Advancing the Clinical Implementation of Clostridial Cancer Therapies

Maria Zygouropoulou, MPharm

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

September 2018

Abstract

Although tumour hypoxia and necrosis are a nemesis to the efficacy of conventional cancer therapies, they can be exploited by anaerobic bacteria for the selective targeting of solid tumours. *Clostridium sporogenes* lends itself well to this purpose, being a non-pathogenic, spore-forming and proteolytic obligate anaerobe. Intravenously injected spores of this organism profusely infiltrate, selectively germinate and naturally thrive in the hypoxic and necrotic regions of tumours, whilst non-germinated spores are eliminated. Capitalising on its inherent tumour-targeting capabilities, *C. sporogenes* can be genetically engineered to enhance its oncolytic properties as well as to incorporate safety features, cooperatively accelerating its translation towards clinical applications.

In the present thesis, *C. sporogenes* NCIMB 10696 was engineered to express nitroreductases which can co-metabolise a clinical-stage, mustard-based prodrug and 2-nitroimidazole PET probe analogues, thus enabling cancer cell killing as well as non-invasive imaging of the clostridial vector. Importantly, a highly expressed nitroreductase candidate was identified and its crystal structure was determined, facilitating the design of optimised prodrug derivatives. The strain was additionally engineered to secrete biologically-active Granulocyte Macrophage-Colony Stimulating Factor in high titres, expanding hitherto explored functionalities and providing proof-of-concept for clostridial-mediated cancer immunotherapy. Furthermore, empowered by the CRISPR/Cas9 technology, a multicopy genome integration strategy was implemented, generating significant improvements in protein yield, beyond those afforded by transcriptional and translational fine-tuning. Novel genomic loci for insertions of therapeutic cargos were characterised, revealing another layer of gene expression control that may be advantageously incorporated in the design of clostridial vectors.

Safety considerations permeated all genetic manipulations undertaken, culminating in the elimination of the biosynthetic gene cluster of the haemolytic clostridiolysin toxin, to render the strain safer and regulatorily compliant. Highlighting the versatility of the vector, intelligibly, safety and efficacy deliverables were closely interlinked (for instance, imaging via the therapeutic functionality, cargo insertion with concomitant aberration of key genes). Overall, the advancements generated here constitute significant progress in the development of a clinically relevant end-product.

Acknowledgements

I am grateful to my supervisor, Nigel Minton, for his trust, his intelligent suggestions and for giving me the opportunity to carry out my PhD in exquisite facilities. Also, to my second supervisor, Alan Cockayne, for being able to address with competence and flair every single one of my peculiar problems, and whose door was always open. I am grateful to Ola Kubiak-Bailey, for introducing me to *C. sporogenes* during my rotation and for the sincere support she offered me throughout. Big part of this work was done thanks to our international collaborators in the Netherlands and New Zealand to who I owe many thanks; particularly to Alex Mowday and Dave Ackerley, whose expertise and character I admire. I am indebted to Ingrid Dreveny who taught me the art of crystallisation and to Karen Robinson and Kazuyo Kaneko for helping me turn Chapter 5 into a more meaningful piece of science; especially to Kazuyo who wholeheartedly accepted me as her protégé in cell culture and immunology. I am extremely thankful to Evangelos Karvonis, my favourite lab technician, who has always accommodated to our insane demands of media (το ευχαριστώ είναι λίγο). I also want to thank my office friends in C02 and my lab friends in B39 and B27 for making the long hours at the University a lot more enjoyable with their (out of hours) presence, their humour and creativity.

To the best colleagues anyone could ask for: Ines, Daphne and Raquel I owe my most special thanks; for making this journey fun, stimulating and worthwhile and whose friendship is after all the greatest outcome of these last four years. I am also grateful to Ines and Daphne for occasionally lending me their brilliant minds and for making valuable contributions to my research. I've been extremely lucky to have the two of you by my side, and although I would change many things if I ever had to do this again, this would stay the same in every possible alternative.

I am forever thankful to my family who is always my reference point, for believing in me unconditionally and for offering me support and understanding when it was most needed. I also need to thank my Albert infinite times for his unfailing patience, love and care and for embracing all versions of myself during these intense years.

Finally, I am very grateful to the BBSRC for funding my research and making all this possible.

To my grandma

Table of Contents

Abstract ..... **i**

Acknowledgements ..... **ii**

Table of Contents ..... **iv**

List of Tables ..... **x**

List of Figures ..... **xi**

List of Abbreviations ..... **xiii**

CHAPTER 1: Introduction **1**

1.1 Cancer 1

1.2 Tumour hypoxia and necrosis: a blessing in disguise? 3

1.3 Clostridia in cancer therapy 5

1.4 Embodiment of clostridial cancer therapies 8

1.4.1 Administration route and form 8

1.4.2 Tumour targeting and colonisation 9

1.4.3 Mechanism of action 10

1.4.4 Termination of treatment 11

1.5 Limitations of clostridial oncolysis 11

1.6 Genetic engineering to improve clostridial oncolytic properties 12

1.6.1 Genetic tools and strain selection 13

1.6.2 Genetic engineering efforts to date 14

1.7 Clostridia in the current landscape of cancer therapy 15

1.8 Objective and thesis organisation 17

CHAPTER 2: Materials and Methods **19**

2.1 Microbiology 19

2.1.1 Bacterial strains 19

2.1.2 Growth media, buffers and supplements 19

2.1.3 Culture conditions 21

2.1.4 Storage conditions 22

2.1.5 Starter cultures 22

2.1.6 Quantification of bacterial growth 22

2.1.7 Preparation of electrocompetent *E. coli* 22

2.1.8 Transformation of *E. coli* by electroporation 23

2.1.9 Transformation of *C. sporogenes* by conjugation 23

2.1.10 Preparation of cell lysates 23

2.1.11 Preparation of cell-free culture supernatants 24
2.1.12 Growth curves 24
2.1.13 Sporulation and germination assay 24
2.1.14 Qualitative assessment of haemolytic activity 24
2.1.15 Preparation of red blood cell stock solution 25
2.1.16 Quantitative assessment of haemolytic activity in co-cultures 25
2.1.17 Quantitative assessment of haemolytic activity in cell-free supernatants 26
2.1.18 Hemin binding assay 26
2.2 Molecular biology 27
2.2.1 Plasmid DNA extraction 27
2.2.2 Genomic DNA extraction 27
2.2.3 Polymerase Chain Reaction (PCR) 28
2.2.4 Agarose gel electrophoresis 29
2.2.5 DNA purification 29
2.2.6 Quantification of DNA 29
2.2.7 Cloning techniques 29
2.2.8 Sanger DNA sequencing 30
2.2.9 Illumina genome sequencing 30
2.2.10 DNA synthesis 30
2.2.11 Plasmid construction 30
2.2.12 Targeted genomic mutagenesis of *C. sporogenes* 36
2.2.13 Plasmid loss 37
2.3 Biochemistry 38
2.3.1 Determination of total protein concentration 38
2.3.2 Precipitation of protein from cell-free culture supernatants 38
2.3.3 Inclusion body purification 39
2.3.4 Sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE) 39
2.3.5 Gel staining and de-staining 40
2.3.6 Western blotting 40
2.3.7 Menadione reductase assay 40
2.3.8 Nitrofurazone reductase assay 41
2.3.9 mGM-CSF Enzyme-Linked Immunosorbent Assay (ELISA) 42
2.4 Crystallography 42
2.4.1 Protein expression and isolation 42
2.4.2 Protein purification 43
2.4.3 Concentration of protein by centrifugation 43

2.4.4 Quantification of protein 44
2.4.5 Crystallisation 44
2.4.6 Crystal storage 44
2.4.7 Molecular visualisation 45
2.5 Eukaryotic cell culture 45
2.5.1 Isolation of murine bone marrow cells 45
2.5.2 Cell culture 45
2.5.3 Preparation of treatment reagents 46
2.5.4 Cell treatment 46
2.5.5 MTT cell viability assay 46
2.5.6 Flow cytometry 47
2.6 Bioinformatic tools 47
2.6.1 Codon optimisation 47
2.6.2 Sequence data handling, plasmid map construction and visualisation 48
2.6.3 Primer design 48
2.6.4 Illumina sequencing data analysis 48
2.6.5 Sequence alignments and database searches 48
2.6.6 Phylogenetic tree construction 48
2.6.7 Promoter prediction 49
2.6.8 Terminator prediction 49
2.6.9 Origin of replication prediction 49
2.6.10 Signal peptide prediction 49
2.6.11 Protease cleavage site prediction 49
2.6.12 Prediction of transmembrane helices in proteins 50
2.6.13 Identification of conserved domains in proteins 50
2.6.14 Data analysis and presentation 50
CHAPTER 3: Clostridial-Directed Enzyme Prodrug Therapy: investigation of suitable nitroreductase candidates 51
3.1 Background 51
3.1.1 Prodrugs as anticancer agents 51
3.1.2 Clostridial-Directed Enzyme Prodrug Therapy (CDEPT) 54
3.1.3 Nitroreductases as the prodrug converting enzymes 56
3.1.4 Nitroreductase-activated prodrugs 58
3.1.5 Imaging of clostridial vectors 59
3.1.6 An international CDEPT collaboration 61
3.2 Objectives 65
3.3 Hypothesis 65
3.4 Study design 66

3.5 Results 67
3.5.1 Construction of nitroreductase-expressing strains 67
3.5.2 Phenotypic characterisation of nitroreductase-expressing strains 70
3.5.3 *In vitro* characterisation of nitroreductase expression 72
3.5.4 Troubleshooting the poorly expressed NfrA_Bs candidate 77
3.5.5 Crystallisation of the highly expressed YfkO_Bl candidate 80
3.6 Discussion 83
3.7 Key Outcomes 87
CHAPTER 4: New strategies for improved gene expression 88
4.1 Background 88
4.1.1 The requirement for overexpression 88
4.1.2 Strategy I: Multicopy integration 90
4.1.3 Strategy II: Implementation of the TcdR-P_{tcdB} expression system 94
4.2 Objectives 97
4.3 Hypothesis 97
4.4 Study design 97
4.4.1 Strategy I 98
4.4.2 Strategy II 98
4.5 Results 100
4.5.1 Identification of genomic loci 100
4.5.2 Characterisation of genomic loci 102
4.5.3 Sequential amplification of gene copy number 108
4.5.4 Preliminary investigation of the T*cdR*-P_{tcdB} system 110
4.5.5 Coupling the TcdR-P_{tcdB} system to a constitutive promoter 113
4.5.6 Mimicking the split configuration on the chromosome 116
4.5.7 Strategy I *vs* Strategy II 121
4.6 Discussion 123
4.7 Key Outcomes 129
CHAPTER 5: Piloting a novel functionality: immunotherapy 130
5.1 Background 130
5.1.1 The immune system in cancer: friend or foe? 130
5.1.2 Cancer immunotherapy using bacteria 132
5.1.3 Granulocyte Macrophage-Colony Stimulating Factor 135
5.1.4 Interleukin 4 136
5.2 Objectives 137
5.3 Hypothesis 138

5.4 Study design 138
5.4.1 Genes 138
5.4.2 Selection of expression signal 138
5.4.3 Selection of secretion signal 140
5.5 Results 142
5.5.1 *C. sporogenes* as an expression host for mGM-CSF and mIL-4 142
5.5.2 Immunoblot detection of mGM-CSF and mIL-4 146
5.5.3 Quantification of secreted mGM-CSF 148
5.5.4 The effect of extracellular proteolysis on protein yield 150
5.5.5 Biological activity of mGM-CSF produced in *C. sporogenes* 153
5.6 Discussion 158
5.7 Key outcomes 163
CHAPTER 6: Towards a safer, non-haemolytic clostridial host 164
6.1 Background 164
6.1.1 Regulatory and safety considerations 164
6.1.2 Safety-enhancing strategies 165
6.1.3 Safety considerations with *C. sporogenes* NCIMB 10696 167
6.1.4 Streptolysin S/clostridiolysin S 168
6.2 Objectives 171
6.3 Hypothesis 172
6.4 Study design 172
6.3.1 Special considerations for the quantitative assessment of SLS/CLS haemolytic activity 173
6.4 Results 174
6.4.1 Characterisation of the haemolytic activity of *C. sporogenes* NCIMB 10696 174
6.4.2 Deletion of the clostridiolysin S gene cluster 176
6.4.3 Deletion of putative haemolysins 180
6.4.4 Inactivation of *closA* and *closC* using ClosTron mutagenesis 186
6.4.5 Whole genome sequencing of Csp-*closC*749a::CT-NH 189
6.4.6 Investigating the role of the sensor histidine kinase *resE4* 191
6.4.7 Haemolytic assays using media containing red blood cells *vs* whole blood 192
6.4.8 Could mechanisms unrelated to haemolysins contribute to a haemolytic phenotype? 195
6.5 Discussion 197
6.6 Key outcomes 201

CHAPTER 7: General discussion 202
7.1 Summary of findings and future work 202
7.1.1 CDEPT: exploration of suitable nitroreducatase candidates 202
7.1.2 New strategies for improved gene expression 204
7.1.3 Piloting a novel functionality: immunotherapy 205
7.1.4 Towards a safer, non-haemolytic clostridial host 207
7.2 Imminent challenges 208
7.2.1 Rigorous preclinical evaluation 208
7.2.2 Quality control 209
7.2.3 Biocontainment 210
7.3 Future perspectives 211
7.3.1 Towards a multifunctional clostridial anticancer vector 211
7.3.2 Expanding the functionality repertoire 213
7.4 Concluding remarks 215
Bibliography 216
Appendix 252
List of Appendix Tables 253
List of Appendix Figures 253
I. Bacterial strains 254
II. Primers 258
III. Plasmids 268
IV. Representative plasmid maps 272
V. Codon Optimisation 276
VI. Codon-optimised nucleotide sequences 287
VII. P_{thl14} promoter/RBS variant for gene expression in *C. sporogenes* 290
VIII. Protease cleavage site prediction in the mGM-CSF protein sequence 294
Appendix Bibliography 299

List of Tables

Table 1.1 Historical milestones of anticancer therapy with Clostridia 6
Table 2.1 Growth media and buffers used in this study 20
Table 2.2 List of supplements used in this study 21
Table 3.1 The selected nitroreductase candidates 64
Table 3.2 Characterisation of nitroreductase candidates towards key substrates 64
Table 3.3 Rank order and cofactor preference of nitroreductase candidates in the enzymatic activity screens 75
Table 4.1 Genomic loci selected for integration 102
Table 4.2 Strains generated for the characterisation of genomic loci 103
Table 4.3 Strains generated for the sequential amplification of gene copy number 108
Table 4.4 Strains generated for the preliminary investigation of the TcdR-P_{tcdB} system 110
Table 4.5 Strains generated for the investigation of the TcdR-P_{tcdb} system under the control of a constitutive promoter 115
Table 4.6 Strains generated for the chromosomal implementation of the TcdR-P_{tcdb} system 118
Table 5.1 The selected signal peptides 141
Table 5.2 Generated strains harbouring mGM-CSF- and mIL-4-expressing plasmids 142
Table 5.3 Predicted cleavage sites of bacterial proteases in the mGM-CSF protein sequence 152
Table 6.1 Putative haemolysins in *C. sporogenes* NCIMB 10696 180
Table 6.2 Strains generated to study the contribution of putative haemolysin genes 182
Table 6.3 Summary of haemolytic activity results 197

List of Figures

Figure 1.1 The principle of clostridial-mediated tumour targeting. 9
Figure 3.1 Clostridial-Directed Enzyme Prodrug Therapy (CDEPT). 55
Figure 3.2 Mechanism of CB1954 activation. 57
Figure 3.3 Mechanism of CP-506 activation. 61
Figure 3.4 The lead PET probe candidates 62
Figure 3.5 Phylogenetic tree of the selected nitroreductase candidates. 65
Figure 3.6 Schematic illustration of allele-coupled exchange (ACE). 69
Figure 3.7 Genomic integration of nitroreductase candidates. 70
Figure 3.8 Growth curves of nitroreductase-expressing strains. 71
Figure 3.9 Sporulation/germination capacity of nitroreductase-expressing strains... 72
Figure 3.10 SDS-PAGE analysis on cell lysates of nitroreductase-expressing strains. 73
Figure 3.11 Enzymatic activity screens of the nitroreductase candidates. 76
Figure 3.12 Genomic integration of NfrA_Bs construct variants..... 78
Figure 3.13 Troubleshooting the low expression levels of NfrA_Bs. 79
Figure 3.14 Purification and crystallisation of YfkO_Bl. 81
Figure 3.15 The crystal structure of YfkO_Bl. 82
Figure 4.1 The *C. difficile* pathogenicity locus. 95
Figure 4.2 The expression cassette for multicopy integration 98
Figure 4.3 The configurations of the P_{lac}-*tcdR*-P_{tcdB}-*yfko_Bl* expression system. 100
Figure 4.4 Genomic integration of the expression cassette in the selected loci..... 104
Figure 4.5 Growth curves of strains with a single copy of the expression cassette in different genomic loci. 105
Figure 4.6 Expression strength of selected genomic loci. 106
Figure 4.7 Expression strength of selected genomic loci relative to their chromosomal location. 107
Figure 4.8 Characterisation of multicopy strains. 109
Figure 4.9 Construction of strains with three configurations of the TcdR-P_{tcdB} system. 111
Figure 4.10 Preliminary investigation of expression levels using the TcdR-P_{tcdB} system. 113
Figure 4.11 Expression of YfkO_Bl using promoters from the in-house promoter library. 114
Figure 4.12 The TcdR-P_{tcdB} system under the control of a constitutive promoter. 116

Figure 4.13 Construction of the chromosomal P_{ptb}-*tcdR* multicopy P_{tcdB}-*yfkO_Bl* strain. 119
Figure 4.14 Investigation of expression of YfkO_Bl using the chromosomal mimic of the split configuration system. 120
Figure 4.15 Direct comparison of the two overexpression strategies. 122
Figure 5.1 Construction of mGM-CSF- and mIL-4- expressing strains. 143
Figure 5.2 Growth curves of mGM-CSF expressing strains. 144
Figure 5.3 Growth curves of mIL-4 expressing strains. 145
Figure 5.4 Acquired mutations in the mIL-4 constructs. 146
Figure 5.5 Western blot analysis of mGM-CSF-expressing strains. 147
Figure 5.6 Quantification of mGM-CSF in culture supernatants of mGM-CSF-expressing strains. 150
Figure 5.7 The effect of extracellular proteolysis on the yield of secreted mGM-CSF. 153
Figure 5.8 MTT cell proliferation assay on mGM-CSF-treated murine bone marrow cells. 155
Figure 5.9 Phenotypic analysis of mGM-CSF-treated murine bone marrow cells. 157
Figure 6.1 The streptolysin S/clostridiolysin S gene clusters 170
Figure 6.2 Characterisation of the haemolytic activity of *C. sporogenes* NCIMB 10696 in solid and liquid media 175
Figure 6.3 Deletion of the clostridiolysin S operon in *C. sporogenes* NCIMB 10696. 177
Figure 6.4 Phenotypic characterisation of Csp-Δ*closA-I*. 178
Figure 6.5 Haemolytic activity of Csp-Δ*closA-I*. 179
Figure 6.6 Sequential deletion of putative haemolysins in *C. sporogenes* NCIMB 10696. 183
Figure 6.7 Haemolytic activity of Csp-Δ1-5 and Csp-Δ*closA-I* Δ1-5. 185
Figure 6.8 ClosTron-mediated inactivation of *closC* and assessment of the strain's haemolytic phenotype on blood agar. 188
Figure 6.9 Schematic diagram showing the group II intron insertion in *closC*. 189
Figure 6.10 Mutations identified in Csp-*closC*749a::CT-NH. 190
Figure 6.11 Haemolytic activity of RW80. 192
Figure 6.12 Comparison of haemolysis in red blood cell-only and whole blood media. 194
Figure 6.13 Haemolytic activity of *C. acetobutylicum* ATCC 824 196
Figure 7.1 A multifunctional clostridial vector engineered for cancer therapy. 214

List of Abbreviations

2HX	2-hydroxylamine
2-NI	2-nitroimidazole
4HX	4-hydroxylamine
5-FOA	5-fluoroorotic acid
5-NI	5-nitroimidazole
ACE	allele - coupled exchange
ADEPT	antibody-directed enzyme prodrug therapy
ATCC	American Type Culture Collection
BM	bone marrow
BMDC	bone marrow-derived dendritic cell
BSA	bovine serum albumin
Cas	CRISPR-associated
CDEPT	clostridial-directed enzyme prodrug therapy
CFU	colony forming unit
CGM	clostridial growth medium
CLS	clostridiolysin S
Cm	chloramphenicol
CRISPR	clustered regularly interspaced palindromic repeats
DC	dendritic cell
dH_2O	distilled water
DMSO	dimethyl sulfoxide
DNBM	dinitrobenzamide mustard
DOC	sodium deoxycholate
DTT	dithiothreitol
EC	expression cassette
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
FBS	foetal bovine serum
FDA	Food & Drug Administration
FMN	flavin mononucleotide
GDEPT	gene-directed enzyme prodrug therapy

GGH	gamma-L-glutamyl hydrazide
GM-CSF	granulocyte macrophage-colony stimulating factor
GMO	genetically modified organism
HAP	hypoxia-activated prodrug
IL-4	interleukin 4
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kan	kanamycin
LB	Luria Bertani
LHA	left homology arm
LPS	lipopolysaccharide
MCS	multiple cloning site
MDSC	myeloid-derived suppressor cells
MTT	3-(4,5-dimethylthiazol- 2-yl)-2,5-diphenyltetrazolium bromide
NADH	nicotinamide adenine dinucleotide, reduced
NADPH	nicotinamide adenine dinucleotide phosphate, reduced
NCIMB	National Collection of Industrial, Marine and Food Bacteria
NEB	New England Biolabs
NIH	National Institutes of Health
NK	natural killer
NTR	nitroreductase
OD	optical density
OS	orthogonal system
PAM	protospacer adjacent motif
PBA	protein blocking agent
PBS	phosphate-buffered saline
PBS-T	phosphate-buffered saline-tween
PCE	prodrug-converting enzyme
PCR	polymerase chain reaction
PET	positron emission tomography
RBC(s)	red blood cell(s)
RBS	ribosome binding site
RHA	right homology arm
RNAP	RNA polymerase
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis

sgRNA	single-guide RNA
SLS	streptolysin S
SN	supernatant
T_H2	type 2 T helper cell
TAE	Tris-acetate EDTA
TAM	tumour associated macrophage
TCA	trichloroacetic acid
TLR	toll-like receptor
Tm	thiamphenicol
Treg	regulatory T cells
TYG	tryptone-yeast extract-sodium thioglycolate
UM	University of Maastricht
UOA	University of Auckland
VDEPT	virus-directed enzyme prodrug therapy
VEGF	vascular endothelial growth factor
VUW	Victoria University of Wellington
WB	whole blood
WT	wildtype

CHAPTER 1

Introduction

1.1 Cancer

The task of defining cancer is challenging and bears the risk of oversimplifying a complex and yet not fully deciphered disease. Although cancer is an umbrella term used to group more than 200 distinct conditions, the underpinning physiology of all cancers involves the transformation of a single normal cell into a malignant one. The cell could be of any type and anywhere in the human body, but depending on the site and tissue of origin, each cancer is accordingly named and histologically classified[1]. In a nutshell, once a cell has been transformed, its subsequent cell divisions generate new malignant cells in an uncontrollable fashion. The accumulation of malignant cells leads to the formation of malignant tumours, which can invade surrounding normal tissue as well as distant parts of the body, disrupting organ architecture and function[2].

The process of transformation is multistep, requiring accumulation of abnormalities in regulatory pathways which govern normal cell proliferation and homeostasis. Cancer is a genomic disease; this acquisition of defects is enabled by an underlying genetic

instability and consequential genetic and/or epigenetic changes in the cell's genome and/or coded mRNAs and proteins[3]. Frequently, these changes manifest as spontaneous mutations in DNA repair genes, oncogenes or tumour suppressor genes as well as functional alterations of their respective products[3].

The genetic instability underpinning all cancers may be the result of factors that are unfortunately unmodifiable, such as inherited mutations in genes associated with increased susceptibility to cancer or advancing age[4]. Chronic infections leading to chronic inflammation or chronic inflammation *per se*, including hereditary inflammatory conditions, may also favour a malignant transformation. Inflammation in particular is becoming an important player in tumour development and it is currently accepted as one of the two tumour-enabling characteristics alongside genetic instability[5]. Environmental, lifestyle and behavioural factors also increase the risk of developing cancer, most prominent amongst them being tobacco smoking, lack of physical exercise, obesity, alcohol intake, sun exposure as well as contact with environmental carcinogens, be it through the atmosphere or the diet[6]. On the basis of current knowledge on external risk factors, it is estimated that up to a half of all cancers could be prevented[7]. Nevertheless, the specific causes/events leading to the development of many cases of cancer still remain unknown.

Although gene-expression profiling of cancers and large-scale systematic sequencing endeavours, such as The Cancer Genome Atlas (TCGA), are aiding the elucidation of the changes that occur in malignant cells[8-10], knowledge of the *precise* 'mix' of these changes is not always attainable. Yet, it is certain that it varies among different cancers, individuals, stages of the disease and even between cells of the same tumour. However, independently of the composition of this medley of changes, their impact is universally reflected on certain aspects of cell behaviour, aptly termed the 'hallmarks of cancer'. Originally, these included self-sufficiency in growth signalling, insensitivity to anti-growth signals, limitless replicative potential, evasion of apoptosis, sustained angiogenesis coupled with tissue invasion, and finally, metastasis, defined as the

development of secondary malignant growths at a distance from a primary site of cancer. In view of more recent findings, another two hallmarks have emerged: the reprogramming of energy metabolism and the evasion of immune destruction[2, 5]. As transformed cells acquire a succession of these eight characteristics they develop selective advantages which allow them to mutate and propagate at an increased rate, outgrow normal cells, and survive despite varying environmental stresses.

Of equal importance to the acquired characteristics of malignant cells that allow tumour formation, is the microenvironment in which they grow. The acquisition of a 'fertile' ground involves the heterotypic interactions of incipient tumour cells with supposedly-normal cells in their proximity, including cells of the recruited vasculature (endothelial), cells of the extracellular matrix (stroma) as well as cells of the innate immune system. Not only do these neighbouring cells provide structural support to the tumour but they also source vital nutrients, necessary for maintaining growth. In essence, normal cells are being deployed in the tumorigenic process by creating a thriving tumour microenvironment and thus helping in the acquisition of the cancer hallmark traits[5, 11].

1.2 Tumour hypoxia and necrosis: a blessing in disguise?

Like healthy tissues, cancer cells need to be supported by a vascular network to fulfil their nutritional and metabolic needs[5]. Tumour cells build and maintain this network via secreted mediators which target microvascular endothelial cells and stimulate vascularisation (a process known as angiogenesis); the predominant angiogenic mediator underpinning this process is the vascular endothelial growth factor (VEGF)[12]. Nevertheless, the resultant tumour-associated neovasculature is highly abnormal, chaotic, tortuous and leaky, with large intercapillary distances and heterogeneous distribution[13]. These defects are further exacerbated by: elevated interstitial fluid pressures which compress the tumour blood vessels and create a barrier to transcapillary

transport; the rapid expansion of the tumour which may obstruct its own vascular supply; and arteriovenous shunts[6]. As a result, blood supply becomes non-uniform, leading to heterogeneously distributed zones of metabolic insufficiency, hypoxia, quiescence and necrosis[13-15]. In fact, these zones constitute a typical histological feature of solid tumours[15]; indicatively, approximately 50-60% of advanced solid tumours may contain up to 75% of hypoxic or anoxic regions depending on their size[16], with oxygen concentrations below 1%, as opposed to 3-15% in normal tissues[16, 17].

These structural and functional defects create a number of protective shields for cancer, justifying the recognition of tumour hypoxia and necrosis as negative prognosis markers, associated with aggressive and metastatic forms of cancer[14, 18]. Crucially, these characteristics undermine the efficacy of conventional cancer treatments, including radiotherapy and chemotherapy. On the one hand, radiotherapy requires a well oxygenated environment for efficacy since oxygen participates directly in radiotherapy-induced DNA damage[19]. On the other hand, as a direct consequence of the aberrant architecture of the vasculature, diffusion of chemotherapeutic agents from the bloodstream into the site of the tumour is impeded, causing uneven drug distribution and regions of sub-therapeutic exposure, which eventually lead to drug-resistance[16, 20]. Furthermore, the presence of hypoxia exacerbates genetic instability and mutability and modulates a number of resistance pathways, including loss of apoptotic potential[21]. Worse yet, cell quiescence leads to attenuation of cellular functions and consequently to the suppression of the drug targets of many chemotherapeutic agents. As a result, quiescent cells escape treatment whilst retaining their ability to regrow if oxygen and nutrient supply is restored[22].

Although the presence of tumour hypoxia, quiescence and necrosis hampers traditional therapies, it nonetheless creates a unique niche within the human body[23]. This offers great opportunities for anaerobic *Clostridium* strains which can exploit these conditions to selectively target solid tumours, thus turning a cancer trait associated with treatment failure into a precision therapy[24, 25]. Necrosis in particular, is not currently being

addressed by any other clinically approved treatment modality, highlighting the opportunity afforded by clostridial strains[26].

1.3 Clostridia in cancer therapy

The genus *Clostridium* comprises a heterogeneous group of Gram-positive, obligate anaerobic and spore-forming bacteria naturally found in soil and aquatic sediments as well as the intestinal tract of animals and humans. The genus has gained considerable infamy owing to a few pathogenic species and their associated diseases, such as *Clostridium perfringens* (gas gangrene), *Clostridium difficile* (pseudomembranous colitis), *Clostridium tetani* (tetanus) and *Clostridium botulinum* (botulism). Nevertheless, it is at the same time the most intensely-studied group of bacteria for application in cancer therapy, with references dating back to the beginnings of the 19th century. Different species have been investigated throughout the years in pursuit of optimum efficacy and safety[26-28]. Table 1.1 presents a historical timeline of important milestones regarding the use of clostridial species in cancer therapy.

Table 1.1 Historical milestones of anticancer therapy with Clostridia

1813	Tumour regression in cancer patients suffering from gas gangrene (caused by *Clostridium perfringens*) [Vautier] cited in [28].	**pathogenic strains**
1927	Intratumoural injection of proteolytic ferments from *Clostridium sporogenes* and *Clostridium histolyticum* results in lysis of Flexner-Jobling rat carcinonas [Torrey and Kahn][29].	
1935	Intratumoural injection of sterile filtrates of *Clostridium histolyticum* results in tumour regression of advanced cancers, **attributed to proteolytic enzymes** [Connel][30].	
1947	Intratumoural injection of *Clostridium histolyticum* spores into transplanted mouse sarcomas results in local **spore germination** and noticeable tumour lysis ('liquefication'). In no instance is the tumour completely destroyed [Parker][31].	
1955	Tumour bearing mice die within 48 hours after intravenous injection with *Clostridium tetani* spores. Death is attributed to production of tetanus toxin by germinated spores in colonised tumours. Healthy controls survive, demonstrating the **exquisite tumour selectivity** of spore germination. **Systemic delivery** is an effective route of administration [Malmgren and Flanigan][32].	
1964	The first study employing a **non-pathogenic strain**. Intravenous injection of *Clostridium sporogenes* M55 spores in mice with solid Erhlich tumours causes tumour lysis ('softening of tumours') without affecting normal tissues. **Tumours regrow from the outer rim of remaining viable cells**. Results are confirmed with different rodent tumour models [Möse and Möse][33].	**non-pathogenic strains**
1967	Möse and Möse inject themselves with spore suspensions of *Clostridium sporogenes* M55, demonstrating the **safety of this strain** [Möse and Möse][34].	
1967	Human trial: 5 patients with advanced cancer receive intravenous injection of *Clostridium sporogenes M55* spores. Oncolysis is observed only in 3 patients and only in **large tumours**. Treatment is **well tolerated**. Surrounding tissues, smaller tumours and metastases are **unaffected by the therapy** [Carey][34].	
1978	Human trial: intracarotid injection of *Clostridium sporogenes M55* spores in patients with inoperable vascular glioblastomas. Glioblastomas of some patients liquefy into brain abscesses which are later operated. Treatment is well tolerated. However, **the rate of recurrence remains uninfluenced** [Heppner and Möse][35].	
1978-1979	Debut of clostridial spore administration **in combination** with irradiation and strategies that intensify tumour hypoxia [Dietzel] cited in [36], [Gericke][37], [Möse][38].	

1988	Debut of **genetic manipulation** of Clostridia for enhancement of their oncolytic properties [Schlechte and Elbe][39].	Non-pathogenic strains
1996	Birth of **Clostridial-Directed Enzyme Prodrug Therapy (CDEPT)**. *Clostridium beijerinckii* NCIMB 805 is engineered to express cytosine deaminase. Murine EMT6 carcinoma cells exhibit up to 500-fold increased sensitivity towards 5-fluorocytosine following treatment with culture supernatants of recombinant strains *in vitro*[Fox *et al.*][40].	
2002	Proof-of-concept of **CDEPT exemplified *in vivo***. Recombinant *Clostridium sporogenes* NCIMB 10696 spores expressing CD are administered intravenously in combination with 5-fluorocytosine, for the treatment of SCCVII tumours in C3H/Km mice. Observed tumour growth delay is greater compared to 5-fluorouracil treatment alone [Liu *et al.*][25] .	
2005	Engineering of Clostridia to express **immunotherapeutic molecules** for the first time. *Clostridium acetobutylicum* DSM792 is engineered to secrete murine Tumour Necrosis Factor (mTNFα) [Theys *et al.*][41]	
2006	**Phase I clinical trial of *Clostridium novyi*-NT spores** in patients with treatment-refractory solid tumour malignancies, given as a single intravenous injection [ClinicalTrials.gov Identifier: NCT00358397].	
2007	Engineering of Clostridia to express **antibodies** for the first time. *Clostridium sporogenes* NCIMB 10696 and *Clostridium novyi*-NT are engineered to secrete an anti-Hypoxia-Inducible Factor 1 alpha (anti-HIF1A) antibody [Groot *et al.*][42].	
2014	*Clostridium sporogenes* NCIMB 10696 is engineered to express a nitroreductase enzyme for use in CDEPT. For the first time, the gene is not carried on a plasmid but **integrated in the chromosome without the use of antibiotic resistance markers**. Treatment with spores of this strain and CB1954 significantly delays growth of human colorectal HCT116 carcinoma in NMRI nu/nu mice. Complete tumour regression is seen in 25% of cases [Heap *et al.*][43].	

The historical clinical findings (Table 1.1) provide proof-of-concept for the use of Clostridia in cancer therapy. More recently, *Clostridium novyi*-NT, a non-toxic (NT) derivative strain of *C. novyi*, has been extensively evaluated in the same context, with outcomes supportive of the predilection of clostridial spores to colonise both animal and human tumours. In particular, it has been shown that intravenous or intratumoural administration of spores can elicit robust tumour lysis in experimental tumours of rabbits, rats and mice[16, 23, 44] as well as in spontaneous tumours of companion dogs[45, 46]. Generally, side-effects occurring following tumour colonisation and lysis can be easily

managed, resembling the common symptoms of bacterial infections, with infrequent occurrence of dose-limiting toxicities in cases of large tumour burdens[44, 46]. Thanks to the collection of encouraging pre-clinical data, *C. novyi*-NT has progressed to Phase I clinical trials in human patients with advanced solid malignancies refractory to standard treatments (NCT00358397, NCT01118819, NCT01924689). The strain has proven effective in the destruction of tumour masses via intravenous or intratumoural injection while displaying an acceptable safety profile, causing only a few common toxicities such as fever and tumour inflammation[47].

Importantly, the spore-forming ability of clostridia has been highlighted as the cornerstone of their safe application in cancer therapy. Clostridial spores cannot germinate in healthy organs and are rapidly cleared from the circulation. They are also less likely to be shed from patients[16]. Additionally, their demonstrated lack of immunogenicity should, if applicable, permit repeated administrations[20, 25, 48, 49]. Remarkably, spore germination occurs uniquely in hypoxic *tumour* tissues, leaving hypoxic non-neoplastic lesions, such as myocardial infarcts, intact[23, 44]. Finally, besides their clinical benefits, spores are considered to have practical advantages in terms of handling, storage, stability and formulation as well as being easier and cheaper to produce[16, 50].

1.4 Embodiment of clostridial cancer therapies

1.4.1 Administration route and form

In a preclinical or clinical setting, a spore dose ranging from 10^6 to 10^{10} colony forming units (CFU) could be administered IV or IT, usually with the former route being favoured[51]. The initial steps of the proposed pathway for tumour entry involve circulating spores entering tumours via the disrupted and leaky vasculature. Their entry may be reinforced by the enhanced permeability and retention (EPR) effect, according

to which particles of certain sizes (typically in the nm-μm range, such as the size of spores) accumulate in tumour tissue as opposed to normal tissue[52].

1.4.2 Tumour targeting and colonisation

Clostridial spores encountering the hypoxic/necrotic areas uniquely found in solid tumours will germinate and grow whilst spores elsewhere in the body will eventually be cleared[53]. Thanks to this natural phenomenon clostridial cancer therapies are inherently selective. A graphical representation of the basis of clostridial tumour targeting is shown in Figure 1.1.

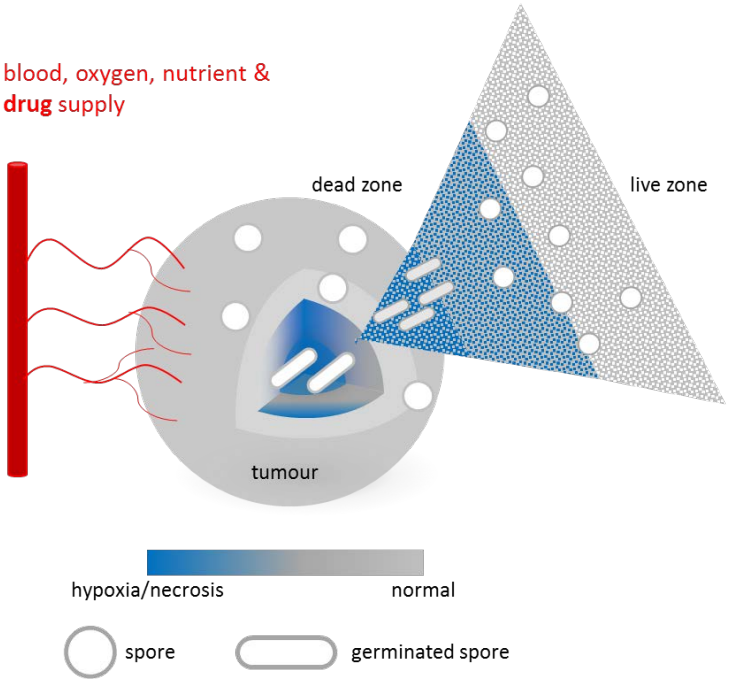


Figure 1.1 The principle of clostridial-mediated tumour targeting. Due to aberrant tumour vasculature, multiple microenvironments are encountered within a tumour. Close to blood vessels, cells are viable and proliferating. Far from blood vessels, hypoxia and nutrient depletion create regions of hypoxia, quiescence or necrosis. The obligate anaerobic nature of clostridial strains implies that clostridial spores can only germinate in these oxygen-deprived regions.

Upon germination, the mobile vegetative clostridial cells can actively migrate and move away from vasculature, penetrating deep into tumour tissue[20], in contrast to drug molecules which are transported passively, dependent on diffusion gradients. Following tumour colonisation, an actively growing bacterial population is established in the tumour site. As a result of their immunoevasive capabilities and the aberrant architecture of their vasculature, tumours are ideal sanctuaries for clostridial replication which to some extent can proceed unhindered by the immune and reticuloendothelial clearance mechanisms. Also, due to increased cell turnover, tumours have an abundance of nutrients which can be utilised by the replicating bacteria[20]. Clostridial growth is only halted by the well-oxygenated outer rim of the tumour; in this manner, potential spread to healthy tissues is prevented (an additional safety feature)[53].

1.4.3 Mechanism of action

Even from the early days of clostridial therapies, it was hypothesized that the inherent oncolytic properties of the species stem from the expression of native cytolytic enzymes upon germination, effectively converting the tumour mass into an infectious ('liquified') lesion[29, 30, 44]. In support of this, it has been shown that proteolytic strains have superior oncolytic properties and are now the preferred strains for cancer therapy[16]. Whole genome sequencing of these strains has confirmed the expression of several secreted proteases, lipases and other degradative enzymes, accounting for their cytolytic properties[27, 54].

Nevertheless, there is currently an increasing body of evidence suggesting that the immune system may also participate in the oncolytic process. Host inflammatory responses may be elicited as a result of the presence of a localised bacterial infection, helping to induce antitumour immunity and subsequent tumour lysis[55]. Moreover, lysis of cancer cells releases new tumour antigens, aiding the eventual recognition of still-surviving cancer cells as 'foreign'[56]. Equally, suppression of tumour growth may also be the result of bacteria competing with tumour cells for available nutrients. In all cases,

it has been shown that bacterial-mediated oncolysis is independent of tumour heterogeneity and genetic background, and is capable of eradicating even the stubborn hypoxic/quiescent cells unlike conventional chemotherapeutics[20].

1.4.4 Termination of treatment

When termination of treatment is desired, vegetative clostridial cells can be easily eliminated with commonly available antibiotics, such as metronidazole[49]. Antibiotic treatment may also constitute a fall-back measure in cases where toxicity has occurred as a result of the therapy, and is a distinctive advantage of clostridial therapies in comparison to conventional chemotherapy[44].

1.5 Limitations of clostridial oncolysis

Although studies to date have supported the propensity of clostridia to colonise and lyse tumours, a number of limitations have transpired, providing an explanation for the failure of preclinical endeavours to cure cancers permanently, despite evident oncolysis[53]. For instance, it has been established that the extent of tumour colonisation is directly proportional to tumour volume and that a threshold size of 3 cm^3 (2-4 g of tumour mass) is required for clostridia to exert their native oncolytic effect[25, 57]. Additionally, clostridial-mediated tumour lysis spares the cancerous cells at the oxygenated periphery of the tumour, presumably because once clostridia have destroyed the hypoxic core of a tumour (and with it their niche), they perish. On the one hand, this might be considered an inherent biosafety control mechanism, as previously noted, but on the other, it allows tumour regrowth to occur from the remaining cancerous cells (the so called 'viable outer rim')[44].

Improvement of the colonisation efficiency and elimination of the outer rim has been the focus of on-going efforts in the development of clostridial cancer therapies. The administration of clostridial spores has been combined with conventional therapies and

various strategies aiming to amplify the hypoxic conditions in the tumour. The expectation was that with the complementation and synergy of different approaches not only would the hypoxic core be lysed but also the viable outer rim ablated. On a pre-clinical level, clostridial spores have been combined with radiation[37, 58], cytotoxic chemotherapeutic drugs[16, 59] (for instance, mitomycin C, vinorelbine and docetaxel), vascular targeting agents (dolastatin D-10[16], combretastatin A-4[24] and vadimezan[53]) as well as reduction of the oxygen levels in the respired air of animals[38] or induction of tumour hyperthermia[60].

For all the above combinations, outcomes were encouraging, with improved tumour regression and even 'cures'[28, 53, 57]. Nevertheless, on some occasions, extreme and even fatal toxicity was observed, attributed to metabolic complications as a result of the rapid breakdown of tumours (the so called tumour lysis syndrome), underscoring the requirement of candidate therapies to strike a balance between effectiveness and safety[16].

1.6 Genetic engineering to improve clostridial oncolytic properties

In this context, genetic engineering of clostridial strains appeared as a reasonable step to further enhance, fine-tune and optimise their oncolytic properties. Genetic engineering would allow the incorporation of different anticancer functionalities in the clostridial host, essentially turning it into a vector for gene therapy. With the strict definition of the term, gene therapy involves the delivery of corrective genetic material *into* target cells (namely tumour cells or 'normal' cells of their microenvironment), with the aim of either modifying/deleting/replacing abnormal gene(s) or introducing new ones of a therapeutic value[61, 62]. In the case of clostridial species (which lack invasive capabilities), uptake *into* the target cells is not realised and the bacteria serve instead as 'Trojan horses', carrying and themselves expressing the therapeutic gene(s) in the tumour region[63]. Genetically engineered clostridial strains are also referred as 'armed'[26].

The delivery and expression of therapeutic genes by the clostridial host in the tumour, offers important advantages which, in a clinical setting, could translate into treatments with larger therapeutic windows; that is, wider margins of administered drug doses to allow effective treatment without causing toxic effects[64]. Importantly, the rapid bacterial replication can amplify the transgene(s) locally, without affecting healthy tissues (lower systemic toxicity) while *in situ* production allows exposure to the therapeutic for longer and at higher concentrations, especially in comparison to the systemic administration of the same therapeutic as a drug molecule (higher effectiveness).

Attractively, in the context of genetic engineering, the use of clostridial hosts can circumvent limitations posed by other gene transfer systems (such as oncolytic viruses), predominantly affording greater genetic flexibility with respect to the size of heterologous DNA that can be incorporated[47]. Furthermore, the clostridial cell represents an ideal environment for gene expression given that any cofactors necessary for protein function would be provided intracellularly, without the need to rely on the unpredictable and heterogeneous tumour microenvironment[65, 66].

1.6.1 Genetic tools and strain selection

Historically, clostridial species have been notoriously difficult to genetically modify; conventional methods had proven ineffective in the genus whilst genus-specific genetic tools were not available. In recent years however, genetic methods have advanced significantly, paving the way for the use of clostridial species in cancer and other fields. These include the formulation of standardised plasmid vector systems[67], procedures for creating directed[68-72] and random mutations[73, 74], methods for bringing about the rapid integration of DNA into the chromosome[75, 76], and most recently CRISPR/Cas9-mediated genome editing[77-79]. The steps and developments needed to bring about desired genetic modifications in any *Clostridium* species have recently been reviewed[80].

Prior to the development of the aforementioned tools, initial genetic engineering efforts were undertaken using those strains for which there were gene transfer protocols

available, including the saccharolytic *Clostridium acetobutylicum* and *Clostridium beijerinckii*[28]. However, although those strains were more amenable to gene transfer, it was quickly established that they had far inferior colonising and therefore oncolytic properties than most proteolytic strains. In fact, cell numbers in experimental tumours colonised by *C. beijerinckii* were two orders of magnitude lower than those in tumours colonised by *Clostridium sporogenes M55* (10^5-10^6 *vs* 1-2 x 10^8 bacteria/g of tumour respectively)[25]. At the same time, in a direct comparison of 26 bifidobacteria, lactobacilli and clostridial strains (including *C. acetobutylicum*), it was found that two proteolytic strains, *C. novyi* and *Clostridium sordelii*, were the only ones able to spread extensively in the poorly vascularized portions of tumours[16].

Thanks to the increasing availability of genetic tools, it is now possible to select those species which are better suited for cancer therapy, based on their colonisation and oncolytic properties. Most prominent among the proteolytic species being investigated for cancer therapy is *C. sporogenes* (previously known as *Clostridium butyricum* M55 or *Clostridium oncolyticum* M55) and *C. novyi*-NT, as previously discussed. The strain of *C. sporogenes* used in current studies (NCIMB 10696 or ATCC 3584) is different to *C. sporogenes* M55 (ATCC 13732) used in earlier experiments, substituted for its greater amenability to DNA transfer and genetic manipulation[81]. In fact, the development of clostridial cancer therapies has been dichotomised between non-engineered and armed strains, with *C. novyi*-NT monopolising the first and *C. sporogenes* the latter.

1.6.2 Genetic engineering efforts to date

To date, clostridial strains have been genetically modified and armed with a variety of anticancer functionalities. Considerable research efforts have been invested into Clostridial-Directed Enzyme Prodrug Therapies (CDEPT), whereby clostridia are equipped with prodrug-converting enzymes (PCEs) which can metabolise non-toxic prodrugs into toxic derivatives[25, 36, 40, 43, 81-86]. Beyond CDEPT, clostridia have also been

engineered to deliver desired antibodies[42] as well as immunotherapeutic messengers[41, 82, 87-89]. Some of these endeavours will be discussed in more detail in subsequent chapters.

1.7 Clostridia in the current landscape of cancer therapy

Although clostridia have been used in cancer therapy since the early 19th century, they now find themselves amongst a growing field of bacterial cancer therapies. Apart from *Clostridium*, species belonging to the *Salmonella*, *Listeria* and *Bifidobacterium* genera represent the most prominently investigated bacterial organisms in this field. Each has distinct characteristics, but they are all unified by their native tumour-homing properties and cytolytic or immunostimulatory capacity[47]. For instance, the facultative anaerobic nature of *Salmonella* may render it useful against all tumour types, including smaller metastases which may not be hypoxic[65]; the potent ability of *Listeria monocytogenes* to invade immune cells makes it an excellent agent for the engagement of the immune system[90]; *Bifidobacterium* strains benefit from an established safety profile and medical use in probiotics, albeit with less powerful colonising properties[16]. In all cases, the genetic plasticity of bacteria is being exploited for the incorporation of additional anticancer functionalities[20]. Notably, recombinant versions of all three of these bacteria are currently in Phase I/II human clinical trials for the treatment of different types of solid tumours[47].

Thanks to their unique properties, bacteria are perceived to hold promise for unmet medical needs of cancer patients, when traditional cancer therapies have not been effective. These include the treatment of refractory metastatic cancers, multidrug-resistant cancers and cancers that evade immune clearance[47]. It is possible that in the future, treatments may consist of the cooperative utilisation of strains with different properties and functionalities, based on the type and stage of cancer, hence justifying the major research efforts with more than one bacterial genus[20]. Importantly, bacterial

therapies would be practical and cost-effective to manufacture, given that the biotechnology industry is sufficiently developed, and suitable infrastructures are already in place.

However, despite a substantial amount of encouraging preclinical data, translation of bacterial-based cancer therapies into the clinic is progressing at a slow pace. The intrinsic toxicity risk associated with bacterial organisms in the context of a human host poses a major concern for therapies with live bacteria[20, 47]. As discussed in more detail in Chapter 6, steps are taken, when necessary, to attenuate the candidate organisms; however, this is not a straightforward task, given that treatment effectiveness is usually intertwined with toxicity. Additionally, the limited clinical experience with bacterial cancer therapies to date provides little information on toxicities and adverse effects associated with their use; therefore, prevention and/or management of potentially elicited clinical symptoms may not be currently feasible[47]. Also, bacterial genetic instability occurring intratumorally poses further challenges, since administered strains may unexpectedly acquire harmful or ineffective phenotypes[91].

Furthermore, the determination of administered doses of bacterial therapies represents a unique problem. Effective dose will depend on the tumour targeting efficiency, extent of bacterial replication and, if applicable, expression levels of the desired therapeutic. Yet, these parameters are likely to vary greatly, according to the type and size of the tumour, the tumour microenvironment as well as the degree of infiltration of immune cells in the tumour[47]. Also, human patient selection warrants special considerations. For instance, bacterial-based therapies may pose greater safety risks in immunocompromised patients, given that a weakened immune system (for instance, due to a haematological cancer or following chemotherapy) could be a predisposing factor to systemic bacterial infections. With regards to the strictly anaerobic clostridial-based therapies, treatment may also be counter-indicated in patients with necrotic tissues other than tumours, such as brain abscesses or diverticulitis[92]. Likewise, the possible

outcomes and risks of co-treating cancer with both bacterial therapies and conventional modalities, require further elucidation. Determining appropriate combinations will be necessary before bacterial cancer therapies can be incorporated in standard treatment regimens[20].

In the case of engineered Clostridia in particular, additional factors have further retarded their clinical implementation in cancer therapy. Amongst these is the hitherto limited availability of genetic tools and the consequent use of inferior strains which impeded the generation of robust preclinical efficacy data. Additionally, the inability, until recently, to introduce heterologous genes in the chromosome of clostridial strains entailed the use of plasmids bearing antibiotic resistance determinants, which would not be suitable in a clinical setting[43]. Finally, with regards to CDEPT, the lack of a kinetically favourable PCE/prodrug combination has been disabling, a problem compounded by the general lack of safe prodrugs currently available.

1.8 Objective and thesis organisation

The generic aim of this thesis was to supply comprehensive improvements in the clostridial vector (*C. sporogenes* NCIMB 10696), encompassing engineering activities to increase its efficacy and safety, and expedite its clinical implementation. Albeit not clear-cut, the open-ended nature of the goal entailed that it could be synchronously pursued from different angles. Inevitably, this led to the development of four somewhat independent projects which are presented herein:

Chapter 3 describes the generation and characterisation of clinical trial-ready strains for use in the well-established context of CDEPT, nonetheless revitalised via the incorporation of selected nitroreductases capable of dually metabolising not only a clinical-stage prodrug, but also imaging agents of interest. Chapter 3 also describes crystallisation studies on a nitroreductase candidate and the determination of its protein sturcture.

Chapter 4 explores advanced methods for the overexpression of genes of interest chromosomally, including multicopy genomic integration and the implementation of an orthogonal expression system (TcdR-P_{tcdB}). Importantly, the power of the CRISPR/Cas9 technology is fully harnessed, heralding ambitious engineering prospects that can be pursued hereafter.

Chapter 5 describes the incorporation of an immunotherapeutic functionality in the clostridial vector, in sync with the promise that immunotherapy holds for the treatment of cancer. Specifically, the expression and secretion of selected cytokines was piloted and optimised.

Chapter 6 embarks on a campaign to ablate the haemolytic activity of *C. sporogenes*, which has emerged as a potentially disabling characteristic. An attenuated version of the organism deficient in clostridiolysin production was successfully created, so is now available for downstream research and clinical use. In the process, certain aspects of the genetic background of haemolysis were also revealed.

Chapter 2 describes the materials and methods that enabled the research undertaken.

Finally, **Chapter 7** discusses exciting opportunities that can be pursued in the future, by bridging the separate outcomes of the present study.

CHAPTER 2

Materials and Methods

2.1 Microbiology

2.1.1 Bacterial strains

All bacterial strains used and generated in this study can be found in Appendix I.

2.1.2 Growth media, buffers and supplements

Except where otherwise specified, media components, antibiotics and other chemicals were supplied by Sigma-Aldrich or Fisher Scientific. Growth media and buffers were prepared in distilled water (dH_2O) and sterilised by autoclaving for 20 min at 121°C or by filtration through a 0.2-µm pore size filter (Minisart®), as required. Their compositions are outlined in Table 2.1. Agar plates were prepared using an additional 1.5% w/v No. 1 Bacteriological Agar (Oxoid). *Escherichia coli* strains were routinely grown in LB medium and *C. sporogenes* strains in TYG medium, except where otherwise stated.

Table 2.1 Growth media and buffers used in this study

Medium/Buffer	Composition	Amount	pH
Luria-Bertani (LB)	Tryptone	10 g/L	7.2
	Yeast Extract	5 g/L	
	NaCl	10 g/L	
Tryptone-Yeast Extract-Sodium Thioglycolate medium (TYG)	Tryptone	30 g/L	7.2
	Yeast Extract	20 g/L	
	Sodium thioglycolate	1 g/L	
2 x YT	Tryptone	16 g/L	5.0/7.0
	Yeast Extract	10 g/L	
	NaCl	5 g/L	
	Glucose	2% w/v	
Clostridial Growth Medium (CGM)	Tryptone	2 g/L	7.0
	Yeast Extract	1 g/L	
	Glucose	5% w/v	
	$(NH_4)_2SO_4$	2 g/L	
	K_2HPO_4	1 g/L	
	KH_2PO_4	0.5 g/L	
	$MgSO_4.7H_2O$	0.1 g/L	
	$MnSO_4.H_2O$	15 mg/L	
	$FeSO_4.7H_2O$	10 mg/L	
	$CaCl_2$	10 mg/L	
	$CoCl_2$	2 mg/L	
	$ZnSO_4$	2 mg/L	
Phosphate-Buffered Saline (PBS)	NaCl	8 g/L	7.4
	KCl	0.2 g/L	
	Na_2HPO_4	1.44 g/L	
	KH_2PO_4	0.24 g/L	

All media supplements utilised in this study and their method of preparation are outlined in Table 2.2.

Table 2.2 List of supplements used in this study

Supplement	Stock concentration	Solvent	Working concentration	Storage	Sterilisation method
Chloramphenicol (Cm)	25 mg/ml	100% ethanol	12.5 μg/ml liquid 25 μg/ml solid	-20°C	-
Thiamphenicol (Tm)	15 mg/ml	50% ethanol	15 μg/ml 30-50 μg/ml for ACE	4°C	filtration
D-cycloserine	50 mg/ml	dH_2O	250 μg/ml	4°C	filtration
Kanamycin (Kan)	50 mg/ml	dH_2O	50 μg/ml	-20°C	filtration
Erythromycin (Erm)	50 mg/ml	100% ethanol	10 μg/ml (in *C. sporogenes*)	4°C	-
Glucose	50% w/v	dH_2O	0.05% w/v	15-25°C	autoclaving
Lactose	500 mM	dH_2O	0.1-10 mM	-	filtration
5-fluoroorotic acid (Europa Bioproducts)	100 mg/ml	DMSO†	4-5 mg/ml	-	-
Theophylline	266 mM	DMSO†	5 mM	-	-
cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche)	25x	PBS	1x	-	-
Bovine serum albumin	100 mg/ml	PBS	10 mg/ml	-	filtration
Horse or sheep blood defibrinated (Oxoid)	-	-	5% v/v	4°C	-
Red blood cells from horse blood	50% v/v	PBS	0.8% v/v 2.0% v/v	-	-

†DMSO: dimethyl sulfoxide; EDTA: ethylenediaminetetraacetic acid

2.1.3 Culture conditions

E. coli strains were grown aerobically at 37°C, with additional agitation at 200 rpm for liquid cultures. *C. sporogenes* strains were grown at 37°C in an anaerobic cabinet (80% N_2, 10% CO_2, 10% H_2; model MG1000 Mark II, Don Whitley Scientific Limited). Liquid and solid media used for the growth of clostridial strains were pre-reduced for a

minimum of 4 hrs in the anaerobic cabinet before use. Supplements were added to the media as required.

2.1.4 Storage conditions

E. coli strains were stored using the Microbank™ Long Term Bacterial and Fungal Storage System (Pro-lab Diagnostics). *C. sporogenes* strains were stored in their media supplemented with 10% v/v glycerol. Culture stocks were stored at -80°C.

2.1.5 Starter cultures

Stains of interest were inoculated from frozen stocks in liquid media with the necessary supplements and grown overnight.

2.1.6 Quantification of bacterial growth

To assess bacterial growth, the optical density (OD) of liquid cultures was measured at 600 nm using a CLARIOstar® microplate reader (BMG LABTECH) and flat bottom 96-well plates (Costar, Corning®). Appropriate dilutions were carried out as necessary. Measurements were taken against a blank of sterile medium.

2.1.7 Preparation of electrocompetent *E. coli*

Sterile, pre-warmed LB was inoculated to an OD_{600} of 0.01 from a starter culture of BL21 (DE3), TOP10 or CA434 cells. Cells were harvested at an OD_{600} of 0.5-0.7 by centrifugation at 5000 ×g for 10 min at 4°C. Cell pellets were subsequently washed three times by gentle re-suspension in ample amount of sterile, ice-cold dH_2O followed by centrifugation at the same conditions. Finally, cells were re-suspended in 1 ml of 1 mM MOPS solution with 10% v/v glycerol, split into 50-µl aliquots and stored at -80°C until further use.

2.1.8 Transformation of *E. coli* by electroporation

For each transformation, a 50 µl aliquot of electrocompetent *E. coli* BL21 (DE3), TOP10 or CA434 cells was thawed on ice and mixed with 200-500 ng of plasmid DNA or 10-20 µl of dialysed ligation solutions in a 0.2-cm electroporation cuvette (Bio-Rad). The mixture was pulsed to 2.5 kV (12.5 kV/cm) using a Bio-Rad MicroPulser™ Electroporator, followed by immediate addition of 400 µl of S.O.C medium (Invitrogen™). Cells were recovered at 30 or 37°C/200 rpm for 1-2 hrs and subsequently plated in suitable dilutions on selective LB agar plates.

2.1.9 Transformation of *C. sporogenes* by conjugation

For each conjugation, donor (*E. coli* CA434 cells harbouring a shuttle plasmid) and recipient (desired C. *sporogenes* strain) cells were grown overnight in broths. One ml of donor cells was centrifuged at 5000 ×g for 1 min, washed with 1 ml of PBS and harvested as before. The resulting pellet was transferred to the anaerobic cabinet and resuspended in 200 µl of the recipient strain. The mixture was spotted on TYG plates supplemented with glucose and incubated inside the cabinet for 7 -24 hrs. Finally, the conjugation mixture was resuspended in 500 µl of PBS and plated in suitable dilutions on TYG plates, supplemented with D-cycloserine to counter-select for *E. coli* donor cells and a second appropriate antibiotic to maintain the conjugated plasmid.

2.1.10 Preparation of cell lysates

Starter cultures were diluted to a starting OD_{600} of 0.01 in fresh broth with the necessary supplements. Two-ml samples were collected at stationary phase (8 hrs) and harvested by centrifugation at 10000 ×g for 10 min at 4°C. Cell pellets were stored at -20°C overnight. Cells were lysed using the BugBuster® Master Mix (Novagen) with addition of cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche) according to the manufacturer's instructions. The soluble fractions were collected by centrifugation at 10000 ×g for 20 min at 4°C and stored at 4°C until further use.

2.1.11 Preparation of cell-free culture supernatants

Starter cultures were diluted to a starting OD_{600} of 0.01 in fresh broth with the necessary supplements. Two-ml samples were collected at stationary phase (8 hrs) and harvested by centrifugation at 10000 ×g for 20 min at 4°C. The supernatants were filtered through a 0.2-μm membrane (Minisart®) and transferred to new tubes. When required, the filtrates were stored at -20°C overnight.

2.1.12 Growth curves

Starter cultures were diluted to a starting OD_{600} of 0.01 in fresh broth with the necessary supplements and growth was monitored over a period of 24 hrs. Samples were taken at 2- or 3-hr intervals, and bacterial growth was measured as described in section 2.1.6.

2.1.13 Sporulation and germination assay

Starter cultures were diluted to a starting OD_{600} of 0.01 in fresh broth with the necessary supplements. Cultures were incubated in the cabinet for a total of 120 hrs. Samples were taken at 0, 24, 48, 72, 96 and 120 hrs and subjected to heat treatment at 80°C for 20 min. Heat-treated samples were serially diluted in PBS as required, and 20-μl aliquots from each dilution were plated in triplicates on agar. Non-heat-treated samples were also plated in a similar fashion, as a control for total cell count. Recovered colonies were enumerated the following day.

2.1.14 Qualitative assessment of haemolytic activity

Starter cultures were diluted 10-fold in PBS and 5 μl of were spotted on TYG/CGM agar containing 5% v/v defibrinated horse blood (Oxoid). Plates were incubated anaerobically for 48 hrs. Haemolytic activity, evident as of zones of clearance around the area of bacterial growth, was visually assessed every 24 hrs.

2.1.15 Preparation of red blood cell stock solution

Red blood cells (RBCs) were isolated from whole blood (WB) by centrifugation at 2000 ×g for 5 min and removal of the resulting plasma and buffy coat layers. RBCs were subsequently washed 3 times in 2 volumes of PBS and diluted with 1 volume of PBS, to yield a 50% v/v RBC solution. For the quantitative haemolytic activity assays, the stock solution was diluted to 0.8% v/v in TYG or 21YT liquid media or in PBS, as required for the co-culture and cell-free supernatant assays respectively.

2.1.16 Quantitative assessment of haemolytic activity in co-cultures

Haemolytic activity in co-cultures was monitored using a modification of the protocol by Jiang *et al.*[93]. Starter cultures were diluted to a starting OD_{600} of 0.01 in fresh broth containing 0.8% v/v RBCs and incubated anaerobically for 48 hrs. To measure haemolytic activity at selected timepoints, 200-µl samples were collected in flat bottom 96-well plates (Costar, Corning®) and centrifuged at 2000 ×g for 10 min to remove bacterial and red blood cells. The negative control consisted of non-inoculated RBC medium and the positive control of RBC medium treated with 1% v/v Triton™ X-100 at each timepoint. Cell-free supernatants (100 µl) were transferred to new 96-well plates and absorbance was measured at 404 nm using a CLARIOstar® microplate reader (BMG LABTECH). Haemolytic activity was expressed as percentage of haemolysis relative to the positive control, calculated according to equation 1, where ABS stands for the absorbance reading at 404 nm:

Equation 1: Calculation of haemolytic activity

$$\%\ \text{haemolysis} = \frac{\text{ABS of sample} - \text{ABS of negative control}}{\text{ABS of positive control} - \text{ABS of negative control}} \times 100$$

2.1.17 Quantitative assessment of haemolytic activity in cell-free supernatants

Haemolytic activity in cell-free supernatants was monitored using a modification of the protocol of Gonzalez *et al.*[94]. Starter cultures were diluted to a starting OD_{600} of 0.01 in fresh broth with or without 10 mg/ml bovine serum albumin (BSA) and incubated for 48 hrs. To measure haemolytic activity at selected timepoints, 200-µl samples were harvested at 2000 ×g for 10 min. Cell-free supernatants (100 µl) were transferred to flat bottom 96-well plates (Costar, Corning®) containing 100 µl of 0.8% v/v RBCs in PBS and incubated at 37°C/180 rpm for 1 hr. The negative control consisted of sterile TYG medium (100 µl) in 100 µl of 0.8% v/v RBCs in PBS and the positive control was prepared as the negative control with the addition of 1% v/v Triton™ X-100. Subsequently, samples were centrifuged at 2000 ×g for 10 min to remove bacterial and red blood cells, 100 µl of the supernatants were transferred to new 96-well plates and absorbance was measured at 404 nm using a CLARIOstar® microplate reader (BMG LABTECH). Haemolytic activity was expressed as percentage of haemolysis relative to the positive control, calculated according to equation 1, as described previously.

2.1.18 Hemin binding assay

C. acetobutylicum starter cultures were diluted to a starting OD_{600} of 0.01 in fresh 21YT broth (pH 5.0) and incubated anaerobically. One-ml samples were collected at 24 hrs and harvested by centrifugation at 10000 ×g for 5 min. The pellets were resuspended in 500 µl of 3 mM hemin dissolved in PBS and incubated at room temperature for 1 hr. Subsequently, bacterial cells were removed by centrifugation under the same conditions and the supernatants were transferred to flat bottom 96-well plates (Costar, Corning®). Absorbance was measured at 404 nm using a CLARIOstar® microplate reader (BMG LABTECH), against the hemin solution as a positive control and PBS as the blank.

2.2 Molecular biology

2.2.1 Plasmid DNA extraction

Plasmid DNA was extracted from overnight cultures using the Monarch® Plasmid Miniprep Kit (New England Biolabs; NEB) as per the manufacturer's instructions and eluted in dH_2O.

2.2.2 Genomic DNA extraction

2.2.2.1 Routine extractions

Genomic DNA was routinely extracted from overnight cultures using the GenElute™ Bacterial Genomic DNA Kit (Sigma-Aldrich) as per the manufacturer's instructions and eluted in dH_2O.

2.2.2.2 Phenol-chloroform extractions

Genomic DNA to be sent for Illumina sequencing was extracted using Phenol: Chloroform: Isoamyl Alcohol (25:24:1). Two-ml of overnight cultures were centrifuged at 12000 ×g for 5 min. The pellets were resuspended in 180 µl of lysis buffer (10 mg/ml lysozyme in PBS) and incubated at 37°C for 30 min. Subsequently, 25 µl of proteinase K (Qiagen), 85 µl of dH_2O and 110 µl of 10 % w/v sodium dodecyl sulfate were added. Following an incubation of 30 min at 65°C, 400 µl of phenol were added and mixed thoroughly. The contents were transferred to Phase Lock Gel Heavy tubes (5PRIME) and centrifuged at 12000 ×g for 3 min; the addition of phenol and centrifugation steps were repeated twice. DNA was precipitated from the top layer with 40 µl of 3 M sodium acetate and 800 µl of ice-cold ethanol. The precipitated DNA was scooped out using a glass Pasteur pipette and washed in 1 ml of 70% v/v ice-cold ethanol. Finally, the DNA pellet was dried at room temperature and dissolved in dH_2O.

2.2.3 Polymerase Chain Reaction (PCR)

DNA polymerases used for the amplification of desired DNA fragments included the DreamTaq Green PCR Master Mix (Thermo Scientific) and the KOD Hot Start Master Mix (Merck Millipore). All PCR reactions were assembled and performed according to manufacturer's instructions, using a thermal cycler block (Applied Biosystems Gene Amp PCR system 2400).

2.2.3.1 Colony PCR

Colony PCR was performed with DreamTaq Green Polymerase for the screening of clones and mutants; the DNA template was 1 µl of a bacterial cell suspension in dH_2O. To facilitate cell lysis, the initial denaturation step was extended to 10 min.

2.2.3.2 Inverse PCR

Inverse PCR was performed according to the procedure of Cartman and Minton[73] to establish whether ClosTron group II introns had been inserted in the genome in a single copy. Genomic DNA of confirmed mutants was extracted as described in section 2.2.2.1, digested with HindIII and, after self-ligation, subjected to inverse PCR with primers inv_F and inv_R. The gel-purified inverse PCR products were sequenced with primer erm_RAM_seq to confirm the site(s) of insertion.

2.2.3.3 Primers

Primers were synthesised by Eurofins Genomics or Sigma-Aldrich and supplied salt-free in lyophilised form. They were reconstituted in dH_2O according to the manufacturer's instructions and diluted to a working concentration of 10 µM. Stock and working solutions were stored at -20°C. Appendix II includes a list of the primers used in this study.

2.2.4 Agarose gel electrophoresis

Analysis of DNA fragments was performed by agarose gel electrophoresis using 1% w/v agarose in 1 x Tris-acetate EDTA (TAE) buffer (VWR International Ltd). For staining, SYBR® Safe DNA Gel Stain (Invitrogen) was diluted 1:10000 in molten agarose. Gels were run in 1 x TAE buffer at 90-120 V for 40-60 min. When required, 6 x purple loading dye (NEB) was added to samples before loading in a 1:5 (dye:sample) volume ratio. A 2-Log DNA ladder (NEB) was used as appropriate. Gels were visualised in a Gel Doc™ XR+ imaging system (Bio-Rad). When required, DNA fragments were extracted from agarose gels under ultraviolet light using a sterile scalpel.

2.2.5 DNA purification

DNA was purified from PCR reaction mixtures using the Monarch® PCR & DNA Cleanup kit and from agarose gels with the Monarch® DNA Gel Extraction kit, as per the manufacturer's instructions. DNA was eluted in dH_2O.

2.2.6 Quantification of DNA

DNA concentration was measured at 260 nm using a NanoDrop Lite Spectrophotometer (Thermo Scientific). Measurements were taken against a blank of dH_2O. The concentration of genomic DNA intended for Illumina sequencing was measured using a Qubit 4 Fluorometer (Invitrogen) according to the manufacturer's instructions.

2.2.7 Cloning techniques

2.2.7.1 Restriction endonuclease digestion of DNA

All restriction endonucleases and accompanying buffers were supplied by NEB. Digestions were set up in accordance to the manufacturer's instructions and incubated at 37°C for a minimum of 15 min.

2.2.7.2 Ligation and membrane dialysis

Digested vector backbones and inserts were ligated together at molar ratios (insert:vector) of 1:1, 3:1, 5:1 or 7:1. Ligation reactions were set up in a final volume of 10 or 20 µl with dH_2O using T4 DNA ligase and buffer (Promega), according to the manufacturer's instructions. The reactions were incubated overnight on melting ice or at room temperature for 3 hrs. Prior to transformation into electrocompetent cells, ligation products were subjected to a 30-min dialysis through an MF-Millipore™ membrane of 0.025-µm pore size (Merck), over an ample volume of dH_2O.

2.2.8 Sanger DNA sequencing

Sanger DNA sequencing was carried out using the Eurofins Genomics' TubeSeq Service for plasmid DNA and purified PCR products.

2.2.9 Illumina genome sequencing

Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk), using their Illumina-based, Standard Whole-Genome Service.

2.2.10 DNA synthesis

When required, DNA fragments were commercially synthesized by ATUM (CA, USA), Biomatik (Ontario, Canada) or Twist Bioscience (CA, USA).

2.2.11 Plasmid construction

A list of plasmids used and constructed in this study can be found in Appendix III and representative plasmid maps are included in Appendix IV. Plasmid assembly was carried out by restriction enzyme digest (subcloning) procedures in *E. coli* TOP10 and confirmed via colony PCR. Correct plasmids were isolated, verified by sequencing and when required electroporated in *E. coli* CA434 cells for conjugation in *C. sporogenes*, or in *E. coli* BL21 (DE3) for the expression and purification of YfkO_Bl.

2.2.11.1 Low copy number plasmids for use in E. coli

To overcome toxicity problems encountered at the cloning stages, selected vectors were modified by replacing the high-copy number Gram-negative replicon ColE1 with the low-copy number p15a replicon. The p15a+*traJ* module was digested out of pMTL85121 via PmeI/SbfI and ligated in the PmeI/SbfI linearised pMTL82151, pMTL-MZ1 and pMTL83151-RECas1 backbones. The new vectors were designated pMTL82121, pMTL-MZ1-p15a and pMTL83121-RECas1 respectively.

2.2.11.2 Modification of pMTL-JH27 (Chapters 3 and 4)

Integrations in the *pyrE* locus of *C. sporogenes* via the means of allele-coupled exchange (ACE) were carried out using the standard pMTL-JH27 vector (Genbank HQ875763.1) or its derivatives, pMTL-MZ1, pMTL-JH27-p15a and pMTL-MZ1-p15a. In the pMTL-MZ1 and pMTL-MZ1-p15a vectors, the right homology arm (RHA) has been elongated from 1.2 to 2.5 kb, to improve the likelihood of homologous recombination. The new RHA was amplified from genomic DNA of *C. sporogenes* with primers NheI-nRHA_F and AscI-nRHA_R and cloned in pMTL-JH27 via NheI/AscI, creating vector pMTL-MZ1. The low copy number vectors pMTL-JH27-p15a and pMTL-MZ1-p15a were created as described in section 2.2.11.1.

2.2.11.3 Plasmids for integration of nitroreductases using ACE (Chapter 3)

P_{thl14}-nitroreductase (NTR) constructs were typically synthesized commercially by Biomatik, ATUM or Twist Bioscience; fragments were digested out of their supplied backbones via NotI/XhoI and subcloned into NotI/XhoI linearised pMTL-JH27 or pMTL-MZ1 vector backbones. P_{thl14}-*yfkO_Bl_362* was assembled in two parts using pMTL-JH27-P_{thl14}-*yfkO_Bl* as the DNA template: the first part was amplified using primers NotI-Pthl_F and BsaI-YfkO-5'-mut_R, with the reverse primer introducing the F84L mutation; the second part was amplified using primers BsaI-YfkO-3'-mut_F

and XhoI-YfkO_R with the forward primer introducing the M112I mutation. The two parts were ligated together via BsaI. The ligated fragment was amplified with NotI-Pthl_F and XhoI-YfkO_R, digested NotI/XhoI and assembled into a NotI/XhoI linearised pMTL-MZ1 vector backbone. For the P_{fdx}-*nfrA_Bs* construct, P_{fdx} was amplified from genomic DNA of *C. sporogenes* using the NotI-Pfdx_F and BsaI-Pfdx_R1 primers, and *nfrA_Bs* was amplified from the pMTL-JH27-P_{thl14}-*nfrA_Bs* vector with BsaI-NfrA_F and XhoI-NfrA_R. The two parts were ligated together via BsaI and the ligated product was amplified with the NotI-Pfdx_F and XhoI-NfrA_R primers, introducing the NotI and XhoI sites respectively. The amplicon was digested NotI/XhoI and assembled into a NotI/XhoI linearised pMTL-JH27 vector backbone.

2.2.11.4 Plasmid for the purification of YfkO_Bl (Chapter 3)

To express YfkO_Bl with an N-terminal His-tag, *yfkO_Bl* was amplified with primers NdeI-YfkO_F and XhoI-YfkO_R from pMTL-JH27-P_{thl14}-*yfkO_Bl* and inserted into a pET-28b vector via NdeI/XhoI, generating vector pET-28b-*yfkO_Bl*.

2.2.11.5 Expression plasmids for TcdR-P_{tcdB} constructs (Chapter 4)

For the pMTL82151-P_{lac}-OS vector, P_{lac}-*tcdR*-P_{tcdB} was amplified as a monomeric fragment with primers NotI-T1bgaR_F and BsaI-PtcdB_R1 using vector pMTL-HZ7 as a template; the *yfkO_Bl* fragment was amplified with primers BsaI-YfkO_F and XhoI-YfkO_R using vector pMTL-JH27-P_{thl14}-*yfkO_Bl* as a template. The amplicons were digested with NotI/BsaI and BsaI/XhoI, respectively and assembled in a NotI/XhoI linearised pMTL82151 backbone in one step. For the pMTL82151-P_{tcdB}-*yfkO_Bl* vector, P_{tcdB} was amplified with primers NotI-PtcdB_F and BsaI-PtcdB_R1 using vector pMTL-HZ7 as a template; the *yfkO_Bl* fragment was amplified as previously described. The amplicons were digested with NotI/BsaI and BsaI/XhoI respectively and assembled in a NotI/XhoI linearised pMTL82151 backbone in one step. Finally, for the pMTL-82151-T1-P_{ptb}-OS vector, terminator T1 was amplified as a primer dimer using NotI-T1_F and XhoI-BsaI-T1_R, P_{ptb} using primers BsaI-Pptb_F

and BsaI-Pptb_R and genomic DNA from *C. acetobutylicum* ATCC 824 as a template, and *tcdR*-P_{tcdB}-*yfkO_Bl* as a monomeric fragment using primers BsaI-tcdR_F and XhoI-YfkO_R and the pMTL82151-P_{lac}-OS vector as a template. The three fragments were digested with NotI/BsaI, BsaI/BsaI and BsaI/XhoI respectively and assembled in a NotI/XhoI linearised pMTL82151 backbone in one step.

2.2.11.6 Plasmids for integration of TcdR-P_{tcdB} constructs using ACE (Chapter 4)

For the pMTL-MZ1-P_{lac}-OS vector, the insert was amplified with the NotI-T1bgaR_F and XhoI-YfkO_R primers using pMTL82151-P_{lac}-OS as a template, and cloned via NotI/XhoI in the pMTL-MZ1 vector. The pMTL-MZ1-p15a-T1-P_{ptb}-OS vector was created in two steps; terminator T1 was amplified as a primer dimer using NotI-T1_F and XhoI-BsaI-T1_R and cloned in the pMTL-MZ1-p15a vector via NotI/XhoI, generating vector pMTL-MZ1-p15a-T1. Subsequently, P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* was amplified from the pMTL-82151-T1-P_{ptb}-OS vector with primers BsaI-Pptb_F and XhoI-YfkO_R and cloned via BsaI/XhoI in pMTL-MZ1-p15a-T1. For the pMTL-MZ1-P_{lac}-*tcdR* vector, the insert was amplified with NotI-T1bgaR_F and XhoI-tcdR_R using pMTL82151-P_{lac}-OS as a template and cloned via NotI/XhoI in the pMTL-MZ1 vector.

2.2.11.7 Plasmids for CRISPR/Cas9 genome editing (Chapters 4 and 6)

CRISPR/Cas9-mediated mutagenesis was carried out using the pMTL83151-RECas1 or pMTL83121-RECas1 vector backbones, containing the *cas9* gene from *Streptococcus pyogenes* (*S. pyogenes*) under the control of the *C. sporogenes* ferredoxin promoter, fused to a theophylline-responsive riboswitch (P_{fdx}-rbE). The single-guide RNA (sgRNA) is driven by the constitutive *araE* promoter (P_{araE}) from *C. acetobutylicum*. sgRNA sequences were selected on the basis of high on- and off- target scores, as calculated on the Benchling platform using the algorithms by Doench *et al.*[95, 96] and Hsu *et al.*[97]. The sequences were further verified using the CRISPy-web platform [98]. sgRNAs were generated as primer dimers using a forward primer, specific for each sgRNA (Appendix

II), and the sgRNA_Rpd reverse primer, inserting the SalI and AsiSI restriction sites, respectively.

For deletions (Chapter 6), editing templates consisted of a left (LHA) and right (RHA) homology arm, homologous to sequences flanking the region to be deleted, amplified from genomic DNA of *C. sporogenes*. The LHA was amplified with a forward primer introducing the AsiSI restriction site and a reverse primer introducing the BsaI site with a specific overhang for the RHA. Likewise, the RHA was amplified with a forward primer introducing the BsaI site and a reverse primer introducing the AscI site. The sgRNA was digested with SalI/AsiSI, the LHA with AsiSI/BsaI and the RHA with BsaI/AscI and all fragments were ligated into SalI/AscI digested pMTL83151-RECas1 or pMTL83121-RECas1 vector backbones in one step.

