Implementation and optimisation of alternative therapeutics for use in *Clostridium sporogenes* as a delivery vehicle

Patrick Gerard Budd, MRes

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

Sep 2015

Abstract

Clostridium sporogenes is part of a highly diverse group of Gram positive, spore forming, anaerobic bacteria. *C. sporogenes* can be used as a delivery vehicle for chemotherapeutics in cancer treatment due to the inactive spore form of *C. sporogenes* only germinating in the microenvironment of the hypoxic tumour. Cancer, despite large investment in treatment and diagnosis, still remains one of the leading causes of death in the world. As such, improvements on current treatments are necessary to improve patient prognosis. Utilising *C. sporogenes* could be a cheap and effective way to do this by utilising their germination properties to deliver anti-cancer therapeutics directly to the hypoxic regions of solid tumours. Through introducing Prodrug Converting Enzymes (PCEs) into *C. sporogenes* when the spores germinate in the tumour the production of the PCE will result in the breakdown of a Prodrug into a toxic product resulting in an anti-cancer effect. This system is known as Clostridial Directed Enzyme Prodrug Therapy (CDEPT). Previous iterations of this system incorporated a nitroreductase gene, used for the breakdown of the prodrug CB1954. During this project alternative prodrugs were investigated, in this case carboxypeptidase G_2. Alternatives into the prodrug and enzyme system are also being investigated in a direct action therapy in the form of the monoclonal anti-VEGF. The aims of this project were to implement the genes and optimise the activity of the drugs if necessary. It was also necessary to confirm that the bacterium was a non-pathogenic group 1 bacterium through sequencing and annotation of the genome.

Acknowledgements

I would like to thank my supervisors Dr Sarah Kuehne, and Professor Nigel Minton, and Dr Mark Collery for all their advice and support throughout the course of the PhD. I would also like to thank the rest of the CRG for the help, in particular, Dr Aleksandra Bailey, who has no only collaborated with me for parts, but offered me advice on *C. sporogenes*.

I would like to say a special thanks to my family, my Mum, my Dad and my two sisters Sinead and Maureen who have been an amazing source of support for my PhD. They have always been available for a chat when I needed it.

Finally I would like to thank my wife, Christine. Without her I would never have had the confidence in my ability to undertake a PhD, and without her I would never have had perseverance to get this far. She has been a constant source of patience, and support.

Table of Contents

Abstract 2
Acknowledgements 3
Table of Contents 4
List of Figures 8
List of Tables 9
Abbreviations 10
Units 13
1 Introduction 15
1.1 Cancer Incidence and Prognosis 15
1.2 Cancer Tumourigenesis 16
1.2.1 Apoptotic Function 19
1.2.2 Proliferative signal 20
1.2.3 Growth Suppressors 20
1.2.4 Vascular growth 21
1.2.5 Replicative Immortality 23
1.3 Solid tumour Physiology 24
1.4 Hypoxia and treatment implications. 26
1.5 Alternative Therapies 29
1.5.1 Nanotherapeutics 29
1.5.2 Antibody therapy 30
1.5.3 Antibody Directed Prodrug Converting Enzymes 31
1.5.4 Gene Directed Prodrug Therapy 34
1.5.5 Virus Directed Prodrug Therapy 35
1.6 *Clostridium sporogenes* 37
1.6.1 A Historical Perspective 40
1.6.2 Current Work 44
1.7 Sporulation 48
1.8 Prodrug Systems 53
1.9 Aims 58
2 Materials and Methods 61
2.1 Chemicals and suppliers 61
2.2 Bacterial strains 62
2.3 Plasmids 63

2.4 Primers 72
2.5 Microbial Methods 78
2.5.1 Media 78
2.5.2 Supplements 80
2.5.3 General culture and Storage conditions 80
2.5.4 Growth Curves 81
2.5.5 Electrocompetant cells 81
2.5.6 Transformation of *E. coli* 82
2.5.7 Electrocompetant *C. acetobutylicum* 82
2.5.8 Preparation of methylated plasmid for *C. acetobutylicum* transformation 83
2.5.9 Transformation of *C. acetobutylicum* 83
2.5.10 Electrocompetant *C. sporogenes.* 84
2.5.11 Transformation of *C. sporogenes* 84
2.5.12 Conjugation of *C. sporogenes* 85
2.5.13 Fractionation of *C. sporogenes* 85
2.6 Molecular Biology techniques 86
2.6.1 Genomic DNA Extraction 86
2.6.2 Plasmid Extraction 89
2.6.3 Amplification by PCR. 91
2.6.4 Restriction Endonuclease Digest 94
2.6.5 Post Digest alteration of DNA 94
2.6.6 Agarose Gel Electrophoresis 94
2.6.7 DNA Gel Extraction 95
2.6.8 Purification of PCR Products. 96
2.6.9 Ligation and Dialysis. 98
2.6.10 Gibson Assembly. 98
2.6.11 Nucleotide Sequencing 99
2.6.12 Oligomelts 99
2.7 RNA 99
2.7.1 RNA Isolation 99
2.7.2 RNA purification 100
2.7.3 Reverse Transcriptase PCR 101
2.8 Protein analysis 101

2.8.1 Preparation of cell lysates 101
2.8.2 Protein precipitation from fluids 102
2.8.3 Preparation of supernatant 102
2.8.4 SDS-PAGE 103
2.8.5 2D SDS-PAGE 103
2.8.6 SDS-PAGE gel staining. 104
2.8.7 Western Blots. 105
2.8.8 His-Tag Purification. 106
2.8.9 Carboxypeptidase G_2 Plate Assay 107
2.8.10 Methotrexate (MTX) enzymatic assay 107
2.9 Statistics 108
3 *C. sporogenes* NCIMB 10696 Genome 110
3.1 Introduction 110
3.2 Results 115
3.2.1 Sequencing 116
3.2.2 Restriction Mapping 120
3.2.3 Annotation 122
3.2.4 Phylogenetic Dendrogram 125
3.2.5 Toxin Identification 127
3.3 Discussion 129
4 Implementation of CPG_2 and Anti-VEGF. 135
4.1 Introduction 135
4.2 Results 140
4.2.1 CPG_2 Cloning into Standardised Plasmid Vectors. 142
4.2.2 Expression of Targets CPG_2 and anti-VEGF 145
4.3 Variovorax paradoxus 153
4.4 Expression of anti-VEGF 155
4.4.1 Anti-VEGF Cloning 158
4.4.2 Expression Part 2 159
4.5 Discussion 164
4.5.1 CPG_2 164
4.5.2 Anti-VEGF 167
5 Optimisation of the CDEPT System. 172
5.1.1 Introduction 172

5.2 Results 178
5.2.1 Identification of Alternative Signal Peptides. 178
5.2.2 Cell Wall Anchoring of Proteins Using the Sortase System 187
5.2.3 The Sortase System in *C. sporogenes* 193
5.3 Conclusions 198
6 Final Conclusions 204
6.1 Aims 204
6.2 Genome 204
6.3 Implementing Anti-VEGF 205
6.4 Implementing Carboxypeptidase G_2 207
6.5 Optimization of Carboxypeptidase G_2 207
6.6 Sortase Research 210
6.7 Carboxypeptidase G_2 in the Context of CDEPT 213
7 Bibliography 218
8 Appendices 232

List of Figures

Figure 1-1 Statistics of cancer. 15
Figure 1-2 The regulatory pathways that are involved in cell homeostasis. 18
Figure 1-3 A tumours effect on vascularisation. 23
Figure 1-4 The physiology of a solid tumour. 26
Figure 1-5 The chemical breakdown of the CPG_2 prodrug 32
Figure 1-6. A scanning electron micrograph of C. sporogenes. 40
Figure 1-7. A schematic representation of the mode of action of CDEPT. 46
Figure 1-8 A step by step process of sporulation. 49
Figure 1-9 Different representations of bacillus spores, which are homologous to clostridial spores. 50
Figure 1-10 Structural diagrams of CPG_2. 54
Figure 1-11 A structural diagram of Anti-VEGF. 56
Figure 3-1 The initial scaffolds orientation, and mapping, and a table of what constitutes the scaffolds. 118
Figure 3-2 The OpGen MapSolver restriction digest maps of C. sporogenes 121
Figure 3-3 The visualization of RATT. 123
Figure 3-4 A circular genome map of C. sporogenes. 125
Figure 3-5 A partial Clostridium dendrogram. 126
Figure 3-6 The toxin type PCR. 129
Figure 4-1 The plasmid maps of the initial storage, and expression vector 142
Figure 4-2 A plasmid map of the inducible expression plasmid. 145
Figure 4-3 Screening plates for CPG_2. 146
Figure 4-4 the same plasmid from fig 2 conjugated into C. sporogenes. 147
Figure 4-5 The enzymatic assay of the CPG_2 producing strains. 148
Figure 4-6 Comparison of CPG_2 versions 150
Figure 4-7 A stained SDS-PAGE of C. sporogenes supernatant 151
Figure 4-8 A comparison of different enzymatic activity in the supernatant 152
Figure 4-9 A) WT Variovorax paradoxus CPG_2 implemented in the pMTL82251:YZ2 plasmid in E. coli. 154
Figure 4-10 Western Blots of the different strains containing anti-VEGF genes. 156
Figure 4-11 Growth curves of strains containing anti-VEGF. 157
Figure 4-12 Western Blots to confirm anti-VEGF production in the new plasmids, through the FLAG-tag 160
Figure 4-13 Western Blot of anti-VEGF in C. sporogenes. 162
Figure 5-1 The metalloprotease locus. 181
Figure 5-2 2d SDS-PAGE of C. sporogenes supernatant. 182
Figure 5-3 MALDI mass spectrometry of the supernatant of Wild type Clostridium sporogenes. 183
Figure 5-4 Units of activity of CPG_2 in different cellular fractions. 186
Figure 5-5 The enzymatic activity of CPG_2 in different cellular fractions. 190
Figure 5-6 A comparison of the different cell wall fractions. 192
Figure 5-7 His-Tag proteins from IMAC chromatography purification. 194
Figure 5-8 Increased concentration of Imidazole on the elution of proteins. 195
Figure 5-9 The RT-PCR gel pictures. 197

List of Tables

Table 2-1 A list of Suppliers and the reagents, kits and equipment they supplied 61
Table 2-2 A Table of the different bacterial strains that have been used 62
Table 2-3 All plasmids that were used in this study, and their reference 63
Table 2-4 The different Oligos which were annealed together for use as signal sequences in cloning 72
Table 2-5 A list of the primers used in this study 74
Table 2-6 A list of Media and Their Composition 79
Table 2-7 Supplements and their concentrations 80
Table 2-8 The solutions needed for 2d SDS-PAGE 104
Table 3-1 Toxin Gene Presence Table 128
Table 4-1 A brief overview of all the plasmids that were created for the expression of the different target genes and the special features of each plasmid, and gene. 141
Table 5-1 The table of the different signal sequences to be implemented 179
Table 5-2 The sortase motifs. 188

Abbreviations

2xYT	- Two-Times Yeast Tryptone
5-FC	- 5-Fluorocytosine
5-FU	- 5-fluorouracil
ACT	- Artemis Comparison Tool
ADCC	- Antibody dependent cellular cytotoxicity
ADEPT	- Antibody Directed Enzyme Prodrug Therapy
BLAST	- Basic Local Alignment Search Tool
BoNT	- Botulinum Neuro Toxin
CDC	- Complement Dependent Cytotoxicity
CDEPT	- Clostridial Directed Enzyme Prodrug Therapy
CDS	- Coding DNA Sequences
CMDA	- 4-[(2-chloroethyl)(2-mesyloxyethyl)amino]benzoyl-L-glutamic acid
COBALT	- Combined Bacteriolytic Therapy
CPG_2	- Carboxypeptidase G_2
CRG	- Clostridia Research Group
DEPC Water	- Diethylpyrocarbonate treated water
DEPT	- Directed Enzyme Prodrug Therapy
DNA	- Deoxyribonucleic acid
DOC/TCA	- Na-Deoxycholate Trichloracetic Acid
DTT	- Dithiothreitol
EDTA	- Ethylenediaminetetraacetic Acid
EPB	- Electroporation Buffer
EPC	- Endothelial Progenitor Cells
EPR	- Enhanced Permeability and Retention
EtBr	- Ethidium Bromide
EtOH	- Ethanol
GDEPT	- Gene Directed Enzyme Pro-drug Therapy
HIF-1	- Hypoxia Inducible Factor-1
IFP	- Interstitial Fluid Pressure

IPTG	- Isopropyl β-D-1-thiogalactopyranoside
LB	- Luria-Bertani
mAbs	-Monoclonal Antibodies
MACC	- Amino Acid Complete Medium
MALDI	- Matrix-Assisted Laser Desorption/Ionization
MOPS	- 3-(N-morpholino)propansulfic acid
mRNA	- messenger Ribonucleic Acid
MRSA	- Methicillin Resistant *Staphylococcus aureus*
MTX	- Methotrexate
MWCO	- Molecular Weight Cut Off
NaAc	- Sodium Acetate
NCIMB	- National Collection of Industrial Food and Marine Bacteria
OD	- Optical Density
ORF	- Open Reading Frame
PBS	- Phosphate Buffered Saline
PCE	- Prodrug Converting Enzymes
PES	- Polyethersulfone
PCR	- Polymerase Chain Reaction
PIP	- Post-translationally Interacting Protein
qRT-PCR	- Quantitative Reverse Transcriptase Polymerase Chain Reaction
RATT	- Rapid Annotation Transfer Tool
RNA	- Ribonucleic Acid
SDS	- Sodium Dodecyl Sulphate
SDS-PAGE	- Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis
SOC	- Super Optimal broth with Catabolite repression
SOE	- Splice Overlap Extension
SRP	- Signal Recognition Particle
TAE	- Tris Base, Acetic Acid and EDTA buffer
TAT	- Twin-Arginine Translocation
TBS	- Tris Buffered Saline
TCA	- Trichloracetic Acid

T_m	- Melting Temperature
TMB	- 3,3’,5,5’-Tetramethylbenzidine
tRNA	- Transfer Ribonucleic Acid
TYG	- Tryptone Yeast extract Sodium Thioglicolate medium
UV	- Ultraviolet
VDEPT	- Virus Directed Enzyme Prodrug Therapy
VEGF	- Vascular Endothelial Growth Factor
WHO	- World Health Organisation
ZMW	- Zero-Mode Waveguide

Units

^{O}C	- Degrees Celsius
Kb	- Kilobase Pairs
Bp	- Base Pairs
kDa	- Kilo Daltons
µg	- microgram
mg	- milligram
g	- gram
sec	- Seconds
min	- minutes
h	- Hours
µl	- microlitre
ml	- millilitre
l	- litre
µM	- Micromolar
mM	- Millimolar
M	- Molar
O/N	- Over Night
W/V	- Weight per volume
V	-Volt
kV	- Kilovolt
cm	- Centimeter
µF	- Microfarads
Ω	- Ohms Resistance
U per ml^{-1}	-Units activity per ml
g	- centrifugal force

Introduction

Chapter 1

1 Introduction

1.1 Cancer Incidence and Prognosis

Cancer is a major clinical issue across the world. In 2012 there were a reported 14 million new cases worldwide, of which 3.44 million cases where in Europe. Of the 14 million reported, there were 8.2 million cancer related deaths. The incidence and death rates of cancer around the world in 2012 are shown in Fig 1-1. According to a 2014 world health organisation report the incidence rate will rise by 70% over the next two decades. Numerous factors can increase the risk of cancer, of which some of the most significant are a high body mass index, low fruit and vegetable intake, tobacco and alcohol abuse, and, finally, lack of physical activity. With increased public awareness of these issues preventative lifestyles will hopefully reduce the incidence rate (WHO, 2014, WHO, 2015).

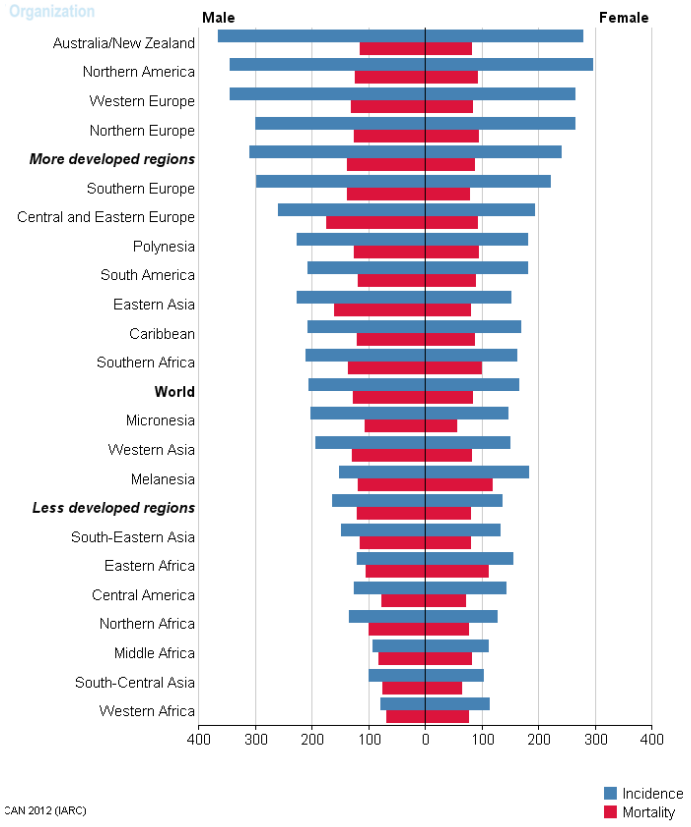


Figure 1-1 Statistics of cancer.

A summation of the incidence and mortality rates of cancer in men and women for the year 2012 (Ferlay J, 2013) Open access for non-commercial replication.

There are many different cancer types, and the incidence rates of different cancers are influenced by different factors. For example external factors such as smoking increases the risk of lung cancer. These factors can also be inherent where incidence rates can be increased by a genetic element, when people have certain mutations in the *BRCA1* gene this makes them more susceptible to breast cancer (Brose et al., 2002, Leongamornlert et al., 2012). The gene *BRCA1* is known to be a tumour suppressor gene, which when functional and active prevents the formation of tumours in specific tissues (such as breast and ovarian tissues) (Zhu et al., 2011, Miki et al., 1994). Alternatively there are mutations in genes (known as proto-oncogenes, when acting normally) which would cause the formation of tumours, such as the *Ras* oncogenes (Rajalingam et al., 2007)

Once diagnosed with cancer, prognosis in patients also be very varied dependent on different factors. The type of cancer you have is a big factor as some cancers can be difficult to detect in early stages (such as pancreatic (Von Hoff et al., 2013)), or are very difficult to treat (such as neuroblastoma (Sung et al., 2013)). Prognosis is also dependent on the range of treatments that are suitable given the cancer type, and patient's background. An early diagnosis for a patient is key to a better prognosis for the patient, and hence early detection is a key focus for healthcare systems. Once diagnosed to ensure that a patient's treatment is optimal a range of drugs need to be available to doctors. As techniques for detection improve, when combined with better treatment options and variety, prognosis and long term survival rates will be improved.

1.2 Cancer Tumourigenesis

Cancer is caused through numerous mutations accumulating in key genes which have specific important roles within a tissue. Mutations occur in the human genome through numerous external factors (e.g. UV light) and through DNA replication and repair errors,

but often in non-coding areas. These mutations are not the only formative factor in tumours as epigenetic factors also play a role in expression of a phenotype.

For the initiation of cancer, mutations occur in one or more key regulatory genes causing amino acid changes and consequently loss of function. When key cellular functions are dependent on the genes which have lost function, it may cause the cells to lose the ability to control cell replication. This finally results in abnormal tissue growth. Each step is a complicated multifactorial event, but will often result in a lump of undifferentiated, rapidly replicating cells (Sherr, 1996). The different regulatory pathways that are involved in the control of cell replication and tumourigenesis are well described, but our understanding of the molecular mechanisms that are involved in each pathway is basic (Sherr, 2000). The development of new treatments for cancer will be improved the more we understand about the regulatory networks involved in tumourigenesis.

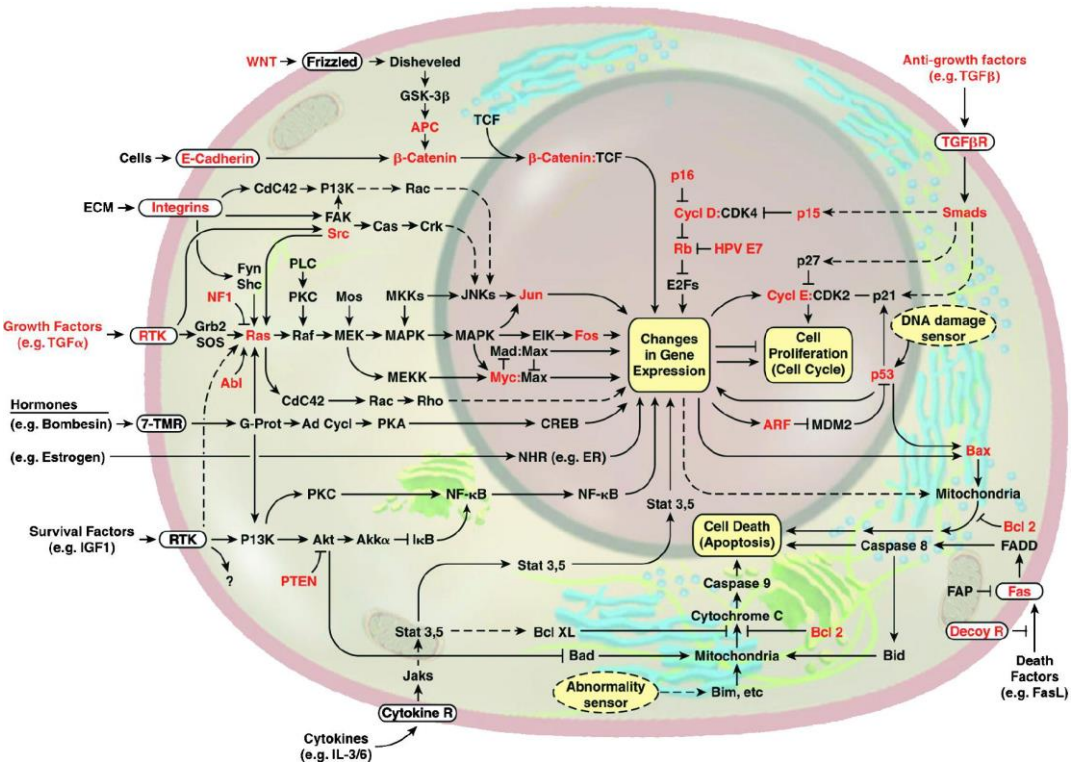


Figure 1-2 The regulatory pathways that are involved in cell homeostasis.

The diagram shows the number of key cell homeostatic genes that when deregulated become oncogenes. All of these genes when incorrect can cause the cell to lose proliferative control (Hanahan and Weinberg, 2000) Permissions for reproduction granted see appendices.

Without proper function the cells may only replicate forming a bundle of cells which constitute a tumour. However this is easily dealt with if the tumour is benign, as it will be self-contained and unable to invade surrounding tissue (McConnell et al., 1998). When it can invade the neighbouring tissue or metastasize to other tissues, it is considered to be malignant and is much more difficult to treat. If the tumour is malignant then it is considered to be cancer (Hamilton et al., 2000).

The regulatory pathways that have been described as being important to the controlled replication of the cell are all illustrated by Hanahan and co-workers (Hanahan and Weinberg, 2011). They described these hallmarks as loss of apoptotic function for programmed cell death of cancer cells, continued vasculature growth for sustained supply of nutrients enabling the cells to function and for cancer tissue to grow, and continued

cell proliferation signal, as well as loss of sensitivity to cell growth repressors so that tissue formation can be halted when necessary (Figure 1-2 shows some of the complex networks involved in normal cell function which can go wrong causing cancer). These different factors all work together, and each factor is controlled by large networks of genes acting in concert to maintain healthy tissue.

There are key genes for each of these factors which when normal function is lost, can cause cancerous tumours to occur. These genes are considered to be either oncogenes or tumour suppressors. Many of these genes have been identified in the pathways described as “hallmarks of cancer”, and often loss of function in these is enough to cause tumourigenesis. The control of expression of these genes, post transcriptional issues, as well as epigenetic factors can all contribute to tumourigenesis.

1.2.1 Apoptotic Function

Apoptotic function is a vital part of the network controlling cell replication, which when activated can, through various signal cascades, instruct cells to die. Induction of this cascade allows the body to remove cells which are faulty or damaged from tissue which keeps the tissue functioning efficiently. With the apoptotic cascade induced the malfunctioning cells will be removed and be replaced by a healthy cell. If function of the apoptotic cascades is lost through mutation, then tissues can end up with cell lines which will not be viable and which will be persistent within the tissue, and therefore difficult to control. The caspase genes are part of the apoptotic pathway and have a key role in both the initiation of, and the execution of apoptotic cascades. Caspases exist within the cytoplasm in zymogen form, when the initiators, such as caspase-8 (Mocarski et al., 2012) and caspase-9 (Brentnall et al., 2013), are activated by stimuli causing oligomerisation, the oligomerised caspase protein in turn initiates the apoptotic cascade. As caspases are part of the cysteine proteases family activation of the zymogens happens through proteolysis.

The caspase act in an activation relay finally activating the caspase executors. When the executors, such as Caspase-3 and Caspase-7 (Brentnall et al., 2013, Mehmet, 2000), are activated they start cleaving hundreds of structurally and functionally important proteins, leading to cell death. If any of the initiating or executing caspases are no longer active then the cascade will fail in the removal of proteins, allowing the cell to survive (McIlwain et al., 2013) These cell lines are a problem, as DNA damage accumulates within the cell line over time, without any programmed cell death to remove these mutations from the tissue. If more mutations occur in key homeostatic genes then tumours could form.

1.2.2 Proliferative signal

Cancer can be caused through an uncontrolled, sustained cell replication signal. Proliferation is normally a controlled process which is switched on or off as needed with proteins to stop and start proliferative signals. Through regulation of the proliferative signal proteins the tissue homeostasis can be maintained. There are several ways in which proliferative signal could be maintained consistently, however the understanding of this system is poor. The mutations may cause the cell to be able regulate the activation of replicative signals leading to the replication process to always be turned on in some cell lines. This would cause uncontrolled replication that would quickly cause a net production of cells, leading to tumours being formed (Pylayeva-Gupta et al., 2011, Feitelson et al., 2015).

1.2.3 Growth Suppressors

The cell replication system is not solely dependent on the presence or absence of the proliferative signal. There are proliferative signals which can have a positive effect, like an accelerator (oncogenes), but there are also growth suppressors which have a negative effect like a brake (tumour suppressors). In cancer cell proliferation cannot always be

stopped due to mutations in the growth suppressor systems causing them to lose function.

Probably the most well-known of the tumour suppressors is the p53 pathway. This pathway coordinates transcription allowing the cell to respond appropriately to all sorts of cellular stresses. Normally this tightly controlled transcription allows control of proliferation, however, mutant p53 behaves in very different ways. It has some very unique characteristics due to its function as a transcriptional regulator and, therefore, causes tumourigenesis in numerous different molecular mechanisms (Muller and Vousden, 2013)

1.2.4 Vascular growth

Vasculature growth is very important to the maintenance of healthy tissue, due to the delivery of oxygen and nutrients. In normal tissue the majority of cells, will be adjacent to, or be close enough to, the vasculature for the easy diffusion of the oxygen and nutrients into the cells. In angiogenesis the blood vessels are responsive to both pro and anti angiogenic signals for normal vascularisation of tissues. Angiogenesis occurs when the pro angiogenic signals outweigh the anti angiogenic signals. In cancer the huge demand for nutrients for maintained tumour growth results in a greater need for vasculature, and therefore the pro angiogenic signals are over expressed.

Angiogenesis occurs through sprouting from pre-existing vessels around the tumour, this is triggered by endogenous pro-angiogenic signals, and/or loss of anti-angiogenic signals. This causes a morphological change whereby the Vascular Endothelial Growth Factor (VEGF) binds to one of its receptors on a vascular endothelial cell. In turn the vascular endothelial cell releases proteases causing the breakdown of the vascular basement membrane. The membrane is used to give form to the vasculature and to separate the vascular endothelial cells from the surrounding tissue and consists. With the removal of

the membrane the endothelial cells are able to dissipate from the vasculature structure causing a break in the wall of the vessel. This break simulates a wound and causes different cells to be recruited to the break in the vessel wall. Cells like Endothelial Progenitor Cells (EPC's) and myeloid cells are recruited and release even more soluble angiogenic factors. The recruited EPCs and stem cells in normal tissue would then replicate out into the surrounding tissue, which forms branches from the existing structures. The replication of endothelial cells into a functional vascular branch is dependent on pericytes, which have numerous proposed functions in the angiogenic process. They have been proposed to distribute angiogenic signals to local endothelial cells to ensure they are all working in unison in the local area, causing either proliferation of the endothelial cells for generation of vasculature, or inhibition to prevent morphogenesis of the vasculature.

For a tumour to maintain growth, the cells need to be able to initiate the production of vasculature, and to produce more pro-angiogenic signals than anti angiogenic signals. In cancerous tissue where the angiogenic signals are not properly regulated, this leads to an abundance of endogenous pro-angiogenic signals (Dameron et al., 1994). However with abnormal regulation of the angiogenic signals this physiological process also occurs in an abnormal manner. This process is shown in Fig 1-3 (Risau, 1997). Irregular formation causes problems with delivery of nutrients and oxygen to the cancerous tumour. The atypical vasculature formation limits the infiltration of blood vessels into the tumour. This allows replicating cells on the outside of the tumour to receive nutrients, or disassociate from the tumour and metastasise to somewhere else in the body. It doesn't allow for the cells in the centre of the tumour to receive enough oxygen and nutrients. For patients this limits the efficacy of chemotherapy due to the lack of effective delivery, but doesn't inhibit tumour growth (Bottos et al., 2012, Rak et al., 2000).

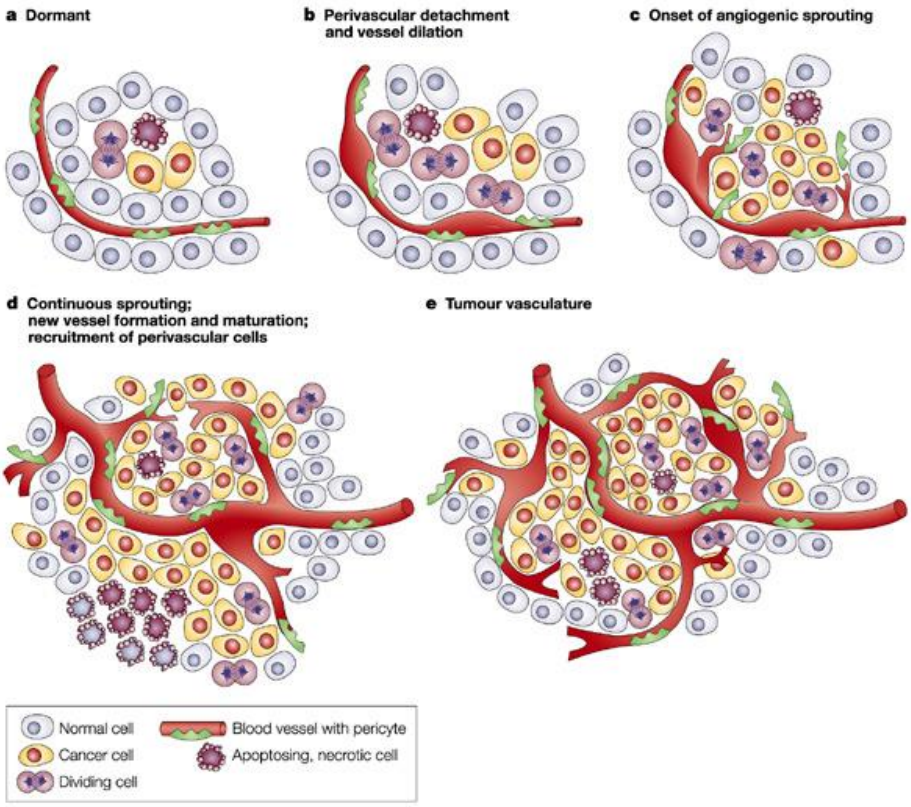


Figure 1-3 A tumours effect on vascularisation.

The graph shows the process that takes place when angiogenesis occurs due to the initiation of a tumour. In a) the tumour does not have enough pro angiogenic factors. In b) the correct amount of factors has been sensed for pericyte detachment and vessel dilation. c) exhibits angiogenic sprouting where endothelial cells replicate to fill gaps where angiogenic sprouting has occurred. In d) the tumour continues to grow exponentially causing blood vessel formation around the tumour which is irregular e)shows the continues formation of vasculature influenced by the tumour (Bergers and Benjamin, 2003) Permissions for reproduction granted, see appendices.

1.2.5 Replicative Immortality

In normal tissue cells are limited in the replicative ability due to senescence and crisis. When cells are replicating there is a limit to the number of times a single cell can replicate. This has been attributed to the telomere chromosomal caps. These caps consist of multiple hexanucleotide repeats which shorten the more often the cell is subject to replication. When the cap goes the cell will enter crisis phase as the chromosome will undergo merging at the ends and the chromosomes will become scrambled (Campisi et al., 2001). To avoid this, the cell first enters senescence, whereby it is functional but has no replicative potential until it undergoes programmed cell death. This allows the tissue

to maintain a functionally homogenous cell type within the tissue. When senescence is avoided the mutations that have occurred in a cell line are not removed, and are passed on to the progeny of the cell containing those mutations. Telomerase allows the addition of the hexanucleotide repeats to the caps of chromosomes, allowing them to continue to replicate long after they would normally have finished. Telomerase is nearly always absent in normal tissue. Its presence in cancerous cell lines shows an ability to evade either senescence or crisis after multiple replications (Campisi et al., 2001, Munoz-Espin and Serrano, 2014).

The model of senescence is still being confirmed, but a delay to senescence is a factor to the development of a tumours. There are potentially many more factors involved in the senescence of cells that could result in tumourigenesis.

1.3 Solid tumour Physiology

When mutations accumulate to the point where there is loss of function in the regulatory pathways, uncontrolled cell replication occurs. This leads to the formation of a mass of undifferentiated cells, known as a tumour. The formation of a tumour occurs in such a way that common characteristics of the mass are seen from patient to patient. One of the characteristics commonly seen is a tumour microenvironment that is hypoxic (Trédan et al., 2007).

In normal tissue the vasculature will develop to allow for all cells to be supplied with oxygen either through direct contact of the vasculature, or through diffusion of the oxygen into the cells. To be sufficiently oxygenated all cells must be in close enough proximity to the vasculature for the efficient diffusion of oxygen, and therefore normal tissue function.

After the uncontrolled replication of cells has occurred, the growth of the tumour is defined by the vasculature that supplies the cancer cells with oxygen (Brown and Giaccia, 1998). The vasculature of a tumour is often irregular and inefficient (Wu et al., 2013), which causes a physiology which encompasses necrotic regions, low oxygen regions, and then slightly more normal regions where the rapid replication of cancer cells occurs. The physiology of a tumour is often necrotic furthest away from the vasculature where the distance from the blood vessels is too large for the diffusion range of the nutrients, therefore the cells die (Brown and Wilson, 2004, Wilson and Hay, 2011). The necrotic cells are surrounded by a hypoxic region where the cells are alive but have an insufficient degree of oxygen to maintain normal function. Where the vasculature has grown abnormally but is still functional enough to supply the immediate tissue there is characteristic uncontrolled proliferation. As the tumour increases in size, the ability of the abnormal vasculature to supply oxygen to the tumour lessens, and more of the tumour is either necrotic or hypoxic. Fig 1-4 shows the average physiology of the tumour.

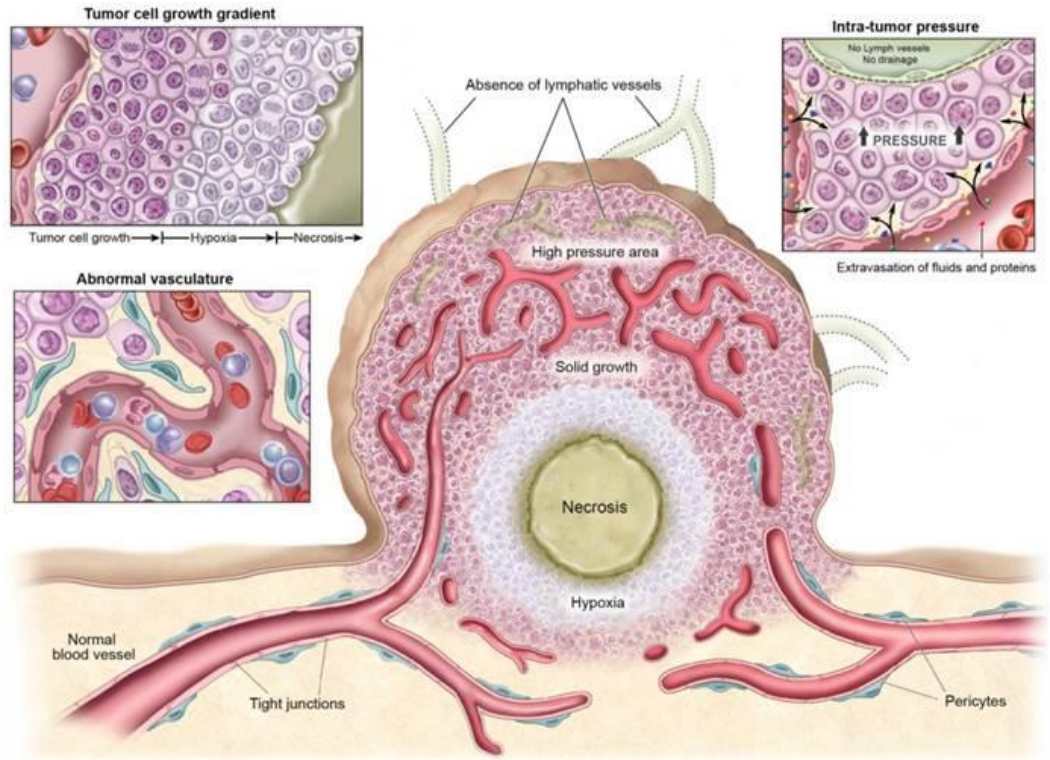


Figure 1-4 The physiology of a solid tumour.

The diagram of the solid tumour shows the different regions of the hypoxic tumour. The first box shows the diffusion gradient of oxygen from the malformed vasculature which is shown in the bottom box. The interstitial tumour pressure is demonstrated in the last box (Kobayashi et al., 2013). Open access for non-commercial replication

Hypoxia in the tumour is aided by the interstitial pressure stopping efficient diffusion of oxygen through the tissue (Wu et al., 2013). High interstitial pressure is caused through blood vessel and lymph vessel abnormalities leading to water and molecules being transported into the tumour but not being efficiently removed (Heldin et al., 2004). This not only prevents oxygen and drugs diffusing through the tumour but the vasculature from forming in these areas, increasing the problems of necrosis and hypoxia.

1.4 Hypoxia and treatment implications.

Traditional treatment options for cancer patients tend to be a combination of surgery, chemotherapy or radiotherapy (http://www.cancerresearchuk.org/about-cancer/cancers-in-general/treatment/, 28/07/2015). The combination chosen depends on the position

and type of tumour that is being treated. Chemotherapy is a range of toxic drugs that are delivered systemically, used to attack rapidly replicating cells, like cancerous cells (Gerber, 2008). Unfortunately through attacking rapidly replicating cells chemotherapeutics also attack cells like stomach epithelial cells, and hair cells (Rajeswaran et al., 2008). There are many different types of chemotherapeutics due to the heterogeneity of cancer and due to patients who can have innate and acquired resistances to the drugs (Donnenberg and Donnenberg, 2005). In conjunction with or instead of chemotherapy, radiotherapy can be performed. The treatments for cancer are getting better with research but currently both radiotherapy and chemotherapy are inefficient in dealing with the hypoxic regions that are commonly found in cancer tumours.

Hypoxia is associated with problems in the treatment of cancer and is therefore associated with a poorer prognosis for many patients. Hypoxia in tumours is caused by the malformed blood vessels which are delivering oxygen and nutrients to the cancer cells. The malformed vessels are very inefficient in the delivery of not only oxygen and nutrients but also chemotherapeutics and internal radiotherapeutics (Brizel et al., 1997, Henderson and Fingar, 1987). So hypoxic regions are symptoms of this malformation, and would indicate a low efficacy of treatment due to the poor delivery of therapeutics (Brizel et al., 1997, Henderson and Fingar, 1987). To exacerbate the issues of delivery, once in and around the tumour what little therapeutic drugs that are delivered have poor pharmacokinetics in the tumour due to high interstitial pressure. This results in the drugs not diffusing into the areas required as they would be moving against a concentration gradient into the tumour. The result of this issue is chemotherapeutics once finally being delivered into the correct area, would only have an effect on the cells at the edge of the tumour and it would leave a tightly packed core of cancerous cells unaffected by the treatment (Jain, 2012).

Hypoxia in the cells is not just an indicator of poor prognosis due to poor efficacy of systemic delivery of the treatments but can also have a direct effect in limiting the efficacy of treatments. As has been mentioned many chemotherapeutics target rapidly replicating cells. However in hypoxic regions, due to the vessels inefficient delivery of nutrients required for replication, cell division happens at a much slower rate. So even if delivered, the chemotherapeutics would not be as efficient as the same drug in a normally oxygenated regions of tissue.

Radiotherapy can be delivered in two ways, internal and external. The external delivery is through the use of focussed high energy radiation beams, directed at the site of treatment. The internal comes through site injections of radioactive elements, or through swallowing of the elements (Moll et al., 2001). Swallowing of a radiosensitiser and radioactive elements will experience the same problems of poor systemic delivery making treatment inefficient again. (Mannaerts et al., 2000, Kupelian et al., 2004) External radiation beam therapy however is also dependent on high levels of oxygen being present in the tissue for it to work well. This is due to the oxygen when being present in the tumour can form free radicals which cause DNA damage. The damage is preferably double stranded breaks which are much harder for cells to repair, especially in the presence of oxygen which binds to the break to prevent binding to the required nucleotide for repair. In the absence of oxygen the free radicals are reduced causing fewer breaks in the DNA, and instead of oxygen thiols are able to bind to the broken DNA strand facilitating its repair. Radiotherapy can be performed in conjunction with immunotherapy, chemotherapy, hormone therapy and surgery.

These issues result in a need for higher doses of chemotherapeutics for patients, which will increase the risk of unwanted side effects, such as acquired resistance, vomiting, hair loss, migraines and fatigue. To improve efficacy of treatment the issue of hypoxia needs

to be dealt with, this calls for alternative treatment routes to be found, either in alternative delivery, targeted drugs or through correction of the vasculature.

1.5 Alternative Therapies.

Due to the heterogeneity of cancer, hypoxia, and the varied response of the patient to traditional treatments, alternatives need to be developed to help improve long term survival rates. For the treatment to be considered effective, the therapy needs to be efficient in killing cancer cells, but ineffective everywhere else in the body. There are many ideas for alternative therapies currently in clinical trials. Many of these will not be affected by a poor vascular system around the tumour, or by the hypoxia in the tumour.

1.5.1 Nanotherapeutics

One such treatment which is being proposed is alternative delivery methods for the drugs, such as nanotherapeutics, liposomal methods, or nanoparticles. Nanoparticles are particles that can exhibit unique properties that can be exploited for the delivery of drugs. They are tiny ranging from 1-100nm in size. The Liposome's are slightly bigger ranging from 400nm - 2.5 µm, they are small vessels constructed of a phospholipid bilayer encapsulating an amount of chemotherapeutics in the centre of a bubble (Sinha et al., 2006). Through the principle of enhanced permeability and retention (EPR) nanotherapeutics accumulate in the tumour as the lymphatic system is generally not present to remove the foreign compound. The retention of nanoparticles in tumours, compared to other chemotherapeutics, is better due to the enhanced retention. The retention is dependent on the increased size of the nanoparticle with regard to other chemotherapeutics. This prevents the normal diffusion of intratumourally injected treatments away from the tumour, which would happen in normal circumstances due to the interstitial pressure and concentration gradient (Ranquin et al., 2005).

1.5.2 Antibody therapy

Antibody therapy is another alternative therapy that can be very effective, without being reduced in efficacy in a hypoxic environment. Antibodies are utilised in several different ways to elicit an anti-tumour effect. Through study of the tumour associated antigens potential antibodies can be found which will target the cancer cells specifically. By creating antibodies that can target cancer cell specific antigens, which would normally be considered to be "self" proteins an immune response or a direct therapeutic response can be elicited. Antibodies are also not affected by the hypoxia in tumours, they have been shown to accumulate in tumours due to the fact that they are able to get into the tumour, and can't be removed. Over time they accumulate increasing the anti-tumour effect (Scott et al., 2012, Adams and Weiner, 2005).

Antibodies can be used as treatments in two distinct ways, they can be used for the activation of the human immune response against cancer cells, or by a direct tumour killing response. This is termed Antibody dependent cellular cytotoxicity (ADCC) or complement dependent cytotoxicity (CDC). ADCC relies on interaction between the antibody and the immune response to obtain cytotoxicity, an antibody will bind to the cancer cell antigens, this bound antibody will in turn allow immune effector cells to recognise the tumour region, and lyse the tumour cells. The lysed cells can also result in tumour antigens being recognised by the B cells of the immune response allowing further recognition and eradication of cancer cells (Iannello and Ahmad, 2005). CDC is a slightly more direct approach, where by an antibody binds to form a complex. This complex triggers a cascade of the events which result in effector cells releasing their proteolytic components which cause the death of cancer cells. The activation of the immune system against cancer cells can result in good tumour regression, and better long term survival rate in patients. However, it can be highly specific, resulting in drugs not being effective in

many people. For example, for ADCC to be effective the cancer has to have the correct antigen, and then be recognised by the Fc receptors. Therefore, the antibody has to be highly specific, this results in it being very narrow spectrum. There is also a direct action antibody therapy that can affect the cell in two ways, agonist and antagonist. When the antibody is an agonist it binds to the cancer antigen, and initiates a cell response, such as apoptosis. The antagonist response would be when the antibody binds to the antigen which causes the inhibition of key processes such as cell replication (Melero et al., 2007, Natsume et al., 2009, Adams and Weiner, 2005)

1.5.3 Antibody Directed Prodrug Converting Enzymes

Another strategy that utilises antibodies is Antibody Directed Enzyme Pro-drug Therapy (ADEPT). In this strategy an enzyme is introduced conjugated to the antibody, which will catalyse the inert compound (a prodrug) into a by-product which is much more toxic than the original, parental compound. The conjugated antibody, like in immunotherapy, will accumulate in the tumour. Once there is a large differential of antibodies present in the cancer tissue compared to that of the normal tissue, then the prodrug can be administered. In clinical trials this differential was determined to be 10,000:1 ratio of antibody in tumour compared to the blood. Once the prodrug is administered, hopefully there will be sufficient concentration of the enzyme present in the tumour to induce a significant anti-tumour effect, not only will this kill the cell that the antibody attaches to but also the surrounding cells in a bystander effect. The antibody and enzyme conjugate can be effective as a single enzyme can cleave many molecules of the prodrug (Napier et al., 2000, Springer and Niculescu-Duvaz, 1997).

This methodology has been tested with regards to several prodrugs for one prodrug converting enzyme. Mouse monoclonal antibody was affinity purified and then an enzyme which had been isolated from *Pseudomonas* sp. strain RS 16 (Carboxypeptidase G_2 or

CPG_2) was conjugated to the monoclonal antibody. This enzyme-antibody conjugate was radiolabelled. A potential prodrug CMDA, which is a benzoic acid mustard, can be broken down into the toxic by-product of 4-[(2 chloroethyl)[2(mesyloxy)ethyl]-amino] benzoic acid (Fig. 1-5). The activity of the conjugated antibody and the unconjugated enzyme was tested *in vitro* using the methotrexate assay. The localisation of the radiolabelled conjugate was measured by measuring radioactivity. CPG_2 has also been tested against an alternative prodrug which is a bis-iodo phenol mustard called ZD2767P. This alternative has a shorter half-life in the body and is still highly potent. The methodology of this approach to treatment gives several advantages over the traditional therapies. The collection of antibodies in the tumour gives some degree of specificity to the treatment, the cells do not need to uptake the antibody conjugate for the enzyme to work, due to the number of prodrug molecules that are able to be cleaved by a single enzyme there is an increase in concentration of the toxic by-product compared to the original conjugate. There is also a bystander effect whereby even if a cancer cell does not have the correct antigen for the conjugate, if it is close to cells which do then the non-antigen containing cell will be exposed to the toxic by-product (Francis et al., 2002).

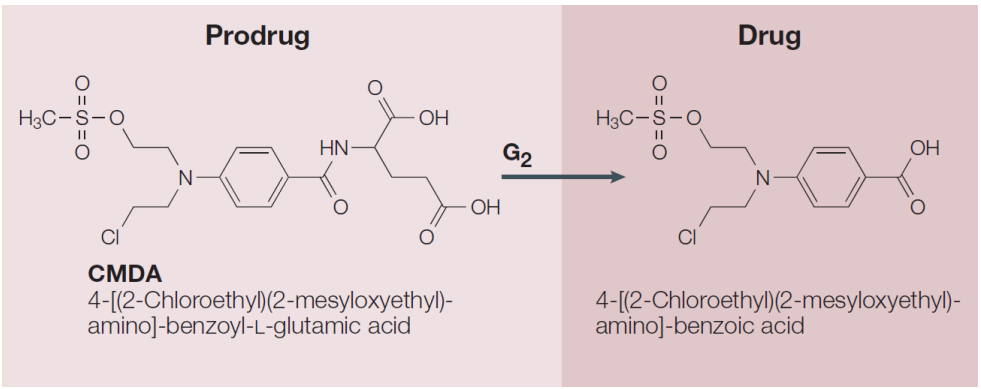


Figure 1-5 The chemical breakdown of the CPG_2 prodrug

This figure shows a chemical representation of carboxypeptidase G_2 breaking down the CMDA prodrug into its toxic product (Minton, 2003). Permissions for reproduction granted see appendices.

As with all treatments there are downsides to the ADEPT approach. While the specificity of the antibody to the antigen is good, not all cancer cells produce the antigens. The number of enzymes conjugated to the antibody is also limited, this results in the level of toxicity in the tumour being limited. There is some leeching of the drug into the surrounding healthy tissue which results in lysis of healthy tissue. When binding the antibody to the enzyme there is a variability in the resulting conjugate, where the enzyme has not bound in the same place every time. This limits the concentration of effective prodrug converting enzymes that can be utilised in the tumour. One of the biggest problems, however, is the use of non-human antibodies. Some of the antibodies that have been used are murine, or rabbit. For some of the trials it was also required for an antibody to be administered which cleared the conjugated antibody out of the human body. This results in an immune response against the antibodies once delivered. Improvements to counter these problems are being looked into like using native, human antibodies, or using abzymes (antibodies which are themselves enzymes to stop the need for the conjugation step) (Springer and Niculescu-Duvaz, 1997, Wentworth et al., 1996).

Of the ten patients that were in the ADEPT trial using CMDA as a prodrug only, one had a response which was partial. The others were either stable or progressive disease. For the trial of the alternative prodrug, 26 patients of the cohort were assessed for clinical response and none was shown in any patient. The best clinical result shown was stable disease. However of the patients treated with ADEPT in this trial none showed any side effects (Francis et al., 2002). In this study the prodrug does not appear to be broken down properly, so while ADEPT has sometimes shown site specificity issues, little prodrug is broken down. Therefore even if the antibody is present in healthy tissue, it will not present side effects.

The Directed Enzyme Prodrug Therapy (DEPT) strategy in principle is a good alternative to the current treatment options, however it is falling down due to its delivery method. In an ideal world there would be a vector for delivery which would exclusively be active in the tumour, limiting the side effects due to the lack of activity in the healthy tissues of the body. This is why it has been implemented with several other delivery vehicles.

1.5.4 Gene Directed Prodrug Therapy

Gene Directed Enzyme Pro-drug Therapy (GDEPT) is a combination of gene therapy and the DEPT strategy. Normally the gene therapy will involve some kind of suicide gene, where by a gene is delivered to the cancer cells, and when produced elicits a terminal response in the cell. The genes picked to do this can be toxic, induce apoptosis, or disrupt metabolic pathways. With GDEPT a gene encoding the enzyme is delivered to the cancerous tumour, and then when produced by the cell, can break down a prodrug to elicit a cytotoxic effect. For this method to be successful, the selection of the correct prodrug and enzyme is essential (Greco and Dachs, 2001).

Delivery of the gene in GDEPT can be done through several ways, one is through a viral approach (Virus Directed Enzyme Prodrug Therapy, or VDEPT), and others are more direct such as cationic lipids, peptides or naked DNA. For example cationic amphiphiles deliver DNA to the cell through coating of a cationic layer. This cationic layer binds to the anionic residue on the cell surface membrane, which will subsequently take up the gene (Behr, 1994). However, by far the most popular way of delivering prodrug converting enzymes to the cells is through a viral vector. When DNA is delivered to the cancer cell, it must be specifically activated in the tumour. To do this the DNA needs to be specifically transcribed and translated in the tumour microenvironment. One way to do that is to use Hypoxia Inducible Factor 1 (HIF-1), a transcription factor which is not present in normal

tissue but highly expressed in tumour cells. HIF-1 binds to sequences called hypoxia responsive elements in the promoter regions of genes causing expression.

When GDEPT is implemented in a direct manner there tends to be a low take up and expression of the gene in cancer cells (10-20% of cells will take the gene up when injected systemically *in vivo*). The toxic product has to have a large bystander effect to consequently overcome this. The enzyme also has to be non-endogenous so that activation of the prodrug is site specific to the tumour. Again, while attractive in principle, there are some issues with GDEPT.

1.5.5 Virus Directed Prodrug Therapy

Virus directed pro-drug enzyme therapy utilizes a viral delivery vector for the release of an enzyme which will break down the pro-drug. Virus Directed Enzyme Prodrug Therapy (VDEPT) is probably the best researched form of delivering a prodrug converting enzyme, and all VDEPT strategies come under the classification of a GDEPT. Many different viruses have been used in the development of VDEPT, such as retroviral RNA viruses, or DNA viruses such as adenovirus, parvovirus and herpes simplex virus (Greco and Dachs, 2001). Neither retrovirus nor adenovirus has distinct advantage over the other, with adenoviruses being able to be produced in higher titres, making production of VDEPT cheaper. It is also possible to alter the capsid of adenoviruses to allow for specified targeting of the virus (although this has not been done for cancer cells). Alternatively retroviruses have a more stable gene transfer method, with a lower chance of an immunogenic response, but cannot be produced in such high titres and will only target replicating cells. The Viruses are modified to produce the prodrug converting enzyme, either inside the cell, or through integration in the host genome. They have often also been modified to be replication deficient. This allows better control of the viral vector once in the body. Once at the tumour site the virus can, using the various mechanisms of

the different viruses, bring about the production of the enzyme. These mechanisms differ from virus to virus, but tend to be about the packaging envelope binding to the cell to delivering the viral DNA or RNA into the transfected cell. The delivered DNA or RNA will then be able to produce the prodrug converting enzyme to break down the prodrug for cytotoxicity, when at the tumour site in sufficient concentration (Xu and McLeod, 2001). However targeting of the VDEPT system is potentially the biggest stumbling block of the system. While some of the viruses are oncotropic most are not. Many steps have been taken to retarget the tropism of the viruses that were used in the VDEPT methodology. When a host cell accepts viral DNA transfer it is due to the virus envelope proteins binding to the cell surface proteins. This binding process results in a hydrophobic transmembrane protein, through which the DNA is transferred. This binding is being modified by creating a biochemical bridge allowing viruses to transfect only cancer cells. The biochemical bridge is a bispecific antibody, which will bind specifically to the antigens of the cancer cells, but also specifically to the virus which has been modified for the production of the prodrug converting enzyme (El-Aneed, 2004, Greco and Dachs, 2001, Grove et al., 1999, Chung-Faye et al., 2001, McNeish et al., 1998).

