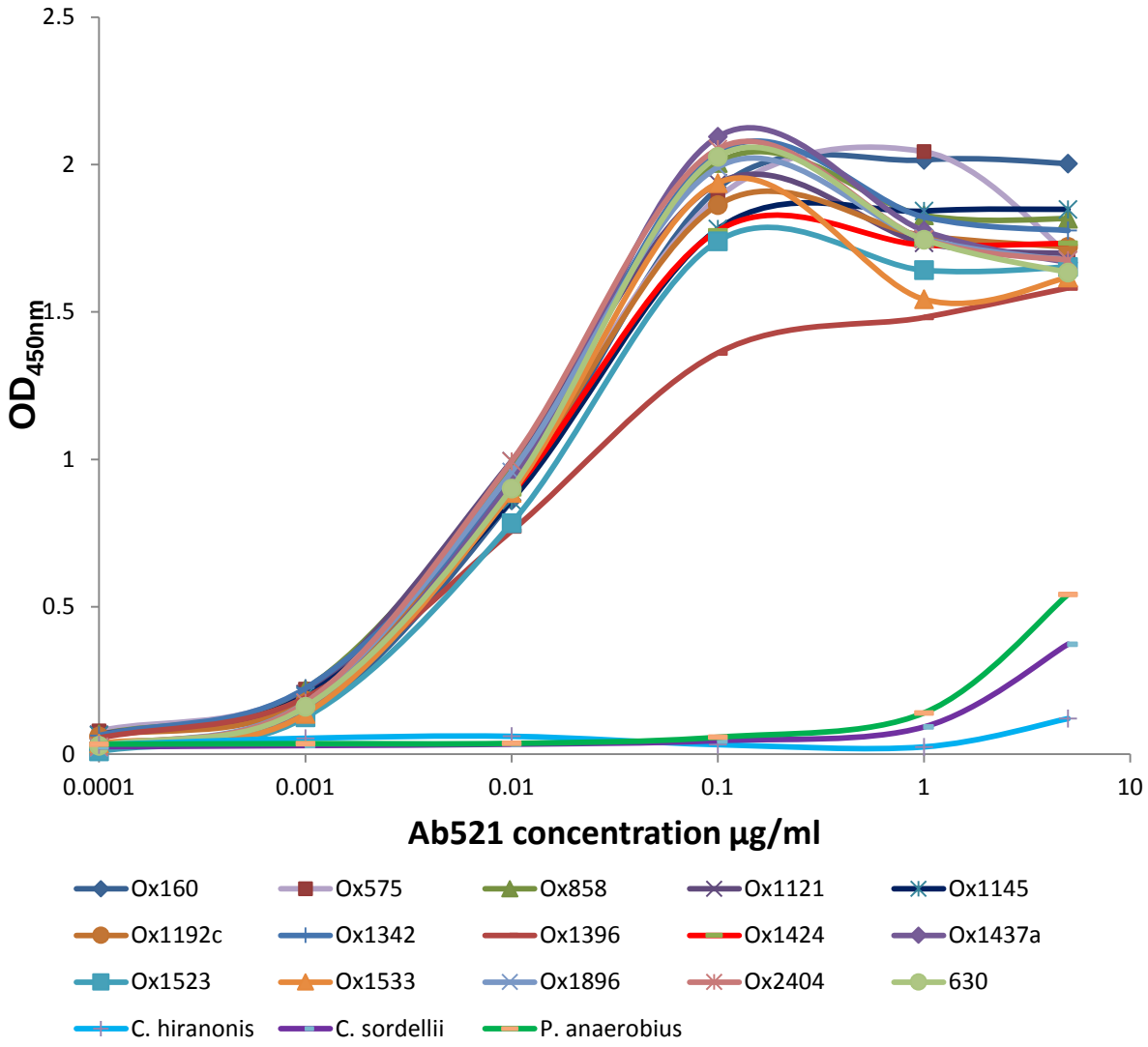


Figure 6-4. Absorbance readings at 450 nm for the Ab521 ELISA from concentrations 5 µg/ml to 0.1ng/ml , with whole cells from all 14 *C. difficile* SLTS (Ox…), *C. difficile* 630 and the closely related species, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The Ab521 concentrations 5 µg/ml to 0.1ng/ml are along the x-axis and each line on the graph denotes the absorbance values for a bacterial strain. The closely related species are grouped at the lower absorbance values whereas, the *C. difficile* strains are grouped and increase in absorbance as the antibody concentration increases. The mean of the two repeats were used and the standard deviation was not shown for ease of visualisation of the curves.

As shown within Figure 3-19 *C. difficile* 630 resides within SLT 7, which is represented by Ox1145. The statistical analysis revealed that there was no significant difference P = > 0.05 between the absorbance values for *C. difficile* 630 and Ox1145 at Ab521 concentrations 1 µg/ml to 0.1 ng/ml. Furthermore, the same lack of significant difference in binding was shown between the two strains that represent SLT clade 2, Ox2404 and Ox858.

The absorbance values for the whole cell ELISA when performed with Ab652, are substantially lower for the *C. difficile* strains than those seen for Ab521 (Figure 6-5). Even at a higher starting concentration of 10 µg/ml, the absorbance readings were around 45 % less than Ab521, with means of lower than one absorbance unit. There was a marked decrease in absorbance for each *C. difficile* strain as the antibody concentration decreased, with values for 100 ng/ml less than half of those seen for 10 µg/ml.

There was a clear difference in values between the controls and the *C. difficile* strains, for 10, 5 and 1 µg/ml, although, the difference was smallest with Ox1396. There was no significant difference between SLTS Ox1396 and *P. anaerobius* at any Ab652 dilution, P = > 0.05. Although, the difference in absorbance for the other *C. difficile* strains was significant at 1-10 µg/ml. When the Ab652 concentration was reduced to 100 ng/ml there was no significant difference in absorbance between the *C. difficile* strains and the closely related species P = > 0.05.

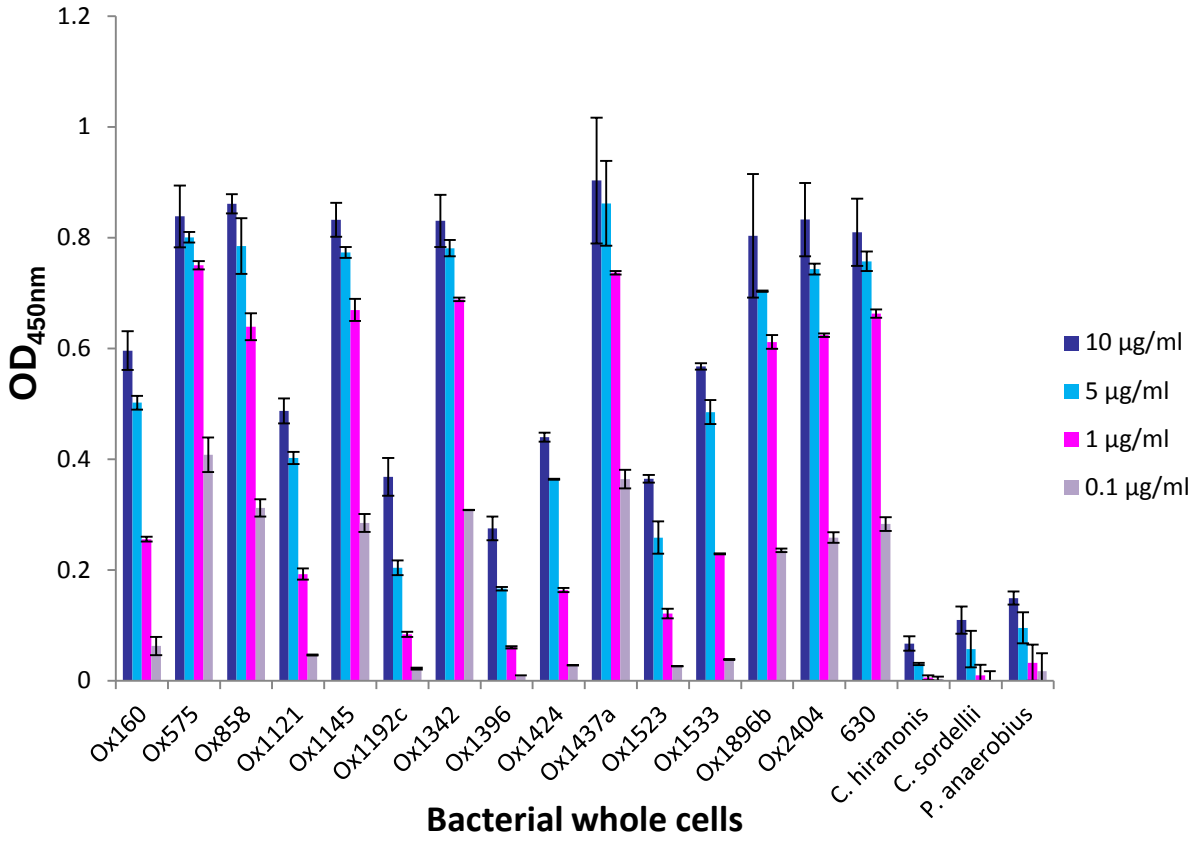


Figure 6-5. Absorbance readings at 450 nm for the Ab652 ELISA at concentrations 0.1-10 µg/ml , with whole cells from all 14 *C. difficile* SLTS, *C. difficile* 630 and the closely related species, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The error bars show the standard deviation seen between the two replicates of each sample.

Similar to Ab521, a graph was produced from the absorbance values for each strain at differing Ab652 concentrations along the x-axis. The three closely related species are consistently at the lowest absorbance values, however, there are also *C. difficile* strains close to these readings. The bacterial strains that are grouped at the higher absorbance readings are all *C. difficile* strains, demonstrating increased antibody binding to these strains. There is not the clear differentiation between all *C. difficile* strains and the closely related species, as displayed by the Ab521 results.

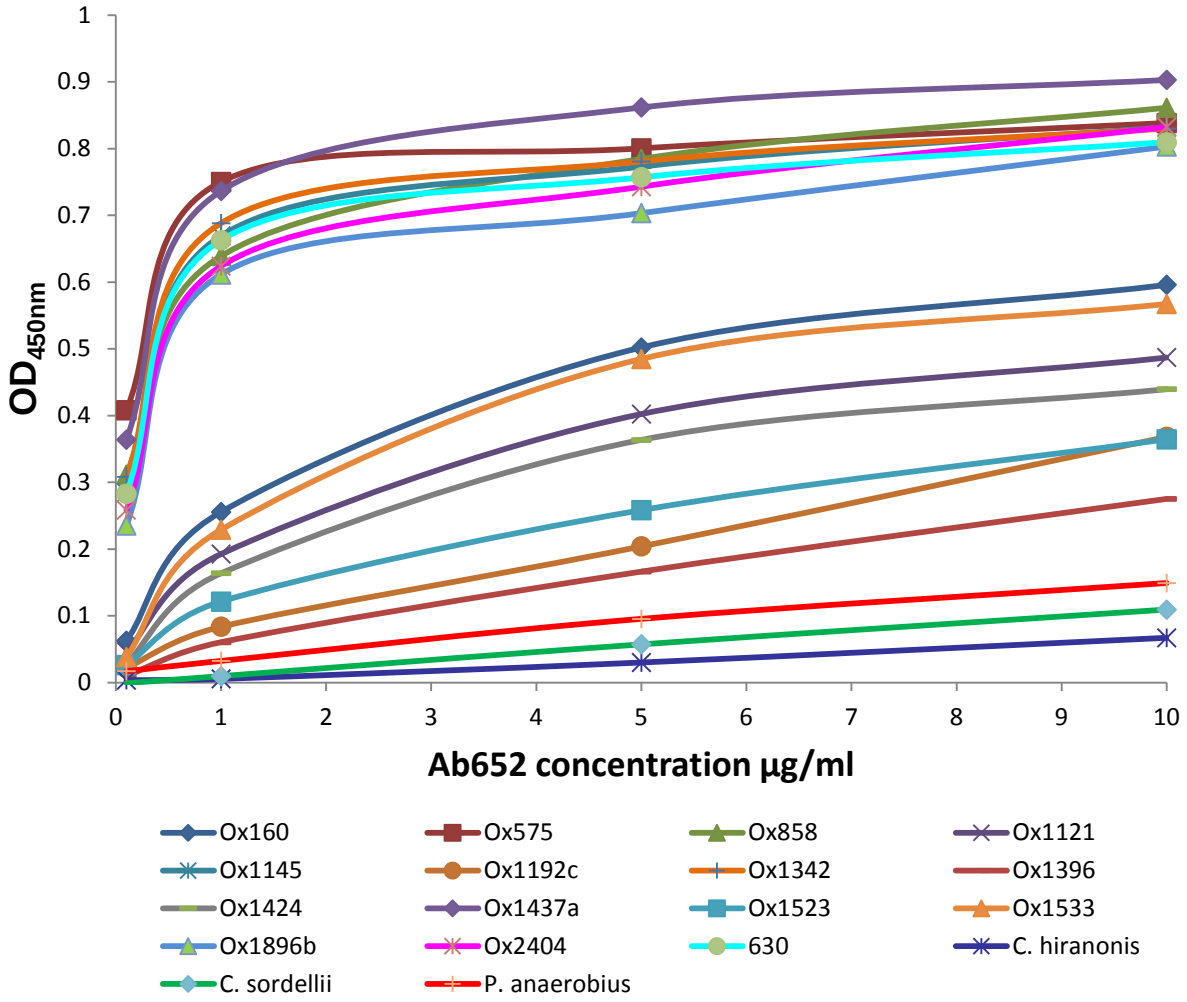


Figure 6-6. Absorbance readings at 450 nm for the Ab652 ELISA for all concentrations, 5 µg/ml to 0.1ng/ml , with whole cells from all 14 *C. difficile* SLTS (Ox…), *C. difficile* 630 and the closely related species, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The Ab652 concentrations 5 µg/ml to 0.1ng/ml are along the x-axis and each line on the graph denotes the absorbance values for a bacterial strain. The closely related species are grouped at the lower absorbance values whereas, the *C. difficile* strains are grouped and increase in absorbance as the antibody concentration increases. The mean of the two repeats were used and the standard deviation was not shown for ease of visualisation of the curves.

6.2.3 Analysing whole cell interactions of Ab521 and Ab652 using flow cytometry

Flow cytometry was applied to further investigate the binding of the mAbs to whole cells. The histogram peaks for *C. difficile* 630 and all 14 SLTS have a marked increase in fluorescence intensity when they are incubated with Ab521 and the secondary antibody, than the cells incubated with the secondary antibody alone (Figure 6-7). All *C. difficile* strains were tested with the secondary antibody only and there was almost no fluorescence for all, with median, as calculated with the FlowJo software, peaks of around zero fluorescence intensity (AU). To avoid repetition, only the peaks for the controls of *C. difficile* 630, Ox1145, Ox1437a (Figure 6-7) are shown.

The *C. difficile* SLTS Ox1192c has the lowest fluorescence intensity peak for both Ab521 and Ab652, with a median fluorescence intensity peak of 730 AU and 264 AU respectively (Table 6-1). For Ab521, the fluorescence intensity peaks ranged from the low Of Ox1192c to the high of Ox1145, which had a median value of 4257 AU, over five times as much as Oc1192c. The flow cytometry results were statistically analysed with IBM SPSS version 21. Mann-Whitney U and Wilcoxon Signed Ranks tests were performed between the fluorescence values of the *C. difficile* strains and the closely related species. When incubated with both Ab521 and the secondary antibody, the P values for both tests were < 0.05 and therefore the results were significantly different. When the same tests were performed with the closely related species, incubated with both antibodies, and the *C. difficile* strains incubated with only the secondary antibody, there was no significant difference between the values P = > 0.05. Therefore, all controls provided similar fluorescence readings which were around 0 AU.

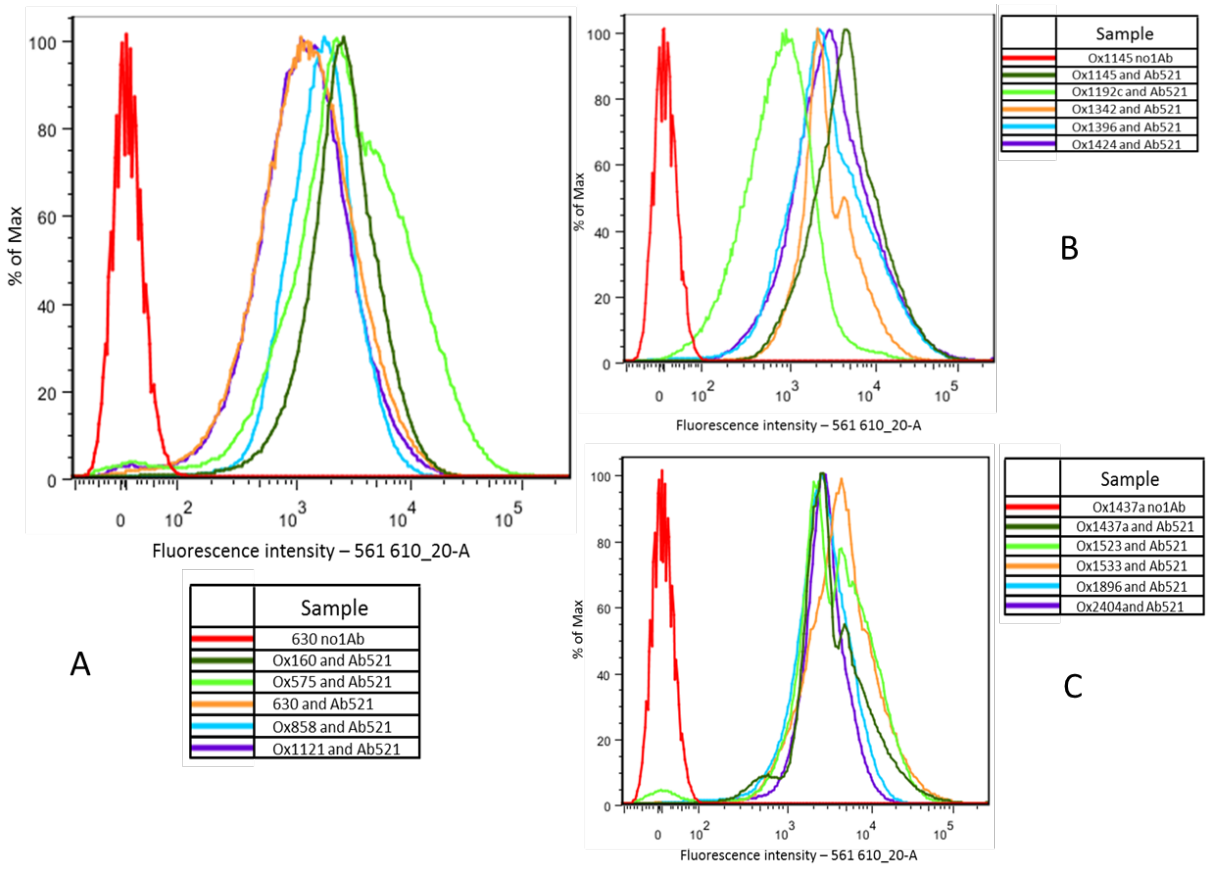


Figure 6-7. Fluorescence intensity histograms of Alexa Fluor 594 conjugated anti-mouse IgG, bound to Ab521 , which in turn was bound to whole *C. difficile* SLTS cells. Controls (red line) are of *C. difficile* cells incubated with the conjugated anti-mouse IgG but not Ab521. All three histograms, A, B and C were performed under the same experimental conditions but with different *C. difficile* strains, which are labelled in the relative key.

The flow cytometry results for whole cells, incubated with Ab652 (Figure 6-8), were similar to the results for incubation with Ab521. The plots displayed a distinct difference in fluorescence values between the *C. difficile* cells incubated with Ab652 and Alexa Fluor 594, anti-mouse IgG, compared to the cells without Ab652 incubation. However, the median fluorescence intensity values for all strains incubated with Ab652 were lower than those incubated with Ab521 with a summary of median fluorescence intensity peaks being displayed in Table 6-1. For example Ox1145 had the highest median fluorescence for both mAbs, yet there was a large difference between them, with Ab521 having a value of 4257 AU and Ab652 just 1660 AU.

Another observation was the difference in the width of the fluorescence peaks for both mAbs, with Ab652 having wider peaks, which are quantitatively displayed by the larger percentage of robust coefficient of variation (%rCV), than Ab521 (Table 6-1). The %rCV,

calculated with FlowJo software, was used because it is not as skewed by outlying values as the coefficient of variation (CV).

"%rCV = 100 * 1/2(Intensity[at 84.13 percentile] – Intensity [at 15.87 percentile]) / Median" (Treister and Roederer, 2015)

All strains incubated with Ab652 had a larger dispersion of fluorescence values as measured by the %rCV, than those with Ab521. For example for Ab652 and Ab521, Ox1396 had 248 AU and 186 AU respectively, Ox1533, 170 AU and 149 AU and finally Ox1896 had 130 AU and 91.2 AU.

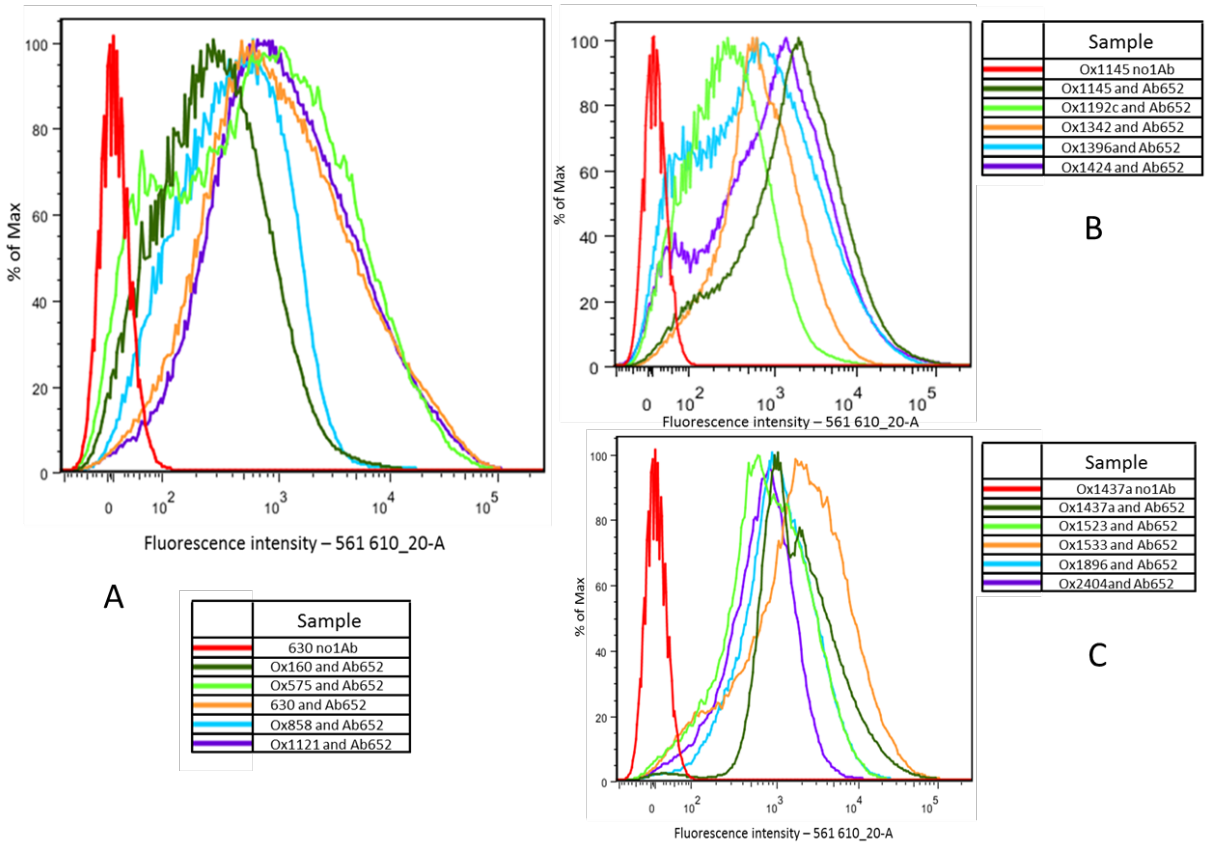


Figure 6-8. Fluorescence intensity histograms of Alexa Fluor 594 conjugated anti-mouse IgG, bound to Ab652 , which in turn was bound to whole *C. difficile* SLTS cells. Controls (red line) are of *C. difficile* cells incubated with the conjugated anti-mouse IgG but no Ab652. All three histograms, A, B and C were performed under the same experimental conditions but with different *C. difficile* strains, which are labelled in the relative key.

The closely related species, *C. hiranonis* and *P. anaerobius* had a median fluorescence of around zero when incubated with either Ab521 (A) or Ab652 (B), and the secondary

antibody (Figure 6-9). Furthermore, Ab521 also had a median fluorescence intensity of zero for *C. sordellii*, whereas Ab652 had a slightly increased positive median value of 26.8 AU and what appeared to be a larger %rCV however, this could not be quantified due the negative fluorescence values within the population. Although the median fluorescence value for Ab652 with *C. sordellii* was positive it remained 10 % less than the *C. difficile* strain with the lowest median value when incubated with Ab652, Ox160. The *C. sordellii* value was less than 1.5 % of the highest median fluorescence intensity displayed by a *C. difficile* strain incubated with Ab652, which was Ox1533.

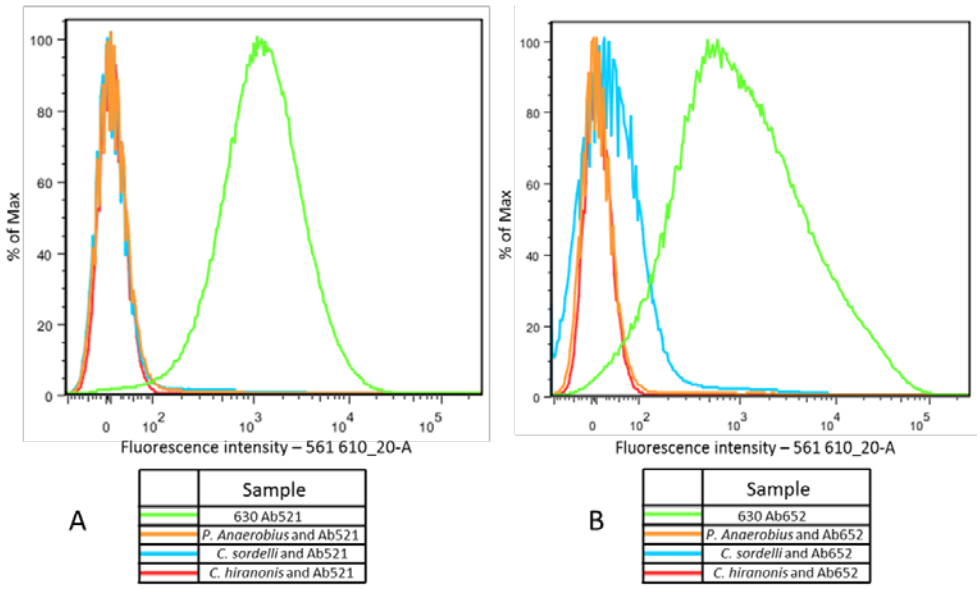


Figure 6-9. Fluorescence intensity histogram of the closely related strains, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714, and *P. anaerobius* VPI 4330 with *C. difficile* 630 as a positive reference. All cells have been incubated with Alexa Fluor 594 anti-mouse IgG and Ab521 (A) or Ab652 (B). *C. difficile* 630 has much higher fluorescence values than the closely related strains. Therefore showing binding of the mAbs to *C. difficile* 630 and not to the closely related strains.

Table 6-1. Fluorescence intensity values provided with flow cytometry for all 14 *C. difficile* SLTS, *C. difficile* 630 and the closely related species *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The bacterial cells

were fixed and incubated with either Ab521 or Ab652 followed by the secondary antibody Alexa Fluor 594 conjugated anti-mouse IgG. The median fluorescence intensity is given as well as the robust coefficient of variance.

	Median fluorescence intensity (AU)		Percentage robust coefficient of variation (%)	
Bacterial strain	Ab521	Ab652	Ab521	Ab652
C. hiranonis DSM-13275	0.3	0.25	N/A	N/A
C. sordellii ATCC 9714	-0.506	26.8	N/A	N/A
P. anaerobius VPI 4330	0.956	-1.15	N/A	N/A
630	1182	903	105	254
Ox160	2419	268	73.1	121
Ox575	2773	685	157	277
Ox858	1582	401	66.1	118
Ox1121	1122	999	96.7	230
Ox1145	4257	1660	109	162
Ox1192c	730	264	89	128
Ox1342	2318	636	92.9	123
Ox1396	2413	558	152	240
Ox1424	2781	941	135	193
Ox1437a	2799	1615	120	140
Ox1523	3413	755	121	138
Ox1533	4047	1899	106	166
Ox1896	2525	1009	73.4	114
Ox2404	2580	675	59	85.5

Statistical analysis was performed on the flow cytometry data and a Mann–Whitney test confirmed that the fluorescence values of all of the *C. difficile* strains, for both Ab521 and Ab652, are significantly difference to the fluorescence values displayed in the closely related species, $P = < 0.05$.

6.2.4 IF imaging of the interaction of Ab521 and Ab652 with whole cells

To gain images of the mAbs interacting with whole cells, IF microscopy was performed. Whole cells were labelled with a fluorescent antibody (see methods section 2.20) using either Ab521 or Ab652 as the primary antibody. The resulting images displayed clear fluorescence at the surface of the *C. difficile* cells, demonstrating positive binding of both Ab521 (B) and Ab652 (F) with Ox1396 (Figure 6-10), however, the fluorescence intensity for

Ab652 (F) was lower than Ab521 (B) for SLTS Ox1342 (Figure 6-11). Both of these figures also show images of the *C. difficile* strains when incubated with the fluorescently labelled secondary antibody only and not the mAbs. The images were displayed with brightfield (C) to reveal presence of the cells and mCherry filter (D), which showed a lack of fluorescence in both *C. difficile* strains without either Ab521 or Ab652. A total of six *C. difficile* strains and three closely related strains were tested for mAb binding using IF microscopy and an overview of the results are provided in Table 6-2. Further IF images of *C. difficile* strains, Ox858, Ox1437a and strains 630 incubated with the mAbs are displayed in the Appendices Figure 9-8, 9-9 and 9-10 respectively.