For integrations (Chapter 4), the editing templates consisted of an additional cargo ligated with the LHA via NotI and with the RHA via XhoI; the LHA and RHA were amplified as previously described, with the reverse and forward primers introducing instead of BsaI, the NotI and XhoI sites respectively. The LHA for integrations in the c14550-3' locus was amplified in two parts, as required for the modification of the editing template to introduce silent mutations in the sgRNA and PAM sequences: part A was amplified using AsiSI-L-c14550_F and BsaI-L-c14550-A_R (with the reverse primer introducing the desired mutations) and part B using BsaI-L-c14550-B_F and NotI-L-c14550_R; the two fragments were ligated together via BsaI and amplified with the flanking primers AsiSI-L-c14550_F and NotI-L-c14550_R, as for all LHA fragments. For the assembly of the editing templates encompassing T1-P_{thl14}-*yfkO_Bl*-T_{fdx} (expression cassette; EC), T1-P_{tcdB}-*yfkO_Bl*-T_{fdx} (orthogonal system; OS) or T1-P_{ptb}-*tcdR*, T1 was generated as a primer dimer using NotI-T1_F and XhoI-BsaI-T1_R; P_{thl14}-*yfkO_Bl*-T_{fdx} was amplified from the pMTL-JH27-P_{thl14}-*yfkO_Bl* vector using primers BsaI-Pthl_F and XhoI-Tfdx-YfkO_R, with the reverse primer introducing the T_{fdx} terminator; P_{tcdB}-*yfkO_Bl*-T_{fdx} was amplified using the pMTL82151-P_{tcdB}-*yfkO_Bl* vector as a template and primers BsaI-PtcdB_F and XhoI-Tfdx-YfkO_R, with the

reverse primer introducing the T_{fdx} terminator; P_{pbt}-*tcdR* was amplififed from the pMTL-82151-T1-Pptb-OS vector using BsaI-Pptb_F and XhoI-tcdR_R. Subsequently, the LHA was digested with AsiSI/NotI, T1 with NotI/BsaI, the P_{thl14}-*yfkO_Bl*-Tfdx, PtcdB-*yfkO_Bl*-T_{fdx} or P_{pbt}-*tcdR* modules with BsaI/XhoI and the RHA with XhoI/AscI. The fragments were linearly ligated and PCR-amplified with flanking primers annealing on the LHA and RHA introducing the AsiSI and AscI restriction sites, respectively. Finally, digested sgRNA (SalI/AsiSI) and editing templates (AsiSI/AscI) were ligated into SalI/AscI linearised pMTL83121-RECas1 backbones in one step.

2.2.11.8 Expression plasmids for mGM-CSF and mIL-4 (Chapter 5)

The *npr* signal peptide was amplified from genomic DNA of *C. sporogenes* using BsaI-npr_F and BsaI-npr_R1 or BsaI-npr_R2 for mGM-CSF and mIL-4, respectively; *lip* was generated as a primer dimer using BsaI-lip_F and BsaI-lip_R1 for mGM-CSF or BsaI-lip_R2 for mIL-4. Likewise, *egla* was generated as a primer dimer with primers BsaI-egla_F and BsaI-egla_R1 or BsaI-egla_R2 for mGM-CSF and mIL-4, respectively. P_{tcdB} was amplified from the pMTL82151-P_{tcdB}-*yfkO_Bl* vector using NotI-PtcdB_F and BsaI-PtcdB_R2. Codon-optimised mGM-CSF and mIL-4 sequences were synthesized by BIOMATIK and amplified with BsaI-GMCSF_F/XhoI-flag-GMCSF_R or BsaI-IL4_F/XhoI-flag-IL4_R respectively, with the reverse primers introducing the FLAG-tag. The promoter modules were digested with NotI/BsaI, the signal peptide modules with BsaI/BsaI and the gene modules with BsaI/XhoI, and all fragments were ligated accordingly (that is, each gene with all three signal peptides) into NotI/XhoI digested pMTL82121 backbones in one step. For the non-secreting controls, P_{tcdB} was amplified from the pMTL82151-P_{tcdB}-*yfkO_Bl* using NotI-PtcdB_F and BsaI-PtcdB_R3 for mGM-CSF or NotI-PtcdB_F and BsaI-PtcdB_R4 for mIL-4, with the reverse primers introducing a start codon. The promoter fragments were digested with NotI/BsaI and ligated into NotI/XhoI digested pMTL82121 backbones, alongside an mGM-CSF or mIL-4 fragment, amplified and digested as described previously.

2.2.12 Targeted genomic mutagenesis of *C. sporogenes*

2.2.12.1 ACE

Genomic integrations in the *pyrE* locus of *C. sporogenes* were carried out using the ACE method described by Heap *et al.*[75] and the pMTL-JH27 vector (Genbank HQ875763.1) or its derivative pMTL-MZ1. Vectors with the cargos of interest were conjugated in *C. sporogenes*, as described in section 2.1.9. Conjugated cells were plated on TYG/D-cycloserine/thiamphenicol plates and incubated for 48 hrs. Morphologically larger colonies were screened via colony PCR with a plasmid-specific (lacZα) and a chromosomal (Csp-3233-SR3 or MZ1-SC_scr) primer to confirm single-crossover mutants. Confirmed single-crossover colonies were directly streaked on TYG plates containing 4.0, 4.5 and 5.0 mg/ml of 5-fluoroorotic acid (5-FOA) to facilitate selection of successful double-crossover Δ*pyrE* mutants. Colonies resistant to 5-FOA were screened via colony PCR using chromosomal primers (Csp-pyrD-sF2 and Csp3233-SR3), to confirm the integration event. Mixed populations (containing both wildtype and mutant cells) were purified on TYG/5-FOA plates and confirmed again via colony PCR.

2.2.12.2 CRISPR/Cas9

CRISPR/Cas9-mediated mutagenesis was carried out for deletion of desired genes as well as ACE-independent genomic integrations. CRISPR/Cas9 plasmids (from section 2.2.11.7) were conjugated in *C. sporogenes* as described in section 2.1.9. Conjugated cells were plated on TYG/D-cycloserine/thiamphenicol plates and incubated for 24-48 hrs. Successful transconjugants were selected based on acquisition of thiamphenicol resistance and re-streaked on TYG/D-cycloserine/thiamphenicol plates containing 5 mM theophylline. Colonies that grew in the presence of the inducer were screened via colony PCR with suitable primers to confirm the desired genomic modification.

2.2.12.3 ClosTron

Target genes of interest were insertionally inactivated using the ClosTron system, as described by Heap *et al.*[68, 99]. Retargeted introns for ClosTron mutagenesis were designed using a freely available tool (at www.clostron.com) and the Perutka algorithm. From the data output, introns were selected based on a high score and predicted site of insertion within or close to the functional domain of the protein. Plasmid construction and synthesis was carried out by ATUM, via incorporation of the selected intron sequences in the standard ClosTron vector, pMTL007C-E2. Plasmids were electroporated unchanged in CA434 cells for conjugation in *C. sporogenes*, carried out as described in section 2.1.9. Conjugated cells were plated on TYG/D-cycloserine/thiamphenicol plates and incubated for 24-48 hrs. Successful transconjugants were selected based on acquisition of thiamphenicol resistance and streaked out to purity on TYG/D-cycloserine/thiamphenicol plates. All growth was subsequently harvested in 1 ml of PBS and serially diluted up to 10^{-6} in PBS. Fifty µl of each dilution were plated on TYG/D-cycloserine/erythromycin plates, to select for the group II intron insertion. Group II insertion in the intended location of the genome was confirmed via colony PCR using gene-specific primers that flank the insertion site and amplify the entire inserted intron. In addition, a gene-specific forward or reverse primer (depending on the orientation of the insertion) and an intron-specific primer (EBS_U) were also used to amplify across the gene-intron junction. Finally, RAM-specific primers (RAM_F, RAM_R) were used to verify the presence of the RAM cassette.

2.2.13 Plasmid loss

Successful *C. sporogenes* mutants no longer requiring the conjugated plasmids underwent plasmid loss. Strains were grown in 10 ml of TYG medium for 48 hrs and then serially diluted in PBS. Fifty µl of appropriate dilutions were plated on TYG/D-cycloserine plates to obtain single colonies. Single colonies were replica-plated on

TYG/D-cycloserine/thiamphenicol and TYG/D-cycloserine plates; colonies that grew exclusively on the TYG/D-cycloserine plates were assumed plasmid-free. Plasmid loss was confirmed via colony PCR with plasmid-specific primers; for ACE vectors plasmid loss was confirmed with primers pIM13_F2 and CatP_R1, for CRISPR/Cas9 vectors with primers CatP_F1 and MCS_R1, and finally for ClosTron vectors with pCB102_F2 and ColE1_R1.

2.3 Biochemistry

2.3.1 Determination of total protein concentration

Protein concentrations were determined using the Pierce™ BCA Protein Assay Kit (Thermo Scientific™) and flat bottom 96-well plates (Costar, Corning®), according to the manufacturer's instructions for the microplate procedure. Albumin standards were used to generate standard curves. Absorbance was measured using a CLARIOstar® microplate reader (BMG LABTECH) set at 562 nm, against a blank of dH_2O.

2.3.2 Precipitation of protein from cell-free culture supernatants

Protein was precipitated from cell-free culture supernatants (obtained as in section 2.1.11), using the DOC/TCA method as described by Schwarz *et al.*[100]. Briefly, 0.03% w/v (final concentration) of sodium deoxycholate (DOC) was added to the filtrates, followed by incubation at room temperature for 10 min and subsequent addition of trichloroacetic acid (TCA) to a final concentration of 10% w/v. The precipitated samples were stored at -20°C overnight. The next day, samples were thawed on ice and centrifuged at 12000 ×g for 20 min at 4°C, at which point the supernatants were discarded. Pellets were washed with 1 ml of ice-cold ethanol and acetone consecutively. For each wash, samples were centrifugated at 12000 ×g for 10 min at 4°C and the supernatants were carefully discarded. Finally, the pellets were air-dried for 1 hr.

2.3.3 Inclusion body purification

Following isolation of the soluble fraction of the lysate, as described in section 2.1.10, the pelleted insoluble debris was further processed for the purification of inclusion bodies, using BugBuster® Master Mix (Novagen), according to the manufacturer's instructions. The purified pellets were air-dried for 1 hr.

2.3.4 Sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE)

The protein content of samples of interest was analysed by SDS-PAGE using an XCell SureLock™ Mini-Cell Electrophoresis System (Thermo Scientific) and 1.0 mm precast Bolt™ 4-12% Bis-Tris Protein Gels (Invitrogen). When required, prior to loading, cell lysates (from section 2.1.10) were standardised according to their total protein content by dilution in dH_2O to the lowest concentration of the set and combined in a 3:1 volume ratio with 4 × NuPAGE® LDS Sample Buffer (Novex™) premixed with 400 mM dithiothreitol (DTT). Likewise, protein pellets (from sections 2.3.2-2.3.3) were resuspended in appropriate volumes of 2 × sample buffer (4 x NuPAGE® LDS Sample Buffer diluted twofold in dH_2O), premixed with 200 mM DTT; for pellets obtained via TCA/DOC precipitation, 0.1 M NaOH was also added prophylactically, to offset any pH shift as a result of residual TCA. All samples were denatured by heating at 85°C for 5 min and briefly vortexed prior to loading. Denatured samples were loaded in equal volumes and Precision Plus Protein™ All Blue Standards (Bio-Rad) was used for molecular weight sizing. For gels that would be subjected to Western blotting, a FLAG-tag protein (Alpha Diagnostics) was mixed 1:1 with 2 × sample buffer, denatured at 85°C for 5 min and loaded alongside as a positive control. Electrophoresis was carried out at 150 V for 60 min in 1 × NuPAGE® MES SDS running buffer (Novex™). Upon completion, gels were briefly rinsed with dH_2O to remove residual SDS.

2.3.5 Gel staining and de-staining

When required, SDS-PAGE gels were stained by immersion in Coomassie brilliant blue stain solution (0.25 % w/v Coomassie Brilliant Blue R-250 (Bio-Rad), 50 % w/v methanol, 10 % w/v acetic acid) for 60 min on a shaking tray. Gels were de-stained overnight in de-staining solution (30 % w/v methanol, 10 % w/v acetic acid). Finally, gels were visualised in a Gel Doc™ XR+ imaging system (Bio-Rad).

2.3.6 Western blotting

When required, SDS-PAGE gels were subjected to Western blotting. Protein transfer was carried out using the TransBlot® Turbo™ transfer system (Bio-Rad) and Trans-Blot® Turbo™ Mini 0.2 µm PVDF Transfer Packs in accordance with the manufacturer's protocol, using the default 'mixed MW' program (1.3 A, 25 V, 7 min). Membranes were blocked by addition of 30 ml blocking buffer, consisting of PBS-T (PBS, 0.05% v/v Tween 20) + 5% w/v milk powder (Marvel Original Dried Skimmed Milk Powder), and incubated at room temperature on a shaker at 45 rpm for 1 hr. Blocked membranes were labelled by addition of monoclonal ANTI-FLAG® M2-peroxidase antibody (Sigma-Aldrich), diluted 1/10000 in 30 ml blocking buffer and incubated overnight at 4°C on a shaker at 45 rpm. Labelled membranes were subjected to 6 × 10-min washes with PBS-T at room temperature on a shaker at 45 rpm. Finally, membranes were developed by addition of 1 ml 3,3′,5,5′-Tetramethylbenzidine (TMB) (Sigma-Aldrich).

2.3.7 Menadione reductase assay

Enzymatic activity in cell lysates was measured as menadione reductase, with the method of Knox *et al.*[101], modified for a microplate format. A volume of 150 µl of a master mix consisting of 10 mM Tris buffer (pH 7.5), 10 µM menadione dissolved in DMSO, 1 mM reduced nicotinamide adenine dinucleotide (NADH) or reduced nicotinamide adenine dinucleotide phosphate (NADPH) and 70 µM cytochrome c from

bovine heart were injected into each well of a 96-well flat-bottom plate (Costar, Corning®) containing 10 µL of cell lysates (or appropriate dilutions of those) and 40 µL of Tris buffer. Blank samples contained 50 µL of Tris buffer. Reduction of cytochrome c was recorded by measuring the increase in absorbance at 550 nm for 1 min, using a CLARIOstar® microplate reader (BMG LABTECH), set at 37°C. Menadione reductase activity was calculated in units/ml (U/ml) of lysate using equation 2, where 0.2 is the total reaction volume (in ml), DF is the dilution factor where applicable, 0.028 is the micromolar extinction coefficient of cytochrome c at 550 nm and 0.01 is the volume (in ml) of cell lysate used.

Equation 2: Calculation of menadione reductase activity

$$\mathrm{U/ml} = \frac{\left(\frac{\Delta A_{550}}{\mathrm{min}}\,\mathrm{sample} - \frac{\Delta A_{550}}{\mathrm{min}}\,\mathrm{blank}\right) \times 0.2 \times \mathrm{DF}}{0.028 \times 0.01}$$

Results are presented in units/mg (U/mg), normalised for total protein concentration using equation 3, where U/ml is the result of equation 2 and mg/ml is the total protein concentration as determined from the BCA assay.

Equation 3: Normalisation of enzymatic activity for protein content

$$\mathrm{U/mg} = \frac{\mathrm{U/ml}}{\mathrm{mg/ml}}$$

2.3.8 Nitrofurazone reductase assay

Enzymatic activity in cell lysates was also measured as nitrofurazone reductase, with the method described by Whiteway *et al.*[102], modified for a microplate format. A volume of 150 µl of a master mix consisting of 10 mM Tris buffer (pH 7.5), 100 µM nitrofurazone dissolved in DMSO and 1 mM NADH or NADPH were injected into each well of a 96-well flat-bottom plate (Costar, Corning®) containing 10 µL of cell lysates (or appropriate dilutions of those) and 40 µL of Tris buffer. Blank samples contained 50 µL of Tris buffer. Reduction of nitrofurazone was determined by a decrease in absorbance at 400 nm for 1 min, using a CLARIOstar® microplate reader (BMG LABTECH), set at 37°C.

Nitrofurazone reductase activity was calculated in units/ml (U/ml) of lysate using equation 4, where 0.2 is the total reaction volume (in ml), DF is the dilution factor where applicable, 0.01296 is the micromolar extinction coefficient of nitrofurazone at 400 nm and 0.01 is the volume (in ml) of cell lysate used.

Equation 4: Calculation of nitrofurazone reductase activity

$$\mathrm{U/ml} = \frac{\left(\frac{|\Delta A_{400}|}{\mathrm{min}}\,\mathrm{sample} - \frac{|\Delta A_{400}|}{\mathrm{min}}\,\mathrm{blank}\right) \times 0.2 \times \mathrm{DF}}{0.01296 \times 0.01}$$

Results are presented in units/mg (U/mg), normalised for total protein concentration using equation 3.

2.3.9 mGM-CSF Enzyme-Linked Immunosorbent Assay (ELISA)

Secreted mGM-CSF from *C. sporogenes* was quantified using the mGM-CSF DuoSet ELISA Development kit (R&D systems) and Nunc MaxiSorp™ flat-bottom 96-well plates (Thermo Scientific), as per the manufacturer's instructions. Cell-free culture supernatants, obtained as in section 2.1.11, were diluted 5000-fold in reagent diluent (1% w/v BSA in PBS) prior to the assay. Likewise, cell lysates, obtained as described in section 2.1.10, were diluted 250-fold. Absorbance was measured at 450 nm with a reference wavelength of 540 nm. Concentrations of mGM-CSF were determined from a standard curve with mGM-CSF standards (1.95-500 pg/ml) using a four-parameter logistic (4-PL) curve-fit.

2.4 Crystallography

2.4.1 Protein expression and isolation

N-terminal His-tagged YfkO_Bl was expressed in *E. coli* BL21 (DE3) from a pET-28b vector. Starter cultures were inoculated in LB/kanamycin broth and grown at 37°C.

The cultures were induced at an OD_{600} of 0.6 - 0.8 with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested 4 hrs post-induction by centrifugation at 4000 ×g for 20 min at 4°C. Cell pellets were resuspended in binding buffer (50 mM Tris, 20 mM imidazole, 200 mM sodium chloride, pH 8.0), and sonicated on ice for 10 min with 30-s cooling pauses at 15 μm. Soluble protein was harvested by centrifugation at 30000 ×g for 35 min at 4°C.

2.4.2 Protein purification

His-tagged YfkO_Bl was purified via Immobilised Metal Affinity Chromatography using the ÄKTAprime system (GE Healthcare Life Sciences) and a 5-ml HiTrap™ Chelating High Performance column (GE Healthcare Life Sciences) preloaded with nickel sulphate. The column was equilibrated with binding buffer, loaded with the protein sample (as obtained in section 2.4.1) and washed again with binding buffer. His-tagged YfkO_Bl was eluted with eluting buffer (50 mM Tris, 500 mM imidazole, 200 mM sodium chloride, pH 8.0). Elution fractions containing protein were pooled together and concentrated as described in section 2.4.3. Subsequently, YfkO_Bl was further purified by size-exclusion chromatography on a HiLoad™ Superdex 75 16/600 gel filtration column (GE Healthcare Life Sciences) using 20 mM Tris, 150 mM sodium chloride, pH 7.5 as running buffer. Protein purity after each purification step was assessed by SDS-PAGE, as described in section 2.3.4.

2.4.3 Concentration of protein by centrifugation

Purified YfkO_Bl was concentrated by repeated centrifugation steps at 4500 ×g for 20 min at 4°C using a 10K Molecular Weight Cut-Off (MWCO) centrifugal filter (Amicon Ultra, Millipore), until the desired concentration was obtained.

2.4.4 Quantification of protein

Protein concentration was measured at 280 nm using a NanoDrop Lite Spectrophotometer (Thermo Scientific). Measurements were taken against a blank of the gel filtration running buffer.

2.4.5 Crystallisation

Pure YfKO_Bl fractions were concentrated to 5.8 and 10.4 mg/ml and subjected to crystallization trials using the SG1 Eco, PACT *premier*™, Morpheus® and JCSG-*plus*™ HT-96 screens (Molecular Dimensions). Crystal screens were performed in MRC 2-drop 96-well plates (Molecular Dimensions) via the sitting drop vapour diffusion technique, using the mosquito® HTS nanolitre handling robot (TTP Labtech) and mixing 0.4 µl of protein with 0.4 µl of reservoir solution, suspended over 70 µl of reservoir solution. Plates were sealed and incubated at 20°C. The condition B11 from the PACT *premier*™ screen was manually optimized by varying the PEG6000 component (Molecular Dimensions) from 15 to 26.5% v/v at increments of 0.5% v/v. Optimisation was carried out in Chryschem 24-well plates (Hampton Research), in sitting drops composed of 1:1 volume ratio of crystallization condition and YfkO_Bl at 5.8 or 10.4 mg/ml. Crystal trays were visually inspected for the presence of crystals using a Leica MZ6 stereomicroscope through a Leica Plan APO 1.0x objective.

2.4.6 Crystal storage

Single crystals were harvested with cryoloops and immersed in their mother solution, with the addition of 10% v/v ethylene glycol as cryoprotectant. Harvested crystals were flash cooled in liquid nitrogen.

2.4.7 Molecular visualisation

Molecular visualisation and structural 3D representations were performed using the PyMOL Molecular Graphics System version 1.74 (Schrodinger LLC). LigPlot^{+} version 2.1 was used for the generation of 2D schematic diagrams of ligand-protein interactions.

2.5 Eukaryotic cell culture

2.5.1 Isolation of murine bone marrow cells

Murine bone marrow (BM) cells were isolated from two female, 7-week old, CD-1 mice (BSU-AMK-187889/90). The femurs and tibiae were removed, cleaned of all connective tissue, rinsed with 70% v/v industrial methylated spirit and stored in RPMI-1640 medium. The ends of each tibia and femur were clipped to expose the interior marrow shaft. The marrow contents were flushed with RPMI-1640 medium, using a 1-ml syringe and a 29G × ½ needle. The contents were collected in a 50-ml falcon tube and diluted with RPMI-1640 to a final volume of 50 ml. Cells were harvested by centrifugation at 1500 ×g for 5 min and washed once again in 50 ml RPMI-1640 followed by centrifugation at the same conditions. Cells were stored at -80°C in RPMI-1640 with 10% v/v DMSO as the cryoprotective agent.

2.5.2 Cell culture

For the cellular proliferation and flow cytometry assays, murine BM cells were recovered from frozen stocks, enumerated using a haemocytometer and seeded in 96-well, tissue culture plates (Nunc™, Thermo Scientific) at a density of 5×10^4 cells per well, in a total volume of 180 µl. Cells were maintained in RPMI-1640 medium supplemented with 10% v/v foetal bovine serum (FBS) and incubated at 37°C in a humidified atmosphere containing 5% CO_2.

2.5.3 Preparation of treatment reagents

Recombinant mGM-CSF (Biolegend) and LEAF™ purified anti-mouse GM-CSF neutralising antibody (BioLegend) were prepared as 20x stocks (400 ng/ml and 200 µg/ml respectively) in RPMI-1640 medium. Cell-free culture supernatants of the mGM-CSF secreting and negative control strains, obtained as in section 2.1.11, were diluted 1.5-fold in RPMI-1640 medium. A 20x stock of recombinant mGM-CSF was also prepared in the diluted cell-free supernatant (SN) of the negative control strain.

2.5.4 Cell treatment

Cells were cultured as described in section 2.5.2 and treated at time zero (T0) with 10 µl cell-free culture supernatants or 10 µl recombinant mGM-CSF in RPMI-1640 medium or in SN (final concentration 20 ng/ml), previously prepared as in section 2.5.3. When required, 10 µl of the neutralising antibody stock solution (final concentration 10 µg/ml) were also added. Untreated controls received no treatment. All well contents were made up to a final volume of 200 µl with RPMI-1640 medium, as appropriate. Each set of treatment conditions was applied in triplicates.

2.5.5 MTT cell viability assay

Following 7 days of exposure to the applied treatments, as described in section 2.5.4, MTT (3-(4,5-dimethylthiazol- 2-yl)-2,5-diphenyltetrazolium bromide; 5 mg/ml in PBS) was added to a final concentration of 400 µg/ml (16 µl per well). The samples were incubated for 3.5 hrs at 37°C, at which point well supernatants were aspirated and formazan was solubilized with 150 µl of DMSO. The plates were incubated at 37°C for 15 min to facilitate formazan solubilisation and absorbance was measured at 540 nm with a CLARIOstar® microplate reader (BMG LABTECH), against a blank of RPMI-1640 + 10% FBS medium (200 µl), containing no cells but likewise treated with MTT.

2.5.6 Flow cytometry

Following 7 days of exposure to the applied treatments, as described in section 2.5.4, presence of differentiated monocytes was assayed by four-color immunofluorescence cell surface staining with a panel of directly conjugated antibodies: hamster anti-mouse CD11c-FITC, rat anti-mouse F4/80-PE, rat anti-mouse CD86-PE/Cy5 and hamster anti-mouse CD80-PE/Cy7 (BioLegend). Isotype-specific immunoglobulin controls were also run for each fluorochrome: hamster IgG-FITC, rat IgG2a-PE, rat IgG2a-PE/Cy5 and hamster IgG-PE/Cy7 for CD11c, F4/80, CD86 and CD80 respectively (BioLegend). To stain, cells were transferred from the 96-well plates to fluorescence activated cell sorter (FACS) tubes and pelleted by centrifugation. The desired antibodies were added directly to the cell pellet and samples were incubated for 30 minutes on ice. Cells were washed twice with protein blocking agent (PBA) buffer (1% w/v BSA + 0.05% w/v sodium azide in PBS) containing 2% v/v FBS, and then fixed with 0.5 ml fixation buffer (BioLegend) for 20 min at room temperature. Subsequently, cells were washed twice with 1 ml PBA. All centrifugation steps were carried out at 300 ×g for 5 min and samples were protected from light at all incubation steps. The final cell pellet was resuspended in 0.5 ml PBA and stored at 4°C. Fifty thousand cells were analysed for all samples on a Beckman Coulter FC500 flow cytometer. Collected data were processed using the Kaluza software (Beckman Coulter; version 2.1).

2.6 Bioinformatic tools

2.6.1 Codon optimisation

Codon optimisation was carried out using ATUM's (formerly DNA2.0) proprietary gene design algorithm, as published previously[103]. The algorithm has been optimised for *C. sporogenes* as part of the PhD project of Dr Aleksandra Kubiak, in collaboration with ATUM[104]. At the time of writing, the PhD thesis of Dr Aleksandra Kubiak was not

publicly available, but the relevant section (section 4.2.4, pp. 122-132) is reproduced in Appendix V.

2.6.2 Sequence data handling, plasmid map construction and visualisation

Routine sequence data handling was performed using SnapGene® (GSL Biotech LLC). The same software was used for plasmid map construction and visualisation.

2.6.3 Primer design

Primers were designed manually on SnapGene® version 3.1.4 (GSL Biotech LLC). Their characteristics were verified using the Eurofins and Sigma-Aldrich online platforms upon ordering.

2.6.4 Illumina sequencing data analysis

Illumina whole-genome sequencing data were assembled and analysed using the CLC Genomics Workbench version 10 (Qiagen). Resequencing analysis for the identification of single nucleotide variants was carried out using the Basic Variant Detection tool and the program's default parameters.

2.6.5 Sequence alignments and database searches

Gene and protein sequence alignments as well as database searches were carried out using the National Centre for Biotechnology Information (NCBI) BLASTn and BLASTp algorithms, respectively.

2.6.6 Phylogenetic tree construction

Multiple protein sequence alignments were carried out using Clustal Omega (EMBL-EBI)[105]. The aligned sequences were analysed for phylogeny using the PhyML version 3.0 software, accessed via www.phylogeny.fr, applying the program's default parameters; statistical analysis was performed using the program's inbuilt approximate likelihood-ratio test (aLRT) with nonparametric Shimodaira–Hasegawa (SH) correction

(SH-aLRT)[106, 107]. The results were generated in Newick format and visualised with the online phylogenetic tree viewer of the Environment for Tree Exploration (ETE) toolkit[108].

2.6.7 Promoter prediction

Promoter recognition was carried out using the BPROM online tool, provided by Softberry programs[109].

2.6.8 Terminator prediction

Rho-independent terminator sequences were predicted using the ARNold web server (Institute for Integrative Biology of the Cell, Université Paris-Sud – Bâtiment) available at rna.igmors.u-psud.fr/toolbox/arnold/index.php.

2.6.9 Origin of replication prediction

The chromosomal origin of replication of *C. sporogenes* NCIMB 10696 (CP009225) was determined using the DoriC database version 10 (TUBIC, Tianjin University, Tianjin, China)[110], with DoriC reference IDs ORI97011244 and ORI97011244.

2.6.10 Signal peptide prediction

The presence of signal peptides and location of signal peptide cleavage sites in protein sequences of interest were predicted using the SignalP 4.1 Server[111].The default parameters for Gram-positive organisms were applied in all cases.

2.6.11 Protease cleavage site prediction

The presence of putative protease cleavage sites in protein sequences of interest were predicted using the ExPASy PeptideCutter tool[112]. For analysis of the mature mGM-CSF protein sequence, the list of available enzymes was narrowed down to those of bacterial origin: Asp-N endopeptidase, Asp-N endopeptidase + N-terminal Glu,

Clostripain, Glutamyl endopeptidase, LysC, Staphylococcal peptidase I, and Thermolysin.

2.6.12 Prediction of transmembrane helices in proteins

The presence of transmembrane helices in given protein sequences was predicted using the TMHMM Server version 2.0[113].

2.6.13 Identification of conserved domains in proteins

Protein functional domains were identified using the NCBI conserved domain database and CD-search online tool[114, 115].

2.6.14 Data analysis and presentation

GraphPad Prism version 7.03 for Windows (GraphPad Software, La Jolla California USA, www.graphpad.com) was used for the design of all graphically presented results as well as all statistical analyses performed.

CHAPTER 3

Clostridial-Directed Enzyme Prodrug Therapy: investigation of suitable nitroreductase candidates

3.1 Background

3.1.1 Prodrugs as anticancer agents

Therapeutic success of cancer chemotherapy relies on a wide therapeutic index which would permit killing of cancer cells without affecting normal tissues[116]. To improve the therapeutic indices of anticancer treatments, the utilisation of prodrugs has been considered. Prodrugs are precursors that lack intrinsic activity, but can be converted to active and potent drugs *in vivo*, at a desired location and/or time[117]. In the context of cancer, prodrug activation can be confined to cancer cells by leveraging unique physiological characteristics of tumours, such as the presence of hypoxia.

Accordingly, the development of hypoxia-activated prodrugs (HAPs; also known as bioreductive prodrugs) has been actively pursued, counting more than 40 years of preclinical and clinical research[118]. The most clinically advanced HAP is TH-302 (evofosfamide), which was recently evaluated in Phase III clinical trials in advanced pancreatic cancer (NCT01746979) and soft tissue sarcoma (NCT01440088), in combination with conventional chemotherapy[119]. Typically, HAPs encompass five main chemical classes: nitro(hetero)cyclic compounds, quinones, aromatic N-oxides, aliphatic N-oxides and transition metal complexes[120]. Although the exact mechanism of prodrug activation varies between the different classes, selective activation under hypoxic conditions occurs via enzymatic reduction by endogenous oxygen-sensitive one-electron oxidoreductases, generating prodrug free radical species which undergo further processing to yield cytotoxic, DNA-damaging derivatives. In healthy, normoxic tissues, prodrug activation is inhibited because the free radical intermediate is immediately re-oxidised (also preventing the generation of downstream products), in what is known as 'futile redox cycling'[21]. Alternatively, bioreductive prodrugs can be activated by two-electron reductases, which can bypass the formation of the oxygen-sensitive free radical intermediate[120]. Since these enzymes are not inhibited by oxygen, selective prodrug activation is founded on the specific localisation or higher abundance of these enzymes in tumours[118], whilst aerobic prodrug activation elsewhere in the body may lead to 'off-target' activity and toxic side effects[121].

In both cases, treatment efficacy is reliant on tumour expression of the appropriate activating one- or two- electron reductases[21]. However, expression levels may be insufficient for prodrug activation and/or vary between different cancers, thus narrowing the applicability of HAPs[120, 122]. Additionally, prodrug reduction requires the presence of active tissue reductases, implying that in the presence of concurrent necrosis, prodrug activation will be limited.

A solution to overcoming some of these limitations is the exogenous provision of suitable prodrug-converting enzymes (PCEs), via gene- (GDEPT) or antibody- (ADEPT)

directed enzyme prodrug therapies. GDEPT employs either viral (in which case it is also known as VDEPT) or non-viral vectors that introduce the desired PCE gene sequence in cancer cells. In ADEPT, the PCE protein product is fused to a monoclonal antibody targeted towards a tumour-specific antigen[123]. These strategies may be applicable to a large repertoire of enzyme/prodrug combinations, extending beyond bioreductive prodrugs and their cognate PCEs[124]. Notable examples of enzyme/prodrug combinations that have also been explored clinically include the pairing of cytosine deaminase with 5-fluorocytosine[125] and of herpes simplex virus thymidine kinase (HSV1-TK) with ganciclovir[126].

Harnessing the selectivity of Clostridia for hypoxic tumour regions and their amenability to be engineered for the expression of heterologous genes, a further variation of DEPT has been born, baptised 'Clostridial-Directed Enzyme Prodrug Therapy' (CDEPT)[81]. CDEPT may be a superior alternative to GDEPT given the advantages that bacterial vectors offer over viral or non-viral methods; viral vectors are associated with sometimes impeding immunogenic and toxic traits, while the better tolerated non-viral vectors often suffer from low transfer efficacies[127, 128]. Importantly, CDEPT does not rely on target cells to express the desired PCE, implying that tumour necrosis, if present, would not compromise therapy effectiveness. CDEPT is also devoid of common ADEPT-related drawbacks, such as poor penetration of antibodies in the centre of the tumour, need for exclusive expression of target antigens by tumour cells, antigenic heterogeneity in tumours, invoked immunogenicity, slow clearance from the body and high costs of production[123]. Beneficially, the incorporation of a PCE in the clostridial vector and its pairing with a suitable prodrug may contribute in overcoming the limitations of incomplete clostridial oncolysis.

3.1.2 Clostridial-Directed Enzyme Prodrug Therapy (CDEPT)

A diagram of CDEPT is shown in Figure 3.1. Following selective spore germination in the tumour, expression of the PCE is highly localised. The prodrug is then systemically administered, but its activation and subsequent anticancer activity take place only inside the tumour where it has encountered the PCE[53]. As with all enzymatic reactions, one molecule of PCE can metabolise large quantities of prodrug, which coupled with bacterial replication in the tumour, can amplify treatment efficacy enormously[129]. CDEPT, by definition, falls under the category of suicide gene therapies[130].

Similarly to other DEPT strategies, the therapeutic success of CDEPT depends heavily on the PCE/prodrug combination. Firstly, the PCE should ideally be absent or otherwise insignificantly expressed in healthy tissues or exclusively expressed in the tumour[131]. Additionally, it should be sufficiently active at body temperature and neutral pH[132]. Importantly, it should be tolerated and expressed well in the desired clostridial host, to maximise the derivable antitumour activity[86].

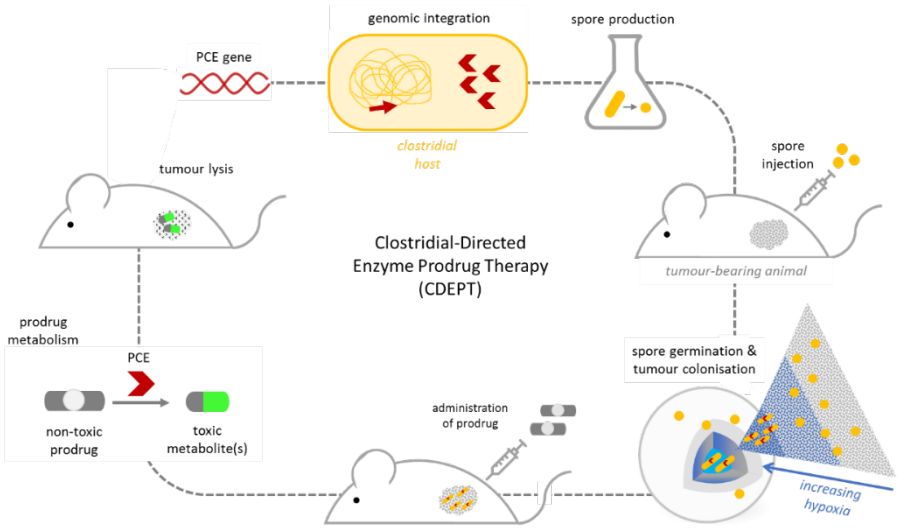


Figure 3.1 Clostridial-Directed Enzyme Prodrug Therapy (CDEPT). From top left corner (clockwise): a gene encoding a suitable prodrug-converting enzyme (PCE) is integrated into the bacterial chromosome. A spore stock of the recombinant cells is prepared, and an appropriate spore dose is administered intravenously into a tumour-bearing animal. The spores germinate only in the hypoxic environment of the tumour; the resulting vegetative cells replicate locally and colonise the tumour. A non-toxic prodrug is then administered systemically. Its conversion into an active drug occurs only within the tumour whereby the PCE is being expressed by the vegetative clostridial cells. The active drug leads to cancer cell death and tumour regression.

As far as the prodrug is concerned, it should have optimal physicochemical properties (such as solubility and stability) with good extravascular penetration into tumours and diffusibility into bacterial cells (if the enzyme is not secreted). A high differential toxicity between the prodrug and its active metabolites is also necessary[131, 132]. Simultaneously, active metabolites should be capable of intercellular diffusion, so that tumour cells in the wider vicinity of clostridial vectors can be targeted. The phenomenon of diffusion-dependent transfer of prodrug cytotoxicity to metabolism-naïve neighbouring cells is defined as the 'bystander effect' and, in an ideal embodiment of CDEPT, it can compensate for a possible heterogenous distribution of clostridial cells within the tumour[131]. Finally, the PCE/prodrug combination should be kinetically favourable, allowing for a high k_{cat} (the turnover number) and low K_M (measure of the

enzyme's affinity for the substrate), so that metabolite production is maximised at any concentration of prodrug[131, 132].

The first enzyme-prodrug combination implemented in CDEPT was the pairing of cytosine deaminase with 5-fluorocytosine; cytosine deaminase catalyses the conversion of 5-fluorocytosine into 5-fluorouracil which is then further metabolised into inhibitors of RNA and DNA biosynthesis[40, 82, 84]. Nevertheless, this combination was gradually relinquished due to the development of tumour resistance against 5-fluorouracil, a previously known shortcoming[25]. As a result, scientific interest shifted predominantly towards systems that employ nitroreductases as the PCEs.

3.1.3 Nitroreductases as the prodrug converting enzymes

Nitroreductases represent a large family of nicotinamide adenine dinucleotide (phosphate) [NAD(P)H]-dependent flavoenzymes that can catalyse the reduction of nitro groups. They have a broad substrate specificity and can accommodate a great variety of polynitrated aromatic compounds, usually generating hydroxylamino or amino derivatives as terminal reduction products. Bacterial nitroreductases are classified on the basis of their oxygen sensitivity into type I (oxygen-insensitive) or type II (oxygen-sensitive), the main difference lying on their catalytic mechanism. Type I nitroreductases perform catalysis via a two-electron sequential mechanism whereas type-II nitroreductases via a single-electron reduction. Type-I nitroreductases are further classified in two groups, represented by the *E. coli* NfsA (group A or 'major') and NfsB (group B or 'minor'). The group A family usually prefers NADPH as the cofactor whereas the group B family may use both NADH and NADPH[133].

In the context of directed enzyme prodrug therapies, nitroreductases have been traditionally paired with CB1954 [tetrazicar; 5-(aziridin-1-yl)-2,4-dinitrobenzamide], a dinitrobenzamide prodrug whose nitroreductase-mediated activation leads to the production of 2- (2HX) and 4-hydroxylamine (4HX) derivatives (Figure 3.2). The 4HX derivative, a bifunctional alkylator, is 10,000-fold more toxic than CB1954 and is further

metabolised to form inter-strand crosslinks in DNA which are poorly repaired, leading to cell death independently of the cell cycle stage[134, 135]. The 2HX product despite being less cytotoxic may also contribute to tumour cell killing because it diffuses readily and displays a more pronounced bystander effect[136].

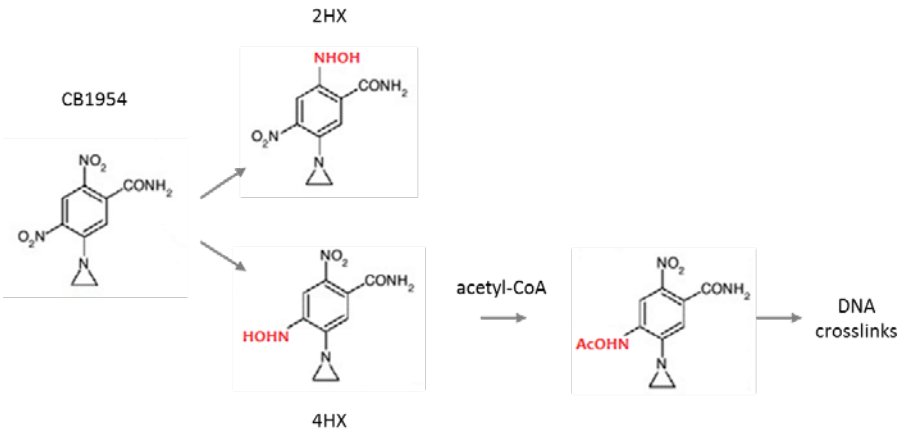


Figure 3.2 Mechanism of CB1954 activation. CB1954 can undergo reduction at the 2-nitro position to generate the 2-hydroxylamine derivative (2HX) and at the 4-nitro position to generate the 4-hydroxylamine derivative (4HX). 4HX may subsequently react with acetyl-CoA to form DNA crosslinks which lead to cell death.

Although clostridial cells natively express a range of nitroreductases, it has been shown that the level of endogenous nitroreductase activity is not adequate to metabolise prodrugs of interest efficiently and generate significant anticancer effects[43, 49, 86], hence necessitating engineering of the clostridial host to (over)express a desired enzyme. *E. coli* NfsB, also known as EcoNTR (WP_000351487.1), was the first bacterial nitroreductase employed for CDEPT, expressed in *C. beijerinckii* and *C. sporogenes* and paired with CB1954[85, 86]. Subsequent studies focused on identifying kinetically superior nitroreductases with respect to CB1954 metabolism, leading to the expression of HinNTR from *Haemophilus influenzae* (WP_005694521.1)[49] and NmeNTR from *Neisseria meningitidis* (WP_002225416.1)[43] in *C. sporogenes*, as alternative NfsB-type nitroreductases with improved CB1954 kinetic parameters and exclusive generation of

the 4HX cytotoxic metabolite. Importantly, the affinity of NmeNTR for CB1954 (K_M = 2.47 μM) was estimated to be below[43] the maximum concentration of CB1954 that can be safely tolerated in a patient's serum (6.3 μM)[137, 138], indicating that enzyme saturation and efficient prodrug activation would be feasible in a clinical setting. This is contrary to most (characterised) nitroreductases to date, with K_M values generally higher than 6.3 μM[139].

Advantageously, incorporation of nitroreductases in the clostridial vector can increase the sensitivity of the strain towards antibiotics routinely used for its clearance (for example, metronidazole), therefore providing an advantageous biosafety control mechanism[140].

3.1.4 Nitroreductase-activated prodrugs

Despite CB1954 being the mainstay prodrug for DEPT strategies with nitroreductases, human clinical trials in cancer patients have revealed dose-limiting hepatotoxicity and diarrhoea associated with its use[137]. Furthermore, subsequent toxicological studies have confirmed metabolism of the prodrug by endogenous human liver enzymes[141]. These findings together with the desire for a proprietary prodrug with protected intellectual rights – as required for the downstream commercialisation of a clinical strain – have led to the exploration of suitable alternatives.

Research has branched towards the discovery of novel prodrugs which are either safer and/or a better match with candidate nitroreductases; amongst these is the family of dinitrobenzamide (DNBM) mustards. These compounds were originally developed as hypoxia-activated cytotoxins, being substrates for oxygen-sensitive human oxidoreductases[120, 142]. Their pairing with type-I nitroreductases, potentially via DEPT strategies, removes the requirement of hypoxia and offers greater flexibility without compromising tumour-specificity[139]. Activation of the latent mustard moiety of DNMB mustards generates extremely potent and stable DNA crosslinking metabolites[143], with pronounced bystander effects[144]. Development of an extensive series of DNBM mustard

analogues (classes A-G) has yielded compounds with improved potency and physicochemical properties[145, 146]. Of interest is PR-104, the lead compound from class B, which based on its preclinical efficacy and overall pharmacokinetic profile[147], was subjected to clinical evaluation in patients with solid tumours[148, 149]. PR-104 is also a good substrate for EcoNTR and currently the only prodrug apart from CB1954 which has been tested in combination with an EcoNTR-armed clostridial strain; in fact, it was more efficient than CB1954 in generating sustained tumour reduction in mice with SiHa xenografts, even at comparatively lower administered doses[86]. Unfortunately, findings from Phase I/II clinical trials employing PR-104 have cast doubts on the clinical utility of this prodrug; dose-limiting myelotoxicity[148, 149] attributed to prodrug metabolism by the human aldo-keto reductase 1C3 (AKR1C3)[121] as well as high levels of clearance[150] were common shortcomings responsible for stalling the clinical implementation of PR-104, inside or outside CDEPT. Consequently, subsequent research has focused on rectifying the PR-104 scaffold towards AKR1C3-resistant analogues[151].

3.1.5 Imaging of clostridial vectors

The ability to monitor the *in vivo* journey of gene therapy vectors (including clostridial ones) in a non-invasive manner has emerged as an important requirement to expedite their preclinical and clinical development[152, 153]. Vector imaging can provide real-time, spatiotemporal information on the distribution of bacteria post-systemic administration, thus enabling monitoring of treatment specificity and efficacy[154]. Importantly, with regards to clostridial therapies in particular, imaging can shed light on the extent of tumour necrosis/hypoxia required for colonisation and, in turn, the degree of colonisation required for therapeutic efficacy. Relevant information is currently incomplete, being inferred from sub-optimal methods such as immunoblotting, Gram-Twort staining and the determination of CFUs per gram of tumour homogenates[26]. In a clinical setting, the imaging functionality can be repurposed towards non-invasive tumour detection, cancer diagnosis and stage classification as well as towards informing

the selection/timing of treatments (such as the administration of a prodrug) and monitoring their effectiveness[17, 26].

In principle, the imaging functionality can be realised by harnessing native genes of the clostridial host[155, 156] or via genetic engineering of the host to express suitable reporter genes[86]. In both cases, detection of the corresponding protein products (and therefore of bacterial cells) is enabled by application of an exogenous imaging probe which is subsequently visualised by the appropriate imaging technology[153]. Evidently, the second approach offers more flexibility with respect to the compatible imageable probes and, consequently, the imaging technique that can be employed. In an ideal embodiment of CDEPT, the incorporated nitroreductase can mediate not only the DEPT functionality but also the imaging one, by being capable of activating the desired prodrug and imaging probe alike[140]; this would enable imaging of the clostridial cell, but, at the same time, direct monitoring of the therapeutic functionality[139]. To this end, proof-of-concept has been demonstrated with an EcoNTR-expressing *C. sporogenes* strain in combination with 6-chloro-9-nitro-5-oxo-5H-benzo[a]phenoxazine (CNOB), whereby CNOB metabolism to a fluorescent derivative coincided spatially with nitroreductase expression in necrotic tumour regions[86]. Alternative nitroreductase-activated probes for optical imaging include the quenched infra-red CytoCy5s[154] and a bioluminescent nitroreductase-caged luciferin (NCL)[157].

Nevertheless, there are currently no nitroreductase-compatible probes that would enable imaging via molecular techniques which offer higher sensitivity and are more suited for clinical applications, such as positron emission tomography (PET)[158]. Therefore, work is underway with [^{18}F]-labelled 2-nitroimidazole (2-NI) PET imaging probes which are being repurposed from hypoxia-activated to nitroreductase-activated equivalents[139]. This could accelerate the translation of CDEPT therapies into clinical trials, given that PET tracers of the 2-NI family have been researched in an oncology context for several years. They are already in preclinical and/or clinical development, implying that considerable time and costs can be saved via this 'piggyback' approach[159, 160].

3.1.6 An international CDEPT collaboration

The various components of CDEPT render it a highly interdisciplinary project. As such, it is currently being pursued via an international collaboration between the University of Nottingham (Clostridia), the University of Auckland (UOA; prodrugs), the Victoria University of Wellington (VUW; PCEs) and the University of Maastricht/Maastro clinic (UM; PET imaging). The collaboration is expected to generate clinical trial-ready CDEPT strains and see them through Phase I.

3.1.6.1 Prodrug selection

Presently, the top prodrug candidate for CDEPT is a novel, patented (patent number WO2014031012A1)[161] AKR1C3-resistant DNBM analogue, CP-506 (previously known as DTP-006 and SN36506), with desirable physicochemical and pharmacokinetic properties, including excellent aqueous solubility and clinically relevant oral bioavailability[162]. CP-506 can be dually activated by hypoxia and bacterial type-I nitroreductases (Figure 3.3). It is currently licensed to Convert Pharmaceuticals for clinical development in the context of hypoxia, while preclinical studies to date have generated encouraging data (clinical trials to commence early 2019)[162-164].

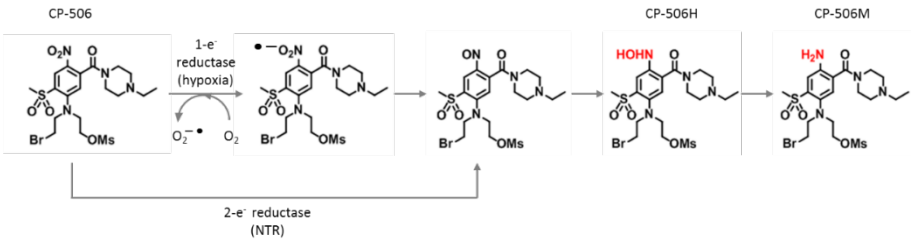


Figure 3.3 Mechanism of CP-506 activation. CP-506 can undergo hypoxia-dependent reduction by one-electron reductases or oxygen-independent reduction by two-electron reductases, such as bacterial nitroreductases (NTR), to produce the cytotoxic amine (CP506-M) and hydroxylamine (CP506-H) species.

3.1.6.2 PET probe selection

The imaging probes of interest are patented (patent number WO2014007650A1)[165], repurposed 2-NI PET tracers, as discussed previously. Specifically, [^{18}F]EF5[166] and [^{18}F]EF1[167] have been optimised to minimise their responsiveness to hypoxia and hence improve the signal of nitroreductase imaging. More specifically, interest is focused on the 5NI-EF5 (SN33623)[168] and 5NI-EF1 (SN37819) (unpublished data; VUW/UOA) analogues (Figure 3.4). It should be noted that EF5 and its analogues require hot $^{18}F_2$ gas for radiolabelling which is not readily available in hospital settings, thus constituting less attractive choices (personal communication; A. Mowday).

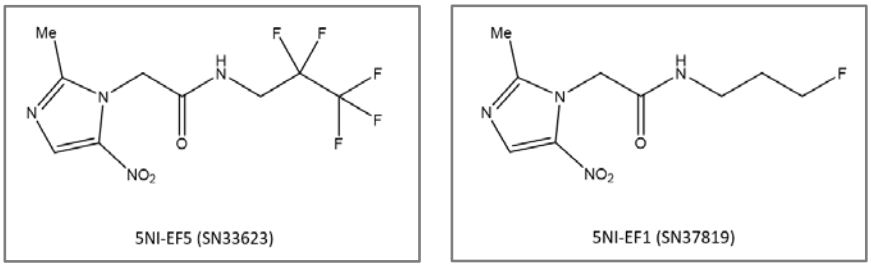


Figure 3.4 The lead PET probe candidates

3.1.6.3 Nitroreductase selection

With a clinical-stage prodrug candidate and promising PET probes at hand, the challenge lies in identifying suitable nitroreductases that can co-activate these substrates as well as be tolerated and expressed at high levels in *C. sporogenes.* To this end, candidates were selected for dual activity towards CP-506 and 5-NI PET probes from an extensive gene library of bacterial oxidoreductases (>58), (developed as described in [169] and expanded thereof [140]), including NfsA and NfsB orthologs as well as engineered variants. Initial screening was carried out as described previously[169], in a nitroreductase-null *E. coli* SOS reporter strain via growth inhibition assays, followed by a series of post-hoc quantitative IC_{50} analyses for the lead candidates. Subsequent steady-state kinetic evaluations towards CP-506 were carried out with purified enzymes.

All aforementioned studies were designed and conducted by Dr Dave Ackerley's team at VUW.

It should be noted that the selection process was iterative, whereby early data on nitroreductase expression in *C. sporogenes* were considered for the selection/engineering of supplementary candidates. Additionally, previous work carried out by Dr. Aleksandra Kubiak (unpublished data) suggested that NfsB-type nitroreductases are more amenable to functional expression in this strain; therefore, attention was given to select top candidates from both families, even though NfsA representatives are generally better metabolisers of the substrates in question. Likewise, YfkO_Bl, a candidate previously considered for CB1954[170], was selected independently based on preliminary work suggesting it is expressed well in *C. sporogenes*.

In total, seven candidates were taken forward for implementation in *C. sporogenes*, encompassing five NfsB-type and two NfsA-type nitroreductases. The variants NfsB_Vv F70A/F108Y (NfsB_Vv_dm) and YfkO_Bl F82L/M110I (YfkO_Bl_362) were rationally engineered to have superior activity with 5-NI PET probes compared to their parent proteins, with retained (or improved) activity towards CP-506. The variant NfsB_Vv S40A/V41T/F70A/N71Q/F108Y/F124V (T500_4) contains four additional mutations with respect to NfsB_Vv_dm which render it more active towards CP-506. The selected candidates are shown in Table 3.1, alongside relevant characterisation data towards the key CDEPT substrates in Table 3.2. A phylogenetic tree of the selected nitroreductases alongside the NfsA_Ec (WP_000189159.1) and NfsB_Ec (WP_000351487.1) representatives from *E. coli* is shown in Figure 3.5, constructed as described in section 2.6.6.

Table 3.1 The selected nitroreductase candidates

Nitroreductase	Originating organism	NCBI reference	Type	Other names
YfkO_Bl	*B. licheniformis*	WP_003179779.1	NfsB (B)	
YfkO_Bl F82L/M110I	*B. licheniformis*	WP_003179779.1[†]	NfsB (B)	YfkO_Bl_362
YdgI_Bs	*B. subtilis*	WP_003225379.1	NfsB (B)	NfsB_Vv_dm
NfsB_Vv F70A/F108Y	*V. vulnificus*	WP_011081872.1[†]	NfsB (B)	
NfsB_Vv S40A/V41T/F70A/N71Q/F108Y/F124V	*V. vulnificus*	WP_011081872.1[†]	NfsB (B)	T500_4
NfrA_Bs	*B. subtilis*	WP_003222161.1	NfsA (A)	
YcnD_Bs	*B. subtilis*	WP_003225013.1	NfsA (A)	

[†] engineered versions of the referenced protein

Table 3.2 Characterisation of nitroreductase candidates towards key substrates

	IC_{50} (μM)			CP-506		
Nitroreductase	CP-506	5NI-EF1	5NI-EF5	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1} s^{-1}$)
NmeNTR[†]	>1000	>250	>150	no activity	no activity	0
YfkO_Bl	>1000	>500	>150	no activity	no activity	0
YfkO_Bl_362	≈1000	43	27	ND	ND	ND
YdgI_Bs	425	35	51	0.29 ± 0.08	400 ± 130	700
NfsB_Vv_dm	478	46	70	2.5 ± 0.3	1897 ± 281	1310
T500_4	407	110	70	0.30 ± 0.01	678 ± 57	440
NfrA_Bs	404	30	34	2.3 ± 0.1	110 ± 20	21000
YcnD_Bs	320	8	17	0.29 ± 0.06	250 ± 40	1200
negative control	>1000	>250	>150			

[†] historical candidate, chosen previously as a CB1954 metaboliser[43]; ND: not determined

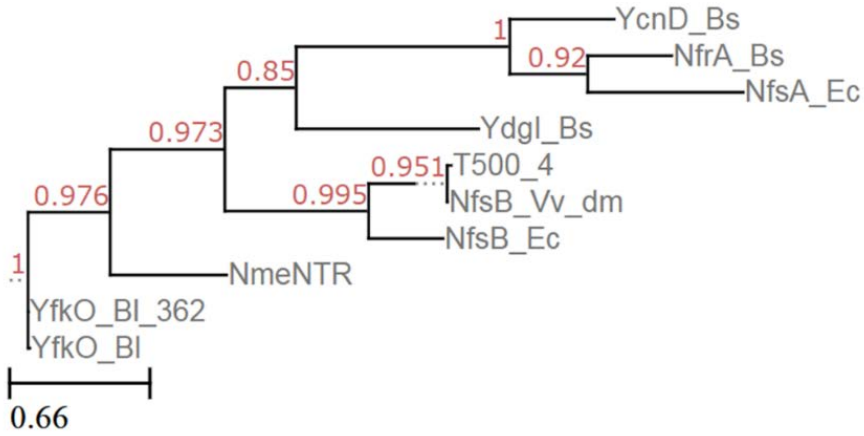


Figure 3.5 Phylogenetic tree of the selected nitroreductase candidates. Maximum likelihood phylogram of the selected nitroreductase candidates alongside NfsA_Ec and NfsB_Ec, constructed using PhyML; node labels in red indicate aLRT SH-like branch support values (>0.90 indicates statistical significance).

3.2 Objectives

- Create a library of *C. sporogenes* strains expressing selected nitroreductase candidates, suitable for downstream *in vivo* studies.
- Conduct a preliminary characterisation of the created strains with regards to nitroreductase expression, growth and sporulation characteristics, to inform subsequent decisions, in coordination with international collaborators.

3.3 Hypothesis

- *C. sporogenes* NCIMB 10696 can express and tolerate any nitroreductase of bacterial origin, without a significant impact on the phenotypic characteristics of the wildtype strain.

3.4 Study design

The nitroreductase-expressing strains were purpose-built for their prospective implementation in a preclinical/clinical setting. Firstly, the nitroreductase genes were integrated in the genome of *C. sporogenes* to ensure stable gene expression without the use of antibiotic resistance markers. Secondly, the constructs were designed for constitutive nitroreductase expression, eliminating the requirement of exogenous administration of an inducer and allowing the simplification of associated trial protocols. Finally, the enzymes were introduced in a non-secreted form, following suit from previous nitroreductase-based CDEPT studies[43, 49, 85, 86], under the assumption that CP-506, based on its lipophilicity (LogP 1.3), can diffuse in the bacterial cell whilst its metabolites can diffuse out.

The gene sequences of selected nitroreductase candidates (Table 3.1), originating from various bacterial strains, were codon-optimised for *C. sporogenes* to eliminate codon bias and ensure efficient translation in this host. Codon optimisation was carried out according to section 2.6.1/Appendix V and the codon-optimised sequences can be found in Appendix VI. In all constructs, nitroreductase expression was driven by a benchmark promoter/ribosome binding site (RBS) combination (P_{thl14}), comprising a shortened version of the thiolase promoter from *C. acetobutylicum* (P_{thl}) and an optimised RBS variant functional in *C. sporogenes* (RBS14). The P_{thl14} nucleotide sequence is shown in Appendix VII. P_{thl14} was previously shown to be very efficient for the expression of NmeNTR[104] and has since been the expression signal of choice for investigating nitroreductase candidates for CDEPT within our group. For comparison purposes, characterisation of the newly constructed nitroreductase-expressing strains was carried out alongside the flagship Csp-Δ*pyrE*::P_{thl14}-*nmeNTR* strain, generated prior to this study.

3.5 Results

3.5.1 Construction of nitroreductase-expressing strains

At the outset of this project, the genetic tool traditionally employed to achieve integration of desired fragments in clostridial genomes was the method of allele-coupled exchange (ACE)[75].

ACE is based on two principles: (i) the integration event yields a selectable phenotype which allows identification of successfully edited cells, and (ii) integration occurs in a two-step process, consisting of two homologous recombination events, the order of which is dictated by homology arms of asymmetric lengths. The first event (single-crossover) occurs at the long homology arm (1200 bp) and results in integration of the entire plasmid. The second event (double-crossover) is mediated by the smaller homology arm (300 bp) and results in plasmid excision, concomitantly with acquisition of the selectable phenotype and integration of the desired insert. The selectable phenotype may be acquired by either inactivation of a counter-selectable gene or by activation of a heterologous positively-selectable marker – for instance, by linking the expression of an antibiotic resistance gene (a promoterless copy of which is originally supplied on the plasmid) to a strong chromosomal promoter. Directing the order of the events via the provision of asymmetric homology arms is crucial in allowing the acquisition of the selectable phenotype to be coupled to the second recombination event, namely the formation of the desired double-crossover mutant.

The configuration of ACE employed here yields a selectable phenotype via inactivation of the *pyrE* gene, as previously described[43, 75]. Truncation of the *pyrE* gene eliminates the ability of the cells to metabolise 5-fluoroorotic acid (5-FOA) into the toxic metabolite 5-fluorouracil, allowing mutant cells to be selected in media supplemented with 5-FOA. Simultaneously, in the absence of a functional *pyrE* gene, cells become auxotrophic for uracil; this serves as an additional safety feature for the clostridial

vector, given that uracil is abundant at the site of the tumour but not readily found in other environments[43]. Consequently, the inactivated *pyrE* allele of the nitroreductase-expressing strains is not repaired after the integration event, contrary to common practice.

The genomic integration of nitroreductase candidates in *C. sporogenes* was carried out as described in section 2.2.12.1. A detailed diagram of the method is shown in Figure 3.6. Integration plasmids were constructed according to section 2.2.11.3 with the general vector design displayed in Appendix IV. The standard ACE vector (pMTL-JH27) employed previously[43, 75], was modified according to section 2.2.11.2, in order to extend the length of the right (long) homology arm; this was done with the aim of overcoming difficulties in obtaining and maintaining colonies that had undergone the single crossover event. In addition, transconjugant colonies were selected on higher concentrations of thiamphenicol (30-50 μg/ml *vs* 15 μg/ml). Given the segregational instability of the pIM13 replicon employed, this modification was expected to intensify the growth advantage of single crossover colonies *vs* colonies where the antibiotic resistance marker was on the plasmid. The above adjustments yielded a modest improvement in integration efficiency and allowed all desired constructs to be integrated successfully.

Genomic integration in the *pyrE* locus was confirmed via colony PCR with flanking chromosomal primers for three biological replicates per nitroreductase construct (Figure 3.7)[a]. The plasmid was lost from selected colonies, as described in section 2.2.13. All integrants were subsequently verified by sequencing.

[a] The P_{thl14}-YfkO_Bl, P_{thl14}-NfrA_Bs and P_{thl14}-YcnD_Bs constructs were integrated by Lissa Sutherland (MSc student) concomitantly with the present study

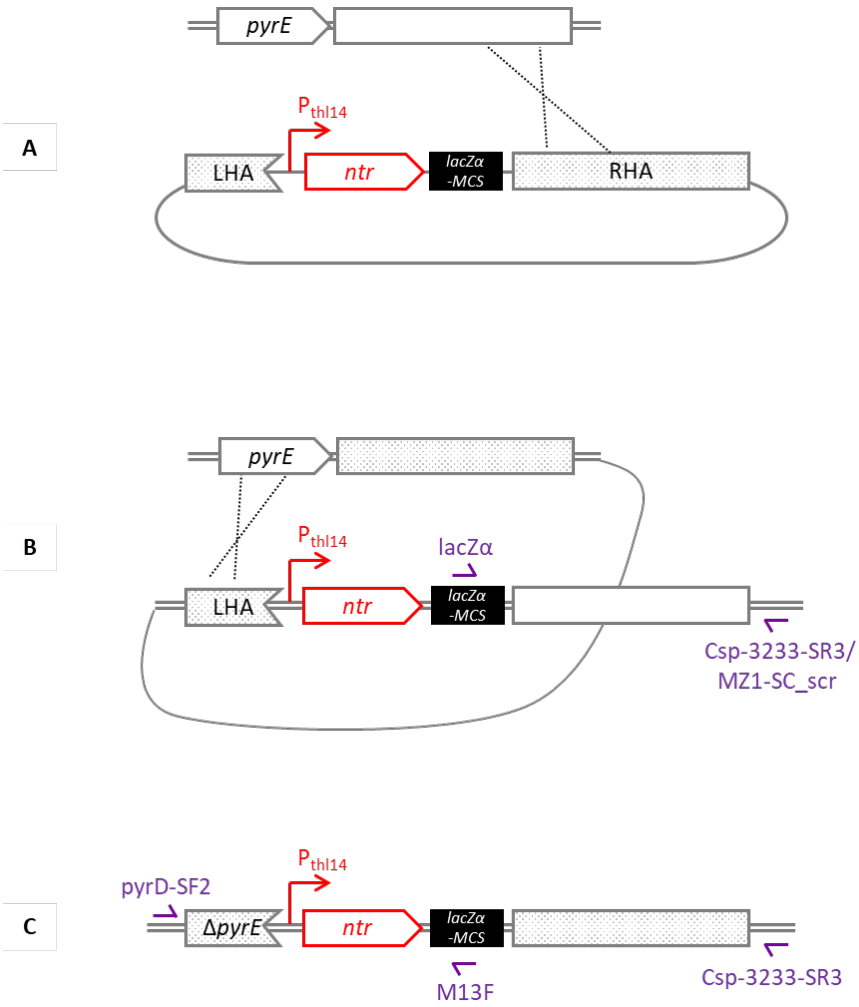


Figure 3.6 Schematic illustration of allele-coupled exchange (ACE). **A.** The construct of interest, P_{thl14}-ntroreductase (*ntr*), is cloned upstream of a *lacZα* multiple cloning site (*lacZα*-MCS), flanked by two arms homologous to genomic regions. Integration occurs in a two-step process consisting of two homologous recombination events. The first recombination event (single-crossover) is mediated by a long region of homology between the vector and 1200 bp (for pMTL-JH27) or 2500 bp (for pMTL-MZ1) immediately downstream of *pyrE* in the bacterial genome (right homology arm, RHA). **B.** This results in genomic integration of the entire plasmid. The single crossover event can be confirmed via colony PCR with primers lacZα and Csp-3233-SR3 or MZ1-SC_scr for pMTL-JH27 and pMTL-MZ1, respectively. The second recombination event (double-crossover) is mediated by a short region of homology between the vector and 300 bp of the *pyrE* gene (left homology arm, LHA). **C.** A successful double crossover event stably integrates the desired sequence in the chromosome whilst excising the plasmid and 273 bp from the 3' end of *pyrE*. This partial deletion leads to inactivation of the gene (Δ*pyrE*), yielding a selectable phenotype by the means of 5-FOA resistance. The double crossover event can be confirmed via colony PCR using primers pyrD-sF2 and Csp-3233-SR3 and integration can be verified by sequencing with pyrD-sF2 and M13F.

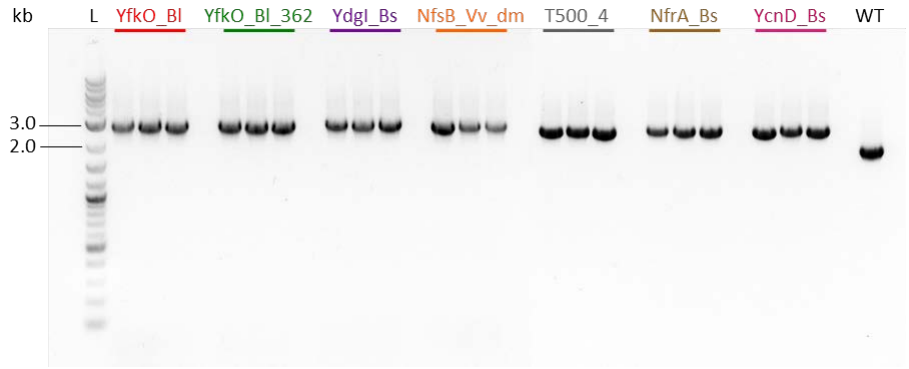


Figure 3.7 Genomic integration of nitroreductase candidates. Colony PCR for three biological replicates of each nitroreductase-expressing strain alongside WT. Bands correspond to products amplified with pyrD-sF2/Csp3233-SR. Csp-Δ*pyrE*::P_{thl14}-*yfkO_Bl* (2785 bp); Csp-Δ*pyrE*::P_{thl14}-*yfkO_Bl_362* (2785 bp); Csp-Δ*pyrE*::P_{thl14}-*ydgI_Bs* (2731 bp); Csp-Δ*pyrE*::P_{thl14}-*nfsB_Vv_dm* (2755 bp); Csp-Δ*pyrE*::P_{thl14}-*t500_4* (2755 bp); Δ*pyrE*::P_{thl14}-*nfrA_Bs* (2851 bp); Csp-Δ*pyrE*::P_{thl14}-*ycnD_Bs* (2851 bp); WT (2024 bp). L: NEB 2-Log DNA ladder.

3.5.2 Phenotypic characterisation of nitroreductase-expressing strains

It was necessary to establish that key characteristics of the wildtype (WT) strain, such as the growth, sporulation and germination capabilities, had not been compromised by the introduction of nitroreductase genes or the manipulations undertaken in the process.

3.5.2.1 Growth characteristics

Growth of all nitroreductase-expressing strains was followed over a period of 24 hrs relative to WT, as described in section 2.1.12.

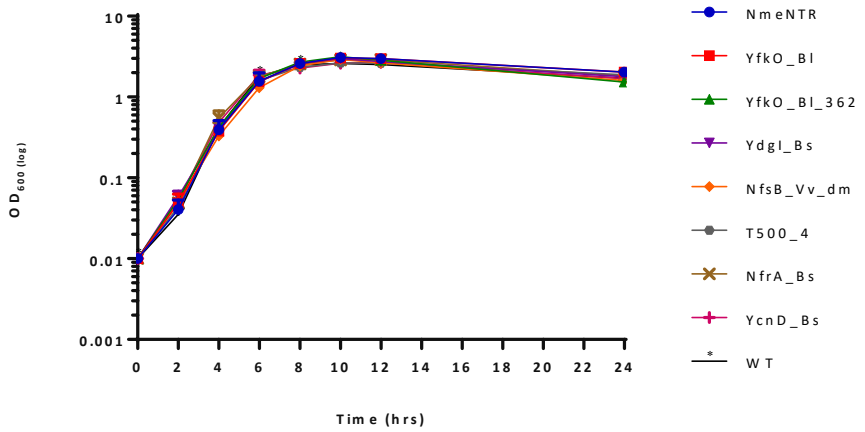


Figure 3.8 Growth curves of nitroreductase-expressing strains. Nitroreductase-expressing strains were grown in liquid TYG medium, alongside WT; OD_{600} was monitored over 24 hrs. Data represent mean values of three independent cultures for WT and of three biological replicates for all nitroreductase-expressing strains ± SD, from a single representative experiment out of three repeats.

All engineered strains were capable of attaining the same total cell density as WT and grew at an identical rate, characterised by a relatively short lag phase (2 hrs) and acquisition of stationary growth at approximately 7 hrs (Figure 3.8).

3.5.2.2 Sporulation and germination characteristics

Sporulation assays were performed as described in section 2.1.13 alongside the asporogenous Csp-Δ*spo0A* strain as a negative control. The sporulating/germinating efficiency of the nitroreductase-expressing strains was the same as WT at each time point assayed. Figure 3.9 shows representative data from the final (120-hr) timepoint, where the ratio of spores (heat-treated cells) relative to the total cell count (non heat-treated cells) is equal for all strains, indicating that all cells in the culture are capable of sporulation and germination. It should be noted that at this timepoint there is approximately a log-fold reduction in the total cell/spore titre for the YdgI_Bs- and NfsB_Vv_dm- expressing strains compared to WT. As expected, the negative control strain did not generate any spores.

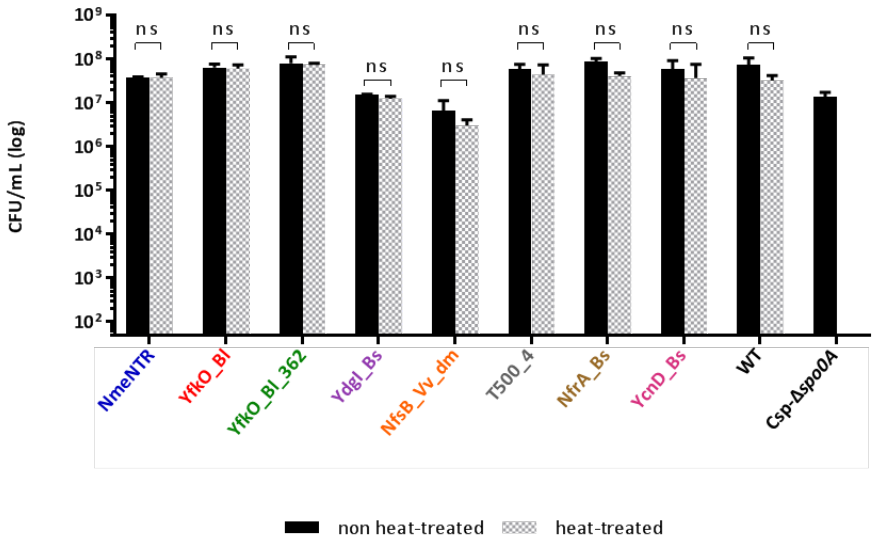


Figure 3.9 Sporulation/germination capacity of nitroreductase-expressing strains. Sporulation and germination capacity of nitroreductase-expressing strains *vs* WT, alongside an asporogenous strain (Csp-Δ*spo0A*) as the negative control, at 120 hrs of incubation; heat-treated CFUs correspond to the number of spores with germinating ability and non heat-treated CFUs to the total cell count. Data represent mean values of three independent cultures for WT and of three biological replicates for all nitroreductase-expressing strains ± SD, from a single representative experiment out of two repeats. The y-axis starts at 50 CFU/ml (detection limit). Statistical analysis was carried out using two-tailed unpaired t-test; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

3.5.3 *In vitro* characterisation of nitroreductase expression

Strong constitutive expression of the PCE in the clostridial host is highly desired to ensure that enzyme levels are not rate-limiting for prodrug conversion. Therefore, the ability of the constructed strains to express the selected nitroreductase candidates was assessed *in vitro*, via SDS-PAGE and enzymatic activity assays.

3.5.3.1 SDS-PAGE

Cell lysates of cultures harvested at the beginning of stationary growth (8 hrs) were subjected to SDS-PAGE analysis, as described in section 2.3.4. For all nitroreductase

candidates, a band of the expected size could be identified in their respective lysate, absent from the lysate of the WT strain (Figure 3.10). The intracellular abundance of the candidates varied, being the highest for YfkO_Bl, YfkO_Bl_362 and YdgI_Bs, followed by NfsB_Vv_dm, T500_4 and YcnD_Bs. Importantly, those candidates attained higher expression levels than the reference nitroreductase, NmeNTR. Conversely, the band for NfrA_Bs was hardly detectable, suggesting that expression of this candidate is not as efficient.

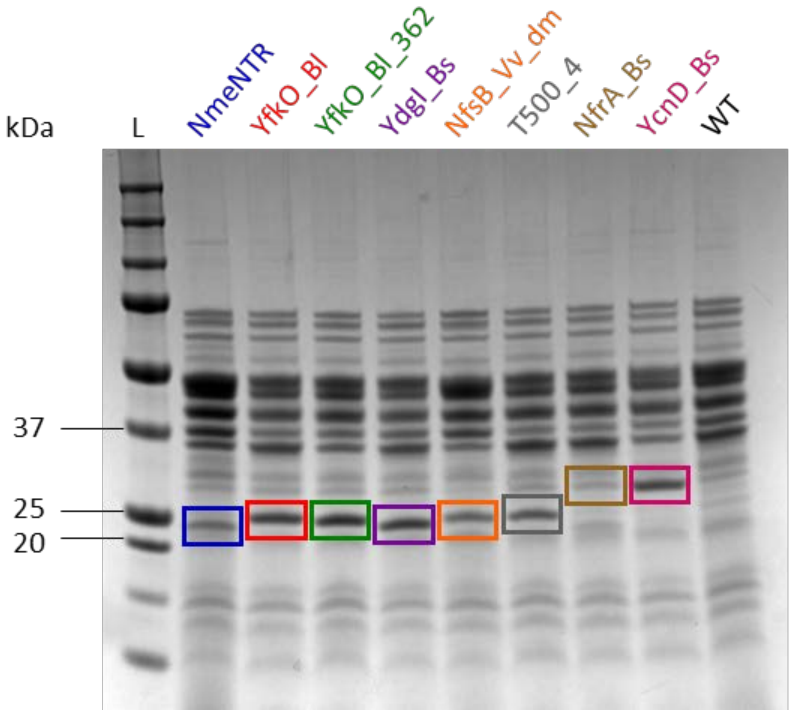


Figure 3.10 SDS-PAGE analysis on cell lysates of nitroreductase-expressing strains. Cell lysates were standardised to an equal total protein content and subjected to SDS-PAGE analysis. Protein bands corresponding to nitroreductases are boxed; NmeNTR: 24.7 kDa; YfkO_Bl: 26.2 kDa; YfkO_Bl_362: 26.1 kDa; YdgI_Bs: 23.4 kDa; NfsB_Vv_dm: 24.1 kDa; T500_4: 24.1 kDa; NfrA_Bs: 28.3 kDa; YcnD_Bs: 27.8 kDa; L: Bio-Rad Precision Plus Protein All Blue Standards ladder.

3.5.3.2 Enzymatic activity assays

Activity of nitroreductases in cell lysates was measured as menadione reductase (also referred as quinone reductase) and nitrofurazone reductase, according to sections 2.3.7-2.3.8, using either NADH or NADPH as a cofactor. Nitroreductase-mediated two-electron reduction of menadione yields its hydroquinone derivative (menadiol) which in turn drives the reduction of cytochrome c, generating a colour change detectable as an increase in absorbance at 550 nm[171, 172]. Likewise, nitrofurazone is reduced to its hydroxylamine derivative and the change can be monitored as a decrease in absorbance at 400 nm[173]. Menadione and nitrofurazone were chosen for their structural disparity to minimise potential substrate bias; menadione is a quinone and does not contain nitro groups while nitrofurazone is nitroaromatic. Specifically, NfsB has been previously reported to reduce menadione about twice as efficiently as NfsA, while NfsA has an up to 3-fold greater activity with nitrofurazone and related nitroaromatic substrates, compared to NfsB[174].

The results of the four screens are depicted in Figure 3.11 and summarised in Table 3.3. All engineered strains had statistically significantly higher enzymatic activity compared to WT in at least one of the four screens, confirming their cellular expression. Nevertheless, the rank order of the candidates fluctuated between the screens, highlighting substrate and cofactor preferences. In general, in 3/4 screens, the lysates of the strains expressing NfsB-like nitroreductases yielded higher activities than those expressing NfsA family representatives (NfrA_Bs and YcnD_Bs); an exception to this was T500_4, which is not a good metaboliser of either substrate (Figure 3.11A-C). NfrA_Bs and YcnD_Bs showed a strong preference (3- to 10- fold higher) for the NADPH cofactor with both substrates, consistent with the family's characteristics[133, 174]. With regards to NfsB family members, cofactor preference was less uniform; while all enzymes showed a modest preference for NADPH in combination with menadione (2- to 4- fold higher), NADH was the preferred cofactor for YdgI_Bs, NfsB_Vv_dm and T500_4 in combination with nitrofurazone (Table 3.3). Independently of the cofactor

employed, it appears that nitrofurazone is a better substrate for NfsA-like nitroreductases than menadione, as reported previously[175] (see Figure 3.11A-D; NfrA_Bs activity is statistically significantly increased relative to WT only in the nitrofurazone screens). Conversely, NfsB-like family members appeared to metabolise both substrates equally well.

Table 3.3 Rank order and cofactor preference of nitroreductase candidates in the enzymatic activity screens

	Rank order				Cofactor preference	
	Menadione		Nitrofurazone		NADPH/NADH ratio	
Nitroreductase	NADH	NADPH	NADH	NADPH	Menadione	Nitrofurazone
NmeNTR	5	4	5	1	3.7	3.9
YfkO_Bl	2	1	1	2	1.5	1.6
YfkO_Bl_362	3	3	4	3	2.3	2.2
YdgI_Bs	4	6	2	7	0.2	1.4
NfsB_Vv_dm	1	2	3	6	0.8	1.3
T500_4	8	8	7	8	0.3	1.9
NfrA_Bs	7	7	8	5	8.2	3.2
YcnD_Bs	6	5	5	4	5.2	9.7

Importantly, YfkO_Bl was consistently in the top 2 candidates in terms of enzymatic activity, in agreement with the SDS-PAGE, which shows high intracellular abundance of this enzyme. Comparably efficient were also the YfkO_Bl's engineered version, YfkO_Bl_362, and NfsB_Vv_dm, with high enzymatic activity in 3/4 screens and correspondingly high expression levels. For other candidates it was hard to draw a correlation between the SDS-PAGE analysis and the enzymatic activity screens; for instance, while YcnD_Bs appears to be expressed at a high level, its activity with the assayed substrates is underwhelming; even with nitrofurazone as the substrate and NADPH as the cofactor (the optimal combination for NfsA nitroreductases), its enzymatic activity was only 4th best. In line with this, T500_4 resulted in strikingly lower nitroreductase activities than its highly related NfsB_Vv_dm despite being

expressed equally well, indicating that the four additional mutations have entirely altered the enzyme's substrate range (rather than its expression level). As anticipated from the results of the SDS-PAGE analysis, NfrA_Bs yielded very low nitroreductase activities; nevertheless, its activity towards nitrofurazone with NADPH as the cofactor was 5th best amongst the set, confirming that some amount of NfrA_Bs is indeed present in the lysate, despite absence of an obvious NfrA_Bs band on the SDS-PAGE gel.

