Development of Novel Molecular Tools for the Identification of Essential Genes of *Clostridium difficile* and a *Clostridium* Tetracycline Inducible Promoter System

David Jeffrey Fraser Walker, BSc(Hons)

Thesis submitted to The University of Nottingham for the degree of

Doctor of Philosophy, September 2011

Declaration

Except where otherwise acknowledged, the work presented in this thesis is my own. No other part has been submitted for another degree at The University of Nottingham or any other institute of learning.

David Jeffrey Fraser Walker

September 2011

Abstract

Clostridium difficile is the leading cause of nosocomial diarrhoea in the world, and a considerable burden to healthcare services. For colonisation of *C. difficile* to occur in the gut of an individual, the resident gut flora must first be quantitatively or qualitatively altered, normally through antibiotic treatment for an unrelated infection. At present, broad-spectrum antibiotics such as vancomycin and metronidazole, the frontline choices for the treatment of *C. difficile* infection; disturb the normal gut flora, suffer from poor recurrence rates, and have received reports of sporadic emergence of resistance. Development of novel narrow-spectrum antimicrobials would circumvent these problems but depend on the identification of novel essential genes. Molecular techniques available to identify and study essential genes in other organisms have not yet been applied to *C. difficile*. In this study, we identified 208 candidate essential genes via *in silico* analysis based upon similarity to known *Bacillus subtilis* essential genes. In order to provide experimental evidence of essentiality, we developed a novel method utilizing partial diploids and functionally characterised three *C. difficile* genes as essential; CD0274, *metK*, and *trpS*. This method not only identified CD0274, a candidate narrow-spectrum drug target and the first essential genes in *C. difficile*, but also provides a reliable method to identify further essential genes for novel antimicrobial targeting. In addition, we developed a *lac* repressor system, a rationally designed theophylline-responsive riboswitch, and most importantly, a Tet inducible promoter system able to conditionally control expression of a *catP* reporter through a wide dynamic range in both *C. difficile* and *C. sporogenes*, to maximum induction factors of 192.89 and 1,275.63, respectively. The combinations of these characteristics make this Tet system not only a powerful tool for identifying essential genes, but bestows a great potential for further analysing gene function far beyond the scope of this project.

Acknowledgments

I would like to thank Nigel Minton and Klaus Winzer for giving me this wonderful opportunity to undertake a PhD with such a prominent group within the *C. difficile* field. Furthermore, their supervision, guidance, and mentoring over the years has helped me to where I am.

I thank all the past and present members of the Clostridia Research Group at the University of Nottingham for providing a friendly and supportive work environment which has made doing this PhD such an enjoyable experience. In particular, I would like to thank Steve Cartman, John Heap, and David Burns for their infinite wisdom and patients in providing excellent training early in the project.

I owe John Heap a special debt of gratitude for being my laboratory supervisor. His practical and most importantly, theoretical training, has been essential in undertaking this project. His continual guidance and support far out with his duties, and work, has given me great confidence in securing and following a research career.

On a personal note, I would like to thank my close friends at The University of Nottingham, Robert Goldstone, Roman Popat, Avika Ruparell and Matt Twigg, who, through many discussions, have given me many great ideas, many comical moments, and have helped through difficult times. I am sure without them; this project would not be what it is.

Last but not least, I would like to thank my parents for their continued support, and my grandmother, who this thesis is dedicated to.

List of Contents

Table of Contents

List of Figures

Chapter One: Introduction

Chapter Three: A Novel Method for the Identification of Essential Genes

Chapter Four: Development of Conditional Expression Systems

List of Tables

Chapter Two: Materials and Methods

Chapter Three: A Novel Method for the Identification of Essential Genes

Abbreviations

CDI	*Clostridium difficile* associated disease
PMC	Pseudomembranous colitis
Da	Dalton
RNA	Ribonucleic acid
mRNA	Messenger ribonucleic acid
DNA	Deoxyribonucleic acid
bp	Basepair(s)
RNP	Ribonucleoprotein
IEP	Intron encoded protein
RAM	Retrotransposition-activated selectable marker
ACE	Allele-coupled exchange
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Tc	Tetracycline
aTc	Anhydrotetracycline
TraDIS	Transposon directed insertion-site sequencing
CDEPT	Clostridial-directed enzyme prodrug therapy
LB	Luria-Bertani
BHIS	Supplemented brain-heart infusion medium
rpm	Revolutions per minute
g	Gravity
PBS	Phosphate buffer solution
ORF	Open reading frame
CAT U	Chloramphenicol acetyltransferase units
CAT U/mg	Chloramphenicol acetyltransferase units per milligram
CAT U/ml	Chloramphenicol acetyltransferase units per millilitre
PCR	Polymerase chain reaction
BLAST	Basic Local Alignment Search Tool
OD	Optical density
MG	Methylglyoxal
PD	Propandiol

Chapter One

Introduction

1. Introduction

Antimicrobial discovery and development was undoubtedly at the forefront of research in the 20th century with a multitude of new products being developed to meet medical demands. With this success came the complacency derived from the belief that "it [was] time to close the book on infectious diseases" (U.S. Surgeon General, 1965). Pharmaceutical companies therefore switched research focus towards other therapeutic areas such as the treatment of chronic conditions[2]. This trend has continued with increased regulatory requirements in manufacturing, safety, and efficacy which have increased the focus on research and development on chronic conditions for a 'commercially attractive' return on investment[2,3]. As pharmaceutical companies headed for the exit, there has been an unrelenting rise in the occurrence of antimicrobial resistance by many common pathogenic bacteria, posing the question; will industry be able to meet future needs for new effective antimicrobial products?[4] At present, physicians commonly have to turn to their second or third choice antibiotics in order to combat resistant bacteria. The effectiveness of vancomycin, which is seen as the last line of defence against many organisms has already started to waiver[5-8]. Therefore, there is a great demand for the discovery of novel antimicrobial targets to add to our arsenal.

The last three decades have presented another major drawback of current antibiotic usage, namely *Clostridium difficile* infection (CDI)[5-13]. *C. difficile*, a Gram-positive, anaerobic spore-forming, multi-drug resistant bacterium is the causative agent of CDI, the leading cause of nosocomial infectious diarrhoea worldwide[5-13]. The pathogen is frequently associated with antibiotic treatment and manifestation of disease range from asymptomatic carriage, mild diarrhoea to life-threatening pseudomembranous colitis (PMC)[5-8]. Over the past few years there has been an exponential increase in the prevalence of *C. difficile* in the UK with severe CDI being identified in originally low risk

groups[6,10,11]. This change in epidemiology and severity of disease could be due to the emergence of hypervirulent and increased microbial resistant strains such as *C. difficile* PCR-ribotype 027[6,8,14].

1.1 *Clostridium difficile:* Emergence of a Significant Human Pathogen

C. difficile was first described in 1935 by Hall and O'Toole as a slow growing anaerobic bacillus that was difficult to isolate from normal new-born infant gut flora[15]. Due to this difficulty in growing the organism in the laboratory, the species name *difficile* was bestowed upon the organism. However, today the name could be construed to represent the difficulty in combating the organism which has grown to epidemic proportions. During the 1950's, pseudomembranous enterocolitis was starting to emerge as a nosocomial disease attributed to either *Staphylococcus aureus*, an organism whose increasing prevalence at the time was associated with hospital in-patients who had undergone antimicrobial therapy, or to *Candida albicans*[5,7]. In 1974, a prospective study of 200 patients treated with clindamycin showed that 21% of the patients developed diarrhoea and 10% pseudomembranous colitis, proposing a direct link between antibiotic usage and colitis[16]. It was not till 1977 that a toxin produced by a *Clostridium* species was proposed as the causative agent of clindamycin-induced ileocecitis in hamsters[16]. However, by 1978 it was firmly established that *C. difficile* was the causative bacterium of antibiotic-associated colitis[5,9,12]. Today *C. difficile* is responsible for up to 20-25% of cases of antibiotic-associated diarrhoea and almost 100% of pseudomembranous colitis (PMC)[17].

According to the Communicable Disease Surveillance Centre (CDSC) for England and Wales, CDI reports of inpatients contracting *C. difficile* have risen from 1,000 in 1990 to 15,000 in 2000, 35,000 in 2003 and to 60,000 in 2007[12]. Of the patients reported to have CDI in 2007, 50,000 were aged 65 years and over. This increase in incidence also

manifests as an increase in mortalities mentioning *C. difficile* on the death certificate. The deaths associated with *C. difficile* have increased steadily from 1999 to 2007, where they peaked at 8,200 deaths. Rates however decreased from 2007 to 2010, where deaths were 2,800, 65.8% lower than 2007 (figure 1.1). This decrease is likely due to increased surveillance and quick containment of infected patients.

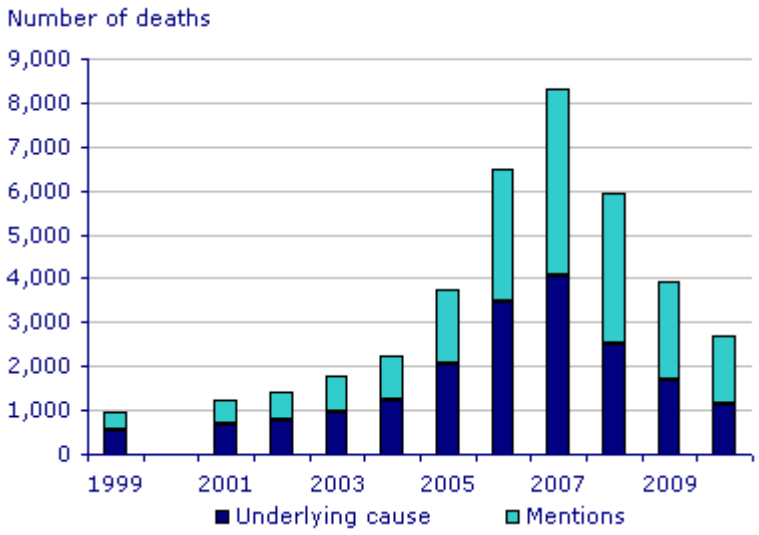


Figure 1.1. Number of death certificates mentioning *C. difficile* as either an underlying cause or directly implicated, in Wales and England between 1999 and 2010. The increase in prevalence in 2004 may be due to a mandatory surveillance scheme being introduced. Taken from the Office of National Statistics (http://www.statistics.gov.uk).

A study by McFarland *et al*, (1989)[18], found that about 20% of patients admitted to a hospital were already positive, or contracted a toxigenic strain of *C. difficile* during their stay[18]. In contrast, it is estimated that only about 1% of the total population of healthy outpatients are positive for the bacterium[11]. Recently, community-acquired CDI has also seen a striking exponential rise in prevalence from less than 1 per 100,000 people in 1996 through to 18 per 100,000 in 2004 (figure 1.2)[11].

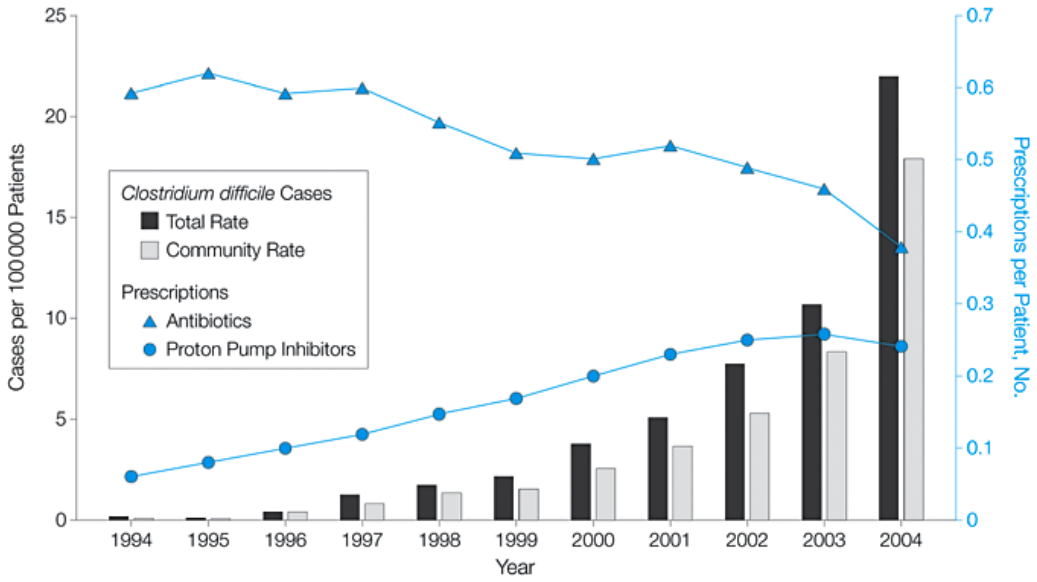


Figure 1.2. Risk of total and community-acquired infection rate of *C. difficile* compared to prescriptions of antibiotic and protein pump inhibitors per 100,000 patients in the UK between the years 1994 and 2004. (Taken from reference [13]).

1.1.1 C. difficile *disease*

C. difficile is able to survive in the environment by producing spores; these spores are highly-resistant to many commonly used cleaning agents and can persist for months in the environment which are subsequently ingested by the patient[19,20]. Interestingly, only about one third of patients infected with a toxigenic strain of *C. difficile* develop symptoms, the remaining two thirds are asymptomatic and are thought to act as a reservoir of infection[21,22]. The organism and its spores can be isolated from a variety of different areas. As expected, places like commodes and bed frames have a very high concentration of the organism, also; toilets, floors, curtains, banisters, door handles, telephones, hands of healthcare workers, and many more places have all been demonstrated to be contaminated[5,17,18].

For colonisation of *C. difficile* to occur in the gut of an individual, the normal gut microflora must be qualitatively or quantitatively altered[17]. Anaerobic microbial flora in a normal individual confers 'colonisation resistance' inhibiting some pathogenic bacteria such as *C. difficile* from colonising the gut, although the precise mechanism involved is unknown[10]. The most common risk factor associated with CDI is exposure to antibiotics[9,10]. A meta-analysis study conducted by Bignardi in 1998 assessing risk factors compared forty-nine published reports based on comparison of *C. difficile* cases against individuals without diarrhoea[9]. It was noted that a high number of antibiotics and classes have a significant statistically positive association with both *C. difficile* carriage and diarrhoea (figure 1.3)[9]. Ironically, vancomycin and metronidazole both used in the treatment of CDI also have a medium risk for predisposing a patient to CDI[9]. Age, in particular being over 65, is also a major risk factor associated with CDI, possibly due to reduced function of the microflora of the gut[9]. There are many more risk factors which can predispose a patient to CDI (figure 1.4). Protein pump inhibitors (PPI's) for example have recently been described to increase the risk three-fold of contracting CDI, especially in a community setting (figure 1.2)[11]. PPI's are drugs presumed to increase gastric pH thereby decreasing the destruction of the spores within the stomach.

High Risk	Medium Risk	Low or Questionable
Aztreonam	Aminoglycosides	Bacitracin
Cefotaxime	Metronidazole	Cisplatin
Nitrofurantoin	Ceftraiaxone	Doxourbicin Hydrochloride
Ceftazidine	Ampicillin or Amox.	Fluorouracil
Amp. or Amox. + ClavulanicAcid	Erythromycin	Methotrexate
Pristinamycin	Co-trimoxazole	Rifampin
Spiramycin	Antistaphylococcal Penicillins	Sulfonamides
Trimethoprim	Vancomycin	Tacrolimus (FK506)
Cefotaxim	1st Gen. Cephalosporins	Teicoplanin
Imipenem	Penicillin	
Cefuroxime	Tetracyclines	
AntipseudomonalPenicillins	Ticarcillin and Calavulanic Acid	
Clindamycin or Lincosamides		
Quinolones		
Cefaclor		

Figure 1.3. Risk of *C. difficile* diarrhoea associated with individual antibiotics or antibiotic classes. Adapted from reference[9].

Risk Factors	*C. difficile* Diarrhoea	*C. difficile* Carriage
Increasing Age (excluding infancy)	✓	
Severity of Underlying Disease	✓	
Malignant Haematologic Disorder		✓
Non-surgical GI procedures	✓	
Nasogastric Tubes	✓	✓
Anti-ulcer Medications	✓	✓
ITU Stay	✓	
Duration of Hospital stay	✓	
Duration of Antibiotic Course	✓	
Multiple Antibiotics	✓	

Figure 1.4. Other risk factors associated with the two main outcomes of colonisation with *C. difficile*. Adapted from reference[9].

The sequence of events that lead to CDI are important: firstly the patient is exposed to a factor or factors that alter the microflora of the gut; secondly, exposure to a toxigenic strain of *C. difficile*, and; finally, the presence of additional factors (such as advanced age) all determine the subsequent outcome of colonisation. CDI used to be seen solely as a healthcare associated infection, however, over the past 6 years the epidemiology of CDI has been changing. Between 2003 and 2005, severe CDI was identified in originally low risk groups such as; peripartum women within 4 weeks after delivery, patients which have not been hospitalised in the previous 3 months which may not have been exposed to antibiotics, and young people with no serious underlying illness[6]. This change in epidemiology and severity of disease could be due to the emergence of hypervirulent and increased antimicrobial resistant ribotypes such as PCR-ribotype 027[6,8].

1.1.2 Emergence of hypervirulent strains

Since the turn of the century, the incidence of CDI has increased dramatically in line with the emergence of so-called hypervirulent strains. These hypervirulent strains are thought to cause more severe disease, higher relapse rates, and increased mortality[6,8,14,23,24]. One of the most prevalent types of hypervirulent strains belong to the restriction endonuclease type BI, North American pulsed-field type I (NAP1), and PCR-ribotype 027 (BI/NAP1/027). In the past, the isolation of such strains was considered rare, however, recently they have become highly represented among clinical isolates, so much so that BI/NAP1/027 is now the most common *C. difficile* type isolated in England and is frequently associated with outbreaks (figure 1.5)[25]. Between 2007 and 2008, the incidence of PCR-ribotype 027 increased by 15% to represent 41% of isolates in 2008, while the incidence of PCR-ribotypes 001 and 106 decreased by 17% and 6%, respectively, during the same time period. The presence of representatives of the BI/NAP1/027 type has been confirmed in at least 16 countries throughout

mainland Europe[12,24]. In North America, the strain is thought to be associated with the increased usage of quinolone class of antibiotics, resulting in greater resistance to floroquinolones[12]. Worryingly, BI/NAP1/027 has also been linked to the emergence of community-acquired CDI, removing the nosocomial label and taking the disease out into the community[11,23].

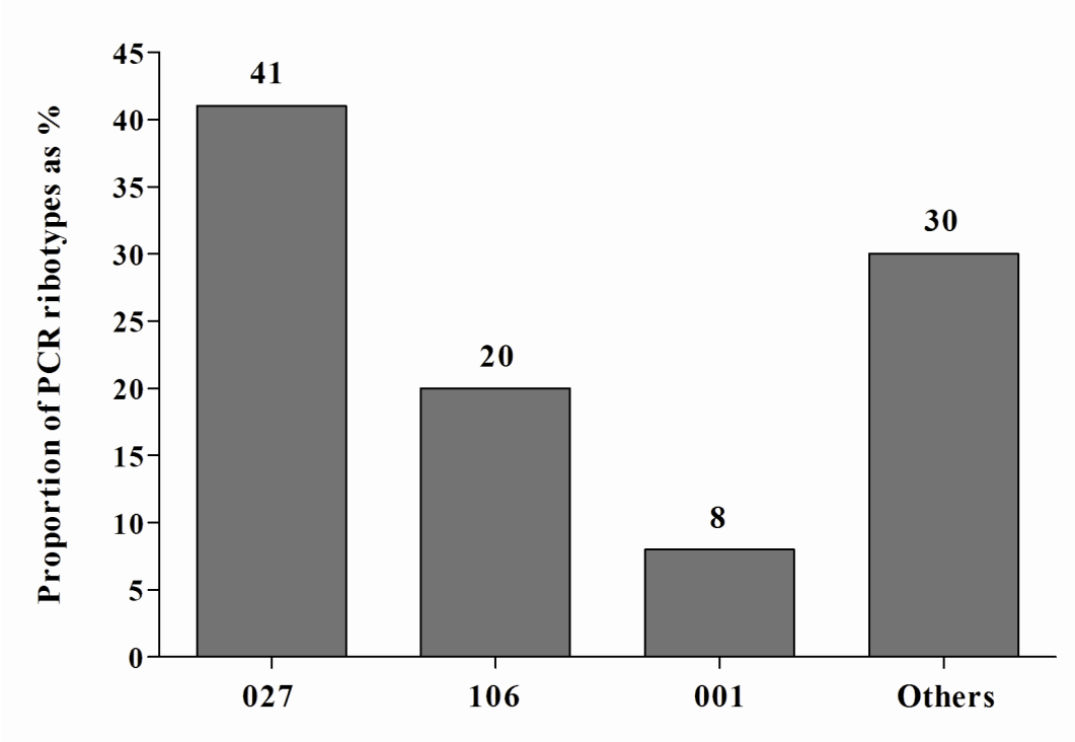


Figure 1.5. Frequency of isolation of *C. difficile* PCR-ribotypes in England during the period of 2007 to 2008. The data represents a total of 677 individual isolates taken nationally. Adapted from reference[25].

The reason why such ribotypes are hypervirulent is unknown. There is evidence to suggest that the BI/NAP1/027 strains produce higher levels of toxins in the laboratory, and it has been suggested that they are more prolific in sporulation than their non-hypervirulent counterparts[13]. Other studies question this latter suggestion, and propose sporulation rates are not associated with type[26]. All in all, the molecular basis of hypervirulence remains unknown, but understanding the basis of the contributing

factors may allow specific countermeasures to be developed to overcome outbreaks and increase the efficacy of treatments.

1.1.3 C. difficile *virulence factors*

Pathogenic strains of *C. difficile* produce two potent toxins, toxin A (TcdA), an enterotoxin, and toxin B (TcdB), a cytotoxin, both of which ultimately mediates the severity of disease[27]. In addition to the two well characterised toxins, a binary toxin, an actin specific adenosine diphosphate-ribosyltransferase, can be found in all PCR-ribotypes of 027 which are thought to make up approximately 41% of clinical isolates of *C. difficile* in 2008[25]. The two large exotoxins, TcdA and TcdB, are some of the largest bacterial toxins known, at 308 kDa and 269 kDa, respectively, and exhibit homology at the amino acid level of over 45% to each other. The encoding genes (*tcdA* and *tcdB)* are localised to a 19.6 kb pathogenicity locus (PaLoc) in *C. difficile*, along with; *tcdE*, a cell wall hydrolase, *tcdR* an alternative sigma factor involved in positive transcriptional regulation, and *tcdC*, a putative negative regulator[28].

When the toxins come into contact with receptors on the luminal side of the colonic epithelium, they are transported into the cytoplasm. Both toxins are glycosyltransferases, and once inside, they interact with the Ras superfamily of small GTPases. TcdA and TcdB catalyse the addition of a glucose moiety to threonine-37 of RhoA, utilising UDP-glucose as a co-substrate[27]. Both toxin A and toxin B are able to target RhoA, -B, -C, Rac, and Cdc42, causing actin condensation resulting in rounding of the cells, membrane bleeding and ultimately apoptosis of the target cell[27].

There is currently some debate as to which toxins are required for virulence. A recent study concluded that toxin B, not toxin A is essential for virulence in a hamster model[29]. The preeminence of toxin B in disease is supported by the isolation of a number of

clinically relevant virulent strains which lack toxin A, and only produce toxin B. However, in contradiction to these results, another study published from our laboratory suggested that both toxin A and toxin B alone are sufficient to cause virulence in a hamster model[30]. There were a number of differences between these studies, such as the strain of *C. difficile* which was used, which could account for the differences seen. It is clear that further research into the role of toxins in virulence is required to elucidate the fundamental function of each toxin in causing CDI.

Apart from the toxins being the main virulence determinants, other virulence factors may have synergistic roles; although the extent of their involvement is unclear. The unique outer surface layer proteins, predominantly composed of SlpA, make up the S-layer of *C. difficile* vegetative cells. It is thought that the S-layer proteins are responsible for adherence of *C. difficile* to host cells which subsequently elicit inflammatory and antibody responses[31-35]. The flagella of *C. difficile*, composed primarily of FliC and FliD, are thought to be involved in adherence to host cells, allowing colonisation of the gut[36]. However, in contrast to the published work, results from our laboratory suggest that flagella may not play a role in adhesion (Soza Baban, Sarah Kuehne, and Nigel Minton, unpublished results).

1.1.4 C. difficile *treatment*

Treatment of CDI has evolved little in recent years, with only a few proven therapeutic options available. At present, oral metronidazole and vancomycin are the antibiotics of choice for the treatment of CDI; however bacitracin methylene disalicylate, fusidic acid, nitazoxanide and teicoplanin have also been shown to be clinically effective[5]. Two more antibiotics; rifaximin, and ramplanin, and 2 non-antimicrobials, CDA1 (MDX-066) and MDX-1388 are currently in developmental phase II or III[14]. However, all the antibiotics mentioned are broad-spectrum and not novel classes so may encounter the same

complications as the other commonly used broad-spectrum antibiotics. *C. difficile* is generally susceptible to vancomycin, but occasionally strains have reported resistance to metronidazole and vancomycin, coupled with the high rate of recurrence following treatment, new treatment options or novel narrow-spectrum antibiotics would be highly advantageous[6].

One such novel antibiotic looks promising, fidaxomicin (PAR-101/OPT-80), the first in a new class of macrocyclic antibiotics. Fidaxomicin, an RNA inhibitor, has a narrow spectrum of activity, almost purely *C. difficile* specific[37]. Louie *et al*, (2011)[37], investigated its efficacy as an alternative to standard CDI treatments[37]. The authors subjected 629 adults to a randomized, double-blind phase II clinical trial. Subjects either were treated with fidaxomicin (200 mg twice a day), or vancomycin (125 mg 4 times a day) orally for 10 days. The experimental endpoint was defined as the patients requiring no further treatment 2 days after completion of the trial, determined by the investigators. To evaluate relapse, the investigators included a secondary endpoint 4 weeks post therapy. The trial determined that fidaxomicin was very similar to vancomycin in treatment of CDI, with 88.2% of patients showing no signs of infection after the first endpoint of treatment with fidaxomicin, compared to 85.8% for vancomycin. In addition, the investigators noted that fidaxomicin treatment resulted in a significantly lower rate of relapse, 15.4% vs. 25.3% (P=.005) for fidaxomicin and vancomycin, respectively[37]. There was also no significant adverse side-effects following fidaxomicin treatment. From these results, fidaxomicin appears to be an extremely promising drug for treating CDI and preventing high relapse rates, outlining the importance and effectiveness of a narrow-spectrum antimicrobial.

1.1.5 A need for novel antimicrobial targets

At present, there is a limited selection of antibiotics available for the treatment of *C. difficile*, many of which target the same conserved cellular process or even the same enzyme as other antimicrobials, increasing the potential for cross-resistance. Over the past 10 years, the search for novel antimicrobial targets and agents against *C. difficile* has acquired a new sense of urgency due to increasing resistance and the chance of recurrence after treatment with current broad-spectrum antibiotics. Development of new antibiotics that circumvent these 'resistance and recurrence problems' will depend on novel species-specific targets being exploited. Early efforts have been limited by the lack of technology to identify targets[38]. With the advent and availability of whole genome sequencing in the mid-1990s, hundreds if not thousands of potential new targets could be identified, leading the way for many novel antimicrobial classes. However, even before the sequencing of many organisms had been undertaken, a number of possible novel target sites have been identified using classical methods in many bacteria; such as CD0274, yet very little, if not no research has been undertaken in *C. difficile*[39,40]. To date, there has only been one candidate essential gene which could be used as a novel narrow-spectrum antibiotic target in *C. difficile* cited in the literature. Liyanage *et al*, (2003)[39,40], hypothesised that CD0274, a suspected alcohol dehydrogenase the authors annotated as *gldA*, was involved in the sole detoxification pathway of the toxic metabolite methylglyoxal (MG). The author's inability to inactivate CD0274 while keeping cellular viability led to the suspicion that it was essential. If true, CD0274 would make a candidate for antimicrobial targeting. However, further validation is required to provide stronger evidence of CD0274's essentiality.

1.2 Selection and Validation of Novel Drug Targets

The ideal microbial enzyme target should have no homology to humans to minimize toxicity, have a fully elucidated mechanism of action and function, and be required for survival of the bacterium to unsure chemical inhibition will ultimately result in cell stasis or death[2-4]. Generally broad-spectrum antibiotics are preferred as they are cost-effective and clinically easy to use, but in the case of *C. difficile*, a narrow-spectrum antibiotic would be of interest due to the complications with gastrointestinal sequelae caused by the antibiotic-associated destruction of the normal gastrointestinal flora. The most important criterion for a novel drug target is the assessment of the target being essential to cell viability during infection, or even better, absolutely required for survival. To assess this viability, a number of genetic tools have been employed in other bacterial species. Unfortunately, many of these tools have not been developed for usage in *C. difficile* for the purpose of essential gene identification, however, with the advancement of current technology and genetic tool developed within our laboratory, it is now possible to combine and develop novel tools to identify essential clostridial genes[38,41,42].

1.2.1 Methods to study essential genes

Essential genes are of significant importance in pathogenic bacteria for the purpose of therapeutic drug targeting, as such, many studies have been carried out using several experimental and computational approaches to identify essential genes in organisms such as; *E. coli*[43,44], *B. subtilis*[45], and *S. aureus*[46,47]. There are two ways which to identify candidate essential genes, through bioinformatic or experimental approaches. The bioinformatic approaches use large-scale *in silico* screening with the assumption that essential genes are likely to be common among different species so will be highly conserved[48]. However, bioinformatic approaches can only identify candidate essential genes, to validate them as essential, experimental approaches must be applied.

There are two broad strategies to identify essential genes by experimentation, by forward genetics, or by reverse genetics. Forward genetics is the term given to an approach used to identify genes, or set of genes, responsible for a particular phenotype. In this method, no possible targets are identified prior to experimentation. The identification of genes is carried out by searching the entire genome through random approaches such as; transposon mutagenesis or shotgun-antisense. In contrast, reverse genetics already knows which gene to target, and is used to identify a phenotype associated to the gene. This targeted approach uses techniques such as the ClosTron, plasmid insertion mutagenesis, allelic exchange, or construction of conditional mutations. These experimental methods can be classified further into direct or indirect. Direct approaches are able to identify essential genes directly, by using conditional mutations, i.e., being able to control expression of a gene by the addition or subtraction of an inducer[49]. Indirect approaches are able to identify essential genes by two ways, the first is they are able to identify non-essential genes directly by the ability to inactivate them; in contrast the genes which cannot be inactivated may be classed as candidate essential genes. To provide stronger evidence of essentiality, you can show that the gene is able to be inactivated if a second copy of the gene is present within the chromosome, thereby constructing a partial diploid and allowing a functional complementation of the native essential gene[50]. However, using reverse genetic techniques such as these, candidate essential genes first need to be identified.

1.2.2 Conservation of essential genes

The strength of negative (purifying) selection is thought to be the main contributing factor in determining the rate of evolution of protein-coding regions[51]. It has been proposed that essential genes should evolve slower than nonessential genes within an organism because they are, by definition, functionally indispensable for the survival of

the organism. Due to the role of purifying selection in determining evolution rates, the selection pressure on essential genes to remain functional manifests as a lower rate of evolution compared to nonessential genes[51]. This theory is called the "knockout-rate" prediction and was established by comparing the synonymous and nonsynonymous nucleotide substitution rates of essential and nonessential genes between *E. coli*, *Helicobacter pylori*, and *Neisseria meningitidis*[51].

In *B. subtilis*, Kobayashi *et al*, (2003)[52], predicted that out of the 271 essential genes identified, 80% have homologues in other bacterial species, and 20% have homologues in all bacterial species[52]. Of those 271 essential genes, 209 have putatively essential orthologues in *Staphylococcus aureus* and 150 are thought to be essential in *E. coli*, although the functions of the genes are unknown[52]. However in *E. coli*, there are 67 genes which are counted as essential whereas they are non-essential in *B. subtilis*, and 86 have no *B. subtilis* orthologs[53]. In the bacterium *H. pylori*, there are 344 identified essential genes, of which only 20 are known to be essential in all three; *E. coli*, *Haemophilus influenza*, and *Mycoplasma genitalium* species[54].

These publications show that although essential genes are conserved more highly than nonessential genes, there can still be variation between species. As such, a genes essentiality in one organism does not automatically predict essentiality in another organism and as such, each gene must be confirmed to be essential by experimental evidence.

1.2.3 In silico *approaches*

In silico analysis has proven to be a powerful tool to compare genomes of different bacteria. In 1995, the first two bacterial genomes were completely sequenced; *H. influenza*[55] and *M.* genitalium[56]. Based upon the assumption that essential genes are conserved across bacterial species, Mushegian and Koonin, (1996)[48], derived the minimal gene set required for cellular life by comparing the complete genomes of both the aforementioned bacteria[48]. Their theory was that both *M. genitalium* and *H. influenza* belong to two ancient lineages, i.e. Gram-positive and Gram-negative bacteria, respectively. Because of this, the genes which are conserved between the two genomes are almost certainly essential for cellular function, as such, it is this category which is most likely to approximate the minimal gene set[48]. Using this theory, the authors speculated that 256 genes is the lowest minimal set required for cell viability[48]. Using an experimental approach, Itaya, (1995)[57], determined the maximum randomly selected genetic loci *in B. subtilis* which could be disrupted without loss of viability[57]. Itaya, (1995)[57], concluded that the estimated minimal genome size of 318 kbp, corresponding to 254 genes (the average gene length *ca.* 1.25 kbp), was the minimal gene set required for life in *B. subtilis*[57]. Considering the differences in approaches of these two studies, both authors predicted a remarkably similar number of genes to be the minimal set required for cellular viability[48,57]. As the publications of complete genomes is increased, other *in silico* predictions to identify essential genes have been made, including comparison of up to 5 bacterial genomes, identifying 89 genes which are conserved between the large phylogenetic distant organisms[58].

While computational approaches are a good starting place, there are significant limitations to using such systems to identify essential genes. When using *in silico* approaches to identify minimal essential gene sets of an organism, they are likely to

underestimate the total number because only the genes that are conserved through the divergence of the species will be recognised, therefore genes which have evolved to become species-specific will not have orthologues in other species, and so will not be recognised. Also, genes which are known to be essential in one organism might not be essential in another organism. *In silico* identification can also be fallible; a study used seven proposed essential genes in *H. influenzae* identified by concordance analysis[48], and experimentally analysed them using transposon mutagenesis[59]. The results showed only three of the predicted seven were essential for viability[59]. As such, using computational approaches will give insight and identify candidate essential genes, but as of now, computational approaches will need to be backed up by reverse genetic methods to confirm essentiality.

1.2.4 ClosTron

Until 2007, one major problem with the genus *Clostridium* was the lack and efficiency of methods for insertionally inactivating specific genes[38]. A few authors had published reports of single crossover knockout mutants obtained from using a replicative-defective plasmid, but the mutants generated were unstable[40]. This was a consequence of excision events between the duplicated regions of DNA that flanked the target site[38,40]. Awad and colleagues[60] reported a mutant generated via allelic exchange in *C. perfringens*. The isolation of such rare mutants in *C. perfringens* was possible because of the high frequency at which DNA can be transformed into this *Clostridium* compared with others. In response to this extremely low integration frequency observed in other clostridial species, a novel method using bacterial group II introns called "TargeTron" has been adapted for use in clostridia by Heap and colleagues and termed "The ClosTron"[38,61].

The ClosTron is a novel reverse genetic method which uses bacterial group II introns, to insertionally inactivate any required target gene. Group II introns are of particular use in targeting genes because they can be retargeted to insert into virtually any target gene with high efficiency and specificity[38,61,62]. The group II introns, from the *ltrB* gene of *lactococcus lactis* (L1.ltrB), mobility is mediated by a site-specific replication mechanism termed retrohoming. Retrohoming is when RNA introns are reverse spliced into DNA target sites and then transcribed with an intron-encoded protein (IEP)[63]. The process is controlled by the production of a ribonucleoprotein (RNP) particle which contains the IEP and excised intron lariat RNA, during RNA splicing. RNPs initiate integration into the specific gene by using the IEP, with base pairing of the intron RNA complementary to the specific DNA target sequence[64]. By modifying the base pairing on the intron RNA, it is possible to retarget the group II intron to integrate into specific complementary DNA sites within the target gene (figure 1.6B). Another advantage of using group II introns is the ability to remove the IEP open reading frame (ORF) from the intron encoding region and placing it on a distal site on the plasmid without affecting intron mobility (figure 1.6A)[62]. Following the IEP-mediated intron integration into the target gene, subsequent loss of the plasmid and the IEP gene, prevents subsequent splicing of the inserted intron and the creation of an unconditional mutant.

In the natural group II intron, the IEP gene occupies a non-structural domain (IV), which plays no role in the retrohoming process. As the IEP gene has been relocated elsewhere on the ClosTron plasmid, this non-structural region can be the location of the insertion of alternative "cargo" DNA, such as a positive selection marker.

To select for integration of the intron within the target gene, an antibiotic selection marker is a quick and easy way to positively select for integration events. Zhong *et al*,

(2003)[62], describe the use of a "TwinTron" in which the L1.LtrB intron domain IV, which previously housed the IEP, now contains an antibiotic resistance marker in which another group I intron interrupts the marker in the opposite orientation (figure 1.6A)[62]. The orientation of the group I intron, group II intron, and the antibiotic resistance marker were designed so when integration of the group II intron into the target gene occurs, the group I intron would self-splice out of the anti-sense antibiotic resistance marker, restoring function. Therefore, insertion of the group II intron into the target gene activates antibiotic resistance to an appropriate antibiotic, allowing positive selection. This type of marker is termed a Retrotransposition-Activated selectable Marker (RAM), because the restoration of antibiotic activity only occurs when the intron has passed through an RNA intermediate[62]. Using this method it is possible to insertionally inactivate genes in clostridia with high frequency and stability[38].

Despite the effectiveness of the ClosTron, if a gene is essential, insertional mutants will not be obtained. Failure to obtain ClosTron insertions if a number of different retargeted plasmids are employed could be taken as evidence that a gene may be essential. However, the design of the retargeted ClosTron plasmids is based on an *in silico* algorithm and is not necessarily 100% reliable. As such, the ClosTron alone is insufficient to identify essential genes.

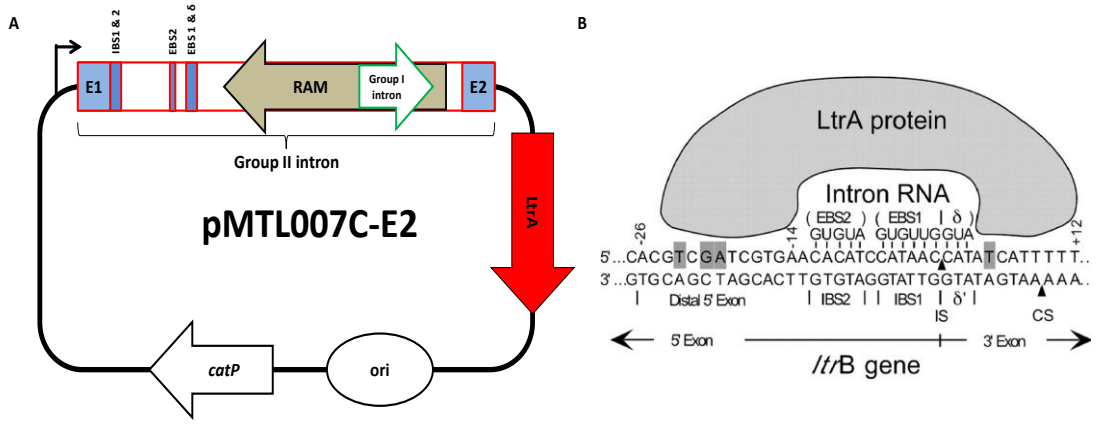


Figure 1.6. The ClosTron system's element orientation on an appropriate vector (pMTL007C-E2) and its site specific recombination mechanism. (**A**) ClosTron construct contained on plasmid pMTL007C-E2, illustrated is the flanking recognition sites (E1 & E2) re-targeted intron region (EBS/IBS), RAM including the group I intron, the LtrA gene, and *catP* for positive selection of the plasmid. (**B**). Mechanism of DNA target site recognition sites; IBS, EBS1d and EBS2 by the *L. lactis* Ll.LtrB group II intron facilitating integration of the intron.

1.2.5 Targeted conditional expression

Targeted conditional expression is a means to control the expression of a target gene by placing it under inducible control, via an inducible promoter or riboswitch. Conditional expression is a powerful tool in analysing gene function, as gene expression can be switched on or off in the presence or absence of an exogenously added inducer. One such application of such systems is the ability to directly identify essential genes[49,65,66]. By placing an essential gene under conditional expression, the resulting strain is dependent upon the inducer for cellular viability[67]. It can also allow the isolation of a conditional mutant genotype which can then be analysed further by defining when the gene is

essential during growth, and what effects different levels of expression of the target gene have upon the cell.

For selection or development of such a system, it is important that the inducible promoter has a dynamic range of expression, tight repression of expression in the absence of an inducer, high levels of expression under induction, the inducer should be nontoxic to the cell under working concentrations, and the inducer should present good bioavailability. Two commonly used inducible promoter systems in prokaryotes are the S_{pac}[68] and $P_{Xyl/tet}$[69] systems for the identification of essential genes. The S_{pac} promoter is a hybrid between the *B. subtilis* SPO1 phage promoter and the operator sequences of the *E. coli lac* operon, the latter element provides the facility to control transcription of the S_{pac} by the addition of the inducer, isopropyl β-D-1-thiogalactopyranoside (IPTG)[68]. The $P_{Xyl/tet}$ system is a hybrid between the xylose inducible system of *B. subtilis* and the operator sequences of the Tn*10 tet* gene. In this case, transcription is controlled via the addition of the inducer, tetracycline (Tc), or its less toxic analogue anhydrotetracycline (aTc)[69]. In addition to these two, there are a range of other inducible promoter systems that have been developed, each with their own characteristics giving them advantages and disadvantages depending on their function, environment, and organism.

There are two ways to identify essential genes by conditional expression. The first uses an integration method to transcriptionally fuse an inducible promoter to an essential target gene, thereby replacing the native promoter[52,67]. This type of system has been used in *B. subtilis* using the P_{Spac} system present upon the non-replication plasmid pMUTIN[52]. The authors cloned the P_{Spac} system at the 5' of the target gene and used the pMUTIN plasmid to undergo a single-crossover recombination event creating an intact copy of the target gene under inducible control of P_{Spac} in the chromosome[52]. If the

recombinant strain is IPTG dependant for cellular viability, the gene can be defined as essential. Using this method, 150 genes have been directly identified as being essential in *B. subtilis*[45]. The big disadvantage to using this system would be if the essential gene is present within an operon. Modifying the transcriptional control of an operon would have an effect on expression of genes downstream and make the interpretation of the results difficult. Also if the target gene is not essential but a gene further downstream is, and it is affected by the insertion, the target gene might be misclassified as essential.

The second strategy to conditionally control the expression of an essential gene is to introduce an ectopically expressed copy of the essential gene into the chromosome under inducible control, creating a partial diploid strain[67]. The native gene can then be removed or disrupted allowing the ectopic copy to take over essentiality in the presence of an inducer. This method circumvents the problems associated with essential genes present within an operon. This approach has been used successfully in *S. aureus* to test genes for essentiality. Fan *et al*, (2001)[67], transcriptionally fused the P_{Spac} or $P_{Xyl/tet}$ system to a candidate essential gene and integrated the construct into the *geh* locus of *S. aureus*[67]. The resulting recombinant strain were then subjected to allelic exchange mutagenesis to replace the native copy with a functionally tetracycline resistance cassette to aid selection. The authors identified two genes to be dependent on the presence of an inducer for cellular viability, providing direct evidence of the genes essential nature[67].

1.2.6 Temperature-sensitive mutants and disruptions

Temperature-sensitive mutants enable conditional control of the activity of a protein via temperature changes, as opposed to controlling transcription or translation. There are a few different ways in which temperature-sensitive mutations can be exploited for the identification of essential genes. One of the first genome wide searches for essential

genes involved the isolation and analysis of temperature-sensitive mutants in *Salmonella enterica* serovar Typhimurium (S. Typhimurium)[70]. The authors subjected the bacterium to chemical mutagenesis and isolated clones which were able to grow at the standard permissive temperature, but not at higher-temperatures. The locations of the mutations were identified as point mutations within key residues in the protein responsible for protein folding, with subsequent destabilisation at the higher temperatures inactivating the protein[70]. However, not all proteins can be destabilised in such ways, and as such not all essential genes can be identified using this method. This type of temperature-sensitive protein destabilisation is commonly found in temperature-sensitive replicons, responsible for the plasmids origin of replication[71].

Another type of temperature-sensitive disruption involves using the group II intron of the TargeTron/ClosTron system[72]. It is different from the aforementioned temperature-sensitive mutations as the temperature sensitivity and disruption is not present within the same protein, rather the protein responsible for the disruption of a target gene is temperature-sensitive. In principle, the ClosTron system can be used for both unconditional and conditional disruptions, including conditional disruptions of essential genes. This is because the group II intron can be inserted into either the sense or antisense orientation relative to the target gene, depending on the group II intron targeting sequence[38]. A group II intron inserted into the antisense strand is transcribed into mRNA by the target genes natural promoter, this mRNA encodes the reverse complement of the group II intron, removing the recognition sequences of the IEP from the mRNA[72]. This reverse group II intron cannot be spliced out again, yielding an unconditional mutation. However, if the group II intron is inserted in the sense orientation within the target gene, the target genes promoter is able to transcribe not only the target gene, but also a functional group II intron[72]. This functional group II

intron can then be spliced out of the target genes mRNA by the IEP, thereby restoring the native mRNA. As the conditional mutation is controlled by the IEP splicing reaction, controlling of the IEP splicing reaction will result in conditional disruptions if the group II intron is inserted in the sense orientation.

During Yao *et al*'s, (2006)[72], study of the group II intron, they identified that the IEP splicing reaction is inherently temperature-sensitive, consequently the TargeTron was able to splice out the group II intron from the target gene at 37°C but not at 43°C[72]. Using this, the authors were able to identify essential genes in *S. aureus* by obtaining viable cells with an insertion of the group II intron in essential genes at 37°C, due to the subsequent splicing of the group II intron from the mRNA, but no viable cells were obtained when the temperature was increased to 42°C, meaning the splicing reaction could not occur resulting in cell death[72]. Although the exact temperature-sensitive mechanism is unknown, it was hypothesised that the inhibition of the RNA splicing was due to the temperature sensitivity of the intron RNA, the IEP, or a combination of them both[64,72]. Employing the temperature sensitive nature of the Group II intron present within the ClosTron in such a way as shown by Yao *et al*, (2006)[72]should allow the identification of essential genes in *Clostridium*.

1.2.7 Partial diploids

A partial diploid is the definition of a bacterial cell which contains a second copy of a particular region of the chromosome. This can be in the form of a complete operon or a single gene, present upon a plasmid or in the chromosome[50,73]. Partial diploids can be useful tools to directly or indirectly identify essential genes when combined with other mutational methods such as; inducible promoters, homologous recombination or the ClosTron[50,67]. In essence, an essential gene cannot be inactivated due to the gene's

product being required for survival; however, if there is a second copy of the essential gene present within the cell, this second copy can functionally complement the native gene's product thereby negating the native gene's essentiality. This substitution of essential function enables the native essential gene to be inactivated using an appropriate method only if the second copy is present.

There are two ways to produce partial diploid *C. difficile* strains, either by supplying a second copy of the candidate essential gene on a plasmid, or by inserting the candidate essential gene into the chromosome using an Allele Chromosomal Exchange (ACE) system or classical homologous recombination. Once the second copy is present, the native gene can be inactivated. Essentiality of the candidate essential gene can be determined by the ability to inactivate the candidate gene in the original strain compared to the partial diploid strain[50,67]. If the gene is not essential, then the candidate gene should be able to be inactivated in both strains, however, if the gene is essential, then only the target gene present within the partial diploid should be able to be inactivated as the inactivation leads to the cell being non-viable in the original strain[50]. Such a system provides indirect evidence to the essentiality of the gene, compared to the direct methods of conditional expression as it is the study of the target genes ability to be inactivated in a partial diploid background.

1.2.8 Transposon mutagenesis

Transposons are naturally occurring elements which are able to transpose themselves from one position within the DNA of a cell to another. There are two different mechanisms in which transposons can move, either by "copy and paste" (Class I) or "cut and paste" (Class II), meaning the DNA is replicated and inserted into a new region or the original DNA element is cut and transposed into a new region of DNA,

respectively[74]. The simplest transposon is a segment of DNA which is flanked by short recognition sequences of DNA, normally inverted repeats. These short recognition sequences are recognised by the transposase enzyme which is able to enact the specific mechanism for that transposase[74].

Transposons are valuable tools for large-scale forward genetic experiments to identify specific genotypes associated with phenotypes or essential genes. Due to the fact that the transposon can insert into DNA sequences, insertion within a coding region can cause gene disruption. If the insertion is within a non-essential gene, the cells viability is not affected, if however the transposon inserts into an essential gene, the genotype is lethal and the cell is non-viable. Using this principle, it is possible to carry out a large-scale experiment to attempt to insertionally inactivate every gene within a genome, thereby creating a mutant library. The genes which cannot be inactivated are candidate essential genes. Such large-scale transposon mutagenesis has been carried out in many bacterial species including; *E. coli*[43], *Mycobacterium tuberculosis*[75], *Pseudomonas aeruginosa*[76], *H. pylori*[54], *Salmonella enterica* serovar Typhi (S. Typhi)[77], *H. infulenzae*[78], and *S. aureus*[79], to name a few.

One of the most interesting studies was carried out by Langridge *et al*, (2009)[77], in which they developed a technique called Transposon Directed Insertion-site Sequencing (TraDIS) to investigate the minimum essential gene set of the bacterium S. Typhi under both standard laboratory and relevant biological conditions[77]. The authors used a derivative of the Tn*5* to create a transposon mutant library of S. Typhi comprising of an estimated 1.1 million individual mutants encompassing 370,000 unique transposon insertion sites, equating to an average of one insertion every 13 bp or 80 unique insertions per gene[77]. From this mutant library, they were able to define an insertion

index based upon the number of insertions within a gene divided by the length of the gene, then using this insertion index to compile a frequency distribution for all annotated genes in the genome[77]. The frequency distribution of insertions for S. Typhi gave clear bimodal distribution results accounting for two types of genes, the first (left-most, figure 7A) peak represents the genes in which an insertion significantly inhibits cellular growth which leads to the absence or severely reduced insertion frequency in these genes (figure 7A)[77]. The second peak (right-most) represents the genes in which an insertion does not significantly affect cellular growth, therefore the transposon can insert with high frequency[77]. From this distribution, they calculated the likelihood of the genes being essential (figure 7B). Using this principle, the authors identified 356 genes which have a likelihood of being essential for growth under standard laboratory conditions[77]. In contrast, they also identified 4162 genes which had low likelihood of being essential, and 19 which could not be placed in either category using their likelihood ratio criteria[77].

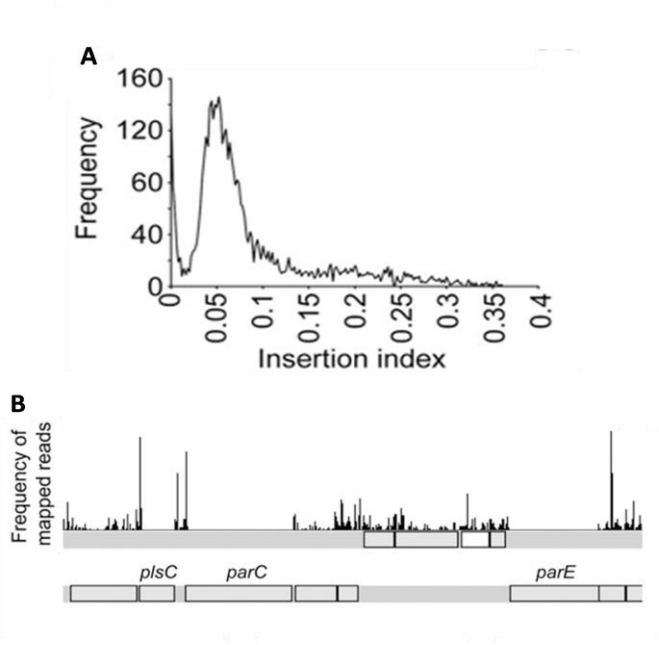


Figure 1.7. TraDIS to identify essential genes in S. Typhi. (**A**) Bimodal frequency distribution of the insertion index (number of inserts per gene

divided by the gene length). The left most peak represent genes which cannot, or have a very low frequency of insertions. The right hand peak (0.05) represents genes which can be readily inserted, i.e. not essential. (**B**) Detailed section of a genome map showing the frequency of insertions for a number of genes. The 3 genes labelled (*plsC*, *parC*, and *parE*) show an absence of insertions, indicating these genes are possibly essential in S. Typhi. Adapted from reference[77].