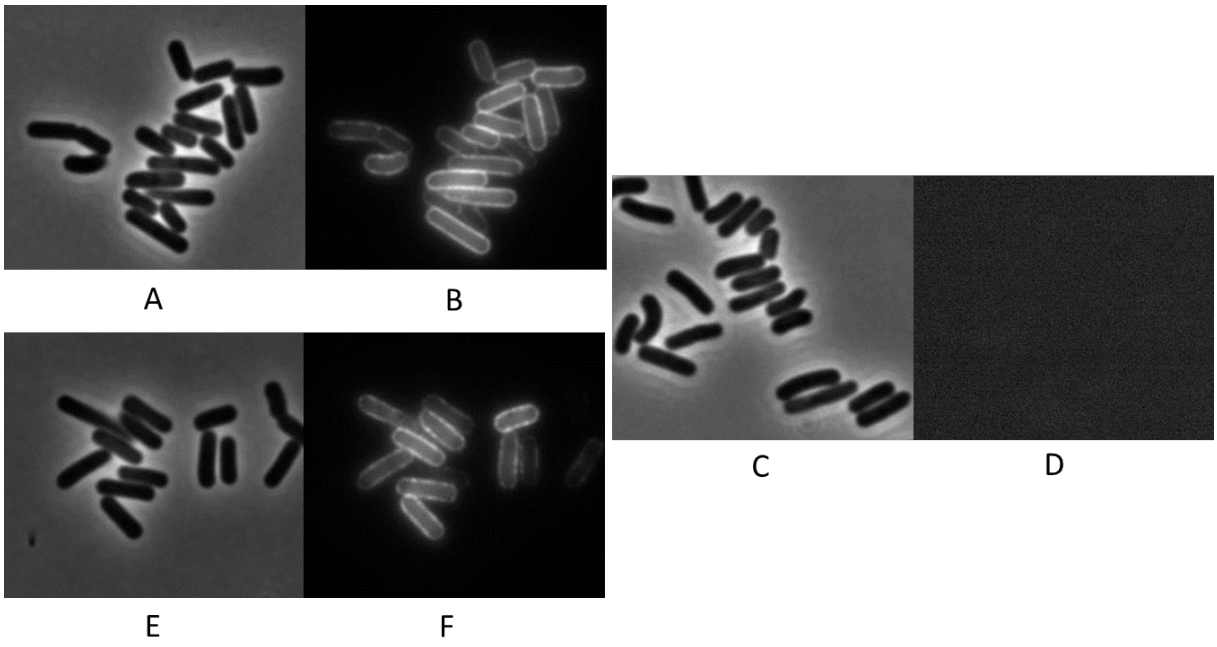


Figure 6-10. Immunofluorescence microscopy images of Ox1396 incubated with Alexa Fluor 594 anti-mouse IgG and Ab521, brightfield image (A), mCherry filter (B). Alexa Fluor 594 anti-mouse IgG and Ab652 brightfield image (E), mCherry filter (F). Control of Ox1396 with Alexa Fluor 594 anti-mouse IgG and no primary antibody brightfield image (C), mCherry filter (D).

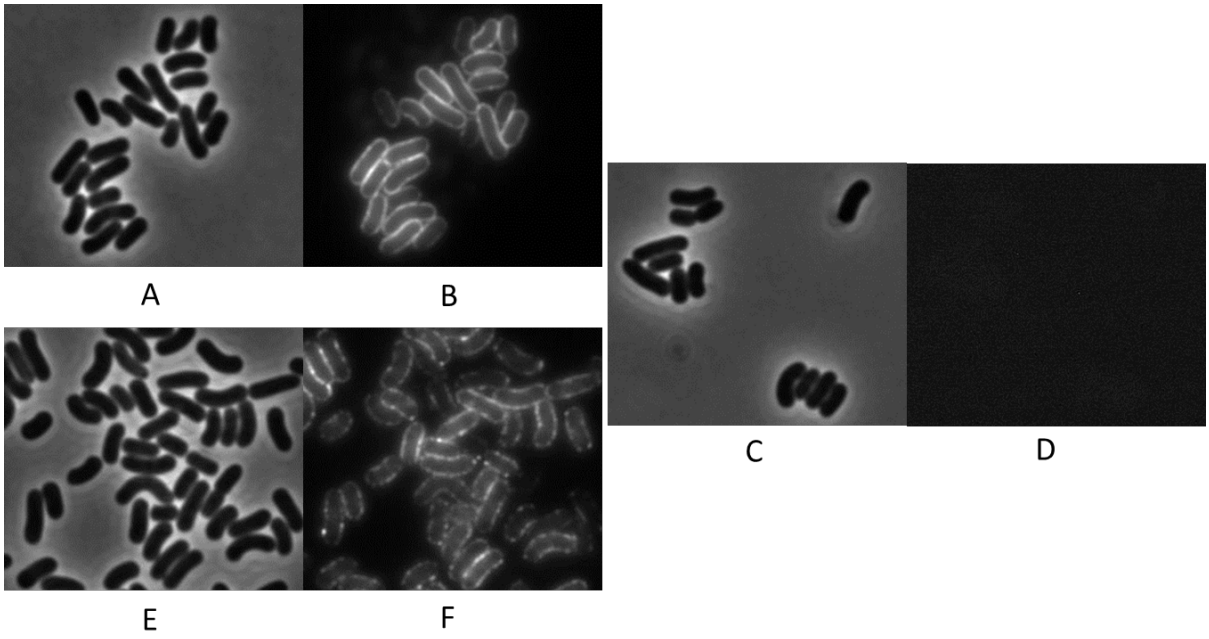


Figure 6-11. Immunofluorescence microscopy images of Ox1342 incubated with Alexa Fluor 594 anti-mouse IgG and Ab521, brightfield image (A), mCherry filter (B). Alexa Fluor 594 anti-mouse IgG and Ab652 brightfield image (E), mCherry filter (F). Control of Ox1342 with Alexa Fluor 594 anti-mouse IgG and no primary antibody brightfield image (C), mCherry filter (D).

The closely related species, *C. sordellii* (Figure 6-12), *C. hiranonis* and *P. anaerobius* (Appendices Figure 9-11 and 9-12) did not display fluorescence when incubated with Ab521 (B) or Ab652 (D) and the secondary antibody (Figure 6-12 D). The actual fluorescent values are provided in For Table 6-2 the average cell fluorescence was calculated using ImageJ. The fluorescence values were used from a minimum of five random cells from *C. difficile* or the closely related species and the average was calculated. The background fluorescence was also calculated from a minimum of five areas, which were the same size as the cells that were analysed. The background for the cells incubated with the secondary antibody only and not the mAbs were calculated, again using ImageJ. The cells were marked on the brightfield image and the markers were transferred to the corresponding fluorescence image, enabling calculation of fluorescence values at the specific cell site.

Table 6-2 and show that the there is some slight binding of Ab652 to *C. sordellii*. However, the binding is negligible, as the value is a minimum of four times lower than the fluorescence value of Ab652 with the *C. difficile* strains. For the other two closely related species and all three for Ab521 the resulting fluorescence is less than 1.5 AU. The non-specific binding is insignificant when compared to binding with *C. difficile*, which the lowest value is seen with Ox8585 and is 366.19 AU for Ab652 and 548.27 AU for Ab521.

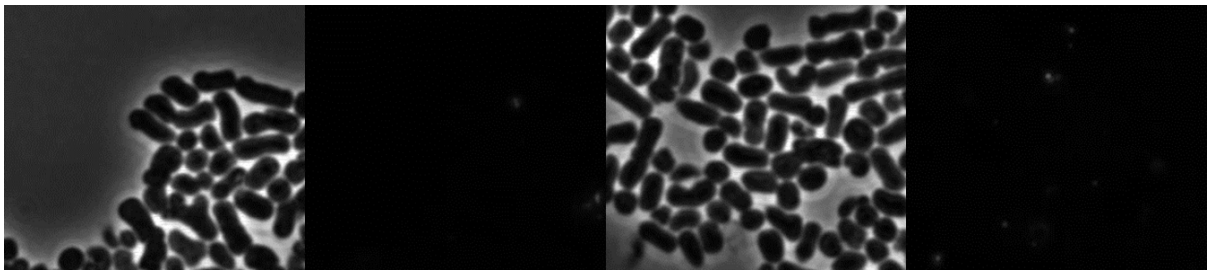


Figure 6-12. Immunofluorescence microscopy images of *C. sordellii* incubated with Alexa Fluor 594 anti-mouse IgG and Ab521 brightfield (A) mCherry (B) or Ab6521 brightfield (C) mCherry (D).

For Table 6-2 the average cell fluorescence was calculated using ImageJ. The fluorescence values were used from a minimum of five random cells from *C. difficile* or the closely related species and the average was calculated. The background fluorescence was also calculated from a minimum of five areas, which were the same size as the cells that were analysed. The background for the cells incubated with the secondary antibody only and not the mAbs were calculated, again using ImageJ. The cells were marked on the brightfield image and the markers were transferred to the corresponding fluorescence image, enabling calculation of fluorescence values at the specific cell site.

Table 6-2. The average fluorescence intensity values provided with immunofluorescence microscopy for *C. difficile* SLTS, Ox858, Ox1396, Ox1342, Ox1533 and Ox1437a, *C. difficile* 630 and the closely related species *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The bacterial cells were fixed and incubated with either Ab521 or Ab652 followed by the secondary antibody Alexa Fluor 594 conjugated anti-mouse IgG. The mean fluorescence intensity is shown as well as the mean background fluorescence, which is subtracted from the former to provide a resulting fluorescence value for the cells. There are values for cells incubated with Ab521 or Ab652, both with the secondary antibody, or there are cells with only the secondary antibody and no primary Ab521 or Ab652 (No 1Ab).

Bacterial Strain	Average cell fluorescence (AU)			Average background (AU)			Resulting fluorescence (AU)		
	Ab521	Ab652	No 1Ab	Ab521	Ab652	No 1Ab	Ab521	Ab652	No 1Ab
630	1572.07	594.05	199.35	211.26	200.85	197.62	1360.81	393.20	1.73
Ox858	748.21	566.92	195.81	199.94	200.73	196.00	548.27	366.19	-0.18
Ox1396	2808.92	1105.05	196.3	205.14	200.08	194.89	2603.78	904.96	1.48
Ox1342	814.80	707.31	194.78	197.62	197.99	196.09	617.18	509.31	-1.31
Ox1533	1525.36	834.97	196.37	254.88	455.46	195.67	1270.48	379.51	0.70
Ox1437a	3711.76	1319.32	223.79	208.31	201.30	222.92	3503.44	1118.03	0.87
C. hiranonis	135.59	133.11	136.58	134.43	132.10	135.16	1.16	1.01	1.42
C. sordellii	197.73	250.35	197.69	196.59	218.92	196.33	1.14	31.42	1.36
P. anaerobius	133.03	132.96	132.89	132.56	132.73	133.41	0.48	0.24	-0.51

6.2.5 TEM analysis of Ab521 and Ab652 with *C. difficile* and *C. sordellii*

TEM enables cells to be seen at great magnification and was used with indirect immunogold labelling to image antibody interaction with whole *C. difficile* and *C. sordellii* cells. The cells were paraformaldehyde fixed and incubated with either Ab521 or Ab652 and the secondary, gold conjugated, anti-mouse IgG.

When incubated with Ab521, both *C. difficile* 630 (A) and Ox575 (B) displayed gold nanoparticles, which can be seen as dense, dark, uniform circles, attached to the surface of the cells (Figure 6-13). To assist with visualisation of some of the more difficult to see nanoparticles, they were marked with white arrows. There are several factors affecting the ability to see the gold nanoparticles firstly, they can only be seen when the focus of the electron beam is nearby. Secondly, the sample being viewed must not be too dense and finally, the negative stain, used to enable visualisation of the cells, must not be too dark.

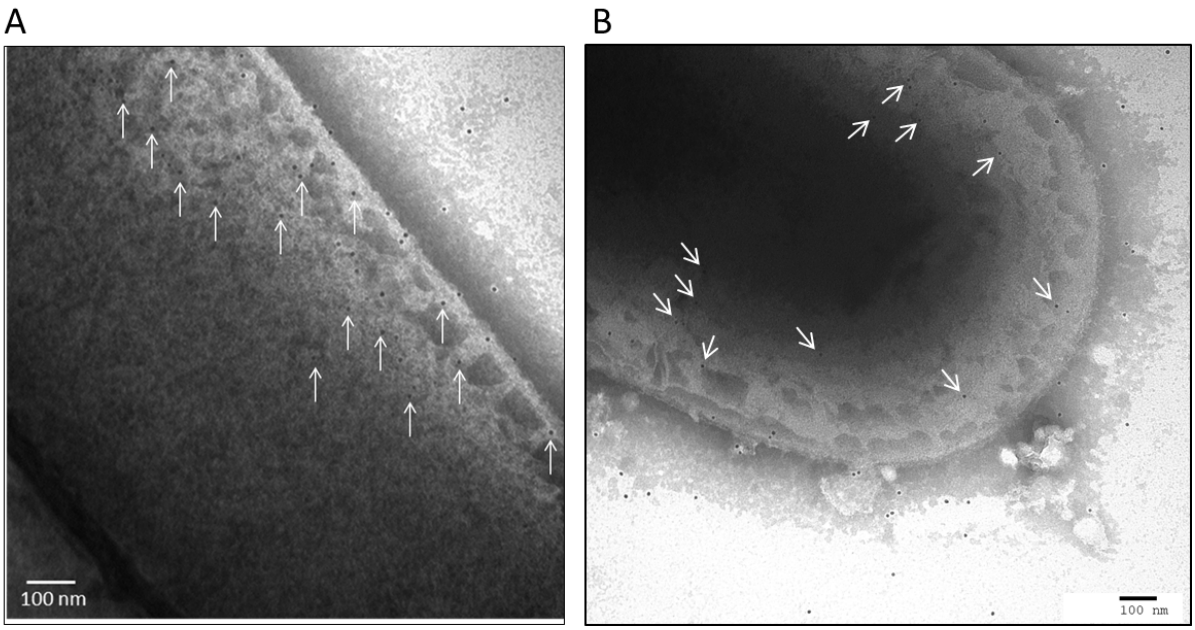


Figure 6-13. Transmission electron microscope image of *C. difficile* 630 (A) and Ox575 (B) incubated with Ab521 and gold conjugated anti-mouse IgG. The white arrows indicate gold nanoparticles that are within the denser region of the cell and are therefore more difficult to see. The S-layer completely encapsulates the bacterium and is at the same position as the cell wall at the surface of the rod shaped cells.

Gold nanoparticles were also seen on *C. difficile* 630 cells when incubated with Ab652 (**Figure 6-14**). However, it was noted that there were fewer particles with Ab652 than on the cells

incubated with Ab521 (Figure 6-13). Quantification of the gold nanoparticles was not performed and consequently, the particle number was purely observational.

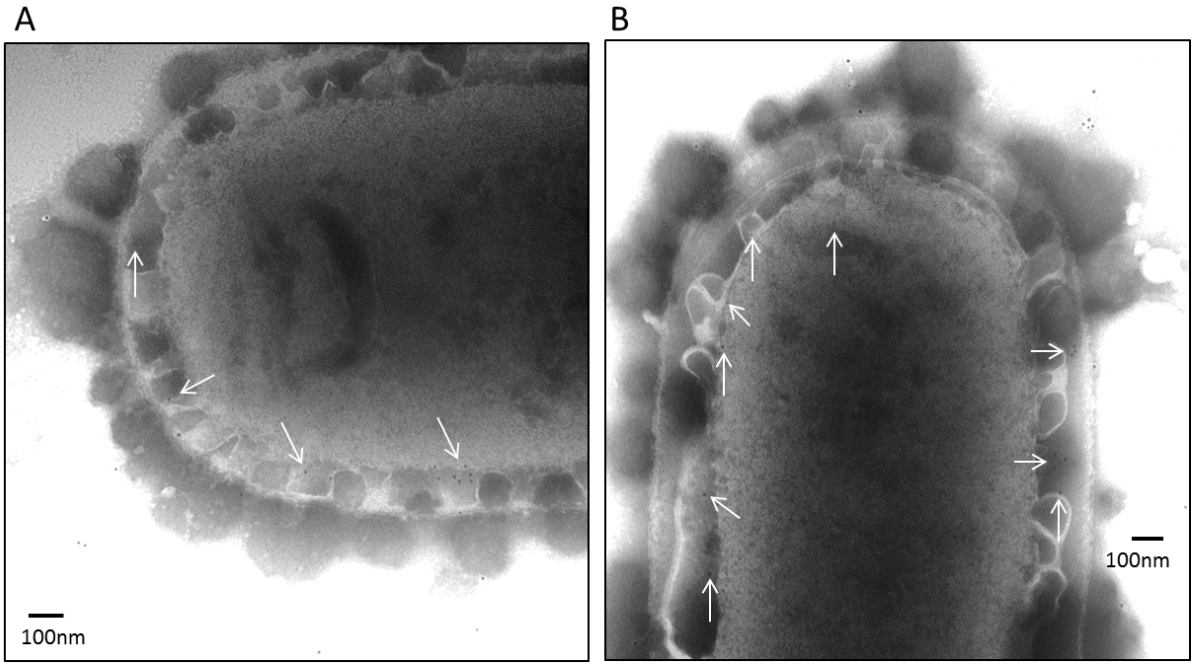


Figure 6-14. Transmission electron microscope image of *C. difficile* 630 incubated with Ab652 and gold conjugated anti-mouse IgG. The white arrows indicate gold nanoparticles that are within the denser region of the cell and are therefore more difficult to see. The S-layer completely encapsulates the bacterium and is at the same position as the cell wall at the surface of the rod shaped cells.

For a negative control, *C. difficile* 630 was incubated with the secondary antibody only and no primary antibody. These cells did not display gold nanoparticles attached to the surface of the cells (Figure 6-15) and therefore showed that the gold conjugated anti-mouse IgG does not non-specifically bind to *C. difficile*.

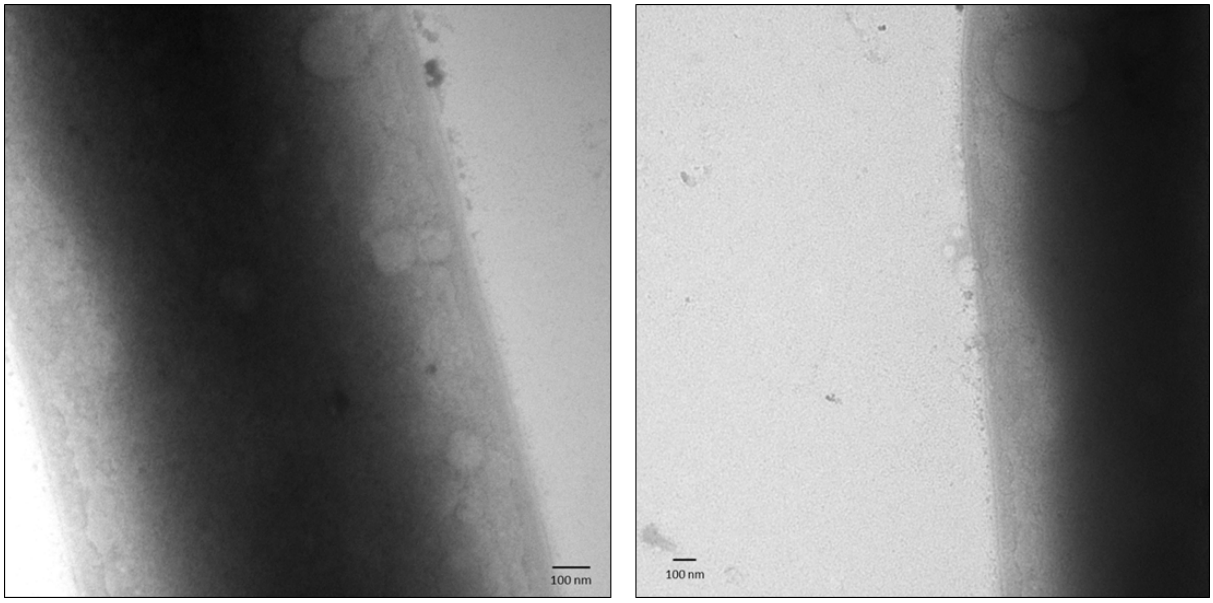


Figure 6-15. Transmission electron microscope image of *C. difficile* 630 without primary antibody and incubated with only gold conjugated anti-mouse IgG. There are no gold nanoparticles displayed on the cells.

C. sordellii was also incubated with either Ab521 (A) or Ab652 (B) and the gold conjugated anti-mouse IgG, and viewed using TEM (**Figure 6-16**). Although one or two gold nanoparticles were seen on the images, they were not on the surface of the organism, which was completely clear of gold particles.

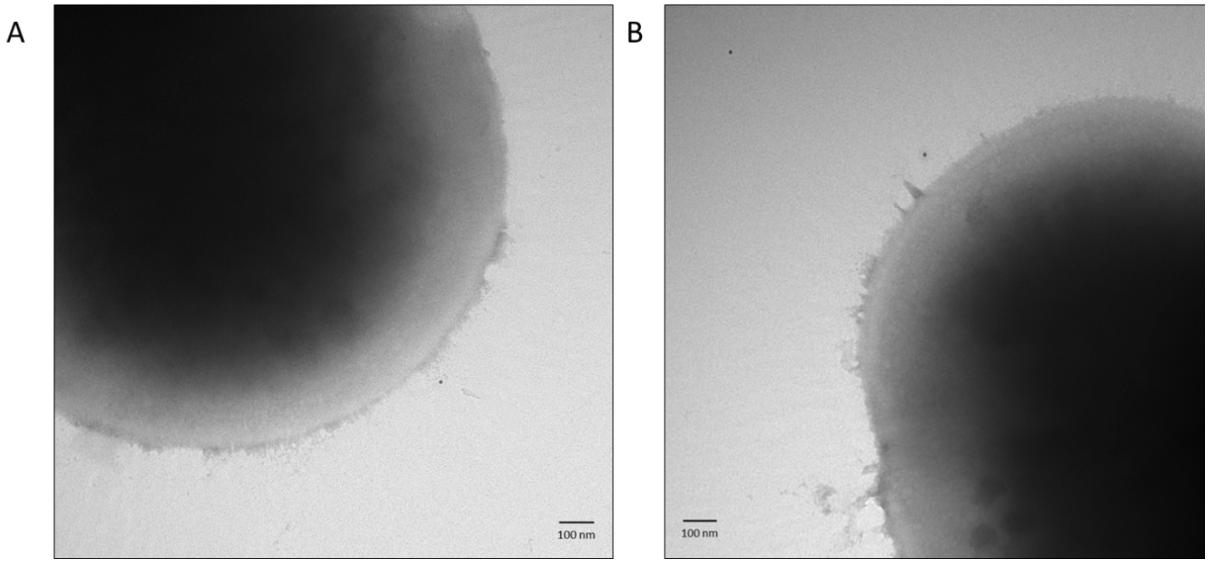


Figure 6-16. Transmission electron microscope image of *C. sordellii* incubated with Ab521 (A) and Ab652 (B) and gold conjugated anti-mouse IgG and

6.2.6 The potential of Ab521 for use in a diagnostic assay or sensor

Initially, the intention of the project was to produce the antibodies for use in a Microelectromechanical systems (MEMS) PoN diagnostic sensor, However, as the project progressed, lateral flow strips were investigated as a potential sensor for use with the *C. difficile* mAbs. Leteral flow strips were selected due to their ease of use and the low production costs. One of the limitations of using lateral flow strips with whole bacteria is the availability of membranes with large enough pore sizes. The largest available pore size for nitrocellulose membrane was found to be 15 µm from Sartorius[28] and strips that were successful at detecting *S. aureus* were made using membranes of this pore size (data not shown). However, after several optimisation attempts, the lateral flow strips were found to be incapable of flowing *C. difficile* down the membrane.

For proof of principle of Ab521 working within lateral flow strips, a test line of purified HMW/LMW complex was bound to the membrane and gold conjugated Ab521 was flowed along the strip (Figure 6-17). As a positive control, anti-mouse IgG was bound to the membrane. With this set up the Ab521 successfully moved through the membrane and bound to the SLP complex and the anti-mouse IgG positive control line (A). As a simple control a gold conjugated prostate-specific antigen (PSA) antibody was also used on a HMW/LMW SLP bound lateral flow strip and did not bind to the complex (B).

[28] https://www.sartorius.co.uk/en/product/product-detail/1un95er100025nt/

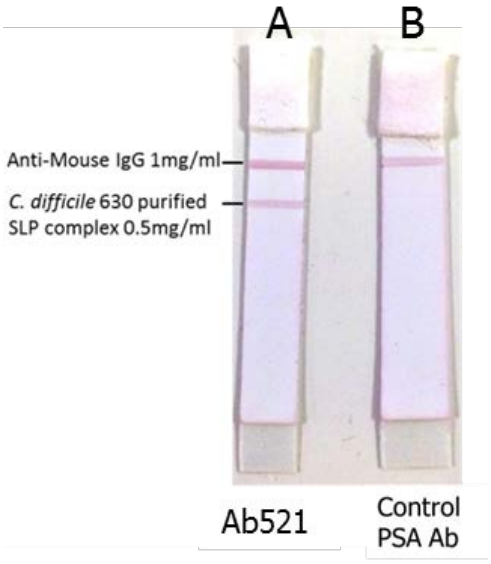


Figure 6-17. Lateral flow strips with a control line (top red line) of anti-mouse IgG 1 mg/ml and a test line of purified *C. difficile* 630 HMW/LMW complex. Gold conjugated Ab521 was then flowed down strip A, providing a positive red line at the HMW/LMW complex. Gold conjugated prostate-specific antigen antibody was flowed down strip B and did not produce a positive signal for the HMW/LMW complex.

6.3 Discussion

6.3.1 Dot blots indicate that mAbs bind to whole *C. difficile* cells

Whole cell binding of the mAbs is advantageous for their integration into rapid PoN sensors (Ahmed *et al.*, 2014) as it would reduce the need for time consuming pre-treatment, such as lysis of cells. The dot blots displayed positive signals for both Ab521 and Ab652, binding to all 14 SLTS and *C. difficile* 630. Since these 14 SLTS are representatives of the 13 SLT clades that all *C. difficile* strains can be clustered into, theoretically, Ab521 and Ab652 should bind to the *C. difficile* species as a whole, at least with dot blot analysis. For SLTS Ox160, the signal for the undiluted cells, with Ab521, was much weaker than the more dilute samples. In all likelihood this was due to the experimental procedure as opposed to the undiluted cells not providing a positive result as further diluted samples were positive. Since the more dilute Ox160 samples provided a positive result, and the mAb binding to this strain was going to be investigated further, it was decided not to repeat the dot blot.

The signal for Ab521 remained strong with the *C. difficile* cells at a dilution of 1 in 1000, whereas, Ab652 displayed poor binding, implying that Ab521 may bind to *C. difficile* with

greater affinity than Ab652. Neither Ab521 nor Ab652 produced a positive signal for any of the three closely related strains, *C. sordellii*, *C. hiranonis* or *P. anaerobius*, indicating that the mAbs do not bind to whole cells from these control species. The closely related species were used to test for non-specific binding of the mAbs, since they have sequences that are similar to *C. difficile* or SlpA, described in detail within the bioinformatics results section 5.2.6 and thus are more likely to interact with the antibodies. To prevent PoN sensors providing false positives, it is imperative that the mAbs do not bind to anything other than *C. difficile*. Subsequently, testing of non-specific binding began with the species that the mAbs are most likely to erroneously bind to, however it is recognised that many more species will require testing. Since dot blots provide only semi-quantitative information and are not sensitive enough to draw conclusions from alone (Jin *et al.*, 2011), complementary methods of examining positive and non-specific binding were pursued.

6.3.2 Ab521 binds to whole *C. difficile* cells with greater efficiency than Ab652

To gain semi-quantitative evidence of whole cell binding, the mAbs were tested using whole cell ELISAs. The results for Ab521 showed that when incubated with each of the 15 *C. difficile* strains, the absorbance values were saturated between with 100 ng/ml of antibody and 5 µg/ml. In order to produce the most cost efficient PoN sensor, it is important to use the minimum amount of detection molecule that still provides clear recognition of its specific antigen. At 0.1 µg/ml, Ab521 distinguished between all *C. difficile* strains and the closely related species with a significance of $P = 0.000$ therefore, providing the specificity required for PoN testing. Since the tested SLTS represent the entire *C. difficile* species, in the context of SlpA, the significant binding of Ab521 to all of these strains indicate that the antibody would bind to all *C. difficile* strains. This binding indicates the high sensitivity of Ab521, which would be useful for a PoN sensor. Even at 1 ng/ml Ab521 there remained a significant difference $P = < 0.05$ between all of the *C. difficile* strains and the closely related species. This antibody concentration enabled clear differentiation between these two groups, therefore, the limit of detection (LOD) of the antibody was 1 ng/ml. However, at this concentration, the difference in absorbance values between the *C. difficile* strains and the closely related species was substantially lower than the difference displayed with the higher concentrations of antibody. The decreased absorbance difference may not be enough to

retain the distinction between the groups once the antibody is integrated in to a PoN sensor. The type of sensor that the antibody is used within, would in-part determine the concentration of the antibody required for specific detection of *C. difficile*. Within the whole cell ELISA results, the optimal antibody concentration for *C. difficile* detection lies between 1 and 10 ng/ml. For future work, dilutions of Ab521 between 1 and 10 ng/ml could be tested and the optimal concentration resolved.

In contrast, Ab652 displayed a significant decline in binding to the *C. difficile* strains as the antibody concentration reduced from 10-0.1 µg/ml. Even at 10 µg/ml, Ab652 bound to the cells with an efficiency that was around 40 % less than Ab521. The lowest concentration of Ab652 which displayed a significant binding difference between most *C. difficile* strains and the controls was 1 µg/ml, ten times more concentrated than Ab521. Furthermore, at all concentrations, Ab652 did not display a significant difference between the binding with *C. difficile* SLTS Ox1396, compared with the closely related species, *P. anaerobius*. The results of the *C. difficile* strains grouped in with closely related species meant that a limit of detection could not be obtained.

These results corroborated the initial findings from the dot blots, that Ab521 has a higher binding efficiency to *C. difficile* than Ab652. Therefore, Ab521 is the more promising detection molecule for use in a PoN sensor.

Within a PoN sensor it is recognised that the samples would not be pure *C. difficile* and would include other organisms and faecal matter. The optimal Ab521 concentration could differ with more complex samples and therefore, further testing of the antibody is required. A useful experiment for future work would be to determine the range at which the antibody still efficiently detected the *C. difficile* cells, by altering the OD or cell number and keeping the antibody concentration constant. It is possible to count the cells, using a bacterial haemocytometer, and calculate roughly how many cells are present. Within this project however, the OD was used because the cell size varied between strains. The larger cells would likely have more HMW SLP epitopes compared to the smaller cells, as they have a larger surface area to cover, so an equal cell count would provide an unequal number of antibody binding sites across the strains. Also, different strains could differentially express SlpA and have a differing density of S-Layer, which could both alter the amount of HMW SLP

epitope that is present within the samples. To test the antibody binding further, mixed samples would be used. These samples would include imitated faecal samples using bacteroidetes and firmicutes (Parks *et al.*, 2013), to ensure the antibodies can still detect *C. difficile* within complex samples. It would be important to first test the individual bacterial species within the 'faecal sample' to confirm that they do not non-specifically bind the mAbs.