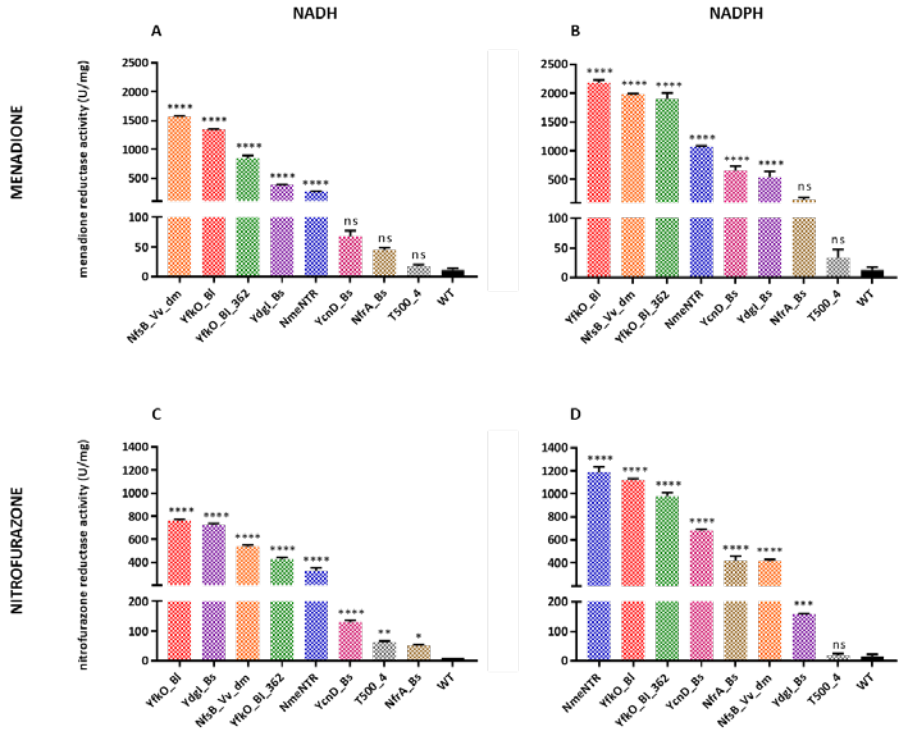


Figure 3.11 Enzymatic activity screens of the nitroreductase candidates. Menadione and nitrofurazone reductase activity in cell lysates of nitroreductase-expressing strains with NADH or NADPH as the cofactor, alongside the lysate of the WT strain as a negative control. **A.** menadione/NADH. **B.** menadione/NADPH. **C.** nitrofurazone/NADH. **D.** nitrofurazone/NADPH. Data represent mean values of three independent cultures for WT and of three biological replicates for all mutant strains, normalised for total protein content ± SD. Depicted data are from a single representative experiment out of more than three repeats. Statistical analysis was carried out using one-way ANOVA with Dunnett's test for multiple comparisons against WT; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

3.5.4 Troubleshooting the poorly expressed NfrA_Bs candidate

The low expression levels of NfrA_Bs in *C. sporogenes* were distinctly disappointing given that NfrA_Bs is currently the top candidate with respect to CP-506 metabolism, as evident from the kinetic parameters in Table 3.2. Despite anecdotal evidence indicating that *C. sporogenes* might be able to express type A nitroreductases less efficiently than type B ones (personal communication; A. Kubiak), it was of interest to investigate potential causes of this problem and explore ways in which expression levels NfrA_Bs could be improved. In the absence of obvious toxicities that could be associated with the specific candidate given the phenotypic conformism of the strain, the possibility was considered that P_{thl14} is not as efficient in the context of the *nfrA_Bs* nucleotide sequence or that codon optimisation might have had an aberrant impact on the translation efficiency of the protein. A further possibility might be that the protein could be aggregating, accounting for the low NfrA_Bs levels in the soluble lysate.

To investigate the above possibilities, *nfrA_Bs* driven by an alternative strong constitutive promoter, that of the ferredoxin gene from *C. sporogenes* (P_{fdx}), was integrated with its native RBS. The gene was additionally integrated as its native nucleotide sequence (from *B. subtilis*), driven again by P_{thl14}. Both integrations were carried out via ACE, as previously described, using the vectors pMTL-JH27-P_{fdx}-*nfrA_Bs* and pMTL-MZ1-P_{thl14}-*nfrA_Bs*-nat (Figure 3.12). Additionally, SDS-PAGE analysis was performed on the whole cell lysate, consisting of both the soluble and insoluble fractions; as a further precaution the insoluble debris obtained after isolation of the soluble lysate was processed for the purification of inclusion bodies that might have been present, as described in section 2.3.3.

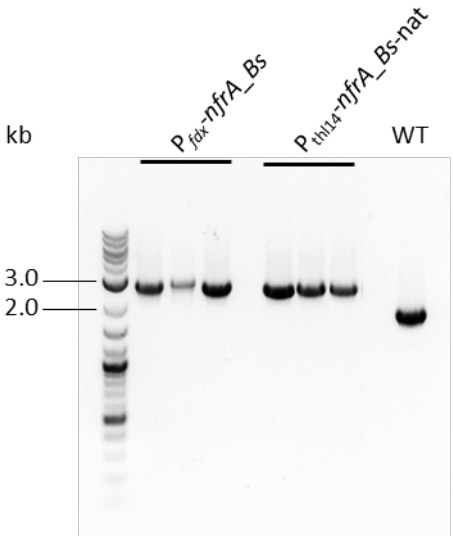


Figure 3.12 Genomic integration of NfrA_Bs construct variants. Colony PCR for three biological replicates of each NfrA_Bs-expressing strain alongside WT. Bands correspond to products amplified with pyrD-sF2/Csp3233-SR. Csp-Δ*pyrE*::P$_{fdx}$-*nfrA_Bs* (2905 bp); Csp-Δ*pyrE*::P$_{thl14}$-*nfrA_Bs*-nat (2851 bp); WT (2024 bp). L: NEB 2-Log DNA ladder.

Nitrofurazone reductase activity was similar in the lysates of the strains carrying the codon-optimised (co) and native (nat) gene sequences (Figure 3.13A) and no differences could be detected in the intensity of the corresponding bands on the SDS-PAGE gel (Figure 3.13B). Likewise, substitution of the promoter did not improve expression (Figure 3.13C). Finally, there was no difference in the band pattern or the band intensity between the three biological replicates of the NfrA_Bs-expressing strains and WT, neither in the whole lysate nor in the purified inclusion body samples, indicating that NfrA_Bs was present in low abundance altogether and not exclusively found in the insoluble fraction (Figure 3.13D).

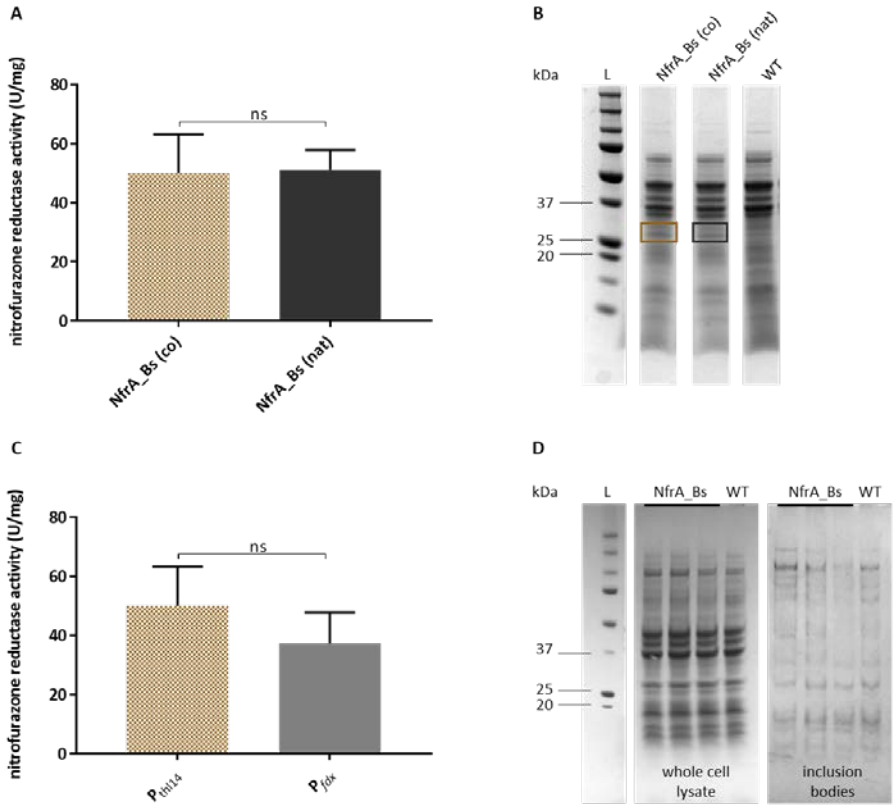


Figure 3.13 Troubleshooting the low expression levels of NfrA_Bs. **A.** Nitrofurazone reductase activity in cell lysates of NfrA_Bs-expressing strains carrying the codon-optimised (co) or native (nat) gene sequence. Data represent mean values of three biological replicates ± SD, from a single representative experiment out of two repeats. **B.** SDS-PAGE analysis on cell lysates of NfrA_Bs-expressing strains carrying the codon-optimised (co; brown box) or native (nat; black box) gene sequence alongside the lysate of the WT strain; NfrA_Bs: 28.3 kDa. L: Bio-Rad Precision Plus Protein All Blue Standards ladder. **C.** Nitrofurazone reductase activity in cell lysates of NfrA_Bs-expressing strains carrying the codon-optimised gene sequence, driven by P_{thl14} or P_{fdx}. Data represent mean values of three biological replicates ± SD, from a single representative experiment out of two repeats. **D.** SDS-PAGE analysis on whole cell lysates and purified inclusion bodies of the three biological replicates of the Csp-Δ*pyrE*::P_{thl14}-*nfrA_Bs* strain, alongside the WT strain. L: Bio-Rad Precision Plus Protein All Blue Standards ladder. Statistical analysis was carried out using two-tailed unpaired t-test; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

3.5.5 Crystallisation of the highly expressed YfkO_Bl candidate

Given the high expression levels of YfkO_Bl in *C. sporogenes*, as verified by the previous screens, but also its inability to metabolise CP-506 (Table 3.2), the desire to match it with a suitable prodrug arose. Initially, YfkO_Bl was subjected to combined semi-rational and random mutagenesis approaches to improve its activity towards CP-506; however, the generated mutant versions gained only trace activity towards the prodrug, with k_{cat}/K_M ratios still 200-fold lower than NfrA_Bs (unpublished data; J. Hyams, D. Ackerley – VUW). Hence, it became of interest to decipher its crystal structure, reasoning that structural knowledge of the protein could facilitate rational directed evolution towards CP-506 activity or alternative drug discovery approaches that might be pursued in the future.

YfkO_Bl was expressed and purified with an N-terminal His-tag, according to sections 2.4.1-2.4.2, using a pET-28b vector assembled as outlined in section 2.2.11.4 and shown in Appendix IV. Crystallisation screens were set up as described in section 2.4.5. Rectangular yellow crystals, characteristic of flavin mononucleotide (FMN)-containing nitroreductases, were observed in multiple conditions of the crystallisation screens employed, after 3-7 days of incubation at 20°C; these include conditions A7, B8, B10, B12 from JCSG-*plus*™, A5, A10, B5, B11, C11, D6, D8, D11, G7, H6 from PACT *premier*™, A2, B3, D11 from Morpheus® and A1, G5 from the SG1 Eco screen. Condition B11 from the PACT *premier*™ screen (0.2 M $CaCl_2.2H_2O$, 0.1 M MES, 20% w/v PEG 6000, pH 6.0) with 10.4 mg/ml of protein generated larger crystals and was therefore optimised further. Diffraction-quality YfkO_Bl crystals were again obtained in identical conditions with B11 as well as in wells containing higher PEG6000 content (23% w/v) (Figure 3.14).

Diffraction data collection and subsequent analyses were carried out by Dr Ingrid Dreveny. The crystal had the symmetry of space group P1 and unit cell dimensions of

a=47.36 Å, b=50.98 Å, c=53.22 Å, α=103.87°, β=105.99°, γ=92.40°. The structure was determined at a resolution of 1.9 Å and was solved by molecular replacement with a putative NAD(P)H-Flavin Oxidoreductase from *Streptococcus pyogenes* M1 (PDB entry: 2hay) as a search model (45% sequence identity with YfkO_Bl).

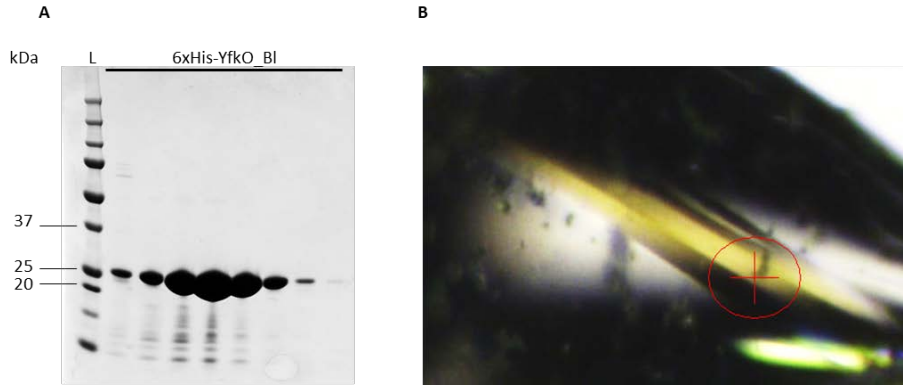


Figure 3.14 Purification and crystallisation of YfkO_Bl. A. SDS-PAGE analysis of YfkO_Bl-containing eluted fractions following purification via Immobilised Metal Affinity Chromatography and size-exclusion chromatography. All depicted fractions were combined and concentrated to set up the crystallisation screens; His-tagged YfkO_Bl: 28.3 kDa. L: Bio-Rad Precision Plus Protein All Blue Standards ladder. **B.** YfkO_Bl crystal.

The functional unit of YfkO_Bl has typical features of the FMN-dependent nitroreductase superfamily, including a homodimeric structure with 2 non-covalently associated FMN molecules within the crevices of the dimer interface, which constitute the active sites of the protein (Figure 3.15A)[176, 177]. The monomers consist of 226 amino acids each and have the typical 5-stranded anti-parallel β-sheet core with surrounding α-helices (Figure 3.15B). The 5-stranded anti-parallel β-sheet motif encompasses four strands from one monomer and one strand from the C-terminus of the other monomer; it should be noted that in the current model the monomers' C-terminal regions showed poor electron density and, as a result, the fifth strand which should have been part of the motif appears disordered (shown in magenta in Figure 3.15B). FMN is bound tightly in the active site in a highly conserved manner, stabilised by hydrogen bonds and

hydrophobic interactions with residues of both monomers (Figure 3.15C-D); the hydrogen bonds encompass 10 bonds with protein residues directly and 4 bonds via 3 highly ordered water molecules. (NB. the structural model is still undergoing refinement and therefore in-depth structural analyses could not be carried out at this stage).

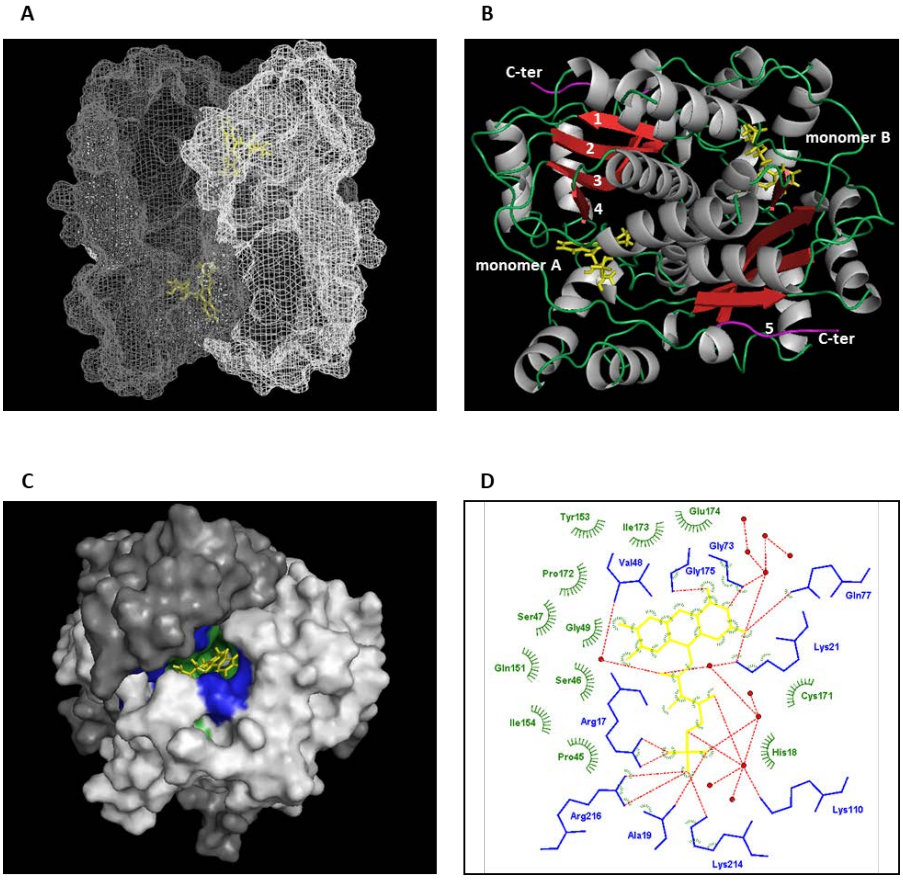


Figure 3.15 The crystal structure of YfkO_Bl. A. Mesh representation of the YfkO_Bl homodimer, showing the FMN cofactors in yellow, bound at the dimer interface. **B.** Cartoon representation, showing the secondary structure; α-helices are shown in grey, β-sheets in red and loops in green. The β-sheets of monomer A are numbered 1-5; no. 5 (magenta) appears disordered due to poor electron density in that region. The FMN cofactors are shown in yellow. **C.** Surface representation showing one of the two FMN binding sites. Residues interacting via hydrogen bonds are depicted in blue and residues interacting via van der Waals forces in green. The FMN cofactor is shown in yellow. **D.** Schematic diagram showing contacts of the FMN cofactor with protein residues, generated using LigPlot^{+}. Hydrogen bonds are shown as broken red lines. The protein residues involved in hydrogen bonds are shown in blue and those involved in hydrophobic interactions in green. The FMN cofactor is shown in yellow and water molecules as red spheres.

3.6 Discussion

Genes of heterologous bacterial nitroreductases were introduced in *C. sporogenes* NCIMB 10696, as a means of narrowing the candidates down to those that could embody clinically relevant options; for selection, important criteria are their ability to be tolerated and expressed at a high level as well as their ability to maintain activity in the host. In light of previous experience in our group suggesting that nitroreductase tolerance and expression in *C. sporogenes* is unpredictable and usually requires substantial fine-tuning, establishment of these parameters is crucial. Additionally, given the high substrate promiscuity of nitroreductases, confirmation of their activity is necessary in establishing that the chosen enzymes are not inhibited by constituents of the clostridial cell cytosol.

In confirmation of the original hypothesis, the strains generated in the present study remained phenotypically identical to WT, suggesting that all candidates are well tolerated in *C. sporogenes* and that cellular functions pertinent to cancer therapy have not been compromised. While i*n vitro* activity of all candidates was confirmed, a clear-cut distinction based on their expression levels was hindered by shortcomings associated with the enzymatic assays employed. Despite efforts to minimise bias against either nitroreductase family (by employing two different substrates and co-factors), it became apparent that direct comparison between NfsA and NfsB family members is not feasible. This is because NfsA nitroreductases usually have higher substrate affinities (K_M) but 1.5- to 5-fold lower overall turnover rates (k_{cat}) than NfsB representatives[139], implying that at conditions of substrate and cofactor excess (such as those employed here), the latter ones will always outperform the former ones (see also: YcnD_Bs enzymatic activity not in accordance with the SDS-PAGE analysis). Furthermore, even between representatives of the same family, enzymatic activity may be more informative of substrate preference rather than of expression levels. This is especially relevant here, where the enzymes have been selected/engineered to have high activity towards

compounds of interest, signifying that their activity towards other substrates may not be optimal; indeed, higher activity towards CP-506 or 5-NI PET probes appears to be at the expense of activity towards menadione and nitrofurazone (by way of illustration, YfkO_Bl_362 *vs* YfkO_Bl, T500_4 *vs* NfsB_Vv_dm). To overcome this limitation, enzymatic screens should ideally have been carried out against the prodrug of interest, CP-506. However, this was not possible at this stage due to limited availability of the compound, added to the fact that CP-506 metabolism is not detectable spectrophotometrically and would, therefore, require development and optimisation of a suitable protocol for liquid chromatography.

In this respect, information inferred from the SDS-PAGE is potentially more accurate with regards to the expression levels of the nitroreductase candidates, yet not quantitative enough for the purpose of narrowing them down. Naturally, scanning densitometry was also considered (and attempted) but was found to be unreliable, given that the enzymes are assayed in crude lysates where native proteins of *C. sporogenes* are also present.

Regardless, the (semi-quantitative) SDS-PAGE results in combination with a cautious consideration of the enzymatic activity screens, conclusively disclosed that the best candidate in terms of CP-506 metabolism (NfrA_Bs) was – inconveniently – poorly expressed in *C. sporogenes*. Initial troubleshooting attempts at the transcriptional (promoter), translational (codon usage) and post-translational (protein aggregation) levels did not reveal a specific cause for this outcome, supporting the erratic nature of nitroreductase expression in *C. sporogenes*, as previously noted (personal communication; A. Kubiak). However, other aspects of gene expression were not investigated, including mRNA stability[178] or protein degradation[179]. To this end, quantitative reverse transcription PCR (RT-qPCR) could elucidate whether the problem is at the mRNA level and inform the implementation of suitable strategies to improve expression of this valuable candidate in the future.

In parallel, YfkO_Bl emerged as an enzyme that can attain high expression levels in *C. sporogenes*, albeit not being a good match neither for CP-506 nor for 5-NI PET probes. In conjunction with our collaborators, it was decided that YfkO_Bl should not be relinquished, provided that a suitable prodrug and imaging agent can be found. Therefore, a new drug screening round was undertaken with compounds acquired from Threshold Pharmaceuticals (CA, USA), successfully identifying lead prodrugs that can be metabolised by YfkO_Bl (unpublished data; A. Patterson – UOA). In this context, structural knowledge derived from the solved crystal structure of the protein could be instrumental in the generation of optimised analogues (for instance, with improved physicochemical properties, toxicity profiles and imaging functional groups) via computer-aided/structure-based drug design[180]. To this end, it would be useful to broaden the crystallisation studies to gain a better insight of the substrate binding site, for example, via co-crystallisation with suitable ligands and/or rational mutagenesis[181, 182].

Nevertheless, in retrospect, given the difficulty in identifying suitable PCE/prodrug or PCE/imaging agent pairs, it would have been more efficient to focus research efforts in optimising expression of the better candidates. To take this further, earlier work should not have skewed the selection of nitroreductase candidates towards the better-expressed but worse prodrug-metabolising NfsB representatives. With the increasing availability of synthetic parts and genetic tools, the task of improving the expression of a given protein is possibly more straight-forward than identifying prodrugs/imaging agents anew. In this respect, the group's strategy for the evaluation of nitroreductase candidates in *C. sprorogenes* should also be revisited. Preliminary screening should involve testing of candidate genes under the control of several expression signals (for example, promoter, RBS and 5'-untranslated region libraries), to account for the inherent unpredictability of expression levels[183]. Additionally, several codon optimised versions of a given protein may be assessed. To enable rapid evaluation, constructs

should be assayed in plasmid-based versions and only top-performing combinations should be integrated in the genome.

An important follow-up to this study should be the assessment of enzymatic activity towards CP-506/5-NI PET probes in culture (that is, with whole cells and not lysates), where reactions will be subject to intracellular concentrations of cofactors, prodrug diffusion rates and varying levels of protein synthesis. This will allow nitroreductase candidates to be compared in a more therapeutically relevant context and facilitate selection of those that can be carried forward to preclinical studies. Importantly, these experiments will provide information on whether the high substrate affinity (K_M) of NfsA-like nitroreductases is advantageous or whether their profound preference for NADPH may pose a problem, given that NADPH has been reported to be less abundant than NADH intracellularly[139].

Finally, a significant outcome of the present study was that most candidates (except for NfrA_Bs) were able to attain higher expression levels than our benchmark NmeNTR nitroreductase (section 3.5.3.1). Crucially, an almost identical NmeNTR-expressing strain (with *nmeNTR* driven by P_{fdx} instead of P_{thl14}) in combination with CB1954 was able to bring about significant growth delay of human colorectal HCT116 tumours in NMRI nu/nu mice as well as generate complete tumour regression in 25% of cases[43]. Therefore, independently of which nitroreductase-expressing strain(s) is selected for implementation in preclinical *in vivo* assays, the expression levels afforded by the present expression strategy could be clinically relevant, in combination with a suitable prodrug.

3.7 Key Outcomes

- A library of nitroreductase-expressing *C. sporogenes* strains was generated, expressing enzymes capable of co-metabolising the CP-506 prodrug and 5-NI PET probes of interest.
- The crystal structure of the well-expressed nitroreductase YfkO_Bl candidate was determined.

CHAPTER 4

New strategies for improved gene expression

4.1 Background

4.1.1 The requirement for overexpression

Special regulatory and safety considerations regarding the use of genetically engineered bacteria in the clinic (discussed in more detail in Chapter 6) imply that the use of plasmid-based DNA is not appropriate for the clinical exploitation of clostridial-based cancer therapies. On the one hand, mobile elements such as plasmids pose the risk of being horizontally transferred to non-target organisms and biotic communities[184]. On the grounds of rising antibiotic resistance, this risk becomes an unsurmountable safety concern given the routine use of antibiotic resistance determinants for plasmid maintenance[185]. On the other hand, the inherent segregational instability of plasmids in the absence of selection, is a hurdle to the sustained and reliable expression of the gene

of interest, especially in the context of intratumoural bacterial colonisation which arises from many generations of bacterial growth[186]. Consequently, integration of desired therapeutic cargos in the bacterial genome represents a suitable alternative[20]. Advantageously, genomic integration offers greater genetic stability and unlimited capacity regarding the size of the heterologous cargo that can be introduced[65]. However, expression of chromosomally located genes is (usually) carried out from a single copy, contrary to plasmid-based systems where the gene may be present in multiple copies in the cell at any one time. Inevitably, this diminishes protein yield significantly and introduces a bottleneck given that the intratumoural concentration of the therapeutic protein is directly linked with the efficacy of the therapy (applicable in CDEPT and elsewhere)[47].

Indeed, difficulty in obtaining sufficient amounts of therapeutic proteins *in vivo* has been the downfall of previous attempts that have failed to demonstrate robust antitumour efficacy[25, 81, 86](also one of the reasons why strains with better colonisation properties were sought). Worse still, given the heterogeneity of human tumours in terms of size and extent of hypoxia, it is likely that bacterial colonisation in a clinical context will be less efficient than in the stringently controlled experimental tumours of animals[187], implying that the capability of the vector to produce copious amounts of the therapeutic agent is all the more critical.

Consequently, genetic engineering strategies that can improve chromosomal gene expression have an important role to play in the construction of clinically relevant strains. While most overexpression needs in clostridia have been typically fulfilled by the use of plasmid-based systems (with high-copy replicons) – addressing the metabolic engineering requirements of strains with industrial rather than clinical interest[188] – some could be germane to chromosomal usage. These include the provision of stronger transcription signals[86], stabilisation of mRNA transcripts via optimisation of 5' untranslated regions[189] and codon optimisation[43, 86].

In this context, preceding work in our group has already investigated and optimised transcriptional and translational parameters, specifically for *C. sporogenes* NCIMB 10696. These include the creation of a large library of constitutive promoters, the rational design of improved RBS sequences and the development of a refined codon optimisation algorithm in collaboration with ATUM (formerly DNA2.0)[104]. Notably, P_{thl14} (deployed for the expression of nitroreductase candidates in Chapter 3), constituted a significant finding of this work, yielding 4-fold higher expression than the transcriptional signal/RBS combination employed theretofore (P_{fdx}). The PhD thesis of Dr Aleksandra Kubiak who undertook this work was not publicly available at the time of writing, therefore relevant details regarding the codon optimisation algorithm and P_{thl14} are included in Appendices V and VII, respectively.

Harnessing these outcomes, it was of interest to explore alternative strategies that could be implemented in conjunction, to further increase gene expression and/or expand the flexibility of the researcher.

4.1.2 Strategy I: Multicopy integration

This strategy involves amplification of the gene dosage by increasing the copy number of the gene of interest on the chromosome; such an approach has been successful in elevating protein yields, adopted primarily in *Bacillus* species[190-192] and the yeast *Pichia pastoris*[193-195]. In theory, copy number amplification could be achieved by integrating tandem repeats of the gene in a single locus or by integrating single gene copies in multiple loci. In view of the intended application, the first option was discarded given that integrated tandem repeats (that is, repetitive homologous regions) are prone to excision via cyclisation[190, 191], thus bearing the risk of altering unpredictably the presumed properties of a therapeutic strain. Consequently, the second option was selected, envisioning the integration of gene copies in pre-designated, rationally selected loci, to avoid interfering with the strain's native characteristics.

4.1.2.1 The prerequisite: availability of pertinent genetic tools

Initially, this strategy presented a significant challenge due to the lack of suitable genetic tools that would allow sequential integrations to be carried out rapidly and in a directed manner. Furthermore, the opposition to use of antibiotic resistance markers, as described previously, extends also to any markers not naturally found on the chromosome[43], thus precluding the use of tools dependent on antibiotic resistance for mutant selection. Most methods theretofore relied on the use of counter-selection markers, such as *pyrE*[71], *mazF*[196] or *codA*[69], carried on replication-defective ('pseudo-suicide') or non-replicating plasmids, as a means of improving the efficiency of selecting rare double-crossover events. Nevertheless, these protocols are lengthy and laborious (up to 4 weeks), of low efficiency or not applicable in a WT background (for instance, in *C. sporogenes*, *pyrE* or *codA* inactivation would be required)[188]. Alternative methods (including ClosTron-mediated knock-in applications[197]) involved the genomic integration of the gene of interest concomitantly with antibiotic resistance markers, where the latter could be subsequently excised via Flp-FRT recombination or equivalent means[196, 198]. Unfortunately, the associated recombinases have been notoriously inefficient in *C. sporogenes* (personal communication; S. Kuehne) while the procedure would also require an additional cycle of transformation and plasmid curing.

In this context, ACE was a seminal addition in the clostridial genetic toolkit, whereby the double-crossover event was directly coupled to a selectable phenotype, for example, via *pyrE* inactivation in a WT host (and acquisition of 5-FOA resistance) – as described in Chapter 3 – or alternatively, via provision of a WT *pyrE* allele in a *pyrE*$^{-}$ host (and reversion to uracil prototrophy)[75]. Although the second approach is applicable genome-wide, provided that the WT *pyrE* allele can be placed downstream of a suitable promoter in the chromosome, it removes the desired characteristic of uracil auxotrophy. Additionally, once the WT *pyrE* allele has been re-introduced, it cannot be used again for selection, thus thwarting the multicopy genome integration strategy. Effectively,

this meant that *pyrE* was, at the time, the only validated locus available for effortless genomic integration in *C. sporogenes*.

Naturally, the identification of genes that can be manipulated to yield a selectable phenotype upon integration, in a similar fashion as *pyrE*, was highly desired. In search for a suitable counter-selection method, the use of gamma-L-Glutamyl hydrazide (GGH) was investigated. GGH is a toxic glutamine analogue which causes cell death via inhibition of the 5-amino imidazole ribonucleotide synthetase and downstream purine biosynthesis[199]. Also, it is a competitive inhibitor of glutamine transport[199-201] and its cell entry can only be realised if the glutamine transport system is intact[199, 200, 202-204]. Hence, it was hypothesized that inactivation of a glutamine transport gene (upon integration) would confer GGH resistance[205], serving as a selectable phenotype for mutant selection. Nevertheless, subsequent experiments revealed that the strain's sensitivity towards GGH could not be improved by inactivation of single genes, thus negating the possibility of using it in an ACE-like manner (results not shown).

4.1.2.2 CRISPR/Cas9: an enabling technology

Clustered regularly interspaced palindromic repeats (CRISPR) are non-contiguous repetitive DNA sequences of up to 50 nucleotides, present in the genomes of diverse prokaryotic organisms. They are separated by non-repetitive sequences of similar length (spacers) and flanked on one side by a common leader sequence of 300-500 nucleotides (the 'long repeat'). CRISPR loci are strictly associated with proteins of DNA-modifying function, aptly termed the CRISPR-associated proteins (Cas)[206, 207]. Although the biological function of CRISPR/Cas systems was not known at the time of their discovery, it has now been established that they mediate immunity against invading extrachromosomal elements, such as bacteriophages and conjugative plasmids[207-209]. The spacers between CRISPR motifs consist of short regions of DNA originating from past invaders, corresponding to 'cellular memory' of previous infections; mRNA transcribed from these sequences can guide Cas proteins to recognise and degrade incoming foreign DNA upon re-entry, thus conferring adaptive immunity to the prokaryotic cell[210].

The last five years have seen CRISPR/Cas systems being repurposed for genome editing applications in both eukaryotes and prokaryotes, with unprecedented success. In particular, the type II system from *Streptococcus pyogenes* has found widespread use due to its simplicity, requiring only one Cas protein for function. This protein is Cas9, an RNA-guided DNA endonuclease that generates site-specific double-stranded breaks (DSBs) in sequences that have been marked for elimination via base pairing with a corresponding guide RNA[211, 212]. DNA cleavage occurs only when the target sequence is also adjacent to a short sequence motif (for example, 5'-NGG-3'), known as the protospacer adjacent motif (PAM)[213]. Cas 9 typically requires a dual guide RNA for target recognition, composed of a precursor CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). For genome editing purposes, this requirement has been simplified with the use of synthetic single-guide RNA (sgRNA) chimeras[214]. Ultimately, modification of 20 bp within the sgRNA sequence can mediate targeting to a complementary 20-bp site in any genomic location of interest (adjacent to a PAM sequence). The facility with which this can be achieved is the biggest advantage of CRISPR/Cas systems compared to other methods of site-specific mutagenesis used previously. For instance, reprogramming of Zinc finger nucleases (ZFNs) and transcription activator–like effector nucleases (TALENs) to target desired sequences typically required engineering of a new enzyme for each new target sequence[215].

In eukaryotic cells, editing occurs as a consequence of Cas9-induced DNA repair (following the DSB), proceeding via non-homologous end joining (NHEJ) and/or homology-directed repair (HDR) pathways[216]. NHEJ is an error-prone mechanism, resulting in imperfect repair of the DSB, often introducing insertions or deletions. Preferably, in the presence of a donor DNA homologous to sequences flanking the targeted site (as anticipated when carrying out genome editing), repair can also proceed via the higher fidelity system of HDR, using the donor DNA as a template for homologous recombination. Essentially, the Cas9-induced DSB promotes homologous recombination, bringing about the introduction of desired alterations provided on the repair template, be it precise mutations, deletions or insertions[212].

In prokaryotic cells, CRISPR/Cas9 is employed as a powerful counter-selection marker, facilitating isolation of allelic exchange (double-crossover) mutants, on the basis that only edited cells are able to escape Cas9's lethal DSB. In edited cells, the target site for Cas9 has presumably been exchanged, via homologous recombination, for a donor DNA template which does not contain the Cas9 target sequence. On the contrary, unedited WT cells still carrying the target sequence, are eliminated, due to the low capacity of bacterial DNA repair mechanisms to cope with DSBs otherwise[217]. Additionally, the donor DNA template may carry the desired genetic edits, as previously described. It follows that identification of successful mutants can be carried out by simple screening and does not require antibiotic resistance marker selection.

During the last year of this project, an in-house CRISPR/Cas9 tool became available, optimised to work in clostridial strains of interest. Thanks to this tool, the multicopy genome integration strategy could be realistically pursued.

4.1.3 Strategy II: Implementation of the TcdR-P$_{tcdB}$ expression system

The second strategy involved the utilisation of an expression system from *C. difficile* R20291, encompassing an alternative RNA polymerase (RNAP) sigma factor (TcdR) and expression cassettes comprising its cognate promoter (P$_{tcdB}$), operably linked to the gene of interest.

TcdR (also known as TcdD or TxeR; 184 amino acids, 22 kDa) belongs to a novel subcategory of the σ^{70} family of sigma factors (designated as group 5)[218] and is indispensable for the expression of toxins A and B in *C. difficile*; it interacts directly with the RNAP core in a dose-depended manner, which permits recognition of the promoters upstream of *tcdA* and *tcdB*, allowing open complex formation and initiation of transcription[219]. Additionally, TcdR mediates recognition of its own promoter (P$_{tcdR}$), positively autoregulating its own expression[220, 221]. TcdR-mediated gene transcription operates with exquisite specificity given that TcdR can only recognise promoters with

conserved non-canonical -35 (and to a lesser extent -10) recognition sequences, while activation of the respective promoters is only feasible in the presence of TcdR[218]. Homologues of TcdR are conserved in other toxinogenic clostridia, including *C. botulinum* (BotR), *C. tetani* (TetR) and *C. perfingens* (UviA), thought to be derived from a common ancestor and exhibiting a degree of functional interchangeability[218, 222]. The *tcdR* gene cluster (the pathogenicity locus; PaLoc) is shown in Figure 4.1, whereby *tcdR* is located immediately upstream of its target genes, similarly to the architecture of toxin clusters found elsewhere[222].

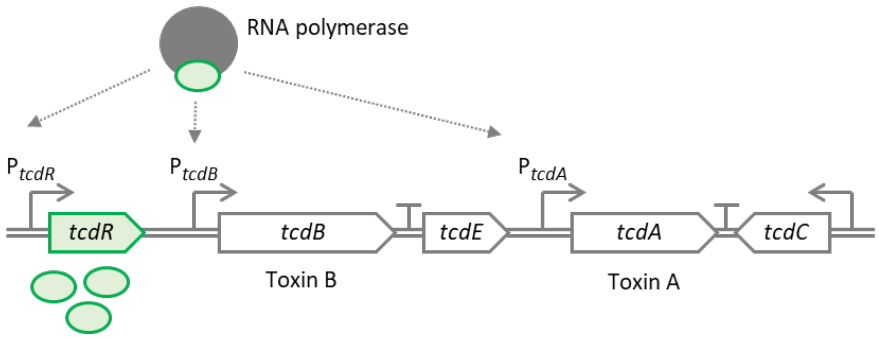


Figure 4.1 The *C. difficile* pathogenicity locus. Genetic organisation of the toxin A and B pathogenicity locus in *C. difficile*; *tcdR* encodes a sigma factor which allows the RNA polymerase to recognise the promoters of the toxin genes *tcdA* and *tcdB* (P_{tcdA} and P_{tcdB} respectively) as well as its own promoter (P_{tcdR}); *tcdE* encodes a holin-like protein and *tcdC* a negative regulator of toxin expression.

Given that *C. sporogenes* does not encode TcdR or TcdR-like sigma factors nor possess native promoters that are recognised by it, repurposing the TcdR-P_{tcdB} system for the transcriptional control of heterologous gene expression in this strain would offer a *bona fida* orthogonal expression pathway. Orthogonal expression systems rely on heterologous components to carry out the conversion of DNA to protein. Consequently, they are generally abstracted from native regulation and crosstalk with endogenous systems, enabling maximal independence from the host[223]. As such, they are considered useful for the assembly of complex synthetic genetic circuits, where insulation from endogenous

interference increases the likelihood of synthetic circuits delivering desired functions, in a robust manner[223]. Additionally, the introduction of heterologous expression pathways allows enlargement of the functional sample space, that is, increase in the number of expressed genes from a given genomic space[224].

With the advent of synthetic biology applications, orthogonal expression components are increasingly becoming available, capable of regulating gene expression on the transcriptional and/or translational level. Importantly, transcriptional control is considered more 'cost-effective' for the cell, preventing the consumption of cellular resources on mRNA synthesis[225]. To this end, the utilisation of bacteriophage-encoded T7 RNAP matched with T7 RNAP-specific promoters has been investigated as a strategy for orthogonal control of transcription. Mutant and fragmented versions of T7 RNAP with altered DNA recognition domains have also been developed, with the aim of expanding the range of T7-pairabe promoters[226-228]. However, the expression of phage-derived RNAP in bacteria is associated with toxicity, even in the absence of a T7 promoter[228], implying that the use of bacterial-derived components, such as heterologous sigma factors, might be preferable[225].

In this context, libraries of sigma factors originating from diverse bacterial strains, including *Bacillus subtilis* and *Lactobacillus plantarum*, have been implemented in *E. coli*; these have been paired with libraries of corresponding promoters, native or modified, in combinations capable of yielding a large range of expression strengths[224, 225, 229]. Regarding the TcdR-P_{tcdB} system in particular, TcdR has been used as a heterologous sigma factor in *C. perfringens*, where P_{tcdA} and P_{tcdB} could bring about expression of a reporter protein (GusA) only in the presence of TcdR[219]. Also, the TcdR-P_{tcdB} system has been successfully employed in *C. sporogenes* NCIMB 10696 for *mariner*-based transposon mutagenesis; in this case, *tcdR* was genomically integrated in the *pyrE* locus, driven by the native P_{pyrE}, while the transposase gene was supplied on a plasmid, driven by P_{tcdB}[74].

In addition to the system's orthogonality in *C. sporogenes*, the P_{tcdB} promoter is also not recognised in *E. coli* (in the absence of TcdR), facilitating cloning and manipulation of potentially toxic genes. Additionally, the system offers a high degree of flexibility, whereby different expression levels can be achieved by varying the configuration/location of the system's components as well as by modulating the dose of TcdR (for example, by employing different promoters), the number or P_{tcdB}-driven gene copies, or both. Finally, the TcdR-P_{tcdB} system as an expression platform is a patented technology (patent number WO2013144647A1)[230], implying that its incorporation in the clostridial vector is commercially desirable.

4.2 Objectives

- Explore the applicability of multicopy integration and the TcdR-P_{tcbB} expression system in improving gene expression in *C. sporogenes*.

4.3 Hypothesis

- Expression of a gene of interest in *C. sporogenes* NCIMB 10696 increases with every additional gene copy integrated in the strain's chromosome.
- In *C. sporogenes* NCIMB 10696, gene expression systems based on a heterologous/orthogonal sigma factor afford higher expression levels than systems utilising the host's native sigma factors.

4.4 Study design

yfkO_Bl (from Chapter 3) was employed as the reporter gene since it is expressed well in *C. sporogenes* and would allow the specific investigation of additional improvements in expression. Additionally, expression levels could be assayed easily via nitrofurazone reductase assays.

4.4.1 Strategy I

The expression cassette (EC) to be integrated in multiple genomic locations was comprised of *yfkO_Bl* driven by P_{thl14}, with strong bidirectional terminators on either side to isolate *yfko_Bl* expression from chromosomal read-through. The cassettes were assembled in CRISPR/Cas9 vectors with flanking arms homologous to regions 5' and 3' of the desired point of insertion, all together comprising the 'editing template' (Figure 4.2).

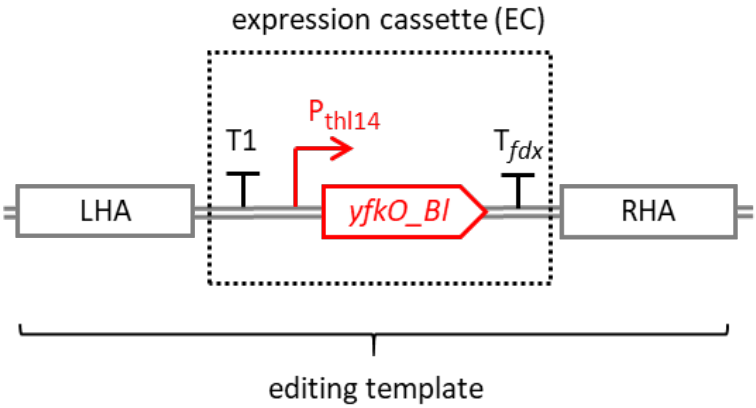


Figure 4.2 The expression cassette for multicopy integration The expression cassette to be integrated in different loci in the genome of *C. sporogenes*. T1: terminator associated with the glutamyl-tRNA synthetase gene (CTC_00977) from *C. tetani* E88; T_{fdx}: terminator associated with the ferredoxin gene (c01220) from *C. pasteurianum*; left (LHA) and right (RHA) homology arms: regions of homology upstream and downstream of the desired point of insertion (1000-1500 bp each).

4.4.2 Strategy II

For the implementation of the TcdR-P_{tcdB} system for gene expression, it was necessary to uncouple *tcdR* expression from its own promoter given that the transcription factors that may be mediating environmental control of *tcdR* expression are unknown[220]. Also, modulating expression of TcdR (for instance, by selection of an independent promoter)

was the most direct way of influencing the expression of the gene of interest. Nevertheless, preliminary work with the TcdR-P_{tcdB} system in *C. acetobutylicum* suggested that there might be a maximum level of TcdR that can be tolerated by the organism, presumably because high amounts may be sequestering available RNAP, outcompeting host-specific sigma factors[229, 231, 232]. For instance, when *tcdR* was placed under the control of the strong $P_{thl,}$ the phenotypic characteristics of the strain changed while weaker expression of the sigma factor was well tolerated (unpublished data; N.P. Minton).

Therefore, for test purposes, *tcdR* expression was linked to a lactose inducible system from *C. perfringens* (P_{lac})[233], whereby up-titration of the dose of the inducer (lactose) would allow determination of the highest level of TcdR expression attainable in *C. sporogenes* without compromising the fitness of the cell. The rationale was that P_{lac} could subsequently be replaced with a promoter of equivalent strength, sourced from an in-house promoter library. To gain a better understanding of the system as well as take advantage of its flexibility, the P_{lac}-*tcdR*-P_{tcdB}-*yfkO_Bl* expression system was studied in three configurations with: a) the entire construct based on a pMTL82151 plasmid ('plasmid-based'), b) the P_{lac}-*tcdR* module on the chromosome and the P_{tcdB}-*yfkO_Bl* module on a pMTL82151 plasmid ('split') and c) the entire construct integrated in the chromosome ('chromosomal'). The three configurations are shown in Figure 4.3.

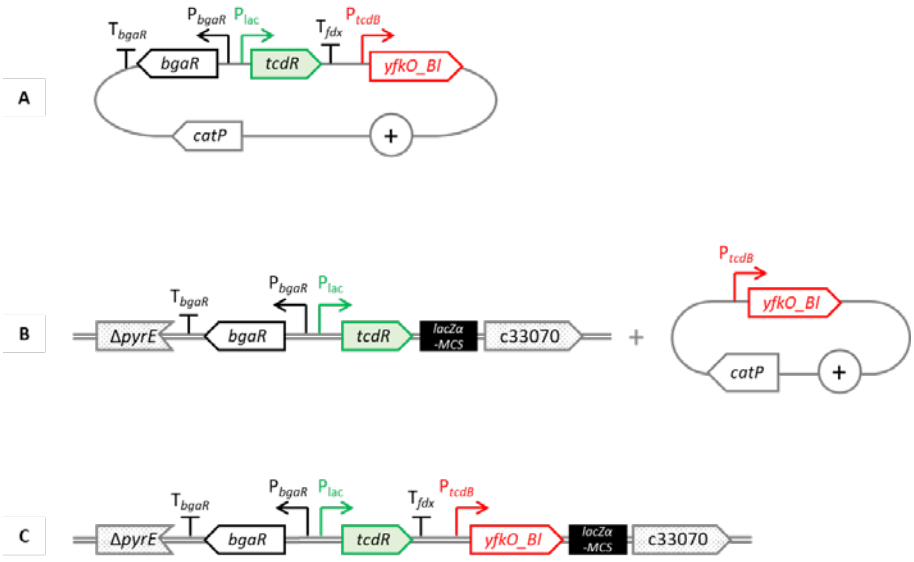


Figure 4.3 The configurations of the P_{lac}-*tcdR*-P_{tcdB}-*yfko_Bl* expression system. Schematic diagram showing the three configurations of the P_{lac}-*tcdR*-P_{tcdB}-*yfko_Bl* expression system. **A.** Plasmid-based. **B.** Split. **C.** Chromosomal. *bgaR*: encodes the transcriptional activator of P_{lac}; P_{bgaR}: the promoter of *bgaR* (divergent from P_{lac}), T_{bgaR}: the terminator associated with *bgaR*; +: pBP1 replicon; *catP*: thiamphenicol resistance marker; *lacZα*-MCS: the *lacZ*-multiple cloning site introduced into the genome with the ACE integration vector; c33070: the gene downstream of the integration locus.

4.5 Results

4.5.1 Identification of genomic loci

As mentioned previously, the genomic loci for integration of the expression cassette ought to be chosen rationally to avoid jeopardising inherent properties of the strain (such as its ability to colonise tumours). Coincidently, it was of interest to target genes whose inactivation/deletion may confer some form of disablement with regards to the capacity of the organism to survive outside of the tumour (namely by compromising its ability to transport or metabolise nutrients), thus enhancing the overall safety of the vector. In this tenet, *glnQ1* was chosen because it encodes a glutamine transport gene (previously investigated with GGH; section 4.1.2.1) while *treBAR* is the operon

encoding the proteins involved in the utilisation of trehalose; interestingly, the case for elimination of this operon is supported by recent findings reporting that *C. difficile* strains with the ability to metabolise low concentrations of this sugar display increased virulence[234]. Furthermore, the *closA-I* operon, encoding a potent haemolytic toxin, was targeted as a means of attenuation (discussed in more detail in Chapter 6).

For exemplification purposes, insertion in a non-coding region (immediately downstream of the c14540-c14550 operon; 1,590,094th bp) was also attempted without deleting or disrupting native sequences. To achieve this, the editing template contained silent mutations in the sgRNA and PAM sequences so that upon homologous recombination the Cas9 target would be eliminated without deletion of the corresponding sequence. The motive for selecting this particular genomic location was that the sgRNA sequence employed had been previously experimentally verified for the deletion of gene c14540, as part of the GGH studies that preceded this work (section 4.1.2.1; results not shown).

Finally, the *pyrE* locus was revisited, to provide a new benchmark, due to the addition of the 5' and 3' terminators which were not included in the ACE construct employed in Chapter 3. The editing template was designed to eliminate 387 bp from the 3' end of *pyrE*, mimicking the outcome of ACE (which eliminates 273 bp) and likewise introducing uracil auxotrophy. In total, 5 genomic loci were targeted for integration of the expression cassette via CRISPR/Cas9; these are listed in Table 4.1.

Table 4.1 Genomic loci selected for integration

Locus	Reason for selection	Substituted region	Coding region	Comments
pyrE	benchmark; prior experience; uracil auxotrophy	387 bp	Y	deletion of the 3' end of *pyrE* to mimic ACE
glnQ1	inactivation of a glutamine transport gene	986 bp	Y	deletion of *glnq1* and 200 bp upstream (putative promoter)
c14550-3'	insertion in non-coding region for exemplification purposes	-	N	insertion downstream of the c14540-c14550 operon
treBAR	inactivation of a carbohydrate (trehalose) utilisation pathway	3597 bp	Y	deletion of *treA* and the biggest parts of *treB* (1004 bp 3') and treR (661 bp 5')
closA-I	removal of a virulent trait (clostridiolysin)	8483 bp	Y	deletion of the entire clostridiolysin operon

4.5.2 Characterisation of genomic loci

To characterise the expression capabilities of the chosen loci, the expression cassette was integrated separately in each locus (in a single copy) using CRISPR/Cas9, as described in section 2.2.12.2, and vectors pMTL83121-RECas1-*pyrE*-EC, pMTL83121-RECas1-*glnQ1*-EC, pMTL83121-RECas1-c14550-3'-EC, pMTL83121-RECas1-*treBAR*-EC and pMTL83121-RECas1-*closA-I*-EC for the *pyrE*, *glnQ1*, c14550-3', *treBAR* and *closA-I* loci, respectively; the vectors were constructed according to section 2.2.11.7 and the general vector design can be found in Appendix IV. The generated strains as part of this work are listed in Table 4.2.

Table 4.2 Strains generated for the characterisation of genomic loci

Strain	Description
Csp-Δ*pyr*E::EC	T1-P_{thl14}-*yfko*_Bl-T_{fdx} integrated in the *pyrE* locus
Csp-Δ*glnQ1*::EC	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated in the *glnq1* locus
Csp-c14550-3'::EC	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated downstream of c14550
Csp-Δ*treBAR*::EC	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated in the *treBAR* locus
Csp-Δ*closA-I*::EC	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated in the *closA-I* locus

Insertions were verified by colony PCR for three biological replicates of each construct (Figure 4.4); flanking chromosomal primers were used for the *pyrE*, c14550-3', *treBAR* and *closA-I* loci while an insert-specific primer in combination with an operon-flanking one was used for the *glnQ1* locus, because the size of the insert was similar to the size of the substituted region; it should be noted that for the *closA-I* locus, the primers employed yield a PCR product only if insertion (and deletion of the operon) has occurred because the WT product is larger than the maximum capacity of the polymerase employed (11.7 kb). The plasmid was lost from confirmed integrants as described in section 2.2.13 and inserted fragments were verified via sequencing with suitable primers (Appendix II).

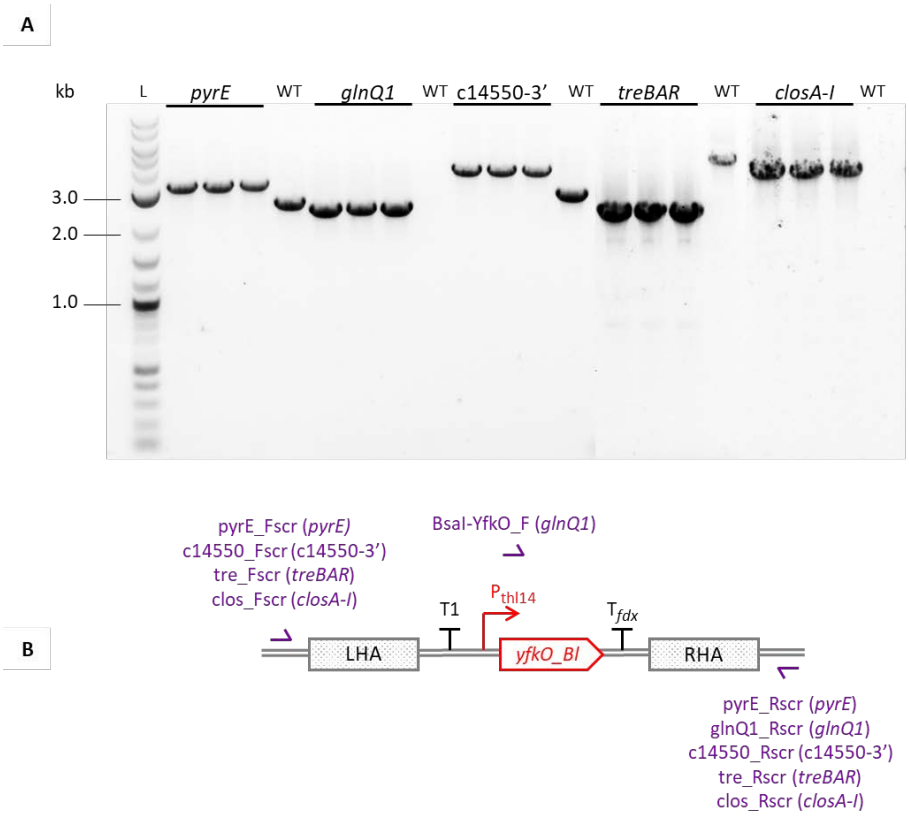


Figure 4.4 Genomic integration of the expression cassette in the selected loci. A. Colony PCR for three biological replicates of each strain with T1-P_{thl14}-*yfko_Bl*-T_{fdx} inserted in a single copy in the indicated locus, alongside the WT strain. Csp-Δ*pyrE*::EC: 3203 bp, WT: 2624 bp; Csp-Δ*glnQ1*::EC: 2428 bp, WT: no band; Csp-c14550-3'::EC: 3905 bp, WT: 2939 bp; Csp-Δ*treBAR*::EC: 2482 bp, WT: 5120 bp; Csp-Δ*closA-I*::EC: 15: 4210 bp, WT: no band (11727 bp). L: NEB 2-Log DNA ladder. **B.** Schematic diagram showing the primers used for colony PCR screening and their binding sites (in purple), as per the indicated locus (in brackets, next to the primer name).

The growth profile of all constructed strains was identical to WT, suggesting that insertions at the chosen loci and concomitant inactivation/elimination of the corresponding genes (where applicable) was well tolerated (Figure 4.5).

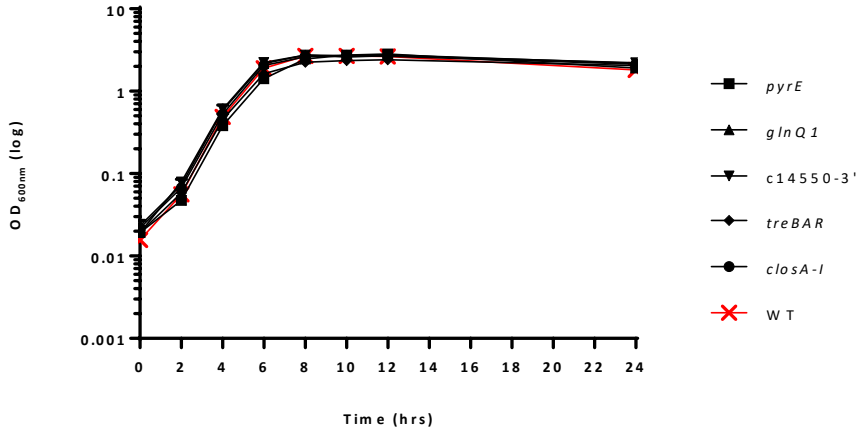


Figure 4.5 Growth curves of strains with a single copy of the expression cassette in different genomic loci. Strains with T1-P_{thl14}-*yfko_Bl*-T_{fdx} inserted in a single copy in the indicated locus were grown in liquid TYG medium, alongside WT; OD_{600} was monitored over 24 hrs. Data represent mean values of three independent cultures for WT and of three biological replicates for all mutant strains ± SD, from a single representative experiment out of three repeats.

Nitrofurazone reductase activity, a proxy for YfkO_Bl expression, was measured in the cell lysates of the constructed strains, as described in section 2.3.8; the Csp-Δ*pyrE*::P_{thl14}-*yfkO_Bl* strain generated previously using ACE was assayed alongside as a reference. As expected, the strain carrying the P_{thl14}-*yfkO_Bl* construct integrated via ACE and the strain carrying the T1-P_{thl14}-*yfko_Bl*-T_{fdx} expression cassette integrated via CRISPR/Cas9 yielded comparable expression levels of YfkO_Bl, confirming that the two genetic tools are equivalent (Figure 4.6). Interestingly, YfkO_Bl expression levels fluctuated between the different genomic locations, with the order of expression strength being as follows: *glnQ1* ≥ *closA-I* ≥ *pyrE* ≥ c14550-3' >> *treBAR* (Figure 4.6). In general, all loci resulted in statistically significant higher expression that *treBAR* while for *glnQ1* the difference was of statistical significance also *vs pyrE* and c14550-3'.

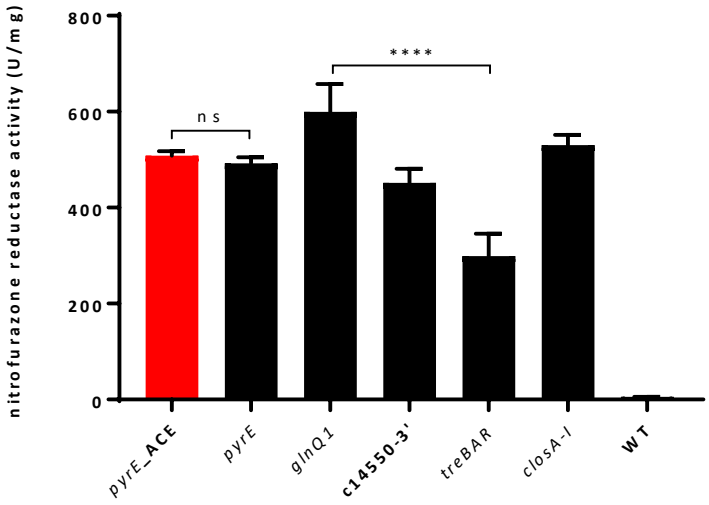


Figure 4.6 Expression strength of selected genomic loci. Nitrofurazone/NADH reductase activity in cell lysates of strains with T1-P_{thl14}-*yfko_Bl*-T_{fdx} inserted in a single copy in the indicated locus, alongside the lysate of the WT strain as a negative control. Data represent mean values of three independent cultures for WT and of three biological replicates for all mutant strains, normalised for total protein content ± SD, from a single representative experiment out of three repeats. Statistical analysis was carried out using one-way ANOVA with Tukey's test for multiple comparisons; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

The results were reanalysed to examine whether expression strength is influenced by the chromosomal position of each locus (Figure 4.7A-B); the observed difference in expression indeed correlated strongly with the location of the expression cassette relative to the origin of replication (*oriC*), whereby the closer the point of insertion to *oriC*, the stronger the expression. As expected, the difference in expression strength was more pronounced between the two extremes, namely, *glnQ1* which is the closest to *oriC* and *treBAR* which is the closest to the terminus (*ter*), with *glnQ1* yielding 2-fold higher expression ($p<0.0001$).

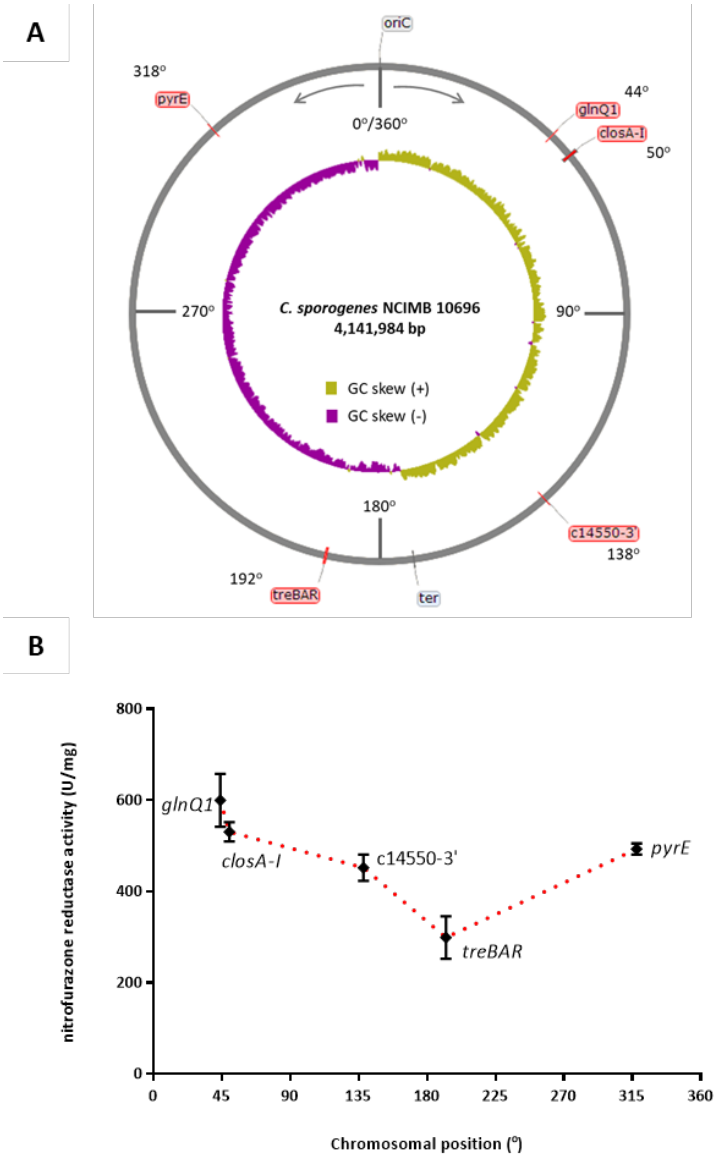


Figure 4.7 Expression strength of selected genomic loci relative to their chromosomal location. A. Chromosomal location of selected loci in degrees (1°≅11,506 bp). The origin of replication (*oriC*), replication terminus (*ter*) and direction of replication (arrows) are indicated; *oriC* was predicted based on the location of the *dnaA* boxes and *ter* based on the point where the GC skew becomes negative between the leading and lagging DNA strand. **B.** Nitrofurazone/NADH reductase activity in cell lysates of strains with T1-P_{thl14}-*yfko_Bl*-T_{fdx} inserted in a single copy in the indicated locus, relative to the chromosomal position of each locus. Data represent mean values of three biological replicates normalised for total protein content ± SD, from a single representative experiment out of three repeats.

4.5.3 Sequential amplification of gene copy number

To investigate whether amplification of the copy number on the chromosome can elevate expression levels, the expression cassette was sequentially integrated in the *pyrE, glnQ1*, c14550-3' and *treBAR* loci; unfortunately, due to time constraints a fifth copy could not be integrated in *closA-I*. Integrations were carried out, confirmed and sequenced as previously described (results not shown). At each stage, the plasmid was lost from confirmed integrants before the next one was introduced. All generated strains are listed in Table 4.3 and Appendix I.

Table 4.3 Strains generated for the sequential amplification of gene copy number

Strain	Description	Gene copy number
Csp-Δ*pyr*E::EC	T1-P_{thl14}-*yfkO*_Bl-T_{fdx} integrated in the *pyrE* locus	1
Csp-EC2	T1-P_{thl14}-*yfkO_Bl*-T_{fdx} integrated in the *pyrE* and *glnQ1* loci	2
Csp-EC3	T1-P_{thl14}-*yfkO_Bl*-T_{fdx} integrated in the *pyrE, glnQ1* and c14550-3' loci	3
Csp-EC4	T1-P_{thl14}-*yfkO_Bl*-T_{fdx} integrated in the *pyrE, glnQ1,* c14550-3' and *treBAR* loci	4

The progressive amplification of the copy number had an incrementally negative impact on the growth rate and total cell density reached by the strains, consistent with an increasing metabolic burden. The difference was more pronounced for the strain carrying 4 copies of the expression cassette *vs* the WT (Figure 4.8A-B). The strain harbouring 3 copies of the expression cassette yielded the highest expression levels of YfkO_Bl (1.6-fold higher *vs* the strain with 1 copy; p=0.0009), while the addition of a further copy caused a significant drop in expression (1.3-fold; p= 0.0241). Further, the expression gain afforded by the sequential additions from 1 to 3 copies did not follow a linear relationship with the copy number; the expression gain was 1.5-fold between 1 and 2 copies (p=0.0033) while being non-significant between 2 and 3 copies (Figure 4.8B).

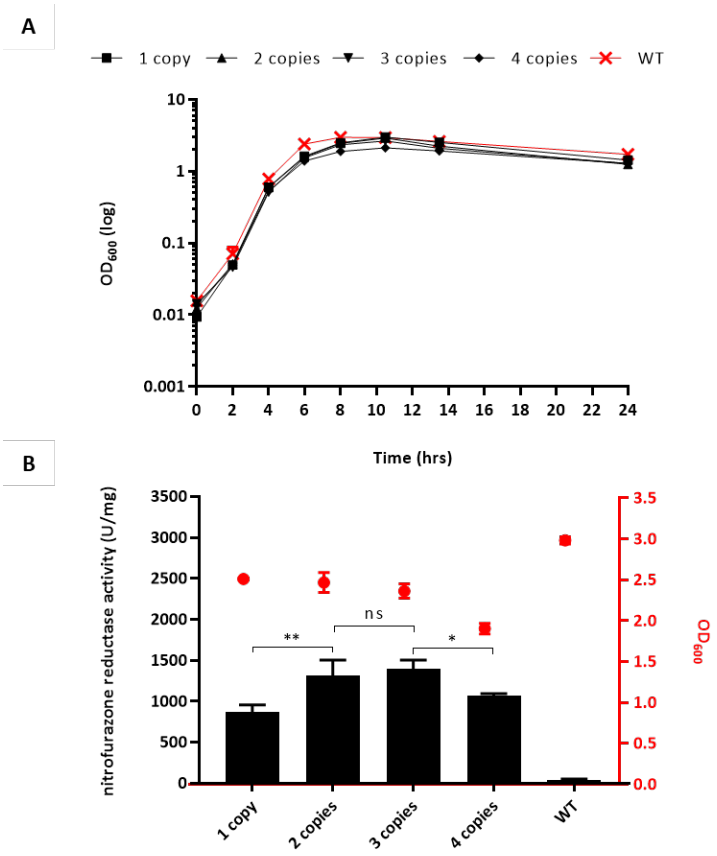


Figure 4.8 Characterisation of multicopy strains. A. Growth curves of strains with increasing copies of the T1-P_{thl14}-*yfko_Bl*-T_{fdx} expression cassette. Strains were grown in liquid TYG medium, alongside WT; OD_{600} was monitored over 24 hrs. Data represent mean values of three independent cultures for WT and of three biological replicates for all mutant strains ± SD, from a single representative experiment out of three repeats. **B.** Nitrofurazone/NADH reductase activity in cell lysates of strains with increasing copies of the T1-P_{thl14}-*yfko_Bl*-T_{fdx} expression cassette, alongside the lysate of the WT strain as a negative control. Data represent mean values of three independent cultures for WT and of three biological replicates for all mutant strains, normalised for total protein content ± SD. OD_{600} of the cultures at the time of harvesting (8 hrs) is also indicated. Depicted data are from a single representative experiment out of three repeats. Statistical analysis was carried out using one-way ANOVA with Tukey's test for multiple comparisons; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

4.5.4 Preliminary investigation of the T*cdR*-P$_{tcdB}$ system

To study the plasmid-based configuration of the TcdR-P$_{tcdB}$ system, plasmid pMTL82151-P$_{lac}$-OS (OS: orthogonal system) was constructed and conjugated into *C. sporogenes* as described in sections 2.2.11.5 and 2.1.9, respectively. For the split configuration, the P$_{lac}$-*tcdR* module was integrated in the *pyrE* locus using ACE and the pMTL-MZ1-P$_{lac}$-*tcdR* vector, constructed as described in section 2.2.11.6. Three plasmid-free biological replicates of Csp-Δ*pyrE*::P$_{lac}$-*tcdR* were subsequently transformed with plasmid pMTL82151-P$_{tcdB}$-*yfkO_Bl* via conjugation; the plasmid was constructed as described in section 2.2.11.5. For the chromosomal configuration, the entire P$_{lac}$-*tcdR*-P$_{tcdB}$-*yfkO_Bl* construct was integrated in the *pyrE* locus via ACE and the pMTL-MZ1-P$_{lac}$-OS vector, assembled as described in section 2.2.11.6. When required, plasmid loss was carried out as detailed in section 2.2.13. Three biological replicates were obtained for all strains and verified via colony PCR with suitable primers (Figure 4.9). For clarity, the aforementioned strains are listed in Table 4.4.

Table 4.4 Strains generated for the preliminary investigation of the TcdR-P$_{tcdB}$ system

Strain	Description	Harboured plasmid	Configuration studied
***C. sporogenes* NCIMB 10696**	wildtype isolate	pMTL82151-Plac-OS	plasmid-based
Csp-Δ*pyrE*::P$_{lac}$-*tcdR*	P$_{lac}$-*tcdR* integrated in the *pyrE* locus	pMTL82151-P$_{tcdB}$-*yfkO_Bl*	split
Csp-Δ*pyrE*::Plac-OS	P$_{lac}$-*tcdR*-P$_{tcdB}$-*yfkO_Bl* integrated in the *pyrE* locus	-	chromosomal

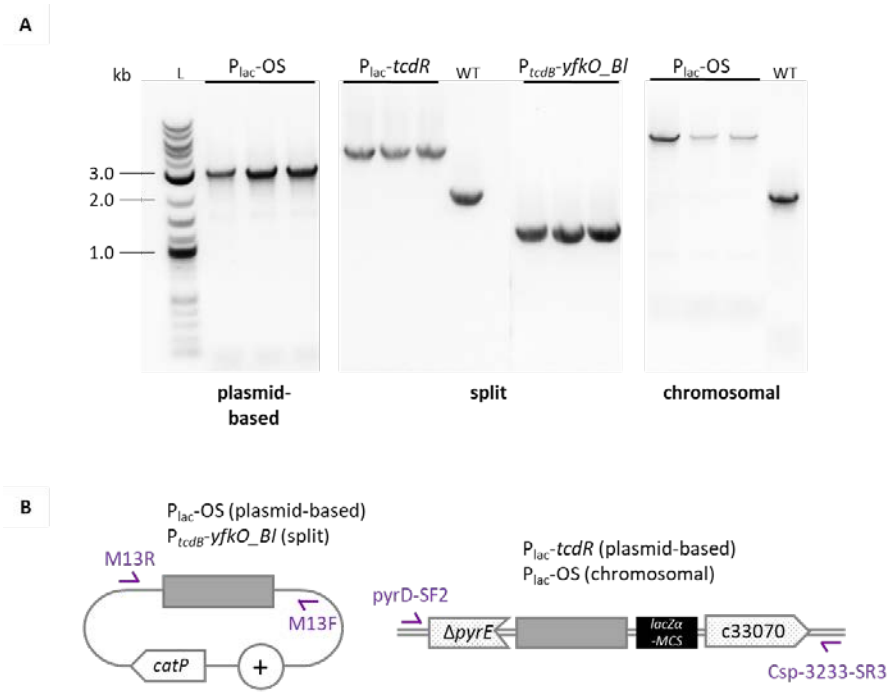


Figure 4.9 Construction of strains with three configurations of the TcdR-P$_{tcdB}$ system. A. Colony PCR for three biological replicates of the strains harbouring different configurations of the P$_{lac}$-*tcdR*-P$_{tcdB}$-*yfkO_Bl* system, as indicated. Plasmid based: Csp-pMTL82151-P$_{lac}$-OS (3149 bp); split: Csp-Δ*pyrE*::P$_{lac}$-*tcdR* (3907 bp), WT (2024 bp), pMTL82151-P$_{tcdB}$-*yfkO*_Bl (1087 bp); chromosomal: Csp-Δ*pyrE*::P$_{lac}$-OS (5027 bp), WT (2024 bp). L: NEB 2-log DNA ladder. **B.** Schematic diagram showing the primers used for colony PCR screening and their binding sites (in purple).

As a negative control, the WT strain was transformed with plasmid pMTL82151-P$_{tcdB}$-*yfkO_Bl*, to establish the background level of promoter activation in the absence of TcdR (neg). To gauge the range of expression levels that can be attained with the TcdR-P$_{tcdB}$ system, a strain harbouring plasmid pMTL82151-Pt$_{hl14}$-*yfkO_Bl* was assayed alongside as a positive control (pos); this plasmid-based version of the P$_{thl14}$-*yfkO_Bl* construct generates the highest expression levels of YfkO_Bl in *C. sporogenes* to date (data from the in-house promoter library, see Figure 4.11).

To measure expression, the strains were cultured as per the usual protocol (section 2.3.8); induction was carried out at the mid exponential phase (4 hrs) with 1 or 10 mM lactose and samples were harvested at the beginning of stationary phase (8 hrs). Cell

lysates were assayed for nitrofurazone reductase activity as previously described. The split configuration yielded the highest expression levels compared to the plasmid-based and chromosomal versions, while under induction with 10 mM lactose, the system also outperformed the positive control (1.2-fold higher; p=0.0108) (Figure 4.10). Importantly, the level of TcdR expression resulting from this version of the TcdR-P_{tcdB} system appeared to be well tolerated at both concentrations of the inducer. As expected, the negative control yielded negligible activity, confirming the orthogonality of the system given that P_{tcdB} is not recognised in *C. sporogenes* unless TcdR is present.

The split and plasmid-based configurations in theory operate with the same number of P_{tcdB}-*yfkO_Bl* copies (both plasmids employ the pBP1 replicon), the difference being that in the split version *tcdR* expression is carried out from a single chromosomal copy. Plausibly, the lower production of TcdR allows the transcriptional/translational machinery to be channelled more efficiently to the production of YfkO_Bl. The lack of dose-dependency in the plasmid-based system between 1 mM and 10 mM of lactose is further indicative of a potential saturation of the expression mechanism, cellular burden and/or toxicity due to excess TcdR. In the plasmid-based version there is also increased leakiness without induction, probably due to plasmid copy-number effects. Compared to the chromosomal configuration, both systems produce, *ceteris paribus*, the same amount of TcdR (expression driven from one chromosomal copy under the control of P_{lac}); presumably, in the split version, higher expression (6.5-fold higher for 1 mM lactose and 2.3-fold higher with 10 mM lactose; p<0.0001) is due to the increased availability of copies of P_{tcdB}-*yfkO-Bl*, as afforded by the pBP1 replicon employed (the exact number is not known). Essentially, it appears that the split configuration operates more effectively (low TcdR, high P_{tcdB}), contrary to the plasmid-based configuration which is 'malfunctioning' (too much TcdR) and to the chromosomal configuration which is functioning below capacity (too little P_{tcdB}). This highlights that pertinent implementation of the system can circumvent TcdR toxicity, while still generating high expression levels of the gene of interest.

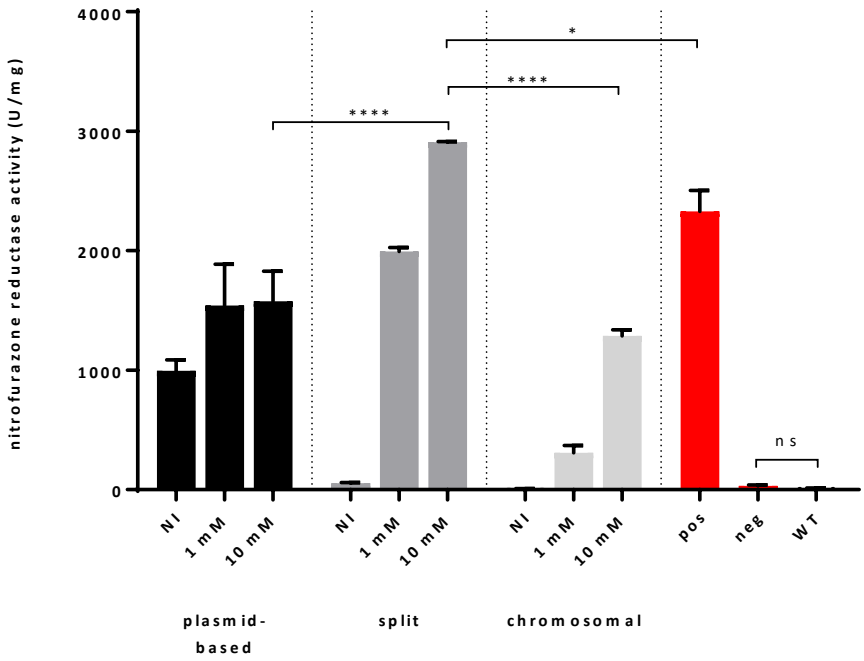


Figure 4.10 Preliminary investigation of expression levels using the TcdR-P$_{tcdB}$ system. Nitrofurazone/NADH reductase activity in cell lysates of strains harbouring different configurations of the P$_{lac}$-*tcdR*-P$_{tcdB}$-*yfkO_Bl* system, as indicated, alongside the lysates of a negative (neg) and positive (pos) control and WT. NI: non-induced, 1 mM: induction with 1 mM lactose, 10 mM: induction with 10 mM lactose. Data represent mean values of three independent cultures for WT and of three biological replicates for all other strains, normalised for total protein content ± SD. Depicted data are from a single representative experiment out of three repeats. Statistical analysis was carried out using one-way ANOVA with Tukey's test for multiple comparisons; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

4.5.5 Coupling the TcdR-P$_{tcdB}$ system to a constitutive promoter

As originally intended, it was of interest to substitute the lactose inducible promoter with a constitutive one of similar strength for the system to be applicable in the design of a clinically relevant vector where the use of exogenous inducers is not favoured. Based on data from the plasmid-based in-house promoter library (Figure 4.11), expression after induction of P$_{lac}$ with 1 mM lactose is approximately equivalent to the

activity of the promoter of the phosphate butyryltransferase gene (*ptb*) from *C. acetobutylicum* (P_{ptb}); while there was no promoter of equal strength to P_{lac} induced with 10 mM lactose, P_{fdx} from *C. sporogenes* was chosen as the next best alternative (albeit 1.5-fold stronger). Nevertheless, upon substitution of P_{lac} with P_{fdx}, no stable transconjugants could be obtained for the plasmid-based version, while the split P_{fdx}-*tcdR* module and the entire chromosomal construct could not be integrated after several attempts. As this was indicative of potential toxicity, the P_{fdx}-driven constructs were not pursued further.

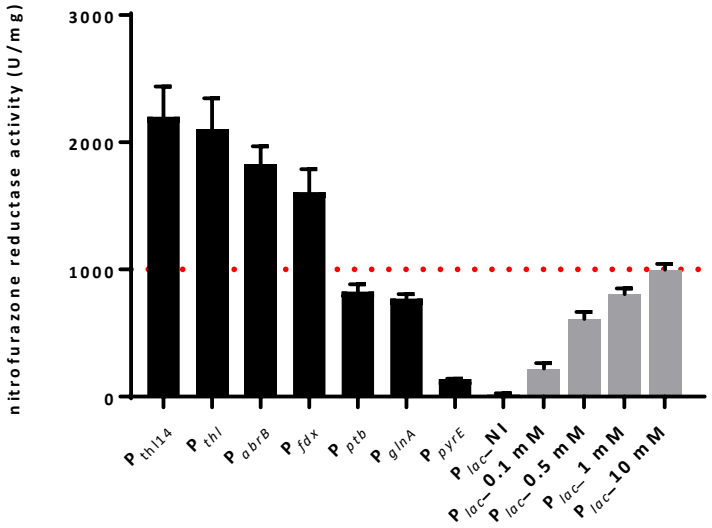


Figure 4.11 Expression of YfkO_Bl using promoters from the in-house promoter library. Nitrofurazone/NADH reductase activity in cell lysates of strains harbouring plasmids expressing YfkO_Bl under the control of different promoters, as indicated, in comparison to the lactose inducible system non-induced (NI) and induced with 0.1, 0.5, 1.0 or 10 mM lactose. Data represent mean values of three biological replicates, normalised for total protein content ± SD. Depicted data are pooled from three independent experiments.

The P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* system was assayed in three configurations, as previously described. Terminator T1 was incorporated upstream of P*ptb* in all constructs. Plasmid-based strains were obtained via transformation with the pMTL82151-T1-P_{ptb}-OS vector. For the split configuration, the T1-P_{ptb}-*tcdR* module was integrated in the *pyrE* locus

using CRISPR/Cas9 and the pMTL83121-RECas1-*pyrE*-T1-P_{ptb}-*tcdR* vector, generating strain Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* which was subsequently transformed with plasmid pMTL82151-P_{tcdB}-*yfkO_Bl*, as for the P_{lac}-driven system. The entire T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* construct was integrated in the *pyrE* locus via ACE using the pMTL-MZ1-p15a-T1-P_{ptb}-OS vector (results not shown). A list of the generated strains is shown in Table 4.5.