A considerable advantage of using TraDIS is the ability to pool each individual mutant and subject the pool to varying conditions or animal models. This allows the simultaneous identification of every candidate essential gene within the pool for each condition and environment by identifying the presence/absence of the mutant in the output pool. This difference in pools under these varying conditions would identify genes which may be identified as "advantageous" for growth, i.e., essential under specific conditions without being absolutely essential[77]. There are a few disadvantageous of using such methods to identify candidate essential genes. Transposon insertions within operons can cause a polar effect downstream of the insertion site, in such cases; an insertion within a nonessential gene upstream may disrupt the expression of an essential gene downstream resulting in a lethal genotype. This scenario would mean that a nonessential gene could be identified as essential due to the low possibility of insertions, when in fact it is due to the disrupted expression of another gene blocking downstream transcription. Therefore, analysing genes within operons must be done with care to rule out possible interference. Also most importantly, this method is only able to identify candidate essential genes, to prove essentiality this method must be complemented with indirect or direct methods such as partial diploids or conditional expression.

1.2.9 Transposon conditional mutagenesis

Transposons can be used to create conditional mutations as well as the classical unconditional mutation. The transposon is able to "carry" cargo which is transposed by the transposase, as long as it is within the flanking recognition sites. To create conditional mutations, an inducible promoter system can be placed within the transposon in such a position that it faces outwards of the transposon (figure 1.8A). If the transposon inserts into the promoter region of a gene in the sense orientation, the inducible promoter will replace the native promoter allowing conditional expression via the presence of an inducer (figure 1.8B). If the gene whose promoter has been replaced is essential, cellular viability is dependent on the inducer to activate the genes transcription. In this instance, the gene can be directly defined as essential as the presence or absence of the inducer will directly correlate with cellular viability. This method has been used to identify essential genes in *Vibrio cholera*[80], *Salmonella enterica* serovar Enteriditis (S. Enteriditis) [81], and *E. coli*[44].

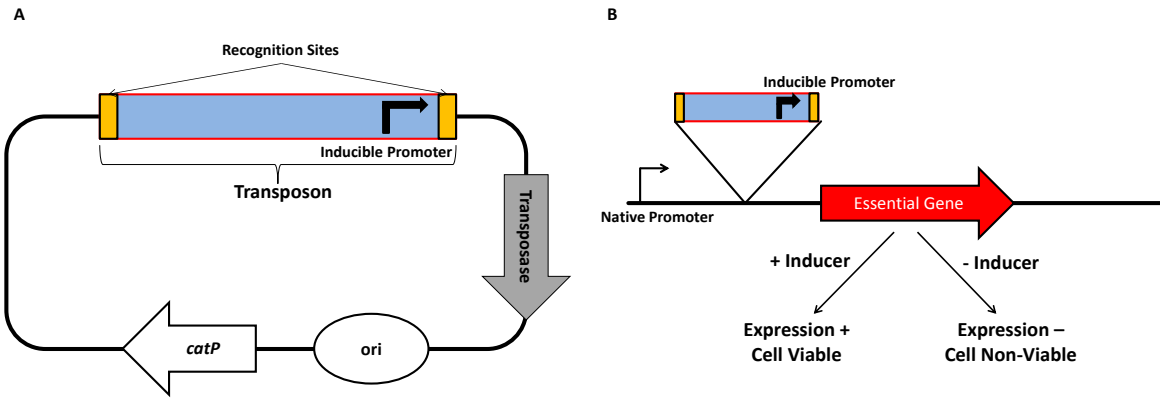


Figure 1.8. Conditional lethal mutation of an essential gene by the transposition of a transposon with outward facing promoters. (**A**) Transposon is present upon a vector containing a transposase to facilitate transposition, as well as *catP* for selection and an origin of replication. (**B**) Following transposition, the transposon moves into the chromosome of the host. If the transposon inserts between an essential

gene (red) and its native promoter, cellular viability is dependent upon the presence of the inducer. If no inducer is present, the cell dies. Once essentiality is identified, DNA sequencing of the transposon's position can be determined with outward facing primers present within the transposon.

In 2000, Judson and Mekalanos[80] developed the TnAraOut system, a transposon-based approach to identify and characterise essential genes in *V. cholera*[80]. The authors constructed a mariner-based transposon with the arabinose-inducible promoter, P_{BAD}, located at the 3' of the transposon facing-outwards[80]. During the random insertions of the transposon, if the transposons inducible promoter transcriptionally fuses with an essential gene, the insertionally mutagenesis should display arabinose dependence. Using this principle, the authors were able to identify 16 mutations which showed varying degrees of arabinose dependence[80]. Of the 16, 4 were shown to carry insertions upstream of known essential genes, whereas the other 12 carried insertions in previously known or hypothetical genes not previously identified as essential[80]. The advantage of using such a system is the ability to identify genes which may be unique to the organism, as the authors showed one of the 16 was not found in another bacterial species, which may represent an example of a narrow-spectrum drug target[80]. Such a system may be developed for use in *C. difficile* as the mariner-based transposon system is already established[42], by combining it with an inducible promoter such as the Tet system; it would provide a powerful tool for large-scale essential gene identification in *C. difficile*.

1.3 Other Relevant Clostridia

Most species of *Clostridium* are saprophytic organisms found ubiquitously in the environment. Certain members may be employed on an industrial scale for the production of biofuels, Botox, or Clostridial-Directed Enzyme Prodrug Therapy

(CDEPT). Commercially relevant *Clostridium* include; *Clostridium theromcellum*, *Clostridium acetobutylicum*, *Clostridium botulinum*, *Clostridium ljungdahlii*, and *Clostridium sporogenes*. However, the genus is most famous for those members that are able to cause disease in humans and domestic animals. Such members include; *Clostridium difficile*, *Clostridium botulinum*, *Clostridium perfringens*, *Clostridium tetani*, and to a lesser extent, *Clostridium sordellii*. All these species of *Clostridium* encounter the same limitation of genetic tools to genetically manipulate and study as *C. difficile*[38]. With the development of new tools become available, the ability to utilize the organisms or develop effective counter measures is becoming possible. Ideally, tools which can be developed for *C. difficile*, will also work in other *Clostridium* for analysis of specific genes or identification of essential genes when relevant.

1.3.1 Clostridium sporogenes

C. sporogenes is a commercially relevant organism due to its potential use in the treatment of cancer using CDEPT, and in canning industry for quality control purposes, as well as medically important as a surrogate for *C. botulinum*. *C. sporogenes* and *C. botulinum* are both genetically related at the species level[82]. Whereas *C. botulinum* is notorious for producing the most potent neurotoxin known, *C. sporogenes* is a non-toxigenic bacterium. The similarities in the strains allow *C. sporogenes* to be used as a surrogate in the study of *C. botulinum*, due to its non-toxic nature, in the laboratory as well as quality control against *C. botulinum* contamination in the food industry[83]. This similarity makes *C. sporogenes* an attractive organism to develop genetic tools for, as they should also work in *C. botulinum*.

As well as being related to *C. botulinum*, *C. sporogenes* is important in its own right due to current research into cancer treatment using CDEPT. In essence, solid cancerous

tumours grow very rapidly often resulting in inadequate blood and oxygen supplies within the tumour[84]. These hypoxic regions are the perfect environment for proliferation of *Clostridium*. If *Clostridium* spores are intravenously administered, they are able to find and colonise, leading to subsequent lysis of the tumours[84]. The key limitation to using *Clostridium* for lysing tumours is the ability of the cancerous cells at the oxygenated periphery to avoid the fate of the cells in the hypoxic core. The lysis of the hypoxic core not only kills the cancerous cells, but also the environment in which *Clostridium* can survive meaning the bacterial cells die[84]. This allows any surviving cells to multiply, replenishing the lost cells. One strategy to overcome this limitation is to engineer *C. sporogenes* to convert a prodrug into an active drug within the tumour, effectively killing all the cancerous cells[84]. Development is underway to engineer *C. sporogenes* for CDEPT; however there are certain aspects that require novel genetic tools, such as a *Clostridium* inducible promoter system, to be able to be used effectively as a cancer treatment.

1.3.2 *Solventogenic* Clostridium acetobutylicum *and* Clostridium beijerinckii

C. acetobutylicum was first described between 1912 and 1914 for their use in the industrial starch-based acetone, butanol, and ethanol (ABE) fermentation process to produce acetone for gunpowder production during World War I[85]. During the 1920s the increase in demand for butanol led to the construction of fermentation plants capable of using processing molasses using the ABE process. However this interest decreased by the 1950s as the production of the chemicals via petrochemical production became a more cost-effective process. As the oil prices rose during the 1970s, there was a renewed interest in the ABE process, leading to the genetic manipulation of the organism to increase yield and purity of the solvents, as well as increasing the range of produced chemicals[85]. The advancement in genetic manipulation over the past two decades has

accelerated the metabolic engineering for novel biologically derived chemicals, combined with the need for renewable energy sources, *C. acetobutylicum* has become one of the most commercially relevant and studied solventogenic organism.

C. beijerinckii (a new species, formally part of C. *acetobutylicum*) is most well known for commercialisation in the same way as *C. acetobutylicum*, in that it is able to metabolise a range of substrates including (but not limited to), starch, glucitol, galactose, and the disaccharides cellobiose, lactose, maltose, pentose, hexose, and sucrose to produce acetone, butanol, and ethanol via the ABE fermentation process[86,87]. However, the big advantage to using *C. beijerinckii* is its ability to be genetically modified with relative ease. With the isolation of hyper-butanol producing strains and the ability to genetically modify *C. beijerinckii* to utilize many substrates and produce a variety of novel solvents, there is great potential of engineering such an organism to produce solvents for attractive renewable biofuels which is economically cheaper and cleaner than petrochemical processes[88].

1.4 Aim of This Project

C. difficile remains the main causative agent of nosocomial diarrhoea in the world. The past two decades have seen an increase in incidence of CDI in the hospital, community and animals used for food. In addition, emergence of hypervirulent PCR-ribotypes associated with; frequent outbreaks, more severe disease, increased mortality rate, higher relapse rate, and increased antimicrobial resistance, has compounded the need to understand this prominent pathogen, in the hopes to treat CDI effectively.

The identification of essential genes for the development of effective novel therapeutic antimicrobials is more important now than ever. Essential genes are of particular importance due to their gene products being required for cellular viability. For an antimicrobial to be effective, it must be able to disrupt this essential gene product, leading the bacterial cell to become non-viable. It is this loss in viability which defines a gene as essential, and a chemical as an antimicrobial. In other bacterial pathogens, there have been countless studies to identify essential genes. However, in *C. difficile*, identification has largely been hindered by the limitation of current techniques. If this limitation was alleviated through the development of a technique capable of identifying essential genes, one hopes that it would lead the way for novel essential gene identification and subsequent antimicrobial development to treat CDI.

The aim of this project was to develop novel genetic tools able to identify and study essential genes in *C. difficile*. Specifically, candidate essential genes were identified using *in silico* approaches which were then confirmed using the indirect experimental approach of partial diploids combined with the ClosTron system. In addition to this, inducible expression systems were developed to directly identify essential genes and modulate expression of target genes in not only *C. difficile*, but other *Clostridium*.

Chapter Two

Materials and Methods

2.1 Materials

Unless stated otherwise, all chemicals and biochemical were supplied by Sigma-Aldrich or Fisher Scientific, enzymes and their buffers used for the purpose of molecular biology were purchased from New England Biolabs (NEB) or Promega, chemicals used in the purification of proteins were purchased from Invitrogen or GE Healthcare Life Sciences, and bacterial growth media was purchased from Oxoid.

Synthetic DNA less than 100 bp were purchased from Eurofins MWG Operon, DNA over 100 bp was purchased from DNA2.0, Mr Gene, or Eurofins MWG Operon.

2.2 List of Bacterial Strains

Strain	Genotype/Properties	Source/Reference
Escherichia coli TOP10	F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Invitrogen
Escherichia coli CA434	Conjugation donor	Laboratory Collection
Escherichia coli C600RK2	F- *tonA21 thi-1 thr-1 leuB6 lacY1 glnV44 rfbC1 fhuA1 λ-*	Laboratory Collection
Escherichia coli K12MG1655	F- *λ- ilvG- rfb-50 rph-1*	Kim. H. Laboratory Collection, Nottingham
Clostridium difficile 630Δ*erm*	PCR-ribotype 012 Erythromycin sensitive strain of *C. difficile*630	Laboratory Collection
Clostridium difficile R20291	BI/NAP1/027 (Stoke Mandeville, UK). Isolated in 2004/5	Anaerobe Reference Laboratory, Cardiff

Clostridium acetobutylicum ATCC 824	Wild-type isolate	Laboratory collection
Clostridium sporogenes NCIMB 10696	Wild-type isolate	Laboratory collection
Clostridium beijerinckii NCIMB 8052	Wild-type isolate	Laboratory collection
Chromosomal Functional Complementation		**Chapter Three**
Clostridium difficile 630Δ*erm pyrE*:: *dhaT*p	ACE *dhaT*p (protected) partial diploid insertion in *pyrE* locus (*pyrE*-)	This work
Clostridium difficile 630Δ*erm pyrE*::*metK*p	ACE *metK*p (protected) partial diploid insertion in *pyrE* locus (*pyrE*-)	This work
Clostridium difficile 630Δ*erm pyrE* ::*trpS*p	ACE *trpS*p (protected) partial diploid insertion in *pyrE* locus (*pyrE*-)	This work
Clostridium difficile 630Δ*erm pyrE*::*dhaTp dhaT*::CT-548a	CT insertion in native *dhaT* at position 548 in strain *Clostridium difficile* 630Δ*erm pyrE*:: *dhaT*p	This work
Clostridium difficile 630Δ*erm pyrE*::*dhaTp dhaT*::CT-623a	CT insertion in native *dhaT* at position 623 in strain *Clostridium difficile* 630Δ*erm pyrE*:: *dhaT*p	This work
Clostridium difficile 630Δ*erm pyrE*::*metK*p *metK*::CT-548s	CT insertion in native *metK* at position 548 in strain *Clostridium difficile* 630Δ*erm pyrE*:: *metK*p	This work
Clostridium difficile 630Δ*erm pyrE*::*metK*p *metK*::CT-676s	CT insertion in native *metK* at position 676 in strain *Clostridium difficile* 630Δ*erm pyrE*:: *metK*p	This work
Clostridium difficile 630Δ*erm pyrE* ::*trpS*p *trpS*::CT-436a	CT insertion in native *trpS* at position 436 in strain *Clostridium difficile* 630Δ*erm pyrE*:: *trpS*p	This work

Clostridium difficile 630Δ*erm pyrE* ::*trpS*p *trpS*::CT-460a	CT insertion in native *trpS* at position 460 in strain *Clostridium difficile* 630Δ*erm pyrE:: trpS*p	This work
Inducible Chromosomal Functional Complementation		**Chapter Five**
Clostridium difficile 630Δ*erm* $P_{fdx\text{-}tetO}$:*metK*p::*pyrE*	Tetracycline inducible promoter $P_{fdx\text{-}tetO}$ controlling expression of protected *metK* inserted into *pyrE* locus using ACE (*pyrE*$^{-}$)	This Work
Clostridium difficile 630Δ*erm* $P_{fdx\text{-}tetO}$:*slpA*p::*pyrE*	Tetracycline inducible promoter $P_{fdx\text{-}tetO}$ controlling expression of protected *slpA* inserted into *pyrE* locus using ACE (*pyrE*$^{-}$)	This Work
Clostridium difficile 630Δ*erm* $P_{fdx\text{-}tetO}$:*trpS*p::*pyrE*	Tetracycline inducible promoter $P_{fdx\text{-}tetO}$ controlling expression of protected *trpS* inserted into *pyrE* locus using ACE (*pyrE*$^{-}$)	This Work
Over production strains		
Escherichia coli BL21*(DE3)	F– *ompT gal dcm lon hsdS*$_B$($r_B^- m_B^-$) *λ(*DE3 *[lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	Invitrogen
Escherichia coli C600	F$^-$ *tonA21 thi-1 thr-1 leuB6 lacY1 glnV44 rfbC1 fhuA1 λ*$^-$	Invitrogen
Escherichia coli Rosetta 2	F$^-$ *ompT hsdS*$_B$($R_B^- m_B^-$) *gal dcm λ(*DE3 *[lacI lacUV5-T7 gene 1 ind1 sam7 nin5])* pLysSRARE (Cam^R)	Novagen

2.3 List of Plasmids

Plasmid	Relevant properties	Source
ClosTron Plasmids		**Chapters Three and Five**
pMTL007C-E2	ClosTron Plasmid	Reference [38]
pMTL007C-E2::Cdi-*dhaT*-213s	pMTL007C-E2 retargeted to Cdi-*dhaT*-213	This work
pMTL007C-E2::Cdi-*dhaT*-548a	pMTL007C-E2 retargeted to Cdi-*dhaT*-548a	This work
pMTL007C-E2::Cdi-*dhaT*-623a	pMTL007C-E2 retargeted to Cdi-*dhaT*-623a	This work
pMTL007C-E2::Cdi-*trpS*-436a	pMTL007C-E2 retargeted to Cdi-*trpS*-436a	This work
pMTL007C-E2::Cdi-*trpS*-460a	pMTL007C-E2 retargeted to Cdi-*trpS*-460a	This work
pMTL007C-E2::Cdi-*metK*-548s	pMTL007C-E2 retargeted to Cdi-*metK*-548s	This work
pMTL007C-E2::Cdi-*metK*-676s	pMTL007C-E2 retargeted to Cdi-*metK*-676s	This work
pMTL007C-E2::Cdi-*slpA*-878a	pMTL007C-E2 retargeted to Cdi-*slpA*-878a	This work
pMTL007C-E2::Cdi-*slpA*-1267s	pMTL007C-E2 retargeted to Cdi-*slpA*-1267s	This work
Modular Plasmids		**All Results Chapters**
pMTL82151	Modular plasmid	Reference [89]
pMTL82254	Modular plasmid including *catP* reporter	Reference [89]

pMTL82151-*dhaT*	pMTL8215x containing *Clostridium difficile* 630Δ*ermdhaT*	This work
Plasmid Complementation		**Chapter Three**
pMTL82151-P*dhaT*-*dhaT*	pMTL8215x containing *Clostridium difficile* 630Δ*erm dhaT* and its native promoter	This work
DNA2.0-*dhaT*-synthfrag	DNA2.0 in-house vector containing protected *Clostridium difficile* 630Δ*erm dhaT* fragment	This work
pMTL-DW2	pMTLDW1 with a protected copy of *Clostridium difficile* 630Δ*erm dhaT* from DNA2.0-*dhaT*-synthfrag	This work
pMTL-DW3	pMTLDW2 derived except *aad9* has replaced *catP* resistance cassette	This work
Chromosomal Complementation		**Chapter Three**
DNA2.0-Pmet-*metK*p	DNA2.0 in house vector containing a protected *Clostridium difficile* 630Δ*erm metK* and its native promoter	This work
DNA2.0-Pthl-*trpS*p	DNA2.0 in house vector containing a protected *Clostridium difficile* 630Δ*erm metK* and the *Clostridium sporogenes* thiolase promoter	This work
pMTL-JH18	ACE plasmid	Heap J.T. Laboratory collection
pMTL-JH18-*dhaT*p	pMTL-JH18 containing insert of pMTL-DW1	This work
pMTL-JH18-*metK*p	pMTL-JH18 containing insert of DNA2.0-P_{metK}-*metK*p	This work
pMTL-JH18-*trpS*p	pMTL-JH18 containing insert of DNA2.0-P_{thl}-*trpS*p	This work
Inducible Chromosomal Functional Complementation		**Chapter Five**
DNA2.0-P_{slpA}-*slpA*p	DNA2.0 in house vector containing a protected *Clostridium difficile* 630Δ*erm slpA* and its native promoter	This Work

pMTL-tet3nO	Tetracycline inducible promoter system containing P_{thl} running *tet*R, P_{fdx} interrupted with a single *tetO*, cloned into pMTL82254	This Work
pMTL-ACEtet-*dhaT*p	Protected essential gene *dhaT* under tetracycline inducible control in vector pMTL-JH18.	This work
pMTL-ACEtet-*trpS*p	Protected essential gene *trpS* under tetracycline inducible control in vector pMTL-JH18.	This work
pMTL-ACEtet-*slpA*p	Protected candidate essential gene *slpA* under tetracycline inducible control in vector pMTL-JH18.	This work
***lac* repressor system Plasmids**		**Chapter Four**
pMTL-DW17	IPTG inducible promoter system containing P_{trnA-1} running *lacI*, P_{fdx} interrupted with single *lacO* cloned into pMTL82254	This work
pMTL-DW17(b)	IPTG inducible promoter system containing P_{trnA-1} running *lacI*, P_{fdx} interrupted with double *lacO* cloned into pMTL82254	This work
pMTL-DW18	IPTG inducible promoter system containing P_{trnA-1} running *lacI* and *ermB*, P_{fdx} interrupted with single *lacO* cloned into pMTL82254	This work
pMTL-DW18(b)	IPTG inducible promoter system containing P_{trnA-1} running *lacI* and *ermB*, P_{fdx} interrupted with double *lacO* cloned into pMTL82254	This work
pMTL-DW18(1)	pMTL-DW18(b) with P_{fdx} from *C. sporogenes* running *lacI*	This work
pMTL-DW18(2)	pMTL-DW18(b) with P_{hydA} from *C. acetobutylicum* running *lacI*	This work
pMTL-DW18(3)	pMTL-DW18(b) with P_{thl} from *C. sporogenes* running *lacI*	This work
pMTL-DW18(4)	pMTL-DW18(b) with P_{trnA} from *C. acetobutylicum* running *lacI*	This work
pMTL-DW18(5)	pMTL-DW18(b) with P_{DNAk} from *C. acetobutylicum* running *lacI*	This work
pMTL-DW18(6)	pMTL-DW18(b) with P_{recA} from *C. acetobutylicum* running *lacI*	This work

pMTL-DW18(7)	pMTL-DW18(b) with P_{gyrA} from *C. acetobutylicum* running *lacI*	This work
pMTL-DW18(8)	pMTL-DW18(b) with P_{adc} from *C. acetobutylicum* running *lacI*	This work
pMTL-DW18(9)	pMTL-DW18(b) with P_{ptb} from *C. acetobutylicum* running *lacI*	This work
pMTL-DW18(10)	pMTL-DW18(b) with P_{AT0200} from *C. acetobutylicum* running *lacI*	This work
pMTL-DW18(0)	pMTL-DW18(b) with no promoter running *lacI*	This work
Riboswitches		**Chapter Four**
pMTL-Thio1	Thio1 riboswitch cloned into pMTL82254	This work
pMTL-Thio2	Thio2 riboswitch cloned into pMTL82254	This work
pMTL-Thio3	Thio3 riboswitch cloned into pMTL82254	This work
pMTL-V1P1	pMTL-Thio1 containing P_{fdx}	This work
pMTL-V1P2	pMTL-Thio1 containing P_{hydA}	This work
pMTL-V1P3n	pMTL-Thio1 containing short P_{thl}	This work
pMTL-V1P5	pMTL-Thio1 containing P_{DNAk}	This work
pMTL-V1P7	pMTL-Thio1 containing P_{gyrA}	This work
pMTL-V1P8	pMTL-Thio1 containing P_{adc}	This work
pMTL-V1P9	pMTL-Thio1 containing P_{ptb}	This work

pMTL-V1P10	pMTL-Thio1 containing P_{AT200}	This work
pMTL-V2P3n	pMTL-Thio2 containing short P_{thl}	This work
pMTL-V2P3n	pMTL-Thio3 containing short P_{thl}	This work
Tet system plasmids		**Chapter Four**
pMTL-*tetO*	Tet system containing no *tet*R promoter, and single *tetO* interrupting P_{fdx}, present in pMTL82254	This work
pMTL-*tetO*1/2	Tet system containing no *tet*R promoter, and double *tetO* interrupting P_{fdx}, present in pMTL82254	This work
pMTL-tet3nO	Tet system containing short P_{thl} running *tet*R, and single *tetO* interrupting P_{fdx}, present in pMTL82254	This work
pMTL-tet3n	Tet system containing short P_{thl} running *tet*R, and no P_{thl} or *tetO*s, present in pMTL82254	This work
Over Expression Plasmids		**T.B.C**
pET-17b	T7 expression vector	Novagen
pET-14b-βgal	Induction control A	Novagen
pET-17b-*dhaT*	*dhaT* over production vector	This work

2.4 List of Oligonucleotides

Oligonucleotide	Sequence 5'-3'
ClosTron Re-targeting	
Cdi-*dhaT*-213s-IBS	AAAAAAGCTTATAATTATCCTTAGAACACAATCCTGTGCGCCCAGATAGGGTG
Cdi-*dhaT*-213s-EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCAATCCTCTTAACTTACCTTTCTTTGT
Cdi-*dhaT*-213s-EBS2	TGAACGCAAGTTTCTAATTTCGATTTGTTCTCGATAGAGGAAAGTGTCT
Cdi-*dhaT*-F3	GAAATAGGAGGTTTTAGAATGAATTTTAATTATAATTTGCC
Cdi-*dhaT*-R3	TAAATATGTAAAGTCAAATTATAAAGACTTATGGTACATTTCC
EBS Universal	
Cdi-*dhaT*-548a-IBS	AAAAAAGCTTATAATTATCCTTAATTATCTGTTTTGTGCGCCCAGATAGGGTG
Cdi-*dhaT*-548a-EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCTGTTTTGGTAACTTACCTTTCTTTGT
Cdi-*dhaT*-548a-EBS2	TGAACGCAAGTTTCTAATTTCGGTTATAATCCGATAGAGGAAAGTGTCT
Cdi-*dhaT*-623a-IBS	AAAAAAGCTTATAATTATCCTTATCTGCCAGAGGGGTGCGCCCAGATAGGGTG

Cdi-*dhaT*-623a-EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCAGAGGGTTTAACTTACCTTTCTTTGT
Cdi-*dhaT*-623a-EBS2	TGAACGCAAGTTTCTAATTTCGGTTGCAGATCGATAGAGGAAAGTGTCT
ClosTron Sequencing	
Spo*fdx*_F1	GATGTAGATAGGATAATAGAATCCATAGAAAATATAGG
pMTL007_R1	AGGGTATCCCCAGTTAGTGTTAAGTCTTGG
Partial Diploid Construction	
dhaT-ORF-F3+RS	GATCGACATATGAATTTTAATTATAATTTGCCAGTAAATATCTTATTTGGGAGAGGC
dhaT-ORF-H6-R3+2RS	GTATAAGGATCCGAATTCTTATTAGTGATGGTGATGGTGATGTAAAGACTTATGGTACATTTCCTTTATTTGTTCCTTATTAAAATACTTTGG
Cdi-*dhaT*-pro-F1	AATATCGCGGCCGCTCTAATAAAAAACTCTATTTAGTTAAATATAGTTAACTTG
Cdi-*dhaT*-pro-R1	AGTAACCATATGAAAACCTCCTATTTCTACTTTAGTAATAATTTTATGTTTTTAACAAAGTG
Cdi630-*dhaT*-R1	CAGCATCCATAATGCTTCCA
Partial Diploid Screening	

Cdi630-CD0189-R2	GGTTCTGGAACCAGAGATTATTTAG
Cdi630-*pyrE*-F2	GATGGGCGGAATAACTAAAGC
*dhaT*_in_F1	TGCAGAAGATTGTTCAGATGC
*dhaT*_in_F2	GTGCAGAAGATTGTTCAGATGC
dhaT-inScaI-F	TTTGATGCACTTGCCCATAA
dhaT-Sil-ScaI-F1	GACGTTCAGGCAATATACGG
Cdi-*trpS*-F1	CCC CTT TGA GAT GAG CAA AA
Cdi-*trpS*-R1	TAA GGC AGA TGG AAG TGC TG
Cdi-*trpS*-iF1	AAA GCA GCA TTT GGA ATT GG
Cdi-*trpS*-iR1	GAT GGC TGT GCA CCA CTA AA
Cdi-*metK*-F1	CAA TAG GAA CAC CCC CTT CA
Cdi-*metK*-R1	GCA ACA TAA GCT CCA GCA AA
Cdi-*metK*-iF1	AAA TAT TGT TGC GGC TGG TC
Cdi-*metK*-iR1	TTG GTC TCC AGC TCC AAC TT
Cdi630-CD0189-R2	GGTTCTGGAACCAGAGATTATTTAG

Cdi630-*pyrE*-F2	GATGGGCGGAATAACTAAAGC
dhaT-inScaI-F	TTTGATGCACTTGCCCATAA
dhaT-Sil-ScaI-F1	GACGTTCAGGCAATATACGG
Cdi-*trpS*-F1	CCCCTTTGAGATGAGCAAAA
Cdi-*trpS*-R1	TAAGGCAGATGGAAGTGCTG
Cdi-*trpS*-iF1	AAAGCAGCATTTGGAATTGG
Cdi-*trpS*-iR1	GATGGCTGTGCACCACTAAA
Cdi-*metK*-F1	CAATAGGAACACCCCCTTCA
Cdi-*metK*-R1	GCAACATAAGCTCCAGCAAA
Cdi-*metK*-iF1	AAATATTGTTGCGGCTGGTC
Cdi-*metK*-iR1	TTGGTCTCCAGCTCCAACTT

lac repressor system

Spo*Fdx*-NotI-F1	GCGGCCGCTATGATGATTATTTTGTAGAT
Spo*Fdx*-NdeI-R2	CATATGTATTTCCTCCTAAAAATTACACAACTTTA
CttRNA-NotI-F1	AAGACGCGGCCGCTTGAATATTAGTATTAGGGG

CttRNA-NdeI-R1	GATCACATATGTTTGGATCCGACTTTTACAG

Riboswitches

ThioF1	GCGGCCGCGAGCTCGGATCCGGTGATACCAGCATCGTATGGATGC CCTTGGCAGCACCCCGCTGCAAGACAACCATATG
ThioR1	CATATGGTTGTCTTGCAGCGGGGTGCTGCCAAGGGCATCCATACG ATGCTGGTATCACCGGATCCGAGCTCGCGGCCGC
ThioF2	GGCCGCGAGCTCGGATCCGGTGATACCAGCATCGTATGGATGCCC TTGGCAGCACCCCGCTGCAAGACAACCA
ThioR2	TATGGTTGTCTTGCAGCGGGGTGCTGCCAAGGGCATCCATACGAT GCTGGTATCACCGGATCCGAGCTCGC
TRS.s-on.F1	GGCCGCGAGCTCGGATCCGGTGATACCAGAGTATGCTCTACTCCC TTGGCAGCACCCCGCTGCAAGAAGAAGAGCA
TRS.s-on.R1	CGCTCGAGCCTAGGCCACTATGGTCTCATACGAGATGAGGGAACC GTCGTGGGGCGACGTTCTTCTTCTCGTAT
TRS.s-off.F1	GGCCGCGAGCTCGGATCCTAGACGAAGAGAACGTCGGCATACGAC GGTTCCCCTCCGTCTAGGAGGACCA
TRS.s-off.R1	CGCTCGAGCCTAGGATCTGCTTCTCTTGCAGCCGTATGCTGCCAAG GGGAGGCAGATCCTCCTGGTAT
Orig.TRS-RBS-F1	GGCCGCGAGCTCGGATCCCCGCTGCAAGACAACCA

Orig.TRS-RBS-R1	CGCTCGAGCCTAGGGGCGACGTTCTGTTGGTAT
TRS.s-on.F1-RBS	GGCCGCGAGCTCGGATCCCCGCTGCAAGAAGAAGAGCA
TRS.s-on.R1-RBS	TATGCTCTTCTTCTTGCAGCGGGGATCCGAGCTCGC
TRS.s-off.F1-RBS	GGCCGCGAGCTCGGATCCCCTCCGTCTAGGAGGACCA
TRS.s-off.R1-RBS	TATGGTCCTCCTAGACGGAGGGGATCCGAGCTCGC
TRS.s-on.R1	TATGCTCTTCTTCTTGCAGCGGGGTGCTGCCAAGGGAGTAGAGCA TACTCTGGTATCACCGGATCCGAGCTCGC
TRS.s-on.R2	TATGTGCTCCTCCTCTTGCAGCGGGGTGCTGCCAAGGGGGTGGAG CACACCCTGGTATCACCGGATCCGAGCTCGC
TRS.s-off.R1	TATGGTCCTCCTAGACGGAGGGGAACCGTCGTATGCCGACGTTCT CTTCGTCTAGGATCCGAGCTCGC

pET vector screening

pTF	CTCGAAAATAATAAAGGGAAAATCAG
pTR	TGGTAGTGTGGGGACTC
dhaT-F-Strep	TGATGCTACATATGAATTTTAATTATAATTTGCCAGTA
*dhaT*S-BamHI-R3	GTCATCGACGGATCCTTATTTTTCGAACTGCGGGTGGCTCCATAAA GACTTATGGTACATTTCCTTTATTTG

Promoter Library	
rnb-F1	GTCAGCGAGCTCAAAAATAAATTTTAAAAAAGTG
rnb-R1	CGCGGATCCCGACTTTTACAGTATACAAGTTT
fdx-F1	GTCAGCGAGCTCGTGTAGTAGCCTGTGAAATA
fdx-R1	CGCGGATCCTGTAACACACCTCCTTAAAAATTACACA
ptb-F1	GTCAGCGAGCTCTAAGTCAGCAGAAAGTATAATGAGA
ptb-R1	CGCGGATCCTTTACTATTTAATTATCTATGTTAAA
thl-F1	GTCAGCGAGCTCATTCTAAAAGAATTTAGAATGAA
thl-R1	CGCGGATCCAAATTTTGATACGGGGTAACAGATAA
hydA-F1	GTCAGCGAGCTCTATTTTTTAATTTAATTATAC
hydA-R1	CGCGGATCCAAAATGTAAAATATAATTAAAATATATTAATAAACTTCGT
adc-F1	GTCAGCGAGCTCGATTATTTATGGAGCTTGTA
adc-R1	CGCGGATCCTTAACATAAAAGTCACCTTCCT
pyrE-F1	GTCAGCGAGCTCCCAATAAAACCCAATCTTTTG
pyrE-R1	CGCGGATCCTCCTACTCCTTTACTTTTACT

DNAk-F1	GTCAGCGAGCTCATCTAATTCTATAATGTGTGAT
DNAk-R1	CGCGGATCCTATTAGCACTCACTCTCTTTGAGTGC
greA-F1	GTCAGCGAGCTCAAATTGTAATTAAAATTATGCACA
greA-R1	CGCGGATCCATTGTTTTTCTCCCCTCTACAAAAA
recA-F1	GTCAGCGAGCTCAATGATTTATTTTGCATAAAAT
recA-R1	CGCGGATCCATATATCCCTTAAGGGTTCACCA
CA0200-F1	GTCAGCGAGCTCAGTCTATATAAAGCTTTAT
CA0200-R1	CGCGGATCCAAACTATCTCCTTTACTTTATAATT
tRNA-F1	GTCAGCGAGCTCGAGAGTTAATGCCAAAGCTTTT
tRNA-R1	CGCGGATCCTCCTCCTTATATATATCTTTGGTATTGTTT

2.5 Bioinformatics Methods

2.5.1 Genome annotation

Bacterial genomes were visualised using Artemis available at http://www.sanger.ac.uk/.

2.5.2 BLAST

Genomic database searches to identify local similarity between sequences was carried out using the Basic Local Alignment Tool (BLAST) available at http://blast.ncbi.nlm.nih.gov/Blast.cgi, nucleotide queries were performed using the

blastn algorithm (nucleotide collection, somewhat similar sequence), translated nucleotide queries were carried out using blastx (non-redundant protein sequences), and protein database searches using blastp (non-redundant protein sequences).

Identification of conserved domains within a protein sequence was carried out using CDD, an algorithm that searches the Conserved Domain Database (CDD), found at http://www.ncbi.nlm.nih.gov/.

2.5.3 Sequence alignment

Alignment of two sequence was carried out using EMBOSS Needle available at http://www.ebi.ac.uk/

2.5.4 Sequence editor and plasmid map designer

Vector map design was carried out using combinations of Vector NTI and GENtle.

2.5.5 Codon editor and analyser

Synthetic genes were designed *in silico* using DNA2.0's Gene Designer found at https://www.dna20.com/genedesigner2/, and analysed for codon usage with Graphical Codon Usage Analyser available at http://gcua.schoedl.de/.

2.5.6 Sequence analysis for sequencing results

Construction and analysis of contigs from sequencing results was performed using the contig express feature of Vector NTI or DNA Baser.

2.5.7 Oligonucleotide design

Oligonucleotides for the purpose of PCR primers or synthetic DNA inserts were designed manually or using Primer3 available at http://frodo.wi.mit.edu/primer3/.

2.5.8 RNA folding prediction

Predictions on RNA folding were performed using *MFOLD* available at http://mfold.rna.albany.edu/.

2.6 Microbiological Materials and Methods

2.6.1 Aerobic bacterial strains and culture conditions

All *E. coli* strains were routinely grown in Luria-Bertani (LB) broth or agar at 37°C except for a few exceptions, with liquid cultures shaken at 200 rpm. For protein over production and when cloning genes thought to be toxic, *E. coli* strains were grown at 30°C. LB broth and agar was supplemented with chloramphenicol (25 μg/ml in agar or 12.5 μg/ml in broth), erythromycin (500 μg/ml), ampicillin (100 μg/ml), spectinomycin (250 μg/ml), Isopropyl β-D-1-thiogalactopyranoside (IPTG 1-3 mM), anhydrotetracycline (aTc 0.00031-1 μg/ml), theophylline (1 μg/ml- 20 μg/ml) or 5-bromo-4-chloro-3-indolyl-β-galactopyranoside (Xgal, 40 μg/ml) when appropriate. *E. coli* TOP10 was used throughout for the purpose of cloning, and performing inducible promoter testing. The conjugation donor strain CA434 was used for plasmid transfer into *Clostridium*. Protein over production was performed in *E. coli* TOP10, C600, Rosetta (DE3), and BL21*(DE3) strains where appropriate.

2.6.2 Anaerobic bacterial strains and culture conditions

All *Clostridium* strains were grown in an anaerobic cabinet at 37°C or in anaerobic gas jars at 30°C or 42°C where appropriate. *C. difficile* strains were routinely grown on BHIS broth or agar, *C. sporogenes* NCIMB 10696 was grown on TYG broth or agar, *C. beijerinckii* NCIMB 8052 was grown on CBM broth or agar, and *C. acetobutylicum* ATCC 824 was grown on CGM broth or agar supplemented with glycerol. All liquid media was

pre-reduced overnight, with agar pre-reduced for a minimum of 4 hours in anaerobic conditions before use. Media was supplemented as appropriate with *C. difficile* supplement (cycloserine 250 μg/ml and cefoxitine 8 μg/ml), cycloserine (250 μg/ml), thiamphenicol (15 μg/ml), erythromycin (2.5 μg/ml or 10 μg/ml), lincomycin (20 μg/ml), spectinomycin (250 μg/ml), or polymixin B (10 μg/ml). The *C. difficile* single-crossover ACE mutants were grown on *C. difficile* minimal media supplemented with 5-Fluoroorotic Acid (5-FOA) (1 mg/ml) and uracil (5 μg/ml). Double-crossover ACE mutants displaying the *pyrE*$^-$ genotype were grown on BHIS supplemented with 5 μg/ml uracil.

2.6.3 Preparation of electro-competent E. coli *cells*

To prepare *E. coli* for transformation by electroporation, a 500 ml conical flask containing 200 mls of LB pre-warmed to 37°C was inoculated with 1/100 volume (2 ml) of fresh *E. coli* overnight culture and shaken at 200 rpm. The cells were grown to an optical density of 0.5 at 600 nm (OD_{600}) representing exponential growth phase and then chilled on ice for 30 minutes. After the set time, the culture was separated into 50 ml falcon tubes and spun at 4000 xg for 20 minutes at 4°C to pellet the cells. The supernatant was poured off and the pellet re-suspended in an equal volume of ice-cold sterile dH_2O and pelleted under the same conditions as before. The supernatant was then discarded and each pellet was re-suspended in 0.6 ml of filter-sterilised 10% glycerol solution, all the samples were pooled into a single falcon and centrifuged again. The final pellet was re-suspended in 0.6 ml sterile 10% glycerol and stored in 50 μl aliquots at -80°C for use.

2.6.4 *Transformation by electroporation of electro-competent* E. coli *cells*

To transform electrocompetent *E. coli* cells, 50 µl aliquot of electro-competent cells was thawed on ice for 5-10 minutes. Plasmid DNA was then mixed with the cells gently and pipetted into a cold 0.2 cm gap electroporation cuvette and pulsed applied using an electroporator at 2.5 kV voltage, 25 µF capacitance and 200 Ω resistance. To recover the electroporated cells, 250 µl of Invitrogen SOC recovery media was added immediately and the cell mixture transferred into a fresh eppendorf tube to recover at 37°C, 200 rpm, for 1 hour. The recovery period is recommended to allow the expression of resistance cassettes present within the transformed cell before the cells are plated onto fresh LB, supplemented with appropriate antibiotics.

2.6.5 *Transfer of plasmid DNA into clostridia by conjugation*

Transfer of shuttle plasmids into *C. difficile*, *C. sporogenes*, and *C. beijerinckii* was carried out by conjugation as published[90]. For conjugation into the aforementioned clostridia, the donor *E. coli* strain CA434 was transformed with the required shuttle vector and left to grow overnight in liquid media in parallel with the recipient under anaerobic conditions. Next day, 1 ml of the donor cell culture was placed into an eppendorf and centrifuged at 4000 xg for 1 ml and resuspended with 0.6 ml sterile Phosphate Buffer Solution (PBS) to wash the cells of any antibiotic. The cells were pelleted again and transferred into an anaerobic cabinet to be mixed with a 5:1 ration of the recipient clostridial cells. The mixture were then plated out onto appropriate media, absent of any antibiotic, in discrete spots of approximately 200 µl and allowed to incubate for 4-24 hours for transfer of plasmids to take place. The conjugation slurry was then picked-up with a 10 µl inoculation loop, resuspended in 500 µl PBS, and plated onto fresh media containing antibiotic selection for the transferred plasmid, and counter-selection against the *E. coli* donor strain. The plates were incubated anaerobically at 37°C for 24-72

hours. Once the transconjugants were large enough to pick, they were purified by sub-culturing onto fresh media for further analysis.

2.6.6 Lysis of E. coli *cells by detergent*

E. coli cells were lysed using Novagen BugBuster Protein Extraction Reagent in accordance with the manufacturer's instructions.

2.6.7 Lysis of clostridial cells by sonification

Clostridial cell pellets were re-suspended in 1 ml of PBS supplemented with Roche Complete protease inhibitor cocktail tablets. The cells were then transferred into a 2 ml boiling eppendorf tube and chilled on ice for 5 minutes. Sonication was carried out on a Heat Systems, Ultrasonic W-280 sonicator. The samples were subjected to sonification by inserting the sonic probe into the eppendorf about half-way down the sample and pulsed for 1 second bursts for 45 seconds at a time at 40% power and power-output rating 3.5, and then allowed to rest for 30 seconds. This was repeated for 3-6 times until the lysates were visibly clear in comparison to the un-sonicated turbid cells. It is important that the samples were kept on ice at all times due to the heat accumulated during sonication having an adverse effect on the proteins in the sample. Once the samples were lysed, the lysates were centrifuged at 13,000 xg for 40 minutes at 4°C and the supernatants removed for further analysis.

2.7 Molecular Biological Materials and Methods

2.7.1 Plasmid DNA extraction and purification

Plasmid DNA was extracted and purified using Qiagen QIAprep Spin Miniprep kit in accordance with the manufacturer's instructions.

2.7.2 Extraction and purification of chromosomal DNA

Chromosomal DNA was extracted and purified using lysozyme and the Qiagen DNeasy Tissue Kit in accordance with the manufacturer's instructions. On occasions where high quality DNA was not required, i.e. PCR screening, a protocol involving chelex 100 was used. In brief, a 1 ml sample was centrifuged at 8000 xg for 10 minutes and the supernatant discarded. The pellet was re-suspended in 0.5 ml 5% chelex 100 and vortexed for 1 minute, the suspension was subjected to 100°C for 10 minutes and centrifuged at 12,000 xg for 10 minutes at 4°C. The resulting supernatant containing the desired chromosomal DNA was removed and stored at -20°C for further screening purposes.

2.7.3 Spectrophotometric quantification of DNA

To quantify the DNA present within a sample, 1.5 µl of sample was placed on a nanodrop ND-1000 spectrophotometer. The nanodrop measues the OD_{260} of the sample and compares it to 50 mg/ml having an OD_{260} of 1, at a path length of 1 cm. Therefore, a sample containing a DNA concentration of 10 mg/ml will have an OD_{260} of 0.2.

2.7.4 Amplification of DNA by Polymerase Chain Reaction (PCR)

Amplification of DNA for screening was carried out using Promega Taq DNA polymerase kit in accordance with the manufacturer's instructions. For cloning where PCR errors must be mimimalised, a high fidelity Taq was used. Fragments of 0-3 kbp, Epicentre Failsafe system in conjunction with Failsafe Premix E buffer, and for fragments over 3 kbp, Novagen KOD DNA polymerase system were used in accordance with the manufacturer's instructions.

2.7.5 Amplification of DNA by Splicing by Overlap Extension (SOEing) PCR

Splicing by Overlap Extension PCR was carried out to splice two DNA fragments together using appropriate primers, to amplify, and splice together the fragment using a either a single or two step reaction mixture. (Horton, *et al.,* 1989; Horton, *et al*., 1990).

2.7.6 Restriction endonuclease digestion of DNA

Complete or partial digests were carried out using NEB restriction enzymes in accordance with the manufacturer's instructions.

2.7.7 Blunting of DNA ends

For blunting DNA to facilitate the ligation of two incompatible DNA ends, NEB T4 polymerase was employed as per the manufacturer's instructions.

2.7.8 De-phosphoylation of DNA

To remove the 5' phosphate groups of digested DNA and limit the ability of the DNA to re-ligate to the undesired DNA end, NEB Antarctic Phosphate was used as per the manufacturer's instructions.

2.7.9 Ligation of DNA

Plasmid DNA fragments, which had been previously digested with restriction endonucleases, were ligated using NEB T4 ligase as per the manufacturer's instructions.

2.7.10 Membrane dialysis of DNA

The product of ligations were dialysed with a Millipore 0.025 μm nitrocellulose membrane over dH_2O for 30 minutes prior to electroporation to remove any unwanted salts from the solution that can arc the electroporator cuvette samples.

2.7.11 Agarose gel electrophoresis

Separation of DNA whether it be, PCR products, restriction fragments, or plasmid DNA, based upon size alone, was carried out in 1% agarose gel in TAE buffer containing 0.5 μg/ml ethidium bromide, electrophorised at 80 V for 30-60 minutes (depending on size of expected fragments). DNA was subsequently visualised under Ultra-Violet (UV light).

2.7.12 Extraction of DNA from agarose gels and reaction mixtures.

PCR products and restriction fragments were purified from their reaction mixtures using Qiagen QIAquick PCR Purification kit, or if the samples were subjected to agarose gel electrophoresis, the desired fragments were excised from the gel and underwent purification using the Qiagen QIAquick Gel Extraction Kit.

2.7.13 Annealing pairs of oligonucleotides

Primers which are complementary to each other were annealed together to generate synthetic DNA inserts with specified cohesive ends. These cohesive ends are designed into the primers so the desired over/underhang is present to use directly for ligating into a plasmid backbone. Each oligonucleotide pair at concentration 100 μM were added equally to a PCR tube and diluted with EB buffer to a ratio of 1:100 to yield 1 μM of each oligonucleotide. This mixture was then heated in a PCR theromcycler to 95°C for 30 seconds and allowed to cool to room temperature at a rate of 1°C per second. This reaction melts the two oligonucleotides together, allowing the resultant insert to be used directly in a ligation reaction.

2.7.14 Nucleotide sequencing

DNA sequencing was carried out by Geneservice Ltd, Nottingham or by The University of Nottingham Biopolymer Synthesis and Analysis Unit, Nottingham.

2.8 Biochemical Materials and Methods

2.8.1 Growth curves and time sampling of clostridial species

An overnight liquid culture of *Clostridium* species was used to inoculate fresh liquid media supplemented with appropriate antibiotics, and aTc if when applicable, to a final concentration of 0.05 OD_{600}. The culture was then grown in a static anaerobic environment at 37°C. Samples were taken for biochemical analysis at appropriate time points and the OD_{600} was measured on a photo spectrophotometer. More detailed protocols for inducible expression systems can be found in section 2.11.

2.8.2 Over-production of proteins in **E. coli**

An appropriate volume of media (normally 200 ml or 1 L) was inoculated at a ratio of 1:100 from an overnight culture of *E. coli* harbouring the appropriate expression vector, and incubated at 27°C, 30°C, or 37°C, shaking at 200 rpm. The cultures were harvested at the appropriate time after inoculation and centrifuged at 10,000 xg for 20 minutes. The supernatants were then discarded and the pellets stored at -20°C for up to 24 hours for analysis.

2.8.3 Poly-Acrylamide Gel Electrophoresis (PAGE)

Cell lysates and purified proteins were analysed by PAGE. Invitrogen NuPage LDS sample buffer and 5% β-mercaptoethanol was added to each sample and boiled at 95°C for 10 minutes to denature the proteins. The samples were then loaded onto Invitrogen 4-12% NuPage Bis-Tris gels in an appropriate PAGE gel tank, and Invitrogen NuPage MES Running Buffer was added, enough to cover the electrodes. The samples were then subjected to electrophoresis at 150 V for 90 minutes to separate the proteins, after which the gel was removed from the cast, washed three times in 100 ml of water for 5 minutes, stained with Invitrogen SimplyBlue SafeStain for one hour shaking slowly, and

finally de-stained in water for 1 hour shaking or static overnight. The resultant gel was then viewed under light or by a densitometer.

2.8.4 Densitometry

Gel scanning was carried out using a Biorad GS-800 Calibrated Densitometer. The data was then analysed and stored with Biorad Quantity One software version 4.2.3, Build 015.

2.8.5 Determination of protein concentration using the Bradford assay

To determine the concentration of protein in samples, the Biorad Quick Start Bradford Protein Assay was used in the 96-well plate format, as per the manufacturer's instructions.

2.8.6 Nickel affinity chromatography purification of histidine-tagged (His-tagged) proteins

His-tagged proteins purified by nickel-affinity chromatography used the Novagen Ni-NTA His·Bind Purification Kit as per the manufacturer's instructions. Purified protein samples were stored at -20 °C and analysed within 24 hours.

2.8.7 Cobalt affinity chromatography purification of histidine-tagged (His-tagged) proteins

His-tagged proteins purified by cobalt-affinity chromatography used Clonetech His-Tag Purification Resins – Talon as per the manufacturer's instructions. Purified protein samples were stored at -20 °C and analysed within 24 hours.

2.8.8 Enzymatic assay of Chloamphenicol AcetylTransferase (CAT)

CAT activity was determined as described by Shaw, (1975)[91], with a few minor differences. Reagents were added to a clean quartz cuvette to give final assay concentrations in a 580 µl reaction mix of 94 mM Tris, 0.083 mM 5,5'-dithio-bis (2-nitobenzoic acid), and 0.17 mM acetyl coenzyme A. This reaction was mixed by inversion and allowed to equilibrate at 37°C for 5 minutes. Once equilibrated, 10 µl of the cell lysate was added to the reaction mixture, inverted, then 10 µl of 0.3% (w/v) chloramphenicol solution was added to start the reaction. A blank containing the reaction mixture minus the cell lysate and chloramphenicol was used as a reference during the reactions. The reaction was left to run for 1 minute in a thermostated photo spectrophotometer during which the absorbance change at 412 nm was recorded. This process was repeated for each sample, with appropriate dilutions made of the lysates when necessary.