Within the ELISAs for both Ab521 and Ab652, the lowest absorbance reading for the *C. difficile* strains was displayed with Ox1396. This decrease could be due to both mAbs binding more weakly to this strain, a lack of epitopes/epitope exposure, or fewer cells bound to the ELISA plate providing less antigen. All three ELISAs were set up with the same samples and as a result from the ELISAs alone, it was impossible to tell which of the three possibilities caused the lack of binding. The difference in absorbance between the control strains and Ox1396 was significant for Ab521 at all concentrations from 5 µg/ml to 0.1 ng./ml, consequently this antibody could accurately differentiate between Ox1396 and the closely related species. On the other hand, the values for all Ab652 concentrations were not significantly different, showing that under these conditions, Ab652 would not differentiate between Ox1396 and *P. anaerobius*. In order to investigate the lower binding to Ox1396 compared to the other SLTS, further examination was performed with different experimental techniques, the results of which will be discussed in detail in the upcoming sections.

Presumably due to IP protection, information regarding the concentration of antibodies used in commercial *C. difficile* diagnostic ELISAs was not found. The recommended starting dilution of the detection antibody for ELISA is 0.5-5 µg/ml (Kemeny and Challacombe, 1988; Thermo, 2010) and since the optimum Ab521 concentration is below that range, it demonstrates the mAb's strong binding ability.

6.3.3 Quantitative evidence that Ab521 binds to whole *C. difficile* cells at a greater efficiency than Ab652

Flow cytometry enables the quantitative analysis of antibody binding to individual cells, whilst also providing information on antibody binding to the population as a whole (Abbas and Lichtman, 2004). Both mAbs bound to all 14 SLTS and *C. difficile* 630 with clear, positive

results when compared to the controls, including *C. difficile* with no mAb and the closely related species with both antibodies. This binding was shown by the significant increase in median fluorescence intensity, which was caused by Ab521/Ab652 binding to whole *C. difficile* cells and the subsequent binding of Alexa Fluor 594 secondary antibody. The lack of fluorescence for the *C. difficile* strains without mAb incubation showed that there was no non-specific binding of the secondary Alexa Fluor 594 antibody to *C. difficile* and that the mAbs are necessary to gain a positive fluorescent signal.

The median fluorescence values for all *C. difficile* strains were greater for Ab521 than Ab652, corresponding with the ELISA and dot blot results. These quantitative flow cytometry results determine that Ab521 binds with higher efficiency to *C. difficile* cells than Ab652. The %rCVs for each *C. difficile* strain were substantially greater for Ab652 than Ab521, indicating a more distributed range of fluorescence intensities across the sample population. On account of this contrast in %rCV value, it can be established that Ab521 binds more reproducibly to each cell type than Ab652.

Unlike the whole cell ELISA results, binding of the mAbs to Ox1396 was not substantially less than the binding to other *C. difficile* strains. In fact the peak fluorescence for Ox1396 was greater than that of five *C. difficile* strains for Ab521 and three for Ab652. As a result of this binding observed in a sample of 1,000,000 cells, it was noted that the lack of binding to Ox1396 within the ELISA results, was likely due to either the technique itself or the poor adhesion of Ox1396 to the ELISA plate.

The three control strains, *C. hiranonis*, *C. sordellii* and *P. anaerobius* were incubated with either Ab521 or Ab652 and anti-mouse Alexa Fluor 594. All three strains for both mAbs were within the fluorescence intensity range of negative binding, close to zero. These results demonstrate the specificity of both Ab521 and Ab652 to *C. difficile* when compared to the three closely related species. When incubated with *C. sordellii*, Ab652 produced a slightly increased median fluorescence intensity of 26.8 AU. Although this fluorescence value is higher than those seen with the other control strains, it remains significantly lower than any fluorescence value observed with *C. difficile*, $P = < 0.05$.

These control species do not provide conclusive evidence that the mAbs do not bind to bacterial strains other than *C. difficile*, however, due to their similarity to *C. difficile*, they were a good starting point. With additional time, perhaps more species would be tested for non-specific binding, to further demonstrate the specific binding of the mAbs to *C. difficile*.

6.3.4 Visualisation of the antibody binding to whole cells using IF microscopy

To enable visualisation of the mAb binding to *C. difficile* cells, IF microscopy was used. The resulting images clearly displayed fluorescence on the surface of the *C. difficile* strains, when incubated with both Ab521 and Ab652. There were six *C. difficile* strains that were tested, the reference strain 630 and the SLTS Ox858, Ox1396, Ox1342, Ox1437a and Ox1533. When incubated with Ab521, all six of these strains displayed greater fluorescence than when incubated with Ab621, providing yet more evidence of the stronger binding affinity of Ab521.

Similarly to the flow cytometry results, IF microscopy also displayed clear binding of both mAbs to Ox1396. From the six *C. difficile* strains, Ox1396 showed the second highest fluorescence value for both Ab521 and Ab652. This binding provided visual confirmation that the ELISA results for Ox1396 were not an accurate depiction of antibody binding to this strain. Furthermore, the IF images also confirmed the flow cytometry results for the closely related species, agreeing that the mAbs did not bind to *C. hiranonis* and *P. anaerobius*, although, Ab652 did show some, albeit slight, binding to *C. sordellii*. The clear binding to *C. difficile* and absence of binding to the closely related strains, allow the specificity of Ab521 and Ab652, and the surface localisation of the binding to be visualised.

6.3.5 Specific surface binding of the mAbs at the epitope site

TEM enables visualisation of the antibody epitope sites, using gold nanoparticles to mark the binding, at greater magnification than is possible with IF microscopy. Although not quantified, a greater number of nanoparticles were observed when *C. difficile* 630 was incubated with Ab521 than with Ab652. This binding further supports the increased binding capability of Ab521 and reveals the specific surface localisation of this binding, which coincides with the, HMW SLP containing, S-Layer.

The control images of *C. difficile* 630 without primary mAb, displayed cells that have not successfully been immunogold labelled, owing to the absence of primary antibody to which the secondary antibody would bind to. The two sets of images are starkly different, with the controls lacking any gold nanoparticles in contrast to the primary incubated *C. difficile* 630, which has an abundance of particles. This disparity in binding exposes the need for the presence of the Ab521 or Ab652 in order to produce a positive signal with *C. difficile*. The lack of binding was also shown with the closely related species, *C. sordellii* which, when incubated with both Ab521 and Ab652 displayed no gold nanoparticles, indicating that the mAbs do not bind non-specifically to *C. sordellii*.

Within the literature on *C. difficile* TEM, cells are always fixed prior to immunogold labelling (Takeoka *et al.*, 1991; Emerson *et al.*, 2009; Piepenbrink *et al.*, 2014). *C. difficile* 630 cells without paraformaldehyde fixing were immunogold labelled, with the same protocol as the fixed cells, and the resulting image showed clear cell lysis and loss of cellular material, data shown in Appendices Figure 9-13.

The TEM images within this project were gained using whole cells, however, thin cross-sections of the cells could be used for future work. As shown by Emerson and colleagues, when immunogold labelling surface proteins in whole *C. difficile* cells, cross sections enable the clear visualisation of the presence of antigen on the external surfaces only (Emerson *et al.*, 2009). This technique would be useful for future TEM analysis of the mAb interactions with whole cells, as it should enable the visualisation of the gold nanoparticles specifically to the cell surface only.

6.3.6 The potential of Ab521 for use in a diagnostic assay or sensor

Lateral flow strips were explored with Ab521 as a possible PoN sensor to detect *C. difficile*, however, the whole cells did not flow freely down the strips. *C. difficile* cells vary in size from 2-8 µm (George *et al.*, 1979), which is smaller than the 15 µm pore size of the membrane, however, *C. difficile* cells can clump and form biofilms (Ethapa *et al.*, 2013; Hammond *et al.*, 2014) perhaps blocking the pores. It may be possible within future work to develop custom membranes with pores large enough to flow whole *C. difficile* cells through.

The strips enabled the flow of gold-conjugated Ab521 to flow freely through the pores and bind to the HMW/LMW SLP complex, providing yet more evidence for the positive binding of this antibody to the HMW SLP. The control lateral flow strip which again used HMW/LMW SLP bound to the membrane and a gold-conjugated prostate-specific antigen (PSA) antibody to not bind to the complex. This simple control was used to show that the SLP complex did not bind all IgGs and that Ab521 had to be specific to the complex. It is recognised that more significant controls could be performed with these strips. For example, closely related whole cells or proteins could be bound to the membrane and the mAbs flowed down the strip, to determine non-specific binding. However, for the brief test to prove that the Ab521 will bind to the complex within lateral flow strips, it was successful.

The promising aspect of lateral flow strips are the associated ease of use and low cost of production (Sajid *et al.*, 2015). Also, it is possible to lyse cells and detect intracellular biomarkers with lateral flow strips, which perhaps would not be applicable for other types of PoN sensors, for example microelectromechanical systems (MEMS) .The bioinformatics approach used within this project was tailored to identify cell-surface associated proteins, as the initial intention was to use any antibodies produced in a MEMS PoN diagnostic sensor. MEMS sensors utilise electrical current to resonate at a particular frequency, and when a significant weight, such as whole bacteria, is bound to the sensor via an antibody, the resonating frequency alters (Arlett *et al.*, 2011). If the unique biomarker was identified to be within the cell and was only accessible post cell lysis, it is unlikely that the additional weight of a protein, rather than the weight of an entire bacterial cell, would be enough to record a significant difference in frequency and provide a positive signal.

6.4 Summary

The five techniques that were used within this chapter to determine antibody binding, verified that both Ab521 and Ab652 bound to whole *C. difficile* cells. Paraformaldehyde-fixed cells were used for flow cytometry, IF microscopy and TEM whereas, native, non-fixed cells were utilised for dot blots and whole cell ELISAs, revealing that binding occurred irrelevant of the fixing status of the cells. The whole cell ELISAs indicated that the mAbs may bind to Ox1396 less than the other SLTS. However, upon further analysis of this binding using flow cytometry and IF microscopy, it was revealed that the mAbs actually bind to

Ox1396 at a more significant rate than some of the other SLTS. Consequently, the poorer binding shown with the ELISAs was not down to the mAbs, having less affinity to this strain, and more likely to have be caused by a lack of cells being bound to the surface of the ELISA plate.

All five techniques determined that both Ab521 and Ab652 bound to all 14 SLTS and *C. difficile* 630, and do not show significant, non-specific binding to the three closely related strains. The results presented in this chapter are not absolute proof that Ab521 and Ab652 are specific to *C. difficile*. However, since the three closely related strains have similar protein sequences to *C. difficile*, the results provide preliminary evidence for *C. difficile* specificity. What is absolutely evident is the ability of Ab521 to positively detect any *C. difficile* strains regardless of SLT clade.

The results from all five methods provided both qualitative and quantitative data, which determined that Ab521 had a stronger binding affinity to whole *C. difficile* cells, than Ab652. Overall, this chapter revealed that Ab521 is the better candidate for use in a PoN sensor as it bound with specificity and sensitivity to whole *C. difficile* cells.

Chapter 7 General discussion

7.1 The bioinformatics approach to biomarker prediction

The results presented within this thesis provide evidence that the bioinformatics approach offers a promising method for the identification of species specific bacterial detection biomarkers. The bioinformatics approach integrated the pre-designed, novel IDRIS system with the manual bioinformatics workflow, the latter being created as part of this work. The IDRIS system, incorporates the TokenDB and subcellular localisation tools, enabling the mining of a large amount of genome data, an exercise which would not be possible to perform manually. With the 2015 IDRIS system and the integrated Web-based browser, all of the annotated whole genome sequence data, within NCBI RefSeq, can be mined for protein sequences that are unique to a bacterial group of interest (GOI), within minutes. The experimental verification of the expression of the epitopes identified and the corresponding development of monoclonal antibodies shows the value of this predictive approach for the production diagnostic detection molecules.

The IDRIS system can be tailored to identify unique biomarkers within any GOI, containing any organisms with whole genome sequences within RefSeq. Furthermore, the system can be used to identify biomarkers within a protein with the subcellular localisation of choice, although only surface associated proteins have so far been investigated. The manual workflow within the bioinformatics approach can also be tailored for user needs. For example, it may be required that the biomarker is unique to the GOI, when compared to only a select choice of other organisms. Or there could be the need to identify biomarkers that are expressed under particular conditions. The manual validation of the IDRIS system and literature search also provides the user with a knowledge base that enables the selection of the most promising biomarker.

Manual confirmation also enables the user to search for biomarker ‘uniqueness’ against the most recent sequence data, including the partial sequences contained within NCBI. This search is vital, especially since the reducing cost of sequencing technology is aiding the exponential increase in bacterial sequence data (Land *et al.*, 2015). As shown in this work, within the space of a few years, the addition of new sequences to the databases can

dramatically alter the predictions about unique biomarkers. However, the accuracy of the approach demonstrated here will only increase as the number of complete and annotated bacterial genome sequences increases.

The top-down approach, designed within this project, of searching the sequence databases prior to laboratory-based procedures, reduces the need for time-consuming and costly laboratory identification of biomarkers by avoiding the conventional approach of large-scale screening of many bacterial strains. Whilst the IDRIS system was previously used to identify a biomarker to differentiate between methicillin-resistant *Staphylococcus aureus* (MRSA) and non-resistant strains (Flanagan *et al.*, 2014), this study was only a proof-of-concept exercise. The IDRIS system prediction of MecA as a target and subsequent identification of a conserved MecA sequence in MRSA, could have been achieved simply by reference to the literature. It is well documented that MecA is modified in MRSA (Llarrull *et al.*, 2009), a simple sequence alignment of MecA protein sequences from all *S. aureus* species could then be used to reveal the section of the MecA protein that is unique to MRSA.

However, in this project, the discovery of peptide sequences unique to *C. difficile* represents a novel finding, since the SlpA biomarker described here has not previously been documented as being unique to *C. difficile*. Nor has the SlpA biomarker been documented as being conserved across all sequenced *C. difficile* strains. Furthermore, the SlpA detection biomarker was not easily identifiable, as the S-layer proteins are not wholly unique to *C. difficile*. In fact there are hundreds of bacteria, within all of the major phylogenetic groups that have S-layers (Sara and Sleytr, 2000) making these proteins one of the most common prokaryote surface structures (Pum *et al.*, 2004).

In this project the IDRIS system was extended with the addition of tools for the systematic manual curation of the biomarkers. With the successful production of antibodies, and their use in proof of concept lateral flow devices, this work therefore represents the first time this full bioinformatics approach has been used to identify and characterise promising biomarkers for PoN sensors. Furthermore, the localisation predictions of the bioinformatics approach were also experimentally tested. The SlpA biomarker was predicted to be on the surface of the protein and the cell, and the successful binding of Ab521 and Ab652 to whole

cells, showed that the epitopes are surface accessible, thus confirming the predictions of the bioinformatics approach.

In summary, the entire bioinformatics approach, incorporating IDRIS, the manual bioinformatics workflow and the laboratory work, is a highly versatile biomarker prediction system. It is envisaged that this approach could be used to predict and characterise biomarkers for many different organisms, not only *C. difficile*.

7.2 Identification of novel monoclonal antibodies

The identification of the 16 amino acid SlpA sequence, demonstrates the power of the system to rapidly sift and highlight promising candidate biomarkers. In the work presented here, these unique linear peptide biomarker sequences were used to develop monoclonal antibodies (mAbs). These biomarkers could also be used to develop other diagnostic reagents such as peptide aptamers. Given the small, linear nature of the peptide epitopes, the mAbs were produced to bind specifically to a sequence and not to a conformation. Whilst the antibodies were developed against the sequence of linear peptide, which is unique to a group of interest, there remains the possibility that the mAbs may bind non-specifically to the conformational structure within other bacteria (Zegers *et al.*, 1995). Additionally, there is the possibility that the mAbs may bind to a bacterial strain that has not yet had its genome sequenced.

Polyclonal antibody sera against the HMW/LMW SLPs has been produced in previous studies, however, it was not identified as specific to *C. difficile*, nor has a study of the binding across the *C. difficile* species been investigated. Antibodies were not purified from the sera and several types of anti-sera were developed within two previous studies, both of which were performed by Calabi and colleagues. Firstly, anti-sera was produced against the HMW/LMW SLP complex from three different *C. difficile* strains, R8366 (ribotype 001), R7404 (ribotype 017) and there reference strain 630 (Calabi *et al.*, 2001). These anti-sera were used to characterise the cleavage of the precursor SlpA protein into the HMW and

LMW SLPs. Secondly anti-sera was produced against the individual HMW and LMW SLPs to determine binding of the SLPs to gastrointestinal tissues (Calabi *et al.*, 2002). Unlike the method used within this project, the antibodies within the anti-sera were polyclonal, and thus were unlikely to bind to the same epitope or have epitopes that are unique to *C. difficile*. Also, unlike this project, the anti-sera were not investigated for their use as *C. difficile* detection molecules. Since the anti-sera were produced against whole proteins, not short linear peptides, any antibodies within the sera could bind via conformational epitopes, unlike the specific linear epitope of Ab521.

Due to the novelty of Ab521 and the successful binding to whole *C. difficile* cells, with *C. difficile* sensitivity and specificity against the bacterial strains tested, a decision was made to patent the antibody. The author of this work is currently working with Newcastle University's Research and Enterprise unit to design a patent application for Ab521.

7.3 Characterisation of monoclonal antibodies produced against the SlpA biomarker

Although there were initially eight mAbs produced against the SlpA biomarker, several of these mAbs did not bind to the whole protein in its native form and since this recognition would be vital in a PoN sensor, those mAbs were disregarded. Two mAbs were deemed to be promising, Ab521 and Ab652, as they bound to the whole cells of representatives from each of the different *C. difficile* surface layer type strain (SLT) groups. There are 13 SLT groups to which all sequenced *C. difficile* strains belong to, with the position of each strain determined by the composition of the SlpA sequence (Dingle *et al.*, 2013). Since both Ab521 and Ab652 mAbs bind to all 13 SLTS then theoretically they will also bind to all *C. difficile* strains. The ability to comprehensively bind all *C. difficile* strains is required to provide high sensitivity in a PoN sensor.

Experimental characterisation of the binding efficiency of Ab521 and Ab652 to a variety of *C. difficile* strains determined that Ab521 bound to *C. difficile* with greater efficiency than Ab652. However, both mAbs bound to all *C. difficile* strains tested, within all techniques used. The efficiency of Ab521 and Ab652 binding to each strain varied according to the method used to estimate it. For example, whole cell ELISAs indicated that both mAbs bound

to Ox1396 with a lower affinity than the other *C. difficile* strains. However, with other techniques, such as flow cytometry, Ab521 and Ab652 were shown to bind efficiently to Ox1396. It is possible that the discrepancies in binding were a result of natural variability in the techniques employed. The ability of ELISA to detect antibody binding to whole cells, depends on the successful adhesion of the bacterial cells to the surface of the ELISA plate (Thermo, 2010). The ELISA plates were observed using an inverted microscope to determine if cells had bound to the plate. The limited capabilities of this kind of microscopy, however, meant that quantification was not possible. The main use for the inverted microscope was to determine that the closely related species had successfully bound to the plate. Therefore, it is possible that Ox1396 did not adhere as efficiently as the other strains. The three different ELISA experiments for both Ab521 and Ab652 determined Ox1396 to be the strain with the lowest absorbance. It could be possible that the *C. difficile* strains differentially express SlpA, accounting for the differences in antibody binding observed with the whole cell biding techniques. However, since Ab521 and Ab652 was determined by flow cytometry and immunofluorescent (IF) microscopy to bind to Ox1396 at a higher affinity than other *C. difficile* strains, it is unlikely that this strain expresses SlpA at a lower rate. The levels of SlpA expression within each *C. difficile* strain could be investigated in future work using techniques such as Northern blotting (Beckmann *et al.*, 2010), two-dimensional gel electrophoresis (Twyman, 2014) or Liquid chromatography–mass spectrometry (LC-MS) (Boetzkes *et al.*, 2012).

When comparing the data from the strains used in both IF microscopy and flow cytometry, Ab521 displayed binding to Ox1437a at the highest level and Ox858 the least with IF microscopy and Ox1533 at the highest level and 630 the least with flow cytometry. These disagreements could have been caused by the limited measuring accuracy of the fluorescence measurements that was performed with the IF microscopy results. For the IF results, ImageJ was used to measure a minimum of five random cells whereas, the flow cytometry measured fluorescence readings from 1,000,000 cells, providing a much greater representation of the sample. The multiple measurements from flow cytometry greatly increased the number of observations and is partly why flow cytometry is stated to be a more sensitive and accurate, quantitative measurement of the activity different antibodies (Bolanos *et al.*, 1988). Although the strains to which Ab521 binds to with the greatest or

least efficiency was found to be variable, the significant binding difference that distinguishes *C. difficile* strains from the closely related species does not alter.

The laboratory validation work was limited by the inability to test all bacterial strains for non-specific binding to the biomarker antibodies. Due to resource availability and time constraints, a set of bacteria had to be selected as an initial screen. The antibodies were produced against small linear peptides and, therefore, were unlikely to recognise epitopes via conformational binding. As a result the bacteria for the screen were chosen via their sequence similarity to *C. difficile* or to the protein in which the biomarker was located, SlpA. Bacteria with similar sequences were chosen since the specific nature of monoclonal antibody binding indicates that the mAbs should not bind to sequences of bacterial strains with decreased similarity. The testing of the three species closely related to *C. difficile*, *C. sordellii*, *C. hiranonis* and *P. anaerobius* showed that the most promising antibody, Ab521 did not show any non-specific binding. Two commensal bacteria were also tested for mAb binding to their denatured proteins, with Ab521 not binding to either *S. aureus* or *B. subtilis*. It is recognised that the size of the bacterial screen for non-specific binding was limited within this project.

There are other species that could be tested, including the species identified in the most recent BLASTP search that contain the most similar sequence to the biomarker, *C. argentinense* and *C. senegalense.* The species that encodes proteins with the closest similarity to the SlpA biomarker was *C. argentinense* of which non-toxigenic samples have been isolated from blood and wounds (Johnson, 2005) but not from faecal samples. Although there is a similarity to the SlpA biomarker, any diagnostic sensor that is designed for faecal samples may not encounter *C. argentinense* and false positives may not be an issue. The genome of *C. senegalense* also encodes a protein with similarity to the SlpA biomarker and was isolated from the faeces of a healthy Senegalese male (Mishra *et al.*, 2012; Ahmed *et al.*, 2014). Since the bacteria were isolated from a faecal sample, a *C. difficile* PoN sensor may come into contact with the species so it requires testing for binding to Ab521. However, there was no evidence to suggest that *C. senegalense* has been isolated from faecal samples in more developed countries and there was very little

literature on this strain. Upon correspondence with Professor John Perry[29], the head of research and development in microbiology at the Newcastle upon Tyne hospitals, it was noted that neither *C. senegalense* no *C. argentinense* had been previously isolated within UK hospitals. Therefore, *C. senegalense* may not be common enough to cause any false positives when detecting *C. difficile*. Nevertheless, if available, this species would require testing with Ab521 for non-specific binding. Further strains for non-specific testing would also include strains such as *C. bartlettii* that are recorded in the pathogen database PARTIC as being closely related to *C. difficile*. *C. bartlettii* has also been isolated from human faeces (Song *et al.*, 2004)and therefore, may be present in the samples being tested for *C. difficile* and could be an important bacterium to test.

Other bacterial species that require testing for non-specific binding with Ab521 includes organisms that may be present in human faeces. These organisms include *Escherichia coli*, *Yersinia enterocolitica*, *Bacillus cereus* and *Aeromonas hydrophila* (Poojary *et al.*, 2014). Since the main symptom for CDI is diarrhoea and it is recommended that only samples exhibiting diarrhoea are tested, therefore, the mAbs will especially require testing against organisms that cause diarrhoea. Organisms that cause diarrhoea include *Shigella*, *Salmonella* and *Campylobacter* (Stallmach *et al.*, 2015). Testing of the Ab521 to these species would be essential to ensure that these samples would not provide false positives, but ideally should be carried out in the form of the final assay in which Ab521 was deployed.

7.4 The binding affinity of Ab521

The K_d for Ab521 binding to the HMW/LMW complex was determined using microscale thermophoresis to be 5.21±0.917 nM. Antibodies with a K_d less than 1 µM have high affinity (Laguna *et al.*, 2015), therefore, Ab521 has a very strong affinity to the HMW/LMW SLP complex. Antibodies described in the literature against *C. difficile* are usually against the *C. difficile* toxins and often, as seen with Ab521, also have binding affinities within the nanomolar range (Demarest *et al.*, 2010; Hussack *et al.*, 2011). Other methods could be used to confirm the binding affinity of Ab521, and also measure a K_d for Ab652, to the

[29] https://www.newcastlelaboratories.com/people/john-d-perry/

HMW/LMW complex. Due to the limited amount of sample available SPR would be an appropriate technique, however, it is unlikely that this method could determine the binding affinity to whole *C. difficile* cells. In order to attain a K_d for the binding of Ab521 or Ab652 to whole cells, different techniques that enable the use of whole cells would be required. The capillaries used with MST are likely to be too thin to allow the free movement of *C. difficile* cells and although whole cell ITC was performed by Howe and colleagues, it was not used to determine a K_d (Howe *et al.*, 2007). Furthermore, SPR is not frequently used with whole cells (Stojanovic *et al.*, 2014) and therefore other techniques would have to be taken into consideration to determine the K_d of Ab521 with whole cells. Flow cytometry has been previously used to determine binding affinities with similar results to those attained with SPR (Geuijen *et al.*, 2005). In this technique, the antibody is titrated against the cells and binding curve is made using the antibody concentration against the mean log fluorescence intensity. ELISAs could also be used to determine a K_d using either traditional or displacement ELISA (Friguet *et al.*, 1985; Orosz and Ovadi, 2002; Bobrovnik, 2003).

7.5 Specific binding to the HMW SLP

Although the Western blots clearly showed Ab521 binding to the HMW SLP, the protein was in its denatured form. The antibody also displayed clear binding to both whole *C. difficile* cells and the HMW/LMW SLP complex, however, further evidence would be needed to definitively prove that Ab521 is binding to the biomarker within the HMW SLP. Since the mAbs were produced against the biomarker, the logical conlusion would be that Ab521 bound to the biomarker epitope within the HMW SLP. Mutant strains could be made that lack the gene, or specific sequence of interest, and used to show that the antibody no longer binds. However, production of SlpA mutants have so far not been successful (Fagan and Fairweather, 2011; Pantaléon *et al.*, 2015), most likely due to the protein being essential for *C. difficile* viability (Dembek *et al.*, 2015). Furthermore, producing any form of *C. difficile* mutant was not possible until 2006 when O'connor and colleagues 'knocked-out' two putative response regulator genes, *rgaR* and *rgbR* (O'Connor *et al.*, 2006). A tool published in 2007, named ClosTron[30] enables the directed construction of stable mutants in *C. difficile*

[30] http://www.clostron.com/clostron1.php

using a bacterial group II intron (Heap *et al.*, 2007). Although a full *C. difficile slpA* mutant has not been successfully produced, perhaps using ClosTron a mutant lacking the specific HMW biomarker could be made. This mutant strains would enable to testing of the antibody to confirm the specific amino acid sequence to which Ab521 binds.

7.6 Using Ab521 for CDI diagnosis

It has been reported that 15-25 % of all antibiotic associated diarrhoea (AAD) cases are caused by CDI (Cohen *et al.*, 2010; Peterson *et al.*, 2011) therefore, it can be deduced that there are up to 85 % of AAD cases that are negative for CDI. The initial step of a two-step algorithm for diagnosing CDI should be an inexpensive and rapid test that accurately reduces the number faecal of samples requiring additional investigation, without providing false negatives. Currently EIAs are often recommended for use as they are inexpensive with GDH EIAs being recommended due to the high sensitivity and low false negative prediction of the test (Cohen *et al.*, 2010; Wilcox, 2013).

There is evidence, that GDH EIAs are not as effective at detecting the *C. difficile* ribotype 027 as other ribotypes (Tenover *et al.*, 2010). In the 2007/08 epidemic, the 027 ribotype was responsible for 40 % of all CDI cases in the UK (Brazier *et al.*, 2008) and within Tenover and co-worker's research, 28.6 % of *C. difficile* isolates were ribotype 027. It is therefore vital that a *C. difficile* PoN sensor is able to detect this ribotype. In the detection systems developed in this work, it was shown that both Ab521 and Ab652 bind to *C. difficile* Ox160, which is strain R027 (Dingle *et al.*, 2013). Both mAbs bound to the denatured HMW SLP within Ox160 cell lysates and also bound to whole Ox160 cells using ELISA and flow cytometry. The flow cytometry results indicated that the Ab521 and Ab652 clearly bind to all *C. difficile* SLTS and to Ox160 with greater efficiency than six of the other strains. Therefore, it is unlikely that Ab521 or Ab652 would be less effective in detecting the epidemic strain, ribotype 027.

Furthermore, GDH EIAs are not specific, that is they do not provide low false positive results, which is likely due to the GDH enzyme being produced by many different bacteria, including *C. sordellii*, which demonstrates cross-reactivity with GDH EIAs developed for *C. difficile* (Commichau *et al.*, 2008; Harper *et al.*, 2010; Burnham and Carroll, 2013). The

bioinformatics approach used within this project promises to identify epitopes that are much less susceptible to cross reactivity. The unambiguous nature of the genome data and the ability to rapidly carry out a systematic, comparative analysis of epitopes from a wide range of strains reduces the likelihood of any cross reactivity. Ab521 was produced against the SlpA biomarker and did not bind to three closely related species, or two commensal species, *S. aureus* and *B. subtilis,* showing specificity to *C. difficile*. Within the closely related species was *C. sordellii* ATCC 9714 to which Ab521 did not bind, either as whole cells, native proteins or denatured cell lysates. Therefore, under these particular conditions, Ab521 is predicted to be more specific than the antibodies used in GDH EIAs. Although a PoN sensor incorporating Ab521 would not differentiate between toxigenic and non-toxigenic *C. difficile* and thus, testing for toxin A and B would still be required, it could reduce the false positives produced from recognising biomarkers from bacteria other than *C. difficile*. However, this would also rely on the sensitivity of the selected sensor. There is therefore, a possible use for the mAb within a PoN sensor, providing the first step of a two-step algorithm, and removing up to 85 % of samples from further testing.

Although Ab521 displays sensitivity by detecting representatives of all *C. difficile* strains, the type of sensor employed would also influence the sensitivity of a *C. difficile* diagnostic test and therefore need to be taken into account. For example, the limit of detection (LOD) of lateral flow strips range from mg/ml to µg/ml (Liu *et al.*, 2006; Posthuma-Trumpie *et al.*, 2009), whereas, ELISAs can detect antigens in the range of ng/ml or even pg/ml (Tang and Hewlett, 2010; Khalil *et al.*, 2011; Chiodo *et al.*, 2013). Many other aspects of the test can alter the sensitivity of sensors, including conjugates used, membrane/plate type and complexity of the sample (Sajid *et al.*, 2015).