There are, however, a few serious problems with VDEPT that limit its efficacy. The transfection rate of an intravenous dose of the modified virus is often low, which can be for several reasons. The first is that generation of high concentrations of the pure virus is difficult. Transfection can also be low due to the quantity of virus that actually reaches the tumour in the first place. If it does not reach the tumour in a high enough concentration it cannot transfect the cells in a high enough concentration to allow an effective anti-cancer response. The majority of the viruses used in VDEPT are non-human because they are the most easily modifiable viruses, however, the immune response of the body clears a high quantity of these foreign viruses before they have a chance to reach the tumour. There

are also some minor problems with the system which could cause issues. There is a chance that the viruses recombine to become replication competent in the environment, meaning that the virus would be difficult to control in the body. There is also a chance that random integration of viral DNA into the human genome could activate dormant oncogenes, to cause cancer rather than curing it, or it could inactivate tumour suppressor genes with the same effect (Xu and McLeod, 2001).

All of these different DEPT methodologies result in some anti-tumour response but none are without their respective issues for serious anticancer effect. However, with research continuing there are many different areas and potential strategies which will result in better treatment, with less side effects, and more long term survival of cancer patients. Having an arsenal of treatments which an oncologist can use is important due to the heterogeneity of cancer. Therefore, no one treatment will make all others redundant. Issues such as acquired and innate immunity, extreme side effects, and poor efficacy of treatments mean that a multitude of options need to be available.

1.6 *Clostridium sporogenes*

As a genus *Clostridium* is very diverse and an important group of bacteria. There are numerous species of note that are either clinically important or industrially important, some are both. They are as a group anaerobic, rod shaped bacteria, that can produce endospores in hostile environments as a survival mechanism (Wells and Wilkins, 1996).

Clostridium botulinum is a very well-known clostridial species due to its importance in both the clinically setting and its use in the cosmetics industry. *C. botulinum* produces the most potent neurotoxin known to man. The neurotoxin needs only tiny quantities for skeletal muscle paralysis to occur. Botulism can occur through several different means, such as food borne, wound or infant. Food borne botulism occurs through the ingestion of *C. botulinum* preformed toxin. Wound and infant botulism occurs when the spores are

able to germinate and produce the toxin. These are all potentially deadly (Sebaihia et al., 2007). While *C. botulinum* toxin is deadly, it has still been utilised for treatment against muscle spasticity (Snow et al., 1990) and has even been used in beauty treatments (Botox). It is closely related to *C. sporogenes* another clostridia which has clinical importance.

Originally all Clostridia that were able to produce paralysis inducing neurotoxins were classed as *C. botulinum.* This however encompassed a huge group of bacteria with very diverse phenotypes not relating to pathogenicity. Due to this the *C. botulinum* group has been further split into 4 classes (Classes I – IV).These four classes have been delineated through the 16S ribosomal RNA sequence and nucleic acid hybridisation studies, which showed distinct lineages for those different classes. Since this classification, strains of Clostridia have been identified which have been shown to produce botulinum neurotoxin like *Clostridium baratii* and *Clostridium butyricum* (Harvey et al., 2002, Aureli et al., 1986). These bacteria don't adhere to the 16S ribosomal RNA profile shown by *C. botulinum* and so they are not grouped together phylogenetically. Conversely to this there are also strains which closely adhere to the characteristics of the different classes that are still distinct phylogenetically, because they have no neurotoxin genes like *Clostridium sporogenes* (Poehlein et al., 2015).

All class I *C. botulinum* strains have a very highly conserved 16s rRNA similarity and the bacteria *C. sporogenes* is also very closely related. The Class I group of *C. botulinum* has a 99.7% 16s rRNA sequence similarity to the proteolytic *C. sporogenes.* The metabolic characteristics of the two species of bacteria are nearly indistinguishable. Studies comparing the two bacteria established that the level of similarity is very high, but that the two species should be kept separate due to their differing pathogenic nature. To this end most bacteria with the Botulinum Neuro Toxin (BoNT) are classified as *C. botulinum*

when they adhere to the 16s ribosomal DNA and nucleic acid hybridisation characteristics displayed by the 4 different classes. When 16s ribosomal DNA and nucleic acid hybridisation studies are similar but no toxins are present the bacteria are classified as *C. sporogenes* (Weigand et al., 2015).

Many of the functional domains of the *C. botulinum* Hall A genome are replicated in *C. sporogenes.* The Class I *C. botulinum* Hall A strain has a group of proteases that break down the surrounding environment into usable components for the large number of transporters to uptake. These consist of several different types of proteases such as clostripains, and thermolysin like metalloproteases. These traits have never been shown in any other Clostridia apart from *C. sporogenes.* This would indicate that *C. sporogenes* like *C. botulinum* uses proteases to break down the surrounding environment for nutrient uptake, often this will be for amino acid fermentation. Through phenotypic and genomic comparisons of *C. sporogenes* and *C. botulinum* they have been shown to be highly similar. For example the Class I proteolytic group of *C. botulinum* and *C. sporogenes* both have the same response mechanism to the germination initiator mix of L-alanine, L-lactate and $NaHCO_3$. This response is due to the same receptor genes (Broussolle et al., 2002).

Of the Coding DNA Sequences (CDS's) 84-87% in the class I proteolytic *C. botulinum* are conserved in *C. sporogenes*.

Like *C. botulinum, C. sporogenes* is a Gram-positive, rod shaped, spore forming, obligate anaerobic bacterium (shown in the scanning electron micrograph in Fig 1-6), which is closely related to the pathogenic bacterium *C. botulinum*. It was originally isolated from a soil sample. (LEE and RIEMANN, 1970). This makes it a class one, non-pathogenic bacteria as it has only loosely been associated with issues, but industrially ideal as it can be used as a surrogate for the study of food borne botulism without any need for the use of

pathogenic bacteria. It is, therefore, used by the food industry to improve sterilisation techniques to ensure that *C. botulinum* does not make its way into any food products.

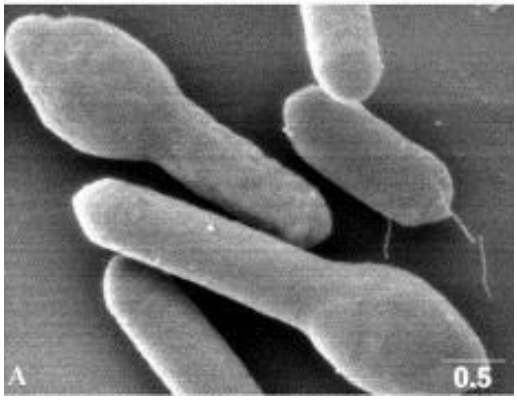


Figure 1-6. A scanning electron micrograph of *C. sporogenes.*

A picture showing both vegetative cells and cells undergoing sporulation of C. sporogenes (Kalchayanand et al., 2004) Permissions for reproduction granted, see appendices.

It is also clinically important for, while not being a confirmed pathogen, it has been associated with gas gangrene of wounds. This has only happened in a limited number of cases. It can be considered to be a clinically important bacterium for another reason, as it has been linked with the DEPT strategy (Kubiak and Minton, 2015). It has many characteristics that would enable it to be the perfect vector for a prodrug converting enzyme to the hypoxic tumour. As it is an anaerobic bacterium it will colonise only the regions of the body that are anaerobic, which is the hypoxic tumour. It can be transported around the body in its spore form as this will be resistant to an immune response. When PCE's are delivered to the hypoxic tumour through use of *C. sporogenes* as a delivery vehicle it is called Clostridial Directed Enzyme Prodrug Therapy (CDEPT)

1.6.1 A Historical Perspective

Clostridium spp. have been well documented to colonise hypoxic tumours exclusively in the human body. Just over 60 years ago, Malmgren *et al* (Malmgren and Flanigan, 1955)

performed a set of experiments that proved this through the injection of isolated *Clostridium tetani* into mice. The mice were separated into 2 different groups, mice with cancerous tumours and mice without. Once the two different sets of mice were injected with *C. tetani* spores the mice that had cancer died from tetanus. This showed that the inactive spores remained inactive in healthy mice, whereas they germinated and exhibited pathogenesis when in mice with tumours. Through biopsies of the tumours it could be seen that the bacteria had germinated and colonised the tissue, resulting in death for the mice (Malmgren and Flanigan, 1955).

This unique characteristic revealed *Clostridium spp.* great potential as a cancer therapeutic, but uncertainty remained as to how best to utilise this feature. Pathogenic bacteria could not be used due to the obvious complications of toxicity. Möse and Möse, however, showed that the *Clostridium* genus had numerous bacteria which exhibited not only colonisation of various types of tumours, but that a lot of these bacteria showed a degree of tumour lysis (Heppner and Möse, 1978). This was thought to be because the bacteria broke down the tumour for metabolic products that were required. This oncolytic effect was studied with numerous variables changed to give a better idea of what is needed for efficient lysis of the tumour by clostridia. Different clostridial spores were used, different types of tumour were tested, as well as supplements to the spore injection to help lytic effect. All these different experiments showed that the use of *Clostridium spp.* in cancer therapy was a real possibility (Möse and Möse, 1964).

After testing in mice the *Clostridium* based cancer treatment was used in clinical trials. The bacterium used was *C. butyricum* M55. These trials showed some oncolysis, but also resulted in low grade fevers in the patients. The fever increased in severity, in line with the lysis of the tumours. The fever tended to appear one week after bacterial infection, but was attributed to the inflammation that occurred in the brain from glioblastoma lysis.

While clinical trials failed to show a clinically relevant level of lysis in the tumour, it did hold promise in the use of *Clostridium spp.* in treatment as there were little side effects, and some native lysis did occur, just not in the scale hoped for from the original researchers (Heppner and Möse, 1978, Barbé et al., 2006).

This led to questions of improving efficacy in this treatment methodology. The main idea was to use a non-pathogenic *Clostridium* not as the treatment in itself, but as a delivery vehicle for a treatment. The previous work into oncolysis caused by *Clostridium* showed a very specific line between the normal healthy tissue, and the malformed cancerous tissue. This makes *Clostridium* an ideal vector for chemotherapeutic delivery, and several methods of utilising this have been proposed. One is a direct method whereby the bacteria have a gene introduced, which when produced, will have a direct effect upon the tumour. The alternative is a more indirect approach which follows the principle described in the DEPT strategies mentioned earlier. This would require an inert prodrug which has no effect on the patient or the bacterial growth, and which can be broken down by an enzyme that it is possible to produce in clostridia (Tirandaz et al., 2006).

These methods were not feasible until reliable reproducible genetic modification techniques were available for *Clostridium,* and the right *Clostridium* species had been picked for the task at hand. Recently clostridial conjugation and transformation methods have been developed, and out of the capability to insert plasmids into many *Clostridium* species has come a whole raft of genetic manipulation techniques. This has allowed researchers to standardise plasmids for numerous tasks (Heap et al., 2009), to knockout genes through targeted insertion (Heap et al., 2007, Heap et al., 2010), to chromosomally integrate genes through homologous cross over (Heap et al., 2012), and to obtain clean deletions of genes through allelic exchange (Ng et al., 2013). These tools have made it possible to implement chemotherapeutics in *Clostridium.*

The characteristics needed for picking the best clostridial delivery vector was for it to be efficient in its colonisation (the colonisation should be site specific, and at a high rate), genetically facile, and non-pathogenic. Due to the high number of clostridial species a bacterium which fitted this description was found in *Clostridium beijerinckii.* It was shown to be able to express a prodrug, cytosine deaminase, for the breakdown of 5-fluorocytosine (5-FC) into 5-fluorouracil (5-FU) which is toxic. It was also shown to be able to produce nitroreductase, an *Escherichia coli* enzyme, which can also break down 5-FC. However, in studies in cancer bearing mice, the *C. beijerinckii* were not shown to have much anti-cancer efficacy (Minton et al., 1995). Consequently closely related *Clostridium* species were looked at for an alternative. *C. sporogenes* was hypothesised to be a suitable alternative as it is closely related to *C. beijerinckii.* When tested for colonisation of tumours, *C. sporogenes* had a greater colonisation rate than that of *C. beijerinckii.* Whereas *C. beijerinckii* produced 10^6 bacteria per gram of tumour at best, *C. sporogenes* produced 10^8 bacteria per gram of tumours (Lambin et al., 1998).

The biggest problem inherent with using *C. sporogenes* as a delivery vector was its resistance to transformation. This was solved in 2002, when Liu *et al* successfully transformed an *E. coli* expression plasmid containing a gene for the expression of cytosine deaminase into *C. sporogenes* (Liu et al., 2002). Expression of the gene was confirmed through western and immunoblots and through the subsequent demonstration of cytosine deaminase activity *in vitro* and *in vivo.* The *in vivo* experiments were performed on tumour bearing mice, where a spore suspension was intravenously injected and then the tumour size was measured over time.

This simple experiment paved the way for future work on using other *Clostridium spp.* as delivery vehicles for prodrug-converting enzymes directed to the site. The specificity would result in numerous upsides for the patient, with little to no side effects that are

normally associated with chemotherapy. Due to the site specificity, the dosages can be lower, and will take less time to administer. These advantages would make it an ideal "Trojan horse" for cancer therapeutics.

1.6.2 Current Work

Current research into *Clostridium sp.* as a treatment option is divided into two camps. The first is people looking into the original idea of using an oncolytic strain of *Clostridium,* this time in the guise of *Clostridium noyvii* NT. This bacterium, like *C. sporogenes* is a spore forming, anaerobic bacterium, but is, however, pathogenic. The first step that was required for clinical application of *C. noyvii* was for it to be attenuated through heating of the bacteria at 70^{o}C for 15 min (Dang et al., 2001). This was performed as the toxin, α-toxin, was coded for within a phage episome. Once the phage elements were inactivated by the heat treatment PCR was performed to ensure loss of the α-toxin gene. Once this attenuated strain of *C. noyvi (C. noyvi* NT) was generated it was looked into for its oncolytic effect. Combinatorial treatment of the *C. noyvi* NT strain for the hypoxic regions within the tumour, and with a chemotherapeutic for the vascularised regions of the tumour was proposed. This method was called Combined Bacteriolytic Therapy (also called COBALT). The COBALT method was used on a cohort of tumour bearing mice to see what possible influence this could have on the efficacy of treatment of hypoxic tumours. Alongside the COBALT treatment, controls of just the *C. noyvi* NT strain, and of just the chemotherapeutics were performed. Of the different treatments only the COBALT method was truly effective at regression of the tumour. While there was slight efficacy of the other treatments the tumour was never completely removed and so would re-establish in a human patient. This is due to the specificity of both chemotherapeutics to only well vascularised areas and of the *C. noyvi* NT to the hypoxic region. The delivery of the drugs was through intravenous injection (Dang et al., 2001).

Through exploration of alternative chemotherapeutics, and alternative animal models, COBALT was continually shown to have an impressive effect on tumours, with complete regression of tumours occurring under the right conditions, leading to cancer free progression of the animal model until the 90 day mark. This led to it being implemented in larger animal models, namely dogs, and eventually 1 human patient. However, due to concerns over intratumoural bacterial concentration being low in larger mammals, the delivery method was changed to an intratumoural injection of spores over an intravenous injection.

The intratumoural injection of spores in 16 dogs was performed in at least 1 of 4 planned cycles of treatment. Most of the cohort in the study received 4 cycles of treatment. Two dogs had the tumours excised before the possible evaluation of their response to the treatment. Of the 14 dogs that could be assessed there was a complete response in three dogs, there was partial response in three further dogs, and the rest showed either stable or progressive disease. Consequently the use of *C. noyvi* NT was used in an intratumoural injection in a human patient. The human patient showed good response to the intratumoural injection over a very short period. However, the side effects of the treatment were quite severe. The treatment caused severe necrosis of the bone tissue that the tumour was attached to, resulted in gas present in the soft tissue. This resulted in difficult to manage pain, and the tumour had to be drained of accumulated fluid and tumour debris. The presence of the *C. noyvi* NT was still found in the cancerous tumours when the patient was last reported upon (Roberts et al., 2014).

Alternatively to the direct use of an attenuated oncolytic bacterium, there is also ongoing research into implementing a prodrug converting enzyme into *C. sporogenes* for clinical use (Liu et al., 2002, Theys et al., 2006a, Heap et al., 2014). Not only are there attempts at

using *C. sporogenes* to produce a Prodrug Converting Enzyme (PCE) but also a direct therapy, such as antibody therapy.

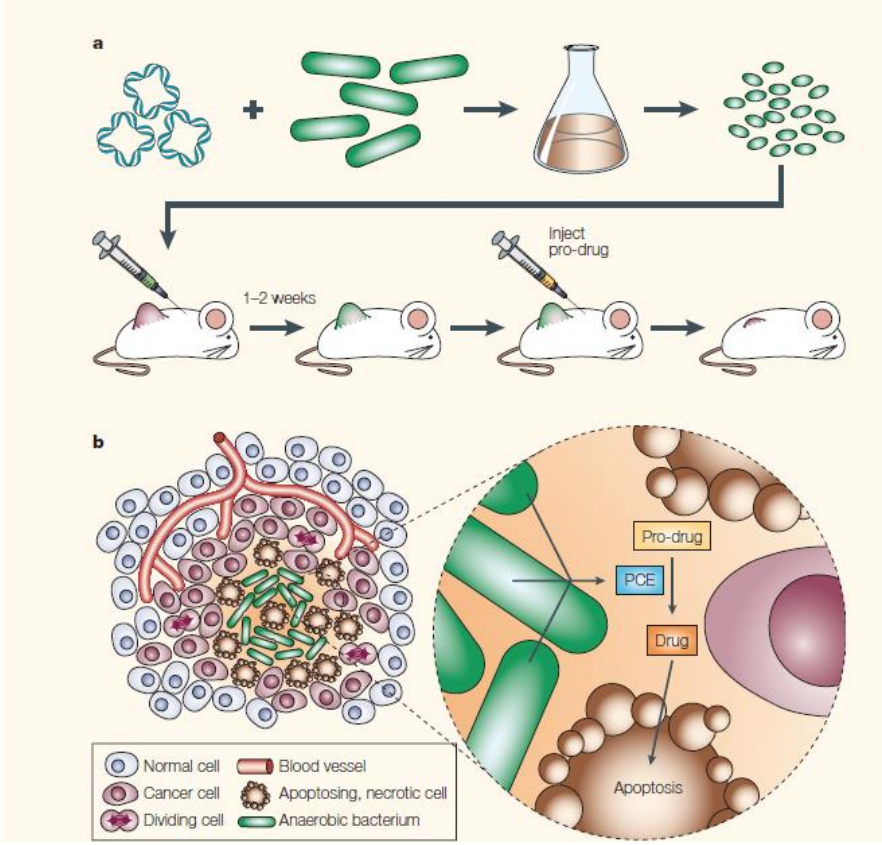


Figure 1-7. A schematic representation of the mode of action of CDEPT.

In a) it is shown the process of the treating a tumour with CDEPT, through the production of the prodrug converting enzyme into C. sporogenes and then growing a spore suspension which is then injected in the mice for the prodrug to be converted into a toxic product, killing the tumour. b) shows what happens in the tumour for the breakdown of the prodrug (Minton, 2003). Permissions for reproduction granted see appendices

C. sporogenes can be seen as an improved vector to the ADEPT, GDEPT and VDEPT systems. It has been shown to be site specific, and to colonise the tumour very well (Theys et al., 2006b, Heap et al., 2014). Due to the fact that the bacteria only colonises the tumour, in its spore form it is not susceptible to the immune system, and so smaller concentrations of spores are required to colonise the tumour making treatment more cost effective. Added to this a lot of the clostridia strains have their own anti-cancer effect on the tumour before PCE's even is implemented. It is ideal as a delivery vehicle for treatments as no toxic effect will be seen until the bacteria germinates and produces the

anti-cancer drug. This limits the side effects exhibited, as unlike chemotherapeutics, no rapidly replicating cells will be harmed unless they are in a solid hypoxic tumour. This model of treatment is shown in Fig 1-7

There are several issues with the use of *C. sporogenes* as a delivery vector for cancer therapeutics. Concerns have been raised about the specificity of the spores to the hypoxic region of the cancer tumour, for example the hypoxia that occurs when engaging in exercise, or in the patients with the condition cortical spreading depression. When suffering from this condition, demand of oxygen exceeds supply causing periods of hypoxia in tissue (Takano et al., 2007). Whether these conditions provide the right microenvironment for the germination of *C. sporogenes* spores, has not been confirmed. Another problem that has been proposed is the spore's removal in the body. In theory the spore suspension, once delivered will use the circulatory system to travel around the body, and once in the right environment germinate. However, of all the spores injected into the body many will not find the right environment to germinate, and due to their resistance to antibiotics, there is an unknown question of what happens to those spores that do not find the correct environment. To avoid possible downstream issues these questions need to be answered. Experiments have been done to assess the presence of clostridial spores in the body over time in rats, however, these experiments are not proven in humans (Lambin et al., 1998). They do show a lack of germination in any non-tumour area in the body, and that the spores left the body after a period of time. This time differed depending on where in the body the spores were looked for but after 25 days no spores were found in the brain or kidney, but some were found in the liver (Lambin et al., 1998).

With the improvement in recombinant methods for *C. sporogenes* a whole range of treatments can be implemented, there have been many different prodrug converting

enzymes implemented in other DEPT systems, which could be transferable to the clostridial based DEPT system. This includes nitroreductase which was used in VDEPT, and CPG_2 which was used in ADEPT. There could also be merit to implementing direct acting treatments such as has been described with suicide gene therapy, only delivered in a *C. sporogenes* vector. These genes could be coding for antibodies which allow for the neutralisation of over-expressed proteins which have a role in tumourigenesis, like the angiogenic regulator VEGF, or could induce apoptosis like a caspase. By having a range of different drugs produced by *C. sporogenes in vivo* an improvement in the prognosis of the patients can be achieved.

1.7 Sporulation

Clostridial species, like some other bacterial species such as *Bacillus subtilis,* have the ability to sporulate. Sporulation is the formation of a metabolically dormant version of the clostridial cell, with a specialised spore wall, which is resistant to numerous physical and chemical stresses. These stresses can include high pressures and temperatures, solvents and lytic enzymes. When clostridia grow to a high enough density that there is nutrient limitation a section of the population begins the process of sporulation.

The sporulation process is a similar morphological process for different bacteria, such as *B. subtilis* and the different clostridial species, as it is hypothesised that they have evolved from the same ancestor. However the molecular process that is involved is far more diverse than was originally thought. The sporulation process consists of seven stages (Fig 1-8). The first stage of the process, Stage 0, is normal vegetative growth of the bacteria. Stage I and Stage II is the beginning of the sporulation process, where an nutrient deficient environment triggers the beginning of sporulation and an axial filament is produced for the asymmetric division of the bacterial cell. Initially about a third of the bacterial DNA is in the spore side of the division but the rest of the DNA is pumped into

the prespore section when the prespore is being formed. The prespore becomes known as the forespore once the prespore is fully enveloped by the mother cell producing the spore. This step is Stage III of the spore formation cycle. For Stages IV and V the forespores germ cell wall is first coated with a peptidoglycan layer, and then the cortex is established. The spore is then coated in the complex protein structure that makes up the coat of the spore. In Stage VI the spore's resistance to UV radiation and heat is established. Finally in stage VII the spore is released through the lysis of the mother cell. This expensive process is morphologically conserved across many of the sporulating bacteria, such as the well characterised *B. subtilis* (Talukdar et al., 2015). However the molecular action in which this process is achieved is different between the Clostridial species and *B. subtilis.*

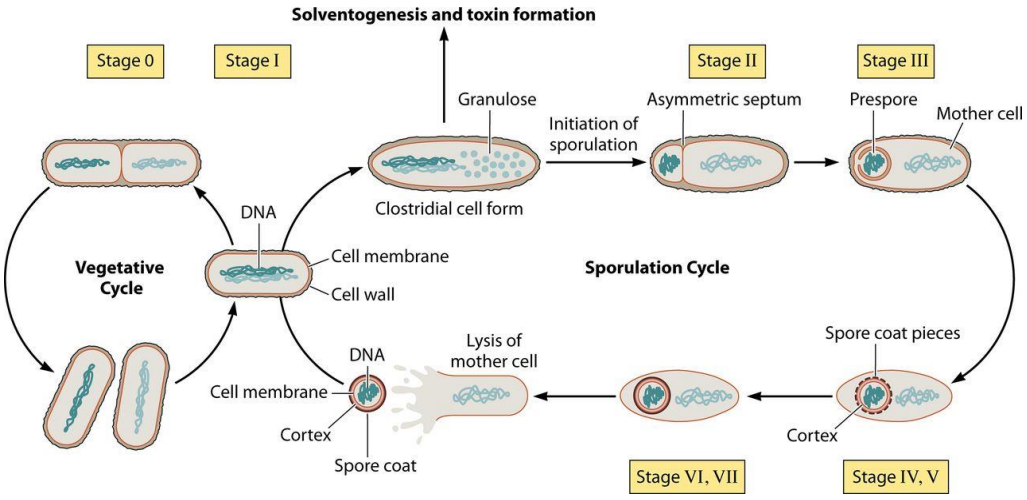


Figure 1-8 A step by step process of sporulation.

A schematic representation of the life cycle of Clostridium bacteria. On the left of the diagram there is a representation of normal growth in Stage 0. In stage 1 the DNA starts to replicate ready for the sporulation process. The sporulation process then continues through to the lysis of the mother cell, and the release of the fully formed spore. (Al-Hinai et al., 2015) Open access for non-commercial replication.

The fully formed spore has a well-defined coat that encapsulates the bacterial chromosome inside the spore core. This is separated from the rest of the spore by the inner forespore membrane which is the initially developed layers. On the outside of the

inner forespore membrane is the cortex, which is subsequently separated from the rest of the spore via the outer forespore membrane. The cortex and the outer forespore membrane are coated in four more layers, the basement layer, the inner coat, the outer coat, and then the final layer of the spore is the crust. This is shown in Fig. 1-9.

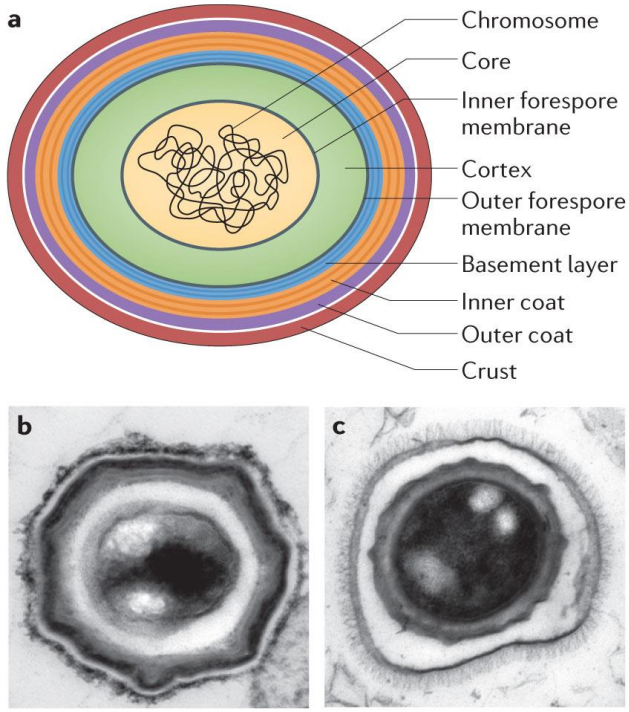


Figure 1-9 Different representations of bacillus spores, which are homologous to clostridial spores.

In a) the structure of the bacillus spore. b) Shows a section transmission micrograph of the B. subtillis stained with ruthenium red. c) shows a section transmission micrograph of B. subtillis stained with osmium tetroxide. (McKenney et al., 2013) Permissions for reproduction granted see appendices.

The initiation of the molecular process is still somewhat unknown. While it is evident that a lack of nutrients or the presence of toxic products can cause cells to start sporulation, the exact way in which this happens is still to be confirmed. In *B. subtillis* this initiation is known to be through the activation of different sensor kinases which once triggered, through either intracellular or extracellular triggers, autophosphorylate, starting a phosphorylation cascade through numerous different sporulation genes, until Spo0A is triggered. This is well characterised in *Bacillus* with the phosphate donor and receiver of

each Spo protein being known, and through this well characterised system, research can be targeted to similar areas in the Clostridial proteome (such as Kinases and Spo proteins). Unfortunately, apart from the key sporulation regulator, Spo0A, no homologous proteins have been found in Clostridial species bar a version of Spo0B in *C. tetani*. The function of this protein remains unknown despite homology to the *Bacillus* Spo0B. However recently it has been confirmed that certain kinases do have a role in the activation of Spo0A in *C. acetobutylicum*, and through the use of BlastP numerous different kinases have been identified in several of the most well-known *Clostridia* species. When the sensor kinases are activated through response to a stimulus, they in turn cause the phosphorylation of Spo0A. Its phosphorylation leads to its activation and the subsequent transcription or repression of those genes important in the sporulation cascade (Durre, 2014).

The Spo0A protein has numerous different roles, and has been associated with functions outside of sporulation such as toxin production. For sporulation Spo0A is a transcription factor that regulates (both positively and negatively) different σ factors which are part of a cascade controlling the different morphological stages in sporulation in both the mother cell and the forespore. The sporulation σ factors recognise and interact with the different promoters of the sporulation genes, such as spore coat proteins like *cotA*, and aid in DNA melting for the RNA polymerase to bind and initiate transcription of those genes. There are four major σ factors which are involved in sporulation (F, E, G, and K) and the genes producing these σ factors have been identified in both *Clostridia* and *Bacillus,* but most extensively characterised in *Bacillus*. The σ factor cascade has been investigated in some *Clostridia,* such as *C. difficile* and *Clostridia botulinum* (Kirk, 2015). The *C. botulinum* sporulation process is the best model for *C. sporogenes* due to the genetic relatedness of the species, until its own sporulation mechanisms can be investigated. In other *Clostridia* this process may be different, but they all require the same σ factors. In *C. botulinum* each

σ factor has a role in activating the transcription of another σ factor, and also activating different genes at the different morphological stages for the formation of a spore (Kirk et al., 2014). For example σ^F has a role in activating σ^E but it is unknown precisely what it does for the morphological changes taking place. The locale of the the different σ factors is known so it is known that σ^E and σ^K are in the mother cell, and σ^G and σ^F are in the prespore. The protein σ^E aids in transcription of the *spoIID*, *spoIIM* and *spoIIP* genes which are responsible for the engulfment of the forespore, not only this but σ^E regulates σ^G. The σ^G factor in turn activates cortex related genes and *spoIVB* which has a role in the production of σ^K. σ^K has a relatively unknown function however it is involved in late stage sporulation, probably being involved in the activation of spore coat proteins. While the σ factors E, F, G and K are used in other Clostridia, in *C. difficile* it has been shown that the activity of the σ factors are not dependent on each other in the same way (Pereira et al., 2013).

Sporulation is a very important part of what makes the Clostridia species important. The clostridial spore allows *Clostridium* pathogens to survive in environments that are hostile to them, so that when the opportunity arises it can come out of this dormant state to grow in its vegetative form. Whether that opportunity arises from infection in a wound, or from colonising a gut that has been cleared from its normal flora through antibiotics, the survival mechanism of sporulation is very important to the *Clostridium* genus. Therefore, it is important that we understand the molecular mechanisms of sporulation so that we may prevent it or utilise it. It is also important to understand the physiology of a spore when produced. Much of what is known about clostridial sporulation is based on the well-studied organism *B. subtillis* which has a similar, but not identical, spore system with Clostridia (Al-Hinai et al., 2015).

1.8 Prodrug Systems

Carboxypeptidase G_2 (CPG_2) is a folate reducing agent that hydrolyses the C-terminal glutamate of folic acid and its analogs. CPG_2 was originally isolated from an unclassified *Pseudomonas* sp. Strain RS-16 (Minton et al., 1983). It has been used clinically already, when patients are suffering from methotrexate toxicity, as methotrexate is a folate analog and would be reduced by the CPG_2. Folic acid is involved in important steps in maintaining cell stability and deficiency in folic acid has been linked with cancer itself. This is because folic acid plays a role in chromosomal stability, and is crucial for DNA synthesis and repair. It is also involved in the production in S-adenosylmethionine, which plays a role in the methylation of genes, allowing for regulation of the transcription of genes. If these genes are oncogenes then this could increase the chances of cancer developing (Duthie and Hawdon, 1998, Kruman et al., 2002).

CPG_2 has also got potential for cancer treatment as a prodrug converting enzyme. It has been implemented previously as part of the ADEPT strategy, however, in clinical studies it showed very little efficacy (Francis et al., 2002). It is unsure as to whether this was down to the prodrug converting enzyme, or to the delivery method. An alternative method for delivery of the prodrug converting enzyme could be investigated to confirm whether the issue was suitable delivery, or suitable therapeutics. Due to the specificity of germination, and the resistance of spores to the human immune system, it could be that the CDEPT strategy is better suited for the CPG_2 prodrug converting system. The structure of the protein CPG_2 and the prodrugs and the drugs chemical structure (after catalysis by CPG_2) have been shown in Fig 1-10

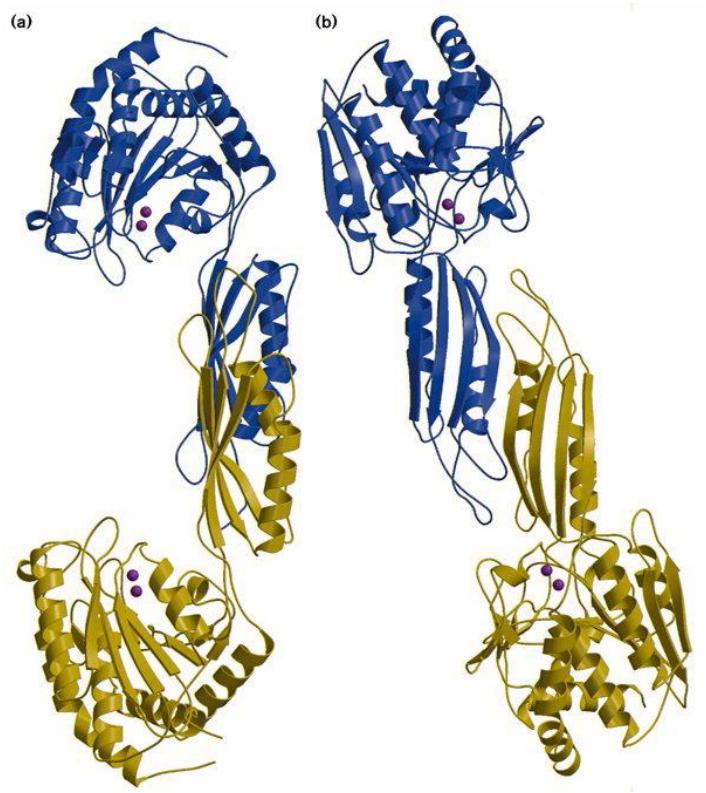


Figure 1-10 Structural diagrams of CPG_2.

The bottom panel shows a view of the CPG_2 protein, and a second orthoganol view of the CPG_2. This view shows the two subunits in blue and yellow (Rowsell et al., 1997, Minton, 2003). Permissions for reproduction granted see appendices

The limits of the CDEPT system are as of yet unknown. There could be a lot of potential for the delivery of drugs in several different and innovative ways. For example, prodrugs' being catalysed by prodrug converting enzymes, was hypothesised as a way of improving specificity to the tumour target site for the toxic treatment, therefore minimising side effects. However, with the research into *C. novyi* NT as an oncolytic bacterium, it has been shown that *Clostridium* can be used in many different ways (Roberts et al., 2014). Direct action therapeutics could be utilised in the systems, such as human specific antibodies. These should not interfere with any systems with the bacterial cell, but once in the cancer tissue, could have a profound effect. Therefore, monoclonal antibodies (mAbs) such as anti-VEGF can be looked into.

Anti-VEGF is a proposed therapy which has been used in clinical trials. It motive action is to neutralise the Vascular endothelial growth factor which is produced in high quantities by cancerous tissue resulting in the production of poorly formed vasculature in and around the tumour tissue. Through neutralization of VEGF it is proposed that there would be two possible outcomes, the first is the neutralization of the VEGF leading to normal vasculature growth, which would result in better efficacy of chemotherapeutics as they would reach the cells better. Alternatively it could result in the neutralization of all VEGF molecules leading to the cancerous tissues having no developing vasculature and, therefore, no nutrients delivered to it. This would result in necrosis of this tissue.

However there are different isoforms of VEGF which have been generated through alternative splicing of RNA. It has been argued that the different isoforms of this gene cannot only play a role in angiogenesis but also conversely prevent it. Isoform VEGF-A_{165}a and VEGF-A_{165}b are supposed to have differing effects on angiogenesis with Isoform VEGF-A_{165}a being involved in pro-angiogenic activity and VEGF-A_{165}b being involved in anti-angiogenic activity according to Bates *et al* (Harper and Bates, 2008). However this is directly in contrast to what has been found by Harris *et al* (Harris et al., 2012), who through their own studies refute the claim that VEGF-A_{165}b is present in tissue at all and that it therefore cannot have an anti-angiogenic effect. If there are pro and anti-angiogenic isoforms of VEGF, then the anti-VEGF antibody would not distinguish between the two isoforms, meaning that the treatment could prevent the anti-angiogenic signals limiting the proposed effect.

Anti-VEGF has been investigated in clinical trials under the guise of bevacizumab. It is a humanized, recombinant monoclonal antibody that can bind to all isoforms of the protein VEGF (the protein structure has been shown in Fig 1-11) (Presta et al., 1997).

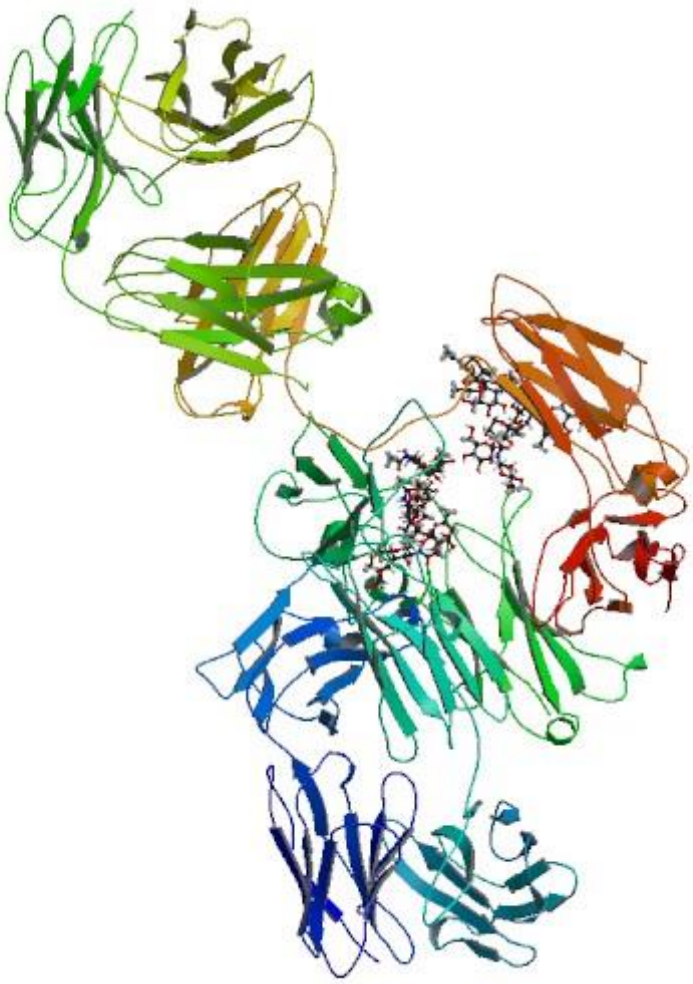


Figure 1-11 A structural diagram of Anti-VEGF.

The structure of bevacizumab, an anti-VEGF monoclonal antibody. This drug is meant to normalize the vasculature around the tumour allowing better delivery of other chemotherapeutics (DrugBank, 2015, Wishart et al., 2006). Open access for non-commercial replication

In this trial 44 patients with adenocarcinomas were treated for two weeks with bevacizumab solely before being treated with bevacizumab combined with the traditional chemotherapy of 5-fluorouracil, beam radiotherapy, and then finally surgery. The effect of vascularisation of the tumour was measured in comparison to that of a patient just receiving 5-fluorouracil, and beam radiotherapy. After 12 days of treatment with bevacizumab, and before the chemotherapy and radiotherapy were implemented, there was some initial regression in tumour size for a single patient. The other patients on the regimen showed no regression of tumour size before the implementation of the rest of

the regimen. At this time point there was also a great reduction in the volume of blood found within the tumour, as well as a much lower density of microvasculature. The interstitial fluid pressure (IFP) was greatly reduced in most patients probably due to the normalisation of the vasculature. The normalisation of tumour vasculature will result in a degree of normalisation of the pressure within the tumour. This should improve the pharmacokinetics within tumours. In clinical surveys the use of bevacizumab has been shown to be a great normaliser of the tumour vasculature, which results in reduced blood flow, blood volume in the tumour, and reduced IFP. All of this will allow for more effective chemotherapy regimens with better prognosis, and lower dosses can be administered. Lower doses will result in a reduction in severe side effects (Shah et al., 2011).

It has been to phase III trials which showed varied results in combinatorial therapies, and in different cancer types. Often bevacizumab was used as a combinatorial therapy with systemically delivered chemotherapeutics. The bevacizumab would act as a tumour vasculature normalizing agent, allowing the chemotherapeutics to have a better effect. Some cancer types have had a much better response to bevacizumab than others. For example one trial has shown to have an improved median survival rate of patients treated with bevacizumab and a standard chemotherapy regimen of irinotecan, fluorouracil and calcium folinate when compared to the control group that was treated with the same standard chemotherapy regimen and a placebo. This strategy was performed on colorectal cancer leading to an improved median survival rate of 15-16 months up to 20-23 months when treated with bevacizumab and the standard regime compared to just the standard regime. This trial of 800 patients was encouraging for the eventual clinical use of the monoclonal antibody, bevacizumab (Deng et al., 2013). However the use of bevacizumab in cases of glioblastoma have not had the same effect on survival rates as with other cancer types. This was shown when patients diagnosed with glioblastoma were

treated with radiotherapy and daily temozolomide. In addition to this they were given either placebo or bevacizumab. Progression free survival of patients was prolonged but not to the extent needed, and overall survival of patients did not increase (Gilbert et al., 2014). For a successful treatment regime with anti-VEGF implemented within a CDEPT strategy, first it would need to be ascertained what sort of cancer the patient was suffering from and whether anti-VEGF would be a suitable strategy for treatment. It would also need to be known whether the combinatorial treatment would be viable with CDEPT. As bevacizumab improves progression free survival often only in combinatorial methods would the drugs be effected by the Clostridia or affect the Clostridia as a treatment. These are questions that would need to be considered for future progress.

1.9 Aims

Through the work done previously on optimisation of the expression of PCE's in *C. sporogenes* the stage is set to start implementing a host of alternative therapeutics such as other PCE's and direct treatments. These can be implemented through the advancement in molecular techniques that are now available to clostridial researchers as well as better knowledge of cancer treatments. It was hypothesised that different therapeutics could be implemented such as alternative PCE's, and that the tumour vasculature could be normalised through implementing the monoclonal antibody anti-VEGF. In using *C. sporogenes* as a delivery vector this could circumvent the issues of interstitial fluid pressure preventing of anti-VEGF that other trials would have to take into account. Utilising this knowledge the aims of this project were to implement efficient expression and exportation of alternative drugs from *C. sporogenes* NCIMB 10696. This was done through:

- Implementing a new prodrug converting enzyme in clostridia, which has already been used in Antibody Directed Enzyme Prodrug Therapy, carboxypeptidase G_2

- Implementing a new direct therapy, in the anti-VEGF antibody, to aid in the normalisation of vasculature, allowing for better chemotherapeutic action.
- Ensuring efficient expression and production of the 2 different genes, and that the two proteins are efficiently exported outside the cell for efficacy of treatment
- Test the functionality of both proteins once efficiently expressed and exported.
- To complete the genome sequence and annotation of *C. sporogenes* NCIMB 10696

Materials and Methods

Chapter 2

2 Materials and Methods

2.1 Chemicals and suppliers

Table 2-1 A list of Suppliers and the reagents, kits and equipment they supplied

Supplier	Chemical Reagents
Sigma Aldrich	Media components and other chemicals unless otherwise stated. Sigma Aldrich GenElute™ HPminiprep kit, GenElute™ PCR cleanup kit, GenElute™ Gel Extraction kit, and GenElute™ Bacterial Genomic Extraction kit.
Oxoid	Media components and other chemicals unless otherwise stated
New England Biolabs	Restriction nucleases and corresponding buffers, Phusion® polymerase, LongAmp® polymerase, and Antartic phosphotase
Novagen	BugBuster, r-Lysozyme™, Benzonase™, pET28b and pEt28a plasmid vectors, His•Bind® purification kit
QIAGEN Ltd	QIAprep spin miniprep kit; QIAquick PCR purification; QIAquick gel extraction kit, QIAamp DNA mini kit.
Novex	*4-12% SDS-PAGE gels, ZOOM 2d SDS-PAGE kit, Anti-His-(C-Term)-HRP, NuPAGE LDS 4x Sample buffer, Nu PAGE Transfer Buffer 20x, NuPAGE MES SDS Running buffer*
BioRad	TransBlot Turbo™ transfer pack
Invitrogen	4-12% SDS-PAGE gels
Fisher Scientific	Media components and other chemicals unless otherwise stated
Promega	T4 DNA ligase and buffer, GoTaq polymerase and buffers
Roche	cOmplete protease inhibitor cocktail tablets
Cambio	Failsafe PCR system

2.2 Bacterial strains

Table 2-2 A Table of the different bacterial strains that have been used

Strain	Description
Escherichia coli TOP10[a]	Plasmid expression and storage strain
Escherichia coli HB101 (CA434)[b]	Conjugative strain
Escherichia coli BL21pLysS (DE3)[c]	Plasmid expression and storage strain
Clostridium acetobutylicum ATCC 824[d]	Wild type isolate
Clostridium sporogenes NCIMB 10696[d]	Wild type isolate
Clostridium sporogenes NCIMB 10696[d] *pyrE* repaired *tcdR*	*Clostridium sporogenes* with *TcdR* gene introduced for induction of genes under the control of *PtcdB*
Variovorax paradoxus ATCC 25301[d]	Wild type isolates
Variovorax paradoxus CPG_2 A[d]	Potential wild type isolate
Variovorax paradoxus CPG_2 B[d]	Potential wild type isolate
Variovorax paradoxus CPG_2 C[d]	Potential wild type isolate

[a]Invitrogen™

[b]M. Young, UCW, Aberystwyth, UK

[c]Novagen®

[d]Clostridia Research Group culture collection.

2.3 Plasmids

Table 2-3 All plasmids that were used in this study, and their reference

Plasmid	Relevant Properties	Source/ Reference
pJ201	Gene storage vector	DNA2.0
pJ201-scFv-antiVEGF_OptA	scFv *antiVEGF* Option A gene, codon optimized for *Clostridium sporogenes,* stored in the pJ201 storage vector	DNA2.0 for this study
pJ201-scFv-antiVEGF_OptB	scFv *antiVEGF* Option B gene, codon optimized for *Clostridium sporogenes,* stored in the pJ201 storage vector	DNA2.0 for this study
pJ201-scFv-antiVEGF_OptC	scFv *antiVEGF* Option C gene, codon optimized for *Clostridium sporogenes,* stored in the pJ201 storage vector	DNA2.0 for this study

pJ201-scFv-antiVEGF_OptD	scFv *antiVEGF* Option D gene, codon optimized for *Clostridium sporogenes,* stored in the pJ201 storage vector	DNA2.0 for this study
pJ201-CPG$_2$OptEC	Carboxypeptidase G$_2$ gene optimized for *E. coli* in the storage vector pJ201	DNA2.0 for this study
pJ221-CPG$_2$OptCspo	Carboxypeptidase G$_2$ gene optimized for *C. sporogenes* in the storage vector pJ221	DNA2.0 for this study
pJ201-SortaseA	*Staphylacoccus aureus SrtA* gene codon optimized for *C. sporogenes* in the storage vector pJ201	DNA2.0 for this study
pJ201-CipAacetoCPG$_2$	*CipA* signal sequence from *Clostridium acetobutylicum* synthesized at the 5' of *CPG$_2$* gene in the storage vector pJ201	DNA2.0 for this study

pJ201-CipCcelluCPG$_2$	*CipC* signal sequence from *Clostridium cellulyticum* synthesized at the 5' of *CPG$_2$* gene in the storage vector pJ201	DNA2.0 for this study
pJ201-CipAthermoCPG$_2$	*CipA* signal sequence from *Clostridium thermocellum* synthesized at the 5' of *CPG$_2$* gene in the storage vector pJ201	DNA2.0 for this study
pJ201-BarnCPG$_2$	*Barn* signal sequence from *Clostridium acetobutylicum* synthesized at the 5' of *CPG$_2$* gene in the storage vector pJ201	DNA2.0 for this study
pJ201-NprCPG$_2$	*Npr* signal sequence from *Clostridium sporogenes* synthesized at the 5' of *CPG$_2$* gene in the storage vector pJ201	DNA2.0 for this study

pJ201-1483CPG$_2$	*1483* signal sequence from *Clostridium sporogenes* synthesized at the 5' of *CPG$_2$* gene in the storage vector pJ201	DNA2.0 for this study
pJ201-1442CPG$_2$	*1442* signal sequence from *Clostridium botulinum* synthesized at the 5' of *CPG$_2$* gene in the storage vector pJ201	DNA2.0 for this study
pJ201-LPRTGC.spor	*LPRTG* signal sequence from *Clostridium sporogenes* synthesized at the 3' of *CPG$_2$* gene in the storage vector pJ201	DNA2.0 for this study
pMTL84151	The modular plasmid with the pCD6 gram positive replicon, catP antibiotic marker, ColE1+tra gram negative replicon and a multiple cloning site	Heap *et al* (Heap et al., 2009)

pMTL84153	The modular plasmid with the pCD6 gram positive replicon, catP antibiotic marker, ColE1+tra gram negative replicon and a *Pfdx* promoter with a multiple cloning site	Heap *et al* (Heap et al., 2009)
pMTL82251	The modular plasmid with the pBP1 gram positive replicon, ermB antibiotic marker, ColE1+tra gram negative replicon and a multiple cloning site	Heap *et al* (Heap et al., 2009)
pMTL84153- scFv-antiVEGF_OptA	scFv *antiVEGF* Option A gene cloned into modular expression vector, pMTL84153	DNA2.0 for this study
pMTL84153- scFv-antiVEGF_OptB	scFv *antiVEGF* Option B gene cloned into modular expression vector, pMTL84153	DNA2.0 for this study
pMTL84153- scFv-antiVEGF_OptC	ScFv *antiVEGF* Option C gene cloned into modular expression vector, pMTL84153	DNA2.0 for this study

pMTL84153- scFv-antiVEGF_OptD	scFv *antiVEGF* Option D gene cloned into modular expression vector, pMTL84153	DNA2.0 for this study
pMTL82251:YZ2-CPG$_2$OptEC	*CPG$_2$* gene optimized for *Escherichia coli* cloned into a modular vector with a *Lac* inducible system	This Study
pMTL82251:YZ2-CPG$_2$CspoV2	*CPG$_2$* gene optimized for *Clostridium sporogenes* cloned into a modular vector with a *Lac* inducible system	This Study
pMTL82251:YZ2-NprCPG$_2$V2	*Npr* signal sequence from *Clostridium sporogenes* synthesized at the 5' of *CPG$_2$* cloned into a modular vector with a *Lac* inducible system	This Study
pMTL82251:YZ2-1483CPG$_2$V2	*1483* signal sequence from *Clostridium sporogenes* synthesized at the 5' of *CPG$_2$* cloned into a modular vector with a *Lac* inducible system	This Study

pMTL82251:YZ2-1442CPG$_2$V2	*1442* signal sequence from *Clostridium botulinum* synthesized at the 5' of *CPG$_2$* cloned into a modular vector with a *Lac* inducible system	This Study
pMTL82251:YZ2-CipAacetoCPG$_2$V2	*CipA* signal sequence from *Clostridium acetobutylicum* synthesized at the 5' of *CPG$_2$* cloned into a modular vector with a *Lac* inducible system	This Study
pMTL82251:YZ2-CipCCelluCPG$_2$V2	*CipC* signal sequence from *Clostridium cellulyticum* synthesized at the 5' of *CPG$_2$* cloned into a modular vector with a *Lac* inducible system	This Study
pMTL82251:YZ2-CipAthermoCPG$_2$V2	*CipA* signal sequence from *Clostridium thermocellum* synthesized at the 5' of *CPG$_2$* cloned into a modular vector with a *Lac* inducible system	This Study
pJ201LPXTG1	*LPXTG* motif identified from the *Clostridium acetobutylicum* gene Phosphohydrolase, Icc family synthesized in the storage vector pJ201	DNA2.0 for this study

pJ201LPXTG2	*LPXTG* motif identified from the *Clostridium acetobutylicum* gene 2,3-Cyclic-nucleotide 2'phosphodiesterase synthesized in the storage vector pJ201	DNA2.0 for this study
pJ201LPXTG3	*LPXTG* motif identified from *Clostridium acetobutylicum* gene Membrane protein synthesized in the storage vector pJ201	DNA2.0 for this study
pMTL82251:YZ2-CPG$_2$LPXTG1	*LPXTG1* motif attached to the 3' *CPG$_2$CspoV2* gene and cloned into a modular vector with a *Lac* inducible system	This Study
pMTL82251:YZ2-CPG$_2$LPXTG2	*LPXTG2* motif attached to the 3' *CPG$_2$CspoV2* gene and cloned into a modular vector with a *Lac* inducible system	This Study

pMTL82251:YZ2-CPG$_2$LPXTG3	*LPXTG3* motif attached to the 3' *CPG$_2$CspoV2* gene and cloned into a modular vector with a *Lac* inducible system	This Study
pMTL82251:YZ2-NprCPG$_2$LPXTG3	*LPXTG3* motif attached to the 3' *NprCPG$_2$* gene and cloned into a modular vector with a *Lac* inducible system	This Study
pMTL82251:YZ2-NprCPG$_2$LPRTG	*LPRTG* motif attached to the 3' *NprCPG$_2$* gene and cloned into a modular vector with a *Lac* inducible system	This Study

2.4 Primers

Oligos were chosen by identification of signal peptides using the SignalP online tool (version 3.0 – 4.0). The nucleotide sequences of these signal peptides were then synthesised in both sense and antisense ready for oligomerisation (Materials and Methods 2.6.12)

Table 2-4 The different Oligos which were annealed together for use as signal sequences in cloning

Signal Sequence	Oligomerisation sequences
Caceto_Barn_rev	TCGAGTGTTTTTGCTTGAACACAAACAACATTATTTAAAAATGCTCCAAA TATCATAA
Caceto_CipC_rev	TCGAGTGTATCTGCAAAAACATGATTTTTACTTATTCCTAAACTTGCTGT CATCATAA
Cbot_1443_for	GATCCATGAAAAGTAAAAAATTATTAACAATATTATTAAGTGCAATAATA ACAGCAAGTAGT
Cspo_1485_for	GATCCATGAAAAGTAAAAAATTATTAGCAACAGTTTTAAGTGCAGTTATA ACATTTAGTGCA
Caceto_CipC_for,	GATCCATGAAAAAAAGAAATATAGCAATATTAGGAATGTTTATGATGAC AGCAAGTTT
Cbot_1445_rev	TCGAGAGGTGCTGCATAAACTGCACTAACTGTACTAAATGTTATAACTG CACTTAAAACTGT
Cbot_1444_rev,	TCGAGAGGTGCTGCACTAACTGCACTAACTGCACTAAATGTTATAACTG CACTTAAAACTGT
Cspo_1485_rev,	TCGAGAGGTGCTGCACTAACTGCACTAACTGCACTAAATGTTATAACTG CACTTAAAACTGT
Cspo_Npr_rev	TCGAGTCCTGCTGCATATATACTACTAACAGGACTAAATGTCATTATTG TACTTAATAATATTG
Cbot_1443_rev	TCGAGTTCTGCTGCATAAACACTACTTATACTACTTGCTGTTATTATTGC ACTTAATAATAT
Cspo_1484_rev	TCGAGAGGTGCTGCACTAACTGCACTAACTGCACTAAATGTTATTATTG CACTTAAAACTGT
Cbot_1442_for,	GATCCATGAAAAGTAAAAAATTATTAACAATATTATTAAGTGCAATAATA ACAGCAAGTAGT
Cbot_1446_for	GATCCATGAAAAGTAAAAAATTATTAGCAACAGTTTTAAGTGCAGTTATA ACATTTAGTACA
Cbot_1445_for	GATCCATGAAAAGTAAAAAATTATTAGCAACAGTTTTAAGTGCAGTTATA ACATTTAGTACA
Cbot_1446_rev	TCGAGAGGTGCTGCACTAACTGCACTAACTGTACTAAATGTTATAACTG CACTTAAAACTGT
Cbot_1444_for	GATCCATGAAAAGTAAAAAATTATTAGCAACAGTTTTAAGTGCAGTTATA ACATTTAGTGCA
Cthermo_CipC_rev	TCGAGTGTTGCTGCACTAACTGTTTGAGGTATCATTGCTGCAAATATTG TTGTTAA
Cspo_1483_fo	GATCCATGAAAAGTAAAAAATTATTAGCAACAGTTTTAAGTGCAGTTATA ACATTAAGTACA
Caceto_Barn_for	GATCCATGAAAAAAAAAAAATTTTGCAGCAATATGTTTTACATTAGTTTTT ATGATATT
Cthermo_CipC_for	GATCCATGAGAAAAGTTATAAGTATGTTATTAGTTGTTGCAATGTTAACA ACAATA
Ccellu_CipC_for	GATCCATGAGAAAAAAAAGTTTAGCATTTTTATTAGCATTAACAATGTTA GTTACATT
Cspo_1483_rev	TCGAGAGGTGCTGCATAAACTGCACTAACTGTACTTAATGTTATAACTG CACTTAAAACTGT
Ccellu_CipC_rev	TCGAGTCCTGCTGCAAATGCTGTTAATTGTGCTCCTAATAATGTAACTA ACATTGTTA
Cspo_1484_for,	GATCCATGAAAAGTAAAAGATTATTAGCAACAGTTTTAAGTGCAATAAT AACATTTAGTGCA
Cbot_1442_rev	TCGAGTTCTGCTGCATAAACACTACTTATACTACTTGCTGTTATTATTGC ACTTAATAATAT
Cspo_Npr_for	GATCCATGAAAAGTAAAAAATTATTAACAATATTATTAAGTACAATAATG ACATTTAGTCCTGT

All primers for PCR were created through identification of high GC areas upstream and downstream of the region that was to be amplified. Approximately 20bp sequences were taken and then modifications were incorporated depending on the purpose of the primer. For cloning at the 5' end restriction enzyme sites were added with a spacer of 3-4 nucleotides between the restriction enzyme recognition site and sequence required for amplification. Complementary overlaps of two PCR products were added so that Splice Overlap Extension PCR might be performed. The primers were then tested for the Melting Temperature (T_m) and therefore GC content,

secondary structure and alternative annealing sites within the template. This was done using online programs such as NEB T_m calculator and primer-BLAST from NCBI. For the primers used in the RT PCR, Primer3Plus was used to design appropriate primers.