Analysis of the data and the subsequent calculation of the CAT Units per ml were carried out by defining the maximum linear rate using Analytic Jena AG WinASPECT software Version 2.3.1.0, and using the equation;

$$Units/ml\ Enzyme = \frac{(\Delta A412/min\ test - \Delta A412/min\ reference)(0.6)(df))}{(0.0136)(0.01)}$$

0.6 = Total Volume (in ml) of assay

df = Dilution factor, where appropriate

0.0136 = Micromolar extinction coefficient for TNB at A_{412}[92]

0.01 = Volume (in ml) of cell lysate

In order to define the U/mg protein, quantification of the protein concentration was carried out using the Bradford Assay (Section 2.7.3), and the U/mg protein calculated using the equation;

$$Units/mg\ protein\ = \frac{Units/ml\ enzyme}{mg\ protein/ml\ enzyme}$$

2.9 Plasmid Construction

All plasmids used in this study are based upon either the ClosTron plasmid pMTL007C-E2[38,41], or the pMTL80000 series of plasmids[89]

2.9.1 pMTL007C-E2 – ClosTron vectors

To re-target the group II intron of the ClosTron to insertionally inactivate the candidate essential genes, the nucleotide sequence of the target genes were input into Sigma's TargeTron algorithm to identify possible target sites within the sequence. This identified a total of 9 possible target sites ranked in order from 1, being the highest score, to 9 being the lowest. The three possible target sites for CD0274; 190a, 213s, and 280s identified by the algorithm were initially chosen, and the primers sequences generated by the algorithm to re-target the group II intron were synthesised. To re-target the group II intron, the primers IBS, EBS1d, and EBS2 specific to each target, EBS primer and the DNA template supplied by Sigma were combined, in varying ratios of outer to inner primers, into a SOEing PCR mixture. The ~350 bp PCR product containing the re-targeted section of the intron was extracted from the gel and purified.

Having purified the re-targeted intron, both the plasmid pMTL007C-E2 and the intron were digested using *Hind*III and *Bsr*G1 restriction endonucleases. The plasmid is linearised with the same pair of restriction enzymes as the re-targeted introns to

generate compatible cohesive ends. Ligation of the digested PCR product into pMTL007C-E2 was then performed. The resultant re-targeted plasmid was transformed into *E. coli* cloning strain TOP10 by electroporation and plated onto LB agar supplemented with chloramphenicol. The ligation plates showed numerous colonies compared to the vector-only control plate, suggesting the digestion and incorporation of the re-targeted intron was successful allowing the expressed *catP* gene to confer resistance to chloramphenicol. To increase the probability of at least one clone chosen to contain the re-targeted intron, 10 clones were chosen at random, re-streaked and were used to inoculate overnights. The newly targeted-region of the plasmid contains different restriction sites than the original *lacZ*α gene; therefore, restriction analysis can be performed to distinguish which clones are recombinant. Restriction endonucleases *Sac*I and *Bgl*II were used to digest the clones and the results analysed using agarose gel electrophoresis. To confirm the intron was correctly inserted and no PCR errors were present, 5 recombinant clones were sent for sequencing. One error-free re-targeted intron was selected for each of the 3 targets and employed for mutant generation.

For further ClosTron plasmids, the group II intron targeting sequence was identified with the Perutka algorithm[93] instead of Sigma's. In-line with the decrease in DNA synthesis, the group II intron targeting sequences were synthesised by DNA2.0, cloned into vector pMTL007C-E2, and delivered ready to use.

2.9.2. Functional complementation plasmid pMTL-DW1

CD0274 of *C. difficile* R20291 was amplified using the primers Cdi-*dhaT*-ORF-F3+RS and Cdi-*dhaT*-ORF-H6-R3+2RS which also include the addition of the restriction sites *Nde*I and *Eco*RI at the 5' and 3', respectively. Both the plasmid pMLT8215x and the excised CD0274 PCR fragment were digested with *Nde*I and *Eco*RI restriction

endonucleases and ligated together. The recombinant plasmid, pMTL8215x-*dhaT*, containing CD0274 was electroporated into *E. coli* TOP10 and incubated overnight on LB agar supplemented with chloramphenicol. The native promoter of CD0274 was amplified using the primers Cdi-*dhaT*-pro-F1 and Cdi-*dhaT*-pro-R1 which introduce the restriction sites *Not*I and *Nde*I into the PCR fragment at the 5' and 3', respectively. Using the restriction endonuclease *Not*I and *Nde*I, the native promoter of CD0274 was cloned into the plasmid pMTL8215x-*dhaT*, yielding pMTL8215x-P*dhaT*-*dhaT*.

To protect the second copy of CD0274 present upon the plasmid from the targeting mechanism of the group II intron, the nucleotide sequence at the recognition sites of the group II intron for plasmids; pMTL007C-E2:Cdi-*dhaT*-548a and pMTL007C-E2:Cdi-*dhaT*-623a, were silently altered between the two native *Sac*I restriction sites so that the intron no longer recognises the "protected" second copy of CD0274, but the amino acid sequence of the gene product is unaltered as outlined in section 3.2.3. This synthetic construct was designed using the gene designer software developed by DNA2.0 and synthesised by the same company. The synthetically altered CD0274 and plasmid pMTL82151-*dhaT* was digested using the restriction endonuclease *Sca*I. Both the plasmid backbone and the synthetic fragment were ligated together to produce a protected CD0274 plasmid-borne copy, representing plasmid pMTL-DW1.

2.9.3 Chromosomal functional complementation plasmids

As was the case with CD0274, a synthetic version of *metK* and *trpS* was constructed containing silent mutations at the recognition sites for the 2 re-targeted group II introns. As the cost of synthesising DNA has decreased significantly, the whole of *metK*, including the native RBS, promoter, and a SAM riboswitch, were synthesised as a single construct to include a *Not*I and *Nhe*I restriction sites at the 5' and 3' for cloning.

Following digestion using the aforementioned restriction endonucleases, the synthetic *metK* gene including the natural promoter was cloned from the courier vector DNA2.0-P_{metK}-*metK*p, into the ACE vector pMTL-JH18 with the restriction sites *Not*I and *Nhe*I, replacing the reporter gene *lacZa* and creating plasmid pMTL-JH18-*metK*p. As *trpS* seems to be a part of an operon, the native promoter could not be reliably identified and therefore was substituted for the thiolase promoter of *C. sporogenes,* and synthesised with the protected *trpS* as a single construct containing restriction sites *Not*I and *Nde*I at the 5' and 3', respectively. The construct was subsequently cloned from the courier vector DNA2.0-P*trpS*-*trpS*p into plasmid pMTL-JH18, giving rise to plasmid pMTL-JH18-*trpS*. In the case of CD0274, the plasmid pMTL-DW1 was subjected to digestion using restriction endonucleases *Not*I and *Nhe*I to remove the protected CD0274 and its natural promoter, and subsequently cloned into pMTL-JH18, giving rise to pMTL-JH18-*dhaT*. Finally, a synthetic version of *slpA* was constructed containing silent mutations at the recognition sites for the 3 re-targeted group II introns. The whole gene and the native promoter were synthesised by DNA2.0 to include a *Not*I and *Nhe*I restriction site at the 5' and 3' for cloning. The gene and its native promoter were then digested with restriction enzymes *Not*I and *Nhe*I and cloned into vector pMTL-JH18, yielding plasmid pMTL-JH18-*slpA*.

Included within the design of all the synthetic genes was an *Nde*I restriction site over the start codon of each synthesised candidate essential gene, enabling the removal/substitution of the included promoter.

2.9.4 lac *repressor system plasmids*

The *Clostridium lac* repressor system was designed and constructed as outlined in section 4.2.1.1.

Switching of the *lacI* promoter was facilitated by using the restriction enzymes *Sac*I and *Bam*HI. All promoters used in this study were amplified, using the appropriate primers named according to the promoter in table 2.4, under the section promoter library, from *C. acetobutylicum,* with the exception of P_{fdx} and P_{thl} which were amplified from *C. sporogenes*. For removal of the *lacI* promoter, the restriction enzymes *Sac*I and *Bam*HI were used in conjunction with T4 polymerase to blunt the ends, and religated using T4 ligase. Switching of the inducible promoter interrupted with the *lac* operators was facilitated with the use of restriction enzymes *Bgl*II and *Sal*I.

2.9.5 Riboswitch harbouring plasmids

The Riboswitches Thio1, Thio2, and Thio3 were designed and constructed as outlined in section 4.2.2.1. Each riboswitch was cloned into plasmid pMTL82254 using the restriction sites *Not*I and *Nde*I.

Switching of the riboswitch promoters was facilitated by using the restriction enzymes *Sac*I and *Bam*HI. All promoters used in this study were amplified, using the appropriate primers named according to the promoter in table 2.4, under the section promoter library, from *C. acetobutylicum,* with the exception of P_{fdx} and P_{thl} which were amplified from *C. sporogenes*.

2.9.6 Tet system plasmids

The Tet system was designed and constructed as outlined in section 4.2.3.1.

Switching of the *tetR* promoter was facilitated by using the restriction enzymes *Sac*I and *Bam*HI. All promoters used in this study were amplified using the appropriate primers named according to the promoter in table 2.4, under the section promoter library, from *C. acetobutylicum,* with the exception of P_{fdx} and P_{thl} which were amplified from *C.*

sporogenes. Switching of the inducible promoter interrupted with the Tet operators was facilitated with the use of restriction enzymes *Cla*I and *Aat*II.

2.9.7 Essential gene conditional expression plasmids.

The Tet system present upon plasmid pMTL-tet3nO and the plasmids; pMTL-JH18-*dhaT*p, pMTL-JH18-*trpS*p, and pMTL-JH18-*slpA*p were digested using *Not*I and *Nde*I. The *tet* inducible promoter was removed and cloned into the functional complementation plasmids, thereby replacing the previous promoter with the tetracycline inducible promoter, yielding plasmids; pMTL-ACEtet-*dhaT*p, pMTL-ACEtet-*trpS*p, and pMTL-ACEtet-*slpA*p.

2.10 Strain Construction

2.10.1. ClosTron integrants

The ClosTron plasmids containing the re-targeted group II intron were electroporated into *E. coli* donor strain CA434 and subsequently conjugated into the recipient host *C. difficile* 630Δ*erm* cell using standard conjugation techniques[90]. Transconjugants were grown on BHIS and selected for using *C. difficile* supplement and thiamphenicol. Conjugation plates were inspected after 3 days incubation, and one large colony was selected, re-suspended in 100 µl of BHIS, 10-fold serial diluted and plated into BHIS agar supplemented with a *C. difficile* supplement and 2.5 µg/ml erythromycin and incubated at 37°C for 3 days under anaerobic conditions. After the incubation period, incubation plates were inspected and erythromycin resistant colonies were picked for each target and screened for integration of the group II intron into the desired target using PCR. PCR screening was carried out in 2 ways; first was using the flanking primers, for example of CD0274; Cdi-*dhaT*-F2 and Cdi-*dhaT*-R2, and the second was using primers that amplified across the intron integration site, for example of CD0274;

Cdi-*dhaT*-F2 for a sense insertion and Cdi-*dhaT*-R2 for an antisense insertion, combined with the EBS universal primer.

2.10.2 Construction of partial diploid C. difficile *630Δ*erm *strains using Allelic Chromosomal Exchange (ACE).*

The plasmids containing the protected copy of the candidate essential genes were conjugated into *C. difficile* 630Δ*erm* via *E. coli* CA434[90]. The resultant transconjugants were selected for on BHIS media supplemented with thiamphenicol and *C. difficile* supplement. After 2-3 days incubation, the single-crossover event can be selected due to the integrants being able to surpass the growth of their replicative-defective free-plasmid neighbours, therefore appear larger in size. Using this principle, 5 individually derived single-crossover colonies were purified and enriched by sub-cloning on fresh BHIS supplemented with thiamphenicol and *C. difficile* supplement. To select for the second recombination event, 5 single-crossover colonies were re-streaked onto *C. difficile* minimal media supplemented with 5-FOA (1 mg/ml) and uracil (5 μg/ml). After 24 hours, 5-FOA resistant colonies are large enough to pick, 50 were selected to undergo patch plating on BHIS or BHIS supplemented with thiamphenicol to identify the cells that are thiamphenicol sensitive, therefore, have lost the ACE plasmid stopping further recombination events. PCR screening to confirm correct insertion was carried out using flanking primers Cdi630-*pyrE*-F2 and Cdi630-CD0189-R2., followed by sequencing.

2.11 Inducible Expression System Testing.

Appropriate plasmids were transformed into *E. coli* or conjugated into *C. difficile* using standard conjugation procedures[90]. The resulting transconjugants/transformants were selected using the appropriate media and antibiotics as previously described. After the

final cultures are obtained, each is normalised to a 10 ml culture of OD_{600} 1, and pelleted by centrifugation at 16,000 rpm for 20 minutes at 4°C. The harvested samples are lysed through sonication as previously described in section 2.6.7 and the samples analysed by the CAT assay to quantify *catP* reporter expression, as described in section 2.7.9.

2.11.1 lac *repressor system in* E. coli *and* C. difficile*.*

For testing the *lac* repressor system in *E. coli*, the subsequent transformants were used to inoculate a 5 ml overnight starter culture of LB supplemented with erythromycin. The next morning, the starter culture was used to inoculate 50 ml of LB supplemented with erythromycin to an OD_{600} of 0.05. Each culture was grown shaking at 200 rpm at 37°C until OD_{600} 0.5 was reached, the cultures was then split into 20 ml aliquots with one being induced with 1 mM IPTG and the other not. The cultures were sampled every 2 hours after induction until 12 hours, then 24 hours, then the results analysed by the CAT assay.

For *C. difficile* testing, the appropriate transconjugants were grown overnight in liquid BHIS, supplemented with erythromycin, and used to inoculate 40 ml BHIS supplemented with erythromycin to an OD_{600} 0.05 starter culture. The samples were grown to an OD_{600}0.5 and split into 20 ml samples, with one being induced with 1 mM IPTG. After 4 hours, the cells analysed by the CAT assay.

2.11.2 Riboswitches in E. coli *and* C. difficile

In *E. coli*, the transformants were used to set up a 5 ml culture of LB supplemented with erythromycin, and incubated overnight at 37°C and 200 rpm shaking. After the transformants had been incubated for at least 13 hours, the cultures were used to inoculate a 20 ml starter culture of LB supplemented with erythromycin to an OD_{600}

0.05 and incubated at 37°C at 200 rpm until an OD_{600} 0.5 was reached. Each culture was split into two 10 ml cultures, with one being induced with 2 mM theophylline, and the other acting as an un-induced control. The cultures were then left to grow for a further 4 hours and analysed by the CAT assay.

Further experiments in *E. coli* were carried out using the riboswitches in which varying concentrations (0, 1, 2, 5, 10, 20 mM) of theophylline were used to induce *catP* expression. For these, the same procedure was carried out as above, except the volume of media for cultures was adjusted as appropriate.

2.11.3 Tet system in C. difficile, C. sporogenes*, C. acetobutylicum, and* C. beijerinckii

For quick testing to ascertain if system works at one concentration and one time point in *C. difficile*, *C. acetobutylicum,* and *C. beijerinckii*, a transconjugants/transformants harbouring the Tet system plasmid pMTL-tet3nO was used to inoculate an overnight culture of appropriate liquid media supplemented with erythromycin. These overnight cultures were used to inoculate starter cultures of 20 ml of liquid media supplemented with erythromycin to an OD_{600} of 0.05. The starter culture was grown to an OD_{600} of 0.5 and split into 10 ml samples, one of the samples was induced with 316 ng/ml aTc, whereas the other was left not induced and left for 4 hours, after which the samples were normalised to an OD_{600} of 1 at 10 ml, and the cells harvested. The harvested cells were lysed by sonication and analysed by the CAT assay, as previously described (Section 2.7.8).

For testing the dynamic range of the Tet system in *C. difficile* and *C. sporogenes*, the plasmid was grown in an overnight culture and used to inoculate a starter culture of 200 ml of appropriate liquid media supplemented with erythromycin to an OD_{600} 0.05. The

culture was left to grow to an OD_{600} 0.5, split into 10 ml aliquots, and induced with varying concentrations of aTc. As it is unknown what concentrations of aTc will induce CatP production, 2-log dilutions of 1 μg/ml of aTc was used to induce *catP* expression, with the lowest concentration being 0.31 ng/ml. The cultures were incubated for 6 hours, after which the cells were harvested, lysed, and analysed.

To ascertain the speed of induction, stability of expression, and bioavailability of inducer, the Tet system (pMTL-tet3nO) was tested at a single concentration of aTc over multiple time points in *C. difficile* and *C. sporogenes*. A transconjugants was used to inoculate an overnight culture and subsequently used to inoculate starter cultures of 200 ml of liquid media supplemented with erythromycin. *C. difficile* and *C. sporogenes* harbouring plasmid pMTL-tet3nO was grown to an OD_{600} 0.5 and induced with 316 ng/ml aTc and sampled every hour after induction for 6 hours, after which the cells were harvested, lysed, and analysed.

Chapter Three

A Novel Method for the Identification of Essential Genes

3.1 Introduction

The treatment of CDI has evolved little in recent years with only a few proven therapeutic options available. Current first choice antimicrobials vancomycin and metronidazole, although effective in clearing an initial infection, can in certain cases exacerbate the infection due to destroying the natural gut microflora, predisposing a patient and inciting a relapse[9]. In addition to this, sporadic reports of resistance to the front-line antimicrobials have arisen, compounding the need for the development of novel antimicrobials.

The development of novel antimicrobials is dependent on the identification of microbial target proteins which are required for survival of the bacterium to unsure chemical inhibition will ultimately result in cell stasis or death[94]. Normally broad-spectrum antibiotics are preferred due to their cost-effectiveness and of clinical use. However, in the case of *C. difficile*, a narrow-spectrum antibiotic would be preferable due to the possibility of the infection relapsing after initial treatment. To assess an antimicrobial target's effectiveness, specific gene disruptions are required to ascertain the effect inhibition has on cellular viability[3,4,94].

3.1.1 Genetic tools in clostridia

Exploitation of clostridial genomes has been hampered by the historical inability to genetically manipulate the *Clostridium* species. With the tools developed or under development such as; directed mutagenesis[38,41], random transposon-mutagenesis[42], and multiple allelic exchange techniques, the secretive world of the clostridial genome is opening up allowing forward and reverse genetic approaches to identify essential genes for the purpose of antimicrobial design.

The ClosTron is a revolutionary system able to insertionally inactivate target genes by using a bacterial group II intron[38,41]. Group II introns are effective because they can be targeted to insert into virtually any gene with high efficiency and specificity[38]. In the case of identification of essential genes, the group II intron can be targeted against the candidate essential gene in an attempt to insertionally inactivate the gene. If the gene is essential, then it should not be able to be insertionally inactivated due to this genotype being lethal. However, this inability to inactivate the gene does not prove it to be essential, only that it cannot be insertionally inactivated. One method which has been previously used to identify essential genes is to provide a functioning second copy of the candidate gene on a plasmid or chromosome, effectively producing a partial diploid[50,95,96]. This second copy is able to functionally complement the native essential gene's product, allowing the native essential gene to be inactivated.

There are effectively two ways to construct a partial diploid, by providing the second copy on either an autonomous plasmid or inserted into the chromosome. Both have their advantageous and disadvantageous which will be discussed within the results section. In the case of the plasmid-based partial diploid, the plasmid can be introduced into the cell using standard conjugation techniques[90], followed by the ClosTron procedure to inactivate the native copy[38,41]. The chromosomal-based partial diploid is more intricate in that the gene needs to be integrated into the chromosome using a homologous recombination technique. This has recently become possible in *Clostridium* species with the advent of ACE (Heap *et al*, unpublished), allowing a stable recombinant chromosome to be constructed, and subsequent inactivation of the native candidate essential gene. These two techniques provide further indirect evidence that a gene is essential.

3.1.2 Aim of this study

Genetic tool development in *C. difficile* has advanced rapidly recently with the advent of the ClosTron[38] and mariner transposon systems[42], although both systems are able to identify genes which cannot be inactivated, they are unable to provide enough evidence to identify essential genes. Development of a new system which is able to use the information from such systems and define genes as essential is paramount in identifying new antimicrobial targets for the treatment of CDI.

The study presented in this chapter aims to develop a novel technique for the identification of essential *C. difficile* genes, and using the technique, identify candidate antimicrobial targets.

3.2 Results

3.2.1 Identification of candidate essential genes

Bioinformatic analysis of genome data assumes that genes which are essential are likely to be common to, and therefore highly conserved among multiple species. To identify possible targets prior to experimentation, all known essential genes in *B. subtilis* were compared to potential homologs in *C. difficile* through *in silico* analysis. In *B. subtilis,* there have been 271 essential genes identified out of 3,830 genes, making up 6.6% of the total genome[52].

Screening was performed *in silico* as follows: The amino acid sequence of each essential protein identified in *B. subtilis* was used to search the *C. difficile* 630Δ*erm* translated nucleotide database with the blastp algorithm. The translated nucleotide sequence was then used to identify the gene which encodes for the similar protein. All the genes which were found to be essential in *B. subtilis* and their *C. difficile* homolog, along with its gene qualifier, amino acid sequence similarity, and gene size, are listed in appendix 1. Out of the 267 genes which could be identified in *B. subtilis*, 59 were not found in *C. difficile* based upon similarity or gene identifier.

To narrow down the 208 *C. difficile* candidate essential genes for use in this study, three criteria were introduced. The criteria in order of importance: translated nucleotide sequence similarity to *B. subtilis* (with a cut-off of 70% positive identity), size of the gene (less than 2.2 kbp for synthesis time and costs), and its genomic context. These three criteria were selected because a higher amino acid similarity could suggest that the related proteins have similar or the same function increasing the chance of the *C. difficile* protein being essential, expense and synthesis time of the synthetic protected gene, and

to minimise potential polar effects of insertionally inactivating the native gene using the ClosTron[38,41].

Table 3.1. List of candidate essential genes of *C. difficile*

Category	***B. subtilis*** **Gene**	***C. difficile*** **CD Number**	**Protein Similarity**	**Size**
DNA metabolism	*dnaA*	CD0001	350/448 (78%)	1.3 kbp
	dnaC	CD3657	325/431 (75%)	1329 bp
	ligA	CD3309	498/667 (74%)	2 kbp
	topA	CD1274	501/691 (72%)	2.1 kbp
RNA metabolism	*sigA*	CD1455	297/371 (80%)	1173 bp
	*csp*R	CD0566	117/159 (73%)	471 bp
	rnpA	CD3679	72/101 (71%)	345 bp
	trmD	CD1256	168/226 (74%)	696 bp
	nusA	CD1307	255/313 (81%)	1115 bp
tRNA synthetases	*cysS*	CD0052	343/463 (74%)	1407 bp
	metS	CD3540	465/660 (70%)	1938 bp
	serS	CD0014	322/418 (77%)	1272 bp
	trpS	CD2610	242/330 (73%)	1002 bp
	frr	CD2137	144/185 (77%)	558 bp

	prfB	CD0144	251/339 (74%)	987 bp
Cell envelope	*accA*	CD1936	225/302 (74%)	945 bp
	gpsA	CD2630	242/337 (71%)	1032 bp
	plsX	CD1178	220/313 (70%)	1023 bp
	murAA	CD0123	331/419 (78%)	1272 bp
Cell shape and division	*ftsZ*	CD2646	293/387 (75%)	1161 bp
	mreB	CD1145	275/332 (82%)	1062 bp
Glycolysis	*eno*	CD3170	364/427 (85%)	1293 bp
	pgm	CD3171	376/516 (72%)	1153 bp
	tpiA	CD3172	193/247 (78%)	744 bp
Nucleotide synthesis	*adk*	CD0091	162/216 (75%)	651bp
	gmk	CD2588	141/199 (70%)	618 bp
	hprT	CD2691	130/165 (78%)	528 bp
	cmk	CD1816	148/204 (72%)	651 bp
	gylA	CD2726	313/414 (75%)	1245 bp
Cofactors	*metK*	CD0130	348/396 (87%)	1194 bp
Other	*era*	CD2437	219/295 (74%)	894 bp
	gcp	CD0152	251/336 (74%)	1017 bp
Unknown	*ymdA*	CD1329	416/498 (83%)	1563 bp

A total of 33 candidate genes which fitted the criteria were identified representing 10 broad categories of function within *B. subtilis* (table 3.1)[52]. To narrow down the selection further for experimental analysis, the potential for therapeutic targeting was taken into account. Due to its high protein homology at 87%, and its known potential as an antimicrobial drug target[97], *metK* was selected. In *S. aureus*, the *metK* gene encodes an *S*-adenosylmethionine (SAM) synthetase, thought to be essential in both survival and virulence[97]. Its translation is in turn controlled by a class 1 SAM-binding riboswitch located between its promoter and ribosome binding site in the 5'-UTR[97]. An antimicrobial agent that targets MetK would, in principle, prevent SAM biosynthesis and inhibit growth and/or virulence in *S. aureus*. There are numerous other class 1 SAM-binding riboswitches found in other bacterial species, *Bacillus anthracis* alone has 17 highly conserved class 1 SAM-binding riboswitches that collectively regulate 36 genes which are required for sulphur metabolism[97]. Although the known number of 1 SAM-binding riboswitches present within *C. difficile* is unknown, *metK* has a high probability of being essential and therefore would make a good target to develop a system for essential gene identification.

The second candidate essential gene which was selected for further analysis was *trpS*. In *S. aureus*, *trpS* encodes for a tryptophanyl-tRNA synthetase, an enzyme that plays a crucial role in protein biosynthesis by catalysing the synthesis of aminoacyl-tRNAs (aa-rRNA) by the following reactions[98]:

1. **AaRS + aa + ATP = AaRS·(aa-AMP) + PPi**

2. **AaRS·(aa-AMP) + tRNA = AaRS + aa-rRNA + AMP**

aa= amino acid, aa-AMP = aminoacyl-AMP

In the first reaction (1), the amino acid and ATP yield an activated, hydrolysable, aminoacyl-adenylate intermediate. The amino acid moiety is then transferred to its cognate tRNA allowing the resultant aminoacyl-tRNA to act as a substrate for protein synthesis on the ribosome. In the second reaction, the amino acid moiety is then transferred to its cognate tRNA, resulting in an aminoacyl-tRNA which can act as a substrate for polypeptide synthesis on the ribosome[98]. If either of these reactions are inhibited, uncharged tRNA molecules are accumulated which bind to the ribosomes and interrupt polypeptide chain elongation. This interruption leads to the stringent response being induced by *relA*[98]. Induction of this pathway leads to the biosynthesis of tetra- (ppGpp) and pentapeptide guanine (pppGpp) which subsequently inhibits RNA polymerase activity and down-regulates a number of high-energy processes including; RNA, DNA, protein, and peptidoglycan biosynthesis[98] (figure 3.1). Therefore, in principle, if you could target an aminoacyl-tRNA synthetase such as *trpS*, you could interfere with synthesis of aminoacyl-tRNAs and induce the stringent response pathway.

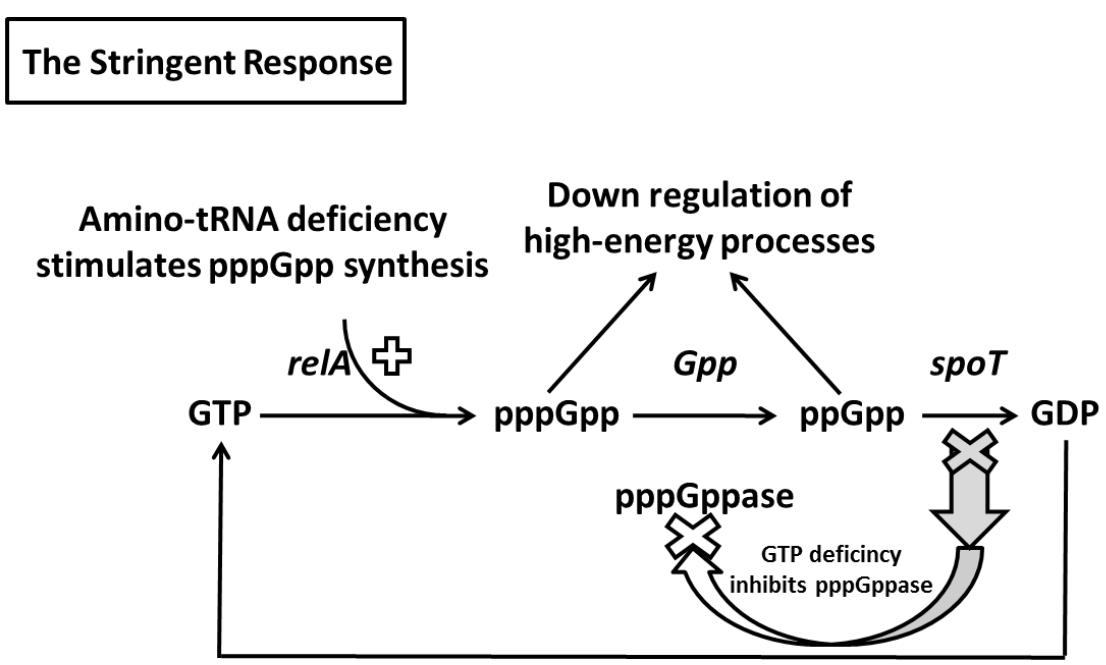


Figure 3.1. Pleiotropic effects in response to *relA* mediated stringent response following amino-tRNA deficiency. Induction of this pathway results in accumulation of tetra- (ppGpp) and penta-peptide guanine (pppGpp), inhibiting RNA polymerase activity and down-regulation other high-energy processes such as; DNA, RNA, and protein metabolism. Adapted from reference[98].

Both of the chosen candidate essential genes were selected due to their high probability of being essential in *C. difficile*. Assuming these candidates are essential in *C. difficile*, an antimicrobial targeted against them would also target a broad-spectrum of bacteria, indicative of a broad-spectrum antibiotic. Generally broad-spectrum antimicrobials are preferred, but in the case of CDI a narrow-spectrum antimicrobial would be of interest. There has only been one candidate essential gene which could be potentially used as a novel narrow-spectrum antibiotic target in *C. difficile* reported in the literature[39,40]. The authors hypothesised that the gene CD0274, a putative alcohol dehydrogenase which they annotated as *gldA,* was involved in the sole detoxification pathway of the toxic metabolite methylglyoxal (MG) and as such was essential[40].

Liyanage *et al*, (2000)[39], first came to study *gldA* in the solventogenic *C. beijerinckii* NCIMB 8052 to isolate and characterise a butanol-tolerant mutant using random-transposon mutagenesis. It had been previously described that some butanol-tolerant *C. acetobutylicum* mutants produced more butanol than their wild-type counterparts[39]. To characterise and isolate a butanol-tolerant mutant in *C. beijerinckii*, they carried out a mutagenesis experiment with transposon Tn*1545* and selected for transposon-inserted butanol-tolerant mutants on 1-1.1% 1-butanol, where wild-type NCIMB 8052 has a tolerance up to 0.4%[39]. It was noted that mutants that were tolerant to butanol had an insertion of Tn*1545* in the intergenic region downstream of *gldA*[39]. The transposon was positioned in such an orientation as the promoter used in selection for transposition was able to transcribe *gldA* in the reverse orientation, downregulating gene expression through antisense RNA production.

In a subsequent study, Liyanage *et al*, (2001)[40], set about creating mutants with an insertionally inactivated *gldA* gene. Using the BLAST sequence analysis for the amino acid translation of the nucleotide sequence using *C. beijerinckii* NCIMB 8052 as the query, they identified CD0274 as the closest match to *gldA* within *C. difficile*. They carried out gene disruption experiments in both organisms using a replicative-defective plasmid that contained an internal fragment of *gldA* and CD0274 lacking both the translational start codon and ADH signature motif 2 which would create single crossover knockout mutants. It was shown that transconjugants which were confirmed to have a mutation in the *gldA* gene, showed limited growth on medium with pin-point sized colonies being formed. When the pin-prick colonies were re-streaked onto fresh media, no viable cells were recovered after 2 days in either *C. beijerinckii* or *C. difficile*[40].

Liyanage *et al*, (2001)[40] also measured free methylglyoxal (MG) levels and sensitivity of wild-type strains of *C. beijerinckii* compared to the transposon-inserted butanol-tolerant

mutant produced in their previous paper[39,40]. It was found that butanol-tolerant mutants contained significantly more free MG than their wild-type counterparts and that the downregulated *gldA* mutants showed increased sensitivity to exogenously added MG. Methylglyoxal is a toxic electrophile inevitably produced as a side-product of glycolysis in both eukaryotic and prokaryotic cells[99]. Under certain conditions, there is an excessive production of toxic sugar phosphates produced during glycolysis[40]. From this evidence they ruled out *gldA* playing a role in central sugar catabolism, in which there are high metabolite fluxes as the observed activity of *gldA* was too low. They also ruled out the participation of *gldA* in another pathway whose flux is lower than that of a primary fermentative pathway. This led them to the hypothesis that GldA plays a primary role in MG detoxification in both *C. beijerinckii* and *C. difficile*[40] (figure 3.2A). However, Liyanage *et al*, (2001)[40], did not go as far as to absolutely prove that *gldA* and CD0274 are essential in *C. beijerinckii* and *C. difficile*.

A BLAST sequence analysis for the amino acid translation of the nucleotide sequence using CD0274 of *C. difficile* 630Δ*erm* as the query, showed that the highest similarity was to a 1,3-propandiol dehydrogenase of *C. beijerinckii* and a 1,3-propandiol dehydrogenase (*dhaT*) of *C. butyricum*. Raynaud *et al*, (2003)[100], noted that DhaT had very high similarity to other sequenced DhaT proteins and has all the common features of class III alcohol dehydrogenases. The authors functionally characterised 1,3-propanediol dehydrogenase of *C. butyricum* and found *dhaT* to be located within a 1,3-propandiole (1,3-PD) operon consisting of three genes; *dhaB1*, *dhaB2* and *dhaT*[100]. Production of 1,3-PD from glycerol involves all three genes; DhaB1 and DhaB2 are responsible for the glycerol dehydratase activity which converts glycerol to 3-hydroxypropionaldehyde, which is subsequently converted to 1,3-PD by *DhaT* (figure 2C). In *C. butyricum* the transcript was only detected and 1,3-PD produced, when the cells where grown on a glucose-glycerol

mixture as substrate. This demonstrates that the operon including *dhaT* is not required for growth under all conditions and as such is not essential in *C. butyricum*.

The best known example of the role of an alcohol dehydrogenase in dihydroxyacetone and methylglyoxal metabolism is in *E. coli* K12[99] (figure 3.2B). Subedi *et al,* 2008[99], in their study characterised the substrate specificity and kinetics of GldA. Enzyme activity was measured for 7 substrates, glycerol and 1,2-PD for oxidation, and hydroacetone, methylglyoxal, DL-gyceraldehyde, glycolaldehyde, and dihydroxyacetone for reduction. The highest specificity of GldA was found to be dihydroxyacetone which was converted to glycerol; next highest activity was the conversion of glycolaldehyde to ethylene glycol. The specificity and specific activity of GldA in converting MG to lactaldehyde was about 6 times lower than the specific activity for dihydroxyacetone[99]. This data suggests that GldA in *E. coli* primarily regulates intracellular levels of dihydroxyacetone by converting it to glycerol[99]. The location within the genome of *E. coli* K12 further supports this by the fact that *gldA* lies immediately downstream of the gene that encodes fructose-6-P aldose (*talC*), overlapping the 3' region by 28 bp. Also the gene encoding a putative phosphotransferase system enzyme (*ptsA*) lies upstream of the *talC* gene forming an operon of the three genes[99]. Due to the complex mechanism of MG detoxification and protective measures in *E. coli*, a disruption of *gldA* in *E. coli* does not result in a lethal phenotype[99]. This would suggest that *gldA* plays a minor if not insignificant role in MG detoxification in *E. coli*.

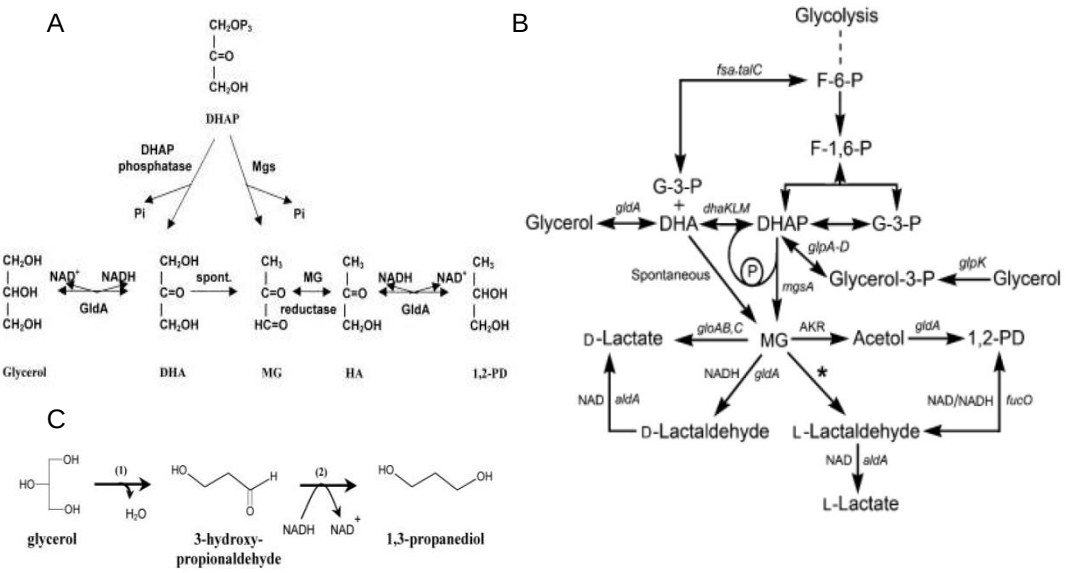


Figure 3.2. (**A**) Hypothetical pathway proposed by Liyanage *et al*, (2001)[40], for the role of GldA in the synthesis and detoxification of methylglyoxal in *C. beijerinckii* and *C. difficile*. Taken from reference[40] (**B**) Pathway for dihydroxyacetone and methylglyoxal metabolism of *E. coli* K12 involving GldA. Taken from reference[99]. (**C**) Pathway for production of 1,3-PD from glycerol in *C. butyricum* involving DhaB1 and DhaB2 at step 1, and DhaT at step 2. Taken from reference[100].

3.2.2 *Attempted gene inactivation of CD0274,* trpS*, and* metK *in* C. *difficile 630Δ*erm *and R20291*

Essential genes are required for survival of the organism, and as such cannot be inactivated without the cell being non-viable. To experimentally test whether CD0274, *trpS* or *metK* can be inactivated, ClosTron technology was employed to insertionally inactivate at specific target sites within the target gene. In the first instance, CD0274, being of high priority due to the possibility of being an essential gene in a narrow range of organisms was targeted. To re-target the group II intron of the ClosTron to

insertionally inactivate CD0274, the nucleotide sequence of CD0274 was entered into Sigma's TargeTron software to identify possible target sites within the sequence. This identified a total of 9 possible target sites each with target scores above the threshold imposed by the software. The targets were designated in order from 1, being the highest score, to 9 being the lowest (figure 3.3). The three possible target sites, 190a, 213s, and 280s identified by the algorithm were initially chosen, and the primers sequences generated by the algorithm to re-target the group II intron were synthesised and constructed, yielding plasmids pMLT007C-E2:Cdi-*dhaT*-190a, pMLT007C-E2:Cdi-*dhaT*-213s, and pMLT007C-E2:Cdi-*dhaT*-280s, respectively.

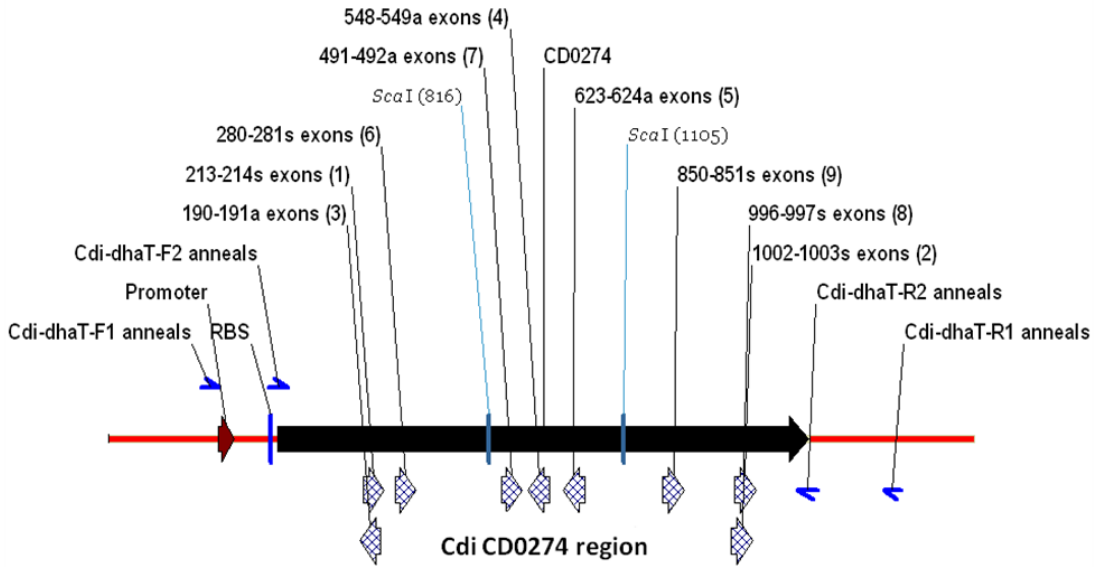


Figure 3.3. ClosTron target sites generated by the Sigma TargeTron algorithm in order of preference. The checked arrows represent the ClosTron targets identified by the algorithm and the orientation of the group II insertion. The red arrow at the 5' of CD0274 indicates the position of the native promoter followed by the RBS downstream. The smaller blue arrows show where the primers anneal for PCR screening flanking CD0274.

Re-targeted plasmids pMLT007C-E2:Cdi-*dhaT*-190a, 213s, and 280s containing the re-targeted group II intron were electroporated into *E. coli* donor strain CA434 and subsequently conjugated into the recipient host *C. difficile* 630Δ*erm* cell using standard conjugation techniques[90]. After the conjugation had taken place, incubation plates were inspected and erythromycin resistant colonies were picked for each CD0274 target and screened for integration of the group II intron into the desired target using PCR. A large number of individual integrants were screened for an insertion of the group II intron into CD0274 but mostly the method did not provide any viable insertions. However, using the flanking primers, 2 out of 40 integrants of target Cdi-*dhaT*-213s were negative for CD0274 suggesting that the intron had inserted within the gene causing the flanking primers to be positioned too far apart to successfully amplify the PCR fragment within the selected annealing time (figure 3.3A). To screen for a correct insertion within CD0274, the primer combination Cdi-*dhaT*-F2 and the EBS universal that amplify across the intron-exon junction was employed. The two integrants for target Cdi-*dhaT*-213s which showed a negative PCR for CD0274 had a PCR fragment of the expected size using the primers which amplified across the gene/intron junction (figure 3.3B). The product was sent for sequencing and was verified to be CD0274 with a successful integration of the group II intron between nucleotides 213 and 214.

Surprisingly, the two *C. difficile* 630Δ*erm* clones which were harbouring the Group II intron within the CD0274 gene were viable after subsequent re-streaking on fresh BHIS agar, indicating either that CD0274 was not essential or that they still produced a fore-shortened but functional DhaT protein. The knock-out protocol was repeated to reproduce the results; however, none of the 20 further integrants screened were positive for insertion of any re-targeted group II intron within CD0274.

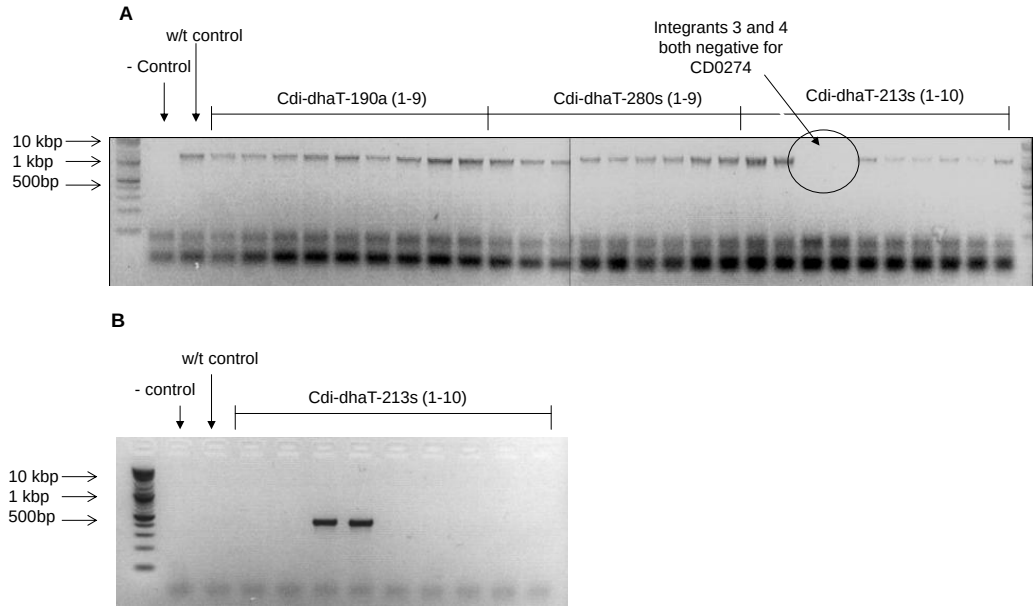


Figure 3.4. PCR screening of erythromycin resistant colonies for the integration of three retargeted group II introns into CD0274. (**A**) Screening for the CD0274 gene of *C. difficile* in integrants of 3 targets, Cdi-*dhaT*-190a (lanes 4-12), -213s (lanes 22-31), and -280s (lanes 13-21) using primers Cdi-*dhaT*-F2 and Cdi-*dhaT*-R1 which flank CD0274. (**B**) Further PCR screening of the Cdi-*dhaT*-213s integrants using the intron-exon junction primers Cdi-*dhaT*-F2 and EBS universal showed a product of ~500 bp.

In the original paper which identified CD0274 as a candidate essential gene, Liyanage *et al*, (2000)[39], noted that when they used a replicative-defective plasmid that contained an internal fragment of CD0274 lacking both the translational start codon and ADH signature motif 2, they could effectively knock-out the gene leading to the hypothesis that CD0274 is essential[40]. However, when they used a replicative-defective plasmid that contained an internal fragment that had both the ADH motifs present, the gene could still produce a functional fore-shortened protein which allowed the cell to be viable[40]. This suggests that if both of the ADH motifs are unaltered, the protein is still functional

and does not cause a lethal phenotype and should have the same phenotypic characteristics as the wild type. To test this, the phenotype of the CD0274 integrant was characterised. The integrant was grown on BHIS agar and CBM media in which glucose was replaced with either glycerol or casein at an equivalent concentration. *C. difficile* 630Δ*erm* was used as a control. As expected, there was no difference in growth between the integrant and the control in either the first plate, or any subsequent re-streaks up to the 4th passage. It was noted that both integrant and control strains grew slower on the CBM and its variations than BHIS agar, of which the casein variation grew the slowest and produced smaller colonies.

As it would seem the ADH motifs need to be disrupted to inactivate the gene, two additional introns were designed, re-targeted to the centre of CD0274 between nucleotides 548-549, and 623-624 to reduce the possibility of a functional fore-shortened protein being produced. As this study continued, the advancement of commercial DNA synthesis technology allowed us to outsource to DNA2.0 for the synthesis and construction of further ClosTron plasmids[41]. The same procedure as described above was repeated to attempt to insertionally inactivate CD0274 with the two new targets. Due to the complications seen with target 213s, a larger number of integrants were screened to confirm that CD0274 could not be insertionally inactivated. After multiple attempts to insertionally inactivate the candidate essential genes, a total of 70 independent erythromycin resistant colonies were screened as above, for each target for the correct integration into the target gene. Out of the 140 possible integrants screened for CD0274, none contained an insertion at the target site. These results could suggest that CD0274 is essential in *C. difficile*, however further evidence was considered necessary to confirm this hypothesis.

In order to strengthen the evidence that the system being developed allows the identification of essential genes, the two other candidate essential genes, *trpS* and *metK*, were additionally targeted using the ClosTron system. In excess of 30 conjugations were required to obtain 70 independent erythromycin resistant colonies for each ClosTron target, and screened as before using the primers EBS and either Cdi-trpS-R1 for *trpS*, or Cdi-metK-F1 for *metK*. The 70 integrants were screened for each target, totalling 140 for each gene. As seen before, none of the colonies screened contained an insertion at the target site. These results suggest that along with CD0274; *metK* and *trpS* cannot be insertionally inactivated, and as such, could be essential for survival.

3.2.3 Testing the temperature sensitive nature of splicing of the group II intron

It was previously noted by Yao *et al*, (2006)[72], that the IED-assisted splicing reaction of the TargeTron system could produce conditional temperature-sensitive mutations in *S. aureus*[72]. It was shown that if the intron was inserted into the host chromosome in the sense orientation, meaning that the target gene's natural promoter was transcribing the intron, the IEP could splice the intron out of the target genes transcript at 32°C, but not at 42°C[72]. Although the exact temperature-sensitive mechanism is unknown, it was hypothesised that the inhibition of the RNA splicing was due to the temperature sensitivity of the intron RNA, the IEP, or a combination of them both[72].

To determine if the group II intron has the same RNA splicing temperature-sensitive characteristics in *C. difficile* as the group II intron in *S. aureus*, the two targets of *metK*; pMTL007C-E2:Cdi-*metK*-548s and pMTL007C-E2:Cdi-*metK*-676s, were employed to attempt to insertionally inactivate *metK* in the sense orientation. Each target was conjugated into the *C. difficile* host strain and transconjugants grown at 32°C or 42°C on erythromycin to select for RAM activation. To ascertain if an integration event had occurred, PCR screening for the correct insertion was carried out on 50 possible

integrants for each target at each temperature, using the same primer combinations as before. The results showed that no PCR positive colonies were obtained for either of the temperatures, and as such any possible insertions into *metK* were not viable (table 3.2). Although the reason for this result is unknown, the only difference between Yao *et al*, (2006)[72] and our study apart from the organism, was the presence of the RAM. The inability of the group II intron to splice out of the precursor mRNA might be due to a decrease in splicing efficiency caused by the increase in the overall group II intron size. This decrease in efficiency might decrease the expression of the essential gene, causing an ineffective protein concentration and subsequent lethal genotype.

Wildtype; Temperature Sensitivity Directed Gene Inactivation

Target Gene	ClosTron Target	Temperature	Colonies Screened	Desired Integrants Obtained	Frequency of desired Integrants
metK	548s	32°C	50	0	0%
		42°C	50	0	0%
	676s	32°C	50	0	0%
		42°C	50	0	0%

Table 3.2. Insertion frequencies of the group II intron re-targeted to *metK* using targets 548s and 676s at 32°C and 42°C in *C. difficile* 630Δ*erm*.

3.2.4 Construction of the CD0274 *partial diploid plasmid*

In the first strategy investigated, an attempt was made to use the conditional temperature-sensitive nature of the IEP-assisted splicing of the group II intron to identify essential genes; unfortunately this method was not possible in *C. difficile,* so an alternative strategy by creating a partial diploid was attempted. A partial diploid is a term used to describe a cell which contains a second copy of only part of its genome[50,73,101]; in this case, the second copy is the target gene being tested. In theory, if CD0274 is essential, the presence of a second copy will allow the chromosomal copy of CD0274 to

be insertionally inactivated. If this disruption of the native CD0274 takes place, then the plasmid borne CD0274 will become the essential copy, and as such the plasmid would become 100% stable and should not be lost over time (figure 3.5).