Using ELISA as the sensor technique, with equal optical densities of bacterial cells, Ab521 was shown to have a LOD of 1 ng/ml however, this was against pure samples of *C. difficile*, the concentration of which would not occur within patient faecal samples. Therefore, the LOD of the mAb would be dependent on the antigen concentration within samples. Attempts could be made to determine the LOD of Ab521 with cells at differing ODs, however, it would not provide an accurate antigen concentration. Known concentrations of the HMW/LMW complex could be used to determine the LOD for each antibody

concentration, although it is unlikely this would translate to whole cells. Ultimately, the LOD of Ab521 would have to be determined within the PoN sensor of choice and with varying numbers of *C. difficile* cells.

Another use for Ab521 would be to incorporate it into a dual PoN test that detects both the organism and the toxins. Currently a diagnostic company called Alere has the European and US patent for the dual use of both GDH and toxin detection together in a single EIA (Boone *et al.*, 2013). Diagnostic companies with their own versions of toxin EIAs (Barbut *et al.*, 2014; Davies *et al.*, 2015) are interested in a second biomarker to detect *C. difficile* cells, to circumvent the patent and use in a dual test of their own. This need for a biomarker specific to *C. difficile* opens up the possibility of licencing Ab521 to these diagnostic companies.

A factor that would affect the sensitivity of any PoN sensor integrating Ab521, is the number of vegetative *C. difficile* cells compared to spores within faecal samples. *C. difficile* is an anaerobic, sporulating organism, yet Ab521 was produced against a biomarker that was prevalent on the surface of vegetative cells and thus would not detect spores. Several studies have shown that there are more vegetative cells than spores within faeces, with one study stating there were ten times more vegetative cells within CDI patients (Jump *et al.*, 2007). A second study performed with mice also found more vegetative cells than spores within faeces (Howerton *et al.*, 2013). Therefore, Ab521 would be likely to successfully detect *C. difficile* within faecal samples.

Identification of a PoN sensor that could successfully incorporate Ab521 would be required if this antibody was going to be used to detect *C. difficile*. The initial work with lateral flow strips had issues with the flow of bacteria, as the *C. difficile* was too large to flow through the pores. If lateral flow strips were to be investigated further, additional product searching would be required to identify a membrane with pores large enough to fit *C. difficile* samples through. Further optimisation of the strips could also be performed with dilution of the sample, perhaps enabling successful use of the current membranes. However, any dilution deemed successful would also have to be noted as a pre-treatment requirement within a PoN sensor, thus decreasing the rapidity of the test.

Other sensors that could be explored with Ab521 include electrochemical, mass-based, magnetic and optical sensors. Electrochemical sensors convert a biological event, such as *C. difficile* cells being captured by Ab521 into an electronic signal (Grieshaber *et al.*, 2008). Mass-based sensors such as Piezoelectric biosensors or microelectromechanical systems operate on the principle that a change in mass can cause a change in resonance frequency (Byrne *et al.*, 2009; Arlett *et al.*, 2011). Magnetic sensors can be used by coating antibodies onto magnetic beads, which bind to the antigen (*C. difficile* cells) and can be separated from samples using magnets (Jaffrezic-Renault *et al.*, 2007). Finally, optical sensors, including surface plasmon resonance can be used as biosensors (Wijaya *et al.*, 2011). Although the work within this project determined Ab521 to be a sensitive recognition molecule, that is it bound to all tested *C. difficile* samples, the choice of sensor used with the mAb could affect the overall sensitivity of the test. It would be interesting to test whether the antibodies produced in this project could be deployed in these additional formats.

7.7 Possible therapeutic use of Ab521

In the work described here, mAbs were investigated for use as *C. difficile* detection molecules. However, antibodies are also used as therapeutic molecules (Chames *et al.*, 2009). There have been several studies that have shown that *C. difficile* SLPs play a critical role in bacterial adherence to host cells (Calabi *et al.*, 2001; Drudy *et al.*, 2001; Merrigan *et al.*, 2013). This adherence promotes *C. difficile* colonisation of the gut and enables the infection to persist (Spigaglia *et al.*, 2013). Additionally SLPs have been shown to modulate immune responses (Ausiello, 2006; Collins *et al.*, 2014). Research has been performed that suggests antibodies against *C. difficile* SLPs could be a potential therapeutic agent. O'Brien and colleagues showed that SLP anti-sera protected hamsters from CDI (O'Brien *et al.*, 2005) and also Kandalaft and co-workers reported that SLP antibodies reduced the mobility of *C. difficile* (Kandalaft *et al.*, 2015). Therefore in future studies, the HMW SLP mAbs could be examined for the ability to reduce *C. difficile* adherence and mobility. The traditional method of treating CDI is the use of antibiotics, which can alter the composition of the gut microbiome and promote antibiotic resistance. Therapeutic antibodies could reduce the need for the use of antibodies in CDI cases.

There are currently no vaccines available for *C. difficile* (Vindigni and Surawicz, 2015), however, there are several groups investigating a possible toxin based vaccine that utilises detoxified toxins (Foglia *et al.*, 2012; Donald *et al.*, 2013; Karczewski *et al.*, 2014). Since the S-layer is the predominant surface antigen, there is potential for its use in a *C. difficile* vaccination (Spigaglia *et al.*, 2013). It is therefore anticipated that a possibility for successful *C. difficile* vaccination could be an integrated S-layer and detoxified toxin vaccine.

7.8 Summary

The bioinformatics approach was identified as a novel system that characterises biomarkers from organisms within a particular group of interest. Laboratory production of antibodies against the *C. difficile* biomarker, which bound to whole *C. difficile* cells with high affinity, confirmed the value of the bioinformatics approach. The most promising mAb, Ab521, successfully identified all strains from representative groups of the entire *C. difficile* species, and did not non-specifically bind to the closely related species that were tested. This antibody displays potential as a diagnostic tool and could also be investigated as a therapeutic agent. To protect both the novelty and value of Ab521, a patent application is currently being written. There is huge promise that the combination of bioinformatics driven biomarker identification and antibody generation, could be applied to PoN sensors for many other pathogenic microorganisms.

Chapter 8 Conclusions and future work

The aim of this project was to establish a bioinformatics approach for the identification of protein sequence biomarkers that are unique to a given strain, or group of strains of interest. The aim was that these biomarkers could act as suitable epitopes for the development of diagnostic reagents, such as antibodies. In the work presented here, the species group for *C. difficile* was chosen as the group of interest to which the system would be applied.

There were two main aspects of the bioinformatics approach for the identification of diagnostic biomarkers. The initial step involved the application of an automated system, named the IDRIS system, which was designed by Dr Keith Flanagan and colleagues in a previous study (Flanagan *et al.*, 2014). The IDRIS system had been used by Keith Flanagan to mine the fully sequenced genomes of 2764 organisms and provide sequence regions that were unique to *C. difficile.* A list of the associated proteins containing these unique regions and predictions from four protein subcellular localisation tools were provided for use in this project. Since an objective of this work was to produce biomarkers for use in rapid point of need (PoN) sensors, only proteins that were predicted to be surface associated were selected from this list. The second step, a manual bioinformatics workflow, was designed within this project to operate on the results from the IDRIS system and used a combination of sequence similarity searches, protein structure prediction and expression data analysis. All of these analysis tools were used to determine biomarkers of at least 15 amino acids that were conserved across the *C. difficile* species. These biomarkers were required to be different to the next most similar sequence, not found within *C. difficile*, by a minimum of four amino acids, in order for the biomarker to be classed as unique to *C. difficile*. Application of this bioinformatics approach successfully identified a very promising, unique, surface associated, peptide sequence biomarker, within the SlpA protein of *C. difficile*.

The second objective for validating the success of the bioinformatics approach was to use laboratory methods, to confirm the SlpA biomarker for potential use in a PoN sensor. Several monoclonal antibodies (mAbs) were produced against the biomarker and two of these successfully bound to all tested *C. difficile* strains. Of these two mAbs, Ab521 was

deemed the most promising as it showed the strongest binding affinity to whole *C. difficile* cells. Furthermore, Ab521 displayed novelty by binding to strains across the *C. difficile* species and specifically to *C. difficile*, amongst the species tested. These results did indeed validate the bioinformatics analysis for three reasons. Firstly, the biomarker was displayed on the surface of *C. difficile* as predicted. Secondly, the biomarker was present in all tested *C. difficile* strains, including the surface layer type strains that represent the entire sequenced *C. difficile* species. Finally, the biomarker was shown to be specific to *C. difficile* when tested against the closely related strains. The novelty of Ab521 highlights the value of the bioinformatics approach, as the system could be used to provide similar results for other bacteria, including other pathogens.

Traditional biomarker identification tends to use laboratory data, from techniques such as protein analysis or microarrays, to perform large-scale searches for unique information within this data, for example a particular protein or increase in gene expression (Freiberg *et al.*, 2006; Natesan and Ulrich, 2010; Yousef *et al.*, 2014). Often this requires the production of a lot of experimental data, which may not identify a unique biomarker. Although bioinformatics has recently emerged as a major tool in bacterial biomarker identification (Li *et al.*, 2002; Zhang *et al.*, 2014), tools are usually used for comparisons of subject genomes to reference genomes (Shao *et al.*, 2010). Whereas, the bioinformatics approach described within this project mines a huge amount of sequence data, with thousands of bacterial genomes, to determine biomarkers that are unique before laboratory work begins. This top-down method increases the likelihood of a discovering a truly unique biomarker and reduces the need for expensive, time-consuming laboratory work. The workflow also searches for fragments as small as 15 amino acids in length unlike traditional methods where entire proteins are searched for (Rabilloud and Triboulet, 2013).

A thorough literature search was performed and a bioinformatics approach of this kind had not been previously described, not including the IDRIS system predecessor (Flanagan *et al.*, 2014). A similar fragmenting technique has been used before for population genomics, however, the fragmenting technique was used with a subject genome sequence against a known reference sequence (Catchen *et al*., 2013). This system is novel due to the integration of mining entire genome data-sets, with the computational prediction of unique

surface biomarkers that are accessible to antibodies and are constitutively expressed. Although this system has been used with *C. difficile*, it would be possible to mine for biomarkers of GOI's containing other organisms.

8.1 Limitations of the bioinformatics approach

Several limitations of the bioinformatics workflow were addressed within this project and will be discussed. Firstly, the automated IDRIS system mines only the fully annotated genome sequences within RefSeq which equates to 2764 organisms. There are many more organisms with partial genomes or smaller sequences, within NCBI, that would require analysis of their sequences. For example, there could be bacterial strains that should ideally be within the group of interest and, therefore, these strains must also be analysed for the presence of the biomarker. Also, the tokens that are classed as unique via the IDRIS system could possibly be present within these other sequences and thus not be unique. The BLASTP searches within the bioinformatics workflow were used to address these two issues, enabling all of the bacterial sequence data within NCBI to be searched. Using BLASTP searches increases the likelihood that the identified biomarkers are truly unique. The issue remains that only data that has been sequenced could be searched for biomarker similarity, however, this is the case with all methods that involve computational mining. Increasing availability of sequence data will enable the constant improvement of this project's bioinformatics workflow, with the more data that is added, the increased likelihood that the predicted biomarkers are truly unique.

8.2 Future work

To further improve the bioinformatics workflow a number of tolerance tests could be performed. Firstly, the IDRIS system currently classes a token (15 amino acid sequence) as unique, even if up to 14 of the 15 amino acids are present within a bacterium outside of the GOI. This 'uniqueness threshold' was selected to provide the largest number of 'unique' tokens for the down-stream analysis. Future work could include determining the maximum number of unique amino acids, within the token, that provides an acceptable number of tokens yet increases the likelihood of them being truly unique.

Secondly, biomarkers are currently required to have a minimum of four amino acids different to the closest related sequence, not in the GOI. This minimum difference was selected to retain specificity of any recognition molecules, such as antibodies, to the strains within the GOI. However, due to the highly specific nature of monoclonal antibodies, it may be possible to decrease this minimum amino acid difference while retaining GOI specificity. Further work could be performed testing antibody binding against sequences of decreasing similarity to the epitope. The minimum number of amino acid changes needed to inhibit mAb binding can then be established. Several other variables to will also have to be taken into account such as position and proximity of the variations in the sequence.

Future work will also require thorough testing of the bioinformatics approach with the prediction and validation of a biomarker for another GOI. However, within the context of this work, the bioinformatics approach produced accurate identification of unique biomarkers for the development of diagnostic reagents for use in PoN sensors.

Chapter 9 Appendices

9.1 Example of a BSA standard curve for a BCA assay

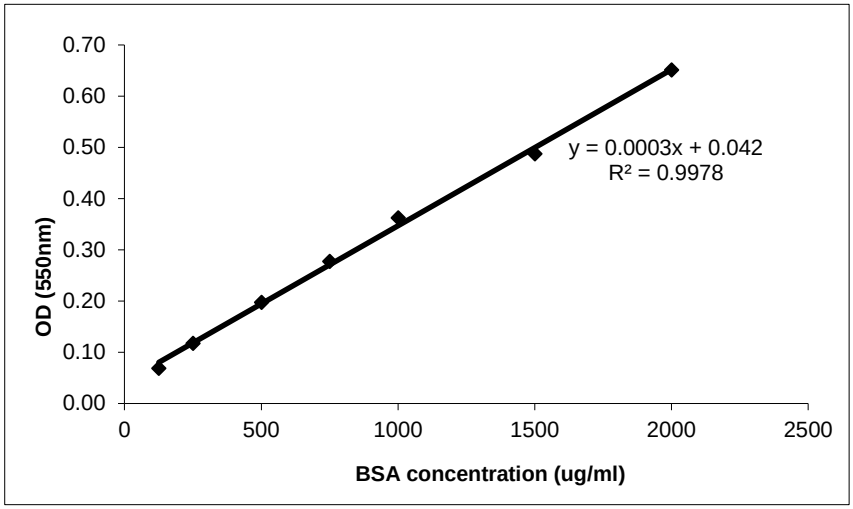


Figure 9-1. An example of a BCA assay, BSA standard curve. Absorbance measured at 550 nm is plotted against concentration of standard BSA samples

9.2 Bioinformatics results

The 28 *C. difficile* proteins that were successful for all tools within the surface association pipeline, SignalP, LipoP, PSORTb and TMHMM, with SlpA highlighted.

Protein ID GI:	NCBI Accession
260683411	YP_003214696.1
260684371	YP_003215656.1
260684408	YP_003215693.1
126700403	YP_001089300.1
126700410	YP_001089307.1
126699809	YP_001088706.1
126699358	YP_001088255.1
260687007	YP_003218140.1
126697956	YP_001086853.1
260684362	YP_003215647.1
260688032	YP_003219166.1
260687059	YP_003218192.1
260688067	YP_003219201.1
260688028	YP_003219162.1
260684369	YP_003215654.1

260688030	YP_003219164.1	
126701015	YP_001089912.1	
260684373	YP_003215658.1	
126700414	YP_001089311.1	
260683463	YP_003214748.1	
260684356	YP_003215641.1	
260683792	YP_003215077.1	
126700409	YP_001089306.1	SlpA
260688015	YP_003219149.1	
126700398	YP_001089295.1	
126700447	YP_001089344.1	
260687452	YP_003218586.1	
126700412	YP_001089309.1	

Table 9-1. The 28 C. difficile proteins that were successful for all tools within the surface association step of the bioinformatics approach . The proteins are listed with the NCBI accession number, which were successful for all tools within the surface association pipeline, SignalP, LipoP, PSORTb and TMHMM, with SlpA highlighted

9.3 *C. difficile* 630 SlpA protein sequence

The *C. difficile* 630 protein sequence, NCBI Accession YP_001089306, with the signal peptide, LMW SLP, HMW SLP and unique biomarker highlighted.

MNKKNIAIAMSGLTVLASAAPVFAATTGTQGYTVVKNDWKKAVKQLQDGLKDNSIGKITVSFNDGVVGEV
APKSANKKADRDAAAEKLYNLVNTQLDKLGDGDYVDFSVDYNLENKIITNQADAEAIVTKLNSLNEKTLI
DIATKDTFGMVSKTQDSEGKNVAATKALKVKDVATFGLKSGGSEDTGYVVEMKAGAVEDKYGKVGDSTAG
IAINLPSTGLEYAGKGTTIDFNKTLKVDVTGGSTPSAVAVSGFVTKDDTDLAKSGTINVRVINAKEESID
IDASSYTSAENLAKRYVFDPDEISEAYKAIVALQNDGIESNLVQLVNGKYQVIFYPEGKRLETKSANDTI
ASQDTPAKVVIKANKLKDLKDYVDDLKTYNNTYSNVVTVAGEDRIETAIELSSKYYNSDDKNAITDKAVN
DIVLVGSTSIVDGLVASPLASEKTAPLLLTSKDKLDSSVKSEIKRVMNLKSDTGINTSKKVYLAGGVNSI
SKDVENELKNMGLKVTRLSGEDRYETSLAIADEIGLDNDKAFVVGGTGLADAMSIAPVASQLKDGDATPI
VVVDGKAKEISDDAKSFLGTSDVDIIGGKNSVSKEIEESIDSATGKTPDRISGDDRQATNAEVLKEDDYF
TDGEVVNYFVAKDGSTKEDQLVDALAAAPIAGRFKESPAPIILATDTLSSDQNVAVSKAVPKDGGTNLVQ
VGKGIASSVINKMKDLLDM

SlpA precursor protein is 719 amino acids, 2157 bp

373 amino acids – HMW SlpA – Mr = 39.45 kDa

322 a.a – LMW SlpA – Mr = 34.3 kDa

24 amino acids - Signal peptide

Amino acid position 641-656 = SlpA biomarker

9.4 DNA sequencing results

C. difficile 630 DNA sequence results (biomarker highlighted)

AATTACTATTATACATTAAAAATCCATAATAGTAGACAATTATTTTGATTTATAGTATAATGATTTTACTTAAAT
AAATTAAACATCTTTTTATTTAGGGATTTCTCACATAAAATAGAGGCTATTTATAGCCTCTATTTTATTTTATATT
TATTTTAAATAAAAAAGACTTCTCATGAGAGAAGCCTTTTTCTATTTAAAGTTTTATTAAAACTTATATTACATAT
CTAATAAATCTTTCATTTTGTTTATAACTGAAGAAGCTATACCTTTACCTACTTGAACTAAGTTAGTTCCACCATC
TTTAGGAACTGCTTTACTTACAGCTACATTTTGGTCAGAAGATAAAGTATCAGTAGCTAGTATGATTGGAGCTG
GAGACTCCTTAAATCTACCTGCTATTGGTGCTGCTGCTAAGGCATCTACTAATTGATCTTCTTTAGTAGAACCAT
CTTTTGCAACAAAGTAATTCACAACTTCACCATCTGTGAAATAATCATCTTCTTTTAAAACTTCAGCATTAGTTG
CTTGTCTATCATCTCCACTTATTCTATCTGGAGTTTTTCCAGTTGCACTATCTATTGACTCTTCAATCTCTTTAGAT
ACGCTATTTTTTCCACCTATTATATCAACATCAGAAGTTCCTAAGAAACTCTTAGCATCATCACTTATTTCTTTTG
CTTTTCCATCTACAACTACTATTGGAGTAGCATCTCCATCTTTAAGTTGAGAAGCAACTGGAGCTATACTCATA
GCATCTGCTAATCCAGTACCACCAACTACAAATGCTTTATCATTATCAAGACCTATTTCATCAGCTATTGCTAAA
GAAGTTTCGTATCTGTCTTCTCCTGATAATCTAGTAACTTTAAGACCCATGTTTTTCAATTCATTTTCTACATCTT
TAGATATAGAATTAACTCCACCAGCTAAATAAACTTTTTTAGAAGTATTTATACCAGTGTCACTCTTTAGTTCAT

ExPASY Translate of *C. difficile* 630 sequencing results (biomarker highlighted)

3'5' Frame 3

E L K S D T G I N T S K K V Y L A G G V N S I S K D V E N E L K N Met G L K V T R L S G E D R Y E T S L A I A
D E I G L D N D K A F V V G G T G L A D A Met S I A P V A S Q L K D G D A T P I V V V D G K A K E I S D D
A K S F L G T S D V D I I G G K N S V S K E I E E S I D S A T G K T P D R I S G D D R Q A T N A E V L K E D D
Y F T D G E V V N Y F V A K D G S T K E D Q L V D A L A A A P I A G R F K E S P A P I I L A T D T L S S D Q N
V A V S K A V P K D G G T N L V Q V G K G I A S S V I N K Met K D L L D Met Stop Y K F Stop Stop N F K
Stop K K A S L Met R S L F Y L K Stop I Stop N K I E A I N S L Y F Met Stop E I P K Stop K D V Stop F I
Stop V K S L Y Y K S K Stop L S T I Met D F Stop C I I V I

9.5 Dot blots

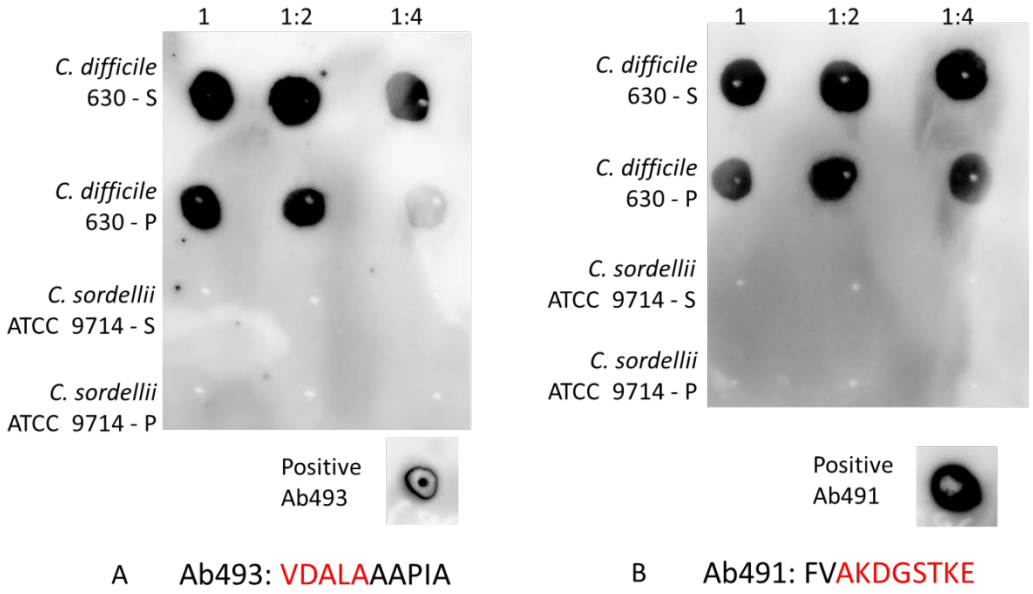


Figure 9-2. Dot blots for Ab493 and Ab491 against *C. difficile* . Samples used were low pH glycine extracted surface proteins (S) and the residual pellet, post extraction (P) for *C. difficile* 630 and *C. sordellii* ATCC 9714 with Ab493 (A) and Ab491 (B). The primary antibody was directly blotted on to the membrane as a positive control. The biomarker regions of the antibody epitopes are displayed in red.

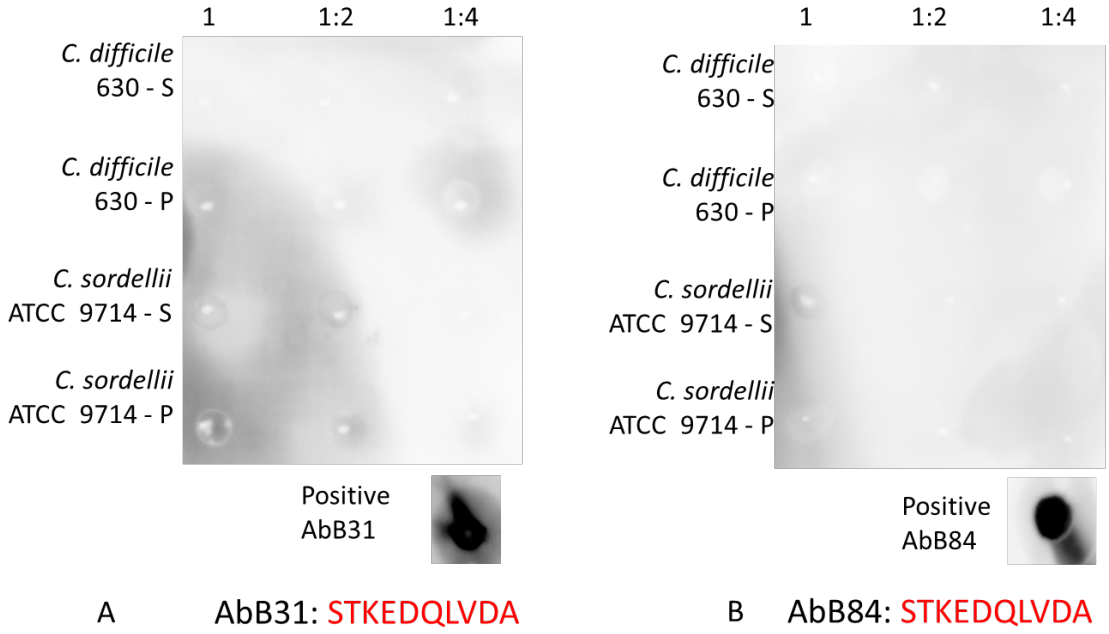


Figure 9-3. Dot blots for AbB31 and AbB84 against *C. difficile* . Samples used were low pH glycine extracted surface proteins (S) and the residual pellet, post extraction (P) for *C. difficile* 630 and *C. sordellii* ATCC 9714 with AbB31 (A) and AbB84 (B). The primary antibody was directly blotted on to the membrane as a positive control. The biomarker regions of the antibody epitopes are displayed in red.

9.6 Binding of the mAbs to proteins within *C. difficile* and the closely related strains

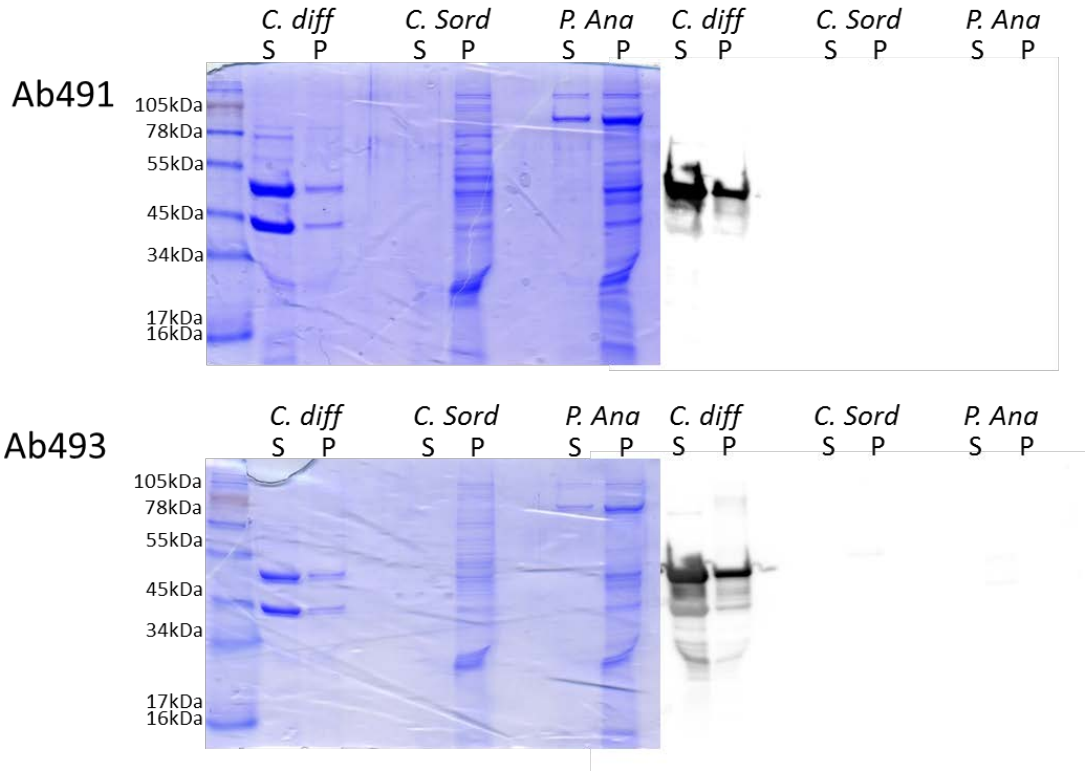


Figure 9-4. Western blot analysis of Ab491 and Ab493 against species closely related to *C. difficile*. Western blot (right) and concomitant SDS-PAGE gel (left) of the low pH glycine extracted surface proteins (S) and residual pellet (P) of *C. difficile* 630 and the closely related strains *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330 with Ab491 (A) and Ab493 (B).

9.7 Binding of the mAbs to proteins within commensal bacteria that are also found within faecal samples

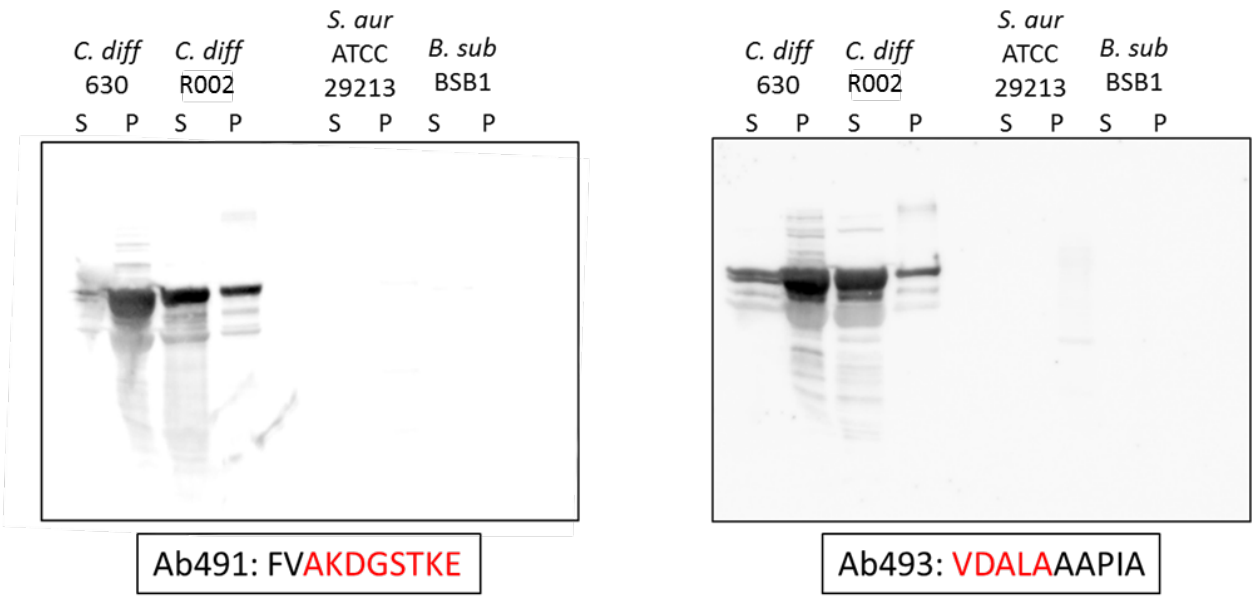


Figure 9-5. Western blot analysis of Ab491 and Ab493 against *S. aureus* and *B. subtilis* . The low pH glycine surface layer proteins (S) and remaining pellet (P) of *C. difficile* 630 and R002, *S. aureus* ATCC 29213 and *B. subtilis* BSB1 were analysed with Ab491 (left) and Ab493 (right).