Table 2-5 A list of the primers used in this study

Primer Name	Primer Sequence	Primer Name	Primer Sequence
CPG2 WT Forward	GTCCATATGTTTTCAAAAATCAAAAAAATTAATTTT	LPRTG_For,	GGTGCGGGAAAATGATTACCACGTACTGGATATAGC
CPG2 WT Reverse	TTTGCTAGCTCACTTGCCGGCGCCCAGAT	LPRTG_Rev_HindIII	GGAAGCTTATTTACTTCTTAATTGCTTTTGTCTTC
CPG2 Synth Forward	GTCCATATGTTCTCAAAGATTAAGAAAATAAATTTT	NPR_CPG2_LPRTG_Rev	GCTATATCCAGTACGTGGTAATCATTTTCCCGCACC
CPG2 Synth Reverse	GACGCTAGCTTATTTACCAGCACCTAGATCCATTA	Npr_CPG2_LPXTG_ rev	CCAGATTTTTGGTAAATCATTTTCCCGCACC
anti-VEGF_A Forward	GCGGAGCTCCATATGTTTTCAAAAATCAAGAA	NprCPG_LPXTG3	GGTGCGGGAAAATGATTTACCAAAAATCTGG
anti-VEGF_A Reverse	GCGCGGCGCCTCTAGACTATTAGTGATGATG	Srt1_F	GAATACCCATCAAAGCATCG
anti-VEGF_B Forward	GCGGAGCTCCATATGTTTAGTAAAATCAAGAAG	rpoA_R	GAGAGCTGAAACGTGATAGTTCTAA
anti-VEGF_B Reverse	GCGCGGCGCCGTCTAGACTATTAGTGATGATGGTGG	rpoA_F	TGCCTTGCCTGCAAATATAA
anti-VEGF_C Forward	GCGGAGCTCGTGTTACATATGTAATGTGGAATTG	Srt2_R	TGCACAAATTTAAATCCCAAA
anti-VEGF_C Reverse	GCGCGGCGCCTCTAGACTATTAGTGATGATG	Srt1_R	GGCCAGAATTTAGAGCTTCG
anti-VEGF_D Forward	GCGGAGCTCCATATGTTTAGCAAGATAAAGAAG	Srt2_F	GGAGAGATTCCGTAGGCTTTTT
anti-VEGF_D Reverse	GCGCGGCGCCGTCGACTCTAGACTATTAGTG	egla_lprtg_Soe_Rev	GCTATATCCAGTACGTGGTATCTAGATTACTTACCAGCGC
cpg_Ec Forward	GCGGCATGCATATGTTTAGCAAAATCAAA	lprtg_egla_Soe_For	GGGCGCTGGTAAGTAATCTAGATACCACGTACTGGATATAGC
cpg_Ec Reverse	GCAAGCTTCTAGATTACTTACCAGCGC	Srt13060_rev	CCATAATGTACTACTTCTCCATTTCCGCTTTAAC
cpg_Cs Forward	GGCTCGAGAAACGAGACAATGTATTGTTTCAAGCAGCAACAGACG	Srt13060_for	GAGATGTTTATAGCGTAGATATG
cpg_Cs Reverse	TCATTTTCCCGCACCTAAATCCATG	Srt22540_rev	GCTATAGCATTGTTGATACTGCAATTCAGC
pMTLA-anti-VEGF_A-rev	CGACGGCCAGTGCCAAGCTTATAAAAATAAGAAGCCTGCAAATGC	srt22540_For	GCTATTCTAAAAGCACCATTTTTTTCAC

pMTLA-anti-VEGF_A-fwd	AATTCGAGCTCGGTACCCGGTGTGTTACATATGTAATGTGGAATTG	Srt_22540_F2	ATAGGAAGGTGCTTTAAATG
pMTLA-82251BB-rev	CAATTTTTTTATCCTGCAGGGGGCCCGATCGGTCTTGCCT	Srt_22540_R2	CTTTCAGGTGCAGATCCTATATATC
pMTLA-82251BB-fwd	TGCAGGCTTCTTATTTTTATAAGCTTGGCACTGGCCGTCG	lprtgegla_Soe_For_cor	GGTGCGGGAAAATGATACCACGTACTGGATATAGC
pMTLA-YIP-rev	CACATTACATATGTAACACACCGGGTACCGAGCTCGAATT	lprtgegla_soe_Rev_Cor,	GCTATATCCAGTACGTGGTATCATTTTCCCGCACC
pMTLA-YIP-fwd	AGGCAAGACCGATCGGGCCCCCTGCAGGATAAAAAATTGTAGAT	EglACPG2_For_BamHI	GCGGGATCCATGTTTTCAAAAATCAAGAAAA
pMTLC-anti-VEGF_C-rev	ATTCAGAACGGTCCTGCAGGATAAAAATAAGAAGCCTGCAAATGC	LPXTG3_Reverse_XhoI	GGCTCGAGTTAAGCACCTCTAATTAATTC
pMTLC-anti-VEGF_C-fwd	AATTCGAGCTCGGTACCCGGTGTGTTACATATGTAATGTGGAATTG	LPRTG_Reverse_XhoI	GGCTCGAGTTATTTACTTCTTAATTGCTTTTGTCTTC
pMTLC-82443BB-rev	CAATTTTTTTATCCTGCAGGGGGCCCTGCGTATTGGGCGC	EglACPG2_Rev_XhoI	GCCTCGAGTCATTTTCCCGCACCTAAATCC
pMTLC-82443BB-fwd	TGCAGGCTTCTTATTTTTATCCTGCAGGACCGTTCTGAAT	NprCPG2_Rev_XhoI	GCCTCGAGTCATTTTCCCGCACCTAAATCC
pMTLC-YIP-rev	CACATTACATATGTAACACACCGGGTACCGAGCTCGAATT	V. Paradoxus_CPG_For	CGCGGATCCACCCGAACGAACAATGCG
pMTLC-YIP-fwd	GCGCCCAATACGCAGGGCCCCCTGCAGGATAAAAAATTGTAGAT	V. Paradoxus_CPG_Rev	CGCAAGCTTGTGGCTTCCTGGTCTGCTGTC
pMTLD-82443BB-rev	CACATTACATATGTAACACAGGGCCCTGCGTATTGGGCGC	V.para_SOE_CPG2 For	CAGAAGCGCGACACAACG
pMTLD-82443BB-fwd	AATTCGAGCTCGGTACCCGGCCTGCAGGACCGTTCTGAAT	v.para_SOE_CPG2 Rev	GCAAGCTTGCCGCCCTTGTCGTCG
pMTLD-anti-VEGF_D-rev	CAATTTTTTTATCCTGCAGGATAAAAATAAGAAGCCTGCAAATGC	V.para SOE_EglA For	GCGGGATCCGGGGGTATTCATATGTTTTC
pMTLD-anti-VEGF_ D-fwd	GCGCCCAATACGCAGGGCCCTGTGTTACATATGTAATGTGGAATTG	V.para SOE_EglA Rev	CGCGCTTCTGAGCTTCAGCTTTATAAGT

pMTLD-YIP-rev	ATTCAGAACGGTCCTGCAGGCCGGGTACCGAGCTCGAATT	pET28b_VEGF_A For	GCGGAGCTCGTGTTACATATGTAATGTGG
pMTLD-YIP-fwd	TGCAGGCTTCTTATTTTTATCCTGCAGGATAAAAAAATTGTAGAT	pET28b_VEGF_A Rev	GCCTCGAGGTCGACTCTAGACTATTAGTGATGA
S-layer protein Rev	GCCTCGAGTTTATTATTTAATTGCATTATTTTACTCTTTTCTTGG	pET28b_VEGF_B For BamHI	GCGGGATCCGTGTTACATATGTAATGTGGAATTG
S-layer protein fwd	GCGTCGACGAGACACCAGAAGGCCAAAAGAATTCAAGCATATTAG	pET28b_VEGF_B For NdeI	GCGCATATGGTGTTACATATGTAATGTGGAATTG
VEGFB_TcdB_NdeI_F	GCGCATATGGTGTTACATATGTAATGTGGAATTG	pET28b_VEGF_B Rev	GCAAGCTTGTCGACTCTAGACTATTAGTGA
CPG_For_BamHI	GCGGGATCCATATGTTTAGCAAAATCAAA	pET28b_VEGF_C for	GCGGAGCTCGTGTTACATATGTAATGTGGAATTG
LPXTG2_Rev_HindIII	GGAAGCTTTTATTGTACTGATTTTTGTTTATTC	pET28b_VEGF_C Rev	GCCTCGAGGTCGACTCTAGACTATTAGTG
LPXTG_Rev_HindIII	GGAAGCTTTTAAGCACCTCTAATTAATTC	pET28b_VEGF_D For	GCGGGATCCGTGTTACATATGTAATGTGGAATTG
LPXTG_For_SOE	GGGCGCTGGTAAGTAATCTAGATTACCAAAAACTGG	pET28b_VEGF_D Rev	GCCTCGAGGTCGACTCTAGACTATTATG
CPG_Rev_SOE, LPXTG1_Rev_HindIII,	CCAGTTTTTGGTAATCTAGATTACTTACCAGCGCGGAAGCTTACTTGGTAGGGGAAGATTTC	ToxA_F	ATTCAGGATGGAAAGTATCACTTAAT
SrtA_Rev_HindIII	GGAAGCTTATTTTACTTCTGTAGCCAC	ToxA_R	TTCTACGCCTGCCTGTGATG
SrtA_For_BamHI	GCGGGATCCATGAAAAAGTGGACTAATAGATTAATG	ToxB_F	CAGGAGAAGTGGAGCGAAAA
CPG2_XhoI_For	GGCTCGAGAAACGAGACAATGTATTGTTTCAAGCAGCAACAGACG	ToxB_R	CTTGCGCCTTTGTTTTCTTG
YZ2_Screening_R	GATTAAGTTGGGTAACGCC	ToxC_F	TCCTCGAGTTACAAGCC
YZ2_Screening_F	ACCATGATTACGAATTCGAG	ToxC_R	CAGGAAAGGGTATATCTG
CCacetoFor_P	GCGGATCCATGAAAAAGAGAAATATCGCGATTTTAGG	ToxD_F	TTAGACTATACAGCATCCC
BarnFor_P,	GCGGATCCATGAAAAAGAAAAACTTTGCGGCTATATGTTTTACC	ToxD_R	TAACTTGTGGACGAATCC
BarnRev_P	GCAAGCTTCATTTTCCCGCACCTAAATCCATGATTAATCTCGC	ToxE_F	CCAAGATTTTCATCCGCCTA
		ToxE_R	GCTATTGATCCAAAACGGTGA

NprFor_p	GCGGATCCATGAAAAGCAAAAAGCTCTTAACCATTTTAC	ToxF_F	CGGCTTCATTAGAGAACGGA
1442For_P	GCGGATCCATGAAATCCAAGAAATTACTTACCATCCTG	ToxF_R	TAACTCCCCTAGCCCCGTAT
CCcelluRev_P	GCAAGCTTCATTTTCCCGCACCTAAATCCATGATTAATCTCGC	ToxG_F	GGAACCATTCAATGACCCAGGGCCAGGAAC
CCthermoRev_P	GCAAGCTTCATTTTCCCGCACCTAAATCCATGATTAATCTCGC	ToxG_R	GCTGCAAATTTGTCGGGCGGTGTAGATGCA
CCacetoRev_P	GCAAGCTTCATTTTCCCGCACCTAAATCCATGATTAATCTCGC	Sc4A6_F	CCACTTCTTCTTCTTGGTAGACATAATGGTAGTTCTAAAGG
CCcelluFor_P	GCGGATCCATGAGAAAGAAAAGCCTGGCATTCCTCTTAGCC	Sc63_R	CCCACGCCTAAATTTTTCGATCCATATTTACTTTTTATGAATATATATCTACCTTTTACCGTTCTAG
CCthermoFor_P	GCGGATCCATGAGAAAAGTAATCTCAATGCTCTTAGTTGTTGCC	Sc32_F	CGGAAGTCCACATCGTTAATATAAAGATGGACTCCGAGAATTTTATCGATATAGGAAATGATACTGG
1442Rev_P	GCAAGCTTCATTTTCCCGCACCTAAATCCATGATTAATCTCGC	Sc32_R	CCCCCAAGACCCTTTGGTATAAGTTCTACCCCGTTATCATCATAAGCAACCC
1483For_p	GCGGATCCATGAGAAAGTGAGAAGCTTCTTGGGGAGGGAAAAATGTGAAAATAGATGAAGATAATTTTATAAGGCAGC	Sc25_F	
NprRev_P	GCAAGCTTTCATTTTCCCGCACCTAAATCCATGATTAATCTCGC	Sc25_R	CCGTACTCTCCTGTTTGTACCAACTTTCGACACTATCTTATAATTAAC
1483Rev_P	GCAAGCTTCATTTTCCCGCACCTAAATCCATGATTAATCTCGC	Sc54B_F	CCAACTGTATTAATTAGGGGGTTAATATATGGGATTCCG
1483_For P	GCGGATCCATGAAAAGTAAGAAACTCCTTGCGACCG	Sc54B_R	CCGTACTCTCCTGTTTGTACCAACTTTCGACACTATCTTATAATTAAC
S-layer_protein_For_BamHI	GCGGATCCGAGACACCAGAAGGCCAAAAGAATTCAAG	Sc4A6_R	CCAACAGGATATTAATCTCTTAGACTACCAAATCGG
S-layer_protein_rev_HindIII	GCAAGCTTCCAAGAAAAGAGTAAAATAATGCAA	Sc63_F	CCGGATACCCTTCTCAACGTTGTCCTCAATTCCCATATTGTAATTCTCTTCTACTATTTCGACATAG
S-layer protein_rev_hindIII_2	GCAAGCTTTTTATTATTTAATTGCATTAT	M13 F	GTAAAACGACGGCCAGT
LPRTG_HindIIIRev	GGAAGCTTATTTACTTCTTAATTGCTTTTGTCTTC	M13 R	AACAGCTATGACCATG

LPRTG_01885_SOEFo	GGTGCGGGAAAATGATTACCACGTACTGGATATAGC	CPG2_NPR_SOERev	GCTATATCCAGTACGTGGTAATCATTTTCCCGCACC

2.5 Microbial Methods

2.5.1 Media

All media components, antibiotics and biochemicals were from Fischer scientific, Oxoid, or Sigma-Aldrich unless otherwise stated. Media was prepared with distilled water, and for agar 1- 1.5% No.1 Bacteriological agar (w/v) was added. Prior to use, all media was autoclaved at 121 ^{o}C for 15 min. Additional components unable to be autoclaved were filter sterilised. Compositions of all types of media are listed in Table

When making plates and cultures they were prepared using aseptic technique.

Basic Media Composition

Table 2-6 A list of Media and Their Composition

Media	Components	Quantity (g/l)
Amino Acid Complete and Definied Medium	The composition of this media is listed in the appendix	The composition of this media is listed in the appendix
E.P. Buffer	Sucrose Na_2HPO_4 HCl	270mM 5mM Adjust to pH 7.4
Luria-Bertani (LB)	Tryptone Yeast Extract NaCl	10 5 10
Phosphate Buffered Saline (PBS)	NaCl KCl Na_2HPO_4 KH_2HPO_4 HCl	8 0.2 1.44 0.24 Adjust to pH 7.4
M1 Minimal Media	The composition of this media is listed in the appendix	The composition of this media is listed in the appendix
Pseaudomonas Minimal Media	The composition of this media is listed in the appendix	The composition of this media is listed in the appendix
S.O.C. Medium (Invitrogen™)	Tryptone Yeast Extract NaCl KCl $MgCl_2$ $MgSO_4$ Glucose	2% w/v 0.5% w/v 10mM 2.5mM 10mM 20mM
TYG	Tryptone Yeast Extract Sodium Thioglycolate NaOH	30 20 1 Adjust to pH 7.2
2xYTG	Tryptone Yeast Extract NaCl Glucose HCl	16 10 5 0.5% Adjust to pH 5.2

2.5.2 Supplements

Table 2-7 Supplements and their concentrations

Supplement	Concentration (μg/ml unless stated)
Ampicillin (Amp)	100
Chloramphenicol (Cm)	12.5 (liquid medium) 25 (solid medium)
Thiamphenicol (Tm)	15
Kanamycin (Kan)	100
Cycloserine (Cs)	250
Erythromiycin (Erm)	50
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	1000mM stock
Folic Acid	10% w/v

2.5.3 General culture and Storage conditions

All strains unless otherwise stated were grown at 37 ^{o}C. Anaerobic bacteria were grown in a MG1000 Mark II anaerobic cabinet from Don Whitely Scientific limited. The gas mix supplied to the anaerobic cabinet was 80% N_2, 10% CO_2, and 10% H_2. When growing aerobic bacteria the cultures were shaken at 200rpm while being maintained at 37 ^{o}C.

Strains were stored in sterile culture media with 10% glycerol. Once grown the cultures are kept at -80 ^{o}C.

All *Clostridium sporogenes* wild type were grown in TYG broth and agar. Wild type was grown in the presence of D-cycloserine at a working concentration of 250 μg/ml, to eliminate the possibility of contamination. Media was further supplemented with antibiotics for plasmid selection when required. For phenotypic selection of *C. sporogenes*, amino acid complete medium (MACC) media was used with appropriate supplements. *Clostridium acetobutylicum* 824 was grown on 2xYT with the relevant antibiotics for the selection of plasmids in correctly transformed bacteria.

For *Escherichia coli,* all strains were grown on Luria-Bertani (LB) media. Antibiotics were added to the media for plasmid selection, and for phenotypic selection the appropriate supplements were added to LB.

2.5.4 Growth Curves

To assess growth of the different strains, optical density of the liquid cultures was measured. All strains were grown in the optimal growth media for the species, with relevant antibiotic selection for the plasmid. Measurements were taken either in a Pharmacia Novaspec II or a GloMax® Multi+ Microplate Multimode reader. Measurements were taken at λ = 600 nm. To begin the growth curve, sterile media was inoculated from a culture which had been grown overnight from a single picked colony. Quantity of culture used to inoculate was worked so that the start point was an OD_{600} of 0.05. Measurements were taken every hour for 24 hours, against a blank measurement of sterile media.

2.5.5 Electrocompetant cells

For transformation of *E. coli* TOP10, *E. coli* CA434 and *E. coli* BL21 (DE3) Star electrocompetant cells were produced.

200 ml of LB broth was pre-warmed to 37 ^{o}C and then inoculated with 1 ml of an overnight culture of *E.* coli from the frozen stocks. Cells were incubated at 37 ^{o}C and kept with constant rotation at 200 rpm. The growth was measured after 3 hours and if the OD_{600} showed a reading of 0.5- 0.7, the cells were harvested through centrifugation at 3700 x *g* for 10 min at 4 ^{o}C. Pellets were kept chilled and the supernatant was discarded. The pellets were resuspended in ½ the volume of ice cold water as was previously in the cultures media, and then centrifuged under the same conditions as the harvesting. This step was repeated twice more using ½ the volume of ice cold dH_2O each time (from 40 ml to 20 ml to 10 ml). After the washing

steps the pellets were resuspended 1 ml of 1mM 3-(N-morpholino)propansulfic acid (MOPS) with 10% v/v glycerol.

Cells were stored in 50 µl aliquots at -80 ^{o}C.

2.5.6 Transformation of *E. coli*

To transform *E. coli* 50 µl of prepared electrocompetant cells were thawed on ice for a maximum of 10 min. An electroporation cuvette which was chilled to -20 ^{o}C was prepared by aliquoting 10 µl of ligated plasmid or 2 µl of purified plasmid and the thawed cells into the cuvette. The cuvette containing the cell and plasmid mix had a pulse of 2250 V (22.5 kV/cm, 200 Ω, 25 µF) applied. The cuvette had 400 µl of SOC media added after the pulse had been applied. Cells were given a recovery time of 60 min at 37 ^{o}C in sterile Eppendorf tubes. To select for correct transformations the recovered cells were plated onto antibiotic plates at three different dilutions (Neat, 1:10 and 1:100) to allow picking of single colonies.

2.5.7 Electrocompetant *C. acetobutylicum*

A loop of the *C. Sporogenes* was inoculated into 10 ml of the appropriate 2xYT media, which was then vortexed and 10 ml volumes of media were serially diluted at 1:10, 1:100 and 1:1000 concentrations from the original inoculum. The lowest concentration of dilution that grows overnight was used to inoculate 60 ml of appropriate media. The broth was grown at 37 ^{o}C until the culture reaches 0.9-1.1 OD when measured at 600nm. The media is divided into 15 ml aliquots in 30 ml falcon tubes, and the cells are harvested by centrifugation at 4 ^{o}C for 10 min at 2600 x *g*. The supernatant was discarded and then the pellets were resuspended into 2.5 ml of ice cold electroporation buffer (EPB). EPB contains 5mM of sodium phosphate (pH 7.4) and 270mM sucrose. After resuspension four pellets are combined, and then the cells present are harvested as before. The resulting supernatant is discarded

and the pellets are resuspended in 2.3 ml of the EPB, which will then be immediately used for transformation.

2.5.8 Preparation of methylated plasmid for *C. acetobutylicum* transformation

Before the plasmid is able to be transformed into *C. acetobutylicum* it needs to be methylated through transformation into *E. coli* TOP10 containing the pAN-2 plasmid. This plasmid has the Φ3T0049 methyltransferase gene which will protect the plasmid form the Cac 8241 DNA restriction system which is present in *C. acetobutylicum*. The pAN-2 plasmid methylates the plasmid which is present in the *E. coli* TOP10 with it, allowing it to be inserted into *C. acetobutylicum* without being affected by the Cac 8241 system (Heap et al., 2007).

2.5.9 Transformation of *C. acetobutylicum*

Once methylated as described in 2.5.9, plasmids were transformed into *C. acetobutylicum*. In a 0.4cm electroporation cuvette, 20 µl of methylated plasmid was added to 570 µl of fresh electrocompetent *C. acetobutylicum* cells. Within the anaerobic cabinet the cuvette was dried and inserted into the electroporation chamber. Under the conditions of 2.0kV, 25 µF and ∞Ω, a pulse of electricity was applied. Immediately 1 ml of pre-warmed, pre-reduced 2xYT media was added, and then samples were incubated for 1-3 h for recovery. Cells were centrifuged at room temperature, at 5000 x *g* for 3 min, and then resuspended in 1 ml of fresh pre-warmed, pre-reduced 2xYT media. To select for correctly transformed *C. acetobutylicum,* bacterial cells were plated on 2xYT agar with an appropriate antibiotic.

2.5.10 Electrocompetant *C. sporogenes*.

Using a full loop of bacteria 15 ml of TYG media was inoculated and left to grow overnight. Inoculate 200 ml of TYG media which has been reduced and warmed in the anaerobic cabinet, with 4 ml of the overnight culture. This can then be grown for up to 3 h until it reaches an OD_{600} of about 0.2-0.3. At this point it can be aliquoted into 50 ml falcon tubes and kept on ice. The aliquoted cells can be centrifuged down at 4000 x *g* for 10 min at 4 ^{o}C. The supernatant can be discarded and the pellets can be resuspended in 20 ml of 0.5M sucrose. Once resuspended, the cells can then be centrifuged again under the same conditions as before. The pellet then can be resuspended in 20 ml 0.5M sucrose with 10% glycerol. Again the resuspended cells were collected through centrifugation at the same conditions as described. After resuspending the cells in 20 ml of 0.5M sucrose with 10% glycerol, the different tubes could be combined and then centrifuged down under the same conditions. The resulting pellets can then be resuspended in 2.5 ml of 0.5M sucrose with 10% glycerol and used immediately in transformation

2.5.11 Transformation of *C. sporogenes*

C. sporogenes does not have a restriction system so no methylation is required for transformation to be successful. Into a 0.2cm electroporation cuvette add 2 µg of plasmid, then add 280 µl of the electrocompetant cells, and mix gently. Ensure the cuvette is dry and add place in the electroporator and pulse with 750-1250V, 200Ω and 25 µF. Immediately after the electroporation, add 3 ml of TYG, to the cells, and allow to recover for 3-4 h. The cells can then be centrifuged down at 4000 x *g* at room temperature for 10 min. Discard the supernatant and resuspend the cell pellets in 1 ml of PBS. The cell mix can then be plated onto TYG plates containing an appropriate antibiotic.

2.5.12 Conjugation of *C. sporogenes*

Conjugation of *C. sporogenes* NCIMB10696 was performed by first transforming the required plasmid into *E. coli* CA434, a donor strain, and selecting for it on antibiotic resistance plates. Overnight cultures of a correct plasmid containing *E. coli* CA434 were grown, and 1 ml samples were centrifuged for 1 min at 2400 x *g*. The pellet was washed in 1 ml of PBS solution and then harvested again in the same way. The pellet was resuspended in 200 µl of *C. sporogenes* NCIMB10696 culture, and the resulting mixture was spotted onto TYG plates without antibiotics. These were allowed to grow for 6-7 hours in an anaerobic cabinet. Conjugated cells were harvested from the plates and resuspended into 500 µl of sterile anaerobic PBS solution. This was then plated out at three different volumes (to allow for single colonies) on TYG media plates with antibiotic supplements for the plasmid and D-cycloserine for selection of only *C. sporogenes* with the correct plasmid.

2.5.13 Fractionation of *C. sporogenes*

Fractionation was performed primarily as described by Nobbs *et al* (Nobbs et al., 2007), but with some modifications. *C. sporogenes* is grown in 10 ml of TYG media for 24 h, after this it is centrifuged down at 3000 x *g* for 20 min. The supernatant is concentrated through ultrafiltration (performed by applying the extracellular fraction to a Polyethersulfone (PES) filter in a column, and then centrifuged until concentrated to an appropriate amount e.g. 10x concentrated) and designated the extracellular fraction. The cell pellets were resuspended in 10 ml of spheroplasting buffer. The spheroplasting buffer is 20mM Tris/HCl pH6.8, 10mM $MgCl_2$ and 26% (w/v) raffinose, with mutanolysin added to a final concentration of 80 U/ml. Once resuspended, the cell suspension was incubated in a rotary shaker for 1 h at 37 ^{o}C. This was then chilled on ice and centrifuged at 12,000 x *g* for 20 min at 4 ^{o}C, the

supernatant was concentrated and designated the cell wall fraction. The resulting pellet was resuspended in 10 ml of 50mM triethanolamine with 250mM Sucrose and then lysed through sonication at 50W for 8 min. This was then centrifuged for 20 min at 12,000 x *g* and at 4 ^{o}C. The supernatant was concentrated through ultrafiltration and designated the protoplast fraction. Alternatively the fractionation protocol was performed as described by Tulli *et al* (Tulli et al., 2013)

2.6 Molecular Biology techniques

2.6.1 Genomic DNA Extraction

2.6.1.1 Kit extraction of genomic DNA.

When performing genomic extraction using the Sigma Aldrich GenElute™ Bacterial Genomic DNA extraction kit, the protocol performed was dependent upon whether the DNA was from a Gram negative or a Gram positive, however this kit was only used for the extraction of Gram positive DNA.

Before beginning a Lysozyme solution was made where 10 mg/ml was dissolved into the Sigma Aldrich supplied Gram Positive Lysis Buffer. The columns that were supplied in the kit were prepared through the addition of 500 µl of Column Preparation Solution to each buffer. The columns were then centrifuged at 15900 x *g* and the flow through discarded. Wash Concentrate was diluted by adding 10ml of 100% Ethanol for every 2.5 ml of Wash Concentrate.

At the same time 4 ml of bacterial culture was pelleted at 6000 x *g* for 3 min. The resulting pellet was resuspended in the premade Lysozyme solution and incubated at 37 ^{o}C for 30 min. After this 20 µl of RNase A was added and incubated at RT for 2 min. Once the incubation was finished 200 µl of the supplied Lysis Solution C and 20 µl of Proteinase K was added to the bacterial cells, and the solution was incubated at

55 °C for 10 min. When the incubation was finished 200 µl of 100% Ethanol was added, and then the entire solution was applied to the pre-prepared columns. This was then centrifuged at 15900 x *g* for 1 min. The flow through was discarded, and the column was washed through the addition of 500 µl of Wash Solution 1. Again the flow through was discarded, and then 500 µl of the dilute Wash Solution Concentrate was added. This was then centrifuged for 1 min at 15900 x *g*, and the flow through was discarded. Without adding anything to the column the spin step was repeated, and then the column was added to a 1.5 ml Eppendorf tube. To the column 55 µl of the Elution buffer was added, and then centrifuged at 15900 x *g* for 1 min. The flow through was then added to the column and the column was centrifuged for a second time. The resulting flow through was then used as the genomic DNA extract.

The Qiagen QIAamp DNA Mini kit was also used only for the isolation of Gram positive bacterial DNA. Pre-use the buffers AW1 and AW2 were diluted with an appropriate volume of 100% ethanol. Bacterial cells were pelleted as described for the Sigma Aldrich genomic DNA extraction kit, and then resuspended in a 10 mg/ml Lysozyme to dH_2O solution. This solution was then incubated at 37 °C for 30 min, and once finished 20 µl of proteinase K and 200 µl of the supplied Buffer AL which was then mixed through vortexing. This was then incubated for 30 min at 56 °C and then for 15 min at 95 °C and then 200 µl of 100% ethanol was added to the mix. This mix was added to the Qiagen supplied columns and then was centrifuged at 6000 x g for 1 min, the flow through was then discarded. To the column 500 µl of the Buffer AW1 was added and then the column was centrifuged again at 6000 x g for 1 min, again the flow through was discarded. After this 500 µl of Buffer AW2 was added to the column and then the column was centrifuged for 3 min at 15900 x *g*. The flow through was discarded and then the column was centrifuged again at full speed for

an additional 1 min. After this the column was inserted into an Eppendorf 1.5 ml tube and then 55 µl of the Buffer AE was applied and then the column was incubated for 1 min at RT. It was then centrifuged at 15900 x *g* and the flow through was collected and reapplied to the column. This was then centrifuged as before and the flow through was used for the genomic DNA.

2.6.1.2 Phenol:Chloroform extraction of genomic DNA

Extraction of higher concentrations of genomic DNA was performed through the Phenol: Chloroform: Isoamyl (25: 24: 1) protocol (Stephen Cartman, CRG). A 5 ml overnight culture of the bacteria was grown, and the cells harvested through centrifugation for 3 min at 6000 x *g*. The pelleted cells were resuspended in 180 µl lysis buffer (10 mg/ml lysozyme in PBS) which was then incubated at 37 ^{o}C and gently agitated periodically. After, 25 µl of Proteinase K solution (Qiagen) was added as well as 85 µl of dH_2O, and 110 µl of 10% Sodium dodecyl sulphate (SDS) w/v solution. This was mixed through inversion, and then incubated at 65 ^{o}C for 30 min. After this incubation step 400 µl of Phenol: Isoamyl: Chloroform alcohol (pH 8.0) was added and mixed thoroughly. The entire solution was added to phase lock tubes, and centrifuged for 3 min at 15900 x *g*, the top layer was then transferred to a fresh phase lock tube and centrifuged in the same manner. This step was repeated a further 2 times with Phenol: Isoamyl: Chloroform alcohol. After the final spin in a phase lock tube the top layer was transferred to a sterile Eppendorf tube which contained 40 µl of 3M NaAc and 800 µl of ice cold 100% EtOH. The samples were gently mixed and incubated at -80 ^{o}C for 30 min, and the resulting mix was centrifuged down for 15 min at 15900 x *g*, after which the resulting pellet was washed in 1 ml of 70% EtOH, and centrifuged down. This wash step was repeated, after which the final pellet was allowed to dry at room temperature. The final dried

pellet was resuspended in 50 µl of dH_2O, ready for quantification or other downstream processes.

2.6.2 Plasmid Extraction

Plasmid extraction was performed using either the Qiagen QIAprep miniprep kit or the Invitrogen HP miniprep kit. When being performed on *C. sporogenes* freeze thaw cycles were performed prior to performing the kit protocol, whereby the cell pellets were incubated at -80 ^oC for 2 min, and then at 37 ^oC for 2 min. This cycle is repeated a further two times.

The GenElute™ Plasmid Miniprep Kit was used for the extraction of plasmids from both Gram positive and Gram negative bacteria. Prior to use 500 µl of RNAase A was added to 100 ml of resuspension solution, and 300 ml of 100% ethanol was added to 75 ml of Wash Solution Concentrate. To extract the plasmid using this kit 4 ml of bacterial culture was pelleted through centrifugation at 6000 x *g* for 3 min. The pelleted cells were kept and the supernatant was discarded. The pellet was resuspended in 200 µl of Resuspension Solution, and then 200 µl of the lysis solution was added. This was mixed gently through inversion, and then 350 µl of Neutralization Solution was added. Again this was mixed gently through inversion, and the resulting mix was centrifuged at 15900 x *g* for 10 min. While the mix was centrifuged, 500 µl of column preparation buffer was added to the columns, and then centrifuged for 1 min at 15900 x *g*. The flow through was then discarded, and the supernatant of the mix which had been centrifuged for 10 min was applied to the column. This was then centrifuged at 15900 x *g* for 1 min and the flow through was discarded. The column was then washed through the application of 500 µl of the Optional Wash Solution to the column which was then centrifuged for 1 min at 15900 x *g*. Finally it was washed through the application 750 µl of the Wash Solution,

and the column was centrifuged for 1 min at 15900 x *g*. The flow through was discarded and the column was centrifuged again at 15900 x *g* for 1 min to remove any residual Wash Solution. The column was then inserted into a 1.5 ml Eppendorf tube, and 50 µl of the Elution solution was added. This was centrifuged for 1 min at 15900 x *g*, and the flow through was re applied to the column and centrifuged at 15900 x *g* for 1 min. The resulting flow through was the extracted plasmid.

Prior to use of the Qiagen QIAprep Spin Miniprep Kit, LyseBlue reagent was added to the buffer P1 in a 1 in 1000 ratio, to the buffer P1 1 mg of RNAase A was added per 10 ml of Buffer was also added. Buffer PE was diluted to the recommended amount with 100% ethanol. Once this was done 4 ml of bacterial culture was pelleted through centrifugation at 6000 x *g* for 3 min. The supernatant was discarded and the pellet was resuspended in Buffer P1, and then an equal amount of Buffer P2 was added. The solution with LyseBlue turned blue at this point, and then 350 µl of Buffer N3 was added, and the blue colour will disappear. At this point the solution was centrifuged for 10 min at 15900 x *g*, and the precipitate was pelleted. The supernatant was then added to the provided columns, and the pelleted precipitate was thrown away. The column was then centrifuged for 1 min at 15900 x *g*, and the flow through was discarded. Then the column was washed by adding 500 µl of Buffer PB, and then centrifuging the column for 1 min at 15900 x *g*. The flow through was discarded, and then 750 µl of Buffer PE was added and the column was centrifuged for 1 min at 15900 x *g*. The flow through was discarded and the column was centrifuged to remove residual wash buffer at 15900 x *g* for 1 min. The column was then inserted into a 1.5 ml Eppendorf tube and 55 µl of Elution Buffer was added. The column was then centrifuged for 1 min at 15900 x *g*, and the flow through was

collected and then added to the column and centrifuged again in the same manner. The resulting flow through is the extracted plasmid.

Storage conditions: short term at 4 ^{o}C and long term at -20 ^{o}C.

2.6.3 Amplification by PCR.

Amplification of all genes and sequences were performed through Phusion PCR, Failsafe PCR system with premix buffer E, and for extended sequence lengths Longamp PCR was performed. All PCR thermocycles were performed in either the Applied Biosystems Gene Amp PCR system 2400, or the Biometra TProfessional TRIO.

2.6.3.1 Phusion PCR

When Phusion High-Fidelity PCR was performed a reaction mix was made where in a 50 µl reaction mix 0.5 µl of the Phusion High-Fidelity DNA polymerase was added to 10 µl of 5x Phusion HF buffer (for a final concentration of 1x), 1 µl of 10mM dNTP's, 2.5 µl of both forward and reverse Primer at a 10 µM concentration, and 2 µl of template (DNA ensuring that the quantity of DNA was above 250ng). To this mix dH_2O was added to a final volume of 50 µl. With this reaction mix a PCR thermocycler was set up where there was an initial Denaturation step of 98 ^{o}C for 30 sec. This was followed by 35 repeated cycles of a secondary denaturation step for 10 sec at 98 ^{o}C, then an annealing step for 30 sec at 5 ^{o}C below the T_m of the primers in use. The cycle was finished with an extension step of 72 ^{o}C for 30 sec per kb of expected product. The PCR was finished with a final extension step of 10 min at 72 ^{o}C.

2.6.3.2 Failsafe PCR

Failsafe PCR was performed through the mix of 2.5 µl of each primer 2 µl of the DNA template and 1 µl of the Failsafe PCR Enzyme Mix. This was mixed with water to a final volume of 25 µl. These 25 µl were then added to an equal amount Failsafe 2x

PreMix E making a PCR solution of 50 µl. PCR solutions were subjected to a thermocycle were initial denaturation was performed for 2 min at 98 ^{o}C, it was followed by repeated cycles of secondary denaturation at 95 ^{o}C for 30 sec, then an annealing step of 5 ^{o}C below the T_m of the primers used and then an extension step of 1 min for every kb expected to be produced at a temperature of 72 ^{o}C. This was repeated 40 times.

2.6.3.3 GoTaq PCR

For a 50 µl reaction of GoTaq polymerase 0.25 µl of the GoTaq DNA polymerase was added to a mix of 2.5 µl of both forward and reverse primer, 10 µl of the 5x Green GoTaq Reaction Buffer1 µl of 10mM Mixed dNTPs and finally 2 µl of the DNA template. This reaction mix was made up to 50 µl with dH_2O. This reaction mix then underwent a PCR thermocycle where there was an initial denaturation step for 2 min at 95 ^{o}C, then the reaction mix underwent a repeated cycle where there was a secondary denaturation step of 95 ^{o}C for 30 sec, then an annealing step which was performed at 5 ^{o}C less than the T_m of the primers for 30 sec, and then an extension step of 72 ^{o}C which was performed for 1 min per kb of expected product. This was repeated 35 times, and was followed by a final extension for 5 min at 72 ^{o}C.

2.6.3.4 LongAmp PCR

LongAmp PCR was performed by making a reaction mix of 10 µl of 5x LongAmp Taq reaction buffer, 1.5 µl of 10mM dNTPs, 2 µl of both forward and reverse primer when they are at a concentration of 10 µM, 2 µl of LongAmp Taq Polymerase and enough template DNA for a quantity less than 1,000ng. To this mix dH_2O was added to a final volume of 50 µl. The mix underwent a thermocycle of 94 ^{o}C for 30 sec for an initial denaturation and then it was followed by 30 cycles of a secondary

denaturation for 30 sec, an annealing step of 5 °C less than the T_m of the primer for 30 sec, and then 65 °C for 50 sec per kb of expected product for the extension step. A final extension step was performed after the repeated cycles were completed of 65 °C for 10 min.

After the final extension for each type of PCR the resulting products were stored short term at 4 °C and long term at -20 °C.

2.6.3.5 Bio-X-Act PCR

For Bio-X-act polymerase from Biozym, the reaction mix was 5 µl of 10x OptiBuffer, 2 µl of 50mM $MgCl_2$, 1 µl of 100mM dNTP, 10 µl of 5 ng/µl template DNA, 1 µl of 100 µM primer mix, 1 µl Bio-X-Act polymerase and then dH_2O to 50 µl. This reaction mix was then implemented in the PCR protocol of an initial 5 min 95 °C denaturation step, then 30 cycles of a secondary denaturation step at 95 °C for 30 sec, an annealing temp appropriate for the primers (5 °C less than the T_m of the primers) for 30 sec, and then a 1 min at 72 °C for every kb of product. A final extension step at 72 °C for 10 min was performed after the repeated cycles had finished.

2.6.3.6 Colony PCR

Colony PCR was performed where quick screening was required of the colonies which had been conjugated, modified or transformed. Phusion PCR was performed for colony PCR. A normal PCR reaction mix was performed, dependent on the polymerase used, with the template being replaced with a single picked colony resuspended in the reaction mix. To break down the cell walls and release the genomic or plasmid DNA an extra pre-step was added to the PCR thermocycle where the reaction mix was heated to 98 °C for 10 min before continuing with a standard PCR thermocycle protocol for Phusion PCR.

After the final extension for each type of PCR the resulting products were stored at -20°c

2.6.4 Restriction Endonuclease Digest

For restriction digests New England Biolabs restriction enzymes were used. For this a restriction mix was made through combining 1 µl of the enzyme or enzymes to be used, with 2 µl of appropriate buffer (Out of Buffer 1, 2, 3, and 4) or the CutSmart buffer and then adding 10 µl of the DNA to be digested. Water was added to a final volume of 20 µl, and incubated at 37 °C for at least 1hr.

2.6.5 Post Digest alteration of DNA

After digests were performed, sticky ends were blunted through the use of T4 DNA Polymerase. The DNA containing sticky ends had 100µM of dNTPs added, with 1 unit T4 DNA polymerase per microgram of DNA to be blunted was added. This was then incubated at 37 °C for 30 min and the T4 DNA polymerase was inactivated.

Dephosphorylation was performed through the addition of 2 µl of 10x Antarctic Phosphatase Reaction Buffer to 1-5 µg of DNA to be dephosphrylated. To this 1 µl of Antarctic Phosphatase enzyme was added. The volume was made up to 20 µl with dH_2O and then incubated at 37 °C for 15 min. The enzyme was then heat inactivated at 70 °C for 5 min.

2.6.6 Agarose Gel Electrophoresis

To separate out DNA fragments of different size, electrophoresis was performed on the different fragments. Agarose was melted in TAE at varying concentrations (0.8-1.5%) and then an electrical current (80-110v) is passed through the gel for a set amount of time. Ethidium Bromide (EtBr) or Sybr® Safe was added to the gels, for visualisation on ultraviolet or blue light

2.6.7 DNA Gel Extraction

Prior to use of Sigma Aldrich GenElute DNA gel extraction kit the wash solution was diluted by adding 4 ml of 100% Ethanol for every 1 ml of wash solution. After the gel was run and the DNA was visualized the DNA was extracted out of the gel through excision of the band that was shown. The band was weighed and 3 times the volume of the gel slice was added of Gel Solubilization Solution. This was then incubated at 50 ^{o}C for 10 min or for long enough for the gel slice to dissolve. To this mix isopropanol was added to the same volume as the gel slice. This was mixed until a homogenous solution was obtained. Column preparation buffer was added to the columns at a 500 µl volume, and the column was then centrifuged at 15900 x *g* for 1 min. The flow through was discarded and the homogenised DNA mix was applied to the column which was then centrifuged again at 15900 x *g*. The volume of mix added was limited to 700 µl. The flow through was discarded and any excess volume of DNA mix was applied to the column and centrifuged again in the same way. The column was then washed with 500 µl of the Gel Solubilization Solution which was passed through by centrifugation at 15900 x *g* for 1 min, the flow through was discarded. The column was then washed with 600 µl the diluted wash solution, which was added to the column and then centrifuged for 1 min at 15900 x *g*. The flow through was discarded and then the column was centrifuged again at 15900 x *g* for 1 min to remove any residual wash solution. After the column was added to an Eppendorf tube and 55 µl of elution buffer was added to the column, and then centrifuged for 1 min at 15900 x *g*. The flow through was reapplied and then the spin was repeated the flow through was kept as the extracted DNA.

Like the GenElute kit prior to using the Qiagen QIAquick Gel Extraction kit, buffer PE must be diluted with 5 parts 100% ethanol to 1 part buffer PE. The gel band is again

excised and weighed, and then the gel band is melted in 3 volumes of buffer QG by heating to 50 °C for 10 min or until the band is dissolved. To this mix 1 volume of isopropanol was added and then the mix was homogenised. The mix was then applied to a QIAquick gel extraction column. It was centrifuged for 15900 x *g* for 1 min, and the flow through was discarded. Any DNA mix which was in excess of 700 µl was kept while the column was centrifuged and then applied to the column after the flow through was discarded. This was repeated until all DNA mix was passed through the column. Once this was performed, 500 µl of buffer QG was applied to the column to remove all excess gel. It was centrifuged for 1 min at 15900 x *g*, the flow through was discarded and then 750 µl of diluted buffer PE was added to the column. The column was then centrifuged for 1 min at 15900 x *g*, the flow through was consequently discarded. The column was centrifuged again under the same circumstances, and was then inserted into a clean Eppendorf 1.5 ml tube. Buffer EB was then added to the column in 55 µl volume. This was then centrifuged at 15900 x *g* for 1 min and the flow through was reapplied to the column and the spin was repeated. The resulting flow through was kept as the extracted DNA.

2.6.8 Purification of PCR Products.

To use the Sigma Aldrich GenElute PCR purification kit first the wash solution concentrate was diluted through the addition of 4 ml of 100% ethanol to 1 ml of Wash Solution Concentrate. Columns were prepared through the addition of 500 µl of Column Preparation Solution and then spinning the columns for 1 min at 13000rpm. The flow through was then discarded. To purify the PCR products from the reaction 5 times the volume of reaction mix in Binding Buffer was added. This was mixed, and then the solution was applied to the pre-prepared columns, after which it was centrifuged at 13000rpm for 1 min. The flow through was discarded and

then the column was washed through the addition of 500 µl of diluted wash solution to the column. This was then centrifuged at 13000rpm for 1 min and the flow through was discarded. The column was then centrifuged again to remove residual wash buffer for 1 min at 13000rpm. This was then discarded and the column was inserted into a 1.5 ml eppendorf tube. The column then had 55 µl of the Elution solution added to it and it was centrifuged for 1 min at 13000rpm. The flow through that resulted from this was then added to the column and centrifuged a second time. What resulted was the purified PCR product.

To use the Qiagen QIAquick PCR purification, first the wash buffer PE must be diluted by 4 parts 100% ethanol for every 1 part buffer PE wash concentrate. After dilution of the wash buffer PE, 5 times the volume of the PCR mix was added in buffer PB to the PCR reaction mix. The mix was then applied to the QIAquick columns and centrifuged at 13000rpm for 1 min. If the mix was more than the column could contain then this step was repeated until all the PCR mix was passed through the column. All flow through was discarded, and the column was subsequently washed. For the washing step 750 µl of diluted Buffer PE was added to the column and then centrifuged at 13000rpm for 1 min. The flow through was discarded and then the column was centrifuged again under the same conditions to remove residual buffer. The column was then placed in a 1.5 ml eppendorf tube and 55 µl of buffer EB was applied to the column. This was then centrifuged at 13000rpm for 1 min and the flow through was collected and reapplied to the column. The column was then re-centrifuged under the same conditions as before and the liquid collected as the purified PCR.

2.6.9 Ligation and Dialysis.

When preparing a plasmid for transformation ligation of the insert to the backbone of the vector needs to be performed. For this Promega T4 DNA ligase was used. When ligating the insert to the backbone the reaction mix can deviate from manufacturer's instructions. When required the reaction mix can be performed with a 6:2 volume ration of insert: backbone rather than a 6:2 molar ratio as stated in manufacturer's instructions. The reaction mix was made up of 1 µl T4 DNA Ligase, 2 µl 10x Ligase Buffer, the 6:2 ratio of insert to backbone and then the mixture was made up to 20 µl with dH_2O. The remainder of the mix is as instructed with all remainder being filled by dH_2O. Ligations were routinely refrigerated at 4 ^oC overnight.

Once incubated the mix was dialysed on the Merck Millipore Dialysis Disks where it was left for 45 min for the salt to be removed before the plasmid was transformed into electrocompetant cells.

2.6.10 Gibson Assembly.

When cloning was either inappropriate (i.e. too many fragments) or ineffective, Gibson assembly was attempted. Gibson assembly was performed by first amplifying all possible fragments through PCR. The primers were designed so that at both 5' and 3' ends of the DNA, overlapping sequences were present between the fragments allowing for precise order of the fragments. Then the fragments were used in the NEB Gibson assembly kit in the following reaction mix: 10 µl of Gibson Assembly Master Mix (2X) was added to a 6:2 insert to backbone mix. Where more than 2 fragments were used the ratio was 3:3:1 for 2 smaller fragments to a large fragment. The mix was then made up to a 20 µl volume using dH_2O. This was then incubated at

50 °C for 15 min minimum. Plasmids that resulted were dialysed and transformed into electrocompetant cells.

2.6.11 Nucleotide Sequencing

When required nucleotide sequencing was performed by Source BioScience. To do this DNA concentration was measured on the Thermo Scientific NanoDrop 1000, and with the correct concentration (100 ng/µl for plasmids, and 1 ng/µl per 100bp for PCR products), 5 µl of the DNA to be sequenced was sent per reaction.

2.6.12 Oligomelts

Oligomelts were performed through resuspending lyophilised oligos to a concentration of 100 µM. From the 100 µM stocks 0.5 µl of the two complementary oligos were mixed together and then diluted 1000 fold. This reaction mix was heated at 95 °C for 20 min. The mix was allowed to cool, and then the resulting double stranded oligos could be used for ligation.

2.7 RNA

2.7.1 RNA Isolation

For RNA isolation bacterial cultures were grown and 10 ml samples were taken at 4 h, 8 h, and 24 h. Samples were inoculated with 1:100 overnight culture to fresh, reduced broth. The overnight culture was inoculated with a loop full of bacteria from plates with correct antibiotic resistance. Once samples were taken 2 volumes of Qiagens RNAprotect were added, and incubated for 5 min at room temperature. Consequently the bacteria were centrifuged at 2600 x *g* for 15 min at 4 °C.

After preparing the cells with RNA protect, RNA extraction was performed. RNA pro was used and 1 ml was used to resuspend the cell pellets. The mixture was then aliquoted into a matrix tube, and homogenisation was performed for 45 sec at

6,400rpm in the Precellys® 24 homogenizer. The homogenised mix was centrifuged at 4 °C for 15 min at 15900 x *g*. The solution was transferred to an RNA free tube, then after 5 min of room temperature incubation 300 µl of cold chloroform was added and mixed by vortexing. Again it was incubated at room temperature for 5 min, after this incubation there was another centrifugation step performed at 15900 x *g* for 15 min at 4 °C. Subsequently the top layer was aliquoted into 500 µl of 100% ethanol, and incubated at -20 °C for a minimum of 30 min, but usually for 2 h. Subsequently the incubated mix is centrifuged again at 4 °C for 15 min at 15900 x *g*, and the resultant supernatant was discarded. The pellet was resuspended in 500 µl of 75% ethanol and then it was centrifuged as before. Once again the pellet was kept and the supernatant was discarded, and the pellet was allowed to air dry. The dry pellet was resuspended in DEPC treated water and stored at -80 °C.

2.7.2 RNA purification

To purify the extracted RNA Turbo DNase was used 5.5 µl of the 10x Turbo DNase buffer and 2 µl of Turbo DNase was added to each sample. The samples were then left for a 30 min incubation step at room temperature. The Turbo DNase process was repeated after the initial attempt.

Once the second incubation is complete the Qiagen RNeasy kit can be used to complete the clean-up and purification of samples. To the RTL buffer from the kit 10 µl/ml of β-mercaptoethanol was added, 350 µl of this solution was added to each sample. In addition to this 200 µl of 100% ethanol was added to the samples and then the solution was applied to the RNeasy columns. The column was centrifuged for 9400 x *g* for 15 sec, and the flow through was discarded. Further to this 500 µl of the RPE buffer was added to the column to wash, and then centrifuged once again as before. Once more 500 µl of RPE was added and then it was centrifuged at 9400 x *g*

for 2 min. Now washed, the samples were eluted from the column by 3 cycles of 1 min spins at 9400 x *g* with 17 µl of DEPC water, until a total volume of 50 µl was collected. Before each spin the column is incubated for 3 min with the DEPC water at room temperature. After all spins were completed the RNA was stored at -80 ^{o}C.

2.7.3 Reverse Transcriptase PCR

Reverse Transcriptase PCR was performed using the Qiagen OneStep RT-PCR Kit. After thawing the different components, 10 µl of the 5x Qiagen OneStep RT-PCR buffer, 2 µl of 10mM dNTP mix (for a final concentration of 400 µM of each dNTP), forward and reverse primer (to a final concentration of 0.6 µM), 2 µl of Qiagen OneStep RT-PCR enzyme mix, 5-10 units of RNase Inhibitor, and template RNA (to a final concentration of 2 µg) were added. This mix was then made up to 50 µl for a single reaction mix with RNase- free water. This was mixed thoroughly and then put in the thermocycler used for PCR (see 2.6.3). A thermocycle of 30 min at 50 ^{o}C for reverse transcription was undertook, followed by a 15 min 95 ^{o}C initial PCR activation step. These were followed by a repeated cycle of 1 min at 94 ^{o}C for denaturation, 1 min at 5 ^{o}C below the T_m of the primers for annealing and then extension for 1 min at 72 ^{o}C for a 1-2kb product (for larger products 1 min per kb at 68 ^{o}C can be used). The cycle was repeated 35 times and the PCR was finished with a 10 min 72 ^{o}C final extension step. Products were then run on an agarose gel.