Table 4.5 Strains generated for the investigation of the TcdR-P_{tcdb} system under the control of a constitutive promoter

Strain	Description	Harboured plasmid	Configuration studied
***C. sporogenes* NCIMB 10696**	wildtype isolate	pMTL82151-T1-P_{ptb}-OS	plasmid-based
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus	pMTL82151-P_{tcdB}-*yfkO_Bl*	split
Csp-Δ*pyrE*:: T1-P_{ptb}-OS	T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* integrated in the *pyrE* locus	-	chromosomal

In agreement with the P_{lac}-driven system, the split configuration of the P_{ptb}-driven system appeared to be very efficient, yielding similar expression levels to the plasmid-based version and 6.3-fold higher than the chromosomal one. Nevertheless, the expression levels attained in the split and chromosomal versions with P_{ptb} were 1.6-fold lower than in the corresponding strains with P_{lac} under induction with 1 mM lactose (indicated as broken lines in Figure 4.12); this could be attributed to differences in the expression strength of P_{lac} *vs* P_{ptb} when integrated in the chromosome (NB. the in-house promoter library is plasmid-based). Alternatively, the system might be able to yield higher expression levels with an inducible promoter, whereby expression of the gene of interest is switched on at the mid-exponential growth phase (*vs* the lag phase for constitutive expression), thus minimising metabolic burden and/or toxic effects[235].

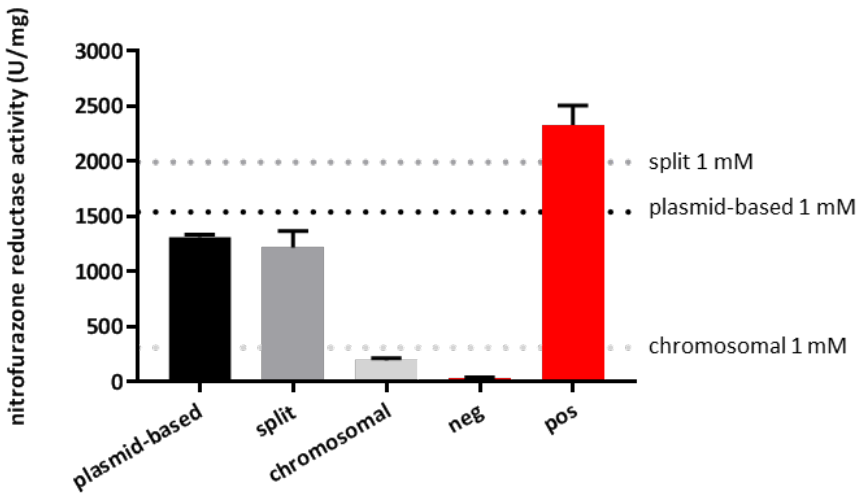


Figure 4.12 The TcdR-P$_{tcdB}$ system under the control of a constitutive promoter. Nitrofurazone/NADH reductase activity in cell lysates of strains harbouring different configurations of the P$_{ptb}$-*tcdR*-P$_{tcdB}$-*yfkO_Bl* system, as indicated, alongside the lysates of a negative (neg) and positive (pos) control. Corresponding expression levels obtained with the P$_{lac}$-driven system under induction with 1 mM lactose in the three configurations are indicated as broken lines. Data represent mean values of three biological replicates, normalised for total protein content ± SD. Depicted data are from a single representative experiment out of three repeats.

4.5.6 Mimicking the split configuration on the chromosome

Although the split configuration is the most favourable in terms of expression levels, it is partly plasmid-based which renders it unsuitable for a clinical strain. Therefore, to increase the efficiency of the chromosomal configuration, a genetic design was devised encompassing a higher number of P$_{tcdB}$-*yfkO_Bl* copies located on the chromosome in combination with one copy of *tcdR*, as a chromosomal 'mimic' of the split configuration. Nevertheless, given that the exact copy number of the pBP1 replicon employed in the pMTL-82151-P$_{tcdB}$-*yfkO_Bl* vector is not known, it was difficult to estimate how many P$_{tcdB}$-*yfkO_Bl* copies would be required for the expression levels to approach those of the split configuration; certainly, it was anticipated that the number of available loci

theretofore (4 additional) would not be sufficient, but the potential expression gain was explored regardless.

Amplification of the number of P_{tcdB}-*yfko_Bl* copies was carried out in the Csp-Δ*pyrE*::T1-P_{ptb}-OS strain generated previously, already containing one P_{tcdB}-*yfko_Bl* copy in the *pyrE* locus. The next copies were sequentially integrated in the *glnQ1*, *treBAR* and c14550-3' loci, using CRISPR/Cas9 and vectors pMTL83121-RECas1-*glnQ1*-OS, pMTL83121-RECas1-*treBAR*-OS and pMTL83121-RECas1-c14550-3'-OS, respectively, assembled as detailed in section 2.2.11.7. Due to time constraints, a fifth copy could not be integrated in the *closA-I* locus. P_{tcdB}-*yfkO_Bl* was also integrated as a single copy in each locus of the Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* strain, carrying only the P_{ptb}-*tcdR* module in the *pyrE* locus, as a reference of the contribution of each copy to the total expression observed. Integrations were confirmed and sequenced as previously described. All generated strains are listed in Table 4.6 and Appendix I. Figure 4.13 shows the colony PCR confirmation of the final strain, containing four P_{tcdB}-*yfko_Bl* copies. The growth profile of all strains was identical to WT (results not shown).

Table 4.6 Strains generated for the chromosomal implementation of the TcdR-P_{tcdb} system

Strain	Description	Gene copy number
Csp-ΔpyrE::T1-P_{ptb}-OS	T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* integrated in the *pyrE* locus	1
Csp-OS2	T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the *glnQ1* locus	2
Csp-OS3	T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the *glnQ1* and *treBAR* loci	3
Csp-OS4	T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the *glnQ1*, *treBAR* and c14550-3' loci	4
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* Δ*glnq1*::P_{tcdB}-*yfkO_Bl*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the *glnQ1* locus	1
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* c14550-3'::P_{tcdB}-*yfkO_Bl*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the c14550-3' locus	1
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* Δ*treBAR*::P_{tcdB}-*yfkO_Bl*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the *treBAR* locus	1
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* Δ*closA-I*::P_{tcdB}-*yfkO_Bl*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the *closA-I* locus	1

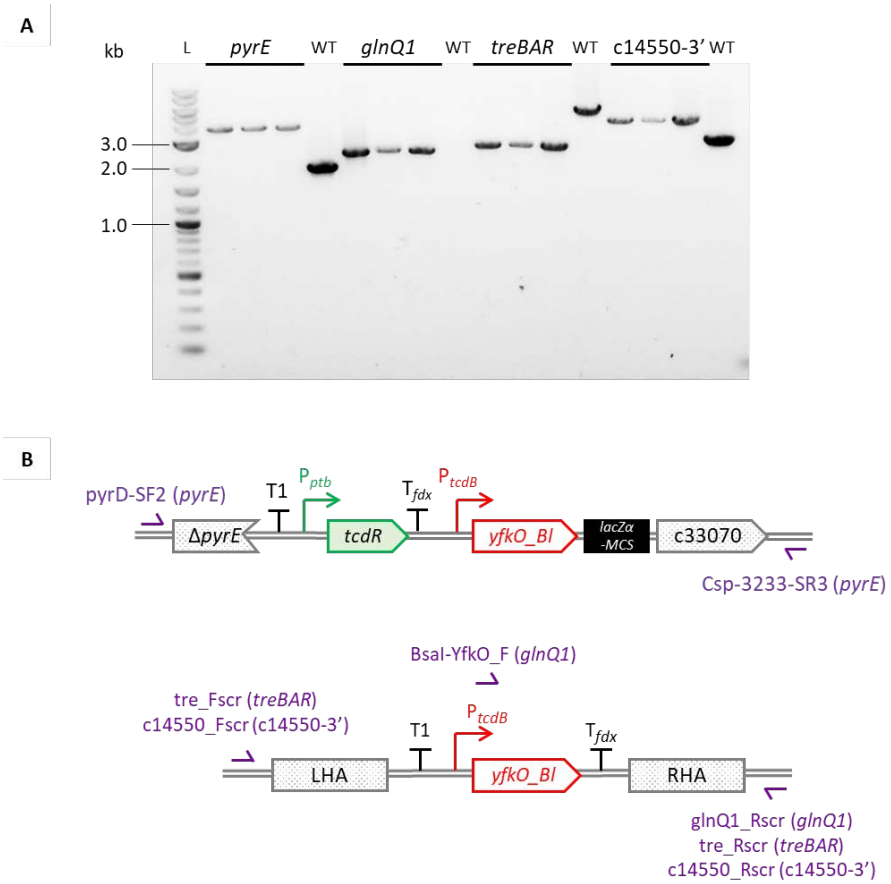


Figure 4.13 Construction of the chromosomal P_{ptb}-*tcdR* multicopy P_{tcdB}-*yfkO_Bl* strain. A. Colony PCR for three biological replicates of the strain containing P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* in the *pyrE* locus and one P_{tcdB}-*yfko_Bl* copy in the rest of the indicated loci, alongside WT; *pyrE*: 3851 bp, WT: 2024 bp; *glnQ1:* 2428 bp, WT: no band; *treBAR*: 2663 bp, WT: 5120 bp; c14550-3':4086 bp, WT: 2939 bp. L: NEB 2-Log DNA ladder. **B.** Schematic diagram showing the primers used for colony PCR screening and their binding sites (in purple), as per the indicated locus (in brackets, next to the primer name).

The cumulative expression gain afforded by 3 additional P_{tcdB}-*yfkO_Bl* copies was 2.3-fold ($p<0.0001$), but total expression was nonetheless 2.7-fold lower than in the split configuration; this highlights that the expression level derived from the split configuration is probably due to a significantly increased availability of P_{tcdB}-driven *yfkO_Bl* copies afforded by the pMTL82151 plasmid and will probably not be realistically achievable via the sequential integration of chromosomal copies of the P_{tcdB}-*yfkO_Bl* module. Additionally, increases in expression levels became progressively

smaller with each additional copy; for example, expression increased 1.8-fold between 1 and 2 copies (p=0.0007), 1.3-fold between 2 and 3 copies (p=0.0336) and not significantly between 3 and 4 copies (Figure 4.14A).

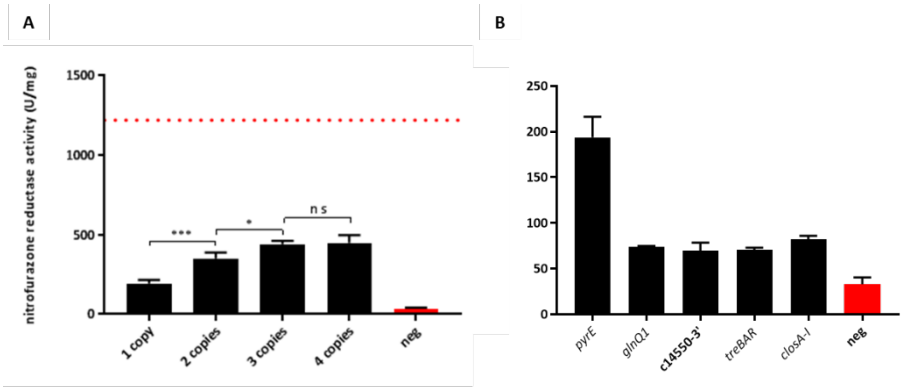


Figure 4.14 Investigation of expression of YfkO_Bl using the chromosomal mimic of the split configuration system. A. Nitrofurazone/NADH reductase activity in cell lysates of strains harbouring P_{ptb}-*tcdR* in the *pyrE* locus and increasing copies of P_{tcdB}-*yfkO_Bl* as indicated, alongside a negative control (neg). Expression levels obtained with the system in the split configuration are indicated as a broken line. **B.** Nitrofurazone/NADH reductase activity in cell lysates of strains with P_{ptb}-*tcdR* in the *pyrE* locus and the P_{tcdB}-*yfkO_Bl* module as a single copy in each indicated locus, alongside a negative control (neg). Data represent mean values of three biological replicates, normalised for total protein content ± SD. Depicted data are from a single representative experiment out of three repeats. Statistical analysis was carried out using one-way ANOVA with Tukey's test for multiple comparisons; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

Interestingly, expression derived from the *glnQ1*, c14550-3', *treBAR* and *closA-I* loci individually is approximately 2.5-fold lower than that derived from the *pyrE* locus (Figure 4.14B), an observation that is not in line with the relative expression strengths of the genomic loci as previously discussed (section 0). A difference that could account for this is that in the Csp-Δ*pyrE*::T1-P_{ptb}-OS strain, the P_{ptb}-*tcdR* module is directly upstream of the P_{tcdB}-*yfkO_Bl* module (both co-localised in the *pyrE* locus) whereas in the rest of the strains they are separated, whereby P_{ptb}-*tcdR* is located in *pyrE* and P_{tcdB}-*yfkO_Bl* in each respective locus. This would suggest that the chromosomal distance

between the sigma factor and the reporter gene impacts on expression significantly, also implying that under the current design, any expression gain resulting from additional copies is suboptimal, given that for 3/4 P_{tcdB}-*yfkO_Bl* modules, TcdR is encoded distantly.

4.5.7 Strategy I *vs* Strategy II

Finally, it was of interest to directly compare the two overexpression strategies; to this end, the best performing chromosomally-engineered strains, namely the strain with 3 copies of the expression cassette (T1-P_{thl14}-*yfko_Bl*-T_{fdx}) (section 4.5.3) and the strain with 4 copies of the P_{tcdB}-*yfko_Bl* module (section 4.5.6), were assayed side-by-side alongside relevant reference strains (Figure 4.15). The strain with 3 copies of the expression cassette yielded 2.7-fold higher expression levels than the strain with 4 copies of the P_{tcdB}-*yfko_Bl* module (p=0.0002), suggesting that the multicopy integration of a gene of interest under the control of a strong constitutive promoter (strategy I) is more efficient than the multicopy integration of the promoter-gene module of the orthogonal system, when TcdR is encoded from a single gene copy under the control of P_{ptb} (strategy II). Additionally, the expression level derived from 3 integrated copies of the expression cassette equalled that of the split configuration of the P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* system, highlighting the superiority of strategy I, in view of the inability of the chromosomal mimic to attain the levels of expression of the split configuration, as previously mentioned. Importantly, the expression levels achieved with strategy I are similar to the levels achieved with the chromosomal configuration of the P_{lac}-TcdR-PtcdB-*yfkO_Bl* system when induced with 10 mM lactose, which represent the highest expression levels derived from a chromosomal construct to date (indicated as a broken red line in Figure 4.15). Nevertheless, this level of expression is strikingly lower than the expression levels achieved by the plasmid-based version of the P_{thl14}-*yfkO_Bl* construct (pos) as well as the expression levels attained with the split configuration of the P_{lac}-TcdR-PtcdB-*yfkO_Bl* system when induced with 10 mM lactose (indicated as

a broken black line in Figure 4.15); the latter represents the maximum expression levels of YfkO_Bl attainable in *C. sporogenes* out of all constructs trialed to date.

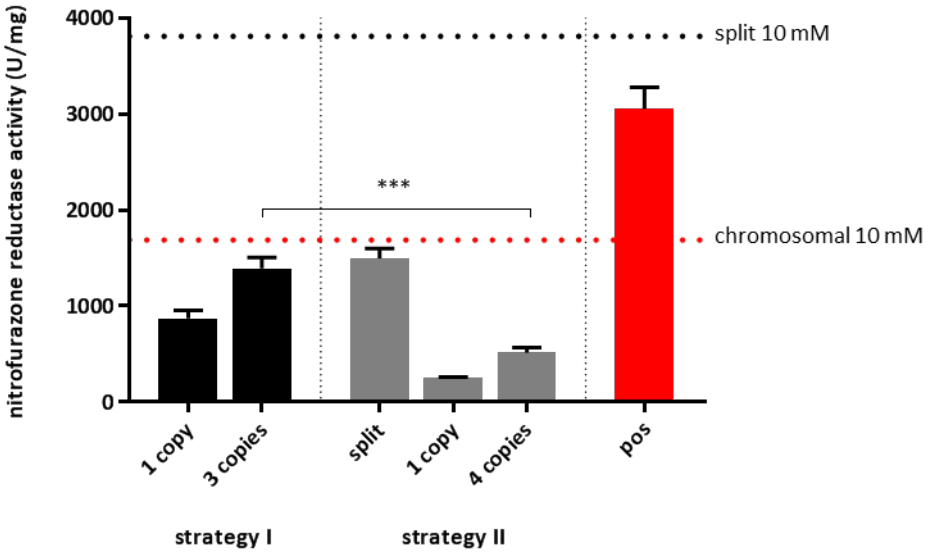


Figure 4.15 Direct comparison of the two overexpression strategies. Nitrofurazone/NADH reductase activity in cell lysates of the best performing strains from strategy I and II, alongside relevant reference strains, as indicated. Expression levels obtained with the P_{lac}-*tcdR*-P_{tcdB}-*yfkO-Bl* system under induction with 10 mM lactose in the split and chromosomal configurations are indicated as broken lines, representing the highest expression levels attainable in *C. sporogenes* with or without the use of a plasmid, respectively. Data represent mean values of three biological replicates, normalised for total protein content ± SD. Depicted data are from a single representative experiment out of three repeats. Statistical analysis was carried out using two-tailed unpaired t-test; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

4.6 Discussion

The prerequisite to the multicopy integration strategy was the identification of suitable genomic loci that can be utilised for the insertion of the expression cassette; this was fulfilled successfully, discovering chromosomal positions capable of yielding higher expression levels than the staple of integration loci in clostridial strains (namely *pyrE*) while imposing no physiological impact. Expression strength was higher for chromosomal positions more proximal to *oriC*, attributable to the simultaneous presence of multiple replication forks at different completion stages, which causes asymmetries in the copy number of genes across the genome; since replication always originates at a fixed position (*oriC*), genes closer to that position are replicated first, implying that their average copy number is, at any given time, likely to be higher than that of genes further away[236]. This gene dosage effect has been previously observed in several bacteria, including *E. coli*[237, 238], *S. typhimurium*[239] and *B. subtilis*[240], whereby expression was shown to vary up to a 5-fold range, depending on the location of the gene. The 2-fold fluctuation observed in the present study between the two chromosomal extremes (*glnQ1 vs treBAR*) is in line with a gene dosage effect, while the difference might be less pronounced given that samples were harvested at the beginning of stationary phase where growth and DNA replication proceed at reduced rates[238]. To a lesser extent, factors relating to local chromosomal DNA organisation, topology or read-through may also be underpinning the observed variability of expression strengths[238, 241, 242]. Importantly, the chromosomal position effect can be incorporated in the design of clostridial anticancer vectors, for instance, by selecting *glnQ1* or other positions closer to *oriC* as a means of achieving higher expression levels. Nevertheless, the extent to which perturbation of the chosen loci imposes the desired fitness disadvantage (section 4.5.1), will have to be explored preclinically.

It appears that under the present transcriptional and translational conditions, the optimal copy number for the expression of a heterologous gene is 2 to 3 copies, with

additional increases having a detrimental effect on expression levels and cell fitness, thus refuting the original hypothesis. While the precise optimal copy number is likely to be product-dependent and inevitably linked to the imposed rate of protein production[243], this trend is in line with outcomes of former studies, disclosing a bell-shaped relationship between gene dosage and protein yield[193, 244] as well as a negative impact on growth beyond a certain copy number[190, 191, 193]. Also consistent with previous studies is the small expression gain afforded by each additional copy, diverging significantly from the theoretical prediction and indicating that expression does not increase linearly with copy number[190, 193, 194, 244].

In this context, it is worth noting that localisation of the same expression cassette on a plasmid (pMTL82151-Pt_{hl14}-*yfkO_Bl*; the positive control) is able to generate significantly higher yields than the multicopy integration strategy (section 4.5.7), as previously commented[190]; this incongruity may be due to the striking difference in copy number (3 *vs* presumably >40 for the pBP1 replicon) or due to host-plasmid interactions which involve dramatic 'reprograming' of the cell metabolism[245-247].

The discrepancies between gene copy number and protein yield may be partly attributed to a gradual saturation of the expression mechanism, with translation being the initial limiting factor, capped by the cell's 'free ribosome pool'[248-251]. This is often reflected as a transient increase in the mRNA levels of the gene of interest but without a corresponding rise in protein yield[194, 195]. Crucially, the marked negative impact on growth, accompanied by the decline in expression levels upon the addition of a 4^{th} copy (Figure 4.8B), is indicative of excessive cellular burden[252]. High burden may be a consequence of the utilisation of the cell's finite gene expression resources, and, depending on the function of the corresponding protein product, of the perturbation of cellular metabolism (for instance, if protein function requires consumption of cellular metabolites or cofactors)[253]. Specifically, it has been shown that high cellular burden depletes the cell's total capacity for gene expression[251], driving re-allocation of cellular resources and consequent downregulation of rRNA synthesis. This process is known as

the 'stringent response' and imposes a negative impact on growth rate[248, 254]. In turn, growth rate reduction causes a decline in the ribosome content of the cell, which decreases the yield of the protein of interest[255, 256].

In addition to the negative impact on cell growth rate, the unnatural burden imposed by the expression of foreign proteins may lead to generalised undesirable physiological changes, including increased mutation rates, unstable performance and high susceptibility to suboptimal growth conditions[257]. Therefore, gene expression of heterologous genes must be subject to cost-benefit optimisation, whereby the establishment of optimal rather than maximal expression is of essence[258]. Importantly, design of heterologous gene expression can be optimised to prevent high-burden systems, for example, by selecting weaker RBS sequences[251, 257].

Nevertheless, tuning of expression should be done in consideration of the desired target expression level; in fact, this may be lower than the theoretical maximum expression that could be afforded by the cell. For instance, upon *in vivo* evaluation of a clostridial anticancer strain, it might be revealed that higher therapeutic benefit can be derived by a population of cells with robust growth rate and modest production of a desired anticancer product *vs* a population of cells engineered to overexpress the anticancer product at maximum capacity, compromising cell fitness and growth rate[259]. It is also possible that the quantity of anticancer product produced by cells carrying, for example, one copy of the gene of interest is adequate for clinical response, thus negating the need for complex genetic engineering.

At any rate, an important consideration with multicopy strains is the segregational stability of their chromosome, even in the absence of integrated tandem repeats; given the presence of multiple homologous sequences, intrachromosomal homologous recombination is still likely, leading to genomic re-organisations, strain instability and/or alteration of the strain's properties[191, 192, 244]. Although integration loci can be prophylactically spread out throughout the chromosome (such as in the present study), it may be difficult to eliminate this risk entirely, given that due to topological effects of

the DNA, chromosomal distance does not always correlate with physical distance[238, 260]. Therefore, for a clinical implementation, extensive testing of the strain over multiple generations may be necessary as a form of quality control[190].

Although the TcdR-P_{tcdB} system did not yield the anticipated expression levels when implemented in a pertinent design, that is, on the chromosome and under the control of a constitutive promoter, its high flexibility implies that there is a large window of optimisation opportunities that can be pursued. Importantly, the WT-like growth profile of the multicopy engineered strains may suggest that the system is not yet operating at full capacity. In this context, a stronger promoter than P_{ptb} could be employed, of an expression strength equivalent to P_{lac} under induction with 10 mM lactose (that is, higher than P_{ptb} but lower than P_{fdx}), or alternatively, the *tcdR* module could be integrated in a locus closer to *oriC*. In the presence of >1 P_{tcdB}-driven gene copies, such a change may generate large improvements without laborious engineering.

However, the location of the sigma factor relative to its target gene(s) emerged as a crucial determinant of expression (section 4.5.6), compromising significantly the derivable expression output of the TcdR-P_{tcdB} system. This is not surprising given that due to the spatial and temporal coupling of transcription and translation in prokaryotes, co-localisation of the transcription factor next to its target gene(s) seems to favour the rapid activation of transcription[261]. This is particularly relevant for small or dedicated regulons (such as the *tcdR*, especially in an orthogonal context) where the relative cellular abundance of the transcription factor is small, meaning that the search for target sites is usually realised via a sliding/hopping mechanism[260] (NB. although not a classical transcription factor, TcdR may be considered as a transcriptional activator). This finding is also supported by the conserved organisation of the *tcdR*-like regulons in clostridial strains whereby *tcdR* is always upstream of its cognate genes (section 4.1.3)[222].

Independently of the system's applicability as a chromosomal overexpression strategy, it may hold promise for applications where chromosomal integration of the gene of

interest is not a strict requirement. Also, utilisation of the P_{tcdB} promoter offers a reliable way of overcoming cloning difficulties in *E. coli*, which have been a substantial caveat of the present study, given the nature of proteins employed and the frequent use of constitutive promoters. Therefore, it would be interesting to study the dynamics of the system in more detail, to inform the design of future implementations. Further experimental evidence is also required to verify the hypothesis proposed in section 4.3 as well as the assertions discussed in section 4.5.4, regarding the different configurations of the system and their corresponding operating modes.

For this purpose, the P_{lac} inducible system could be utilised again, to permit easy adjustment of the TcdR dose. P_{tcdB} could be coupled, instead of *yfkO_Bl*, to an innocuous reporter gene functional in Clostridia (such as *catP*), isolating the system's performance from potential YfkO_Bl-associated toxicity. Additionally, TcdR in cell lysates could be quantified using a suitable ELISA assay, to enable direct measurement of TcdR expression in parallel to the downstream reporter. These experiments would also benefit from concomitant RT-qPCR studies for the determination of *tcdR* and *catP* reporter transcript levels; having an insight in both transcription (via RT-qPCR) and translation (via reporter expression assays) processes would be highly informative of the dynamics of the system as well as its associated bottlenecks for expression. Furthermore, in the split and chromosomal configurations, inducing concentrations higher than 10 mM of lactose could be employed to identify the point at which additional supply of TcdR stops conferring higher expression, relative to the availability of P_{tcdB} binding sites. In this respect, it would be useful to identify, via RT-qPCR, the exact copy number of the replicon employed (pBP1) in *C. sporogenes*.

Finally, it would be useful to test the functionality of different TcdR homologues in *C. sporogenes*, as well as their cognate promoters; for instance, BotR from *C. botulinum* in combination with P_{ha33} or P_{ntnh} might result in higher expression levels than the TcdR-P_{tcdB} combination. Further variations could arise from creating chimeric promoters; for example, the TATA boxes of characterised constitutive promoters (such as P_{fdx} and P_{thl})

could be substituted for those recognised specifically by sigma factors of the TcdR family, generating expression signals with varied expression strengths. Furthermore, randomization of the spacer between the −10 and −35 conserved nucleotide regions could also yield variability in expression[225].

A predictable outcome of the present study was that protein yield may increase following induction of expression at the mid-exponential phase, compared to constitutive expression (section 4.5.5). This could be further explored experimentally (for example, by inducing cultures at the time of inoculation), as it might transpire that the benefit is exclusive to a TcdR context, given that competition between TcdR and endogenous sigma factors for RNAP[249] may be imposing an undue burden upon early TcdR expression, but might not be relevant otherwise. Nevertheless, decoupling of cell growth from foreign protein production is known to minimise cellular burden, allowing cells to accumulate biomass prior to redirecting cellular resources to the biosynthesis of desired products[262-264]. To this end, auto-inducing media with defined compositions of sugars have also been developed, taking advantage of the diauxic growth characteristics of bacterial strains carrying the lac operon regulatory elements (lacO and LacI)[265-267].

However, given that the precise composition of the tumour microenvironment cannot be predicted or manipulated, inducibility without the need of exogenous inducers would have to be introduced on the genetic level. In this case, expression of the gene of interest could be coupled to a promoter which is natively switched on at the mid log phase in *C. sporogenes*. A suitable promoter could be identified from transcriptomic data, but extensive characterisation would be required to establish environmental factors involved in its regulation, especially in the environment of a tumour. Recently, more sophisticated dynamic sensor and regulatory systems have also been designed, enabling precise and reliable growth-dependent switching of protein production. For example, growth phase-sensor and substrate-sensor modules have been combined in a circuit that switches on gene expression only when high biomass has been acquired and a target substrate required for biosynthesis is also available[264]. Elsewhere, the cell's native

quorum sensing mechanism has been re-wired and coupled to the production of target proteins[263]. In the future, similar autonomous designs could be implemented in *C. sporogenes*, allowing the benefits of inducible expression to be harnessed, without the need for exogenous induction.

4.7 Key Outcomes

- Two alternative overexpression strategies were devised and implemented in *C. sporogenes*.
- Multicopy integration of a gene of interest elevated significantly the yield of an already well-expressed protein, when three copies of the gene were present on the chromosome. A further increase in gene copy number was not beneficial.
- Utilisation of the orthogonal TcdR-P_{tcdB} system was inferior to multicopy integration when implemented chromosomally but may be valuable for research purposes or applications where chromosomal integration of the gene of interest is not required.

CHAPTER 5

Piloting a novel functionality: immunotherapy

5.1 Background

5.1.1 The immune system in cancer: friend or foe?

It is now accepted that the immune system, and particularly the inflammatory state, can significantly impact on tumour development and progression. On the one hand, chronic unresolved inflammation has a tumour-promoting role via the channelling of wound-healing growth and angiogenic factors that (unintentionally) also nurture the tumour. On the other hand, acute inflammatory processes, such as those associated with pathogen clearance, activate immune effector mechanisms that can eliminate cancer cells[268].

The process of 'immunoediting' describes the dynamic interactions between the immune system and the tumour, and provides an explanation for the escalating capability of

tumours to evade the immune system. Immunoediting proceeds in three stages: elimination, equilibrium and escape. At the early stages of cancer, transformed cells may be identified and eliminated via the routine immunosurveillance processes and, as a result, tumour formation could be entirely prevented (elimination). Unfortunately, some – craftier – cancer cells with diminished immunogenicity are bound to survive the immunoediting process and be left behind (equilibrium). Under this selective pressure, new populations of tumour cells derived from these renegade cells propagate the ability to evade the immune system and, as a result, growing tumours progressively become invisible to the immune system (escape)[269]. Reasonably, evasion from immune destruction has recently emerged as a hallmark of cancer[5].

The immunoediting theory is grounded on the intrinsic ability of the immune system to operate as a barrier in cancer formation and progression, as evidenced by preclinical and clinical research findings[5]. Importantly, the presence of tumour-infiltrating cytotoxic T cells and natural killer (NK) cells is considered a positive prognostic marker in many human cancers[270]. Nevertheless, the ability of cancer cells to evade the immune system eventually outweighs the ability of the immune system to eliminate them; cancer cells, with their ever-changing phenotype, manage to 'manipulate' and recruit immune cells to their advantage, via the secretion of appropriate immunomodulators. Under the cancer's influence, cells of both the innate and adaptive arms of the immune system commence inefficient and/or irrelevant functions, consistent with chronic inflammation and wound healing rather than acute inflammatory processes[5, 271]. This also modulates the behaviour of non-immune tumour companion cells (such as stromal cells, endothelial cells, fibroblasts and others), always in favour of tumour progression[272]. Certainly, tumour-promoting and tumour-antagonising processes can take place concomitantly, with the eventual outcome being dictated by the combination/domination of different effectors and messengers[273]. This dichotomy is evident in all aspects of the immune system whereby subsets of cells (such as macrophages) as well as individual cytokines (such as TNF-α) can exert either pro- or anti-tumoural functions (or both)[87, 274].

Cancer-imposed immune dysfunction is reflected in the behaviour of many cell subtypes; for instance, tumour associated macrophages (TAMs) are skewed towards the pro-tumoural M2-like phenotype[274], while tumour-resident dendritic cells (DCs) experience maturation defects and are unable to mediate antigen-specific activation of effector T cells[275]. Additionally, tumours recruit regulatory T cells (Treg) and myeloid-derived suppressor cells (MDSCs) which are actively immunosuppressive, while potentially useful cells including NK and cytotoxic T cells are not engaged[275]. The immunosuppressive microenvironment is underpinned by an array of secreted cytokines, chemokines and chemicals (or their absence) that not only divert the immune system, but also directly promote tumour growth. These include the production of nitric oxide which downregulates T-cell signalling[274], the generation of reactive oxygen species which amplify mutagenicity[276], the supply of growth and survival factors that sustain proliferation and inhibit apoptosis, and the secretion of matrix-modifying enzymes and pro-angiogenic factors that promote invasion and metastasis[5]. Some of these factors have pleiotropic pro-tumoural functions including VEGF, that enhances angiogenesis and vascular permeability and at the same time attracts Treg cells and impairs DC maturation[275, 277]. Also, IL-6, IL-10, and transforming growth factor β (TGF-β) are pro-tumoural cytokines found in abundance, while beneficial cytokines such as IL-2, IL-12 and IFN-γ are depleted[274, 275].

5.1.2 Cancer immunotherapy using bacteria

Since all tumours are potentially immunogenic, the problem culminates in a 'paralysed' or ineffective endogenous immune response. Consequently, the organising principle of cancer immunotherapy is that the immune system can be stimulated by an exogenous intervention and driven against the tumour, rectifying the pre-existing immune status, and tipping the balance in favour of tumour-destructive processes[273]. On a cellular level, effective antitumour immunity, involving the participation of the adaptive immune system, must proceed in three stages. Firstly, DCs need to acquire tumour antigens and be exposed to a strong maturation signal. Secondly, mature DCs need to navigate to

the lymph nodes and activate protective T-cell responses, including cytotoxic $CD8^+$ T cells; NK and B-cell responses may also be elicited following DC maturation. Finally, activated lymphocytes need to enter the tumour bed and mediate their actions against cancer cells, overcoming the immunosuppressive environment[275, 278].

Considering this biological landscape, the use of bacteria in cancer immunotherapy seems an obvious one if one considers them in their intrinsic role as immune system 'intruders'. In fact, the concept of cancer immunotherapy finds its roots in William B. Coley who, in 1891, inspired by the spontaneous remission of a sarcoma in a patient who was also suffering from erysipelas, devised intratumoural injections of live or attenuated *Streptococcus pyogenes* and *Serratia marcescens* (known as 'Coley's toxins') as a means of treating cancer[279]. In the same tenet, the intravesical administration of the Bacillus Calmette-Guérin (BCG) vaccine, originally developed for the prevention of tuberculosis, has now been established as the standard of care for superficial bladder tumours[280].

More recently, the immunotherapeutic potential of a wide range of bacterial strains, including *Salmonella*, *Listeria* and *Clostridium* species, has been closely investigated. It is now understood that the intratumoural presence of bacteria can generate an acute inflammatory response, potentially perturbing many aspects of the immunosuppressive environment. In fact, the elicited immune response is perceived as the principal mechanism of action of some bacterial anticancer therapies[55, 281]. The enhanced anti-tumour immunity is consistent with the recruitment of lymphocytes including $CD4^+$ (helper) and $CD8^+$ (cytotoxic) T cells, an M2 to M1 phenotype switch in TAMs and the stimulation of TLR signalling pathways which go together with the maturation of DCs, activation of NKs and downregulation of immunosuppressive cytokines[282]. The acute inflammatory reaction is a consequence of the inherent immunogenicity of foreign bacteria, being plentiful sources of structurally-conserved pathogen-associated molecular patterns (PAMPs) which serve as toll-like receptor (TLR) agonists[283].

Nevertheless, activation of the immune system poses a trade-off for bacterial cancer therapies, given that the immune system will be stimulated not only against the cancer but also against the bacterial vector itself, leading to bacterial clearance or inactivation and limiting the possibility of repeated administration[20]. However, the degree to which this occurs and its impact on the treatment outcome needs to be further researched. For instance, it has been shown that *C. sporogenes* and *C. novyi*-NT vegetative cells can maintain high densities of tumour colonisation for an extensive period of time (up to 14 days) in mice, rats and dogs[24, 25, 46].

The natural ability of bacteria to activate the immune system can be augmented via genetic engineering[284], using invasive bacteria (such as *Salmonella* or *Listeria*) which introduce the DNA into the host cells, or non-invasive strains, for example *Clostridium*, *Bifidobacterium* and *E. coli* Nissle, which express the DNA autonomously. Nevertheless, plasmid DNA carried by non-invasive bacteria may also be transferred to immune, phagocytic cells of the host, following bacterial engulfment[285]. The delivery of immunomodulating DNA by bacterial cells offers many advantages. Firstly, the natural tumour-homing preferences of bacteria will localise expression of the immunomodulator in the tumour, without the need for intratumoural administration. Importantly, *in situ* production of the therapeutic agent can bypass toxicities that might arise from systemic delivery (particularly relevant in the case of cytokines)[273], without compromising efficacy[286]. Indeed, local immunotherapy has been shown to induce strong anti-tumour immunity even against metastatic deposits where immunotherapy has not been applied (the abscopal effect)[287]. Also, production of the immunomodulator by the bacterial or host cells, is advantageous for agents that would be difficult and/or expensive to produce and purify for administration as protein drugs[288]. Finally, other bacterial products (for example, exotoxins, proteases) may mediate direct cell killing of cancer cells, acting synergistically with the delivered immunomodulator[282]. In view of the above advantages, bacteria have been extensively engineered, predominantly for the delivery of relevant cytokines.

Salmonella strains have been engineered to express a range of immunomodulators including IL-2[289-291], IFN-γ [292], TNF-α[284, 293], LIGHT (tumour necrosis factor superfamily member 14)[288], IL-18[294], chemokine (C-C motif) ligand 1 (CCL1)[295, 296] and CD40 ligand (CD40L)[297]. *Listeria* strains, considered the prototype bacteria for cancer immunotherapy, can mediate strong antitumour immunity of their own accord but have similarly been exploited for the delivery of tumour associated antigens, including EGFRvIII[298], mesothelin[299], HPV-16-E7[300], NY-ESO1[298], PSA[301] and HER2/neu[302] epitopes. Importantly, the IL-2 expressing *Salmonella* strain and most *Listeria* strains are already undergoing clinical evaluation.

Comparatively, engineering efforts for the incorporation of immunotherapeutics in *Clostridium* have been less sophisticated (probably due to the hitherto limited availability of genetic tools). These include the genetic engineering of *C. acetobutylicum* to secrete mouse TNF-α (mTNF-α)[41, 82, 87-89] and rat IL-2 (rIL-2)[303] as well as of *C. sporogenes* to secrete mouse IL-12 (mIL-12)[304].

Albeit limited, the above studies provide encouragement for the introduction of an immunotherapeutic functionality in *C. sporogenes*. Stemming from the interest of one of the group's collaborators from Mayo clinic (Dr Khashayarsha Khazai; Minnesota, USA), the incorporation of GM-CSF and IL-4 in the clostridial vector was investigated.

5.1.3 Granulocyte Macrophage-Colony Stimulating Factor

Granulocyte Macrophage-Colony Stimulating Factor (GM-CSF) is a haematopoietic growth factor that promotes the differentiation of BM myeloid progenitor cells into monocytes (macrophages and dendritic cells) and granulocytes (neutrophils, eosinophils, and basophils), as well as their mobilisation into the bloodstream[305, 306]. The homeostatic basal levels of GM-CSF are low, but they increase rapidly in response to infection or injury[306, 307], being produced by a variety of cells including leukocytes, endothelial cells and fibroblasts[308].

In the context of cancer, GM-CSF (sargramostim; LEUKINE®) has been in clinical use for almost three decades, with its initial therapeutic indications including the acceleration of myeloid recovery following chemotherapy or bone marrow transplantation and the mobilisation of peripheral hematopoietic progenitor cells for stem cell transplantation[309]. More recently, GM-CSF has been incorporated in various anticancer vaccines as an adjuvant, being a component of the marketed sipuleucel-T (Provenge®)[310] and talimogene laherparepvec (IMLYGIC®)[311] treatments for advanced prostate cancer and unresectable melanoma, respectively.

The antitumour effect of GM-CSF is primarily attributed to the local recruitment and activation of DCs, facilitating tumour-specific antigen presentation, T-cell activation and secretion of immunostimulatory cytokines[312]. Immune-independent tumour suppressive effects have also been documented, including cell cycle progression blockage from G0/G1 to the S phase[305]. Nevertheless, the clinical findings to date suggest that the therapeutic benefit derived by GM-CSF is maximised in conjunction with other therapeutic modalities[313], justifying its incorporation in the clostridial vector. Additionally, the inclusion of GM-CSF in already marketed anticancer vaccines, could facilitate regulatory claims in the future, by virtue of a similar mechanism of action.

5.1.4 Interleukin 4

Interleukin 4 (IL-4) is a pleiotropic cytokine, primarily recognised as a B, T and mast cell growth factor and a key mediator of type 2 T helper cell (T_H2) differentiation[314]. It is also involved in the proliferation and differentiation of myeloid progenitors into mature DCs[315]. IL-4 is largely produced by T_H2 cells (thus acting in a positive feedback loop), but also by NK cells, basophils, mast cells and eosinophils[314]. Given that T_H2-polarised immunity generally favours tumour growth[316], it may seem paradoxical that IL-4 is considered for cancer therapy. In fact, IL-4 has been associated with pro-tumoural mechanisms, including impairment of the cytolytic capacity of $CD8^+$ T cells[317], promotion of MDSC survival[274] as well as TAM reprogramming towards pro-tumoural

phenotypes involving the production of WNT7β[274] and cathepsins[318]. Critically, the main source of IL-4 in tumours is cancer cells themselves, whereby IL-4 acts in a paracrine manner[318].

Nevertheless, IL-4 is a good example of the dichotomous role of the immune system in cancer, as it has also been shown to reduce the tumorigenicity of tumour cell lines in *in vivo* transplantation assays as well as to inhibit the growth of established tumours[319, 320]. Its antitumour activity was initially attributed to the generation of a rapid (innate) inflammatory reaction and the influx of cytocidal eosinophils at the site of the tumour[320]. Although the role of eosinophils *per se* became the subject of subsequent conflicting findings, it was recently established that eosinophil-dependent antitumour immunity is mediated indirectly, through the facilitation of $CD8^+$ T cell influx, normalisation of the tumour vasculature and reprograming of TAMs towards the M1 phenotype[321]. Additionally, IL-4 has been shown to mediate the activation of tumour-infiltrating DCs[322], inhibit cancer cell growth through p21-mediated activation of STAT6[323] and exert a negative role on angiogenesis[324].

It is now believed that the overall pro- or anti-tumoural effects are dependent on the timing, concentration, location and source of the cytokine; in particular, early administration of high doses of exogenous IL-4 (as opposed to cancer-cell derived IL-4) prevents the differentiation of stoma-resident myeloid progenitor cells into MDSCs, thereby attenuating their associated immunosuppressive capabilities[315, 317]. Plausibly, the delivery of IL-4 using a clostridial vector may be an appropriate strategy in exploiting exclusively the anti-tumoural effects of the cytokine.

5.2 Objectives

- Determine the capability of *C. sporogenes* to tolerate, express and secrete GM-CSF and IL-4.

- Test the efficiency of different secretory signals and assess the extent of extracellular proteolysis.
- Confirm the biological activity of *C. sporogenes*-derived GM-CSF and IL-4.

5.3 Hypothesis

- *C. sporogenes* NCIMB 10696 can express and secrete biologically active GM-CSF and IL-4.

5.4 Study design

Expression and secretion efficiencies were assayed using the expression plasmid pMTL82121. The general vector design, as employed for the experiments in this chapter, can be found in Appendix IV. The expression plasmids were assembled as described in section 2.2.11.8 and the host strains were created as described in Chapter 4.

5.4.1 Genes

To enable downstream preclinical assays, the gene sequences of mouse GM-CSF (*mGMCSF*) (NP_034099.2) and mouse IL-4 (*mIL4*) (NP_067258.1) were employed, with the intention of assaying the constructed strains *in vivo*, in mice. The gene sequences corresponding to the mature protein product (without the native signal peptide) were codon-optimised for *C. sporogenes* to ensure efficient translation, according to section 2.6.1. The codon-optimised sequences can be found in Appendix VI. A FLAG (DYKDDDDK)-tag was also fused at the 3' end of the genes, before the stop codon, to enable immunoblot detection.

5.4.2 Selection of expression signal

Given the desire for high constitutive expression of the immunomodulator but also the lack of prior knowledge regarding the potential toxicity of these heterologous proteins

when expressed in *C. sporogenes*, the use of the medium strength promoter, P_{ptb}, was favoured, as chosen from the in-house promoter library. Unfortunately, some of the desired constructs could not be cloned successfully in *E. coli* when driven by P_{ptb}, despite the use of the p15a low-copy number Gram-negative replicon. To overcome the cloning problems, P_{ptb} was replaced by P_{tcdB} which is not recognised in *E. coli*. The P_{tcdB}-driven mGM-CSF constructs were conjugated in the Csp-Δ*pyrE*::T1-P_{ptb}-tcdR strain which expresses the TcdR sigma factor, required for the initiation of transcription, constitutively. The P_{tcdB}-driven mIL-4 constructs were conjugated into the Csp-Δ*pyrE*::P_{lac}-*tcdR* strain, in which *tcdR* is minimally expressed in the absence of lactose; this approach was taken due to problems encountered in obtaining stable transconjugant colonies in the Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* strain, presumably due to IL-4 intolerance.

It should be noted that *E. coli* has been previously employed for the expression recombinant GM-CSF of different origins (for example, ovine and human), including molgamostrim, a marketed version of human GM-CSF. In published examples, expression has been engineered to occur under induction, in response to exogenous stimuli, such as nitric oxide[325], IPTG[326-328], salt[329] and auto-induction media[330]. Occasionally, expression levels have been reported to be low, due to poor cytoplasmic stability of the protein and/or toxicity-associated cell lysis following induction[326]. Additionally, GM-CSF is known to aggregate in insoluble form, requiring the use of tags (for instance thioredoxin) to enhance protein solubility[328] or specialised strains[329]. Likewise, recombinant IL-4 has been expressed in *E. coli* strains primarily under induction, in response to nalidixic acid[331], IPTG[332, 333] and temperature[332], but also constitutively[334]. In all cases, the protein was expressed as inactive insoluble aggregates, requiring protein refolding. To our knowledge, GM-CSF and IL-4 expression has not been reported in any other bacterial strain to date. Nevertheless, commercially available preparations of GM-CSF and IL-4 (for instance, by Sigma-Aldrich) may have been expressed in different (undisclosed) bacterial expression hosts.

5.4.3 Selection of secretion signal

For an effective immune response, the immunotherapeutic effector must leave the bacterial cell and exert its actions in the tumour microenvironment and/or join the systemic or lymphatic circulation. Although it can be safely assumed that a degree of bacterial cell lysis will be occurring in the tumour, to achieve reliable exit of the molecule throughout the bacterial growth cycle, the molecule needs to be secreted.

Secretion via the general bacterial secretory pathway (Sec) is achieved by incorporation of suitable signal peptides at the amino terminal of proteins. Proteins destined for secretion are thus synthesised as pre-proteins, with their mature form being released upon cleavage of the signal peptide by a membrane-bound signal peptidase during protein translocation across the cell membrane[335]. Employing a Gram-positive organism such as *C. sporogenes* is advantageous given that the secreted protein is directly released in the extracellular environment (as opposed to the periplasmic space)[336]. The signal peptide has a decisive role on the final extracellular yield of the secreted protein; a given signal peptide may be less or more efficient in combination with different proteins or when utilised in different expression hosts[337]. In view of this, identification of suitable signal peptides for a protein of interest in a desired host is best achieved via high-throughput screening of large signal peptide libraries[337]. Such an endeavour was out of the scope of the present thesis, and instead, the signal peptides employed were selected rationally. To guarantee secretory efficiency in *C. sporogenes*, the native murine signal peptides of mGM-CSF and mIL-4 were replaced by signal peptides of clostridial origin.

Npr is a signal peptide from *C. sporogenes* associated with the secretion of an M4-like metalloprotease (AKC62190.1), an abundant extracellular protein in the supernatant of WT *C. sporogenes* NCIMB 10696[338]. Earlier in-house use of this signal peptide with heterologous proteins in *C. sporogenes* had yielded good secretory efficiencies[338]. Lip is the signal peptide of a highly-secreted lipase (ABK60864.1) from *C. novyi*-NT[54], but has not been employed for the secretion of heterologous proteins before. Finally, EglA

is the signal peptide of a beta-1,4-endoglucanase precursor (AAA23230.1) from *C. saccharobutylicum* P262 and has been the signal peptide of choice in the majority of engineered clostridial strains for cancer therapy, whenever secretion was desired[41, 42, 82, 87, 88, 303, 304]. To assess the functionality of the chosen signal peptides in combination with the mGM-CSF and mIL-4 downstream protein sequences, the signal peptide prediction program SignalP 4.1 was used as described in section 2.6.10. In a given presecretory protein sequence, SignalP 4.1 can discriminate signal peptide from non-signal peptide sequences as well as identify the signal peptide cleavage site[111]. The reported D-score indicates the likelihood of a given sequence to contain a signal peptide; the cut-off value below which a signal peptide cannot be assigned is 0.450. The D-scores of the selected signal peptides and the predicted cleavage sites are summarised in Table 5.1.

Table 5.1 The selected signal peptides

Signal peptide	Originating organism	Associated protein (NCBI reference)	Size (aa)	Native cleavage site	Native D-score†	New cleavage site	New D-score†
Npr	*C. sporogenes* NCIMB 10696	M4 peptidase (AKC62190.1)	25	A-A	0.794	mGM-CSF: A-A mIL-4: A-H	0.737 0.611
Lip	*C. novyi*-NT	lipase (ABK60864.1)	28	A-T	0.805	mGM-CSF: A-A mIL-4: A-H	0.828 0.731
EglA	*C. saccharo-butylicum* P262	beta-1,4-endoglucanase precursor (AAA23230.1)	38	A-A	0.628	mGM-CSF: A-A mIL-4: A-H	0.570 0.498

aa: amino acids; D-score: indicates the likelihood of a given sequence to contain a signal peptide, where D>0.450 for signal peptide assignment

Given that the selected signal peptides originated from closely-related clostridial strains with AT-rich genomes and similar codon preferences, their sequences were used without codon optimisation. Still, it is worth noting that codon usage in signal peptides of the Sec pathway is crucial, dictating translational kinetics which allow the secretory machinery to interact efficiently with presecretory proteins[339]. Non-secreted mGM-CSF and mIL-4 controls were also constructed by fusing the P_{tcdB} promoter with the gene sequences of the mature proteins, with the addition of a start codon (AUG) between them.

5.5 Results

5.5.1 *C. sporogenes* as an expression host for mGM-CSF and mIL-4

The mGM-CSF and mIL-4 expression plasmids were introduced in the host strains, Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* and Csp-Δ*pyrE*::P_{lac}-*tcdR* respectively, by conjugation. Successful transconjugants were confirmed by colony PCR with the M13F and M13R primers (Figure 5.1). The resulting mGM-CSF-expressing strains will be referred to as: npr-G, lip-G, egla-G and non-sec-G for the Npr, Lip, EglA signal peptides and the non-secreted control, respectively. Likewise, the mIL-4-expressing strains will be referred as: npr-I, lip-I, egla-I and non-sec-I. All generated strains are listed in Table 5.2. Only one stable transconjugant colony could be obtained for npr-I even after several conjugation attempts.

Table 5.2 Generated strains harbouring mGM-CSF- and mIL-4-expressing plasmids

Strain	Description	Harboured plasmid	Abbreviated name
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus	pMTL83121-P_{tcdB}-*npr-mGMCSF*	npr-G
		pMTL83121-P_{tcdB}-*lip-mGMCSF*	lip-G
		pMTL83121-P_{tcdB}-*egla-mGMCSF*	egla-G
		pMTL83121-P_{tcdB}-*mGMCSF*	non-sec-G
Csp-Δ*pyrE*:: P_{lac}-*tcdR*	P_{lac}-*tcdR* integrated in the *pyrE* locus	pMTL83121-P_{tcdB}-*npr-mIL4*	npr-I
		pMTL83121-P_{tcdB}-*lip-mIL4*	lip-I
		pMTL83121-P_{tcdB}-*egla-mIL4*	egla-I
		pMTL83121-P_{tcdB}-*mIL4*	non-sec-I

Given the cloning problems encountered in *E. coli*, the difficulty in obtaining mIL-4 transconjugants, the specific nature of the proteins as well as the lack of prior experience in our group regarding the expression of proteins of non-bacterial origin in clostridia, confirmation that mGM-CSF and mIL-4 expression in *C. sporogenes* is not toxic to the

host was sought. Therefore, the growth of transformed strains was monitored over a period of 24 hrs, reasoning that any potential toxicity would primarily impact cell growth[340]. The strains were assayed against a negative control, consisting of *C. sporogenes* transformed with an empty pMTL82121 plasmid (neg). In the case of mIL-4, the cultures were also induced with varying concentrations of lactose (0.1 – 10 mM). Additionally, the integrity of the constructs was confirmed by colony PCR and sequencing before and after growing the cells, to rule out toxicities manifesting as plasmid instability or sequence alterations.

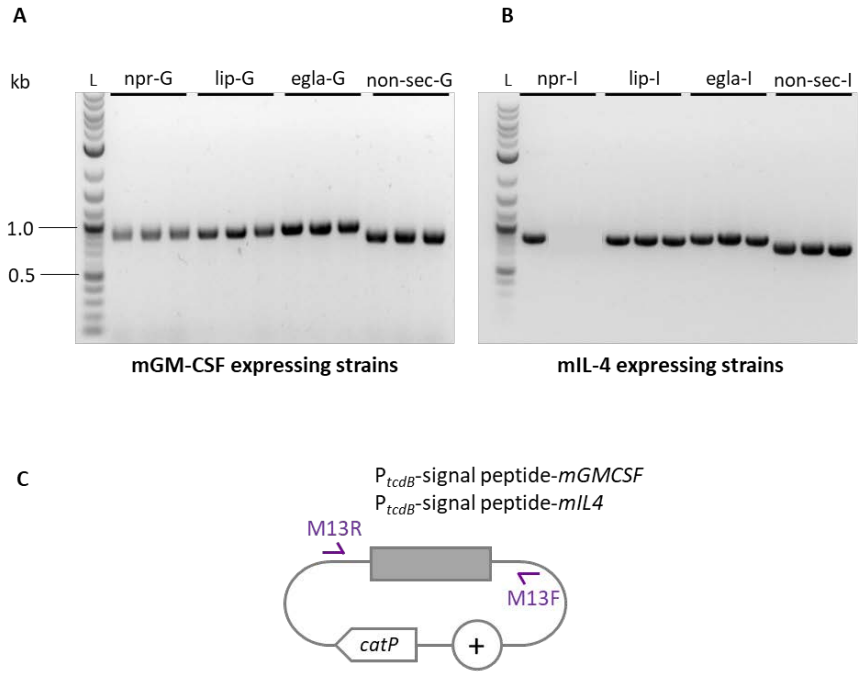


Figure 5.1 Construction of mGM-CSF- and mIL-4- expressing strains. Colony PCR on strains harbouring the mGM-CSF and mIL-4 expressing plasmids**. A.** mGM-CSF expressing strains: npr-G (877 bp); lip-G (886 bp); egla-G (916 bp); non-sec-G (805 bp). **B.** mIL-4 expressing strains: npr-I (865 bp); lip-I (874 bp); egla-I (904 bp); non-sec-I (793 bp). L: NEB 2-log DNA ladder. **B.** Schematic diagram showing the M13F/M13R primers used for colony PCR screening and their binding sites (in purple).

The growth rate and total cell density of the mGM-CSF-expressing strains were identical to those of the negative control (Figure 5.2). Additionally, the constructs were stable after being grown for 24 hrs and no sequence alterations occurred, suggesting that mGM-CSF is well tolerated at the given expression level.

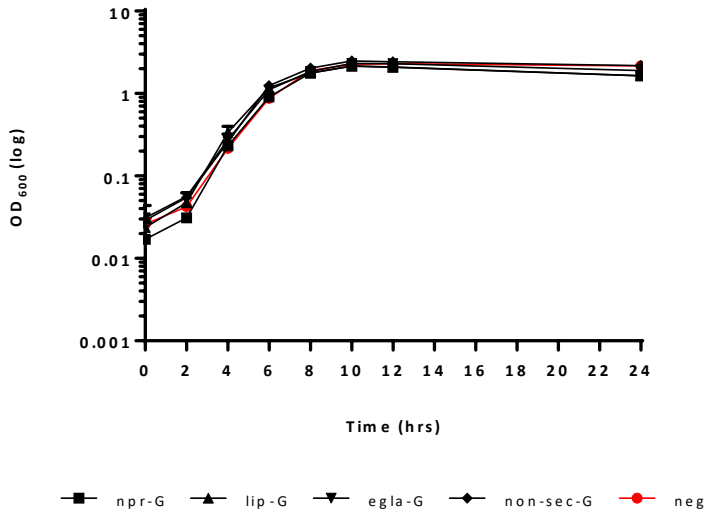


Figure 5.2 Growth curves of mGM-CSF expressing strains. mGM-CSF-expressing strains were grown in liquid TYG/thiamphenicol medium, alongside a strain carrying an empty plasmid (negative control; neg). OD_{600} was monitored over 24 hrs. Data represent mean values of three biological replicates ± SD, from a single representative experiment out of three repeats.

In contrast, the mIL-4-expressing strains exhibited erratic behaviours on different occasions. It was observed that when the nucleotide sequence of the constructs was preserved, there was a negative impact on the growth rate and total cell density, particularly with the Npr and Lip signal peptides (Figure 5.3A). However, the same strains were able to maintain a growth rate and final cell density similar to the negative control if mutations had been acquired (Figure 5.3B). The mutations consisted of point-nonsense mutations in the *mIL4* coding sequence (npr-I_1, lip-I_2) and base substitutions in the highly conserved TATA boxes of the P_{tcdB} promoter[218] (lip-I_1, lip-I_3) (Figure 5.4). Interestingly, the mutations were present in cultures that had not

undergone induction, implying that even the low level of mIL-4 expression/secretion resulting from the 'leakiness' of the lactose promoter in the chromosome is poorly tolerated. In the cases where sequence integrity was preserved (npr-I, lip-I, egla-I in Figure 5.3C and egla-I in Figure 5.3D), induction with lactose (1.0 mM) was detrimental, manifesting as a temporary but sharp drop in OD_{600}, consistent with cell death. Induction with lower concentrations of lactose (0.01 - 0.05 mM) was also attempted, but a portion of the cell population always perished (results not shown).

Finally, the growth rate and final cell density of the non-secreted control (non-sec-I) was similar to the negative control in all cases (Figure 5.3A-D), suggesting that mIL-4-associated toxicity is probably linked to the secretion process rather than the protein itself. It is also possible, but less likely, that the protein exerts a harmful effect extracellularly, so only when it is secreted.

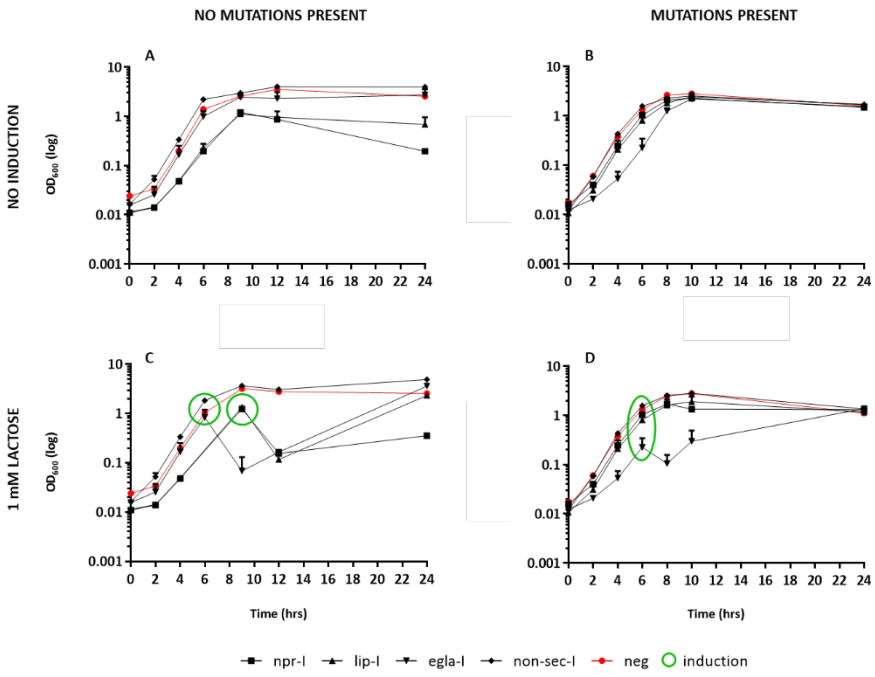
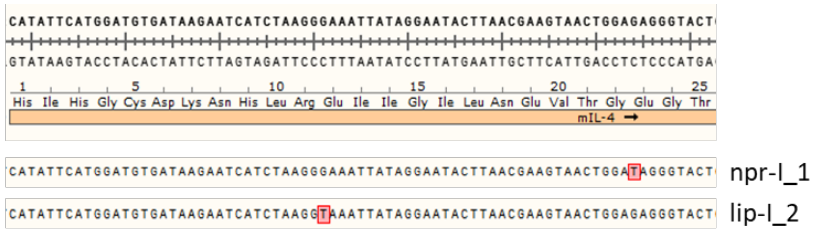


Figure 5.3 Growth curves of mIL-4 expressing strains. mIL-4-expressing strains were grown in liquid TYG/thiamphenicol medium, alongside a strain carrying an empty plasmid (negative control; neg). OD_{600} was monitored over 24 hrs. Data represent mean values of three biological replicates ± SD. **A.** Non-mutated strains **B.** Mutated strains **C.** Non-mutated strains induced with 1 mM lactose **D.** Mutated strains induced with 1 mM lactose.

A



B



C

strain	position	type	mutation	coding region change
npr-I_1	mIL-4	SNP	67G>T	E23*
lip-I_1	P_{tcdB}	SNP	121C>T	N/A
lip-I_2	mIL-4	SNP	34G>T	E12*
lip-I_3	P_{tcdB}	SNP	142C>T	N/A

Figure 5.4 Acquired mutations in the mIL-4 constructs. mIL-4-expressing strains were screened by colony PCR with M13F and M13R after 24 hrs of growth; purified PCR products were sequenced with the same primers. Sequence alignments were performed using SnapGene. **A.** Point non-sense mutations in npr-I_1 and lip-I_2. **B.** Mutations in the -30/-10 P_{tcdB} TATA boxes in lip-I_1 and lip-I_3. **C.** Mutation analysis. SNP: single nucleotide polymorphism, N/A: not applicable

5.5.2 Immunoblot detection of mGM-CSF and mIL-4

The lysates and supernatants of mGM-CSF- and mIL-4-expressing cultures were analysed by Western blot as a qualitative confirmation of protein expression and secretion. Samples were harvested at the early stationary growth phase (8 hrs); lysates and supernatants were prepared according to sections 2.1.10 and 2.1.11, respectively, and Western blotting was carried out as described in section 2.3.6.

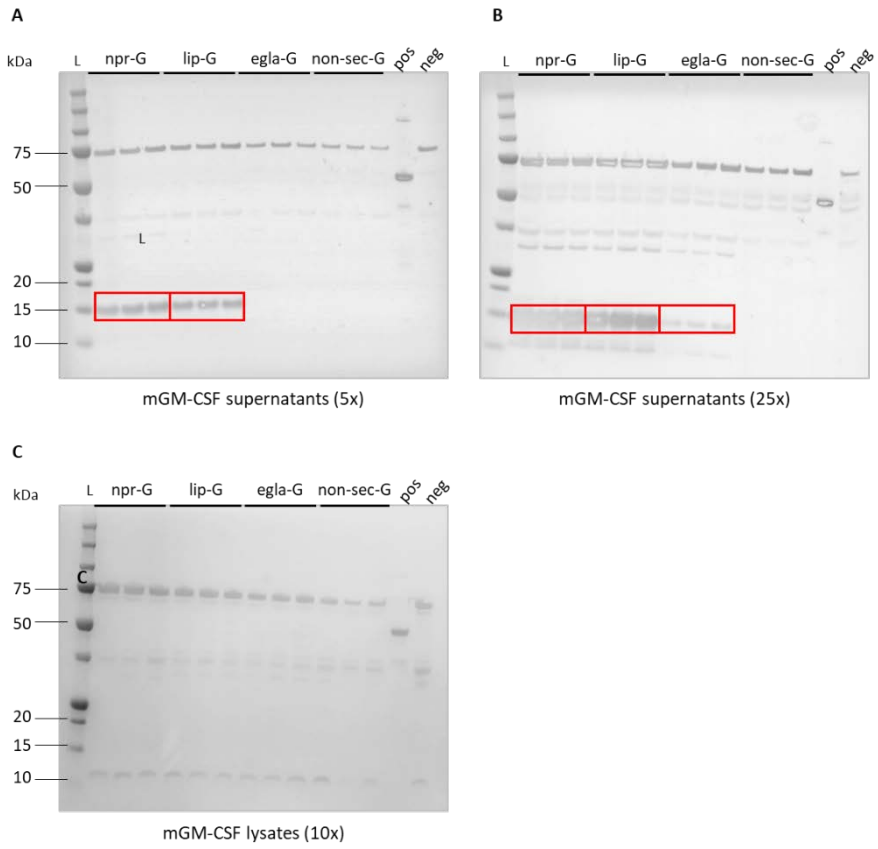


Figure 5.5 Western blot analysis of mGM-CSF-expressing strains. The supernatants and lysates of mGM-CSF-expressing strains were subjected to SDS-PAGE/Western blot with monoclonal ANTI-FLAG M2-peroxidase antibody. **A.** Supernatants concentrated 5-fold. **B.** Supernatants concentrated 25-fold. **C.** Lysates concentrated 10-fold. L: Bio-Rad Precision Plus Protein All Blue Standards ladder; pos: FLAG-tagged positive control; neg: supernatant of a strain carrying an empty plasmid (negative control).

Bands corresponding to FLAG-tagged mature mGM-CSF (15.1 kDa) were visible in the culture supernatants of the strains carrying a signal peptide, but absent in those of the non-secreted and negative controls (Figure 5.5A-B). Secretion with the Npr and Lip signal peptides appears to be efficient, as shown by the strong bands in both 5-fold and 25-fold concentrated supernatants (relative to the original volume of sample harvested). Conversely, bands corresponding to EglA-mediated secretion were only visible for the 25-fold concentrated samples and were less intense than those generated using the Npr and Lip signal peptides. Bands smaller than 15 kDa, present only in the supernatants of the secreting strains (Figure 5.5B), may correspond to mGM-CSF degradation

products, a result of proteolysis from native extracellular proteases or as a consequence of the heat denaturation step during sample preparation for SDS-PAGE.

No bands corresponding to the presecretory mGM-CSF products (sizes 17.9 – 19.5 kDa) could be detected in the cell lysates, presumably due to low intracellular abundance at the concentration assayed; nevertheless, the lysate concentration could not be increased, limited by the volume of Bugbuster reagent that was required to sufficiently lyse the given quantity of cells. It should be noted that due to non-specific binding of the anti-FLAG antibody to endogenous proteins, only those bands that were absent from the negative control sample (lane 14) were considered for the interpretation of the immunoblotting results.

No bands of the expected size could be detected in the supernatants and lysates of mIL-4 expressing cultures (non-induced, mutation-free), attributable to the low expression levels that can be attained in the non-induced state (results not shown).

5.5.3 Quantification of secreted mGM-CSF

In view of the mIL-4 associated toxicity and the inability to detect secreted mIL-4 in culture supernatants, only the mGM-CSF-expressing strains were subjected to further characterisation. To obtain a quantitative insight in the secretory efficiencies of the signal peptides employed, an ELISA assay was carried out as described in section 2.3.9, with capture and detection antibodies specific for mGM-CSF; it was assumed that the FLAG-tag would not interfere significantly with the native folding of the protein.

Mean extracellular mGM-CSF concentrations in cell-free culture supernatants at the time of harvesting (8 hrs) were 576.5 ng/ml for the Npr, 510.3 ng/ml for the Lip and 29.3 ng/ml for the EglA signal peptide constructs (Figure 5.6). As expected, no mGM-CSF could be detected in the supernatants of the non-secreted and negative control samples (values outside of the standard range). After normalisation for the OD_{600} of the cultures at the time of harvesting, use of the Npr or Lip signal peptides yields

approximately 20-fold higher extracellular levels of mGM-CSF compared to EglA, in agreement with the Western blot results. Although Npr appears to be more efficient than Lip, the difference is not statistically significant (p=0.0636).

Quantification of intracellular mGM-CSF was also attempted as a means of asserting that observed differences in the quantities of extracellular mGM-CSF are solely due to different secretory capacities of the signal peptides employed and not due to differences in expression levels. Nevertheless, this was done in the knowledge that intracellular mGM-CSF would likely be less readily detected by ELISA given the presence of the signal peptide, the association of the presecretory product with the secretory machinery, as well as the absence of disulphide bondsin the cytoplasm[341]; these factors would impede folding of the protein to its native conformation and, therefore, its binding to the ELISA antibodies. As anticipated, minimal amounts of mGM-CSF were detected in the cell lysates, including the non-secreted control (results not shown). In view of the aforementioned limitations, it was decided to disregard these results and evaluate secretory efficiency based only on the mGM-CSF extracellular quantities.

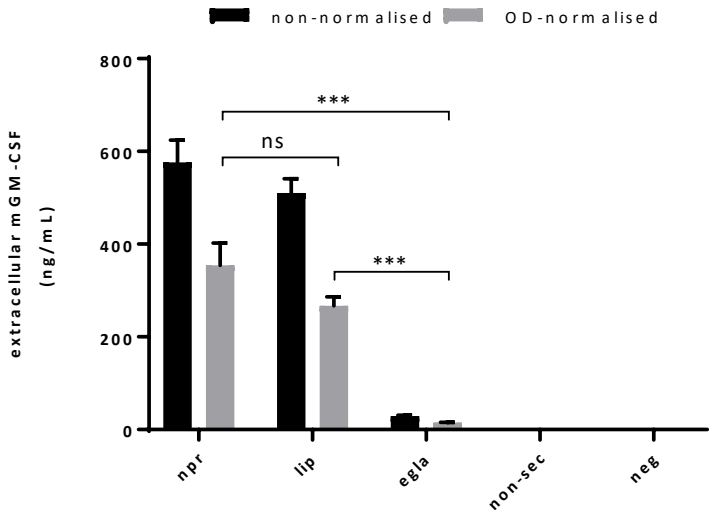


Figure 5.6 Quantification of mGM-CSF in culture supernatants of mGM-CSF-expressing strains. Secreted mGM-CSF was quantified in supernatants of mGM-CSF-expressing strains with a commercial mGM-CSF ELISA kit. Secretion is under the control of the indicated signal peptide. Data represent mean values of three biological replicates ± SD, from a single representative experiment out of three repeats. OD-normalised values are obtained by dividing non-normalised values with the OD_{600} of the cultures at the time of harvesting. Statistical analysis was carried out using one-way ANOVA with Tukey's test for multiple comparisons; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

5.5.4 The effect of extracellular proteolysis on protein yield

C. sporogenes has been typically classified as proteolytic, based on its ability to hydrolyse proteins and ferment amino acids[342, 343]. Genome sequencing has also confirmed the presence of numerous genes encoding proteases; as an indication, the *MEROPS* peptidase database assigns 158 known and putative peptidases to the species[344]. Amongst these is an operon of 6 thermolysin-like metalloproteases of the M4 family (loci c14680-c14730 in *C. sporogenes* NCIMB 10696)[27]. The same operon has been previously identified in the proteolytic (Group I) *C. botulinum* Hall A strain, a neurotoxigenic counterpart of *C. sporogenes*[345]. In fact, the protein products of this operon were found to be the most abundant extracellular proteins in the cell-free supernatant of WT *C. sporogenes* NCIMB 10696[338], underscoring the highly proteolytic

nature of the strain. (As mentioned previously, the Npr signal peptide, hereby implemented for the secretion of mGM-CSF, is the native signal peptide of a protease from this operon.)

Protease secretion poses a trade-off in the context of *C. sporogenes*-mediated cancer therapy. On the one hand, it underpins the strain's superior colonising and oncolytic abilities, especially in comparison to saccharolytic representatives of the genus[25, 54]. On the other hand, proteolysis might compromise the yield of secreted recombinant proteins, thus curtailing the therapeutic doses that could be achieved otherwise[346]. Importantly, this effect may be compounded by the intratumoural presence of proteases of eukaryotic origin, expressed by cancer cells, such as matrix metalloproteinases (MMPs)[347]. Certainly, given the instrumental role of clostridial proteases in the colonisation and degradation of tumour tissue, it would be counterintuitive to pursue protease knockout strategies implemented elsewhere, merely to increase protein yield[348].

In silico analysis of the mGM-CSF protein sequence revealed the presence of several putative sites cleavable by bacterial proteases. The analysis was carried out as described in section 2.6.11 and the potential cleavage sites are displayed in Table 5.3. The complete report can be found in Appendix VIII. Amongst the predicted sites are 30 putatively cleaved by a thermolysin-like metalloprotease of the M4 family. However, considering that *C. sporogenes* NCIMB 10696 proteases are poorly characterised to date, it is difficult to assert the significance of these predictions. Therefore, it would be beneficial to gauge the relevance of proteolysis *in vitro*, as a guide for future engineering efforts

Table 5.3 Predicted cleavage sites of bacterial proteases in the mGM-CSF protein sequence

Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N endopeptidase	11	26 27 91 101 104 112 124 127 128 129 130
Asp-N endopeptidase + N-terminal Glu	21	16 20 26 27 34 35 37 42 59 89 91 93 101 104 112 116 124 127 128 129 130
Clostripain	4	4 11 55 64
Glutamyl endopeptidase	10	17 21 35 36 38 43 60 90 94 117
LysC	13	14 20 47 48 57 69 71 108 119 120 124 127 132
Staphylococcal peptidase I	9	17 21 35 38 43 60 90 94 117
Thermolysin	30	6 8 15 18 22 24 25 30 32 39 45 48 51 55 57 58 62 66 69 72 73 75 77 96 100 103 106 109 110 115

To this end, the highly mGM-CSF secreting strain, npr-G, was grown with (+PI) and without (-PI) protease inhibitors (cOmplete™, EDTA-free Protease Inhibitor Cocktail at 1x concentration) added at the time of inoculation. According to the manufacturer (Roche), the selected protease inhibitor cocktail is an effective inhibitor of a broad spectrum of serine and cysteine proteases, but not of metalloproteases. For complete inhibition of the latter, an EDTA-containing inhibitor should have been used, able to sequester the zinc ion required for enzyme function[343]. In this case, where the inhibitor was added during cell culture (and not following cell harvest as typically done), the use of EDTA was not considered appropriate, as it could interfere with the function of other metal-dependent proteins. Therefore, only partial protease protection was expected, possibly arising from inhibition of yet uncharacterised proteases, other than the metalloproteases discussed previously.

Samples were harvested after 8 hrs, at the early stationary growth phase, when extracellular proteases were expected to be present in the culture supernatants[343, 349]. The cell-free supernatants were subjected to Western blot analysis and their mGM-CSF content was quantified via ELISA. Mean extracellular mGM-CSF concentration increased to 695.2 ng/ml in the presence of protease inhibitors (*vs.* 514.9 ng/ml in their absence). Nevertheless, co-culture with protease inhibitors had a minor impact on total

cell density; OD_{600} at the time of harvesting (8 hrs) was 1.5 with protease inhibitors *vs.* 2.0 without. After OD_{600}-normalisation of extracellular mGM-CSF levels, the yield gain afforded by protease inhibition was 1.8-fold (p= 0.0032) (Figure 5.7).

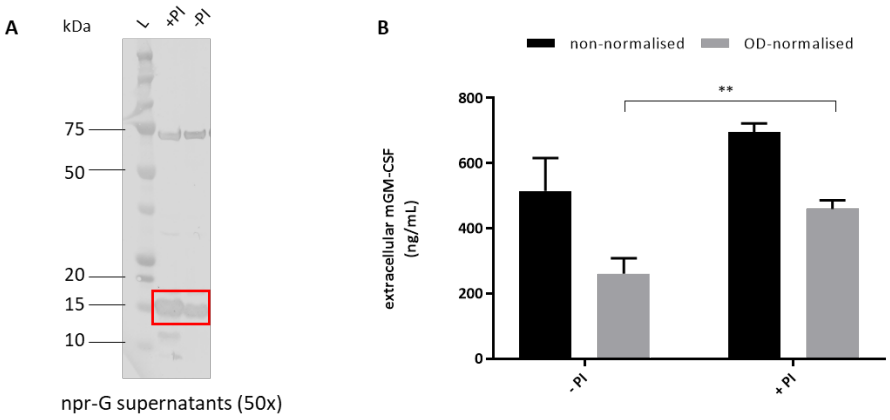


Figure 5.7 The effect of extracellular proteolysis on the yield of secreted mGM-CSF. npr-G was grown in TYG/thiamphenicol medium with (+PI) or without (-PI) protease inhibitor cocktail. **A.** Supernatants were concentrated 50-fold and subjected to SDS-PAGE/Western blot with monoclonal ANTI-FLAG M2-peroxidase antibody. L: ladder. **B.** Secreted mGM-CSF was quantified in supernatants with a commercial mGM-CSF ELISA kit. Data represent mean values of three biological replicates ± SD, from a single representative experiment out of two repeats. OD-normalised values are obtained by dividing non-normalised values with the OD_{600} of the cultures at the time of harvesting. Statistical analysis was carried out using two-tailed unpaired t-test; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

5.5.5 Biological activity of mGM-CSF produced in *C. sporogenes*

Having established that *C. sporogenes* produces mGM-CSF efficiently, it was necessary to ascertain that the secreted protein is also biologically active. Therefore, the present thesis investigated the ability of murine BM cells, isolated as described in section 2.5.1, to proliferate and differentiate *in vitro*, in response to mGM-CSF present in the cell-

free supernatant of the engineered npr-G strain. Treatment reagents were prepared as described in section 2.5.3 while cell culture and treatment were carried out according to sections 2.5.2 and 2.5.4, respectively. Briefly, the supernatants were applied at a final concentration of approximately 19.2 ng/ml of mGM-CSF (10 μl/well of a 1.5-fold diluted supernatant in a final volume of 200 μl/well), assuming an initial extracellular concentration of 576.5 ng/ml, as measured previously (section 5.5.3). The supernatant of the negative control strain (neg) was likewise diluted 1.5-fold and used as a negative control. Purified recombinant GM-CSF (20 ng/ml) was assayed alongside as a positive control, prepared in two alternative ways: a) in the culture medium of BM cells (mGM-CSF) and b) in the 1.5-fold diluted supernatant of the negative control strain (mGM-CSF_SN); the latter would permit a more realistic comparison against mGM-CSF produced by *C. sporogenes*.

Cellular proliferation was assayed by an MTT assay, as described in section 2.5.5, whereby cellular metabolic activity, estimated by the conversion of MTT to formazan, was used as a proxy for cell viability. *C. sporogenes*-derived mGM-CSF was able to induce statistically significant proliferation of BM cells compared to the untreated control ($p<0.0001$)(Figure 5.8A). Inhibition of proliferation (67%) in the presence of an anti-mGM-CSF neutralising antibody (Figure 5.8B) and lack of proliferation following treatment with the neg supernatant (Figure 5.8A) indicate proliferation is specifically induced by mGM-CSF. Interestingly, proliferation in response to the positive control, mGM-CSF, was approximately 1.5-fold higher than that in response to either mGM-CSF_SN or npr-G ($p<0.0001$), despite no differences in the mGM-CSF dose. This may suggest that certain components of the bacterial supernatant could be downregulating BM cell proliferation and/or mediating toxic effects.

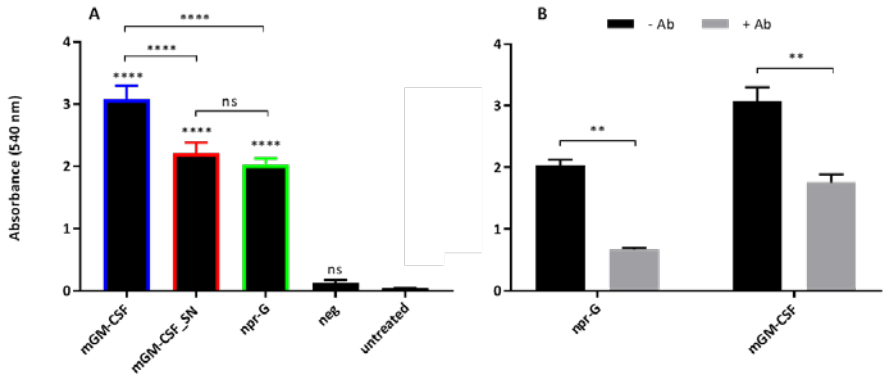


Figure 5.8 MTT cell proliferation assay on mGM-CSF-treated murine bone marrow cells. Proliferation of murine BM cells was measured via an MTT assay. **A.** Cell proliferation in response to treatment with purified mGM-CSF dissolved in culture medium (mGM-CSF), purified mGM-CSF dissolved in bacterial supernatant (mGM-CSF_SN), supernatant of the npr-G strain (npr-G) and supernatant of the negative control strain (neg), in comparison to untreated cells. Depicted data are from a single representative experiment out of two repeats. Statistical analysis was carried out using one-way ANOVA with Tukey's test for multiple comparisons. **B.** Cell proliferation in response to treatment with purified mGM-CSF dissolved in culture medium (mGM-CSF) or supernatant of the npr-G strain (npr-G) in the presence (+Ab) and absence (-Ab) of anti-mouse GM-CSF neutralising antibody. Depicted data are from a single representative experiment out of two repeats. Statistical analysis was carried out using two-tailed unpaired t-test. Data represent mean values of three replicates ± SD. p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

To further characterise the GM-CSF-derived BM cells, treated cells were analysed by flow cytometry, as outlined in section 2.5.6. The flow cytometry studies, including data collection and analysis, were directly supervised by Dr Kazuyo Kaneko.

In vitro culture of murine BM cells in the presence of GM-CSF is expected to yield a heterogeneous cellular population, comprised mainly of macrophages and DCs[350]. The surface markers CD11c and F4/80 have been typically used for the distinction between murine DCs and macrophages, respectively[351]. CD80 and CD86 are co-stimulatory molecules required for effective antigen presentation to T-cells and are usually indicative of the maturation state of both cell types[352, 353].