9.8 BLAST results of the mAb epitopes against the closely related species, *P. anaerobius*

Top BLASTP result for Ab491-FVAKDGSTKE against *P. anaerobius*

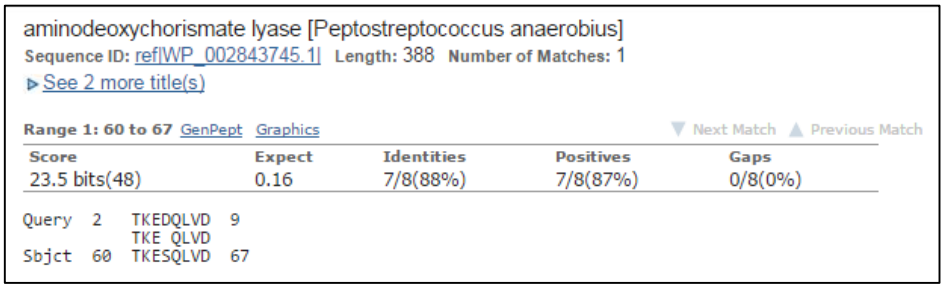
MgtE intracellular domain protein [Peptostreptococcus anaerobius VPI 4330 = DSM 2949]
Sequence ID: gb|EKX94484.1| Length: 236 Number of Matches: 1

Range 1: 193 to 202 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
21.4 bits(43)	0.72	6/10(60%)	7/10(70%)	0/10(0%)

```
Query  1    FVAKDGSTKE  10
            F+ KD ST E
Sbjct  193  FISKDSSTME  202
```

Top BLASTP result for Ab521-STKEDQLVDA against *P. anaerobius*

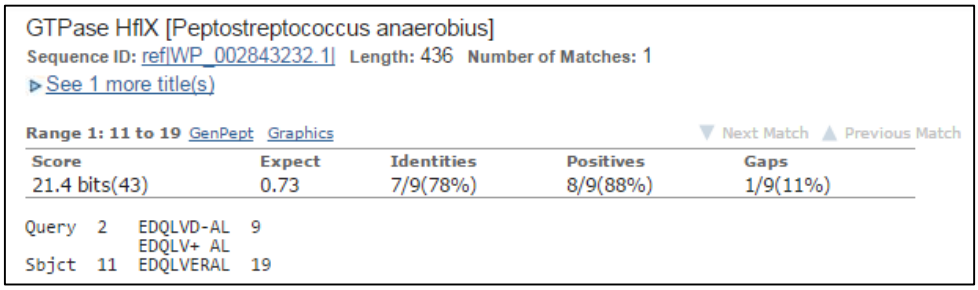

aminodeoxychorismate lyase [Peptostreptococcus anaerobius]
Sequence ID: ref|WP_002843745.1| Length: 388 Number of Matches: 1
See 2 more title(s)

Range 1: 60 to 67 GenPept Graphics Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
23.5 bits(48)	0.16	7/8(88%)	7/8(87%)	0/8(0%)

```
Query  2   TKEDQLVD  9
           TKE QLVD
Sbjct  60  TKESQLVD  67
```

Top BLASTP result for Ab652-KEDQLVDALA against *P. anaerobius*


GTPase HflX [Peptostreptococcus anaerobius]
Sequence ID: ref|WP_002843232.1| Length: 436 Number of Matches: 1
See 1 more title(s)

Range 1: 11 to 19 GenPept Graphics Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
21.4 bits(43)	0.73	7/9(78%)	8/9(88%)	1/9(11%)

```
Query  2   EDQLVD-AL  9
           EDQLV+ AL
Sbjct  11  EDQLVERAL  19
```

9.9 Specific whole cell binding of Ab521 and Ab652 to the *C. difficile* SLTS

Whole cell ELISA, raw data.

<>	Ox1424	Ox1437a	Ox1523	Ox1533	Ox1896	Ox2404	630	no cells	*C. hiranoni*	*C. sordelli*	*P. anaerob*	Controls	
10µg 521	1.8495	1.8989	1.872	1.8964	1.9021	1.9746	1.8914	0.049	0.4196	0.9925	1.0294	0.0506	No 1AB
10µg 521	1.7932	1.6908	1.6974	1.6713	1.6753	1.7257	1.5786	0.0419	0.3483	0.8934	1.2017	0.0519	No 1AB
5µg 521	1.7547	1.7156	1.6896	1.6904	1.7263	1.7258	1.7000	0.105	0.2158	0.5737	0.6337	0.0456	No 2Ab
5µg 521	1.7982	1.7057	1.7016	1.6294	1.7111	1.7119	1.6544	0.1438	0.1642	0.5821	0.6702	0.0428	No 2Ab
1µg 521	1.7553	1.8271	1.6853	1.5708	1.7723	1.8197	1.7733	0.0801	0.0773	0.1752	0.2074	0.0349	No OPD
1µg 521	1.7854	1.8106	1.6843	1.6017	1.8128	1.7601	1.8037	0.0535	0.0714	0.1888	0.2149	0.034	No OPD
0.1µg 521	1.654	1.7101	1.6649	1.6017	1.6642	1.7788	1.6475	0.0552	0.0422	0.0586	0.0803	1.2528	1Ab
0.1µg 521	1.6934	1.8992	1.81	1.6204	1.7386	1.8784	1.8341	0.0428	0.0525	0.0681	0.091	1.2964	2Ab

Figure 9-6. Raw data of whole cell ELISA with Ab521. The ELISA used whole cells of *C. difficile* SLTS and 630, and the closely related species, *C. hiranonis* DSM-13275, *C. sordellii* ATCC 9714 and *P. anaerobius* VPI 4330. The cells were incubated with Ab521 at varying concentrations.

Statistical data of the one-way ANOVA analysis of the Ab652 5 µg/ml whole cell ELISA data.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Absorbance

	(I) Bac_No	(J) Bac_No	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	630.0	P. anaerobius	.6619500*	.0186898	.000	.589345	.734555
		C. hiranonis	.7273500*	.0186898	.000	.654745	.799955
		C. sordellii	.7001750*	.0186898	.000	.627570	.772780
		Ox1121	.3555000*	.0215811	.000	.271663	.439337
		Ox1145	-.0151500	.0215811	1.000	-.098987	.068687
		Ox1192c	.5538500*	.0215811	.000	.470013	.637687
		Ox1342	-.0229000	.0215811	1.000	-.106737	.060937
		Ox1396	.5916000*	.0215811	.000	.507763	.675437
		Ox1424	.3934000*	.0215811	.000	.309563	.477237
		Ox1437a	-.1046000*	.0215811	.006	-.188437	-.020763
		Ox1523	.4986500*	.0215811	.000	.414813	.582487
		Ox1533	.2721000*	.0215811	.000	.188263	.355937
		Ox160	.2558000*	.0215811	.000	.171963	.339637
		Ox1896	.0537000	.0215811	.558	-.030137	.137537
		Ox2404	.0140000	.0215811	1.000	-.069837	.097837
		Ox575	-.0426000	.0215811	.855	-.126437	.041237
		Ox858	-.0267000	.0215811	.997	-.110537	.057137
	P. anaerobius	630.0	-.6619500*	.0186898	.000	-.734555	-.589345
		C. hiranonis	.0654000*	.0152602	.021	.006118	.124682
		C. sordellii	.0382250	.0152602	.547	-.021057	.097507
		Ox1121	-.3064500*	.0186898	.000	-.379055	-.233845
		Ox1145	-.6771000*	.0186898	.000	-.749705	-.604495
		Ox1192c	-.1081000*	.0186898	.001	-.180705	-.035495
		Ox1342	-.6848500*	.0186898	.000	-.757455	-.612245
		Ox1396	-.0703500	.0186898	.065	-.142955	.002255
		Ox1424	-.2685500*	.0186898	.000	-.341155	-.195945
		Ox1437a	-.7665500*	.0186898	.000	-.839155	-.693945
		Ox1523	-.1633000*	.0186898	.000	-.235905	-.090695
		Ox1533	-.3898500*	.0186898	.000	-.462455	-.317245
		Ox160	-.4061500*	.0186898	.000	-.478755	-.333545
		Ox1896	-.6082500*	.0186898	.000	-.680855	-.535645
		Ox2404	-.6479500*	.0186898	.000	-.720555	-.575345
		Ox575	-.7045500*	.0186898	.000	-.777155	-.631945
		Ox858	-.6886500*	.0186898	.000	-.761255	-.616045

Figure 9-7. One-way ANOVA analysis, performed with IBM SPSS version 21, of the whole cell ELISA data from Ab652 at 5 µg/ml, with post hoc Tukey test. The significance value was set to P = < 0.05.

9.10 Immunofluorescent microscopy results of *C. difficile* strains and closely related species incubated with the mAbs

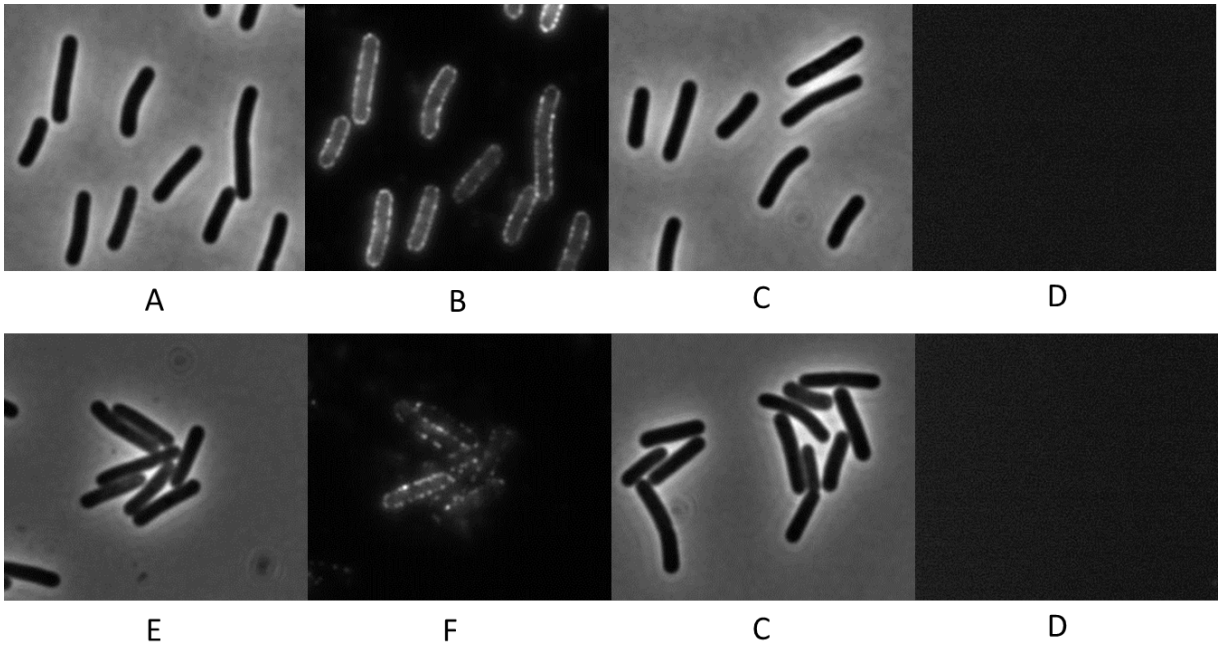


Figure 9-8. Immunofluorescence microscopy images of Ox858 incubated with Alexa Fluor 594 anti-mouse IgG and Ab521, brightfield image (A), mCherry filter (B). Alexa Fluor 594 anti-mouse IgG and Ab652 brightfield image (E), mCherry filter (F). Control of Ox858 with Alexa Fluor 594 anti-mouse IgG and no primary antibody brightfield image (C), mCherry filter (D).

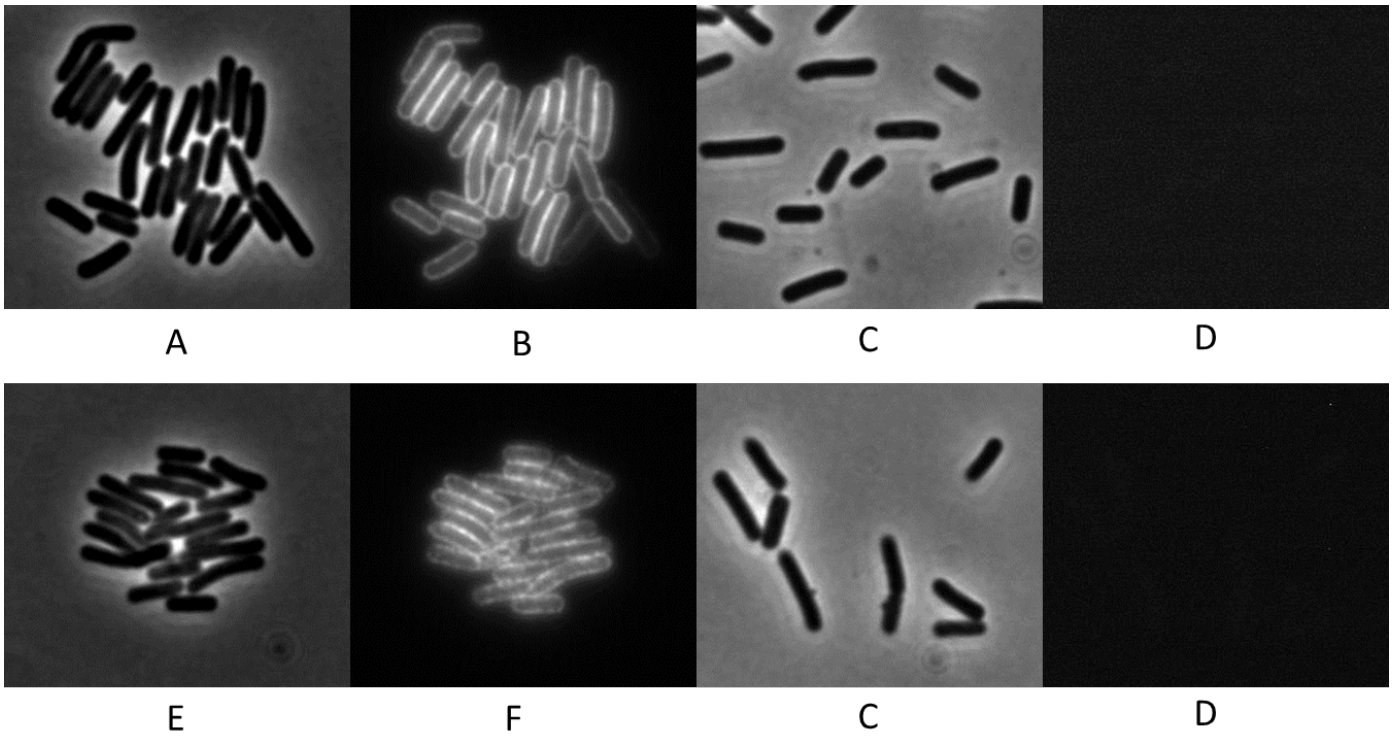


Figure 9-9. Immunofluorescence microscopy images of Ox1437a incubated with Alexa Fluor 594 anti-mouse IgG and Ab521, brightfield image (A), mCherry filter (B). Alexa Fluor 594 anti-mouse IgG and Ab652 brightfield image (E),

mCherry filter (F). Control of Ox1437a with Alexa Fluor 594 anti-mouse IgG and no primary antibody brightfield image (C), mCherry filter (D).

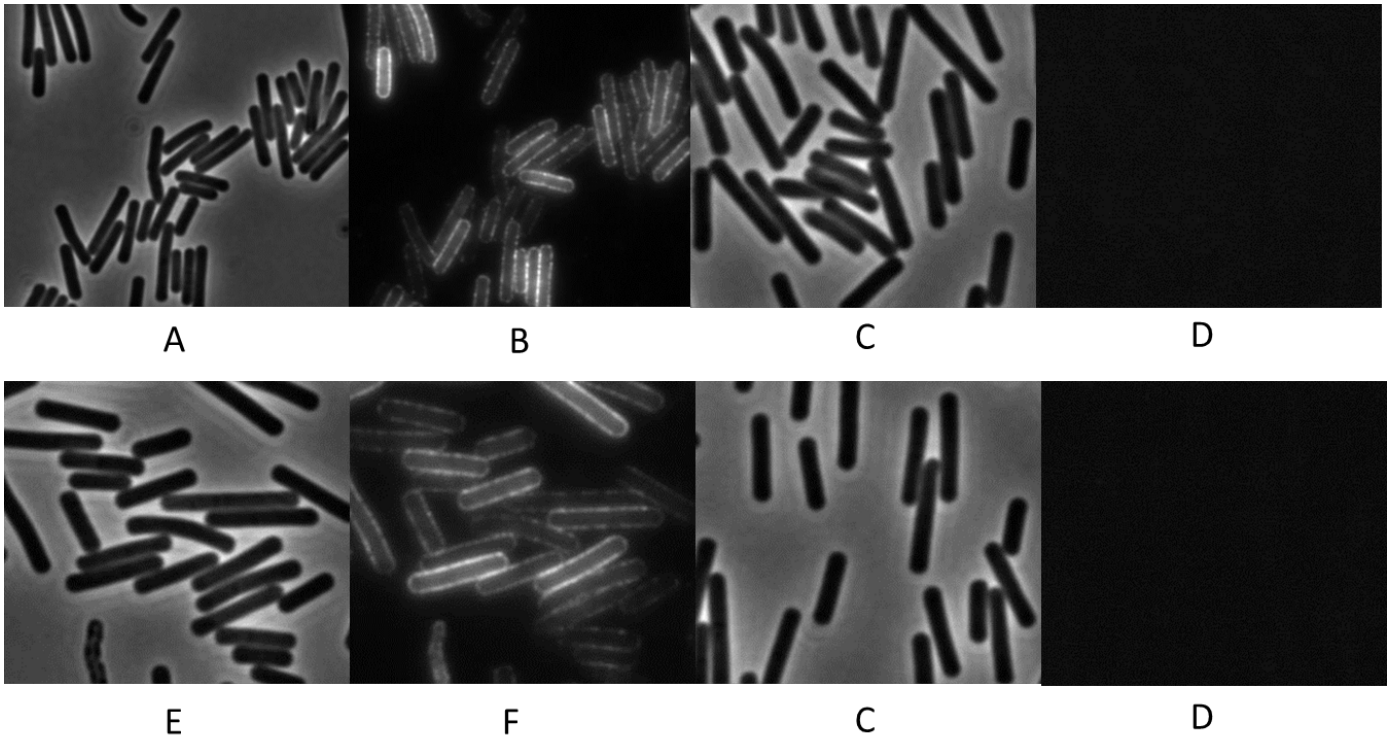


Figure 9-10. Immunofluorescence microscopy images of 630 incubated with Alexa Fluor 594 anti-mouse IgG and Ab521, brightfield image (A), mCherry filter (B). Alexa Fluor 594 anti-mouse IgG and Ab652 brightfield image (E), mCherry filter (F). Control of 630 with Alexa Fluor 594 anti-mouse IgG and no primary antibody brightfield image (C), mCherry filter (D).

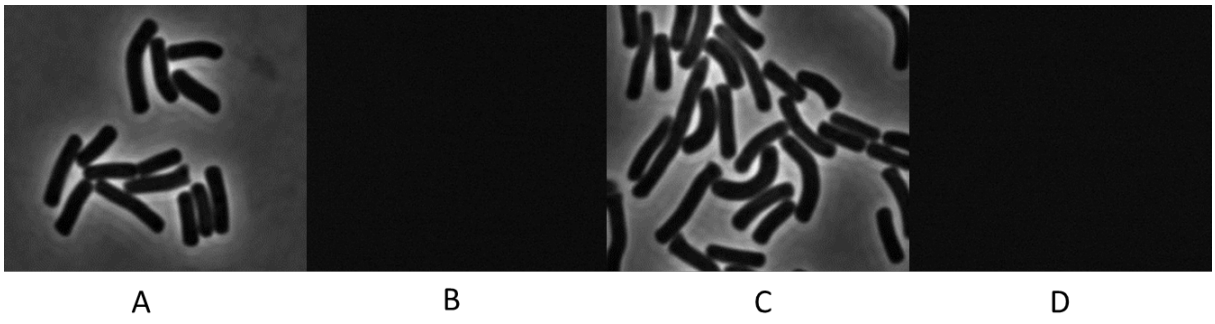


Figure 9-11. Immunofluorescence microscopy images of *C. hiranonis* incubated with Alexa Fluor 594 anti-mouse IgG and Ab521 brightfield (A) mCherry (B) or Ab6521 brightfield (C) mCherry (D).

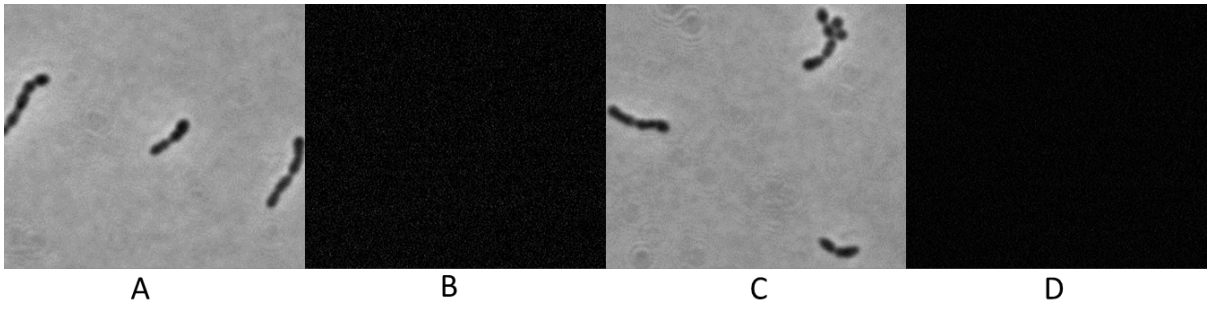


Figure 9-12. Immunofluorescence microscopy images of *P. anaerobius* incubated with Alexa Fluor 594 anti-mouse IgG and Ab521 brightfield (A) mCherry (B) or Ab6521 brightfield (C) mCherry (D).

9.11 Transmission electron microscopy

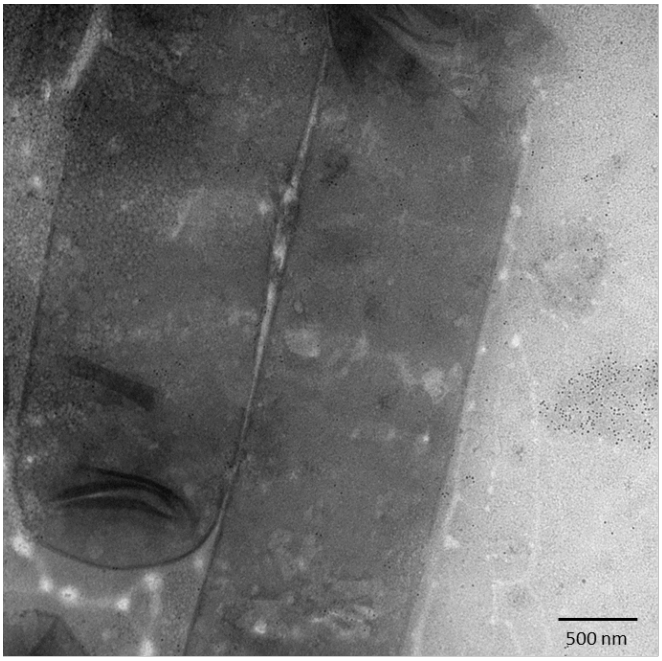


Figure 9-13.Transmission electron microscope image of non-fixed C. difficile 630 incubated with Ab521 and gold conjugated anti-mouse IgG.

Chapter 10 Publications, patents and attended meetings

10.1 Publications

Flanagan, K., Cockell, S., Harwood, C., Hallinan, J., Nakjang, S., Lawry, B. and Wipat, A. (2014) 'A distributed computational search strategy for the identification of diagnostics targets: application to finding aptamer targets for methicillin-resistant staphylococci', *J Integr Bioinform*, 11(2), p. 242.

10.2 Publications in writing

'Determining unique biomarkers that are conserved across the *C. difficile* species using a bioinformatics approach'. Author list and journal to be determined.

10.3 Patents in writing

A patent is currently in writing for the specific *C. difficile* antibody Ab521.

10.4 Attended meetings

ICDS – May 2015 – Poster presentation

i-sense IRC meeting - June 2014 – Oral presentation

i-sense IRC meeting – Jan 2016 – Poster presentation

Chapter 11 References

Abbas, A.K. and Lichtman, A.H. (2004) *Basic Immunology: functions and disorders of the immune system*. 2nd edn. Saunders.

Ahmed, A., Rushworth, J.V., Hirst, N.A. and Millner, P.A. (2014) 'Biosensors for Whole-Cell Bacterial Detection', *Clinical Microbioly Reviews*, 27(3), pp. 631-646.

Akerstrom, B., Brodin, T., Reis, K. and Bjorck, L. (1985) 'Protein G: A powerful tool for binding and detection of monoclonal and polyclonal antibodies', *Journal of Immunology*, 135(4), pp. 2589-2592.

Akhter, S., Aziz, R.K. and Edwards, R.A. (2012) 'PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies', *Nucleic Acids Res*, 40(16), p. e126.

Al-Nassir, W.N., Sethi, A.K., Li, Y., Pultz, M.J., Riggs, M.M. and Donskey, C.J. (2008) 'Both Oral Metronidazole and Oral Vancomycin Promote Persistent Overgrowth of Vancomycin-Resistant Enterococci during Treatment of Clostridium difficile-Associated Disease', *Antimricrobial agents and Chemotherapy*, 52(7), pp. 2403-6.

Alfa, M.J., Lo, E., Wald, A., Dueck, C., DeGagne, P. and Harding, G.K. (2010) 'Improved eradication of Clostridium difficile spores from toilets of hospitalized patients using an accelerated hydrogen peroxide as the cleaning agent. ', *BMC Infect Dis.* , 10, pp. 268-77.

Allison, D.B., Cui, X., Page, G.P. and Sabripour, M. (2006) 'Microarray data analysis: from disarray to consolidation and consensus', *Nat Rev Genet*, 7(1), pp. 55-65.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) 'Basic local alignment search tool', *Journal of molecular biology*, 215(3), pp. 403-410.

Arlett, J., Myers, E. and Roukes, M. (2011) 'Comparative advantages of mechanical biosensors', *Nat Nanotechnol*, 6(4).

Ausiello, C.M., Cerquetti, M., Fedele, G., Spensieri, F., Palazzo, R., Nasso, M., Frezza, S., Mastrantonio, P. (2006) 'Surface layer proteins from Clostridium difficile induce

inflammatory and regulatory cytokines in human monocytes and dendritic cells.', *Microbes and Infection*, 8(11), pp. 2640-6.

Babakhani, F., Bouillaut, L., Gomez, A., Sears, P., Nguyen, L. and Sonenshein, A.L. (2012) 'Fidaxomicin Inhibits Spore Production in Clostridium difficile', *Clin Infect Dis.* , 55(2), pp. S162-9.

Barbut, F., Surgers, L., Eckert, C., Visseaux, B., Cuingnet, M., Mesquita, C., Pradier, N., Thiriez, A., Ait-Ammar, N., Aifaoui, A., Grandsire, E. and Lalande, V. (2014) 'Does a rapid diagnosis of Clostridium difficile infection impact on quality of patient management?', *Clinical Microbiology and Infection*, 20(2), pp. 136-144.

Barra-Carrasco, J., Hernández-Rocha, C., Miranda-Cárdenas, C., Álvarez-Lobos, M., Guzmán Durán, A.M. and Paredes-Sabja, D. (2013) 'Diagnostic accuracy of a multiplex real-time PCR to predict Clostridium difficile ribotype 027', *Anaerobe*, (0).

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Muertter, R.N., Holko, M., Ayanbule, O., Yefanov, A. and Soboleva, A. (2011) 'NCBI GEO: archive for functional genomics data sets--10 years on', *Nucleic Acids Res*, 39(Database issue), pp. D1005-10.

Bartlett, J.G. (2007) 'Clostridium difficile: Old and New Observations', *Journal of Clinical Gastroenterology*, 41, pp. S24-S29.

Bartlett, J.G., Chang, T.W., Gurwith, M., Gorbach, S.L. and Onderdonk, A.B. (1978a) 'Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia', *N Engl J Med*, 298(10), pp. 531-4.

Bartlett, J.G. and Gerding, D.N. (2008) 'Clinical Recognition and Diagnosis of Clostridium difficile Infection ', *Clin Infect Dis.*, 46(1), pp. S12-28.

Bartlett, J.G., Moon, N.T.W., Chang, T.W., Taylor, N. and Onderdonk, A.B. (1978b) 'Role of Clostridium difficile in antibiotic-associated pseudomembranous colitis', *Gastroenterology*, 75, pp. 778-782

Bartsch, S.M., Umscheid, C.A., Fishman, N. and Lee, B.Y. (2013) 'Is Fidaxomicin worth the cost? An economic analysis.', *Clin Infect Dis.*, Epub ahead of print, p. PMID:

23704121.

Beckmann, B.M., Grünweller, A., Weber, M.H.W. and Hartmann, R.K. (2010) 'Northern blot detection of endogenous small RNAs (∼14 nt) in bacterial total RNA extracts', *Nucleic Acids Res*, 38(14), p. e147.

Berglund, L., Andrade, J., Odeberg, J. and Uhlén, M. (2008) 'The epitope space of the human proteome', *Protein Science*, 17(4), pp. 606–613.

Bettelheim, F., Brown, W., Campbell, M., Farrell, S. and Torres, O. (2015) *Introduction to General, Organic and Biochemistry*. 11th Edition edn. Cengage Learning.

Bhalla, A., Aron, D.C. and Donskey, C.J. (2007) 'Staphylococcus aureus intestinal colonization is associated with increased frequency of S. aureus on skin of hospitalized patients', *BMC Infect Dis*, 7, p. 105.

Bidet, P.B., F; Lalande, V; Burghoffer, B and Petit, JC. (1999) 'Development of a new PCR-ribotyping method for Clostridium difficile based on ribosomal RNA gene sequencing', *FEMS Microbiology Letters*, 175, pp. 261-266.