2.8 Protein analysis

2.8.1 Preparation of cell lysates

For both *E. coli* and *C. sporogenes* frozen -80 ^{o}C stocks were revived in overnight cultures, with appropriate media and antibiotic selective supplements to maintain plasmids. The overnight cultures were then used to inoculate more appropriate media with selective antibiotics in a 1:100 v/v ratio. Samples were taken at certain

defined time points (e.g. exponential growth phase). 1 ml samples were taken of the cultures for treatment and the pellets were harvested through centrifugation at 9400 x *g* for 10 min at 4 °C. Pellets could be stored at -20 °C.

The pellets were lysed with Novagen® BugBuster Protein Extraction Reagent, using the protocol recommended by Novagen®. The soluble fraction obtained by this protocol was stored at -20 °C, and used in further experimentation.

2.8.2 Protein precipitation from fluids

TCA Protein precipitation was used to purify proteins out of the supernatant for western blot analysis. Samples were taken in the same way as described for preparation of cell lysates. Once centrifuged at 9400 x *g* for 10 min at 4 °C the supernatant was obtained. To the supernatant 1 volume of 100% trichloracetic acid (TCA) was added to the supernatant and was incubated on ice for 30 min. After this incubation period the samples were centrifuged down at 9400 x *g* for 5 min at 4 °C. The supernatant was discarded and the resulting pellet was washed in ice cold acetone. This step was repeated a further 2 times. The final pellet was incubated in a sonicating water bath for 20 min. The pellet is then allowed to air dry for 10 min to remove remaining acetone. Once air dried the sample was resuspended in 0.1M TrisHCl buffer, pH 7.3

2.8.3 Preparation of supernatant

Ultrafiltration was used to concentrate supernatant through corning ultrafiltration columns. The columns PES molecular weight cut off was picked based upon the protein that was to be concentrated. A 50k MWCO to concentrate CPG_2 was picked and supernatant from cultures was passed through by centrifugation. The flow through was discarded and the concentrated protein remnants stored at -20 °C

2.8.4 SDS-PAGE

Samples were prepared for SDS-PAGE by boiling 12 µl of sample containing 1X NuPAGE LDS Loading Buffer (Life Technologies) and a final solution of 0.4M Dithiothreitol (DTT) for 5 min at 80 ^{o}C. ColorPlus protein ladder was also boiled for 5 min at 80 ^{o}C, however it was not mixed with the loading buffer and DTT. Samples were loaded onto a 4-12% Bis-Tris Protein Gel from Novagen. The gels were run for 1 h and 30 min at 150v or until the loading dye reached the bottom of the gel.

2.8.5 2D SDS-PAGE

2D SDS-PAGE was performed using the Thermo-Fisher Scientific Zoom® IPGRunner™ system. This was modified to be appropriate for the samples being looked at. Sample solutions were made up as shown in the protocol, however there were some slight alterations made. Unlike the user protocol there is no pre-treatment of the sample, as the pre-treatments was for lysis of cells. Supernatant samples were taken and then concentrated through ultrafiltration, this concentrated sample was used in the solution shown in Table 2-8 for samples.

Once samples solutions were made, they were applied to dehydrated broad range (pH3-10) ZOOM® strips in the IPGRunner™ Mini-Cell. The Mini-Cell was sealed, and then incubated at room temperature for 1 h to rehydrate the isoelectric strips. The Mini-Cell was then used in the ZOOM® IPGRunner™ for isoelectric focussing. Proper connections of the cathode and anode of the ZOOM® IPGRunner™ to the mini-cell were ensured and then deionised water was used for the heat sink. A variable voltage program was then applied for successful focussing, where 200V was applied for 20 min, 450V for 15 min, 750V for 15 min, and then finally 1000V for 45 min. This was performed on the Bio Rad PowerPac 1000.

After focussing the strips were removed ZOOM® IPGRunner™, equilibriation was performed before running the second dimension. The strips were incubated for 15 min in the reducing solution and then 15 min with the alkylating solution. Once equilibrated the strip could be inserted into the precast 4-12% Bis-Tris ZOOM® Gels, with the wellbeing filled with the ZOOM® strip and then 0.5% melted, 50 ^{o}C agarose in the running buffer which is to be used. Once set the gel can be run as described in the SDS-PAGE section

Table 2-8 The solutions needed for 2d SDS-PAGE

Buffer	Buffer constituents	Quantity
Protein Solution	1X ZOOM® 2D Protein Solubilizer 1 or 2	128 µL
	Lysate	10–12 µL (0.25–1.0%)
	2 M DTT	0.7 µL
	Appropriate ZOOM® Carrier Ampholytes (v/v)	0.4–1.6 µL
	Bromophenol Blue	Trace
	Deionized water	to 140 µL
Reducing Buffer	4X NuPAGE® LDS Sample Buffer	2.25ml
	NuPAGE® Sample Reducing Agent (10X)	1.0ml
	Deionized water	6.75ml
Alkylating Buffer	4X NuPAGE® LDS Sample Buffer	2.5ml
	NuPAGE® Sample Reducing Agent (10X)	7.5ml
	Iodoacetamide	232mg

2.8.6 SDS-PAGE gel staining.

If needed the protein gels were stained using colloidal coomassie stain. The protein gel is fixed in a fixing solution consisting 10% Acetic acid, 40% dH_2O, and 50% ethanol, this is performed for 1-24 h. After the fixing process, the gel is stained using the staining solution which consists 0.075% (w/v) Coomassie Brilliant Blue G250, 1.5% (v/v) o-Phosphoric Acid, 7.5% (w/v) Ammonium sulphate, 25% (v/v) Methanol, the remaining amount is made up of dH_2O. The protein gel is then shaken in the staining solution for 24-48 h, then destained with dH_2O. Alternatively SimplyBlue™ SafeStain was used according to manufacturer's protocol.

2.8.7 Western Blots.

Western Blots were either chemoluminescence or colourimetric. Both western blots were performed in the same way with differing development methods. To perform the western blot, first an SDS-PAGE must be performed as described, without staining of the gel. Once the gel has completely run, proteins were transferred onto a nitrocellulose membrane, using either the Trans -Blot® Turbo™ Blotting system by BioRad, or the XCell II™ Blot Module by Invitrogen. The BioRad Trans-Blot® Turbo™ system was performed using the Trans-Blot Transfer pack, as instructed, and then underwent the Pre-programmed Mixed MW programme. This transferred the proteins from the pre-run gel to the nitrocellulose membrane in the Trans-Blot Transfer pack. The XCell II™ Blot was performed with the membrane, and then the gel, being sandwiched by blotting paper and sponge either side of the gel. These different layers are compressed in the XCell II™ Blot module, by locking the anode and cathode side of the XCell II™ together in a gel tank. Then the module was filled with 1x NuPage Transfer buffer supplied by Invitrogen. A low volt (30v) current was passed through the module for 1 h.

Once transferred the nitrocellulose membrane is blocked using 1x Tris Buffered Saline with 3% powdered milk. A 10x TBS stock was made of 88g of NaCl, and 60g of Tris buffer, dissolved in 1l of water. The solution is then adjusted to pH8 for a 10x concentrated TBS solution. The blocking process is done for 1 h, where the nitrocellulose membrane is immersed in 25 ml of the 1x TBS with 3% powdered milk and shaken. After 1 h, the 1x TBS with 3% powdered milk is discarded and 25 ml of the same solution this time containing antibody for the protein being looked for, at an appropriate concentration. The membrane and antibody solution can be shaken at room temperature for 1 h, or at 4 ^{o}C overnight. The membrane is washed with 1x

TBS with 0.1% (v/v) Tween for 5 min on a rotary shaker, 3 times, at room temperature. Development of the membrane can then be performed. When the antibody specific for the protein being looked for is conjugated to a horse radish peroxidise, a colorimetric western can be performed, whereby the membrane is developed with 3,3',5,5'-Tetramethylbenzidine (TMB). The membrane should be washed of the TMB to avoid burning of the membrane, by shaking with dH_2O for 5 min, and then repeating this twice more.

2.8.8 His-Tag Purification.

For purification of a protein containing a His Tag, immobilized metal ion chromatography (IMAC) was performed. This was done using the His•Bind® protein purification kit, provided by Merck Millipore. Samples were prepared as recommended by the kit through use of BugBuster Protein Extraction Reagent. The column was prepared as instructed by the manufacturers, where all buffers were made up to a 1x solution, and then passed through the settled uncharged IDA agarose resin in the column. The sample was then passed through the prepared resin, and washed with the kit supplied wash. However when required the elution buffer was changed from the kit provided elution buffer to alternatives. The kit provided elution buffer is a 4x solution of 4 M imidazole, 2 M NaCl, 80 mM Tris-HCl, pH 7.9. This is diluted down to 1x elution buffer and passed through the column, however it is a strong solution to strip all the remaining proteins off the column after washing. When there is low affinity to the column there is a problem with the protein being stripped off with other proteins in the mix as well. To counter this, the imidazole concentration can be altered in a buffer to reduce the number of different proteins in the elution buffers. The different was fractions and the elution buffer can then be tested to see where the His Tagged protein has being eluted into.

2.8.9 Carboxypeptidase G_2 Plate Assay

For confirmation of bacterial production of CPG_2 colonies were picked from transformation plates, and then patch plated onto appropriate phenotype plates with supplements of IPTG and folic acid (as described in Table 2-6). Colonies were picked with sterile toothpicks and then using a plate grid, streaked 1 colony per square on the plate grid. This was done in duplicate one stroke of the picked colony on the phenotype plate, and then one stroke of the same toothpick on a non-phenotype plate which is the same media without the supplements. This is done to ensure that any phenotype shown is down to the introduction of the folic acid and IPTG.

2.8.10 Methotrexate (MTX) enzymatic assay

To assay the protein CPG_2 quantitatively the methotrexate assay was performed. Samples were prepared through either fractionation as described, through lysis, or supernatant concentration as described.

Methotrexate was dissolved in 0.1M Tris-HCl pH7.3 containing 0.2mM $ZnSO_4$ to a concentration of 0.6mM. In a quartz cuvette 0.1 ml of this solution was mixed with 0.9 ml of the 0.1M Tris-HCl containing 0.2mM $ZnSO_4$. After pre-warming the cuvette to 37 ^oC, the initial A_{320} was taken to zero all results. Then 10 µl of sample was added to the methotrexate solution and mixed well before decrease in absorbance was measured. The value for the decrease in absorbance was then put into this equation to calculate the units of activity per ml:

$$\frac{\Delta A_{320} \times 100 \times Dilution\ factor}{8.3}$$

The ε_{max} for methotrexate is 8300.

2.9 Statistics

All data was input into GraphPad (v6). The graphs were generated from the data tables, and using the stastical analysis functions of the program, p-value was obtained. For comparisons between groups of data students t-test was performed, where data was unpaired. For significant difference between two groups a p-value of less than 0.05 was required.

C. sporogenes NCIMB 10696 Genome

Chapter 3

3 *C. sporogenes* NCIMB 10696 Genome

3.1 Introduction

To confirm *C. sporogenes* suitability as a potential vector for the delivery of prodrug converting enzymes and direct action drugs, the genome was sequenced. Genome sequencing has come a very long way in the past 10 years going from something that was a very costly to a relatively inexpensive process (Liu et al., 2012, Loman et al., 2012). This has opened the door to huge amounts of knowledge about the way organisms function and many advancements in the field of genomics (Shendure et al., 2004).

There are two main platforms for the sequencing of genomes, both based on PCR, but utilising this key tool in different ways. The systems are 454 sequencing and Illumina sequencing. For 454 sequencing good quality DNA is broken down into fragment libraries. The fragments are attached to the aptamers and are biotinylated, and made into single stranded DNA. The aptamers are short DNA sequences with complementary oligos immobilized on to the sequencing beads. The aptamers allow the fragments in the libraries to attach to sequencing beads, which are then emulsified. The emulsification mixture contains all of the reagents necessary for amplification of the aptamers that have already been attached to the beads. After emulsion PCR the beads are covered in single stranded DNA sequences, which can then be used for sequencing. These beads go into a Pico Titer Plate, consisting of small wells that accommodate a single bead and in which all sequence reactions take place. Once the single strand containing beads are introduced onto the plates and all set up for the sequencing run, the 454 sequencing machine washes reagents containing buffers and nucleotides. Each of the four nucleotides is sequentially added. The polymerase that is in the emulsification mix will allow the nucleotide to

bind to its complementary base when the correct nucleotide is dispensed by the 454 machine. The annealing process triggers a light emission through an enzymatic reaction triggered by some of the chemicals emitted when a nucleotide is incorporated. This is then detected by the cameras in the 454 sequencing machine. This light signal gets more intense for multiple single base regions in the genome, allowing the intensity of the light signal to indicate the number of sequential nucleotides which are annealed (Margulies et al., 2005, Morozova and Marra, 2008). Each nucleotide has modifications that allow it to give off the luminescence, but these modifications make the nucleotides terminal. These modifications are reversible allowing the continued sequencing of the fragments. This process is called Cyclic Reversible Terminators (Metzker, 2010).

Illumina Solexa sequencing works in a similar way. First fragment libraries need to be created from high quality genomic DNA. The fragment library has two aptamers ligated at both the 5' and 3' ends of the fragments from the library which are complementary versions of the same nucleotide sequence. These same aptamers have been immobilized on the flow cell from which the sequencing will take place. After the DNA has been attached to the aptamers, the single stranded fragment will anneal to the complementary aptamer on the flow cell at both ends forming a bridge. These aptamers will then be used as PCR primers and will extend along the fragment forming double stranded bridges. The double stranded bridges are then denatured to single strands and the ends bind to the complementary aptamer elsewhere on the plate forming new bridges to be amplified. This happens repeatedly until you have high density regions of the fragments all single stranded ready to undergo the sequencing process (Metzker, 2010, Holt and Jones, 2008).

To sequence the fragments that are isolated on the flow cell polymerases are used to anneal nucleotides on to the single strands that are fixed to the flow cell. These nucleotides, like in 454 sequencing, use a Cyclic Reversible Terminator method for sequencing. However, rather than fluorescence to indicate which nucleotide is annealing to the single strand, there is a colour emission, one colour for each different nucleotide. As each nucleotide is bound to the single strand the colour emission is picked up by the Illumina machine (Mardis, 2008, Ansorge, 2009).

The results from both the 454 and Illumina sequencing can then be cleaned up, and mapped. This is the biggest task of generating an accurate genome, and a good bioinformatics pipeline is needed for the clean-up of reads. Only with good quality information can an accurate genome be assembled. Once the raw sequencing data is obtained there are a lot of reads which will be incorrect due to incorrect binding of non-complementary bases, false signals being detected or DNA being damaged. This will cause reads to be incorrect, and the quality of the data produced will reflect this. Fortunately due to the nature of this method there will be hundreds of reads of each locus, which will compensate for these errors, and allow any suspect reads to be discarded. If there is the odd single nucleotide change in sequences this will not affect the overall sequence obtained from the sequencers.

When assembly is performed, the multiple reads of each locus of the genome give immense amounts of data to be analysed. The numerous reads of each locus gives confidence in the sequence, but sorting through the data to eliminate inaccurate reads, and to create an accurate assembly through overlap consensus sequence takes a great deal of processing power. This is why when doing the aptamer libraries pair ended libraries are commonly used. By creating paired reads which have a set number of base pairs between the first read and its pair we can better map the reads

through positional knowledge. While generation of the pair ended read library from the genome is more difficult, the additional information that it provides for the positions of the sequenced reads makes it very worthwhile. The generation of a paired end library can be done through the addition of an EcoP151 restriction site at the 5' and 3' ends of the cDNA due for sequencing. This is done by ligation of aptamers containing the EcoP151 site to either end of the DNA. The EcoP151 site is recognised by the enzyme, and then cuts 27bp into the DNA sequence that has been annealed to the DNA to be sequenced. This leaves a 54bp sequence between two aptamers that can then be mapped easier due to the fact that the length of sequence between the two short reads is known.

Assembly can be done against a reference genome, or *de novo.* Depending on the type of assembly being performed different read types can be used. The *de novo* assembly will require the added information of the pair-ended reads, but some might prefer to have single reads if they are performing assembly against a reference genome. There are numerous ways for the genome to be assembled, however, the most common way is for the reads to be aligned to form an overlap layout consensus. When combined with the positional information garnered from the paired end reads this allows blocks of sequence consensus to be formed, these are known as the contigs (Pop et al., 2004, Havlak et al., 2004, Pop, 2009). Improvements on this methodology, such as de bruijn graphs, allow assemblies to be performed with improved efficiency when dealing with issues of replicated sequence throughout the genome. The de bruijn graphs treat each read as a group called a k-mer. These k-mers are then treated as single reads where the ends are nodes. Through identifying matching nodes the assembly is produced (Zerbino and Birney, 2008),

Once the genome has been sequenced and is assembled there will need to be further analysis, depending on the state of the genome this could simply be annotation. It is more likely that the genome will be in a highly fragmented state requiring further sequencing and experimental analysis (Pop and Salzberg, 2008). The further sequencing is slightly simpler, however, and can be performed through Sanger sequencing which will give much longer reads. Through PCR of the gaps between the contigs it is possible to close some of those gaps, and end up with a more complete genome.

To help with mapping of assembled contigs, optical genome mapping or restriction maps can be performed. This is performed through the isolation of a single molecule of chromosomal DNA onto a charged glass slide, which stretches the genome flat. The DNA is then covered with an enzyme which cuts the DNA, and it consequently contracts away from the cut site leaving a gap. As the DNA molecule is isolated on the charged glass slide the fragment order is maintained. The DNA is then stained with a fluorescent dye, and the intensity of fluorescence is measured to show the length of the DNA strand (i.e. the greater the intensity, the longer the strand). This is repeated numerous times to give a consensus restriction map, and account for any inefficiencies in the restriction digest. This can then be compared to an *in silico* restriction digest of other genome data such as the illumina or 454 sequencing and through this we can gain a much greater understanding of the positions of contigs

(OpGen, 2015).

With the decreasing cost of sequencing a genome, it has driven a wealth of new bioinformatic tools to deal with immense amounts of data that can be generated. With the increasingly facile programs that have been created increasing amounts of information can be elucidated from the reads that are gained from sequencing. Not

only can more be done with the sequencing data, but sequencing data can be generated quicker with fewer errors in new machines, such as a HiSeq, and ion torrent. However, as it stands the gold standard of sequencing is still the Illumina Solexa sequencer, the Roche 454 sequencer, and more recently the PacBio sequencer which works through a process called Single Molecule Real-Time (SMRT) sequencing. A single molecule of DNA is attached with a DNA polymerase into a chamber with a special censor called the Zero-Mode Waveguide (ZMW). The molecule is allowed to replicate as normal, but with specially linked fluorescent dyes. The dyes are different for each of the four nucleotides, and when the nucleotide bonds, the dye is released. This is detected by the ZMW, giving the DNA molecules sequence.

For this study initial Roche 454 sequencing was performed, followed by assembly. For the purposes of CDEPT, it is required that there is a single complete DNA sequence, to confirm that the bacterium is a Biosafety Level 1. To this end numerous further strategies were applied to the *C.* sporogenes genome for gap closure and confirmation that the bacterium was non-pathogenic. Some of these tactics included resequencing on the Illumina platform, creating an *in vitro* restriction map, and performing numerous *in silico* techniques.

3.2 Results

The here presented results were obtained through a collaborative effort between Dr Aleksandra Kubiak, Dr Anja Pohlain and myself. All collaborative work is presented in this chapter to give a clear idea of what was undertaken to achieve the genome in its published form.

To this end Dr Kubiak performed the initial genomic DNA extraction, which was sent for 454 sequencing, performed gap closure which resulted in 6 different scaffolds.

She also performed the genomic DNA extractions. After completion of the genome she performed the genome analysis such as genome map, and phylogenetic tree. She performed the PCR for identification of the potential *C. botulinum* toxin genes.

As part of my MRes I performed the original annotation and aided Dr Kubiak in attempts to close the scaffold gaps. For this PhD I performed the Illumina sequence genome assembly (both de novo and against a reference), and then analysed the optical mapping data. Using the bioinformatics data I attempted to close the scaffold gaps. I also continued the gap closure with PCR to attempt to bridge the gaps between the six scaffolds.

Dr Anja Pohlain performed the final gap closures of the remaining scaffolds, and then performed the re-annotation of the complete sequence.

3.2.1 Sequencing

The bacterial DNA was prepared as described in the materials and methods Section 2.6.1.2. Sequencing of the *C. sporogenes* genome was performed by preparing 3 separate libraries which were paired end reads with a 3kb insert. Then using the 454 platform the libraries were sequenced.

Sequencing data was cleaned up by removing the sequencing aptamers, and disregarding the reads which were not reliable. With the reads remaining an assembly was performed, using the software GS assembler 2.3 (Newbler 2.3). This work was also performed at the DeepSeq facility at the University of Nottingham. This assembly resulted in 45 contigs, which had the gaps between these contigs closed through various different PCR methodologies to form scaffolds. The most common one performed was standard failsafe PCR. The Sanger sequencing data of the PCRs was integrated into the sequence data to give six final scaffolds. The PCRs

to close the contig gaps were mostly for products of about 1kb and were performed by Aleksandra Kubiak. The gaps between the scaffolds were attempted to be closed in the same manner as the contig gaps were closed. This was, however, unsuccessful, and it was suspected that these gaps were too large to be closed from standard PCR. The contigs which could be grouped into the 6 scaffolds are shown in Fig 3-1.

The scaffolds were mapped against an already sequenced alternative strain, *C. botulinum* F Langeland. This showed the *C. sporogenes* genome to have about 98% coverage if homologous to the reference strain. The remaining 2%, which were the areas in between the scaffolds, were shown to be too large to be sequenced with conventional PCR through comparison with the reference genome. It was suspected that the reason that there was such difficulty in originally amplifying those gaps was due to the fact there were multiple ribosomal sub units that were situated in the gaps, and that these regions have a high AT content. As the sub units are homologous within an organism this would explain the difficulty in assembling multiple loci of the exact same sequence.

a)

Scaffold Number	Number of Contigs	Size (bp)	% coverage
Scaffold 01	2	4,563	0.1116
Scaffold 02	1	4,156	0.1017
Scaffold 03	2	28,249	0.6912
Scaffold 04	36	3,442,410	84.2249
Scaffold 05	2	417,351	10.2113
Scaffold 06	2	143,313	3.5064
Gaps	x	47,122	1.1529
Total		4 087 164	100

b)

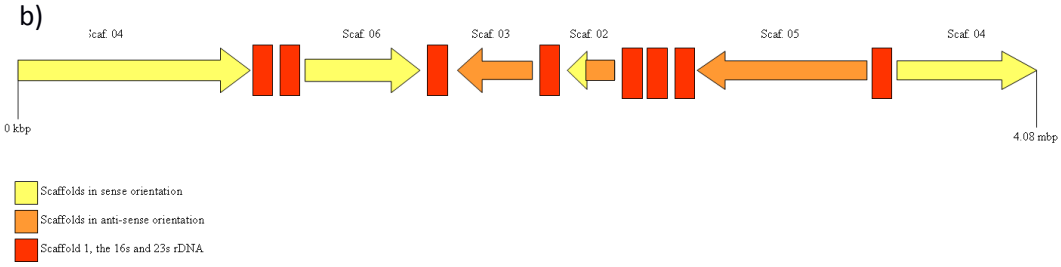


Figure 3-1 The initial scaffolds orientation, and mapping, and a table of what constitutes the scaffolds.

A) shows the different scaffolds and the number of contigs that make them up. It also shows the length of the scaffolds and what percentage of the genome that is. B) shows the different scaffolds and their orientation when mapped against the C. botulinum F Langeland.

The gaps between the scaffolds were attempted to be closed using numerous different techniques. Primarily the gaps were attempted to be closed through use of specialist PCR polymerase, LongAmp polymerase, specialising in regions greater than usual. The LongAmp polymerase can amplify up to 40kb products, while the gaps in between the different scaffolds were not that big, therefore primers were designed to bridge those gaps. Some of the PCRs performed resulted in DNA amplification, however these bands were not in accordance with the size expected from the maps produced. The products were sent for sequencing, however, the sequencing failed most likely due to the high AT concentration. Consequently attempts were made to use the 16s and 23s rDNA ribosomal sub units as stepping points as the products would be smaller and, therefore, more stable. However this also failed.

The genomic DNA was extracted and resequenced using the alternative sequencing platform, Illumina. Using the reads from this a *de novo* assembly was performed, as well as an assembly that was performed against the previous model organism *C. botulinum* F Langeland. The two assemblies were compared for accuracy, and for any possible gap closure using the previously assembled *C. sporogenes* scaffolds. When the genome was assembled against the *C. botulinum* F Langeland this resulted in over a hundred contigs. When a de novo assembly was performed there was a much reduced number of contigs (94), of a greater size. The assembly when done with the reference *C. botulinum* was disregarded due to the poor quality of the assembly. By comparison of the two different assemblies, we were not immediately able to increase the level of coverage for the *C. sporogenes* genome however confidence in the layout of the assembly was increased. Finally to see if the original assembly could be improved upon the reads of all the sequencing runs were combined for a single assembly. Again there was no improvement on the sequences that were already known.

After alternative attempts to close the genome had failed, Dr Anja Poehlain, used her pipeline to assemble the genome. This pipeline consisted of assembling the Roche 454 sequencing reads using the Roche Newbler software (version 2.8), a hybrid assembly was also performed with both the Illumina and the Roche 454 reads. The hybrid assembly was performed using the program Mira (version 3.5). The contigs which were obtained were then aligned against closed genomes of the closest organisms to *C. sporogenes* NCIMB 10696, namely strains like the *Clostridium botulinum* Hall A. This created a similar result to what had been seen previously, where the gaps which were left were rDNA repeat regions. To resolve this issue a single read library was created from the 454 sequences and then along with the

allignment results the gaps were closed. For *in vitro* closing of the gaps Bio-X-act from Biozym as described in the methods 2.6.3.5. These methods resulted in one contiguous genome for the strain *C. sporogenes* NCIMB 10696

3.2.2 Restriction Mapping

While the amount of information gives a degree of confidence in the assembly, due to the inability to PCR any sequences between the scaffolds the assembly of the genome could still only be considered hypothetical. Therefore OpGen®, a company which performs restriction mapping, did a restriction map of the isolated *C. sporogenes* NCIMB 10696 genome. The restriction map was performed through the extraction of genomic DNA. Single molecules of the genomic DNA were then stretched and imobilized onto a charged glass surface. The genomic DNA is then digested while immobilized on the glass plate, at the points were the restriction enzyme (NcoI) cut, the DNA recoils leaving a gap. After digestion for visualization and establishing the size of the fragments the DNA was stained with a fluorescent dye. The level of fluorescence of the fragments is directly proportional to the length of the fragment that has been stained. Through multiple repeats of this procedure the restriction sites are developed into a "barcode" like map shown in Figure 3-2. The OpGen MapSolver software was used to create an *in silico* map of the genome sequence that had been obtained from the Illumina and 454 sequencing. This was then compared to the physical restriction map that had been developed using NcoI. These were then aligned to show whether the assembly was poor, or whether our hypothesised scaffold order was correct. The 6 scaffold assembly showed, when aligned against the *in vitro* digest of the genome, that there was an area of Scaffold 04 which was not correct, and that there was a section missing from the genome assembly that lay in the centre of the scaffold. When performing the alignment between the *in silico* scaffolds and the digested genome due to ambiguity in the

restriction profile of the smallest scaffolds they could not be aligned to one specific place in the digested genome.

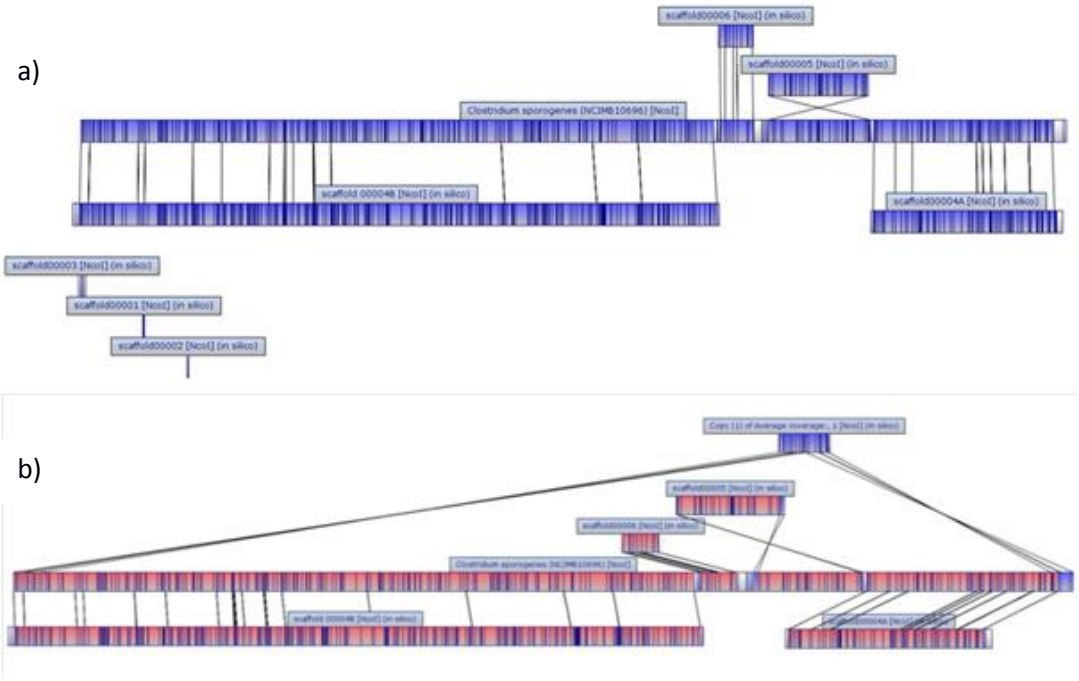


Figure 3-2 The OpGen MapSolver restriction digest maps of C. sporogenes

A) *Shows the first genome assembly aligned against the restriction map of the single molecule of genomic DNA. The Scaffolds 01, 02 and 03 as can be seen don't have a very distinctive restriction profile, and therefore could not be aligned to the restriction map. B) Shows the alternative restriction map alignment with the added re-sequencing data. The original two restriction maps that were aligned are now highlighted red, with the extra contig from the re-sequencing added in blue.*

When analysing the OpGen mapping in the MapSolver many different genomic sequences can be compared to the *in vitro* produced map. Each different genome sequence will be digested *in silico* once imported in to the MapSolver software and can then be compared to the other digested genomes already displayed. This technique was performed by introducing all the different assemblies performed on the genome of *C. sporogenes.* By introducing some of the other genome assemblies we were able to see in greater detail the differences and similarities between the

different assemblies. It was also possible to quickly assess whether there were some contigs which covered any gaps in the assemblies, giving a better understanding of what needed to be done to complete the genome. This method allowed us to identify based on the restriction digest profile that there was one contig that crossed a gap in the genome. This closed the gap that had been opened by the restriction digest profile.

The genome was closed finally by Dr Anja Poehlain who has experience in completing entire genomes. The pipeline that she followed was similar to that which had been tried with a few alterations in the programs used. The sequencing data was taken and a *de novo* assembly was performed using Newbler 2.8 and Mira 3.4. The resulting gaps were closed with further information gained through Sanger dideoxy sequencing using Bio-X-Act polymerase.

3.2.3 Annotation

The resulting genome in its numerous renditions has been annotated several times, through various methods. For each rendition of the genome, there was an annotation and as the assemblies improved so too did the annotation methods. The original methodology used to annotate the genome was through identifying homologous genes in two different genomes, one of which is annotated. The annotation is then transferred from the annotated genome and applied to the un-annotated genome. This process is performed by a tool developed by the Sanger Centre in Cambridge called Rapid Annotation Transfer Tool (RATT) (represented in fig 3-3). This tool is a Perl-based command line program. The program takes into account issues such as single nucleotide polymorphisms (SNPs), translocations of DNA, and other deletions when transferring the annotations. This annotation was then checked through performing a Glimmer based identification of the open

reading frames, which uses hidden Markov models to identify all possible start and stop codons, and then rule out possibilities that are not realistic. Once the Open Reading Frames (ORFs) were identified if they did not match an already annotated gene, the sequence of the ORF was blasted to identify if there were any possible genes which matched the sequence. All hypothetical proteins were checked in this way also, in case there was an alternative annotation that wasn't present in the reference genome. This methodology provided a rough annotation that was used for the original 6 scaffolds. This annotated genome was used until better tools were available for the improvement of the genome and its annotation.

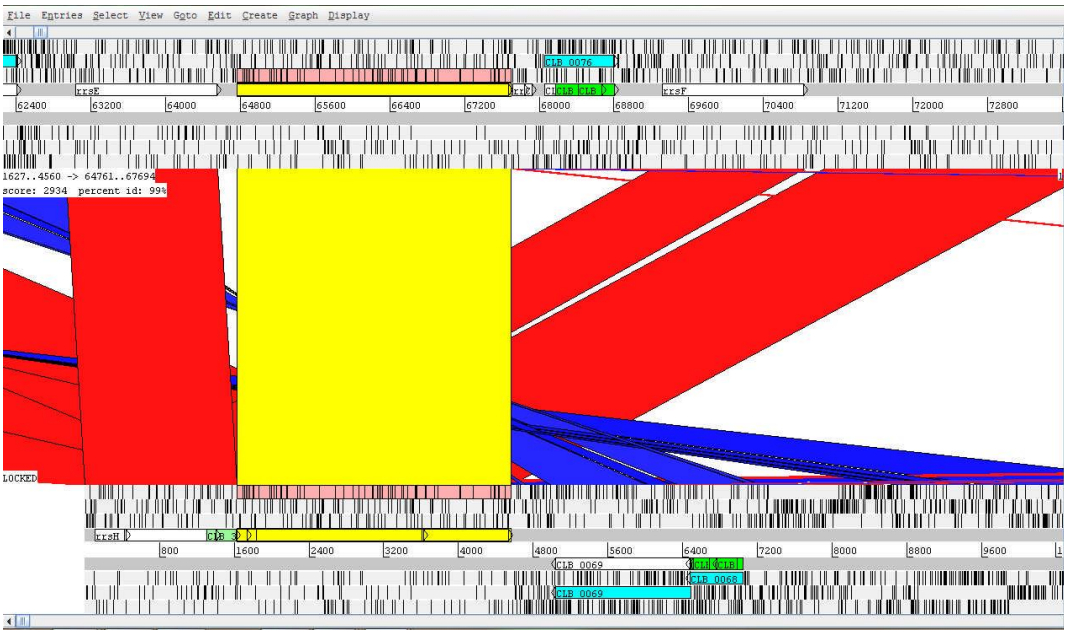


Figure 3-3 The visualization of RATT.

Artemis Comparison Tool (ACT) can visualise how the RATT tool worked in drawing a comparison between an already annotated genome and the de novo annotation. On the top is the C. sporogenes genome and the bottom is C. botulinum and in between the genomes shows blocks of homology that will allow annotation to be transferred between the annotated genome to the de novo genome.

The final annotation was performed using the program PRODIGAL on the complete genome to predict gene ORFs, and RNA elements such as rRNA and tRNA were identified using the RNAmmer and tRNAscan software. The IMG/ER (Integrated

Microbial Genomes/ Expert Review) system was used to annotate the ORFs. The annotated ORFs were curated through a variety of databases, such as Swiss-Prot, TREMBL and InterPro.

The sequence assembly and annotations gave us 3,732 coding sequences (CDS), 2,956 ORFs which are protein encoding, 776 putative genes for 60 hypothetical proteins, 4 pseudo genes (genes which are functionless, but a copy of a gene), 9 rRNA clusters which total 27 genes and 80 tRNA's. The genome consists of a GC content of 27.98% and 82% of the genome was found to be encoding. In terms of similarity the genome has 97-100% similarity to other *C. sporogenes* strains, and 91-99% to the proteolytic class 1 *C. botulinum* strains. The similarity between these two species rRNA was higher at 99-100%. The genome of *C. sporogenes* and its features have been represented in a circular map shown in fig 3-4

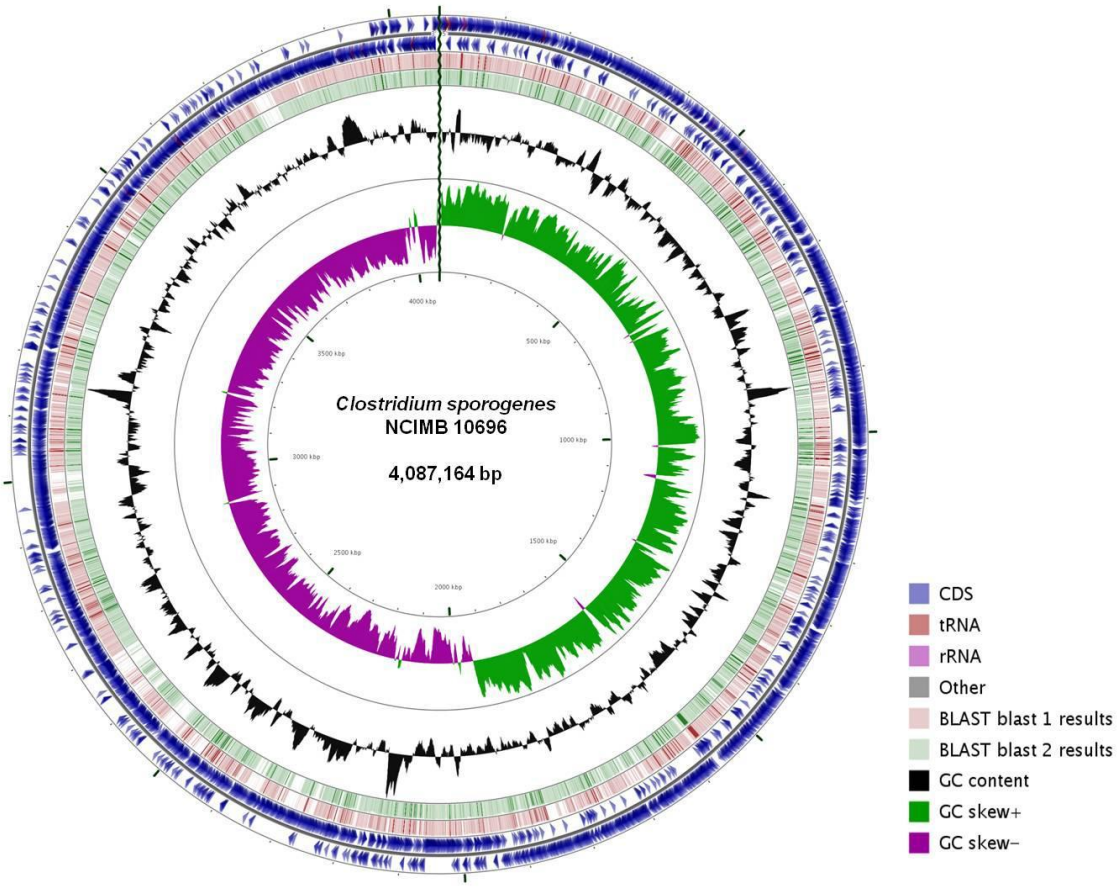


Figure 3-4 A circular genome map of C. sporogenes.

The genome shows from the inside out GC deviation (where green values >1 and purple <1), GC content for the black ring, Blast results of C. sporogenes ATCC 15579 (pale blue), and Blast results of C. botulinum Hall A (pale green), then finally the coding sequence for C. sporogenes NCIMB 10696 in both orientations.(Kubiak, 2013)

3.2.4 Phylogenetic Dendrogram

The *C. sporogenes* genome sequence has been completed and can therefore be put into context of the genus through constructing a phylogenetic tree. The 16s rDNA data was taken from the *C. sporogenes* genome and then was used to perform a discontiguous megablast to generate 101 16s rDNA sequences, of similar sequences. These were then used in a neighbour joining method to create a phylogenetic dendrogram in the FigTree program (shown in fig 3-5). This was performed together with by Aleksandra Bailey.

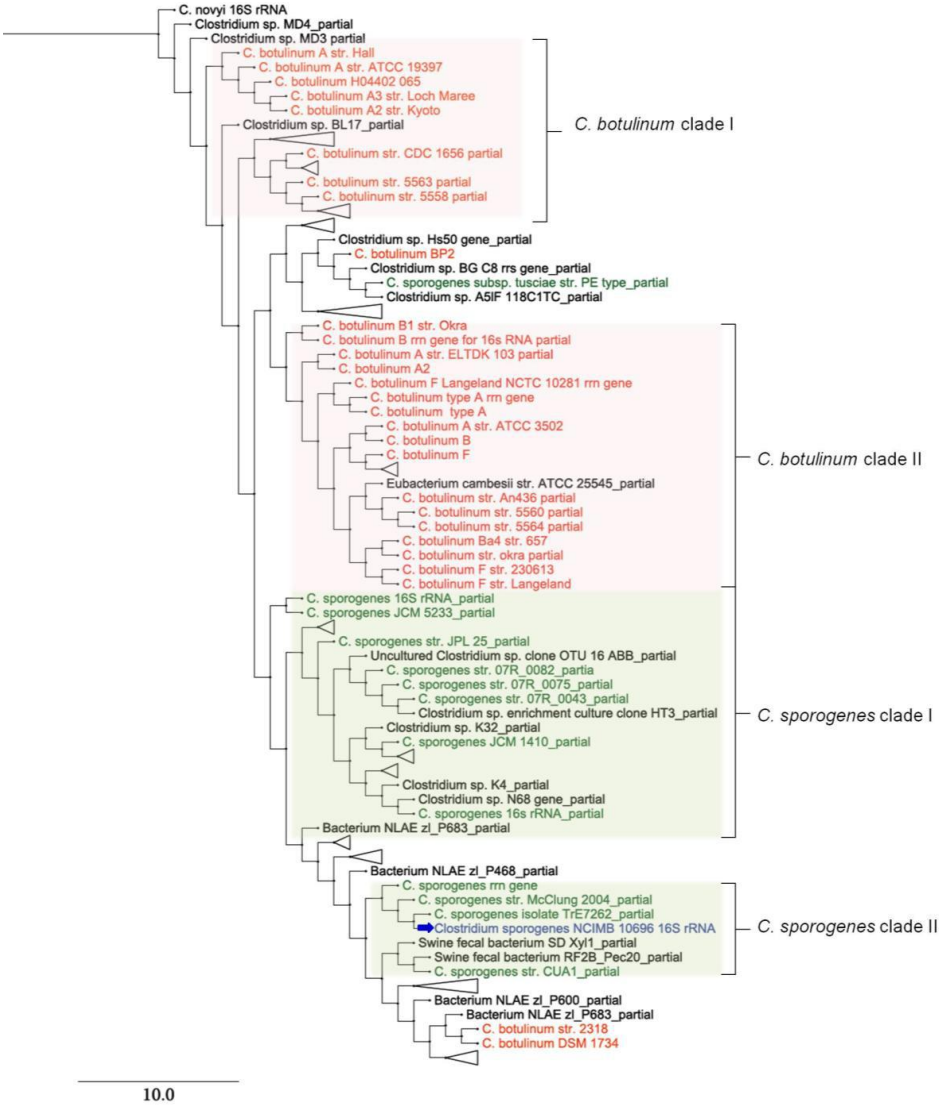


Figure 3-5 A partial Clostridium dendrogram.

The dendrogram here shows the relationship between the 16s rDNA of numerous different organisms that were identified through a discontiguous megablast of the C. sporogenes NCIMB 10696 16s rDNA. The results of this megablast were consequently made into a dendrogram with the initial organism marked in blue. The scale depicts percentage divergence (i.e. if a branch of the tree is that distance from another branch it is 10% divergent) (Kubiak, 2013)

As can be seen the *C. sporogenes* bacteria, highlighted in green, form two distinct clades which are closely related. It can be seen that both *C. sporogenes* and *C. botulinum* form two distinct clades. Of those four clades there was a distinct division between the two species with their clades being placed together. There were a few outlying strains of both *C. sporogenes* and *C. botulinum* which sit on branches which

are not associated to the clades of their species. Through this dendrogram we can ascertain that the *C. sporogenes* and *C. botulinum* species are completely separate.

3.2.5 Toxin Identification

With the genome and annotation complete post genomic analysis was performed. To confirm that the proposed host *C. sporogenes* was suitable for cancer therapy it was essential to ensure that there were no pathogenic elements. To do this the literature for *C. sporogenes* was researched to identify whether it had been linked to disease or toxin production. This identified some association with three different potential conditions which could be cause for concern if *C. sporogenes* is to be used as a vehicle for cancer therapy. The *C. botulinum* neurotoxin was looked for, due to *C. sporogenes* and *C. botulinum*'*s* very close genetic similarity. Despite never showing any evidence of this potent neurotoxin it was deemed important to show that there was no remnant of this gene cluster even if it was not functional. There is also another haemolytic toxin that has been associated with both *C. sporogenes* and *C. botulinum* which is the streptolysin like toxin, which has been named clostridiolysin (Gonzalez et al., 2010). The final thing to be considered is the link between *C. sporogenes* and gas gangrene. Through culturing bacteria out of the different wounds that have resulted in gas gangrene, *C. sporogenes* has been isolated several times.

After identifying these toxin genes, the sequences for the different genes were collated. The DNA sequence was BLASTed against a local database which consisted of the sequenced *C. sporogenes.* This resulted in confirmation that there was no toxin genes present in *C. sporogenes* for either the botulinum neurotoxin, or for the streptolysin toxin. However, there are some issues for proving that *C. sporogenes* does not cause gas gangrene. The first is that there is no known cause of gas

gangrene, however, there are some commonly associated toxin genes. The commonly associated gas gangrene toxin genes, which are the α, θ and δ toxins (Stevens and Bryant, 1993), were not found in the *C. sporogenes* genome. All toxin genes identified and searched for in *C. sporogenes* have been listed in table 3-1

Table 3-1 Toxin Gene Presence Table

Gene Name	Gene function	Present in *C. sporogenes*
atx	Neurotoxin Type A	Not Found
boNT/A1	Neurotoxin Type A1	Not Found
boNT/A3	Neurotoxin Type A3	Not Found
boNT/B	Neurotoxin Type B	Not Found
boNT/BvA4	Neurotoxin Type BvA4	Not Found
boNT/BvB	Neurotoxin Type BvB	Not Found
boNT/E3	Neurotoxin Type E3	Not Found
CbC4_4096	Neurotoxin Type C	Not Found
ntnh	nontoxic-nonhemagglutinin component of B	Not Found
ntnH	nontoxic-nonhemagglutinin component of A1	Not Found
ntnH_1	nontoxic-nonhemagglutinin component of BvA4	Not Found
ntnH_2	nontoxic-nonhemagglutinin component of BvB	Not Found
ha-17	Hemagglutinin component of the Neurotoxin	Not Found
ha-33	Hemagglutinin component of the Neurotoxin	Not Found
ha-70	Hemagglutinin component of the Neurotoxin	Not Found
CLJ B0563	Bacteriocin Protoxin Streptolysin	Not Found
CLM 0572	Bacteriocin Protoxin Streptolysin	Not Found
sagB	Streptolysin Associated Protein B	Not Found
sagC	Streptolysin Associated Protein C	Not Found
sagD	Streptolysin Associated Protein D	Not Found

The table showing the different genes identified in the literature that could be present in C. sporogenes. The gene sequences were all obtained from the NCBI database, and blasted against the C. sporogenes genome to ascertain their presence or not.

While the *C. botulinum* neurotoxin genes were not detected through BLAST, due to the genetic relatedness of *C. sporogenes* and *C. botulinum* it was deemed prudent to show *in vitro* that there are no toxin genes present in the *C. sporogenes* NCIMB 10696 strain through performing PCR of the toxin genes. The PCR's were performed by Aleksandra Kubiak, using primers designed by Dr Chris Humphreys both of the CRG. The primers were designed to amplify only part of the toxin genes. There were

no fragments of expected size in the *C. sporogenes* genome as shown by the PCRs (Fig 3-6) however there were some faint bands shown.

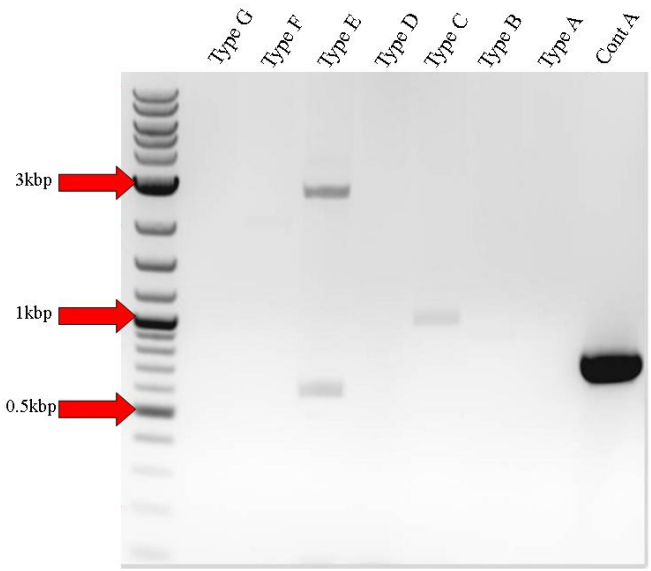


Figure 3-6 The toxin type PCR.

The gel picture shows the attempts to amplify the different toxin types out of the genome of C. sporogenes the control was performed for toxin type A, showing that the PCR works. When looking for the other toxin type the band sizes where expected to be for Type B 205bp, Type C 169bp, Type D 264bp, Type E 389, Type F 543bp and Type G 326bp (Kubiak, 2013).

These bands were probably down to un-optimised conditions, resulting in nonspecific binding of the primers and therefore false bands. This suggests that there are no potential *C. botulinum* toxin genes in the *C. sporogenes* genome.

3.3 Discussion

The importance of confirming a safe non-pathogenic group-1 bacterial strain for the use as a delivery vehicle for the cancer therapeutics is important. To this end a stringent genome sequencing was performed of *C. sporogenes* NCIMB 10696. The methodology of sequencing this genome was performed to ensure that there was complete genome coverage, and that there was a good degree of confidence in the genome. The bacteria was sequenced on both the Roche 454 platform initially and

then resequenced on the Illumina Solexa platform. The data this generated was used to form 45 contigs which were joined together to form the 6 scaffolds. The gaps between those scaffolds were then closed completely by Dr Anja Poehlin, leaving us with a genome of the size of 4,141,984 bp (Kubiak et al., 2015). Gaps often consisted of ribosomal DNA (rDNA) repeats. The rDNA and the high AT content made closing the genome gaps difficult. The sequencing of these AT rich regions can be problematic due to DNA instability. When sequencing the addition of new nucleotides forms hydrogen bonds, two for bonding AT, and three for bonding GC. By having high numbers of repeated AT's sequentially, it greatly reduces the number of hydrogen bonds, causing instability. This causes difficulty in sequencing as the strong bond between the base pairs is how the sequencing is reliable. Also when there are large numbers of AT bonds polymerase slippage occurs (Aird et al., 2011). The repeated rDNA sequences cause problems due to the fact that the rDNA sequences are highly conserved. The highly conserved gene is replicated throughout the genome resulting in lots of reads which are for the same sequence at different loci in the genome. The loci where the rDNA are, are difficult to identify when doing an assembly due to this as there appears to be a particular well covered single locus, not numerous loci (Treangen and Salzberg, 2012).

The positioning of those scaffolds was initially established through the assembly being performed against the closest related strain, and was later given confidence through the OpGen optical mapping. Finally this was annotated several times over, using increasingly efficient and up to date methods. The final result was a complete genome with 100% coverage that was fully annotated. This genome was then published (Kubiak et al., 2015).

Due to the attempts at getting the complete 100% coverage of the genome, a lot of different checks have been performed that would not be normally done to perform a genome sequencing. Due to the reduced cost of sequencing a bacterium, it was viable to perform a resequencing of the genome to ensure that there was no genetic information that was missing from the initial set of reads that were obtained from the Roche 454 sequencing. To also get the best of the two market leading sequencing platforms one run of each platform was performed. Both times the library that was made for the *C. sporogenes* genome was a paired end library to ensure that there was good positional information allowing assemblies to be better. Then to further confirm this, the OpGen optical mapping was performed. This resulted in good confidence in the confirmation of the scaffolds. Finally with the ever improving platforms more sequencing data can be produced for cheaper. With this comes a demand for better downstream tools and software. Therefore the initial annotation of the genome performed on Rapid Annotation Transfer Tool, was cast aside in favour of the new improved techniques allowing better confidence in what each genes annotation. Using this strong genome basis we can confirm the non-pathogenic status of *C. sporogenes*

As had been established the genetic relatedness of *C. sporogenes* to *C. botulinum* made it essential to understand the genetic composition of *C. sporogenes* and ensure that there were no homologous toxin genes. Despite only limited association with other potential toxins, all other potential side effects must be known before implementation in a patient, as they will have to be managed. So due to the relatedness of *C. sporogenes* to *C. botulinum* and the links that have been talked about in *C. sporogenes* literature a complete genome was needed to ensure that *C. sporogenes* is not a pathogen. While some side effects might be manageable in the

context of treatment, they will have to be compared to be understood to ascertain whether CDEPT is a viable treatment option. To ensure that CDEPT is an improvement on the current chemotherapeutics, a genome of the strain used in CDEPT must be performed to ensure that the toxicity that has been associated with the bacteria is not relevant for this strain. For CDEPT to be considered a success it has to either improve upon the efficacy of the best treatments available or, have the same efficacy without the debilitating side effects.

With many *C. sporogenes* strains now sequenced, we can ascertain that the species so far does not appear to possess any toxin producing genes (Kubiak et al., 2015, Bradbury et al., 2012). This gives confidence that *C. sporogenes* is a class 1 non-pathogenic bacterium. There are, however, some questions to be raised about the sequencing platforms and the error margin for the different platforms. While both platforms sequence the majority of the DNA sequence correctly, neither are perfect and as a result errors will occur in the genome. The single base pair misreads could result in an amino acid change within the genome that is not actually present *in vivo.* As such the level of confidence in the reads might not be as high as hoped. However, hopefully by performing the resequencing these misreads will be accounted for. Alternatively newer and better platforms could be used to improve upon this, such as ion torrent, or PacBio sequencing (Merriman et al., 2012, Quail et al., 2012).

While very thorough genome sequencing has been performed there will always be room for improvement and this more often than not can be found in the annotation. The annotation techniques, as has been mentioned, are getting better as demand for simple effective annotation tools increases. These better tools however still only assign function to the genes found through *in silico* analysis. Without performing *in vivo* testing of the genes to confirm the assigned function, these annotations are only

hypothetical. This results in many of the genes ORFs being annotated as hypothetical proteins with no known function. While the function is probably not known of these genes, it could be that there is a false open reading frame being identified by the programs. This can only be confirmed or denied through *in vitro* research of the genes. So while we could see from the annotation that there were no toxin genes, it was important to try to confirm through PCR of the potential toxin genes. This was why the PCR looking for the *C. botulinum* neurotoxins was performed. With this helping to confirm the lack of neurotoxins, it was felt that the *C. sporogenes* genome has been developed to a high confidence, and has shown that there is a very low chance of toxin genes being present. To help add confidence a southern blot for the different toxin genes could be performed.