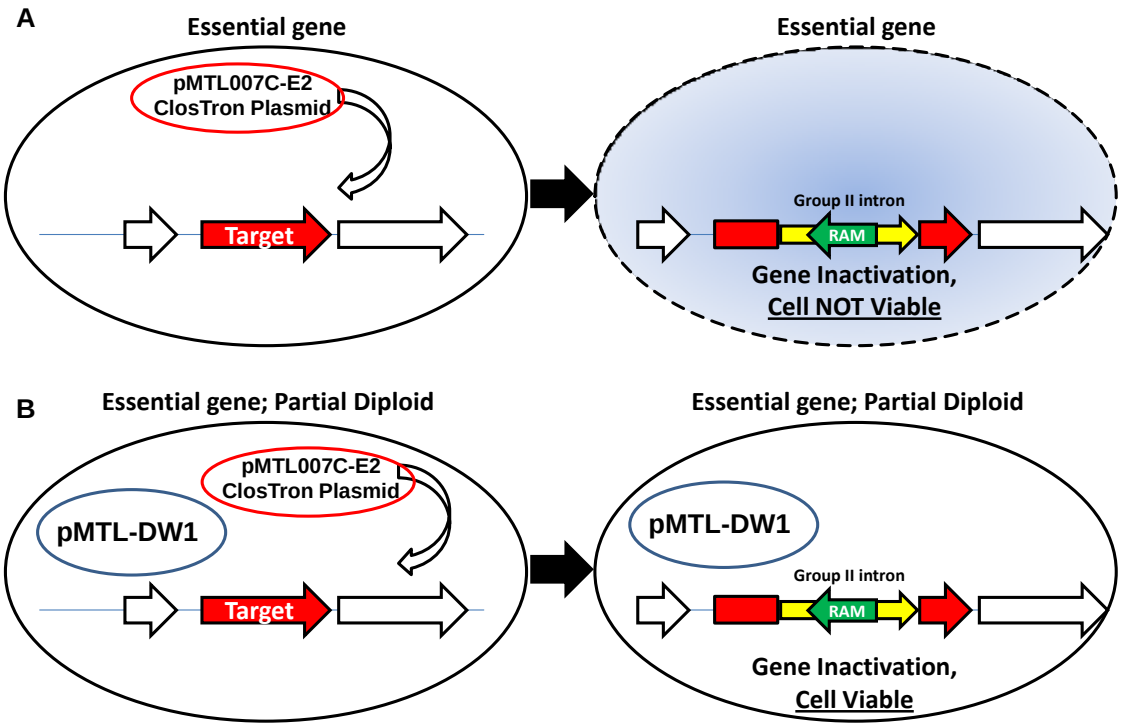


Figure 3.5. Graphical representation of the partial diploid method. (**A**) In the case of an essential gene, if the group II intron insertionally inactivates the essential target gene, the genotype is lethal and the cell is no longer viable. Due to this, the insertion cannot be identified as the cell does not grow. (**B**) If there is a second copy of the essential gene present upon plasmid pMTL-DW1, the group II intron can successfully insertionally inactivate the target gene, with the second copy functionally complementing the essential function, allowing the recombinant cell to grow. If the cell is able to grow, the integrated intron can be detected by PCR.

Implementation of the partial diploid requires three different antibiotic resistance markers; thiamphenicol for pMTL007C-E2:cdi-*dhaT*-548a and -623a, erythromycin for selection of integration using the RAM, and a third for positive selection for the plasmid containing CD0274, thus *C. difficile* strain R20291 was used for this method. *C. difficile* R20291 is naturally sensitive to spectinomycin 750 μg/ml, whereas *C. difficile* 630Δ*erm* is not[41]. Therefore plasmid pMTL82151 was chosen, into which CD0274 and its native promoter of *C. difficile* R20291 were cloned, yielding plasmid pMTL82151-P_{dhaT}-*dhaT*.

Within plasmid pMTL82151-P_{dhaT}-*dhaT*, CD0274 contains the exact same nucleotide sequence as the native copy, therefore there needs to be a way to protect the gene from disruption, i.e. prevent the group II intron insertion. The group II intron is able to insert within a target site by recognising a 45 bp sequence. Each retargeted group II intron recognises a unique, highly specific sequence. It follows that, if the sequence is altered, the group II intron will not be able to recognise the altered target site, preventing insertion. To protect the second copy from the group II intron, the nucleotide sequence at the recognition sites of plasmids; pMTL007C-E2:Cdi-*dhaT*-548a and pMTL007C-E2:Cdi-*dhaT*-623a were silently altered, i.e. the changes made did not cause any alteration to the amino acid sequence of the encoded CD0274 gene product (figure 3.6). The site between the two *Sca*I restriction sites was chosen as it could be switched with the synthetically altered fragment with ease, and includes the two target sites for the plasmids pMTL007C-E2:cdi-*dhaT*-548a and -623a. This synthetic construct was designed using the gene designer software developed by DNA2.0 and synthesised by the same company and cloned into plasmid pMTL82151-P_{dhaT}-*dhaT*, yielding pMTL-DW1.

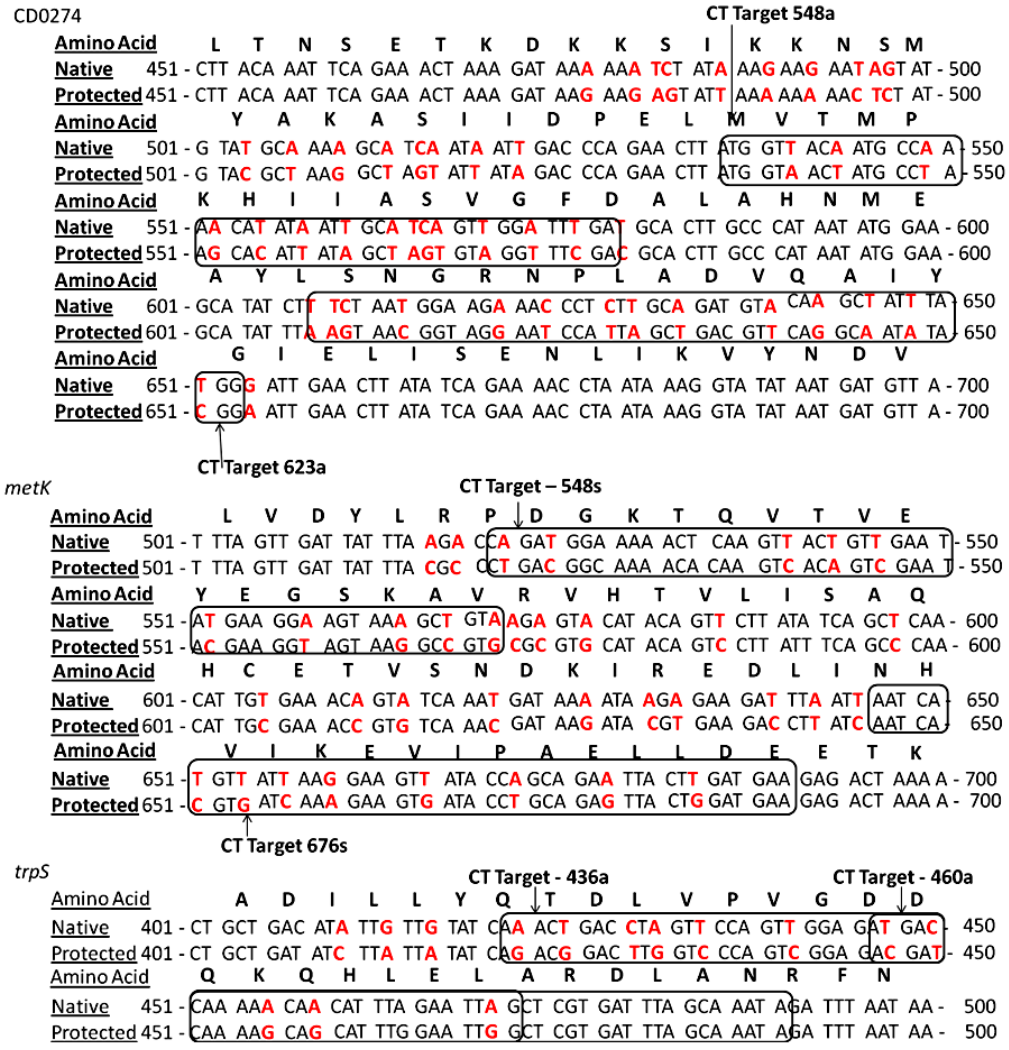


Figure 3.6. Protection of the synthetic partial diploid copy of the candidate essential gene. The group II intron recognises a 45 mer target site displayed by a box encompassing the two target sites for each candidate essential gene. In bold red is the silent nucleotide changes made in the synthetic protected copy of each candidate essential gene compared to the native gene, to alter the group II introns recognition sequence and block integration, but allow the same translated nucleotide sequence.

Plasmid pMTL-DW1 was conjugated into *C. difficile* R20291, followed by the ClosTron plasmids and subsequent disruption experiments were carried out to inactivate the native CD0274. It was noted that the conjugation efficiency of pMTL-DW1 was very low compared to the pMTL84151 empty vector control, and that the resultant pMTL-DW1 transconjugants grew very slowly in comparison to the control, and could not be propagated onto fresh solid media, only liquid culture, growing to low cell densities overnight (OD_{600} 0.1, compared to OD_{600}1.2-1.6 for the parent strain). The transconjugants containing pMTL-DW1 were subjected to another round of conjugation to conjugate the ClosTron plasmids into the cell to allow subsequent insertional inactivate the native CD0274. This proved to be problematic, with no positive double-transconjugants being obtained after 20 independent conjugation attempts. Transconjugants containing the ClosTron plasmids were obtained, but without the pMTL-DW1 plasmid being identified, suggesting either the double-transconjugants had not been detected or the pMTL-DW1 plasmid had been lost during the second conjugation.

The four antibiotics used for selection and the presence of two plasmids could cause the double-transconjugants to take longer to grow than the single transconjugants, allowing background growth to dominate. *C. difficile* supplement was being used for negative selection against the *E. coli* after plating out the conjugation, but background growth was found to predominate after about 60 hours. Cycloserine and cefoxitin (*C. difficile* supplement) are both bacteriostatic antibiotics therefore once the antibiotic activity is decreased over time, the background growth becomes visible. Due to this, another antibiotic which lasts longer or is bactericidal to select against the *E. coli* donor would be advantageous. One such antibiotic is polymyxin B. To test the efficacy of polymyxin B compared to, and in combinations with, *C. difficile* supplement in selecting for double-

transconjugants, conjugations were carried out using *E. coli* C600RK2 harbouring the ClosTron plasmids and the recipient *C. difficile* R20291 harbouring pMTL-DW1. It was found that polymyxin B was able to suppress background growth for a longer period of time, but it was not effective in combination with the *C. difficile* supplement. The reason for this is unknown. However, no double-transconjugants were obtained after further conjugations and screening. These results suggest that the plasmid pMTL-DW1 is being lost during the subsequent conjugation transfer of the ClosTron plasmid; therefore the order in which the plasmids were conjugated into *C. difficile* R20291 was reversed. Unfortunately this didn't circumvent the problems as no double-transconjugants were obtained. Other attempts at improving the stability of the plasmid by changing the Gram-positive replicon and changing the antibiotic resistance marker also proved to be ineffective.

It would seem the presence of CD0274 upon a plasmid within *C. difficile* R20291 effects the growth of the cell dramatically, inhibiting the passage of pMTL-DW1 transconjugants onto fresh agar, ultimately resulting in cells that rapidly loose the plasmid when selective pressure is removed. This apparent instability suggests that the presence of CD0274 leads to detrimental level of the DhaT protein.

3.2.5 Construction of chromosomal partial diploid strains

As it was not feasible to use the plasmid-borne CD0274 method to construct a plasmid-based CD0274 partial diploid strain, an alternative strategy was investigated in which the gene was introduced into the *C. difficile* 630Δ*erm* chromosome using ACE (figure 3.7). ACE is a new method developed in our lab which allows integration of plasmid DNA into the chromosome without the need for a counter-selection marker (figure 3.8A). In conventional plasmid-based allelic exchange systems, the use of a positive selection marker is required to select for a single-crossover integration event which inserts the

plasmid DNA into the target chromosome, then the presence of a negative-selection marker within the integrated DNA is used to select for the second-crossover event and loss of the exchanged allele-containing plasmid, culminating in allelic exchange. ACE has a few important differences which circumvent the use of a plasmid-borne negative-selection marker, allowing integration of cargo DNA into the *pyrE* locus, and producing only the genotype desired. The *pyr* operon of *C. difficile* is responsible for the biosynthesis of pyrimidine nucleotides with *pyrE* encoding an orotate phosphoribosyl-transferase resulting in sensitivity to 5-fluoroorotic acid (5-FOA)[102,103]. Therefore, if the *pyrE* gene becomes inactivated, the mutant strain becomes resistant to 5-FOA as it can no longer utilise 5-FOA[103].

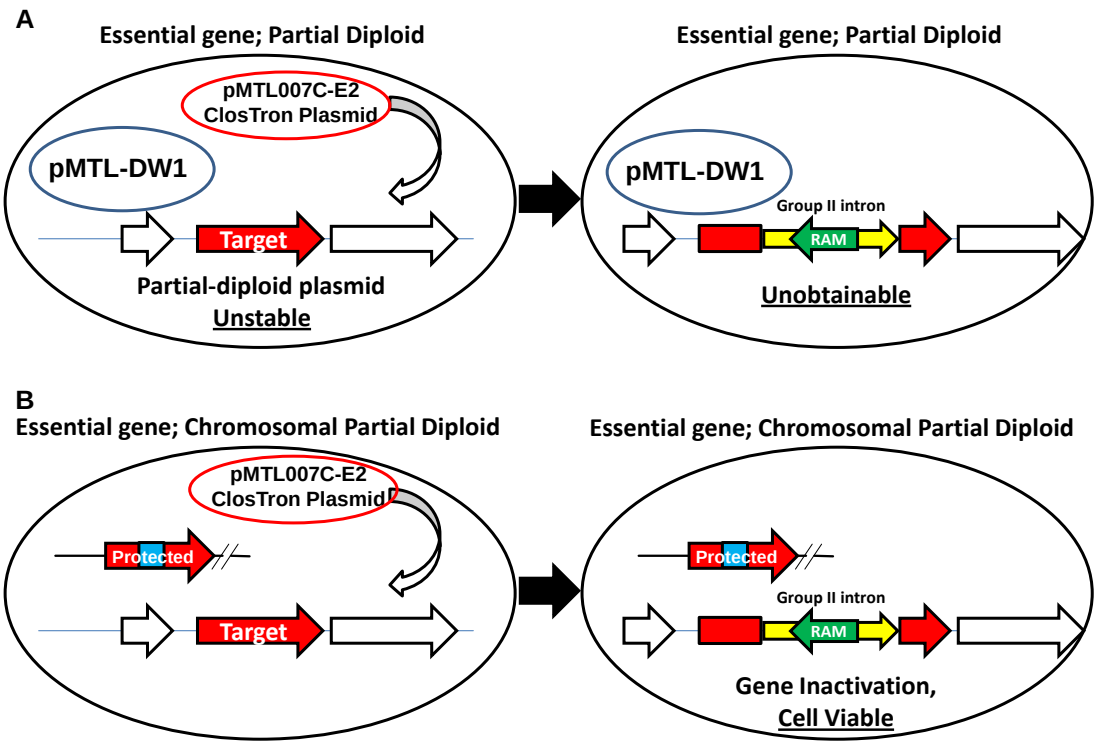


Figure 3.7. Comparison of the plasmid and chromosomal partial diploid strategies (**A**) The strategy used in chapter 3.2.3. The plasmid pMTL-DW1 containing a protected copy of CD0274 able to functionally complement the native copy, allowing the native copy to be inactivated

via ClosTron (**B**) Due to the strategy in (A) not working, a second strategy of placing the protected second copy within the chromosome was undertaken. This second copy functions in the same way as the native copy, allowing the native to be inactivated only in the presence of the ectopic protected copy.

In the first instance, *metK* was used to test the chromosomal-borne partial diploid system. As was the case with CD0274, a synthetic version of *metK* was constructed which contains silent mutations at the recognition sites for the two re-targeted group II intron ClosTron plasmids; pMTL007C-E2:Cdi-*metK*-548s and pMTL007C-E2:Cdi-*metK*-676s, between nucleotides 516 and 686**,** but keeping the same translated nucleotide sequence (figure 3.6). This construct was subsequently cloned into plasmid pMTL-JH18, yielding plasmid pMTL-JH18-*metK*. The synthetic construct is now flanked by a 300 bp internal fragment of *pyrE* at the 5' and a 1200 bp DNA fragment homologous to the sequence downstream of *pyrE* at the 3'. Also present on the plasmid is a *catP* chloramphenicol/thiamphenicol antibiotic resistance marker, and the pIM13 replicon.

The *E. coli* strain CA434 was used as a conjugative donor strain to transfer the pIM13 based plasmid pMTL-JH18-*metK* into *C. difficile* using standard conjugation techniques[90]. The pIM13 replicon is replication defective in *C. difficile*[89]. Consequently, as the *catP* gene responsible for thiamphenicol resistance is carried by the pMTL-JH18-*metK*, the growth rate of the transconjugants in the presence of thiamphenicol is restrained by the rate at which the plasmid replicates. Integration of the plasmid by single-crossover integration and concomitant relocation of the *catP* gene to a chromosomal location removes this constraint, allowing the cells to grow faster in the presence of thiamphenicol. Such single-crossover integrants therefore have a growth advantage in antibiotic

supplemented media. The single-crossover event can, therefore, be selected through the isolation of faster growing colonies.

The integration of the plasmid is dependent on homologous recombination of the two homology arms which flank the gene (*metK*) to be integrated. In classical methods of allele exchange, the two homology arms are of equal length to stop bias of which arm recombines in the first crossover event. In ACE the homology arms are asymmetric, with the left-hand homology arm (LHA) being 300 bp long, and the right-hand homology arm (RHA) being 1200 bp in length. This 3:1 ratio of right-hand to left-hand length difference allows the RHA to recombine in the first cross-over event at a much greater frequency than the shorter LHA. Integrants at the RHA, therefore predominate. Such cells retain sensitivity to 5-FOA as a functional, intact copy of *pyrE* is still present. Using this principle, 5 individually derived single-crossover colonies were purified and enriched by sub-cloning on fresh BHIS supplemented with thiamphenicol and *C. difficile* supplement.

Next, the second-crossover event occurs between the smaller 300 bp LHA. There are theoretically two possible outcomes of this recombination event, the type strain in which the original integrated plasmid carrying the mutant allele is excised, or the desired genotype in which the plasmid excises together with the type strain allele. The net result is allelic-exchange. To select for the desired event, single-crossover colonies are re-streaked onto *C. difficile* minimal media supplemented with 5-FOA to select for the creation of the mutant cell. As *pyrE* becomes inactivated, the media is also supplemented with uracil to allow the growth of the created auxotroph. After 24hours, 5-FOA resistant colonies are large enough to pick, and undergo patch plating on BHIS or BHIS supplemented with thiamphenicol to identify the colonies that are thiamphenicol sensitive due to the loss of the excised plasmid. Loss of the plasmid is

important to prevent further recombination events. The 5-FOA resistant colonies are then subjected to PCR screening using primers that amplified across the insertion locus to confirm allelic exchange has taken place at the desire site (figure 3.8C).

As this method of introducing the *metK* into the chromosome at the *pyrE* locus was shown to be successful, the other candidate essential genes, CD0274 and *trpS*, were also delivered into the chromosome using an identical procedure. As with *metK* and CD0274, *trpS* also needs to be protected from group II intron targeting. The nucleotides around the two ClosTron target sites were silently altered for each target gene(figure 3.6). In addition, as *trpS* forms part of an operon, and therefore lacks a promoter immediately upstream of its 5'-end, the thiolase promoter of *C. sporogenes* was positioned immediately upstream of the coding region. The candidate essential gene was subsequently cloned into plasmid pMTL-JH18 giving rise to plasmid pMTL-JH18-*trpS*. Both candidate essential genes, CD0274 and *trpS*, were introduced into the *C. difficile* 630Δ*erm* chromosome as before, and their expected insertion confirmed by PCR(figure 3.8D).

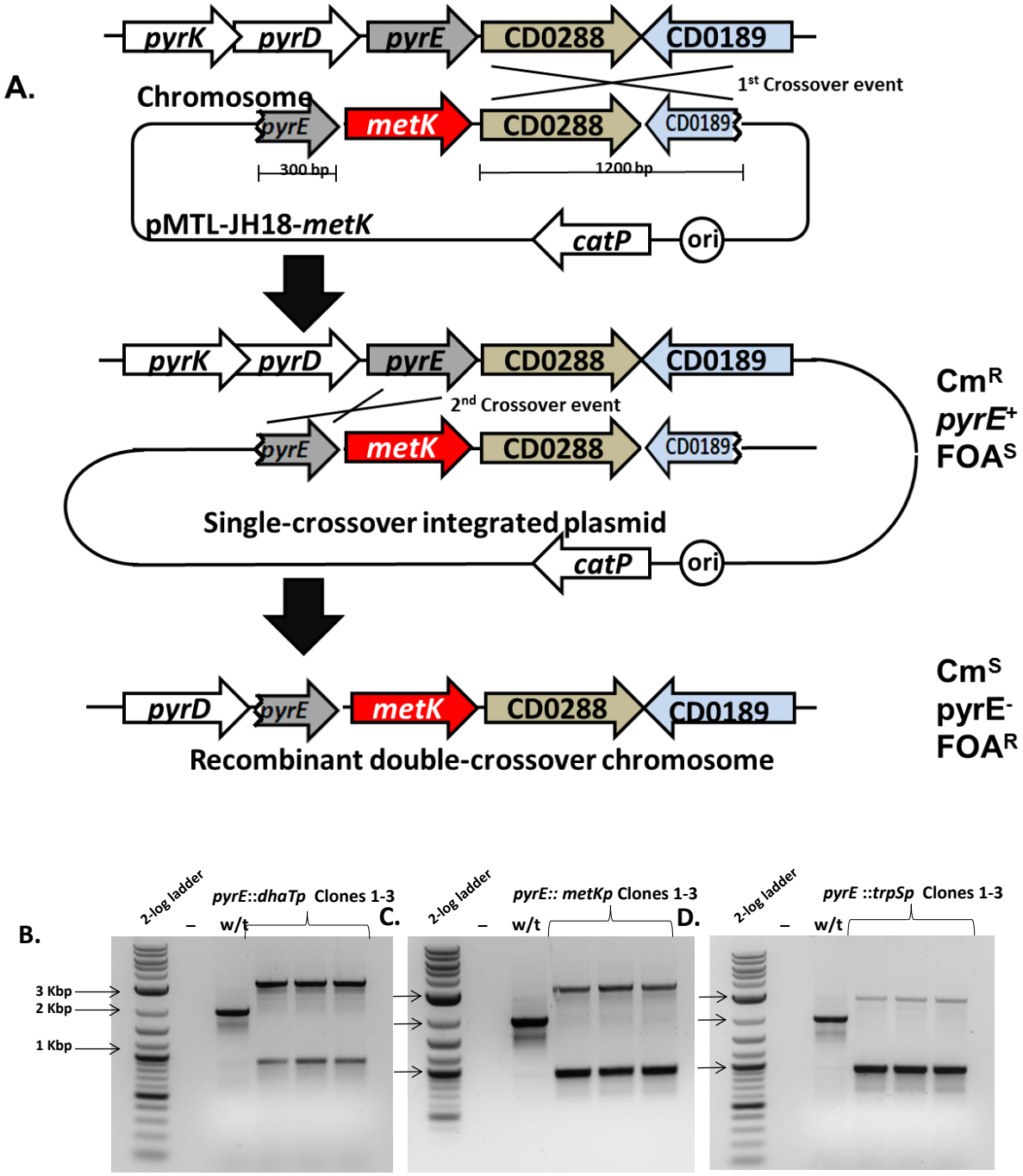


Figure 3.8. DNA integration at the *pyrE* locus of *C. difficile* 630Δ*erm* (**A**) Selection of double-crossover clones using plasmid pMTL-JH18. The first recombination event is directed by the long right-hand homology arm consisting of 1500 bp downstream of *pyrE* on the *C. difficile* 630Δ*erm* chromosome. The resulting single-crossover clones have their replication deficiency conferred by pIM13 alleviated allowing them to grow quicker

than the freely-replicating plasmid containing cells. Subsequently, the second crossover event between the 300 bp left homology arm of pMTL-JH18 containing an internal fragment of *pyrE* completes the allelic-exchange. Two possible outcomes can occur, the parent strain or the double-crossover clone. To select only for the desired event, selection on 5-FOA allows only *pyrE*- strains (double-crossover clones) to grow. (**B**) PCR screening of 3 CD0274 double crossover clones using primers Cdi630-pyrE-F2 and Cdi630-CD0189-R2 which anneal on the parent strain chromosome flanking the insertion site. MW is a 2-log DNA Ladder (NEB) molecular weight marker, lane 1 is water only negative control, lane 2 is (expected size ~2 Kbp) *C. difficile* 630Δ*erm* DNA control, lanes 3-5 are candidate clones (expected size ~ 3.3 Kbp). (**C**) PCR screening of 3 *metK* clones using the same primers and order as (B) (expected size ~ 3.3 Kbp). (**D**) PCR screening of 3 *trpS* clones using the same primers and order as (B) (expected size ~ 3 Kbp).

3.2.6 Inactivation of CD0274, **trpS***, and* **metK** *in a partial diploid strain*

The insertion of the candidate essential genes in to the chromosome allows the functional complementation of the native candidate essential genes. This in theory means that the native candidate essential genes are no longer essential due to the second ectopic copy taking over the essential function[96], allowing the native copy to be insertionally inactivated using the ClosTron[38]. To test this, each partial diploid strain was subjected to the ClosTron procedure to disrupt the native copy.

After the selecting for erythromycin resistant colonies representing integration of the group II intron, 20 colonies for each ClosTron target was screened by PCR using the same primers as used before which amplify across the intron-exon junction of the

intron insertion. Each target showed a high frequency of integration ranging from 75% for the 623a target of CD0274, to 100% for targets: 548a of CD0274 460a of *trpS*, and both targets of *metK* (table 3.3B). This result demonstrates that; the targets used in this study can efficiently target the group II intron to insert at its respective target site, and that these candidate essential genes can only be insertionally inactivated in the presence of a second protected copy, identifying them as essential for viability.

A

Wildtype; Directed Gene Inactivation

Target Gene	ClosTron Target	Colonies Screened	Desired Integrants Obtained	Frequency of desired Integrants
metK	548s	70	0	0%
	676s	70	0	0%
dhaT	548a	70	0	0%
	623a	70	0	0%
trpS	436a	70	0	0%
	460a	70	0	0%

B

Partial-Diploid; Directed Gene Inactivation

Colonies Screened	Desired Integrants Obtained	Frequency of desired Integrants
20	20	100%
20	20	100%
20	20	100%
20	18	90%
20	15	75%
20	20	100%

Table 3.3. Insertion frequencies of the group II intron re-targeted to 2 target sites for each of the 3 candidate essential genes in *C. difficile* 630Δ*erm* at 37°C. (**A**) Attempted candidate essential gene inactivation using the group II intron in *C. difficile* 630Δ*erm* parent strain. (**B**) Attempted candidate essential gene inactivation using the group II intron in the ACE created partial diploid.

3.3 Discussion

Over the past 10 years, the search for novel antimicrobial targets and agents to treat CDI has acquired an increased sense of urgency. Current front-line broad-spectrum antimicrobials are starting to waiver, with the sporadic emergence of resistant strains and increasing rates of relapse following successful treatment. The identification of novel antimicrobial targets is essential if CDI is to be controlled in the coming years[5-10,13,14,38,104-112]. The reason for our lack of novel drug targets and essential genes is largely a consequence of the availability of woefully inadequate methods for generating clostridial mutants[38,41]. To date, only a single putative gene has been identified as essential in *C. difficile*[39,40]. CD0274 was identified through the generation of a single-crossover mutant using a replication-deficient plasmid. Cells with the desired mutant genotype appeared to be lethal, with the mutant cells being unable to be propagated[40]. Whilst this strategy appears to have provided reasonable preliminary evidence that CD0274 is essential, it was reliant on the transient, detectable growth of the mutant, in the form of 'pin-prick' colonies. Thus, in this instance, growth of the organism was possible until the build-up of a toxic metabolite curtailed further multiplications. Such a strategy could, therefore, only be applied to similar circumstances and would not be applicable when a particular gene and its encoded product was absolutely required for growth. Moreover, the use of single-crossover mutants can complicate interpretation of data, due to the reversion to the type strain following excision of the plasmid as a consequence of recombination between duplicated DNA.

In this study, we wanted to develop a method to more definitely identify essential genes in *C. difficile*, and a procedure that could be more universally applied. After demonstrating that the candidate essential genes were not able to be insertionally inactivated using the ClosTron, we tested the temperature sensitivity of the group II

intron as it had been previously shown that the TargeTron system could be used to identify essential genes in *S. aureus*[72]. To test out this ability in *C. difficile* using *metK* as the target gene and the pMTL007C-E2 ClosTron plasmids targeted to *metK*, we attempted to insertionally inactivate *metK* at both temperatures. Unfortunately this technique did not provide any integrants with an insertion in the target gene, even at the permissive temperature. The reason for this is unknown but suggests that intron splicing of the *metK* mRNA is insufficient for the production of MetK at a viable level. The big difference between the method employed here and Yao *et al*'s (2007), with the exception of the organism, was that they did not use a RAM to positively select for integrants, whereas we did[72]. This could mean that the increase in size of the intron decreases mobility frequency of the IEP-assisted intron splicing to a level below effective expression for viability of the cell[62]. If the RAM was removed from the intron to counteract this, the proportion of cells in which an integration event had taken place would be very low compared to the population. This low frequency would make the successful isolation of an integrant challenging.

In other bacterial species such as *M. tuberculosis*, a partial diploid in which a second copy of the essential gene is present within the cell is used to functionally complement the native essential gene, allowing the native copy to be inactivated[95,96]. Unfortunately our attempts at this method with CD0274 were unsuccessful. Multiple attempts at conjugating the partial diploid plasmid into *C. difficile* failed to provide stable transconjugants which could be propagated. Efforts to overcome this instability by changing the Gram-negative replicon and/or resistance marker were unsuccessful. It was hypothesised that the increased gene dosage of CD0274 that results from its localisation to a plasmid is detrimental to the cell.

In order to overcome the instability seen in the plasmid-based partial diploid method, we used a second protected copy of a candidate essential gene and created a stable partial diploid by inserting the gene into the chromosome using ACE. Using these strains, we were able to insertionally inactivate the native copy of the essential genes using the ClosTron system with high frequency and efficiency only when the second copy was present. Using this approach, we identified three genes which are essential for viability in *C. difficile*; CD0274, *trpS*, and *metK*. Although the reason why CD0274 is essential is unknown, it has been hypothesised that it is involved in detoxification of toxic metabolite[39,40]. At present, this is the only known reported method to functionally demonstrate essentiality of a gene in *C. difficile*. However, this novel exploitation of the group II intron could be used in any bacteria that the TargeTron/ClosTron system is viable in. Further research on the subject could be very advantageous in identifying and characterising potential drug targets for the treatment of CDI and other bacterial infections.

Chapter Four

Development of *Clostridium* Conditional Expression Systems

4.1 Introduction

The combination of biochemical, physiological and genetics studies in genetically engineered bacteria have provided a wealth of information for studying essential biochemical processes[113-115]. Such examples include; enhanced or ectopic transcription and translation of enzymes and regulatory proteins, and the expression of anti-sense RNA which interferes by sequestering mRNA[116], decreasing translation of a gene product.

Conditional expression of a gene or operon is a means to control transcription or translation by the addition of an external stimulus (inducer)[113,116-119]. Conditional expression systems are of great importance due to their ability to control any desired gene expression throughout the cells growth. One such application for conditional expression systems is the means to identify essential genes[67,72,80,116,120]. By placing a candidate essential gene under conditional control, cell viability is dependent upon an inducer being present[46,47,65,115,121]. If no inducer is present, the candidate essential gene's expression is turned off, and as a result the genotype is lethal[45,46,52]. This is direct evidence that defines a gene as essential, rather than non-essential. Using conditional expression also allows the strain to be used for further analysis, as opposed to the genotype being lethal.

There are two main ways to control gene expression in prokaryotes, by controlling transcription[67] or translation[1,122]. Current methods to control transcription use a class of promoter which is able to modulate levels of RNA polymerase II-dependent transcription in response to a signal[113]. Controlling translation is not as common, but recently advances in riboswitch[117,122,123] and anti-sense RNA[39,47,116,124] have allowed researchers to harness the potential of using such regulatory systems.

4.1.1 Inducible promoter systems

A broad range of promoters that differ in their ability to modulate the temporal and spatial expression patterns of target genes are currently available in other bacteria[119]; unfortunately, no such appropriate systems had been developed in *Clostridium* species at the start of this project. Recently however, a tetracycline inducible promoter system has been used in *C. difficile*[120]. As the name suggests, inducible promoters are able to modulate transcription by the presence or absence of an external stimulus. There are two main groups of inducible promoters, the physically-regulated promoters; whose promoters are regulated by physical stimuli such as light or temperature[125], and chemically-regulated promoters; whose promoters are regulated by a variety of chemical compounds[126-128]. Due to our requirements, chemically-regulated promoters would be more suitable as varied concentration of inducer can be applied easily, a favourable characteristic. In other bacterial species, there are a number of small molecule chemically-regulated promoters, including the *lac* repressor system, and the Tet system[126-128].

The *lac* repressor system is based upon the natural regulation of the *lac* operon of *E. coli*, a paradigm for negative control of gene expression at the transcription level. The *lac* operon is required for the transport and metabolism of lactose[68]. It consists of three structural genes; *lacZ* (β-galactosidase), *lacY* (β-galactoside permease), and *lacA* (β-galactoside transacetylase), co-transcribed from the promoter P_{lac} immediately upstream into a polycistronic mRNA molecule (figure 4.1)[129]. Contained within the promoter are three operator sequences which are capable of binding the repressor molecule LacI, transcribed upstream of the operon[130]. In the absence of lactose, LacI is produced and binds to the operator sequences of P_{lac}, inhibiting RNA polymerase II-dependant transcription[130]. RNA polymerase II-dependant transcription resumes when the

repressor molecule is removed from the operator binding sites by an inducer[130]. The inducer binds to the LacI resulting in a conformational change which decreases it's affinity to the operators significantly, thereby inhibiting the binding of the repressor to the operators, allowing transcription of the operon[68]. In the natural system, lactose is converted to allolactose by β-galactosidase (the product of the *lacZ* gene) which acts upon LacI. However, as allolactose can be metabolised by the cell, the levels of inducer decrease over time, therefore the non-metabolized lactose analogue isopropyl-thio-b-D-galactoside (IPTG) is used (Figure 4.2)[131].

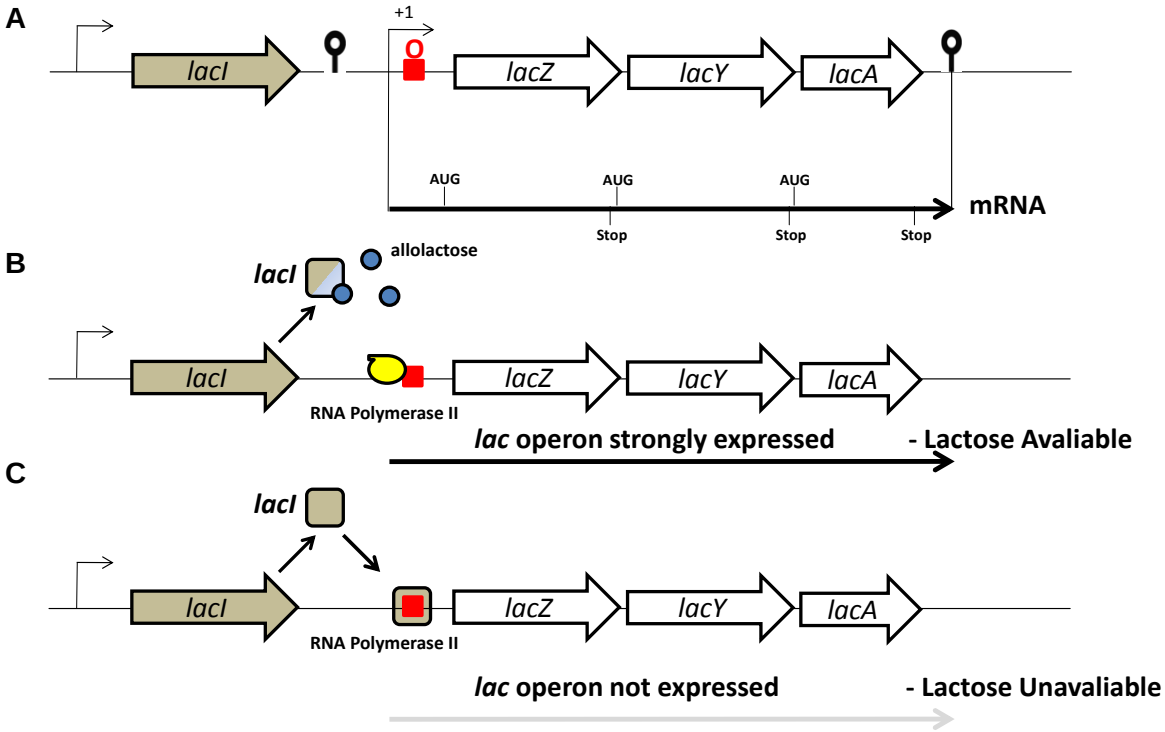


Figure 4.1. Natural regulation of the *lac* operon of *E. coli*. (**A**) The *lac* operon of *E. coli* contains three genes; *lacZ*, *lacY*, *lacA*, under transcriptional control of a promoter upstream of *lacZ* containing three *lac* operators. Upstream of the operon is *lacI* whose gene product LacI is able to bind to the *lacO* (red square designated with O above) thereby inhibiting RNA polymerase II-dependant transcription (yellow thought

bubble). (**B**) In the presence of lactose, LacZ converts lactose to Allolactose (blue circle) which is able to able to bind to and cause a conformational change in the structure of LacI (diagonally shaded brown/blue square), disabling its ability to bind to *lacO*, allowing RNA polymerase II to bind to the *lac* operon promoter and transcribe the operon, which in turn allows *E. coli* to utilize lactose as an energy source. (**C**) In the absence of lactose, LacI (brown square) is able to bind to *lacO*, inhibiting transcription of the operon.

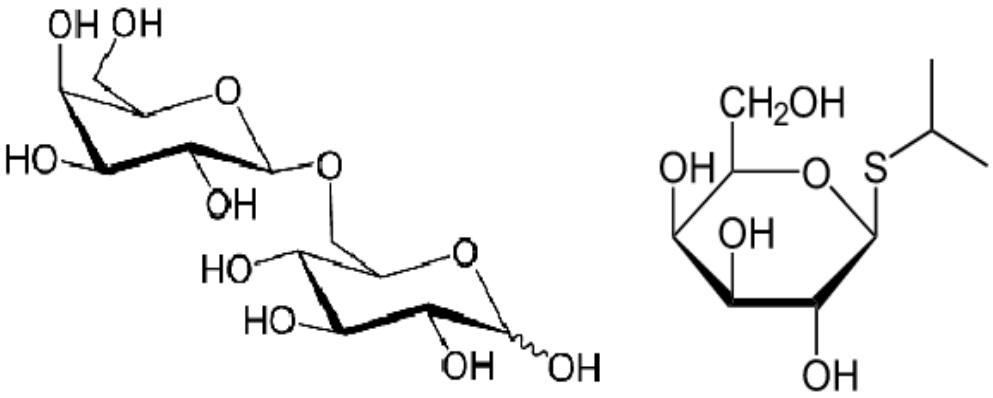


Figure 4.2. Diagram showing the structural differences between allolactose (left), and its structural non-metabolised analogue, IPTG (right).

Another commonly used inducible promoter system is the Tet system[69,114,118,119,126,132,133]. Tetracyclines are among the most commonly used class of broad-spectrum antibiotics. They are able to inhibit polypeptide elongation by binding to the small ribosomal subunit, although the exact mechanism is unknown[118]. As with most antibiotics, many Gram-negative bacterial species have developed resistance mechanisms to overcome

tetracycline's (Tc) lethality[134]. The most commonly used mechanism employs *tetA*, a membrane-associated protein that is able to actively export the antibiotic out of the cell, thereby stopping its action against the small ribosomal subunit[118]. High levels of *tetA* expression is lethal to the bacterial cell due to the nonspecific cation transport leading to a collapse of the cytoplasmic membrane potential, therefore regulation of *tetA* is under tight control[118,119]. Up-stream of *tetA* and divergently orientated is *tet*R, whose transcription is mediated by divergent promoters containing two palindromic *tet*R operators (*tetO1 and tetO2*) 11 bp apart[118,119] (figure 4.3). Expression of both *tetA* and *tet*R is controlled by the *tet*R gene product, a homodimer capable of binding to the palindromic DNA operator sequence which inhibits RNA polymerase II-mediated transcription[113,114,118,119,133,135]. As the mode of action of Tc is to inhibit protein synthesis, the *tet*R/*tetA* switch has to be particularly sensitive; meaning very low levels of Tc can induce *tet*R/*tetA* expression. In the absence of Tc, the *tet*R gene product is expressed and binds to the *tetO*s, inhibiting *tetA* and *tet*R expression (Figure 4.3A). However, in the presence of Tc, Tc coupled with Mg^{2+} binds to TetR causing a conformational change, decreasing the affinity of TetR to the *tetO*s by nine-orders of magnitude compared to TetR alone (Figure 4.3B)[118]. This difference in affinity allows the silent (TetR bound to *tetO*) and the active (TetR bound to $[MgTc]^{+}$) to be delineated and sub-inhibitory concentrations of Tc allow transcription to proceed. Due to the nature of Tc being an antibiotic, the Tc analogue anhydrotetracycline (aTc) is commonly used as a replacement. It has the higher affinity for TetR compared to Tc, but has a decreased toxicity[136].

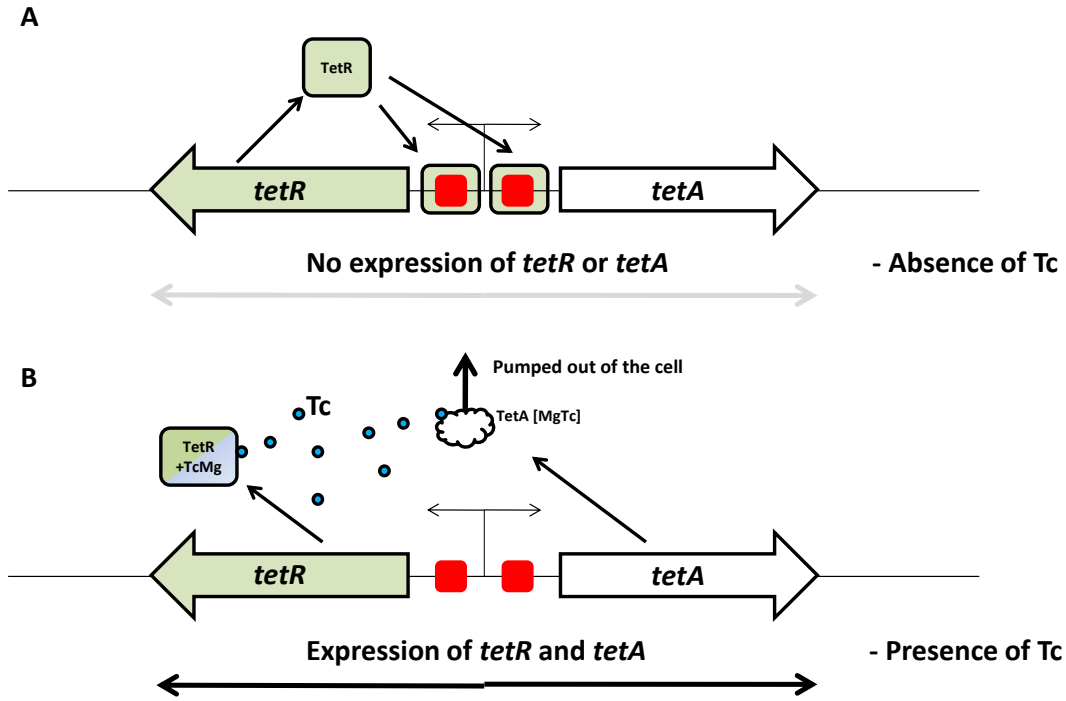


Figure 4.3. The natural *tetA* and *tetR* repressor system of *E. coli*. (**A**) In the absence of Tc, the TetR repressor (olive square) is able to bind to the two *tetO* operator sites (red squares) inhibiting RNA polymerase binding to the divergent promoters which control transcription of *tetA* and *tetR*. (**B**) In the presence of Tc (blue circles), Tc binds to Mg^{2+} which subsequently binds to TetR triggering a conformational change which decreases TetR's affinity to the *tetOs* by nine orders of magnitude, allowing binding of RNA polymerase II and transcription of *tetR* and *tetA*. As a consequence, TetA (white cloud) binds to $[MgTc]^{+}$and pumps the Tc out of the cell, thereby conferring resistance to Tc.

The Tet system is unique in that with a few base pair mutations of the class B TetR, TetR shows an increased affinity to *tetO*, instead of the aforementioned decrease upon binding of the inducer to the effector protein[133,137]. This reverse Tet repressor depends

on the specific mutations, originally identified using random mutagenesis, causing TetR to form a non-inducible conformation when no Tc is present (Figure 4.4A)[137]. However, when $[MgTc]^+$ binds to TetR, the protein undergoes a conformational change which increases its affinity to *tetO* (Figure 4.4B). Due to these characteristics, the Tet- and Tet-reverse systems are the most efficient inducible systems for transcriptional regulation currently known; as such they are commonly used for targeted gene regulation in prokaryotes and eukaryotes[118].

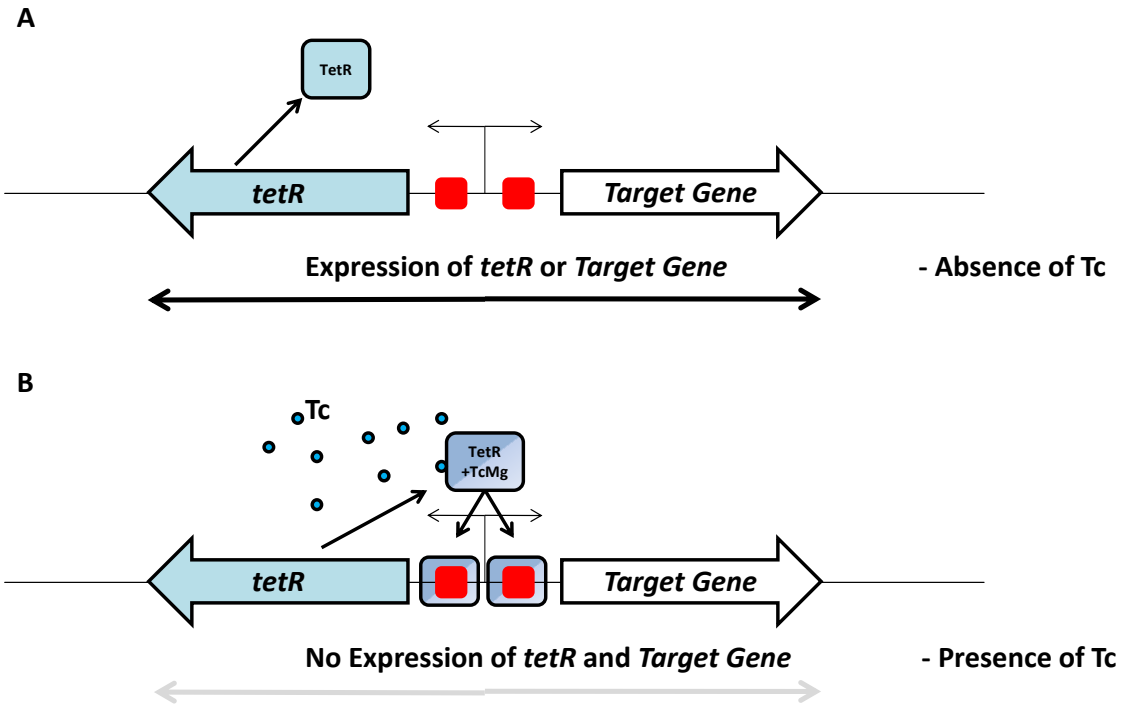


Figure 4.4. The synthetic revTetR system. (**A**) In the absence of Tc, the revTetR repressor (light blue square) is unable to bind to the *tetO* (red squares) sequences. (**B**) In the presence of Tc (blue circles), Mg^{2+} binds to Tc which subsequently binds to the revTetR protein causing a conformational change (two tone blue square) which increases revTetR affinity to *tetO*, bringing about binding. This binding of the revTetR repressor to its complementary *tetO* sequence inhibits transcription of the target genes.

4.1.2 Riboswitches

Riboswitches are natural RNA aptamers that are able to modulate translation of a high number of bacterial metabolic genes in response to small molecule ligands without the need for proteins[1,97,122]. There are two structurally significant features of riboswitches; the aptamer domain which recognises the ligand, and the expression platform which links the ligand-binding to a change in gene expression[1]. Recent studies have identified a variety of synthetic riboswitches which respond to nonendogenous small molecule ligands[138]. Synthetic riboswitches can be, in principle, engineered to respond to any permeable, non-toxic molecule that can interact with RNA. As such, synthetic riboswitches have great potential to detect and respond to theoretically any small molecules in a ligand-dependant fashion to modulate gene expression levels of any target gene.

Identification of aptamers that bind to small molecules have been previously demonstrated[123], however, developing the aptamers into a functional riboswitch has proved difficult in the past. A procedure called SELEX (systematic evolution of ligands by exponential enrichment) allows isolation and amplification of RNA or DNA nucleotides with selective affinity for specific small molecule ligands[117,138]. This technique has allowed the identification of RNA oligomers that are capable of binding with high affinity and specificity to both proteins and small molecule targets including; T4 DNA polymerase, adenosine-5'-triphosphate (ATP), theophylline, and several amino acids[117]. The theophylline ligand is an unusual ligand compared to the others as theophylline is not found in prokaryotes, as such; there are no natural aptamers in the bacterial genome. Identification of the theophylline ligand was carried out by generating a pool of 10^{14} RNA molecules that contain a 40-nucleotide region of random sequence[117]. Affinity chromatography was then carried out using theophylline cross-

linked to a sepharose column. Any RNA bound to the theophylline was converted to DNA, PCR amplified, then sequenced to ascertain the sequence of the bound RNA. From this the authors identified a number of RNA ligands capable of binding theophylline with an affinity of 10,000 times greater than caffeine, which only differs in its structure by a single methyl group at the N-7 position[117].

The ligand is only the first part of a riboswitch, the second is expression platform. Soukup and Breaker, (1999)[123], were able to bind a pre-existing catalytic and aptamer domain via a structural bridge. In this case, the authors bound aptamers of FMN (flavin mononucleotide), theophylline or ATP to a ^{32}P-labeled self-cleaving ribozyme with varying lengths of bridge sequence connecting the two[123]. Using these constructs the authors showed it was possible to construct precision RNA molecular switches that trigger translation only in the presence of their corresponding ligand[123]. This simultaneous use of rational and combinational approaches to enzyme engineering will prove to be a powerful approach in the design of new ribozymes and riboswitches with enhanced kinetic characteristics. This would be hugely advantageous for applications in directed evolution or metabolic engineering.

4.1.3 Aim of this study

Research into the molecular microbiology of *Clostridium* species has been limited by the lack of tools that most bacterial geneticists take for granted. Inducible systems are one of the most fundamental tools used in the biochemical, physiological and genetic studies, providing a wealth of information in a vast array of areas. However, inducible systems have remained problematic in *Clostridium* species, with very few shown to work or the inducers being metabolised, therefore inappropriate for our work.

The aim of this chapter was to develop a genetic tool able to modulate transcription or translation in the presence of a small molecule ligand that meets the criteria of tight repression in the absence of an inducer, low likelihood of gratuitous induction, and good bioavailability.