Bobrovnik, S.A. (2003) 'Determination of antibody affinity by ELISA. Theory', *J Biochem Biophys Methods*, 57(3), pp. 213-36.

Boetzkes, A., Felkel, K.W., Zeiser, J., Jochim, N., Just, I. and Pich, A. (2012) 'Secretome analysis of Clostridium difficile strains', *Arch Microbiol*, 194(8), pp. 675-87.

Bolanos, B., Bodon, Q., Jimenez, T., Garcia-Mayol, D., Lavergne, J.A. and Diaz, A.M. (1988) 'Analysis by fluorescence microscopy and flow cytometry of monoclonal antibodies produced against cell surface antigens', *P R Health Sci J*, 7(1), pp. 35-8.

Boone, J.H., Lyerly, D.M. and Carman, R.J. (2013) *Clostridium difficile dehydrogenase and toxin as a biomarker*

Bradshaw, C.S., Morton, A.N., Garland, S.M., Horvath, L.B., Kuzevska, I. and Fairley, C.K. (2005) 'Evaluation of a point-of-care test, BVBlue, and clinical and laboratory criteria for diagnosis of bacterial vaginosis', *J Clin Microbiol*, 43(3), pp. 1304-8.

Bradshaw, W.J., Kirby, J.M., Thiyagarajan, N., Chambers, C.J., Davies, A.H., Roberts, A.K., Shone, C.C. and Acharya, K.R. (2014) 'The structure of the cysteine protease and lectin-like domains of Cwp84, a surface layer-associated protein from Clostridium difficile', *Acta Crystallogr D Biol Crystallogr*, 70(Pt 7), pp. 1983-93.

Brandt, L.J. (2012) 'Fecal Transplantation for the Treatment of Clostridium difficile Infection', *Gastroenterol Hepatol (N Y)*, 8(3), pp. 191-4.

Bravo, A., Cases, M., Queralt-Rosinach, N., Sanz, F. and Furlong, L.I. (2014) 'A knowledge-driven approach to extract disease-related biomarkers from the literature', *Biomed Res Int*, 2014, p. 253128.

Brazier, J.S., Raybould, R., Patel, B., Duckworth, G., Pearson, A., Charlett, A., Duerden, B.I. and Network., H.R.M. (2008) 'Distribution and antimicrobial susceptibility patterns of Clostridium difficile PCR ribotypes in English hospitals, 2007-08.', *Euro Surveillance*, 13(41), p. pii:19000.

Brecher, S.M., Novak-Weekley, S.M. and Nagy, E. (2013) 'Laboratory diagnosis of Clostridium difficile infections: there is light at the end of the colon', *Clin Infect Dis*, 57(8), pp. 1175-81.

Brocchieri, L. and Karlin, S. (2005) 'Protein length in eukaryotic and prokaryotic proteomes', *Nucleic Acids Res*, 33(10), pp. 3390-400.

Bryant, K. and McDonald, L.C. (2009) 'Clostridium difficile Infections in Children', *PEDIATRIC INFECTIOUS DISEASES*, 28, pp. 145-146.

Buchan, B.W. and Ledeboer, N.A. (2014) 'Emerging Technologies for the Clinical Microbiology Laboratory', *Clin Microbiol Review*, 27(4), pp. 783-822.

Buckley, A.M., Spencer, J., Candlish, D., Irvine, J.J. and Douce, G.R. (2011) 'Infection of hamsters with the UK Clostridium', *J Med Microbiol*, 60(Pt 8), pp. 1174-80.

Buffie, C.G., Jarchum, I., Equinda, M., Lipuma, L., Gobourne, A., Viale, A., Ubeda, C., Xavier, J. and Pamer, E.G. (2012) 'Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to Clostridium difficile-induced colitis', *Infect Immun.*, 80(1), pp. 62-73.

Burnham, C.D. and Carroll, K.C. (2013) 'Diagnosis of Clostridium difficile Infection: an Ongoing Conundrum for Clinicians and for Clinical Laboratories', *American society for microbiology*, 26(3), pp. 604-630.

Bustin, S.A. (2002) 'Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems', *J Mol Endocrinol*, 29(1), pp. 23-39.

Byrne, B., Stack, E., Gilmartin, N. and O'Kennedy, R. (2009) 'Antibody-Based Sensors: Principles, Problems and Potential for Detection of Pathogens and Associated Toxins', *Sensors (Basel)*, 9(6), pp. 4407-45.

Calabi, E., Calabi, F., Phillips, A.D. and Fairweather, N.F. (2002) 'Binding of Clostridium difficile Surface Layer Proteins to Gastrointestinal Tissues ', *Infection and Immunology*, 70(10), pp. 5770-5778.

Calabi, E. and Fairweather, N. (2002) 'Patterns of sequence conservation in the S-Layer proteins and related sequences in Clostridium difficile', *J Bacteriol*, 184(14), pp. 3886-97.

Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H.R., Dell, A., Dougan, G. and Fairweather, N. (2001) 'Molecular characterization of the surface layer proteins from Clostridium difficile.', *Molecular Microbiology*, 40(5), pp. 1187-99.

Calderaro, A., Buttrini, M., Martinelli, M., Gorrini, C., Montecchini, S., Medici, M.C., Arcangeletti, M.C., De Conto, F., Covan, S. and Chezzi, C. (2013) 'Comparative analysis of different methods to detect Clostridium difficile infection.', *New Micorbiologica*, 36(1), pp. 57-63.

Cammarota, G., Ianiro, G. and Gasbarrini, A. (2014) 'Fecal microbiota transplantation for the treatment of Clostridium difficile infection: a systematic review.', *Journal of Clinical Gastroenterology*, 48(8), pp. 693-702.

Carroll, K.C. and Bartlett, J.G. (2011) 'Biology of Clostridium difficile: implications for epidemiology and diagnosis', *Annu Rev Microbiol*, 65, pp. 501-21.

Carter, G.P., Douce, G.R., Govind, R., Howarth, P.M., Mackin, K.E., Spencer, J., Buckley, A.M., Antunes, A., Kotsanas, D., Jenkin, G.A., Dupuy, B., Rood, J.I. and Lyras, D. (2011) 'The Anti-Sigma Factor TcdC Modulates Hypervirulence in an Epidemic BI/NAP1/027 Clinical Isolate of Clostridium difficile', *PLoS Pathog*, 7(10), p. e1002317.

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

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G. and Parkhill, J. (2005) 'ACT: the Artemis Comparison Tool', *Bioinformatics*, 21(16), pp. 3422-3.

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

Cerquetti, M., Serafino, A., Sebastianelli, A. and Mastrantonio, P. (2002) 'Binding of Clostridium difficile to Caco-2 epithelial cell line and to extracellular matrix proteins

', *FEMS Immunol. Med. Microbiol*, 32, pp. 211-218.

Chames, P., Van Regenmortel, M., Weiss, E. and Baty, D. (2009) 'Therapeutic antibodies: successes, limitations and hopes for the future', *Br J Pharmacol*, 157(2), pp. 220-33.

Chang, J.Y., Antonopoulos, D.A., Kalra, A., Tonelli, A., Khalife, W.T., Schmidt, T.M. and Young, V.B. (2008) 'Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea.', *Journal of infectious disease*, 197(3), pp. 435-8.

ChapetonMontes, D., Candela, T., Collignon, A. and Janoir, C. (2011) 'Localization of the Clostridium difficile cysteine protease Cwp84 and insights into its maturation process', *J Bacteriol*, 193(19), pp. 5314-21.

Chiodo, F., Marradi, M., Tefsen, B., Snippe, H., van Die, I. and Penades, S. (2013) 'High sensitive detection of carbohydrate binding proteins in an ELISA-solid phase assay based on multivalent glyconanoparticles', *PLoS One*, 8(8), p. e73027.

Choudhuri, S. (2014) *Bioinformatics for Beginners: Genes, Genomes, Molecular Evolution, Databases and Analytical Tools*. Elsiever.

Clabots, C.R., Johnson, S., Olson, M.M., Peterson, L.R. and Gerding, D.N. (1992) 'Acquisition of Clostridium difficile by Hospitalized Patients: Evidence for Colonized New Admissions as a Source of Infection', *The Journal of Infectious Diseases*, 166(3), pp. 561-597.

Clark, L.C. and Lyons, C. (1962) 'ELECTRODE SYSTEMS FOR CONTINUOUS MONITORING IN CARDIOVASCULAR SURGERY', *Annals of the New York academy of sciences*, 102, pp. 29-45.

Clarke, S. and Foster, J.R. (2012) 'A history of blood glucose meters and their role in self-monitoring of diabetes mellitus', *British Journal of Biomedical Science*, 69(2), pp. 83-93.

Cohen, S.H., Gerding, D.N., Johnson, S., Kelly, C.P., Loo, V.G., McDonald, L.C., Pepin, J. and Wilcox, M.H. (2010) 'Clinical practice guidelines for Clostridium difficile infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA).', *Infection Control Hospital Epidemiology*, 31(5), pp. 431-55.

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

Commichau, F.M., Gunka, K., Landmann, J.J. and Stülke, J. (2008) 'Glutamate Metabolism in Bacillus subtilis: Gene Expression and Enzyme Activities Evolved To Avoid Futile Cycles and To Allow Rapid Responses to Perturbations of the System', *J Bacteriol*, 190(10), pp. 3557-64.

Coons, A.H., Creech, H.J. and Jones, R.N. (1941) ' Immunological properties of an antibody containing a fluorescent group', *Experimental Biology and Medicine*, 47(2), pp. 200-202.

Coons, A.H. and Kaplan, M.H. (1950) 'Localization of antigen in tissue cells. II. Improvements in a method for the detection of antigen by means of fluorescent antibody', *The Journal of Experimental Medicine*, 91(1), pp. 1-13.

Couchman, E.C., Browne, H.P., Dunn, M., Lawley, T.D., Songer, J.G., Hall, V., Petrovska, L., Vidor, C., Awad, M., Lyras, D. and Fairweather, N.F. (2015) 'Clostridium sordellii genome analysis reveals plasmid localized toxin genes encoded within pathogenicity loci', *BMC Genomics*, 16, p. 392.

Craig, N.L., Cohen-Fix, O., Green, R., Greider, C.W., Storz, G. and Wolberger, C. (2010) *Molecular Biology. Principles of Genome function*. Oxford: Ocford University Press.

Crobach, M.J., Dekkers, O.M., Wilcox, M.H. and Kuijper, E.J. (2009) 'European Society of Clinical Microbiology and Infectious Diseases (ESCMID): data review and recommendations for diagnosing Clostridium difficile-infection (CDI)', *Clin Microbiol Infect*, 15(12), pp. 1053-66.

Crook, D.W., Walker, A.S., Kean, Y., Weiss, K., Cornely, O.A., Miller, M.A., Esposito, R., Louie, T.J., Stoesser, N.E., Young, B.C., Angus, B.J., Gorbach, S.L. and Peto, T.E. (2012) 'Fidaxomicin versus vancomycin for Clostridium difficile infection: meta-analysis of pivotal randomized controlled trials.', *Clin Infect Dis.*, 55(2), pp. S93-103.

Culbreath, K., Ager, E., Nemeyer, R.J., Kerr, A. and Gilligan, P.H. (2012) 'Evolution of Testing Algorithms at a University Hospital for Detection of Clostridium difficile Infections', *Journal of Clinical Microbiology*, 50(9), pp. 3073-6.

Dalbey, R.E., Wang, P. and van Dijl, J.M. (2012) 'Membrane proteases in the bacterial protein secretion and quality control pathway', *Microbiol Mol Biol Rev*, 76(2), pp. 311-30.

Dang, T.H., de la Riva, L., Fagan, R.P., Storck, E.M., Heal, W.P., Janoir, C., Fairweather, N.F. and Tate, E.W. (2010) 'Chemical probes of surface layer biogenesis in Clostridium difficile', *ACS Chem Biol*, 5(3), pp. 279-85.

Dang, T.H.T., Fagan, R.P., Fairweather, N.F. and Tate, E.W. (2012) 'Novel inhibitors of surface layer processing in Clostridium difficile', *Bioorganic and Medicinal Chemistry* 20, pp. 614-621.

Davies, A.H., Roberts, A.K., Shone, C.C. and Acharya, K.R. (2011) 'Super toxins from a super bug: structure and function of Clostridium difficile toxins.', *The biochemical journal*, 436(3), pp. 517-26.

Davies, K.A., Berry, C.E., Morris, K.A., Smith, R., Young, S., Davis, T.E., Fuller, D.D., Buckner, R.J. and Wilcox, M.H. (2015) 'Comparison of the Vidas C. difficile GDH Automated Enzyme-Linked Fluorescence Immunoassay (ELFA) with Another Commercial Enzyme Immunoassay (EIA) (Quik Chek-60), Two Selective Media, and a PCR Assay for gluD for Detection of Clostridium difficile in Fecal Samples.', *Journal of Clinical Microbiology*, 53(6), pp. 1931-4.

Dawson, L.F., Valiente, E., Donahue, E.H., Birchenough, G. and Wren, B.W. (2011) 'Hypervirulent Clostridium difficile PCR-ribotypes exhibit resistance to widely used disinfectants', *PLoS One.* , 6(10), p. e25754.

De Gruttola, V.G., Clax, P., DeMets, D.L., Downing, G.J., Ellenberg, S.S., Friedman, L., Gail, M.H., Prentice, R., Wittes, J. and Zeger, S.L. (2001) 'Considerations in the evaluation of surrogate endpoints in clinical trials. summary of a National Institutes of Health workshop', *Control Clin Trials*, 22(5), pp. 485-502.

de la Riva, L., Willing, S.E., Tate, E.W. and Fairweather, N.F. (2011) 'Roles of Cysteine Proteases Cwp84 and Cwp13 in Biogenesis of the Cell Wall of Clostridium difficile', *Journal of Bacteriology*, 193(13), pp. 3276-85.

DeJesus, M.A. and Ioerger, T.R. (2013) 'A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data', *BMC Bioinformatics*, 14, p. 303.

Delmee, M., Van Broeck, J., Simon, A., Janssens, M. and Avesani, V. (2005) 'Laboratory diagnosis of Clostridium difficile–associated diarrhoea: a plea for culture.', *Journal of Medical Microbiology*, 54, pp. 187-91.

Demarest, S.J., Hariharan, M., Elia, M., Salbato, J., Jin, P., Bird, C., Short, J.M., Kimmel, B.E., Dudley, M., Woodnutt, G. and Hansen, G. (2010) 'Neutralization of Clostridium difficile toxin A using antibody combinations', *MAbs*, 2(2), pp. 190-8.

Dembek, M., Barquist, L., Boinett, C.J., Cain, A.K., Mayho, M., Lawley, T.D., Fairweather, N.F. and Fagan, R.P. (2015) 'High-throughput analysis of gene essentiality and sporulation in Clostridium difficile', *MBio*, 6(2), p. e02383.

Dethlefsen, L., Huse, S., Sogin, M.L. and Relman, D.A. (2008) 'The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing.', *PLoS Biology*, 6(11), p. e280.

Dingle, K.E., Didelot, X., Ansari, M.A., Eyre, D.W., Vaughan, A., Griffiths, D., Ip, C.L., Batty, E.M., Golubchik, T., Bowden, R., Jolley, K.A., Hood, D.W., Fawley, W.N., Walker, A.S., Peto, T.E., Wilcox, M.H. and Crook, D.W. (2013) 'Recombinational Switching of the Clostridium difficile S-Layer and a Novel Glycosylation Gene Cluster Revealed by Large-Scale Whole-Genome Sequencing ', *Journal of infectious disease*, 207(4), pp. 675-86.

Donald, R.G.K., Flint, M., Kalyan, N., Johnson, E., Witko, S.E., Kotash, C., Zhao, P., Megati, S., Yurgelonis, I., Lee, P.K., Matsuka, Y.V., Severina, E., Deatly, A., Sidhu, M., Jansen, K.U., Minton, N.P. and Anderson, A.S. (2013) 'A novel approach to generate a recombinant toxoid vaccine against Clostridium difficile', *Microbiology*, 159(Pt 7), pp. 1254-66.

Drekonja, D., Reich, J., Gezahegn, S., Greer, N., Shaukat, A., MacDonald, R., Rutks, I. and Wilt, T.J. (2015) 'Fecal Microbiota Transplantation for Clostridium Difficile Infection: A Systematic Review of the Evidence', *Ann Intern Med*, 162(9), pp. 630-8.

Drudy, D., Calabi, E., Kyne, L., Sougioultzis, S., Kelly, E., Fairweather, N. and Kelly, C.P. (2004) 'Human antibody response to surface layer proteins in Clostridium difficile infection', *FEMS Immunol Med Microbiol*, 41(3), pp. 237-42.

Drudy, D., O'Donoghue, D.P., Baird, A., Fenelon, L. and O'Farrelly, C. (2001) 'Flow cytometric analysis of Clostridium difficile adherence to human intestinal epithelial cells.', *Journal of Medical Microbiology*, 50(6), pp. 526-34.

Eastwood, K., Else, P., Charlett, A. and Wilcox, M. (2009) '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*, 47(10), pp. 3211-17.

Ecker, D.J., Sampath, R., Massire, C., Blyn, L.B., Hall, T.A., Eshoo, M.W. and Hofstadler, S.A. (2008) 'Ibis T5000: a universal biosensor approach for microbiology', *Nat Rev Microbiol*, 6(7), pp. 553-8.

Eckert, C., Emirian, A., Le Monnier, A., Cathala, L., De Montclos, H., Goret, J., Berger, P., Petit, A., De Chevigny, A., Jean-Pierre, H., Nebbad, B., Camiade, S., Meckenstock, R., Lalande, V., Marchandin, H. and Barbut, F. (2014) 'Prevalence and pathogenicity of binary toxin-positive Clostridium difficile strains that do not produce toxins A and B.', *New microbes and new infections*, 3, pp. 12-17.

Edgar, R., Domrachev, M. and Lash, A.E. (2002) 'Gene Expression Omnibus: NCBI gene expression and hybridization array data repository', *Nucleic Acids Res*, 30(1), pp. 207-10.

Edgar, R.C. (2004) 'MUSCLE: a multiple sequence alignment method with reduced time and space complexity', *BMC Bioinformatics*, 5, p. 113.

Eglow, R., Pothoulakis, C., Itzkowitz, S., Israel, E.J., O'Keane, C.J., Gong, D., Gao, N., Xu, Y.L., Walker, W.A. and LaMont, J.T. (1992) 'Diminished C. difficile toxin A sensitivity in newborn rabbit ileum is associated with decreased toxin A receptor', *J Clin Invest*, 90(3), pp. 822-829.

Eidhin, D.N., Ryan, A.W., Doyle, R.M., Walsh, J.B. and Kelleher, D. (2006) 'Sequence and phylogenetic analysis of the gene for surface layer protein, slpA, from 14 PCR ribotypes of Clostridium difficile', *J Med Microbiol*, 55(Pt 1), pp. 69-83.

Elgert, K.D. (2009) *Immunology: Understanding The Immune System*. 2nd edn. Wiley-Blackwell.

EMBL-EBI (2015) *European Nucleotide Archive - Statisctics*. Available at: http://www.ebi.ac.uk/ena/about/statistics (Accessed: 14th December).

Emerson, D., Agulto, L., Liu, H. and Liu, L. (2008) 'Identifying and Characterizing Bacteria in an Era of Genomics and Proteomics', *BioScience*, 58(10), pp. 925-936.

Emerson, J. and Fairweather, N. (2009) *Surface structures of C. difficile and other clostridia: implications for pathogenesis and immunity*. UK: Caister Academic Press.

Emerson, J.E., Reynolds, C.B., Fagan, R.P., Shaw, H.A., Goulding, D. and Fairweather, N.F. (2009) 'A novel genetic switch controls phase variable expression of CwpV, a Clostridium difficile cell wall protein', *Mol Microbiol*, 74(3), pp. 541-56.

Engvall, E. and Perlmann, P. (1971) 'Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G.', *Immunochemistry* 8, pp. 871-874.

Enright, A.J. and Ouzounis, C.A. (2001) 'BioLayout--an automatic graph layout algorithm for similarity visualization', *Bioinformatics*, 17(9), pp. 853-4.

Entzian, C. and Schubert, T. (2015) 'Studying small molecule–aptamer interactions using MicroScale Thermophoresis (MST)', *Methods*, p. doi:10.1016/j.ymeth.2015.08.023.

Ethapa, T., Leuzzi, R., Ng, Y.K., Baban, S.T., Adamo, R., Kuehne, S.A., Scarselli, M., Minton, N.P., Serruto, D. and Unnikrishnan, M. (2013) 'Multiple factors modulate biofilm formation by the anaerobic pathogen Clostridium difficile', *J Bacteriol*, 195(3), pp. 545-55.

Evans, C.T. and Safdar, N. (2015) 'Current Trends in the Epidemiology and Outcomes of Clostridium difficile Infection', *Clin Infect Dis.*, 60 (suppl 2), pp. S66-S71.

Eyre, D.W., Walker, A.S., Griffiths, D., Wilcox, M.H., Wyllie, D.H., Dingle, K.E., Crook, D.W. and Peto, T.E. (2012) ' Clostridium difficile mixed infection and reinfection. ', *J Clin Microbiol.*, 50(1), pp. 142-4.

Fagan, R.P., Albesa-Jové, D., Qazi, O., Svergun, D.I., Brown, K.A. and Fairweather, N.F. (2009) 'Structural insights into the molecular organization of the S-layer from Clostridium difficile', *Molecular Microbiology*, 71(5), pp. 1308-22.

Fagan, R.P. and Fairweather, N.F. (2011) 'Clostridium difficile Has Two Parallel and Essential Sec Secretion Systems.', *The Journal of Biological Chemistry.*, 286(31), pp. 27483-27493.

Fagan, R.P., Janoir, C., Collignon, A., Mastrantonio, P., Poxton, I.R. and Fairweather, N.F. (2011) 'A proposed nomenclature for cell wall proteins of Clostridium difficile', *J Med Microbiol*, 60(Pt 8), pp. 1225-8.

Fagerquist, C.K. (2013) 'Top-down proteomic identification of bacterial protein biomarkers and toxins using MALDI-TOF-TOF-MS/MS and post-source decay', *Reviews in Analytical Chemistry*, 32(2), pp. 127–133.

Flanagan, K., Cockell, S., Harwood, C., Hallinan, J., Nakjang, S., Lawry, B. and Wipat, A. (2014) 'A distributed computational search strategy for the identification of diagnostics targets: application to finding aptamer targets for methicillin-resistant staphylococci', *J Integr Bioinform*, 11(2), p. 242.

Foglia, G., Shah, S., Luxemburger, C. and Pietrobon, P.J. (2012) 'Clostridium difficile: Development of a novel candidate vaccine.', *Vaccine*, 30(29), pp. 4307-9.

Fordtran, J.S. (2006) 'Colitis due to Clostridium difficile toxins: underdiagnosed, highly virulent, and nosocomial', *Proc (Bayl Univ Med Cent)*, 19(1), pp. 3-12.

Fraser, T.G. and Swiencicki, J.F. (2013) *Clostridium difficile*. Available at: http://www.clevelandclinicmeded.com/medicalpubs/diseasemanagement/infectious-disease/clostridium-difficile-infection/ (Accessed: 26/5/13).

Freiberg, C., Brunner, N., Macko, L. and Fischer, H.P. (2006) 'Discovering antibiotic efficacy biomarkers: toward mechanism-specific high content compound screening', *Mol Cell Proteomics*, 5(12), pp. 2326-35.

Freifeld, A.G., Simonsen, K.A., Booth, C.S., Zhao, X., Whitney, S.E., Karre, T., Iwen, P.C., Viljoen, H.J. (2012) 'A new rapid method for Clostridium difficile DNA extraction and detection in stool: toward point-of-care diagnostic testing.', *Journal of molecular diagnostics*, 14(3), pp. 274-9.

Friguet, B., Chaffotte, A.F., Djavadi-Ohaniance, L. and Goldberg, M.E. (1985) 'Measurements of the true affinity constant in solution of antigen-antibody complexes by enzyme-linked immunosorbent assay', *J Immunol Methods*, 77(2), pp. 305-19.

Galardini, M., Biondi, E.G., Bazzicalupo, M. and Mengoni, A. (2011) 'CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes', *Source Code Biol Med*, 6, p. 11.

Galdys, A.L., Nelson, J.S., Shutt, K.A., Schlackman, J.L., Pakstis, D.L., Pasculle, A.W., Marsh, J.W., Harrison, L.H. and Curry, S.R. (2014) 'Prevalence and duration of asymptomatic Clostridium difficile carriage among healthy subjects in Pittsburgh, Pennsylvania', *The Journal of Clinical Microbiology*, 52(7), pp. 2406-2409.

Gan, S.D. and Patel, K.R. (2013) 'Enzyme Immunoassay and Enzyme-Linked Immunosorbent Assay', *J Invest Dermatol*, 133(9), p. e12.

Gardy, J.L. and Brinkman, F.S. (2006) 'Methods for predicting bacterial protein subcellular localization', *Nat Rev Microbiol*, 4(10), pp. 741-51.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. (2003) 'ExPASy: the proteomics server for in-depth protein knowledge and analysis', *Nucleic Acids Research*, 31, pp. 3784-3788.

George, W.L., Sutter, V.L., Citron, D. and Finegold, S.M. (1979) 'Selective and differential medium for isolation of Clostridium difficile', *J Clin Microbiol*, 9(2), pp. 214-9.

Gerding, D.N., Johnson, S., Peterson, L.R., Mulligan, M.E. and Silva, J. (1995) 'Clostridium difficile-Associated Diarrhea and Colitis', *Infection Control and Hospital Epidemiology*, 16(8), pp. 459-477.

Gerding, D.N., Johnson, S., Peterson, L.R., Mulligan, M.E. and Silva, J. (2012) 'Clostridium difficile Infection Prevention: Biotherapeutics, Immunologics, and Vaccines', *Discovery Medicine*, 13(68), pp. 75-83.

Gerding, D.N., Muto, C.A. and Owens Jr, R.C. (2008) 'Measures to control and prevent Clostridium difficile infection', *Clin Infect Dis.*, 46, pp. S43-S49.

Gerver, S., Mihalkova, M., Abernethy, J., Bou-Antoun, S., Nsonwu, O., Kausar, S., Wasti, S., Apraku, D., Davies, J. and Hope, R. (2015) *Annual Epidemiological Commentary: Mandatory MRSA, MSSA and E. coli bacteraemia and C. difficile infection data, 2014/15*. PHE publications gateway number 2015167.

Geuijen, C.A., Clijsters-van der Horst, M., Cox, F., Rood, P.M., Throsby, M., Jongeneelen, M.A., Backus, H.H., van Deventer, E., Kruisbeek, A.M., Goudsmit, J. and de Kruif, J. (2005)

'Affinity ranking of antibodies using flow cytometry: application in antibody phage display-based target discovery', *J Immunol Methods*, 302(1-2), pp. 68-77.

Ghose, C. (2013) 'Clostridium difficile infection in the twenty-first century', *Emerging Microbes and Infections*, 2(62), p. doi: 10.1038/emi.2013.62. .

Giljohann, D.A. and Mirkin, C.A. (2009) 'Drivers of biodiagnostic development', *Nature*, 462(7272), pp. 461-4.

Gimenez, D.F. and Ciccarelli, A.S. (1970) 'Another type of Clostridium botulinum', *Zentralbl Bakteriol Orig*, 215(2), pp. 221-4.

Girinathan, B.P., Braun, S.E. and Govind, R. (2014) 'Clostridium difficile glutamate dehydrogenase is a secreted enzyme that confers resistance to H(2)O(2)', *Microbiology*, 160(Pt 1), pp. 47-55.

Godfrey, W.L., Hill, D.M., Kilgore, J.A., Buller, G.M., Bradford, J.A., Gray, D.R., Clements, I., Oakleaf, K., Salisbury, J.J., Ignatius, M.J. and Janes, M.S. (2005) 'Complementarity of Flow Cytometry and Fluorescence Microscopy', *Microscopy and Microanalysis*, 11, pp. 246-247.

Goldenberg, S.D., Bisnauthsing, K.N., Patel, A., Postulka, A., Wyncoll, D., Schiff, R. and French, G.L. (2014) 'Point-of-Care Testing for Clostridium Difficile Infection: A Real-World Feasibility Study of a Rapid Molecular Test in Two Hospital Settings', *Infectious diseases and therapy*, 3(2), pp. 295-306.

Goldenberg, S.D., Cliff, P.R. and French, G.L. (2010) 'Glutamate Dehydrogenase for Laboratory Diagnosis of Clostridium difficile Infection', *J Clin Microbiol*, 48(8), pp. 3050-1.

Goldenberg, S.D. and French, G.L. (2011) 'Lack of association of tcdC type and binary toxin status with disease severity and outcome in toxigenic Clostridium difficile.', *The Journal of Infection*, 62(5), pp. 355-62.

Goldenberg, S.D., Volpe, H. and French, G.L. (2012) 'Clinical negligence, litigation and healthcare-associated infections', *J Hosp Infect*.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P. and Tiedje, J.M. (2007) 'DNA-DNA hybridization values and their relationship to whole-genome sequence similarities', *Int J Syst Evol Microbiol*, 57(Pt 1), pp. 81-91.

Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001) 'Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure', *J Mol Biol*, 313(4), pp. 903-19.

Graille, M., Stura, E.A., Corper, A.L., Sutton, B.J., Taussig, M.J., Charbonnier, J.B. and Silverman, G.J. (2000) 'Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity', *Proc Natl Acad Sci U S A*, 97(10), pp. 5399-404.

Grein, J.D., Ochner, M., Hoang, H., Jin, A., Morgan, M.A. and Murthy, A.R. (2014) 'Comparison of testing approaches for Clostridium difficile infection at a large community hospital', *Clinical Microbiol and Infection*, 20(1), pp. 65-9.

Grieshaber, D., MacKenzie, R., Vörös, J. and Reimhult, E. (2008) 'Electrochemical Biosensors - Sensor Principles and Architectures', *Sensors (Basel)*, 8(3), pp. 1400-58.

Group., B.D.W. (2001) 'Biomarkers and surrogate endpoints: preferred definitions and conceptual framework', *Clin Pharmacol Ther*, 69(3), pp. 89-95.

Guarner, F. (2006) ' Enteric Flora in Health and Disease', *Digestion*, 73, pp. 5-12.

Guerrero, D.M., Nerandzic, M.M., Jury, L.A., Chang, S., Jump, R.L. and Donskey, C.J. (2011) 'Clostridium difficile infection in a Department of Veterans Affairs long-term care facility', *Infect Control Hosp Epidemiol*, 32, pp. 513-515.

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

Hall, I.C. and O'Toole, E. (1935) 'INTESTINAL FLORA IN NEW-BORN INFANTSWITH A DESCRIPTION OF A NEW PATHOGENIC ANAEROBE, BACILLUS DIFFICILIS', *American Journal of Diseases of Children*, 49(2), pp. 390-402.