Implementation of CPG_2 and Anti-VEGF

Chapter 4

4 Implementation of CPG_2 and Anti-VEGF.

4.1 Introduction

When introducing a foreign gene into a bacterium it is no longer sufficient for the bacteria to be shown to be transcribing and translating the foreign DNA. Understanding bacterial protein expression, and the mechanisms involved, can give us new and innovative ways of not only incorporating and expressing the genes but also expressing the genes at the highest level possible. The optimisation of gene expression is desirable in numerous different ways but can also cause many unforeseen issues that need to be addressed.

There are many problems that could arise from the expression of novel DNA in a new organism. The expression levels could be very low, which could be due to poor regulation and sub-optimal conditions. Factors such as the ribosomal binding site, promoter, and the potential toxicity of the encoded protein can all have an effect and many of these issues will not be known until attempts to express the gene are made in the new host.

When dealing with a potentially toxic gene one way forward is the use of an inducible system, containing a switch which will allow initiation of the expression of the gene and production of the encoded protein. There are many inducible systems which can be used but probably the most common one, and best understood, is the Lac inducible system. The *lac* operon, comprising the *lacZ, Y* and *A* genes, is preceded by a single promoter and followed by a transcriptional terminator. The *lac* promoter is characterised by the presence of an operator sequence (*lacO*) that immediately follows the +1 to which the LacI repressor binds. Expression of the *lacI* gene is controlled by its own promoter and terminator binding of LacI to *lacO* blocks

transcriptional from the *lac* promoter, preventing the production of the *lacZYA* mRNA. When the organism encounters lactose the product of the *lacZ* gene, β-galactosidase processes the sugar to alolactose which binds to LacI. The ensuing conformational change causes the repressor to dissociate from *lacO*, allowing production of mRNA. Exogenous Isopropyl β-D-1-thiogalactopyranoside (IPTG) fulfils the same function as alolactose, binding to LacI and preventing its interaction with *lacO.* To implement this in a novel host aiming to control expression requires positioning of *lacO* in an equivalent position upstream of the gene of interest, and then the *lacI* gene to be expressed in the host. By having a constitutively expressed *lacI* gene and an operator, the transcription of the gene will be repressed until the introduction of IPTG (Fuerst et al., 1989, Brown et al., 1987, Hu and Davidson, 1987).

However, there are other inducible systems that can also be used such as the rhamnose inducible system or the T7 inducible system. The T7 inducible system works through the highly specific relationship between the T7 polymerase and T7 promoter. The T7 promoter will only drive expression of the novel gene when in the presence of the T7 polymerase (Tabor and Richardson, 1985). The rhamnose system works through the presence of L-rhamnose. When the gene of interest has transcription driven by the *rhaB* promoter the absence of L-rhamnose causes no transcription of the gene. When L-rhamnose is present there is *rhaR* and *rhaS* transcription, and *rhaS* accumulates triggering the transcription of *rhaB* (Terpe, 2006, Giacalone et al., 2006). Some systems combine two together to ensure the tight control of expression as some of these methods can be leaky. For example the pET28b expression plasmid uses what is called a *T7Lac* promoter. Most pET plasmids use the T7 system for inducible transcription, which is a standard utilisation of the T7 expression system. However the pET28b plasmid has been adapted to increase

repression for particularly difficult to express genes. The T7 promoter is still present, meaning that T7 RNA polymerase is still required for expression however, a *lacI* gene has also been added under its own promoter. This means that even if the T7 RNA polymerase were to be bound to the promoter, unless alolactose or IPTG were to be bound to the LacI protein, no expression could take place. In the host *E. coli* BL21 (DE3) star, the T7 RNA polymerase genes is under the promoter *lacUV5* which requires IPTG for induction, as does the T7*lac* promoter for the production of the gene, meaning that the gene should be tightly repressed until the addition of IPTG which will not only induce the production of the T7 RNA polymerase, but also allow it to bind to the T7*lac* promoter which will be free to transcribe the gen of interest because the LacI protein will be bound to the IPTG that has been added rendering it incapable of preventing transcription (Peränen et al., 1996)

For Clostridial Directed Enzyme Prodrug Therapy (CDEPT) the yield of the prodrug converting enzyme (PCE) should be maximised to limit the number and length of treatments. This is a trade-off between minimal side effects, while maximising drugs in the body, without adversely affecting the normal function of the cell. This can be achieved in several ways, such as changing the promoter, the ribosomal binding site, or through codon optimisation (Burgess-Brown et al., 2008).

As the genetic code is degenerate, there are more variations of the base pairs coding for the amino acid than there are amino acids. This results in more than one three base code coding for a single amino acid. Each organism has a codon bias, leading to optimal expression levels of genes when codons are optimised for the expressing organism. Codon optimisation is often performed to try to ensure that the ideal amino acid code for the expression organism will be used.

To utilise this codon bias in an organism an ideal gene should be designed using the correct codon for the amino acid being coded for. However this is not an easy task as there are millions of possible DNA sequences coding for the same amino acid sequence, even for relatively short sequences. To develop the best possible sequence *in silico* processes can be applied to the genomic data available and a codon usage table can be generated. The codon usage table defines which codons are used most frequently for that particular amino acid in that organism. By picking the most commonly used codons for each amino acid in the codon usage table, it is thought that an optimal sequence will be created (Gustafsson et al., 2004). However this does not always work, and further studies have shown that the preferred codon usage system is actually more complicated.

There will be several different versions of the sequence due to the fact that there might not be one ideal codon for each amino acid. Therefore the different optimised sequences will go through a process of clean up and optimisation to ensure that, of the potential sequences, only the best possible options can be used *in vivo*. The clean-up process consists of ensuring the correct GC content of the gene, elimination of repetitive sequences, and avoiding mRNA secondary structures which would cause issues of limited translation, and eliminating any other undesirable features for the downstream processes such as undesirable restriction sites (Gustafsson et al., 2012).

For this study two different potential therapeutics were studied, a PCE which could be implemented in a traditional CDEPT manner, and a direct action scFv antibody. The PCE to be implemented is CPG_2, a folate reducing agent which has been shown to break down methotrexate, and used as treatment in cases of methotrexate toxicity. The monoclonal antibody to be implemented is anti-VEGF, an antibody which neutralises the Vascular Endothelial Growth Factor and so allows

normalisation of the vasculature in and around the tumour. This anti-angiogenic strategy, as has been mentioned, is seen to be a balancing act between vascular normalisation or vascular pruning (Jain, 2014). Too much anti-angiogenic signals would lead to vascular pruning, which would prevent delivery of chemotherapeutics to the tumour. While it was thought that vascular pruning would decrease the level of oxygen and increase cell death, it seems unlikely this effect would be enough to allow it to be used as a standalone treatment. With vascular normalisation it would allow chemotherapeutics to be delivered resulting in a better prognosis for the patient.

Neither option is bad for a CDEPT strategy. The increased hypoxia would provide an environment which would better suit the germination of more spores after the initial anti-angiogenic effect, or the normalisation would make an environment that is less suitable for germination, but better for efficient delivery of chemotherapeutics.

4.2 Results

Apart from when stated as otherwise, all genes in this study were codon optimised by DNA2.0 to maximise potential yield of the PCE and of the direct action antibody, anti-VEGF. The anti-VEGF monoclonal antibody is a humanized single chain variable fragment antibody. Previously a codon optimisation pipeline was developed and tested for the expression of genes in *C. sporogenes*, where a codon optimisation algorithm was developed and then a clean-up process implemented. This work was done by DNA2.0 and Aleksandra Kubiak (Kubiak, 2013). Using their pipeline, DNA2.0 generated different versions of the two therapeutic genes that were to be implemented in *C. sporogenes*.

To begin there were four different DNA2.0 generated versions of the anti-Vascular Endothelial Growth Factor (anti-VEGF) antibody encoding gene, and two different versions of the carboxypeptidase G_2 (CPG_2). There was two versions of CPG_2 that were not optimised by DNA2.0, one that had been isolated from wild type *Variovorax paradoxus cpg_WT*, and a codon optimised version of that which had been performed by Entelechon (Germany), *cpg_Synth*. The codon optimised version was optimised for the original proposed vector organism *Clostridium acetobutylicum* ATCC 824 for research done previously in this laboratory. All different versions of the genes are shown in table 1. Each gene had a signal sequence shown to work in *C. acetobutylicum*, the *eglA* signal sequence (Theys et al., 1999), and the anti-VEGF genes due to lack of observable phenotype were produced to incorporate a FLAG-tag and a His Tag at the c-terminus so that they could be purified.

Table 4-1 A brief overview of all the plasmids that were created for the expression of the different target genes and the special features of each plasmid, and gene.

Plasmid	Gene of Interest	Antibiotic selection	Promoter/Inducible System	Other features of note
pJ201	*anti-VEGF_A, anti-VEGF_B, anti-VEGF_C anti-VEGF_D, cpg_EC, cpg_CS*	Kanamycin	No promoter system, storage plasmid	All inserted anti-VEGF genes have Flag-tag and His-tags
pMTL84153	*anti-VEGF_A, anti-VEGF_B, anti-VEGF_C anti-VEGF_D*	Chloramphenicol	ferodoxin promoter, *Pfdx*	All inserted anti-VEGF genes have Flag-tag and His-tags
pMTL82251:YZ2	*cpg_EC, cpg_CS, anti-VEGF_A, anti-VEGF_B, anti-VEGF_C, anti-VEGF_D*	Erythromycin	A *lacI* driven by an independent promoter, then after an alternate promoter a *lacO*	None
pMTL550fm	*cpg_wt, cpg_synth*	Chloramphenicol	Facoid promoter	None
pMTL82251:PtcdB	*anti-VEGF_A, anti-VEGF_B, anti-VEGF_C, anti-VEGF_D*	Erythromycin	*TcdB* promoter	None
pET28a	*cpg_EC, cpg_CS*	Kanamycin	T7 promoter,	His tag

4.2.1 CPG_2 Cloning into Standardised Plasmid Vectors.

With all genes either a codon optimised or wild type version, finding a suitable expression system was researched. The table 4-1 also shows what plasmids the gene was originally stored in, and what features the plasmids had. These different versions of the genes were originally on different plasmids, and to minimise variables in the expression plasmids, a standardised plasmid was proposed. This was derived from the modular plasmid system previously developed (Heap et al., 2009). The plasmids proposed were one expression plasmid, and a control, with no promoter. The plasmid maps are shown below in Figure 4-1.

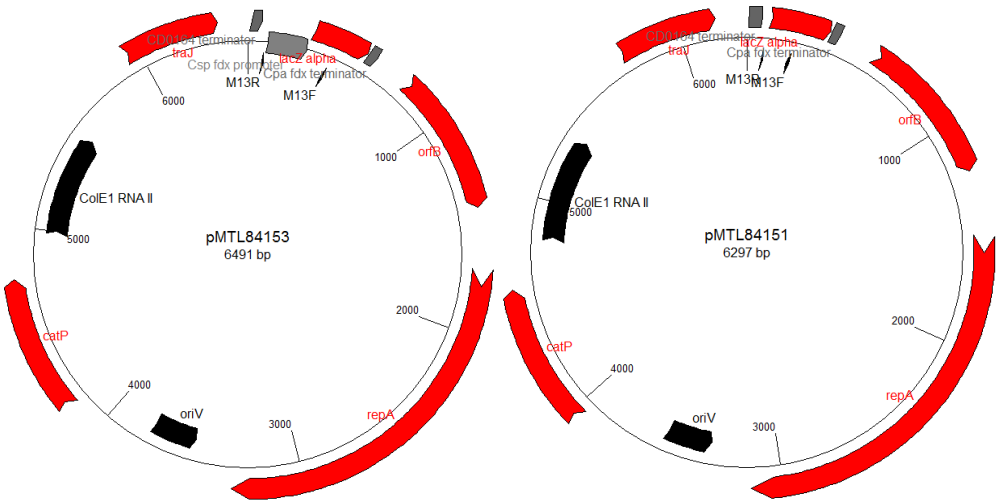


Figure 4-1 The plasmid maps of the initial storage, and expression vector

The two standardised plasmids that the various drugs were attempted to be cloned into. They are identical bar the use of a promoter before the multiple cloning site in the pMTL84153 plasmid. The lacZ alpha gene is used as both a multiple cloning site, and for blue white screening.

As the anti-VEGF genes were already on the pMTL84153 plasmid and an alternative promoter-less plasmid (DNA2.0's storage plasmid pJ201) no cloning needed to be done. However, the original CPG_2_WT and CPG_2_Synth were on plasmids that were not ideal for *C. sporogenes*. They therefore needed to be subcloned into the

proposed modular plasmids. Due to the age of the stock cultures of the plasmids for CPG_2 WT and Synth, there were numerous issues with the growth of the bacteria. Attempts were made to revive and purify strains in the hope that would improve growth. Typical OD readings of overnight cultures were low. From purified strains growth was still relatively poor, so alternative attempts at optimising plasmid cloning were undertaken. These included increasing the volume of overnight culture used in plasmid minipreps, digesting a larger volume of plasmid, and using a smaller amount of elution buffer per mg of excised gel band. When CPG_2_WT or CPG_2_Synth plasmids were obtained the plasmids were digested using NdeI and PstI restriction enzymes. However, due to a lack of PstI site on the modular pMTL84153 plasmid vector backbone, the PstI sticky-end was blunt-ended, and the plasmid pMTL84153 was digested using the blunt-end enzyme SmaI. This simple cut and paste method did not result in correct plasmids in the transformed colonies so the cloning strategy was changed to a PCR based strategy. It was found after numerous attempts at moving the gene from one plasmid to another using various cloning strategies that the concentrations of the insert were too low to be successfully ligated into the backbone. The insertion of a functional CPG_2 could never be confirmed in these plasmids. When testing the original plasmids for expression through plating the transformants onto plates with folic acid complement, they did not give a phenotype. If phenotype had been observed there would have been a yellow halo around the colonies as folic acid was broken down, and the non-soluble by-product pteroate precipitates out. Consequently these two old versions of the gene were not used in this study.

The WT *cpg* gene sequence was therefore sent to DNA2.0 for codon optimisation. Two versions of the *cpg* gene were synthesised by DNA2.0 (opt_EC and opt_CS).

When synthesised, due to difficulties with cloning, the genes were received in the storage plasmid pJ201, and not in the pMTL84153 plasmid like the anti-VEGF genes. This resulted in further attempts to clone the *cpg* gene, however, these attempts at cloning were performed with different recipient plasmids to the modular pMTL84153 and pMTL84151 plasmids. It was hoped that an inducible plasmid would prevent any potential toxicity. The alternative recipient plasmids were the model *E. coli* expression vector pET28a, which has a *lac* inducible system, and an alternative clostridial inducible system pMTL82251:YZ2 shown in Fig 4-2 (Zhang et al., 2015). The clostridial inducible system consisted of a *lacI* gene being driven by an ideal promoter (*Pptb* from the phosphate butyryltransferase gene in *C. acetobutylicum*) in antisense, and then the *Pfac* promoter, a *lacO* and the *cpg* genes on the sense strand. The cloning strategy was also improved upon to remedy the issues experienced with the first strategy. This improved strategy was based on PCR of the respective inserts introducing at the 5' and 3' sticky ended restriction sites (BamHI and HindIII). The PCR was optimised, cleaned up, and then digested, and resultant digests were run on an agarose gel, and cleaned to ensure the highest quality and concentration of insert DNA.

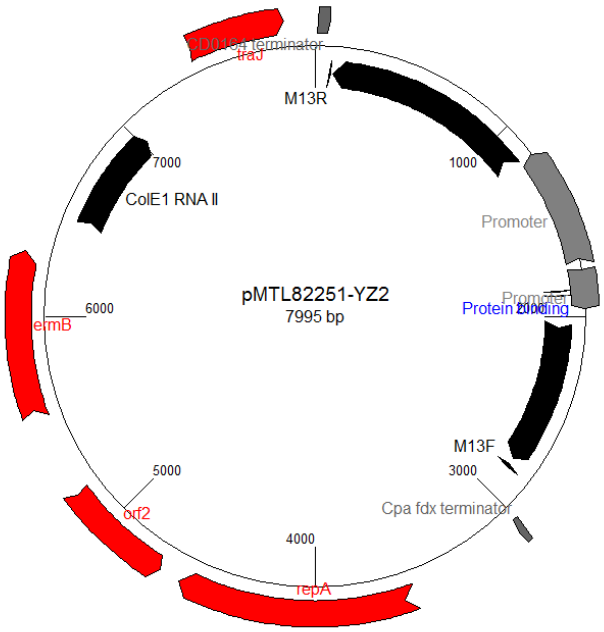


Figure 4-2 A plasmid map of the inducible expression plasmid.

The plasmid map of the inducible plasmid that was used for expression of CPG_2

4.2.2 Expression of Targets CPG_2 and anti-VEGF

Once the new strategy was implemented with the alternative expression plasmids, screening resulted in colonies which showed phenotype for expressed *cpg_Ec* in the *E. coli* BL21 (DE3) Star host as it is known as a good expression host (shown in Fig 4-3). In both the pET28a and the pMTL82251:YZ2 plasmid the expression of the *cpg* gene was under the control of the *lac* inducible system. The phenotype is as a result of the CPG_2 in the plates not only being able to digest the prodrug, but also being able to cleave folic acid into glutamate and the insoluble pteroate. This results in the pteroate precipitating in the media, resulting in the yellow halos that are visible in Fig 4-3. Colonies with halos were picked and plasmids were sequenced to confirm the presence of the *cpg_Ec* gene.

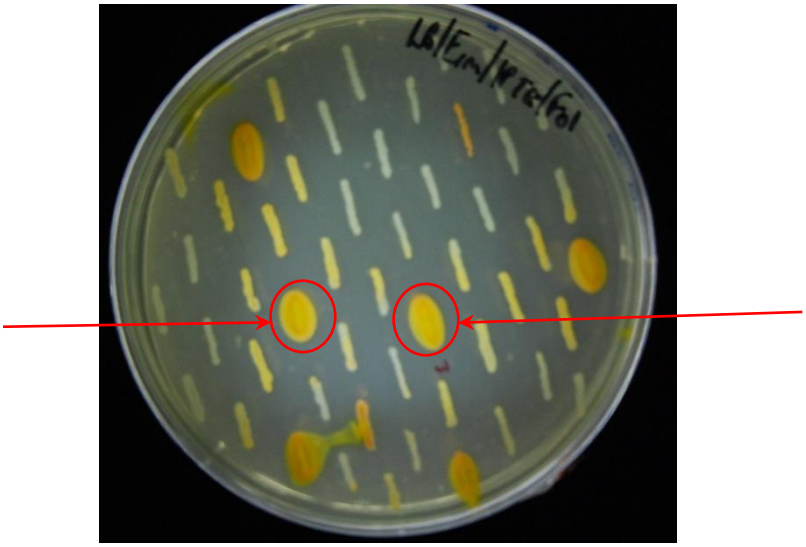


Figure 4-3 Screening plates for CPG_2.

The screening for the pMTL82251:YZ2:cpg_Ec plasmid electrotransformed into the E. coli BL21 (DE3) Star in the photo. Examples of the positive phenotype (where the pteroate precipitating out of the media) have been circled and highlighted

With the enzyme successfully cloned into the *E. coli* expression plasmid (pET28a), and the *C. sporogenes* expression plasmid (pMTL82251:YZ2) the pMTL82251:YZ2:CPG_2_Ec plasmid was conjugated into *C. sporogenes* using the standard protocol. The colonies were screened as before using a folic acid and IPTG supplement. However, the normal TYG media could not be used for this screening as it is too dark to distinguish the colour of the pteroate. Therefore, the chemically defined Lovett medium (see methods) was used. This is not a minimal medium, and so growth of *C. sporogenes* is not affected. It is translucent and so shows the colour change of the folic acid breakdown well. This is shown in Fig 4-4.

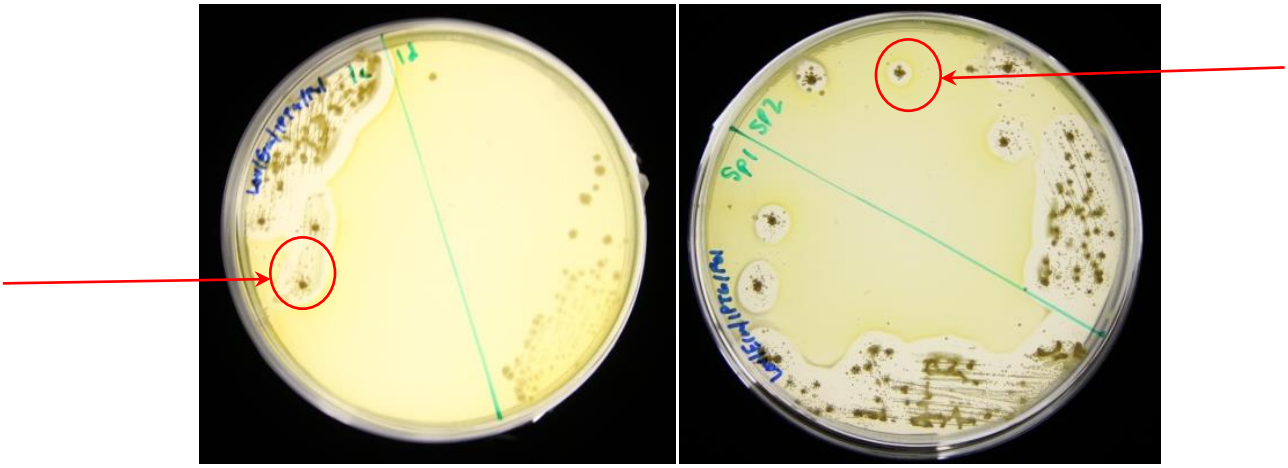


Figure 4-4 the same plasmid from fig 2 conjugated into C. sporogenes.

You can see there are 4 streaked strains of which 3 have an active cpg_Ec gene the 4th clearly shows no clearance, and so is not active. Again examples of the CPG_2 activity have been circled and highlighted to show what was expected.

Once expression had been confirmed in *C. sporogenes* and the plasmid had been sequenced to confirm the presence of the *cpg_2_Ec* gene, a random selection of 4 colonies containing the plasmid were enzymatically assayed for activity. After 3 h of growth IPTG was added to the culture. The bacterial cultures were sampled at 4 h, 8 h, and 24 h, 1 ml of culture was taken each time. The samples were originally processed into lysates, and the supernatant was tested after TCA protein precipitation (see methods). The results are shown in fig 4-5

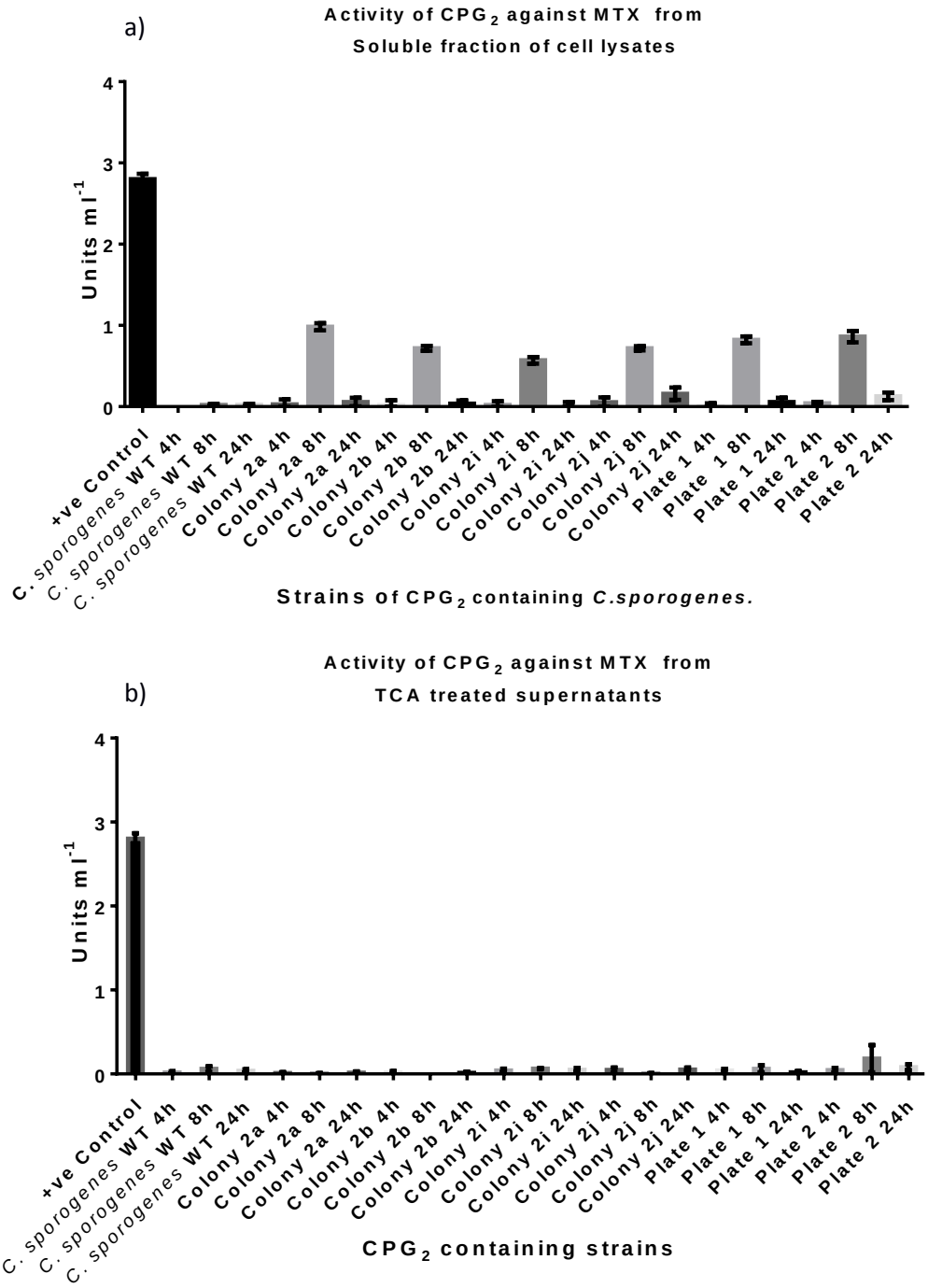


Figure 4-5 The enzymatic assay of the CPG2 producing strains.

Several random colonies were picked, as a cross section of the conjugated colonies, and the plates were scraped if remaining colonies to ensure no uniquely high producing colonies were missed in the screening. a) The top graph shows the enzymatic activity of the lysates of these colonies, and b) The bottom graph shows the activity of the TCA precipitated supernatants. For each sample biological duplicates and technical triplicates were performed. Standard Error bars are shown.

The lysate showed a peak in activity at the 8 h timepoint, and a reduction past that point. Due to low activity observed, the supernatant was processed to see whether it was being exported at all from the bacterial cell. To begin with the processing was performed through protein precipitation, but no activity was seen.

Due to the lack of significant difference between the selection of colonies with the *cpg_Ec* gene one colony was picked to represent all the different biological replicates that had been originally assayed. This colony was 2a, and will now be referred to as the CPG_2_Ec strain.

The alternate version of the codon optimised *cpg* (*cpg_Cs*) was synthesised and was cloned in the pET28a and pMTL82251:YZ2 plasmids. It was cloned into the same plasmid for expression in *C. sporogenes* as the *cpg_Ec* gene. This plasmid was then conjugated into *C. sporogenes,* and colonies which gave a yellow phenotype on a folic acid supplemented solid media chosen for further analysis. A panel of these *cpg_Cs* strains were assayed again showing no real difference. Therefore, a single representative colony was used referred to as the *cpg_Cs* strain. The activity of the *cpg_Cs* and the *cpg_Ec* were compared to assess which one was better. Due to the lack of activity in the supernatant it was hypothesised that the extremely high TCA concentration was denaturing any protein that was exported in the supernatant, alternative methods of protein concentration were researched and ultrafiltration with a molecular weight cut off (MWCO) of 50,000 was performed. The resulting concentrate was subjected to the MTX assay, but while it showed a slight increase in activity, the majority of the protein activity was still found in the soluble fraction of the lysates. This is shown in fig 4-6.

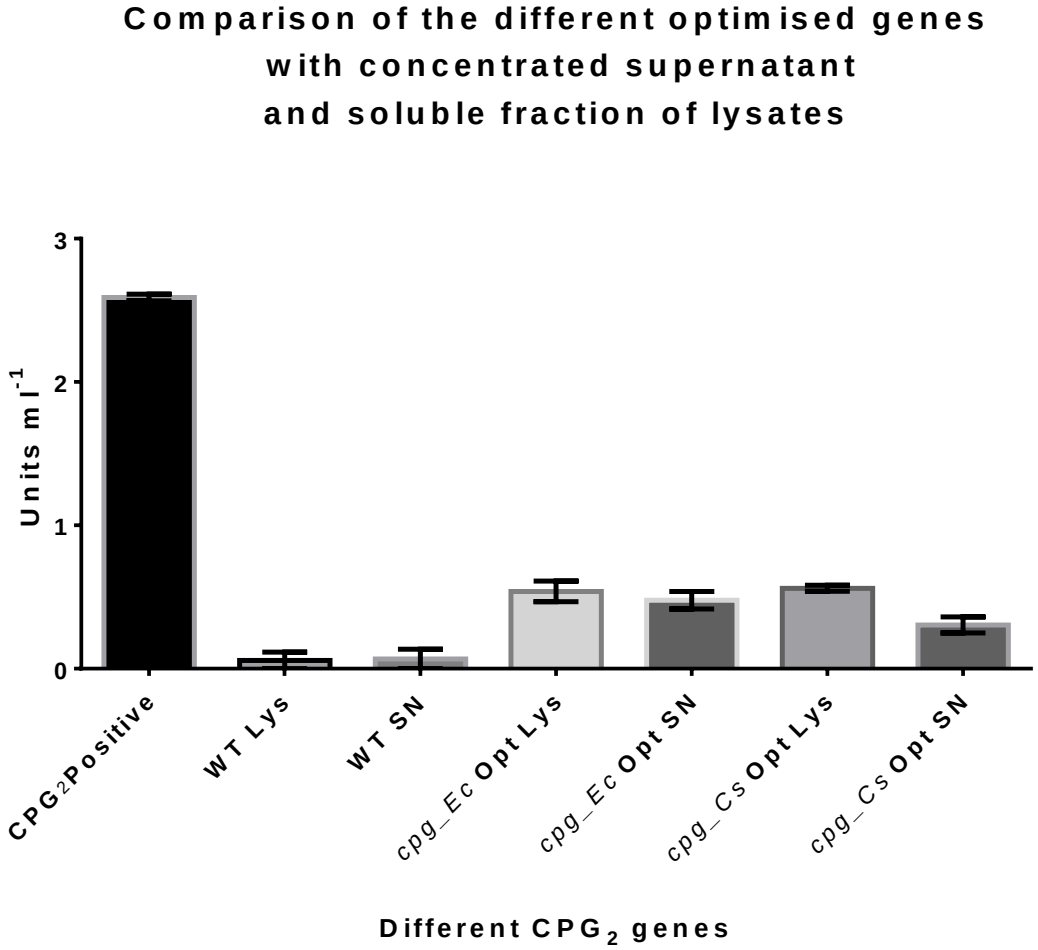


Figure 4-6 Comparison of CPG_2 versions

A comparison of the two optimised gene versions being expressed in C. sporogenes. The 1 ml samples were taken after growth for 8 h and lysates were treated as described for bugbuster, and supernatants were ultrafiltrated. Technical triplicates were performed and standard error bars are shown.

Reasons to account for the apparent lack of protein outside of the cell were investigated. The genome of *C. sporogenes* has shown that there is an operon encoding 6 metalloproteases that could be expressed. These metalloproteases could inhibit CPG_2 activity by breaking down the enzyme before being assayed. The ultrafiltrated supernatant of the *C. sporogenes* wild type, and the *C. sporogenes* containing the pMTL82251:YZ2:CPG_2_Ec were therefore analysed by SDS-PAGE. If the protein was present in the supernatant you would expect to see an extra band on

the gel for the pMTL82251:YZ2:CPG_2_Ec samples. The SDS-PAGE was stained with colloidal coomassie, and is shown in Fig 4-7.

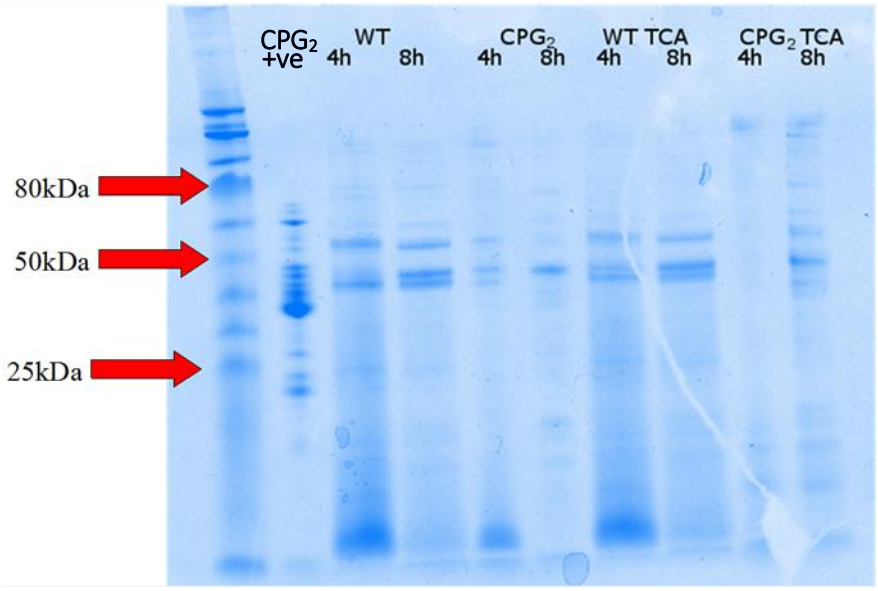


Figure 4-7 A stained SDS-PAGE of C. sporogenes supernatant

A Coomassie stained gel of the supernatants from both a WT and a strain containing the cpg_Ec gene. These were then concentrated with TCA or left untreated and then all samples were run against purified CPG_2 as a positive control.

Concurrently, different *C. sporogenes* strains were grown in the presence of an EDTA-free protease inhibitor cocktail (cOmplete EDTA free Protease Inhibitor Cocktail, Roche). This inhibitor should inhibit metalloproteases by at least 10% when the metalloprotease concentration is 0.5 µg/ml, and all other proteases by a minimum of 74% when the concentration is 20 µg/ml or higher (https://www.lifescience.roche.com/shop/en/uk/products/complete-edta-free-3271402-1, 14/10/2016). Growth was unaffected by this addition to the culture media when measured for OD_{600}. Samples were then taken from liquid culture at 8 h. The *C. sporogenes* were treated with BugBuster to lyse the cell pellets, and through ultrafiltration for concentration of the supernatant. The samples were then subjected to a MTX assay. The results are shown in Fig 4-8.

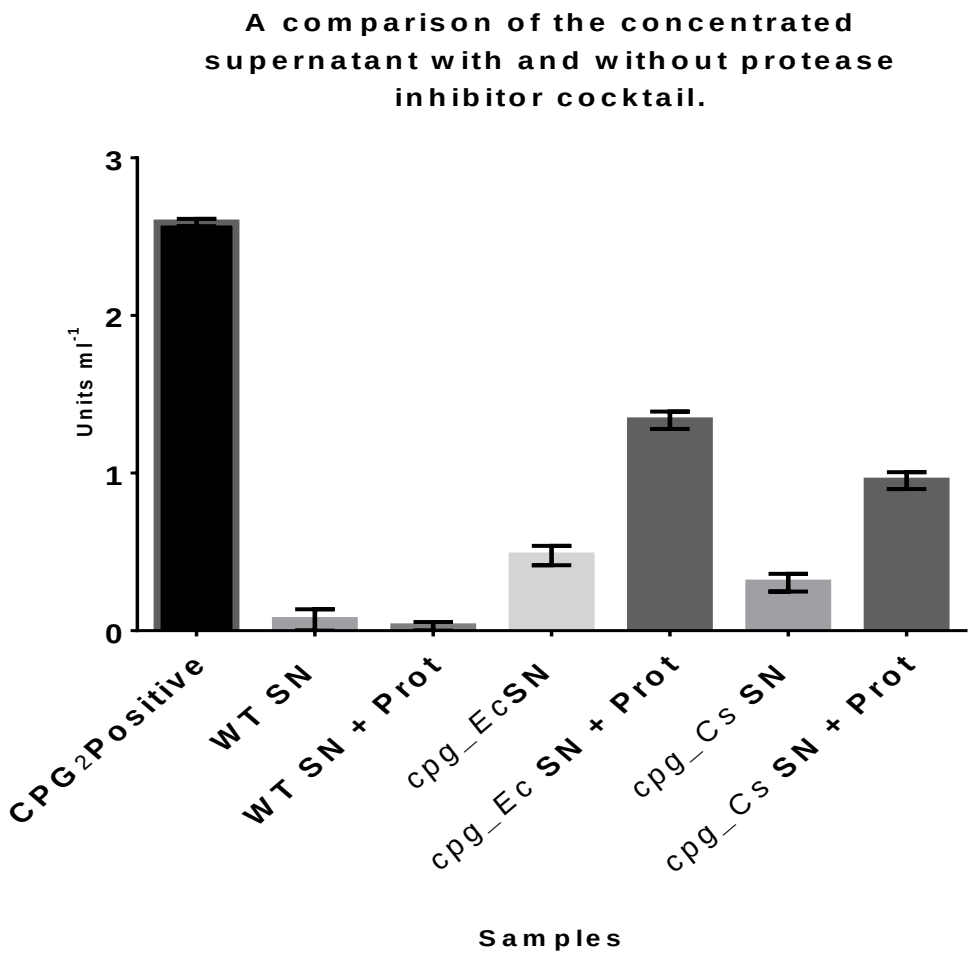


Figure 4-8 A comparison of different enzymatic activity in the supernatant

A comparison of the supernatant of the strains producing different cpg genes, Cpg_Ec and cpg_Cs with and without the Roche cOmplete protease inhibitor cocktail the strains with protease inhibitor cocktail are labelled +Prot. A wild type strain, and purified CPG_2 were used as negative and positive controls respectively. Technical triplicates were performed and standard error bars are shown

Through the enzymatic activity assay we can see that there is a significant increase (p is 0.0005 for *cpg_Ec* and 0.0011 for *cpg_Cs* with protease) in the level of CPG_2 activity when the protease inhibitor cocktail is present in the growth media. The inhibitors may be preventing the breakdown of CPG_2 when it is free in the media, ensuring more active molecules.

4.3 Variovorax paradoxus

Due to the inability to clone the original wild type from the p550FM plasmids, work was done to try and re-isolate the original wild type version of *cpg* from the original bacterium *Variovorax paradoxus*. *V. paradoxus* strains were isolated originally from soil samples by growing on plates with folic acid as the sole carbon source. When the folic acid is the sole carbon source only strains containing CPG_2 will grow, as they break down the folic acid making it useable as a carbon source. Alternatively if folic acid is used as a supplement, then colonies were picked when the colony exhibited CPG_2 phenotype. Strains from the CRG culture collection were revived on LB media, and grown on M9 minimal media. The M9 minimal medium was supplemented with folic acid to ensure that the strains that were growing had a functional *cpg* gene present. There were three strains which were isolated from soil samples by previous researchers at the CRG, and one strain from the ATCC centre. Of the *V. paradoxus* strains that were in the CRG culture collection, only the strain from the ATCC exhibited any functional CPG_2 activity. Before incorporation into the expression plasmids, SOE PCR was used to incorporate the signal peptide onto the CPG_2 that was isolated from *V. paradoxus*. This gene was cloned into the standard *lac* inducible systems that were used for the expression of *cpg*, pMTL82251:YZ2, and pET28a. The plasmids were transformed into *E. coli* BL21 (DE3) Star cells, and expression was confirmed through patch plate screening on LB with folic acid and IPTG supplements. This resulted in colonies displaying the characteristic yellow phenotype. The plasmids were sequenced to confirm the presence of the wild type *cpg* gene. These correct plasmids were consequently conjugated from *E. coli* CA434 to *C. sporogenes* NCIMB 10696. To confirm the plasmid in *C. sporogenes* the colonies obtained from conjugation were patch plated onto Lovitt media again with IPTG and folic acid supplements. This resulted in colonies producing the CPG_2 protein, and giving the

characteristic yellow halo. Again the plasmid was sequenced to confirm the presence of the *cpg* gene. The bacteria that were confirmed to have the correct plasmid present were grown in culture and samples were taken from a 24 h culture at the 4 h, 8 h, and 24 h marks. These samples were tested for enzymatic activity in a quantitative assay on the lysate and the supernatant (shown in Fig 4-9). When the results were compared to the level of expression of the codon optimised genes, no difference was seen.

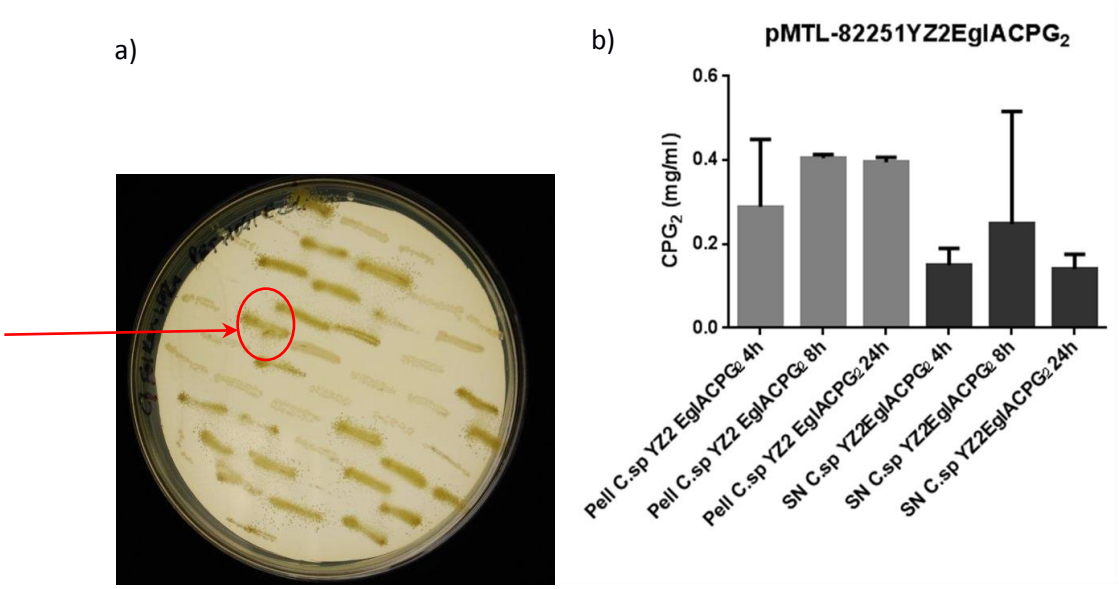


Figure 4-9 A) WT Variovorax paradoxus CPG_2 implemented in the pMTL82251:YZ2 plasmid in E. coli.

The screening plate showed phenotype of folic acid breakdown by CPG_2 from numerous colonies. These colonies are circled on the plate. B) An enzymatic assay of the CPG_2 after being attached to the EglA signal peptide and conjugated into C. sporogenes. Samples were taken at the 4 h, 8 h and 24 h time point for both supernatant and lysate. Samples were performed in biological and technical triplicate with standard deviation bars.

The expression is not different from that shown by the codon optimised genes, but it does show again that there is a much lower level of expression in the supernatant than the lysate. It can be seen that there was a slight peak in production at the 8 h mark, but due to high degree of standard error the confidence in this being true is

low. This slight fluctuations could be attributed to many things, such as slight temperature variations when growing, slight oxygen levels in the anaerobic cabinets. All would account for the slight increase in at the time point.

4.4 Expression of anti-VEGF

DNA2.0 was responsible for synthesising codon optimised versions of the anti-VEGF antibody, and cloning them into an expression vector for insertion into *C. sporogenes*. They were synthesised with a FLAG-tag and a His-Tag at the C-terminus. Due to the success in synthesising the anti-VEGF genes and cloning them into the originally intended plasmid of pMTL84153, these plasmids were conjugated straight into *C. sporogenes* from the *E. coli* CA434 donor strain. The conjugated colonies were screened and sequenced to show correct plasmid insertion. Consequently the conjugated strains were tested for the production of the 4 different versions of the anti-VEGF gene. The activity of the 4 different anti-VEGF genes was assayed through a western blot, using an anti-FLAG tag antibody to show correct activity. This is shown in Fig 4-10.

The western blots were performed on *C. sporogenes* containing either the plasmid pMTL84153:antiVEGF_A, pMTL84153:antiVEGF_B, pMTL84153:antiVEGF_C, or pMTL84153:antiVEGF_D. To initiate production of the anti-VEGF antibodies, IPTG was introduced to the media at 3 h. Samples were then taken at 4 h, 8 h, and 24 h. Due to presence of the *eglA* signal sequence in the constructs we looked for the protein in both the cell lysates and the supernatant of the culture after it had been collected. The lysates of *C. sporogenes* showed a band for a protein with a FLAG-tag but not at the expected molecular weight of 31kDa. The detected protein seemed about 2.5 times larger in size (Fig 4-10).

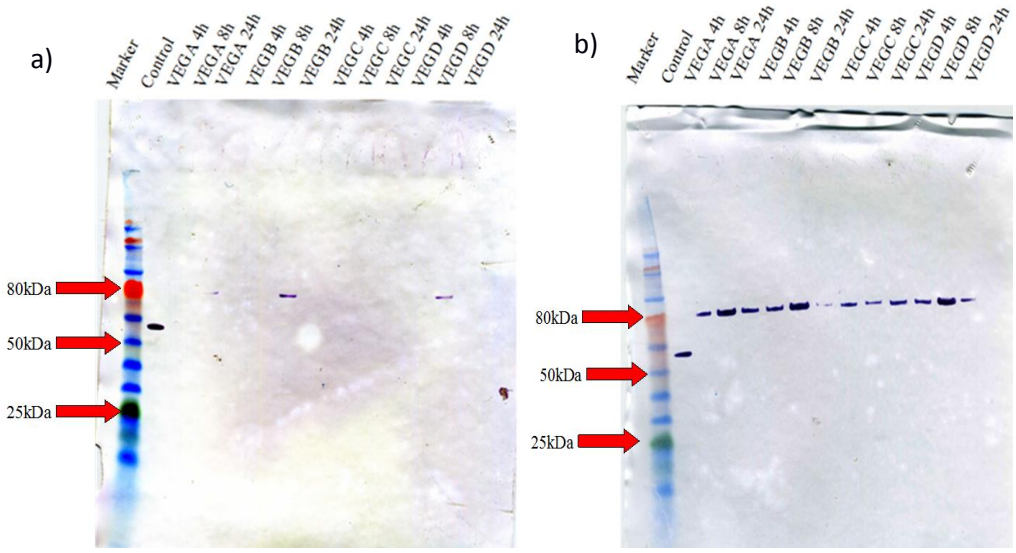


Figure 4-10 Western Blots of the different strains containing anti-VEGF genes.

Shows the activity of the pMTL84153:anti-VEGF plasmids in C. sporogenes through western blot of the Flag-Tag attached to the various anti-VEGF genes. The anti-VEGF protein, with the added flag and his tags is expected to be about 31kDa. The protein shown in the western blots is about 55kDa. Shown in blot A is the supernatant of the C. sporogenes cells and B is the lysates. The control was a purchased protein of known size with a FLAG-tag

To ensure that this was due to a naturally occurring FLAG-tag, and not down to the anti-VEGF antibodies dimerising, stringent denaturing techniques were applied, by boiling the sample for 5 min longer before the western blot, using denaturing bis-tris gels, and ensuring that DTT was in the sample buffer. To further ensure that there was no anti-VEGF being expressed, but not detected due to low concentrations an alternative Na-Deoxycholate TCA (DOC/TCA) precipitation was performed. The higher concentration of protein was used to perform the western blots shown in Fig 4-10. This showed no difference in the outcome of the western blot. Western blots using multiple colonies of the wild type *C. sporogenes* were performed to show whether the bacteria produce a native FLAG-tag. The western blot showed the presence of a false positive for the FLAG-tag and BLASTp showed some closely related amino acid sequences which could account for the potential false positive.

When taking samples for the western blots, the cultures for the different strains were also measured for OD_{600}, to see whether growth differed between the constructs. The OD_{600} of the cultures was greatly reduced compared to WT *C. sporogenes* at the 4 h, 8 h, and 24 h time points. Growth curves of all strains were performed with IPTG, comparing to *C. sporogenes* WT (Fig 4-11).

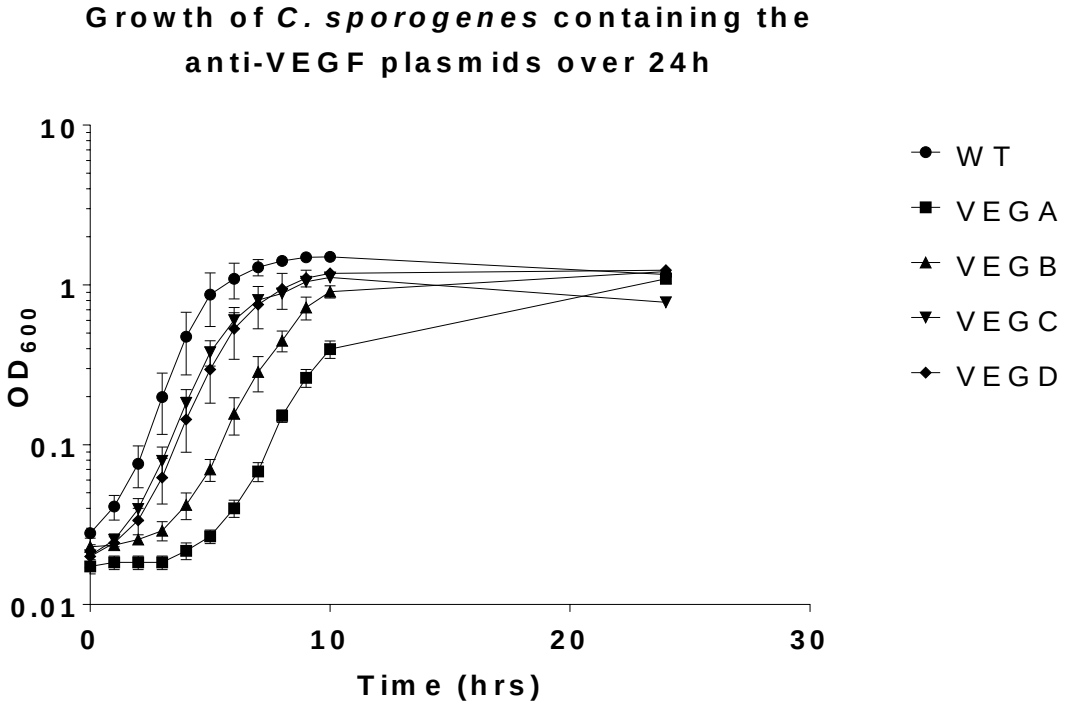


Figure 4-11 Growth curves of strains containing anti-VEGF.

The growth curve of all the different C. sporogenes strains with the inserted pMTL84153:anti-VEGF plasmids. The growth was measured hourly for 12 h, and then a final time point at 24 h. C. sporogenes usually reaches stationary phase within 12-15 h. All time points were read in biological triplicate standard error bars are shown.

The growth curves showed a slightly reduced growth for all strains when compared to the WT strain, which was confirmed to be significant (p value 0.0016) through one way ANOVA. This lack of consistency across the different anti-VEGF containing strains led us to believe that the anti-VEGF variations were differentially toxic, and so to improve upon the system for successful production the genes should be cloned into

the pET28a plasmid, for *E. coli* expression (as a model), and the inducible pMTL82251:YZ2 plasmid for *C. sporogenes* expression.

4.4.1 Anti-VEGF Cloning

The cloning strategy for the new anti-VEGF plasmids was the same as the one applied for the *cpg* genes. The restriction enzymes were changed to ensure no cuts occurred in the open reading frame of the insert. The growth of the transformed bacteria was carried out at both 30^{o}C and 37^{o}C to improve the chances of obtaining a correct plasmid. While many colonies were obtained for the four different anti-VEGF plasmids, once screened and sequenced none of these plasmids were correct, bar for the anti-VEGF_B.

Alternative attempts to produce a plasmid using Gibson assembly were performed. Primers were designed that would both amplify the regions of importance and introduce overlaps in each oligo. These oligos were then used for Gibson assembly whereby the overlaps were digested at the 3' end by an exonuclease and then used as sticky ends for ligation. The 3' ends are extended by a polymerase and then the nicks in the DNA are closed by using a DNA ligase (Gibson et al., 2009). The Gibson assembly consisted of three main cassettes, one with the origins of replication for both *C. sporogenes* and *E. coli* host, and an antibiotic resistance gene for selection. The other cassette consisted of the inducible system that has been described for the *cpg* genes containing a promoter, a *lacI* gene, and a terminator on the antisense strand and on the sense strand further downstream a promoter, *lacO*. The anti-VEGF genes (A, C, D) along with a terminator were the final cassettes. Each different variation of the anti-VEGF genes was incorporated once with the origin of replication cassette, and the *lac* cassette. Through screening and sequencing of the *lacI* cassette and the anti-VEGF_A, anti-VEGF_C. and anti-VEGF_D cassettes, it was shown that

there were two more successful constructs. The correct plasmids (pMTL82251:YZ2:anti-VEGF_A, pMTL82251:YZ2:anti-VEGF_B, and pMTL82251:YZ2:anti-VEGF_D) were transformed into both *E. coli* BL21 (DE3) Star, to study the expression of the anti-VEGF antibody in *E. coli*, and the conjugation donor strain *E. coli* CA434.

4.4.2 Expression Part 2

Expression studies of the anti-VEGF genes in *E. coli* were performed in the same manner as has been described before. Samples were taken at 4 h, 8 h, and 24 h and the lysates of the cells and the supernatant of the culture were both used in western blots. The supernatant was first concentrated through Trichloroacetic acid (TCA) protein precipitation. The different samples were then used in a colourimetric western blot (shown in Fig 4-12)

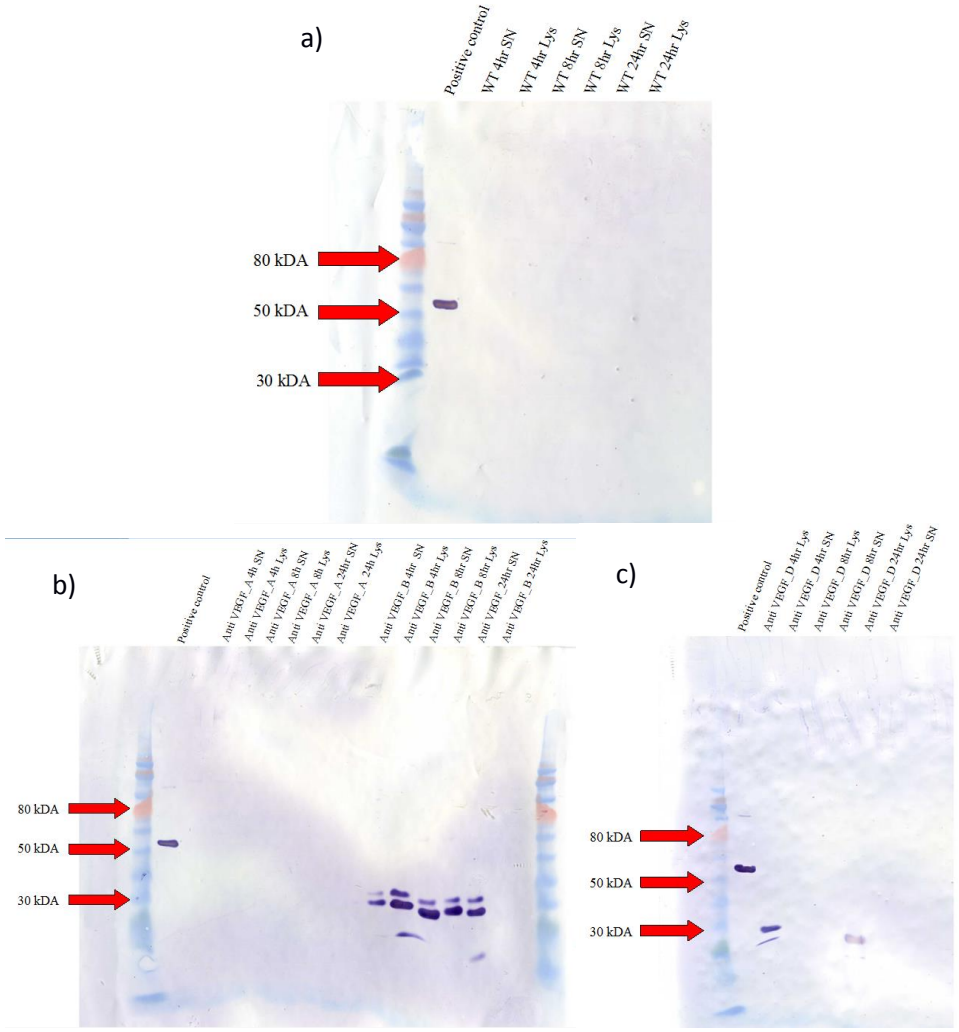


Figure 4-12 Western Blots to confirm anti-VEGF production in the new plasmids, through the FLAG-tag

The western blots of the E. coli strains containing pMTL82251:YZ2:anti-VEGF plasmids. Samples were taken at 4 h, 8 h, and 24 h to show early exponential, late exponential and stationary phase. Wild type E. coli were used as a negative control for 4 h, 8 h and 24 h time points, and a standardized Flag-Tag protein was used for a positive control. a) shows the wild type controls b)shows anti-VEGF_A and anti-VEGF_B c)shows anti-VEGF_D. All anti-VEGF genes were produced with a flag tag and murine Monoclonal ANTI-FLAG® M2-peroxidase (HRP) antibody was used with HRP substrate 3,3',5,5'-Tetramethylbenzidine.