Following treatment with mGM-CSF, the resulting cell population, after normalisation for the isotype controls, consisted of 58% $CD11c^+$, 54% $F4/80^+$, 51% $CD80^+$ and 22% $CD86^+$ cells. Likewise, treatment with mGM-CSF_SN generated 5% $CD11c^+$, 50% $F4/80^+$, 43% $CD80^+$ and 19% $CD86^+$ cells, and with npr-G, 3% $CD11c^+$, 49% $F4/80^+$, 36% $CD80^+$ and 15% $CD86^+$ cells (Figure 5.9). It is evident that while the population composition with respect to the F4/80, CD80 and CD86 markers is similar, the reduction in CD11c expression in the presence of bacterial supernatant is striking. As expected, isotype controls generated very little signal.

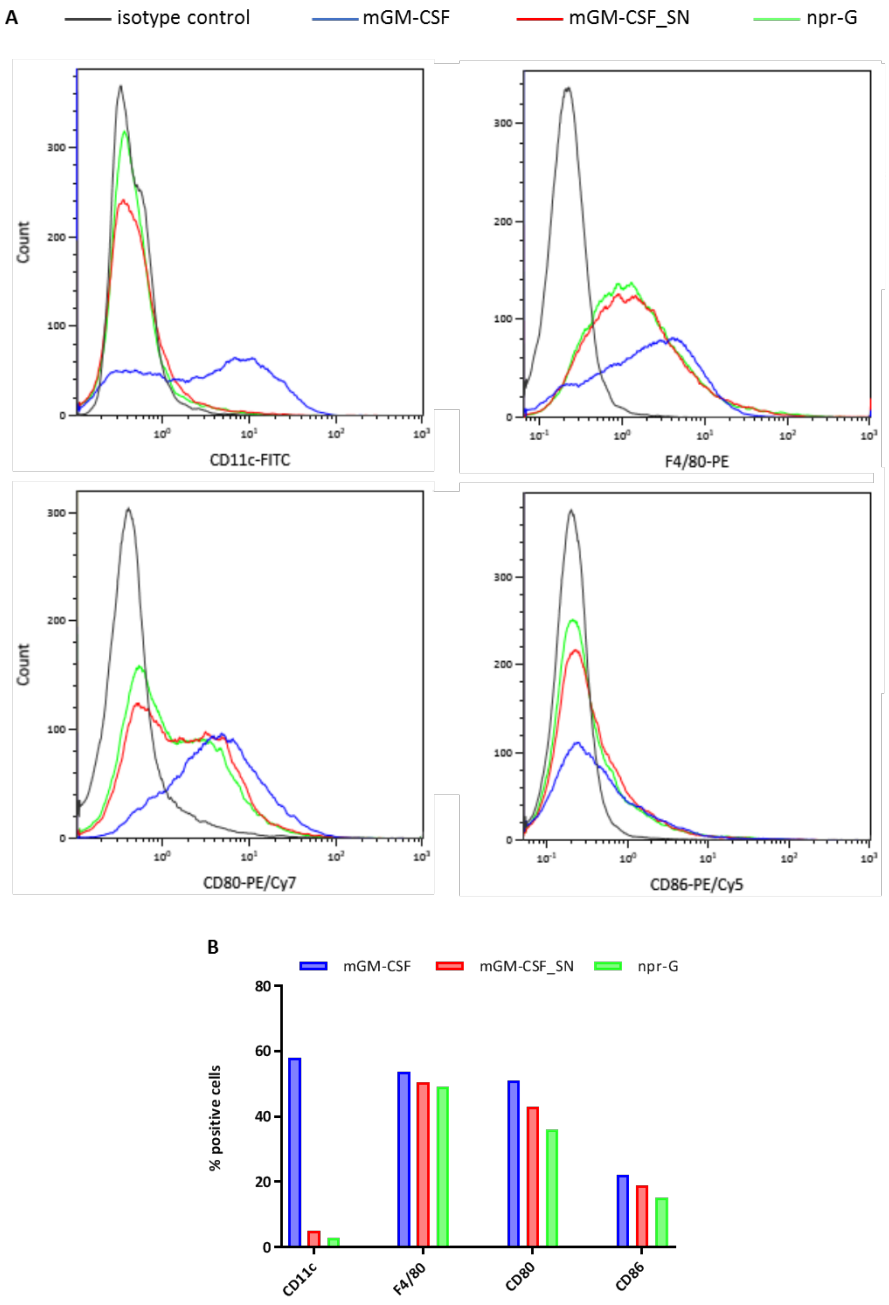


Figure 5.9 Phenotypic analysis of mGM-CSF-treated murine bone marrow cells. mGM-CSF-treated BM cells were examined for expression of the indicated markers using four colour flow cytometry. **A.** Histogram overlays of BM cells subjected to the annotated treatments, with respect to each marker. mGM-CSF: purified mGM-CSF dissolved in culture medium; mGM-CSF_SN: purified mGM-CSF dissolved in bacterial supernatant; npr-G: supernatant of the npr-G strain; neg: supernatant of the negative control strain. **B.** Quantification of $CD11c^+$, $F4/80^+$, $CD80^+$ and $CD86^+$ cells as a percentage of the total gated cell population, adjusted for the corresponding isotype control; no error bars are depicted as the measurements were carried out once.

5.6 Discussion

C. sporogenes was demonstrated to be an appropriate host for the expression and secretion of mGM-CSF, but not of mIL-4, thus contradicting the hypothesis proposed in section 5.3. Expression and/or secretion of mIL-4 appeared to be toxic to the cells, manifesting as growth defects or acquisition of compensatory mutations, despite the utilisation of a relatively tight inducible system. It can be argued that toxicity is linked to saturation of the Sec-translocon capacity, previously shown to cause growth defects as well as cell size and granularity alterations[354, 355]; this is supported by the fact that toxicity was more pronounced in the mIL-4 strain carrying the highly efficient Npr signal peptide as evidenced by the de facto inability to obtain stable transconjugants and the strain's predisposition for the acquisition of mutations. Given that signal peptides may behave differently in combination with different proteins[337], it is no surprise that toxicity was limited to mIL-4, despite utilisation of the same secretion signals in the mGM-CSF constructs. It should also be noted that the first amino acid of mature mIL-4 is a histidine (H), in contrast to that of mature mGM-CSF which is an alanine (A) (Table 5.1), alanine being the most favoured amino acid at the +1 position of secreted proteins[356]. Hence, it is possible that expression and/or secretion fine-tuning (possibly via the optimisation of the signal peptide cleavage site) could eliminate toxicity presently associated with mIL-4. Importantly, the ability of *C. sporogenes* to tolerate mGM-CSF, but not mIL-4 at the given expression/secretion level, highlights that no one size fits all and that expression of heterologous proteins should be studied and optimised individually.

The expression and secretion strategy employed in this study was successful, yielding approximately 3000-fold higher extracellular levels of recombinant protein in *C. sporogenes* than those previously obtained with rIL-12 (576.5 ng/ml for mGM-CSF *vs.* 18 pg/ml for IL-12)[304]. Even use of the same signal peptide, EglA, yielded more than 1000-fold higher secretion (29.3 ng/ml), possibly due to the superior stability of the

replicon employed (pBP1 *vs.* pIM13)[67, 75], the efficiency of the expression system or the protein itself. Concurrently, the yield of mGM-CSF obtained in the presence of protease inhibitors (695.2 ng/ml) was close to the highest reported yield of a secreted cytokine in a clostridial host, namely that of rIL-2 in the saccharolytic *C. acetobutylicum* DSM 792 strain (800 ng/ml)[303].

Furthermore, the use of novel signal peptides (Npr and Lip) was exemplified, expanding the available repertoire for secretion in clostridial hosts. Encouragingly, both Npr and Lip outperformed EglA which has been the default option thus far. It is possible that EglA has been used recurrently in the absence of alternatives, but is probably far from ideal, being a long signal peptide of 38 amino acids (close to the usual maximum size of 40 for Gram-positive organisms[356]) and employing a rare start codon (UUG)[357]. Accordingly, the predicted D-scores for EglA were lower than those of Npr and Lip with both mGM-CSF and mIL-4 (Table 5.1).

Although the proteolytic character of *C. sporogenes* was shown to compromise the yield of extracellular mGM-CSF, the relevance of this needs to be evaluated *in vivo*. Provided that expression and secretion processes are adequately efficient, it is likely that the intratumoural concentrations achieved can mediate the desired effects in spite of the competing proteolysis, as it was evident with the mIL-12-secreting strain[304]. Still, proteolysis may become relevant if the cytokine gene is integrated in the genome, whereby the secreted dose will be critically diminished. Nevertheless, it is expected that the tumour environment will be rich in alternative substrate candidates, which might be preferentially metabolised by the secreted proteases. In any case, the protein of interest could be mutated in order to improve its proteolytic resistance, whilst taking care not to compromise its biological function[358]. This strategy can be undertaken to circumvent proteolysis both arising from proteases secreted by Clostridia and cancer cells alike. For this purpose, detailed characterisation of the proteolytic activity of *C. sporogenes* would be beneficial.

Biological activity of *C. sporogenes*-derived mGM-CSF was confirmed *in vitro* by an MTT cell viability assay. Considering the time limitations of the present study, the metabolism-proxy MTT assay offered a commonly used and high-throughput way of measuring cell proliferation[359]. However, its use for this purpose is founded on the assumption that there is a linear relationship between cellular metabolism of tetrazolium by living cells (quantified by the resulting absorbance signal) and cell number, which may not always be justifiable. Therefore, interpretation of MTT assay-derived data in isolation may generate misleading conclusions[360]. To increase confidence in the results obtained here, a complementary colony forming assay should have been carried out, which would enable direct enumeration and characterisation of treated cells[361].

It must also be emphasised that the flow cytometry results presented in section 5.5.5 should be interpreted with extreme caution, given the lack of experimental replicates at the time of writing. Although this is a major obstacle in making firm conclusions, some initial ideas can be drawn regarding the cell populations derived post-treatment with *C. sporogenes*-derived mGM-CSF. As observed elsewhere, treatment of BM cells with GM-CSF *in vitro* generated cell populations positive for both CD11c and F4/80[350, 362, 363]. These may consist of F4/80$^+$ immature DCs[350], CD11c$^+$ monocyte-derived macrophages[350, 363] and F4/80$^+$ monocytes transitioning to either macrophages or DCs[363]. More detailed differentiation is usually carried out by investigating marker expression levels (high/low), but such analysis was not undertaken here. The small CD86$^+$ cell population in combination with the higher percentage of CD80$^+$ relative to CD86$^+$ cells are indicative of a yet immature state, as one would expect, given that no exogenous maturation triggers were (intentionally) applied. This agrees with previous findings where CD80 is frequently the most abundant ligand of the two when myeloid precursors are grown *in vitro* into an immature state[364]. Elsewhere, CD80 expression was pronounced in cells corresponding to DCs (not macrophages)[363], suggesting that the CD80$^+$ cells in the present study are most likely to represent (immature) DCs. However, in the same study, higher doses of GM-CSF (>10 ng/ml) and identical culture periods (7 days) favoured the differentiation of BM cells into macrophages, in approximate

ratios (macrophages:DCs) of 3:1-4:1[363], thus reducing the chance of $CD80^+$ cells being only DCs.

Unexpectedly, the bacterial supernatant exerted an independent, negative impact on cellular proliferation and expression of CD11c. CD11c is a surface integrin involved in cell adhesion and migration, pathogen sensing and interactions with components of the complement system[365]. Downregulation of CD11c has been previously observed on murine BM-derived DCs (BMDCs) following TLR triggering induced by lipopolysaccharide (LPS) or CpG motif-containing oligonucleotides[365, 366]; importantly, CD11c downregulation did not affect activation, maturation and functionality aspects, as evidenced by the upregulation of the major histocompatibility complex II (MHC II) and CD86, IL-6 and IL-12 cytokine secretion, and the ability of $CD11c^-$ to stimulate allogenic spleen cells in a mixed lymphocyte reaction assay[366]; elsewhere, it has also been shown that the endocytic capacity of $CD11c^-$ cells is not affected[365]. Although *C. sporogenes* does not contain LPS (being Gram-positive), it is possible that other bacterial components present in the cell-free supernatant can serve as TLR agonists, including lipoteichoic acid (TLR2), flagellin (TLR5) and nucleic acids (TLR9)[367]. Additionally, small chain fatty acids, and particularly butyrate and propionate, have been shown to exert immunomodulatory effects on DCs, altering the profile of secreted cytokines[368]. Given that butyric and propionic acids are known by-products present in the supernatant of *C. sporogenes*[369], it is plausible that they also could be interfering with the DC differentiation or proliferation pathways.

Considering the above findings, further investigation in the biological effects mediated by the bacterial supernatant is required; gene expression profiling would allow precise identification of the derived cell types[350] whilst analysis of secreted cytokines coupled with suitable immunoassays could confirm their functionality. Additionally, the mGM-CSF-expressing vector should be evaluated in an appropriate, immunocompetent mouse model, to establish the biological activity of the cytokine *in vivo* as well as its effectiveness in combination with clostridia. *In vivo* assessment of the engineered strains

will also provide an insight on the desired expression level, in relation to the type of tumour, spore dose, degree of colonisation as well as extent of bacterial replication. Conveniently, the lack of glycosylation is not a concern given that the aglycone version of GM-CSF can fulfil the full biological functions of the protein (albeit with different pharmacokinetics) and is supported by the approved clinical use of molgamostim (non-glycosylated GM-CSF produced in *E. coli*)[370-373].

Incorporation of GM-CSF in the clostridial vector is expected to be synergistic, the combined effectiveness being greater than GM-CSF or clostridia as separate modalities. On the one hand, intratumoural delivery of GM-CSF (equivalent to the clostridial-mediated tumour-specific expression) has been shown to be the most effective administration route[374]. Additionally, a prolonged and sustained intratumoural presence of the cytokine at a dose of 90 to 300 ng/10^6 cells per 24 hours generates a stronger immune response compared to high-dose single administration[312], suggesting that GM-CSF expression in the clostridial vector would be appropriate. Importantly, this dose could, in theory, be effectively delivered by the npr-G strain; the estimated *C. sporogenes* CFU count corresponding to the OD of the culture at the time of harvesting (8 hrs) is approximately 1.6 x 10^8 CFU/mL (results not shown), inferring that this amount of bacterial cells secrete 576 ng/mL of mGM-CSF, as quantified in section 5.5.3. Assuming that *C. sporogenes* tumour colonisation can result in up to 1-2 x 10^8 CFU/g of tumour, as previously reported[25], it follows that sufficient quantity (>300 ng) of GM-CSF could be delivered to the tumour by clostridial cells. Given also that the intratumoural CFU number can remain stable over long periods of time[25], it is anticipated that GM-CSF will be present at that dose continuously. However, it is not currently known what proportion of intratumoural CFUs exists as spores at any given time. Considering that spores are inert and would not be expressing the cytokine, it is possible that the actual GM-CSF quantity produced will be lower than the one predicted; yet, it might still be within the dose required to ensue a strong immune response.

Furthermore, clostridial-mediated delivery will provide additional DC maturation signals via the means of clostridial and tumour-associated (following oncolysis) antigens, thereby ensuring that antitumour responses take place[282, 375]. On the other hand, GM-CSF will enhance clostridial-mediated oncolysis, via the involvement of the immune system. In particular, the hypoxic environment required for clostridial germination favours the acquisition of a DC 'tolerogenic' phenotype, namely a cell behaviour that would lead to immunological tolerance rather than immunological antitumoural responses[275, 278]. This serves to highlight the cogency of incorporating a DC-activating cytokine, such as GM-CSF, in the clostridial vector.

Effectively, the present study was a pilot trial for the incorporation of an immunotherapeutic functionality in *C. sporogenes*. Recently, concerns have emerged regarding coincident pro-tumoural functions of GM-CSF, associated with its ability to induce immune-independent proliferation and/or migration of cancer cells; as such, GM-CSF secretion has been linked to poorer outcomes in various cancer types, including bladder, head, colorectal and lung cancer[305]. In view of the successful expression and secretion of biologically active mGM-CSF in *C. sporogenes*, mGM-CSF could be replaced by a less controversial immunotherapeutic mediator in the future, provided that an acceptable toxicity profile can be afforded whilst employing similar expression and secretion tools.

5.7 Key outcomes

- *C. sporogenes* was successfully engineered to express and secrete high levels of biologically active mGM-CSF, despite its proteolytic nature.
- The use of new signal peptides (Npr, Lip) was exemplified, yielding high secretory efficiencies.

CHAPTER 6

Towards a safer, non-haemolytic clostridial host

6.1 Background

6.1.1 Regulatory and safety considerations

According to the European Medicines Agency (EMA), an engineered clostridial vector would fall under the definition of gene therapy medicinal products (GTMPs) – a subcategory of advanced therapy medicinal products (ATMPs) – being of biological origin and containing recombinant nucleic acid responsible for its therapeutic mechanism of action (Directive 2001/83/EC). More specifically, it would be classified as a genetically modified organism (GMO), namely, a biological entity capable of replication or DNA transfer, whose genetic material has been modified in a way that does not occur naturally by mating and/or natural recombination (Directive 2001/18/EC). Likewise, according to the U.S. Food & Drug Administration (FDA),

engineered clostridia for cancer therapy would be classified as 'Cellular & Gene Therapy Products', whereby "gene therapy is a medical intervention based on modification of the genetic material of living cells" and "recombinant DNA materials used to transfer genetic material for such therapy are considered components of gene therapy and as such are subject to regulatory oversight"[376].

By means of their classification, engineered clostridia warrant special regulatory considerations and must fulfil rigorous safety criteria before they can enter clinical trials and, subsequently, obtain marketing authorisation[377]. Worse yet, regulatory hurdles are compounded by the novelty of bacterial-based therapies and the small foothold they currently hold in the clinic[378].

The special regulatory considerations that apply to gene therapies span all aspects from preclinical development to manufacture of the final product. Of great importance to bacterial-based gene therapies in particular are the risk of aberrant localisation to non-target cells/tissues, the possibility of antibiotic/virulence gene transfer to host microbiota or pathogenic species, and the potential of horizontal transmission from the patient to family members and health care providers (also known as shedding), given their competency of replication[185, 376, 378]. The theoretical risk posed by the dissemination of bacteria used as treatment delivery vectors to non-target tissues and the possibility of them causing undue disease, is particularly relevant in the context of cancer therapy, where bacterial therapies may be administered to immunocompromised patients following chemotherapy[47, 379]. Crucially, this 'theoretical' risk has unfortunately materialised in cancer clinical trials, with patients developing systemic listeriosis following administration of recombinant *Listeria monocytogenes* strains[379, 380].

6.1.2 Safety-enhancing strategies

Considering the above challenges associated with bacterial-based gene therapies, bacterial anticancer delivery vectors are designed with inbuilt safety precautions. These often include attenuation of the strains employed to remove potential pathogenic traits

or improve their tumour specificity, as well as the incorporation of desirable safety features (for example, auxotrophy, imaging functionality)[47]. Alternative safety-enhancing strategies relate specifically to the recombinant DNA or the adopted method of genetic engineering; for instance, expression of genetic cargos may be placed under the control of external or tumour-specific stimuli such as radio- (P_{recA}, P_{recN}), hypoxia- (HIP-1), SOS- or L-arabinose- (P_{BAD}) inducible promoters, to prevent activation in systemic circulation and healthy tissues[53, 284, 381, 382]. Importantly, the therapeutic genes should, whenever possible, be introduced in the bacterial genome. This circumvents the use of mobile genetic elements carrying antibiotic resistance markers (such as plasmids), while securing the genetic stability of the cargos and, therefore, their predicted/safe function[20].

Accordingly, most bacterial strains currently being investigated for cancer therapy have undergone safety-enhancing alterations. Indicatively, *C. novyi*-NT has been obtained after heat-treatment of the pathogenic WT *C. novyi* strain and consequent removal of a lethal toxin carried on a phage[16]. Likewise, the superoxide dismutase gene (*sod*) of *C. perfingens* has been deleted in order to minimise the oxygen tolerance of the strain and improve its tumour selectivity[383-385]. *Salmonella typhimurium* has been attenuated by deletions of the *msbB* and *purI* genes (strain VNP20009), whereby *msbB* knockout renders the strain deficient in lipid A biosynthesis and, therefore, reduces the risk of lipopolysaccharide-induced septic shock, while deletion of *purI* creates a requirement for an external source of adenine for the strain's survival[187]. A strain auxotrophic for leucine and arginine (leu/arg) has also been created, with improved tumour-seeking properties (strains A1/A1-R)[386, 387]. Alternative attenuation strategies of *Salmonella* strains include the deletion of *aroA* (strain SL5000)[297], *aroA/aroD* (strain BRD509)[284], *htrA*[388], *cya/crp/asd* (strain χ4550)[289] and other genes, with an unprecedented level of sophistication. Similarly, *L. monocytogenes* has been attenuated by deletion of two virulence genes *actA* and *inlB* (strain LADD)[90], downregulation of the expression of the global virulence regulator *prfA* (complemented XFL-7 strain)[389], as well as by rendering

the strain replication-deficient, but metabolically active, by deletions in the *uvrAB* genes followed by treatment with psoralen and ultraviolet radiation (KBMA strain)[390].

Nevertheless, the challenge of creating safer strains is to strike a balance between attenuation and effectiveness. Attenuation strategies should not compromise the immunogenic potential of the vector (especially when the principal mechanism of action is immune-related)[391] nor interfere with the anti-cancer capacities of the strains employed[392, 393].

6.1.3 Safety considerations with *C. sporogenes* NCIMB 10696

The clostridial host employed in the present study, *C. sporogenes* NCIMB 1096, was chosen not only thanks to its favourable colonising and oncolytic properties, but also thanks to its non-pathogenic nature, being classified as a group 1 organism by the UK Advisory Committee on Dangerous Pathogens (ACDP)[a] as well as the American Type Culture Collection (ATCC). Nevertheless, this classification is not unanimously accepted and *C. sporogenes* is considered a group 2 organism[b] (or equivalent) elsewhere, including in the Netherlands, Belgium, Germany, Switzerland and Japan.

According to the Netherlands Commission on Genetic Modification (COGEM) (personal communication; J. Theys), classification of *C. sporogenes* as a group 2 organism is founded on: a) the close relation of the organism with proteolytic *C. botulinum* group A, E and F strains[27, 345] together with a lack of standard laboratory diagnostic tools that can discriminate between them, b) the possession of a streptolysin S-like gene cluster, encoding a haemolytic toxin and virulence factor in group A Streptococci[94, 345], c) *in vivo* evidence of haemolysis in rabbits[394-396], d) *in vitro* evidence of cytotoxicity against animal cell lines[395, 397, 398], and e) sporadic clinical cases where *C. sporogenes* was found to be the causative agent of disease, suggesting occasional pathogenicity[397, 399].

[a] ACDP hazard group 1: Unlikely to cause human disease.

[b] ACDP hazard group 2: Can cause human disease and may be a hazard to employees; it is unlikely to spread to the community and there is usually effective prophylaxis or treatment available.

Crucially, the presence of the streptolysin S-like gene cluster was regarded as a feature that might impede the clinical translation of clostridial therapies employing *C. sporogenes*.

6.1.4 Streptolysin S/clostridiolysin S

Streptolysin S (SLS) is a potent, non-immunogenic, cytolytic toxin secreted typically by group A streptococci (GAS) strains, mediating complete breakdown of erythrocytes and, therefore, being responsible for the signature beta-haemolytic phenotype of the organism on blood agar plates[400]. Nevertheless, the repertoire of SLS-susceptible targets is wide, and extends to lymphocytes, polymorphonuclear leukocytes, platelets, tissue culture cell lines, tumour cells as well as intracellular organelles, prokaryotic protoplasts and spheroplasts[401]. Thanks to its stability in the presence of oxygen it has been named 'S', to differentiate it from an oxygen-labile toxin of the same organism (streptolysin 'O')[402]. A distinctive characteristic of SLS is that it is active only when cell-bound or in association with suitable high-molecular-weight carrier molecules (also called inducers), such as RNA core or serum albumin, being rapidly inactivated otherwise[401-403].

Although the precise chemical structure of SLS has not been elucidated[400], it was recently established that it acts via disruption of the major erythrocyte anion exchange protein (band 3), causing rapid influx of Cl^- ions into erythrocytes and thus leading to osmotic imbalance and cell rupture[404]. SLS is considered a virulence factor[400, 405] and has been shown to promote disease progression *in vivo*, particularly in the context of skin and soft tissue infections[406-408]. Apart from exerting a direct cytotoxic effect on target tissue cells, SLS can also mediate killing of immune cells recruited at the site of infection (such as neutrophils and macrophages), thus interfering with phagocytic clearance and indirectly contributing to virulence[400, 407].

The biosynthetic SLS gene cluster in *Streptococcus pyogenes* (designated as *sag* for SLS-associated-genes) consists of nine contiguous genes (*sagA-I*) organised in an operon; *sagA* is now accepted as the structural gene for SLS, encoding a small peptide (53 amino

acids; 2.7 kDa) with a putative signal sequence[409]. The remaining genes encode for a dehydrogenase (*sagB*), a cyclodehydratase (*sagC*), a 'docking' scaffold (*sagD*), a putative membrane-spanning peptidase/immunity protein (*sagE*), a membrane-associated protein of unknown function (*sagF*) and an ABC transporter complex (*sagG-I*)[400, 409, 410]. Importantly, inactivation of each gene individually has been shown to ablate the haemolytic phenotype of GAS strains on blood agar[407, 409], indicating that an intact operon is required for haemolytic activity.

The small size, amino acid composition and cysteine-rich nature of SagA, in combination with the *sag* operon architecture, have long been suspected as reminiscent characteristics of a bacteriocin-like toxin[407, 409]; subsequent studies have now confirmed that SagA is produced in a manner analogous to microcin B17 from *E. coli*, whereby a non-toxic precursor of SagA (prepropeptide) undergoes extensive post-translational modifications by SagBCD and is converted to a toxic form via the cooperative installation of heterocyclic rings on serine, threonine and cysteine residues[408, 410]. Subsequently, export of the mature toxin is realised via the dedicated ABC transporter complex (SagG-I)[409] while SagE is likely to be mediating proteolytic cleavage of the signal peptide concomitant with export[400]. Also consistent with bacteriocin operons elsewhere, there is a leaky rho-independent terminator downstream of *sagA* which permits the polycistronic transcription of *sagA-I*; the terminator is thought to contribute to a stoichiometric excess of the *sagA* mRNA relative to the other gene transcripts, establishing an energetically favourable state for the cell given that many molecules of SagA can be processed for activity and export by fewer molecules of the associated gene products[400, 409].

Comparative genomics have revealed the conservation of an SLS-type gene cluster in several prokaryotes including *L. monocytogenes*, *S. aureus* and *C. botulinum*/*C. sporogenes*. In fact, amongst the organisms investigated, the gene clusters of *C. botulinum* and *C. sporogenes* (*closA-I*) bear the highest homology to the *sagA-I* operon in terms of content and organisation (Figure 6.1)[410]. To reflect this homology, the

clostridial toxin has been named 'clostridiolysin S' (CLS)[94]. Crucially, *closA*, *closC* and *closD* are functionally equivalent to their streptococcal counterparts, being able to restore haemolytic activity when complemented in the corresponding SLS$^-$ GAS strains[94], while ClosA can be activated by SagBCD *in vitro* and generate equal cytotoxicity to SagA[408, 410]. Additionally, the mechanism of ClosA activation has been studied in detail and falls under the classical paradigm of heterocycle incorporation, as described previously[94].

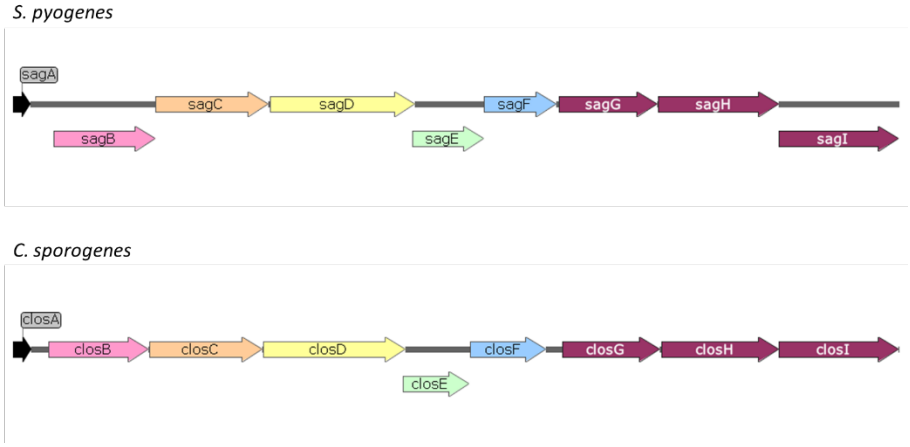


Figure 6.1 The streptolysin S/clostridiolysin S gene clusters. Genetic organization of the streptolysin S gene cluster (*sagA–I*) from *S. pyogenes* (8290 bp) and the clostridiolysin S gene cluster (*closA–I*) from *C. sporogenes* (8483 bp).

Nevertheless, the potential contribution of CLS in the pathogenicity of *C. sporogenes* (or *C. botulinum*) has not been investigated. *C. sporogenes* is a natural inhabitant of the gut[411, 412] and, in that context, it is plausible that CLS fulfils the formal role of bacteriocins in controlling competing bacterial flora, without causing collateral damage to eukaryotic cells[94]. Additionally, preclinical studies with this strain have not so far revealed pathogenic traits *in vivo*[25, 43, 49]. It should be noted that in cases where *in vivo* or *in vitro* cytotoxicity has been documented with clinical isolates of *C. sporogenes* (not necessarily NCIMB 10696), as discussed in section 6.1.3, the described characteristics of the causative toxin(s) differ significantly from those of CLS; these include the small

repertoire of toxin-susceptible cells (as an illustration, haemorrhagic activity was exhibited in rabbits but not mice, rats and guinea pigs[395, 396]), a preference for endothelial cells[395], a size of 65-80 kDa[413], type III/IV collagenase activity[413] as well as the ease with which active protein was purified[413]. The possibility of a different toxin having a contributory role in pathogenicity (or the degree to which this should be a cause of concern) will have to be rigorously investigated preclinically.

In the context of cancer therapy, creation of a CLS-deficient strain might be considered a double-edged sword. On the one hand, given the toxin's activity against cells of the immune system[401], its removal can be seen as beneficial towards the generation of an effective anti-tumour immune response (as discussed in Chapter 5). On the other hand, given its activity against tumour cells[414], its elimination might compromise the native colonising or oncolytic properties of the strain. Yet, in view of the safety concerns and regulatory hurdles discussed previously, creation of a CLS-deficient strain was sought regardless. Furthermore, classification of *C. sporogenes* as a hazard group 2 organism goes hand in hand with corresponding physical containment measures; as a result, the pursuit of a non-haemolytic strain was also dictated by practical considerations (particularly relevant to our collaborators in the Netherlands) with the aim of down-classifying *C. sporogenes* to hazard group 1 and allowing laboratory handling of the organism without the need for specialised containment facilities[c].

6.2 Objectives

- Create a non-haemolytic (attenuated) strain of *C. sporogenes* NCIMB 10696, by removal of the CLS biosynthetic gene cluster, to be used subsequently as the parent strain for the generation of clinically applicable strains.

[c] Where a strain is attenuated or has lost known virulence genes, then the containment required by the classification of its parent strain need not necessarily apply, subject to assessment appropriate for risk in the workplace (Classification of biological agents, RIVM Letter report 205084002/2012, M. R. Klein).

6.3 Hypothesis

- Deletion of the CLS biosynthetic gene cluster (genes *closA-I*) abolishes the haemolytic phenotype of *C. sporogenes* NCIMB 10696.

6.4 Study design

The haemolytic phenotype of the NCIMB 10696 strain had not been previously investigated (let alone confirmed), neither in our group nor elsewhere. Therefore, a basic characterisation of the strain's haemolytic activity was first carried out which also informed the design of downstream experiments.

Based on previous findings with *C. sporogenes* ATCC 15579, it was assumed that inactivation of single genes (*closA* or *closC*) could abolish haemolytic activity[94]. However, with the CRISPR/Cas9 technology available, deletion of the entire 8.5 kb CLS operon was pursued, foreseeing that this strategy would minimise the risk of the strain reverting to a haemolytic phenotype outside of the laboratory (for instance, by lateral gene transfer)[185] and be more easily upheld in regulatory claims.

Nevertheless, when interest in creating a non-haemolytic strain first emerged, the CRISPR/Cas9 technology had not yet been optimised in our group. Therefore, ClosTron mutagenesis was initially deployed to investigate the contribution of individual genes in haemolysis, in knowledge that this tool (introducing an erythromycin resistance gene) is not appropriate for the creation of a clinical strain. The results of these initial experiments brought to light some interesting and unexpected findings, which form part of this thesis and are presented herein.

6.3.1 Special considerations for the quantitative assessment of SLS/CLS haemolytic activity

Haemolytic activity is typically quantified in liquid media by measuring spectrophotometrically the amount of haemoglobin released following erythrocyte lysis, relative to lysis by a positive control (here 1% v/v Triton X-100). This usually involves co-incubation of cell-free haemolysin-containing supernatants (or other haemolytic compounds) with red blood cells (RBCs), while the RBC content (0.2-2% v/v) and incubation time (30 min – 4 hrs) of the reaction may vary widely. In the case of SLS, prior to supernatant isolation, cells are treated by addition of albumin, serum, RNA core or other suitable carriers, to allow extraction of the cell-bound toxin into an active extracellular complex; treatment can be applied to pelleted bacterial cells[415, 416] or in liquid bacterial cultures (usually 1 hr before cell harvesting)[94], followed by isolation of the haemolysin-containing supernatant by centrifugation. Importantly, the resulting haemolytic activity is directly linked to the carrier molecule employed, due to variations in toxin affinity and required cell killing time[415]. However, given that SLS is most active in its cell-bound form[415] and preferably in direct contact with targeted cells[407], SLS activity is also measured by incubating washed bacterial cells directly with RBCs, in a suitable buffer[416, 417].

In the present study, to measure haemolytic activity of carrier-bound CLS in cell-free supernatants, *C. sporogenes* was cultured in media with BSA, aliquots were harvested at desired timepoints and the resultant supernatants were used to set up haemolytic reactions as described in section 2.1.17. Considering the limitations imposed by working anaerobically (such as the confined space of the anaerobic cabinet), the protocol of measuring cell-bound haemolytic activity discussed previously would not be easily executed without exposure to atmospheric oxygen. Therefore, *C. sporogenes* was cultured in its usual medium, but supplemented with RBCs (0.8% v/v), as detailed in section 2.1.16; aliquots were retrieved and centrifuged at desired timepoints, followed by direct measurement of haemoglobin in the resultant supernatant. A potential

limitation of this method (hereby designated as 'co-culturing') is that erythrocytes are constantly exposed to a growing bacterial culture, meaning that measured haemolysis is cumulative rather than being due to the amount of toxin present at the specific timepoint.

6.4 Results

6.4.1 Characterisation of the haemolytic activity of *C. sporogenes* NCIMB 10696

The haemolytic activity of *C. sporogenes* NCIMB 10696 was first assessed on solid TYG medium containing 5% v/v of defibrinated horse or sheep blood. The strain was directly streaked on agar plates from a frozen stock. Haemolytic zones appeared after 48 hrs of incubation and were visible as a clearing of the agar around the area of bacterial growth, indicative of beta haemolysis and consistent with CLS expression. The haemolytic activity was more pronounced against horse blood (Figure 6.2A); as such, horse blood was employed in all subsequent solid and liquid media assays.

In the co-culturing assay, haemolytic activity was low during the lag phase of growth (2 hrs), peaking at the transition to stationary phase (6 hrs; >100% compared to the positive control) and remaining maximal thereafter (Figure 6.2B). A cell-density dependent pattern of SLS expression has been observed previously, with SLS activity peaking similarly at the late log phase[406, 418], reflected genetically in a progressive increase of *sagA* expression with growth[419]; in fact, growth-dependency has been attributed to SLS autoinduction, in a quorum-sensing analogous manner[419].

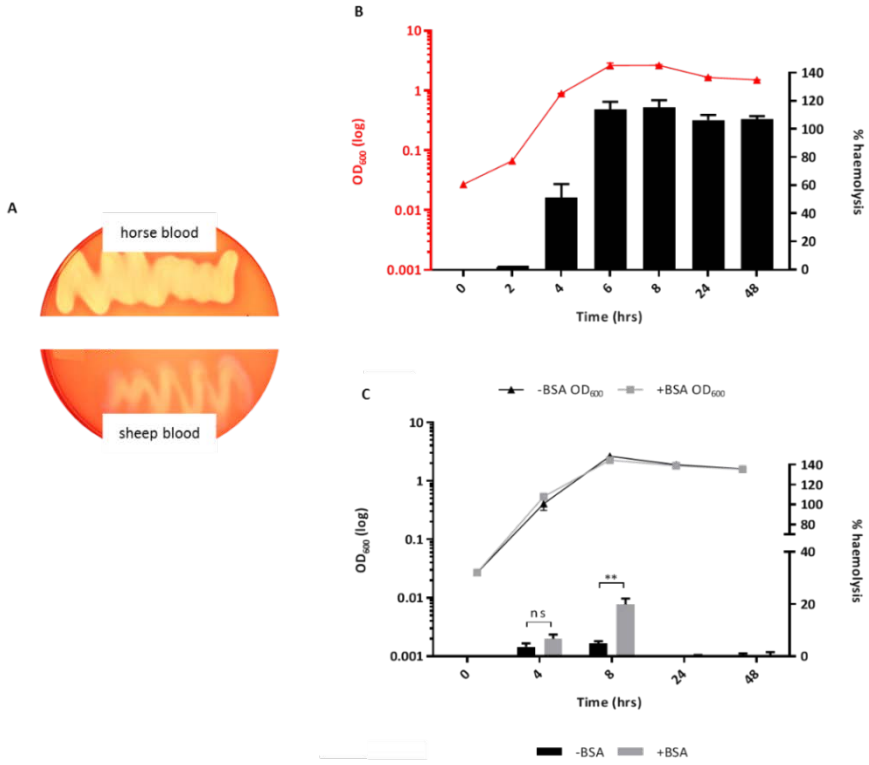


Figure 6.2 Characterisation of the haemolytic activity of *C. sporogenes* NCIMB 10696 in solid and liquid media. A. *C. sporogenes* on TYG agar with 5% v/v horse or sheep blood at 48 hrs of incubation. **B.** Haemolytic activity of cell-bound CLS; *C. sporogenes* was cultured in liquid TYG/0.8% v/v RBC (co-culturing method). OD_{600} was monitored in replicate cultures with no RBCs. Data represent mean values of three independent cultures ± SD, relative to haemolysis detected in the presence of 1% v/v Triton X-100. Depicted data are from a single representative experiment out of two repeats. **C.** Haemolytic activity of carrier-bound CLS, in cell-free culture supernatants with (+) or without (-) BSA as the molecular carrier; *C. sporogenes* was cultured in liquid TYG with/without BSA and retrieved supernatants were reacted in a 1:1 vol ratio with 0.8% v/v RBC at 37°C for 1 hr. Data represent mean values of three reactions ± SD, relative to haemolysis detected in the presence of 1% v/v Triton X-100. Depicted data are from a single representative experiment out of two repeats. Statistical analysis was carried out using two-tailed unpaired t-test; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

Cell-free supernatants were haemolytic at 4 and 8 hrs but no activity could be detected thereafter. As expected, haemolytic activity was higher in the presence of BSA (4-fold higher at 8 hrs), consistent with the requirement of a carrier molecule to prevent CLS

inactivation (Figure 6.2C). Unsurprisingly, the haemolytic activity of cell-free supernatants (carrier-bound CLS) was 6- to 8- fold lower than the activity derived from co-cultures (cell-bound CLS), at 4 and 8 hrs respectively, confirming the superior cytotoxicity of cell-bound toxin. It is also possible that BSA is not an efficient carrier for CLS, or that the implemented reaction time (1 hr) was not sufficient for detectable erythrocyte lysis[415]. Failure to detect haemolytic activity from 8 hrs onwards coincides with the commencement of protease secretion in *C. sporogenes*[343, 349] and may indicate proteolytic inactivation of CLS[402], in absence of a sufficient/efficient carrier. Given the aforementioned limitations, it was reasoned that the co-culturing protocol is more reliable for assessing haemolytic activity in *C. sporogenes* and was therefore employed in all liquid assays thereafter.

6.4.2 Deletion of the clostridiolysin S gene cluster

CRISPR/Cas9-mediated deletion of *closA-I* was carried out as described in section 2.2.12.2, using the pMTL83121-RECas1-*closA-I* plasmid, whose construction is outlined in section 2.2.11.7. Deletion was verified by colony PCR in three biological replicates (Figure 6.3). Operon-flanking primers (clos_Fscr, clos_Rscr) yield a PCR product only if the operon is deleted, the WT product being larger than the maximum capacity of the polymerase employed (11.7 kb). Concurrently, to confirm absence of the WT product, an operon-internal and an operon-flanking primer were used (clos_Fscr, closC_R). The plasmid was lost from confirmed Csp-Δ*closA-I* mutants, as described in section 2.2.13.

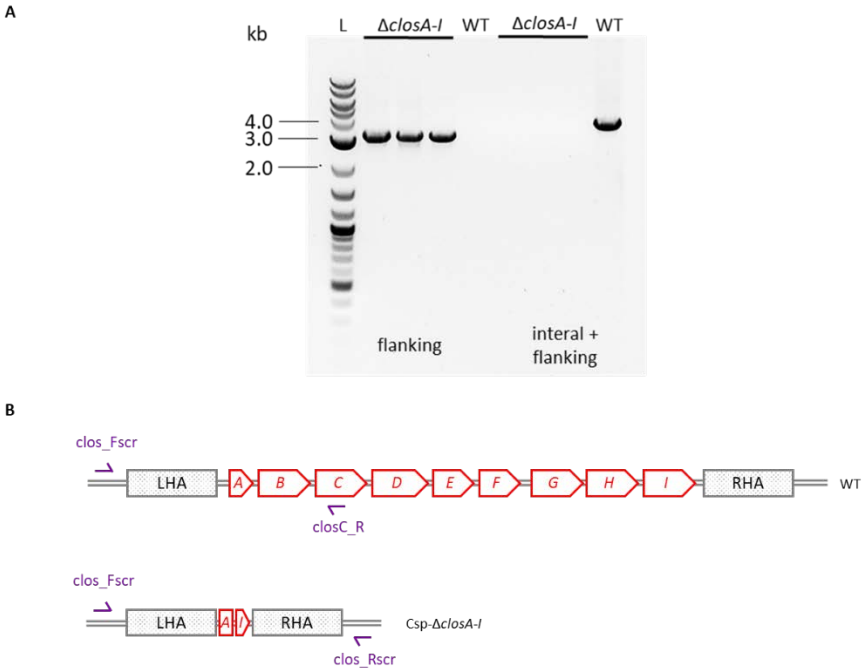


Figure 6.3 Deletion of the clostridiolysin S operon in *C. sporogenes* NCIMB 10696. A. Colony PCR for three biological replicates of Csp-Δ*closA-I* and WT. Products amplified with flanking primers (clos_Fscr/clos_Rscr): Csp-Δ*closA-I*: 3259 bp, WT: no band (11727 bp). Products amplified with one internal and one flanking primer (clos_Fscr/closC): Csp-Δ*closA-I*: no band, WT: 3954 bp. **B.** Schematic diagram showing the primers used for colony PCR screening and their binding sites (in purple) in the WT and Csp-Δ*closA-I* strains.

To confirm that relevant phenotypic characteristics of the parent strain had not been compromised by deletion of the operon, the growth and sporulation/germination capacities of Csp-Δ*closA-I* were assessed as described in sections 2.1.12 and 2.1.13, respectively; the growth and sporulation profiles of Csp-Δ*closA-I* were identical to those of WT (Figure 6.4). Furthermore, the genomes of the three biological replicates of Csp-Δ*closA-I* were sequenced in their entirety, as described in section 2.2.9, confirming that the CLS gene cluster had been deleted and that no additional alterations were present (results not shown).

Finally, the haemolytic activity of Csp-Δ*closA-I* was assessed in solid and liquid media as described previously. Surprisingly, all three biological replicates generated haemolytic zones of clearance on agar plates, similar in size to that of WT (Figure 6.5). The haemolytic zones were visible at 48 hrs and appeared simultaneously for WT and the Csp-Δ*closA-I* mutants. However, the results of the co-culture liquid haemolysin assay paint a different picture; the Csp-Δ*closA-I* strain exhibited significantly reduced haemolytic activity compared to WT at all time points assayed. Nevertheless, the strain remained moderately haemolytic, its haemolytic activity increasing with incubation time, being the highest at 48 hrs (42% relative to the positive control). Based on these results, it was assumed that contrary to what was seen with ATCC 15579, the CLS operon, although a major contributor to haemolytic activity in NCIMB 10696, is not the sole one.

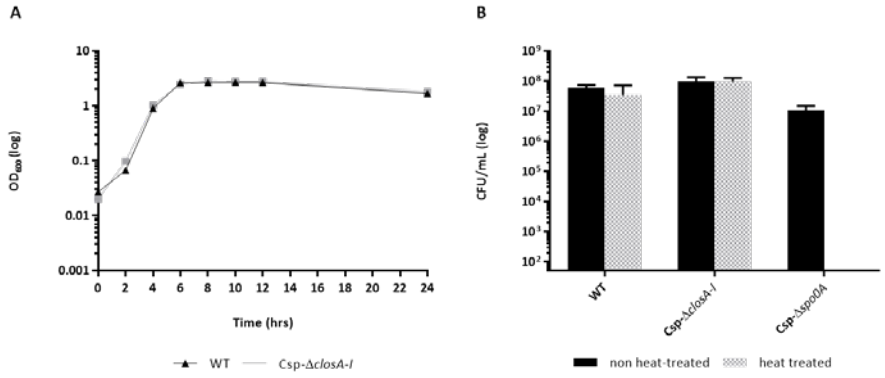


Figure 6.4 Phenotypic characterisation of Csp-Δ*closA-I*. A. Growth profile of Csp-Δ*closA-I vs* WT; strains were cultured in liquid TYG and OD_{600} was monitored over 24 hrs. Data represent mean values of three independent cultures for WT and of three biological replicates for Csp-Δ*closA-I* ± SD, from a single representative experiment out of three repeats. **B.** Sporulation and germination capacity of Csp-Δ*closA-I vs* WT, alongside an asporogenous strain (Csp-Δ*spo0A*) as the negative control, at 120 hrs of incubation; heat-treated CFUs correspond to the number of spores with germinating ability and non heat-treated CFUs to the total cell count. Data represent mean values of three independent cultures for WT and of three biological replicates for Csp-Δ*closA-I* ± SD. The y-axis starts at 50 CFU/ml (detection limit). Depicted data are from a single representative experiment out of two repeats.

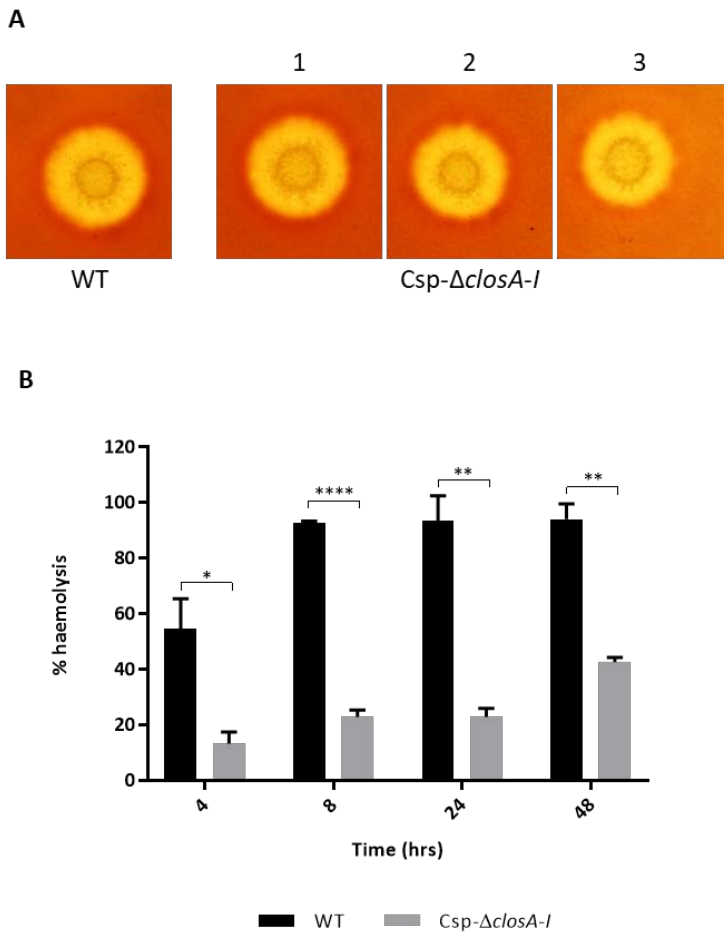


Figure 6.5 Haemolytic activity of Csp-Δ*closA-I*. **A.** Haemolytic phenotype on 5% v/v horse blood agar at 48 hrs of incubation **B.** Haemolytic activity in liquid TYG/0.8% v/v RBC. Data represent mean values of three independent cultures for WT and of three biological replicates for Csp-Δ*closA-I* ± SD, relative to haemolysis detected in the presence of 1% v/v Triton X-100. Depicted data are from a single representative experiment out of four repeats. Statistical analysis was carried out using two-tailed unpaired t-tests; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

6.4.3 Deletion of putative haemolysins

To identify genes responsible for the residual haemolytic activity of Csp-Δ*closA-I*, the genome of *C. sporogenes* NCIMB 10696 was searched for putative haemolysins, using the two available annotations (CP009225.1/NZ_CP009225.1). The protein products of the putative genes were further analysed for the presence of signal peptides, transmembrane helices and conserved functional domains using the SignalP 4.1, TMHMM version 2.0 and NCBI CD-search online tools respectively, as described in sections 2.6.10-2.6.13 (Table 6.1).

Table 6.1 Putative haemolysins in *C. sporogenes* NCIMB 10696

no.	gene identifier	annotation		NCBI protein code	size (aa)	signal peptide	TMDs	CDs
		CP009225.1	NZ_CP009225.1					
1	c14780	haemolysin III/D	haemolysin III/D	WP_003491043.1	212	N	7	YqFa/HlyIII
2	c16380	hypothetical protein UPF0073	haemolysin	WP_033059280.1	216	N	6	YqFa/HlyIII
3	c01750	putative bacteriocin	haemolysin D	WP_033058096.1	476	N	1	HlyD
4	c31080	haemolysin XhlA-like protein	putative bacteriocin ABC transporter	WP_033061004.1	84	N	1	XhlA
5	c29800	HlyC-CorC family protein	haemolysin	WP_033060876.1	441	N	4	TlyC/DUF21/ CorC_HlyC

aa: amino acids, TMDs: transmembrane domains, CDs: conserved domains, N: not present

None of the identified genes is predicted to be a secreted protein, while 3/5 genes appear to contain a significant number of transmembrane helices, indicating they could be cell membrane-associated haemolysins. Genes c01750 and c31080 appear to be part of operons, with genes c01740 encoding a peptidase domain-containing ABC transporter and c31070 encoding a hypothetical protein, respectively. Genes homologous to c14780 and c16380 have been formerly identified as putative haemolysins in the *C. botulinum* ATCC 3502 Hall A strain[345]; the functional domains HlyIII and YqFa and the predicted

membrane localisation of the proteins suggest they might be functionally related to pore-forming haemolysins from *Bacillus cereus*[420, 421] and *Vibrio vulnificus*[422]. The functional domain of c31080 (XhlA) has been previously associated with haemolysins from *Clostridium chauvoei*[423], *Xenorhabdus nematophila*[424] and *Bacillus* strains[425]. The assigned functional domain for the protein product of c01750 (HlyD) corresponds to the membrane fusion protein of the prototypical alpha haemolysin type I secretion system in *E. coli*[426], while the TlyC domain of c29800 is found in a haemolysin from Rickettsia typhi[427]. Building on the functional domain analysis, it was decided to investigate the role of these putative haemolysins in *C. sporogenes* experimentally.

All five genes were knocked out individually as well as sequentially (1 to 5) in the WT and Csp-Δ*closA-I* strains, using CRISPR/Cas9 and plasmids pMTL83151-RECas1-c14780, pMTL83151-RECas1-c16380, pMTL83151-RECas1-c01750, pMTL83151-RECas1-c31080 and pMTL83151-RECas1-c29800, constructed as described in section 2.2.11.7. All generated strains can be found in Table 6.2 and Appendix I. The plasmid was lost from all strains, as described in section 2.2.13. Gene deletions were confirmed by colony PCR for three biological replicates of each strain, using suitable chromosomal primers annealing outside of the editing templates. Figure 6.6 shows representative data from the colony PCR confirmation of the final strain (Csp-Δ1-5), generated after the sequential knockout of all five putative haemolysins in the WT parent strain. The growth profile of all generated strains was similar to WT and Csp-Δ*closA-I* (results not shown).

Table 6.2 Strains generated to study the contribution of putative haemolysin genes

Strain	Description
Csp-Δc14780	c14780 deletion
Csp-Δc16380	c16380 deletion
Csp-Δc01750	c01750 deletion
Csp-Δc31080	c31080 deletion
Csp-Δc29800	c29800 deletion
Csp-Δ*closA-I* Δc14780	*closA-I* and c14780 deletion
Csp-Δ*closA-I* Δc16380	*closA-I* and c16380 deletion
Csp-Δ*closA-I* Δc01750	*closA-I* and c01750 deletion
Csp-Δ*closA-I* Δc31080	*closA-I* and c31080 deletion
Csp-Δ*closA-I* Δc29800	*closA-I* and c29800 deletion
Csp-Δ1-2	c14780 and c16380 deletion
Csp-Δ1-3	c14780, c16380 and c01750 deletion
Csp-Δ1-4	c14780, c16380, c01750 and c31080 deletion
Csp-Δ1-5	c14780, c16380, c01750, c31080 and c29800 deletion
Csp-Δ*closA-I* Δ1-2	*closA-I*, c14780 and c16380 deletion
Csp-Δ*closA-I* Δ1-3	*closA-I*, c14780, c16380 and c01750 deletion
Csp-Δ*closA-I* Δ1-4	*closA-I*, c14780, c16380, c01750 and c31080 deletion
Csp-Δ*closA-I* Δ1-5	*closA-I*, c14780, c16380, c01750, c31080 and c29800 deletion

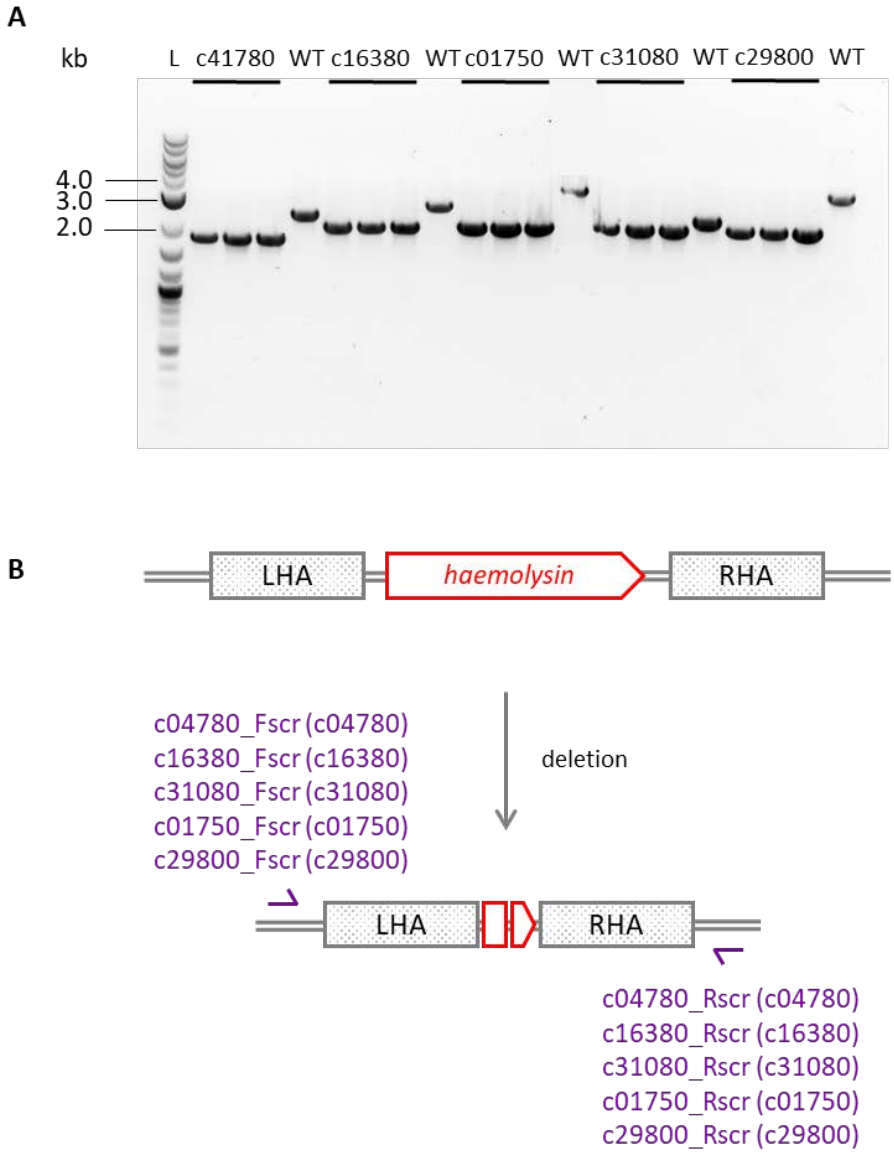


Figure 6.6 Sequential deletion of putative haemolysins in *C. sporogenes* NCIMB 10696. A. Colony PCR for three biological replicates of Csp-Δ1-5 alongside WT, to confirm the deletion of each indicated haemolysin gene. Δc14780: 1850 bp, WT: 2474 bp**;** Δc16380: 2135 bp, WT: 2771 bp; Δc01750: 2119 bp, WT: 3538 bp**;** Δc31080: 2121 bp, WT: 2361 bp; Δc29800: 2107 bp, WT: 3418 bp. L: NEB 2-Log DNA ladder. **B.** Schematic diagram showing the primers used for colony PCR screening and their binding sites (in purple), as per the indicated locus (in brackets, next to the primer name).

The haemolytic activity of the generated strains was assessed on solid and liquid media as described previously. For clarity, Figure 6.7 shows only the results of the final strains Csp-Δ1-5 and Csp-Δ*closA-I* Δ1-5. On solid medium, the strains generated haemolytic zones similar to WT and Csp-Δ*closA-I*, as observed earlier. In liquid media, the haemolytic activity of Csp-Δ1-5 was in general lower than its parent strain (WT), reaching statistical significance at 8, 24 and 48 hrs. Nevertheless, although the cumulative knockout of the putative haemolysins reduced the haemolytic activity of the WT strain, it is evident that CLS remains the major contributor of haemolytic activity; indeed, in the absence of the CLS operon, the additional reduction in haemolytic activity brought about from the deletion of all five putative haemolysins is not of statistical significance. Importantly, the Csp-Δ*closA-I* Δ1-5 strain with the CLS operon and all five putative haemolysins knocked out remains fully haemolytic on solid medium and moderately haemolytic in liquid medium at 48 hrs (44% relative to the positive control), suggesting that other genes in addition to those deleted and/or other mechanisms might be involved.

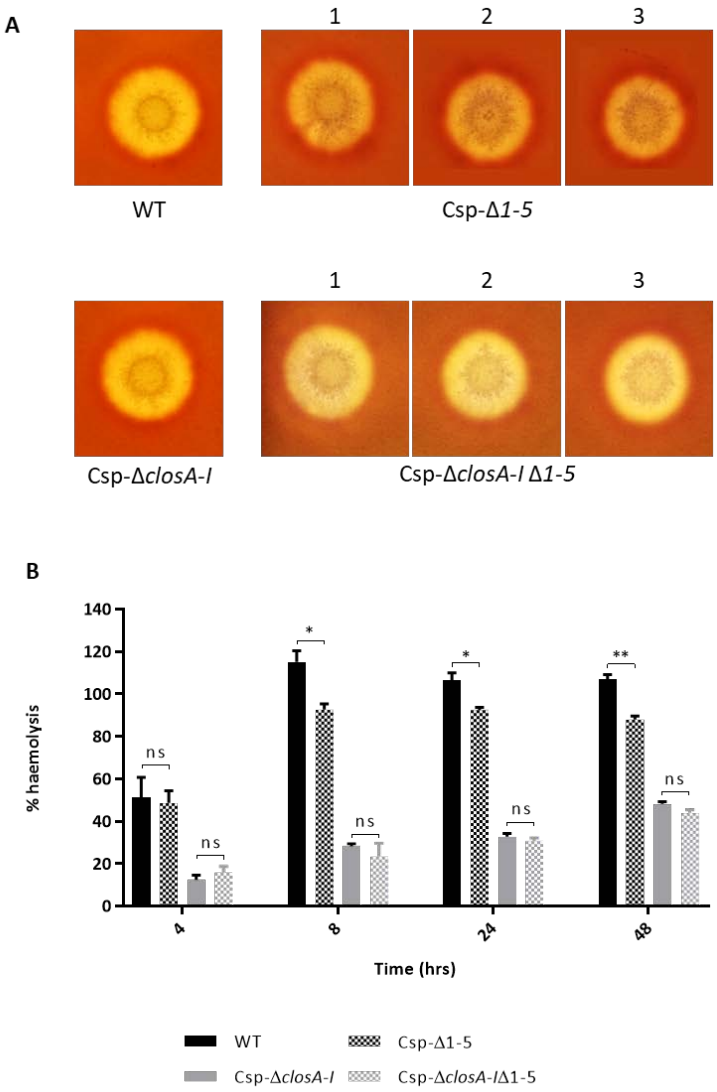


Figure 6.7 Haemolytic activity of Csp-Δ1-5 and Csp-Δ*closA-I* Δ1-5. **A.** Haemolytic phenotype on 5% v/v horse blood agar at 48 hrs of incubation **B.** Haemolytic activity in liquid TYG/0.8% v/v RBC. Data represent mean values of three independent cultures for WT and of three biological replicates for all mutant strains ± SD, relative to haemolysis detected in the presence of 1% v/v Triton X-100. Depicted data are from a single representative experiment out of three repeats. Statistical analysis was carried out using two-tailed unpaired t-tests; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

6.4.4 Inactivation of *closA* and *closC* using ClosTron mutagenesis

As mentioned previously, prior to the circulation of the CRISPR/Cas9 tool in our group, the role of the CLS gene cluster was initially investigated using ClosTron mutagenesis. Based on the study by Gonzalez *et al.*, this thesis sought to inactivate *closA* and *closC*, given that plasmid integrational mutants of the same genes abolished the haemolytic activity of *C. sporogenes* ATCC 15579 on blood agar plates[94]. ClosTron mutagenesis of *closA* and *closC* was carried out as described in section 2.2.12.3, using plasmids pMTL007C-E2-*closA*-101s and pMTL007C-E2-*closC*-749a, respectively. ClosTron-mediated gene inactivation works by propagation of a mobile group II intron, derived from the *ltrB* gene of *Lactococcus lactis*, into a specific DNA target site, mediated by base-pairing between the target site DNA and the excised intron lariat RNA. To facilitate isolation of successful mutants, the group II intron contains an erythromycin resistance gene (*ermB*) interrupted by a self-splicing group I intron, together comprising the retrotransposition-activated marker (RAM); e*rmB* is expressed only upon successful genomic integration of the group II intron, which allows the nested group I intron to be spliced out. Hence, acquisition of erythromycin resistance can be used to phenotypically select successful mutants [68, 99]. According to the existing convention, retargeted introns should be identified by the position (specific base) and orientation, sense (s) or antisense (a), of their insertion[99]. For example, the intron for inactivation of *closC*, Csp-*closC*-749a, should be inserted in the antisense orientation after base 749 of the *closC* open reading frame (ORF) in *C. sporogenes*.

For *closA*, none of the screened erythromycin-resistant colonies had the group II intron inserted correctly. Those analysed further had the intron inserted 192s in the c28980 gene encoding a MarR family transcriptional regulator (results not shown). No further attempts were made at inactivating *closA*, given its small size (174 bp) which constituted it a difficult target. Nevertheless. identification of *closA* ClosTron mutants might have benefitted from plating obtained transconjugants on blood agar plates, given

the anticipated non-haemolytic mutant phenotype. A similar phenotypic screen, not relying on the acquisition of erythromycin resistance, was employed to obtain *C. perfringens* mutant colonies with a group II insertion in the *plc* gene (encoding alpha toxin); the screen was based on the disappearance of the white halo around mutant colonies when plated on media containing egg yolk[428]. However, in view of the knowledge obtained retrospectively regarding the inability to abolish the haemolytic phenotype of the strain by deletion of the CLS gene cluster, this would have been fruitless effort. -

For *closC*, 13% of the screened erythromycin-resistant colonies had the group II intron inserted in the intended target, whereas in the rest, the intron was inserted 513s in the thymidine kinase gene (*tdk*) (results not shown). Figure 6.8A shows the results of colony PCR screening for three biological replicates of Csp-*closC*749a::CT with three sets of primers, confirming insertion of the intron in *closC*, presence of the RAM cassette and the *closC*-intron junction. The plasmid was lost from the confirmed biological replicates, as described in section 2.2.13. Inverse PCR and sequencing of the resulting PCR products was carried out as described in section 2.2.3.2, to confirm intron insertion at the desired target gene and as a single copy in the entire genome (Figure 6.8B). Finally, the haemolytic activity of the plasmid-free Csp-*closC*749a::CT mutants was assessed on blood agar plates as described previously. Based on the predicted point of insertion, integration of the group II intron in *closC* is not expected to interfere with the expression of downstream genes of the operon (*closD-I*), as shown by their undisturbed ORFs in Figure 6.9.

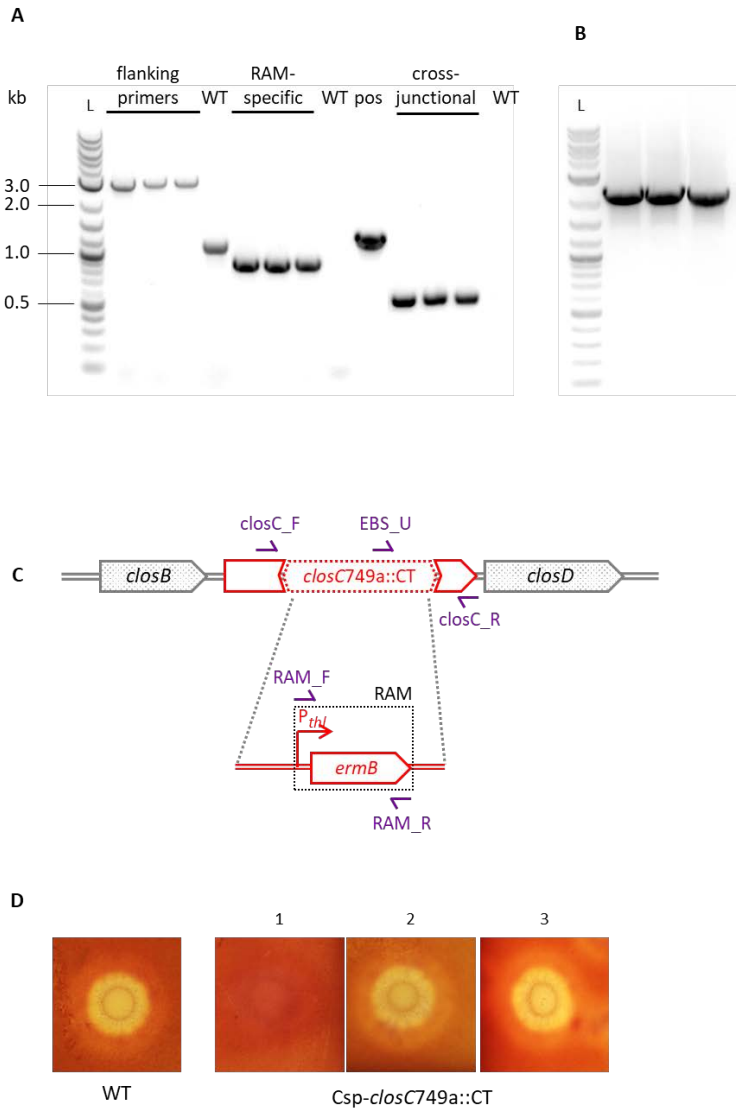


Figure 6.8 ClosTron-mediated inactivation of *closC* and assessment of the strain's haemolytic phenotype on blood agar. A. Colony PCR for three biological replicates of Csp-*closC*749a::CT alongside WT. Gene-flanking closC_F/closC_R primers: 2890 bp, WT: 1041 bp; RAM-specific RAM_F/RAM_R primers: 860 bp (group I intron spliced out), WT: no band, pos: 1253 bp (pMTL007C-E2-*closC*-749a used as the DNA template; contains the group I intron; positive control); cross-junctional EBS_U/closC_R primers: 540 bp, WT: no band. L: NEB 2-Log DNA ladder. **B.** Inverse PCR to confirm single intron insertions in the genomes of the three biological replicates of Csp-*closC*749a::CT. L: NEB 2-Log DNA ladder. Bands correspond to products amplified with inv_F and inv_R; single band confirms insertion in a single copy. **C.** Schematic diagram showing the primers used for colony PCR screening and their binding sites (in purple). **D.** Haemolytic phenotype of WT and Csp-*closC*749a::CT (1-3) strains on 5% v/v horse blood agar at 48 hrs of incubation.

Surprisingly, replicate no. 1 of Csp-*closC*749a::CT had no haemolytic activity on solid medium, while replicates no. 2 and 3 generated haemolytic zones of clearance, similar in size to that of WT (Figure 6.8C) and, retrospectively, in agreement with the results obtained with the Csp-Δ*closA-I* strain. Given that the group II intron was inserted as a single copy, at the same (predicted) location in all three replicates and that 2/3 replicates did not abolish haemolytic activity, it was assumed that the non-haemolytic phenotype of replicate no. 1 could only be accounted for by the presence of randomly acquired mutations elsewhere in the genome. The strain was renamed as Csp-*closC*749a::CT-NH (non-haemolytic), to differentiate it from the other two haemolytic biological replicates.

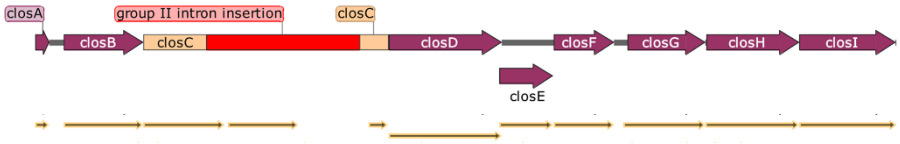


Figure 6.9 Schematic diagram showing the group II intron insertion in *closC*. The group II intron is inserted in the antisense orientation after the 749th bp of the *closC* ORF; the predicted ORFs for the genes of the CLS cluster, following insertion of the group II intron, are depicted as orange arrows.

6.4.5 Whole genome sequencing of Csp-*closC*749a::CT-NH

Intrigued by the non-haemolytic phenotype of Csp-*closC*749a::CT-NH, the strain's genome was sequenced in its entirety, according to section 2.2.9. The WT, Csp-*closC*749a::CT_2 and Csp-*closC*749a::CT_3 strains were sequenced alongside for comparison purposes. Sequencing data were processed and analysed for the presence of mutations, as described in section 2.6.4. Two additional mutations were identified in Csp-*closC*749a::CT-NH, which were absent in WT and the haemolytic replicates of Csp-*closC*749a::CT (Figure 6.10).

The first mutation consists of a 3-bp insertion (TAT) in the *resE4* gene (c03700) which encodes an orphan sensor histidine kinase. The mutation results in an in-frame insertion of a leucine (L) in the MARVEL domain of the protein. The second mutation consists of a 1-bp alteration in the *flhA* gene (c27230) which creates a premature stop codon in the mRNA of the encoded protein, involved in flagellar biosynthesis.



C

Reference Position	Type	Length	Reference	Allele	Count	Coverage	Frequency	Coding region change	Amino acid change
389232	Insertion	3	-	TAT	21	23	91	resE4:c.157_158insTAT	resE4:p.L55_S56insL
2980538	SNV	1	G	A	103	103	100	flhA:c.1210C>T	flhA:p.Q404*

Figure 6.10 Mutations identified in Csp-*closC*749a::CT-NH. Mutations were identified in the non-haemolytic Csp-*closC*749a::CT-NH strain via Illumina whole genome sequencing. Data were assembled and analysed using the CLC Genomics Workbench against the sequenced WT genome as a reference. The depicted sequence alignments were performed using SnapGene. **A.** Mutations in the *resE4* gene **B.** Mutations in the *flhA* gene **C.** Mutation analysis.

To study the effect of these mutations separately, it was attempted to re-create them in the WT strain using CRISPR/Cas9, whereby the editing template should contain the desired (point) mutation(s) as well as silent mutations in the sgRNA and PAM sequences, so that upon homologous recombination the Cas9 target is removed (as required for cell survival) concurrently with the introduction of the desired mutations. Given that precise mutations do not generate obvious size alterations, screening of successful mutants is feasible with a primer annealing on the mutated sgRNA/PAM sequence (23 bp) in combination with another suitable chromosomal primer, whereby

the generation of a PCR product is a positive indicator of successful mutagenesis. However, for both *resE4* and *flhA*, this technique was unsuccessful in introducing the desired mutations; it appeared that in colonies that had successfully undergone homologous recombination (based on the presence of the mutated sgRNA/PAM sequence), the recombined region was limited to 200-650 bp proximal to the sgRNA, leaving out the region where the mutations were supposed to be introduced. This was a shortcoming that had not been foreseen but could be overcome in the future by choosing sgRNAs as close as possible to the desired mutations.

Instead, due to time limitations, it was attempted to study these mutations by gene complementation in *trans*, on pMTL82121 plasmids. However, cloning problems were encountered in *E. coli* with *resE4* and transformation problems in *C. sporogenes* with *flhA*; it was assumed that the latter were due to the choice of a constitutive promoter (P*ptb*), as opposed to the gene's native promoter, or inappropriate gene dosage arising from the high-copy plasmid deployed. Therefore, a previously generated *C. sporogenes* NCIMB 10696 strain was employed, with a transposon insertion in the *resE4* gene (RW80), which would allow the investigation of *resE4* in haemolysis; unfortunately, the role of *flhA* could not be investigated at this stage.

6.4.6 Investigating the role of the sensor histidine kinase *resE4*

The haemolytic activity of RW80 was assessed on solid and liquid media, as described previously, alongside WT, Csp-closC749a::CT-NH and Csp-Δ*closA-I* for comparison purposes.

RW80, similarly to Csp-*closC*749a::CT-NH, was non-haemolytic on solid media; however, while the haemolytic activity of Csp-*closC*749a::CT-NH in liquid media was markedly reduced (similar to Csp-Δ*closA-I* at all time points assayed), the haemolytic activity of RW80 was the same as that of WT (Figure 6.11).

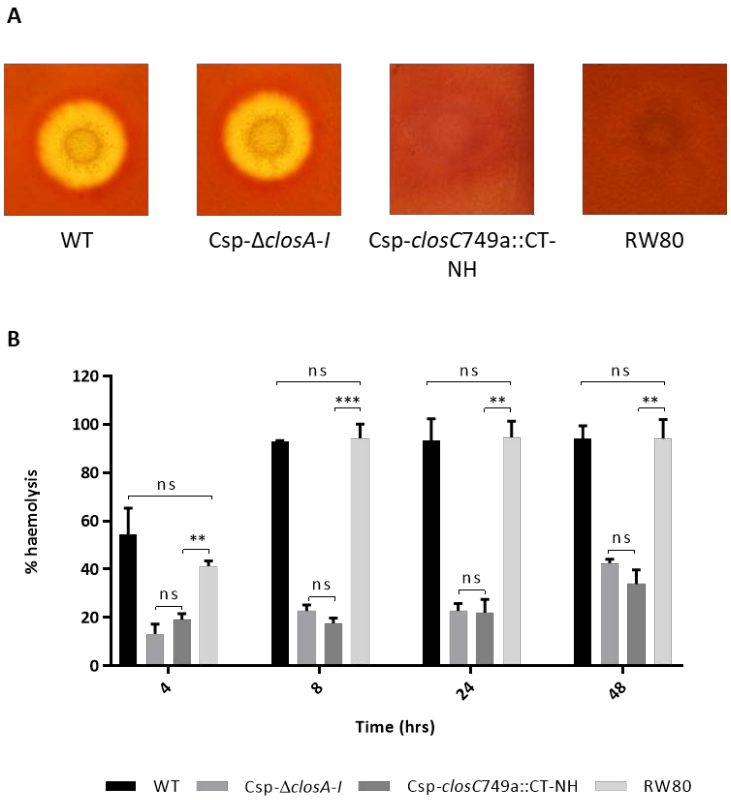


Figure 6.11 Haemolytic activity of RW80. A. Haemolytic phenotype on 5% v/v horse blood agar at 48 hrs of incubation **B.** Haemolytic activity in liquid TYG/0.8% v/v RBC. Data for WT, RW80 and Csp-clsC749a::CT-NH represent mean values of three independent cultures ± SD, relative to haemolysis detected in the presence of 1% v/v Triton X-100. Data for Csp-Δ*closA-I* represent mean values of three biological replicates ± SD, relative to haemolysis detected in the presence of 1% v/v Triton X-100. Depicted data are from a single representative experiment out of three repeats. Statistical analysis was carried out using two-tailed unpaired t-tests; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

6.4.7 Haemolytic assays using media containing red blood cells *vs* whole blood

Liquid haemolytic assays typically employ washed RBCs in contrast to agar plates which are usually prepared with whole blood (WB). This difference probably arises due to the need of eliminating residual haemoglobin released from the spontaneous

breakdown of RBCs during storage, to minimise the absorbance background in the liquid assays. To rule out that differences in the consistency of blood contribute to the discrepancies observed in the haemolytic profiles between solid and liquid media, the liquid co-culture assay was carried out in 2% v/v WB (equivalent to 0.8% v/v RBC, assuming 40% v/v RBC content in WB) and the strains were also assayed on agar plates containing 2% v/v (equivalent to 5% v/v WB) and 0.8% v/v (equivalent to 2% v/v WB) RBCs only.

WB liquid medium generally resulted in higher haemolytic activities than RBC-only medium, the difference being of significance for Csp-Δ*closA-I* (Figure 6.12B). On solid medium containing only RBCs no visible haemolytic zones could be identified for any of the strains assayed, even after prolonged incubation times (Figure 6.12A); if haemolytic zones were present, there was no observable difference between the different strains.

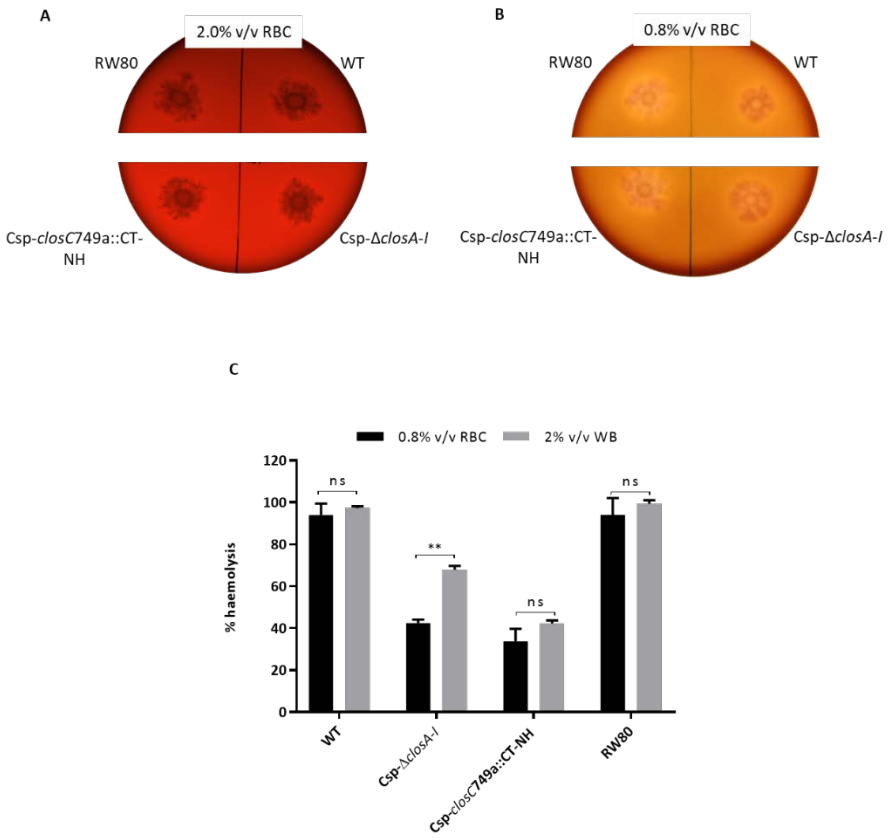


Figure 6.12 Comparison of haemolysis in red blood cell-only and whole blood media. A. Haemolytic phenotype on 2% v/v RBC agar at 48 hrs of incubation **B.** Haemolytic phenotype on 0.8% v/v RBC agar at 48 hrs of incubation. **C.** Haemolytic co-culture activity in liquid TYG with 0.8% v/v RBC or 2% v/v WB at 48 hrs of incubation. Data for WT, RW80 and Csp-clsC749a::CT-NH represent mean values of three independent cultures ± SD, relative to haemolysis detected in the presence of 1% v/v Triton X-100. Data for Csp-Δ*closA-I* represent mean values of three biological replicates ± SD, relative to haemolysis detected in the presence of 1% v/v Triton X-100. Depicted data are from a single representative experiment out of two repeats. Statistical analysis was carried out using two-tailed unpaired t-tests; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

6.4.8 Could mechanisms unrelated to haemolysins contribute to a haemolytic phenotype?

Given the unsuccessful efforts to convert *C. sporogenes* NCIMB 10696 to a non-haemolytic phenotype, the possibility was considered that other mechanisms unrelated to the expression of haemolysins *per se* may be contributing to haemolysis. To this end, the haemolytic activity of *C. acetobutylicum* ATCC 824 was assayed in liquid (21YT) and solid (CGM) media with 0.8% v/v RBC and 5% v/v WB, respectively. *C. acetobutylicum* is non-pathogenic and universally classified as a group 1 organism, with no haemolysin genes in its genome or megaplasmid[429]. The pH of liquid media for *C. acetobutylicum* is typically adjusted to 5.0, but the organism is also able to grow at higher pH. Hereby, such an adjustment was avoided to prevent spontaneous erythrocyte lysis which has been reported at acidic pH[430]; for this reason, growth was slightly delayed, and haemolytic activity was measured from 8 hrs of incubation onwards.