4.2 Results

4.2.1.1 Design and construction of a Clostridium lac *repressor system*

Over the years, a multitude of inducible expression systems have been developed for use in controlling gene expression in many bacterial species, unfortunately there had been no system developed in *C. difficile* that have met the required criteria for our use at the time of this study. There are several characteristics which would make an inducible promoter system desirable. Ideally, the promoter system would work in a number of *C. difficile* strains, and hopefully different species of *Clostridium* as there is a wide variety of *Clostridium* that are medically or industrially important. It would be easy to add and remove the inducer *in vitro* or *in vivo* depending on the environment of the *Clostridium* (mammalian cells or bioreactor). It would have a wide range of activity based upon the concentration of inducer added, combined with tight repression in the absence of the inducer. The bioavailability of the inducer must be high, i.e., the administered dose of the inducer that is inside the cell must remain unchanged and not metabolised by the cell. These characteristics would allow high production of the target gene's product for potential purification or analysis of the effects on phenotype or over production. Addition of the inducer would increase/decrease target gene expression rapidly for short-term studies on phenotype. The range of response to inducer would allow a specific controlled dose-responsive range of target gene expression, and tight repression would limit the 'leakiness' of the system and remove the possibility of toxic gene products becoming toxic to the cell, and in the case of essential genes, removal of the inducer would be lethal to the cell. Finally, the inducer concentration would remain unchanged over time allowing long-term studies of potential phenotypes.

A classical-example of a plasmid-based system that met the above requirements is the *lac* repressor system[68,126,129,139]. The *lac* repressor system is currently used in many other

bacterial species, but has not been adapted for use in *Clostridium* effectively[68,121,126,140]. The addition of the non-metabolised inducer IPTG, is easy, subsequent effects are rapid, and the system is known to be dose-dependent, meaning an intermediate expression level can be obtained accurately (figure 4.5)[126].

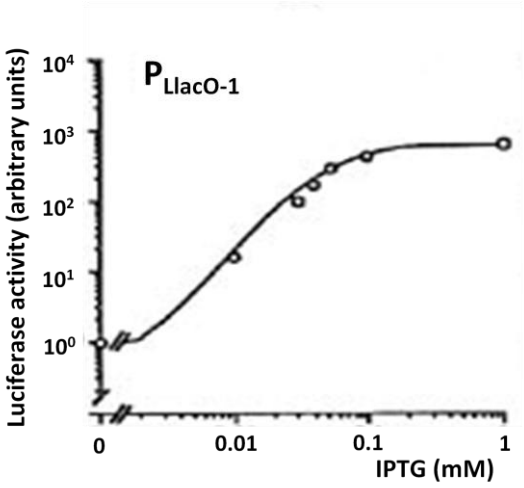


Figure 4.5. Example of dose-dependent curve of the IPTG *lacI* inducible promoter system ($P_{LlacO-1}$) from *E. coli* in response to varying concentrations of IPTG. Production is quantified as luciferase activity (arbitrary units). Adapted from reference[126].

Construction of the IPTG inducible promoter was carried out *in silico* using *Clostridium* components or components optimized for use in *C. difficile*. The functionality of the *lac* repressor system is based upon two main components; the LacI repressor, and the *lac*Os interrupting a promoter, which is placed in such a way that inhibition of transcription occurs upon repressor binding. As there is no natural *lacI* gene or *lac*Os present within *C. difficile*'s genome, the *lacI* gene from *E. coli* MG1655 was chosen to fulfil the expression of the repressor, as it is the best characterised. To increase the efficiency of translation in *C. difficile*, *lacI* was codon optimized by DNA2.0 using their proprietary algorithm.

It has been shown that a high ratio of repressor expression to operators is required for tight repression of the target gene[116]. Due to this requirement, an *rrnA* promoter from *C. sporogenes* was chosen due to its function; it should, and has been shown to be constitutively expressed throughout growth[41]. The usage of a *C. sporogenes* promoter over a *C. difficile* promoter was selected to decrease the possibility of homology to the chromosome, which may lead to homologous recombination between the plasmid and the chromosome. It is thought high-expression of LacI is toxic to the cell; therefore the *rrnA* promoter was 'weakened' by the removal of 1 or 2 bp from the spacer sequence between the -35 and -10 of the promoter (figure 4.6). The removal of the basepairs of the spacer sequence was based upon the findings of Jensen and Hammer, (1998)[141], that promoters with a 16 bp spacer were significantly weaker than their 17 bp counterparts. The basepair(s) selected for removal were chosen at random. However, during synthesis of the two constructs, it was found that it was not possible to synthesise pMTL-DW17 or pMTL-DW18 without errors in the *rrnA* promoters, possibly due to the high volume of LacI production causing toxicity. Because of this, we selected to use the mutant variations of the two promoters isolated by DNA2.0 (shown in figure 4.6.) to transcribe *lacI*.

PrrnA -35 -10
AAAAAAGTG**TTGACA**AGTTTAAAAAAAACTTG**TATACT**GTAAAAGTCG

PrrnA-1 -35 -10
AAAAAAGTG**TTGACA**AGTTTAAAAAAACTTG**TATACT**GTAAAAGTCG

PrrnA-2 -35 -10
AAAAAAGTG**TTGACA**AGTTTAAAAAACTTG**TATACT**GTAAAAGTCG

PrrnA-1* -35 -10
AAAAAAGTG**TGGACA**AGTTTAAAAAAACTTG**TATACT**GTAAAAGTCG

PrrnA-2* -35 -10
AAAAAAGTG**TTGACA**AGTTTAAAAAACTTG**TATTCA**GTAAAAGTCG

Figure 4.6. The original P_{rrnA} sequence and the two variations P_{rrnA-1} and P_{rrnA-2} showing the removed basepairs in the spacer sequence between the -35 and -10. Below those marked with asterisks are the two mutated promoters isolated with the mutations shown in red.

To obtain high levels of target gene transcription modulation, the ferredoxin promoter of *C. sporogenes* was chosen for the same reasons as the previous promoter. The positioning of the *lacO* within the context of the promoter is crucial for inhibition of RNA polymerase II binding and subsequent transcription of the target gene, because of this, two different *lacO* configurations were designed (Figure 4.7). The first *lacO* configuration is the perfect palindromic synthetic operator O_{ied}[142]. It has previously been shown that the synthetic O_{ied} has a higher affinity to LacI resulting in higher repression than the natural *lacO*s[139,142]. Due to this knowledge, the first inducible promoter was designed to have a single O_{ied} at the 3' of P_{fdx}, as this has been previously demonstrated to work effectively[68]. The second *lacO* configuration contains two *lacO*s, the O_1

(strongest natural *lacO*) in the same position as the other configurations O_{ied}, and the O_1 positioned 95 bp upstream (This was a mistake, the distance separating the two operators should have been 70.5 bp. It was not corrected due to reasons that become apparent in section 4.2.1.2) due to a study conducted by Muller *et al*, (1996)[139]., that showed that the highest levels of repression was identified when the inter-operator distance was 70.5 bp[139].

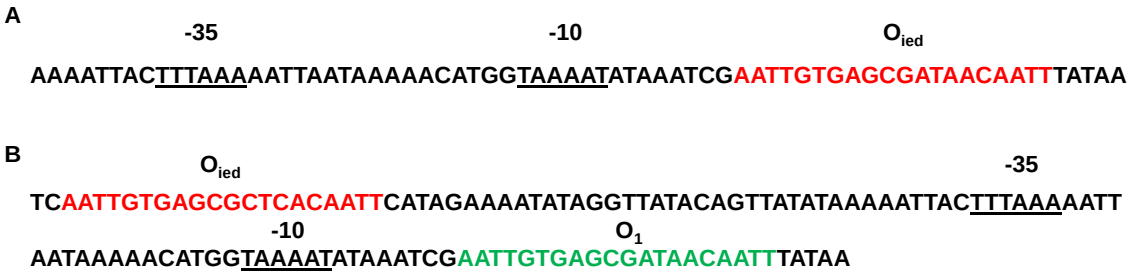


Figure 4.7. P_{thl} inducible promoter sequences interrupted with *lacO*s. (**A**) P_{thl} interrupted with the synthetic O_{ied} downstream from the -1 transcriptional start site. (**B**) P_{thl} interrupted with two *lacO* sequences; the first O_{ied} placed 95 bp upstream of the second O_1 operator, which is positioned in the same place as the O_{ied} from (A).

The consensus RBS (AGGAGGA) from the alanyl-tRNA synthetase of *C. sporogenes* was selected for translation initiation of both *lacI* and the target gene to be induced, and placed 7 bp upstream of the start codon of each gene. As the orientation of the *lacI* and the inducible promoter are on the same strand, this increases the probability of read through interfering with the transcription of *lacI* from a gene upstream, depending on the genetic context, and the target gene from *lacI*'s promoter. Therefore, two transcriptional terminators were identified from *C. tetani* using WebGeSTer DB – A Transcriptional Terminator Database, and were positioned at the 5' of the *lacI* P_{rrnA}, and 3' of *lacI*. To facilitate easy removal or substitution of each component in the system,

appropriate restriction sites were introduced between the components to allow easy cloning of new components, and to be compatible with the pMTL80000 modular vector series of plasmids[89] (Figure 4.8).

Finally, the purpose of this system is the systematic functional analysis of *C. difficile* genes using conditional mutations similar to the pMUTIN strategy previously published[45]. In this strategy, the inducible system is inserted into the chromosome between the target gene's promoter and start codon, thereby allowing the inducible system to modulate expression in response to a ligand[45]. Integration of the system into the chromosome of *C. difficile* requires employing an allelic exchange system using a negative selection marker such as *codA*. There are theoretically two possible outcomes after the second-crossover event; the allelic exchange vector being excised resulting in the parent genotype, or the desired allelic exchange event resulting in a recombinant chromosome. To ensure that the recombinant genotype is selected, a positive selection marker was inserted upstream of *lacI* to form an operon. This allows the rapid identification of the recombinant cells through the selection of *ermB* by growing on erythromycin after the loss of the integration plasmid. This was originally designed in one of the two constructs, however, with the restriction sites present; the *ermB* can be cloned into the other construct with ease (Figure 4.7B). The two different constructs shown in Figure 4.8 were synthesised by DNA2.0 and cloned into pMTL82254 using the restriction endonuclease *Not*I and *Nde*I, to create plasmids pMTL-DW17 and pMTL-DW18.

Once the constructs were synthesised, the double operator configuration of pMTL-DW18 was digested using the restriction endonucleases *Sal*I and *Aat*II, and cloned in to plasmid pMTL-DW17, using standard cloning techniques, to create plasmid pMTL-

DW17b. This allows the direct comparison of the operator configuration efficiency at repressing the *catP* reporter gene present upon pMTL82254.

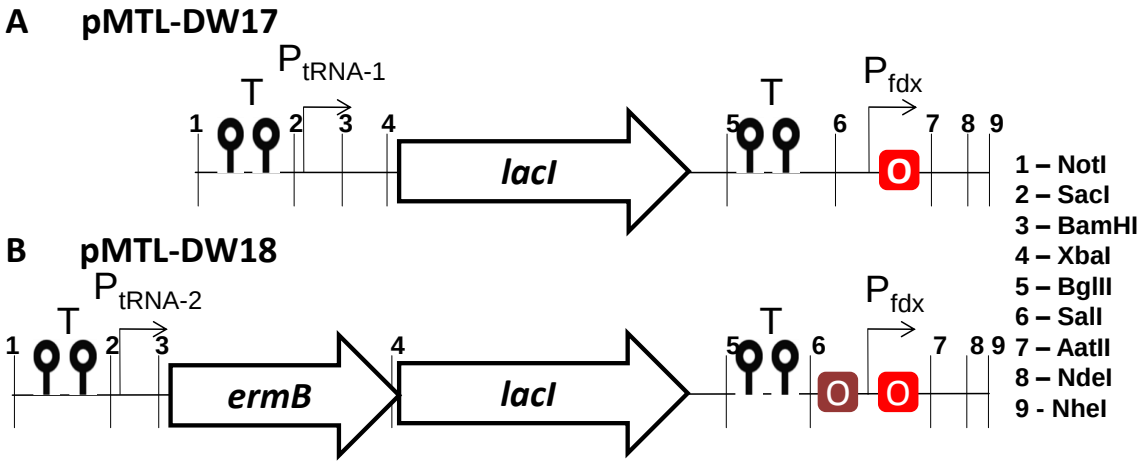


Figure 4.8. Schematic diagram showing the main features of the plasmids (**A**) pMTL-DW17 containing *lacI* and (**B**) pMTL-DW18 containing both *lacI* and *ermB*. Each of the vertical lines above which a number (1-9) is placed represents the placement of restriction sites (listed on the far right) designed into the system for quick switching of components. The double hooped structures labelled T are 2 bidirectional transcriptional terminators positioned to inhibit any read-through from natural promoters upstream or into the inducible promoter. The small arrows indicate the position of the two promoters (P_{tRNA-1}/P_{tRNA-2} and P_{fdx}) in each construct, and finally the red boxes with the letter O represent the positioning of the *lac*Os relative to the P_{fdx}.

4.2.1.2 Testing the Clostridium lac repressor system

In order to test if the synthetic constructs were constructed correctly and the system functions as expected, the plasmids pMTL-DW17, pMTL-DW17b, and negative control pMTL82254, were transformed into *E. coli* TOP10. The reason for carrying out the

initial testing in *E. coli* is because classical IPTG inducible promoter systems are known to work in *E. coli*, and therefore in theory, this system should also function in this host.

The quantification of CatP production showed that there was no large difference in expression or repression of *catP* between the plasmids pMTL-DW17 and pMTL-DW17b (Figure 4.8). It was noted that plasmid pMTL-DW17b had an increase in expression compared to pMTL-DW17 until 6 hours after IPTG induction, when expression levels started to decrease from its peak of 94,000 CAT U/mg protein, to 90,000 CAT U/mg protein, whereas pMTL-DW17 peaked at 8 hours with a CAT level of 102,000 U/mg protein. After 8 hours, both plasmids showed a sharp decrease in CAT U/mg protein. Each plasmid showed its greatest induction at 8 hours after induction; pMTL-DW17 with a maximum induction factor (expression/repression) of 8.45, and pMTL-DW17b with a maximum induction factor of 6.19. The downfall of this current system is the leakiness of *catP* expression without induction. However this does not equate to the systems characteristics in *C. difficile*.

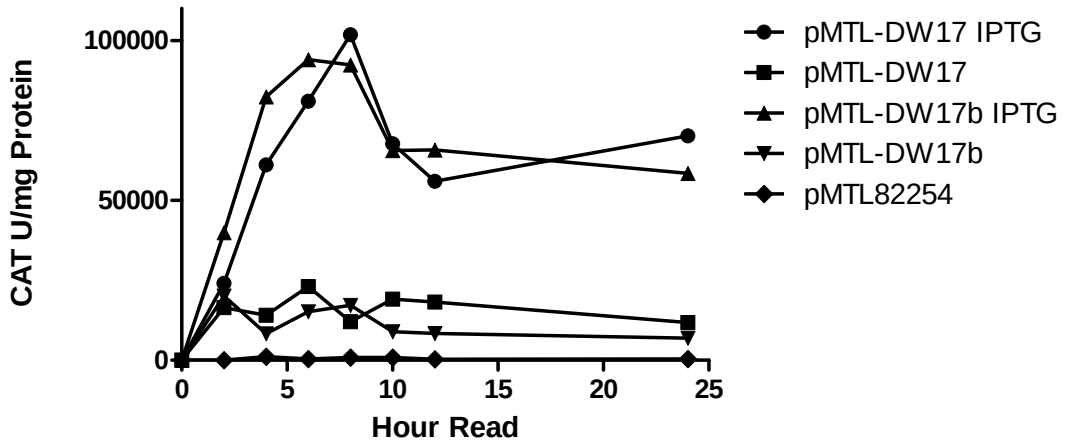


Figure 4.9. The production of CatP produced by plasmids pMTL-DW17 and pMTL-DW17b in *E. coli* with and without 1 mM IPTG over 24 hours. *E. coli* cultures harbouring the plasmids pMTL-DW17 and

pMTL-DW17b were grown to an OD_{600} 0.5 and either, allowed to continue growing or were induced with 1 mM IPTG. Each time point after induction (hour 0), CatP production was quantified (CAT U/mg protein). pMTL82254 is an empty vector control harbouring but not expressing *catP*. This data is only from a single experimental.

Using plasmids pMTL-DW17 and pMTL-DW17b, it has been shown that the system works in *E. coli*, however surprisingly there is no difference in induction factors between the single and double operator configuration as expected[139]. As the system has been shown to work, the decision to test it in *C. difficile* was taken. Due to the requirement of a positive selection marker present within the inducible promoter system as described above, the *ermB* gene was cloned into plasmid pMTL-DW17 yielding plasmid pMTL-DW18. To ascertain if the presence of the *ermB* gene upstream of *lacI* affects the production of CatP, the same experiment with pMTL-DW17b was carried out. The results showed that the presence of the *ermB* gene did not affect the modulated production of CatP, with no significant differences in induction factors recorded for either; pMTL-DW18, max induction factor 8.42, or pMTL-DW18b, max induction factor 6.45, compared to the absence of the *ermB* gene.

To test the effectiveness of the systems in *C. difficile*, plasmids pMTL-DW17 and pMTL-DW18 were conjugated into *C. difficile* using the standard conjugation procedure[90] and tested appropriately (Materials and Methods, section 2.11.1). The results showed the system did not repress in the absence of an inducer, with the expression of *catP* in the presence and absence of IPTG being equal for both pMTL-DW17 and pMTL-DW18 (figure 4.10). This suggests that the system is not being repressed, possibly due to low or no production of the repressor LacI. To test if the $P_{rrnA\text{-}1*}$ or $P_{rrnA\text{-}2*}$ are active in *C. difficile*, the promoters were amplified using primers PrrnA-F1 and PrrnA-R1,

introducing restriction sites *Not*I and *Nde*I at the 5' and 3', respectively. The PCR fragments were cloned into pMTL82254 and conjugated into *C. difficile* and analysed. The results showed that the mutations acquired during synthesis of pMTL-DW17 and pMTL-DW18 was deleterious to the promoter being viable in *C. difficile*, and as such, they are no longer active in the organism (Figure 4.10). This could account for the no repression identified in using pMTL-DW17 and pMTL-DW18.

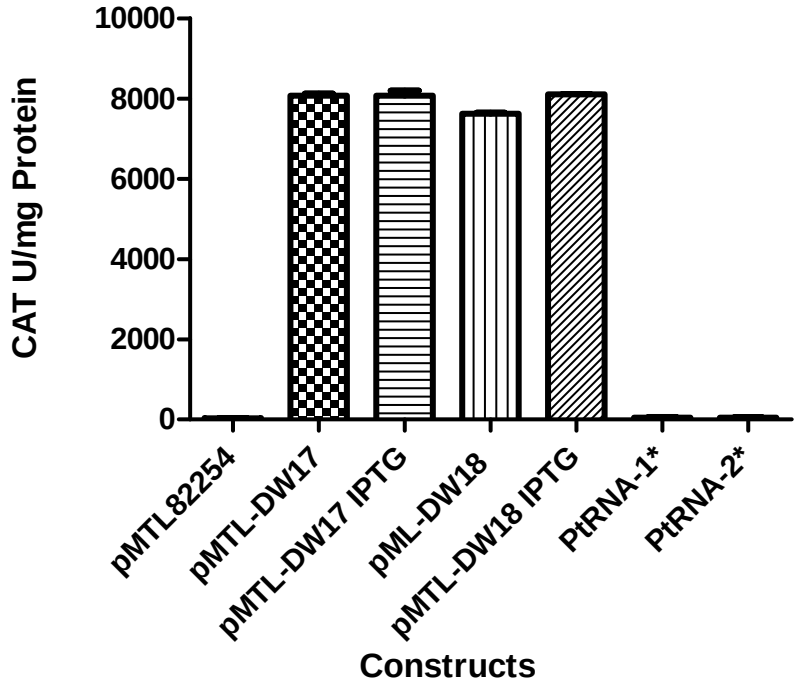


Figure 4.10. Comparison of the levels of CatP production between different promoters transcribing *lacI* in plasmids pMTL-DW17 and pMTL-DW18. *C. difficile* 630Δ*erm* harbouring the palsmids was grown to an OD_{600} 0.5 and either, induced with 1 mM IPTG or let to grow, for 4 hours, afterwhich CatP production was quantified (CAT U/mg protein). Also, the strengths of P_{rrnA-1} and P_{rrnA-2} transcriptionally fused to *catP* in pMTL82254 were accessed after 4 hours growth past OD_{600} 0.5 in *C.*

difficile 630Δ*erm*. The error bars indicate standard error of the means from three experiments.

To circumvent this problem, another set of promoters were selected to act as the promoter for *lacI* transcription. The *C. sporogenes* promoters; P_{fdx}, P_{hydA}, P_{recA} and P_{pyrE} were amplified using the forward and reverse primers specific to each promoter, named according to the promoter, e.g. P_{fdx} was amplified using *fdx*-F1 and *fdx*-R1. The promoters were then cloned into the construct pMTL-DW18 to replace the P_{rrnA-1}, using standard cloning techniques (Materials and Methods, section 2.9.4). To act as a positive control for reporter expression and inducible control for *lacI* expression, the aforementioned restriction enzymes were used to digest pMTL-DW18, and blunted with T4 polymerase to remove the promoter present, followed by a blunt end ligation. Each construct was conjugated into *C. difficile* and analysed as previously described, with an induction of 1 mM of IPTG for 4 hours, after which the cells were harvested and analysed using the CAT assay. The promoters used to transcribe *lacI* were also amplified from *C. sporogenes* and cloned into pMTL82254 at the 5' of *catP* to test for transcription activity in *C. difficile* and tested the same way as the promoter variations of pMTL-DW18.

Following analysis of the pMTL-DW18 *lacI* promoter variations, the results shown in figure 4.11 were unexpected. There was no measurable level of CatP production when the *lacI* repressor was expressed with the promoters: P_{fdx}, P_{hydA}, or P_{pyrE}, in the absence or presence of IPTG. However, when P_{recA} transcribes *lacI*, measurable levels of 1,562 CAT U/mg protein without IPTG, and 1,356 CAT U/mg protein induced with IPTG were detected, compared to no promoter transcribing *lacI* showing previously seen maximum levels of 7,256 CAT U/mg protein. The promoters were then cloned in front of *catP* to test the strength of each promoter. The strongest *C. sporogenes* promoter was P_{fdx}

showing 8,141 CAT U/mg protein, followed by P_{hydA} at 4,968 CAT U/mg protein, P_{pyrE} at 1,341 CAT U/mg protein, and finally P_{recA} at 146 CAT U/mg protein. The comparison of promoter strengths transcribing *lacI* shows that P_{fdx}, P_{hydA}, and P_{pyrE} are strong promoters in comparison to P_{recA} (Figure 4.12). This result suggests a super-repressed state is being observed when a mid to strong promoter is expressing *LacI*, and the super-repression is not able to be reversed with the addition of IPTG. However, when the weaker P_{recA} promoter is initiating transcription of *lacI*, there is a low expression of *catP* which again, is not affected with the addition of IPTG.

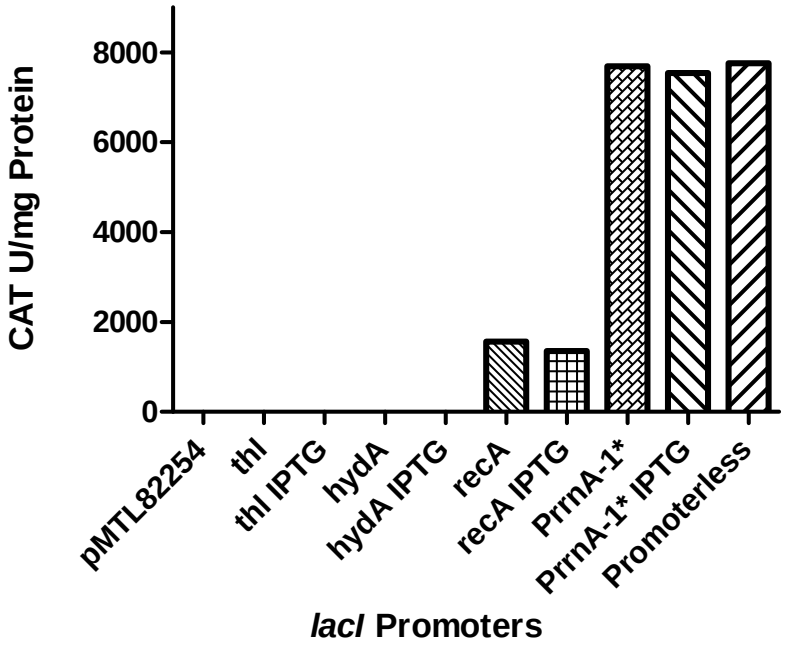


Figure 4.11. Comparison between the strengths of the different promoters transcribing *lacI* in their ability to produce CatP in *C. difficile* 630Δ*erm* in repsonse to IPTG. Each promoter was transcriptionally fused to *lacI* in vector pMTL-DW18 and grown in *C. difficile* 630Δ*erm* to an OD_{600} 0.5. Once the OD had been reached, 1 mM of IPTG was added to induce CatP production for 4 hours. Subsequently, CatP

production (CAT U/mg protein) was compared between cultures. This data is only from a single experiment.

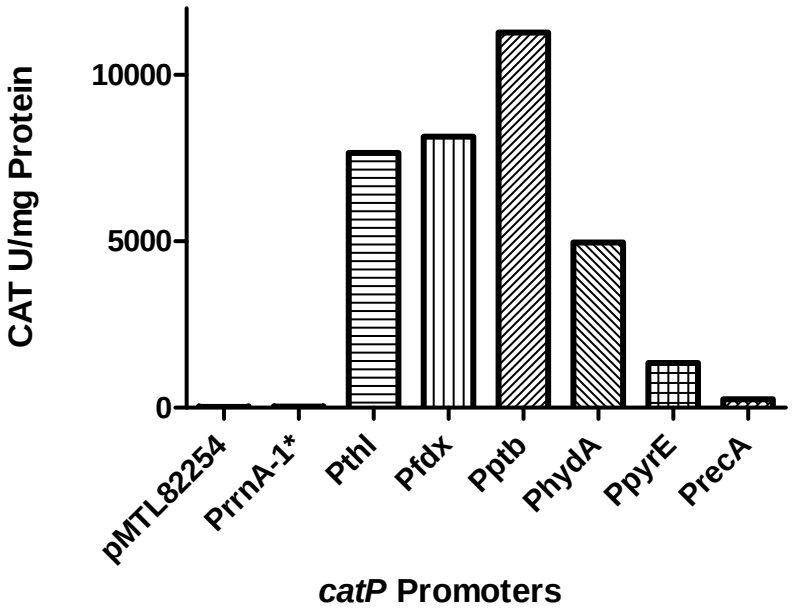


Figure 4.12. The strength of different promoters in their ability to produce CatP in *C. difficile* 630Δ*erm* Each promoter was transcriptionally fused to *catP* in vector pMTL82254 and CatP production was quatified (CAT U/mg protein) 4 hours after growth from OD_{600} 0.5. This data is only from a single experiment.

4.2.2.1 Design and construction of a Clostridium *theophylline-responsive riboswitch*

Riboswitches are RNA-encoded inducible elements, able to modulate gene translation in response to a ligand without the need for protein interaction[117,123,143]. The combination of a ligand recognition domain (aptamer) and an expression platform allows changes in gene translation in response to a ligand[123]. In nature, riboswitches are

widespread in both prokaryote and eukaryote genomes, being responsible for regulating gene expression for a number of metabolic processes[97]. In addition to naturally occurring riboswitches, the rise in synthetic riboswitches, identified through high-throughput screening techniques and rational design, to respond to endogenous and non-endogenous metabolites have provided great tools for directed evolution and metabolic engineering[117,123,143].

To choose a riboswitch for modulation of gene expression in *Clostridium* species, the same criteria that were required for the *lac* repressor system were sought. The riboswitch would ideally work in a number of *Clostridium* species, display good bioavailability, be easy to add/remove inducer, and addition of inducer would result in a subsequent increase/decrease in target gene expression. One drawback to using a riboswitch is there is no wide dynamic range of dose-dependent expression in response to a ligand due to each switch only being able to turn gene expression on or off. The small range available is due to the difference in number of mRNA molecules that are exhibiting the on or off conformation within a cell.

The synthetic theophylline-responsive riboswitch clone 8.1 developed by Lynch *et al*, (2006)[1], was chosen for testing in *C. difficile* due to its induction factor of 36 being the most efficient riboswitch published[1]. Their riboswitch was identified by composing a library using oligonucleotide-based cassette mutagenesis[1]. Mutagenic primers with degenerate regions were designed to create randomised spacer sequences between the previously published theophylline aptamer mTCT8-4, and the RBS of the IS*10-lacZ* reporter gene (figure 4.13)[1]. Using the sequencing data from the high-throughput screen, the authors were able to make presumptions on the possible mechanism of action for the riboswitch clone 8.1. Using the program *mFOLD*, they predicted the secondary structures of clone 8.1 in the region extending from the 5' of the aptamer, to

the 3' of the start codon shown in figure 4.13. They hypothesised that in the absence of the ligand, the aptamer is paired with the RBS and start codon, thereby inhibiting translation[1]. In the presence of the ligand theophylline, the experimentally determined secondary structure appears to free the RBS and the start codon, allowing translation to proceed. This appears in the calculated differences in Gibbs free energy between the two conformations, which is less than the difference of theophylline binding to the ligand. The difference suggests the conversion between the two conformations is thermodynamically favourable in the presence of theophylline. The mechanism of action is consistent with previous studies by de Smit and van Duin[144-146], who noted that secondary structure in close proximity to the RBS drastically reduces translation efficiency of the downstream mRNA[144-146].

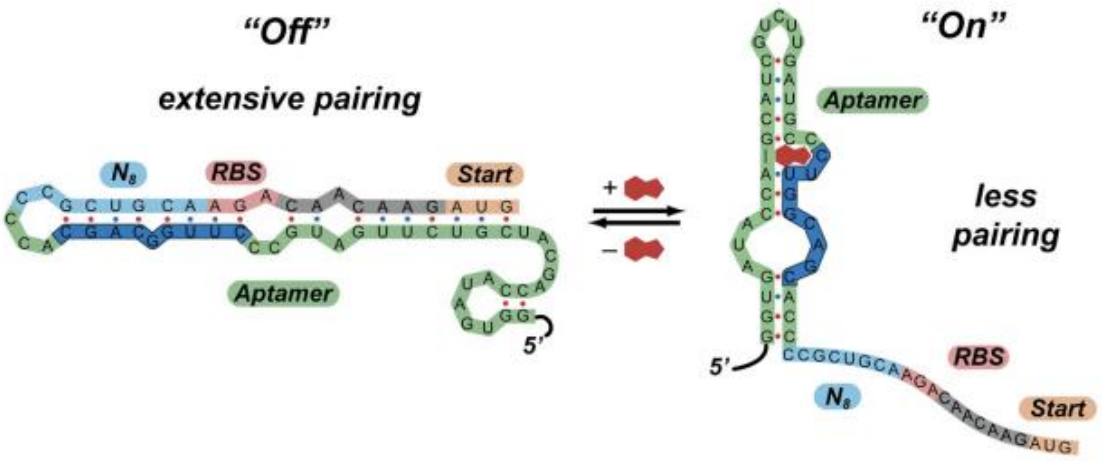


Figure 4.13. Predicted mechanism of action of the synthetic riboswitch clone 8.1 developed by Lynch *et al*, (2007)[1]. In the absence of theophylline (left) the secondary structure of the 5' mRNA binds extensively to 3' region which includes the RBS and start codon. In the presence of theophylline (right) the structure of the riboswitch shifts in such a way as to release the RBS and start codon, allowing translation to

initiate. The structure was predicted by *mFOLD* and confirmed by NMR[1]. Modified from reference[1].

Using this predicted mechanism of action, it is theoretically possible to make modifications to the sequence as long as the nucleotide secondary structure is complementary in both configurations, and the ligand is still able to bind to the aptamer. Using this theory, the sequence of the clone 8.1 riboswitch[1] at nucleotides 54 and 56 were changed from A and G to C and T, respectively, to introduce an *Nde*I restriction site for cloning. Due to the requirement of modifying one side of the secondary structure, the complementary nucleotides 18 and 20 were changed from C and T to A and G respectively to create Thio1 (figure 4.14). The restriction sites *Not*I, *Sac*I, and *Bam*HI were introduced at the 5' so the riboswitch could be cloned into the pMTL80000 series of modular plasmids[89] using *Not*I and *Nde*I, and an appropriate promoter could be cloned into the restriction sites *Sac*I and *Bam*HI. The primers ThioF2 and ThioR2 were synthesised and annealed together using the method described in section 2.7.13. The oligonucleotide fragment was then cloned into to vector pMTL82254 using the restriction sites *Not*I and *Nde*I. Subsequently, the thiolase and ferredoxin promoters from *C. sporogenes*, and *ptb*, *hydA*, *adc*, *DNAk*, *gyrA*, and CD0200 from *C. acetobutylicum*, were amplified using primers the appropriate primers and cloned into the pMTL82254 vector containing the riboswitch, using the restriction sites *Sac*I and *Bam*HI, yielding plasmids: pMTL-V1P3n, pMTL-V1P1, pMTL-V1P2, pMTL-V1P4, pMTL-V1P5, pMTL-V1P6, pMTL-V1P7, pMTL-V1P8, respectively.

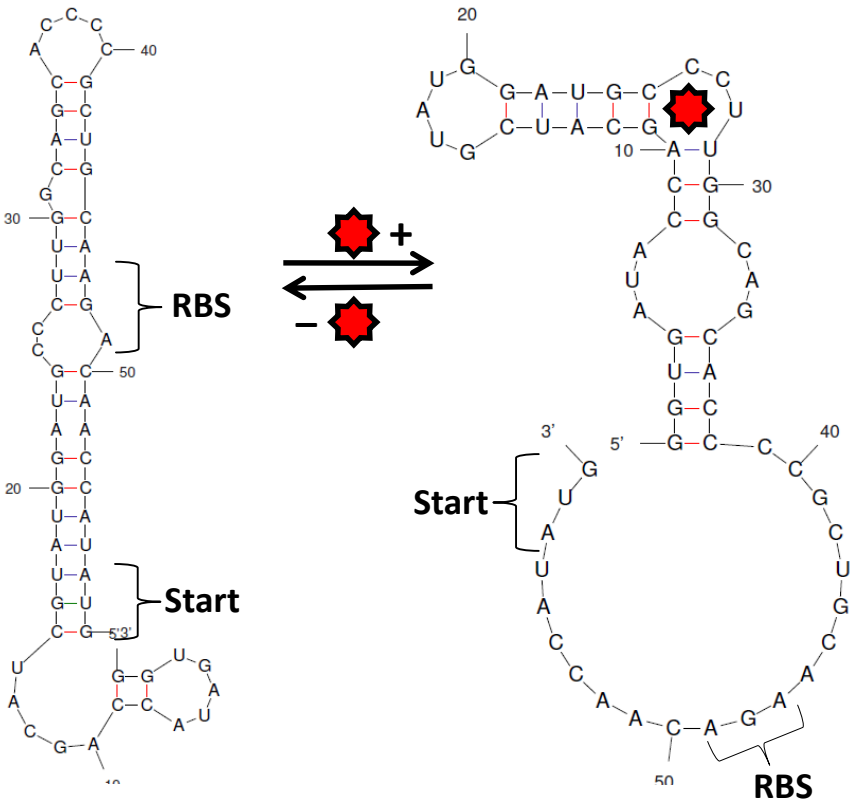


Figure 4.14. Predicted mechanism of action of Thio1. The mechanism of action is proposed to be the same as clone 8.1 shown in figure 4.13. The red star with black circumference represents theophylline, and aptamer binding location on the right conformation. Secondary structure predicted by *mFOLD*.

The RBS present within the Thio1 riboswitch, is to our knowledge, not found in *C. difficile*, as such, it is possible it might not be able to initiate translation of CatP. Therefore, the decision was taken to develop two variations of the riboswitch which contain different RBSs. In the first instance, the RBS AGAAGAA was chosen as it is most likely initiates transcription for *pyc* in *C. difficile*. To build the new RBS into the riboswitch, a total of 22 nucleotides were changed to keep the secondary structure the same as riboswitch Thio1, without changing the sequence or structure of the proposed aptamer. This also included the addition of three basepairs at nucleotides 54, 55, and 56.

To counter the addition of the three basepairs, the last complementary sequence between nucleotides 53 and 62 bp was increased by 2 bp, and the last nucleotide was not bound to a complementary nucleotide. The proposed structure and mechanism of action are shown in figure 4.15A, and the sequence alignment for Thio1 and Thio2 are shown in figure 4.15B.

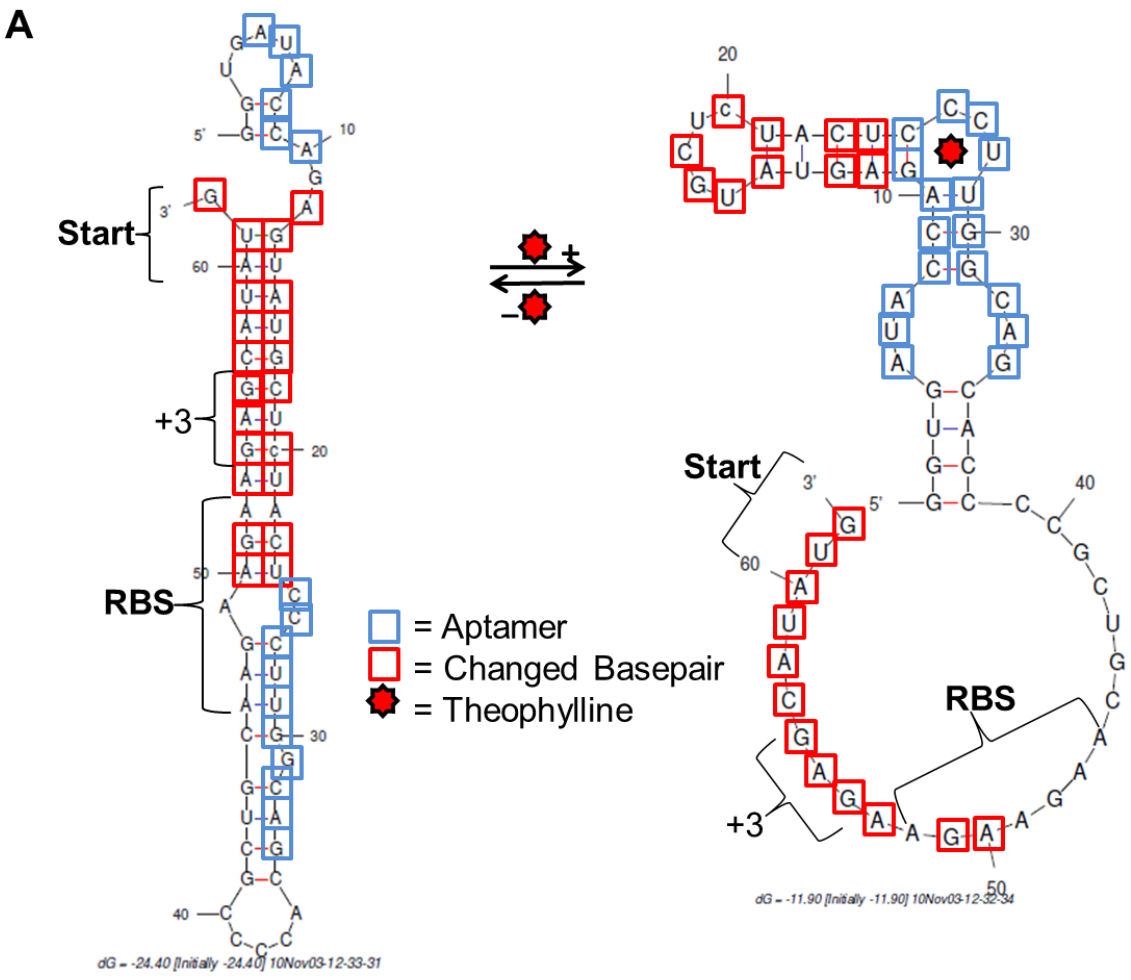
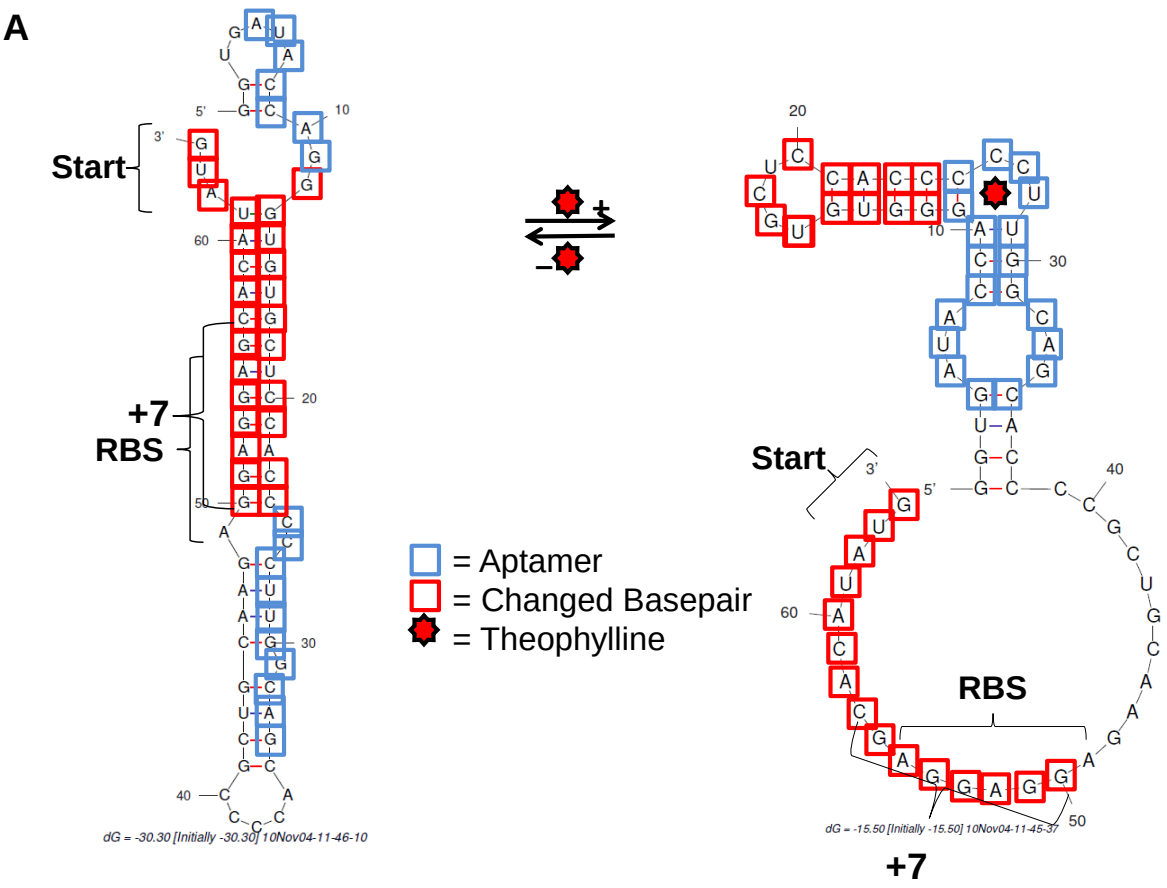


Figure 4.15. Predicted mechanism of action of Thio2. (**A**) The mechanism of action is proposed to be the same as clone 8.1 shown in

figure 4.13. The basepairs outlined in blue are the proposed theophylline aptamer. The basepairs outlined in red are the basepairs which were changed to introduce a new RBS and to keep the same complementary bases resulting in the same secondary structure. (**B**) An alignment of the riboswitch comparing Thio1, our original riboswitch, with the new RBS present upon Thio2. In total 13 basepairs were changed and 3 added to Thio2. Secondary structure predicted by *mFOLD*.

Another variation of Thio1 was developed using the RBS AGGAGGA, as this RBS is known to initiate translation within *C. difficile*[38,89]. The introduction of the new RBS into the riboswitch was based upon the same theory as the design of Thio2, where the RBS was designed around the aptamer and mechanism of action being unchanged, but in order to conform to the secondary structure, a total of 26 nucleotides were changed, including the addition of 7 nucleotides between nucleotides 50 and 57. This was required to allow the riboswitch to conform to the same secondary structure in the absence of theophylline as it is the predicted lowest energy conformation the riboswitch can fold into. This increases the complementary bound nucleotides at the 3' to 9, compared with 6 for Thio1, which also includes a mismatch at nucleotide 58. This results in the RBS not being bound into the theophylline free mRNA riboswitch structure, however, the RBS is, and should be sufficient to inhibit translation. The proposed structure and mechanism of action are shown in figure 4.16A, and the sequence alignment for Thio1 and Thio2 are shown in figure 4.16B.

Each newly designed riboswitch was cloned into pMTL82254 using the restriction sites *Not*I and *Nde*I and subsequently the same thiolase promoter used to promote transcription of pMTL-V1P3n was cloned into Thio2 and Thio3, making plasmids pMTL-V2P3n and pMTL-V3P3n.



B

```
Thio1    1 GGTGATACCAGCATCGTATG---GATGCCCTTGGCAGCACCCCGCTGCAA     47
           ||||||||||||   .||.||   .|..||||||||||||||||||||||
Thio3    1 GGTGATACCAG---GGTGTGCTCCACCCCCTTGGCAGCACCCCGCTGCAA     47

Thio1   48 GA-------CAACCATATG     59
           ||       ||  ||||||
Thio3   48 GAGGAGGAGCA--CATATG     64
```

Figure 4.16. Predicted mechanism of action of Thio3. (**A**) The mechanism of action is proposed to be the same as clone 8.1 shown in figure 4.13. The basepairs outlined in blue are the proposed theophylline aptamer. The basepairs outlined in red are the basepairs which were changed to introduce the 2nd RBS variation but keeping the same complementary bases resulting in the same secondary structure. (**B**) An alignment of the riboswitch comparing Thio1, our original riboswitch, with the new RBS present upon Thio3. In total 26 basepairs were

changed and 7 added to Thio3. Secondary structure predicted by *mFOLD*.

4.2.2.2 Testing the Clostridium *theophylline-responsive riboswitch*

In the first instance, we decided to test Thio1 with a variety of different promoters to identify the induction factors of each promoter in *E. coli*. To do this, we transformed each of the 7 promoter variations of Thio1 into *E. coli* TOP10, along with negative control pMTL82254. The results showed that the Thio1 promoter variants were able to repress production of CatP tightly in the absence of theophylline, and induced it in response to addition of theophylline (figure 4.17). The strongest CatP production was pMTL-V1P5 (*DNAk);* with a peak production of 3,955.27 CAT U/mg protein, an induction factor of around 20.5, came as no surprise as it has been shown to be an extremely powerful promoter compared to other clostridial promoters, however, it has been suggested that *DNAk* may not be constitutively expressed in *C. acetobutylicum* (Hengzheng Wang, Unpublished). Thus, the second strongest promoter, thiolase, was selected for studying the other variations of riboswitches as it is known to be constitutively expressed[41].

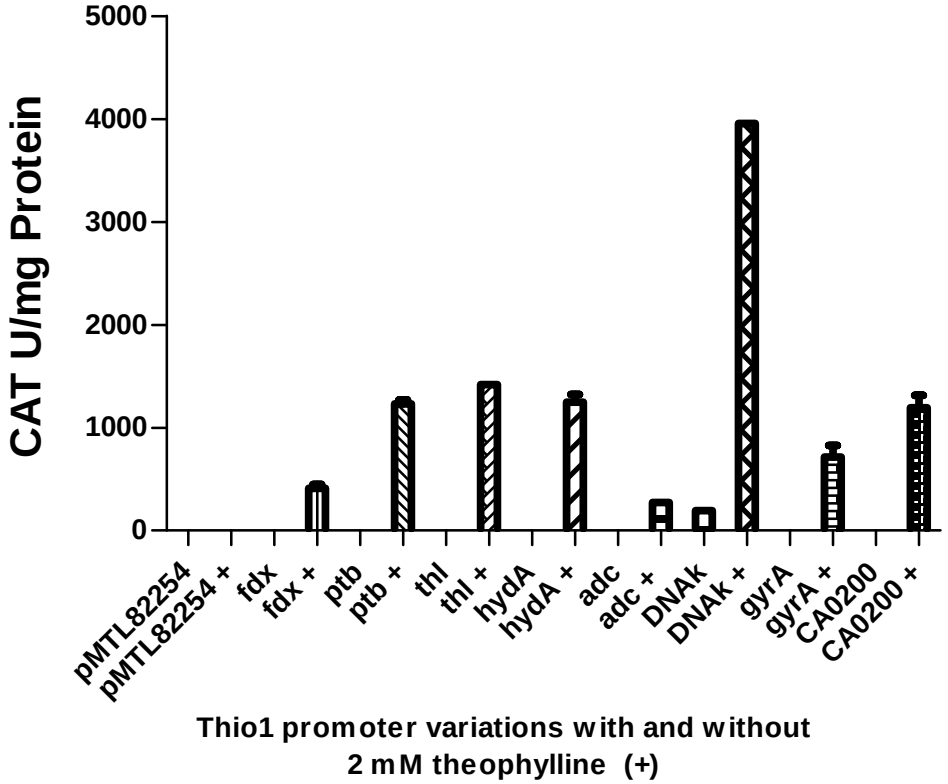


Figure 4.17. Comparison of the strength of different promoters placed upstream of the Thio1 riboswitch which is transcriptionally fused to the *catP* reporter system present on pMTL82254 within *E. coli*. Each Thio1 promoter variation was accessed in its ability to produce CatP (CAT U/mg protein) in the presence (+) or absence of 2 mM theophylline after 4 hours from induction at OD_{600} 0.5. The error bars indicate standard error of the means from three experiments.

To test whether the two rationally designed riboswitches would work in *E. coli*, Thio2 and Thio3 were tested using the thiolase promoter in the same way as the promoter variations in Thio1. The results showed that pMTL-V2P3n was able to repress CatP production to levels below the detection of the photo spectrometer, and upon addition of 2 mM theophylline was able to produce CatP to a level higher than the original pMTL-V1P3n (figure 4.18). Unfortunately, pMTL-V3P3n was not as successful and was

not able to repress CatP production in the absence of theophylline. As Thio2 was able to achieve greater induction than Thio1, the dynamic range of *catP* expression was characterised in response to varying concentrations of theophylline. To do this, pMTL-V1P3n and pMTL-V2P3n were selected and treated in the same way as previously, except at OD_{600} 0.5, the cultures were induced with either: 0, 1, 2, 5, 10, or 20 mM theophylline, and incubated for 4 hours, after which CatP production was quantified. The results were surprising in the fact that the riboswitches had a degree of dose-dependence in response to increased concentration of theophylline up to 10 mM, which delivered the maximum production of CatP. At 20 mM, the production of CatP was reduced by around 50% compared to 10 mM. In light of these results for both Thio1 and Thio2 in *E. coli*, the decision was taken to test the riboswitches in *C. difficile*.