Of all the different constructs at the different time points, only pMTL82251:YZ2:anti-VEGF_B showed expression, which was present in both the supernatant and the lysate. It was shown in the lysate from 4 h, and the supernatant from 8 h Anti-VEGF_D containing plasmids did show some activity, but its presence was erratic.

The band was exhibited at the correct size for the monomeric version of the protein (about 30kDa).

After expression had been assayed in *E. coli* BL21 (DE3) Star for the numerous plasmids, all of three anti-VEGF plasmids were conjugated into *C. sporogenes* from *E. coli* CA434. All the different constructs exhibited antibiotic resistance once conjugated into *C. sporogenes*, so the plasmids were extracted and were sequenced to ensure that the anti-VEGF genes were correct. Despite exhibiting antibiotic resistance from the *ermB* gene on the plasmid backbone there were no correct anti-VEGF sequences found in *C. sporogenes.* The plasmid was extracted transformed into *E. coli* to confirm the plasmid was present.

The evidence for the anti-VEGF gene variants being toxic to *C. sporogenes* was strengthened through the difficulty in conjugating the plasmids into *C. sporogenes*. The fact that no correct insert could be found in the colonies that were exhibiting antibiotic resistance on the selective plates, even after correct sequencing in the conjugation donor, left two potential sources for what was causing the issue. The first was the conjugation process, and the second was *C. sporogenes.* To remedy a strain specific issue an alternative clostridial host was selected, *C. acetobutylicum*. To ensure that the process was not the issue the plasmid was inserted into both *C. sporogenes* and *C. acetobutylicum* by electrotransformation, and through conjugation. All of these attempts still did not result in any correct, viable colonies.

Concurrently to the sequencing of the plasmids from *C. sporogenes,* western blots were performed on samples taken at 4 h, 8 h and 24 h after introduction of IPTG in the culture on 3 h time point. This was performed on colonies, which were yet unconfirmed by sequencing. The samples were treated in the same manner as with

E. coli, so that lysates of the cells, and proteins precipitated out of the supernatant were run on the western blot (Fig 4-13).

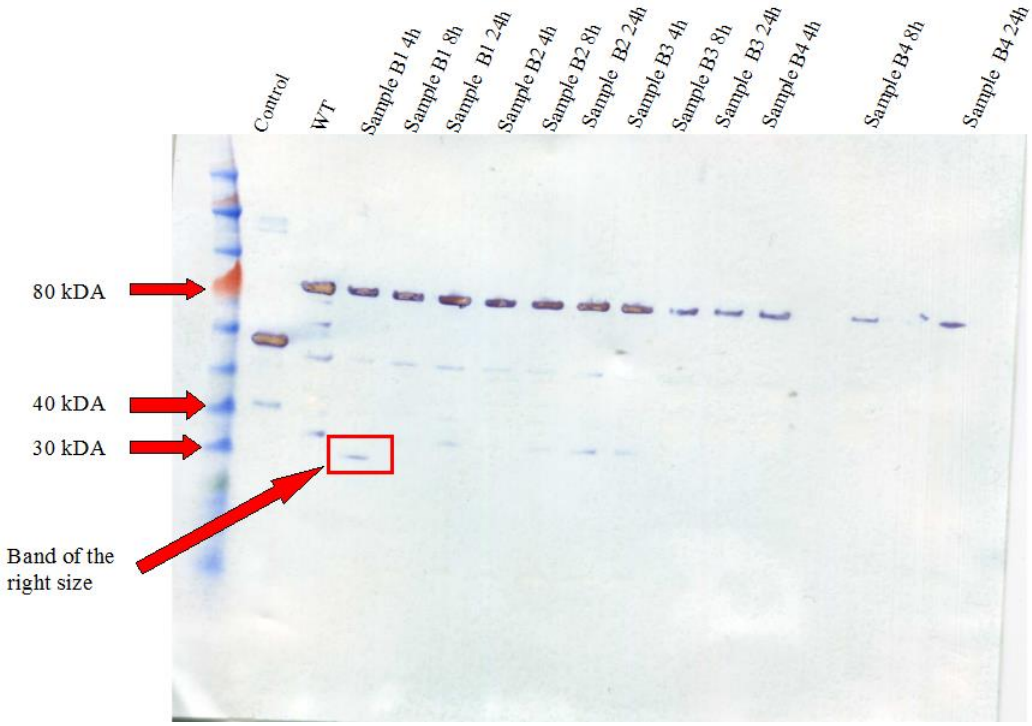


Figure 4-13 Western Blot of anti-VEGF in *C. sporogenes.*

The western blot of the C. sporogenes strains that contain the pMTL82251:YZ2 plasmid containing the various versions of the anti-VEGF gene which had been successfully cloned at the normal three time points of 4 h, 8 h and 24 h.

As can be seen in Fig 4-13, there is a very faint band of the right size for the monomeric version of the anti-VEGF antibody. This is exhibited once again by the anti-VEGF_B plasmid. Due to the incorrect sequencing data obtained of this plasmid, it was extracted and transformed back into the *E. coli* BL21 (DE3) Star host, and once again tested for expression by western blotting of the protein precipitated supernatant, and the lysates of the cells. This did not result in a band for any of the different time points. Plasmids of the back transformed *E. coli* were also screened and sequenced, and while the antibiotic selection was evident there was no correct sequence for the anti-VEGF_B gene. While it was clear that the correct plasmid was

present in the *E. coli* CA434, once conjugated into *C. sporogenes* there was no evidence of the anti-VEGF being maintained in the bacteria.

The final attempt to express the anti-VEGF genes was through the utilisation of the *tcdR* regulator. A strain of *C. sporogenes* had been engineered where by using the protocol Allelic Chromosomal Exchange, a copy of the *tcdR* gene had been incorporated into the chromosome. The gene when expressed in *C. difficile,* its native organism, activates the toxin genes *tcdA* and *tcdB*. It is a specific sigma factor for the promoters of toxin A and B, *PtcdA* and *PtcdB,* respectively. By binding downstream of the promoters it allows transcription of the *tcdA* and *tcdB* genes. The system utilises this by having the anti-VEGF being driven by the *PtcdB* promoter, and the *tcdR* gene being expressed from the chromosome.

A single anti-VEGF gene was tested in this system, anti-VEGF_B. The plasmid with the *PtcdB* promoter had the anti-VEGF_B gene cloned into it. This was confirmed through sequencing. There was no way of confirming expression in *E. coli*, due to the lack of the *tcdR* gene in these bacteria. The plasmid was transformed into *E. coli* CA434 for the conjugation into *C. sporogenes*. This was successful and colonies exhibited antibiotic selection. These colonies were picked and as before, western blots were performed on *C. sporogenes* lysates and supernatants. Sequencing was performed to confirm the presence of the gene and the expression. Unlike previous attempts there was no issue with the sequencing in plasmids extracted from *C. sporogenes*, however there was still no expression of the gene detected by western blot in this organism.

4.5 Discussion

4.5.1 CPG_2

In both *C. sporogenes* NCIMB 10696 and *E. coli* BL21 (DE3) Star, CPG_2 protein has been shown to be expressed. This was shown through the phenotype exhibited where a ring of pteroate precipitate is shown around bacterial colonies when there is Folic acid in the media. Successful expression was achieved using a *lac* inducible system, with expression only being exhibited once CPG_2 was cloned in the pMTL82251:YZ2 or the pET28a plasmids. The quantitative assays of the bacterial lysate and the bacterial supernatants showed most of the enzymatic activity in the cell lysates, there appeared to be some activity in the supernatant but only in very low quantities. This could be explained through the death and degradation of cells in the culture before the samples were taken, rather than the export of protein out of the cells. Unfortunately this situation is not ideal for the CDEPT system as the activity of CPG_2 is required outside of the cell. Most of the activity was also seen at the 8 h time point with a reduction at later points. The increase between 4 h and 8 h can be explained by the relatively small amount of time for an adequate number of cells to produce CPG_2 before sampling was performed, at 4 h. However the decrease between 8 h and 24 h is harder to explain. It could be down to a reduction in the number of viable cells being able to produce CPG_2 as the cell culture will have entered into the stationary phase of growth by this point. What was produced when in exponential growth phase may have been degraded to the point of non-functionality by this point resulting in a low assay result. To clarify the quantity and timing of production of CPG_2 qRT-PCR could be performed. This would show whether transcription of the gene was time dependent and whether it was transcribed at a

high level, giving an idea of why the protein activity is mostly seen at the 8 hour time point.

To accurately assay the supernatant concentration it had to be concentrated. Normal protein precipitation methods denature the protein, rendering it non-functional, and therefore quantitative measurement is difficult. For that reason concentration through ultrafiltration was performed. This gave a better methodology for the quantification of the enzyme in the supernatant. With the improved methodology it was confirmed through the assays of the supernatant that only a small percentage was being exported outside of the bacterial cells.

The lack of enzymatic activity in the supernatant could be caused by two different issues, first what is being exported outside of the cell could be degraded by secreted *C. sporogenes* proteases. *C. sporogenes* has long been known to be a non-pathogenic counterpart to the Class I proteolytic *C. botulinum,* and through study of *C. botulinum* we know that there is an operon of six proteolytic enzymes present. Due to the degree of homology between the two species of bacteria it was hypothesised that the CPG_2 protein once produced, is exported alongside the metalloproteases that *C. sporogenes* produces. To this end the bacterial genome of *C. sporogenes* was searched for the metalloprotease operon that is functional in *C. botulinum* and it was found. This situation is problematic as the metalloproteases are suspected as being key for *C. sporogenes* to colonise the hypoxic regions of the tumour. To counter the proteases *in vitro* and to show that it was these proteins inhibiting the CPG_2, *C. sporogenes* strains were grown in the presence of protease inhibitor cocktails. The protease inhibitor cocktails should inhibit some of the metalloproteases that are expected to be produced by *C. sporogenes* allowing for an increase in the level of CPG_2 protein seen active in the enzymatic assays. The assays of these cultures clearly

show a significant difference in the amount of CPG_2 activity present in the supernatant.

As the proteolytic enzymes are produced after about 4 h of growth there might not be enough present in the bacterial culture to really break down the CPG_2 when 8 h samples are taken. However the production if left uninhibited would result in enough accumulated metalloproteases to digest enough CPG_2 that by the 24 h timepoint not enough was present for a positive result in the methotrexate assay. With the addition of the Protease inhibitor cocktails the metalloproteases were inhibited enough for a CPG_2 to accumulate and remain in the medium until the 24 h mark where originally it would have been degraded.

The second issue could be the signal sequences. The *eglA* signal sequence has been confirmed to be active in other clostridial species (Theys et al., 1999), but never been confirmed in *C. sporogenes.* While the transferral of *Clostridial* sequences between different *Clostrial* species often works, this is not always the case, and it might not be the optimal signal sequence. The signal sequence is, therefore, a key factor in having an efficient system. This is why the original *cpg* genes isolated from *V. paradoxus* were attached to a clostridial signal sequence through splice overlap extension (SOE) PCR. When attempts were made to isolate a natural alternative *cpg* gene as a comparison to the codon optimised versions they were attached to the same signal sequence through the same method. However the assays show that the level of protein in the supernatant could be increased through inhibition of the proteases in the media, and through the amount of activity in the cell lysates. This would indicate that the signal sequence should be looked into, as a better signal sequence would export more from the cell.

There is a great deal of optimisation required for the improvement of the system, and the development of this system could potentially set the stage for numerous different prodrug and prodrug converting enzymes being implemented into the CDEPT system. It could also be used to implement a clostridial based direct action therapy, when a suitable drug can be expressed in *C. sporogenes*.

4.5.2 Anti-VEGF

After numerous different attempts, and many different protocols being attempted to clone anti-VEGF into a suitable plasmid for expression, it was thought that the production of anti-VEGF in *C. sporogenes* was unobtainable. Expression of the four anti-VEGF genes in *E. coli,* a model expression system, with a model expression vector for this organism in the pET28a plasmid was incredibly difficult. Only one gene was produced of the four variants in *E. coli* and this gives an indication that this antibody may in some way be toxic to the bacterial cell. Potentially anti-VEGF_B, being codon optimised for *C. sporogenes*, and in an inducible system reduces the level of expression enough so that it might be produced without too many issues. When introduced into *C. sporogenes* the level of expression should be higher due to the optimisation of the system (i.e. more suited ribosomal binding site, promoter and codon optimisation) and therefore results in issues with toxicity.

When the genes were sequenced from *C. sporogenes* they showed truncated versions of the gene at best. This indicates that while the bacteria needed the plasmid to maintain antibiotic resistance, it also could not at the same time have the anti-VEGF genes present, potentially these genes are toxic to the bacterial cell. When looking at the bacterial colonies on a selective plate, the morphology was always irregular, indicating that the cells were not healthy. The fact that the plasmid was sequenced correctly prior to being inserted into the *C. sporogenes* cell, and then was

never able to be sequenced correctly from plasmids exhibiting antibiotic resistance, indicates that it was either a *C. sporogenes* defence mechanism, or an issue with conjugation of this plasmid. It can be confirmed that this is down to the bacteria rejecting the gene in some way, due to the experiments performed on transformation and conjugation of two different clostridial strains*, C. acetobutylicum* and *C. sporogenes*. We can confirm that the bacteria are not the causing the truncation of the anti-VEGF genes due to the fact that it occurs when in alternative bacteria, shows that there must be a common cause. Without studying the cause of the loss of function of the gene it is unknown how the truncation occurs. To elucidate whether it was the bacteria or the plasmid insertion protocol there were attempts made at transforming the plasmid into *C. sporogenes.* Attempts were also made to conjugate, and to transform the plasmids into an alternative clostridial host *C. acetobutylicum*. These were all unsuccessful, leaving the only option being that the gene is toxic for clostridium. It is unknown how the anti-VEGF antibody is causing bacterial toxicity, it could be that despite attempts to integrate the gene into an inducible system the *lac* inducible system is too leaky to allow correct constructs to be maintained in the *C. sporogenes* host.

To attempt to express the anti-VEGF_B gene in *C. sporogenes* an orthologous expression system based on the *tcdR* and *tcdB* promoter system was thought to be a good alternative to the *lac* system, as within this system the gene would be firmly repressed until in the presence of the TcdR sigma factor, something which is only naturally found in *C. difficile*. This allows for correct insertion of the gene, however, when inserted there is no production. Potentially the level of expression of *tcdR* is not high enough to cause effective expression of the anti-VEGF_B gene. This

alternative inducible system has also failed in stable production of anti-VEGF, this leaves very little that can be done to implement anti-VEGF.

There are alternative inducible systems that could be looked into for further attempts at insertion of the anti-VEGF and consequent production, such as the T7 Promoter system, the tetR system (Fagan and Fairweather, 2011), and the Rhamnose system. These alternative inducible systems could provide a way to tightly control the production of the toxic anti-VEGF until required. This is however unlikely given that the alternative attempts have already failed.

This would still leave the problem of production optimisation. If the production of the anti-VEGF gene is put into context of treatment, having a drug product that is so toxic to the cell being produced is problematic. No sooner than the expression is induced it will cause mass bacterial death and therefore the expression of the gene will cease. This situation would therefore require multiple doses of a spore suspension for the production of the any anti-tumour effect and the efficacy of the drug would be low. This would consequently have very little effect on the vasculature and would result in the patients' combinatorial treatments like chemotherapy still not being very effective. With this in mind it was deemed that the use of anti-VEGF as a direct action therapy would not be feasible.

However this does not mean that the use of small direct action molecules is not a feasible strategy. Due to the complexity of the cancer cells vascular growth system there are numerous other potential angiogenic targets that could be used to the same effect as anti-VEGF, such as Platelet Derived Growth Factor-B (PDGF-B). There are also factors that can be used in treatments that are not angiogenic factors such as apoptotic factors, by incorporating one of the caspase initiators, like caspase-8 or caspase-9, into the bacterial host it might be possible to initiate the cancer cells

apoptotic process. Having a range of possible mammalian molecules implemented in the clostridia would be beneficial as it would give a range of treatment regimens which could be used in the case of innate and acquired resistance to therapeutics

Optimisation of CDEPT for Anti-VEGF and CPG_2

Chapter 5

5 Optimisation of the CDEPT System.

5.1.1 Introduction

With the expression of *cpg* gene confirmed in the proposed host *C. sporogenes*, optimisation of the system was required. It had been shown that there were some important issues with the way that the system was being implemented that would limit efficacy. The key problem was the Carboxypeptidase G_2 protein folding within the cytoplasm, and not being exported, where it could have an effect upon the prodrug. This could be caused by the signal peptide, EglA from *C. acetobutylicum* (Theys et al., 1999), which has not been implemented in *C. sporogenes* before and might not be appropriate. This issue is compounded by the proteolytic nature of the bacteria as when any CPG_2 protein is found outside of the cell, it is cleaved by the metalloproteases produced by *C. sporogenes*.

Recombinant proteins can be targeted to multiple different cellular locations, when required. For the purpose of this study it was proposed that the best place for the protein to be functionally active is in the extracellular fraction. To do this the original sequence sent to DNA2.0 for codon optimisation was amended to include a signal sequence, from the *EglA* gene, which had been shown to work in other clostridia, namely *C. acetobutylicum* (Theys et al., 1999). However the experiments performed in *Chapter 3* of this work suggested that this signal peptide is not efficient in *C. sporogenes*. For optimisation of the system alternative signal peptides must be found. There are several types of bacterial secretion systems for Gram-positive bacteria, some of which are only poorly understood (Freudl, 2013).

There are several methods for the translocation of proteins such as the Sec pathway, and the Twin-Arginine Translocation (TAT) pathway. The best known is the Sec

pathway, which has been extensively studied for Gram-positive bacteria in *Bacillus subtillis*. Through high levels of conservation between different Sec systems, it is hypothesised to be relevant to *C. sporogenes.* The Sec pathway has two ways of working, a ribosomal co-translational manner, or a post-translational manner. In a post-translational process, the translated proteins are unable to fold correctly due to post-translationally interacting proteins (PIPs) such as SecA which recognizes the signal peptide to bind to the amino acid chain. This incomplete state allows the protein to pass through the membrane channels of the bacteria. The SecA protein serves a second purpose from inhibiting the folding of the protein as it also recruits the protein to the SecYEG protein which is a conducting channel for translocation through the membrane. The movement of the unfolded protein through the SecYEG channel is powered by the energy provided through SecA, which as an ATPase cleaves ATP and provides energy for translocation of the protein. The unfolded protein will pass through the channel in an ATP dependent manner. In addition to this the SecDF protein, a transmembrane protein located adjacent to the SecYEG pore, acts as a proton motive force giving extra influence in the movement of the protein through the SecYEG channel.

In a co-translational process the signal peptide is recognised before complete translation of the protein has occurred. The co-translational signal peptide, which is highly hydrophobic typically, is recognised by the Signal Recognition Particle (SRP) and binds. This developing chain-SRP complex binds to another protein, FtsY. Consequently the protein complex chaperones the ribosome and the developing amino acid chain to the SecYEG translocational channel where the exit site of the ribosome is close to the entrance of the SecYEG protein. The energy required to produce the translation provides the energy for the translocation to occur.

The components of this system have been well characterised, and so looking for homologs of the *sec* genes in other organisms can give an idea of whether there is a functional system present. Using the genome of *C. sporogenes* all components of the sec system, such as the genes responsible for the transmembrane proteins SecYEG, were identified as present bar the *secB* gene. As the *secB* gene codes for the protein chaperone this could limit the efficacy of the system as there would be no way of preventing the folding of the protein before it is translocated, and no way for the transmembrane proteins to identify the protein to be translocated. As there are many proteins that have no assigned function within the *C. sporogenes* genome one of these may be a surrogate for the *secB* gene, allowing the system to function without a recognised form of *secB*.

For both processes, either after or during the proteins passage through the channel, the signal peptide is cleaved by a signal peptidase (SPase). With the signal peptide gone, and the PIPs no longer attached to the amino acid chain, it is free to fold correctly and become the active protein. There are many different SPases for the different signal peptides, however, the signal peptides follow a typical structure comprising of an N-terminal positive charged sequence, a hydrophobic region, and a polar carboxyl terminal region. The average length of the peptide is about 32 amino acids (Kudva et al., 2013, Freudl, 2013, Fagan and Fairweather, 2011).

The Twin-arginine translocation (Tat) pathway is an alternative secretory system to the Sec pathway. This pathway utilises the same type of signal peptides that the Sec pathway does with the same charged, hydrophobic, and a polar carboxyl terminal functional domains. The Tat pathway is different from the Sec pathway as rather than the channel through the membrane recruiting proteins to translocate, the protein here has a pore created for it. It also differs as the protein can be folded

before movement through the pore. The Tat system has two protein complexes, a docking complex and a membrane pore complex. There are two main groups of proteins involved in the Tat system, the TatA, TatB and TatE proteins are one group, and then TatC which is a large transmembrane protein is the other. The docking complex consists of a TatA like protein, and TatC. When the protein that is to be translocated across the membrane is recruited to the docking complex, numerous TatA like proteins are also brought to the area. These bind together to form an appropriate pore that will allow the protein that has formed a complex with the docking proteins to be translocated across the membrane (Goosens et al., 2014). Once the protein has passed through the newly formed pore, the signal peptide is removed.

Very little is known about the *C. sporogenes* secretory systems, but from the genome sequence we know that the *sec* genes are present. While this is promising for the utilisation of the system, it should not be taken for granted that the genes that are present are functional in *C. sporogenes*. Unlike what has been shown in *C. difficile* (Fagan and Fairweather, 2011), *C. sporogenes* has just one gene encoding SecA. The tat system was searched for in the complete *C. sporogenes* genome sequence and no genes matching either the gene ontology or the tat gene sequences were identified in the genome. From this it is hypothesised that there is no functional Tat system.

Due to the similarity in signal peptide functional domains, prediction of the signal peptide can be automated through the use of neural networks. One of the most common tools for the prediction of signal peptides is the SignalP program. The SignalP4.0 program uses a neural network to identify the signal sequence, where negative feedback of transmembranous proteins was used to improve upon the specificity of identification of the signal peptides (Petersen et al., 2011).

With signal peptide identification software, and knowledge of what type of systems typically secrete proteins in a Gram-positive bacteria, we can utilise these systems to ensure that the CPG_2 PCE is exported out of the bacteria into the extracellular fraction of cultures where it would be required to work in an *in vivo* experiment.

When the protein is successfully exported outside of the bacterial cell the next potential issue with CDEPT arises. The organism *C. sporogenes* shows great homology to the organism *C. botulinum*. In sequencing of the *C. sporogenes* genome the metalloprotease locus of the *C. botulinum* bacteria was found to be present in part in the *C. sporogenes* genome. The metalloproteases are excreted proteins in *C. botulinum*, and is likely to be the same in *C. sporogenes* (Sebaihia et al., 2007). The metalloproteases are endopeptidases cleaving non-terminal amino acid sequences (Matthews, 1988). This would lead to a breakdown of any CPG_2 protein when it is efficiently exported which would inhibit the efficacy of any *in vivo* studies.

To limit this, the best option would be to knockout the proteases; however this produces an alternative problem. The proteases are hypothesised to be behind the high levels of colonisation of hypoxic tumours by *C. sporogenes*, by knocking these out in order to implement a PCE would be counterintuitive. They also potentially cause the oncolytic effect that is exhibited by some clostridial species. Therefore an alternative route to protecting the PCE must be found. One proposal was to anchor the protein to the cell wall. This would help through making the PCE less available to being broken down by the metalloproteases. One of the most extensively studied ways of doing this would be to utilise the sortase system of *C. sporogenes*.

The sortase system is a well-studied system, in particular in the organism *Staphylococcus aureus.* In *S. aureus* the sortase system is used as a facilitator to binding to human tissues. During infection, *S. aureus* binds to squamous, keratinized

cells, at the anterior of the nares. This binding is facilitated through some of the proteins that are attached to the cell wall of the bacteria such as clumping factor B (*ClfB*) (Weidenmaier et al., 2008). These proteins are bound through the sortase system, and as such the system was studied as an alternate target for treatment of *S. aureus* bacteria that can also be methicillin resistant (MRSA).

The sortase system works through cleavage of a specific amino acid sequence, LPXTG, where the X can be any amino acid. The cleavage occurs between the threonine-glycine bond (Paterson and Mitchell, 2004), and leaves a c-terminal cleavage site which can consequently bind to the cell wall. The binding to the cell wall occurs through the formation of an amide bond between the C-terminus of the newly cleaved threonine, and the available amino group found at the end of the wall crossbridge (Navarre and Schneewind, 1999).

While this standardised view of how the sortase cell wall anchoring works, it is based on the model organism for *S. aureus* sortases. The cell walls of bacteria can be highly dissimilar even between Gram-positives and therefore this standardised view of what happens might not be applicable for other bacteria. There are alternative factors which can also have a negative impact on the efficacy of the sortase system from one protein to the next. The Sortase LPXTG signal is known to have very similar functional domains to that of signal peptides, in that it has a recognition sequence for the sortase enzyme to cleave, a hydrophobic domain and is followed by a charged tail. The charged tail and the hydrophobic domain can vary from protein to protein and are primarily used to maintain the protein at the cell wall, until the amide bond can be formed. So the size, charge and nature of the protein in relation to the LPXTG motif will have a role in efficacy (Mazmanian et al., 2002).

Normally the proteins that have been anchored to the cell wall have functions involved in virulence, toxicity and colonisation. In *S. aureus* the use of the sortase system is for facilitating colonization of the nares, which also has a toxic effect. There have been papers describing sortases and sortase substrates in other *Clostridium* species, but never in *C. sporogenes* (Donahue et al., 2014, Pallen et al., 2001). While there has been some research done into the sortases, the research is limited and the role of sortases in *Clostridial* species is not fully understood. While in some species like *C. difficile* there are surface layer proteins that allow adherence to the mucosal lining of the gut, the proteins that aid in this binding have no LPXTG motif, and are not associated with the sortase system (Fagan et al., 2011a, Waligora et al., 2001).

This study attempts to establish a better export protocol for the CPG_2 protein in *C. sporogenes,* to characterise and utilise the *C. sporogenes* sortase system for the anchoring of the CPG_2 protein to the cell wall, and to assay the effect this has on the enzymatic activity of CPG_2 in the presence of the metalloproteases that are produced by *C. sporogenes*.

5.2 Results

5.2.1 Identification of Alternative Signal Peptides.

From numerous sources a selection of highly exported proteins was compiled. Highly exported proteins such as Barnase, a ribonuclease which is only functional in the extracellular fraction due to the tight binding to barstar in the intracellular fraction, the CipA proteins from *C. thermocellum* and *C. acetobutylicum,* CipC from *C. cellulyticum* which are all scaffolding proteins from the cellulosomes in these organisms were identified (Drs Katrin Schwarz and Katalin Kovacs, personal communication). Added to the collection were the signal peptides from the highly exported metalloproteases of *C. botulinum* Hall A strain. It was hypothesised that

what might work best would be a native signal sequence from *C. sporogenes*. To this end the identification of exported proteins in *C. sporogenes* was attempted.

Table 5-1 The table of the different signal sequences to be implemented

Signal Sequence	Function of Gene	Bacteria
Barnase	Ribonuclease	*C. acetobutylicum*
CipA	Scaffolding protein	*C. acetobutylicum*
CipA	Scaffolding protein	*C. thermocellum*
CipC	Scaffolding protein	*C. cellulyticum*
Npr	Peptidase Propeptide	*C. sporogenes*
1442	Metalloprotease	*C. botulinum*
1483	Metalloprotease	*C. sporogenes*

To achieve the identification of *C. sporogenes* highly exported proteins a study of the extracellular fraction of *C. sporogenes* NCIMB 10696 was performed. Before beginning it was hypothesised that there would be a highly exported locus of metalloproteases. This hypothesis was based upon the Class I proteolytic *C. botulinum* strains which possess a group of six thermolysin-like metalloproteases, and the genetic similarities of the *C. botulinum* strains to *C.* sporogenes.

The effects of the metalloproteases had been seen in *C. sporogenes* from previous studies where activity of the CPG_2 protein was highly increased in the extracellular fraction when the host was grown in the presence of the Roche cOmplete EDTA free protease inhibitor cocktail (chapter 3). Through study of the completed *C. sporogenes* genome it was found that there was a metalloprotease locus which had five genes coding for the similar metalloproteases (annotated as bacillolysin “like”), however, there appeared to be some truncated versions of the gene and there were less than the six exhibited in *C. botulinum*. There were three large bacillolysin genes and two smaller versions, with a peptidase propeptide and YPEB domain protein positioned in-between the truncated versions of the bacillolysin and the larger versions. The *Npr* signal sequence from *C. sporogenes* which was used, can be seen

replicated throughout the *C. botulinum* metalloprotease locus at the 3' start of all the metalloprotease genes. As this region is where the signal sequence has been identified in the *C. botulinum* metalloprotease genes it could be that the level of exportation would be consistent between the signal peptides identified for *Npr* and the metalloprotease genes (shown in Fig 5-1).

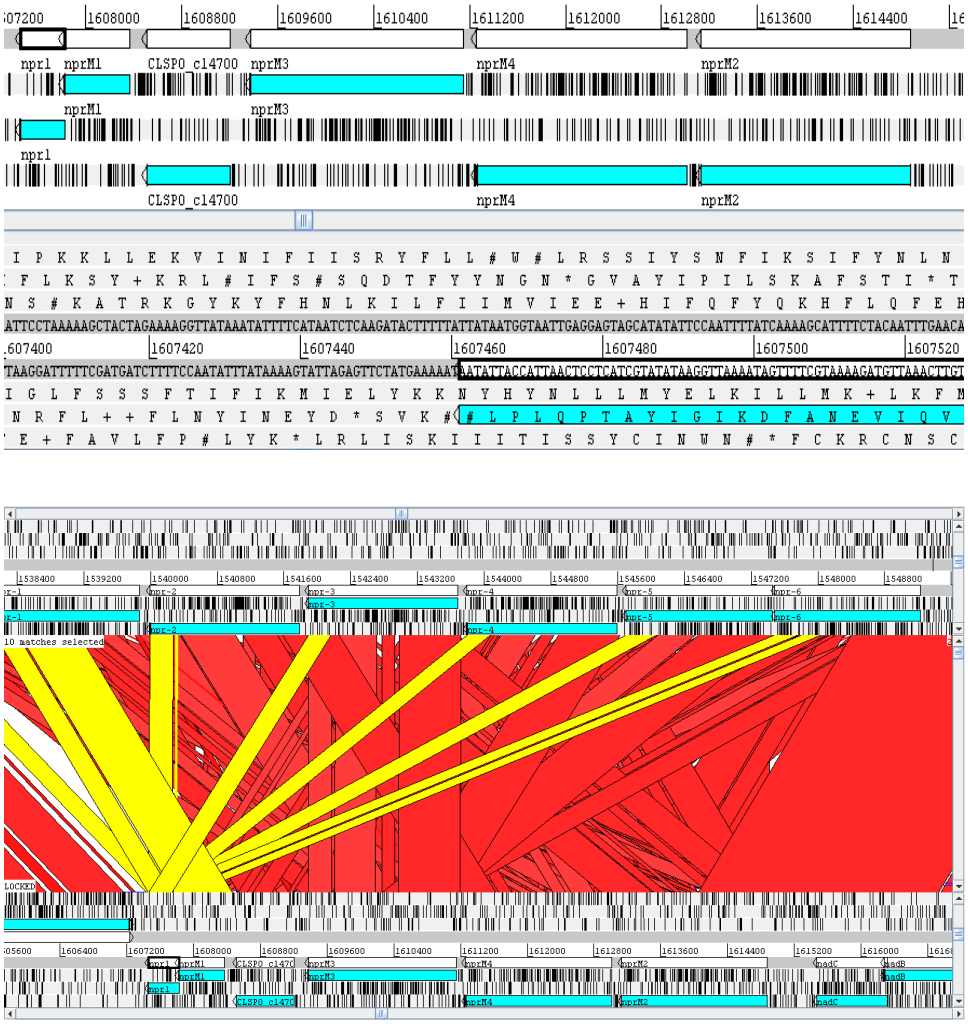


Figure 5-1 The metalloprotease locus.

a) A snapshot of the gene cluster that codes for the metalloproteases, which are causing the CPG_2 to be broken down. The CLSPO_c14700 had the signal sequence identified and has been added to the signal sequence database. B) An artemis comparison was performed of the metalloprotease locus in both the C. botulinum genome and the C. sporogenes genome. This was visualized in the Artemis Comparison Tool (ACT) where the blast file generates bands for each homologous region of genome. The number of bands show that there are many regions of homology between all the genes with several C. botulinum genes matching to any single C. sporogenes gene and vice versa. The truncated Npr gene (highlighted) was identified in all C. botulinum metalloprotease genes at the 3' end of the gene where the signal peptide can be found.

With this knowledge, 2D SDS-PAGE was performed on treated samples of the extracellular fraction of *C. sporogenes* to confirm what was present in the supernatant of the bacterial culture. The ZOOM® system from life technologies was

performed as described in section 2.8.5 of the materials and methods. The sample treatments had to be optimised as early attempts of the 2D SDS-PAGE showed issues, such as, overloaded proteins and lack of proteins solubility. Protein precipitation and ultrafiltration were both performed and serial dilutions from 0 to 10^{-2} were performed (optimisation data not shown). It was found a 10^{-1} dilution, with ZOOM® Protein Solubilizer 1, of the ultrafiltrated samples lead to the best 2D SDS-PAGE when stained with colloidal coomassie. This is shown in Fig 5-2

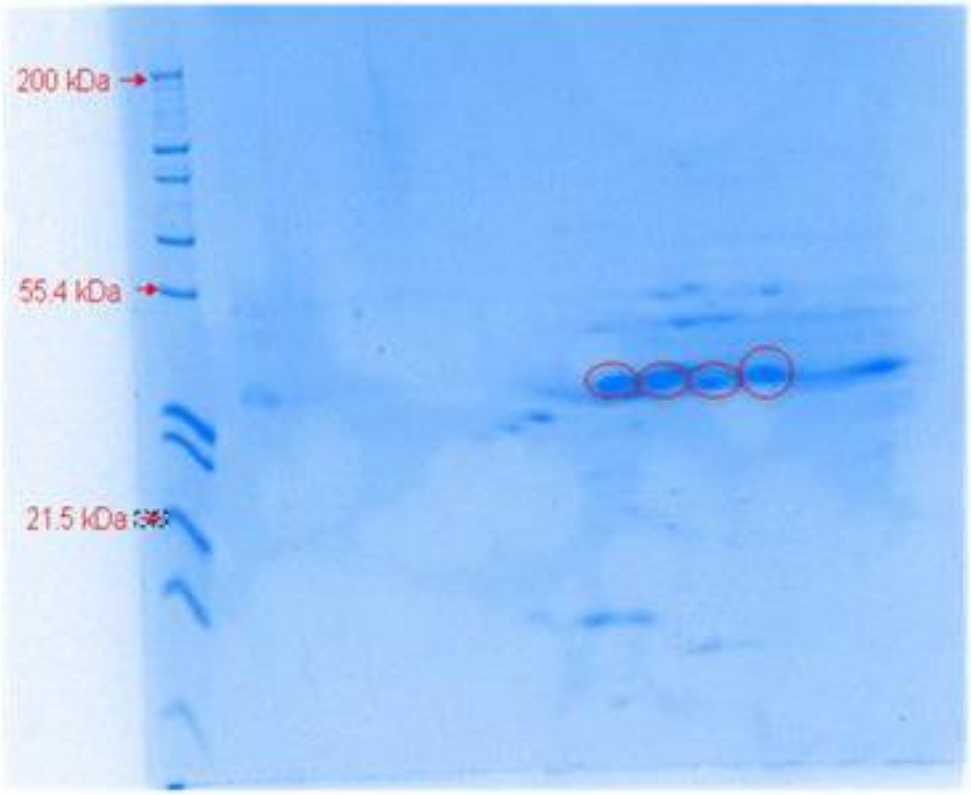


Figure 5-2 2d SDS-PAGE of C. sporogenes supernatant.

To identify any alternative signal sequences that might be used for the exportation of the CPG₂ proteins. This gel shows the results for the optimised zoom protocol which gave the clearest result

The spots seen on the 2D SDS-PAGE correspond to the predicted size of the bacillolysin proteins, and those spots are the most concentrated spots of the supernatant, so these proteins were added to the database of proteins for the signal

peptide analysis. When an SDS-PAGE gel was sent for MALDI mass spectrometry (samples were sent to the PNAC facility at the University of Cambridge) the main components identified were proteases (see fig 5-3). With the database of proteins complete, SignalP 3.0 was used to identify the signal peptides of the different amino acid sequences of the proteins in the database. This was done on the majority of the proteins in the database, however, only one representative signal peptide was taken from the different metalloproteases of *C. botulinum* and of *C. sporogenes* because the amino acid sequence was 100% homologous for all the different proteins.

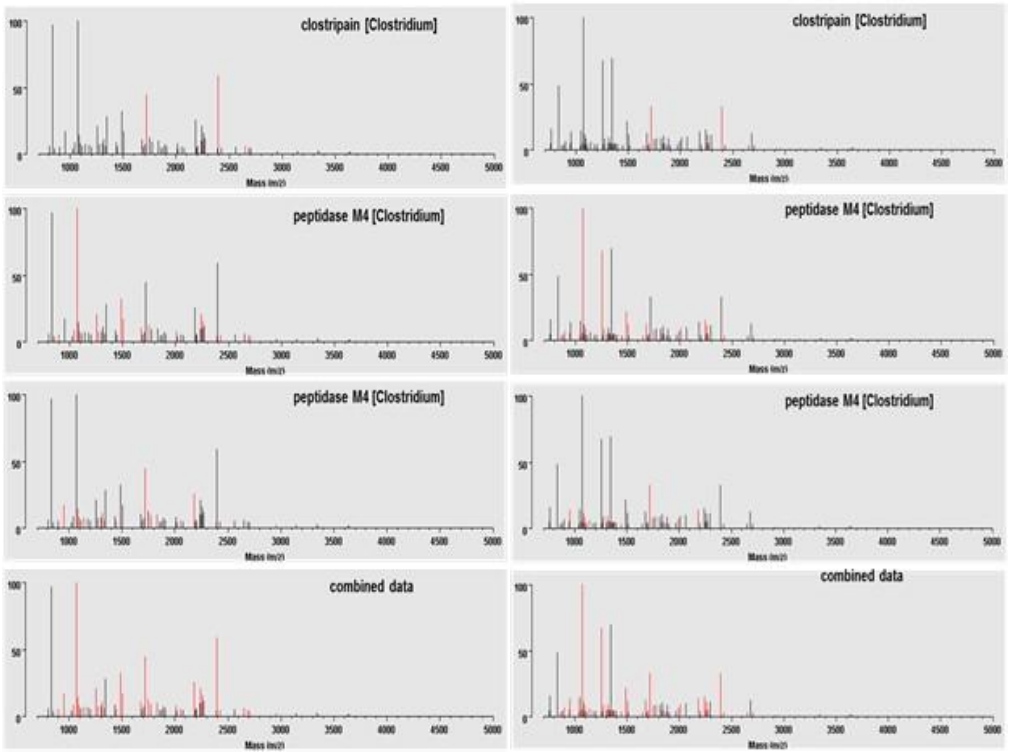


Figure 5-3 MALDI mass spectrometry of the supernatant of Wild type Clostridium sporogenes.

The profiles of the different proteins that were identified from the MALDI mass spectrometry and what protein each profile represents. These proteins were all preoteases such as Clostripains, and peptidase M4. Full MALDI profiles are in the appendix.

The signal sequences were synthesised as oligomers in both the 5' – 3' sense, and 3' – 5' antisense, and restriction enzymes were introduced to allow sticky ends at the 5'

end complementary to the sticky end of the plasmid backbone, and at 3' end for the *cpg_Cs* gene. Oligo melts were performed as described in Materials and Methods 2.6.12. The original *cpg_Cs* was excised from its plasmid leaving behind the *eglA* signal sequence so that there was just the enzyme coded for in the section of DNA obtained, with sticky ends that correspond to the sticky end of the signal sequence at the 5' end, and of the plasmid backbone at the 3' end. All cut sites were designed to be different so that a 3-way ligation of the signal sequences, *cpg_Cs* gene, and the plasmid could be performed. The completed plasmids were transformed into *E. coli* CA434 and colonies obtained were tested for enzymatic activity on phenotype plates. However, no phenotype was observed after repeated attempts, so the seven different signal peptides were codon optimized for expression in *C. sporogenes*, and synthesised onto the *cpg_Cs* gene by DNA2.0. These modified genes were all cloned into the same plasmid host which had been shown to transcribe and translate *cpg_Cs* effectively, pMTL82251:YZ2. There was one signal sequence that was proving difficult to clone into the pMTL82251:YZ2 plasmid which was the native *C. sporogenes* metalloprotease signal sequence 1483, which took much longer to clone and transform. The plasmids where transformed into *E. coli* CA434, and expression tested on phenotype plates. With expression confirmed the plasmids were sequenced and found to have the correct sequence. All bar one of the plasmids were quick to be conjugated into *C. sporogenes*. The one that was not easily cloned or conjugated was the same native signal sequence obtained from one of the *C. sporogenes* metalloprotease genes 1483 that had been problematic to clone.

The 1483 gene and surrounding cluster was re-examined to check for any mistakes in the genome assembly or annotation. Through comparisons of this locus between different assemblies of the genome it was found that there was a degree of

ambiguity about the sequence in this area, probably due to the genes being replicated six times in one area of the genome. Assembler programs will have trouble in identifying the difference between the six versions of the same gene. As the annotation of the genome is still based around the identification of open reading frames and then BLASTing those ORFs for identification, the 1483 signal peptide was still included in any further attempts to conjugate and assess the fraction of enzymatic activity. Conjugation of the plasmid had appeared to happen, however expression of the *cpg_Cs* gene with this signal peptide was never seen. Through the lack of effective production and exportation of the protein, and the reassessment of the genome the 1483 signal sequence was proven to be false.

The other conjugated plasmids were assessed for expression in two fractions the soluble proteins of the intracellular fraction, and the extracellular fraction. They were grown for 24 h in TYG broth, with protease inhibitor cocktail added at 1x concentration. After 3 h IPTG was added, and after 24 h 1 ml of sample was taken. The samples were centrifuged for 10 min at 9400 x *g* and the supernatant was aliquoted out. The supernatant was then ultrafiltrated to a 10x concentration, and the cell pellet was lysed using BugBuster™. The samples were then assayed for activity in both the supernatant and the soluble intracellular fraction. The results are shown in Fig 5-4.

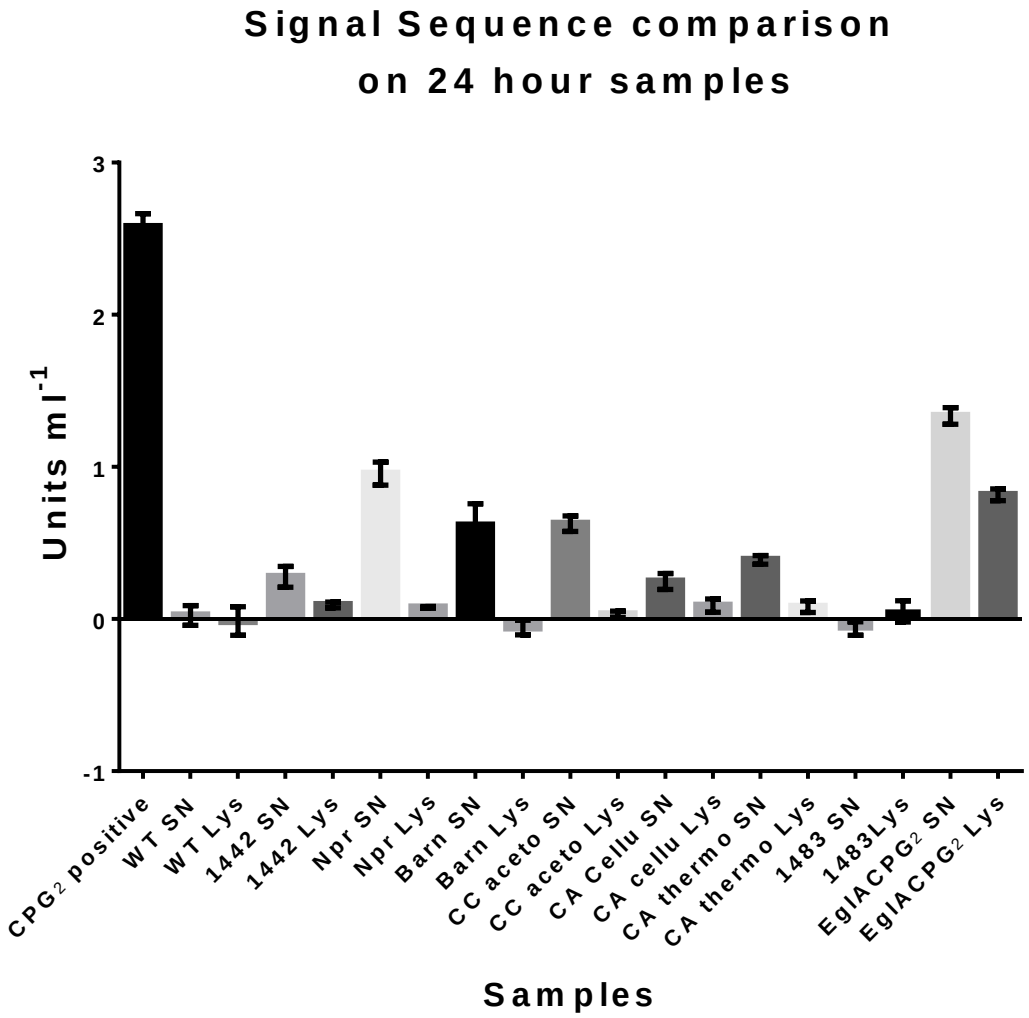


Figure 5-4 Units of activity of CPG_2 in different cellular fractions.

The soluble intracellular proteins were assayed in the MTX assay. The extracellular proteins were concentrated through ultrafiltration, and assayed without any chemical modification. A comparison of these different fractions was performed against the wild type bacteria as a negative control. Each fraction was performed in technical triplicate and standard error bars are shown.

The different signal peptides were compared through a standard student's t-test. Statistical analysis showed that there was a significant increase in activity in the extracellular fraction for all signal peptides against the wild type except the 1442, 1483 and CAcellu signal peptides (p-value between wild type supernatant and Npr was 0.0007, Barn was 0.0196, CC aceto was 0.0019, CA thermo was 0.0068 and EglA

was 0.0001). The activity of CPG_2 in the supernatant is significantly higher in the extracellular fraction for the Npr signal peptide against all other signal peptides bar the Barn and EglA signal peptide, however, due to the large amount of activity for the EglA construct in the lysate it is thought that expression might be higher, rather than a more efficient exportation.

5.2.2 Cell Wall Anchoring of Proteins Using the Sortase System

The sortase system in *C. sporogenes* was studied for the anchoring of the CPG_2 protein onto the cell wall. This was initially looked at through an *in silico* approach. The genome was studied looking for two things, a functional sortase, and any sortase substrates. There were two sortase enzymes, CLSPO_c13060 and CLSPO_c22540, identified in the *C. sporogenes* genome, although one appears to be a truncated version of the other.

The identification of any possible sortase substrates is slightly more difficult, as the key amino acid sequence is LPXTG, which has a variable amino acid for the X. However, as there are just 22 amino acid variations at this point, each variation of the LPXTG motif was blasted against the completed *C. sporogenes* genome. Only one sequence matched any of the possible LPXTG variations. Using the PSI Blast (Position Specific Iterated BLAST) tool, which looks for functional domains, the LPXTG motif found in the genome was input into the PSI Blast tool. All *C. sporogenes* strains sequenced were used as the database. The web tool results were then fed back into the algorithm, to narrow down the set of parameters that the Blast was operating with. This resulted in the same protein that had been input, from all other *C. sporogenes* which had been used in the database. Therefore, it appeared there were no other sortase substrates in the *C. sporogenes* genome. This method was performed by Pallen *et al* (Pallen et al., 2001) when identifying different LPXTG

motifs in *C. acetobutylicum*, where they identified 3 probable sortase substrates and one non substrate which carried the motif LPXTG, however it was unlikely to be a true sortase substrate, due to its short hydrophobic domain and the charged tail.

Due to the sortase substrate efficacy at maintaining the protein at the cell wall before the amide bond is created being variable, it was deemed sensible to develop a motif database to see which one would work the best. Therefore the sortase motifs that were identified from the Pallen *et al* paper were also included in this study (table 1). As before, with all the different variations of the *cpg* gene, the LPXTG motifs were codon optimised for expression in *C. sporogenes*. They were then attached to the *cpg_Cs* gene through SOE PCR. The SOE PCR also introduced the restriction sites that were used to clone the *cpg_Cs* with the LPXTG motifs into the pMTL82251:YZ2 plasmid. The signal peptides used were those of EglA and Npr on the basis that the expression of CPG_2 with the EglA signal peptide appeared to be increased, and Npr was an efficient exporter. The gene inserts which were cloned into the pMTL82251:YZ2 plasmid are shown in table 5-2.

Table 5-2 The sortase motifs.

The amino acid sequence of all the different sortase motifs that were used with the cpg_2 genes, and from which organism the motif originated. The motifs were all codon optimised for C. sporogenes by DNA2.0.

Motif	Amino Acid Sequence	Organism
LPXTG1	LPKTGEFFDATMLLSIALICLASGAILIFVNKKKSSPTK.	*C. acetobutylicum*
LPXTG2	LPKTGSMIDSTVLLIIGTLLLLLGLAFIIWNKFKNKQKSVQ.	*C. acetobutylicum*
LPXTG3	LPKSGGILINLWSLFIVFVIILFFTFLIFKNILRILKKYQPIELIRGA.	*C. acetobutylicum*
LPRTG	LPRTGYSFNFIMLGIGFILISIASIFLIRRQKQLRSK.	*C. sporogenes*

The plasmids once cloned were transformed into *E. coli* CA434 and the bacteria which exhibited the CPG_2 phenotype were sequenced. Once the bacteria were confirmed to have the correct gene, they were used to conjugate the plasmid into *C.*

sporogenes. The colonies that were obtained from this were then put through the same review of phenotypic plates and sequencing as the *E. coli* CA434. The confirmed bacteria were then grown in the same manner as before where samples of a 24 h culture were taken. The culture had IPTG supplement added at a final concentration of 1 μM. They were also supplemented with 1x protease inhibitor cocktail, which was added to the broth before inoculation. The samples were centrifuged down as with the signal peptide samples, and then treated for the separation of the different fractions. This was undertaken as described by Nobbs *et al* (Nobbs et al., 2007), with some alterations. Rather than 100 ml, smaller 10 ml cultures were was grown. Once proteins were resuspended in the suitable buffers they were ultrafiltrated to a 10x concentration to account for the reduced culture volume. The resulting fractions were extracellular, cell wall and protoplast fractions. The different fractions were then tested in an enzymatic assay to assess where the CPG_2 protein was localising. It was first performed with the constructs containing the EglA signal peptide, and then used with the improved Npr signal peptide. This resulted in the results shown below in Fig 5-5.

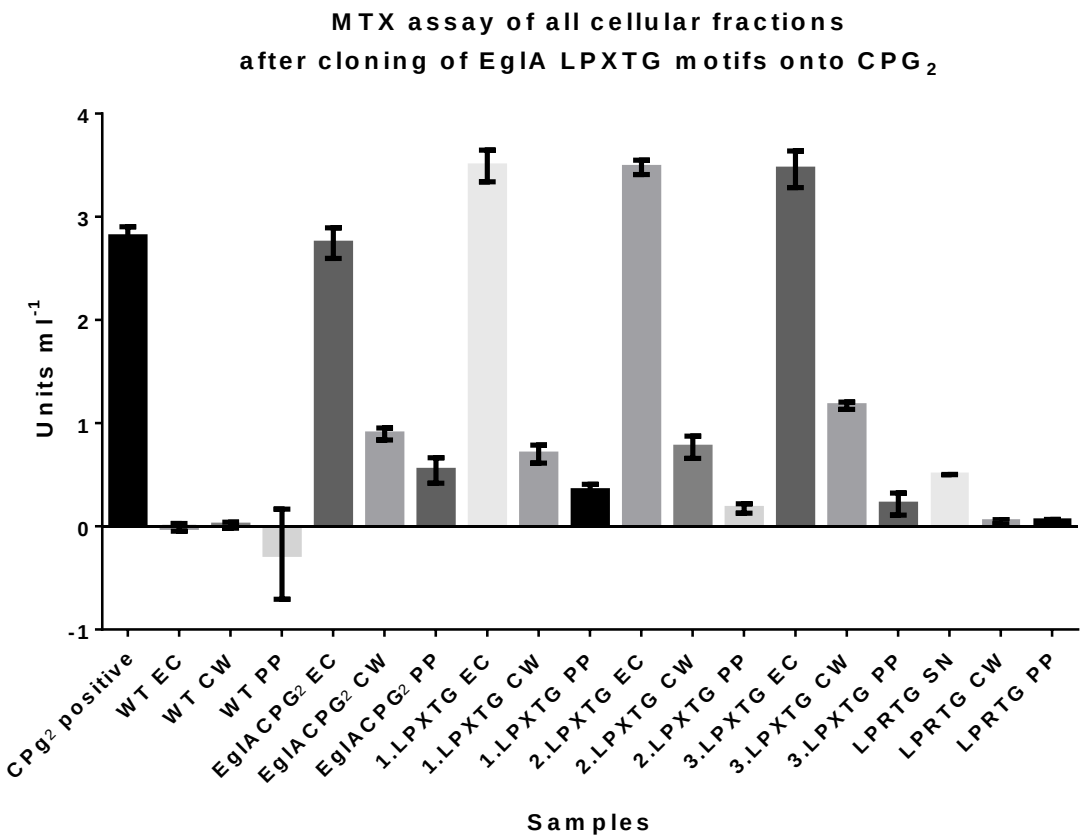


Figure 5-5 The enzymatic activity of CPG_2 in different cellular fractions.