C. acetobutylicum was haemolytic in both solid and liquid media; surprisingly, in liquid media its haemolytic activity peaked at 24 hrs (17% relative to the positive control) but decreased thereafter (Figure 6.13B). Importantly, this level of haemolysis is comparable to that generated by Csp-Δ*closA-I* at the same time point (23% relative to the positive control; Figure 6.5). On solid medium, the strain generated large zones of haemolysis, evident after 24 hrs of incubation; interestingly, cell colour changed from the typical creamy white to brown (Figure 6.13A), suggesting that *C. acetobutylicum* might be accumulating released haemoglobin following erythrocyte lysis. This hypothesis was confirmed with a hemin binding assay (Figure 6.13C), carried out as described in section 2.1.18, and a gene encoding a hemin permease was found in the strain's genome (CAC2442)[431]. Taken together, although the unexpected haemoglobin-binding capacity of *C. acetobutylicum* interfered with the quantitative measurement of its haemolytic activity in liquid media, these data confirm that a strain which is, in theory, considered non-haemolytic may also cause erythrocyte lysis via haemolysin-unrelated mechanisms.

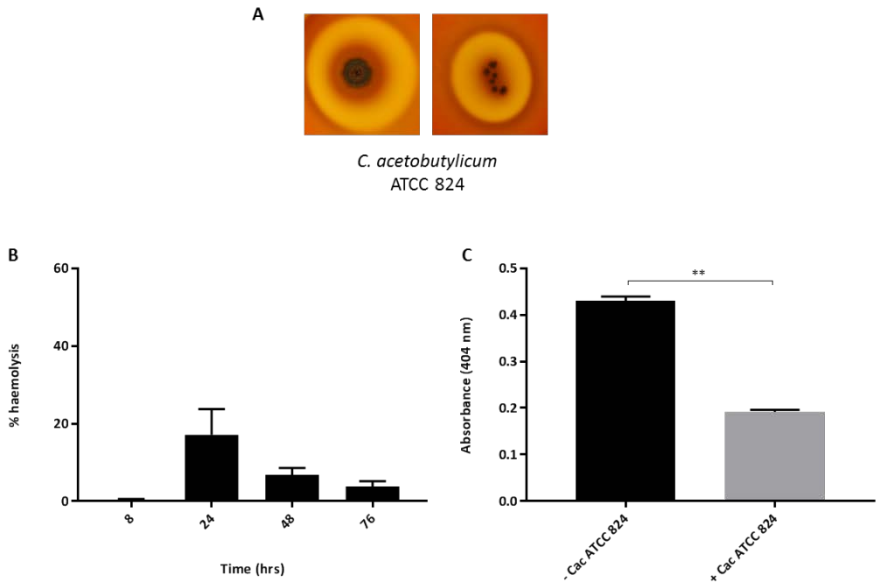


Figure 6.13 Haemolytic activity of *C. acetobutylicum* ATCC 824. A. Haemolytic phenotype on 5% v/v horse blood CGM agar at 48 hrs of incubation **B.** Haemolytic activity in liquid 21TY with 0.8% v/v RBC. Data represent mean values of three independent cultures ± SD, relative to haemolysis detected in the presence of 1% v/v Triton X-100. Depicted data are from a single representative experiment out of three repeats. **C.** Hemin binding assay; reduction in absorbance corresponds to a decrease in the amount of free hemin in the medium following its accumulation by bacterial cells (+Cac ATCC 824). Depicted data are from a single experiment. Statistical analysis was carried out using two-tailed unpaired t-test; p-values are indicated as: 0.1234 (ns), 0.0332 (*), 0.0002 (***), <0.0001 (****).

6.5 Discussion

To aid the present discussion, the haemolytic activity results obtained with the various strains are summarised qualitatively in Table 6.3.

Table 6.3 Summary of haemolytic activity results

Strain	Solid medium	Liquid medium
WT	++++	++++
Csp-Δ*closA-I*	++++	++
Csp-Δ1-5	++++	++++
Csp-Δ*closA-I* Δ1-5	++++	++
Csp-*closC*749a::CT-NH	-	++
RW80	-	++++
C. acetobutylicum	++++	++

In view of the profound lack of a correlation between the 'quantity' of expressed haemolysin(s) and the size/timing of haemolytic zone formation on solid medium (best exemplified by the haemolytic zones of the WT and Csp-Δ*closA-I* strains) as well as the inability to ablate the haemolytic phenotype even after the sequential deletion of all putative haemolysins, the following hypotheses were formulated: a) CLS exerts only a small (if any) contribution on the haemolytic phenotype of the strain on agar, contrary to what happens in liquid, and that b) the apparent haemolytic phenotype of the strain on solid medium is probably not mediated by haemolysins *per se*.

Regarding the first point, discrepancies in the haemolytic phenotype of bacterial strains between solid and liquid media have been observed previously[432, 433], but no precise cause has been stated. Plausibly, they can be attributed to differential gene expression between the two media phases. In support of the second point, the only non-haemolytic strains on solid media (Csp-*closC*749a::CT-NH and RW80) bore mutations in the *resE4* gene, strongly suggesting that the encoded orphan sensor histidine kinase is implicated

in the manifestation of a haemolytic phenotype on agar; interestingly, *resE4* is located immediately downstream of one of the two putative accessory gene regulator (*agr*) loci in *C. sporogenes*[434]. The Agr system is considered the classical quorum-sensing mechanism in Gram-positive organisms, positively regulating in a cell density-dependent manner a plethora of genes, including toxins and degradative exoenzymes[435]. Accordingly, the *agr* locus upstream of *resE4* (*agrBD2*), has been shown to regulate neurotoxin production in *C. botulinum* and, to a lesser extent, sporulation[434]. Although *resE4* was not experimentally proven to be functionally linked with *agrBD2*, the authors speculate that it may play a role in sensing the autoinducing peptide (AgrD2). If this holds true, it can be logically assumed that any genes involved in haemolysis on solid media are under the control of the Agr quorum sensing system and particularly *agrBD2*. This is also in line with the appearance of haemolytic zones after an extended time of incubation (48 hr), when usually overnight incubation suffices for detectable SLS-mediated haemolysis[436].

The question that arises next concerns the identity of these genes. To this end, the superior haemolytic activity observed on horse *vs* sheep blood agar (section 6.46.4.1) provides a tantalising clue; such a preference is uncommon for SLS-mediated haemolysis[437], but could indicate that the protein(s) responsible has phospholipase-like activity. Phospholipases represent a diverse family of esterases capable of hydrolysing membrane glycerophospholipids and many of them are known to possess haemolytic activity[438]. Distinctively, the degree of susceptibility of erythrocytes of different mammalian origins towards a certain phospholipase is indicative of the phospholipid composition of the outer leaflet of their membrane and the substrate preference of the phospholipase in question[439]. Horse and sheep erythrocytes differ significantly in their relative phosphatidylcholine (PC) and sphingomyelin (SM) contents, with horse erythrocytes being richer in PC and sheep erythrocytes in SM, correlating with their different susceptibilities towards a putative phospholipase[440-442]. By extrapolation, it is possible that a similar protein is responsible for the residual haemolytic activity observed in liquid media, after deletion of the CLS operon and all putative haemolysins.

Indeed, *C. sporogenes* encodes a putative Zn-dependent phospholipase C (PLC) (c27100), two patatin-like (PLP)/cytosolic phospholipase 2 (cPLA2)-related proteins (c04730, c15170) and a lysophospholipase L1 (c27530). Importantly, the encoded PLC is 35% identical and 45% similar with the N-terminal (PLC) domain of the alpha toxin from *C. perfringens,* which is essential for the toxin's haemolytic activity[439]. Additionally, there is evidence to support that the *agr* system is implicated in the control of *plc* gene expression in *C. perfingens*[443, 444].

It should be noted that the above points could also provide an explanation for the contrasting phenotypes of RW80 in solid (-) and liquid (++++) media. Given that the only strains with reduced haemolytic activity in liquid had the CLS operon deleted or inactivated (Csp-Δ*closA-I*, Csp-Δ*closA*-IΔ1-5, Csp-*closC*749a::CT-NH) (Table 6.3), it can be assumed that CLS is actively expressed in RW80. Hence, RW80 is fully haemolytic in liquid because of CLS, but not haemolytic on agar as a result of the inactivated *resE4* (which conceivably controls a putative phospholipase) and concomitantly the lack of a contribution from CLS in the haemolytic phenotype on agar, as previously discussed.

Nevertheless, simultaneously with an *agr*-mediated response, haemolysis may also be iron-regulated. It is possible that blood components, such as transferrin, ferritin or lactoferrin, are actively sequestering iron from the media, hence depleting the available iron source for bacterial growth[445]. Iron is a crucial micronutrient and therefore bacteria have developed mechanisms that allow them to fulfil their nutritional requirements when it is not readily accessible; amongst these is the upregulation of genes encoding degradative/haemolytic enzymes (for example phospholipases), for the purpose of gaining access to intracellular sources of iron such as haemoglobin[440, 445]. A regulatory role of iron in the haemolytic phenotype is supported by the lack of haemolysis on RBC-only plates (section 6.4.7), in which iron-sequestering components have been presumably washed off. Furthermore, considering that any potential iron depletion is amplified in solid media (given that nutrient accessibility is restricted on a local scale), it fits that

the difference between WB and RBC-only media is more pronounced on agar. In this context, the similarity in zones of clearance between the WT and Csp-Δ*closA-I* strains may also be justified, given that on horse blood agar WT probably procures iron via CLS-mediated haemolysis[419, 446] while Csp-Δ*closA-I* via iron acquisition systems, with no observable difference in the overall haemolytic phenotype.

Certainly, based on the existing data, the hypothesis of an *agr*- and/or iron- controlled phospholipase (or other degradative enzyme) being responsible for haemolysis on solid (and liquid) media is highly speculative and would have to be explored experimentally. Perhaps transposon-mediated random mutagenesis[74] or transcriptomic analysis in the WT and *resE4*$^{-}$ (RW80) strains would be appropriate methods for unveiling responsible genes in a high-throughput manner.

The probability of a gene in the CLS operon being a transcriptional repressor of a different haemolysin should also be contemplated. In this tenet, upon deletion of the operon, the gene of the second haemolysin would be expressed constitutively while its protein product would generate WT-like haemolytic zones on agar (++++), and significantly reduced activity in liquid (++), presumably because it is a less potent haemolysin than CLS (again, with no evidence of dose-dependency on solid media). If this is true, the regulator would have to be ClosC given that deletion of the whole operon and ClosTron-mediated inactivation of *closC* (in Csp-*closC*749a::CT 2 and 3) resulted in similar phenotypes. However, this is highly unlikely, because SagC (ClosC) is a well characterised protein and there is no literature to support a regulatory function.

Given the similarity of the NCIMB 10696 and ATCC 15579 strains, it is bizarre that the haemolytic phenotype of ATCC 15579 on agar was ablated by single inactivation of *closA* or *closC*, assuming that no random mutations had been acquired[94]. In fact, ATCC 15579, in addition to CLS, encodes 4/5 putative haemolysins detected in *C. sporogenes* NCIMB 10696 (all except c01750) as well as 4/4 putative phospholipases (genome in contigs; GenBank accession: GCA_000155085.1). However, ATCC 15579 has a third putative *agr* locus[434], indicating that despite homology at the nucleotide

level, regulation of gene expression in the two strains might differ (also providing supporting evidence that *agr* might be implicated in haemolysis). It should be noted that the authors did not measure haemolytic activity of the mutant strains in liquid and, therefore, no further conclusions can be drawn.

Independently of the (complex) genetic background, the existing data refute the original hypothesis and pose a challenge with regards to the intended goal, which was abolishment of the haemolytic activity of *C. sporogenes* NCIMB 10696. In view of the results obtained with *C. acetobutylicum* as well as the gradual increase in haemolytic activity with incubation time, it is questionable whether any residual activity (particularly in liquid) is the result of true haemolysis, mediated by individual proteins, or whether it is the collective outcome of co-culturing bacteria with erythrocytes; certainly, the likelihood of the latter being relevant in an *in vivo* context is very small. It is plausible that erythrocyte lysis is caused by pH drifts during culture (particularly relevant with *C. acetobutylicum*[447]) or by secondary bacterial products that are released upon cell lysis and accumulate thereafter. In view of such a possibility, measurement of haemolytic activity using the cell-bound method described in section 6.3.1 might be more appropriate.

6.6 Key outcomes

- Deletion of the CLS operon in *C. sporogenes* NCIMB 10696 failed to ablate the haemolytic activity of the strain; haemolysis was greatly reduced in liquid medium but remained unaltered on solid medium.
- Putative additional haemolysins are not responsible for the residual haemolytic phenotype of the CLS-deficient strain.
- A better understanding of underlying genetic mechanisms that might be involved in haemolysis was obtained.

CHAPTER 7

General discussion

7.1 Summary of findings and future work

This thesis sought to deliver improvements in efficacy (CDEPT, immunotherapy) as well as safety (imaging, attenuation) aspects of the clostridial vector, required to expedite its preclinical and clinical development.

7.1.1 CDEPT: exploration of suitable nitroreducatase candidates

Building on previous work originating from a long-standing collaboration, a library of nitroreductase-producing *C. sporogenes* strains was generated, expressing enzymes capable of co-metabolising a prodrug (CP-506) and PET probes (5NI-EF5 and 5NI-EF1) of interest. Relevant phenotypic characteristics of the parent strain were not affected by the engineering activities undertaken and expression of functional nitroreductases was confirmed for all candidates. To our knowledge, this is the first time that *C. sporogenes* is engineered to express nitroreductases that could realistically fulfil the tripartite (PCE, prodrug/imaging agent, clostridial host) requirement of

CDEPT in the clinic, utilising a prodrug and imaging agents of established clinical relevance and safety (to date).

The strains were engineered to be suitable for a prospective implementation in preclinical and clinical settings, harbouring features that would permit their safe use, in line with regulatory guidance. These include the stable integration of the gene of interest in the genome without the use of antibiotic markers, the strains' dependence on exogenous uracil which is abundant only in tumours, and finally, their increased sensitivity to antibiotics, due to enhanced antibiotic metabolism by the incorporated nitroreductase.

Significantly, YfkO_Bl, although not a good match for CP-506 and the chosen 5-NI imaging agents, emerged as a promising nitroreductase candidate in terms of its ability to attain high expression levels in *C. sporogenes* NCIMB 10696. This prompted the search for new prodrugs, leading to the identification of a promising lead compound – TH-2672 (Threshold Pharmaceuticals). Analogues of TH-2672 are currently being generated, with improved safety and pharmacokinetic profiles as well as PET capabilities (unpublished data; A. Patterson – UOA), providing an alternative option to CP-506.

Consequently, the work presented in Chapter 3 illuminated two dominant CDEPT strategies; these should be pursued in parallel, in order to maximise the chances of preclinical/clinical success:

- NfrA_Bs-expressing strain + CP-506, offering the benefits of an already clinically established prodrug but hampered by low expression of the PCE in the clostridial host, and,
- YfkO_Bl-expressing strain + YfkO_Bl prodrug (TH-2672 or analogues thereof), benefitting from high expression levels of the PCE without laborious engineering, but requiring significant time and resources to be dedicated in prodrug development.

To this end, overexpression strategies such as those explored in Chapter 4 may be of relevance in the case of NfrA_Bs, while structural knowledge derived from the elucidated crystal structure of YfkO_Bl may facilitate the rational design of suitable prodrug analogues.

In the future, nitroreductase activity in clostridial cell lysates should also be assessed towards the prodrug(s) of interest (CP-506 or other). This is because enzymatic screens employing different substrates (such as menadione and nitrofurazone) were found to be misleading, complicating the task of narrowing down the nitroreductase candidates. Additionally, nitroreductase-expressing strains should be evaluated in anaerobic cell cultures (as opposed to aerobic cell lysates), representing more accurately the *in vivo* therapeutic context.

7.1.2 New strategies for improved gene expression

To fulfil the requirement for high expression of the therapeutic of interest, anticipated to be necessary for anticancer efficacy, two different overexpression strategies were devised and implemented in *C. sporogenes* NCIMB 10696. These improved upon earlier work on the fine-tuning of transcription and translation processes and circumvented the use of extrachromosomal elements.

Multicopy chromosomal integration of the gene of interest appeared to hold promise for overexpression of genes chromosomally, managing to elevate significantly the yield of an already well-expressed protein. As such, it may represent a viable solution for improving the expression of heterologous genes when other approaches are not adequate, for instance, in the case of the NfrA_Bs nitroreductase candidate, as discussed previously. In the present study, the optimal gene copy number for *yfko_Bl* under the control of P_{thl14} was 3 copies, while addition of a 4^{th} copy had a negative impact on the growth rate and protein yield.

The utilisation of the orthogonal TcdR-P_{tcdB} system from *C. difficile* did not yield the anticipated expression levels when implemented chromosomally, but characterisation data obtained from its implementation in three different configurations and with two different promoters (inducible and constitutive), helped establish it as a useful expression tool in other contexts. For instance, implementation of the system in the split configuration (with the *tcdR* module located on the chromosome and the P*tcdB*-driven gene copy located on a multicopy plasmid), resulted in impressively high expression levels, with no evidence of imposed burden. Importantly, the system was exemplified in this configuration for the investigation of mGM-CSF expression in *C. sporogenes*, as described in Chapter 5, yielding high quantities of secreted cytokine.

Overall, the findings presented in Chapter 4 institute higher flexibility for the operator, thanks to the increased availability of genomic loci, the in-depth exemplification of the CRISPR/Cas9 tool for various genomic alterations as well as the improved understanding of recombinant protein expression in *C. sporogenes* NCIMB 10696. As a result of these improvements, more ambitious genetic engineering activities can be undertaken in the future, discussed in more detail in section 7.3.1.

For the interested researcher, both strategies studied in Chapter 4 lend themselves well to further investigation and fine-tuning. By way of illustration, future work could include the characterisation of additional genomic loci, the expansion of the TcdR-P_{tcdB} system, and the design of an autonomous inducible system that can afford high expression levels whilst minimising cellular burden.

7.1.3 Piloting a novel functionality: immunotherapy

C. sporogenes NCIMB 10696 was successfully engineered for the first time in our group for the incorporation of an immunotherapeutic functionality, establishing that the strain can tolerate, express and secrete high levels of mGM-CSF, despite its proteolytic nature. To our knowledge, this is the first time that GM-CSF of any eukaryotic origin is expressed in a bacterial species other than *E. coli*; crucially, the protein was detectable

via ELISA, implying that mGM-CSF is expressed and secreted in soluble form and correctly folded.

Preliminary studies also established that *C. sporogenes*-derived mGM-CSF is biologically active, able to bring about the proliferation and differentiation of murine BM cells. Furthermore, based on the current findings, it appears that components of the clostridial supernatant may be exerting an independent negative impact on cell proliferation and expression of the CD11c DC surface marker. At the time of writing, it is imperative that these results are validated by additional experimental repeats as well as by supplementary investigations, including colony-forming assays, gene expression profiling and immunosecretome analysis following cell treatment. Nevertheless, the above outcomes in their totality encourage the incorporation of alternative cytokines or other immunotherapeutic molecules in the future.

In parallel to the establishment of *C. sporogenes* as a suitable host for the expression of immunotherapeutic proteins, the use of new signal peptides was exemplified. Npr and Lip yielded significantly higher secretory efficiencies than the signal peptide commonly used until now (Egla), expanding the range of signal sequences that can be employed in the future, whenever secretion of the protein of interest is required. The signal peptide library could be further improved by characterising additional signal peptides of clostridial origin and/or by optimising their cleavage sites. Alternatively, signals that localise proteins of interest on the cell wall, such as sortases, could also be investigated[448]. Anchoring of the protein of interest on the cell wall may be suitable for immunotherapeutics that need to act only at a local level, simultaneously providing partial protection against proteolytic cleavage by clostridial or tumour proteases.

In the present study, it was not possible to demonstrate expression and secretion of mIL-4 in *C. sporogenes*, due to unforeseen toxic effects of the cytokine, plausibly attributed to inefficient secretion. This may be circumvented by pairing the secretion of mIL-4 to alternative signal peptides.

7.1.4 Towards a safer, non-haemolytic clostridial host

Deletion of the CLS operon in *C. sporogenes* NCIMB 10696 was sought, to remove a potentially pathogenic train of the strain and improve its safety profile, ahead of clinical studies. However, contrary to what was anticipated, elimination of the entire CLS biosynthetic gene cluster failed to ablate the haemolytic activity of the strain entirely. Subsequent deletion of all identified putative haemolytic genes also failed to abolish the haemolytic activity of the strain. While the objective of creating a strain devoid of haemolytic activity was not accomplished, a serendipitous mutation in the *resE4* histidine kinase led to a better understanding of underlying mechanisms that might be involved in haemolysis. In particular, it is plausible that expression of genes with haemolytic activity (other than those of the CLS cluster) is under the control of the Agr quorum sensing mechanism. The identity of those genes is yet unknown, but it is speculated that the protein(s) involved might be phospholipase(s).

Further steps should await *in vivo* evaluation of the CLS-deficient strain, to establish: (i) whether the residual haemolytic activity of the strain constitutes a pathogenic trait, and (ii) whether the colonising/oncolytic capabilities of the strain have been compromised by removal of the CLS gene cluster. Provided that the CLS-deficient strain exhibits an improved safety profile *vs* the WT strain and that relevant capabilities of the strain persist, its downgrading to a Group 1 (or equivalent) organism could be solicited, where applicable, in dialogue with the relevant regulation authorities. In turn, the reclassified strain should be established as the parent strain of engineered *C. sporogenes* NCIMB 10696 strains thereof. Alternatively, if complete abolishment of the strain's haemolytic activity is pursued, identification and directed knock-out of responsible genes will be necessary. To identify such genes, a high-throughput approach should be employed, such as transposon mutagenesis.

7.2 Imminent challenges

7.2.1 Rigorous preclinical evaluation

In view of the regulatory hurdles faced by bacterial cancer therapies, owing to their novelty and inherent toxicity risks, the generation of robust safety and efficacy preclinical data is a prerequisite for their translation into the clinic. In particular, preclinical investigations will be insightful towards aspects that are not currently well-understood, as discussed in section 1.7. Briefly, these regard potential adverse effects associated with the therapies, the establishment of administered/effective doses, appropriate patient selection, the combination of clostridial cancer therapies with conventional treatment modalities, and the risk of genetic instability in the tumour.

In parallel, preclinical data are urgently required to inform genetic engineering endeavours and optimisations. Currently, the inability to draw correlations between the production capacity of a strain *in vitro* and its efficacy *in vivo* implies that the true relevance of improvements/modifications on the genetic level cannot be appreciated. For example, as discussed in Chapter 3, the expression of the top candidate NfrA_Bs was considered disappointing relative to the rest of the enzymes; however, in an *in vivo* context, it is possible that the given quantity of NfrA_Bs will be enough for a clinical response. Also, as commented in Chapter 4, when high expression of a desired product is sought, one needs to be aware of the intended target expression levels. Nevertheless, given that therapeutic efficacy will probably depend on several factors (for example, cancer type, extent of hypoxia, spore dose, tumour targeting efficiency, tumour colonisation efficiency, bacterial replication), different parameters should be carefully evaluated to provide the required level of guidance.

A considerable obstacle hitherto has been the inability to achieve reproducible clostridial colonisation in animal tumours[26, 47, 92]. As a result, preclinical studies should give emphasis to the establishment of suitable tumour models and to the in-depth

characterisation of parameters that are involved in this process. In fact, the colonisation process is a 'black box', with, until now, very little insight into the events that occur intratumorally; for example, details of whether clostridial cells go through cycles of sporulation and germination or how fast they divide, remain unknown. Assessment of a variety of tumour models can contribute towards a rational and hierarchical paradigm, useful for patient selection in the future[92]. To this end, the ability to image the clostridial vector *in vivo* will be instrumental in providing relevant information as well as to verify safety.

Additionally, it is necessary to reconsider the suitability of animal models that have been typically employed to date, given that preceding studies (mainly in the context of CDEPT) were predominantly carried out in immunodeficient mice (for example, in the NMRI nu/nu strain)[43, 49]. In view of the incorporated immunotherapeutic functionality, but also the native immunostimulatory benefits of the clostridial host, evaluation should be carried out in immune-competent mice (or companion animals). These experiments will help elucidate the anticipated role of the immune response as well as rule out selective elimination of clostridial vegetative cells.

7.2.2 Quality control

As with all clinical products intended for use in humans, information relating to the sterility, purity, identity, potency and stability of bacterial cancer therapies is of relevance, all together feeding into the chemistry, manufacturing and control (CMC) regulatory package of a given product[449]. Nevertheless, since bacterial cancer therapies represent a new category of 'live therapeutics' and differ significantly from conventional drugs, commonly used procedures of asserting the above characteristics may not be applicable in their case.

For instance, unlike conventional therapeutic molecules, clostridial spores cannot be sterilised by traditional methods, such as autoclaving or filtering. High temperature autoclaving would render the spores non-viable, while contaminating particles in the

same size range as spores would remain even after filtration. This entails that spore production for clinical use should be subject to good manufacturing practice (GMP) standards, requiring the use of dedicated clean rooms and strict aseptic protocols to ensure that resulting products are free from pathogenic agents[47, 92]. In this context, spore production protocols should be optimised to enable large-scale production of clostridial spores to high purity, and at an affordable cost.

Likewise, instability of the final vector at the genetic level poses an 'unfamiliar' problem; genetic mutations could alter the predicted therapeutic function of the product and/or create a harmful phenotype[20]. Therefore, robustness of engineered strains should be thoroughly investigated prior to commercialisation, for instance, by challenging the strains under different conditions and tumour environments. Tests should be done over prolonged periods of times as well as periodically (to confirm appropriateness of storage conditions and 'shelf life'); whole genome sequencing and appropriate activity/functionality assays might be useful ways of verifying strain stability.

7.2.3 Biocontainment

The present thesis has highlighted several safeguards, inherent or engineered, that are expected to facilitate the clinical implementation of clostridial vectors in the current regulatory framework. These include: a) the natural predilection of clostridial spore germination for hypoxic/necrotic tumour regions; b) the inertness of therapeutic cargos in spore form, thus preventing off-target effects; c) the attenuation of the strain by uracil auxotrophy and by interfering with its ability to acquire or metabolise nutrients outside of the tumour, functioning as a precaution in the case of shedding (Chapters 3 and 4); d) the elimination of the potent CLS toxin (Chapter 6); e) the ability to stably integrate therapeutic cargos in the genome (Chapters 3 and 4); f) the absence of mobile genetic elements or non-natural antibiotic resistance markers (Chapters 3 and 4); g) the ability of non-invasive imaging of the clostridial vector *in vivo*, and; h) the increased

susceptibility to antibiotics, the last two points being conditional to the incorporation of a suitable nitroreductase (Chapter 3).

Although these features represent an extensive network of *inbuilt* safety precautions, the use of clostridial therapies warrants the implementation of yet more sophisticated biocontainment strategies. While the sporulating ability of clostridia is fundamental to their safe and practical application in cancer therapy, it is also, paradoxically, a cause of concern. In theory, the inherent resilience of spores would allow GMOs to survive indefinitely in the environment[450] (for example, following shedding from human patients); this poses an uncertain risk that should, at least, be managed. Although several bio-strategies have been developed that can limit the spread of *vegetative* cells (for instance, controlled expression of essential genes, auxotrophy and kill switches)[451], none of them has specifically addressed the containment of recombinant spores. A solution germane to this problem would be the engineering of a strain which is conditional for sporulation – that is, a strain that can only sporulate when a given inducer is supplied, in a laboratory or industrial setting. Such a strain would be strictly confined to the tumour: following intratumoural spore germination, clostridial vegetative cells would not be able to sporulate again and would hence die upon encountering normoxic conditions outside of the tumour. Work on generating a conditionally sporulating strain is currently in progress (unpublished data; Ines Cañadas Blasco). The likely availability of such a strain is expected to greatly facilitate the clinical implementation of *C. sporogenes*–based cancer therapies.

7.3 Future perspectives

7.3.1 Towards a multifunctional clostridial anticancer vector

At heart, *C. sporogenes* has been envisioned as a multimodal delivery platform, that can simultaneously convey more than one functionality. Considering the protean nature of cancer, multi-functionalisation of the clostridial vector confers it an advantage,

enabling the co-targeting of mechanisms responsible for cancer growth and/or treatment resistance, possibly also alleviating the drug burden of cancer patients[452]. In particular, new evidence suggests chemotherapy and immunotherapy may act synergistically, contrary to the conventional view that perceives the two modalities as incompatible (due to chemotherapy-induced myelosuppression)[278]. The biological basis of their synergy appears to be the ability of chemotherapy to occasionally induce immunological *vs* programmed cancer cell death (namely death that releases tumour antigens) or its ability to interfere directly (molecularly) with the tumour's immunoevasive propensity[453]. Current understanding is that the partial immunological effect of classical chemotherapy can be enforced/harnessed by its combination with additional immune stimulation, provided certain conditions are fulfilled, such as the spatial and temporal co-existence of TLR ligands and antigens[454]. In this context, the coordinated delivery of both treatment modalities via the clostridial vector will be particularly advantageous, complemented by the immunostimulatory and oncolytic effects that the vector exerts naturally.

Intriguingly, a combination of the separate outcomes of this work could pave the way for the realisation of a multifunctional clostridial vector. The demonstration of the capability of *C. sporogenes* to express functional nitroreductases and mGM-CSF as well as the generation of a strain with foreign DNA in multiple genomic locations, imply that *C. sporogenes* can now be feasibly engineered to deliver multiple therapeutic genes. For instance, both the CDEPT (chemotherapeutic) and immunotherapeutic functionalities could be conveyed simultaneously. Adding to this, the availability of a PET-capable probe metabolisable by the nitroreductase, would automatically enable an imaging capability. Crucially, owing to the improved understanding on recombinant protein expression and secretion in this host, production of the therapeutic cargos can be optimised in a logical and clinically relevant way.

7.3.2 Expanding the functionality repertoire

Certainly, the precise therapeutic molecules incorporated in the clostridial vector could be diversified in the future. On the basis of work in this thesis and previous studies, it appears that *C. sporogenes* is potentially capable of tolerating and expressing/secreting a variety of proteins, thus, in principle, it could be engineered to deliver *any* desired anticancer agent and turned into a platform technology. In the era of precision and personalised medicine[455] enabled by the plummeting costs of DNA sequencing[10], the clostridial vector can be further envisioned as a *personalised* platform technology. The cell can be rapidly customised to produce therapeutics on demand, whereby different strains are generated for different patients, in the paradigm set by personalised licensed cancer vaccines and engineered chimeric antigen receptors (CAR) T cell therapies (Oncophage®, sipuleucel-T, Kymriah™, Yescarta™)[310, 456, 457].

Specifically, it would be of interest to engineer *C. sporogenes* to express proteins (antibodies or small peptides) that could mediate immune checkpoint blockade. Immune checkpoint blockade refers to the inhibition of pathways implicated in the downregulation of T-cell responses. Thanks to encouraging clinical outcomes, predominantly in the treatment of advanced melanoma, immune checkpoint inhibition has attracted significant attention as a therapeutic strategy[458]. Marketed immune checkpoint inhibitors include the monoclonal antibodies ipilimumab (anti-CTLA-4; cytotoxic T-lymphocyte-associated protein 4), nivolumab and pembrolizumab (anti-PD-1; programmed cell death protein 1)[459]. Engineering of the clostridial vector to express such inhibitors is anticipated to enhance the effectiveness of both modalities: on the one hand, checkpoint blockade inhibition would exert a powerful co-immunostimulatory effect, additional to that exerted by the clostridial vector; on the other hand, clostridial presence will assist in sensitising the tumour towards immune checkpoint blockade, by provision of tumour-associated neoantigens (following oncolysis) and stimulation of immune cell infiltration into the tumour (a consequence of the localised bacterial infection)[92]. Crucially, local production of immune checkpoint inhibitors could

circumvent adverse effects associated with their systemic administration, such as diarrhoea, enterocolitis and intestinal perforation[460].

Finally, it would appear wise to utilise the clostridial vector to target pathways relevant to the microenvironment of a hypoxic tumour. Therefore, *C. sporogenes* could be engineered to express protein molecules that can mediate VEGF inhibition, Hypoxia Inducible Factor 1A (HIF1A) blockade, or mitochondrial shutdown[461]. Initially, incorporation of agents which are already in clinical trials or clinical use may be preferable.

Figure 7.1 shows a schematic diagram of the different functionalities/properties that the envisioned clostridial anticancer vector could convey in the future.

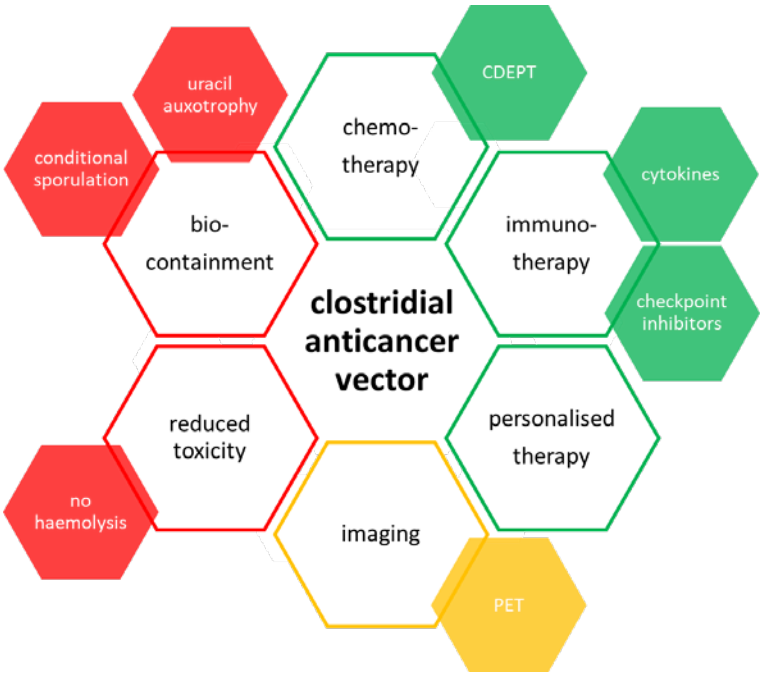


Figure 7.1 A multifunctional clostridial vector engineered for cancer therapy. The clostridial vector could be engineered to convey various therapeutic functionalities (in green), such as chemotherapy and immunotherapy; additionally, it could be tailored ('personalised') for the treatment of a specific tumour in a specific patient. The vector could be further engineered to enable *in vivo* imaging applications (in yellow). Finally, safety features (in red) could also be incorporated. The emergent ability to manipulate the strain's genome easily and at will, is central to the multi-functionalisation of the envisioned vector.

7.4 Concluding remarks

The recent licensing of genetically modified viruses for human gene therapy in advanced melanoma (IMLYGIC®) and inherited retinal disease (LUXTURNA™) heralds the arrival of an opportune time for microbial gene therapies. Momentously, this is also mirrored in the recent proposal on behalf of the National Institutes of Health (NIH) and the FDA to streamline the review of investigational gene therapy products, by eliminating additional controls imposed formerly, since "there is no longer sufficient evidence to claim that the risks of gene therapy are entirely unique and unpredictable"[462].

In parallel, new funding opportunities on behalf of the NIH for 'Microbial-based Cancer Therapy - Bugs as Drugs', instil further enthusiasm in the development of bacterial/clostridial-based cancer therapies (funding opportunity announcements PAR-19-193 and PAR-19-194). As outlined in the announcement, the funding aims "to stimulate exploratory development of novel microbial-based cancer therapies, especially for conditions where conventional cancer therapies are inadequate, such as poorly vascularized, hypoxic, solid tumors, dormant or slowly dividing cells resistant to current interventions, and brain tumors".

In this changing environment, it is realistic to anticipate that sophisticated clostridial cancer therapies will indeed find their place in the repertoire of anticancer treatments, an advent expected to be facilitated by work presented in this thesis.

Bibliography

1. Song, Q., Merajver, S.D., and Li, J.Z., *Cancer classification in the genomic era: five contemporary problems.* Hum Genomics, 2015. **9**: p. 27.
2. Hanahan, D. and Weinberg, R.A., *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
3. Stratton, M.R., Campbell, P.J., and Futreal, P.A., *The cancer genome.* Nature, 2009. **458**: p. 719.
4. Garber, J.E. and Offit, K., *Hereditary cancer predisposition syndromes.* J Clin Oncol, 2005. **23**(2): p. 276-92.
5. Hanahan, D. and Weinberg, Robert A., *Hallmarks of Cancer: The Next Generation.* Cell, 2011. **144**(5): p. 646-674.
6. World Health Organization (WHO), *World Cancer Report 2014*, Bernard W. Stewart and C.P. Wild, Editors. 2014.
7. Vineis, P. and Wild, C.P., *Global cancer patterns: causes and prevention.* Lancet, 2014. **383**(9916): p. 549-557.
8. Van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., and Friend, S.H., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. **415**(6871): p. 530-536.
9. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., and Lander, E.S., *Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring.* Science, 1999. **286**(5439): p. 531-537.
10. McDermott, U., *Next-generation sequencing and empowering personalised cancer medicine.* Drug Discovery Today, 2015. **20**(12): p. 1470-1475.
11. Visvader, J.E. and Lindeman, G.J., *Cancer stem cells in solid tumours: accumulating evidence and unresolved questions.* Nat Rev Cancer, 2008. **8**(10): p. 755-68.
12. Weis, S.M. and Cheresh, D.A., *Tumor angiogenesis: molecular pathways and therapeutic targets.* Nature Medicine, 2011. **17**: p. 1359.
13. Nagy, J.A., Chang, S.-H., Shih, S.-C., Dvorak, A.M., and Dvorak, H.F., *Heterogeneity of the Tumor Vasculature.* Seminars in Thrombosis and Hemostasis, 2010. **36**(3): p. 321-331.
14. Vaupel, P. and Mayer, A., *Hypoxia in cancer: significance and impact on clinical outcome.* Cancer Metastasis Rev, 2007. **26**(2): p. 225-239.
15. Bertout, J.A., Patel, S.A., and Simon, M.C., *The impact of O2 availability on human cancer.* Nat Rev Cancer, 2008. **8**(12): p. 967-75.

16. Dang, L.H., Bettegowda, C., Huso, D.L., Kinzler, K.W., and Vogelstein, B., *Combination bacteriolytic therapy for the treatment of experimental tumors.* Proc Natl Acad Sci U S A, 2001. **98**(26): p. 15155-15160.
17. Cronin, M., Akin, A.R., Collins, S.A., Meganck, J., Kim, J.-B., Baban, C.K., Joyce, S.A., van Dam, G.M., Zhang, N., van Sinderen, D., O'Sullivan, G.C., Kasahara, N., Gahan, C.G., Francis, K.P., and Tangney, M., *High Resolution In Vivo Bioluminescent Imaging for the Study of Bacterial Tumour Targeting.* PloS one, 2012. **7**(1): p. e30940.
18. Richards, C.H., Mohammed, Z., Qayyum, T., Horgan, P.G., and McMillan, D.C., *The prognostic value of histological tumor necrosis in solid organ malignant disease: a systematic review.* Future Oncol, 2011. **7**(10): p. 1223-35.
19. Rockwell, S., Dobrucki, I.T., Kim, E.Y., Marrison, S.T., and Vu, V.T., *Hypoxia and radiation therapy: Past history, ongoing research, and future promise.* Current molecular medicine, 2009. **9**(4): p. 442-458.
20. Forbes, N.S., *Engineering the perfect (bacterial) cancer therapy.* Nat Rev Cancer, 2010. **10**(11): p. 785-794.
21. Wilson, W.R. and Hay, M.P., *Targeting hypoxia in cancer therapy.* Nat Rev Cancer, 2011. **11**(6): p. 393-410.
22. Cosse, J.P. and Michiels, C., *Tumour hypoxia affects the responsiveness of cancer cells to chemotherapy and promotes cancer progression.* Anticancer Agents Med Chem, 2008. **8**(7): p. 790-7.
23. Staedtke, V., Bai, R.-Y., Sun, W., Huang, J., Kibler, K.K., Tyler, B.M., Gallia, G.L., Kinzler, K., Vogelstein, B., Zhou, S., and Riggins, G.J., *Clostridium novyi-NT can cause regression of orthotopically implanted glioblastomas in rats.* Oncotarget, 2015. **6**(8): p. 5536-46.
24. Theys, J., Landuyt, W., Nuyts, S., Van Mellaert, L., Bosmans, E., Rijnders, A., Van Den Bogaert, W., van Oosterom, A., Anne, J., and Lambin, P., *Improvement of Clostridium tumour targeting vectors evaluated in rat rhabdomyosarcomas.* FEMS Immunol Med Microbiol, 2001. **30**(1): p. 37-41.
25. Liu, S.C., Minton, N.P., Giaccia, A.J., and Brown, J.M., *Anticancer efficacy of systemically delivered anaerobic bacteria as gene therapy vectors targeting tumor hypoxia/necrosis.* Gene Ther, 2002. **9**(4): p. 291-6.
26. Mowday, A., Guise, C., Ackerley, D., Minton, N., Lambin, P., Dubois, L., Theys, J., Smaill, J., and Patterson, A., *Advancing Clostridia to Clinical Trial: Past Lessons and Recent Progress.* Cancers, 2016. **8**(7): p. 63.
27. Kubiak, A.M., Poehlein, A., Budd, P., Kuehne, S.A., Winzer, K., Theys, J., Lambin, P., Daniel, R., and Minton, N.P., *Complete Genome Sequence of the Nonpathogenic Soil-Dwelling Bacterium Clostridium sporogenes Strain NCIMB 10696.* Genome Announc, 2015. **3**(4).
28. Mengesha, A., Dubois, L., Paesmans, K., Wouters, B., Lambin, P., and and Theys, J., *Chapter 11: Clostridia in Anti-tumour Therapy*, in *Clostridia: Molecular Biology in the Post-genomic Era*, H. Brüggemann and G. Gottschalk, Editors. 2009, Caister Academic Press. p. 199-214.

29. Torrey, J.C. and Kahn, M.C., *The Treatment of Flexner-Jobling Rat Carcinomas with Bacterial Proteolytic Ferments.* Am J Cancer Res, 1927. **11**(4): p. 334-376.
30. Connell, H.C., *The Study and Treatment of Cancer by Proteolytic Enzymes: Preliminary Report.* CMAJ, 1935. **33**(4): p. 364-370.
31. Parker, R.C., Plummer, H.C., Siebenmann, C.O., and Chapman, M.G., *Effect of Histolyticus Infection and Toxin on Transplantable Mouse Tumors.* Exp Biol Med, 1947. **66**(2): p. 461-467.
32. Malmgren, R.A. and Flanigan, C.C., *Localization of the Vegetative Form of Clostridium tetani in Mouse Tumors Following Intravenous Spore Administration.* Cancer Res, 1955. **15**(7): p. 473-478.
33. Möse, J.R. and Möse, G., *Oncolysis by Clostridia. I. Activity of Clostridium Butyricum (M-55) and Other Nonpathogenic Clostridia against the Ehrlich Carcinoma.* Cancer Res, 1964. **24**(2 Part 1): p. 212-216.
34. Carey, R.W., Holland, J.F., Whang, H.Y., Neter, E., and Bryant, B., *Clostridial oncolysis in man.* Eur J Cancer, 1967. **3**(1): p. 37-46.
35. Heppner, F. and Möse, J.R., *The liquefaction (oncolysis) of malignant gliomas by a non pathogenic Clostridium.* Acta neurochir, 1978. **42**(1-2): p. 123-125.
36. Pennington, O., *The development of molecular tools for the expression of Prodrug Converting Enzymes in Clostridium sporogenes*. 2006, University of Nottingham.
37. Gericke, D., Dietzel, F., Konig, W., Ruster, I., and Schumacher, L., *Further progress with oncolysis due to apathogenic clostridia.* Zentralbl Bakteriol Orig A, 1979. **243**(1): p. 102-12.
38. Mose, J.R., *Experiments to improve the oncolysis-effect of clostridial-strain M55.* Zentralbl Bakteriol Orig A, 1979. **244**(4): p. 541-5.
39. Schlechte, H. and Elbe, B., *Recombinant plasmid DNA variation of clostridium oncolyticum — model experiments of cancerostatic gene transfer.* Zentralbl Bakteriol Mikrobiol Hyg A, 1988. **268**(3): p. 347-356.
40. Fox, M.E., Lemmon, M.J., Mauchline, M.L., Davis, T.O., Giaccia, A.J., Minton, N.P., and Brown, J.M., *Anaerobic bacteria as a delivery system for cancer gene therapy: in vitro activation of 5-fluorocytosine by genetically engineered clostridia.* Gene Ther, 1996. **3**(2): p. 173-8.
41. Theys, J., Nuyts, S., Landuyt, W., Van Mellaert, L., Dillen, C., Bohringer, M., Durre, P., Lambin, P., and Anne, J., *Stable Escherichia coli-Clostridium acetobutylicum shuttle vector for secretion of murine tumor necrosis factor alpha.* Appl Environ Microbiol, 1999. **65**(10): p. 4295-300.
42. Groot, A.J., Mengesha, A., Wall, E.v.d., Diest, P.J.v., Theys, J., and Vooijs, M., *Functional antibodies produced by oncolytic clostridia.* Biochem Biophys Res Commun, 2007. **364**(4): p. 985-989.
43. Heap, J.T., Theys, J., Ehsaan, M., Kubiak, A.M., Dubois, L., Paesmans, K., Mellaert, L.V., Knox, R., Kuehne, S.A., Lambin, P., and Minton, N.P., *Spores*

of Clostridium engineered for clinical efficacy and safety cause regression and cure of tumors in vivo. Oncotarget, 2014. **5**(7): p. 1761-9.

44. Diaz, L.A., Cheong, I., Foss, C.A., Zhang, X., Peters, B.A., Agrawal, N., Bettegowda, C., Karim, B., Liu, G., Khan, K., Huang, X., Kohli, M., Dang, L.H., Hwang, P., Vogelstein, A., Garrett-Mayer, E., Kobrin, B., Pomper, M., Zhou, S., Kinzler, K.W., Vogelstein, B., and Huso, D.L., *Pharmacologic and Toxicologic Evaluation of C. novyi-NT Spores.* Toxicol Sci, 2005. **88**(2): p. 562-575.
45. Krick, E.L., Sorenmo, K.U., Rankin, S.C., Cheong, I., Kobrin, B., Thornton, K., Kinzler, K.W., Vogelstein, B., Zhou, S., and Diaz, L.A., Jr., *Evaluation of Clostridium novyi-NT spores in dogs with naturally occurring tumors.* Am J Vet Res, 2012. **73**(1): p. 112-8.
46. 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., Jr., Riggins, G.J., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., Bettegowda, C., Huso, D.L., Varterasian, M., Saha, S., and Zhou, S., *Intratumoral injection of Clostridium novyi-NT spores induces antitumor responses.* Sci Transl Med, 2014. **6**(249): p. 249ra111.
47. Forbes, N.S., Coffin, R.S., Deng, L., Evgin, L., Fiering, S., Giacalone, M., Gravekamp, C., Gulley, J.L., Gunn, H., Hoffman, R.M., Kaur, B., Liu, K., Lyerly, H.K., Marciscano, A.E., Moradian, E., Ruppel, S., Saltzman, D.A., Tattersall, P.J., Thorne, S., Vile, R.G., Zhang, H.H., Zhou, S., and McFadden, G., *White paper on microbial anti-cancer therapy and prevention.* Journal for ImmunoTherapy of Cancer, 2018. **6**(1): p. 78.
48. Fabricius, E.M., Schneeweiss, U., Schau, H.P., Schmidt, W., and Benedix, A., *Quantitative investigations into the elimination of in vitro-obtained spores of the non-pathogenic Clostridium butyricum strain CNRZ 528, and their persistence in organs of different species following intravenous spore administration.* Res Microbiol, 1993. **144**(9): p. 741-53.
49. Theys, J., Pennington, O., Dubois, L., Anlezark, G., Vaughan, T., Mengesha, A., Landuyt, W., Anné, J., Burke, P.J., Dûrre, P., Wouters, B.G., Minton, N.P., and Lambin, P., *Repeated cycles of Clostridium-directed enzyme prodrug therapy result in sustained antitumour effects in vivo.* Br J Cancer, 2006. **95**(9): p. 1212-1219.
50. Wei, M.Q., Ellem, K.A., Dunn, P., West, M.J., Bai, C.X., and Vogelstein, B., *Facultative or obligate anaerobic bacteria have the potential for multimodality therapy of solid tumours.* Eur J Cancer, 2007. **43**(3): p. 490-6.
51. Zu, C. and Wang, J.S., *Tumor-colonizing bacteria: A potential tumor targeting therapy.* Crit Rev Microbiol, 2014. **40**(3): p. 225-235.

52. Fang, J., Liao, L., Yin, H., Nakamura, H., Shin, T., and Maeda, H., *Enhanced bacterial tumor delivery by modulating the EPR effect and therapeutic potential of Lactobacillus casei.* J Pharm Sci, 2014. **103**(10): p. 3235-43.
53. Minton, N.P., *Clostridia in cancer therapy.* Nat Rev Microbiol, 2003. **1**(3): p. 237-42.
54. Bettegowda, C., Huang, X., Lin, J., Cheong, I., Kohli, M., Szabo, S.A., Zhang, X., Diaz, L.A., Velculescu, V.E., Parmigiani, G., Kinzler, K.W., Vogelstein, B., and Zhou, S., *The genome and transcriptomes of the anti-tumor agent Clostridium novyi-NT.* Nat Biotechnol, 2006. **24**(12): p. 1573-1580.
55. Agrawal, N., Bettegowda, C., Cheong, I., Geschwind, J.F., Drake, C.G., Hipkiss, E.L., Tatsumi, M., Dang, L.H., Diaz, L.A., Jr., Pomper, M., Abusedera, M., Wahl, R.L., Kinzler, K.W., Zhou, S., Huso, D.L., and Vogelstein, B., *Bacteriolytic therapy can generate a potent immune response against experimental tumors.* Proc Natl Acad Sci U S A, 2004. **101**(42): p. 15172-7.
56. Albert, M.L., *Death-defying immunity: do apoptotic cells influence antigen processing and presentation?* Nat Rev Immunol, 2004. **4**(3): p. 223-231.
57. Bernardes, N., Seruca, R., Chakrabarty, A.M., and Fialho, A.M., *Microbial-based therapy of cancer: Current progress and future prospects.* Bioeng Bugs, 2010. **1**(3): p. 178-190.
58. Bettegowda, C., Dang, L.H., Abrams, R., Huso, D.L., Dillehay, L., Cheong, I., Agrawal, N., Borzillary, S., McCaffery, J.M., Watson, E.L., Lin, K.-S., Bunz, F., Baidoo, K., Pomper, M.G., Kinzler, K.W., Vogelstein, B., and Zhou, S., *Overcoming the hypoxic barrier to radiation therapy with anaerobic bacteria.* Proc Natl Acad Sci U S A, 2003. **100**(25): p. 15083-15088.
59. Cheong, I., Huang, X., Bettegowda, C., Diaz, L.A., Jr., Kinzler, K.W., Zhou, S., and Vogelstein, B., *A bacterial protein enhances the release and efficacy of liposomal cancer drugs.* Science, 2006. **314**(5803): p. 1308-11.
60. Dietzel, F., Gericke, D., and Konig, W., *[Tumor hyperthermia using high frequency for increase of oncolysis by Clostridium butyricum (M 55)].* Strahlentherapie, 1976. **152**(6): p. 537-41.
61. Verma, I.M. and Somia, N., *Gene therapy -- promises, problems and prospects.* Nature, 1997. **389**(6648): p. 239-42.
62. Amer, M.H., *Gene therapy for cancer: present status and future perspective.* Mol Cell Ther, 2014. **2**: p. 27.
63. Wei, M.Q., Ren, R., Good, D., and Anne, J., *Clostridial spores as live 'Trojan horse' vectors for cancer gene therapy: comparison with viral delivery systems.* Genet Vaccines Ther, 2008. **6**: p. 8.
64. Brody, T., *Chapter 4 - Dose Modification and Dose Titration*, in *FDA's Drug Review Process and the Package Label*, T. Brody, Editor. 2018, Academic Press. p. 101-151.
65. Wei, M.Q., Mengesha, A., Good, D., and Anne, J., *Bacterial targeted tumour therapy-dawn of a new era.* Cancer Lett, 2008. **259**(1): p. 16-27.

66. Lehouritis, P., Springer, C., and Tangney, M., *Bacterial-directed enzyme prodrug therapy.* J Control Release, 2013. **170**(1): p. 120-131.
67. Heap, J.T., Pennington, O.J., Cartman, S.T., and Minton, N.P., *A modular system for Clostridium shuttle plasmids.* J Microbiol Methods, 2009. **78**(1): p. 79-85.
68. Heap, J.T., Kuehne, S.A., Ehsaan, M., Cartman, S.T., Cooksley, C.M., Scott, J.C., and Minton, N.P., *The ClosTron: Mutagenesis in Clostridium refined and streamlined.* J Microbiol Methods, 2010. **80**(1): p. 49-55.
69. Cartman, S.T., Kelly, M.L., Heeg, D., Heap, J.T., and Minton, N.P., *Precise manipulation of the Clostridium difficile chromosome reveals a lack of association between the tcdC genotype and toxin production.* Appl Environ Microbiol, 2012. **78**(13): p. 4683-90.
70. Kuehne, S.A., Heap, J.T., Cooksley, C.M., Cartman, S.T., and Minton, N.P., *ClosTron-mediated engineering of Clostridium.* Methods Mol Biol, 2011. **765**: p. 389-407.
71. Ng, Y.K., Ehsaan, M., Philip, S., Collery, M.M., Janoir, C., Collignon, A., Cartman, S.T., and Minton, N.P., *Expanding the repertoire of gene tools for precise manipulation of the Clostridium difficile genome: allelic exchange using pyrE alleles.* PloS one, 2013. **8**(2): p. e56051.
72. Ehsaan, M., Kuit, W., Zhang, Y., Cartman, S.T., Heap, J.T., Winzer, K., and Minton, N.P., *Mutant generation by allelic exchange and genome resequencing of the biobutanol organism Clostridium acetobutylicum ATCC 824.* Biotechnol Biofuels, 2016. **9**(1): p. 4.
73. Cartman, S.T. and Minton, N.P., *A mariner-Based Transposon System for In Vivo Random Mutagenesis of Clostridium difficile.* Appl Environ Microbiol, 2010. **76**(4): p. 1103-1109.
74. Zhang, Y., Grosse-Honebrink, A., and Minton, N.P., *A universal mariner transposon system for forward genetic studies in the genus Clostridium.* PloS one, 2015. **10**(4): p. e0122411.
75. Heap, J.T., Ehsaan, M., Cooksley, C.M., Ng, Y.-K., Cartman, S.T., Winzer, K., and Minton, N.P., *Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker.* Nucleic Acids Res, 2012. **40**(8): p. e59-e59.
76. Kovács, K., Willson, B.J., Schwarz, K., Heap, J.T., Jackson, A., Bolam, D.N., Winzer, K., and Minton, N.P., *Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in Clostridium acetobutylicum ATCC 824.* Biotechnol Biofuels, 2013. **6**(1): p. 117.
77. Wang, Y., Zhang, Z.T., Seo, S.O., Choi, K., Lu, T., Jin, Y.S., and Blaschek, H.P., *Markerless chromosomal gene deletion in Clostridium beijerinckii using CRISPR/Cas9 system.* J Biotechnol, 2015. **200**: p. 1-5.
78. Huang, H., Chai, C., Li, N., Rowe, P., Minton, N.P., Yang, S., Jiang, W., and Gu, Y., *CRISPR/Cas9-Based Efficient Genome Editing in Clostridium*

ljungdahlii, an Autotrophic Gas-Fermenting Bacterium. ACS Synth Biol, 2016. **5**(12): p. 1355-1361.

79. Li, Q., Chen, J., Minton, N.P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X., Yang, J., Gu, Y., Jiang, W., Jiang, Y., and Yang, S., *CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii.* Biotechnol J, 2016. **11**(7): p. 961-72.
80. Minton, N.P., Ehsaan, M., Humphreys, C.M., Little, G.T., Baker, J., Henstra, A.M., Liew, F., Kelly, M.L., Sheng, L., Schwarz, K., and Zhang, Y., *A roadmap for gene system development in Clostridium.* Anaerobe, 2016. **41**: p. 104-112.
81. Minton, N.P., Mauchline, M.L., Lemmon, M.J., Brehm, J.K., Fox, M., Michael, N.P., Giaccia, A., and Brown, J.M., *Chemotherapeutic tumour targeting using clostridial spores.* FEMS microbiology reviews, 1995. **17**(3): p. 357-64.
82. Theys, J., Landuyt, A.W., Nuyts, S., Van Mellaert, L., Lambin, P., and Anne, J., *Clostridium as a tumor-specific delivery system of therapeutic proteins.* Cancer Detect Prev, 2001. **25**(6): p. 548-57.
83. Tirandaz, H., Salehi-Najafabadi, A., and Eslamifar, A., *Cytosine deaminase producing Clostridium may be used in detection of tumors.* Med Hypotheses, 2006. **67**(5): p. 1257-8.
84. Theys, J., Landuyt, W., Nuyts, S., Van Mellaert, L., van Oosterom, A., Lambin, P., and Anne, J., *Specific targeting of cytosine deaminase to solid tumors by engineered Clostridium acetobutylicum.* Cancer Gene Ther, 2001. **8**(4): p. 294-7.
85. Lemmon, M.J., van Zijl, P., Fox, M.E., Mauchline, M.L., Giaccia, A.J., Minton, N.P., and Brown, J.M., *Anaerobic bacteria as a gene delivery system that is controlled by the tumor microenvironment.* Gene Ther, 1997. **4**(8): p. 791-6.
86. Liu, S.-C., Ahn, G.-O., Kioi, M., Dorie, M.-J., Patterson, A.V., and Brown, J.M., *Optimized Clostridium-Directed Enzyme Prodrug Therapy Improves the Antitumor Activity of the Novel DNA Cross-Linking Agent PR-104.* Cancer Res, 2008. **68**(19): p. 7995-8003.
87. Nuyts, S., Van Mellaert, L., Theys, J., Landuyt, W., Bosmans, E., Anne, J., and Lambin, P., *Radio-responsive recA promoter significantly increases TNFalpha production in recombinant clostridia after 2 Gy irradiation.* Gene Ther, 2001. **8**(15): p. 1197-201.
88. Nuyts, S., Van Mellaert, L., Barbe, S., Lammertyn, E., Theys, J., Landuyt, W., Bosmans, E., Lambin, P., and Anne, J., *Insertion or deletion of the Cheo box modifies radiation inducibility of Clostridium promoters.* Appl Environ Microbiol, 2001. **67**(10): p. 4464-70.
89. Barbe, S., Van Mellaert, L., and Anne, J., *Clostridium-mediated transfer system of therapeutic proteins to solid tumors: increased heterologous protein production.* Commun Agric Appl Biol Sci, 2004. **69**(2): p. 49-52.
90. Brockstedt, D.G., Giedlin, M.A., Leong, M.L., Bahjat, K.S., Gao, Y., Luckett, W., Liu, W., Cook, D.N., Portnoy, D.A., and Dubensky, T.W., Jr., *Listeria-*

based cancer vaccines that segregate immunogenicity from toxicity. Proc Natl Acad Sci U S A, 2004. **101**(38): p. 13832-7.

91. Chien, T., Doshi, A., and Danino, T., *Advances in bacterial cancer therapies using synthetic biology.* Curr Opin Syst Biol, 2017. **5**: p. 1-8.
92. Zhou, S., Gravekamp, C., Bermudes, D., and Liu, K., *Tumour-targeting bacteria engineered to fight cancer.* Nat Rev Cancer, 2018. **18**(12): p. 727-743.
93. Jiang, N., Tan, N.S., Ho, B., and Ding, J.L., *Measurement of the red blood cell lysis by bacterial hemolysin.* 2007.
94. Gonzalez, D.J., Lee, S.W., Hensler, M.E., Markley, A.L., Dahesh, S., Mitchell, D.A., Bandeira, N., Nizet, V., Dixon, J.E., and Dorrestein, P.C., *Clostridiolysin S, a post-translationally modified biotoxin from Clostridium botulinum.* J Biol Chem, 2010. **285**(36): p. 28220-8.
95. Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E., *Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation.* Nat Biotechnol, 2014. **32**: p. 1262.
96. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H.W., Listgarten, J., and Root, D.E., *Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.* Nat Biotechnol, 2016. **34**: p. 184.
97. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., and Zhang, F., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**: p. 827.
98. Blin, K., Pedersen, L.E., Weber, T., and Lee, S.Y., *CRISPy-web: An online resource to design sgRNAs for CRISPR applications.* Synthetic and Systems Biotechnology, 2016. **1**(2): p. 118-121.
99. Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P., and Minton, N.P., *The ClosTron: a universal gene knock-out system for the genus Clostridium.* J Microbiol Methods, 2007. **70**(3): p. 452-64.
100. Schwarz, K., Fiedler, T., Fischer, R.-J., and Bahl, H., *A Standard Operating Procedure (SOP) for the preparation of intra- and extracellular proteins of Clostridium acetobutylicum for proteome analysis.* J Microbiol Methods, 2007. **68**(2): p. 396-402.
101. Knox, R.J., Boland, M.P., Friedlos, F., Coles, B., Southan, C., and Roberts, J.J., *The nitroreductase enzyme in Walker cells that activates 5-(aziridin-1-yl)-2,4-dinitrobenzamide (CB 1954) to 5-(aziridin-1-yl)-4-hydroxylamino-2-nitrobenzamide is a form of NAD(P)H dehydrogenase (quinone) (EC 1.6.99.2).* Biochem Pharmacol, 1988. **37**(24): p. 4671-7.
102. Whiteway, J., Koziarz, P., Veall, J., Sandhu, N., Kumar, P., Hoecher, B., and Lambert, I.B., *Oxygen-insensitive nitroreductases: analysis of the roles of nfsA and nfsB in development of resistance to 5-nitrofuran derivatives in Escherichia coli.* J Bacteriol, 1998. **180**(21): p. 5529-39.

103. Welch, M., Govindarajan, S., Ness, J.E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C., *Design Parameters to Control Synthetic Gene Expression in Escherichia coli.* PloS one, 2009. **4**(9): p. e7002.
104. Kubiak, A.M., *The use of Clostridium sporogenes in directed enzyme pro-drug therapy.* 2013, University of Nottingham.
105. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol Syst Biol, 2011. **7**: p. 539.
106. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., and Gascuel, O., *Phylogeny.fr: robust phylogenetic analysis for the non-specialist.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W465-9.
107. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O., *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.* Syst Biol, 2010. **59**(3): p. 307-21.
108. Huerta-Cepas, J., Serra, F., and Bork, P., *ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data.* Molecular Biology and Evolution, 2016. **33**(6): p. 1635-1638.
109. Solovyev, V. and Salamov, A., *Automatic Annotation of Microbial Genomes and Metagenomic Sequences*, in *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, R.W. Li, Editor. 2011, Nova Science Publishers. p. 67-78.
110. Luo, H., Quan, C.L., Peng, C., and Gao, F., *Recent development of Ori-Finder system and DoriC database for microbial replication origins.* Brief Bioinform, 2018.
111. Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H., *SignalP 4.0: discriminating signal peptides from transmembrane regions.* Nature Methods, 2011. **8**: p. 785.
112. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.e., Wilkins, M.R., Appel, R.D., and Bairoch, A., *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomics Protocols Handbook*, J.M. Walker, Editor. 2005, Humana Press: Totowa, NJ. p. 571-607.
113. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* J Mol Biol, 2001. **305**(3): p. 567-80.
114. Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Geer, L.Y., and Bryant, S.H., *CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.* Nucleic Acids Res, 2017. **45**(D1): p. D200-d203.

115. Marchler-Bauer, A. and Bryant, S.H., *CD-Search: protein domain annotations on the fly.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W327-31.
116. Chatelut, E., Delord, J.P., and Canal, P., *Toxicity patterns of cytotoxic drugs.* Invest New Drugs, 2003. **21**(2): p. 141-8.
117. Connors, T., *Prodrugs in Cancer Chemotherapy*, in *Enzyme-Prodrug Strategies for Cancer Therapy*, R.G. Melton and R.J. Knox, Editors. 1999, Springer US: Boston, MA. p. 11-37.
118. Phillips, R.M., *Targeting the hypoxic fraction of tumours using hypoxia-activated prodrugs.* Cancer chemotherapy and pharmacology, 2016. **77**(3): p. 441-457.
119. Spiegelberg, L., Houben, R., Niemans, R., de Ruysscher, D., Yaromina, A., Theys, J., Guise, C.P., Smaill, J.B., Patterson, A.V., Lambin, P., and Dubois, L.J., *Hypoxia-activated prodrugs and (lack of) clinical progress: The need for hypoxia-based biomarker patient selection in phase III clinical trials.* Clinical and translational radiation oncology, 2019. **15**: p. 62-69.
120. Guise, C.P., Mowday, A.M., Ashoorzadeh, A., Yuan, R., Lin, W.-H., Wu, D.-H., Smaill, J.B., Patterson, A.V., and Ding, K., *Bioreductive prodrugs as cancer therapeutics: targeting tumor hypoxia.* Chinese Journal of Cancer, 2014. **33**(2): p. 80-86.
121. Guise, C.P., Abbattista, M.R., Singleton, R.S., Holford, S.D., Connolly, J., Dachs, G.U., Fox, S.B., Pollock, R., Harvey, J., Guilford, P., Donate, F., Wilson, W.R., and Patterson, A.V., *The bioreductive prodrug PR-104A is activated under aerobic conditions by human aldo-keto reductase 1C3.* Cancer Res, 2010. **70**(4): p. 1573-84.
122. Guise, C.P., Abbattista, M.R., Tipparaju, S.R., Lambie, N.K., Su, J., Li, D., Wilson, W.R., Dachs, G.U., and Patterson, A.V., *Diflavin oxidoreductases activate the bioreductive prodrug PR-104A under hypoxia.* Mol Pharmacol, 2012. **81**(1): p. 31-40.
123. Xu, G. and McLeod, H.L., *Strategies for enzyme/prodrug cancer therapy.* Clin Cancer Res, 2001. **7**(11): p. 3314-24.
124. Niculescu-Duvaz, I., Spooner, R., Marais, R., and Springer, C.J., *Gene-Directed Enzyme Prodrug Therapy.* Bioconjugate Chemistry, 1998. **9**(1): p. 4-22.
125. Freytag, S.O., Stricker, H., Pegg, J., Paielli, D., Pradhan, D.G., Peabody, J., DePeralta-Venturina, M., Xia, X., Brown, S., Lu, M., and Kim, J.H., *Phase I study of replication-competent adenovirus-mediated double-suicide gene therapy in combination with conventional-dose three-dimensional conformal radiation therapy for the treatment of newly diagnosed, intermediate- to high-risk prostate cancer.* Cancer Res, 2003. **63**(21): p. 7497-506.
126. Fillat, C., Carrio, M., Cascante, A., and Sangro, B., *Suicide gene therapy mediated by the Herpes Simplex virus thymidine kinase gene/Ganciclovir system: fifteen years of application.* Curr Gene Ther, 2003. **3**(1): p. 13-26.
127. Wang, D. and Gao, G., *State-of-the-art human gene therapy: part I. Gene delivery technologies.* Discov Med, 2014. **18**(97): p. 67-77.

128. Wang, D. and Gao, G., *State-of-the-art human gene therapy: part II. Gene therapy strategies and applications.* Discov Med, 2014. **18**(98): p. 151-161.
129. Theys, J. and Lambin, P., *Clostridium to treat cancer: dream or reality?* Ann Transl Med, 2015. **3**(Suppl 1): p. S21.
130. Zarogoulidis, P., Darwiche, K., Sakkas, A., Yarmus, L., Huang, H., Li, Q., Freitag, L., Zarogoulidis, K., and Malecki, M., *Suicide Gene Therapy for Cancer – Current Strategies.* J Genet Syndr Gene Ther, 2013. **4**: p. 16849.
131. Portsmouth, D., Hlavaty, J., and Renner, M., *Suicide genes for cancer therapy.* Molecular Aspects of Medicine, 2007. **28**(1): p. 4-41.
132. Knox, R.J., *Enzymes and Prodrugs Used for ADEPT*, in *Enzyme-Prodrug Strategies for Cancer Therapy*, R.G. Melton and R.J. Knox, Editors. 1999, Springer US: Boston, MA. p. 97-131.
133. Roldan, M.D., Perez-Reinado, E., Castillo, F., and Moreno-Vivian, C., *Reduction of polynitroaromatic compounds: the bacterial nitroreductases.* FEMS microbiology reviews, 2008. **32**(3): p. 474-500.
134. Roberts, J.J., Friedlos, F., and Knox, R.J., *CB 1954 (2,4-dinitro-5-aziridinyl benzamide) becomes a DNA interstrand crosslinking agent in Walker tumour cells.* Biochem Biophys Res Commun, 1986. **140**(3): p. 1073-8.
135. Knox, R.J., Friedlos, F., Jarman, M., and Roberts, J.J., *A new cytotoxic, DNA interstrand crosslinking agent, 5-(aziridin-1-yl)-4-hydroxylamino-2-nitrobenzamide, is formed from 5-(aziridin-1-yl)-2,4-dinitrobenzamide (CB 1954) by a nitroreductase enzyme in Walker carcinoma cells.* Biochem Pharmacol, 1988. **37**(24): p. 4661-9.
136. Helsby, N.A., Ferry, D.M., Patterson, A.V., Pullen, S.M., and Wilson, W.R., *2-Amino metabolites are key mediators of CB 1954 and SN 23862 bystander effects in nitroreductase GDEPT.* Br J Cancer, 2004. **90**(5): p. 1084-1092.
137. 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., and Kerr, D.J., *Virus-directed, enzyme prodrug therapy with nitroimidazole reductase: a phase I and pharmacokinetic study of its prodrug, CB1954.* Clin Cancer Res, 2001. **7**(9): p. 2662-8.
138. Patel, P., Young, J.G., Mautner, V., Ashdown, D., Bonney, S., Pineda, R.G., Collins, S.I., Searle, P.F., Hull, D., Peers, E., Chester, J., Wallace, D.M., Doherty, A., Leung, H., Young, L.S., and James, N.D., *A phase I/II clinical trial in localized prostate cancer of an adenovirus expressing nitroreductase with CB1954 [correction of CB1984].* Mol Ther, 2009. **17**(7): p. 1292-9.
139. Williams, E.M., Little, R.F., Mowday, A.M., Rich, M.H., Chan-Hyams, J.V., Copp, J.N., Smaill, J.B., Patterson, A.V., and Ackerley, D.F., *Nitroreductase gene-directed enzyme prodrug therapy: insights and advances toward clinical utility.* Biochem J, 2015. **471**(2): p. 131-53.
140. Copp, J.N., Mowday, A.M., Williams, E.M., Guise, C.P., Ashoorzadeh, A., Sharrock, A.V., Flanagan, J.U., Smaill, J.B., Patterson, A.V., and Ackerley, D.F., *Engineering a Multifunctional Nitroreductase for Improved Activation of*

Prodrugs and PET Probes for Cancer Gene Therapy. Cell Chemical Biology, 2017. **24**(3): p. 391-403.

141. Tang, M.H.Y., Helsby, N.A., Wilson, W.R., and Tingle, M.D., *Aerobic 2- and 4-nitroreduction of CB 1954 by human liver.* Toxicology, 2005. **216**(2–3): p. 129-139.
142. Denny, W.A. and Wilson, W.R., *Considerations for the design of nitrophenyl mustards as agents with selective toxicity for hypoxic tumor cells.* Journal of Medicinal Chemistry, 1986. **29**(6): p. 879-887.
143. Mattes, W.B., Hartley, J.A., and Kohn, K.W., *DNA sequence selectivity of guanine-N7 alkylation by nitrogen mustards.* Nucleic Acids Res, 1986. **14**(7): p. 2971-2987.
144. Wilson, W.R., Hicks, K.O., Pullen, S.M., Ferry, D.M., Helsby, N.A., and Patterson, A.V., *Bystander effects of bioreductive drugs: potential for exploiting pathological tumor hypoxia with dinitrobenzamide mustards.* Radiat Res, 2007. **167**(6): p. 625-36.
145. Palmer, B.D., Wilson, W.R., Atwell, G.J., Schultz, D., Xu, X.Z., and Denny, W.A., *Hypoxia-Selective Antitumor Agents. 9. Structure-Activity Relationships for Hypoxia-Selective Cytotoxicity among Analogs of 5-[N,N-Bis(2-chloroethyl)amino]-2,4-dinitrobenzamide.* Journal of Medicinal Chemistry, 1994. **37**(14): p. 2175-2184.
146. Palmer, B.D., Wilson, W.R., Anderson, R.F., Boyd, M., and Denny, W.A., *Hypoxia-Selective Antitumor Agents. 14. Synthesis and Hypoxic Cell Cytotoxicity of Regioisomers of the Hypoxia-Selective Cytotoxin 5-[N,N-Bis(2-chloroethyl)amino]-2,4-dinitrobenzamide.* Journal of Medicinal Chemistry, 1996. **39**(13): p. 2518-2528.
147. Patterson, A.V., Ferry, D.M., Edmunds, S.J., Gu, Y., Singleton, R.S., Patel, K., Pullen, S.M., Hicks, K.O., Syddall, S.P., Atwell, G.J., Yang, S., Denny, W.A., and Wilson, W.R., *Mechanism of action and preclinical antitumor activity of the novel hypoxia-activated DNA cross-linking agent PR-104.* Clin Cancer Res, 2007. **13**(13): p. 3922-32.
148. Jameson, M.B., Rischin, D., Pegram, M., Gutheil, J., Patterson, A.V., Denny, W.A., and Wilson, W.R., *A phase I trial of PR-104, a nitrogen mustard prodrug activated by both hypoxia and aldo-keto reductase 1C3, in patients with solid tumors.* Cancer Chemother Pharmacol, 2010. **65**(4): p. 791-801.
149. McKeage, M.J., Gu, Y., Wilson, W.R., Hill, A., Amies, K., Melink, T.J., and Jameson, M.B., *A phase I trial of PR-104, a pre-prodrug of the bioreductive prodrug PR-104A, given weekly to solid tumour patients.* BMC Cancer, 2011. **11**: p. 432.
150. Gu, Y., Tingle, M.D., and Wilson, W.R., *Glucuronidation of anticancer prodrug PR-104A: species differences, identification of human UDP-glucuronosyltransferases, and implications for therapy.* J Pharmacol Exp Ther, 2011. **337**(3): p. 692-702.

151. Mowday, A.M., Ashoorzadeh, A., Williams, E.M., Copp, J.N., Silva, S., Bull, M.R., Abbattista, M.R., Anderson, R.F., Flanagan, J.U., Guise, C.P., Ackerley, D.F., Smaill, J.B., and Patterson, A.V., *Rational design of an AKR1C3-resistant analog of PR-104 for enzyme-prodrug therapy.* Biochem Pharmacol, 2016. **116**: p. 176-87.
152. Fass, L., *Imaging and cancer: a review.* Mol Oncol, 2008. **2**(2): p. 115-52.
153. Waerzeggers, Y., Monfared, P., Viel, T., Winkeler, A., Voges, J., and Jacobs, A.H., *Methods to monitor gene therapy with molecular imaging.* Methods, 2009. **48**(2): p. 146-160.
154. Bhaumik, S., Sekar, T.V., Depuy, J., Klimash, J., and Paulmurugan, R., *Noninvasive optical imaging of nitroreductase gene-directed enzyme prodrug therapy system in living animals.* Gene Ther, 2012. **19**(3): p. 295-302.
155. Bettegowda, C., Foss, C.A., Cheong, I., Wang, Y., Diaz, L., Agrawal, N., Fox, J., Dick, J., Dang, L.H., Zhou, S., Kinzler, K.W., Vogelstein, B., and Pomper, M.G., *Imaging bacterial infections with radiolabeled 1-(2'-deoxy-2'-fluoro-beta-D-arabinofuranosyl)-5-iodouracil.* Proc Natl Acad Sci U S A, 2005. **102**(4): p. 1145-50.
156. Stanton, M., Cronin, M., Lehouritis, P., and Tangney, M., *In Vivo Bacterial Imaging without Engineering; A Novel Probe-Based Strategy Facilitated by Endogenous Nitroreductase Enzymes.* Curr Gene Ther, 2015. **15**(3): p. 277-88.
157. Vorobyeva, A.G., Stanton, M., Godinat, A., Lund, K.B., Karateev, G.G., Francis, K.P., Allen, E., Gelovani, J.G., McCormack, E., Tangney, M., and Dubikovskaya, E.A., *Development of a Bioluminescent Nitroreductase Probe for Preclinical Imaging.* PloS one, 2015. **10**(6): p. e0131037.
158. Condeelis, J. and Weissleder, R., *In Vivo Imaging in Cancer.* Cold Spring Harb Perspect, 2010. **2**(12): p. a003848.
159. Peeters, S.G., Zegers, C.M., Yaromina, A., Van Elmpt, W., Dubois, L., and Lambin, P., *Current preclinical and clinical applications of hypoxia PET imaging using 2-nitroimidazoles.* Q J Nucl Med Mol Imaging, 2015. **59**(1): p. 39-57.
160. Fleming, I.N., Manavaki, R., Blower, P.J., West, C., Williams, K.J., Harris, A.L., Domarkas, J., Lord, S., Baldry, C., and Gilbert, F.J., *Imaging tumour hypoxia with positron emission tomography.* Br J Cancer, 2015. **112**(2): p. 238-250.
161. Smaill, J.B., Patterson, A.V., Ashoorzadeh, A., Guise, C.P., Mowday, A.M., Ackerley, D.F., Williams, E.M., and Copp, J.N., *Novel prodrugs and methods of use thereof.* 2013.
162. Niemans, R., Yaromina, A., Theys, J., Ashoorzadeh, A., Anderson, R., Bull, M., Guise, C., Hsu, H.L., Abbattista, M., Mowday, A., Patterson, A.V., Smaill, J.B., Ackerley, D.F., Dubois, L., and Lambin, P., *OC-0236: DTP-006: a novel, orally bioavailable hypoxia-activated prodrug.* Radiotherapy and Oncology, 2016. **119**: p. S107.

163. Niemans, R., Yaromina, A., Theys, J., Ashoorzadeh, A., Anderson, R., Bull, M., Guise, C., Hsu, H.L., Abbattista, M., Mowday, A., Patterson, A.V., Smaill, J.B., Dubois, L., and Lambin, P., *OC-0591: Hypoxic cell killing by SN36506, a novel hypoxia-activated prodrug.* Radiotherapy and Oncology, 2017. **123**: p. S311.

164. Niemans, R., Yaromina, A., Theys, J., Marcus, D., Ashoorzadeh, A., Abbattista, M., Mowday, A., Biemans, R., Lieuwes, N., Deschoemaeker, S., Heyerick, A., Guise, C., Smaill, J.B., Patterson, A.V., Dubois, L., and Lambin, P., *EP-2327: Hypoxic cell killing by CP-506, a novel hypoxia-activated prodrug.* Radiotherapy and Oncology, 2018. **127**: p. S1283-S1284.

165. Anderson, R.F., Smaill, J.B., Patterson, A.V., Ashoorzadeh, A., Ackerley, D.F., Copp, J.N., Mowday, A.M., Williams, E.M., Guise, C.P., Koch, C., Kachur, A., and Dolbier, W.R.J., *Compounds and methods for selective imaging and/or ablation.* 2012.

166. Dolbier, W.R., Li, A.-R., Koch, C.J., Shiue, C.-Y., and Kachur, A.V., *[18F]-EF5, a marker for PET detection of hypoxia: synthesis of precursor and a new fluorination procedure.* Applied Radiation and Isotopes, 2001. **54**(1): p. 73-80.

167. Kachur, A.V., Dolbier, W.R., Jr., Evans, S.M., Shiue, C.Y., Shiue, G.G., Skov, K.A., Baird, I.R., James, B.R., Li, A.R., Roche, A., and Koch, C.J., *Synthesis of new hypoxia markers EF1 and [18F]-EF1.* Appl Radiat Isot, 1999. **51**(6): p. 643-50.

168. Williams, E.M., *Development of bacterial nitroreductase enzymes for noninvasive imaging in cancer gene therapy.* 2013, Victoria University of Wellington.

169. Prosser, G.A., Copp, J.N., Mowday, A.M., Guise, C.P., Syddall, S.P., Williams, E.M., Horvat, C.N., Swe, P.M., Ashoorzadeh, A., Denny, W.A., Smaill, J.B., Patterson, A.V., and Ackerley, D.F., *Creation and screening of a multi-family bacterial oxidoreductase library to discover novel nitroreductases that efficiently activate the bioreductive prodrugs CB1954 and PR-104A.* Biochem Pharmacol, 2013. **85**(8): p. 1091-103.

170. Emptage, C.D., Knox, R.J., Danson, M.J., and Hough, D.W., *Nitroreductase from Bacillus licheniformis: a stable enzyme for prodrug activation.* Biochem Pharmacol, 2009. **77**(1): p. 21-9.