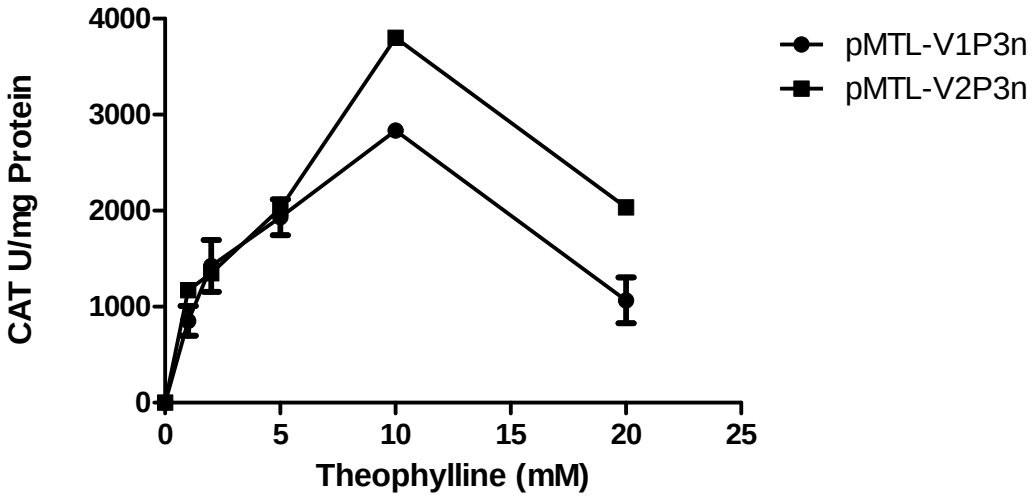


Figure 4.18. Comparison of pMTL-V1P3n and pMTL-V2P3n in their ability to produce CatP in response to varying concentrations of theophylline in *E. coli*. The P_{thl} was placed at the 5' of Thio1 and Thio2 riboswitches which are transcriptionally fused to *catP* present on plasmid pMTL82254. At OD_{600} 0.5, the cultures were induced with varying concentrations of theophylline (0, 1, 2, 5, 10, 20 mM) for 4 hours, after

which CatP production was quantified (CAT U/mg protein). The error bars indicate standard error of the means from three experiments.

In the first instance, the thiolase promoter was retained and a simple experiment to ascertain whether the riboswitches would work was carried out by growing a starter culture to OD_{600} 0.5 and inducing with 10 mM theophylline for 4 hours, after which CatP was quantified. Unfortunately, neither riboswitches were able to elicit a response in the presence or absence of theophylline. In case the problem was promoter related, a selection of promoters including the ferredoxin promoter and DNAk were cloned into the riboswitches and varying concentrations of theophylline was used to induce the cultures, as well as carrying out longer induction times. Thio1 and Thio2 were still, however, unable to express CatP to a detectable level. As it could be possible the RBSs are not able to initiate translation in *C. difficile*, the RBSs alone for Thio1 and Thio2 were cloned downstream of the thiolase promoter and placed in the appropriate position to express the reporter of pMTL82254. In the case of Thio1's RBS, no production of CatP was detected; however, the RBS of Thio2 was able to produce 7,200 CAT U/mg protein in combination with the thiolase promoter, showing that Thio2's RBS does work as expected in *C. difficile*, and in-line with previous results (figure 4.11).

4.2.3.1 Design and construction of a Clostridium Tet inducible promoter system

Our preceding attempts to develop a *Clostridium* genetic tool able to control transcription using a *lac* repressor system, and translation using a theophylline-responsive riboswitch, had proven challenging. A number of other inducible systems were considered for use in *C. difficile*, including the $P_{Xyl\text{-}xylR}$-inducible promoter from *B. subtilis*, and the arabinose-inducible P_{BAD} promoter system from *E. coli*. Both, however, have deficiencies. The xylose-inducible promoter system is able to be repressed by the presence of glucose, which is commonly found within mammalian cells, therefore

invalidating its usage for *in vivo* studies[114]. This repression cannot be overcome easily by a higher concentration of the inducer, making the system un-controllable. The arabinose-inducible P_{BAD} promoter system has been shown to be ineffective in *S. aureus*[114] and *B. subtilis*, possibly due to poor penetration of arabinose into the Gram-positive cells. The possibility of poor penetration does not bode well for the probability that the same occurrence would happen in *C. difficile*; therefore the selection of a system used commonly in Gram-positive bacteria would be advantageous.

The Tet inducible promoter system, developed for use in *B. subtilis* and *E. coli* has been shown to work in many organisms including; S. *aureus*[114], *M. tuberculosis*[113], *Streptococcus intermedius*[147], and most recently *C. difficile*[120], it has even been shown to work in transgenic plants[148] and mammalian cells[149]. The broad-range of hosts which this system is known to work in, the knowledge that tetracycline does in fact enter into the *Clostridium* cell, and has recently been shown to work in *C. difficile*[120], means the system should work efficiently in *Clostridium*.

Design of the *Clostridium* Tet inducible promoter system was based upon the architecture of the pRMC2 plasmid[132], which in turn is based upon pALC2073[114], derived from pWH353[69], the original tetracycline-inducible promoter system. The tetracycline-inducible $P_{xyl/tetO}$ promoter from pWH353 was derived from the Tet regulatory element originating from the transposon Tn*10*[69].

All of these derivatives contain the $P_{xyl/tetO}$ inducible promoter which transcribes the target gene, and the *tet*R repressor gene transcribed from the divergently orientated P_{tetR}. One interesting point is the positioning of a second tetracycline operator between the -35 and -10 of the P_{tetR}. This means that in the absence of Tc, there is a negative feedback loop, repressing both the *tet*R gene and the gene of interest, however, in the presence of low levels of Tc/aTc, the repression is alleviated quickly due to the low expression of

TetR, allowing expression of both TetR and the gene of interest. This placement of the second *tetO* within P_{tetR} means that the induction of $P_{xyl/tetO}$ is very sensitive, giving the inducible promoter a limited intermediate range (figure 4.19)[126].

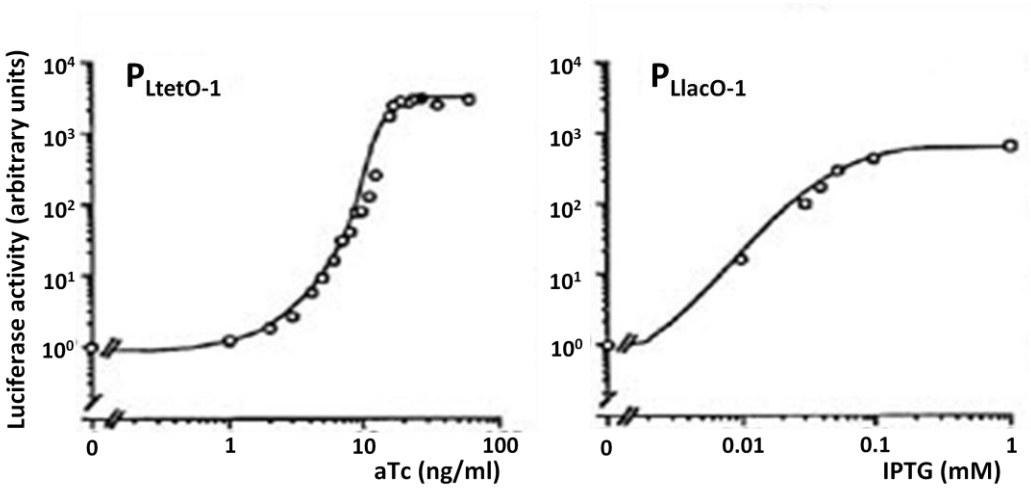


Figure 4.19. A comparison between the dose-dependent activation of transcription and the dynamic range of the Tet inducible promoter system (left) and the *lac* repressor system (right) in response to varying concentrations of aTc and IPTG in *E. coli*, respectively. Transcription was quantified by luciferase activity (arbitrary units). Adapted from reference[126].

As previously discussed, the ideal inducible promoter system should possess a wide dynamic range. Consequently the intermediate induction capabilities of the $P_{xyl/tetO}$ would be in appropriate. To circumvent this limitation, the Tet inducible promoter system was redesigned using *Clostridial* components and incorporating components that could be easily changed, in keeping with the compatibility of the pMTL80000 modular vector system[89].

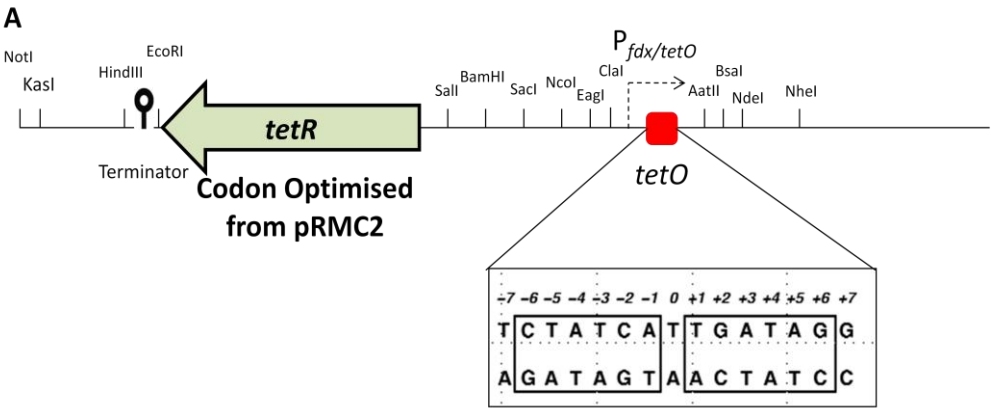
The Tet system is constructed from three main components, the inducible promoter which will be interrupted with a *tetO*, the *tetR* repressor, and the *tetR* repressor's

promoter. For construction of the *Clostridium* Tet inducible promoter system, the same design principles used for the *lac* repressor system were adopted. For the inducible promoter, the *C. sporogenes* ferredoxin promoter, previously used in our *lac* repressor system, was selected. To make the promoter inducible, two different promoter/*tetO* configurations were developed; the first consisted of inserting a 15 bp palindromic *tetO* fragment between the -35 and -10 of the ferredoxin promoter, as this positioning has been previously published as being effective in other bacteria[118]. The second was to introduce two *tetO*s, the first *tetO1* was inserted in between the -35 and -10 of the ferredoxin promoter, and the second *tetO2* was inserted downstream of the -10 (Figure 4.20B). The double operator configuration was chosen to be tested in parallel to the single operator configuration, as the double operator configuration has been previously described to increase the repression of the inducible promoter, thereby reducing the leakiness in the absence of an inducer[69]. As there is no need to synthesis the whole construct to include only the second operator configuration, the *tetO1/2* was synthesised as a separate construct with the restriction endonucleases *Cla*I and *Aat*II present at the 5' and 3' respectively, to enable easy switching of $P_{fdx/tetO}$.

For the repressor protein, the *tet*R gene of pRMC2 was selected and codon optimised for usage within *C. difficile* by DNA2.0[132]. Initially, no promoter was selected to initiate transcription of *tet*R for two reasons, the first being the companies constructing the synthetic constructs have encountered problems in the past when synthesising constructs which contain repressors that are expressed at high levels, and secondly, it was predicted that testing if the inducible promoters are able to initiate transcription of the reporter gene, it would be more straightforward in the absence of a repressor.

To stop unwanted transcription from promoters upstream of the inducible promoter system interfering with the regulation of the system, a dual transcriptional terminator

identified from *C. tetani* was positioned upstream of the *tetR* gene, at the beginning of the inducible promoter construct. Finally, to enable quick introduction and switching of components of the system, a number of restriction sites were introduced into the designs which are not found within the pMTL80000 series of plasmids[89] (figure 4.20A).



B

Original P_{fdx} – AAATTACTTTAAAAATTAATAAAAACATGGTAAAATATAAATCGTATAAAGTTGT

$P_{fdx/tetO}$ – AAATTACTTTAAAA**TCTATCATTGATAGG**GTAAAATATAAATCGTATAAAGTTGT

$P_{fdx/tetO1/2}$ – AAATTACTTTAA**ACTCTATCATTGATAGAGT**AAAA**TCCCTATCAGTGATAGAGA**

Figure 4.20. Schematic diagram of the main features of the tetracycline inducible promoter system and its operator configuration placement. (**A**) The construct pMTL-*tetO* synthesised by DNA2.0 with the *tetR* gene on the reverse strand codon optimized for *C. difficile* and taken from pRMC2, also shown is the $P_{fdx/tetO}$ configuration (shown magnified) and the placement of the restriction endonuclease sites shown above their respective place. The hoped structure at the 5' of *tetR* is a bidirectional transcriptional terminator to inhibit transcriptional read-through interfering with genes upstream. (**B**) The original P_{fdx} sequence at the top followed by the two *tetO* configurations. The bold red text represents the *tetO* present within $P_{fdx/tetO}$, and the *tetO1* and *tetO2* present within

$P_{fdx/tetO1/2}$. Underlined are the proposed -35 and -10 sequences of the P_{fdx} and its variations.

Another application that was built into the design, but unfortunately not tested due to time constraints, was the ability to reverse the function of the Tet system. It has been previously found that a few mutations in the TetR-based eukaryotic transactivator can reverse the function of the inducer aTc, from an inducer into a co-repressor[150,151]. To explore if the TetR protein can be converted to a revTetR protein by mutagenesis, Scholz *et al*, (2004)[137], set about creating a *tet*R mutant library in *E. coli*, from which more than 100 TetR mutants were identified that required aTc to function as a repressor[137]. Their most notable TetR mutant, r1.7, was capable of reaching induction factors of 102[137]. It was found that the mutations involved in the reversal of TetR were substitutions at amino acids, 15, 17, and 25, from glutamic acid, leucine, leucine, substituted to, alanine, glycine, and valine, respectively. These changes caused the native TetR conformation to have a low affinity to *tetO*; however, the binding of aTc increased the affinity to *tetO*, increasing binding and subsequently repressing transcription[137]. Incorporation of this into the revTetR system was carried out by synthesising a fragment of DNA which could be cloned into the *tet*R gene between the restriction endonuclease sites *Bpu*10I and *Xba*I. Within this fragment were 4 bp changes which encode for the substitute amino acids, changing the original TetR into Scholz's clone r1.7 revTetR[137] (figure 4.21).

A. TetR

TCTAGACTTGATAAAAGTAAAGTTATCAATAGTGCATTGG**A**ATTA**CT**AAATGAAGTTGGAATAGAAGGA**C**TAACTACAAGAAA
GTTAGCTCAGAAACTTGGAGTAGAACAACCTACGTTGTATTGGCATGTGAAGAATAAGAGAGCTTTACTTGAC**GCCTTAGC**

TTG GAA TTA CTA AAT GAA GTT GGA ATA GAA GGA CTA
L E L L N E V G I E G L

B. revTetR

TCTAGACTTGATAAAAGTAAAGTTATCAATAGTGCATTGG**C**ATTA**GG**AAATGAAGTTGGAATAGAAGGA**G**TAACTACAAGAA
AGTTAGCTCAGAAACTTGGAGTAGAACAACCTACGTTGTATTGGCATGTGAAGAATAAGAGAGCTTTACTTGAC**GCCTTAGC**

TTG G**C**A TTA **GG**A AAT GAA GTT GGA ATA GAA GGA **G**TA
L **A** L **G** N E V G I E G **V**

Figure 4.21. Overview of the changes to the DNA sequence to reverse the function of TetR into a revTetR system. (**A**) The original TetR DNA sequence with the changes underlined and red. In bold black are the recognition sites for the restriction enzymes *Bpu*10I and *Xba*I which can be used to replace the fragment. Underneath is a close-up of the DNA sequence on top with the required changes in red, and the original amino acid sequence in red below. (**B**) The synthesised fragment which can be switch with the original to reverse the function of TetR. On the top is the DNA sequence outlining the changes undertaken to reverse the function flanked by the appropriate restriction recognition sites from the original (A). Below is a close-up of the DNA sequence with the implemented changes showing the subsequent change in codons; E, L, L, to A, G, V, respectively.

The final construct was synthesised by DNA2.0, which includes the $P_{fdx/tetO}$ configuration, termed DNA2.0-*tetO*. The second $P_{fdx/tetO1/2}$configuration and revTetR fragments were synthesised separately. Once the constructs were received, the construct was removed from plasmid DNA2.0-*tetO* and introduced into plasmid pMTL82254

using restriction endonucleases *Not*I and *Nde*I, yielding plasmid pMTL-*tetO*. For comparison of the two difference operator configurations, the *tetO*1/2 configuration was digested from the DNA2.0-*tetO*1/2 vector and cloned into pMTL-*tetO* using restriction endonucleases *Cla*I and *Aat*II, yielding plasmid pMTL-*tetO*1/2.

The plasmids now have the inducible promoter able to transcribe the reporter gene *catP*; however, at this stage the inducible promoter is not inducible as the TetR repressor is not expressed. To allow expression of TetR, the thiolase promoter from *C. sporogenes* was selected to transcribe *tetR*, due to its known strong constitutive properties. The resulting DNA fragment was cloned into pMTL-*tetO* and pMTL-*tetO*1/2 using restriction endonucleases *Sac*I and *Bam*HI, yielding the final plasmids pMTL-tet3nO and pMTL-tet3nO1/2, respectively. Nomenclature: tet represents the tetracycline inducible construct, 3n represents $P_{thl,}$ and finally the O represents the single operator configuration, and *tetO*1/2 the double.

4.2.3.2 Toxicity of anhydrotetracycline in Clostridium *species*

Before testing of the tetracycline inducible promoter system can be undertaken, the bacteriostatic activity of aTc on the *Clostridium* species used in this study must be evaluated. Tetracycline is a known bacteriostatic antibiotic; however, aTc is thought to be less toxic to cells than its original counterpart[136]. To test if this is the case, the plasmid pMTL-tet3nO was digested with restriction enzymes *Cla*I and *Aat*II, blunted with T4 polymerase, and re-ligated together using T4 ligase. This yields plasmid pMTL-tet3n, which includes the tetracycline inducible promoter system minus the inducible promoter upstream of the *catP* reporter to be used as a control. *C. difficile*, *C. sporogenes*, and *C. beijerinckii* were then conjugated with *E. coli* CA434 harbouring the plasmid to transfer pMTL-tet3nO and were subsequently subjected to varying concentrations of

aTc; no aTc, 100 ng/ml, 310 ng/ml, and 1000 ng/ml, at the start of a growth curve and their growth measured using a photo spectrophotometer at OD_{600nm}.

The results for *C. difficile* showed that after the initial addition of the varying concentrations of aTc to the starter culture, the OD dropped from the initial starting OD_{600} of 0.05 during the first 2 hours, compared to the culture with no aTc added which showed minimal change in OD (figure 4.22). At the 4 hour reading, concentrations 100 and 310 ng/ml aTc showed the same growth as the *C. difficile* grown in no aTc, however, the 1000 ng/ml culture showed a noticeable decrease in OD. This trend continued throughout growth until the 24 hour reading where all cultures were of comparable density. These results suggest that the growth of *C. difficile* is not affected by concentrations of up to 310 ng/ml aTc, but is affected by 1000 ng/ml aTc. This effect slows down the growth during the 8 hours which were monitored, but the same final OD can be reached after 24 hours.

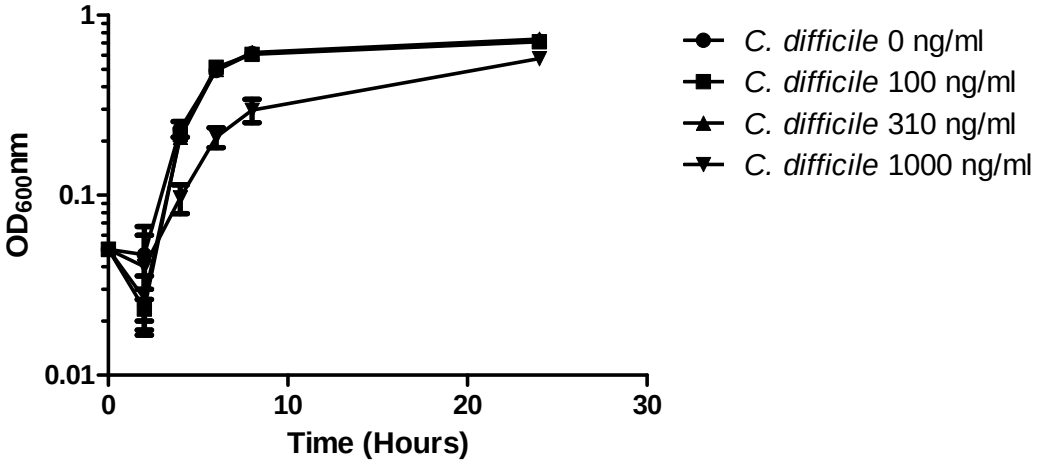


Figure 4.22. Growth curve studying the ability of *C. difficile* 630Δ*erm* to grow (OD_{600}) in the presence of varying concentrations of aTc over a 24 hour period. An overnight culture of *C. difficile* was used to inoculate a starter culture of 10 ml LB to an OD_{600} 0.05. The appropriate aTc

concentration was added to the starter culture (0, 100, 310, or 1000 ng/ml) and the OD_{600} was measured at specific time points (2, 4, 6, 8, and 24 hours) after inoculation. The error bars indicate standard error of the means from three experiments.

The same experiment was repeated for *C. sporogenes* and *C. beijerinckii* harbouring pMTL-82254-no$P_{thl/tetO}$. The results for *C. sporogenes* showed that after 2 hours, all four concentrations showed growth reflective of the concentration of aTc, meaning the fastest growing culture had no aTc present, the 2nd was supplemented with 100 ng/ml aTc, 3rd was *C. sporogenes*, and the slowest was 1000 ng/ml aTc (figure 4.23). After 6hours, the same order of growth speed was noted, the culture with no aTc reached stationary phase with an OD of 1.1, whereas 100 ng/ml and 310 ng/ml aTc were still growing with an OD of 1 and 0.35 respectively, however the culture supplemented with 1000 ng/ml had decreased in OD to 0.03, lower than the initial starting culture. By 10 hours, the cultures containing 100 ng/ml and 310 ng/ml had reached stationary phase with OD of 1.1 and 1.03 equal to that of no aTc, in contrast, the culture supplemented with 1000 ng/ml decreased in OD to 0 suggesting that the cells had succumbed to the aTc overtime. As for *C. beijerinckii*, the results were considerably different (figure 4.24). After the first 2hours, there was no notable difference with growth between each culture, with all showing OD's between 0.12 and 0.18. After 4 hours, the difference in growth rates between the cultures became obvious, the culture without aTc increased in OD to 0.35 whereas all the other decreased to 0.07, 0.1, and 0.06 for 100, 310, and 1000 ng/ml aTc, respectively. After 6 hours growth, unexpectedly the OD600's for both 310 and 1000 ng/ml decreased to 0 compared to the OD of the cultures without aTc increasing up to 0.78, this continued with the 8 hour reading with the cultures without aTc reaching an OD of 1.12 and 100 ng/ml aTc increasing slightly to 0.2, however the

other two aTc concentrations remained at an OD of 0. The final reading taken after 24 hours showed that all cultures had grown to what is assumed as stationary phase with an OD of 2, 2, 1.28 and 0.75 for 0, 100, 310, and 1000 ng/ml aTc.

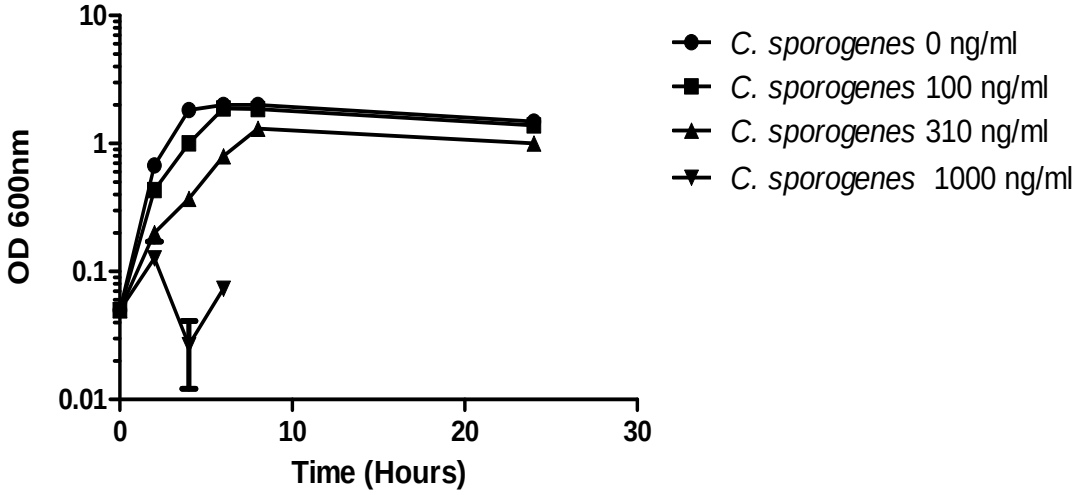


Figure 4.23. Growth curve studying the ability of *C. sporogenes* to grow (OD_{600}) in the presence of varying concentrations of aTc over a 24 hour period. An overnight culture of *C. sporogenes* was used to inoculate a starter culture of 10 ml LB to an OD_{600} 0.05. The appropriate aTc concentration was added to the starter culture (0, 100, 310, or 1000 ng/ml) and the OD_{600} was measured at specific time points (2, 4, 6, 8, and 24 hours) after inoculation. The error bars indicate standard error of the means from three experiments.

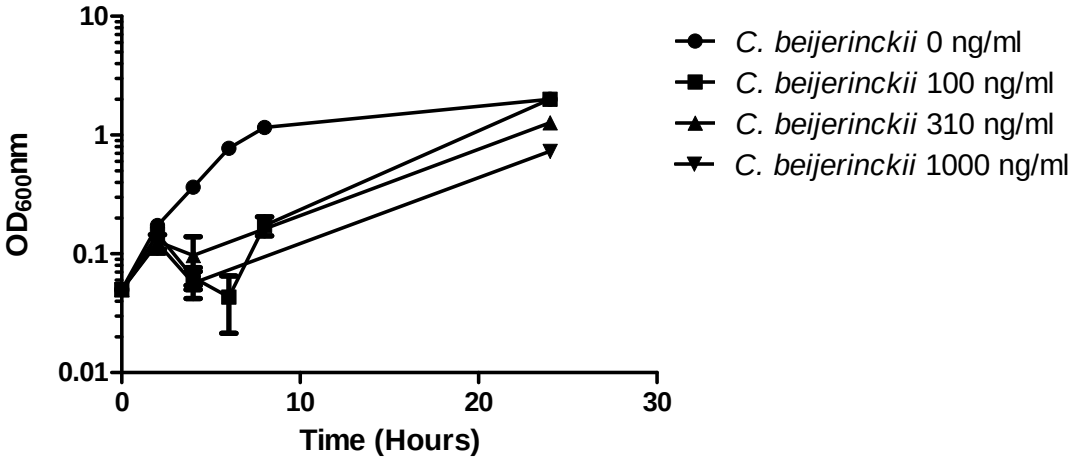


Figure 4.24. Growth curve studying the ability of *C. beijerinckii* to grow (OD_{600}) in the presence of varying concentrations of aTc over a 24 hour period. An overnight culture of *C. beijerinckii* was used to inoculate a starter culture of 10 ml LB to an OD_{600} 0.05. The appropriate aTc concentration was added to the starter culture (0, 100, 310, or 1000 ng/ml) and the OD_{600} was measured at specific time points (2, 4, 6, 8, and 24 hours) after inoculation. The error bars indicate standard error of the means from three experiments.

4.2.3.2 Testing the Clostridium *Tet inducible expression system in* C. difficile *630Δ*erm

In the first instance, it was necessary to ensure that the introduction of the *tetO* sequences into the P_{fdx} did not interfere with its ability to initiate RNA polymerase II-dependant transcription of the reporter, *catP*. To do this, plasmids pMTL-*tetO* and pMTL-*tetO1/2* were transformed into the conjugation donor CA434, and from there conjugated into *C. difficile* using the standard conjugation technique[90]. If the inducible promoters are active, they should confer resistance to thiamphenicol due to $P_{fdx/tetO}$ and $P_{fdx/tetO/1}$ expressing *catP*. Therefore, as a quick test to ensure they are still working in *C. difficile*, a transconjugant of each plasmid was streaked onto BHIS supplemented with

thiamphenicol. The results showed that only plasmid pMTL-tet3nO containing the single operator configuration conferred resistance to thiamphenicol, with no colonies obtained for *C. difficile* harbouring pMTL-*tetO1/2*.

To quantify expression of *catP*, which is directly proportional to the strength of each promoter, the CAT assay was implemented. The results showed that the single operator configuration was able to produce CatP to a level of 7,520 CAT U/ml enzyme, compared to very low detected expression with the double *tetO* configuration (Figure 4.25).

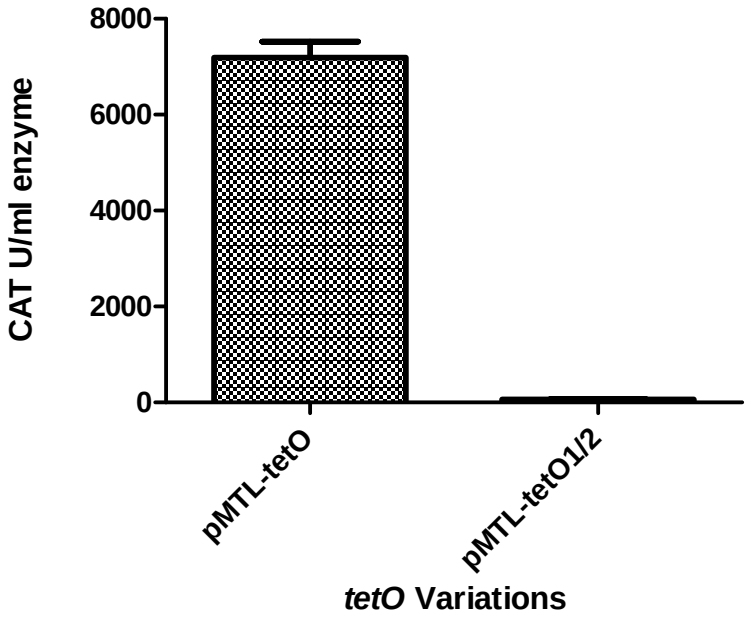


Figure 4.25. Comparison of the two *tetO* operator configurations, present within P_{fdx} outlined in figure 4.18(B), to access their ability to produce CatP in *C. difficile* 630Δ*erm*. Each *tetO* construct was transcriptionally fused to *catP* in the tetracycline inducible promoter constructs lacking TetR production in plasmid pMTL82254. CatP production was quantified (CAT U/mg enzyme) after 4 hours of growth

past OD_{600} 0.5. The error bars indicate standard error of the means from three experiments.

As it is now known the single operator configuration of pMTL-*tetO* can express the reporter in the absence of TetR, next it was desirable to know whether the inducible promoter can be repressed, and induced with the addition of aTc. To do this, the plasmid pMTL-tet3nO was conjugated into *C. difficile* 630Δ*erm* and quantification of *catP* expression was repeated with a 4 hour induction time. The results showed that the single operator was able to repress the production of CatP to around 39 CAT U/mg protein when *tetR* was expressed from the P_{thl} promoter. When 316 ng/ml aTc was added, the repression was alleviated causing an induction of CatP production of up to 7,523 CAT U/mg protein (figure. 4.26). This difference between expression and repression equates to an induction factor (IF) of 192.89. It was also noted that even with induction of 316 ng/ml, the construct could only reach around 76% the CatP production levels previously seen when no repressor molecule was being expressed. These results show that the construct pMTL-tet3nO is able to act as an inducible promoter system in *C. difficile*. To further analyse the Tet inducible promoter system to understand its characteristics, two more experiments were carried out to determine its ability to produce a dynamic range of CatP, and speed of induction to enable short-term studies on phenotype.

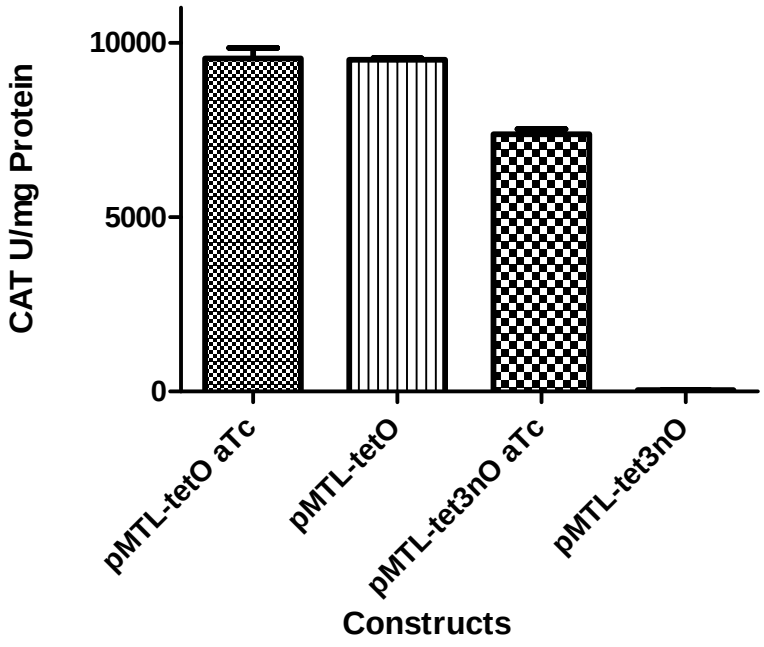


Figure 4.26. Comparison of the two constructs; pMTL-*tetO* and pMTL-tet3nO, in their ability to produce CatP in the presence or absence of 316 ng/ml aTc in *C. difficile* 630Δ*erm*. The two plasmids, differing in the presence of the P_{thl} transcribing *tetR* (pMTL-tet3nO) or absence of P_{thl} transcribing *tetR* (no TetR production) (pMTL-tetO), are transcriptionally fused to *catP* on plasmid pMTL82254 and were grown to an OD_{600} 0.5 and subsequently induced for 4 hours with 316 ng/ml aTc, after which CatP production was quantified (CAT U/mg protein). The error bars indicate standard error of the means from three experiments.

4.2.3.2.1 Dynamic induction range and time course of inducible CatP production in C. difficile *630Δ*erm

The ability to control gene expression through a range of expression levels is an important characteristic and one of the most important desirable characteristics. The

original tetracycline system is known to be very sensitive to concentrations of tetracycline, and as such intermediate induction levels are difficult to achieve[126]. To combat this, the system was re-designed to remove the P_{tetR} promoter which contains *tetO1*, and replaced it with P_{thl}. To establish if this has affected its induction kinetics, the pMTL-tet3nO vector was subjected to varying concentrations of aTc.

The results showed that CatP production is surprisingly dose-dependent in response to aTc induction (Figure 4.27). Without aTc, CatP production was very tightly repressed to levels indistinguishable from background, with the addition of concentrations as low as 0.31 ng/ml aTc induction of CatP can be detected to a level of 345 CAT U/mg protein. Increasing the concentration of aTc in 2-log doses increases the production levels of CatP in a dose-dependent fashion. Around half maximum CatP production was seen with the addition of 10 ng/ml aTc, producing 4,100 CAT U/mg protein, compared to full induction with 316 ng/ml aTc at 7,000 CAT U/mg protein. Increasing the concentration of aTc from 316 ng/ml to 1 ng/ml does not increase the production of CatP, as the system seems to be saturated. These results show that the system seems to be more dose-dependent than the original Tet system[126], with intermediate ranges of expression obtainable through a range of aTc concentrations between 0.31 and 316 ng/ml.

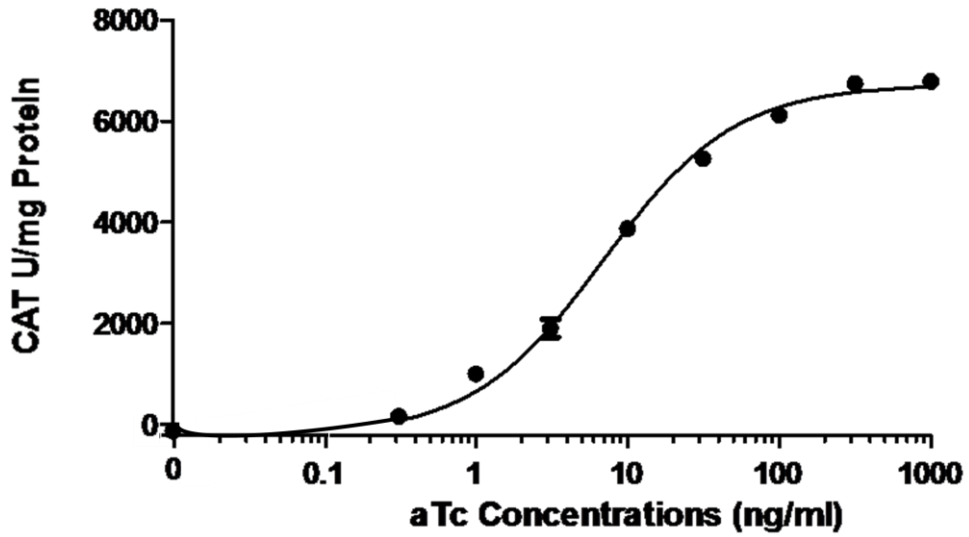


Figure 4.27. Comparison of CatP production at varying 2-log concentrations of aTc in *C. difficile* 630Δ*erm* harbouring plasmid pMTL-tet3nO. *C. difficile* 630Δ*erm* harbouring plasmid pMTL-tet3nO was grown to an OD_{600} 0.5 and induced with varying concentrations of aTc (0.31, 1, 3.1, 10, 31.6, 100, 316.4 or 1000 ng/ml) for 4 hours and subsequently analysed for CatP production (CAT U/mg protein). The error bars indicate standard error of the means from three experiments.

To test if the system induces quickly, has a stable expression level, and the inducers bioavailability does not decrease over 6 hours, *C. difficile* harbouring plasmid pMTL-tet3nO was grown to an OD_{600} 0.5 and induced with 316 ng/ml aTc and sampled every hour after induction for 6 hours. The results showed that after 1 hour, pMTL-tet3nO was able to produce CatP quickly and strongly, to a level of 3,695 CAT U/mgprotein, compared to very tight repression of CatP production in the absence of aTc (figure 4.28). The production of CatP increased over 4 hours, until expression levels reached maximum levels of 6,794 CAT U/mg protein. After 4 hours, the production of CatP stabilised and stayed within the same vicinity of production levels as the 4 hour reading.

These results pMTL-tet3nO is able to induce quickly, with a maximum level of production reached by 4 hours, and no decrease in production over time associated with aTc degradation.

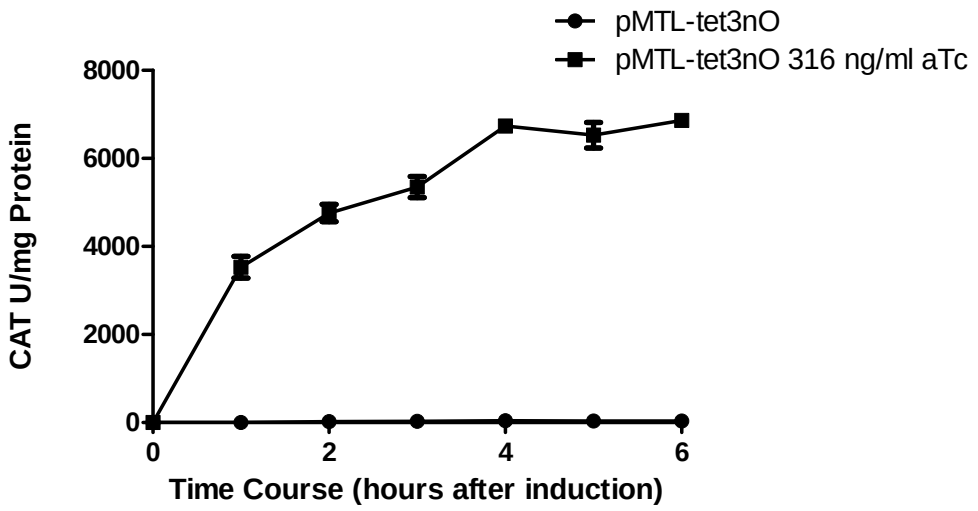


Figure 4.28. Time course of CatP production after induction of *C. difficile* 630Δ*erm* harbouring pMTL-tet3nO at OD_{600} 0.5. *C. difficile* 630Δ*erm* containing plasmid pMTL-tet3nO was grown to an OD_{600} 0.5 and either induced with 316 ng/ml aTc, or not induced. Samples were taken every hour for 6 hours after induction for each culture and CatP production was quantified (CAT U/mg protein). The error bars indicate standard error of the means from three experiments.

4.2.3.3 Testing the Clostridium Tet *inducible expression system in* C. sporogenes *NCIMB 10969*

As the Tet system was shown to work exceedingly well in *C. difficile* 630Δ*erm*, it was valuable to ascertain if the system functions in other *Clostridium* species due to their importance as pathogens or in industrial applications. *C. sporogenes* is an organism of

significant interest due to its similarity to the human pathogen *C. botulinum*, and its application in CDEPT research.

In the first instance, we carried out the dose-dependent experiment to identify the dynamic range of *catP* expression in response to varying concentrations of aTc, which was also performed in *C. difficile*. The results showed that there was a good intermediate range of CatP production achieved (figure 4.29). These results were comparable to the same construct in *C. difficile*, although the system was not as sensitive to low levels of aTc and the strength of expression at its peak was over 4 times stronger than *C. difficile*. This increase in strength is comparable with previously seen results due to the strength of the *C. sporogenes* ferredoxin promoter being around 4 times higher in *C. sporogenes* compared to *C. difficile* (O. Pennington, unpublished). Without the addition of aTc, tight repression of CatP was seen with only 86 CAT U/mg protein being detected. The detected units of CAT/mg protein consistently remained below 1,000 until 10 ng/ml of aTc was added, which increased production from 786 CAT U/mg protein obtained with 3.1 ng/ml aTc, to 5,681 CAT U/mg protein. As a higher concentration of aTc was added, a higher level of induction was detected. At 31.6 ng/ml aTc, 17,751 CAT U/mg proteinwas detected, increasing to 26,564 and finally to the highest production of 37,180 CAT U/mg protein for induction with 100 ng/ml and 316.4 ng/ml, respectively. It was also noted that after induction with 1 μg/ml of aTc, the levels of production of CatP decreased from 37,180 CAT U/mg protein for 316.4 ng/ml, to 16,121 CAT U/mg protein, and that the OD_{600} of the culture at time of sampling after induction with 1 μg/ml aTc was lower than all other aTc concentrations, although this difference was accounted for in the calculation of CAT U/mg. This decrease in CatP production and growth seems to correspond to the sensitivity of *C. sporogenes* to 1000 ng/ml aTc (figure 4.21).

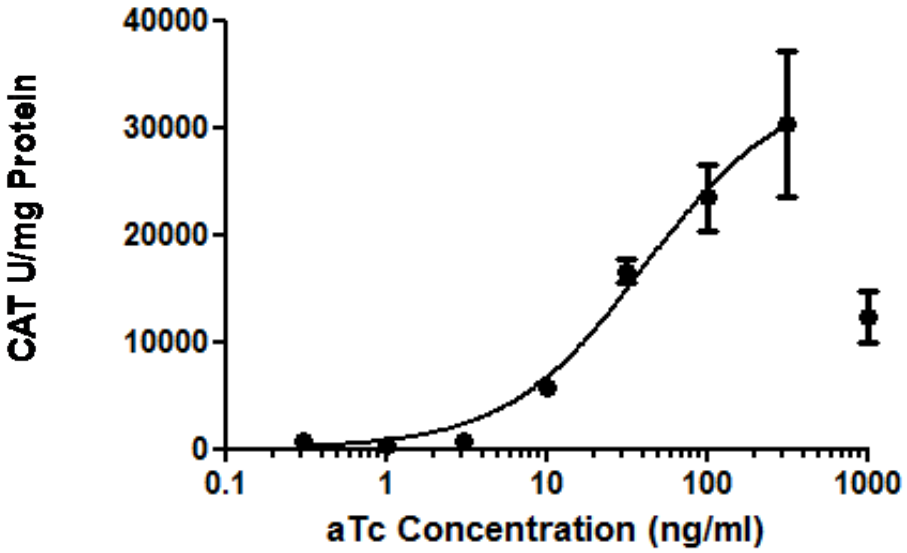


Figure 4.29. Comparison of CatP production after induction at OD_{600} 0.5 with varying 2-log concentrations of aTc in *C. sporogenes NCIMB 10969* harbouring plasmid pMTL-tet3nO. *C. sporogenes* harbouring plasmid pMTL-tet3nO was grown to an OD_{600} 0.5 and induced with varying concentrations of aTc (0.31, 1, 3.1, 10, 31.6, 100, 316.4 or 1000 ng/ml) for 4 hours and subsequently analysed for CatP production (CAT U/mg protein). The error bars indicate standard error of the means from three experiments.

To test if the system induces CatP production quickly, has a stable expression level, and the inducers bioavailability does not change over 6 hours, the plasmid pMTL-tet3nO was tested in the same way as it was in *C. difficile*, with samples being taken every 1 hour after induction with 316.4 ng/ml aTc. The results were very similar to that of *C. difficile*, in that the system was able to produce CatP quickly and strongly (figure 4.30). Upon induction with aTc, the production of CatP increased from 106 CAT U/mg protein, to a very respectable 16,012 CAT U/mg protein within 1 hour. After 2 hours, production of CatP had increased to 24,317 CAT U/mg protein, with fewer than 50 CAT U/mg

protein produced in the un-induced samples. This trend increased until maximum production was obtained at 6 hour after induction, with 31,586 CAT U/mg protein. The un-induced/repressed samples remained below 160 CAT U/mg protein throughout the time course study, with the maximum induction factor noted at 1,275.63 for the 3 hour samples. This induction factor, if correct, represents one of the highest noted induction factors obtained for an inducible promoter system. This high induction factor is partly due to the strength of the P_{fdx} promoter in *C. sporogenes* compared to *C. difficile*. It would seem that the promoter is on average around 3-4 times stronger in *C. sporogenes*, however, is able to repress to the same levels as *C. difficile*. This functionality makes this Tet system a very powerful tool for controlling gene expression in *C. sporogenes*, as well as *C. difficile*.

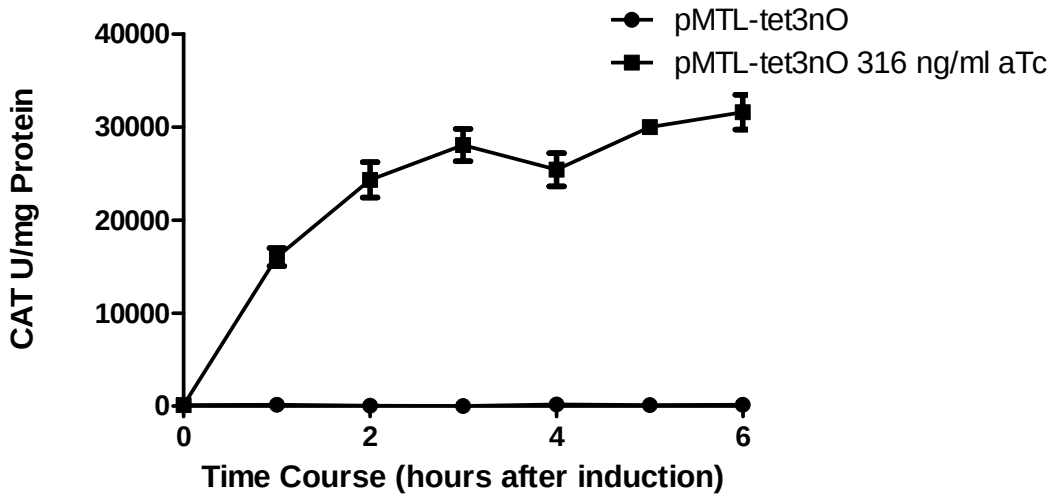


Figure 4.30. Comparison of CatP production at varying time points after induction of *C. sporogenes NCIMB 10969* harbouring plasmid pMTL-tet3nO. *C. sporogenes* harbouring plasmid pMTL-tet3nO was grown to an OD_{600} 0.5 and either induced with 316 ng/ml aTc, or not induced. Samples were taken every hour for 6 hours after induction for each

culture and CatP production was quantified (CAT U/mg protein). The error bars indicate standard error of the means from three experiments.

4.2.3.4 Testing the Clostridium *Tet inducible expression system in* C. beijerinckii *NCBI 8052 and* C. acetobutylicum *ATCC 824*

The last two organisms chosen to test the system in were *C. beijerinckii* NCBI 8052 and *C. acetobutylicum* ATCC 824. These two clostridia were chosen due to their usage in the bio-energy field. As a quick test, the plasmids pMTL-*tetO* and pMTL-tet3nO were conjugated into *C. beijerinckii* and transformed into *C. acetobutylicum*, and the performance of the Tet system was analysed.

The results showed that in both organisms, the absence of *tet*R expression allows the P_{fdx} to promote high levels of *catP* transcription to between 4,500 and 5,000 CAT U/mg protein (figure 4.31). In the presence of a promoter transcribing the repressor, *tet*R, *catP* is tightly repressed to levels below 160 CAT U/mg protein of CatP production, however, the repression cannot be alleviated to a reasonable level, which was possible in *C. difficile* and *C. sporogenes*, with the Tet system showing induction factors of around 2 and 4 for both *C. beijerinckii* and *C. acetobutylicum*, respectively (figure 4.31). It is unknown why the Tet system is unable to be induced in these organisms. It would seem the inducible promoter is active due to CatP production in the absence of TetR; however, the addition of the inducer seems to have little effect on the alleviation of the repression..

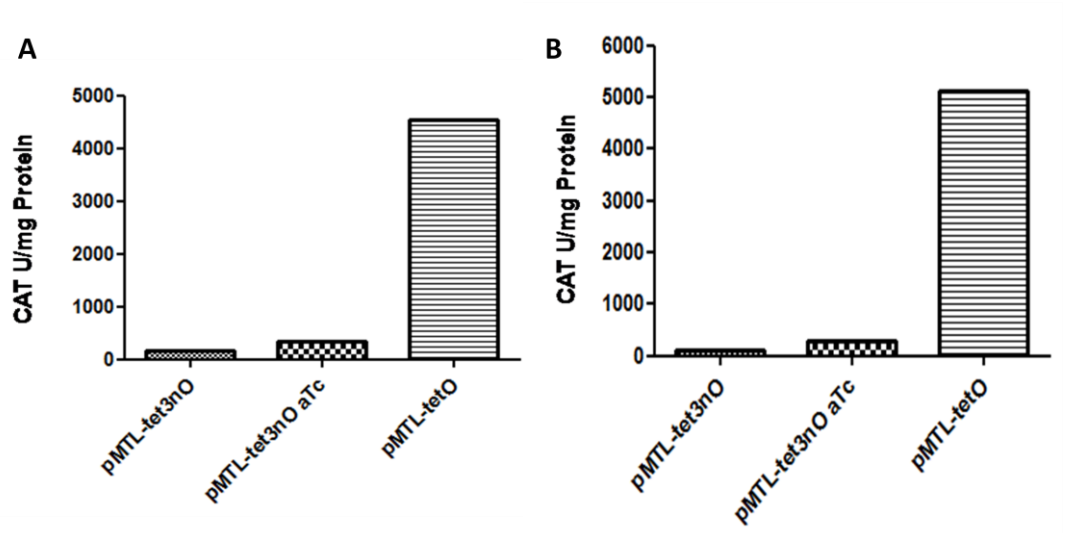


Figure 4.31. Comparison of CatP production at 6 hours after induction with (**A**) C. *acetobutylicum* and (**B**) C. *beijerinckii*, harbouring plasmids pMTL-tet3nO and pMTL-*tetO*. Cultures of C. *acetobutylicum* and C. *beijerinckii* containing plasmid pMTL-tet3nO were grown to an OD_{600} 0.5 and induced with 316 ng/ml aTc for 6 hours, after which CatP production was quantified (CAT U/mg protein). In parallel, the plasmids pMTL-tet3nO and pMTL-tetO were grown 6 hours past the same OD_{600} and CatP production quantified (CAT U/mg protein). The error bars indicate standard error of the means from three experiments.

4.3 Discussion

One of the major gaps in the repertoire of genetic tools available for *Clostridium* was the absence of an easily usable inducible promoter system which can control gene expression through a wide dynamic range. There have been a number of systems developed for *Clostridium*, however, each of them has their advantages and disadvantageous which would limit the systems versatility. The desirable characteristics that were sought in this study were; the ability of the system to work in multiple clostridia, wide dynamic range of activity, high bioavailability of the inducer, high target gene expression, quick induction, and tight repression.