The different bacterial cell fractions, which were the supernatant (EC), cell wall (CW), and protoplast (PP) fractions, were all enzymatically assayed. These fractions were assayed for the different sortase signal motifs (see Table 5-2) which had been associated with Clostridia. All different sortase signals were attached through SOE PCR to the CPG_2 gene with EglA as the signal peptide. Therefore as negative controls both C. sporogenes WT and C. sporogenes containing CPG_2 with the EglA signal peptide and no sortase motif were used. Technical triplicates were performed and standard error bars were used.

The level of expression was limited in the cell wall fraction but there was some indication that it was the sortase motif which was not working, and that the system was fine. Student's t-tests were performed to compare levels of enzymatic activity in the different fractions. When comparing the localisation of the CPG_2 protein at the cell wall there was a significant increase (p value of LPXTG3 against $EglACPG_2$ 0.0154, against 1.LPXTG 0.0002, against 2.LPXTG 0.024, and against LPRTG <0.0001) in the amount of enzymatic activity shown for the 3.LPXTG motif. However to confirm this

an alternative fractionation protocol was used from Tulli *et al* (Tulli et al., 2013). The alternative fractionation protocol resulted in no difference in the localisation of the CPG_2 protein than seen with the original approach.

With the improved signal peptide identified, the 3.LPXTG motif, which had appeared to work best with the previous signal peptide EglA and the *cpg_*Cs gene were attached to the Npr signal peptide. The native LPXTG motif was also attached to the *cpg_Cs* gene and the Npr signal peptide. These constructs were cloned into the pMTL82251:yz2 plasmid, and conjugated into *C. sporogenes*.

The enzymatic assay results showed that the majority of the protein for the Npr signal peptide based constructs was in the supernatant. The EglA signal peptide based constructs did not exhibit as much expression, however, they did show a marked increase in the level of expression in the cell wall fraction. Consequently enzymatic activity of the CPG_2 was assayed in the cell wall for the different sortase constructs and the activity was compared (in Fig 5-6)

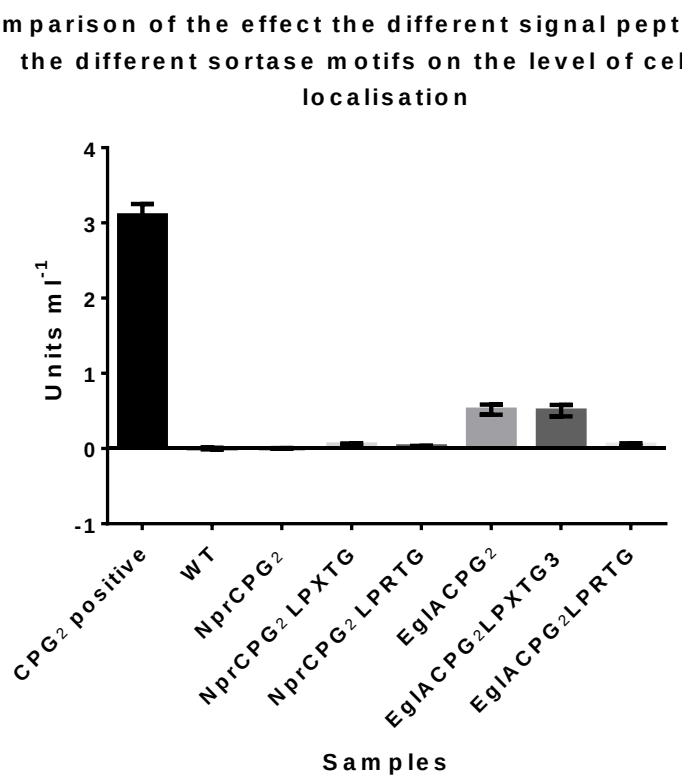


Figure 5-6 A comparison of the different cell wall fractions.

The different cell wall fractions were all assayed for CPG_2 activity to compare what combination of signal peptide and sortase motif exhibited the most localisation of CPG_2 at the cell wall fraction. Samples were performed in technical triplicate and standard error bars are shown.

While the Npr signal peptide was shown to be the best method for exporting CPG_2, when a LPXTG motif is incorporated the localisation is not to the cell wall. With the EglA signal peptide it does not show any evidence of being exported. Thus, when the cell fractionation is performed the activity is seen at the cell wall fraction. This effect is seen regardless of the attachment of an LPXTG motif. Analysis of the localisation of the CPG_2 protein at the cell wall through a student's t-test showed that there was a significant increase in its presence at the cell wall when the $NprCPG_2$ protein was compared to the $NprCPG_2LPRTG$ protein. The level of localisation was not significantly different between the $EglACPG_2$ and $EglACPG_2LPXTG3$ after the addition

of additional data. The localisation of the CPG_2 was, however, significantly higher for the $EglACPG_2$ and $EglACPG_2LPXTG$ to all other variations.

There was a significant increase in the amount of CPG_2 at the cell wall for all different strains which had a *cpg* gene present bar the *NprCPG₂* gene when compared to the wild type bacteria which did not possess the *cpg* gene.

5.2.3 The Sortase System in *C. sporogenes*

The functionality of the sortase system was investigated in *C. sporogenes.* Three methods were looked into, the first was an attempt to assay the transcription of the two sortases, the second was to purify the native S-Layer linked protein that had been identified for the native LPXTG motif, raise antibodies for it, and see where this protein is localised. This protein was annotated as an S-layer linked protein from the genome annotation that was performed, but had no experimental data to confirm it as a S-layer linked protein.

The purification of the S-layer protein was attempted through the cloning of the S-layer protein into the pET28a plasmid, to add a his-tag. The S-layer protein sequence to be cloned was modified to exclude the LPXTG motif, and the signal sequence as once exported and anchored the signal peptide and LPXTG motif would have been cleaved off the protein. Therefore, the antibody would need to be targeted to the rest of the protein to bind after anchoring. The cloning of the modified sequence was performed, and the plasmid sequenced, which confirmed the correct sequence had been introduced into the multiple cloning site. The plasmid was introduced into the expression host *E. coli* BL21 (DE3) Star. Once the sequence was confirmed the production of the protein was induced. A 24 h 10 ml culture was grown, and IPTG was added after 3 h to induce production. At the 24 h point the culture was centrifuged down, and the cell pellet was lysed as described (in methods). The

pET28a plasmid supplies the promoter, start codon and a terminator after introducing a His-Tag to the protein. The His-Tag was used to purify the protein using the His-Bind kit from Merck Millipore. This uses immobilized metal affinity chromatography, due to the His-Tag binding to the divalent Ni^{2+} cation. The beads of the resin, are charged, and then the His-Tag binds to those beads where all other proteins will flow through the column. Then imidazole is used to strip the columns of the His-tagged protein. To begin with the standard protocol as described by the manufacturer was used. This however resulted in numerous proteins being eluted in the His-Tag fraction (shown in Fig 5-7).

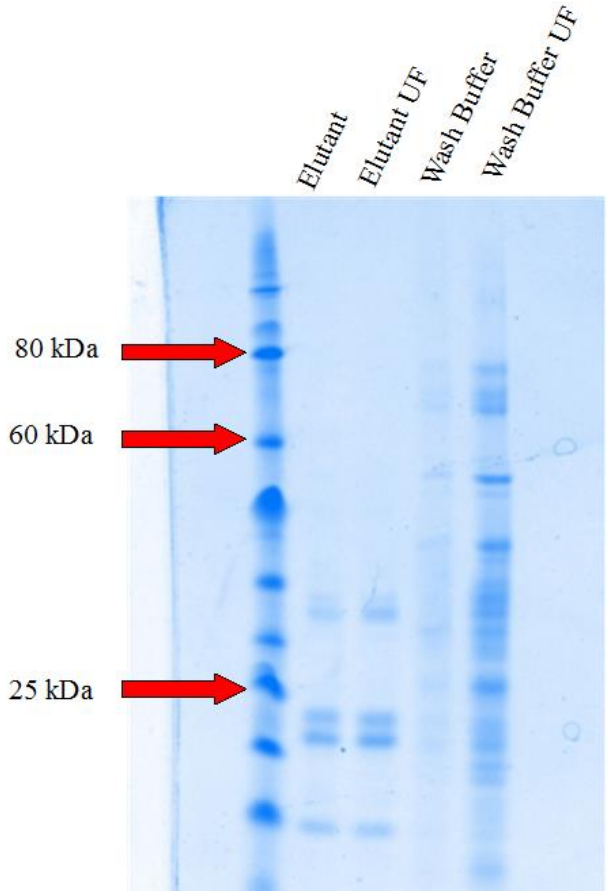


Figure 5-7 His-Tag proteins from IMAC chromatography purification.

The different flow through fractions of the IMAC chromatography, neat and ultrafiltrated. Ultrafiltration was performed in case protein concentrations were too low. The elutant buffer should have only the his-tagged S-layer protein, wash buffer was also looked at in case the his-tagged protein was eluted in the wash buffer rather than the elutant buffer.

Therefore a range of elutant buffers were made up with gradually increasing concentrations of imidazole. This would separate out the different proteins until there was one elutant with one protein of the right size (58kDa) which could be confirmed to be the S-layer protein.

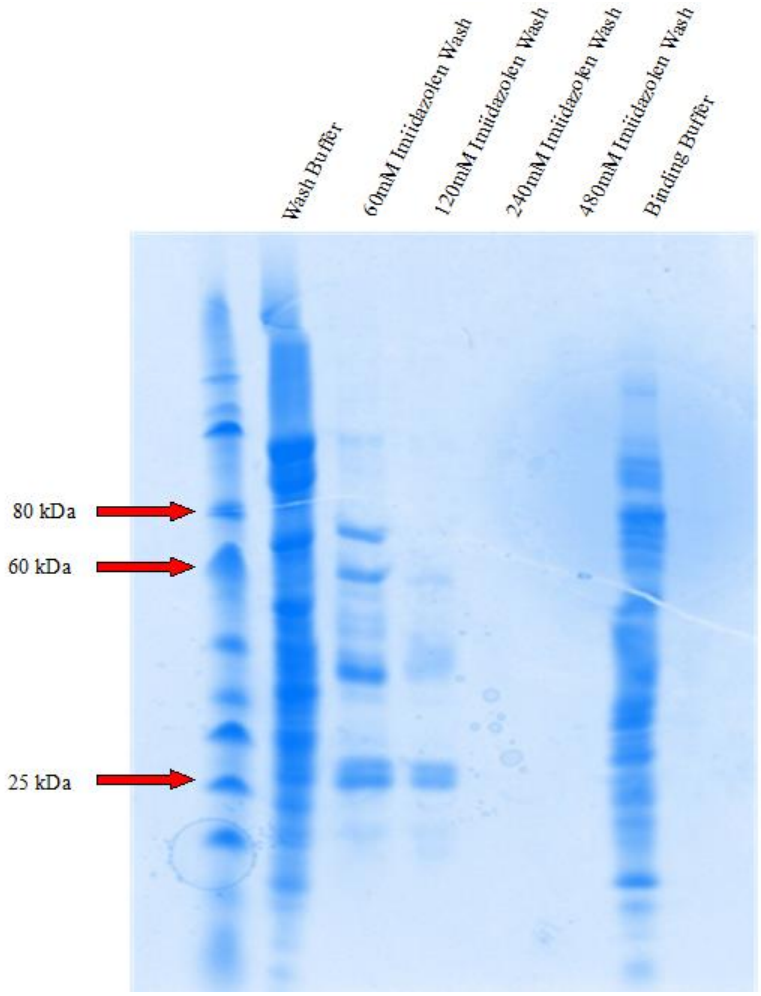


Figure 5-8 Increased concentration of Imidazole on the elution of proteins.

The gradually increasing concentrations of imidazole were run through the IMAC column for better separation of the His-tagged S-layer protein from all other proteins. The proteins were eluted from the column and the different concentrations were looked at to see if a single fraction had the his-tagged protein within.

As can be seen in Fig 5-8, despite using gradually increasing amounts of imidazole there was no separation of the numerous proteins observed in the elutant. At the 240mM concentration of imidazole all remaining proteins that were bound to the

charged resin in the column were eluted. This did not give the single protein in elutant required to generate an antibody to be used with *C. sporogenes*. Other methods were attempted to identify whether the sortase system was active in *C. sporogenes*.

For assaying the transcription of the sortases qRT-PCR was performed. Primers were designed through the online program Primer 3 Plus. Primers were designed for the two sortase genes, and a constitutively expressed housekeeping gene, *rpoA.* The RNA was extracted from *C. sporogenes* as described by 2.7.1. The RNA was tested for DNA contamination through PCR of the *spo0A* gene. The pure RNA was consequently nanodropped to confirm the concentration. The concentration was high, leading to attempts to perform One Step RT-PCR as described in materials and methods. The results of the One Step RT-PCR are shown in Fig 5-9.

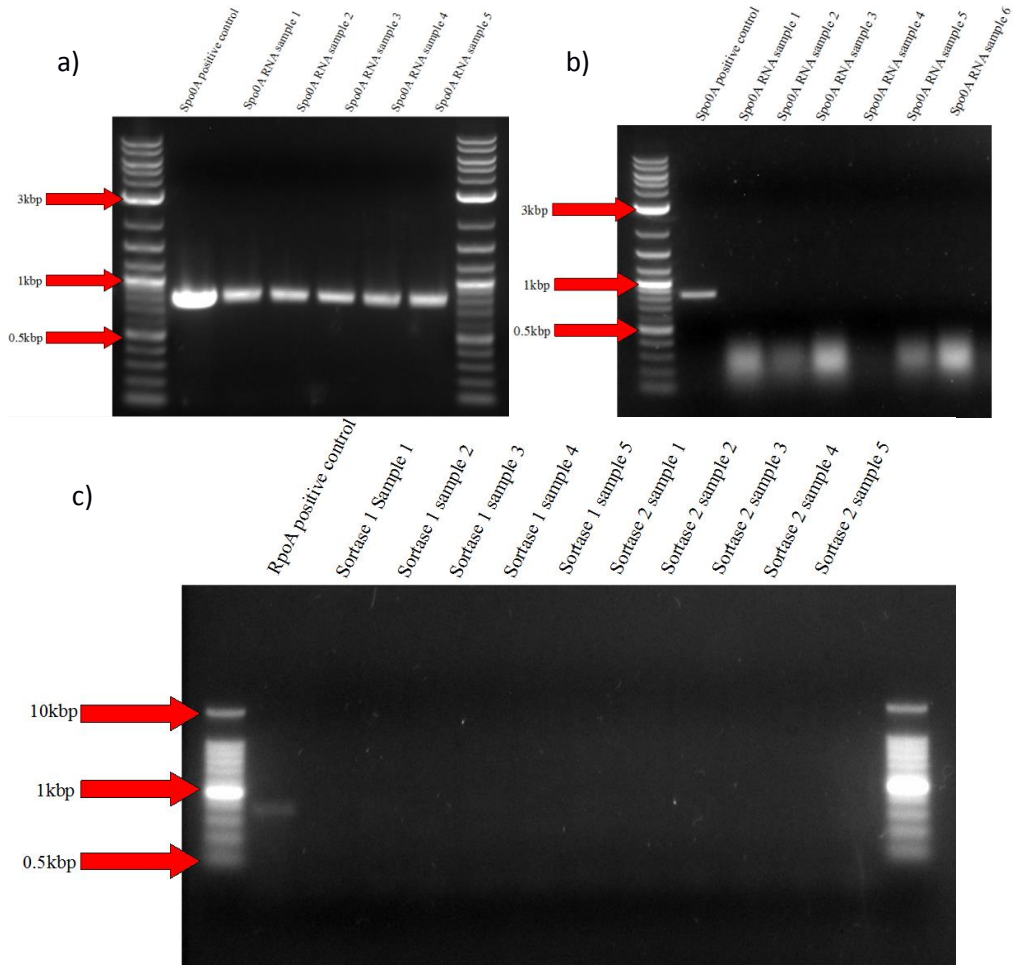


Figure 5-9 The RT-PCR gel pictures.

A) The original RNA extractions which were performed were tested for DNA contamination through PCR of the Spo0A gene. This was performed against a genomic DNA extraction as a positive control. The bands show that there was contamination. B) Shows the repeated PCR on fresh RNA extractions, giving confirmation that the RNA was pure. C) was the RT PCR on the RNA extractions for the two different sortase enzymes, as a positive control the rpoA gene which should be constitutively expressed was also included in the RT-PCR.

Due to the lack of time and results in the One Step RT-PCR, the level of expression of the sortase enzymes was not quantified during this study.

5.3 Conclusions

Progress was made at optimising the system through improving the signal peptide. However the failure to properly utilise the *C. sporogenes* sortase system was disappointing as the major obstacle in developing a useful therapeutic is the activity of the bacillolysin like metalloproteases against any other exported extracellular protein. It was hoped through the anchoring of the protein to the cell wall there would be no free floating CPG_2 and therefore there would be less interaction between the CPG_2 and the metalloproteases. The improvement of the signal sequences was shown through the significant increase in the level of CPG_2 activity in the supernatant. The signal peptide Npr, from the bacillolysin like gene cluster, was well exported in its native organism *C. sporogenes.*

The expression of the *cpg* genes with the *eglA* signal sequence resulted in high levels of CPG_2 in cell lysates, however, this could be seen as bad for the sortase system. The sortase system has been shown to be dependent on the proteins which are to be cell wall linked being exported out of the cell through the Sec system (Schneewind and Missiakas, 2014). The Sec system appears to be in *C. sporogenes* through examination of the genome, however it has not been experimentally confirmed. The activity of the CPG_2 protein when attached to the EglA signal peptide is seen in the intracellular fraction. This shows that if the protein is being exported then it is not exported through the sec system and therefore will not be optimal for use in the sortase system. If it was being exported efficiently through the sec system then the protein would not be functionally active in the cell, and would only fold outside of the cell. However, despite folding in the cell, the level of CPG_2 activity that was seen was the highest when the EglA signal peptide was included. The Npr signal peptide worked well in the expected manner, as it was not functionally active within the cell,

but was exported and folded outside of the cell more efficiently than any other signal sequence, however, the level of expression was not as high as that of the EglA constructs.

This lead to both signal peptides being used for the cell wall linking of the CPG_2. The different LPXTG identified motifs were attached to these two signal peptides to see whether the sec system was important for sortase function in *C. sporogenes* and whether the higher expression will aid the increase in the localisation of the protein to the cell wall. The levels of expression in the cell wall, however, gave inconsistent results. The level of localisation for the proteins $EglACPG_2$ and $EglACPG_2$LPXTG was significantly increased compared to the others. This is probably not down to the signal peptides, and might not even be down to the sortase motif. So little is known about how this system works in clostridia, that fundamental research is necessary to fully utilise it.

To further understand how best to utilise the sortase system fundamental research was performed into *C. sporogenes* sortase genes. To do this, the S-layer linked protein which had a sortase LPXTG motif, implying it was a sortase substrate was isolated from the genome and attempts were made to purify the protein. The protein was then to be used to raise antibodies so it could be seen where the protein was localising in the cell. Attempts were also made to perform RT-PCR of the sortase enzymes to confirm their expression. The inability to produce results for the fundamental research into the mechanisms behind the experimental procedures leaves us with a lack of understanding in how to fully utilise it. To find out more about the system the sortase enzyme genes could be knocked out, through the targeted intron mutagen, ClosTron. However as has been shown through genome examination there are two different sortases. This causes a problem, as ClosTron

knockout of two genes in a single organism is problematic. A flip recombinase is required to remove the selective marker, and allows selection of the intron at the second insertional site. The flip recombinase is, however, untested in *C. sporogenes* and might not be functional. So for the knockout of two genes in *C. sporogenes* a clean deletion of at least one of the sortase genes using allelic exchange needs to be performed. Unfortunately this is also problematic due to the very high AT concentration of the non-coding regions of *C. sporogenes*. The high AT content of the non-coding regions lead to great difficulty in achieving successful PCR or synthesis of these products, and in this case there were also secondary DNA structures which would inhibit successful cloning of the homology arms. The relevant regions have been sent off for synthesis and delays were seen due to these issues

For the isolation of the S-layer protein it is unknown why there were several different proteins in elutant fraction of the isolated proteins. To attempt to isolate the protein a greater range of imidazole concentrations could be used to find the correct concentration point at which the S-layer protein is eluted without all other bound proteins in the column. There are numerous different proteins in *E. coli* which have a similar his-tag motif, however, these should not bind due to the positioning of the motif in the protein and their non-exact motifs. There are other issues which can result in nonspecific binding of proteins to the column resin. For example if binding between the S-layer protein and other native proteins occurs, it can result in those proteins being striped alongside the his-tagged protein. Non-specific hydrophobic interactions can occur between non his-tagged proteins and the resin within the column. There are alternative purification methods which can be used to improve upon this system.

If possible qRT-PCR should be performed to give a good idea as to the level of expression of the sortases enzymes. It has been shown in experiments in *C. acetobutylicum* that the limiting factor for the level of cell wall localisation was the level of expression of the sortase enzymes rather than the motif of signal peptide. Through incorporating of the gene to be cell wall linked onto the chromosome over expression of the sortase enzyme could then be performed through an expression plasmid. This would lead to an increase in the amount of cell wall linked protein.

The reasons for failure in the fundamental research is not clearly understood as there may be many factors working together, but can be circumvented through the alteration of the techniques that were used to investigate the *C. sporogenes* sortase system.

There are numerous further experiments which could be done to improve our understanding. For example ClosTron could be attempted and there are other ways of knocking out the activity of the sortase, such as allelic chromosomal exchange to form clean deletions. By using flanking regions of DNA either side of the sortase open reading frames with selective markers in a plasmid, homologous recombination events can be screened for through looking for marker selection. Once the sortases are knocked out we can look for a loss of phenotype with the S-layer protein, through stripping the cell wall linked proteins of the cell using the glycine method (described by (Calabi et al., 2001)) we could see whether a protein is lost from the stripped protein profile on an SDS-PAGE. It could also be looked to see whether any of the LPXTG motifs that were implemented on the CPG_2 protein were significantly less localised at the cell wall without functional sortases. Or inversely to this, the *cpg_Cs* gene with the correct Npr signal sequence and the varying LPXTG motifs could be integrated onto the chromosome and then sortases could be overexpressed

on a plasmid to see if this upregulates the level of CPG_2 protein in the cell wall fraction.

Further to this we could identify whether one or both of the sortases are functional and what sort of sortase system is active in *C. sporogenes*. There are several different types of sortase system which while operating under the same basic principle, have subtle differences which could have an effect on the level of binding to the cell wall that is achieved. It is suspected that like other clostridia*, C. sporogenes* has a class C sortase, which is the least studied of the four classes. This is suspected to utilise the same motif as the well-studied class A sortase of *S. aureus*. This is down to the sortase enzyme that cleaves the LPXTG motif at T-G bond and while the sortase C enzyme appears to be different to the sortase A enzyme it appears to cleave the same motif, just in a different way.

As very little is known about the structure and function of the cell wall of *C. sporogenes* the lack of binding of the CPG_2 could be down to the structure of the cell wall. If the structure deviates from what is suspected this could limit the amide bonds of the carboxyl group available after cleavage of the LPXTG motif to the amino group of the pentaglycine crossbridge. If for whatever structural reason this amino group is not available the amide bond will not be formed and therefore the protein will not be cell wall linked.

Final Conclusions

Chapter 6

6 Final Conclusions

6.1 Aims

The aims of this project were to endow *C. sporogenes* with the ability to produce both a direct action antibody therapeutic, and a prodrug converting enzyme. These were the prodrug converting enzyme Carboxypeptidase G_2 (CPG_2) and the monoclonal antibody anti-VEGF. The encoded genes would need to be tested for expression and functionality when cloned into the host organism, and thereafter their expression optimised to improve the efficiency of the treatment *in vivo*. It was also investigated whether *C. sporogenes* was suitable as a vector for cancer therapeutics by completing the genome and ensuring it is safe for use as a delivery method for treatments.

6.2 Genome

The completion of the *C. sporogenes* genome was performed with a final genome coverage of 100%. The route taken to achieve this included some novel and new techniques which enabled a complete genome, with a great deal of confidence in the positioning of the contigs. It also confirmed that no associated toxin genes were present in the *C. sporogenes* NCIMB 10696 strain (Kubiak et al., 2015). There had been suggestions that there was the presence of some streptolysin like toxins in *C. sporogenes* (Gonzalez et al., 2010), however, no such genes were found in the genome. The effects of the streptolysin genes were also absent. Due to the genetic homology between *C. sporogenes* and *C. botulinum* the *C. botulinum* neurotoxins were also looked for in the bacteria, all of this confirming *C. sporogenes* NCIMB 10696 to be non-pathogenic and should, therefore, be safe for the treatment of cancer. However, there is one issue which might cause some concern for the

implementation of *C. sporogenes* and that is it being linked as a causative agent for gas gangrene from wounds (Gorbach and Thadepalli, 1975). This issue has been linked with numerous *Clostridium* species, such as *C. novyi* which is also being explored as the basis of a possible anti-cancer therapeutic. The single patient treated with *C. novyi* exhibited a gas build up in the soft tissue where the hypoxic tissue showed colonisation from the *C. novyi* (Roberts et al., 2014). This side effect was consistent with gas gangrene, where the gas in the tissue comes about due to the bacteria releasing gas. The patient's side effects were treated with a course of antibiotics.

As *C. sporogenes* also releases gas when growing, and has been linked with gas gangrene, it could be a side effect which occurs in *C. sporogenes* also. Gas gangrene can be a serious condition, which causes necrotic damage to soft tissues, and is difficult to treat. Treatment would not be limited to antibiotics as the level of penetration is limited in ischemic tissue conditions displayed by gas gangrene. Ischemia results in tissue hypoxia which would already be present due to the nature of the cancerous tumour, and as the CDEPT treatment is designed to kill hypoxic tissues a course of antibiotics would limit any bacterial activity to the tumour site until such time as the tumour is destroyed.

6.3 Implementing Anti-VEGF

Bringing about the production of human monoclonal antibody anti-VEGF in *C. sporogenes* proved to be very problematic. Indeed, the successful production of this protein was never conclusively confirmed in *C. sporogenes.* It was suspected that this could be due to toxicity to the host bacterium. To overcome this, an inducible system was used, however, there were still issues with expression due to suspected toxicity (Biebinger et al., 1997). The reason for this toxicity is unknown but could be caused

by any number of things. Issues such as high copy number plasmids could lead to the over expression of the plasmid genes causing over recruitment of ribosomes. Another possible issue could be that the anti-VEGF monoclonal antibody causes an unknown interaction with some proteins that will prevent key processes in the cell working. To account for the issue of over recruitment of ribosomes a simple substitution of the Gram-positive replicon within the plasmid would allow for a lower copy number, and therefore reduce the number of ribosomes required. However, if the toxicity were to persist then protein-protein interaction studies would have to be performed to ascertain how the anti-VEGF was causing toxicity. A ProteoChip microarray experiment could be performed which would show interactions between anti-VEGF and the native proteins (Lee et al., 2003). The protein interaction, if identified, might not however give a solution to the issue of toxicity. The protein that anti-VEGF might interact with could be a transcription factor and therefore there are many possible downstream interactions which could cause the toxicity. So even by identifying the protein-protein interaction this does not necessarily result in an obvious solution to the toxicity of the anti-VEGF. To implement anti-VEGF protein production in *C. sporogenes* appears to be a task that is far beyond the potential benefits of achieving it. Current medical trials of the anti-VEGF protein have so far confirmed that it works, but only in collaboration with other chemotherapeutics (Jain et al., 2006, Kabbinavar et al., 2003), so with individual efficacy of anti-VEGF being low it appears alternative direct action molecules would be better suited to this task. The complexity of the cancer cells molecular systems, while making cancer heterogeneous, also gives plenty of potential targets for direct action therapies. The drugs that could be implemented could focus on many different pathways, one which could be implemented would be a biological molecule that would induce the

apoptotic cascade, this could be an ideal way to kill of tumour cells, such as initiator caspases (e.g. caspase-8).

6.4 Implementing Carboxypeptidase G_2

Once incorporated in the pMTL82251:YZ2 *lacI* inducible system expression of Carboxypeptidase G_2 (CPG_2) was quickly confirmed. The localisation of the produced enzyme was not ideal for Clostridial Directed Enzyme Prodrug Therapy (CDEPT) with most of the enzyme activity occurring inside the cell. For CDEPT to work efficiently, the enzyme would have to be active outside of the *C. sporogenes* cell, this would be best as it is not known yet which prodrug might be used, and therefore whether the bacterium can take it up. Improvement in exporting the protein would be required for the activity to occur in the tumour *in vivo*. When looking into the level of enzymatic activity outside of the cell it was also shown that there was a great degree of degradation of the CPG_2 enzyme, which was most likely due to the proteolytic activity of the host. This proteolytic activity and effect was confirmed through the MALDI mass spectrometry of the *C. sporogenes* extracellular fraction identifying proteases, and through the inhibition of CPG_2 degradation when *C. sporogenes* were grown in media containing protease inhibitors. Therefore, the issues for implementing an effective CDEPT system with CPG_2 were inefficient exportation and proteolytic action against the CPG_2 prodrug converting enzyme.

6.5 Optimization of Carboxypeptidase G_2

There has been progress made in the attempts to optimise the export of CPG_2 to the external medium. When comparing the different levels of activity in the extracellular fraction there were several different signal peptides that resulted in the same level of the activity in the extracellular fraction. The `level of activity overall was lower than that of the original signal sequence tested EglA, which was originally not

considered to be a viable option. The EglA signal peptide was considered to be faulty due to a considerable amount of activity of CPG_2 within the cell. The protein when produced, if translated with an efficient signal peptide, will be inhibited from folding for efficient export. The activity detected within the cell shows the protein to be folding and, therefore, the signal peptide is not working efficiently. It could be that the EglA signal peptide is not utilising the Sec exportation pathway and therefore can fold before being exported (Quixley and Reid, 2000, Zappe et al., 1988).

With regards to the other signal peptides, the best was Npr, obtained from a gene encoding one of several metalloproteases. With the alternative signal peptides no enzymatic activity was seen in the intracellular fraction, however, the level of activity in the extracellular fraction was no improvement on the EglA peptide. If expression of the CPG_2 is consistent across the signal peptide library, then there is a considerable amount of CPG_2 protein in the *C. sporogenes* that is unfolded, but not being exported. For the signal peptide studies it is hard to say to what extent the best signal peptide is, it could be that one signal peptide works better with a certain protein and therefore other assayable proteins should be attached to the signal peptides. This would show which signal peptide might be the most consistent across numerous proteins. However, as signal peptide's efficiency can be linked to the protein that is being exported then some peptides might function better for CPG_2 but the same peptides might not work for alternative proteins. As expression is also variable, even when using the same ribosomal binding site, and promoter between different proteins attributing the enzymatic activity levels to either the exportation of the protein or to the level of expression is very difficult (Terpe, 2006).

When compared to other studies that are on-going in CDEPT treatments there is no need for the prodrug converting enzymes to be exported, and these studies have a

distinct advantage. The most common prodrug which is used is the nitroreductase protein. This protein is folded in the cell and is active there, without any exportation necessary for breakdown of the prodrug, CB1954 (Minton, 2003). This is in contrast to most alternative PCE's and their prodrugs, where the prodrug would not be taken up by the bacteria to be cleaved inside the cell

Through the sequencing of the bacteria, the improved production of the CPG_2 protein when in the presence of the Roche cOmplete™ protease inhibitor cocktail and the MALDI mass spectrometry of the extracellular fraction we can say with confidence that the metalloproteases that are produced, are exported and present in the media when *C. sporogenes* is grown. It is suspected that these metalloproteases are the oncolytic effector when *C. sporogenes* colonises a hypoxic tumour. It is also hypothesised that these metalloproteases are part of the reason that the bacteria can colonise the tumour so well. Proteases can be utilised to break down their surrounding environment to release nutrients which can then be taken up by the bacteria and used for growth and metabolism (Weiser et al., 2003, Kavermann et al., 2003). However due to the need to export the protein out of the system into the surrounding environment to breakdown the prodrug, this creates a dilemma. For the system to work the bacteria must be able to export the metalloproteases, however, for the system to work the CPG_2 must also be exported without being digested by the same proteases. Without either one the system will not cause a significant anti-cancer effect, but they both cannot be active and free floating in the environment. The only proposed solution was to anchor the CPG_2 to the bacterial cell wall through exploitation off the endogenous *C. sporogenes* sortase system. By anchoring the protein to the cell wall this would make the CPG_2 less available to being digested by the metalloproteases. It has also been shown that anchoring of proteins to cell walls

prevents autolysis occurring of the cell, this could be true for *C. sporogenes* (Morgan and Meri, 1994). While this approach might not be perfect, it might improve the situation to a workable point.

6.6 Sortase Research

The identification of the different components of the sortase system was achieved quickly through the completed genome and the software tools such as PSI-BLAST. With the identification of different components completed it showed that there could be a functional sortase present in *C.* sporogenes. It was thought that the best way for the anchoring of proteins to *C. sporogenes* cell wall would be to utilise the native sortase system. Therefore research was performed into the sortase system of *C. sporogenes,* which, despite implementation of the sortase substrate motifs with CPG_2, was inconclusive. Activity of the CPG_2 was present but no noticeable improvement could be seen in the cell wall fraction of the *C. sporogenes* cell. The identification of sortase motifs in *C. acetobutylicum* by Desvaux *et al* (Desvaux et al., 2005) provided a methodology for identifying motifs in *C. sporogenes*, and some potential non-native motifs for utilisation if necessary. Using this method the identity of the only sortase substrate native to *C. sporogenes* was confirmed. This was disappointing as it lowered the chances of utilisation, and when combined with the information that there are two sortase enzymes might hint at the system being vestigial, and no longer active. All three motifs from *C. acetobutylicum* and the native *C. sporogenes* motif were used with both the EglA and the Npr signal peptides to anchor CPG_2. There was a negligible difference between the different sortase motifs that were used, even when the native signal was utilised. Conversely, there was an apparent increase in the amount of CPG_2 that was localised at the cell wall for the EglA signal peptide, which is true whether there was an LPXTG motif attached to the

gene or not. This indicates that the increase being seen is not due to the sortase anchoring system, and is down to the signal peptide. However, when the native *C. sporogenes* LPXTG motif was used, there was a significant decrease in the level of CPG_2 that was localised to the cell wall. There is little or no correlation between the different variables, which makes an underlying cause very difficult to identify.

Due to the lack of utilisation of the sortase system, underlying fundamental research into the *C. sporogenes* system was attempted. To establish sortase functionality qRT-PCR of the sortase enzymes could be done, this should give levels of transcription of the gene at certain time points. However there are several different issues that can affect activity post transcription such as epigenetic factors. Alternatively knockout studies can be performed. Mutations of the two sortase enzymes can be made, then the cell wall proteins stripped from the cell using a method described in Calabi *et al* (Calabi et al., 2001). The proteins stripped from the cell wall of the wild type *C. sporogenes* and the *ΔSortase C. sporogenes* strain can then be compared to see if the *ΔSortase* strain shows loss of proteins on a protein gel.

To properly understand the mechanism of action more information about key parts of the sortase system is needed. Before the protein can be cleaved at the LPXTG motif, and anchored to the cell wall, it must be exported out of the cell. This is accomplished through the signal sequences that were identified in the optimisation of expression chapter. However, while we have been able to identify signal peptides that properly export it is not known which of the secretory systems are utilised. This can make a difference in the efficacy of the sortase system, as the method for exportation of a sortase cell linked system should be through the Sec system and not the alternative secretory systems such as the TAT system.

Once the protein has been cleaved and exported out of the cell it should form an amide bond to the amino group of the cell wall peptidoglycans. However, the cell wall of the *C. sporogenes* is not well characterised. There have been some investigations into clostridial cell walls, such as that research undertaken in *Clostridium difficile* (Fagan et al., 2009, Fagan and Fairweather, 2011, Fagan et al., 2011a). The cell wall has been recognised to consist of the cytoplasmic membrane, with the peptidoglycan layer outside of that, and then the two S-layer cell wall proteins. The high molecular weight S-layer protein surrounds the peptidoglycan layer, and then the low molecular weight protein is attached to that (Fagan et al., 2009). In *C. difficile* the S-layer proteins are formed from the precursor protein produced by the s*lpA* gene (Calabi et al., 2001). These S-layer proteins are anchored to the cell walls of *C. difficile* through an alternative binding motif called the pfam 04122. Through studies of this motif (Fagan et al., 2009) it has been shown to contain a putative cell wall binding repeat. In *C. difficile* all cell wall proteins (cwp's) contain three repeats of the pfam 04122 motif (Fagan et al., 2011b).

As sortases are used in colonisation techniques, which are particularly well documented for *S. aureus*, it is hypothesised that the alternative colonisation methods of *C. sporogenes* has reduced the necessity for a fully functional sortase system. Sortase has also been heavily linked with virulence in the species in which the sortase has been studied. Again this is no longer a factor for *C. sporogenes* as, while it has a common ancestor with *C. botulinum,* the toxicity is either lost now, or was gained in *C. botulinum* after speciation. This indicates that the *C. sporogenes* sortase system may at best supplement its colonisation ability but with the limited number of sortase substrates found and the truncated version of the sortase enzyme it indicates that the sortase system is actually vestigial. Potentially the best way to

utilise this system is to implement a completely foreign, but functioning system through not only the introduction of the *cpg* gene with an appropriate sortase motif but also a functional sortase enzyme, which would function for the cell wall that *C. sporogenes* has. To confirm a non-functional sortase system in *C. sporogenes* would be interesting as to date, any Gram-positive which has been studied for the sortase system has had a functional one.

6.7 Carboxypeptidase G_2 in the Context of CDEPT

The use of Clostridia Directed Enzymatic Prodrug Therapy (CDEPT) is getting closer to being brought to clinical trial, but only with the nitroreductase prodrug converting enzyme. This treatment is unique for CDEPT in that the nitroreductase is not required to be exported outside of the cell to break down the prodrug that will provide the toxic effect. This provides a distinct advantage to the other prodrug and prodrug converting enzyme combinations which have been tested, as the nitroreductase will be active but also protected from degradation by the metalloproteases. As has been shown by this study the issue of extracellular degradation of PCE's is potentially one of the biggest stumbling blocks to getting other prodrug and PCE combinations implemented in *C. sporogenes* and ready for clinical trials.

While there are no CDEPT approaches in clinical trial, there is a *clostridial* directed therapy which is going into its first human trial. This therapy has been described by Roberts *et al* (Roberts et al., 2014) who have been injecting attenuated *C. novyi* into not only mice but also dogs and a single human who have been suffering from cancer. The *C. novyi* NT strain has not been genetically modified to produce a prodrug converting enzyme as the bacterium itself is oncolytic, and so the attenuated strain is the treatment in itself. In the human patient this resulted in

tumour regression, but a big build-up of gas in that region. The cause of the anti-cancer effect of *C. novyi* NT growth in a tumour is unknown (Dang et al., 2001), but could be the same as the anti-cancer effect exhibited by wild type *C. sporogenes* (proteolysis). The *C. novyi* NT strain has genetic similarities to *C. sporogenes* and seen as *C. novyi* NT has been brought to clinical trial the results from that trial allows us to infer what might happen for CDEPT when used in human patients. Both strains have extracellular proteases and both strains probably owe their high levels of colonisation to those proteases. Both strains have some oncolytic effect when no PCE is implemented which has been shown in tumour bearing mice for *C. sporogenes* (Heap et al., 2014) and in humans for *C. novyi*, again this is probably down to the extracellular proteases. The side effect of the proteases oncolytic effect could be gas gangrene.

Both *C. sporogenes* and *C. novyi* have been clinically associated with gas gangrene (Gorbach and Thadepalli, 1975). Gas gangrene does not have a known mechanism, but there are some contributing factors, such as the θ toxin and α toxin of *Clostridium perfringens* (Bryant et al., 2000). These toxins both play a key role in the rapid necrosis of the flesh that occurs in gas gangrene. However, *C. sporogenes* does not have these genes, so for it to be a cause of gas gangrene it must be associated with the tissue necrosis in an alternative manner. This could be down to the proteolytic nature of *C. sporogenes.* Through investigation of the *C. novyi* NT genome, it can be seen that there are also thermolysin like exported metalloproteases, and other zinc dependent peptidases much like *C. botulinum*. It has been demonstrated that extracellular enzymes aid colonisation for numerous different bacteria (Weiser et al., 2003, Valiente et al., 2008, Kavermann et al., 2003). These bacterial proteases tend to break down proteins that inhibit bacterial

colonisation, and release key amino acids which can then be taken up for use by bacteria. This could be the case for both *C. sporogenes* and *C. novyi* NT not only having some necrotic effect on solid tumours, but their increased ability to colonise these regions and even the gas gangrene. These proteases could have a common purpose to destroy the surrounding area, which would release nutrients for the uptake by *C. sporogenes* causing the tissue necrosis. When *C. sporogenes* grows it produces gas, which would result in gas inflammation. The proteases in *C. novyi* could be the cause of the gas gangrene like symptoms exhibited in the singular patient that was treated with *C. novyi* NT, as it was attenuated and the only toxin normally present was removed (Dang et al., 2001).

The issue of gas gangrene could be circumvented, but it might be at the cost of the system. To use an alternative non proteolytic bacterium for the delivery of PCE's would be one such approach, or to use strains with the metalloprotease system knocked out would be another, but this might be at the detriment of the levels of colonisation or native oncolytic effect. A balancing act of these characteristics must be performed for optimum efficiency. However, whether this optimum would be an improvement on current treatments or the treatments now in late stage clinical trials remains to be seen.

The metalloproteases could be investigated further for a greater understanding as to how the potential toxic side effects of the CDEPT occur, and also how the unique level of colonisation by the proteolytic bacteria occurs. By removing the proteases it would show whether these extracellular proteins are essential for colonisation. If these proteins are not essential for colonisation this will be important for the implication of the export of the prodrug converting enzyme, as the metalloproteases digest the extracellular CPG_2. This could be the issue which decides whether having

alternative prodrug converting enzymes and therapies to nitroreductase is a realistic option.

Bibliography

7 Bibliography

ADAMS, G. P. & WEINER, L. M. 2005. Monoclonal antibody therapy of cancer. *Nat Biotech,* 23**,** 1147-1157.

AIRD, D., ROSS, M. G., CHEN, W. S., DANIELSSON, M., FENNELL, T., RUSS, C., JAFFE, D. B., NUSBAUM, C. & GNIRKE, A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome biology,* 12**,** R18.

AL-HINAI, M. A., JONES, S. W. & PAPOUTSAKIS, E. T. 2015. The Clostridium Sporulation Programs: Diversity and Preservation of Endospore Differentiation. *Microbiology and Molecular Biology Reviews,* 79**,** 19-37.

ANSORGE, W. J. 2009. Next-generation DNA sequencing techniques. *New Biotechnology,* 25**,** 195-203.

AURELI, P., FENICIA, L., PASOLINI, B., GIANFRANCESCHI, M., MCCROSKEY, L. M. & HATHEWAY, C. L. 1986. Two Cases of Type E Infant Botulism Caused by Neurotoxigenic *Clostridium butyricum* in Italy. *Journal of Infectious Diseases,* 154**,** 207-211.

BARBÉ, S., VAN MELLAERT, L. & ANNÉ, J. 2006. The use of clostridial spores for cancer treatment. *Journal of Applied Microbiology,* 101**,** 571-578.

BEHR, J. P. 1994. Gene transfer with synthetic cationic amphiphiles: prospects for gene therapy. *Bioconjugate chemistry,* 5**,** 382-9.

BERGERS, G. & BENJAMIN, L. E. 2003. Tumorigenesis and the angiogenic switch. *Nature reviews. Cancer,* 3**,** 401-410.

BIEBINGER, S., ELIZABETH WIRTZ, L., LORENZ, P. & CHRISTINE, C. 1997. Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. *Molecular and Biochemical Parasitology,* 85**,** 99-112.

BOTTOS, A., MARTINI, M., DI NICOLANTONIO, F., COMUNANZA, V., MAIONE, F., MINASSI, A., APPENDINO, G., BUSSOLINO, F. & BARDELLI, A. 2012. Targeting oncogenic serine/threonine-protein kinase BRAF in cancer cells inhibits angiogenesis and abrogates hypoxia. *Proceedings of the National Academy of Sciences,* 109**,** E353–E359.

BRADBURY, M., GREENFIELD, P., MIDGLEY, D., LI, D., TRAN-DINH, N., VRIESEKOOP, F. & BROWN, J. L. 2012. Draft Genome Sequence of *Clostridium sporogenes* PA 3679, the Common Nontoxigenic Surrogate for Proteolytic Clostridium botulinum. *Journal of bacteriology,* 194**,** 1631-1632.

BRENTNALL, M., RODRIGUEZ-MENOCAL, L., DE GUEVARA, R. L., CEPERO, E. & BOISE, L. H. 2013. Caspase-9, caspase-3 and caspase-7 have distinct roles during intrinsic apoptosis. *Bmc Cell Biology,* 14.

BRIZEL, D. M., SIBLEY, G. S., PROSNITZ, L. R., SCHER, R. L. & DEWHIRST, M. W. 1997. Tumor hypoxia adversely affects the prognosis of carcinoma of the head and neck. *International Journal of Radiation Oncology*Biology*Physics,* 38**,** 285-289.

BROSE, M. S., REBBECK, T. R., CALZONE, K. A., STOPFER, J. E., NATHANSON, K. L. & WEBER, B. L. 2002. Cancer Risk Estimates for *BRCA1* Mutation Carriers Identified in a Risk Evaluation Program. *Journal of the National Cancer Institute,* 94**,** 1365-1372.

BROUSSOLLE, V., ALBERTO, F., SHEARMAN, C. A., MASON, D. R., BOTELLA, L., NGUYEN-THE, C., PECK, M. W. & CARLIN, F. 2002. Molecular and Physiological Characterisation of Spore Germination in *Clostridium botulinum* and *C. sporogenes*. *Anaerobe,* 8**,** 89-100.

BROWN, J. M. & GIACCIA, A. J. 1998. The Unique Physiology of Solid Tumors: Opportunities (and Problems) for Cancer Therapy. *Cancer research,* 58**,** 1408-1416.

BROWN, J. M. & WILSON, W. R. 2004. Exploiting tumour hypoxia in cancer treatment. *Nature reviews. Cancer,* 4**,** 437-447.

BROWN, M., FIGGE, J., HANSEN, U., WRIGHT, C., JEANG, K.-T., KHOURY, G., LIVINGSTON, D. M. & ROBERTS, T. M. 1987. Lac repressor can regulate expression from a hybrid SV40 early promoter containing a lac operator in animal cells. *Cell,* 49**,** 603-612.

BRYANT, A. E., CHEN, R. Y. Z., NAGATA, Y., WANG, Y., LEE, C. H., FINEGOLD, S., GUTH, P. H. & STEVENS, D. L. 2000. Clostridial Gas Gangrene. I. Cellular and Molecular Mechanisms of Microvascular Dysfunction Induced by Exotoxins of *Clostridium perfringens*. *Journal of Infectious Diseases,* 182**,** 799-807.

BURGESS-BROWN, N. A., SHARMA, S., SOBOTT, F., LOENARZ, C., OPPERMANN, U. & GILEADI, O. 2008. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expression and Purification,* 59**,** 94-102.

CALABI, E., WARD, S., WREN, B., PAXTON, T., PANICO, M., MORRIS, H., DELL, A., DOUGAN, G. & FAIRWEATHER, N. 2001. Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Molecular Microbiology,* 40**,** 1187-1199.

CAMPISI, J., KIM, S. H., LIM, C. S. & RUBIO, M. 2001. Cellular senescence, cancer and aging: the telomere connection. *Experimental gerontology,* 36**,** 1619-37.

CHUNG-FAYE, G., PALMER, D., ANDERSON, D., CLARK, J., DOWNES, M., BADDELEY, J., HUSSAIN, S., MURRAY, P. I., SEARLE, P., SEYMOUR, L., HARRIS, P. A., FERRY, D. & KERR, D. J. 2001. Virus-directed, enzyme prodrug therapy with nitroimidazole reductase: a phase I and pharmacokinetic study of its prodrug, CB1954. *Clinical cancer research : an official journal of the American Association for Cancer Research,* 7**,** 2662-8.

DAMERON, K. M., VOLPERT, O. V., TAINSKY, M. A. & BOUCK, N. 1994. Control of angiogenesis in fibroblasts by p53 regulation of thrombospondin-1. *Science,* 265**,** 1582-4.

DANG, L. H., BETTEGOWDA, C., HUSO, D. L., KINZLER, K. W. & VOGELSTEIN, B. 2001. Combination bacteriolytic therapy for the treatment of experimental tumors. *Proceedings of the National Academy of Sciences,* 98**,** 15155-15160.

DENG, T., ZHANG, L., LIU, X.-J., XU, J.-M., BAI, Y.-X., WANG, Y., HAN, Y., LI, Y.-H. & BA, Y. 2013. Bevacizumab plus irinotecan, 5-fluorouracil, and leucovorin (FOLFIRI) as the second-line therapy for patients with metastatic colorectal cancer, a multicenter study. *Medical Oncology,* 30**,** 1-5.

DESVAUX, M., KHAN, A., SCOTT-TUCKER, A., CHAUDHURI, R. R., PALLEN, M. J. & HENDERSON, I. R. 2005. Genomic analysis of the protein secretion systems in *Clostridium acetobutylicum* ATCC 824. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1745**,** 223-253.

DONAHUE, E., DAWSON, L., VALIENTE, E., FIRTH-CLARK, S., MAJOR, M., LITTLER, E., PERRIOR, T. & WREN, B. 2014. *Clostridium difficile* has a single sortase, SrtB, that can be inhibited by small-molecule inhibitors. *BMC Microbiology,* 14**,** 219.

DONNENBERG, V. S. & DONNENBERG, A. D. 2005. Multiple Drug Resistance in Cancer Revisited: The Cancer Stem Cell Hypothesis. *The Journal of Clinical Pharmacology,* 45**,** 872-877.

DRUGBANK. 2015. *Bevacizumab structure* [Online]. Available: http://www.drugbank.ca/drugs/DB00112.

DURRE, P. 2014. Physiology and Sporulation in Clostridium. *Microbiology spectrum,* 2.

DUTHIE, S. J. & HAWDON, A. 1998. DNA instability (strand breakage, uracil misincorporation, and defective repair) is increased by folic acid depletion in human lymphocytes in vitro. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology,* 12**,** 1491-7.

EL-ANEED, A. 2004. An overview of current delivery systems in cancer gene therapy. *Journal of Controlled Release,* 94**,** 1-14.

FAGAN, R. P., ALBESA-JOVÉ, D., QAZI, O., SVERGUN, D. I., BROWN, K. A. & FAIRWEATHER, N. F. 2009. Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. *Molecular Microbiology,* 71**,** 1308-1322.

FAGAN, R. P. & FAIRWEATHER, N. F. 2011. *Clostridium difficile* Has Two Parallel and Essential Sec Secretion Systems. *Journal of Biological Chemistry,* 286**,** 27483-27493.

FAGAN, R. P., JANOIR, C., COLLIGNON, A., MASTRANTONIO, P., POXTON, I. R. & FAIRWEATHER, N. F. 2011a. A proposed nomenclature for cell wall proteins of *Clostridium difficile*. *Journal of medical microbiology,* 60**,** 1225-8.

FAGAN, R. P., JANOIR, C., COLLIGNON, A., MASTRANTONIO, P., POXTON, I. R. & FAIRWEATHER, N. F. 2011b. A proposed nomenclature for cell wall proteins of *Clostridium difficile*. *Journal of Medical Microbiology,* 60**,** 1225-1228.

FEITELSON, M. A., ARZUMANYAN, A., KULATHINAL, R. J., BLAIN, S. W., HOLCOMBE, R. F., MAHAJNA, J., MARINO, M., MARTINEZ-CHANTAR, M. L., NAWROTH, R., SANCHEZ-GARCIA, I., SHARMA, D., SAXENA, N. K., SINGH, N., VLACHOSTERGIOS, P. J., GUO, S., HONOKI, K., FUJII, H., GEORGAKILAS, A. G., BILSLAND, A., AMEDEI, A., NICCOLAI, E., AMIN, A., ASHRAF, S. S., BOOSANI, C. S., GUHA, G., CIRIOLO, M. R., AQUILANO, K., CHEN, S., MOHAMMED, S. I., AZMI, A. S., BHAKTA, D., HALICKA, D., KEITH, W. N. & NOWSHEEN, S. 2015. Sustained proliferation in cancer: Mechanisms and novel therapeutic targets. *Seminars in Cancer Biology,* 35, Supplement**,** S25-S54.

FERLAY J, S. I., ERVIK M, DIKSHIT R, ESER S, MATHERS C, REBELO M, PARKIN DM, FORMAN D, BRAY, F. 2013. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet].

FRANCIS, R. J., SHARMA, S. K., SPRINGER, C., GREEN, A. J., HOPE-STONE, L. D., SENA, L., MARTIN, J., ADAMSON, K. L., ROBBINS, A., GUMBRELL, L., O'MALLEY, D., TSIOMPANOU, E., SHAHBAKHTI, H., WEBLEY, S., HOCHHAUSER, D., HILSON, A. J., BLAKEY, D. & BEGENT, R. H. 2002. A phase I trial of antibody directed enzyme prodrug therapy (ADEPT) in patients with advanced colorectal carcinoma or other CEA producing tumours. *British journal of cancer,* 87**,** 600-7.

FREUDL, R. 2013. Leaving home ain't easy: protein export systems in Gram-positive bacteria. *Research in Microbiology,* 164**,** 664-674.

FUERST, T. R., FERNANDEZ, M. P. & MOSS, B. 1989. Transfer of the inducible lac repressor/operator system from *Escherichia coli* to a vaccinia virus expression vector. *Proceedings of the National Academy of Sciences,* 86**,** 2549-2553.

GERBER, D. E. 2008. Targeted therapies: a new generation of cancer treatments. *American family physician,* 77**,** 311-9.

GIACALONE, M. J., GENTILE, A. M., LOVITT, B. T., BERKLEY, N. L., GUNDERSON, C. W. & SURBER, M. W. 2006. Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. *BioTechniques,* 40**,** 355-64.

GIBSON, D. G., YOUNG, L., CHUANG, R.-Y., VENTER, J. C., HUTCHISON, C. A. & SMITH, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth,* 6**,** 343-345.

GILBERT , M. R., DIGNAM , J. J., ARMSTRONG , T. S., WEFEL , J. S., BLUMENTHAL , D. T., VOGELBAUM , M. A., COLMAN , H., CHAKRAVARTI , A., PUGH , S., WON , M., JERAJ , R., BROWN , P. D., JAECKLE , K. A., SCHIFF , D., STIEBER , V. W., BRACHMAN , D. G., WERNER-WASIK , M., TREMONT-LUKATS , I. W., SULMAN , E. P., ALDAPE , K. D., CURRAN , W. J. J. & MEHTA , M. P. 2014. A Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma. *New England Journal of Medicine,* 370**,** 699-708.

GONZALEZ, D. J., LEE, S. W., HENSLER, M. E., MARKLEY, A. L., DAHESH, S., MITCHELL, D. A., BANDEIRA, N., NIZET, V., DIXON, J. E. & DORRESTEIN, P. C. 2010. Clostridiolysin S, a Post-translationally Modified Biotoxin from *Clostridium botulinum*. *Journal of Biological Chemistry,* 285**,** 28220-28228.

GOOSENS, V. J., MONTEFERRANTE, C. G. & VAN DIJL, J. M. 2014. The Tat system of Gram-positive bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1843**,** 1698-1706.

GORBACH, S. L. & THADEPALLI, H. 1975. Isolation of Clostridium in Human Infections: Evaluation of 114 Cases. *Journal of Infectious Diseases,* 131**,** S81-S85.

GRECO, O. & DACHS, G. U. 2001. Gene directed enzyme/prodrug therapy of cancer: historical appraisal and future prospectives. *Journal of cellular physiology,* 187**,** 22-36.

GROVE, J. I., SEARLE, P. F., WEEDON, S. J., GREEN, N. K., MCNEISH, I. A. & KERR, D. J. 1999. Virus-detected enzyme prodrug therapy using CB1954. *Anti-Cancer Drug Design,* 14**,** 461-472.