Hammond, E.N., Donkor, E.S. and Brown, C.A. (2014) 'Biofilm formation of Clostridium difficile and susceptibility to Manuka honey', *BMC Complement Altern Med*, 14, p. 329.

Hancock, D.C. and O'Reilly, N.J. (2005) 'Synthetic peptides as antigens for antibody production.', *Methods in Molecular Biology*, 295, pp. 13-26.

Hansen, G., Blatt, S., Brecher, S.M., Dubberke, E. and Dorsett, P. (2010) 'Clostridium difficile. Navigating the Testing Options for Diagnosis ', *Clinical Laboratory News: Clostridium difficile* 36(7), pp. 10-13.

Hansen, L.B., Buus, S. and Schafer-Nielsen, C. (2013) 'Identification and Mapping of Linear Antibody Epitopes in Human Serum Albumin Using High-Density Peptide Arrays', *PLoS One*, 8(7), p. e68902.

Harper-Owen, R., Dymock, D., Booth, V., Weightman, A.J. and Wade, W.G. (1999) 'Detection of unculturable bacteria in periodontal health and disease by PCR', *J Clin Microbiol*, 37(5), pp. 1469-73.

Harper, C.J., Hayward, D., Kidd, M., Wiid, I. and van Helden, P. (2010) 'Glutamate dehydrogenase and glutamine synthetase are regulated in response to nitrogen availability in Myocbacterium smegmatis', *BMC Microbiol*, 10, p. 138.

He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D.J., Martin, M.J., Connor, T.R., Harris, S.R., Fairley, D., Bamford, K.B., D'Arc, S., Brazier, J., Brown, D., Coia, J.E., Douce, G., Gerding, D., Kim, H.J., Koh, T.H., Kato, H., Senoh, M., Louie, T., Michell, S., Butt, E., Peacock, S.J., Brown, N.M., Riley, T., Songer, G., Wilcox, M., Pirmohamed, M., Kuijper, E., Hawkey, P., Wren, B.W., Dougan, G., Parkhill, J. and Lawley, T.D. (2013) 'Emergence and global spread of epidemic healthcare-associated Clostridium difficile', *Nat Genet*, 45(1), pp. 109-13.

Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P. and Minton, N.P. (2007) 'The ClosTron: a universal gene knock-out system for the genus Clostridium', *J Microbiol Methods*, 70(3), pp. 452-64.

Heeg, D., Burnd, D.A., Cartman, S.T. and Minton, N.P. (2012) 'Spores of Clostridium difficile Clinical Isolates Display a Diverse Germintaion Response to Bile Salts', *PLoS One.*, 7(2), p. e32381.

Hernández-Rocha, C., Barra-Carrasco, J., Álvarez-Lobos, M., Paredes-Sabja, D. and Guzmán-Durán, A.M. (2013) 'Prospective comparison of a commercial multiplex real-time polymerase chain reaction and an enzyme immunoassay with toxigenic culture in the diagnosis of Clostridium difficile–associated infections', *Diagnostic Microbiology and Infectious Disease*, 75(4), pp. 361-365.

Higdon, R. and Kolker, E. (2007) 'A predictive model for identifying proteins by a single peptide match', *Bioinformatics*, 23(3), pp. 277-80.

Holub, M., Lawrence, D.A., Andersen, N., Davidova, A., Beran, O., Maresova, V. and Chalupa, P. (2013) 'Cytokines and chemokines as biomarkers of community-acquired bacterial infection', *Mediators Inflamm*, 2013, p. 190145.

Hong, H.A., Khaneja, R., Tam, N.M., Cazzato, A., Tan, S., Urdaci, M., Brisson, A., Gasbarrini, A., Barnes, I. and Cutting, S.M. (2009) 'Bacillus subtilis isolated from the human gastrointestinal tract', *Res Microbiol*, 160(2), pp. 134-43.

Hoppert, M. and Holzenburg, A. (1998) *Electron microscopy in microbiology*. UK: BIOS Scientific Publishers.

Howe, J., Hammer, M.U. and Brandenburg, K. (2007) 'Calorimetric investigations of the effect of polymyxin B on different Gram-negative bacteria', *Thermochimica Acta*, 458(1–2), pp. 34-37.

Howerton, A., Patra, M. and Abel-Santos, E. (2013) 'A new strategy for the prevention of Clostridium difficile infection', *J Infect Dis*, 207(10), pp. 1498-504.

Hudson, M. (2010) *Deaths involving Clostridium difficile: England and Wales, 2009* (Accessed: 27th August. Available at: http://www.ons.gov.uk/ons/rel/subnational-health2/deaths-involving-clostridium-difficile/2009/index.html).

Hussack, G., Arbabi-Ghahroudi, M., van Faassen, H., Songer, J.G., Ng, K.K., MacKenzie, R. and Tanha, J. (2011) 'Neutralization of Clostridium difficile toxin A with single-domain antibodies targeting the cell receptor binding domain', *J Biol Chem*, 286(11), pp. 8961-76.

Ibarra, J.A., Perez-Rueda, E., Carroll, R.K. and Shaw, L.N. (2013) 'Global analysis of transcriptional regulators in Staphylococcus aureus', *BMC Genomics*, 14, p. 126.

Issarachaikull, R., Khantipong, M., Sawatpanich, A. and Suankratay, C. (2015) 'PROSPECTIVE EVALUATION OF A NOVEL TWO-STEP PROTOCOL FOR SCREENING OF CLOSTRIDIUM DIFFICILE INFECTION IN HOSPITALIZED ADULT PATIENTS', *Southeast Asian J Trop Med Public Health*, 46(6), pp. 1037-48.

Jaffrezic-Renault, N., Martelet, C., Chevolot, Y. and Cloarec, J.P. (2007) 'Biosensors and Bio-Bar Code Assays Based on Biofunctionalized Magnetic Microbeads', *Sensors (Basel)*, 7(4), pp. 589-614.

Jakupciak, J.P. and Colwell, R.R. (2009) 'Biological agent detection technologies', *Mol Ecol Resour*, 9 Suppl s1, pp. 51-7.

Janda, J.M. and Abbott, S.L. (2002) 'Bacterial Identification for Publication: When Is Enough Enough?', *J Clin Microbiol*, 40(6), pp. 1887-91.

Janeway, C., A, Jr. , Travers, P., Walport, M. and Shlomchik, M. (2001) *Immunobiology: The Immune System in Health and Disease*. New York: Garland Science.

Janezic, S. and Rupnik, M. (2010) 'Molecular typing methods for Clostridium difficile: pulsed-field gel electrophoresis and PCR ribotyping.', *Methods of Molecular Biology*, 646, pp. 55-65.

Jarvinen, A.K., Laakso, S., Piiparinen, P., Aittakorpi, A., Lindfors, M., Huopaniemi, L., Piiparinen, H. and Maki, M. (2009) 'Rapid identification of bacterial pathogens using a PCR- and microarray-based assay', *BMC Microbiol*, 9, p. 161.

Jemal, A., Thomas, A., Murray, T. and Thun, M. (2002) 'Cancer statistics, 2002', *CA Cancer J Clin*, 52(1), pp. 23-47.

Jerabek-Willemsen, M., Wienken, C.J., Braun, D., Baaske, P. and S., D. (2011) 'Molecular Interaction Studies Using Microscale Thermophoresis', *Assay and Drug Development Technologies*, 9(4), pp. 342-53.

Jerabek-Willemsena, M., Andréa, T., Wannera, R., Rotha, H.M., Duhra, S., Baaskea, P. and Breitsprecher, D. (2014) 'MicroScale Thermophoresis: Interaction analysis and beyond', *Journal of Molecular Structure*, 1077, pp. 101–113.

Jin, S.G., Wu, X., Li, A.X. and Pfeifer, G.P. (2011) 'Genomic mapping of 5-hydroxymethylcytosine in the human brain', *Nucleic Acids Res*, 39(12), pp. 5015-24.

Johnson, E.A. (2005) *Clostridial Neurotoxins*. Taylor and Francis Group.

Johnson, M. (2013) *Mouse Antibody* (Accessed: 07 October).

Johnson, S., Kent, S.A., O'Leary, K.J., Merrigan, M.M., Sambol, S.P., Peterson, L.R. and Gerding, D.N. (2001) 'Fatal pseudomembranous colitis associated with a variant clostridium difficile strain not detected by toxin A immunoassay', *Ann Intern Med*, 135(6), pp. 434-8.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R. and Hunter, S. (2014) 'InterProScan 5: genome-scale protein function classification', *Bioinformatics*, 30(9), pp. 1236-40.

Jump, R.L., Pultz, M.J. and Donskey, C.J. (2007) '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?', *Antimicrob Agents Chemother*, 51(8), pp. 2883-7.

Juncker, A.S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. and Krogh, A. (2003) 'Prediction of lipoprotein signal peptides in Gram-negative bacteria', *Protein Sci*, 12(8), pp. 1652-62.

Kachrimanidou, M. and Malisiovas, N. (2011) 'Clostridium difficile infection: a comprehensive review.', *Critical reviews in Microbiology*, 37(3), pp. 178-87.

Kall, L., Krogh, A. and Sonnhammer, E.L. (2004) 'A combined transmembrane topology and signal peptide prediction method', *J Mol Biol*, 338(5), pp. 1027-36.

Kandalaft, H., Hussack, G., Aubry, A., van Faassen, H., Guan, Y., Arbabi-Ghahroudi, M., MacKenzie, R., Logan, S.M. and Tanha, J. (2015) 'Targeting surface-layer proteins with single-domain antibodies: a potential therapeutic approach against Clostridium difficile-associated disease', *Applied Microbiology and Biotechnology*, 99(20), pp. 8549-62. doi.

Karas, J.A., Enoch, D.A. and Aliyu, S.H. (2010) 'A review of mortality due to Clostridium difficile infection', *Journal of Infection*, 61(1), pp. 1-8.

Karczewski, J., Zorman, J., Wang, S., Miezeiewski, M., Xie, J., Soring, K., Petrescu, I., Rogers, I., Thiriot, D.S., Cook, J.C., Chamberlin, M., Xoconostle, R.F., Nahas, D.D., Joyce, J.G., Bodmer, J.L., Heinrichs, J.H. and Secore, S. (2014) 'Development of a recombinant toxin fragment vaccine for Clostridium difficile infection', *Vaccine*, 32(24), pp. 2812-8.

Karjalainen, T., Saumier, N., Barc, M.C., Delmée, M. and Collignon, A. (2002) 'Clostridium difficile genotyping based on slpA variable region in S-layer gene sequence: an alternative to serotyping.', *J Clin Microbiol.*, 40(2452-2458).

Kaur, R., Dikshit, K.L. and Raje, M. (2002) 'Optimization of immunogold labeling TEM: an ELISA-based method for evaluation of blocking agents for quantitative detection of antigen', *J Histochem Cytochem*, 50(6), pp. 863-73.

Kelly, C.P., Pothoulakis, C. and LaMont, J.T. (1994) 'Clostridium difficile Colitis', *N Engl J Med*, 330, pp. 257-262.

Kemeny, D.M. and Challacombe, S.J. (1988) *ELISA and Other Solid Phase Immunoassays: Theoretical and Practical Aspects*. John Wiley and Sons.

Khalil, I.F., Alifrangis, M., Recke, C., Hoegberg, L.C., Ronn, A., Bygbjerg, I.C. and Koch, C. (2011) 'Development of ELISA-based methods to measure the anti-malarial drug chloroquine in plasma and in pharmaceutical formulations', *Malar J*, 10, p. 249.

Kim, N.O., Jung, S.M., Na, H.Y., Chung, G.T., Yoo, C.K., Seong, W.K. and Hong, S. (2015) 'Enteric Bacteria Isolated from Diarrheal Patients in Korea in 2014', *Osong Public Health Res Perspect*, 6(4), pp. 233-40.

King, A.M., Mackin, K.E. and Lyras, D. (2015) 'Emergence of toxin A-negative, toxin B-positive Clostridium difficile strains: epidemiological and clinical considerations.', *Future Microbiology*, 10(1), pp. 1-4.

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

Koo, H.L., Van, J.N., Zhao, M., Ye, X., Revell, P.A., Jiang, Z.D., Grimes, C.Z., Koo, D.C., Lasco, T., Kozinetz, C.A., Garey, K.W. and DuPont, H.L. (2014) 'Real-time polymerase chain reaction detection of asymptomatic Clostridium difficile colonization and rising C. difficile-associated disease rates.', *Infect Control Hosp Epidemiol*, 35(6), pp. 667-73.

Kotiranta, A., Haapasalo, M., Lounatmaa, K. and Kari, K. (1995) 'Crystalline surface protein of Peptostreptococcus anaerobius', *Microbiology*, 141 (Pt 5), pp. 1065-73.

Kovacs-Simon, A., Leuzzi, R., Kasendra, M., Minton, N., Titball, R.W. and Michell, S.L. (2014) 'Lipoprotein CD0873 is a novel adhesin of Clostridium difficile', *J Infect Dis*, 210(2), pp. 274-84.

Kovacs-Simon, A., Titball, R.W. and Michell, S.L. (2011) 'Lipoproteins of bacterial pathogens', *Infect Immun*, 79(2), pp. 548-61.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. (2001) 'Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes', *J Mol Biol*, 305(3), pp. 567-80.

Kronvall, G., Grey, H.M. and Williams, R.C.J. (1970) 'Protein A reactivity with mouse immunoglobulins. Structural relationship between some mouse and human immunoglobulins.', *Journal of Immunology*, 105(5), pp. 1116-23.

Kuehne, S.A., Cartman, S.T., Heap, J.T., Kelly, M.L., Cockayne, A. and Minton, N.P. (2010) 'The role of toxin A and toxin B in Clostridium difficile infection.', *Nature*, 467(7316), pp. 711-3.

Kuehne, S.A., Collery, M.M., Kelly, M.L., Cartman, S.T., Cockayne, A. and Minton, N.P. (2014) 'Importance of Toxin A, Toxin B, and CDT in Virulence of an Epidemic Clostridium difficile Strain', *The Journal of Infectious Diseases*, 209(1), pp. 83-6.

Kufelnicka, A.M. and Kirn, T.J. (2011) 'Effective Utilization of Evolving Methods for the Laboratory Diagnosis of Clostridium difficile Infection', *Clin Infect Dis.*, 52(12), pp. 1451-7.

Kurka, H., Ehrenreich, A., Ludwig, W., Monot, M., Rupnik, M., Barbut, F., Indra, A., Dupuy, B. and Liebl, W. (2014) 'Sequence Similarity of Clostridium difficile Strains by Analysis of Conserved Genes and Genome Content Is Reflected by Their Ribotype Affiliation', *PLoS One*, 9(1), p. e86535.

Kvach, E.J., Ferguson, D., Riska, P.F. and Landry, M.L. (2010) 'Comparison of BD GeneOhm Cdiff real-time PCR assay with a two-step algorithm and a toxin A/B enzyme-linked immunosorbent assay for diagnosis of toxigenic Clostridium difficile infection.', *Journal of Clinical Microbiology*, 48(1), pp. 109-14.

Laguna, M., Holgado, M., Hernandez, A.L., Santamaria, B., Lavin, A., Soria, J., Suarez, T., Bardina, C., Jara, M., Sanza, F.J. and Casquel, R. (2015) 'Antigen-Antibody Affinity for Dry Eye Biomarkers by Label Free Biosensing. Comparison with the ELISA Technique', *Sensors (Basel)*, 15(8), pp. 19819-29.

Lalande, V., Barrault, L., Wadel, S., Eckert, C., Petit, J.C. and Barbut, F. (2011) 'Evaluation of a loop-mediated isothermal amplification assay for diagnosis of Clostridium difficile infections', *J Clin Microbiol*, 49(7), pp. 2714-6.

Land, M., Hauser, L., Jun, S.R., Nookaew, I., Leuze, M.R., Ahn, T.H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S. and Ussery, D.W. (2015) 'Insights from 20 years of bacterial genome sequencing', *Funct Integr Genomics*, 15(2), pp. 141-61.

Lanis, J.M., Barua, S. and Ballard, J.D. (2010) 'Variations in TcdB activity and the hypervirulence of emerging strains of Clostridium difficile', *PLoS Pathog*, 6(8), p. e1001061.

Lanzas, C. and Dubberke, E.R. (2014) 'Effectiveness of Screening Hospital Admissions to Detect Asymptomatic Carriers of Clostridium difficile: A Modeling Evaluation', *Infection Control and Hospital Epidemiology*, 35(8), pp. 1043-1050.

Larson, H.E., Pricea, A.B., Honoura, P. and Borriellob, S.P. (1978) 'CLOSTRIDIUM DIFFICILE AND THE ÆTIOLOGY OF PSEUDOMEMBRANOUS COLITIS', *The Lancet*, 1(8073), pp. 1063–1066.

Lawley, T.D., Clare, S., Walker, A.W., Stares, M.D., Connor, T.R., Raisen, C., Goulding, D., Rad, R., Schreiber, F., Brandt, C., Deakin, L.J., Pickard, D.J., Duncan, S.H., Flint, H.J., Clark, T.G., Parkhill, J. and Dougan, G. (2012) 'Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice.', *PLoS Pathog*, 8(10), p. 1002995. doi:10.1371/journal.ppat.1002995.

Leadley, D. (2010) *Transmission Electron Microscopy (TEM)*. Available at: https://www2.warwick.ac.uk/fac/sci/physics/current/postgraduate/regs/mpags/ex5/techniques/structural/tem/ (Accessed: 13/10/2015).

Leekha, S., Aronhalt, K.C., Sloan, L.M., Patel, R. and Orenstein, R. (2013) 'Asymptomatic Clostridium difficile colonization in a tertiary care hospital: Admission prevalence and risk factors', *American Journal of Infection Control*, 41(5), pp. 390-393.

Leenaars, M. and Hendriksen, C.F.M. (2005) 'Critical Steps in the Production of Polyclonal and Monoclonal Antibodies: Evaluation and Recommendations', *ILAR Journal*, 46(3), pp. 269-279.

Leirdal, M., Shadidy, M., Rosok, O. and Sioud, M. (2004) 'Identification of genes differentially expressed in breast cancer cell line SKBR3: potential identification of new prognostic biomarkers', *Int J Mol Med*, 14(2), pp. 217-22.

Li, C., Vandenber, K., Prabhulkara, S., Zhua, X., Schneperb, L., Metheeb, K., Rosserd, C.J. and Almeidec, E. (2011) 'Paper based point-of-care testing disc for multiplex whole cell bacteria analysis', *Elsevier*, 26, pp. 4342-4348.

Li, J., Zhang, Z., Rosenzweig, J., Wang, Y.Y. and Chan, D.W. (2002) 'Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer', *Clin Chem*, 48(8), pp. 1296-304.

Liang, Y. (2008) 'Applications of isothermal titration calorimetry in protein science', *Acta Biochimica et Biophysica Sinica*, 40(7), pp. 565-576.

Lieberman, M., Marks, A.D., Smith, C.M. and Marks, D.B. (2007) *Marks' Essential Medical Biochemistry. A Clinical Approach*. Lippincott Williams & Wilkins.

Liu, H., White, J., Crawford, F., Jin, N., Ju, X., Liu, K., Jiang, C., Marrack, P., Zhang, G. and Kappler, J.W. (2015a) 'A Rapid Method to Characterize Mouse IgG Antibodies and Isolate Native Antigen Binding IgG B Cell Hybridomas.', *PLoS One*, 10(8), p. e0136613.

Liu, J. (2014) 'The history of monoclonal antibody development – Progress, remaining challenges and future innovations', *Annals of Medicine and Surgery*, 3, pp. 113-116.

Liu, J., Mazumdar, D. and Lu, Y. (2006) 'A simple and sensitive "dipstick" test in serum based on lateral flow separation of aptamer-linked nanostructures', *Angew Chem Int Ed Engl*, 45(47), pp. 7955-9.

Liu, Y., Liu, N., Ma, X., Li, X., Ma, J., Li, Y., Zhou, Z. and Gao, Z. (2015b) 'Highly specific detection of thrombin using an aptamer-based suspension array and the interaction analysis via microscale thermophoresis.', *The Analyst*, 140(8), pp. 2762-70.

Llarrull, L.I., Fisher, J.F. and Mobashery, S. (2009) 'Molecular basis and phenotype of methicillin resistance in Staphylococcus aureus and insights into new beta-lactams that meet the challenge', *Antimicrob Agents Chemother*, 53(10), pp. 4051-63.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D. and Darnell, J. (2000) *Molecular Cell Biology. 4th edition.* New York: W.H. Freeman.

Loo, V.G., Poirier, L., Miller, M.A., Oughton, M., Libman, M.D., Michaud, S., Bourgault, A.M., Nguyen, T., Frenette, C., Kelly, M., Vibien, A., Brassard, P., Fenn, S., Dewar, K., Hudson, T.J., Horn, R., Rene, P., Monczak, Y. and Dascal, A. (2005) 'A predominantly clonal multi-institutional outbreak of Clostridium difficile-associated diarrhea with high morbidity and mortality', *N Engl J Med*, 353(23), pp. 2442-9.

Louie, T.J., Miller, M.A., Mullane, K.M., Weiss, K., Lentnek, A., Golan, Y., Gorbach, S., Sears, P. and Shue, Y. (2011) 'Fidaxomicin versus Vancomycin for Clostridium difficile Infection', *The New England Journal of Medicine*, 364(422-431).

Lowy, I., Molrine, D.C., Leav, B.A., Blair, B.M., Baxter, R., Gerding, D.N., Nichol, G., Thomas, W.D.J., Leney, M., Sloan, S., Hay, C.A. and Ambrosino, D.M. (2010) 'Treatment with monoclonal antibodies against Clostridium difficile toxins.', *362*, 3, pp. 197-205.

Lyerly, D.M., Ball, D.W., Toth, J. and Wilkins, T.D. (1988) 'Characterization of cross-reactive proteins detected by Culturette Brand Rapid Latex Test for Clostridium difficile', *J Clin Microbiol*, 26(3), pp. 397-400.

Lyras, D., O'Connor, J.R., Howarth, P.M., Sambol, S.P., Carter, G.P., Phumoonna, T., Poon, R., Adams, V., Vedantam, G., Johnson, S., Gerding, D.N. and Rood, J.I. (2009) 'Toxin B is essential for virulence of Clostridium difficile.', *Nature*, 458(7242), pp. 1176-9.

Mak, T.W. and Saunders, M.E. (2001) *The Immune Response: Basic and Clinical Principles*. Academic Press.

Martin, A.C.R. (2010) *'Protein Sequence and Structure Analysis of Antibody Variable Domains'*. Berlin: Springer.

McDonald, L.C., Killgore, G.E., Thompson, A., Owens, R.C., Kazakova, S.V., Sambol, S.P., Johnson, S. and Gerding, D.N. (2005) 'An epidemic, toxin gene-variant strain of Clostridium difficile.', *N Engl J Med*, 353(23), pp. 2433-41.

McMaster-Baxter, N.L. and Musher, D.M. (2007) 'Clostridium difficile: recent epidemiologic findings and advances in therapy', *Pharmacotherapy* 27, pp. 1029-1039.

Merrigan, M., Venugopal, A., Mallozzi, M., Roxas, B., Viswanathan, V.K., Johnson, S., Gerding, D.N. and Vedantam, G. (2010) 'Human hypervirulent Clostridium difficile strains exhibit increased sporulation as well as robust toxin production.', *Journal of Bacteriology*, 192(19), pp. 4904-11.

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

Mettler, M., Grimm, F., Capelli, G., Camp, H. and Deplazes, P. (2005) 'Evaluation of Enzyme-Linked Immunosorbent Assays, an Immunofluorescent-Antibody Test, and Two Rapid Tests (Immunochromatographic-Dipstick and Gel Tests) for Serological Diagnosis of Symptomatic and Asymptomatic Leishmania Infections in Dogs', *Journal of Microbiology*, 43(11), pp. 5515-5519.

Michaelsen, T.E., Kolberg, J., Aase, A., Herstad, T.K. and Høiby, E.A. (2004) 'The four mouse IgG isotypes differ extensively in bactericidal and opsonophagocytic activity when reacting with the P1.16 epitope on the outer membrane PorA protein of Neisseria meningitidis.', *Scandinavian Journal of Immunology*, 59(1), pp. 34-9.

Miller, M.A., Hyland, M., Ofner-Agostini, M., Gourdeau, M. and Ishak, M. (2002) 'Morbidity, Mortality, and Healthcare Burden of Nosocomial Clostridium difficile–Associated Diarrhea in Canadian Hospitals', *Infection Control and Hospital Epidemiology*, 23(3).

Miniño, A.M., Xu, J.Q. and Kochanek, K.D. (2010) 'Deaths: preliminary data for 2008', *National Vital Statistics Report*, 59(2), p. http://www.cdc.gov/nchs/data/nvsr/nvsr59/nvsr59_02.pdf. Accessed 27 August 2015.

Mishra, A.K., Lagier, J.C., Robert, C., Raoult, D. and Fournier, P.E. (2012) 'Non-contiguous finished genome sequence and description of Clostridium senegalense sp. nov', *Stand Genomic Sci*, 6(3), pp. 386-95.

Mitchell, A., Chang, H.Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S.Y., Bateman, A., Punta, M., Attwood, T.K., Sigrist, C.J., Redaschi, N., Rivoire, C.,

Xenarios, I., Kahn, D., Guyot, D., Bork, P. and Letunic, I. (2015) 'The InterPro protein families database: the classification resource after 15 years', 43(Database issue), pp. D213-21.

Mitterer, G., Huber, M., Leidinger, E., Kirisits, C., Lubitz, W., Mueller, M.W. and Schmidt, W.M. (2004) 'Microarray-based identification of bacteria in clinical samples by solid-phase PCR amplification of 23S ribosomal DNA sequences', *J Clin Microbiol*, 42(3), pp. 1048-57.

Mohan, A. and Harikrishna, J. (2015) 'Biomarkers for the diagnosis of bacterial infections: in pursuit of the 'Holy Grail'', *Indian J Med Res*, 141(3), pp. 271-3.

Monaghan, T.M., Robins, A., Knox, A., Sewell, H.F. and Mahida, Y.R. (2013) 'Circulating antibody and memory B-Cell responses to C. difficile toxins A and B in patients with C. difficile-associated diarrhoea, inflammatory bowel disease and cystic fibrosis', *PLoS One*, 8(9), p. e74452.

Natesan, M. and Ulrich, R.G. (2010) 'Protein Microarrays and Biomarkers of Infectious Disease', *Int J Mol Sci*, 11(12), pp. 5165-83.

National Statistics, O.f. (2013) *Deaths Involving Clostridium difficile, England and Wales, 2012*. http://www.ons.gov.uk/ons/rel/subnational-health2/deaths-involving-clostridium-difficile/2012/stb-deaths-involving-clostridium-difficile-2012.html: Crown.

Negri, P., Chen, G., Kage, A., Nitsche, A., Naumann, D., Xu, B. and Dluhy, R.A. (2012) 'Direct Optical Detection of Viral Nucleoprotein Binding to an Anti-Influenza Aptamer', *Anal Chem*, 84(13), pp. 5501-8.

Newman, J.D. and Setford, S.J. (2006) 'Enzyme Biosensors', *Molecular Biotechnology*, 32(3), pp. 249-68.

Nielsen, H., and Krogh, A. (1998) 'Prediction of signal peptides and signal anchors by a hidden Markov model.', *Proc Int Conf Intell Syst Mol Biol*, 6, pp. 122-130.

Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) 'Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites', *Protein Engineering*, 10(1), pp. 1-6.

Nilsson, H.O., Taneera, J., Castedal, M., Glatz, E., Olsson, R. and Wadstrom, T. (2000) 'Identification of Helicobacter pylori and other Helicobacter species by PCR, hybridization, and partial DNA sequencing in human liver samples from patients with primary sclerosing cholangitis or primary biliary cirrhosis', *J Clin Microbiol*, 38(3), pp. 1072-6.

O'Brien, J.A., Lahue, B.J., Caro, J.J. and Davidson, D.M. (2007) 'The emerging infectious challenge of clostridium difficile-associated disease in Massachusetts hospitals: clinical and economic consequences', *Infect Control Hosp Epidemiol*, 28(11), pp. 1219-27.

O'Brien, J.B., McCabe, M.S., Athié-Morales, V., McDonald, G.S., Ní Eidhin, D.B. and Kelleher, D.P. (2005) 'Passive immunisation of hamsters against Clostridium difficile infection using antibodies to surface layer proteins ', *FEMS Microbiology Letters*, 246(2), pp. 119-205.

O'Connor, J.R., Johnson, S. and Gerding, D.N. (2009) 'Clostridium difficile Infection Caused by the Epidemic BI/NAP1/027 Strain', *Gastroenterology*, 136(6), pp. 1913-1924.

O'Connor, J.R., Lyras, D., Farrow, K.A., Adams, V., Powell, D.R., Hinds, J., Cheung, J.K. and Rood, J.I. (2006) 'Construction and analysis of chromosomal Clostridium difficile mutants', *Mol Microbiol*, 61(5), pp. 1335-51.

O'Donoghue, C. and Kyne, L. (2011) 'Update on Clostridium difficile infection', *Current Opinion Gastroenterology*, 27, pp. 38-47.

Odell, I.D. and Cook, D. (2013) 'Immunofluorescence Techniques', *J Invest Dermatol*, 133(1), p. e4.

Orosz, F. and Ovadi, J. (2002) 'A simple method for the determination of dissociation constants by displacement ELISA', *J Immunol Methods*, 270(2), pp. 155-62.

Owen, R. (2003) *Techniques for Detection of Specific Organisms and Potential for DNA-Based Technologies*. Interparm/CRC.

Ozdemir, M.S., Marczak, M., Bohets, H., Bonroy, K., Roymans, D., Stuyver, L., Vanhoutte, K., Pawlak, M. and Bakker, E. (2013) 'A label-free potentiometric sensor principle for the detection of antibody-antigen interactions.', *Analytical Chemistry*, 85(9), pp. 4770-6.

Pandey, S. (2010) 'HYBRIDOMA TECHNOLOGY FOR PRODUCTION OF MONOCLONAL ANTIBODIES', *International Journal of Pharmaceutical Sciences Review and Research* 1(2), pp. 88-94.

Pantaléon, V., Soavelomandroso, A.P., Bouttier, S., Briandet, R., Roxas, B., Chu, M., Collignon, A., Janoir, C., Vedantam, G. and Candela, T. (2015) 'The Clostridium difficile Protease Cwp84 Modulates both Biofilm Formation and Cell-Surface Properties', *PLoS One*, 10(4), p. e0124971.

Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G. and Thomas, R. (2008) 'Understanding and using sensitivity, specificity and predictive values', *Indian J Ophthalmol*, 56(1), pp. 45-50.

Parks, B.W., Nam, E., Org, E., Kostem, E., Norheim, F., Hui, S.T., Pan, C., Civelek, M., Rau, C.D., Bennett, B.J., Mehrabian, M., Ursell, L.K., He, A., Castellani, L.W., Zinker, B., Kirby, M., Drake, T.A., Drevon, C.A., Knight, R., Gargalovic, P., Kirchgessner, T., Eskin, E. and Lusis, A.J. (2013) 'Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice', *Cell Metab*, 17(1), pp. 141-52.