To achieve these characteristics, a *lac* repressor system was first developed which used LacI as a repressor. The original system had been shown to work in many organisms and is widely thought of as a paradigm for inducible promoters. Unfortunately the original system does not seem to work efficiently in *C. difficile*[38] (Personal communication with Prof. Minton). To overcome this, the system was re-designed using clostridial components. Upon testing the system it was noticed that the stronger the promoter transcribing *lacI*, the higher the repression of our reporter CatP, however, the repression could not be alleviated with the addition of the inducer. This inability to induce the production of CatP would suggest that the inducer is not acting upon the LacI repressor. Although the exact mechanism is unknown, a possible reason for this inability to induce the inducible system could be very limited diffusion or import rate of IPTG into the *C. difficile* cells, resulting in a significantly lower intracellular IPTG concentration compared to extracellular levels. It is not, to our knowledge, established whether *C. difficile* can actively transport lactose, and therefore IPTG, into the cells. There does not seem to be a LacY homologue present within the genome, in addition to this, *C. difficile* has a Gram-positive cell wall which may limit or inhibit IPTG to enter

the cell via diffusion. This problem has been encountered in other bacteria such as, *Synechocystis*[152], *E. coli*[153] and *Pseudomonas fluorescens*[153] and subsequently solved by the introduction of the *E. coli* lactose permease gene, *lacY*[153]. Other ways to circumvent this problem could be to develop *lac* repressor mutants which have a high sensitivity to IPTG; this approach has been used previously by employing error-prone PCR and selecting for IPTG sensitive mutants[154]. A combination of these two approaches could provide *C. difficile* with a very efficient system able to actively import IPTG, and able to sense very small concentrations, even as low as 1 μM of IPTG[154].

In parallel to the IPTG inducible promoter system, the utility of riboswitches, able to control translation through varying mRNA secondary folding conformations in response to specific ligands, was tested in their ability to control gene expression in *C. difficile*. In the first instance, clone 8.1 developed by Lynch et al, (2007)[1] was selected, to which was added an *Nde*I site at the start codon for ease of cloning, yielding Thio1. It was found that Thio1 was able to repress well in *E. coli* and respond with an increase in expression when the inducer theophylline was added; however, when Thio1 was tested in *C. difficile*, no CatP production was detected. The same absence of production was noted when the riboswitch was removed and replaced with just the RBS. This suggested that the RBS employed was not active in *C. difficile*, in contrast to *E. coli*. In response to this, two novel riboswitches were developed by rational design to include two RBSs likely to function in *C. difficile*. Out of these two riboswitches, Thio3 was not found to be able to repress CatP production in the absence of theophylline, possibly due to an error in the construction inhibiting its ability to form its conformations effectively. The second riboswitch Thio2 was able to repress and express *catP* in response to theophylline to a higher induction factor than the original Thio1. The subsequent testing of the riboswitches Thio2 and Thio3 in *C. difficile* produced disappointing results.

Both the riboswitches were unable to express *catP* in the absence or presence of the inducer, theophylline. However, compared to Thio1, the RBSs alone were able to express *catP* showing the RBSs are in fact active in *C. difficile*. The reason for Thio1 and Thio2's inability to work in *C. difficile* is not clear. It could be due to the decrease in the efficiency of translation in the presence of the riboswitches. In *E. coli* it was noted that the presence of the riboswitch in Thio3 decreased translation around 10-30 times compared to the RBS alone. In using these plasmids in *C. difficile*, the lower copy number means lower overall production of CatP; although the RBSs alone can lead to detectable production of CatP, the presence of the riboswitch might lower the translation efficiency to a level below the detectable limit of the older model photo spectrophotometer used for the assay.

In the last instance, the development of a tetracycline inducible promoter system was investigated as it is known that tetracycline is able to enter the cell (sensitive to the antibiotic), thereby avoiding the hypothesised problems seen with the IPTG system. The system pMTL-tet3nO was very successful in repressing *catP* expression in the absence of aTc, to a level that was almost undetectable above background in *C. difficile* and *C. sporogenes*. Possibly due to the exclusion of the *tetO1* present in the P_{tetR}, the dose-dependent response to varying concentrations of aTc was remarkable, with a dynamic range of intermediate production levels able to be achieved. This dynamic range is a big advantage of the system over the original tetracycline system which is known to have a limited intermediate range[126] (figure 4.16). The Tet system also filled the other required characteristics of rapid induction with high levels of production seen within 1 hour, high bioavailability, and very tight repression in the absence of the inducer. The surprising result was the ~76% decrease in production of CatP in the presence of TetR. This decrease is more than likely due to the high constitutive expression of TetR and

subsequent increase in binding to the *tetO* which cannot be overcome with increased concentrations of aTc added. Although this decrease in expression is not favourable, the levels of CatP production is still very high and an acceptable trade-off for increased repression, particularly in the case of using the inducible promoter system to conditionally control essential genes for identification purposes.

As a universal inducible promoter system was being sought, the construct was tested in *C. acetobutylicum* and *C. beijerinckii*. Surprisingly the system didn't work as expected in either. The control plasmid pMTL-*tetO* produced high levels of CatP, and in the absence of the inducer aTc, the full tet3nO construct repressed to very high levels. However, it could not be induced to levels about 3 times above the repressed state. The reason for this is unknown and has similarities with the problems encountered with IPTG in the *lac* repressor system. However, aTc is known to be able to enter the cell due to the sensitivity of exogenously added aTc. On closer inspection, a BLAST search using the translated nucleotide sequence of *tetA*(B) (P02980) from *E. coli* shows that both organisms, as well as *Clostridium ljungdahlii* and *Clostridium perfringens*, have what appears to be a natural tetracycline inducible promoter system including a *tetA* metal-tetracycline/H(+) antiporter. This possible natural tetracycline inducible promoter including *tetA* could be responsible for actively transporting aTc out of the cell, resulting in decreased availability of aTc within the cell. If this is the case, aTc could be removed before it is able to bind to TetR, thereby inhibiting transcription as if no/very low concentrations of aTc are present intracellularly. To counter this problem, it may be possible to inactive the hypothesised *tetA* in both organisms and test whether it has an effect on induction of CatP. Unfortunately due to time constraints, it was not possible to test if this would work, but would make an excellent experiment to undertake in the future. In conclusion, it is intended to implement this tetracycline inducible promoter

system to develop a method to identify essential genes using a conditional lethal genotype.

Chapter Five

Use of a Tet Inducible Promoter System to Conditionally Control Lethal Phenotypes for the Identification of Essential Genes

5.1 Introduction

The identification of essential genes in *C. difficile* has been prevented by the limitation of current techniques[38,41]. Molecular tools such as the ClosTron are capable of insertionally inactivating genes, however essential genes are unable to be inactivated, making this system alone insufficient to identify essential genes[38,41]. In chapter three, a system was developed which employs the use of an ectopically located partial diploid strain to functionally complement a native essential gene, thereby negating the essentiality of the native gene and allowing it to be insertionally inactivated using the ClosTron. For the first time three essential genes were identified in *C. difficile*, however the identification was through the indirect evidence of only being able to inactivate the gene in a partial diploid background. For a direct method to identify essential genes, one has to use conditional mutational systems such as an inducible promoter[46,47,72,116].

5.1.2 Conditional mutations

Conditional expression is achieved by placing a gene of interest under the control of an inducible promoter[69,113,114,118,133]. There are two main ways in which to control an essential genes transcription using an inducible promoter. The first uses either a single-crossover integration or double allele-exchange process to replace the essential genes native promoter with an inducible promoter. This has been used successfully in *S. aureus*[49,140,155] and *B. subtilis*[45] to identify essential genes. However, there is a significant limitation to using such a strategy. In the case of the essential gene being present within an operon transcribed from a polycistronic mRNA, the introduction of an inducible promoter into the operon could potentially co-regulate genes downstream, thereby changing multiple parameters and making interpretation of the results difficult. This is particularly apparent in complex operons housing multiple essential genes and secondary transcriptional control systems.

Another strategy to circumvent this limitation is to introduce an ectopically expressed protected copy of the essential target gene under inducible control into the chromosome. This means that the native essential gene can be inactivated either by ClosTron or allelic exchange mutagenesis in the presence of an inducer. Once the native gene is inactivated, the inducible second copy is now the essential copy, similar to the strategy undertaken in chapter three. However in this strategy; only in the presence of an inducer will the cell be viable. If the essential gene is switched off due to the absence of an inducer, then the cell will die. This would allow direct identification of essential genes by studying viability in the presence or absence of an inducer[45]. A study like this has not been undertaken within *C. difficile* due to the lack of an effective inducible promoter system. With the development of the Tet system in chapter four, it is now possible to undertake this type of study.

5.1. Aim of this study

The use of inducible promoter systems to create conditional lethal phenotypes will allow the direct identification of essential genes within *C. difficile*. The advantage of using such a system is the ability to further analysis the mutation through a dynamic range of expression levels, as well as throughout the life cycle of the organism. Gathering far more information and evidence as to the function and nature of the essential gene compared to the method developed in chapter three.

The aim of this chapter is to develop a genetic tool able to modulate transcription of ectopically located genes for the identification of essentiality and analysis of potential as therapeutic drug targets.

5.2 Results

5.2.1 Construction of conditional lethal recombinant strains

Conditional mutations and control of gene expression have been frequently used in other bacterial species for identification of essential genes, unfortunately systems such as; amber stop codons[66,156,157], temperature sensitive mutations[72,101], riboswitches[1,122,143], and most inducible promoters have not been developed to meet our requirements or published for use in *C. difficile*[38,120]. In chapter four, a tetracycline inducible promoter system was developed which meets the requirements for conditional mutations.

Three genes were previously identified as essential in *C. difficile*; CD0274, *trpS*, and *metK*. In order to further study the survival kinetics and nature of essentiality, CD0274 was chosen for the conditional lethal experiment due to the probability of the gene being essential to few bacterial species, making it a possible antimicrobial target[100,158,159]. The previously identified essential gene *trpS* was also chosen to gather further evidence that the system functions as would be expected. Finally, a third new target was chosen, *slpA*. In previous work unpublished in our lab, *slpA* was not able to be insertionally inactivated using the ClosTron system, the first indication that a gene could be essential. In *C. difficile*, *slpA* is known to encode for a surface-exposed proteinaceous layer termed the S-layer[31,32,120]. Previous studies have shown that the S-layer forms a regular two-dimensional array which is visible by electron microscope[32]. This S-layer is constructed from a higher-molecular-mass protein (P46) of between 48 to 56 kDa, and a lower-molecular-mass protein (P36) of between 36 to 45 kDa, encoded by the C-terminal and variable N-terminal of *slpA*, respectively[31,32]. It is suggested that P36, the lower-molecular-mass protein is expressed on the exterior of the S-layer, of which it is shown to have adhesive properties to Caco-2 cell lines[31,32]. Due to its exclusive usage in *C. difficile* and its location on the exterior of the cell, it would make an attractive target for

development of a vaccine or a novel narrow-spectrum antimicrobial agent against *C. difficile*.

As was the case with the three targets in chapter three, if *slpA* is to be under inducible expressed at an ectopic site, meaning the native copy is to be insertionally inactivated, the nucleotide sequence must be changed to protect the second copy from inactivation by the ClosTron system, as were the previous candidates in chapter three. To do this, the nucleotide sequence around three ClosTron target sequences were altered, between nucleotides; 861 and 909, 1043 and 1089, and finally 1243 and 1284, to inhibit the recognition and subsequent integration of the group II intron targets pMTL007C-E2:Cdi-*slpA*-878a, pMTL007C-E2:Cdi-*slpA*-1072s, and pMTL007C-E2:Cdi-*slpA*-1267s, respectively (figure 5.1). These changes are silent, meaning that the amino acid sequence is unchanged, and the subsequent gene product is unaltered from the native. The synthetic protected *slpA*p was designed using the gene designer software developed by DNA2.0 and synthesised by the same company. During the construction of the construct by DNA2.0, the restriction sites *Nde*I and *Nhe*I were introduced at the 5' and 3' of the gene to enable easy cloning into our Tet system and modular vector plasmids[89].

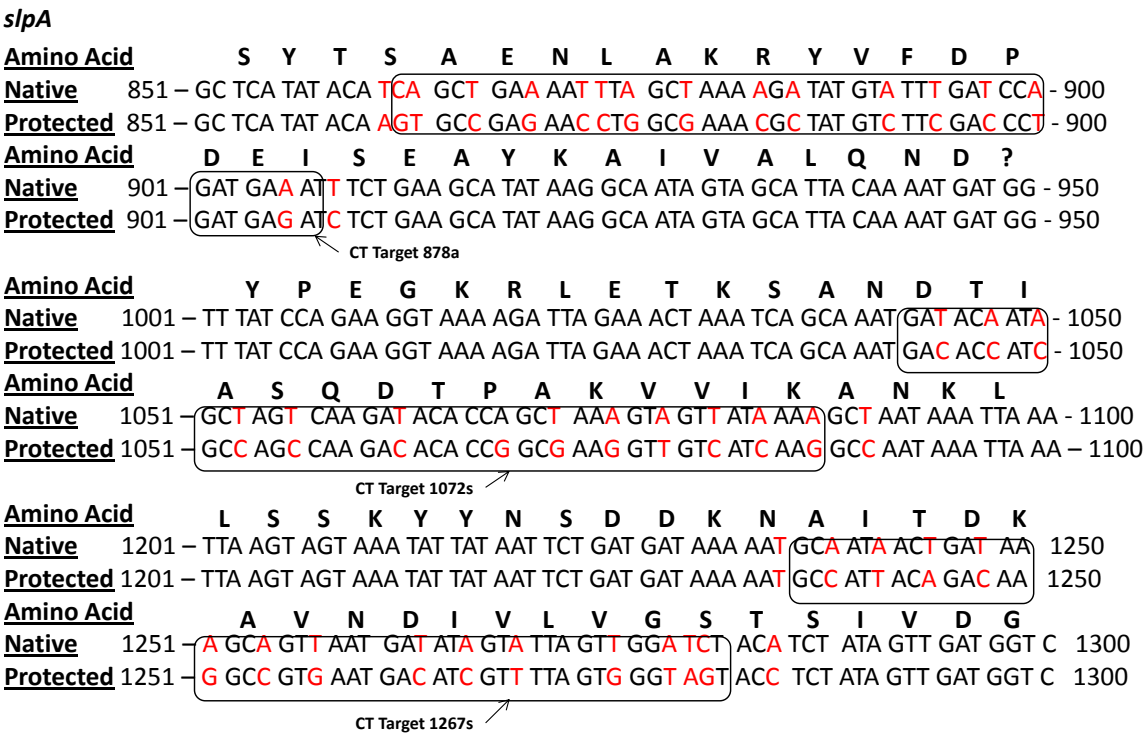


Figure 5.1. Protection of the synthetic partial diploid copy of the candidate essential gene *slpA*. The black box outlines the 45 bp mer target sequence of the three retargeted group II introns. In bold red is the silent nucleotide changes made in the synthetic protected copy of each candidate essential gene compared to the native gene required to protect the gene from the group II intron's targeting mechanism, but keeping the same translated amino acid sequence.

In order to insert the essential gene under inducible control into the chromosome, the ACE system was used (unpublished, Heap *et al*). The essential genes were digested from plasmids; DNA2.0-*slpA*p, DNA2.0-P_{thl}-*trpS*p, and pMTL-DW1, cloned into intermediary pMTL-tet3nO, and finally into the ACE plasmid pMTL-JH18, yielding plasmids pMTL-ACEtet-*dhaT*p, pMTL-ACEtet-*slpA*p, and pMTL-ACEtet-*trpS*p, relating to the protected genes *slpA*, *metK*, and *trpS* present under tetracycline inducible control, respectively.

Construction of the inducible partial diploid strain was carried out using ACE as previously described in section 3.2.4. Each plasmid was conjugated into *C. difficile* 630Δ*erm* using the donor strain *E. coli* CA434 and standard conjugation techniques[90]. The resulting transconjugants were selected for on BHIS supplemented with thiamphenicol and *C. difficile* supplement. After 24 hours, transconjugants started to become large enough to pick, therefore two transconjugants for each plasmid were re-streaked onto fresh media. For each plasmid, three individual colonies with the larger phenotype representing individually derived single-crossover recombinants were purified and enriched by sub-cloning on fresh BHIS supplemented with thiamphenicol and *C. difficile* supplement (figure 2A).

The next step was selection of the second-crossover event occurring between the smaller right hand homology arms. Each of the 3 individually derived single-crossover recombinants were re-streaked onto *C. difficile* minimal media supplemented with 5-FOA and uracil. After 24 hours, 5-FOA resistant colonies were large enough to pick, and were selected to undergo patch plating on BHIS or BHIS supplemented with thiamphenicol to identify the cells that are thiamphenicol sensitive, therefore, have lost the knock-in plasmid stopping further recombination events. Each of the colonies was screened by PCR using primers that amplified across the insertion locus to confirm allelic exchange has taken place at the desire site. PCR screening confirmed that all colonies screened had undergone the desired allele exchange (figure 2B).

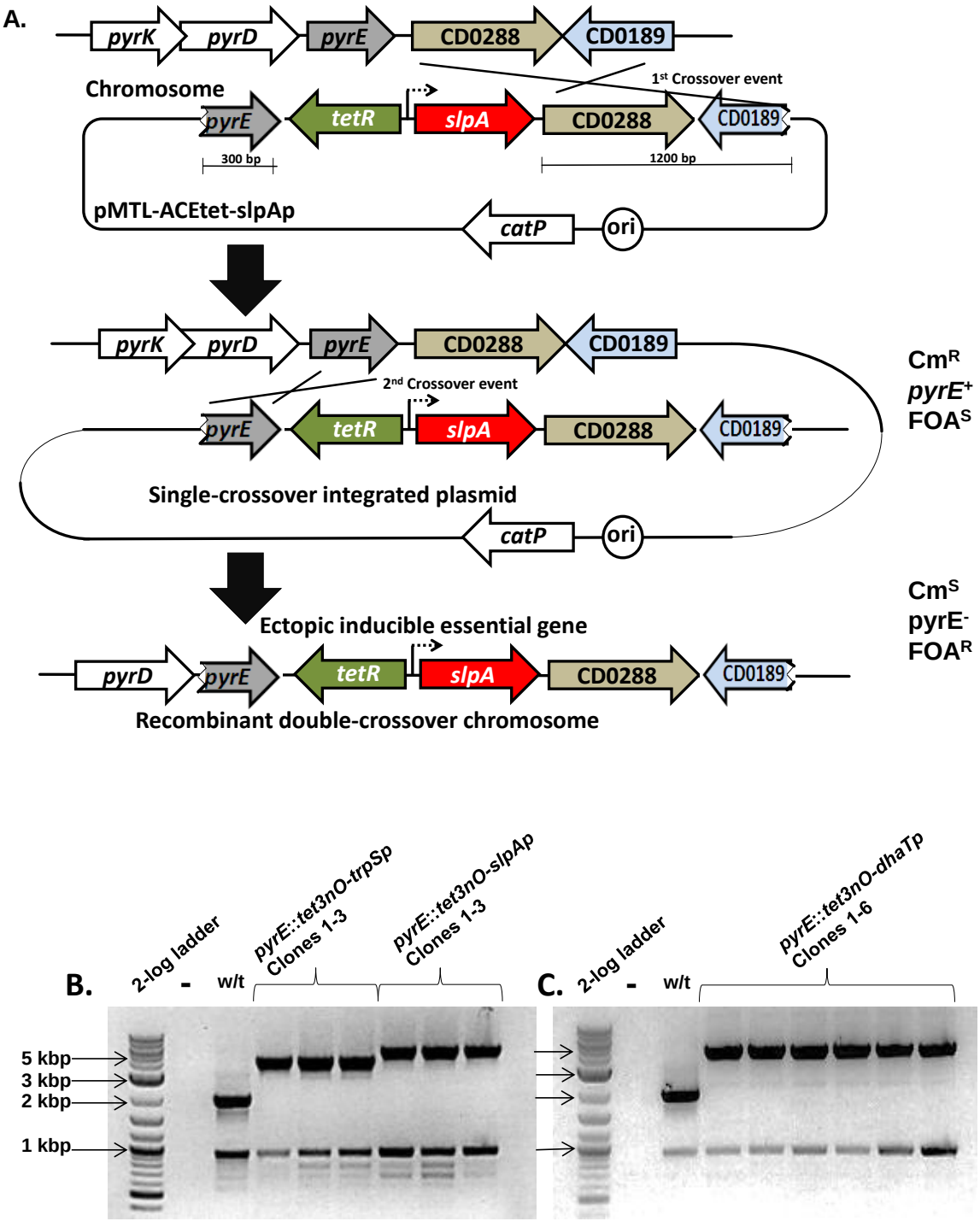


Figure 5.2. DNA integration of candidate essential gene *slpA* under inducible control in the *pyrE* locus of *C. difficile* 630Δ*erm* (**A**) Selection of the double-crossover clones using plasmid pMTL-ACEtet-*slpA*p. The first recombination event is directed by the long right-hand homology arm consisting of 1500 bp downstream of *pyrE* on the *C.*

*difficile*630Δ*erm*chromosome. The resulting single-crossover clones have their replication deficiency conferred by pIM13 alleviated allowing them to grow quicker than the freely-replicating plasmid containing cells. Subsequently, the second crossover event between the 300 bp left homology arm of pMTL-ACEtet-*slpA*p containing an internal fragment of *pyrE* completes the allelic-exchange. Two possible outcomes can occur, the parent strain or the double-crossover clone. To select only for the desired event, selection on 5-FOA allows only *pyrE*$^{-}$ strains (double-crossover clones) to grow. (**B**) PCR screening of 3 clones of *pyrE*::tet3nO-*trpS* and *pyrE*::tet3nO-*slpA* using primers Cdi630-pyrE-F2 and Cdi630-CD0189-R2 which anneal on the parent chromosome flanking the insertion site. MW is a 2-log DNA Ladder (NEB) molecular weight marker, lane 1 is water only negative control, lane 2 is parent (expected size ~2 Kbp) *C. difficile* 630Δ*erm* DNA control, lanes 3-5 are *pyrE*::tet3nO-*trpS* candidate clones (expected size ~ 4.2 Kbp), and lanes 6-8 are *pyrE*::tet3nO-*slpA* candidate clones (expected size ~ 5.4 kbp). (**C**) PCR screening of6 *pyrE*::tet3nO-*dhaT* candidate clones using the same primers and order as (B) (expected size ~4.6 kbp).

Having generated inducible partial diploid strains, to ensure that the ectopically located essential genes essentiality is under inducible control, the native copy must be inactivated. There are two ways in which to do this, the first uses an allele exchange method to in-frame delete the native gene (unpublished, Cartman *et al*). However due to time restrictions of this project, it was decided to use the alternative quicker method, the ClosTron system[38,41], to insertionally inactivate the native copy as was done in chapter three. As none of the three essential genes are present in the middle of an

operon, the insertion of the group II intron should limit polar effects on other genes; therefore the in-frame deletion method has no significant advantage over the ClosTron method in this instance.

To insertionally inactivate *slpA*, three target sites; 878a, 1072s, and 1267s were identified with their appropriate group II intron retargeting sequences. Two targets, 878a and 1267s were constructed by DNA2.0 and cloned into the ClosTron plasmid pMTL007C-E2, yielding ClosTron target plasmids; pMTL007C-E2:Cdi-*slpA*-878a and pMTL007C-E2:Cdi-*slpA*-1267s. Two targets were chosen in case one target was not able to insertionally inactivate *slpA*. In the first instance, plasmid pMTL007C-E2:Cdi-*slpA*-1267s was chosen. For the two other target genes, the ClosTron plasmids pMTL007C-E2:Cdi-*dhaT*-548a and pMTL007C-E2:Cdi-*trpS*-436a were chosen to insertionally inactivate targets CD0274 and *trpS*, respectively.

Each ClosTron plasmid was conjugated into its appropriate strain, i.e. pMTL007C-E2:Cdi-*slpA*-1267s was conjugated into strain *C. difficile* 630Δ*erm*::tet-*slpA*p which contains a second protected copy of *slpA* under tetracycline inducible control, using standard conjugation techniques[90]. After the conjugation time had elapsed, the conjugation slurry was plated onto BHIS supplemented with thiamphenicol, *C. difficile* supplement and most importantly, 316 μg/ml aTc. aTc is added due to the ClosTron being able to insertionally inactivate the native essential gene at this stage; therefore the protective second copy must be expressed allowing the insertion to not be lethal to the cell. After 24 hours, the transconjugants were visible and large enough to pick, so independent colonies were sub-cultured to ensure purity and re-streaked onto BHIS supplemented with erythromycin, *C. difficile* supplement, and 316 μg/ml aTc. Within 24 hours erythromycin resistance colonies were large enough to pick, suggesting that the group II intron has inserted into the chromosome thereby activating the RAM[38,41]. To

ensure that that the insertion was into the desired gene, PCR screening for the desired ClosTron insertion using primers which flank the target gene, appropriate to each strain was carried out. Unfortunately, all the amplicons were of the native genes size, suggesting the insertion into the target gene had not taken place.

The conjugations were repeated a total of 5 times, and a total of 5 erythromycin resistant colonies representative of possible ClosTron integrants were screened for each of the three targets mentioned above, including three more ClosTron plasmids; pMTL007C-E2:Cdi-*slpA*-878a, pMTL007C-E2:Cdi-*dhaT*-623a, and pMTL007C-E2:Cdi-*trpS*-460a, to increase the possibility of integration. Again no integrants into the desired gene were obtained. This result suggests that the target genes cannot be inactivated in these strains. It could be possible the inducible promoter system is not able to express the second protected target gene to a high enough level to functionally complement the native, thereby allowing the native gene to be inactivated.

In order to overcome this deficiency, the inducible promoter system including each of the target genes was cloned into plasmid pMTL82151, and conjugated into *C. difficile* 630Δ*erm*. Placing the inducible promoter system and target gene on a plasmid should multiply the expression of the gene by the copy number of the plasmid, thereby increasing the overall expression of the target genes. The ClosTron plasmids; pMTL007C-E2:Cdi-*slpA*-1267s, pMTL007C-E2:Cdi-*dhaT*-548a, and pMTL007C-E2:Cdi-*trpS*-436a were subsequently conjugated into the appropriate *C. difficile* transconjugants, and the ClosTron procedure was carried out with minor modifications[38,41]. As the second copy must be expressed before the native can be inactivated, aTc must be supplemented into the media when selecting for integrants to induce transcription. However, as was noted in chapter three, over-production of CD0274 produced instability in the cell, therefore a range of aTc; 10 ng/ml, 31 ng/ml,

and 310 ng/ml, was added to the media to deliver a range of expression for each target gene. Double-transconjugants were obtained for both 10 ng/ml and 31 ng/ml aTc and verified by PCR, but unfortunately at the time of writing, no integrants into the desired target gene had been obtained. Due to time constraints, not enough work has been carried out on the method to establish if this method is viable.

5.2.2. Problems with culture collection C. difficile *630Δ*erm *strain*

It must be noted, at the end of this work which includes certainly the double-transconjugants attempts, the *C. difficile* 630Δ*erm* strain from our culture collection stock had started to act peculiarly. The strain took over 8 hours to grow to an OD_{600} of 0.5, as opposed to 4 hour normally, and reached a maximum OD_{600} of 0.6 after 24 hours, as opposed to normally in excess of OD_{600}1.1, also the plasmid pMTL-tet3nO did not produce CatP at all when induced the cells from the same stock. After conferring with my colleagues, we noticed that other colleagues had noted this stock had a decreased sporulation rate, growth rate, and was unable to grow in some media. It might be possible that the strain in our collection had accumulated mutations which have caused these multiple deficiencies. This could account for the inability to insertionally inactivate the native essential genes as the Tet system might not work as would be expected. In light of these observations, the stock which was used was discarded, and a backup stock was used as a replacement. Unfortunately, I was unable to repeat this work due to time constraints.

5.3 Discussion

Evaluation of antimicrobial targets has historically been carried out by screening compounds that inhibit target bacteria[3,160-163]. However, the drive for novel targets has led to new methods being developed to identify novel essential genes, and subsequent rational screening or design to inhibit there product[46,59,78,164,165]. This process of reverse targeting takes into account specific mechanisms of action more than blindly screening for compounds and allows specific metabolic pathways to be targeted. The ability to control essential genes by inducible promoter systems is a powerful tool in not only identifying essential genes, but also to differentiate bacteriostatic and bactericidal antimicrobial targets[116]. Conditional expression allows the target genes transcription to be controlled by the addition or subtraction of an inducer. Gene transcription, and subsequent protein expression, can be decreased or turned off via an inducible promoter, mimicking the binding of an antimicrobial agent to the essential genes product. The resulting conditional mutants can then be monitored by survival kinetics to identify if the target gene is an essential gene, and what type of target it is (figure 5.3)[116].

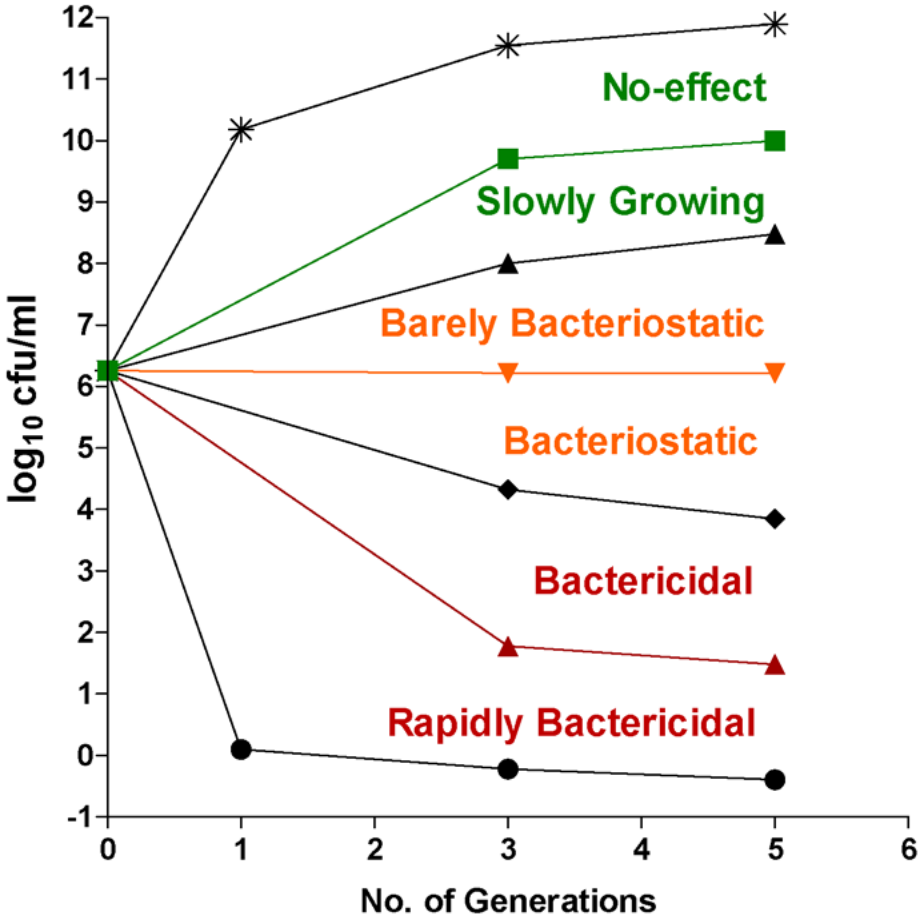


Figure 5.3. Hypothesised survival kinetics of bacteria after exposure by different types of antimicrobial targets. The different types of responses are indicated, with the bacteriostatic and bactericidal responses (Log_{10} cfu/ml) required for an effective antimicrobial. Taken from reference[116].

In this chapter, it was attempted to introduce the Tet system developed in chapter four, in combination with a conditionally expressed second protected copy of a target essential gene inserted into the chromosome in the *pyrE* locus. This second copy allows subsequent inactivation of the native copy, and therefore functional complementation of the native copy, a strategy which was shown to work with a constitutively expressed essential gene in chapter three, except in this strategy the target gene is under inducible control. Unfortunately, the desired procedure was not shown to work within the time frame of this project. It proved possible to construct the three conditional partial

diploid strains containing; *slpA*, *trpS*, and CD0274 under tetracycline inducible control, but a ClosTron integrant in the native target gene of each respective strain could not be obtained. The reason for this inability to inactivate the native gene is unknown but could be due to the inducible promoter system not being able to express each target gene to a high enough level to be able to functionally complement the essential native copy if it were to be inactivated. There are a few different ways to ascertain if this is the reason for the system not working. In the previous chapter, the Tet system was not integrated into the chromosome to test its characteristics in a single copy context. It could be possible the system does not function as well in the chromosome as it does on a plasmid. Additionally, it was not shown that the target genes were in fact being expressed in the cell. One way to achieve this would be to tag the target genes in the chromosome (His·Tag or Streptavidin) and show by western blot that the candidate essential proteins are being produced. This was not carried out at the time due to time constraints, but would be an excellent experiment to narrow down the reasons for this experiment not working, and possibly modifying the Tet system to overcome it.

Working on the assumption that the Tet system was not strong enough to express the target genes to a high enough level to functionally complement the native gene, we decided to place each target gene controlled by the Tet system onto a plasmid, and use a plasmid based system to create a partial diploid. The plasmid based system could allow the copy number of the plasmid to multiply the expression level of each target gene, allowing the second copy to functionally complement the native. This approach was not possible in chapter three with CD0274; however, it might have been due to high expression levels expression, which can be controlled in this system with the addition of varying concentrations of the inducer. The three target genes were easily conjugated into the cell, followed by their respective ClosTron plasmids. Double-transconjugants

containing the partial diploid and the ClosTron plasmid were obtained for each strain, but the subsequent integration into the native target genes were not. The reason for this is unknown but could be linked to the peculiarities noticed in the strain of *C. difficile* 630Δ*erm* being used.

Peculiarities were noticed in the phenotype of the culture collection strain which were used in [parts] of this study. In later experiments, including the last work in this chapter, the culture collection strain of *C. difficile* 630Δ*erm* behaved differently. These differences could be due to accumulation of mutations in the strain caused by constant use within the group. Our group keeps all strains in triplicate with only one being used at a time, and the other two as backups in the eventuality something like this happens. As such the first backup stock was tested by comparing the growth speed between the two stocks, with the backup acting as would be expected. As such, the backup stock was used to replace the current stock. However, this replacement happened after this work had been completed. It is unknown if/when these mutations arose, and as such, it is not known how much of this work is affected. If this is the reason for the experiment not working, repeating the experiment with fresh stocks should overcome this problem.

With further work, this type of system could be a powerful tool to identify not only essential genes, but also the type of antimicrobial drug target. This extra information would be hugely advantageous in applying an efficient antibiotic regime for treating CDI. In most cases, bactericidal antimicrobials outperform bacteriostatic antimicrobials[166]. It has also been suggested that the use of bactericidal antimicrobials decreases the emergence of drug resistance and bacteriostatic drugs, to have efficacy, must be augmented by host defences to clear the infection[167]. These advantageous over bacteriostatic targets would point to targeting and development of bactericidal targets, as they would be the choicest targets for the treatment of CDI.

Chapter Six

General Discussion

6.1. Key Findings of This Work

6.1.1. Identification of C. difficile *essential genes*

To-date, there has not been a single *C. difficile* essential gene identified in the literature. The reason for this shortcoming is partly due to the limitations of current genetic tools. In 2007, Heap *et al*[98], developed the first genetic tool to efficiently and consistently inactivate genes in *Clostridium*, however this tool, the ClosTron, is insufficient alone to identify essential genes. In order to identify essential genes in *C. difficile*, a novel method must be developed. This requirement for a novel tool combined with the importance of identifying *C. difficile* essential genes forms the basis of this study.

To identify candidate essential genes with which to develop a novel method to confirm gene essentiality in *C. difficile*, we first used bioinformatic methods to screen the *C. difficile* genome. Based upon the knockout-rate prediction[51], essential genes evolve at lower rates compared to nonessential genes; we compared all the known essential genes in *B. subtilis* to identify homologues, and as such candidates, in *C. difficile*. Using the translated amino acid sequence of *B. subtilis* essential genes[52], we used the BLAST tool to identify 208 candidate essential genes in *C. difficile*. To narrow down the candidate essential genes for experimental analysis, the introduction of four criteria; sequence similarity, size of the gene, genomic context, and potential for antimicrobial targeting, was taken into account. Although these strict criteria were used for this study, it does not limit the potential of the method to identify essential genes, only increases the possibility of successfully identifying essential genes against which to develop a method. Based upon this, we selected the genes *metK* and *trpS* to analyse experimentally as they are known potential broad-spectrum drug targets, increasing the possibility of being essential in *C. difficile*[97,98].

The main limitation of *in silico* screening is the inability of the method to identify candidate essential genes which have evolved to be essential within a limited number of species. Due to the nature of *C. difficile* infection and the high recurrence rate after treatment being associated with an altered gut microflora caused by broad-spectrum antimicrobials, the development of a narrow-spectrum antimicrobial would be highly advantageous[37]. Therefore, the only narrow-spectrum candidate essential gene identified in the literature was also selected. CD0274 has been proposed to be only essential in *C. difficile* and *C. butyricum*[39,40]. It was previously found that inactivation of the candidate essential gene CD0274, a proposed 1,3 propandiol dehydrogenase, was associated with a lethal phenotype, the first sign that a gene may be essential[39,40]. However, confirmation of essentiality requires further evidence, and a tool able to provide this.

In other bacterial species, a common way to identify essential genes is to employ the use of a partial diploid system. By combining already established tools in *C. difficile*, it was possible to develop a similar novel strategy to functionally characterise essentiality of candidate essential genes, CD0274, *metK*, and *trpS*.

The simplest way to construct a partial diploid is to deliver the second copy of the gene on a plasmid into the cell via conjugation[50]. Unfortunately, the simplest approach to constructing the *C. difficile* partial diploid also proved to be the most troublesome. We noted that it was difficult to produce stable CD0274 transconjugants, with transconjugants showing very restricted growth in liquid media, and pin-prick size colonies on agar with an inability to passage the transconjugants on fresh agar. The reason for this instability is unknown, but could be due to a high level of expression of CD0274 within the cell. CD0274 is thought to be involved in detoxification of methylglyoxal (MG), a toxic metabolite produced as a side-product of glycolysis[40,168,169]. Under certain conditions there is an excessive production of toxic sugar phosphates

produced during glycolysis[40,168,169]. To prevent accumulation of these toxic sugar phosphates, carbon flow is diverted via the MG bypass which allows the organism to survive. Regulation of the toxic metabolite is kept under strict control in most bacteria; however, *C. difficile* lacks many of these protective features making it a possible drug target[99,169,170]. It is this lack of control which may cause this instability in *C. difficile*. A high level of expression of CD0274 may disrupt the natural regulation and subsequent MG detoxification, causing toxic side effects. To circumvent this instability, the alternative method to produce a partial diploid was undertaken, and the candidate essential gene was introduced into the chromosome using the newly developed allele exchange method, ACE (Heap *et al,* unpublished).

The alternative ACE method was successful in producing a stable partial diploid CD0274 strain, as such; the native CD0274 was insertionally inactivated using the ClosTron system with ease. This ability to inactivate CD0274 a total of 140 times demonstrates the gene can be inactivated, only in the presence of the second copy, providing strong evidence of CD0274's essentiality. To exemplify this method further, the two other candidate essential genes were analysed in the same way. The results showed that both the genes, *trpS* and *metK*, were able to be inactivated only in the presence of the second copy. This result not only provides strong evidence that the candidates are essential, but also provides evidence that this method is able to identify essential genes with relative ease. An advantage of using this system is its simplicity in constructing the strains and analysing the results. Because the methods used in this study, i.e. ACE and the ClosTron, are viable in not only other *Clostridium*, but a wide range of bacterial species, this novel method could be used to identify essential genes in any species in which the established techniques will work, these currently include; *C.*

perfringens, *C. botulinum*, and *C. tetani* to name but a few pathogenic *Clostridium* where no essential genes have currently been identified.

There are two disadvantageous of using such a system over other methods. The first is the evidence acquired is indirect, i.e., the measure of essentiality of the gene in the presence of a second copy. There could be an, albeit low, possibility, that the presence of the second copy allows the native gene to be inactivated when the gene is in fact not essential. The second limitation is the static ectopic gene expression level. This static expression limits the amount of information which can be acquired about the function of the essential gene, and what kind of antimicrobial target inhibition of the gene product would pertain to. It has been suggested that bactericidal antimicrobials outperform their bacteriostatic counterparts in their ability to clear an infection, and decrease the emergence of drug resistance[167]. This therefore means the ability to differentiate between bacteriostatic and bactericidal targets would be valuable to prioritise essential genes for antimicrobial targeting.

6.1.2. Development of a tetracycline inducible promoter to identify and differentiate bacteriostatic and bactericidal antimicrobial targets.

The conditional modulation of gene expression levels is proven to be a powerful tool in determining the function of a gene product in bacterial growth and pathogenesis[49,59,116,121]. This ability allows valuable information to be extrapolated by modulating gene expression up or down via the addition or subtraction of an inducer. In contrast to using the ClosTron system to inactivate a gene, resulting in a total loss of function, modulating expression can provide important information on the functional importance of a gene product[116]. In the case of essential genes, it can identify conditional lethal phenotypes and provide information on functional importance or viability and evaluation of inhibition for therapeutic targeting.

Inducible promoters are well characterised in many organisms, unfortunately at the start of this study, there was no suitable inducible promoter system for *C. difficile*. In order to select an inducible promoter system to identify and characterise essential genes in *C. difficile*, there were certain characteristics which would be required; wide dynamic range of activity, tight repression, good inducer bioavailability, high target gene expression, quick induction, and the ability of the system to work in multiple *Clostridium*. In *C. perfringens*, a lactose inducible promoter system was developed, however, the inducer lactose was metabolised after 75 minutes leading to subsequent decrease in induction[171]. This poor bioavailability of lactose excludes the possibility of using such a system in *C. difficile* for essential gene identification; therefore, a new inducible promoter system must be developed. We selected the Tet inducible promoter system as it is known to meet all the requirements which we were looking for, with the exception of a wide dynamic range[126], and was unknown if it would work in *Clostridium* at the start of this study.

We redesigned the Tet system using clostridial components mirroring the architecture of the Tet system present upon pRMC2 with one fundamental difference[132]. In pRMC2's Tet system, there is two *tetO*s present within divergent promoters, one promoter transcribing the target gene and the other *tetO* present within the promoter transcribing *tetR*, the gene encoding the tetracycline repressor[132]. Due to this operator/promoter configuration, in the absence of an inducer, TetR is able to repress not only the target gene, but also production of TetR. This means that there is very low levels of TetR present within the cell. This very low base level of repressor allows very small concentrations of inducer to activate target gene expression, making the system very sensitive to induction, manifesting as a limited dynamic range. To extend the dynamic range of expression, we separated the target gene and *tetR* promoters and removed the *tetO* from the *tetR* promoter.

This new configuration allowed high independent constitutive expression of *tet*R and subsequent tight repression of the target gene in *C. difficile* and *C. sporogenes*. In response to varying concentrations of aTc, the system also showed a good dynamic range of expression in *C. difficile* and *C. sporogenes* compared to previously published Tet systems. In addition, we were able to elicit high levels of induction quickly, enabling short term studies to be carried out. These results indicate that the Tet system (tet3nO) meets all the required characteristics to be used for essential gene identification in *C. difficile*.

In the final study in chapter five, we transcriptionally fused the inducible promoter system to two of the previously identified essential genes, CD0274 and *trpS*. These genes were chosen as CD0274, if essential in very few organisms, would make an excellent candidate narrow-spectrum antimicrobial target. On the other hand, *trpS*, a tryptophanyl-tRNA synthetase, is found in all bacterial species and is known to be a good antimicrobial target, therefore would be an auspicious example to exemplify the method[98]. In addition to these two, we selected a third and final gene, *slpA*. Previous studies in our laboratory failed to inactivate *slpA*, the first indication a gene may be essential. SlpA is of interest as it is the main component of *C. difficile*'s S-layer and unique to *C. difficile*, representing another candidate narrow-spectrum antimicrobial target[31,32].

To introduce the transcriptionally fused inducible candidate essential genes into the *C. difficile* genome, the ACE method which was successful in chapter three was selected. The three genes were introduced into the *pyr* locus easily, however, the attempts at inactivating the native candidate genes was unsuccessful. It seemed as if the native copies were still essential and an inactivation provided a nonviable cell. The reason for this is unknown, but could be due to the inducible promoter system not working as would be expected, stopping functional complementation. Another possible reason

could be peculiarities with our laboratory strain of *C. difficile* 630Δ*erm*, which are outlined in chapter five. Either way, further characterisation of the inducible promoter system's properties within the chromosome and its ability to express the candidate essential genes will identify the problem with the system, which hopefully should be able to be rectified.

Although we have developed this system to be used to identify essential genes in *C. difficile*, the system is not limited to such studies. Inducible promoters can be used in any study which requires the conditional/modulation of expression of any target gene.

6.2. Potential of CD0274, *metK* and *trpS* as Novel Therapeutic Targets.

Essential gene products represent candidate antimicrobial targets as the loss of functionality leads to a loss of cell viability. In order to be an effective antimicrobial target, a gene must be required for growth *in vivo*, as well as *in vitro*. An example is the fatty acid synthesis genes *fabF* and *fabM* of *Staphylococcus agalactiae* which are essential for growth in laboratory media, but only confer an advantage in human serum and *in vivo* infection due to the presence of fatty acids, limiting their ability to be used as targets for the treatment of septicaemia[172]. Unfortunately, our studies only identified CD0274, *metK*, and *trpS*, to be essential for *C. difficile* in laboratory media, however, the function of the products of these essential genes suggests that they are required for survival, and as such, may have potential as targets for the treatment of CDI.

We chose *trpS* and *metK* to exemplify the method because they are known to be essential in other bacterial species, and have been identified to make good novel antimicrobial targets[97,98]. The prospects of using these two broad-spectrum targets of other bacterial species increase the possibility that these two essential genes will also provide good antimicrobial targets against *C. difficile*. However, they might be effective targets against

C. difficile; they might not be effective targets for the treatment of CDI. The main predisposing factor for CDI is broad-spectrum antimicrobial therapy, which often eradicates beneficial flora in the gut, allowing *C. difficile* to flourish. Following treatment for CDI with therapeutic antimicrobials such as; vancomycin and metronidazole, the gut flora is subsequently destroyed again, allowing *C. difficile* to relapse and cause disease again.

There has been considerable interest in a narrow-spectrum antimicrobial target which is able to act upon *C. difficile* alone, allowing the native gut microflora to replenish and inhibit CDI relapse[5,12,111]. The most promising of the novel *C. difficile* antimicrobials is fidaxomicin, an RNA inhibitor which is almost purely specific for *C. difficile*. In comparison against vancomycin, fidaxomicin showed similar efficacy in treating a patient, and significantly lower rate of relapse, highlighting the effectiveness of a narrow-spectrum antimicrobial[37]. In previous studies in *C. difficile*, a species-specific candidate essential gene was serendipitously identified[39,40].

Liyanage et al, (2001)[40], identified CD0274 (they annotated as *gldA*) as a candidate essential gene, thought to be involved in detoxification of toxic metabolite MG. MG is produced from dihydroxyacetone phosphate (DHAP) via MG synthase (MGS), encoded for by *mgsA*[168-170]. There are other proposed routes in higher organisms but variations in bacteria are very poorly documented[173]. Hopper and Cooper were the first to purify MGS from *E. coli* in 1971 and found it to be homotropically activated by DHAP and allosterically inhibited by its other product, inorganic phosphate[174]. The intricacy of MG regulation mechanisms is increased by various glycolytic intermediates being able to regulate MG levels through changes in concentration, as commonly seen with many other catabolic pathways. A significant characteristic of MG is its cytotoxic effect[169,170]. The exact mechanism of action of MG in inhibiting bacterial cell growth or

causing death is unclear, but it has been proposed that MG interferes with protein synthesis through interactions with 7-methylguanosine which is a residue present within 16s and 23s RNA, thereby preventing DNA replication initiation[169,170]. The electrophile can react with the nucleophilic centres of proteins, DNA, and RNA. It has also been shown to react with the side-chains of the amino acids; cysteine, lysine, arginine, and the base guanine, and to a lesser extent with the bases adenine and cytosine[169,170]. Detoxification of MG would seem to be disrupted by the inactivation of CD0274, causing the loss of cell viability through the aforementioned toxicity of MG accumulation. It is this loss of viability which would be an attractive attribute combined, with the specificity of CD0274's essentiality, CD0274 would make a strong candidate novel antimicrobial target for the treatment of CDI.

6.3. The Future of Essential Gene Identification in *C. difficile*

The identification of essential genes is paramount for the development of antimicrobial drugs for the treatment of CDI. In this study, we used bioinformatic approaches to identify candidate essential genes, and subsequently developed a novel method to provide strong evidence of essentiality of three *C. difficile* genes, CD0274, *trpS*, and *metK*. However, these are only three from potentially hundreds of essential genes present within the *C. difficile* genome. The ability to use forward genetic studies to screen the entire genome for essential genes has become popular and has been successful in other bacterial species[44,59,75,78,80,175]. With the advent of a mariner-based transposon system for *C. difficile*, it is now possible to carry out such a system in *C. difficile*[42]. There is two main ways by which to approach this, by using a TraDIS system[77] or a Tn*TetO*ut system[80].

The TraDIS system, which was described in chapter one, has proven to be a powerful method to identify candidate essential genes in S. Typhi by their inability to be inactivated[77]. This type of method, although time consuming, would not only provide

the frequency of insertion for every gene, but also produce a mutant in every gene within the *C. difficile* genome. The main limitation of using this system is that a gene's inability to be inactivated does not equate to the gene being essential, it only defines the gene as a candidate. To provide further evidence for essentiality, each gene identified as a candidate must be analysed by either the strategy taken in chapter three, or alternatively chapter five.

To circumvent this limitation, an adaptation of the TnAraOut transposon system could be adopted. Instead of the P_{BAD} inducible promoter system, it would be possible to substitute the P_{BAD} system for the Tet inducible promoter system developed in chapter four. By using the mariner-based transposon system harbouring the Tet system at the 5' facing outwards, any transcriptionally fused essential genes would be dependent on the addition of aTc for cellular viability[80,81]. This dependence allows high through-put screening of a transposon mutant library to directly identify every possible essential gene within *C. difficile*, with no need for any further methods to be used.

6.4. Proposed Future Work

6.4.1. Conditional lethal phenotypes

The methods developed in this project for identifying essential genes, and for further analysing gene function, have great potential far beyond the scope of this project. In my opinion, the inducible promoter system developed in chapter four is the best outcome of this work, allowing any target gene's expression to be modulated to study the gene's function. In the terms of using conditional expression for essential gene identification as employed in chapter five, if the complications of conditionally controlling lethal phenotypes could be fixed, the method would provide strong direct evidence of a gene's

essentiality and its nature as an antimicrobial target. Therefore, continuation of this study would be extremely valuable.

To complete the development of conditional lethal phenotypes, the following issues must be addressed;

- Characterisation of the Tet system in the genome to ensure the system works; this can be accomplished by introducing the tet3nO construct expressing *catP* in the *pyr* locus using ACE.

- Determine if the essential genes are being expressed within the chromosome in chapter five by tagging each gene with a His·Tag and demonstrating expression using Western blots.

- If the inducible promoter is not strong enough to express the essential genes to a high enough level, either replace the P_{thl} running *tetR* to decrease the concentration of TetR, thereby allowing a lower concentration of aTc to induce a maximum response without being toxic to the cell, or by moving the functional complementation system onto a plasmid, allowing the plasmid copy number to multiply expression of the essential gene to achieve the required concentration to functionally complement.

6.4.2. Broad-host range Clostridium *inducible promoter system*

The development of a broad-host range *Clostridium* vector would still be a very powerful tool for analysing gene function in other *Clostridium*, as one currently does not exist. The Tet system we developed worked very well in *C. difficile* and *C. sporogenes*; however, we were unable to elicit an induction of the system in *C. beijerinckii* and *C. acetobutylicum*. The reason for this is unknown, but could be due to the inducer being actively transported

out of the cell by a native *tetA*(B) gene present within construct *C. acetobutylicum* and *C. beijerinckii*.

To determine if the presence of the *tetA* gene is responsible for the system not working in both these organism, and to further analyse the Tet system as a *Clostridium* inducible promoter system, the following actions must be undertaken;

- Inactivation of the native *tetA* gene in *C. beijerinckii* and *C. acetobutylicum* using the ClosTron system would determine *tetA*'s effect on the Tet system's ability to control the expression of *catP*.

- Further analysis of the inducible promoter system's characteristics in other *Clostridium* would determine its ability to act as a universal *Clostridium* inducible promoter system.