GUSTAFSSON, C., GOVINDARAJAN, S. & MINSHULL, J. 2004. Codon bias and heterologous protein expression. *Trends in Biotechnology,* 22**,** 346-353.

GUSTAFSSON, C., MINSHULL, J., GOVINDARAJAN, S., NESS, J., VILLALOBOS, A. & WELCH, M. 2012. Engineering genes for predictable protein expression. *Protein Expression and Purification,* 83**,** 37-46.

HAMILTON, S. R., AALTONEN, L. A., CANCER, I. A. F. R. O. & ORGANIZATION, W. H. 2000. *Pathology and genetics of tumours of the digestive system*, IARC press Lyon:.

HANAHAN, D. & WEINBERG, R. A. 2000. The Hallmarks of Cancer. *Cell,* 100**,** 57-70.

HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

HARPER, S. J. & BATES, D. O. 2008. VEGF-A splicing: the key to anti-angiogenic therapeutics? *Nat Rev Cancer,* 8**,** 880-887.

HARRIS, S., CRAZE, M., NEWTON, J., FISHER, M., SHIMA, D. T., TOZER, G. M. & KANTHOU, C. 2012. Do Anti-Angiogenic VEGF (VEGFxxxb) Isoforms Exist? A Cautionary Tale. *PLoS ONE,* 7**,** e35231.

HARVEY, S. M., STURGEON, J. & DASSEY, D. E. 2002. Botulism Due to *Clostridium baratii* Type F Toxin. *Journal of Clinical Microbiology,* 40**,** 2260-2262.

HAVLAK, P., CHEN, R., DURBIN, K. J., EGAN, A., REN, Y., SONG, X.-Z., WEINSTOCK, G. M. & GIBBS, R. A. 2004. The Atlas Genome Assembly System. *Genome research,* 14**,** 721-732.

HEAP, J. T., EHSAAN, M., COOKSLEY, C. M., NG, Y.-K., CARTMAN, S. T., WINZER, K. & MINTON, N. P. 2012. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic acids research,* 40**,** e59.

HEAP, J. T., KUEHNE, S. A., EHSAAN, M., CARTMAN, S. T., COOKSLEY, C. M., SCOTT, J. C. & MINTON, N. P. 2010. The ClosTron: Mutagenesis in Clostridium refined and streamlined. *Journal of microbiological methods,* 80**,** 49-55.

HEAP, J. T., PENNINGTON, O. J., CARTMAN, S. T., CARTER, G. P. & MINTON, N. P. 2007. The ClosTron: A universal gene knock-out system for the genus Clostridium. *Journal of microbiological methods,* 70**,** 452-464.

HEAP, J. T., PENNINGTON, O. J., CARTMAN, S. T. & MINTON, N. P. 2009. A modular system for Clostridium shuttle plasmids. *Journal of microbiological methods,* 78**,** 79-85.

HEAP, J. T., THEYS, J., EHSAAN, M., KUBIAK, A. M., DUBOIS, L., PAESMANS, K., VAN MELLAERT, L., KNOX, R., KUEHNE, S. A., LAMBIN, P. & MINTON, N. P. 2014. Spores of Clostridium engineered for clinical efficacy and safety cause regression and cure of tumors in vivo. *Oncotarget,* 5**,** 1761-1769.

HELDIN, C.-H., RUBIN, K., PIETRAS, K. & OSTMAN, A. 2004. High interstitial fluid pressure [mdash] an obstacle in cancer therapy. *Nat Rev Cancer,* 4**,** 806-813.

HENDERSON, B. W. & FINGAR, V. H. 1987. Relationship of Tumor Hypoxia and Response to Photodynamic Treatment in an Experimental Mouse Tumor. *Cancer research,* 47**,** 3110-3114.

HEPPNER, F. & MÖSE, J. R. 1978. The liquefaction (oncolysis) of malignant gliomas by a non pathogenic Clostridium. *Acta neurochirurgica,* 42**,** 123-125.

HOLT, R. A. & JONES, S. J. M. 2008. The new paradigm of flow cell sequencing. *Genome research,* 18**,** 839-846.

HTTP://WWW.CANCERRESEARCHUK.ORG/ABOUT-CANCER/CANCERS-IN-GENERAL/TREATMENT/. 28/07/2015. *Cancer Treatment Webpage* [Online]. Cancer Research UK. Available: http://www.cancerresearchuk.org/about-cancer/cancers-in-general/treatment/.

HTTPS://WWW.LIFESCIENCE.ROCHE.COM/SHOP/EN/UK/PRODUCTS/COMPLETE-EDTA-FREE-3271402-1. 14/10/2016. *cOmplete, EDTA Free Protease Inhibitor Cocktail Tablets Supporting Snformation* [Online]. Available: https://www.lifescience.roche.com/shop/en/uk/products/complete-edta-free-3271402-1.

HU, M. C. T. & DAVIDSON, N. 1987. The inducible lac operator-repressor system is functional in mammalian cells. *Cell,* 48**,** 555-566.

IANNELLO, A. & AHMAD, A. 2005. Role of antibody-dependent cell-mediated cytotoxicity in the efficacy of therapeutic anti-cancer monoclonal antibodies. *Cancer and Metastasis Reviews,* 24**,** 487-499.

JAIN, R. K. 2012. Delivery of molecular and cellular medicine to solid tumors(). *Advanced drug delivery reviews,* 64**,** 353-365.

JAIN, RAKESH K. 2014. Antiangiogenesis Strategies Revisited: From Starving Tumors to Alleviating Hypoxia. *Cancer Cell,* 26**,** 605-622.

JAIN, R. K., DUDA, D. G., CLARK, J. W. & LOEFFLER, J. S. 2006. Lessons from phase III clinical trials on anti-VEGF therapy for cancer. *Nat Clin Prac Oncol,* 3**,** 24-40.

KABBINAVAR, F., HURWITZ, H. I., FEHRENBACHER, L., MEROPOL, N. J., NOVOTNY, W. F., LIEBERMAN, G., GRIFFING, S. & BERGSLAND, E. 2003. Phase II, Randomized Trial Comparing Bevacizumab Plus Fluorouracil (FU)/Leucovorin (LV) With FU/LV Alone in Patients With Metastatic Colorectal Cancer. *Journal of Clinical Oncology,* 21**,** 60-65.

KALCHAYANAND, N., DUNNE, C. P., SIKES, A. & RAY, B. 2004. Germination induction and inactivation of Clostridium spores at medium-range hydrostatic pressure treatment. *Innovative Food Science & Emerging Technologies,* 5**,** 277-283.

KAVERMANN, H., BURNS, B. P., ANGERMÜLLER, K., ODENBREIT, S., FISCHER, W., MELCHERS, K. & HAAS, R. 2003. Identification and Characterization of Helicobacter pylori Genes Essential for Gastric Colonization. *The Journal of experimental medicine,* 197**,** 813-822.

KIRK, D. G. 2015. *Alternative sigma factors F, E, G, and K in Clostridium botulinum sporulation and stress response.* Doctor of Philosophy, University of Helsink.

KIRK, D. G., ZHANG, Z., KORKEALA, H. & LINDSTRÖM, M. 2014. Alternative Sigma Factors SigF, SigE, and SigG Are Essential for Sporulation in Clostridium botulinum ATCC 3502. *Applied and Environmental Microbiology,* 80**,** 5141-5150.

KOBAYASHI, H., WATANABE, R. & CHOYKE, P. L. 2013. Improving conventional enhanced permeability and retention (EPR) effects; what is the appropriate target? *Theranostics,* 4**,** 81-9.

KRUMAN, II, KUMARAVEL, T. S., LOHANI, A., PEDERSEN, W. A., CUTLER, R. G., KRUMAN, Y., HAUGHEY, N., LEE, J., EVANS, M. & MATTSON, M. P. 2002. Folic acid deficiency and homocysteine impair DNA repair in hippocampal neurons and sensitize them to amyloid toxicity in experimental models of Alzheimer's disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 22**,** 1752-62.

KUBIAK, A. M. 2013. *The use of Clostridium sporogenes in directed enzyme pro-drug therapy.* PhD Thesis (PhD)-, University of Nottingham.

KUBIAK, A. M. & MINTON, N. P. 2015. The potential of clostridial spores as therapeutic delivery vehicles in tumour therapy. *Research in Microbiology,* 166**,** 244-254.

KUBIAK, A. M., POEHLEIN, A., BUDD, P., KUEHNE, S. A., WINZER, K., THEYS, J., LAMBIN, P., DANIEL, R. & MINTON, N. P. 2015. Complete Genome Sequence of the Nonpathogenic Soil-Dwelling Bacterium *Clostridium sporogenes* Strain NCIMB 10696. *Genome Announcements,* 3.

KUDVA, R., DENKS, K., KUHN, P., VOGT, A., MÜLLER, M. & KOCH, H.-G. 2013. Protein translocation across the inner membrane of Gram-negative bacteria: the Sec and Tat dependent protein transport pathways. *Research in Microbiology,* 164**,** 505-534.

KUPELIAN, P. A., POTTERS, L., KHUNTIA, D., CIEZKI, J. P., REDDY, C. A., REUTHER, A. M., CARLSON, T. P. & KLEIN, E. A. 2004. Radical prostatectomy, external beam radiotherapy <72 Gy, external beam radiotherapy ≥72 Gy, permanent seed implantation, or combined seeds/external beam radiotherapy for stage T1–T2 prostate cancer. *International Journal of Radiation Oncology*Biology*Physics,* 58**,** 25-33.

LAMBIN, P., THEYS, J., LANDUYT, W., RIJKEN, P., VAN DER KOGEL, A., VAN DER SCHUEREN, E., HODGKISS, R., FOWLER, J., NUYTS, S., DE BRUIJN, E., VAN MELLAERT, L. & ANNÉ, J. 1998. Colonisation of Clostridium in the body is restricted to hypoxic and necrotic areas of tumours. *Anaerobe,* 4**,** 183-188.

LEE, W. H. & RIEMANN, H. 1970. The Genetic Relatedness of Proteolytic *Clostridium botulinum* Strains. *Microbiology,* 64**,** 85-90.

LEE, Y., LEE, E. K., CHO, Y. W., MATSUI, T., KANG, I.-C., KIM, T.-S. & HAN, M. H. 2003. ProteoChip: A highly sensitive protein microarray prepared by a novel method of protein immobilization for application of protein-protein interaction studies. *PROTEOMICS,* 3**,** 2289-2304.

LEONGAMORNLERT, D., MAHMUD, N., TYMRAKIEWICZ, M., SAUNDERS, E., DADAEV, T., CASTRO, E., GOH, C., GOVINDASAMI, K., GUY, M., O'BRIEN, L., SAWYER, E., HALL, A., WILKINSON, R., EASTON, D., GOLDGAR, D., EELES, R. & KOTE-JARAI, Z. 2012. Germline BRCA1 mutations increase prostate cancer risk. *British journal of cancer,* 106**,** 1697-1701.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012. Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology,* 2012**,** 251364.

LIU, S. C., MINTON, N. P., GIACCIA, A. J. & BROWN, J. M. 2002. Anticancer efficacy of systemically delivered anaerobic bacteria as gene therapy vectors targeting tumor hypoxia/necrosis. *Gene therapy,* 9**,** 291-6.

LOMAN, N. J., MISRA, R. V., DALLMAN, T. J., CONSTANTINIDOU, C., GHARBIA, S. E., WAIN, J. & PALLEN, M. J. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotech,* 30**,** 434-439.

MALMGREN, R. A. & FLANIGAN, C. C. 1955. Localization of the vegetative form of *Clostridium tetani* in mouse tumors following intravenous spore administration. *Cancer research,* 15**,** 473-8.

MANNAERTS, G. H. H., MARTIJN, H., CROMMELIN, M. A., DRIES, W., REPELAER VAN DRIEL, O. J. & RUTTEN, H. J. T. 2000. Feasibility and first results of multimodality treatment, combining ebrt, extensive surgery, and ioert in locally advanced primary rectal cancer. *International Journal of Radiation Oncology*Biology*Physics,* 47**,** 425-433.

MARDIS, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics,* 24**,** 133-141.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y.-J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L. I., JARVIE, T. P., JIRAGE, K. B., KIM, J.-B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-380.

MATTHEWS, B. W. 1988. Structural basis of the action of thermolysin and related zinc peptidases. *Accounts of Chemical Research,* 21**,** 333-340.

MAZMANIAN, S. K., TON-THAT, H., SU, K. & SCHNEEWIND, O. 2002. An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. *Proceedings of the National Academy of Sciences,* 99**,** 2293-2298.

MCCONNELL, J. D., BRUSKEWITZ, R., WALSH, P., ANDRIOLE, G., LIEBER, M., HOLTGREWE, H. L., ALBERTSEN, P., ROEHRBORN, C. G., NICKEL, J. C., WANG, D. Z., TAYLOR, A. M. & WALDSTREICHER, J. 1998. The Effect of Finasteride on the Risk of Acute Urinary Retention and the Need for Surgical Treatment among Men with Benign Prostatic Hyperplasia. *New England Journal of Medicine,* 338**,** 557-563.

MCILWAIN, D. R., BERGER, T. & MAK, T. W. 2013. Caspase Functions in Cell Death and Disease. *Cold Spring Harbor Perspectives in Biology,* 5.

MCKENNEY, P. T., DRIKS, A. & EICHENBERGER, P. 2013. The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. *Nat Rev Micro,* 11**,** 33-44.

MCNEISH, I. A., GREEN, N. K., GILLIGAN, M. G., FORD, M. J., MAUTNER, V., YOUNG, L. S., KERR, D. J. & SEARLE, P. F. 1998. Virus directed enzyme prodrug therapy for ovarian and pancreatic cancer using retrovirally delivered *E. coli* nitroreductase and CB1954. *Gene therapy,* 5**,** 1061-9.

MEHMET, H. 2000. Apoptosis: Caspases find a new place to hide. *Nature,* 403**,** 29-30.

MELERO, I., HERVAS-STUBBS, S., GLENNIE, M., PARDOLL, D. M. & CHEN, L. 2007. Immunostimulatory monoclonal antibodies for cancer therapy. *Nature reviews. Cancer,* 7**,** 95-106.

MERRIMAN, B., D TEAM, I. T. & ROTHBERG, J. M. 2012. Progress in Ion Torrent semiconductor chip based sequencing. *ELECTROPHORESIS,* 33**,** 3397-3417.

METZKER, M. L. 2010. Sequencing technologies [mdash] the next generation. *Nat Rev Genet,* 11**,** 31-46.

MIKI, Y., SWENSEN, J., SHATTUCK-EIDENS, D., FUTREAL, P., HARSHMAN, K., TAVTIGIAN, S., LIU, Q., COCHRAN, C., BENNETT, L., DING, W. & ET, A. 1994. A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science,* 266**,** 66-71.

MINTON, N. P. 2003. Clostridia in cancer therapy. *Nature reviews. Microbiology,* 1**,** 237-42.

MINTON, N. P., ATKINSON, T. & SHERWOOD, R. F. 1983. Molecular cloning of the Pseudomonas carboxypeptidase G2 gene and its expression in *Escherichia coli* and *Pseudomonas putida*. *Journal of bacteriology,* 156**,** 1222-7.

MINTON, N. P., MAUCHLINE, M. L., LEMMON, M. J., BREHM, J. K., FOX, M., MICHAEL, N. P., GIACCIA, A. & BROWN, J. M. 1995. Chemotherapeutic tumour targeting using clostridial spores. *FEMS Microbiology Reviews,* 17**,** 357-364.

MOCARSKI, E. S., UPTON, J. W. & KAISER, W. J. 2012. Viral infection and the evolution of caspase 8-regulated apoptotic and necrotic death pathways. *Nat Rev Immunol,* 12**,** 79-88.

MOLL, S., NICKELEIT, V., MUELLER-BRAND, J., BRUNNER, F. P., MAECKE, H. R. & MIHATSCH, M. J. 2001. A new cause of renal thrombotic microangiopathy: Yttrium 90-DOTATOC internal radiotherapy. *American Journal of Kidney Diseases,* 37**,** 847-851.

MORGAN, B. P. & MERI, S. 1994. Membrane proteins that protect against complement lysis. *Springer Seminars in Immunopathology,* 15**,** 369-396.

MOROZOVA, O. & MARRA, M. A. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics,* 92**,** 255-264.

MÖSE, J. R. & MÖSE, G. 1964. Oncolysis by Clostridia. I. Activity of *Clostridium Butyricum* (M-55) and Other Nonpathogenic Clostridia against the Ehrlich Carcinoma. *Cancer research,* 24**,** 212-216.

MULLER, P. A. J. & VOUSDEN, K. H. 2013. p53 mutations in cancer. *Nat Cell Biol,* 15**,** 2-8.

MUNOZ-ESPIN, D. & SERRANO, M. 2014. Cellular senescence: from physiology to pathology. *Nature reviews. Molecular cell biology,* 15**,** 482-96.

NAPIER, M. P., SHARMA, S. K., SPRINGER, C. J., BAGSHAWE, K. D., GREEN, A. J., MARTIN, J., STRIBBLING, S. M., CUSHEN, N., O'MALLEY, D. & BEGENT, R. H. 2000. Antibody-directed enzyme prodrug therapy: efficacy and mechanism of action in colorectal carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research,* 6**,** 765-72.

NATSUME, A., NIWA, R. & SATOH, M. 2009. Improving effector functions of antibodies for cancer treatment: Enhancing ADCC and CDC. *Drug design, development and therapy,* 3**,** 7-16.

NAVARRE, W. W. & SCHNEEWIND, O. 1999. Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope. *Microbiology and Molecular Biology Reviews,* 63**,** 174-229.

NG, Y. K., EHSAAN, M., PHILIP, S., COLLERY, M. M., JANOIR, C., COLLIGNON, A., CARTMAN, S. T. & MINTON, N. P. 2013. Expanding the Repertoire of Gene Tools for Precise Manipulation of the *Clostridium difficile* Genome: Allelic Exchange Using *pyrE* Alleles. *PloS one,* 8**,** e56051.

NOBBS, A. H., VAJNA, R. M., JOHNSON, J. R., ZHANG, Y., ERLANDSEN, S. L., OLI, M. W., KRETH, J., BRADY, L. J. & HERZBERG, M. C. 2007. Consequences of a sortase A mutation in *Streptococcus gordonii*. *Microbiology,* 153**,** 4088-97.

OPGEN. 2015. *Whole genome mapping:* [Online]. Available: http://opgen.com/genomic-services/what-is-whole-genome-mapping.

PALLEN, M. J., LAM, A. C., ANTONIO, M. & DUNBAR, K. 2001. An embarrassment of sortases - a richness of substrates? *Trends in Microbiology,* 9**,** 97-102.

PATERSON, G. K. & MITCHELL, T. J. 2004. The biology of Gram-positive sortase enzymes. *Trends in Microbiology,* 12**,** 89-95.

PERÄNEN, J., RIKKONEN, M., HYVÖNEN, M. & KÄÄRIÄINEN, L. 1996. T7 Vectors with a Modified T7lacPromoter for Expression of Proteins in *Escherichia coli*. *Analytical Biochemistry,* 236**,** 371-373.

PEREIRA, F. C., SAUJET, L., TOMÉ, A. R., SERRANO, M., MONOT, M., COUTURE-TOSI, E., MARTIN-VERSTRAETE, I., DUPUY, B. & HENRIQUES, A. O. 2013. The Spore Differentiation Pathway in the Enteric Pathogen *Clostridium difficile*. *PLoS Genet,* 9**,** e1003782.

PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth,* 8**,** 785-786.

POEHLEIN, A., RIEGEL, K., KÖNIG, S. M., LEIMBACH, A., DANIEL, R. & DÜRRE, P. 2015. Genome sequence of Clostridium sporogenes DSM 795(T), an amino acid-degrading, nontoxic surrogate of neurotoxin-producing *Clostridium botulinum*. *Standards in genomic sciences,* 10**,** 40.

POP, M. 2009. Genome assembly reborn: recent computational challenges. *Briefings in Bioinformatics,* 10**,** 354-366.

POP, M., PHILLIPPY, A., DELCHER, A. L. & SALZBERG, S. L. 2004. Comparative genome assembly. *Briefings in Bioinformatics,* 5**,** 237-248.

POP, M. & SALZBERG, S. L. 2008. Bioinformatics challenges of new sequencing technology. *Trends in Genetics,* 24**,** 142-149.

PRESTA, L. G., CHEN, H., O'CONNOR, S. J., CHISHOLM, V., MENG, Y. G., KRUMMEN, L., WINKLER, M. & FERRARA, N. 1997. Humanization of an Anti-Vascular Endothelial Growth Factor Monoclonal Antibody for the Therapy of Solid Tumors and Other Disorders. *Cancer research,* 57**,** 4593-4599.

PYLAYEVA-GUPTA, Y., GRABOCKA, E. & BAR-SAGI, D. 2011. *RAS* oncogenes: weaving a tumorigenic web. *Nature reviews. Cancer,* 11**,** 761-774.

QUAIL, M. A., SMITH, M., COUPLAND, P., OTTO, T. D., HARRIS, S. R., CONNOR, T. R., BERTONI, A., SWERDLOW, H. P. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics,* 13**,** 341.

QUIXLEY, K. W. & REID, S. J. 2000. Construction of a reporter gene vector for *Clostridium beijerinckii* using a Clostridium endoglucanase gene. *Journal of molecular microbiology and biotechnology,* 2**,** 53-7.

RAJALINGAM, K., SCHRECK, R., RAPP, U. R. & ALBERT, Š. 2007. *Ras* oncogenes and their downstream targets. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1773**,** 1177-1195.

RAJESWARAN, A., TROJAN, A., BURNAND, B. & GIANNELLI, M. 2008. Efficacy and side effects of cisplatin- and carboplatin-based doublet chemotherapeutic regimens versus non-platinum-based doublet chemotherapeutic regimens as first line treatment of metastatic non-small cell lung carcinoma: A systematic review of randomized controlled trials. *Lung cancer,* 59**,** 1-11.

RAK, J., YU, J. L., KLEMENT, G. & KERBEL, R. S. 2000. Oncogenes and Angiogenesis: Signaling Three-Dimensional Tumor Growth. *J Investig Dermatol Symp Proc,* 5**,** 24-33.

RANQUIN, A., VERSEES, W., MEIER, W., STEYAERT, J. & VAN GELDER, P. 2005. Therapeutic nanoreactors: combining chemistry and biology in a novel triblock copolymer drug delivery system. *Nano letters,* 5**,** 2220-4.

RISAU, W. 1997. Mechanisms of angiogenesis. *Nature,* 386**,** 671-674.

ROBERTS, N. J., ZHANG, L., JANKU, F., COLLINS, A., BAI, R.-Y., STAEDTKE, V., RUSK, A. W., TUNG, D., MILLER, M., ROIX, J., KHANNA, K. V., MURTHY, R., BENJAMIN, R. S., HELGASON, T., SZVALB, A. D., BIRD, J. E., ROY-CHOWDHURI, S., ZHANG, H. H., QIAO, Y., KARIM, B., MCDANIEL, J., ELPINER, A., SAHORA, A., LACHOWICZ, J., PHILLIPS, B., TURNER, A., KLEIN, M. K., POST, G., DIAZ, L. A., RIGGINS, G. J., PAPADOPOULOS, N., KINZLER, K. W., VOGELSTEIN, B., BETTEGOWDA, C., HUSO, D. L., VARTERASIAN, M., SAHA, S. & ZHOU, S. 2014. Intratumoral injection of *Clostridium novyi*-NT spores induces antitumor responses. *Science translational medicine,* 6**,** 249ra111-249ra111.

ROWSELL, S., PAUPTIT, R. A., TUCKER, A. D., MELTON, R. G., BLOW, D. M. & BRICK, P. 1997. Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure,* 5**,** 337-347.

SCHNEEWIND, O. & MISSIAKAS, D. 2014. Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1843**,** 1687-1697.

SCOTT, A. M., WOLCHOK, J. D. & OLD, L. J. 2012. Antibody therapy of cancer. *Nature reviews. Cancer,* 12**,** 278-287.

SEBAIHIA, M., PECK, M. W., MINTON, N. P., THOMSON, N. R., HOLDEN, M. T. G., MITCHELL, W. J., CARTER, A. T., BENTLEY, S. D., MASON, D. R., CROSSMAN, L., PAUL, C. J., IVENS, A., WELLS-BENNIK, M. H. J., DAVIS, I. J., CERDEÑO-TÁRRAGA, A. M., CHURCHER, C., QUAIL, M. A., CHILLINGWORTH, T., FELTWELL, T., FRASER, A., GOODHEAD, I., HANCE, Z., JAGELS, K., LARKE, N., MADDISON, M., MOULE, S., MUNGALL, K., NORBERTCZAK, H., RABBINOWITSCH, E., SANDERS, M., SIMMONDS, M., WHITE, B., WHITHEAD, S. & PARKHILL, J. 2007. Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. *Genome research,* 17**,** 1082-1092.

SHAH, M. A., JHAWER, M., ILSON, D. H., LEFKOWITZ, R. A., ROBINSON, E., CAPANU, M. & KELSEN, D. P. 2011. Phase II Study of Modified Docetaxel, Cisplatin, and Fluorouracil With Bevacizumab in Patients With Metastatic Gastroesophageal Adenocarcinoma. *Journal of Clinical Oncology,* 29**,** 868-874.

SHENDURE, J., MITRA, R. D., VARMA, C. & CHURCH, G. M. 2004. Advanced sequencing technologies: methods and goals. *Nat Rev Genet,* 5**,** 335-344.

SHERR, C. J. 1996. Cancer cell cycles. *Science,* 274**,** 1672-7.

SHERR, C. J. 2000. The Pezcoller Lecture: Cancer Cell Cycles Revisited. *Cancer research,* 60**,** 3689-3695.

SINHA, R., KIM, G. J., NIE, S. & SHIN, D. M. 2006. Nanotechnology in cancer therapeutics: bioconjugated nanoparticles for drug delivery. *Molecular Cancer Therapeutics,* 5**,** 1909-1917.

SNOW, B. J., TSUI, J. K. C., BHATT, M. H., VARELAS, M., HASHIMOTO, S. A. & CALNE, D. B. 1990. Treatment of spasticity with botulinum toxin: A double-blind study. *Annals of Neurology,* 28**,** 512-515.

SPRINGER, C. J. & NICULESCU-DUVAZ, I. I. 1997. Antibody-directed enzyme prodrug therapy (ADEPT): a review. *Advanced drug delivery reviews,* 26**,** 151-172.

STEVENS, D. L. & BRYANT, A. E. 1993. Role of θ Toxin, a Sulfhydryl-Activated Cytolysin, in the Pathogenesis of Clostridial Gas Gangrene. *Clinical Infectious Diseases,* 16**,** S195-S199.

SUNG, K. W., SON, M. H., LEE, S. H., YOO, K. H., KOO, H. H., KIM, J. Y., CHO, E. J., LEE, S. K., CHOI, Y. S., LIM, D. H., KIM, J. S. & KIM, D. W. 2013. Tandem high-dose chemotherapy and autologous stem cell transplantation in patients with high-risk neuroblastoma: Results of SMC NB-2004 study. *Bone marrow transplantation,* 48**,** 68-73.

TABOR, S. & RICHARDSON, C. C. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences,* 82**,** 1074-1078.

TAKANO, T., TIAN, G.-F., PENG, W., LOU, N., LOVATT, D., HANSEN, A. J., KASISCHKE, K. A. & NEDERGAARD, M. 2007. Cortical spreading depression causes and coincides with tissue hypoxia. *Nat Neurosci,* 10**,** 754-762.

TALUKDAR, P. K., OLGUÍN-ARANEDA, V., ALNOMAN, M., PAREDES-SABJA, D. & SARKER, M. R. 2015. Updates on the sporulation process in Clostridium species. *Research in Microbiology,* 166**,** 225-235.

TERPE, K. 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology,* 72**,** 211-222.

THEYS, J., NUYTS, S., LANDUYT, W., VAN MELLAERT, L., DILLEN, C., BÖHRINGER, M., DÜRRE, P., LAMBIN, P. & ANNÉ, J. 1999. Stable *Escherichia coli-Clostridium acetobutylicum* Shuttle Vector for Secretion of Murine Tumor Necrosis Factor Alpha. *Applied and Environmental Microbiology,* 65**,** 4295-4300.

THEYS, J., PENNINGTON, O., DUBOIS, L., ANLEZARK, G., VAUGHAN, T., MENGESHA, A., LANDUYT, W., ANNE, J., BURKE, P. J., DURRE, P., WOUTERS, B. G., MINTON, N. P. & LAMBIN, P. 2006a. Repeated cycles of Clostridium-directed enzyme prodrug therapy result in sustained antitumour effects in vivo. *British journal of cancer,* 95**,** 1212-1219.

THEYS, J., PENNINGTON, O., DUBOIS, L., ANLEZARK, G., VAUGHAN, T., MENGESHA, A., LANDUYT, W., ANNE, J., BURKE, P. J., DURRE, P., WOUTERS, B. G., MINTON, N. P. & LAMBIN, P. 2006b. Repeated cycles of Clostridium-directed enzyme prodrug therapy result in sustained antitumour effects in vivo. *Br J Cancer,* 95**,** 1212-1219.

TIRANDAZ, H., SALEHI-NAJAFABADI, A. & ESLAMIFAR, A. 2006. Cytosine deaminase producing Clostridium may be used in detection of tumors. *Medical hypotheses,* 67**,** 1257-1258.

TREANGEN, T. J. & SALZBERG, S. L. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet,* 13**,** 36-46.

TRÉDAN, O., GALMARINI, C. M., PATEL, K. & TANNOCK, I. F. 2007. Drug Resistance and the Solid Tumor Microenvironment. *Journal of the National Cancer Institute,* 99**,** 1441-1454.

TULLI, L., MARCHI, S., PETRACCA, R., SHAW, H. A., FAIRWEATHER, N. F., SCARSELLI, M., SORIANI, M. & LEUZZI, R. 2013. CbpA: a novel surface exposed adhesin of *Clostridium difficile* targeting human collagen. *Cellular Microbiology,* 15**,** 1674-1687.

VALIENTE, E., LEE, C.-T., HOR, L.-I., FOUZ, B. & AMARO, C. 2008. Role of the metalloprotease Vvp and the virulence plasmid pR99 of *Vibrio vulnificus* serovar E in surface colonization and fish virulence. *Environmental Microbiology,* 10**,** 328-338.

VON HOFF, D. D., ERVIN, T., ARENA, F. P., CHIOREAN, E. G., INFANTE, J., MOORE, M., SEAY, T., TJULANDIN, S. A., MA, W. W., SALEH, M. N., HARRIS, M., RENI, M., DOWDEN, S., LAHERU, D., BAHARY, N., RAMANATHAN, R. K., TABERNERO, J., HIDALGO, M., GOLDSTEIN, D., VAN CUTSEM, E., WEI, X., IGLESIAS, J. & RENSCHLER, M. F. 2013. Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine. *New England Journal of Medicine,* 369**,** 1691-1703.

WALIGORA, A. J., HENNEQUIN, C., MULLANY, P., BOURLIOUX, P., COLLIGNON, A. & KARJALAINEN, T. 2001. Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. *Infection and immunity,* 69**,** 2144-53.

WEIDENMAIER, C., KOKAI-KUN, J. F., KULAUZOVIC, E., KOHLER, T., THUMM, G., STOLL, H., GÖTZ, F. & PESCHEL, A. 2008. Differential roles of sortase-anchored surface proteins and wall teichoic acid in *Staphylococcus aureus* nasal colonization. *International Journal of Medical Microbiology,* 298**,** 505-513.

WEIGAND, M. R., PENA-GONZALEZ, A., SHIREY, T. B., BROEKER, R. G., ISHAQ, M. K., KONSTANTINIDIS, K. T. & RAPHAEL, B. H. 2015. Implications of Genome-Based Discrimination between *Clostridium botulinum* Group I and *Clostridium sporogenes* Strains for Bacterial Taxonomy. *Applied and Environmental Microbiology,* 81**,** 5420-5429.

WEISER, J. N., BAE, D., FASCHING, C., SCAMURRA, R. W., RATNER, A. J. & JANOFF, E. N. 2003. Antibody-enhanced pneumococcal adherence requires IgA1 protease. *Proceedings of the National Academy of Sciences,* 100**,** 4215-4220.

WELLS, C. L. & WILKINS, T. D. 1996. Clostridia: Sporeforming Anaerobic Bacilli. *In:* BARON, S. (ed.) *Medical Microbiology.* 4th ed. Galveston (TX).

WENTWORTH, P., DATTA, A., BLAKEY, D., BOYLE, T., PARTRIDGE, L. J. & BLACKBURN, G. M. 1996. Toward antibody-directed "abzyme" prodrug therapy, ADAPT: carbamate prodrug activation by a catalytic antibody and its in vitro application to human tumor cell killing. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 799-803.

WHO. 2014. *World Health Organisation Cancer Fact Sheets* [Online]. World Health Organisation. Available: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.

WHO 2015. Cancer Factsheet No297.

WILSON, W. R. & HAY, M. P. 2011. Targeting hypoxia in cancer therapy. *Nature reviews. Cancer,* 11**,** 393-410.

WISHART, D. S., KNOX, C., GUO, A. C., SHRIVASTAVA, S., HASSANALI, M., STOTHARD, P., CHANG, Z. & WOOLSEY, J. 2006. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research,* 34**,** D668-D672.

WU, M., FRIEBOES, H. B., MCDOUGALL, S. R., CHAPLAIN, M. A. J., CRISTINI, V. & LOWENGRUB, J. 2013. The effect of interstitial pressure on tumor growth: coupling with the blood and lymphatic vascular systems. *Journal of theoretical biology,* 320**,** 131-151.

XU, G. & MCLEOD, H. L. 2001. Strategies for enzyme/prodrug cancer therapy. *Clinical cancer research : an official journal of the American Association for Cancer Research,* 7**,** 3314-24.

ZAPPE, H., JONES, W. A., JONES, D. T. & WOODS, D. R. 1988. Structure of an endo-beta-1,4-glucanase gene from *Clostridium acetobutylicum* P262 showing homology with endoglucanase genes from Bacillus spp. *Applied and Environmental Microbiology,* 54**,** 1289-92.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome research,* 18**,** 821-829.

ZHANG, Y., GROSSE-HONEBRINK, A. & MINTON, N. P. 2015. A Universal *Mariner* Transposon System for Forward Genetic Studies in the Genus *Clostridium*. *PloS one,* 10**,** e0122411.

ZHU, Q., PAO, G. M., HUYNH, A. M., SUH, H., TONNU, N., NEDERLOF, P. M., GAGE, F. H. & VERMA, I. M. 2011. BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature,* 477**,** 179-184.

Appendices

8 Appendices

Lovitt *et al.* – MACC (AMINO ACID COMPLETE MEDIA)

LPBM Media	1L	400 ML
Component	Quantity	
KH_2PO_4	2.0 g/l	800 mg
K_2HPO_4	2.0 g/l	800 mg
$MgCl_2$	93 mg/l	37 mg
$(NH_4)_2SO_4$	5.0 g/l	2 g
10% $NaHCO_3$ 0.01% w/v	25 ml added after sterilisation	10 ml
Resazurin solution	1 ml	0.4 ml

General Vitamin solution 10 ml after sterilisation (FS) 4 ml

		100x Stock in 100ml
Biotin	2 mg/l	20 mg
Folic acid	2 mg/l	20 mg
Pyridoxamine HCL	10 mg/l	100 mg
Thiamine HCL	5 mg/l	50 mg
Riboflavin HCL	5 mg/l	50 mg
Nicotinic acid	5 mg/l	50 mg
Pantothenic acid	5 mg/l	50 mg
Cyanocobalamin	1 mg/l	10 mg
p-aminobenzoic acid	5 mg/l	50 mg
lipoic acid	5 mg/l	50 mg

Trace element solution	10 ml	4ml
		10x Stock in 100 ml
Nitrilotriacetic acid 12.8 g/l	12.8 g	
$FeSO_4.7H_2O$	0.1 g/l	100 mg
$MnCl_2$	63 mg/l	63 mg
$CoCl_2.6H_2O$	0.24 g/l	240 mg
$CaCl_2$	75 mg/l	75 mg
$ZnCl_2$	0.1 g/l	100 mg
$CuCl_2.2H_2O$	13 mg/l	13 mg
H_3BO_4	0.01 g/l	10 mg
$NaMoO_4.2H_2O$	0.01 g/l	10 mg
NaCl	1.0 g/l	1 g
$NaSeO_3$	0.017 g/l	17 mg
$NiSO_4.6H_2O$	0.026 g/l	26 mg
$NaWO_4.2H_2O$	0.1 g/l	100 mg

After all supplements were added, *5 %* w/v cysteine HCl (pH *6)* was added until reduction was effected (as judged by the discolouration of the resazurin); this required *ca* 5-10 ml/l medium.

ACC (amino acid complete) – make stock, autoclave all except tryptophan (FS)

Glycine	3 mmol/l	
Valine	2 mmol/l	
Isoleucine	2 mmol/l	
Arginine	2 mmol/l	
Leucine	1 mmol/l	
Histidine	1 mmol/l	
Methionine	1 mmol/l	
Phenylalanine	1 mmol/l	
Tryptophan	1 mmol/l	
Tyrosine	0.125 mmol/l	0.5% glucose added to final conc.

M9 minimal media for Pseudomonas

M9 minimal media was prepared as follows 500 ml of 1 x M9 salts were supplemented with 87 % (w/v) glycerol to a final concentration of 2 %, $MgSO_4$ (2 mM), $CaCl_2$ (0.1 mM) and Glutamate (0.05 %).

M9 Minimal Media Salts

In 0.46 l water:

7.5 g	$Na_2HPO_4 \cdot 12H_2O$,
3.0 g	KH_2PO_4,
0.25g	NaCl, 5.0 g

Signal Sequences of the different proposed exported proteins.

Botulinum Signal Peptides

1442 MKSKKLLTILLSAIITASSISSVYA AESVG

>For Primer

GATCCATGAAAAGTAAAAAATTATTAACAATATTATTAAGTGCAATAATAACAGCAAGTAGTATA
AGTAGTGTTTATGCAGCAGAAC

>Rev Primer

TCGAGTTCTGCTGCATAAACACTACTTATACTACTTGCTGTTATTATTGCACTTAATAATATTG
TTAATAATTTTTTACTTTTCATG

1443 MKSKKLLTILLSAIITASSISSVYA AESVG

>For Primer

GATCCATGAAAAGTAAAAAATTATTAACAATATTATTAAGTGCAATAATAACAGCAAGTAGTATA
AGTAGTGTTTATGCAGCAGAAC

>Rev Primer

TCGAGTTCTGCTGCATAAACACTACTTATACTACTTGCTGTTATTATTGCACTTAATAATATTG
TTAATAATTTTTTACTTTTCATG

1444 MKSKKLLATVLSAVITFSAVSAVSA APVGK

>For Primer

GATCCATGAAAAGTAAAAAATTATTAGCAACAGTTTTAAGTGCAGTTATAACATTTAGTGCAGTT
AGTGCAGTTAGTGCAGCACCTC

>Rev Primer

TCGAGAGGTGCTGCACTAACTGCACTAACTGCACTAAATGTTATAACTGCACTTAAAACTGT
TGCTAATAATTTTTTACTTTTCATG

1445 MKSKKLLATVLSAVITFSTVSAVYA APVGK

>For Primer

GATCCATGAAAAGTAAAAAATTATTAGCAACAGTTTTAAGTGCAGTTATAACATTTAGTACAGTT
AGTGCAGTTTATGCAGCACCTC

>Rev Primer

TCGAGAGGTGCTGCATAAACTGCACTAACTGTACTAAATGTTATAACTGCACTTAAAACTGT
TGCTAATAATTTTTTACTTTTCATG

1446 MKSKKLLATVLSAVITFSTVSAVSA APVGK

>For Primer

GATCCATGAAAAGTAAAAAATTATTAGCAACAGTTTTAAGTGCAGTTATAACATTTAGTACAGTT
AGTGCAGTTAGTGCAGCACCTC

>Rev Primer

TCGAGAGGTGCTGCACTAACTGCACTAACTGTACTAAATGTTATAACTGCACTTAAAACTGT
TGCTAATAATTTTTTACTTTTCATG

Sporogenes Signal Peptides

1483 MKSKKLLATVLSAVITLSTVSAVYA APVGK

>For Primer

GATCCATGAAAAGTAAAAAATTATTAGCAACAGTTTTAAGTGCAGTTATAACATTAAGTACAGTT
AGTGCAGTTTATGCAGCACCTC

>Rev Primer

TCGAGAGGTGCTGCATAAACTGCACTAACTGTACTTAATGTTATAACTGCACTTAAAACTGTT
GCTAATAATTTTTTACTTTTCATG

1484 MKSKRLLATVLSAIITFSAVSAVSA APVGK
>For Primer
GATCCATGAAAAGTAAAAGATTATTAGCAACAGTTTTAAGTGCAATAATAACATTTAGTGCAGTTAGT
GCAGTTAGTGCAGCACCTC
>Rev Primer
TCGAGAGGTGCTGCACTAACTGCACTAACTGCACTAAATGTTATTATTGCACTTAAAACTGTT
GCTAATAATCTTTTACTTTTCATG

1485 MKSKKLLATVLSAVITFSAVSAVSA APVGK
>For Primer
GATCCATGAAAAGTAAAAAATTATTAGCAACAGTTTTAAGTGCAGTTATAACATTTAGTGCAGTT
AGTGCAGTTAGTGCAGCACCTC
>Rev Primer
TCGAGAGGTGCTGCACTAACTGCACTAACTGCACTAAATGTTATAACTGCACTTAAAACTGT
TGCTAATAATTTTTTACTTTTCATG

Npr MKSKKLLTILLSTIMTFSPVSSIYA AGPVG
>For Primer
GATCCATGAAAAGTAAAAAATTATTAACAATATTATTAAGTACAATAATGACATTTAGTCCTGTT
AGTAGTATATATGCAGCAGGAC
>Rev Primer
TCGAGTCCTGCTGCATATATACTACTAACAGGACTAAATGTCATTATTGTACTTAATAATATT
GTTAATAATTTTTTACTTTTCATG

C. cellulolyticum Cip C>MRKKSLAFLLALTMLVTLLGAQLTAFAAG
>For Primer
GATCCATGAGAAAAAAAAGTTTAGCATTTTTATTAGCATTAACAATGTTAGTTACATTATTAGGA
GCACAATTAACAGCATTTGCAGCAGGAC
>Rev Primer
TCGAGTCCTGCTGCAAATGCTGTTAATTGTGCTCCTAATAATGTAACTAACATTGTTAATGCT
AATAAAAATGCTAAACTTTTTTTTCTCATG

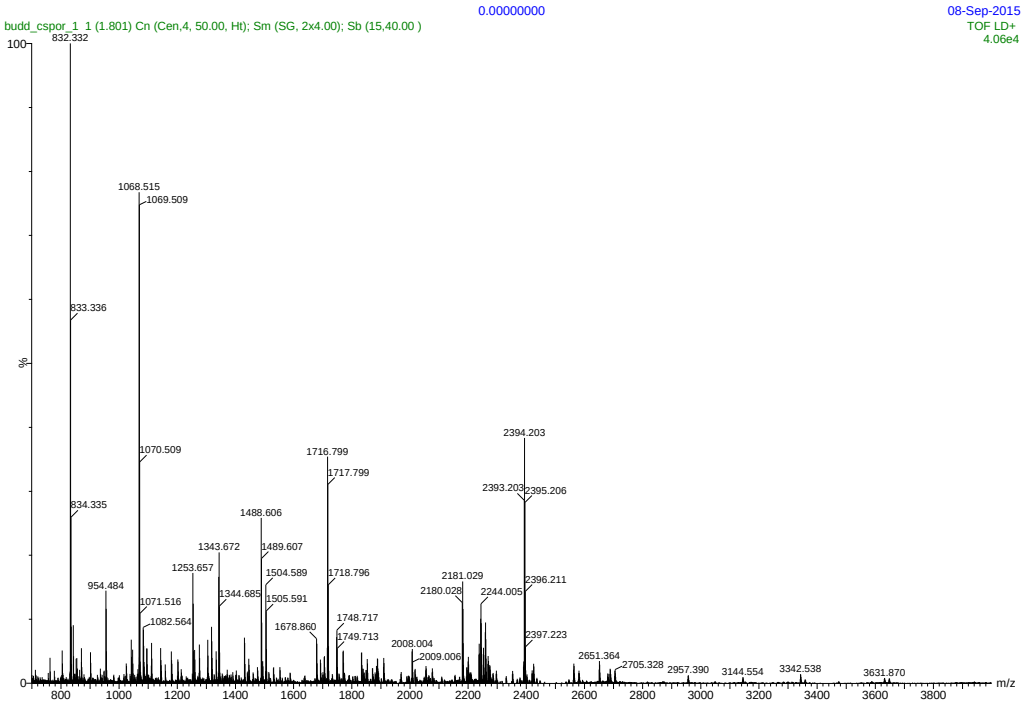
C. thermocellum Cip C>MRKVISMLLVVAMLTTIFAAMIPQTVSAAT
>For Primer
GATCCATGAGAAAAGTTATAAGTATGTTATTAGTTGTTGCAATGTTAACAACAATATTTGCAGCA
ATGATACCTCAAACAGTTAGTGCAGCAACA
>Rev Primer
TCGAGTGTTGCTGCACTAACTGTTTGAGGTATCATTGCTGCAAATATTGTTGTTAACATTGCA
ACAACTAATAACATACTTATAACTTTTCTCATG

C. acetobutylicum Cip C>MKKRNIAILGMFMMTASLGISKNHVFADT
>For Primer
GATCCATGAAAAAAAGAAATATAGCAATATTAGGAATGTTTATGATGACAGCAAGTTTAGGAATA
AGTAAAAATCATGTTTTTGCAGATACAC
>Rev Primer
TCGAGTGTATCTGCAAAAACATGATTTTTACTTATTCCTAAACTTGCTGTCATCATAAACATTC
CTAATATTGCTATATTTCTTTTTTTCATG

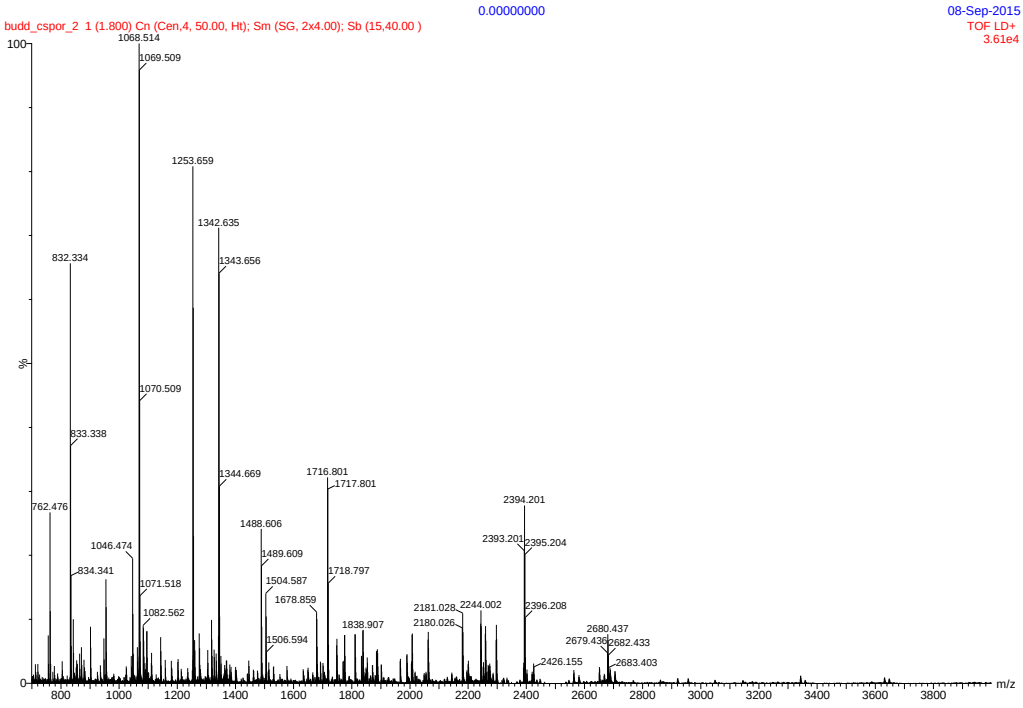
Barnase>MKKKNFAAICFTLVFMIFGAFLNNVVCVQAKT
>For Primer
GATCCATGAAAAAAAAAAAATTTTGCAGCAATATGTTTTACATTAGTTTTTATGATATTTGGAGCA
TTTTTAAATAATGTTGTTTGTGTTCAAGCAAAAACAC
>Rev Primer
TCGAGTGTTTTTGCTTGAACACAAACAACATTATTTAAAAATGCTCCAAATATCATAAAAACT
AATGTAAAACATATTGCTGCAAAATTTTTTTTTTTCATG

Maldi Mass Fingerprint of *C. Sporogenes* WT Supernatant SDS-PAGE bands

Band 1



Band 2


budd_cspor_2 1 (1.800) Cn (Cen,4, 50.00, Ht); Sm (SG, 2x4.00); Sb (15,40.00)
0.00000000
08-Sep-2015
TOF LD+
3.61e4
100
%
0
m/z
1068.514
1069.509
1253.659
1342.635
832.334
1343.656
1070.509
833.338
1344.669
1716.801
1717.801
762.476
2394.201
1488.606
2393.201
2395.204
1046.474
1489.609
834.341
1504.587
1718.797
1071.518
1678.859
2181.028
2244.002
2396.208
1082.562
2180.026
1838.907
2680.437
1506.594
2679.436
2682.433
2426.155
2683.403
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
3200
3400
3600
3800

ELSEVIER LICENSE
TERMS AND CONDITIONS

Apr 19, 2016

This is an Agreement between Patrick Budd ("You") and Elsevier ("Elsevier"). It consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Patrick Budd
Customer address	Centre for Biomolecular Sciences Nottingham, None NG7 2RD
License number	3803051153179
License date	Feb 06, 2016
Licensed content publisher	Elsevier
Licensed content publication	Cell
Licensed content title	The Hallmarks of Cancer
Licensed content author	Douglas Hanahan,Robert A Weinberg
Licensed content date	7 January 2000
Licensed content volume number	100
Licensed content issue number	1
Number of pages	14
Start Page	57
End Page	70
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	**1**
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 2
Title of your	Implementation and optimisation of alternative therapeutics for use in Clostridium

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at htto://myaccount.coovright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. Posting licensed content on any Website: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: **All** content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. For journal authors: the following clauses are applicable in addition to the above:
Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which may include any author-incorporated changes suggested through the processes of submission, peer-review and editor-author communication after peer review.

Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - bylupdating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do

research accessed under that agreement. This includes use for classroom teaching and internal training at the

institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. For book authors the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.
19. Thesis/Dissertation: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article

be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

20. **Other Conditions:**

v1.8

Questions? customercare Ca **or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.**

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Apr 19, 2016

This is an Agreement between Patrick Budd ("You") and Nature Publishing Group ("Nature Publishing Group"). It consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3803060626090
License date	Feb 06, 2016
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature Reviews Cancer
Licensed Content Title	Tumorigenesis and the angiogenic switch
Licensed Content Author	Gabriele BergersandLaura E. Benjamin
Licensed Content Date	Jun 1, 2003
Volume number	3
Issue number	6
Type of Use	reuse in a dissertation /thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 2 The classical angiogenic switch
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Implementation and optimisation of alternative therapeutics for use in Clostridium sporogenes as a delivery vehicle
Expected completion date	May 2016
Estimated size (number of pages)	220
Total	**0.00 USD**

Terms and Conditions

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use,subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you

paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   For AOP papers, the credit line should read:

   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   Note: For republication from the *British Journal of Cancer*, the following credit lines apply. Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the credit line should read: Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   Note: For adaptation from the *British Journal of Cancer*, the following credit line applies.

   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information.Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the *British Journal of Cancer*, the following credit line applies.

   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Apr 19, 2016

This is an Agreement between Patrick Budd ("You") and Elsevier ("Elsevier"). It consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Patrick Budd
Customer address	Centre for Biomolecular Sciences Nottingham, None NG7 2RD
License number	3803070028670
License date	Feb 06, 2016
Licensed content publisher	Elsevier
Licensed content publication	Innovative Food Science & Emerging Technologies
Licensed content title	Germination induction and inactivation of Clostridium spores at medium-range hydrostatic pressure treatment
Licensed content author	N. Kalchayanand,C.P. Dunne,A. Sikes,B. Ray
Licensed content date	September 2004
Licensed content volume number	5
Licensed content issue number	3
Number of figures/tables/illustrations	1
	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 1 part a)
Title of your thesis/dissertation	Implementation and optimisation of alternative therapeutics for use in Clostridium sporogenes as a delivery vehicle
Expected completion date	May 2016
Estimated size (number of pages)	220

VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	**0.00 USD**

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier

The following terms and conditions apply only to specific license types:

15. Translation: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. Posting licensed content on any Website: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. For journal authors: the following clauses are applicable in addition to the above:
Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author-incorporated changes suggested during submission, peer review and

Authors can share their accepted author manuscript:

- immediately

  via their non-commercial person homepage or blog

  bpupdating a preprint in arXiv or RePEc with the accepted manuscript

  via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group

  dioectly by providing copies to their students or to research collaborators for their personal use

  foo private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement

- after the embargo period

  via non-commercial hosting platforms such as their institutional repository
  via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. For book authors the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.
19. Thesis/Dissertation: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article

be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/bv-nc-sa/4.0.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions:**

v1.8

Questions? customercareAcopyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Apr 19, 2016

This is an Agreement between Patrick Budd ("You") and Elsevier ("Elsevier"). It consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Patrick Budd
Customer address	Centre for Biomolecular Sciences Nottingham, None NG7 2RD
License number	3803110631816
License date	Feb 06, 2016
Licensed content publisher	Elsevier
Licensed content publication	Structure
Licensed content title	Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy
Licensed content author	Sian Rowsell,Richard A Pauptit,Alec D Tucker,Roger G Melton,David M Blow,Peter Brick
Licensed content date	15 March 1997
Licensed content volume number	5
Licensed content issue number	3
Number of pages	11
Start Page	337
End Page	347
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 3

Total **0.00 GBP**

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

15. Translation: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. Posting licensed content on any Website: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. For journal authors: the following clauses are applicable in addition to the above:
Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author-managed copy of the manuscript of an article that has been accepted for publication which may include any author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - bpupdating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. For book authors the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. Thesis/Dissertation: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article

be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

CC BY NC ND: The CC BY-NC-ND license alloWs users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

20. **Other Conditions:**

v1.8

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Apr 19, 2016

This is an Agreement between Patrick Budd ("You") and Nature Publishing Group ("Nature Publishing Group"). It consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3804281351259
License date	Feb 08, 2016
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature Reviews Microbiology
Licensed Content Title	Clostridia in cancer therapy
Licensed Content Author	Nigel P. Minton
Licensed Content Date	Dec 1, 2003
Volume number	1
Issue number	3
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
High-res required	no
Figures	Figure 1 Clostridial Directed Enzyme-Prodrug Therapy (CDEPT) Box **1** Carboxypeptidase G2 chemical breakdown diagram
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Implementation and optimisation of alternative therapeutics for use in Clostridium sporogenes as a delivery vehicle
Expected completion date	May 2016
Estimated size (number of pages)	220
Total	**0.00 GBP**

Terms and Conditions

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use,subject to the conditions below:

paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   For AOP papers, the credit line should read:

   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   Note: For republication from the *British Journal of Cancer*, the following credit lines apply. Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the credit line should read: Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   Note: For adaptation from the *British Journal of Cancer*, the following credit line applies.

   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information.Translations of up to a 400 words do not require **NPG** approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the *British Journal of Cancer*, the following credit line applies.

   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1

Questions? customercareRcopyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.