171. Anlezark, G.M., Melton, R.G., Sherwood, R.F., Coles, B., Friedlos, F., and Knox, R.J., *The bioactivation of 5-(aziridin-1-yl)-2,4-dinitrobenzamide (CB1954)--I. Purification and properties of a nitroreductase enzyme from Escherichia coli--a potential enzyme for antibody-directed enzyme prodrug therapy (ADEPT).* Biochem Pharmacol, 1992. **44**(12): p. 2289-95.

172. Sanchez, L.B., Elmendorf, H., Nash, T.E., and Muller, M., *NAD(P)H:menadione oxidoreductase of the amitochondriate eukaryote Giardia lamblia: a simpler homologue of the vertebrate enzyme.* Microbiology, 2001. **147**(Pt 3): p. 561-70.

173. Race, P.R., Lovering, A.L., White, S.A., Grove, J.I., Searle, P.F., Wrighton, C.W., and Hyde, E.I., *Kinetic and structural characterisation of Escherichia*

coli nitroreductase mutants showing improved efficacy for the prodrug substrate CB1954. J Mol Biol, 2007. **368**(2): p. 481-92.

174. Vass, S.O., Jarrom, D., Wilson, W.R., Hyde, E.I., and Searle, P.F., *E. coli NfsA: an alternative nitroreductase for prodrug activation gene therapy in combination with CB1954.* Br J Cancer, 2009. **100**(12): p. 1903-11.
175. Zenno, S., Koike, H., Kumar, A.N., Jayaraman, R., Tanokura, M., and Saigo, K., *Biochemical characterization of NfsA, the Escherichia coli major nitroreductase exhibiting a high amino acid sequence homology to Frp, a Vibrio harveyi flavin oxidoreductase.* J Bacteriol, 1996. **178**(15): p. 4508-14.
176. Parkinson, G.N., Skelly, J.V., and Neidle, S., *Crystal structure of FMN-dependent nitroreductase from Escherichia coli B: a prodrug-activating enzyme.* J Med Chem, 2000. **43**(20): p. 3624-31.
177. Lovering, A.L., Hyde, E.I., Searle, P.F., and White, S.A., *The structure of Escherichia coli nitroreductase complexed with nicotinic acid: three crystal forms at 1.7 A, 1.8 A and 2.4 A resolution.* J Mol Biol, 2001. **309**(1): p. 203-13.
178. Hui, M.P., Foley, P.L., and Belasco, J.G., *Messenger RNA degradation in bacterial cells.* Annu Rev Genet, 2014. **48**: p. 537-59.
179. Battesti, A. and Gottesman, S., *Roles of adaptor proteins in regulation of bacterial proteolysis.* Curr Opin Microbiol, 2013. **16**(2): p. 140-7.
180. Macalino, S.J., Gosu, V., Hong, S., and Choi, S., *Role of computer-aided drug design in modern drug discovery.* Arch Pharm Res, 2015. **38**(9): p. 1686-701.
181. Johansson, E., Parkinson, G.N., Denny, W.A., and Neidle, S., *Studies on the nitroreductase prodrug-activating system. Crystal structures of complexes with the inhibitor dicoumarol and dinitrobenzamide prodrugs and of the enzyme active form.* J Med Chem, 2003. **46**(19): p. 4009-20.
182. Grove, J.I., Lovering, A.L., Guise, C., Race, P.R., Wrighton, C.J., White, S.A., Hyde, E.I., and Searle, P.F., *Generation of Escherichia coli nitroreductase mutants conferring improved cell sensitization to the prodrug CB1954.* Cancer Res, 2003. **63**(17): p. 5532-7.
183. Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D., and Church, G.M., *Composability of regulatory sequences controlling transcription and translation in Escherichia coli.* Proc Natl Acad Sci U S A, 2013. **110**(34): p. 14024-14029.
184. Harrison, E. and Brockhurst, M.A., *Plasmid-mediated horizontal gene transfer is a coevolutionary process.* Trends Microbiol, 2012. **20**(6): p. 262-7.
185. Morrissey, D., O'Sullivan, G.C., and Tangney, M., *Tumour targeting with systemically administered bacteria.* Curr Gene Ther, 2010. **10**(1): p. 3-14.
186. Danino, T., Lo, J., Prindle, A., Hasty, J., and Bhatia, S.N., *In Vivo Gene Expression Dynamics of Tumor-Targeted Bacteria.* ACS Synth Biol, 2012. **1**(10): p. 465-470.
187. Toso, J.F., Gill, V.J., Hwu, P., Marincola, F.M., Restifo, N.P., Schwartzentruber, D.J., Sherry, R.M., Topalian, S.L., Yang, J.C., Stock, F.,

Freezer, L.J., Morton, K.E., Seipp, C., Haworth, L., Mavroukakis, S., White, D., MacDonald, S., Mao, J., Sznol, M., and Rosenberg, S.A., *Phase I study of the intravenous administration of attenuated Salmonella typhimurium to patients with metastatic melanoma.* J Clin Oncol, 2002. **20**(1): p. 142-52.

188. Pyne, M.E., Bruder, M., Moo-Young, M., Chung, D.A., and Chou, C.P., *Technical guide for genetic advancement of underdeveloped and intractable Clostridium.* Biotechnology Advances, 2014. **32**(3): p. 623-641.
189. Lee, J., Jang, Y.S., Papoutsakis, E.T., and Lee, S.Y., *Stable and enhanced gene expression in Clostridium acetobutylicum using synthetic untranslated regions with a stem-loop.* J Biotechnol, 2016. **230**: p. 40-3.
190. He, W., Mu, W., Jiang, B., Yan, X., and Zhang, T., *Food-Grade Expression of d-Psicose 3-Epimerase with Tandem Repeat Genes in Bacillus subtilis.* Journal of Agricultural and Food Chemistry, 2016. **64**(28): p. 5701-5707.
191. Huang, K., Zhang, T., Jiang, B., Yan, X., Mu, W., and Miao, M., *Overproduction of Rummeliibacillus pycnus arginase with multi-copy insertion of the arg (R.pyc) cassette into the Bacillus subtilis chromosome.* Appl Microbiol Biotechnol, 2017. **101**(15): p. 6039-6048.
192. Yomantas, Y.A., Abalakina, E.G., Golubeva, L.I., Gorbacheva, L.Y., and Mashko, S.V., *Overproduction of Bacillus amyloliquefaciens extracellular glutamyl-endopeptidase as a result of ectopic multi-copy insertion of an efficiently-expressed mpr gene into the Bacillus subtilis chromosome.* Microb Cell Fact, 2011. **10**: p. 64.
193. Zhu, T., Guo, M., Tang, Z., Zhang, M., Zhuang, Y., Chu, J., and Zhang, S., *Efficient generation of multi-copy strains for optimizing secretory expression of porcine insulin precursor in yeast Pichia pastoris.* J Appl Microbiol, 2009. **107**(3): p. 954-963.
194. Yang, H., Zhai, C., Yu, X., Li, Z., Tang, W., Liu, Y., Ma, X., Zhong, X., Li, G., Wu, D., and Ma, L., *High-level expression of Proteinase K from Tritirachium album Limber in Pichia pastoris using multi-copy expression strains.* Protein Expression and Purification, 2016. **122**: p. 38-44.
195. Shu, M., Shen, W., Yang, S., Wang, X., Wang, F., Wang, Y., and Ma, L., *High-level expression and characterization of a novel serine protease in Pichia pastoris by multi-copy integration.* Enzyme Microb Technol, 2016. **92**: p. 56-66.
196. Al-Hinai, M.A., Fast, A.G., and Papoutsakis, E.T., *Novel System for Efficient Isolation of Clostridium Double-Crossover Allelic Exchange Mutants Enabling Markerless Chromosomal Gene Deletions and DNA Integration.* Appl Environ Microbiol, 2012. **78**(22): p. 8112.
197. Kuehne, S.A., Heap, J.T., Cooksley, C.M., Cartman, S.T., and Minton, N.P., *ClosTron-Mediated Engineering of Clostridium*, in *Strain Engineering: Methods and Protocols*, J.A. Williams, Editor. 2011, Humana Press: Totowa, NJ. p. 389-407.

198. Lee, S.H., Kim, H.J., Shin, Y.A., Kim, K.H., and Lee, S.J., *Single Crossover-Mediated Markerless Genome Engineering in Clostridium acetobutylicum.* J Microbiol Biotechnol, 2016. **26**(4): p. 725-9.
199. Masters, P.S. and Hong, J.S., *Genetics of the glutamine transport system in Escherichia coli.* J Bacteriol, 1981. **147**(3): p. 805-19.
200. Wu, L. and Welker, N.E., *Cloning and characterization of a glutamine transport operon of Bacillus stearothermophilus NUB36: effect of temperature on regulation of transcription.* J Bacteriol, 1991. **173**(15): p. 4877-88.
201. Weiner, J.H., Furlong, C.E., and Heppel, L.A., *A binding protein for L-glutamine and its relation to active transport in E. coli.* Arch Biochem Biophys, 1971. **142**(2): p. 715-7.
202. Tamura, G.S., Nittayajarn, A., and Schoentag, D.L., *A glutamine transport gene, glnQ, is required for fibronectin adherence and virulence of group B streptococci.* Infect Immun, 2002. **70**(6): p. 2877-85.
203. Boos, W., *Bacterial Transport.* Annu Rev Biochem, 1974. **43**(1): p. 123-146.
204. Betteridge, P.R. and Ayling, P.D., *The regulation of glutamine transport and glutamine synthetase in Salmonella typhimurium.* J Gen Microbiol, 1976. **96**(2): p. 324-34.
205. Tamura, G.S., Bratt, D.S., Yim, H.H., and Nittayajarn, A., *Use of glnQ as a counterselectable marker for creation of allelic exchange mutations in group B streptococci.* Appl Environ Microbiol, 2005. **71**(1): p. 587-90.
206. Jansen, R., Embden, J.D.A.v., Gaastra, W., and Schouls, L.M., *Identification of genes that are associated with DNA repeats in prokaryotes.* Molecular Microbiology, 2002. **43**(6): p. 1565-1575.
207. Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D., *Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.* Microbiology, 2005. **151**(Pt 8): p. 2551-61.
208. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.* J Mol Evol, 2005. **60**(2): p. 174-82.
209. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
210. Bhaya, D., Davison, M., and Barrangou, R., *CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation.* Annu Rev Genet, 2011. **45**: p. 273-97.
211. Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V., *The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli.* Nucleic Acids Res, 2011. **39**(21): p. 9275-9282.
212. Sternberg, Samuel H. and Doudna, Jennifer A., *Expanding the Biologist's Toolkit with CRISPR/Cas9.* Molecular Cell, 2015. **58**(4): p. 568-574.

213. Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C., *Short motif sequences determine the targets of the prokaryotic CRISPR defence system.* Microbiology, 2009. **155**(3): p. 733-740.
214. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E., *A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity.* Science, 2012. **337**(6096): p. 816.
215. Sander, J.D. and Joung, J.K., *CRISPR-Cas systems for editing, regulating and targeting genomes.* Nat Biotechnol, 2014. **32**(4): p. 347-55.
216. Hsu, Patrick D., Lander, Eric S., and Zhang, F., *Development and Applications of CRISPR-Cas9 for Genome Engineering.* Cell, 2014. **157**(6): p. 1262-1278.
217. Selle, K. and Barrangou, R., *Harnessing CRISPR–Cas systems for bacterial genome editing.* Trends Microbiol, 2015. **23**(4): p. 225-232.
218. Dupuy, B. and Matamouros, S., *Regulation of toxin and bacteriocin synthesis in Clostridium species by a new subgroup of RNA polymerase σ-factors.* Res Microbiol, 2006. **157**(3): p. 201-205.
219. Mani, N. and Dupuy, B., *Regulation of toxin synthesis in Clostridium difficile by an alternative RNA polymerase sigma factor.* Proc Natl Acad Sci U S A, 2001. **98**(10): p. 5844-5849.
220. Mani, N., Lyras, D., Barroso, L., Howarth, P., Wilkins, T., Rood, J.I., Sonenshein, A.L., and Dupuy, B., *Environmental response and autoregulation of Clostridium difficile TxeR, a sigma factor for toxin gene expression.* J Bacteriol, 2002. **184**(21): p. 5971-8.
221. Karlsson, S., Dupuy, B., Mukherjee, K., Norin, E., Burman, L.G., and Akerlund, T., *Expression of Clostridium difficile toxins A and B and their sigma factor TcdD is controlled by temperature.* Infect Immun, 2003. **71**(4): p. 1784-93.
222. Martin-Verstraete, I., Peltier, J., and Dupuy, B., *The Regulatory Networks That Control Clostridium difficile Toxin Synthesis.* Toxins (Basel), 2016. **8**(5).
223. An, W. and Chin, J.W., *Synthesis of orthogonal transcription-translation networks.* Proc Natl Acad Sci U S A, 2009. **106**(21): p. 8477-82.
224. Gaida, S.M., Sandoval, N.R., Nicolaou, S.A., Chen, Y., Venkataramanan, K.P., and Papoutsakis, E.T., *Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries.* Nature Communications, 2015. **6**: p. 7045.
225. Bervoets, I., Van Brempt, M., Van Nerom, K., Van Hove, B., Maertens, J., De Mey, M., and Charlier, D., *A sigma factor toolbox for orthogonal gene expression in Escherichia coli.* Nucleic Acids Res, 2018. **46**(4): p. 2133-2144.
226. Meyer, A.J., Ellefson, J.W., and Ellington, A.D., *Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for in Vivo or in Vitro Synthetic Circuitry.* ACS Synth Biol, 2015. **4**(10): p. 1070-1076.
227. Shis, D.L. and Bennett, M.R., *Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants.* Proc Natl Acad Sci U S A, 2013. **110**(13): p. 5028-5033.

228. Temme, K., Hill, R., Segall-Shapiro, T.H., Moser, F., and Voigt, C.A., *Modular control of multiple pathways using engineered orthogonal T7 polymerases.* Nucleic Acids Res, 2012. **40**(17): p. 8773-81.
229. Rhodius, V.A., Segall-Shapiro, T.H., Sharon, B.D., Ghodasara, A., Orlova, E., Tabakh, H., Burkhardt, D.H., Clancy, K., Peterson, T.C., Gross, C.A., and Voigt, C.A., *Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters.* Molecular systems biology, 2013. **9**: p. 702-702.
230. Minton, N.P., Zhang, Y., Schwarz, K., and Heap, J., *Bacterial expression system.* 2012.
231. Ishihama, A., *Functional Modulation of Escherichia Coli RNA Polymerase.* Annual Review of Microbiology, 2000. **54**(1): p. 499-518.
232. Goldfarb, A., Zalenskaya, K., and Malik, S., *Competition between sigma factors for core RNA polymerase.* Nucleic Acids Res, 1987. **15**(20): p. 8521-8530.
233. Hartman, A.H., Liu, H., and Melville, S.B., *Construction and characterization of a lactose-inducible promoter system for controlled gene expression in Clostridium perfringens.* Appl Environ Microbiol, 2011. **77**(2): p. 471-8.
234. Collins, J., Robinson, C., Danhof, H., Knetsch, C.W., van Leeuwen, H.C., Lawley, T.D., Auchtung, J.M., and Britton, R.A., *Dietary trehalose enhances virulence of epidemic Clostridium difficile.* Nature, 2018. **553**: p. 291.
235. Heiss, S., Hormann, A., Tauer, C., Sonnleitner, M., Egger, E., Grabherr, R., and Heinl, S., *Evaluation of novel inducible promoter/repressor systems for recombinant protein expression in Lactobacillus plantarum.* Microb Cell Fact, 2016. **15**: p. 50.
236. Couturier, E. and Rocha, E.P., *Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes.* Mol Microbiol, 2006. **59**(5): p. 1506-18.
237. Sousa, C., de Lorenzo, V., and Cebolla, A., *Modulation of gene expression through chromosomal positioning in Escherichia coli.* Microbiology, 1997. **143 (Pt 6)**: p. 2071-8.
238. Block, D.H., Hussein, R., Liang, L.W., and Lim, H.N., *Regulatory consequences of gene translocation in bacteria.* Nucleic Acids Res, 2012. **40**(18): p. 8979-92.
239. Schmid, M.B. and Roth, J.R., *Gene location affects expression level in Salmonella typhimurium.* J Bacteriol, 1987. **169**(6): p. 2872.
240. Sauer, C., Syvertsson, S., Bohorquez, L.C., Cruz, R., Harwood, C.R., van Rij, T., and Hamoen, L.W., *Effect of Genome Position on Heterologous Gene Expression in Bacillus subtilis: An Unbiased Analysis.* ACS Synth Biol, 2016. **5**(9): p. 942-947.
241. Bryant, J.A., Sellars, L.E., Busby, S.J., and Lee, D.J., *Chromosome position effects on gene expression in Escherichia coli K-12.* Nucleic Acids Res, 2014. **42**(18): p. 11383-92.
242. Kelly, C.L., Taylor, G.M., Hitchcock, A., Torres-Mendez, A., and Heap, J.T., *A Rhamnose-Inducible System for Precise and Temporal Control of Gene Expression in Cyanobacteria.* ACS Synth Biol, 2018. **7**(4): p. 1056-1066.

243. Hohenblum, H., Gasser, B., Maurer, M., Borth, N., and Mattanovich, D., *Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant Pichia pastoris.* Biotechnol Bioeng, 2003. **85**(4): p. 367-375.
244. Wang, J.-J., Rojanatavorn, K., and Shih, J.C.H., *Increased production of Bacillus keratinase by chromosomal integration of multiple copies of the kerA gene.* Biotechnol Bioeng, 2004. **87**(4): p. 459-464.
245. Birnbaum, S. and Bailey, J.E., *Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant Escherichia coli.* Biotechnol Bioeng, 1991. **37**(8): p. 736-45.
246. Wang, Z., Xiang, L., Shao, J., Wegrzyn, A., and Wegrzyn, G., *Effects of the presence of ColE1 plasmid DNA in Escherichia coli on the host cell metabolism.* Microb Cell Fact, 2006. **5**: p. 34.
247. Walter, K.A., Mermelstein, L.D., and Papoutsakis, E.T., *Host-plasmid interactions in recombinant strains of Clostridium acetobutylicum ATCC 824.* FEMS Microbiol Lett, 1994. **123**(3): p. 335-341.
248. Sanden, A.M., Prytz, I., Tubulekas, I., Forberg, C., Le, H., Hektor, A., Neubauer, P., Pragai, Z., Harwood, C., Ward, A., Picon, A., De Mattos, J.T., Postma, P., Farewell, A., Nystrom, T., Reeh, S., Pedersen, S., and Larsson, G., *Limiting factors in Escherichia coli fed-batch production of recombinant proteins.* Biotechnol Bioeng, 2003. **81**(2): p. 158-66.
249. Mauri, M. and Klumpp, S., *A model for sigma factor competition in bacterial cells.* PLoS Comput Biol, 2014. **10**(10): p. e1003845.
250. Vind, J., Sorensen, M.A., Rasmussen, M.D., and Pedersen, S., *Synthesis of proteins in Escherichia coli is limited by the concentration of free ribosomes. Expression from reporter genes does not always reflect functional mRNA levels.* J Mol Biol, 1993. **231**(3): p. 678-88.
251. Ceroni, F., Algar, R., Stan, G.-B., and Ellis, T., *Quantifying cellular capacity identifies gene expression designs with reduced burden.* Nature Methods, 2015. **12**: p. 415.
252. Shachrai, I., Zaslaver, A., Alon, U., and Dekel, E., *Cost of Unneeded Proteins in E. coli Is Reduced after Several Generations in Exponential Growth.* Molecular Cell, 2010. **38**(5): p. 758-767.
253. Borkowski, O., Bricio, C., Murgiano, M., Rothschild-Mancinelli, B., Stan, G.-B., and Ellis, T., *Cell-free prediction of protein expression costs for growing cells.* Nature Communications, 2018. **9**(1): p. 1457.
254. Dong, H., Nilsson, L., and Kurland, C.G., *Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction.* J Bacteriol, 1995. **177**(6): p. 1497.
255. Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T., *Interdependence of cell growth and gene expression: origins and consequences.* Science, 2010. **330**(6007): p. 1099-102.

256. Klumpp, S. and Hwa, T., *Bacterial growth: global effects on gene expression, growth feedback and proteome partition.* Current opinion in biotechnology, 2014. **28**: p. 96-102.
257. Wu, G., Yan, Q., Jones, J.A., Tang, Y.J., Fong, S.S., and Koffas, M.A.G., *Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications.* Trends Biotechnol, 2016. **34**(8): p. 652-664.
258. Dekel, E. and Alon, U., *Optimality and evolutionary tuning of the expression level of a protein.* Nature, 2005. **436**(7050): p. 588-92.
259. Borkowski, O., Ceroni, F., Stan, G.-B., and Ellis, T., *Overloaded and stressed: whole-cell considerations for bacterial synthetic biology.* Current Opinion in Microbiology, 2016. **33**: p. 123-130.
260. Janga, S.C., Salgado, H., and Martinez-Antonio, A., *Transcriptional regulation shapes the organization of genes on bacterial chromosomes.* Nucleic Acids Res, 2009. **37**(11): p. 3680-8.
261. Kolesov, G., Wunderlich, Z., Laikova, O.N., Gelfand, M.S., and Mirny, L.A., *How gene order is influenced by the biophysics of transcription regulation.* Proc Natl Acad Sci U S A, 2007. **104**(35): p. 13948-53.
262. Ceroni, F., Blount, B.A., and Ellis, T., *Sensing the Right Time to Be Productive.* Cell Systems, 2016. **3**(2): p. 116-117.
263. Tsao, C.-Y., Hooshangi, S., Wu, H.-C., Valdes, J.J., and Bentley, W.E., *Autonomous induction of recombinant proteins by minimally rewiring native quorum sensing regulon of E. coli.* Metabolic Engineering, 2010. **12**(3): p. 291-297.
264. Lo, T.-M., Chng, S.H., Teo, W.S., Cho, H.-S., and Chang, M.W., *A Two-Layer Gene Circuit for Decoupling Cell Growth from Metabolite Production.* Cell Systems, 2016. **3**(2): p. 133-143.
265. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. **41**(1): p. 207-34.
266. Blommel, P.G., Becker, K.J., Duvnjak, P., and Fox, B.G., *Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition.* Biotechnol Prog, 2007. **23**(3): p. 585-98.
267. Fox, B.G. and Blommel, P.G., *Autoinduction of protein expression.* Current protocols in protein science, 2009. **Chapter 5**: p. Unit-5.23.
268. Mantovani, A., Romero, P., Palucka, A.K., and Marincola, F.M., *Tumour immunity: effector response to tumour and role of the microenvironment.* Lancet, 2008. **371**(9614): p. 771-783.
269. Dunn, G.P., Old, L.J., and Schreiber, R.D., *The immunobiology of cancer immunosurveillance and immunoediting.* Immunity, 2004. **21**(2): p. 137-48.
270. Pagès, F., Galon, J., Dieu-Nosjean, M.C., Tartour, E., Sautès-Fridman, C., and Fridman, W.H., *Immune infiltration in human tumors: a prognostic factor that should not be ignored.* Oncogene, 2009. **29**: p. 1093.
271. Yang, Y., *Cancer immunotherapy: harnessing the immune system to battle cancer.* J Clin Invest, 2015. **125**(9): p. 3335-7.

272. Balkwill, F.R., Capasso, M., and Hagemann, T., *The tumor microenvironment at a glance.* Journal of Cell Science, 2012. **125**(23): p. 5591.
273. Dougan, M. and Dranoff, G., *Immune therapy for cancer.* Annu Rev Immunol, 2009. **27**: p. 83-117.
274. Ugel, S., De Sanctis, F., Mandruzzato, S., and Bronte, V., *Tumor-induced myeloid deviation: when myeloid-derived suppressor cells meet tumor-associated macrophages.* The Journal of Clinical Investigation, 2015. **125**(9): p. 3365-3376.
275. Motz, G.T. and Coukos, G., *Deciphering and reversing tumor immune suppression.* Immunity, 2013. **39**(1): p. 61-73.
276. Grivennikov, S.I., Greten, F.R., and Karin, M., *Immunity, Inflammation, and Cancer.* Cell, 2010. **140**(6): p. 883-899.
277. Goel, H.L. and Mercurio, A.M., *VEGF targets the tumour cell.* Nat Rev Cancer, 2013. **13**(12): p. 871-882.
278. Mellman, I., Coukos, G., and Dranoff, G., *Cancer immunotherapy comes of age.* Nature, 2011. **480**: p. 480.
279. Hoption Cann, S.A., van Netten, J.P., and van Netten, C., *Dr William Coley and tumour regression: a place in history or in the future.* Postgraduate Medical Journal, 2003. **79**(938): p. 672.
280. Herr, H.W. and Morales, A., *History of bacillus Calmette-Guerin and bladder cancer: an immunotherapy success story.* J Urol, 2008. **179**(1): p. 53-6.
281. Wood, L.M. and Paterson, Y., *Attenuated Listeria monocytogenes: a powerful and versatile vector for the future of tumor immunotherapy.* Front Cell Infect Microbiol, 2014. **4**: p. 51.
282. Kaimala, S., Al-Sbiei, A., Cabral-Marques, O., Fernandez-Cabezudo, M.J., and Al-Ramadi, B.K., *Attenuated Bacteria as Immunotherapeutic Tools for Cancer Treatment.* Front Oncol, 2018. **8**: p. 136.
283. Mogensen, T.H., *Pathogen recognition and inflammatory signaling in innate immune defenses.* Clin Microbiol Rev, 2009. **22**(2): p. 240-73, Table of Contents.
284. al-Ramadi, B.K., Adeghate, E., Mustafa, N., Ponery, A.S., and Fernandez-Cabezudo, M.J., *Cytokine expression by attenuated intracellular bacteria regulates the immune response to infection: the Salmonella model.* Mol Immunol, 2002. **38**(12-13): p. 931-40.
285. Byrne, W.L., Murphy, C.T., Cronin, M., Wirth, T., and Tangney, M., *Bacterial-mediated DNA delivery to tumour associated phagocytic cells.* J Control Release, 2014. **196**: p. 384-93.
286. Nelson, D., Fisher, S., and Robinson, B., *The "Trojan Horse" approach to tumor immunotherapy: targeting the tumor microenvironment.* J Immunol Res, 2014. **2014**: p. 789069.
287. Ngwa, W., Irabor, O.C., Schoenfeld, J.D., Hesser, J., Demaria, S., and Formenti, S.C., *Using immunotherapy to boost the abscopal effect.* Nat Rev Cancer, 2018. **18**(5): p. 313-322.

288. Loeffler, M., Le'Negrate, G., Krajewska, M., and Reed, J.C., *Attenuated Salmonella engineered to produce human cytokine LIGHT inhibit tumor growth.* Proc Natl Acad Sci U S A, 2007. **104**(31): p. 12879-12883.
289. Sorenson, B.S., Banton, K.L., Frykman, N.L., Leonard, A.S., and Saltzman, D.A., *Attenuated Salmonella typhimurium with interleukin 2 gene prevents the establishment of pulmonary metastases in a model of osteosarcoma.* J Pediatr Surg, 2008. **43**(6): p. 1153-8.
290. al-Ramadi, B.K., Fernandez-Cabezudo, M.J., El-Hasasna, H., Al-Salam, S., Bashir, G., and Chouaib, S., *Potent anti-tumor activity of systemically-administered IL2-expressing Salmonella correlates with decreased angiogenesis and enhanced tumor apoptosis.* Clin Immunol, 2009. **130**(1): p. 89-97.
291. Ha, X.Q., Yin, Q., Zhao, H.B., Hui, L., Wang, M.L., Peng, J.H., Dong, J.Z., Deng, Z.Y., Zhao, Y., and Zhang, Y.Y., *Inhibitory effects of the attenuated Salmonella typhimurium containing the IL-2 gene on hepatic tumors in mice.* J Biomed Biotechnol, 2012. **2012**: p. 946139.
292. Al-Ojali, S.M., Tara Moore, C.B., Fernandez-Cabezudo, M.J., and Al-Ramadi, B.K., *IFNgamma expression by an attenuated strain of Salmonella enterica serovar Typhimurium improves vaccine efficacy in susceptible TLR4-defective C3H/HeJ mice.* Med Microbiol Immunol, 2013. **202**(1): p. 49-61.
293. al-Ramadi, B.K., Al-Dhaheri, M.H., Mustafa, N., Abouhaidar, M., Xu, D., Liew, F.Y., Lukic, M.L., and Fernandez-Cabezudo, M.J., *Influence of vector-encoded cytokines on anti-Salmonella immunity: divergent effects of interleukin-2 and tumor necrosis factor alpha.* Infect Immun, 2001. **69**(6): p. 3980-8.
294. Loeffler, M., Le'Negrate, G., Krajewska, M., and Reed, J.C., *IL-18-producing Salmonella inhibit tumor growth.* Cancer Gene Ther, 2008. **15**(12): p. 787-94.
295. Xiang, R., Mizutani, N., Luo, Y., Chiodoni, C., Zhou, H., Mizutani, M., Ba, Y., Becker, J.C., and Reisfeld, R.A., *A DNA vaccine targeting survivin combines apoptosis with suppression of angiogenesis in lung tumor eradication.* Cancer Res, 2005. **65**(2): p. 553-61.
296. Loeffler, M., Le'Negrate, G., Krajewska, M., and Reed, J.C., *Salmonella typhimurium engineered to produce CCL21 inhibit tumor growth.* Cancer Immunol Immunother, 2009. **58**(5): p. 769-75.
297. Urashima, M., Suzuki, H., Yuza, Y., Akiyama, M., Ohno, N., and Eto, Y., *An oral CD40 ligand gene therapy against lymphoma using attenuated Salmonella typhimurium.* Blood, 2000. **95**(4): p. 1258-63.
298. Crittenden, M., Bahjat, K.S., Li, R., Gore, P., Fountain, C., Hanson, B., Skoble, J., Lauer, P., Murphy, A.L., Dubensky, T., Brockstedt, D.G., and Urba, W., *Phase I study of safety and immunogenicity of ADU-623, a live-attenuated listeria monocytogenes vaccine (ΔactA/ΔinlB) expressing EGFRVIII and NY-ESO-1, in patients with who grade III/IV astrocytomas.* Journal for Immunotherapy of Cancer, 2015. **3**(Suppl 2): p. P162-P162.
299. Le, D.T., Brockstedt, D.G., Nir-Paz, R., Hampl, J., Mathur, S., Nemunaitis, J., Sterman, D.H., Hassan, R., Lutz, E., Moyer, B., Giedlin, M., Louis, J.L., Sugar,

E.A., Pons, A., Cox, A.L., Levine, J., Murphy, A.L., Illei, P., Dubensky, T.W., Jr., Eiden, J.E., Jaffee, E.M., and Laheru, D.A., *A live-attenuated Listeria vaccine (ANZ-100) and a live-attenuated Listeria vaccine expressing mesothelin (CRS-207) for advanced cancers: phase I studies of safety and immune induction.* Clin Cancer Res, 2012. **18**(3): p. 858-68.

300. Jia, Y.Y., Tan, W.J., Duan, F.F., Pan, Z.M., Chen, X., Yin, Y.L., and Jiao, X.A., *A Genetically Modified attenuated Listeria Vaccine Expressing HPV16 E7 Kill Tumor Cells in Direct and Antigen-Specific Manner.* Frontiers in Cellular and Infection Microbiology, 2017. **7**: p. 279.
301. Hannan, R., Zhang, H., Wallecha, A., Singh, R., Liu, L., Cohen, P., Alfieri, A., Rothman, J., and Guha, C., *Combined immunotherapy with Listeria monocytogenes-based PSA vaccine and radiation therapy leads to a therapeutic response in a murine model of prostate cancer.* Cancer Immunol Immunother, 2012. **61**(12): p. 2227-38.
302. Mason, N.J., Gnanandarajah, J.S., Engiles, J.B., Gray, F., Laughlin, D., Gaurnier-Hausser, A., Wallecha, A., Huebner, M., and Paterson, Y., *Immunotherapy with a HER2-Targeting Listeria Induces HER2-Specific Immunity and Demonstrates Potential Therapeutic Effects in a Phase I Trial in Canine Osteosarcoma.* Clin Cancer Res, 2016. **22**(17): p. 4380-90.
303. Barbe, S., Van Mellaert, L., Theys, J., Geukens, N., Lammertyn, E., Lambin, P., and Anne, J., *Secretory production of biologically active rat interleukin-2 by Clostridium acetobutylicum DSM792 as a tool for anti-tumor treatment.* FEMS Microbiol Lett, 2005. **246**(1): p. 67-73.
304. Zhang, Y.L., Lu, R., Chang, Z.S., Zhang, W.Q., Wang, Q.B., Ding, S.Y., and Zhao, W., *Clostridium sporogenes delivers interleukin-12 to hypoxic tumours, producing antitumour activity without significant toxicity.* Lett Appl Microbiol, 2014. **59**(6): p. 580-6.
305. Hong, I.S., *Stimulatory versus suppressive effects of GM-CSF on tumor progression in multiple cancer types.* Exp Mol Med, 2016. **48**(7): p. e242.
306. Gasson, J.C., *Molecular physiology of granulocyte-macrophage colony-stimulating factor.* Blood, 1991. **77**(6): p. 1131-45.
307. Ushach, I. and Zlotnik, A., *Biological role of granulocyte macrophage colony-stimulating factor (GM-CSF) and macrophage colony-stimulating factor (M-CSF) on cells of the myeloid lineage.* J Leukoc Biol, 2016. **100**(3): p. 481-9.
308. Francisco-Cruz, A., Aguilar-Santelises, M., Ramos-Espinosa, O., Mata-Espinosa, D., Marquina-Castillo, B., Barrios-Payan, J., and Hernandez-Pando, R., *Granulocyte-macrophage colony-stimulating factor: not just another haematopoietic growth factor.* Med Oncol, 2014. **31**(1): p. 774.
309. Armitage, J.O., *Emerging applications of recombinant human granulocyte-macrophage colony-stimulating factor.* Blood, 1998. **92**(12): p. 4491-508.
310. Kawalec, P., Paszulewicz, A., Holko, P., and Pilc, A., *Sipuleucel-T immunotherapy for castration-resistant prostate cancer. A systematic review and meta-analysis.* Archives of Medical Science : AMS, 2012. **8**(5): p. 767-775.

311. Johnson, D.B., Puzanov, I., and Kelley, M.C., *Talimogene laherparepvec (T-VEC) for the treatment of advanced melanoma.* Immunotherapy, 2015. **7**(6): p. 611-9.
312. Zarei, S., Schwenter, F., Luy, P., Aurrand-Lions, M., Morel, P., Kopf, M., Dranoff, G., and Mach, N., *Role of GM-CSF signaling in cell-based tumor immunization.* Blood, 2009. **113**(26): p. 6658-68.
313. Yan, W.L., Shen, K.Y., Tien, C.Y., Chen, Y.A., and Liu, S.J., *Recent progress in GM-CSF-based cancer immunotherapy.* Immunotherapy, 2017. **9**(4): p. 347-360.
314. Paul, W.E., *History of interleukin-4.* Cytokine, 2015. **75**(1): p. 3-7.
315. Jiang, J., Wang, Z., Li, Z., Zhang, J., Wang, C., Xu, X., and Qin, Z., *Early exposure of high-dose interleukin-4 to tumor stroma reverses myeloid cell-mediated T-cell suppression.* Gene Ther, 2010. **17**: p. 991.
316. Shurin, M.R., Lu, L., Kalinski, P., Stewart-Akers, A.M., and Lotze, M.T., *Th1/Th2 balance in cancer, transplantation and pregnancy.* Springer Semin Immunopathol, 1999. **21**(3): p. 339-59.
317. Li, Z., Chen, L., and Qin, Z., *Paradoxical roles of IL-4 in tumor immunity.* Cell Mol Immunol, 2009. **6**(6): p. 415-22.
318. Gocheva, V., Wang, H.W., Gadea, B.B., Shree, T., Hunter, K.E., Garfall, A.L., Berman, T., and Joyce, J.A., *IL-4 induces cathepsin protease activity in tumor-associated macrophages to promote cancer growth and invasion.* Genes Dev, 2010. **24**(3): p. 241-55.
319. Tepper, R.I., Pattengale, P.K., and Leder, P., *Murine interleukin-4 displays potent anti-tumor activity in vivo.* Cell, 1989. **57**(3): p. 503-512.
320. Tepper, R.I., Coffman, R.L., and Leder, P., *An eosinophil-dependent mechanism for the antitumor effect of interleukin-4.* Science, 1992. **257**(5069): p. 548-51.
321. Carretero, R., Sektioglu, I.M., Garbi, N., Salgado, O.C., Beckhove, P., and Hammerling, G.J., *Eosinophils orchestrate cancer rejection by normalizing tumor vessels and enhancing infiltration of CD8(+) T cells.* Nat Immunol, 2015. **16**(6): p. 609-17.
322. Eguchi, J., Kuwashima, N., Hatano, M., Nishimura, F., Dusak, J.E., Storkus, W.J., and Okada, H., *IL-4-transfected tumor cell vaccines activate tumor-infiltrating dendritic cells and promote type-1 immunity.* J Immunol, 2005. **174**(11): p. 7194-201.
323. Lee, H.L., Park, M.H., Song, J.K., Jung, Y.Y., Kim, Y., Kim, K.B., Hwang, D.Y., Yoon do, Y., Song, M.J., Han, S.B., and Hong, J.T., *Tumor growth suppressive effect of IL-4 through p21-mediated activation of STAT6 in IL-4Ralpha overexpressed melanoma models.* Oncotarget, 2016. **7**(17): p. 23425-38.
324. Dehne, N., Tausendschon, M., Essler, S., Geis, T., Schmid, T., and Brune, B., *IL-4 reduces the proangiogenic capacity of macrophages by down-regulating HIF-1alpha translation.* J Leukoc Biol, 2014. **95**(1): p. 129-37.
325. McKay, R., Ghodasra, M., Schardt, J., Quan, D., Pottash, A.E., Shang, W., Jay, S.M., Payne, G.F., Chang, M.W., March, J.C., and Bentley, W.E., *A*

platform of genetically engineered bacteria as vehicles for localized delivery of therapeutics: Toward applications for Crohn's disease. Bioengineering & translational medicine, 2018. **3**(3): p. 209-221.

326. Berges, H., Joseph-Liauzun, E., and Fayet, O., *Combined effects of the signal sequence and the major chaperone proteins on the export of human cytokines in Escherichia coli.* Appl Environ Microbiol, 1996. **62**(1): p. 55-60.
327. O'Brien, P.M., Seow, H.F., Entrican, G., Coupar, B.E.H., and Wood, P.R., *Production and characterisation of ovine GM-CSF expressed in mammalian and bacterial cells.* Veterinary Immunology and Immunopathology, 1995. **48**(3): p. 287-298.
328. Das, K.M.P., Banerjee, S., Shekhar, N., Damodaran, K., Nair, R., Somani, S., Raiker, V.P., Jain, S., and Padmanabhan, S., *Cloning, soluble expression and purification of high yield recombinant hGMCSF in Escherichia coli.* International journal of molecular sciences, 2011. **12**(3): p. 2064-2076.
329. Srinivasa Babu, K., Muthukumaran, T., Antony, A., Prem Singh Samuel, S.D., Balamurali, M., Murugan, V., and Meenakshisundaram, S., *Single step intein-mediated purification of hGMCSF expressed in salt-inducible E. coli.* Biotechnology Letters, 2009. **31**(5): p. 659-664.
330. Malekian, R., Jahanian-Najafabadi, A., Moazen, F., Ghavimi, R., Mohammadi, E., and Akbari, V., *High-yield Production of Granulocyte-macrophage Colony-stimulating Factor in E. coli BL21 (DE3) By an Auto-induction Strategy.* Iranian Journal of Pharmaceutical Research, 2019. **18**(1): p. 469-478.
331. Levine, A.D., Rangwala, S.H., Horn, N.A., Peel, M.A., Matthews, B.K., Leimgruber, R.M., Manning, J.A., Bishop, B.F., and Olins, P.O., *High Level Expression and Refolding of Mouse Interleukin 4 Synthesized in Escherichia coli.* Journal of Biological Chemistry, 1995. **270**(13): p. 7445-7452.
332. Lundell, D., Greenberg, R., Alroy, Y., Condon, R., Fossetta, J.D., Gewain, K., Kastelein, R., Lunn, C.A., Reim, R., Shah, C., van Kimmenade, A., and Narula, S.K., *Cytoplasmic and periplasmic expression of a highly basic protein, human interleukin 4, inEscherichia coli.* Journal of Industrial Microbiology, 1990. **5**(4): p. 215-227.
333. Wondimu, A., Veit, M., Kohn, B., Kaul, S., Hoffmann, A., Brunnberg, L., and Schmidt, M.F.G., *Molecular cloning, expression and characterization of the Canis familiaris Interleukin-4.* Cytokine, 2001. **16**(3): p. 88-92.
334. van Kimmenade, A., Bond, M.W., Schumacher, J.H., Laquoi, C., and Kastelein, R.A., *Expression, renaturation and purification of recombinant human interleukin 4 from Escherichia coli.* Eur J Biochem, 1988. **173**(1): p. 109-14.
335. Tsirigotaki, A., De Geyter, J., Sostaric, N., Economou, A., and Karamanou, S., *Protein export through the bacterial Sec pathway.* Nat Rev Microbiol, 2017. **15**(1): p. 21-36.
336. Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodriguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M.L., Mattanovich, D., and Villaverde, A., *Protein folding*

and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact, 2008. **7**: p. 11.

337. Freudl, R., *Signal peptides for recombinant protein secretion in bacterial expression systems.* Microbial Cell Factories, 2018. **17**(1): p. 52.
338. Budd, P.G., *Implementation and optimisation of alternative therapeutics for use in Clostridium sporogenes as a delivery vehicle.* 2017, University of Nottingham.
339. Zalucki, Y.M., Beacham, I.R., and Jennings, M.P., *Biased codon usage in signal peptides: a role in protein export.* Trends Microbiol, 2009. **17**(4): p. 146-150.
340. Rosano, G.L. and Ceccarelli, E.A., *Recombinant protein expression in Escherichia coli: advances and challenges.* Front Microbiol, 2014. **5**: p. 172.
341. de Marco, A., *Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli.* Microbial Cell Factories, 2009. **8**: p. 26-26.
342. Andreesen, J.R., Bahl, H., and Gottschalk, G., *Introduction to the Physiology and Biochemistry of the Genus Clostridium*, in *Clostridia*, N.P. Minton and D.J. Clarke, Editors. 1989, Springer US: Boston, MA. p. 27-62.
343. Allison, C. and Macfarlane, G.T., *Physiological and nutritional determinants of protease secretion by Clostridium sporogenes: characterization of six extracellular proteases.* Appl Microbiol Biotechnol, 1992. **37**(2): p. 152-156.
344. Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., and Finn, R.D., *The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database.* Nucleic Acids Res, 2017. **46**(D1): p. D624-D632.
345. Sebaihia, M., Peck, M.W., Minton, N.P., Thomson, N.R., Holden, M.T., Mitchell, W.J., Carter, A.T., Bentley, S.D., Mason, D.R., Crossman, L., Paul, C.J., Ivens, A., Wells-Bennik, M.H., Davis, I.J., Cerdeno-Tarraga, 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., and Parkhill, J., *Genome sequence of a proteolytic (Group I) Clostridium botulinum strain Hall A and comparative analysis of the clostridial genomes.* Genome Res, 2007. **17**(7): p. 1082-92.
346. Van Mellaert, L., Barbé, S., and Anné, J., *Clostridium spores as anti-tumour agents.* Trends Microbiol, 2006. **14**(4): p. 190-196.
347. Gialeli, C., Theocharis, A.D., and Karamanos, N.K., *Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting.* Febs j, 2011. **278**(1): p. 16-27.
348. Westers, L., Westers, H., and Quax, W.J., *Bacillus subtilis as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2004. **1694**(1): p. 299-310.
349. Allison, C. and Macfarlane, G.T., *Regulation of protease production in Clostridium sporogenes.* Appl Environ Microbiol, 1990. **56**(11): p. 3485-3490.

350. Helft, J., Bottcher, J., Chakravarty, P., Zelenay, S., Huotari, J., Schraml, B.U., Goubau, D., and Reis e Sousa, C., *GM-CSF Mouse Bone Marrow Cultures Comprise a Heterogeneous Population of CD11c(+)MHCII(+) Macrophages and Dendritic Cells.* Immunity, 2015. **42**(6): p. 1197-211.
351. Murray, P.J. and Wynn, T.A., *Protective and pathogenic functions of macrophage subsets.* Nat Rev Immunol, 2011. **11**(11): p. 723-37.
352. Dudek, A.M., Martin, S., Garg, A.D., and Agostinis, P., *Immature, Semi-Mature, and Fully Mature Dendritic Cells: Toward a DC-Cancer Cells Interface That Augments Anticancer Immunity.* Front Immunol, 2013. **4**: p. 438.
353. Harris, N.L. and Ronchese, F., *The role of B7 costimulation in T-cell immunity.* Immunol Cell Biol, 1999. **77**(4): p. 304-11.
354. Tan, N.S., Ho, B., and Ding, J.L., *Engineering a novel secretion signal for cross-host recombinant protein expression.* Protein Engineering, Design and Selection, 2002. **15**(4): p. 337-345.
355. Schlegel, S., Rujas, E., Ytterberg, A.J., Zubarev, R.A., Luirink, J., and de Gier, J.-W., *Optimizing heterologous protein production in the periplasm of E. coli by regulating gene expression levels.* Microbial Cell Factories, 2013. **12**: p. 24-24.
356. Zhang, X., Li, Y., and Li, Y., *Evaluating signal peptide prediction methods for Gram-positive bacteria.* Biologia, 2009. **64**(4): p. 655-659.
357. Belinky, F., Rogozin, I.B., and Koonin, E.V., *Selection on start codons in prokaryotes and potential compensatory nucleotide substitutions.* Sci Rep, 2017. **7**(1): p. 12422.
358. Markert, Y., Köditz, J., Mansfeld, J., Arnold, U., and Ulbrich-Hofmann, R., *Increased proteolytic resistance of ribonuclease A by protein engineering.* Protein Engineering, Design and Selection, 2001. **14**(10): p. 791-796.
359. van Tonder, A., Joubert, A.M., and Cromarty, A.D., *Limitations of the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) assay when compared to three commonly used cell enumeration assays.* BMC research notes, 2015. **8**: p. 47-47.
360. Chan, G.K.Y., Kleinheinz, T.L., Peterson, D., and Moffat, J.G., *A simple high-content cell cycle assay reveals frequent discrepancies between cell number and ATP and MTS proliferation assays.* PloS one, 2013. **8**(5): p. e63583-e63583.
361. Barreda, D.R., Hanington, P.C., and Belosevic, M., *Regulation of myeloid development and function by colony stimulating factors.* Dev Comp Immunol, 2004. **28**(5): p. 509-54.
362. Yamazaki, T., Nagata, K., and Kobayashi, Y., *Cytokine production by M-CSF- and GM-CSF-induced mouse bone marrow-derived macrophages upon coculturing with late apoptotic cells.* Cell Immunol, 2008. **251**(2): p. 124-30.
363. Na, Y.R., Jung, D., Gu, G.J., and Seok, S.H., *GM-CSF Grown Bone Marrow Derived Cells Are Composed of Phenotypically Different Dendritic Cells and Macrophages.* Mol Cells, 2016. **39**(10): p. 734-741.
364. Sansom, D.M., Manzotti, C.N., and Zheng, Y., *What's the difference between CD80 and CD86?* Trends Immunol, 2003. **24**(6): p. 314-9.

365. Griffiths, K.L., Tan, J.K., and O'Neill, H.C., *Characterization of the effect of LPS on dendritic cell subset discrimination in spleen.* J Cell Mol Med, 2014. **18**(9): p. 1908-12.

366. Singh-Jasuja, H., Thiolat, A., Ribon, M., Boissier, M.C., Bessis, N., Rammensee, H.G., and Decker, P., *The mouse dendritic cell marker CD11c is down-regulated upon cell activation through Toll-like receptor triggering.* Immunobiology, 2013. **218**(1): p. 28-39.

367. Vacchelli, E., Eggermont, A., Sautès-Fridman, C., Galon, J., Zitvogel, L., Kroemer, G., and Galluzzi, L., *Trial Watch: Toll-like receptor agonists for cancer therapy.* OncoImmunology, 2013. **2**(8): p. e25238.

368. Nastasi, C., Candela, M., Bonefeld, C.M., Geisler, C., Hansen, M., Krejsgaard, T., Biagi, E., Andersen, M.H., Brigidi, P., Ødum, N., Litman, T., and Woetmann, A., *The effect of short-chain fatty acids on human monocyte-derived dendritic cells.* Sci Rep, 2015. **5**: p. 16148.

369. Montville, T.J., Parris, N., and Conway, L.K., *Influence of pH on organic acid production by Clostridium sporogenes in test tube and fermentor cultures.* Appl Environ Microbiol, 1985. **49**(4): p. 733-736.

370. Cebon, J., Nicola, N., Ward, M., Gardner, I., Dempsey, P., Layton, J., Duhrsen, U., Burgess, A.W., Nice, E., and Morstyn, G., *Granulocyte-macrophage colony stimulating factor from human lymphocytes. The effect of glycosylation on receptor binding and biological activity.* J Biol Chem, 1990. **265**(8): p. 4483-91.

371. Denzlinger, C., Tetzloff, W., Gerhartz, H.H., Pokorny, R., Sagebiel, S., Haberl, C., and Wilmanns, W., *Differential activation of the endogenous leukotriene biosynthesis by two different preparations of granulocyte-macrophage colony-stimulating factor in healthy volunteers.* Blood, 1993. **81**(8): p. 2007-13.

372. Guillermo, M., Guillermina, F., Ricardo, K., and Marina, E., *Recombinant human granulocyte-macrophage colony-stimulating factor: effect of glycosylation on pharmacokinetic parameters.* Electronic Journal of Biotechnology; Vol 10, No 2 (2007), 2007.

373. Zhang, Q., Johnston, E.V., Shieh, J.-H., Moore, M.A.S., and Danishefsky, S.J., *Synthesis of granulocyte–macrophage colony-stimulating factor as homogeneous glycoforms and early comparisons with yeast cell-derived material.* Proc Natl Acad Sci U S A, 2014. **111**(8): p. 2885-2890.

374. Andtbacka, R.H.I., Collichio, F.A., Amatruda, T., Senzer, N.N., Chesney, J., Delman, K.A., Spitler, L.E., Puzanov, I., Doleman, S., Ye, Y., Vanderwalde, A.M., Coffin, R., and Kaufman, H., *OPTiM: A randomized phase III trial of talimogene laherparepvec (T-VEC) versus subcutaneous (SC) granulocyte-macrophage colony-stimulating factor (GM-CSF) for the treatment (tx) of unresected stage IIIB/C and IV melanoma.* J Clin Oncol, 2013. **31**(18_suppl): p. LBA9008-LBA9008.

375. Ni, D., Qing, S., Ding, H., Yue, H., Yu, D., Wang, S., Luo, N., Su, Z., Wei, W., and Ma, G., *Biomimetically Engineered Demi-Bacteria Potentiate Vaccination against Cancer.* Adv Sci (Weinh), 2017. **4**(10): p. 1700083.

376. Halioua-Haubold, C.L., Peyer, J.G., Smith, J.A., Arshad, Z., Scholz, M., Brindley, D.A., and MacLaren, R.E., *Regulatory Considerations for Gene Therapy Products in the US, EU, and Japan.* Yale J Biol Med, 2017. **90**(4): p. 683-693.
377. Husain, S.R., Han, J., Au, P., Shannon, K., and Puri, R.K., *Gene therapy for cancer: regulatory considerations for approval.* Cancer Gene Ther, 2015. **22**(12): p. 554-563.
378. Lehouritis, P., Hogan, G., and Tangney, M., *Designer bacteria as intratumoural enzyme biofactories.* Adv Drug Deliv Rev, 2017. **118**: p. 8-23.
379. Denham, J.D., Lee, D.H., Castro, M., Pandya, S., Aslam, S., Nanjappa, S., and Greene, J.N., *Two cases of disseminated infection following live organism anti-cancer vaccine administration in cancer patients.* Int J Infect Dis, 2018. **72**: p. 1-2.
380. Sacco, J.J., Evans, M., Harrington, K.J., Man, S., Powell, N., Shaw, R.J., and Jones, T.M., *Systemic listeriosis following vaccination with the attenuated Listeria monocytogenes therapeutic vaccine, ADXS11-001.* Human Vaccines & Immunotherapeutics, 2016. **12**(4): p. 1085-1086.
381. St Jean, A.T., Zhang, M., and Forbes, N.S., *Bacterial therapies: completing the cancer treatment toolbox.* Curr Opin Biotechnol, 2008. **19**(5): p. 511-517.
382. Mengesha, A., Dubois, L., Lambin, P., Landuyt, W., Chiu, R.K., Wouters, B.G., and Theys, J., *Development of a flexible and potent hypoxia-inducible promoter for tumor-targeted gene expression in attenuated Salmonella.* Cancer Biol Ther, 2006. **5**(9): p. 1120-8.
383. Li, Z., Fallon, J., Mandeli, J., Wetmur, J., and Woo, S.L., *A genetically enhanced anaerobic bacterium for oncopathic therapy of pancreatic cancer.* J Natl Cancer Inst, 2008. **100**(19): p. 1389-400.
384. Li, Z., Fallon, J., Mandeli, J., Wetmur, J., and Woo, S.L., *The oncopathic potency of Clostridium perfringens is independent of its alpha-toxin gene.* Hum Gene Ther, 2009. **20**(7): p. 751-8.
385. Li, Z., Fallon, J., Mandeli, J., Wetmur, J., and Woo, S.L., *Retraction: A genetically enhanced anaerobic bacterium for oncopathic therapy of pancreatic cancer.* J Natl Cancer Inst, 2010. **102**(4): p. 283.
386. Zhao, M., Yang, M., Li, X.-M., Jiang, P., Baranov, E., Li, S., Xu, M., Penman, S., and Hoffman, R.M., *Tumor-targeting bacterial therapy with amino acid auxotrophs of GFP-expressing Salmonella typhimurium.* Proc Natl Acad Sci U S A, 2005. **102**(3): p. 755.
387. Zhao, M., Yang, M., Ma, H., Li, X., Tan, X., Li, S., Yang, Z., and Hoffman, R.M., *Targeted therapy with a Salmonella typhimurium leucine-arginine auxotroph cures orthotopic human breast tumors in nude mice.* Cancer Res, 2006. **66**(15): p. 7647-52.
388. Xu, W., Zhou, T., Zhou, J., Qiang, Z., Zhang, J., and Hua, Z., *Attenuated Salmonella VNP20009 mutant (DeltahtrA) is a promising candidate for*

bacteria-mediated tumour therapy in hosts with TNFR1 deficiency. Lett Appl Microbiol, 2018. **67**(1): p. 97-103.

389. Gunn, G.R., Zubair, A., Peters, C., Pan, Z.K., Wu, T.C., and Paterson, Y., *Two Listeria monocytogenes vaccine vectors that express different molecular forms of human papilloma virus-16 (HPV-16) E7 induce qualitatively different T cell immunity that correlates with their ability to induce regression of established tumors immortalized by HPV-16.* J Immunol, 2001. **167**(11): p. 6471-9.
390. Brockstedt, D.G., Bahjat, K.S., Giedlin, M.A., Liu, W., Leong, M., Luckett, W., Gao, Y., Schnupf, P., Kapadia, D., Castro, G., Lim, J.Y.H., Sampson-Johannes, A., Herskovits, A.A., Stassinopoulos, A., Bouwer, H.G.A., Hearst, J.E., Portnoy, D.A., Cook, D.N., and Dubensky Jr, T.W., *Killed but metabolically active microbes: a new vaccine paradigm for eliciting effector T-cell responses and protective immunity.* Nature Medicine, 2005. **11**: p. 853.
391. Felgner, S., Kocijancic, D., Frahm, M., Heise, U., Rohde, M., Zimmermann, K., Falk, C., Erhardt, M., and Weiss, S., *Engineered Salmonella enterica serovar Typhimurium overcomes limitations of anti-bacterial immunity in bacteria-mediated tumor therapy.* Oncoimmunology, 2018. **7**(2): p. e1382791.
392. Zhang, X., Xu, Q., Yang, L., Lai, Y., Zhang, Z., Han, C., Jiang, C., Li, J., Shi, Y., and Hua, Z.-C., *The genes slyA, STM3120 and htrA are required for the anticancer ability of VNP20009.* Oncotarget, 2016. **7**(49): p. 81187-81196.
393. Tangney, M., van Pijkeren, J.P., and Gahan, C.G.M., *The use of Listeria monocytogenes as a DNA delivery vector for cancer gene therapy.* Bioeng Bugs, 2010. **1**(4): p. 284-287.
394. Hara, Y., Igimi, S., Kasuga, F., Ogura, A., Nagaoka, Y., Takahashi, S., and Mizuno, S., *Isolation of enterotoxin-producing Clostridium sporogenes from cefmetazole-associated diarrheic rabbit.* J Vet Med Sci, 1991. **53**(3): p. 531-2.
395. Hara-Kudo, Y., Ogura, A., Noguchi, Y., and Kumagai, S., *Characteristics of toxicity and haemorrhagic toxin produced by Clostridium sporogenes in various animals and cultured cells.* J Med Microbiol, 1997. **46**(4): p. 270-5.
396. Hara-Kudo, Y., Ogura, A., Noguchi, Y., Terao, K., and Kumagai, S., *Effect of hemorrhagic toxin produced by Clostridium sporogenes on rabbit ligated intestinal loop.* Microb Pathog, 1997. **22**(1): p. 31-8.
397. Marrie, T.J., Faulkner, R.S., Badley, B.W., Hartlen, M.R., Comeau, S.A., and Miller, H.R., *Pseudomembranous colitis: isolation of two species of cytotoxic clostridia and successful treatment with vancomycin.* Can Med Assoc J, 1978. **119**(9): p. 1058-60.
398. Schallehn, G. and Wolff, M.H., *Morphological changes in human embryonic lung fibroblasts caused by cytotoxins of various Clostridium species.* Zentralbl Bakteriol Mikrobiol Hyg A, 1988. **267**(3): p. 367-78.
399. Inkster, T., Cordina, C., and Siegmeth, A., *Septic arthritis following anterior cruciate ligament reconstruction secondary to Clostridium sporogenes; a rare clinical pathogen.* J Clin Pathol, 2011. **64**(9): p. 820-1.

400. Molloy, E.M., Cotter, P.D., Hill, C., Mitchell, D.A., and Ross, R.P., *Streptolysin S-like virulence factors: the continuing sagA.* Nature reviews. Microbiology, 2011. **9**(9): p. 670-681.
401. Alouf, J.E., *Streptococcal toxins (streptolysin O, streptolysin S, erythrogenic toxin).* Pharmacol Ther, 1980. **11**(3): p. 661-717.
402. Duncan, J.L. and Mason, L., *Characteristics of streptolysin S hemolysis.* Infect Immun, 1976. **14**(1): p. 77-82.
403. Wessels, M.R., *Streptolysin S.* J Infect Dis, 2005. **192**(1): p. 13-15.
404. Higashi, D.L., Biais, N., Donahue, D.L., Mayfield, J.A., Tessier, C.R., Rodriguez, K., Ashfeld, B.L., Luchetti, J., Ploplis, V.A., Castellino, F.J., and Lee, S.W., *Activation of band 3 mediates group A Streptococcus streptolysin S-based beta-haemolysis.* Nature Microbiology, 2016. **1**: p. 15004.
405. Ginsburg, I., *Is streptolysin S of group A streptococci a virulence factor?* Apmis, 1999. **107**(12): p. 1051-9.
406. Betschel, S.D., Borgia, S.M., Barg, N.L., Low, D.E., and De Azavedo, J.C.S., *Reduced Virulence of Group A Streptococcal Tn916 Mutants That Do Not Produce Streptolysin S.* Infect Immun, 1998. **66**(4): p. 1671-1679.
407. Datta, V., Myskowski, S.M., Kwinn, L.A., Chiem, D.N., Varki, N., Kansal, R.G., Kotb, M., and Nizet, V., *Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection.* Mol Microbiol, 2005. **56**(3): p. 681-95.
408. Mitchell, D.A., Lee, S.W., Pence, M.A., Markley, A.L., Limm, J.D., Nizet, V., and Dixon, J.E., *Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S.* J Biol Chem, 2009. **284**(19): p. 13004-12.
409. Nizet, V., Beall, B., Bast, D.J., Datta, V., Kilburn, L., Low, D.E., and De Azavedo, J.C., *Genetic locus for streptolysin S production by group A streptococcus.* Infect Immun, 2000. **68**(7): p. 4245-54.
410. Lee, S.W., Mitchell, D.A., Markley, A.L., Hensler, M.E., Gonzalez, D., Wohlrab, A., Dorrestein, P.C., Nizet, V., and Dixon, J.E., *Discovery of a widely distributed toxin biosynthetic gene cluster.* Proc Natl Acad Sci U S A, 2008. **105**(15): p. 5879-84.
411. Wikoff, W.R., Anfora, A.T., Liu, J., Schultz, P.G., Lesley, S.A., Peters, E.C., and Siuzdak, G., *Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites.* Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3698.
412. Dodd, D., Spitzer, M.H., Van Treuren, W., Merrill, B.D., Hryckowian, A.J., Higginbottom, S.K., Le, A., Cowan, T.M., Nolan, G.P., Fischbach, M.A., and Sonnenburg, J.L., *A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites.* Nature, 2017. **551**(7682): p. 648-652.
413. Hara-Kudo, Y., Yamakawa, Y., and Kumagai, S., *Purification and some properties of Clostridium sporogenes hemorrhagic toxin.* Biochem Biophys Res Commun, 1996. **227**(2): p. 413-8.

414. Okamoto, H., Shoin, S., Koshimura, S., and Shimizu, R., *Studies on the Anticancer and Streptolysin S-Forming Abilities of Hemolytic Streptococci.* Japanese Journal of Microbiology, 1967. **11**(4): p. 323-336.
415. Ofek, I., Bergner-Rabinowitz, S., and Ginsburg, I., *Oxygen-stable hemolysins of group A streptococci. 8. Leukotoxic and antiphagocytic effects of streptolysins S and O.* Infect Immun, 1972. **6**(4): p. 459-64.
416. Ginsburg, I., Bentwich, Z., and Harris, T.N., *Oxygen-stable hemolysins of Group A Streptococci. 3. The relationship of the cell-bound homolysin to streptolysin S.* J Exp Med, 1965. **121**: p. 633-45.
417. Cotter, P.D., Draper, L.A., Lawton, E.M., Daly, K.M., Groeger, D.S., Casey, P.G., Ross, R.P., and Hill, C., *Listeriolysin S, a Novel Peptide Haemolysin Associated with a Subset of Lineage I Listeria monocytogenes.* PLOS Pathogens, 2008. **4**(9): p. e1000144.
418. Lyon, W.R., Madden, J.C., Levin, J.C., Stein, J.L., and Caparon, M.G., *Mutation of luxS affects growth and virulence factor expression in Streptococcus pyogenes.* Mol Microbiol, 2001. **42**(1): p. 145-57.
419. Salim, K.Y., de Azavedo, J.C., Bast, D.J., and Cvitkovitch, D.G., *Role for sagA and siaA in Quorum Sensing and Iron Regulation in Streptococcus pyogenes.* Infect Immun, 2007. **75**(10): p. 5011.
420. Baida, G.E. and Kuzmin, N.P., *Cloning and primary structure of a new hemolysin gene from Bacillus cereus.* Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 1995. **1264**(2): p. 151-154.
421. Baida, G.E. and Kuzmin, N.P., *Mechanism of action of hemolysin III from Bacillus cereus.* Biochim Biophys Acta, 1996. **1284**(2): p. 122-4.
422. Chen, Y.C., Chang, M.C., Chuang, Y.C., and Jeang, C.L., *Characterization and virulence of hemolysin III from Vibrio vulnificus.* Curr Microbiol, 2004. **49**(3): p. 175-9.
423. Frey, J. and Falquet, L., *Patho-genetics of Clostridium chauvoei.* Res Microbiol, 2015. **166**(4): p. 384-92.
424. Cowles, K.N. and Goodrich-Blair, H., *Expression and activity of a Xenorhabdus nematophila haemolysin required for full virulence towards Manduca sexta insects.* Cell Microbiol, 2005. **7**(2): p. 209-19.
425. Zhu, L., Peng, D., Wang, Y., Ye, W., Zheng, J., Zhao, C., Han, D., Geng, C., Ruan, L., He, J., Yu, Z., and Sun, M., *Genomic and transcriptomic insights into the efficient entomopathogenicity of Bacillus thuringiensis.* Sci Rep, 2015. **5**: p. 14129.
426. Gentschev, I., Dietrich, G., and Goebel, W., *The E. coli α-hemolysin secretion system and its use in vaccine development.* Trends Microbiol, 2002. **10**(1): p. 39-45.
427. Radulovic, S., Troyer, J.M., Beier, M.S., Lau, A.O., and Azad, A.F., *Identification and molecular analysis of the gene encoding Rickettsia typhi hemolysin.* Infect Immun, 1999. **67**(11): p. 6104-8.

428. Chen, Y., McClane, B.A., Fisher, D.J., Rood, J.I., and Gupta, P., *Construction of an alpha toxin gene knockout mutant of Clostridium perfringens type A by use of a mobile group II intron.* Appl Environ Microbiol, 2005. **71**(11): p. 7542-7547.

429. Nolling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I., Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M.J., Bennett, G.N., Koonin, E.V., and Smith, D.R., *Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.* J Bacteriol, 2001. **183**(16): p. 4823-38.

430. Ivanov, I.T., *Low pH-induced hemolysis of erythrocytes is related to the entry of the acid into cytosole and oxidative stress on cellular membranes.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 1999. **1415**(2): p. 349-360.

431. Tomas, C.A., Beamish, J., and Papoutsakis, E.T., *Transcriptional Analysis of Butanol Stress and Tolerance in Clostridium acetobutylicum.* J Bacteriol, 2004. **186**(7): p. 2006-2018.

432. Goluszko, P. and Nowacki, M.R., *Extracellular haemolytic activity of Serratia marcescens.* FEMS Microbiol Lett, 1989. **52**(1-2): p. 207-11.

433. DeBoy, J.M., 2nd, Wachsmuth, I.K., and Davis, B.R., *Hemolytic activity in enterotoxigenic and non-enterotoxigenic strains of Escherichia coli.* J Clin Microbiol, 1980. **12**(2): p. 193-8.

434. Cooksley, C.M., Davis, I.J., Winzer, K., Chan, W.C., Peck, M.W., and Minton, N.P., *Regulation of neurotoxin production and sporulation by a Putative agrBD signaling system in proteolytic Clostridium botulinum.* Appl Environ Microbiol, 2010. **76**(13): p. 4448-60.

435. Le, K.Y. and Otto, M., *Quorum-sensing regulation in staphylococci-an overview.* Front Microbiol, 2015. **6**: p. 1174.

436. Gera, K. and McIver, K.S., *Laboratory Growth and Maintenance of Streptococcus pyogenes (The Group A Streptococcus, GAS).* Current protocols in microbiology, 2013. **30**: p. 9D.2.1-9D.2.13.

437. Asam, D., Mauerer, S., and Spellerberg, B., *Streptolysin S of Streptococcus anginosus exhibits broad-range hemolytic activity.* Med Microbiol Immunol, 2015. **204**(2): p. 227-37.

438. Flores-Diaz, M., Monturiol-Gross, L., Naylor, C., Alape-Giron, A., and Flieger, A., *Bacterial Sphingomyelinases and Phospholipases as Virulence Factors.* Microbiol Mol Biol Rev, 2016. **80**(3): p. 597-628.

439. Titball, R.W., Naylor, C.E., and Basak, A.K., *The Clostridium perfringens alpha-toxin.* Anaerobe, 1999. **5**(2): p. 51-64.

440. Fiester, S.E., Arivett, B.A., Schmidt, R.E., Beckett, A.C., Ticak, T., Carrier, M.V., Ghosh, R., Ohneck, E.J., Metz, M.L., Sellin Jeffries, M.K., and Actis, L.A., *Iron-Regulated Phospholipase C Activity Contributes to the Cytolytic Activity and Virulence of Acinetobacter baumannii.* PloS one, 2016. **11**(11): p. e0167068.

441. Ochi, S., Oda, M., Nagahama, M., and Sakurai, J., *Clostridium perfringens alpha-toxin-induced hemolysis of horse erythrocytes is dependent on Ca2+ uptake.* Biochim Biophys Acta, 2003. **1613**(1-2): p. 79-86.
442. Titball, R.W., *Bacterial phospholipases C.* Microbiol Rev, 1993. **57**(2): p. 347-66.
443. Vidal, J.E., Chen, J., Li, J., and McClane, B.A., *Use of an EZ-Tn5-Based Random Mutagenesis System to Identify a Novel Toxin Regulatory Locus in Clostridium perfringens Strain 13.* PloS one, 2009. **4**(7): p. e6232.
444. Ohtani, K., Yuan, Y., Hassan, S., Wang, R., Wang, Y., and Shimizu, T., *Virulence gene regulation by the agr system in Clostridium perfringens.* J Bacteriol, 2009. **191**(12): p. 3919-27.
445. Krewulak, K.D. and Vogel, H.J., *Structural biology of bacterial iron uptake.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2008. **1778**(9): p. 1781-1804.
446. Eichenbaum, Z., Muller, E., Morse, S.A., and Scott, J.R., *Acquisition of iron from host proteins by the group A streptococcus.* Infect Immun, 1996. **64**(12): p. 5428-5429.
447. Thorn, G.J., King, J.R., and Jabbari, S., *pH-induced gene regulation of solvent production by Clostridium acetobutylicum in continuous culture: Parameter estimation and sporulation modelling.* Mathematical Biosciences, 2013. **241**(2): p. 149-166.
448. Paterson, G.K. and Mitchell, T.J., *The biology of Gram-positive sortase enzymes.* Trends Microbiol, 2004. **12**(2): p. 89-95.
449. FDA, U.S., *Recommendations for Microbial Vectors used for Gene Therapy - Guidance for Industry.* 2016 Available from: https://www.fda.gov/downloads/biologicsbloodvaccines/guidancecompliancere gulatoryinformation/guidances/cellularandgenetherapy/ucm466625.pdf [Accessed 16/04/2019]
450. Setlow, P., *Spore Resistance Properties.* Microbiol Spectr, 2014. **2**(5).
451. Torres, L., Krüger, A., Csibra, E., Gianni, E., and Pinheiro, V.B., *Synthetic biology approaches to biological containment: pre-emptively tackling potential risks.* Essays in biochemistry, 2016. **60**(4): p. 393-410.
452. Hanahan, D., *Rethinking the war on cancer.* Lancet, 2013. **383**(9916): p. 558-563.
453. Emens, L.A. and Middleton, G., *The Interplay of Immunotherapy and Chemotherapy: Harnessing Potential Synergies.* Cancer immunology research, 2015. **3**(5): p. 436-443.
454. Lesterhuis, W.J., Haanen, J.B., and Punt, C.J., *Cancer immunotherapy--revisited.* Nat Rev Drug Discov, 2011. **10**(8): p. 591-600.
455. Verma, M., *Personalized Medicine and Cancer.* Journal of Personalized Medicine, 2012. **2**(1): p. 1-14.
456. Wood, C.G. and Mulders, P., *Vitespen: a preclinical and clinical review.* Future Oncol, 2009. **5**(6): p. 763-774.

457. Zheng, P.-P., Kros, J.M., and Li, J., *Approved CAR T cell therapies: ice bucket challenges on glaring safety risks and long-term impacts.* Drug Discovery Today, 2018. **23**(6): p. 1175-1182.

458. Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J.J., Cowey, C.L., Lao, C.D., Schadendorf, D., Dummer, R., Smylie, M., Rutkowski, P., Ferrucci, P.F., Hill, A., Wagstaff, J., Carlino, M.S., Haanen, J.B., Maio, M., Marquez-Rodas, I., McArthur, G.A., Ascierto, P.A., Long, G.V., Callahan, M.K., Postow, M.A., Grossmann, K., Sznol, M., Dreno, B., Bastholt, L., Yang, A., Rollin, L.M., Horak, C., Hodi, F.S., and Wolchok, J.D., *Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma.* N Engl J Med, 2015. **373**(1): p. 23-34.

459. Sharma, P. and Allison, J.P., *The future of immune checkpoint therapy.* Science, 2015. **348**(6230): p. 56-61.

460. Kroschinsky, F., Stölzel, F., von Bonin, S., Beutel, G., Kochanek, M., Kiehl, M., Schellongowski, P., on behalf of the Intensive Care in, H., and Oncological Patients Collaborative, G., *New drugs, new toxicities: severe side effects of modern targeted and immunotherapy of cancer and their management.* Critical Care, 2017. **21**: p. 89.

461. Paolicchi, E., Gemignani, F., Krstic-Demonacos, M., Dedhar, S., Mutti, L., and Landi, S., *Targeting hypoxic response for cancer therapy.* Oncotarget, 2016. **7**(12): p. 13464-13478.