6.5. Concluding Remarks

It is clear that our knowledge of essential genes in *C. difficile* is elementary at best. However, now that we have the tools available to us, we can start identifying and understanding the nature and effects essential genes have upon a cell's viability. In addition, the work carried out in this study has identified a potential novel narrow-spectrum antimicrobial drug target for the treatment of CDI. With the range of tools at our disposal constantly expanding, the path is opening up to thoroughly understand this "essential" aspect of *C. difficile*, and the subsequent development of effective therapeutic antimicrobial drugs is sure to follow.

Bibliography

1 Liyanage, H., Kashket, S., Young, M. & Kashket, E. R. *Clostridium beijerinckii* and *Clostridium difficile* detoxify methylglyoxal by a novel mechanism involving glycerol dehydrogenase. *Applied and Environmental Microbiology* **67**, 2004-2010 (2001).

2 Lynch, S. A., Desai, S. K., Sajja, H. K. & Gallivan, J. P. A high throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chem Biol* **14**, 173-184, (2007).

3 Chalker, A. F. & Lunsford, R. D. Rational identification of new antibacterial drug targets that are essential for viability using a genomics-based approach. *Pharmacology & Therapeutics* **95**, 1-20, (2002).

4 McDevitt, D., Payne, D. J., Holmes, D. J. & Rosenberg, M. Novel targets for the future of antimicrobial agents. *J Appl Microbiol* Vol. 92 Suppl 28S-34S (2002).

5 Black, M. T. & Hodgson, J. Novel target sites in bacteria for overcoming antibiotic resistance. *Advanced Drug Delivery Reviews* **57**, 1528-1538, (2005).

6 Aslam, S., Hamill, R. & Musher, D. M. Treatment of *Clostridium difficile*-associated disease: old therapies and new strategies. *Lancet Infectious Diseases* **5**, 549-557, (2005).

7 Razavi, B., Apisarnthanarak, A. & Mundy, L. M. *Clostridium difficile*: Emergence of hypervirulence and fluoroquinolone resistance. *Infection* **35**, 300-307, (2007).

8 Mylonakis, E., Ryan, E. T. & Calderwood, S. B. *Clostridium difficile*-associated diarrhea - A review. *Archives of Internal Medicine* **161**, 525-533, (2001).

9 Cookson, B. Hypervirulent strains of *Clostridium difficile*. *Postgraduate Medical Journal* **83**, 291-295, (2007).

10 Bignardi, G. E. Risk factors for *Clostridium difficile* infection. *Journal of Hospital Infection* **4 0**, 1-15, (1998).

11 Thomas, C., Stevenson, M. & Riley, T. V. Antibiotics and hospital-acquired *Clostridium difficile*-associated diarrhoea: a systematic review. *Journal of Antimicrobial Chemotherapy* **51**, 1339-1350, (2003).

12 Dial, S., Delaney, J. A. C., Barkun, A. N. & Suissa, S. Use of gastric acid-suppressive agents and the risk of community-acquired *Clostridium difficile*-associated disease. *Jama-Journal of the American Medical Association* **294**, 2989-2995, (2005).

13 Kuijper, E. J., Coignard, B. & Tull, P. Emergence of *Clostridium difficile*-associated disease in North America and Europe. *Clin Microbiol Infect* **12 Suppl 6**, 2-18, (2006).

14 Warny, M. *et al.* Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* **366**, 1079-1084, (2005).

15 Monaghan, T., Boswell, T. & Mahida, Y. R. Recent advances in *Clostridium difficile*-associated disease. *Gut***57**, 850-860, (2008).

16 HALL, I. C. & O'TOOLE, E. Intestinal flora in new-born Iinfants: With a description of a new pathogenic anaerobe, *Bacillus difficilis*. *Am J Dis Child* **49**, 390-402, (1935).

17 Bartlett, J. G., Onderdonk, A. B., Cisneros, R. L. & Kasper, D. L. Clindamycin-associated colitis due to a toxin-producing species of *Clostridium* in hamsters. *J Infect Dis* **136**, 701-705, (1977).

18 Kelly, C. P. & LaMont, J. T. *Clostridium difficile* infection. *Annu Rev Med* **49**, 375-390, (1998).

19 McFarland, L. V., Mulligan, M. E., Kwok, R. Y. & Stamm, W. E. Nosocomial acquisition of *Clostridium difficile* infection. *N Engl J Med* **320**, 204-210, (1989).

20 Burns, D. A., Heap, J. T. & Minton, N. P. *Clostridium difficile* spore germination: an update. *Res Microbiol* **161**, 730-734, (2010).

21 Burns, D. A., Heap, J. T. & Minton, N. P. SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate. *J Bacteriol* **192**, 657-664, (2010).

22 Fawley, W. N. & Wilcox, M. H. Molecular epidemiology of endemic *Clostridium difficile* infection. *Epidemiology and Infection* **126**, 343-350, (2001).

23 Wilcox, M. H. Gastrointestinal disorders and the critically ill. *Clostridium difficile* infection and pseudomembranous colitis. *Best Pract Res Clin Gastroenterol* **17**, 475-493, (2003).

24 Kuijper, E. J., van Dissel, J. T. & Wilcox, M. H. *Clostridium difficile*: changing epidemiology and new treatment options. *Curr Opin Infect Dis* **20**, 376-383, (2007).

25 Kuijper, E. J. *et al.* Update of *Clostridium difficile* infection due to PCR ribotype 027 in Europe, 2008. *Euro Surveill* **13,** (2008).

26 Brazier, J. S. *et al.* Distribution and antimicrobial susceptibility patterns of *Clostridium difficile* PCR ribotypes in English hospitals, 2007-08. *Euro Surveill* **13,** (2008).

27 Burns, D. A., Heap, J. T. & Minton, N. P. The diverse sporulation characteristics of *Clostridium difficile* clinical isolates are not associated with type. *Anaerobe* **16**, 618-622, (2010).

28 Voth, D. E. & Ballard, J. D. *Clostridium difficile* toxins: Mechanism of action and role in disease. *Clinical Microbiology Reviews* **18**, 247-260, (2005).

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

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

31 Kuehne, S. A. *et al.* The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature* **467**, 711-713, (2010).

32 Karjalainen, T. *et al.* Molecular and genomic analysis of genes encoding surface-anchored proteins from *Clostridium difficile*. *Infect Immun* **69**, 3442-3446, (2001).

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

34 Rupnik, M., Wilcox, M. H. & Gerding, D. N. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol* **7**, 526-536, (2009).

35 Drudy, D. *et al.* Human antibody response to surface layer proteins in *Clostridium difficile* infection. *FEMS Immunol Med Microbiol* **41**, 237-242, (2004).

36 Tasteyre, A., Barc, M. C., Collignon, A., Boureau, H. & Karjalainen, T. Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect Immun* **69**, 7937-7940, (2001).

37 Ausiello, C. M. *et al.* Surface layer proteins from *Clostridium difficile* induce inflammatory and regulatory cytokines in human monocytes and dendritic cells. *Microbes Infect* **8**, 2640-2646, (2006).

38 Louie, T. J. *et al.* Fidaxomicin versus vancomycin for *Clostridium difficile* infection. *N Engl J Med* **364**, 422-431, (2011).

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

40 Liyanage, H., Young, M. & Kashket, E. R. Butanol tolerance of *Clostridium beijerinckii* NCIMB 8052 associated with down-regulation of *gldA* by antisense RNA. *Journal of Molecular Microbiology and Biotechnology* **2**, 87-93 (2000).

41 Heap, J. T. *et al.* The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *Journal of Microbiological Methods* **80**, 49-55, (2010).

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

43 Gerdes, S. Y. *et al.* Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* **185**, 5673-5684, (2003).

44 Serina, S. *et al.* Scanning the *Escherichia coli* chromosome by random transposon mutagenesis and multiple phenotypic screening. *Res Microbiol* **155**, 692-701, (2004).

45 Ogasawara, N. Systematic function analysis of *Bacillus subtilis* genes. *Research in Microbiology* **151**, 129-134, (2000).

46 Forsyth, R. A. *et al.* A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol Microbiol* Vol. 43 1387-1400, (2002).

47 Ji, Y., Woodnutt, G., Rosenberg, M. & Burnham, M. K. R. Identification of essential genes in *Staphylococcus aureus* using inducible antisense RNA. *Methods in Enzymology* Vol. Volume 358 (ed Patrik M. Bavoil Virginia L. Clark) 123-128, (Academic Press, 2002).

48 Mushegian, A. R. & Koonin, E. V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA* **93**, 10268-10273, (1996).

49 Zhang, L. *et al.* Regulated gene expression in *Staphylococcus aureus* for identifying conditional lethal phenotypes and antibiotic mode of action. *Gene* **255**, 297-305, (2000).

50 Boyle, D. S. & Donachie, W. D. *mraY* is an essential gene for cell growth in *Escherichia coli*. *Journal of Bacteriology* **180**, 6429-6432, (1998).

51 Jordan, I. K., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res* **12**, 962-968, (2002).

52 Kobayashi, K. *et al.* Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA* Vol. **100**, 4678-4683, (2003).

53 Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006-2008, (2006).

54 Salama, N. R., Shepherd, B. & Falkow, S. Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. *J Bacteriol* **186**, 7926-7935, (2004).

55 Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science* **269**, 496-512, (1995).

56 Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397-403, (1995).

57 Itaya, M. An estimation of minimal genome size required for life. *FEBS Lett* **362**, 257-260, (1995).

58 Bruccoleri, R. E., Dougherty, T. J. & Davison, D. B. Concordance analysis of microbial genomes. *Nucleic Acids Research* **26**, 4482-4486, (1998).

59 Akerley, B. J. *et al.* Systematic identification of essential genes by *in vitro* mariner mutagenesis. *Proc Natl Acad Sci USA* **95**, 8927-8932, (1998).

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

61 Karberg, M. *et al.* Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat Biotech* **19**, 1162-1167, (2001).

62 Zhong, J., Karberg, M. & Lambowitz, A. M. Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Research* **31**, 1656-1664, (2003).

63 Cousineau, B. *et al.* Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* **94**, 451-462, (1998).

64 Yao, J. & Lambowitz, A. M. Gene targeting in Gram-negative bacteria by use of a mobile group II intron ("TargeTron") expressed from a broad-host-range vector. *Appl Environ Microbiol* Vol. **73,** 2735-2743, (2007).

65 Yin, D. & Ji, Y. Identification of essential genes in *Staphylococcus aureus* by construction and screening of conditional mutant library. *Methods Mol Biol* **16**, 297-305, (2008).

66 Herring, C. D. & Blattner, F. R. Conditional lethal amber mutations in essential *Escherichia coli* genes. *Journal of Bacteriology* **186**, 2673-2681, (2004).

67 Fan, F. *et al.* Regulated Ectopic Expression and Allelic-Replacement Mutagenesis as a Method for Gene Essentiality Testing in *Staphylococcus aureus*. *Plasmid* **46**, 71-75, (2001).

68 Yansura, D. G. & Henner, D. J. Use of the *Escherichia coli lac* repressor and operator to control gene expression in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **81**, 439-443, (1984).

69 Geissendorfer, M. & Hillen, W. Regulated expression of heterologous genes in *Bacillus subtilis* using the Tn*10* encoded tet regulatory elements. *Appl Microbiol Biotechnol* **33**, 657-663, (1990).

70 Schmid, M. B., Kapur, N., Isaacson, D. R., Lindroos, P. & Sharpe, C. Genetic analysis of temperature-sensitive lethal mutants of *Salmonella typhimurium*. *Genetics* **123**, 625-633, (1989).

71 Blomfield, I. C., Vaughn, V., Rest, R. F. & Eisenstein, B. I. Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. *Mol Microbiol* **5**, 1447-1457, (1991).

72 Yao, J. *et al.* Use of targetrons to disrupt essential and nonessential genes in *Staphylococcus aureus* reveals temperature sensitivity of Ll.LtrB group II intron splicing. *RNA-a Publication of the Rna Society* **12**, 1271-1281, (2006).

73 Franklin, F. C. H. & Williams, P. A. Construction of a partial diploid for the degradative pathway encoded by the Tol plasmid (Pwwo) from *Pseudomonas Putida* Mt-2 - Evidence for the positive nature of the

regulation by the *xylR* gene. *Molecular & General Genetics* **177**, 321-328, (1980).

74 Hayes, F. Transposon-based strategies for microbial functional genomics and proteomics. *Annu Rev Genet* **37**, 3-29, (2003).

75 Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**, 77-84, (2003).

76 Jacobs, M. A. *et al.* Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* **100**, 14339-14344, (2003).

77 Langridge, G. C. *et al.* Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res* **19**, 2308-2316, (2009).

78 Akerley, B. J. *et al.* A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc Natl Acad Sci USA* **99**, 966-971, (2002).

79 Bae, T. *et al. Staphylococcus aureus* virulence genes identified by bursa aurealis mutagenesis and nematode killing. *Proc Natl Acad Sci USA* **101**, 12312-12317, (2004).

80 Judson, N. & Mekalanos, J. J. TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. *Nat Biotechnol* **18**, 740-745, (2000).

81 Kim, J. N., Youm, G. W. & Kwon, Y. M. Essential genes in *Salmonella enteritidis* as identified by TnAraOut mutagenesis. *Curr Microbiol* **57**, 391-394, (2008).

82 Bull, M. K., Olivier, S. A., van Diepenbeek, R. J., Kormelink, F. & Chapman, B. Synergistic inactivation of spores of proteolytic *Clostridium botulinum* strains by high pressure and heat is strain and product dependent. *Appl Environ Microbio* **75**, 434-445, (2009).

83 Ocio, M. J., Sanchez, T., Fernandez, P. S., Rodrigo, M. & Martinez, A. Thermal resistance characteristics of PA 3679 in the temperature range of

110-121 degrees C as affected by pH, type of acidulant and substrate. *Int J Food Microbiol* **22**, 239-247, (1994).

84 Minton, N. P. Clostridia in cancer therapy. *Nat Rev Microbiol* **1**, 237-242, (2003).

85 Nolling, J. *et al.* Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J Bacteriol* **183**, 4823-4838, (2001).

86 Ezeji, T. C., Qureshi, N. & Blaschek, H. P. Butanol fermentation research: upstream and downstream manipulations. *Chem Rec* **4**, 305-314, (2004).

87 George, H. A., Johnson, J. L., Moore, W. E., Holdeman, L. V. & Chen, J. S. Acetone, Isopropanol, and Butanol Production by *Clostridium beijerinckii* (syn. *Clostridium butylicum*) and *Clostridium* aurantibutyricum. *Appl Environ Microbiol* **45**, 1160-1163 (1983).

88 Qureshi, N. & Blaschek, H. P. ABE production from corn: a recent economic evaluation. *J Ind Microbiol Biotechnol* **27**, 292-297, (2001).

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

90 Purdy, D. *et al.* Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Molecular Microbiology* **46**, 439-452, (2002).

91 Shaw, W. V. Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria. *Methods Enzymol* **43**, 737-755, (1975).

92 Ellman, G. & Lysko, H. A precise method for the determination of whole blood and plasma sulfhydryl groups. *Anal Biochem* **93**, 98-102, (1979).

93 Perutka, J., Wang, W., Goerlitz, D. & Lambowitz, A. M. Use of a computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J Mol Biol* Vol. **336**. 21-439, (2004).

94 Breithaupt, H. The new antibiotics. *Nature Biotechnology* **17**, 1165-1169, (1999).

95 Parish, T. & Stoker, N. G. *glnE* is an essential gene in *Mycobacterium tuberculosis*. *J Bacteriol* **182**, 5715-5720 (2000).

96 Parish, T. *et al.* Functional complementation of the essential gene *fabG1* of *Mycobacterium tuberculosis* by *Mycobacterium smegmatis fabG* but not *Escherichia coli fabG*. *Journal of Bacteriology* **189**, 3721-3728, (2007).

97 Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nat Biotech* **24**, 1558-1564, (2006).

98 Hurdle, J. G., O'Neill, A. J. & Chopra, I. Prospects for Aminoacyl-tRNA Synthetase Inhibitors as New Antimicrobial Agents. *Antimicrob Agents Chemother* **49**, 4821-4833, (2005).

99 Subedi, K. P., Kim, I., Kim, J., Min, B. & Park, C. Role of GldA in dihydroxyacetone and methylglyoxal metabolism of *Escherichia coli* K12. *Fems Microbiology Letters* **279**, 180-187, (2008).

100 Raynaud, C., Sarcabal, P., Meynial-Salles, I., Croux, C. & Soucaille, P. Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. *Proc Natl Acad Sci USA* **100**, 5010-5015, (2003).

101 Budd, M. & Campbell, J. L. Temperature-sensitive mutations in the yeast DNA-Polymerase-I gene. *PROC NATL ACAD SCI USA* **84**, 2838-2842, (1987).

102 Poulsen, P., Bonekamp, F. & Jensen, K. F. Structure of the *Escherichia coli pyrE* operon and control of *pyrE* expression by a UTP modulated intercistronic attentuation. *EMBO J* **3**, 1783-1790, (1984).

103 Yamagishi, A., Tanimoto, T., Suzuki, T. & Oshima, T. Pyrimidine biosynthesis genes (*pyrE* and *pyrF*) of an extreme thermophile, *Thermus thermophilus*. *Appl Environ Microbiol* **62**, 2191-2194, (1996).

104 Bartlett, J. G. Narrative review: The new epidemic of *Clostridium difficile*-associated enteric disease. *Annals of Internal Medicine* **145**, 758-764, (2006).

105 Noren, T. Outbreak from a high-toxin intruder: *Clostridium difficile*. *Lancet* **366**, 1053-1054, (2005).

106 Owens, R. C. *Clostridium difficile* - Associated disease: An emerging threat to patient safety - Insights from the society of infectious diseases pharmacists. *Pharmacotherapy* **26**, 299-311, (2006).

107 Pepin, J. *et al. Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. *Canadian Medical Association Journal* **171**, 466-472, (2004).

108 Sebaihia, M. *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genetics* **38**, 779-786, (2006).

109 Wilcox, M. H. Cleaning up *Clostridium difficile* infection. *Lancet* **348**, 767-768, (1996).

110 Wilcox, M. H. *Clostridium difficile* - setting the scene. *Journal of Antimicrobial Chemotherapy* **41**, 1-3, (1998).

111 Wilcox, M. H. Treatment of *Clostridium difficile* infection. *Journal of Antimicrobial Chemotherapy* **41**, 41-46, (1998).

112 Wilcox, M. H. & Fawley, W. N. Hospital disinfectants and spore formation by *Clostridium difficile*. *Lancet* **356**, 1324-1324, (2000).

113 Blokpoel, M. C. *et al.* Tetracycline-inducible gene regulation in mycobacteria. *Nucleic Acids Res* **33**, e22, (2005).

114 Bateman, B. T., Donegan, N. P., Jarry, T. M., Palma, M. & Cheung, A. L. Evaluation of a tetracycline-inducible promoter in *Staphylococcus aureus in vitro* and *in vivo* and its application in demonstrating the role of *sigB* in microcolony formation. *Infect Immun* **69**, 7851-7857, (2001).

115 Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **72**, 211-222, (2006).

116 Kaur, P., Agarwal, S. & Datta, S. Delineating bacteriostatic and bactericidal targets in mycobacteria using IPTG inducible antisense expression. *PLoS ONE* **4**, e5923, (2009).

117 Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. *Science* **263**, 1425-1429, (1994).

118 Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol* **7**, 215-219, (2000).

119 Ramos, J. L. *et al.* The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* **69**, 326-356, (2005).

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

121 Shang, Y. *et al.* A robust inducible-repressible promoter greatly facilitates gene knockouts, conditional expression, and overexpression of homologous and heterologous genes in *Tetrahymena thermophila*. *Proc Natl Acad Sci USA* **99**, 3734-3739, (2002).

122 Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res* **32**, 1610-1614, (2004).

123 Soukup, G. A. & Breaker, R. R. Engineering precision RNA molecular switches. *Proc Natl Acad Sci USA* **96**, 3584-3589, (1999).

124 Jayasuriya, H. *et al.* Discovery and antibacterial activity of glabramycin A-C from *Neosartorya glabra* by an antisense strategy. *J Antibiot* **62**, 265-269, (2009).

125 Levskaya, A. *et al.* Synthetic biology: engineering *Escherichia coli* to see light. *Nature* **438**, 441-442, (2005).

126 Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**, 1203-1210, (1997).

127 Siegele, D. A. & Hu, J. C. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc Natl Acad Sci USA* **94**, 8168-8172, (1997).

128 Lutz, R., Lozinski, T., Ellinger, T. & Bujard, H. Dissecting the functional program of *Escherichia coli* promoters: the combined mode of action of Lac repressor and AraC activator. *Nucleic Acids Res* **29**, 3873-3881, (2001).

129 Beckwith, J. R. Regulation of the *lac* operon. *Science* **156**, 597-604, (1967).

130 Romanuka, J. *et al.* Specificity and affinity of Lac repressor for the auxiliary operators O2 and O3 are explained by the structures of their protein-DNA complexes. *J Mol Biol* **390**, 478-489, (2009).

131 Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588-592, (2005).

132 Corrigan, R. M. & Foster, T. J. An improved tetracycline-inducible expression vector for *Staphylococcus aureus*. *Plasmid* **61**, 126-129, (2009).

133 Stary, E. *et al.* New Architectures for Tet-On and Tet-Off Regulation in *Staphylococcus aureus*▽. *Appl Environ Microbiol* **76**, 680-687, (2010).

134 Levy, S. B. & McMurry, L. Plasmid-determined tetracycline resistance involves new transport systems for tetracycline. *Nature* **276**, 90-92, (1978).

135 Rodríguez-García, A., Combes, P., Pérez-Redondo, R., Smith, M. C. A. & Smith, M. C. M. Natural and synthetic tetracycline-inducible promoters for use in the antibiotic-producing bacteria *Streptomyces*. *Nucleic Acids Research* **33**, e87, (2005)

136 Ehrt, S. *et al.* Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res* **33**, e21, (2005).

137 Scholz, O. *et al.*Activity reversal of Tet repressor caused by single amino acid exchanges. *Mol Microbiol* Vol. **53**, 777-789, (2004).

138 Sinha, J., Topp, S. & Gallivan, J. P. From SELEX to cell dual selections for synthetic riboswitches. *Methods Enzymol* **497**, 207-220, (2011).

139 Müller, J., Oehler, S. & Müller-Hill, B. Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator. *Journal of Molecular Biology* **257**, 21-29, (1996).

140 Jana, M., Luong, T. T., Komatsuzawa, H., Shigeta, M. & Lee, C. Y. A method for demonstrating gene essentiality in *Staphylococcus aureus*. *Plasmid* **44**, 100-104, (2000).

141 Jensen, P. R. & Hammer, K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl Environ Microbiol* **64**, 82-87, (1998).

142 Sadler, J. R., Sasmor, H. & Betz, J. L. A perfectly symmetric *lac* operator binds the *lac* repressor very tightly. *Proc Natl Acad Sci USA* **80**, 6785-6789, (1983).

144 de Smit, M. H. & van Duin, J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc Natl Acad Sci USA* **87**, 7668-7672, (1990).

145 de Smit, M. H. & van Duin, J. Control of Translation by mRNA Secondary Structure in *Escherichia coli*: A Quantitative Analysis of Literature Data. *Journal of Molecular Biology* **244**, 144-150, (1994).

146 de Smit, M. H. & van Duin, J. Translational initiation on structured messengers: Another role for the shine-dalgarno interaction. *Journal of Molecular Biology* **235**, 173-184, (1994).

147 Liu, D. *et al.* The essentiality and involvement of *Streptococcus intermedius* histone-like DNA-binding protein in bacterial viability and normal growth. *Mol Microbiol* **68**, 1268-1282, (2008).

148 Padidam, M. Chemically regulated gene expression in plants. *Curr Opin Plant Biol* **6**, 169-177, (2003).

149 Blau, H. M. & Rossi, F. M. Tet B or not tet B: advances in tetracycline-inducible gene expression. *Proc Natl Acad Sci USA* **96**, 797-799, (1999).

150 Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766-1769, (1995).

151 Urlinger, S. *et al.* Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc Natl Acad Sci USA* **97**, 7963-7968, (2000).

152 Huang, H. H., Camsund, D., Lindblad, P. & Heidorn, T. Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Res* **38**, 2577-2593, (2010).

153 Hansen, L. H., Knudsen, S. & Sorensen, S. J. The effect of the *lacY* gene on the induction of IPTG inducible promoters, studied in *Escherichia coli* and *Pseudomonas fluorescens.Curr Microbiol* **36**, 341-347, (1998).

154 Satya Lakshmi, O. & Rao, N. M. Evolving Lac repressor for enhanced inducibility. *Protein Eng Des Sel* **22**, 53-58, (2009).

155 Rohrer, S., Ehlert, K., Tschierske, M., Labischinski, H. & Berger-Bachi, B. The essential *Staphylococcus aureus* gene *fmhB* is involved in the first step of peptidoglycan pentaglycine interpeptide formation. *Proc Natl Acad Sci USA* **96**, 9351-9356, (1999).

156 Herring, C. D., Glasner, J. D. & Blattner, F. R. Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. *Gene* **311**, 153-163, (2003).

157 Herring, C. D. & Blattner, F. R. Global transcriptional effects of a suppressor tRNA and the inactivation of the regulator *frmR*. *Journal of Bacteriology* **186**, 6714-6720, (2004).

158 Daniel, R., Boenigk, R. & Gottschalk, G. Purification of 1,3-Propanediol Dehydrogenase from *Citrobacter freundii* and Cloning, Sequencing, and Overexpression of the Corresponding Gene in *Escherichia coli*. *Journal of Bacteriology* **177**, 2151-2156, (1995).

159 Luers, F., Seyfried, M., Daniel, R. & Gottschalk, G. Glycerol conversion to 1,3-propanediol by *Clostridium pasteurianum*: cloning and expression of the gene encoding 1,3-propanediol dehydrogenase. *Fems Microbiology Letters* **154**, 337-345, (1997).

160 Collins, L. & Franzblau, S. G. Microplate alamar blue assay versus BACTEC 460 system for high-throughput screening of compounds against *Mycobacterium tuberculosis* and *Mycobacterium avium*. *Antimicrob Agents Chemother* **41**, 1004-1009, (1997).

161 De La Fuente, R., Sonawane, N. D., Arumainayagam, D. & Verkman, A. S. Small molecules with antimicrobial activity against *E. coli* and *P. aeruginosa* identified by high-throughput screening. *Br J Pharmacol* **149**, 551-559, (2006).

162 Rathinakumar, R., Walkenhorst, W. F. & Wimley, W. C. Broad-spectrum antimicrobial peptides by rational combinatorial design and high-throughput screening: the importance of interfacial activity. *J Am Chem Soc* **131**, 7609-7617, (2009).

163 Trias, J. & Gordon, E. M. Innovative approaches to novel antibacterial drug discovery. *Current Opinion in Biotechnology* **8**, 757-762, (1997).

164 Knuth, K., Niesalla, H., Hueck, C. J. & Fuchs, T. M. Large-scale identification of essential *Salmonella* genes by trapping lethal insertions. *Mol Microbiol* **51**, 1729-1744, (2004).

165 Amini, S. & Tavazoie, S. Antibiotics and the post-genome revolution. *Curr Opin Microbiol* **14**, 512-518, (2011).

166 Levison, M. E. Pharmacodynamics of antimicrobial drugs. *Infectious Disease Clinics of North America* **18**, 451-465, (2004).

167 Stratton, C. W. Dead bugs don't mutate: susceptibility issues in the emergence of bacterial resistance. *Emerg Infect Dis* **9**, 10-16, (2003).

168 Freedber.Wb, Kistler, W. S. & Lin, E. C. C. Lethal synthesis of methylglyoxal by *Escherichia coli* during unregulated glycerol metabolism. *Journal of Bacteriology* **108**, 137-142, (1971).

169 Ferguson, G. P., Totemeyer, S., MacLean, M. J. & Booth, I. R. Methylglyoxal production in bacteria: suicide or survival? *Archives of Microbiology* **170**, 209-219, (1998).

170 Braun, L., Garzo, T., Riba, P., Mandl, J. & Kalapos, M. P. Methylglyoxal and cell viability. *International Journal of Biochemistry* **26**, 987-990, (1994).

171 Hartman, A. H., Liu, H. & Melville, S. B. Construction and characterization of a lactose-inducible promoter system for controlled gene expression in *Clostridium perfringens*. *Appl Environ Microbiol* **77**, 471-478, (2011).

172 Brinster, S. *et al.* Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens. *Nature* **458**, 83-86, (2009).

173 Kalapos, M. P. Methylglyoxal toxicity in mammals. *Toxicology Letters* **73**, 3-24, (1994).

174 Hopper, D. J. & Cooper, R. A. Regulation of *Escherichia coli* methylglyoxal synthase - New control site in glycolysis. *Febs Letters* **13**, 213-219, (1971).

175 Judson, N. & Mekalanos, J. J. Transposon-based approaches to identify essential bacterial genes. *Trends Microbiol* Vol. **8,** 521-526, (2000).

Appendix

Appendix 1. All candidate essential genes of *C. difficile* based upon *in silico* analysis of similarity to *B. subtilis* essential genes[52].

NF = Not Found

Gene	CD Number	Protein Homology	Size	Annotation
dnaA	CD0001	Identities = 278/448 (62%), Positives = 350/448 (78%)	1317 bp	
dnaB	NF			
dnaC	CD3657	Identities = 224/431 (51%), Positives = 325/431 (75%)	1329 bp	*dnaB*
dnaD	NF			
dnaE	CD3397	Identities = 473/1198 (39%), Positives = 699/1198 (58%)	3570 bp	
dnaG	CD1454	Identities = 186/606 (31%), Positives = 342/606 (56%)	1827 bp	
dnaI	CD0410	Identities = 50/162 (31%), Positives = 85/162 (52%)	852 bp	
dnaN	CD0002	Identities = 124/360 (34%), Positives = 215/360 (60%)	1209 bp	
dnaX	CD0016	Identities = 192/480 (40%), Positives = 303/480 (63%)	1.6Kbp	
yqeN	NF			
holB	CD3549	Identities = 95/304 (31%), Positives = 161/304 (52%)	936 bp	
ligA	CD3309	Identities = 365/667 (54%), Positives = 498/667 (74%)	2Kbp	
pcrA	CD0328	Identities = 355/754 (47%), Positives = 496/754 (66%)	2265 bp	
polC	CD1305	Identities = 678/1420 (48%), Positives = 942/1420 (66%)	4299 bp	
priA	CD2586	Identities = 324/830 (39%), Positives = 501/830 (60%)	2490 bp	
ssb	CD3662	Identities = 41/105 (39%), Positives = 61/105 (58%)	435 bp	
gyrA	CD0006	Identities = 471/811 (58%), Positives = 638/811 (79%)	2427 bp	
gyrB	CD0005	Identities = 407/634 (64%), Positives = 508/634 (80%)	1902 bp	
hbs	CD3496	Identities = 50/92 (54%), Positives = 69/92 (75%)	276 bp	*hupA*
parC	CD0006	Identities = 322/731 (44%), Positives = 490/731 (67%)	2427 bp	*gyrA*
parE	CD0005	Identities = 341/633 (54%), Positives = 452/633 (71%)	1902 bp	*gyrB*
smc	CD1250	Identities = 449/1222 (37%), Positives = 734/1222 (60%)	3555 bp	
topA	CD1274	Identities = 390/691 (56%), Positives = 501/691 (72%)	2.1KBp	
ypuG	NF			
ypuH	NF			
ydiO	CD3147	Identities = 56/218 (26%), Positives = 90/218 (41%)	1626 bp	
ydiP	CD0927	Identities = 63/171 (37%), Positives = 93/171 (54%)	1074 bp	
rpoA	CD0098	Identities = 209/311 (67%), Positives = 255/311 (82%)	948 bp	
rpoB	CD0066	Identities = 783/1183 (66%), Positives = 957/1183 (81%)	3717 bp	
rpoC	CD0067	Identities = 785/1184 (66%), Positives = 940/1184 (79%)	3486 bp	
sigA	CD1455	Identities = 248/371 (66%), Positives = 297/371 (80%)	1173bp	
cca	CD2466	Identities = 90/202 (45%), Positives = 135/202 (67%)	1347 bp	
cspR	CD0566	Identities = 89/159 (55%), Positives = 117/159 (73%)	471bp	
(m)rnc	CD0053	Identities = 50/120 (42%), Positives = 80/120 (67%)	396 bp	
rnpA	CD3679	Identities = 45/101 (44%), Positives = 72/101 (71%)	345bp	
trmD	CD1256	Identities = 121/226 (53%), Positives = 168/226 (74%)	696bp	

trmU	CD1281	Identities = 152/360 (42%), Positives = 223/360 (62%)	1086 bp	
yycF	NF			
yycG	NF			
ybdL	CD0355	Identities = 15/56 (27%), Positives = 31/56 (55%)	1191 bp	
nusA	CD1307	Identities = 193/313 (61%), Positives = 255/313 (81%)	1115bp	
rplA	CD0062	Identities = 157/228 (69%), Positives = 186/228 (82%)	699 bp	
rplB	CD0076	Identities = 189/274 (69%), Positives = 222/274 (81%)	831 bp	
rplC	CD0073	Identities = 126/206 (61%), Positives = 160/206 (78%)	630 bp	
rplD	CD0074	Identities = 117/207 (57%), Positives = 148/207 (71%)	654 bp	
rplE	CD0084	Identities = 124/178 (70%), Positives = 153/178 (86%)	543 bp	
rplF	CD0086	Identities = 106/179 (59%), Positives = 135/179 (75%)	543 bp	
rplI	CD3658	Identities = 75/147 (51%), Positives = 108/147 (73%)	450 bp	
rplJ	CD0063	Identities = 78/165 (47%), Positives = 110/165 (67%)	507 bp	
rplL	CD0064	Identities = 77/121 (64%), Positives = 98/121 (81%)	366 bp	
rplM	CD0104	Identities = 72/139 (52%), Positives = 98/139 (71%)	432 bp	
rplN	CD0082	Identities = 98/122 (80%), Positives = 111/122 (91%)	369 bp	
rplO	CD0089	Identities = 93/147 (63%), Positives = 113/147 (77%)	444 bp	
rplP	CD0080	Identities = 103/143 (72%), Positives = 122/143 (85%)	432 bp	
rplQ	CD0099	Identities = 70/118 (59%), Positives = 84/118 (71%)	342 bp	
rplR	CD0087	Identities = 79/122 (65%), Positives = 97/122 (80%)	369 bp	
rplS	CD1257	Identities = 76/114 (67%), Positives = 95/114 (83%)	351 bp	
rplT	CD0687	Identities = 73/117 (62%), Positives = 91/117 (78%)	357 bp	
rplU	CD1161	Identities = 61/103 (59%), Positives = 79/103 (77%)	312 bp	
rplV	CD0078	Identities = 67/111 (60%), Positives = 85/111 (77%)	336 bp	
rplW	CD0075	Identities = 60/95 (63%), Positives = 70/95 (74%)	291 bp	
rplX	CD0083	Identities = 64/101 (63%), Positives = 82/101 (81%)	309 bp	
rpmA	CD1163	Identities = 68/93 (73%), Positives = 78/93 (84%)	291 bp	
rpmB	CD2562A	Identities = 30/61 (49%), Positives = 42/61 (69%)	189 bp	
rpmC	CD0080A	Identities = 39/62 (63%), Positives = 51/62 (82%)	204 bp	
rplD	CD0074	Identities = 117/207 (57%), Positives = 148/207 (71%)	654 bp	
rplE	CD0084	Identities = 124/178 (70%), Positives = 153/178 (86%)	543 bp	
rplF	CD0086	Identities = 106/179 (59%), Positives = 135/179 (75%)	543 bp	
rpmGA	CD0058A	Identities = 32/49 (65%), Positives = 39/49 (80%)	150 bp	*rpmG*
rpmGB	CD0058A	Identities = 25/49 (51%), Positives = 32/49 (65%)	150 bp	*rpmG*
rpmH	CD3680	Identities = 27/42 (64%), Positives = 31/42 (74%)	138 bp	
rpmI	CD0686	Identities = 40/62 (65%), Positives = 47/62 (76%)	195 bp	
rpmJ	CD0094A	Identities = 36/37 (97%), Positives = 37/37 (100%)	114 bp	
rpsB	CD2140	Identities = 168/236 (71%), Positives = 205/236 (87%)	714 bp	
rpsC	CD0079	Identities = 123/212 (58%), Positives = 168/212 (79%)	816 bp	
rpsD	CD0097	Identities = 102/208 (49%), Positives = 139/208 (67%)	624 bp	
rpsE	CD0088	Identities = 122/168 (73%), Positives = 142/168 (85%)	510 bp	
rpsF	CD3663	Identities = 37/91 (41%), Positives = 62/91 (68%)	279 bp	
rpsG	CD0069	Identities = 117/156 (75%), Positives = 138/156 (88%)	471 bp	
rpsH	CD0085	Identities = 90/132 (68%), Positives = 114/132 (86%)	199 bp	
rpsI	CD0105	Identities = 90/130 (69%), Positives = 101/130 (78%)	393 bp	
rpsJ	CD0072	Identities = 83/103 (81%), Positives = 95/103 (92%)	312 bp	

rpsK	CD0096	Identities = 102/132 (77%), Positives = 118/132 (89%)	399 bp	
rpsL	CD0068	Identities = 103/136 (76%), Positives = 116/136 (85%)	423 bp	
rpsM	CD0095	Identities = 75/122 (61%), Positives = 101/122 (83%)	372 bp	
rpsN	CD0084A	Identities = 26/43 (60%), Positives = 34/43 (79%)	186 bp	
rpsO	CD1316	Identities = 57/82 (70%), Positives = 67/82 (82%)	258 bp	
rpsP	CD1253	Identities = 56/90 (62%), Positives = 74/90 (82%)	273 bp	
rpsQ	CD0081	Identities = 59/80 (74%), Positives = 63/80 (79%)	255 bp	
rpsR	CD3661A	Identities = 37/61 (61%), Positives = 48/61 (79%)	228 bp	
rpsT	CD2473	Identities = 35/88 (40%), Positives = 51/88 (58%)	267 bp	
rpsU	CD2446a	Identities = 34/55 (62%), Positives = 45/55 (82%)	180 bp	
alaS	CD1282	Identities = 418/880 (47%), Positives = 593/880 (67%)	2640 bp	
argS	CD0711	Identities = 170/602 (28%), Positives = 285/602 (47%)	1701 bp	
asnS	CD2245	Identities = 161/455 (35%), Positives = 245/455 (54%)	1398 bp	*asnC*
aspS	CD2739	Identities = 330/590 (56%), Positives = 434/590 (74%)	1788 bp	
cysS	CD0052	Identities = 258/463 (55%), Positives = 343/463 (74%)	1407 bp	
gltX	CD0051	Identities = 218/485 (44%), Positives = 324/485 (66%)	1482 bp	
glyQ	CD2433	Identities = 201/288 (70%), Positives = 245/288 (85%)	879 bp	
glyS	CD2432	Identities = 293/694 (42%), Positives = 442/694 (64%)	2067 bp	
hisS	CD2740	Identities = 195/415 (47%), Positives = 283/415 (68%)	1263 bp	
ileS	CD2618	Identities = 265/822 (32%), Positives = 418/822 (50%)	3108 bp	
leuS	CD2521	Identities = 388/821 (47%), Positives = 537/821 (65%)	2421 bp	
lysS	CD3552	Identities = 287/497 (58%), Positives = 384/497 (77%)	1530 bp	
metS	CD3540	Identities = 354/660 (53%), Positives = 465/660 (70%)	1938 bp	
pheS	CD0699	Identities = 195/341 (57%), Positives = 259/341 (76%)	1029 bp	
pheT	CD0700	Identities = 319/813 (39%), Positives = 479/813 (59%)	2394 bp	
proS	CD0049	Identities = 279/547 (51%), Positives = 380/547 (69%)	1716 bp	
serS	CD0014	Identities = 258/418 (61%), Positives = 322/418 (77%)	1272 bp	
trpS	**CD2610**	**Identities = 181/330 (54%), Positives = 242/330 (73%)**	**1002 bp**	
tyrS	CD1521	Identities = 107/386 (27%), Positives = 194/386 (50%)	1209 bp	
valS	CD3256	Identities = 467/891 (52%), Positives = 630/891 (70%)	2667 bp	
gatA	CD3117	Identities = 35/152 (23%), Positives = 57/152 (38%)	861 bp	*bglG*
gatB	CD1041	Identities = 17/61 (28%), Positives = 27/61 (44%)	3828 bp	*aadA LamB/YcsF*
gatC	CD1384	Identities = 15/43 (35%), Positives = 27/43 (63%)	792 bp	*family protein*
fmt	CD2584	Identities = 152/311 (48%), Positives = 216/311 (69%)	930 bp	
frr	CD2137	Identities = 101/185 (54%), Positives = 144/185 (77%)	558 bp	
fusA	CD0070	Identities = 486/685 (71%), Positives = 568/685 (83%)	2067 bp	
infA	CD0094	Identities = 56/72 (78%), Positives = 64/72 (89%)	219 bp	
infB	CD1309	Identities = 360/584 (62%), Positives = 452/584 (77%)	1941 bp	
prfA	CD3484	Identities = 205/351 (58%), Positives = 278/351 (79%)	1065 bp	
prfB	CD0144	Identities = 179/339 (52%), Positives = 251/339 (74%)	987 bp	
tsf	CD2139	Identities = 162/302 (54%), Positives = 208/302 (69%)	912 bp	
tufA	CD0058	Identities = 309/397 (78%), Positives = 340/397 (86%)	1194 bp	
spoVC	NF			
groEL	CD0194	Identities = 388/526 (74%), Positives = 453/526 (86%)	1629 bp	
groES	CD0193	Identities = 55/92 (60%), Positives = 74/92 (80%)	285 bp	

map	CD0092	Identities = 162/248 (65%), Positives = 201/248 (81%)	747 bp	*map1*
ffh	CD1252	Identities = 287/450 (64%), Positives = 358/450 (80%)	1353 bp	
ftsY	CD1251	Identities = 158/303 (52%), Positives = 226/303 (75%)	1281 bp	
prsA	CD2263	Identities = 87/228 (38%), Positives = 128/228 (56%)	1143 bp	annotated wrong, *prsA* CD3500 Also highly similar to
secA	CD2792	Identities = 394/784 (50%), Positives = 558/784 (71%)	2346 bp	CD0143, *secA1*
secE	CD0059	Identities = 18/54 (33%), Positives = 33/54 (61%)	222 bp	
secY	CD0090	Identities = 181/427 (42%), Positives = 265/427 (62%)	1269 bp	*prlA*
accA	CD1936	Identities = 164/302 (54%), Positives = 225/302 (74%	945 bp	
accB	CD1939	Identities = 62/156 (40%), Positives = 93/156 (60%)	453 bp	
accC	CD1938	Identities = 245/443 (55%), Positives = 323/443 (73%)	1362 bp	
accD	CD1937	Identities = 139/276 (50%), Positives = 186/276 (67%)	855 bp	
acpA	CD1062	Identities = 36/74 (49%), Positives = 51/74 (69%)	228 bp	
acpS	CD3466	Identities = 50/122 (40%), Positives = 74/122 (60%)	381 bp	
birA	CD3558	Identities = 125/323 (39%), Positives = 200/323 (62%)		
fabD	CD1181	Identities = 142/290 (49%), Positives = 197/290 (68%)	951 bp	
fabF	CD1184	Identities = 258/409 (63%), Positives = 327/409 (80%)	1239 bp	
fabG	CD1182	Identities = 142/245 (58%), Positives = 188/245 (77%)	750 bp	
cdsA	CD2135	Identities = 96/273 (35%), Positives = 142/273 (52%)	780 bp	
gpsA	CD2630	Identities = 177/337 (52%), Positives = 242/337 (71%)	1032 bp	*glyC*
pgsA	CD0173	Identities = 78/250 (31%), Positives = 120/250 (48%)	1068 bp	
yhdO	NF			
yerQ	CD2060	Identities = 89/282 (31%), Positives = 158/282 (56%)	900 bp	
plsX	CD1178	Identities = 145/313 (46%), Positives = 220/313 (70%)	1023 bp	
gcaD	CD3515	Identities = 234/451 (52%), Positives = 330/451 (73%)	1380 bp	*glmU*
glmS	CD0120	Identities = 267/613 (43%), Positives = 413/613 (67%)	1833 bp	
ybbT	NF			
yvyH	NF			
asd	CD3224	Identities = 153/331 (46%), Positives = 216/331 (65%)	1002 bp	
dapA	CD3223	Identities = 125/277 (45%), Positives = 186/277 (67%)	882 bp	*dapA1*
dapB	CD3229	Identities = 78/270 (29%), Positives = 136/270 (50%)	756 bp	*dapB2*
dapF	CD2590	Identities = 91/277 (32%), Positives = 158/277 (57%)	831 bp	
ykuQ	NF			
ykuR	NF			
alr	NF			
ddl	CD1626	Identities = 134/360 (37%), Positives = 199/360 (55%)	1101 bp	*vanG*
racE	CD3562	Identities = 114/268 (42%), Positives = 174/268 (64%)	807 bp	*murI*
mraY	CD2654	Identities = 144/311 (46%), Positives = 215/311 (69%)	969 bp	
murAA	CD0123	Identities = 250/419 (59%), Positives = 331/419 (78%)	1272 bp	*murA*
murB	CD3402	Identities = 111/282 (39%), Positives = 177/282 (63%)	915 bp	
murC	CD3518	Identities = 147/441 (33%), Positives = 226/441 (51%)	1353 bp	
murD	CD2653	Identities = 170/454 (37%), Positives = 273/454 (60%)	1356 bp	
murE	CD2664	Identities = 206/471 (44%), Positives = 293/471 (62%)	1455 bp	
murF	CD2655	Identities = 169/456 (37%), Positives = 257/456 (56%)	1374 bp	
murG	CD2651	Identities = 141/359 (39%), Positives = 224/359 (62%)	1230 bp	

tagA	NF			
tagB	NF			
tagD	NF			
tagF	NF			
tagG	NF			
tagH	NF			
tagO	NF			
divIB	NF			
ftsA	NF			
ftsL/divIC	NF			
ftsW	CD1152	Identities = 113/350 (32%), Positives = 183/350 (52%)	1131 bp	*mrdB*
ftsZ	CD2646	Identities = 230/387 (59%), Positives = 293/387 (75%)	1161 bp	
pbpB	CD2656	Identities = 217/676 (32%), Positives = 335/676 (49%)	1980 bp	*spoVD*
rodA	CD2652	Identities = 103/285 (36%), Positives = 151/285 (52%)	1131 bp	*spoVE*
mreB	CD1145	Identities = 209/332 (62%), Positives = 275/332 (82%)	1062 bp	*mreB2*
mreC	CD1146	Identities = 69/246 (28%), Positives = 110/246 (45%)	900 bp	
eno	CD3170	Identities = 322/427 (75%), Positives = 364/427 (85%)	1293 bp	
fbaA	CD3448	Identities = 124/285 (43%), Positives = 178/285 (62%)	852 bp	*gatY*
pfkA	CD3395	Identities = 191/319 (60%), Positives = 251/319 (79%)	960 bp	
pgm	CD3171	Identities = 286/516 (55%), Positives = 376/516 (72%)	1153 bp	*gpmL*
prsA	CD2263	Identities = 87/228 (38%), Positives = 128/228 (56%)	1143 bp	
tkt	CD3459	Identities = 93/253 (36%), Positives = 139/253 (54%)	828 bp	
tpiA	CD3172	Identities = 150/247 (60%), Positives = 193/247 (78%)	744 bp	
dxr	CD2130	Identities = 203/383 (53%), Positives = 281/383 (73%)	1155 bp	
dxs	CD1207	Identities = 306/623 (49%), Positives = 417/623 (66%)	1866 bp	
ispE	CD3566	Identities = 124/254 (49%), Positives = 175/254 (69%)	891 bp	
yacM	NF			
yacN	NF			
yqfY	NF			
yqiD	NF			
hepS	NF			
hepT	CD1205	Identities = 70/226 (30%), Positives = 114/226 (50%)	888 bp	*ispA*
menA	NF			
menB	CD0800	Identities = 89/256 (34%), Positives = 137/256 (53%)	783 bp	*crt1*
menC	CD1389	Identities = 101/361 (27%), Positives = 179/361 (49%)	1068 bp	
menD	NF			
menE	CD2852	Identities = 123/510 (24%), Positives = 221/510 (43%)	1515 bp	*dltB*
menH	NF			
resA	CD3010	Identities = 49/113 (43%), Positives = 65/113 (57%)	bp	
resB	NF			
resC	NF			
trxA	CD1690	Identities = 44/102 (43%), Positives = 70/102 (68%)	318 bp	*trxA1* multiple present
trxB	CD2356	Identities = 144/312 (46%), Positives = 207/312 (66%)	948 bp	*trxB3* multiple present
yumC	NF			*trxB*
adk	CD0091	Identities = 123/216 (56%), Positives = 162/216 (75%)	651 bp	

gmk	CD2588	Identities = 97/199 (48%), Positives = 141/199 (70%)	618 bp	
guaB	CD2335	Identities = 170/493 (34%), Positives = 250/493 (50%)	1500 bp	
hprT	CD2691	Identities = 93/165 (56%), Positives = 130/165 (78%)	528 bp	*hpt*
nrdE	CD2995	Identities = 357/697 (51%), Positives = 492/697 (70%)	2100 bp	
nrdF	CD2994	Identities = 173/318 (54%), Positives = 230/318 (72%)	972 bp	
ymaA	NF			
cmk	CD1816	Identities = 102/204 (50%), Positives = 148/204 (72%)	651 bp	
pryG	NF			
tmk	NF			
tyaG	NF			
dfrA	NF			
folD	CD0720	Identities = 121/276 (43%), Positives = 180/276 (65%)	873 bp	
gylA	CD2726	Identities = 262/414 (63%), Positives = 313/414 (75%)	1245 bp	
nadE	CD0794	Identities = 95/256 (37%), Positives = 141/256 (55%)	756 bp	
ppnK	CD1049	Identities = 70/270 (25%), Positives = 138/270 (51%)	801 bp	
yqeJ	NF			
yueK	CD3543	Identities = 232/473 (49%), Positives = 329/473 (69%)	1431 bp	
metK	**CD0130**	**Identities = 294/396 (74%), Positives = 348/396 (87%)**	**1194 bp**	
csd	NF			
yurU	NF			
yurV	NF			
yurX	NF			
yurY	NF			
yrvO	NF			
mrpA	NF			
mrpB	NF			
mrpC	NF			
mrpD	NF			
mrpF	NF			
ppaC	CD0333	Identities = 108/237 (45%), Positives = 154/237 (64%)	1599 bp	
era	CD2437	Identities = 165/295 (55%), Positives = 48/295 (74%)	894 bp	
obg	CD1164	Identities = 219/429 (51%), Positives = 310/429 (72%)	1278 bp	
ylqF	NF			
yphC	NF			
yqeH	NF			
ysxC	NF			
gcp	CD0152	Identities = 183/336 (54%), Positives = 251/336 (74%)	1017 bp	
odhB	CD0038	Identities = 101/306 (33%), Positives = 157/306 (51%)	1047 bp	*acoC*
pdhA	CD0036	Identities = 100/324 (30%), Positives = 164/324 (50%)	969 bp	*acoA*
ydiC	CD0149	Identities = 73/234 (31%), Positives = 127/234 (54%)	450 bp	
ykqC	NF			
yneS	CD2631	Identities = 80/183 (43%), Positives = 110/183 (60%)	615 bp	
ymdA	CD1329	Identities = 327/498 (65%), Positives = 416/498 (83%)	1563 bp	
yloQ	CD2576	Identities = 140/301 (46%), Positives = 189/301 (62%)	885 bp	
yqjK	CD2539	Identities = 90/308 (29%), Positives = 149/308 (48%)	918 bp	*elaC*
ywlC	CD3482	Identities = 164/333 (49%), Positives = 227/333 (68%)	1062 bp	

yacA	NF		
ydiB	CD0149	Identities = 69/145 (47%), Positives = 92/145 (63%)	453 bp
ylaN	NF		
yqeI	CD1165	Identities = 38/96 (39%), Positives = 59/96 (61%)	369 bp