Payne, S.H., Bonissone, S., Wu, S., Brown, R.N., Ivankov, D.N., Frishman, D., Pasa-Tolic, L., Smith, R.D. and Pevzner, P.A. (2012) 'Unexpected diversity of signal peptides in prokaryotes', *MBio*, 3(6).

Péchiné, S., Janoir, C. and Collignon, A. (2005) 'Variability of Clostridium difficile surface proteins and specific serum antibody response in patients with Clostridium difficile-associated disease.', *Journal of Clinical Microbiology*, 43(10), pp. 5018-25.

Peck, M.W. (2014) *Clostridium botulinum*. Elsevier publications.

Pedruzzi, I., Rivoire, C., Auchincloss, A.H., Coudert, E., Keller, G., de Castro, E., Baratin, D., Cuche, B.A., Bougueleret, L., Poux, S., Redaschi, N., Xenarios, I. and Bridge, A. (2015) 'HAMAP in 2015: updates to the protein family classification and annotation system', *Nucleic Acids Res*, 43(Database issue), pp. D1064-70.

Peng, Z., Soper, S.A., Pingle, M.R., Barany, F. and Davis, L.M. (2010) 'Ligase Detection Reaction Generation of Reverse Molecular Beacons for Near Real-Time Analysis of Bacterial

Pathogens Using Single-Pair Fluorescence Resonance Energy Transfer and a Cyclic Olefin Copolymer Microfluidic Chip', *Analytical Chemistry*, 82(23), pp. 9727-35.

Pépin, J., Valiquette, L., Alary, M.E., Villemure, P., Pelletier, A., Forget, K., Pépin, K. and Chouinard, D. (2004) 'Clostridium difficile-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity.', *Canadian Medical Association Journal*, 171(5), pp. 466-72.

Peters, A. (2008) 'The Diagnosis of Pregnancy', *Global Library of Womens Medicine*, pp. (ISSN: 1756-2228).

Petersen, T.N., Brunak, S. von Heijne, H. (2011) 'SignalP 4.0: discriminating signal peptides from transmembrane regions', *Nature Methods*, 8, pp. 785-786.

Peterson, L.R., Manson, R.U., Paule, S.M., Hacek, D.M., Robicsek, A., Thomson, R.B. and Jr, K., K.L. (2007) 'Detection of toxigenic Clostridium difficile in stool samples by real-time polymerase chain reaction for the diagnosis of C. difficile-associated diarrhea.', *Clin Infect Dis.*, 54(9), pp. 1152-60.

Peterson, L.R., Mehta, M.S., Patel, P.A., Hacek, D.M., Harazin, M., Nagwekar, P.P., Thomson, R.B., Jr. and Robicsek, A. (2011) 'Laboratory testing for Clostridium difficile infection: light at the end of the tunnel', *Am J Clin Pathol*, 136(3), pp. 372-80.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) 'UCSF Chimera--a visualization system for exploratory research and analysis', *J Comput Chem*, 25(13), pp. 1605-12.

Piepenbrink, K.H., Maldarelli, G.A., de la Pena, C.F., Mulvey, G.L., Snyder, G.A., De Masi, L., von Rosenvinge, E.C., Gunther, S., Armstrong, G.D., Donnenberg, M.S. and Sundberg, E.J. (2014) 'Structure of Clostridium difficile PilJ exhibits unprecedented divergence from known type IV pilins', *J Biol Chem*, 289(7), pp. 4334-45.

Planche, T.D., Davies, K.A., Coen, P.G., Finney, J.M., Monahan, I.M., Morris, K.A., O'Connor, L., Oakley, S.J., Pope, C.F., Wren, M.W., Shetty, N.P., Crook, D.W. and Wilcox, M.H. (2013) 'Differences in outcome according to Clostridium difficile testing method: a prospective

multicentre diagnostic validation study of C difficile infection.', *Lancet Infectious Diseases*, 13, pp. 936-945.

Poojary, N.S., Ramlal, S., Urs, R.M., Sripathy, M.H. and Batra, H.V. (2014) 'Application of monoclonal antibodies generated against Panton-Valentine Leukocidin (PVL-S) toxin for specific identification of community acquired methicillin resistance Staphylococcus aureus', *Microbiol Res*, 169(12), pp. 924-30.

Posthuma-Trumpie, G.A., Korf, J. and van Amerongen, A. (2009) 'Lateral flow (immuno)assay: its strengths, weaknesses, opportunities and threats. A literature survey', *Anal Bioanal Chem*, 393(2), pp. 569-82.

Poutanen, S.M. and Simor, A.E. (2004) 'Clostridium difficile-associated diarrhea in adults ', *CMAJ*, 171(1), pp. 51-58.

Prassas, I., Chrystoja, C.C., Makawita, S. and Diamandis, E.P. (2012) 'Bioinformatic identification of proteins with tissue-specific expression for biomarker discovery', *BMC Med*, 10, p. 39.

Prevention, T.C.f.D.C.a. (2008) 'Surveillance for community-associated Clostridium difficile – Connecticut, 2006.', *Morbidity and Mortality Weekly Report*, 57, pp. 340-343.

Pruitt, K., Brown, G., Tatusova, T. and Maglott, D. (2002) *The NCBI Handbook*.

Pum, D., Schuster, B., Sara, M. and Sleytr, U.B. (2004) 'Functionalisation of surfaces with S-layers', *IEE Proc Nanobiotechnol*, 151(3), pp. 83-6.

Qazi, O., Hitchen, P., Tissot, B., Panico, M., Morris, H.R., Dell, A. and Fairweather, N. (2009) 'Mass spectrometric analysis of the S-layer proteins from Clostridium difficile demonstrates the absence of glycosylation.', *Journal of mass spectrometry*, 44(3), pp. 368-74.

Rabilloud, T. and Triboulet, S. (2013) 'Two-dimensional SDS-PAGE fractionation of biological samples for biomarker discovery', *Methods Mol Biol*, 1002, pp. 151-65.

Rahman, O., Cummings, S.P., Harrington, D.J., Sutcilffe, I.C. (2008) 'Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of gram-positive bacteria. ', *World Journal of Microbiology and Biotechnology* 24, pp. 2377-2382.

Ramirez-Rosales, A. and Cantu-Llanos, E. (2012) '[Intrahospital mortality in patients with Clostridium difficile-associated diarrhea infection.]', *Rev Gastroenterol Mex*.

Ramnani, P., Saucedob, N.M. and Mulchandania, A. (2015) 'Carbon nanomaterial-based electrochemical biosensors for label-free sensing of environmental pollutants', *Chemosphere*, doi:10.1016/j.chemosphere.2015.04.063.

Rea, M.C., O'Sullivan, O., Shanahan, F., O'Toole, P.W., Stanton, C., Ross, P.H. and Colin Hill, C. (2012) 'Clostridium difficile Carriage in Elderly Subjects and Associated Changes in the Intestinal Microbiota', *J. Clin. Microbiol.*, 50(3), pp. 867-875.

Reller, M.E., Alcabasa, R.C., Lema, C.A. and Carroll, K.C. (2010) 'Comparison of two rapid assays for Clostridium difficile Common antigen and a C difficile toxin A/B assay with the cell culture neutralization assay', *Am J Clin Pathol*, 133(1), pp. 107-9.

Remes-Troche, J.M. (2012) '[Clostridium difficile-associated diarrhea infection: Is it time for us to start worrying in Mexico?]', *Rev Gastroenterol Mex*.

Restrepo-Montoya, D., Vizcaino, C., Nino, L.F., Ocampo, M., Patarroyo, M.E. and Patarroyo, M.A. (2009) 'Validating subcellular localization prediction tools with mycobacterial proteins', *BMC Bioinformatics*, 10, p. 134.

Reynolds, C.B., Emerson, J.E., de la Riva, L., Fagan, R.P. and Fairweather, N.F. (2011) 'The Clostridium difficile Cell Wall Protein CwpV is Antigenically Variable between Strains, but Exhibits Conserved Aggregation-Promoting Function', *PLoS Pathog*, 7(4), p. e1002024.

Riggs, M.M., Sethi, A.K., Zabarsky, T.F., Eckstein, E.C., Jump, R.L. and Donskey, C.J. (2007) 'Asymptomatic carriers are a potential source for transmission of epidemic and nonepidemic Clostridium difficile strains among long-term care facility residents.', *Clinical infectious diseases*, 45(8), pp. 992-8.

Robinson, J.P., Sturgis, J. and Lumar, G.L. (2009) *Immunofluorescence*. 5th edn.

Robotham, J. and Wilcox, M. (2012) *Update Guidance on the Diagnosis and Report of Clostridium Difficile*. https://www.gov.uk/government/publications/updated-guidance-on-the-diagnosis-and-reporting-of-clostridium-difficile: Gov.UK.

Rodriguez-R, L.M. and Konstantinidis, K.T. (2014) 'Bypassing Cultivation To

Identify Bacterial Species', *Microbe*, 9(3), pp. 111-118.

Rodriguez, C., Taminiau, B., Van Broeck, J., Delmee, M. and Daube, G. (2015) 'Clostridium difficile infection and intestinal microbiota interactions', *Microb Pathog*, 89, pp. 201-209.

Rohlke, F. and Stollman, N. (2012) 'Fecal microbiota transplantation in relapsing Clostridium difficile infection', *Therap Adv Gastroenterol*, 5(6), pp. 403-20.

Rupnik, M., Wilcox, M.H. and Gerding, D.N. (2009) 'Clostridium difficile infection: new developments in epidemiology and pathogenesis', *Nat Rev Micro*, 7(7), pp. 526-536.

Saade, E., Deshpande, A., Kundrapu, S., Sunkesula, V.C., Guerrero, D.M., Jury, L.A. and Donskey, C.J. (2013) 'Appropriateness of empiric therapy in patients with suspected Clostridium difficile infection.', *Current Medical Research and Opinion* 29(8), pp. 985-8.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) 'Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase', *Science*, 239(4839), pp. 487-91.

Sajid, M., Kawde, A.N. and Daud, M. (2015) 'Designs, formats and applications of lateral flow assay: A literature review', *Journal of Saudi Chemical Society*, 19(6), pp. 689-705.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) 'DNA sequencing with chain-terminating inhibitors. ', *Proceedings of the National Academy of Sciences of the U.S.A*, 74(12), pp. 5463-7.

Saper, C.B. (2009) 'A Guide to the Perplexed on the Specificity of Antibodies', *Journal of Histochemistry and Cytochemistry* 57(1), pp. 1-5.

Sara, M. and Sleytr, U.B. (2000) 'S-Layer proteins', *J Bacteriol*, 182(4), pp. 859-68.

Saujet, L., Monot, M., Dupuy, B., Soutourina, O. and Martin-Verstraete, I. (2011) 'The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and virulence factor expression in Clostridium difficile', *J Bacteriol*, 193(13), pp. 3186-96.

Savariau-Lacomme, M.-P., et al., (2003) 'Transcription and Analysis of Polymorphism in a Cluster of Genes Encoding Surface-Associated Proteins of Clostridium difficile', *Journal of Bacteriology*, 185(15), p. 4461.

Schwan, C., Kruppke, A.S., Nölke, T., Schumacher, L., Koch-Nolte, F., Kudryashev, M., Stahlberg, H. and Aktories, K. (2014) 'Clostridium difficile toxin CDT hijacks microtubule organization and reroutes vesicle traffic to increase pathogen adherence.', *Proceedings of the National Academy of Sciences of the U.S.A*, 111(6), pp. 2313-8.

Schwan, C., Stecher, B., Tzivelekidis, T., van Ham, M., Rohde, M., Hardt, W.D., Wehland, J. and Aktories, K. (2009) 'Clostridium difficile toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria.', *PLoS Pathogen*, 5(10).

Schwarz, J.A., Contescu, C.I. and Putyera, K. (2004) *Dekker Encyclopedia of Nanoscience and Nanotechnology*. New York: Marcel Dekker, Inc.

Sebaihia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeño-Tárraga, A.M., Wang, H., Holden, M.T., Wright, A., Churcher, C., Quail, M.A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B. and Parkhill, J. (2006) 'The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome.', *Nature Genetics*, 38(7), pp. 779-86.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S. and Huttenhower, C. (2011) 'Metagenomic biomarker discovery and explanation', *Genome Biol*, 12(6), p. R60.

Seidel, S.A., Dijkman, P.M., Lea, W.A., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J.S., Srinivasan, P., Baaske, P., Simeonov, A., Katritch, I., Melo, F.A., Ladbury, J.E., Schreiber, G., Watts, A., Braun, D. and Duhr, S. (2013) 'Microscale thermophoresis quantifies

biomolecular interactions under previously challenging conditions.', *National Insituties of Health*, 59(3), pp. 301-315.

Sekirov, I., Russell, S.L., Caetano, L., Antunes, M. and Finlay, B.B. (2010) 'Gut Microbiota in Health and Disease', *American physiological society*, 90(3), pp. 859-904.

Serdyuk, I.N., Zaccai, N.R. and Zaccai, J. (2007) *Methods in Molecular biophysics. Structure, Dynamics, Function.* Cambridge University Press.

Servin, A.L. (2004) 'Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens.', *FEMS Microbiology reviews*, 28(4), pp. 405-40.

Sewell, B., Rees, E., Thomas, I., Ch'ng, C.L., Isaac, M. and Berry, N. (2014) 'Impact on patient length of stay and costeffectiveness of rapid molecular testing for Clostridium difficile', *Infectious diseases and therapy*, 3(2), pp. 281-293.

Shafique, S. (2012) *Polymerase Chain Reaction*. Lap Lambert Academic Publishing.

Shanholtzer, C.J., Willard, K.E., Holter, J.J., Olson, M.M., Gerding, D.N. and Peterson, L.R. (1992) 'Comparison of the Vidas Clostridium difficile toxin A immunoassay with C. difficile culture and cytotoxin and latex tests.', *Journal of Clinical Microbiology*, 30, pp. 1837-40.

Shao, Y., He, X., Harrison, E.M., Tai, C., Ou, H.Y., Rajakumar, K. and Deng, Z. (2010) 'mGenomeSubtractor: a web-based tool for parallel in silico subtractive hybridization analysis of multiple bacterial genomes', *Nucleic Acids Res*, 38(Web Server issue), pp. W194-200.

Sharma, S., Zapatero-Rodríguez, J., Estrela, P. and O'Kennedy, R. (2015) 'Point-of-Care Diagnostics in Low Resource Settings: Present Status and Future Role of Microfluidics.', *Biosensors*, 5(3), pp. 577-601.

Shearer, A., Boehmer, M., Closs, M., Dela Rosa, R., Hamilton, J., Horton, K., McGrath, R. and Schulman, C. (2009) 'Comparison of glucose point-of-care values with laboratory values in critically ill patients', *Am J Crit Care*, 18(3), pp. 224-30.

Shetty, N., Wren, M.W. and Coen, P.G. (2011) 'The role of glutamate dehydrogenase for the detection of Clostridium difficile in faecal samples: a meta-analysis', *J Hosp Infect*, 77(1), pp. 1-6.

Sigrist, C.J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A. and Bucher, P. (2002) 'PROSITE: a documented database using patterns and profiles as motif descriptors', *Brief Bioinform*, 3(3), pp. 265-74.

Simon, R. and Roychowdhury, S. (2013) 'Implementing personalized cancer genomics in clinical trials', *Nat Rev Drug Discov*, 12(5), pp. 358-69.

Simor, A.E., Bradley, S.F., Strausbaugh, L.J., Crossley, K. and Nicolle, L.E. (2013) 'Clostridium difficile in Long-Term–Care Facilities

for the Elderly', *INFECTION CONTROL AND HOSPITAL EPIDEMIOLOGY*, 23(11), pp. 696-703.

Sjöbring, U., Björck, L. and Kastern, W. (1991) 'Streptococcal protein G. Gene structure and protein binding properties.', *The Journal of Biological Chemistry.*, 266(1), pp. 399-405.

Sleytr, U.B. and Messner, P. (1983) 'Crystalline surface layers on bacteria', *Annu Rev Microbiol*, 37, pp. 311-39.

Snape, A., Papachristodoulou, D., Elliott, W.H. and Elliott, D.C. (2014) *Biochemistry and Molecular Biology. 5th Edition*. Oxford: Oxford University Press.

Song, L., Zhao, M., Duffy, D.C., Hansen, J., Shields, K., Wungjiranirun, M., Chen, X., Xu, H., Leffler, D.A., Sambol, S.P., Gerding, D.N., Kelly, C.P. and Pollock, N.R. (2015) 'Development and validation of digital ELISAs for ultrasensitive detection and quantification of C. difficile toxins in stool.', *Journal of Clinical Microbiology*, pp. JCM.01334-15. [Epub ahead of print].

Song, Y.L., Liu, C.X., McTeague, M., Summanen, P. and Finegold, S.M. (2004) 'Clostridium bartlettii sp. nov., isolated from human faeces', *Anaerobe*, 10(3), pp. 179-84.

Song, Z., Chen, L. and Xu, D. (2010) 'Bioinformatics methods for protein identification using Peptide mass fingerprinting', *Methods Mol Biol*, 604, pp. 7-22.

Soreide, K. (2009) 'Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research', *J Clin Pathol*, 62(1), pp. 1-5.

Spigaglia, P., Barketi-Klai, A., Collignon, A., Mastrantonio, P., Barbanti, F., Rupnik, M., Janezic, S. and Kansau, I. (2013) 'SURFACE-LAYER (S-LAYER) OF HUMAN AND ANIMAL CLOSTRIDIUM DIFFICILE STRAINS AND THEIR BEHAVIOUR IN ADHERENCE TO EPITHELIAL CELLS AND INTESTINAL COLONIZATION.', *Journal of Medical Microbiology*, p. PMID 23518658.

Stallmach, A., Hagel, S. and Lohse, A.W. (2015) '[Diagnostic workup and therapy of infectious diarrhea : Current standards]', *Internist (Berl)*.

Stojanovic, I., Schasfoort, R.B. and Terstappen, L.W. (2014) 'Analysis of cell surface antigens by Surface Plasmon Resonance imaging', *Biosens Bioelectron*, 52, pp. 36-43.

Storch, W.B. (2000) *Immunofluorescence in Clinical Immunology: A Primer and Atlas*. Springer Basel AG.

Suarez, S., Ferroni, A., Lotz, A., Jolley, K.A., Guerin, P., Leto, J., Dauphin, B., Jamet, A., Maiden, M.C., Nassif, X. and Armengaud, J. (2013) 'Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory', *J Microbiol Methods*, 94(3), pp. 390-6.

Sugiura, T., Imagawa, H. and Kondo, T. (2000) 'Purification of horse immunoglobulin isotypes based on differential elution properties of isotypes from protein A and protein G columns', *Journal of Chromatography B: Biomedical Sciences and Applications*, 742(2), pp. 327-334.

Sundberg, E.J., Anderson, P.S., Gorshkova, I.I. and Schuck, P. (2007) *Surface Plasmon Resonance Biosensing in the Study of Ternary Systems of Interacting Proteins*. Maryland: Springer.

Szafranski, S.P., Wos-Oxley, M.L., Vilchez-Vargas, R., Jauregui, R., Plumeier, I., Klawonn, F., Tomasch, J., Meisinger, C., Kuhnisch, J., Sztajer, H., Pieper, D.H. and Wagner-Dobler, I. (2015) 'High-resolution taxonomic profiling of the subgingival microbiome for biomarker discovery and periodontitis diagnosis', *Appl Environ Microbiol*, 81(3), pp. 1047-58.

Taban, B.M., Aytac, S.A., Akkoc, N. and Akcelik, M. (2013) 'Characterization of antibiotic resistance in Salmonella enterica isolates determined from ready-to-eat (RTE) salad vegetables.', *Brazilian Journal of Microbiology*, 44, pp. 385-391.

Takeoka, A., Takumi, K., Koga, T. and Kawata, T. (1991) 'Purification and characterization of S layer proteins from Clostridium difficile GAI 0714', *J Gen Microbiol*, 137(2), pp. 261-7.

Tam Dang, T.H., Fagan, R.P., Fairweather, N.F. and Tate, E.W. (2012) 'Novel inhibitors of surface layer processing in Clostridium difficile', *Bioorganic & Medicinal Chemistry*, 20(2), pp. 614-621.

Tang, S. and Hewlett, I. (2010) 'Nanoparticle-based immunoassays for sensitive and early detection of HIV-1 capsid (p24) antigen', *J Infect Dis*, 201 Suppl 1, pp. S59-64.

Tang, Y.W., Ellis, N.M., Hopkins, M.K., Smith, D.H., Dodge, D.E. and Persing, D.H. (1998) 'Comparison of Phenotypic and Genotypic Techniques for Identification of Unusual Aerobic Pathogenic Gram-Negative Bacilli', *Journal of Clinical Microbiology*, 36(12), pp. 3674-9.

Taslim, H. (2009) 'Clostridium difficile infection in the elderly.', *Acta Medica Indonesia*, 41(3), pp. 148-51.

Tattevin, P., Buffet-Bataillon, S., Donnio, P., Revest, M. and Michelet, C. (2013) 'Clostridium difficile infections: do we know the real dimensions of the problem?', *International Journal of Antimicrobial Agents*, (0).

Taylor, L.H., Latham, S.M. and Woolhouse, M.E. (2001) 'Risk factors for human disease emergence', *Philos Trans R Soc Lond B Biol Sci*, 356(1411), pp. 983-9.

Tenover, F.C., Baron, E.J., Peterson, L.R. and Persing, D.H. (2011) 'Laboratory diagnosis of Clostridium difficile infection can molecular amplification methods move us out of uncertainty?', *Journal of molecualr diagnostics*, 13(6), pp. 573-82.

Tenover, F.C., Novak-Weekley, S., Woods, C.W., Peterson, L.R., Davis, T., Schreckenberger, P., Fang, F.C., Dascal, A., Gerding, D.N., Nomura, J.H., Goering, R.V., Akerlund, T., Weissfeld, A.S., Baron, E.J., Wong, E., Marlowe, E.M., Whitmore, J. and Persing, D.H. (2010) 'Impact of

strain type on detection of toxigenic Clostridium difficile: comparison of molecular diagnostic and enzyme immunoassay approaches', *J Clin Microbiol*, 48(10), pp. 3719-24.

Thermo, S. (2010) 'ELISA technical guide and protocols ' *Tech Tip #65*. Thermo Fisher Scientific Inc. . Available at: https://tools.thermofisher.com/content/sfs/brochures/TR0065-ELISA-guide.pdf.

Tiwari, A.K. and Srivastava, R. (2014) 'A survey of computational intelligence techniques in protein function prediction', *Int J Proteomics*, 2014, p. 845479.

Tonyushkina, K. and Nichols, J.H. (2009) 'Glucose Meters: A Review of Technical Challenges to Obtaining Accurate Results', *J Diabetes Sci Technol*, 3(4), pp. 971-980.

Trayhurn, P. (1996) 'Northern blotting', *Proc Nutr Soc*, 55(1b), pp. 583-9.

Treister, A. and Roederer, M. (2015) *FAQ FlowJo.com*. Available at: http://www.flowjo.com/v6/html/faq.html#2.4.4 (Accessed: 13/10/15).

Troy, D.B. and Beringer, P. (2006) *Remington: The Science and Practice of Pharmacy*. 21st edn. Lippincott Williams & Wilkins.

Twyman, R.M. (2014) *Principles of Proteomics*. Second edn. Garland Science.

Ubeda, C., Taur, Y., Jenq, R.R., Equinda, M.J., Son, T., Samstein, M., Viale, A., Socci, N.D., van den Brink, M.R., Kamboj, M. and Pamer, E.G. (2010) 'Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans ', *Journal of Clinical Investigation*, 120(12), pp. 4332-41.

Uniprot-Consortium. (2012) 'Reorganizing the protein space at the Universal Protein Resource (UniProt)', *Nucleic Acids Res*, 40(Database issue), pp. D71-5.

Vaishnavi, C. (2011) 'Clostridium difficile infection: clinical spectrum and approach to management', *Indian J Gastroenterology*, 30(6), pp. 245-54.

Vedantam, G., Clark, A., Chu, M., McQuade, R., Mallozzi, M. and Viswanathan, V.K. (2012) 'Clostridium difficile infection: Toxins and non-toxin virulence factors, and their contributions to disease establishment and host response', *Gut Microbes*, 3(2), pp. 121-34.

Veigas, B., Pedrosa, P., Carlos, F.F., Mancio-Silva, L., Grosso, A.R., Fortunato, E., Mota, M.M. and Baptista, P.V. (2015) 'One nanoprobe, two pathogens: gold nanoprobes multiplexing for point-of-care', *Journal of Nanobiotechnology*, 13(48), pp. doi: 10.1186/s12951-015-0109-1.

Vindigni, S.M. and Surawicz, C.M. (2015) 'C. difficile Infection: Changing Epidemiology and Management Paradigms', *Clin Transl Gastroenterol*, 6, p. e99.

Voth, D.E. and Ballard, J.D. (2005) 'Clostridium difficile Toxins: Mechanism of Action and Role in Disease', *Clin Microbiol Review*, 18(2), pp. 247-263.

Waligora, A.J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A. and Karjalainen, T. (2001) 'Characterization of a cell surface protein of Clostridium difficile with adhesive properties', *Infect Immun*, 69(4), pp. 2144-53.

Warny, M., Pepin, J., Fang, A., Killgore, G., Thompson, A., Brazier, J., Frost, E. and McDonald, L.C. (2005) 'Toxin production by an emerging strain of Clostridium difficile associated with outbreaks of severe disease in North America and Europe', *The Lancet*, 366(9491), pp. 1079-1084.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) 'Jalview Version 2--a multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9), pp. 1189-91.

Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., Machi, D., Mao, C., Nordberg, E.K., Olson, R., Overbeek, R., Pusch, G.D., Shukla, M., Schulman, J., Stevens, R.L., Sullivan, D.E., Vonstein, V., Warren, A., Will, R., Wilson, M.J.C., Seung Yoo, H., Zhang, C., Zhang, Y. and Sobral, B.W. (2014) ' “PATRIC, the bacterial bioinformatics database and analysis resource.” ', *Nucleic Acids Research*, 42(D1), pp. D581-D591.

Wienken, C.J., Baaske, P., Rothbauer, U., Braun, D. and S., D. (2010) 'Protein-binding assays in biological liquids using microscale thermophoresis.', *Nature Communications*, 1(100), p. doi: 10.1038/ncomms1093.

Wijaya, E., Lenaerts, C., Maricot, S., Hastanin, J., Habraken, S., Vilcot, J.-P., Boukherroub, R. and Szunerits, S. (2011) 'Surface plasmon resonance-based biosensors: From the development of different SPR structures to novel surface functionalization strategies', *Current Opinion in Solid State and Materials Science*, 15(5), pp. 208-224.

Wilcox, M. (2013) *Updated guidance on the management and treatment of Clostridium difficile infection*.Crown.

Wilcox, M.H. and Fawley, W.N. (2000) 'Hospital disinfectants and spore formation by Clostridium difficile', *The Lancet*, 356(9238), p. 1324.

Wilson, K.H., Silva, J. and Fekety, R. (1982) 'Fluorescent-Antibody Test for Detection of Clostridium difficile in Stool Specimens', *Journal of Clinical Microbiology*, 16(3), pp. 464-468.

Woese, C.R. and Fox, G.E. (1977) 'Phylogenetic structure of the prokaryotic domain: The primary kingdoms', *PNAS*, 74(11), pp. 5088–5090.

Wu, S. and Zhang, Y. (2007) 'LOMETS: a local meta-threading-server for protein structure prediction', *Nucleic Acids Res*, 35(10), pp. 3375-82.

Wüst, J., Sullivan, N.M., Hardegger, U. and Wilkins, T.D. (1982) 'Investigation of an outbreak of antibiotic-associated colitis by various typing methods.', *Journal of Clinical Microbiology*, 16(6), pp. 1096-101.

Yang, Z., Ramsey, J., Hamza, T., Zhang, Y., Li, S., Yfantis, H.G., Lee, D., Hernandez, L.D., Seghezzi, W., Furneisen, J.M., Davis, N.M., Therien, A.G. and Feng, H. (2015) 'Mechanisms of protection against Clostridium difficile infection by the monoclonal antitoxin antibodies actoxumab and bezlotoxumab.', *Infection and Immunity*, 83(2), pp. 822-31.

Yousef, M., Najami, N., Abedallah, L. and Khalifa, W. (2014) 'Data Mining for Bioinformatics Applications', *Journal of Intelligent Learning Systems and Applications*, 6(4), pp. 153-161.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J. and Brinkman, F.S. (2010) 'PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes', *Bioinformatics*, 26(13), pp. 1608-15.

Zegers, N.D., Boersma, W.J.A. and Claassen, E. (1995) *Immunological Recognition of Peptides in Medicine and Biology*. CRC Press.

Zhang, F., Wu, X. and Chen, J.Y. (2014) 'Computational Biomarker Discovery. Towards the Virtual Cell' Chen, M. and Hofestädt, R. *Approaches in Integrative Bioinformatics*. Springer, pp. 355-386.

Zhang, S., Xing, P., Guo, G., Liu, H., Lin, D., Dong, C., Li, M. and Feng, D. (2015) 'Development of microbeads of chicken yolk antibodies against Clostridium difficile toxin A for colonic-specific delivery.', *Drug delivery*, pp. 1-8. [Epub ahead of print].

Zhang, Y. (2008) 'I-TASSER server for protein 3D structure prediction', *BMC Bioinformatics*, 9, p. 40.

Zhang, Y. and Skolnick, J. (2004) 'SPICKER: a clustering approach to identify near-native protein folds', *J Comput Chem*, 25(6), pp. 865-71.

Zhao, W., Chen, J.J., Foley, S., Wang, Y., Zhao, S., Basinger, J. and Zou, W. (2015) 'Biomarker identification from next-generation sequencing data for pathogen bacteria characterization and surveillance', *Biomark Med*, 9(11), pp. 1253-64.

Zheng, Z., Katoh, S., He, Q., Oritani, K., Miyake, K., Lesley, J., Hyman, R., Hamik, A., Parkhouse, R.M., Farr, A.G. and Kincade, P.W. (1995) 'Monoclonal antibodies to CD44 and their influence on hyaluronan recognition', *J Cell Biol*, 130(2), pp. 485-95.

Zhou, Q., Son, K., Liu, Y. and Revzin, A. (2015) 'Biosensors for Cell Analysis', *Annual Review of Biomedical Engineering*, 17, pp. DOI: 10.1146/annurev-bioeng-071114-040525.

Zillner, K., Jerabek-Willemsen, M., Duhr, S., Braun, D., Längst, G. and P., B. (2012) 'Microscale Thermophoresis as a Sensitive Method to Quantify Protein: Nucleic Acid Interactions in Solution', *Methods in Molecular Biology*, 815, pp. 241-52.