462. Collins, F.S. and Gottlieb, S., *The Next Phase of Human Gene-Therapy Oversight.* N Engl J Med, 2018.

Appendix

List of Appendix Tables

Table A1 List of bacterial strains used in this study 254
Table A2 List of primers used in this study 258
Table A3 List of plasmids used in this study 268
Table A4 Nucleotide sequences of codon-optimised genes employed in this study . 287

List of Appendix Figures

Figure A1 Plasmid for genomic integrations via ACE 272
Figure A2 Plasmid for CRISPR/Cas9-mediated genome editing 273
Figure A3 Expression plasmid 273
Figure A4 Plasmid for ClosTron-mediated mutagenesis 274
Figure A5 Plasmid for the purification of YfkO_Bl 275

I. Bacterial strains

Table A1 List of bacterial strains used in this study

Strain	Relevant features	Source
	General use	
E. coli TOP10	cloning and plasmid storage	Life Technologies
E. coli CA434	conjugative donor	[1]
E. coli BL21 (DE3)	recombinant protein production	NEB
C. sporogenes NCIMB 10696 (Csp)	wildtype isolate	laboratory collection
Csp-Δ*spo0A*	*spo0A* deletion; asporogenous strain	Ines Cañadas Blasco
	Chapter 3	
Csp-Δ*pyr*E::P_{thl14}-*nmeNTR*	P_{thl14}-*nmeNTR* integrated in the *pyr*E locus via ACE	Dr Aleksandra Kubiak
Csp-Δ*pyr*E::P_{thl14}-*yfkO_Bl*	P_{thl14}-*yfkO_Bl* integrated in the *pyr*E locus via ACE	Lissa Sutherland
Csp-Δ*pyr*E::P_{thl14}-*yfkO_Bl_362*	P_{thl14}-*yfkO_Bl_362* integrated in the *pyr*E locus via ACE	this study
Csp-Δ*pyr*E::P_{thl14}-*ydgI_Bs*	P_{thl14}-*ydgI_Bs* integrated in the *pyr*E locus via ACE	this study
Csp-Δ*pyr*E::P_{thl14}-*nfsB_Vv_dm*	P_{thl14}-*nfsB_Vv_dm* integrated in the *pyr*E locus via ACE	this study
Csp-Δ*pyr*E::P_{thl14}-*t500*	P_{thl14}-*t500* integrated in the *pyr*E locus via ACE	this study
Csp-Δ*pyr*E::P_{thl14}-*nfrA_Bs*	P_{thl14}-*nfrA_Bs* integrated in the *pyr*E locus via ACE	Lissa Sutherland
Csp-Δ*pyr*E::P_{thl14}-*ycnD_Bs*	P_{thl14}-*ycnD_Bs* integrated in the *pyr*E locus via ACE	Lissa Sutherland
Csp-Δ*pyr*E::P_{fdx}-*nfrA_Bs*	P_{fdx}-*nfrA_Bs* integrated in the *pyr*E locus via ACE	this study

Strain	Relevant features	Source
Csp-Δ*pyr*E::P_{thl14}-*nfrA_Bs*-nat	P_{thl14}-*nfrA_Bs*-nat integrated in the *pyr*E locus via ACE	this study
	Chapter 4	
Csp-Δ*pyr*E::EC	T1-P_{thl14}-*yfko*_Bl-T_{fdx} integrated in the *pyrE* locus via CRISPR/Cas9 (1 copy)	this study
Csp-Δ*glnQ1*::EC	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated in the *glnq1* locus via CRISPR/Cas9	this study
Csp-c14550-3'::EC	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated downstream of c14550 via CRISPR/Cas9	this study
Csp-Δ*treBAR*::EC	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated in the *treBAR* locus via CRISPR/Cas9	this study
Csp-Δ*closA-I*::EC	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated in the *closA-I* locus via CRISPR/Cas9	this study
Csp-EC2	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated in the *pyrE* and *glnQ1* loci via CRISPR/Cas9 (2 copies)	this study
Csp-EC3	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated in the *pyrE, glnQ1* and c14550-3' loci via CRISPR/Cas9 (3 copies)	this study
Csp-EC4	T1-P_{thl14}-*yfko_Bl*-T_{fdx} integrated in the *pyrE, glnQ1,* c14550-3' and *treBAR* loci via CRISPR/Cas9 (4 copies)	this study
Csp-Δ*pyrE*::P_{lac}-*tcdR*	Plac-*tcdR* integrated in the *pyrE* locus via ACE	this study
Csp-Δ*pyrE*::P_{lac}-OS	P_{lac}-*tcdR*-P_{tcdB}-*yfkO_Bl* integrated in the *pyrE* locus via ACE	this study
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus via CRISPR/Cas9	this study
Csp-ΔpyrE::T1-P_{ptb}-OS	T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* integrated in the *pyrE* locus via ACE (1 copy)	this study
Csp-OS2	T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* integrated in the *pyrE* locus via ACE and P_{tcdB}-*yfkO_Bl* in the *glnQ1* locus via CRISPR/Cas9 (2 copies)	this study
Csp-OS3	T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* integrated in the *pyrE* locus via ACE and P_{tcdB}-*yfkO_Bl* in the *glnQ1* and *treBAR* loci via CRISPR/Cas9 (3 copies)	this study

Strain	Relevant features	Source
Csp-OS4	T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl* integrated in the *pyrE* locus via ACE and P_{tcdB}-*yfkO_Bl* in the *glnQ1*, *treBAR* and c14550-3' loci via CRISPR/Cas9 (4 copies)	this study
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* Δ*glnq1*::P_{tcdB}-*yfkO_Bl*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the *glnQ1* locus via CRISPR/Cas9	this study
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* c14550-3'::P_{tcdB}-*yfkO_Bl*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the c14550-3' locus via CRISPR/Cas9	this study
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* Δ*treBAR*::P_{tcdB}-*yfkO_Bl*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the *treBAR* locus via CRISPR/Cas9	this study
Csp-Δ*pyrE*::T1-P_{ptb}-*tcdR* Δ*closA-I*::P_{tcdB}-*yfkO_Bl*	T1-P_{ptb}-*tcdR* integrated in the *pyrE* locus and P_{tcdB}-*yfkO_Bl* in the *closA-I* locus via CRISPR/Cas9	this study
	Chapter 6	
Csp-Δ*closA-I*	*closA-I* deletion	this study
Csp-Δc14780	c14780 deletion	this study
Csp-Δc16380	c16380 deletion	this study
Csp-Δc01750	c01750 deletion	this study
Csp-Δc31080	c31080 deletion	this study
Csp-Δc29800	c29800 deletion	this study
Csp-Δ*closA-I* Δc14780	*closA-I* and c14780 deletion	this study
Csp-Δ*closA-I* Δc16380	*closA-I* and c16380 deletion	this study
Csp-Δ*closA-I* Δc01750	*closA-I* and c01750 deletion	this study
Csp-Δ*closA-I* Δc31080	*closA-I* and c31080 deletion	this study
Csp-Δ*closA-I* Δc29800	*closA-I* and c29800 deletion	this study

Strain	Relevant features	Source
Csp-Δ1-2	c14780 and c16380 deletion	this study
Csp-Δ1-3	c14780, c16380 and c01750 deletion	this study
Csp-Δ1-4	c14780, c16380, c01750 and c31080 deletion	this study
Csp-Δ1-5	c14780, c16380, c01750, c31080 and c29800 deletion	this study
Csp-Δ*closA-I* Δ1-2	*closA-I*, c14780 and c16380 deletion	this study
Csp-Δ*closA-I* Δ1-3	*closA-I*, c14780, c16380 and c01750 deletion	this study
Csp-Δ*closA-I* Δ1-4	*closA-I*, c14780, c16380, c01750 and c31080 deletion	this study
Csp-Δ*closA-I* Δ1-5	*closA-I*, c14780, c16380, c01750, c31080 and c29800 deletion	this study
Csp-*closC*749a::CT-NH	group II intron directionally inserted in *closC*; resE4 and *flhA* mutated, non-haemolytic	this study
Csp-*closC*749a::CT	group II intron directionally inserted in *closC*	this study
RW80	transposon inserted in *res4*	laboratory collection
C. acetobutylicum ATCC 824	wildtype isolate	laboratory collection

II. Primers

Table A2 List of primers used in this study

Primer	Sequence	Use
	Plasmid construction; section 2.2.11.1	
p15a_R1	CAGAGAACCTTCGAAAAAC	screening
catP_F1	GATAAATAGTTAACTTCAGGTTTGTC	screening/sequencing
	Plasmid construction; section 2.2.11.2	
NheI-nRHA_F	GCTAGCATTTAAAAATAAGGAGTGTCTCAAAATAGATTTAATTTTG	RHA
AscI-nRHA_R	GGCGCGCCTTCTACTTCATATGAAACTTTTTTAAC	RHA
MCS_F1_alt	CTTCCCAACAGTTGCGC	screening
pIM13_R1	CTTTTTGCTGCTTTTCC	screening
	Plasmid construction; section 2.2.11.3	
NotI-Pthl_F	TATAGCGGCCGCGGGGGTTAGCATATGCATAAGTTTAA	P_{thl14}-*yfko_Bl_362*
BsaI-YfkO-5'-mut_R	TATAGGTCTCTTCCACTATCTCTTCTCATTTCTTTAGCTGTTCTACCTAATAAAAGTACTAAATGAGATGC	P_{thl14}-*yfko_Bl_362*
BsaI-YfkO-3'-mut_F	TATAGGTCTCATGGATATGTAGCTGATCAATTGAAACATGTTAAGAAAATTCCTGAAGATATAATAG	P_{thl14}-*yfko_Bl_362*
XhoI-YfkO_R	AAAAACTCGAGTTATTCAACCCATTTTAC	P_{thl14}-*yfko_Bl_362*
NotI-Pfdx_F	ATATGCGGCCGCGTGTAGTAGCCTGTGAAATAAGTAA	P_{fdx}-*nfrA_Bs*
BsaI-Pfdx_R1	TATAGGTCTCTTCATATGTAACACACCTCCTTAAAAATTACACAAC	P_{fdx}-*nfrA_Bs*
BsaI-NfrA_F	TATAGGTCTCAATGAACAATACCATAGAAACTATTCT TAATC	P_{fdx}-*nfrA_Bs*
XhoI-NfrA_R	TATATCTCGAGTTAATTTTTGTTAAATCCCTTTTCTTTAACATAATC	P_{fdx}-*nfrA_Bs*

Primer	Sequence	Use
JH27_F	GATGGAATCATGTGTATACGAAGAGG	screening/sequencing
M13F	TGTAAAACGACGGCCAGT	screening/sequencing
	Plasmid construction; section 2.2.11.4	
NdeI-YfkO_F	TATACATATGACAGAACAAAGCAAGAAACAA	*yfkO_Bl*
XhoI-YfkO_R	AAAAACTCGAGTTATTCAACCCATTTTAC	*yfkO_Bl*/screening
pET-28b_seq	CGATGGTGTCCGGGATCTCGACGCT	screening
	Plasmid construction; section 2.2.11.5	
NotI-T1bgaR_F	AAAAAGCGGCCGCTTTAGATATTAATTC	P_{lac}-*tcdR*-P_{tcdB}
BsaI-PtcdB_R1	TATAGGTCTCTTCATAAAATTTTCTCCTTTACTATAATATTTTTATTG	P_{lac}-*tcdR*-P_{tcdB}/ P_{tcdB}
BsaI-YfkO_F	TATAGGTCTCAATGACAGAACAAAGCAAGAAAC	*yfkO_Bl*
XhoI-YfkO_R	AAAAACTCGAGTTATTCAACCCATTTTAC	*yfkO_Bl*
NotI-PtcdB_F	AAAAAGCGGCCGCTTAATGAATTTAAAGAAATATTTAC	P_{tcdB}
NotI-T1_F	TATAGCGGCCGCCTGCCAATAGATAAAATAAAGTCTGCCACAATTGTGGCAGACTTTATTTTATCTATTGATTTATTTCATTTTTTAATAGAGACCCTCGAG	T1
XhoI-BsaI-T1_R	TATACTCGAGGGTCTCTATTAAAAAATGAAATAAATCAATAGATAAAATAAAGTCTGCCACAATTGTGGCAGACTTTATTTTATCTATTGGCAGGCGGCCGC	T1
BsaI-Pptb_F	AAAAAGGTCTCATAATAAGTCAGCAGAAAGTATAATG	P_{ptb}
BsaI-Pptb_R	AAAAAGGTCTCTTGGTCGTACACTCCCTTTTAC	P_{ptb}
BsaI-tcdR_F	AAAAAGGTCTCAACCAATGCAAAAGTCTTTTTATG	*tcdR*-P_{tcdB}-*yfkO_Bl*
M13F	TGTAAAACGACGGCCAGT	screening/sequencing
M13R	CAGGAAACAGCTATGACC	screening/sequencing
	Plasmid construction section 2.2.11.6	
NotI-T1bgaR_F	AAAAAGCGGCCGCTTTAGATATTAATTC	P_{lac}-*tcdR*-P_{tcdB}-*yfkO_Bl*

Primer	Sequence	Use
XhoI-YfkO_R	AAAAACTCGAGTTATTCAACCCATTTTAC	P_{lac}-*tcdR*-P_{tcdB}-*yfkO_Bl*
NotI-T1_F	TATAGCGGCCGCCTGCCAATAGATAAAATAAAGTCTGCCACAATTGTGGCAGACTTTATTTTATCTATTGATTTATTTCATTTTTTAATAGAGACCCTCGAG	T1
XhoI-BsaI-T1_R	TATACTCGAGGGTCTCTATTAAAAAATGAAATAAATCAATAGATAAAATAAAGTCTGCCACAATTGTGGCAGACTTTATTTTATCTATTGGCAGGCGGCCGC	T1
BsaI-Pptb_F	AAAAAGGTCTCATAATAAGTCAGCAGAAAGTATAATG	P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl*
XhoI-tcdR_R	AAAAACTCGAGTTACAAGTTAAAATAATTTTCATAGTCTTTTTTTATTTTTTTAAATG	P_{lac}-*tcdR*
JH27_F	GATGGAATCATGTGTATACGAAGAGG	screening/sequencing
M13F	TGTAAAACGACGGCCAGT	screening/sequencing
	Plasmid construction section; 2.2.11.7	
sgRNA_Rpd	CGCGCGCGGCGATCGCATAAAAATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTTATAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTG	sgRNAs
sgRNA_pyrE_F	TGCTGTCGACATAGTAGGACCAGCTATGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAA	*pyrE* sgRNA
sgRNA_glnQ1_F	TATAGTCGACGGTTTAGAAGAAATAGCGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAA	*glnQ1* sgRNA
sgRNA_c14550_F	TATAGTCGACGACAAAACAAGTCAAATAGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAA	c14550-3'sgRNA
sgRNA_tre_F	TATAGTCGACATACCTCCACAGAACATAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAA	*treBAR* sgRNA
sgRNA_clos_F	TATAGTCGACGGACTCTTAACCCCAACTAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAA	*closA-I* sgRNA
sgRNA_c14780_F	TGCTGTCGACCCAAATATAGTCCAACCTAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAA	c14780 sgRNA
sgRNA_c163880_F	TGCTGTCGACACAGTTTTTTCAATTTTCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAA	c16380 sgRNA
sgRNA_c01750_F	TGCTGTCGACAAGTACTCCGATTAAAACAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAA	c01750 sgRNA

Primer	Sequence	Use
sgRNA_c31080_F	TGCTGTCGACGTTGTTTTAATTGATCACAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAA	c31080 sgRNA
sgRNA_c29800_F	TGCTGTCGACACTTGTAGCGGCAGAAGCACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTATAAAAA	c29800 sgRNA
AsiSI-L-pyrE_F	TGCTTAGCGATCGCAATAAAGTTTTGCCATCTGCAATTACTGTAAG	*pyrE* LHA
NotI-L-pyrE_R	TGCTTAGCGGCCGCATTTAAAAATAAGGAGTGTCTCAAAATAGATTTAATTTTG	*pyrE* LHA
XhoI-R-pyrE_F	TGCTTACTCGAGTATATTAAAGTCTACATTTTTAAGCTTCTGTGC	*pyrE* RHA
AscI-R-pyrE_R	TGCTTAGGCGCGCCTTCAAGTAAAGGATTGACCTTAAATCCTAAAGAAGG	*pyrE* RHA
AsiSI-L-gnQ1_F	TATAGCGATCGCGCTTAACTCCTTGT GATGAAAAATTC	*glnQ1* LHA
NotI-L-glnQ1_R	TATATGCGGCCGCATTTATATCTAATTTAAAGGTAATAGC	*glnQ1* LHA
XhoI-R-glnQ1_F	TATATCTCGAGAAATAAAATTTAATTTATGAAATATTTTCACCAATATAC	*glnQ1* RHA
AscI-R-glnQ1_R	TATAGGCGCGCCCTCCTCCCATTATAG CTTGATAAC	*glnQ1* RHA
AsiSI-L-c14550_F	TGCTTAGCGATCGCTGGTCAGAAAACGTCTTCTTCTAATGAAAAAG	c14550-3’ LHA
BsaI-L-c14550-A_R	TGCTTAGGTCTCTCAAATCCTTCAATCTGTGAAGTCTTATCATTTCTTGACTCGAAGGGTGGATATTGAGCACATAACCCTACTAC	c14550-3’ LHA part A
BsaI-L-c14550-B_F	TGCTTAGGTCTCATTTGATGTAGACTTAGCTAAAGAATTAGG	c14550-3’ LHA part B
NotI-L-c14550_R	TGCTTAGCGGCCGCCTAATTCCAATTTTCATTTCTTTTTTCTATAAATTTTG	c14550-3’ LHA
XhoI-R-c14550_F	TGCTTACTCGAGATTAATAAGTGAGTCTATTTGTTTGTAAAAT	c14550-3’ RHA
AscI-R-c14550_R	TGCTTAGGCGCGCCAATTCATCTACTAATATATATTTAAATAATTTTTTATAACC	c14550-3’ RHA
AsiSI-L-tre_F	TATAGCGATCGCGATAAGTCCTATTATTATTAAAATTTCCTATAAGAAAATATTAATG	*treBAR* LHA
NotI-L-tre_R	TGCTTAGCGGCCGCCATTTTCAACTAATTTAATATCTCCAATAATATTACGG	*treBAR* LHA
XhoI-R-tre_F	TGCTTACTCGAGGATTTAAATTTGTAGAATTTGCACGAAGAAAATAATACTG	*treBAR* RHA
AscI-R-tre_R	TATAGGCGCGCCCAATAATTTACAAGGTATTTAAGATAAG	*treBAR* RHA
AsiSI-L-clos_F	TATAGCGATCGCTATCATGCTACAATATGATGCATATAC	*closA-I* LHA

Primer	Sequence	Use
NotI-L-clos_R	TGCTTAGCGGCCGCATCTTTCACCCCTTACTATTAAATATTAAATTAAATG	*closA-I* LHA
XhoI-R-clos_F	TGCTTACTCGAGGTTATAAAAAATAAATGTAGATAAAAATTTTTTAATAGG	*closA-I* RHA
AscI-R-clos_R	TATAGGCGCGCCTACTAAAAATTCCTTGGCTATTTCATAAG	*closA-I* RHA
BsaI-L-clos_R	CTTAAAAAATTAACATATCTTTCACCCCTTACTATTAAATATTAAATTAAATG	*closA-I* LHA
BsaI-R-clos_F	AGATATGTTAATTTTTTAAGTTATAAAAAATAAATGTAGATAAAAATTTTTTAATAG	*closA-I* RHA
AsiSI-L-c14780_F	TGCTTAGCGATCGCCCAGGTACTATGCCTAAGTTCCTTGGATTT	c14780 LHA
BsaI-L-c14780_R	TGCTTAGGTCTCTAAATAGGTAACAAATAGAACACTATGAACTATATA	c14780 LHA
BsaI-R-c14780_F	TGCTTAGGTCTCAATTTAACATAAACATATCTCCTTTATAATATATATATTATATT	c14780 RHA
AscI-R-c14780_R	TGCTTAGGCGCGCCTGATAACAATGGATAAAAAACTTATTCAGAAAAG	c14780 RHA
AsiSI-L-c16380_F	TGCTTAGCGATCGCATGGTATAAATCATTCCATAGCAGCTTTATCTA	c16380 LHA
BsaI-L-c16380_R	TGCTTAGGTCTCTTAAAATGTAAACTCCAAAAGTGCTACTGGTAAATTGATTAC	c16380 LHA
BsaI-R-c16380_F	TGCTTAGGTCTCATTTAAACATAAAAAACATCCTCTCTAAATTGCATTG	c16380 RHA
AscI-R-c16380_R	TGCTTAGGCGCGCCATTCTCTTTTCCATATTAGCCTTACTAATTAAAG	c16380 RHA
AsiSI-L-c01750_F	TGCTTAGCGATCGCGAACACATCAAGTTCTAAAAGGATCTATGTC	c01750 LHA
BsaI-L-c01750_R	TGCTTAGGTCTCTACTCCTTCATTAGTTAATCCCTCCCTTATCCTTTGATG	c01750 LHA
BsaI-R-c01750_F	TGCTTAGGTCTCAGAGTAATTTTATAGTATTTTTTATTTTAAAGAG	c01750 RHA
AscI-R-c01750_R	TGCTTAGGCGCGCCATTTTGTATCCATTATAAAACTCTCCTTAATCTAATTAC	c01750 RHA
AsiSI-L-c31080_F	TGCTTAGCGATCGCTCTTTTCCCTAACATTTACACCACTTGGAGTAC	c31080 LHA
BsaI-L-c31080_R	TGCTTAGGTCTCTGTTAAAATAGAAAGTGAAGTGGAGTACATTGGAG	c31080 LHA
BsaI-R-c31080_F	TGCTTAGGTCTCATAACTCCACATTATACCTCCATTAAATAAAACACATC	c31080 RHA
AscI-R-c31080_R	TGCTTAGGCGCGCCCTAATCAAATAAACTTTGGTAATGTTGGAGCAATTGGG	c31080 LHA
AsiSI-L-c29800_F	TGCTTAGCGATCGCATGTATTTTTACATTACTCTTATTATTTTTAAAATAATC	c29800 LHA

Primer	Sequence	Use
BsaI-L-c29800_R	TGCTTAGGTCTCTACTCATTATTCATAAAAATTCACTCCTTCTTATATAGTG	c29800 LHA
BsaI-R-c29800_F	TGCTTAGGTCTCAGAGTAAAAAAAGTAGAACCTCTATTTAGAGGTTC	c29800 RHA
AscI-R-c29800_R	TGCTTAGGCGCGCCGAAGGAATAATAGCATCAAAGATAGCAGCCCATGC	c29800 RHA
NotI-T1_F	TATAGCGGCCGCCTGCCAATAGATAAAATAAAGTCTGCCACAATTGTGGCAGACTTTATTTTATCTATTGATTTATTTCATTTTTTAATAGAGACCCTCGAG	T1
XhoI-BsaI-T1_R	TATACTCGAGGGTCTCTATTAAAAAATGAAATAAATCAATAGATAAAATAAAGTCTGCCACAATTGTGGCAGACTTTATTTTATCTATTGGCAGGCGGCCGC	T1
BsaI-Pthl_F	TATAGGTCTCATAATGGGGGTTAGCATATGCATAAGTTTAATTTTTTTG	P_{thl14}-*yfkO_Bl*-T_{fdx}
XhoI-Tfdx-YfkO_R	TATACTCGAGATAAAAATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTTATTTATTCAACCCATTTTACAATTTGATCTAAAGCTCTTCTAGTTTTGGGTCTAG	P_{thl14}-*yfkO_Bl*-T_{fdx}/ P_{tcdB}-*yfkO_Bl*-T_{fdx}
BsaI-PtcdB_F	TATAGGTCTCATAATTTAATGAATTTAAAGAAATATTTAC	P_{tcdB}-*yfkO_Bl*-T_{fdx}
BsaI-Pptb_F	AAAAAGGTCTCATAATAAGTCAGCAGAAAGTATAATG	P_{pbt}-*tcdR*
XhoI-tcdR_R	AAAAACTCGAGTTACAAGTTAAAATAATTTTCATAGTCTTTTTTTATTTTTTTAAATG	P_{pbt}-*tcdR*
CRISPR_Fscr	ATATATTCTAGATTTATATTTAGTCCCTTGCC	screening/sequencing
pCB102_R1	CTGTTATGCCTTTTGACTATC	screening/sequencing
Plasmid construction; section 2.2.11.8		
BsaI-npr_F	TATAGGTCTCATTTTATGAAAAGTAAAAAATTATTAACTATACTTTTAAG	*npr*
BsaI-npr_R1	TATAGGTCTCTGCGCAGCATAAATTGATGAAACAGGTG	*npr*
BsaI-npr_R2	TATAGGTCTCTTATGAGCATAAATTGATGAAACAGGTG	*npr*
BsaI-lip_F	TATAGGTCTCATTTTATGAATAAGTTTTTTAAAAAAGTTATGAGTAAGTTATTATTAGTACCGTTCTTAGTTTTATCTTTGGCAACACC	*lip*
BsaI-lip_R1	TATAGGTCTCTGCGCAGCAAATACAGGTGTTGCCAAAGATAAAACTAAGAACGGTACTAATAATAACTTACTCATAAC	*lip*
BsaI-lip_R2	TATAGGTCTCTTATGAGCAAATACAGGTGTTGCCAAAGATAAAACTAAGAACGGTACTAATAATAACTTACTCATAAC	*lip*
BsaI-egla_F	TATAGGTCTCATTTTTTGTTTTCAAAAATCAAAAAAATTAATTTTTTTAAAAAAACATTTTCTTTTTTAATTGCTGTTGTAATGATGTTGTTTACAGTATTAGG	*egla*

Primer	Sequence	Use
BsaI-egla_R1	TATAGGTCTCTGCGCAGCTTCAGCTTTATAAGTATTTGTTCCTAATACTGTAAACAACATCATTACAACAGCAATTAAAAAAGAAAATG	*egla*
BsaI-egla_R2	TATAGGTCTCTTATGAGCTTCAGCTTTATAAGTATTTGTTCCTAATACTGTAAACAACATCATTACAACAGCAATTAAAAAAGAAAATG	*egla*
NotI-PtcdB_F	AAAAAGCGGCCGCTTAATGAATTTAAAGAAATATTTAC	P_{tcdB}
BsaI-PtcdB_R2	TATAGGTCTCTAAAATTTTCTCCTTTACTATAATATTTTTATTG	P_{tcdB}
BsaI-GMCSF_F	TATAGGTCTCAGCGCCTACAAGGAGCCCAATAACTG	mGM-CSF
XhoI-flag-GMCSF_R	TATACTCGAGTTATTTATCATCATCATCTTTATAATCTTTTTGACCTGGTTTTTTAC	mGM-CSF
BsaI-IL4_F	TATAGGTCTCACATATTCATGGATGTGATAAGAATCATC	mIL-4
XhoI-flag-IL4_R	TATACTCGAGTTATTTATCATCATCATCTTTATAATCAGAATAATCCATTTG	mIL-4
BsaI-PtcdB_R3	TATAGGTCTCTGCGCCATAAAATTTTCTCCTTTACTATAATATTTTTATTGAATATTTTTAC	P_{tcdB}
BsaI-PtcdB_R4	TATAGGTCTCTTATGCATAAAATTTTCTCCTTTACTATAATATTTTTATTGAATATTTTTAC	P_{tcdB}
M13F	TGTAAAACGACGGCCAGT	screening/sequencing
M13R	CAGGAAACAGCTATGACC	screening/sequencing
	Plasmid loss section 2.2.13	
pIM13_F2	CACCACGTAGTTATTGGG	ACE
CatP_R1	CAAGTTTATCGCTCTAATGAAC	ACE
CatP_F1	GATAAATAGTTAACTTCAGGTTTGTC	CRISPR/Cas9
MCS_R1	GGTCATAGCTGTTTCCTG	CRISPR/Cas9
pCB102_F2	GTCAAGTATGAAATCATAAATAAAG	ClosTron
ColE1_R1	CTTTTCTACGGGGTCTGAC	ClosTron
	Chapter 3	
lacZα	CTGGCCGTCGTTTTACAACGTCGTG	screening (single crossover)

Primer	Sequence	Use
Csp-3233-SR3	CAAAAGCACTTGAAAATGCATG	screening
MZ1-SC_scr	GCTTCCCCACCTTCAAAAGCTTCTCC	screening (single crossover)
pyrD-sF2	GAAGACTTAGAAAATTATATGAAAGAAGAAGGT	screening/sequencing
M13F	TGTAAAACGACGGCCAGT	sequencing
	Chapter 4	
pyrE_Fscr	CTTTAGGATACAAACAAGCAGAAGAC	screening
pyrE_Rscr	GGAACTTTTGGATTTGGAGAAGAATATG	screening
c14550_Fscr	ATGTTCAAGAAAAAAGTATTAAGTG	screening
c14550_Rscr	CTGCAAATAAAGAATTACCCTTAGATAAC	screening
tre_Fscr	GACTTATTTGGGTTTATTTAAACTACCCCG	screening
tre_Rscr	TAGCATTGGACATTATTGTTAAGCATAAGAC	screening
clos_Fscr	GACAAAGAAAAGATGGAGCTATAC	screening
clos_Rscr	CTCCAATTGTGACAAATATAAATTATAATTTTC	screening
BsaI-YfkO_F	TATAGGTCTCAATGACAGAACAAAGCAAGAAAC	screening
glnQ1_Rscr	ATTCCTGTAATCCAAACATCTATAGC	screening
pyrE_Fseq	CTATTATAACTCAAAACAATAAAAATTAATAACC	sequencing
pyrE_Rseq	GAAGACTTAGAAAATTATATGAAAGAAGAAGGT	sequencing
glnQ1_Fseq	TATATAATTTTAAATTAGCTATTAC	sequencing
glnQ1_Rseq	CCCTGTAAAAATCCATTAAAAATC	sequencing
c14550_Fseq	GCCCTTTGAAGTATATTTAACTTTAGG	sequencing
c14550_Rseq	TTTACTTCTTTTCATTATCTAATTCCTCC	sequencing
tre_Fseq	CACCTCTTATCCCTGCCCTTATAGTTGGAGG	sequencing

Primer	Sequence	Use
tre_Rseq	CTGTCTTAAAGTAGGTTTAATTTTAAGACAG	sequencing
clos_Fseq	AAGCTATTCGATACTATAAATATAAAGAAG	sequencing
clos_Rseq	TATTTTAACTCACTAATTTTAATAAAAACC	sequencing
lacZα	CTGGCCGTCGTTTTACAACGTCGTG	screening (single crossover)
MZ1-SC_scr	GCTTCCCCACCTTCAAAAGCTTCTCC	screening (single crossover)
pyrD-sF2	GAAGACTTAGAAAATTATATGAAAGAAGAAGGT	screening/sequencing
Csp-3233-SR3	CAAAAGCACTTGAAAATGCATG	screening
M13F	TGTAAAACGACGGCCAGT	screening/sequencing
M13R	CAGGAAACAGCTATGACC	screening/sequencing
YfkO_Rseq	CAAGTATGAATTGGAAATCTTC	sequencing
PbgaL_Rseq	CTTTTAGAATAGTTAAAAAGTTC	sequencing
tcdR_Fseq	ATGCAAAAGTCTTTTTATGAATTAATTGTTTTAG	sequencing
tcdR_Rseq	TTTAGCCTTATTAACAGCTTGTCTAG	sequencing
	Chapter 5	
M13F	TGTAAAACGACGGCCAGT	screening/sequencing
M13R	CAGGAAACAGCTATGACC	screening/sequencing
	Chapter 6	
clos_Fscr	GACAAAGAAAAGATGGAGCTATAC	screening
clos_Rscr	CTCCAATTGTGACAAATATAAATTATAATTTTC	screening
closC_R	CTTCTAAAATTTCATCCACAATATTT	screening
c14780_Fscr	GGATATGAAACAGAGATAATACAAGGG	screening
c14780_Rscr	GTGATTGTAATATAATCATGTTTATTTTCG	screening

Primer	Sequence	Use
c16380_Fscr	AACATTATAATTATATAAGTAATC	screening
c16380_Rscr	GTATTTTAACAATCATTTATTATC	screening
c01750_Fscr	AGGGAAGTGTTAAGGCGGTATTTGCTG	screening
c01750_Rscr	AAATGCAGCTTTTTTTACTATAGGTG	screening
c31080_Fscr	TGTACTTTAACGCCATTGGGAAGAG	screening
c31080_Rscr	ATTGGCGACATAGCACATATCCTTGG	screening
c29800_Fscr	CCTTCTAATAATGCTGTTAAAAAGC	screening
c29800_Rsc	CAGCTATAGGAGGAGCTATTGCAGC	screening
closC_Fscr	GCAATAATTTAAGAATATATGATAACG	screening
RAM_F	ACGCGTTATATTGATAAAAATAATAATAGTGGG	screening
RAM_R	ACGCGTGCGACTCATAGAATTATTTCCTCCCG	screening
EBS_U	CGAAATTAGAAACTTGCGTTCAGTAAAC	screening
inv_F	GAAGCACCAGCATCTCTTG	inverse PCR
inv_R	CACTCAAGTCTCGATTCAGCA	inverse PCR
erm_RAM_seq	GGAAATGGCAATGATAGCGA	inverse PCR sequencing

III. Plasmids

Table A3 List of plasmids used in this study

Plasmid	Relevant features	Source
	General vector backbones	
pMTL82151	expression plasmid, ColE1/pBP1 replicons, Cm^R/Tm^R	laboratory collection
pMTL85121	expression plasmid, p15a/pIM13 replicons, Cm^R/Tm^R; used as a source of p1a+*traJ*	laboratory collection
pMTL82121	expression plasmid, p15a/pBP1 replicons, Cm^R/Tm^R	this study
pMTL-JH27	ACE vector, 1.2 kb RHA, ColE1/pIM13 replicons, Cm^R/Tm^R	laboratory collection
pMTL-MZ1	ACE vector, 2.5 kb RHA, ColE1/ pIM13 replicons, Cm^R/Tm^R	this study
pMTL-MZ1-p15a	ACE vector, 1.2 kb RHA, p15a/pIM13 replicons, Cm^R/Tm^R	this study
pMTL-MZ1-p15a-T1	ACE vector, 2.5 kb RHA, p15a/pIM13 replicons, Cm^R/Tm^R, contains T1 terminator	this study
pMTL83151-RECas1	CRISPR/Cas9 vector, *cas9* driven by P_{fdx}-rb3, ColE1/pCB102 replicons, Cm^R/Tm^R	Ines Cañadas Blasco
pMTL83121-RECas1	CRISPR/Cas9 vector, *cas9* driven by P_{fdx}-rb3, p15a/pCB102 replicons, Cm^R/Tm^R	this study
pMTL-HZ7	contains P_{lac}-*tcdR*-P_{tcdB}-*catP*; used as a DNA template	laboratory collection
pMTL007C-E2	ClosTron vector, ColE1/pCB102 replicons, Cm^R/Tm^R	laboratory collection
pET-28b	recombinant protein expression vector for *E. coli* BL21 (DE3), pBR322/ROP replicon, Kan^R	laboratory collection
	Chapter 3	
pMTL-JH27-P_{thl14}-*yfkO_Bl*	ACE vector, contains P_{thl14}-*yfkO_Bl*	Lissaa Sutherland
pMTL-MZ1-P_{thl14}-*yfkO_Bl_362*	ACE vector, contains P_{thl14}-*yfkO_Bl_362*	this study
pMTL-JH27- P_{thl14}-*ydgI_Bs*	ACE vector, contains P_{thl14}-*ydgI_Bs*	this study
pMTL-MZ1-P_{thl14}-*nfsB_Vv_dm*	ACE vector, contains P_{thl14}-*nfsB_Vv_dm*	this study

Plasmid	Relevant features	Source
pMTL-MZ1-P_{thl14}-*t500_4*	ACE vector, contains P_{thl14}-*t500_4*	this study
pMTL-JH27-P_{thl14}-*nfrA_Bs*	ACE vector, contains P_{thl14}-*nfrA_Bs*	Lissaa Sutherland
pMTL-JH27-P_{thl14}-*ycnD_Bs*	ACE vector, contains P_{thl14}-*ycdN_Bs*	Lissaa Sutherland
pMTL-JH27-P_{fdx}-*nfrA_Bs*	ACE vector, contains P_{fdx}-*nfrA_Bs*	this study
pMTL-JH27-P_{thl14}-*nfrA_Bs*-nat	ACE vector, contains P_{thl14}-*nfrA_Bs*-nat	this study
pET-28b-*yfkO_Bl*	recombinant protein expression vector, expresses *yfko_Bl* with an N-terminal His-tag	this study
	Chapter 4	
pMTL83121-RECas1-*pyrE*-EC	CRISPR/Cas9 vector, targeting the *pyrE* locus and inserting T1-P_{thl14}-*yfkO_Bl*-T_{fdx}	this study
pMTL83121-RECas1-*glnQ1*-EC	CRISPR/Cas9 vector, targeting the *glnQ1* locus and inserting T1-P_{thl14}-*yfkO_Bl*-T_{fdx}	this study
pMTL83121-RECas1-c14550-3'-EC	CRISPR/Cas9 vector, targeting the c14550-3' locus and inserting T1-P_{thl14}-*yfkO_Bl*-T_{fdx}	this study
pMTL83121-RECas1-*treBAR*-EC	CRISPR/Cas9 vector, targeting the *treBAR* locus and inserting T1-P_{thl14}-*yfkO_Bl*-T_{fdx}	this study
pMTL83121-RECas1-*closA-I*-EC	CRISPR/Cas9 vector, targeting the *closA-I* locus and inserting T1-P_{thl14}-*yfkO_Bl*-T_{fdx}	this study
pMTL82151-P_{lac}-OS	expression vector, contains P_{lac}-*tcdR*-P_{tcdB}-*yfkO_Bl*	this study
pMTL-MZ1-P_{lac}-*tcdR*	integration vector, contains P_{lac}-*tcdR*	this study
pMTL82151-P_{tcdB}-*yfkO_Bl*	expression vector, contains P_{tcdB}-*yfkO_Bl*	this study
pMTL-MZ1-P_{lac}-OS	integration vector, contains P_{lac}-*tcdR*-P_{tcdB}-*yfkO_Bl*	this study
pMTL82151-P_{thl14}-*yfkO_Bl*	expression vector, contains P_{thl14}-*yfkO_Bl*; used as a positive control	laboratory collection
pMTL82121-P_{thl}-*yfkO_Bl*	expression vector, contains P_{thl}-*yfkO_Bl*; in-house promoter library	laboratory collection
pMTL82121-P_{abrB}-*yfkO_Bl*	expression vector, contains P_{abrB}-*yfkO_Bl*; in-house promoter library	laboratory collection
pMTL82151-P_{fdx}-*yfkO_Bl*	expression vector, contains P_{fdx}-*yfkO_Bl*; in-house promoter library	laboratory collection
pMTL82151-P_{ptb}-*yfkO_Bl*	expression vector, contains P_{ptb}-*yfkO_Bl*; in-house promoter library	laboratory collection
pMTL82121-P_{glnA}-*yfkO_Bl*	expression vector, contains P_{glnA}-*yfkO_Bl*; in-house promoter library	laboratory collection

Plasmid	Relevant features	Source
pMTL82151-P_{pyrE}-*yfkO_Bl*	expression vector, contains P_{pyrE}-*yfkO_Bl*; in-house promoter library	laboratory collection
pMTL82151-P_{lac}-*yfkO_Bl*	expression vector, contains P_{lac}-*yfkO_Bl*; in-house promoter library	laboratory collection
pMTL-82151-T1-P_{ptb}-OS	expression vector, contains T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl*	this study
pMTL83121-RECas1-*pyrE*-T1-P_{ptb}-*tcdR*	CRISPR/Cas9 vector, targeting the *pyrE* locus and inserting T1-P_{ptb}-*tcdR*	this study
pMTL-MZ1-p15a-T1-P_{ptb}-OS	integration vector, contains T1-P_{ptb}-*tcdR*-P_{tcdB}-*yfkO_Bl*	this study
pMTL83121-RECas1-*glnQ1*-OS	CRISPR/Cas9 vector, targeting the *glnQ1* locus and inserting T1-P_{tcdB}-*yfkO_Bl* -T_{fdx}	this study
pMTL83121-RECas1-*treBAR*-OS	CRISPR/Cas9 vector, targeting the *treBAR* locus and inserting T1-P_{tcdB}-*yfkO_Bl* -T_{fdx}	this study
pMTL83121-RECas1-c14550-3'-OS	CRISPR/Cas9 vector, targeting the c14550-3' locus and inserting T1-P_{tcdB}-*yfkO_Bl* -T_{fdx}	this study
pMTL83121-RECas1-*closA-I* -OS	CRISPR/Cas9 vector, targeting the *closA-I* locus and inserting T1-P_{tcdB}-*yfkO_Bl* -T_{fdx}	this study
Chapter 5		
pMTL83121-P_{tcdB}-*npr-mGMCSF*	expression vector, contains P_{tcdB}-*npr-mGMCSF*	this study
pMTL83121-P_{tcdB}-*lip-mGMCSF*	expression vector, contains P_{tcdB}-*lip-mGMCSF*	this study
pMTL83121-P_{tcdB}-*egla-mGMCSF*	expression vector, contains P_{tcdB}-*egla-mGMCSF*	this study
pMTL83121-P_{tcdB}-*mGMCSF*	expression vector, contains P_{tcdB}-*mGMCSF*; used as a non-secreted control	this study
pMTL83121-P_{tcdB}-*npr-mIL4*	expression vector, contains P_{tcdB}-*npr-mIL4*	this study
pMTL83121-P_{tcdB}-*lip-mIL4*	expression vector, contains P_{tcdB}-*lip-mIL4*	this study
pMTL83121-P_{tcdB}-*egla-mIL4*	expression vector, contains P_{tcdB}-*egla-mIL4*	this study
pMTL83121-P_{tcdB}-*mIL4*	expression vector, contains P_{tcdB}-*mIL4*; used as a non-secreted control	this study
Chapter 6		
pMTL83121-RECas1-*closA-I*	CRISPR/Cas9 vector, targeting and deleting *closA-I*	this study
pMTL83151-RECas1-c14780	CRISPR/Cas9 vector, targeting and deleting c14780	this study
pMTL83151-RECas1-c16380	CRISPR/Cas9 vector, targeting and deleting c16380	this study

Plasmid	Relevant features	Source
pMTL83151-RECas1-c01750	CRISPR/Cas9 vector, targeting and deleting c01750	this study
pMTL83151-RECas1-c31080	CRISPR/Cas9 vector, targeting and deleting c31080	this study
pMTL83151-RECas1-c29800	CRISPR/Cas9 vector, targeting and deleting c29800	this study
pMTL007C-E2-*closA*-101s	ClosTron vector, contains Csp-*closA*-101s intron-targeting sequence	this study
pMTL007C-E2-*closC*-749a	ClosTron vector, contains Csp-*closC*-749a intron-targeting sequence	this study

IV. Representative plasmid maps

Gram-positive replicon in orange; Gram-negative replicon in blue; Standard plasmid modules in white; lacZα-MCS in black; Application-specific features in grey, red and green; Relevant primer biding sites are shown in purple and important restriction sites are annotated.

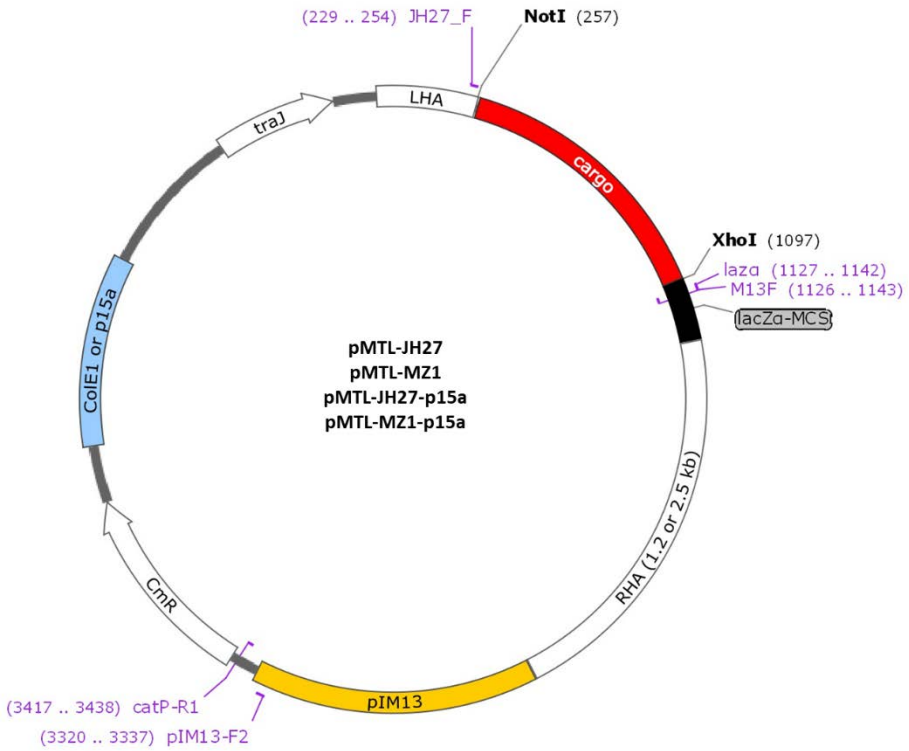


Figure A1 Plasmid for genomic integrations via ACE

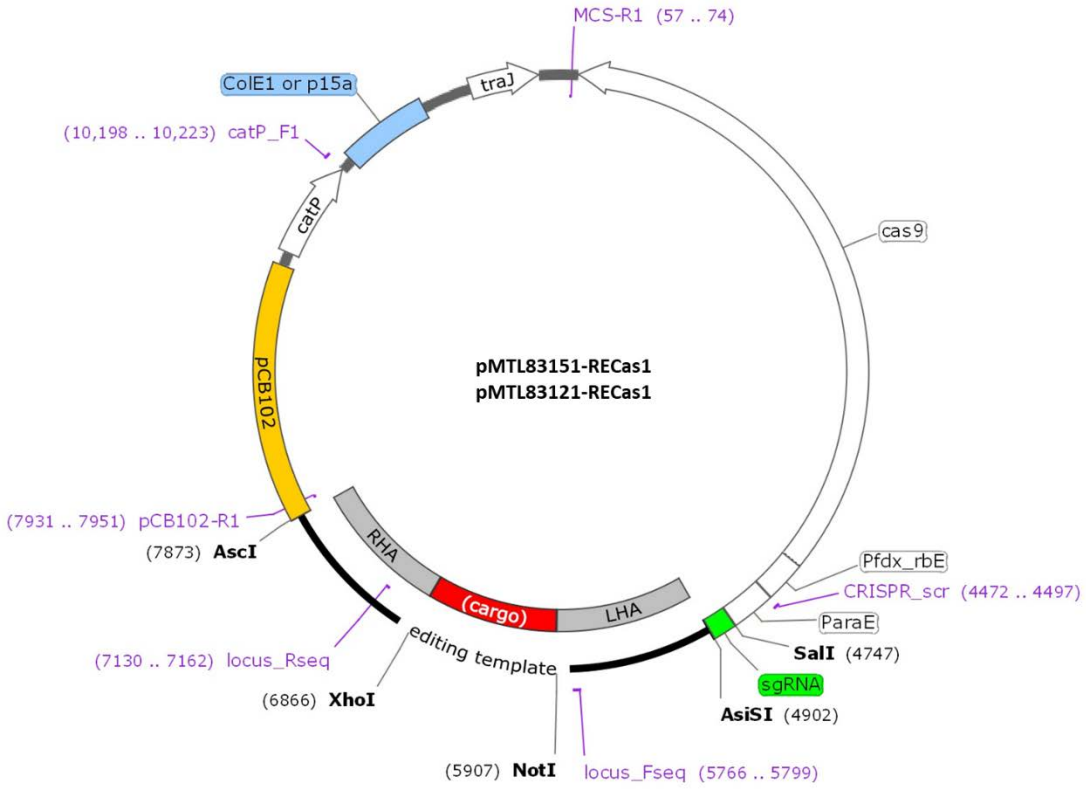


Figure A2 Plasmid for CRISPR/Cas9-mediated genome editing

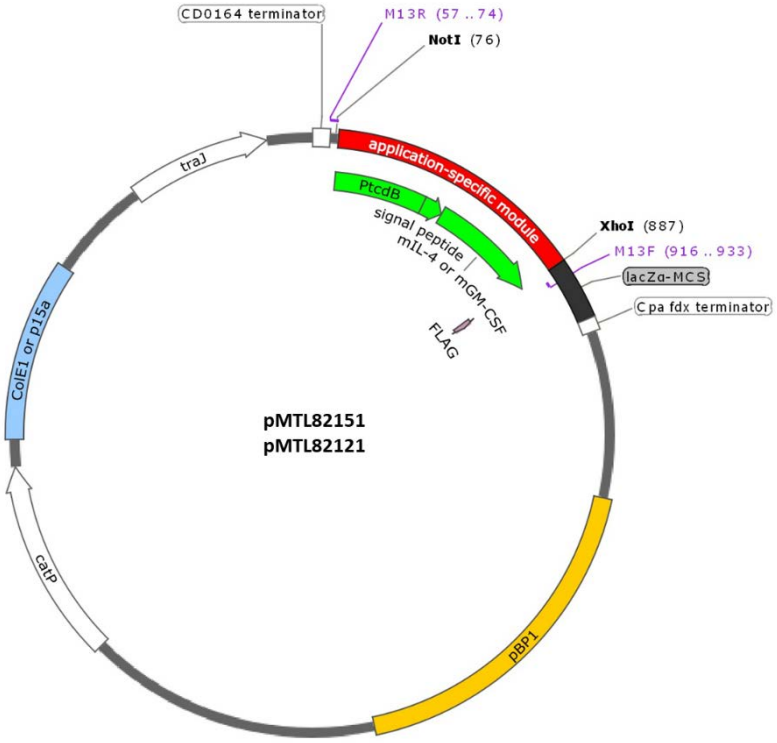


Figure A3 Expression plasmid

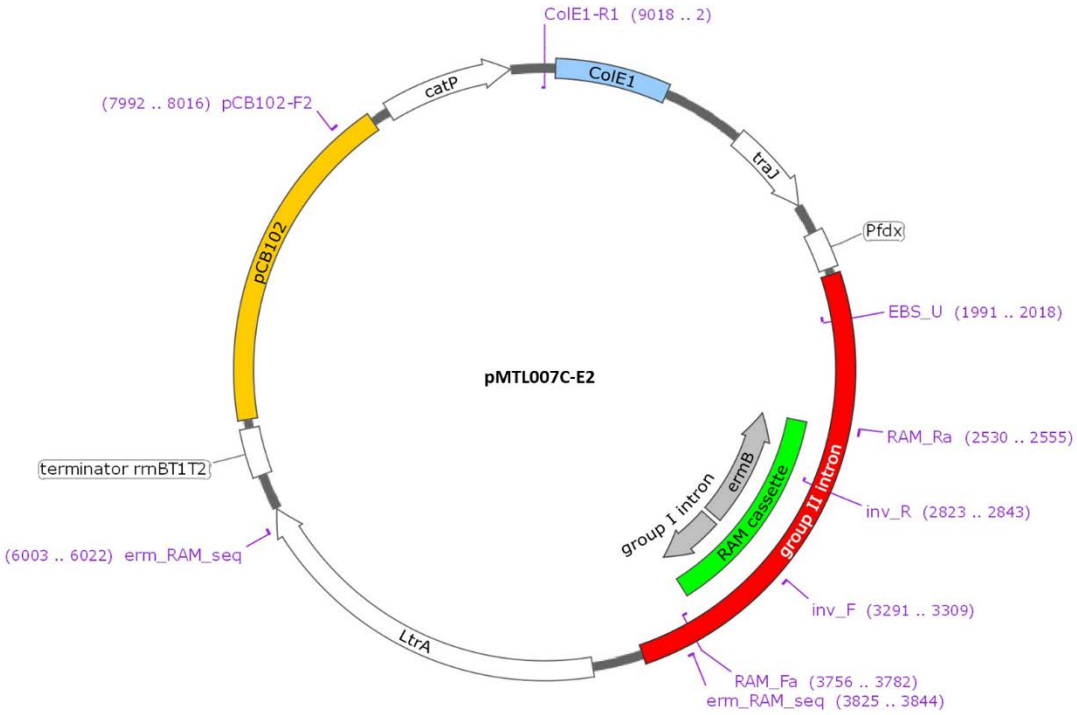


Figure A4 Plasmid for ClosTron-mediated mutagenesis

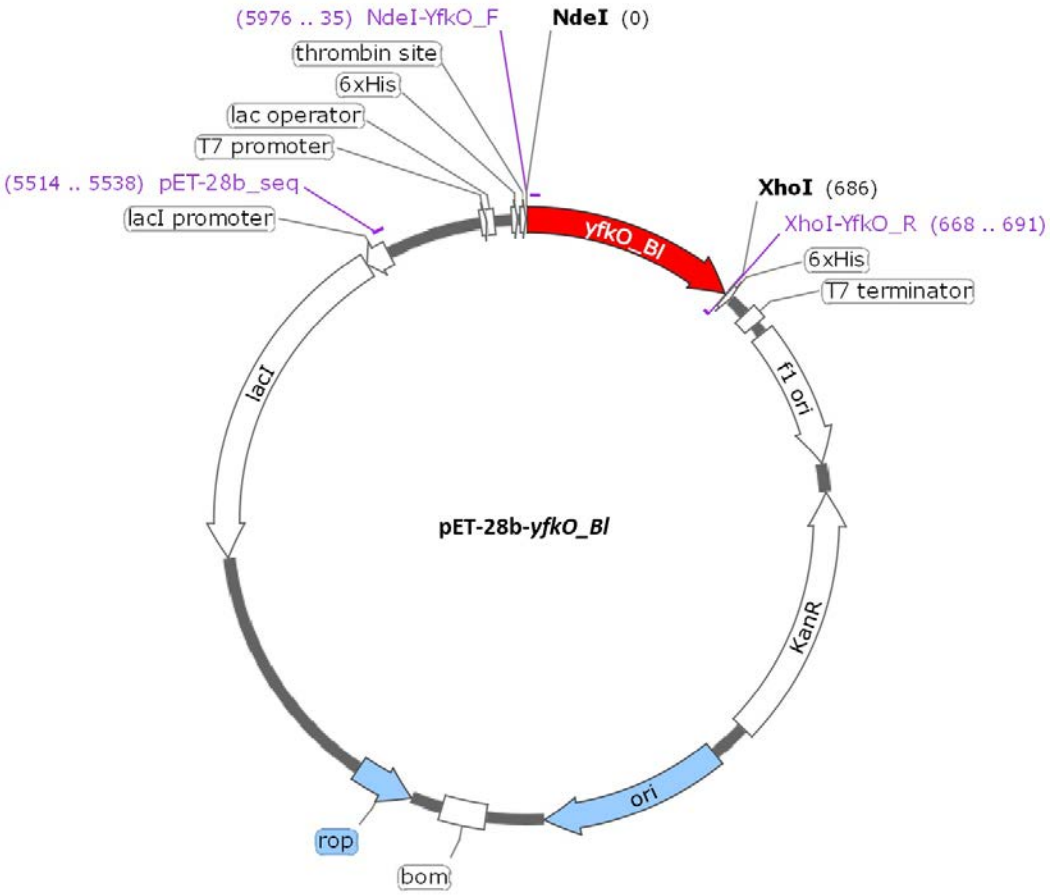


Figure A5 Plasmid for the purification of YfkO_Bl

V. Codon Optimisation

Below is a section from the PhD thesis of Dr Aleksandra Kubiak (pp. 122-132 of the original document) describing the adaptation of ATUM's proprietary codon gene design (codon optimisation) algorithm for *C. sporogenes* NCIMB 10696.

4.2.4 Codon optimisation of the nmeNTR gene

Codon usage bias refers to differences in the frequency of occurrence of synonymous codons in coding DNA. Codon (triplets) encodes a specific amino acid residue in a polypeptide chain or for termination of translation. The degeneracy of the genetic code allows amino acids to be encoded by more than one codon. However, different organisms show particular preferences for one of the several codons. It is thought that optimal codons help to achieve faster translation rates and high accuracy. In this respect, codon usage is considered to be an important contributor to setting the level of gene expression. In the two highly studied microbial species, *E. coli* and *B. subtilis*, highly expressed genes are transcribed from optimal codons that are recognised more efficiently by the most abundant tRNA[2]. It is believed that many factors contribute to codon usage bias, G+C composition, amino acids conservation or RNA stability.

In all the clostridial genomes sequenced so far it was found that the percentage of G+C content varies from 29 to 31. Further examination showed that highly expressed genes with around 13-18 % GC content were related to significant A+T predominance in terms of codon bias and A/T „silent" third position in the triplet codon[3-6]. From these studies it can be concluded that for clostridial genomes the strong mutational bias towards A and T could be the main factor influencing codon usage.

This is why, one of the major interests was to examine codon usage in respect to nitroreductase gene from *N. meningitidis* (*nmeNTR*) and expressed in *C. sporogenes*.

DNA2.0 has developed a technology to identify and quantify non-coding variables within structural genes and how to combine these variables to generate algorithms to predictably design genes for heterologous protein over expression in *C. sporogenes* and other clostridial species. The previously constructed, integrated and tested version of NmeNTR was manually codon optimised according to the most frequent codon usage in clostridia strains. This however was not the most optimal procedure and for that reason is the subject of much more thorough and detailed investigation, using DNA2.0™ technology.

The NmeNTR variants design and synthesis have been performed by DNA2.0™ and their cloning and testing was a subject of the investigation in this work.

4.2.4.1 Design of experiment, cloning and expression of *nmeNTR* gene in *C. sporogenes*

A first set of 48 systematically varied variants of the *nmeNTR* gene was designed by DNA 2.0™ using Gene Designer software and GeneGPS™ technology (schematically represented in Figure 4.9).

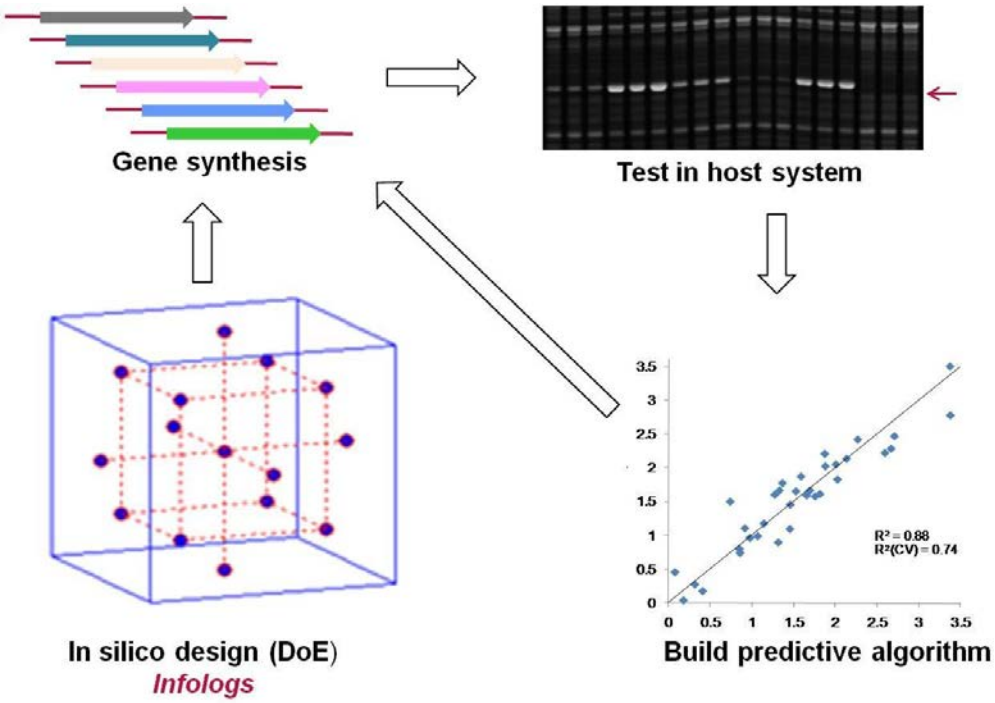


Figure 4.9: Schematic representation of the DNA2.0™ methodology and its application for the codon optimisation study.

Gene variants were designed by back-translating the protein sequence using a Monte Carlo repeated random sampling algorithm[7]. This algorithm selects a codon for each position at a probability defined in a codon frequency lookup table. The amino acid sequence has been kept constant and the coding DNA sequence is systematically varied. The codon bias starting point was that observed in the *C. sporogenes* genome. Variation includes bias of the codons for each of the amino acids with multiple codon options, rare codon threshold, and mRNA secondary structures. Design of Experiment (DoE) methods were used to maximize independent sampling of the variables. Known or strongly suspected deleterious motifs, such as splice sites and RNA destabilizing sequences, were avoided throughout the genes. Strong mRNA structures, as predicted by RNA folding algorithms were eliminated.

Each variant was synthesized as an independent and unique sequence and cloned into plasmid pMTL84151:Pcs*fdx* between *Nde*I and *Nhe*I restriction sites. Additionally a control of a native, non-codon optimised version of *nmeNTR* gene, named Q9K022 (derived from *N. meningitides* MC58; Accession number: AAF41217.1) was prepared. To do this, the Q9K022 enzyme previously cloned into pCR2.1 expression vector (CRG group) was digested and inserted into prepared pMTL84151:Pcs*fdx* vector via *Nde*I and *Xho*I sites. All NmeNTR variants and Q9K022 control were conjugated into *C. sporogenes* prior to transformation in *E. coli*. The variables named: 25, 27 and 29 proved difficult for the expression in *E. coli* host and were omitted from the study.

Cell lysates were prepared from 3 individual transformants and menadione assays were conducted in triplicates to estimate level of nitroreductase produced (Figure 4.10).

From Figure 4.10 it appears clear that all tested constructs exhibit a broad spectrum of nitroreductase expression levels in *C. sporogenes*, ranging from 0.31 to 3.98 U per mg. Many of them showed a significant improvement compared to the Q9K022 control (non-optimised NmeNTR). All collected expression data was transferred to DNA2.0™ for further analysis which involved the application, selection and multivariate regression to identify significant variables and design a second set of gene variants.

Accordingly, a second set of 9 NmeNTR variables were designed and synthesised, which consisted of already assessed codons hybrids preferentially used by *C. sporogenes* (Table 4.5).

Table 4.5: The list of codon variants selected and assayed in the second round of codon optimisation study.

Variable	Codon combination
85	Hyb_11_48_48
86	Hyb_48_11_48
87	Hyb_48_48_11
88	Hyb_17_20_20
89	Hyb_20_17_20
90	Hyb_20_20_17
91	Hyb_46_4_4
92	Hyb_4_46_4
93	Hyb_4_4_46

Following the steps as previously described, the synthesized variables were cloned into the pMTL84153 expression vector, conjugated into *C. sporogenes.* The level of nitroreductase produced in the resultant transconjugants were assessed using the menadione reductase assay. The results are presented in Figure 4.11.

From Figure 4.11 it can be concluded that 5 out of 9 tested variables of hybridised codons showed improvement in *nmeNTR* gene expression when compared to the native version of the same gene. The decreased expression levels exhibited by other variants also provided valuable information and allowed DNA2.0™ to further optimise and refine the algorithm for clostridial codon optimisation.

All data obtained throughout this study was used by DNA2.0™ to design an initial model of the algorithm for the enzyme expression of desired protein in clostridial hosts. By using the technologies developed by DNA2.0, it is expected to establish a systematic understanding of the factors controlling protein expression in the *Clostridium* and implement it in number of future studies on enzyme expression.

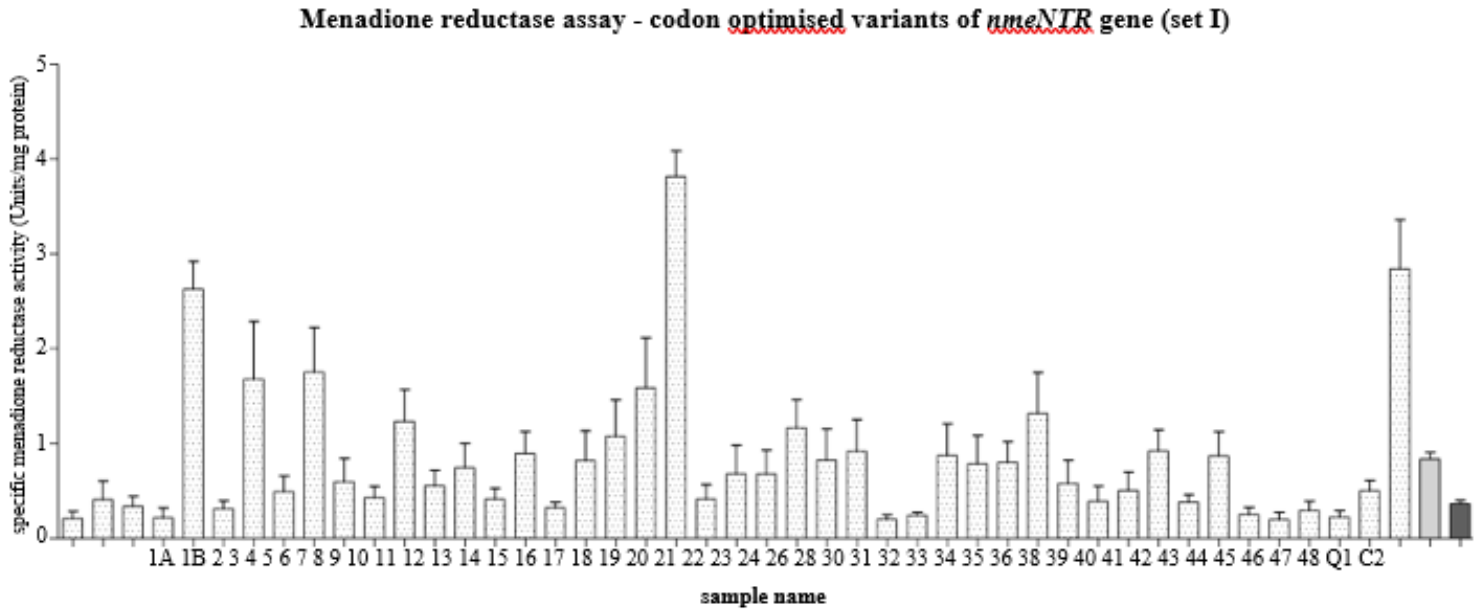


Figure 4.10: Menadione reductase assay results carried out on *C. sporogenes* cell lysates harbouring plasmids with codon optimised NmeNTR enzyme variants (the first round of optimisation study).

From 1A to 48: NmeNTR codon variants cloned into pMTL84153(Pcsp*fdx*) vector, Q1: pMTL84153:Q9K022 (native version of NmeNTR); C2: pMTL84153 empty vector control; all constructs were expressed in *C. sporogenes* host). Soluble fractions of lysates were prepared as described in 2.5.14 and obtained from 7-hour time point from the start of culture incubation. Activities were normalised to total protein concentrations, determined using the Bradford protein assay. Activities are expressed in units per mg. The data represent the average of six independent experiments and error bars indicate the standard errors of the means. Statistical analysis was not performed for this experiment.

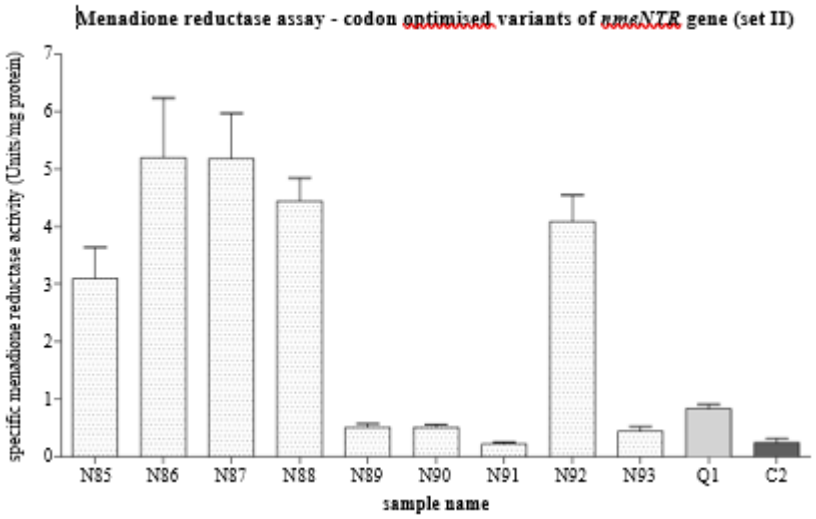


Figure 4.11: Menadione reductase assay results carried out on *C. sporogenes* cell lysates harbouring plasmids with codon optimised NmeNTR enzyme variants (the second round of optimisation study).

From 1A to 48: NmeNTR codon variants cloned into pMTL84153(Pcsp*fdx*) vector, Q1: pMTL84153:Q9K022 (native version of NmeNTR); C2: pMTL84153 empty vector control; all constructs were expressed in *C. sporogenes* host. Soluble fractions of lysates were prepared as described in 2.5.14 and obtained from 7-hour time point from the start of culture incubation. Activities were normalised to total protein concentrations, determined using the Bradford protein assay. Activities are expressed in units per mg. The data represent the average of six independent experiments and error bars indicate the standard errors of the means. Statistical analysis was not performed for this experiment.

4.2.4.2 Building a predictive model for the expression of *C. sporogenes*

By applying the DNA2.0™ multivariate regression analysis technology, (schematically presented in Figure 4.12) the company searched the parameter space for correlation with expression levels of the *nmeNTR* gene. This regression system selected the optimal weights for each variable to create a model for the *C. sporogenes* strain.

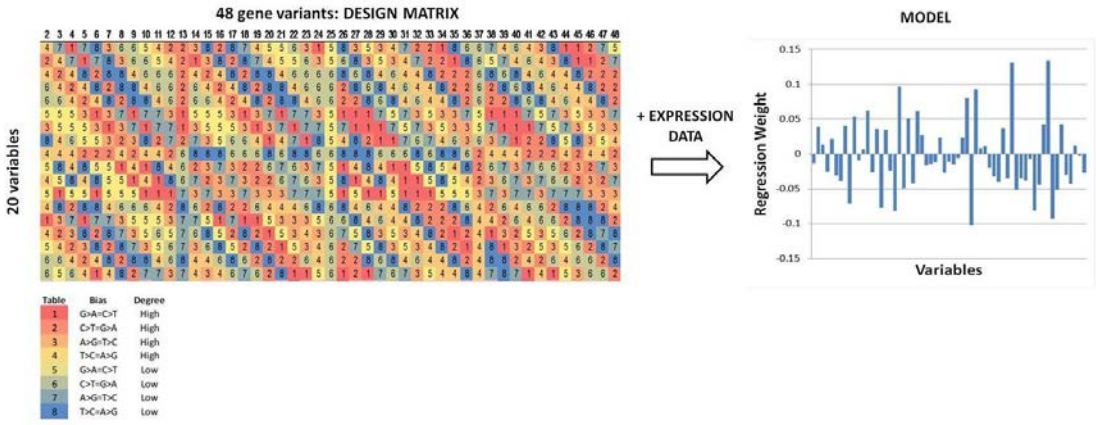


Figure 4.12: The schematic representation of the codon optimisation data analysis (PLS) carried out by DNA2.0™.

DNA2.0™ designed and synthesized a pool of 48 gene variants. This design was validated according to expression data and streamlined to predict an expression model.

In the final stage of the analysis, the protein expression levels were checked for the correlation with the Codon Adaptation Index (CAI). As a result, it was concluded that expression level of the *nmeNTR* gene variants in *C. sporogenes* do not correlate with the CAI for the host, when the evaluation is based on genomic codon frequencies. Partial Least Squares (PLS) analysis of the same data can, however, identify a significant correlation between gene expression and codon usage frequencies (Figure 4.13). This model for *C. sporogenes* created the bio- mathematical over-expression algorithm that will be used to synthesise the final codon optimised version of the *nmeNTR* gene.

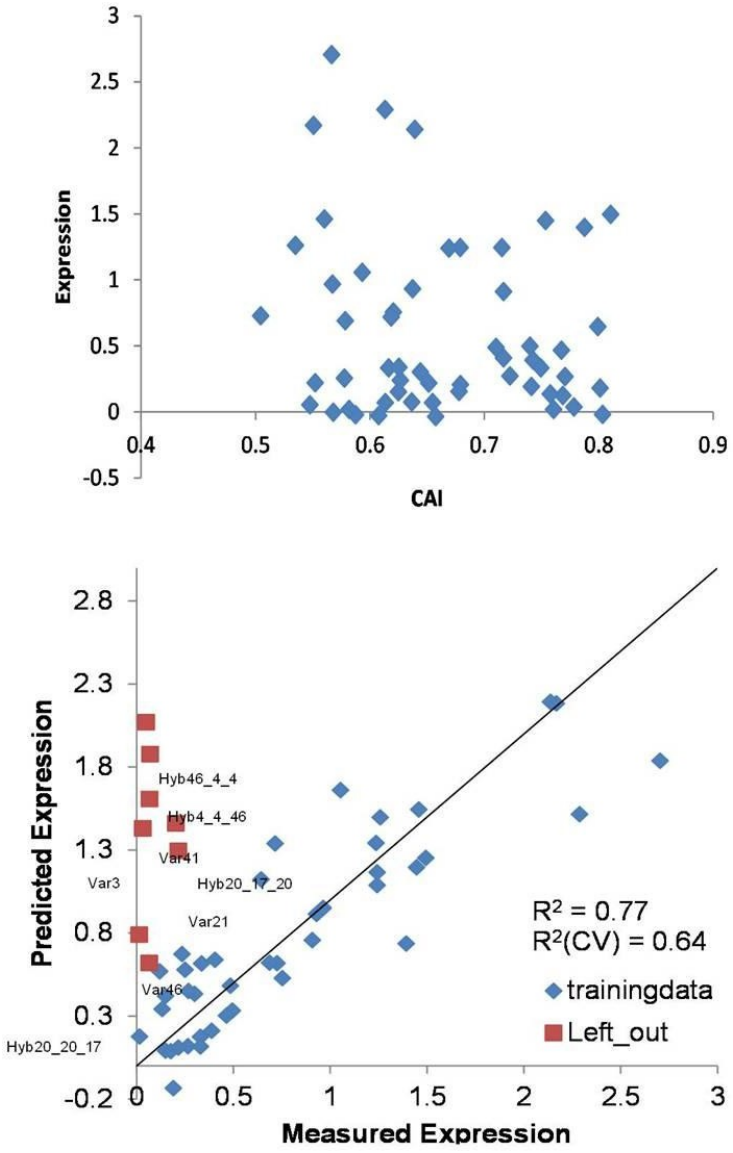


Figure 4.13: Predicted model for codon usage in *C. sporogenes.*

Upper panel – expression levels compared with CAI; Lower panel - correlation of expression levels and CAI when PLS analysis adapted. Blue diamonds indicate the 42 gene training set used to create model. Red squares are variants which were not used for model building due to undetectable expression. R^2(CV) indicates the correlation coefficient for the fit of the model in cross-validation.

4.2.4.3 Codon optimisation study for other clostridia

In order to evaluate whether the designed algorithm would allow the over-expression of any desired enzyme and by universally applicable across all clostridia, it was of an interest to carry out the similar expression study in other species.

A total of 8 previously studied variants were selected and the activity levels of

nmeNTR gene measured in *Clostridium acetobutylicum, Clostridium difficile, Clostridium beijerinckii,* and *Clostridium ljungdahlii*. The cloning procedures were followed as previously described and conjugation or transformation procedures applied as menadione reductase assays and sample preparation was carried out by Ying Zhang, Hengzheng Wang and Louise Sewell (CRG). The manadione activity results are shown in Figure 4.14 (A –D).

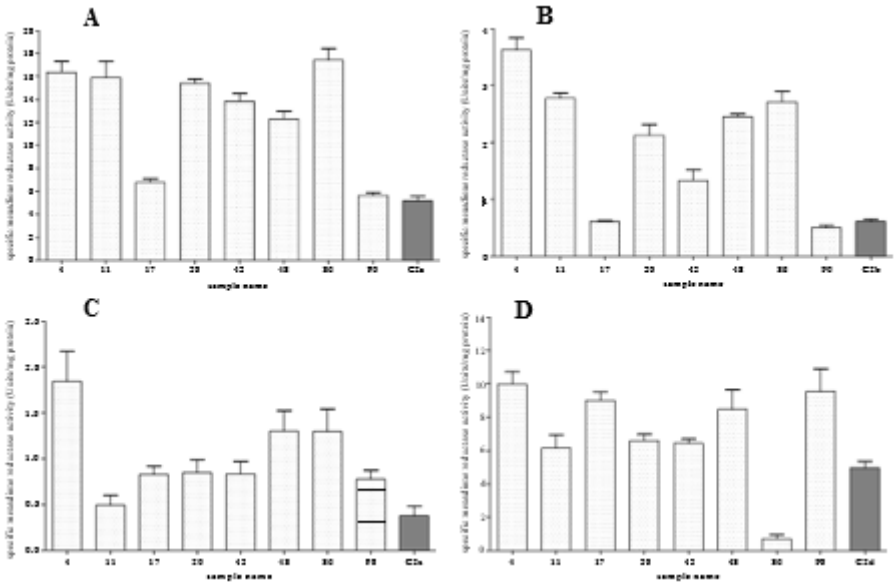


Figure 4.14: Menadione reductase assay results performed on selected plasmid-based variants of NmeNTR enzyme expressed in: A) *C. difficile*; B) *C. acetobutylicum*;, C) *C. ljungdahlii* and D) *C. beijerinckii*.

C2a, C2b, C2c and C2d: pMTL84153 negative controls conjugated (or transformed) into subsequent strains. Soluble fractions of lysates were prepared as described in 2.5.14 and 2.5.15 and obtained from 7-hour time point from the start of culture incubation. Activities were normalised to total protein concentrations, determined using the Bradford protein assay. Activities are expressed in units per mg. The data represent the average of nine independent experiments and error bars indicate the standard errors of the means. Statistical analysis was not performed for this experiment.

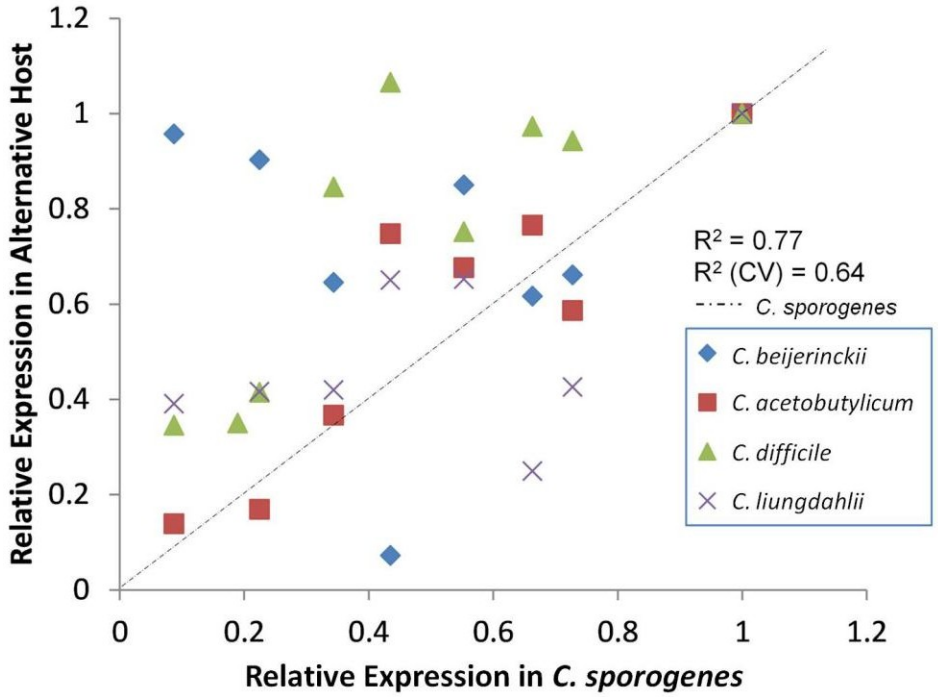


Figure 4.15: Comparative summary of the expression levels in different clostridia.

Expression levels of nitroreductase gene expressed in 4 different clostridia were evaluated by DNA2.0™. Figure indicates the correlation of all data sets with the initial predictive model built for *C. sporogenes* (dotted line). All strains, with exception of *C. beijerinckii* (blue diamonds), were found to correlate with the algorithm. R^2(CV) indicates the correlation coefficient for the fit of the model in cross-validation.

All data gathered from the expression studies in the four other *Clostridium* species were analysed by DNA2.0™ and compared to the initial algorithm model derived for *C. sporogenes*. This analysis revealed that with the exception of *C. beijerinckii*, the expression data of the variants correlated with the algorithm. This indicated that codon preferences in these clostridial species mirror that of *C. sporogenes* (Figure 4.15). At this stage, however, the sample size is quite limited, and analysis of further gene variants is required before definitive conclusions can be drawn.

VI. Codon-optimised nucleotide sequences

Table A4 Nucleotide sequences of codon-optimised genes employed in this study

Gene	DNA sequence	Chapter
yfkO_Bl	ATGACAGAACAAAGCAAGAAACAAGAAATACTAGATGCTTTTCAGTTTAGACATGCAACAAAAGAGTTTGATCCAGATAGGAAAATATCAGATG AAGATTTCCAATTCATACTTGAAGCTGGTAGATTATCACCATCTTCTGTAGGTTTAGAACCTTGGCAATTTGTAGTTGTACAAAATAAAGAACTAA GAGAAAAGTTAAGACAAGTTTCTTGGGGAGCTCAAGGACAATTACCAACTGCATCTCATTTTGTACTTTTATTAGGTAGAACAGCTAAAGAAATG AGAAGAGATAGTGGATATGTAGCTGATCAATTGAAACATGTTAAGAAAATGCCTGAAGATATAATAGAAAATATGCTAAAAGAAGATGGAGTTT TAGAAAGTTTTCAAGATGGTGATTTTCACTTGTATGAATCAGATAGAGCAATGTTTGATTGGGTTAGTAAACAAACTTACATTGCATTAGCAAATA TGATGACTGCAGCTGCTTTAATAGGAATTGATTCTTGTCCAATTGAAGGATTTAACTATGACAAAGTACATGACATACTTGAGAAAGAAGGTGTA TTAGAAGATGGAAGATTTGACATATCAGTAATGGCAGCTTTTGGATATAGAGTAAAAGAACCTAGACCCAAAACTAGAAGAGCTTTAGATCAAA TTGTAAAATGGGTTGAATAA	3-4
yfkO_Bl_362	ATGACAGAACAAAGCAAGAAACAAGAAATACTAGATGCTTTTCAGTTTAGACATGCAACAAAAGAGTTTGATCCAGATAGGAAAATATCAGATG AAGATTTCCAATTCATACTTGAAGCTGGTAGATTATCACCATCTTCTGTAGGTTTAGAACCTTGGCAATTTGTAGTTGTACAAAATAAAGAACTAA GAGAAAAGTTAAGACAAGTTTCTTGGGGAGCTCAAGGACAATTACCAACTGCATCTCATTTAGTACTTTTATTAGGTAGAACAGCTAAAGAAATG AGAAGAGATAGTGGATATGTAGCTGATCAATTGAAACATGTTAAGAAAATTCCTGAAGATATAATAGAAAATATGCTAAAAGAAGATGGAGTTT TAGAAAGTTTTCAAGATGGTGATTTTCACTTGTATGAATCAGATAGAGCAATGTTTGATTGGGTTAGTAAACAAACTTACATTGCATTAGCAAATA TGATGACTGCAGCTGCTTTAATAGGAATTGATTCTTGTCCAATTGAAGGATTTAACTATGACAAAGTACATGACATACTTGAGAAAGAAGGTGTA TTAGAAGATGGAAGATTTGACATATCAGTAATGGCAGCTTTTGGATATAGAGTAAAAGAACCTAGACCCAAAACTAGAAGAGCTTTAGATCAAA TTGTAAAATGGGTTGAATAA	3

Gene	DNA sequence	Chapter
ydgI_Bs	ATGATTAAGACTAATGACTTTATGGAAATAATGAAAGGTAGACGATCTATTAGAAATTATGATCCCACTGTCAAAATTAGTAAGGAAGAAATGAC AGAGATATTAGAAGAAGCAACTACCGCTCCTTCCAGCGTAAACGCTCAACCATGGCGTTTTCTTGTAATAGACTCTCCTGAGGGTAAAGAAAAGT TAGCGCCATTAGCTTCATTTAACCAAACTCAAGTTACCACTAGTAGCGCAGTTATAGCGGTTTTTGCAGATATGAACAATGGTGACTATCTAGAAG AAATCTACTCTAAAGCCGTTGAACTTGGTTATATGCCTCAAGAAGTAAAAGATAGACAAATAGCAGCATTAACCGCACATTTTGAAAAGTTGCCA GCTCAGGTGAATAGAGAAACTATTTTAATAGATGGGGGACTTGTTTCAATGCAGTTAATGCTAACTGCTAGAGCTCACGGATATGATACCAATCC GATTGGTGGATATGATAAAGAAAACATAGCTGAAACCTTCGGCTTAGACAAAGAAAGATATGTGCCCGTCATGTTATTGTCAATTGGTAAAGCT GCGGATGAAGGGTATGCTTCCTATAGATTACCAATCGATAAGATTGCCGAATGGAAATAG	3
nfsB_Vv_dm	ATGACTATTGTTCAAGCTGCCCAATCAAGATATTCAACTAAGGCTTTTGATGCGTCAAGAAAACTACCAGAAGAAAAAGTCGCAGCCGTCAAAGA ATTGATTCGTATGTCCGCTAGTAGCGTTAATTCACAACCTTGGCATTTCATAGTAGCATCTAGTGAAGAAGGCAAGGCAAGAATTGCCAAGGCTA CCCAGGGCGGATTTGCAGCTAACGAAAGAAAAATATTAGATGCATCCCATGTAGTAGTGTTTTGTGCAAAAACTGCTATAGATGAAGCTTATTTA TTAGATTTACTTGAATCTGAGGACAAAGACGGTAGATATGCTGATGTGGAAGCGAAAAATGGTATGCACGCGGGAAGATCATTCTTTGTTAACA TGCATAGATTTGACTTAAAAGATGCACATCATTGGATGGAAAAACAAGTATATCTAAATGTTGGTACCTTGTTACTTGGGGCTAGCGCTATGGAA ATAGATGCAGTACCAATCGAAGGGTTTGACGCTAAAGTCCTTGATGAAGAATTTGGTTTAAGAGAAAAGGGTTTTACATCAGTTGTTATAGTTCC CTTAGGTTATCACAGCGAAGATGATTTCAACGCGAAATTACCTAAGTCCCGTTGGTCTGCTGAAACTGTGTTTACCGAGATTTAG	3
t500_4	ATGACTATTGTTCAAGCTGCCCAATCAAGATATTCAACTAAGGCTTTTGATGCGTCAAGAAAACTACCAGAAGAAAAAGTCGCAGCCGTCAAAGA ATTGATTCGTATGTCCGCTAGTGCCACAAATTCACAACCTTGGCATTTCATAGTAGCATCTAGTGAAGAAGGCAAGGCAAGAATTGCCAAGGCTA CCCAGGGCGGATTTGCAGCTCAAGAAAGAAAAATATTAGATGCATCCCATGTAGTAGTGTTTTGTGCAAAAACTGCTATAGATGAAGCTTATTTA TTAGATTTACTTGAATCTGAGGACAAAGACGGTAGATATGCTGATGTGGAAGCGAAAAATGGTATGCACGCGGGAAGATCATTCGTAGTTAACA TGCATAGATTTGACTTAAAAGATGCACATCATTGGATGGAAAAACAAGTATATCTAAATGTTGGTACCTTGTTACTTGGGGCTAGCGCTATGGAA ATAGATGCAGTACCAATCGAAGGGTTTGACGCTAAAGTCCTTGATGAAGAATTTGGTTTAAGAGAAAAGGGTTTTACATCAGTTGTTATAGTTCC CTTAGGTTATCACAGCGAAGATGATTTCAACGCGAAATTACCTAAGTCCCGTTGGTCTGCTGAAACTGTGTTTACCGAGATTTAG	3

Gene	DNA sequence	Chapter
nfrA_Bs	ATGAACAATACCATAGAAACTATTCTTAATCATAGATCAATAAGATCATTTACCGACCAACTATTGACTGCTGAAGAAATTGATATCTTAGTGAAA TCTGCGCAAGCTGCTAGTACCTCTAGTTATGTACAGGCTTATAGCATAATAGGGGTATCTGATCCCGAGAAAAAGCGAGAATTAAGTGTATTGGC TGGTAATCAACCTTATGTAGAAAACAACGGTCATTTCTTCGTCTTTTGTGCAGATCTACATAGACATCAAAAACTTGCAGAAGAAAAGGGTGAAA ACATATCAGAACTATTAGAAAATACAGAAATGTTTATGGTTTCCTTGATTGACGCAGCATTAGCCGCACAGAATATGTCAGTTGCTGCTGAATCTA TGGGTTTAGGTATATGCTATATTGGTGGGATTCGTAACGAATTAGATAAAGTTACCGAAGTGTTACAAACTCCAGATCATGTCTTACCACTTTTTG GCTTAGCGGTTGGACACCCGGCTAACTTATCCGGTAAAAAGCCCAGATTACCTAAACAAGCCGTCTATCACGAGAATACTTATAACGTTAATGCA GATGATTTTAGAGATACTATGAATGCTTATGATCAAACTATATCAGACTATTATAGAGAAAGAACTAATGGACAGAGAGAAGAAACTTGGAGCG ACCAAATTCTTAACTTTATGAAACAAAAGCCAAGAACCTACTTAAATGATTATGTTAAAGAAAAGGGATTTAACAAAAATTAA	3
ycnD_Bs	ATGAATGAAGTGATTAAGTCATTAACTGATCATAGAAGTATTAGATCCTATACCGATGAACCCGTTGCAAAAGAACAATTAGATCAAATCATACA AGCTGTTCAATCAGCTCCAACTTCTATAAACGGACAACAAGTAACTGTTATTACCGTTCAGGATAAGGAACGAAAGAAAAAGATTTCCGAATTAT CTGGTGGCCAACCTTGGATTGACCAGGCACCAGTATTTCTTCTATTTTGTGCTGATTTCAACAGAGCCAAAATTGCGCTTGAAGATTTGAATGACA TTAAAATGGAAATAACCAATGGGTTAGAATCAGTGTTGGTCGGTGCAGTAGACGCGGGTATAGCTTTAGGTACTGCAACAGCCGCTGCAGAATC TTTAGGTCTTGGTACTGTTCCAATCGGAGCAGTCAGAGGAAACCCACAAGAACTTATAGAATTACTAGAATTACCGAAATATGTATTTCCTGTTAG TGGTTTAGTTATTGGGCACCCCGCTGACAGAAGTGCTAAGAAACCTAGATTACCCCAAGAAGCTGTAAACTATCAAGAAACCTATCTAAACCAAG ATGAATTAACCAGCTATATACAGGCTTATGATGAGAAAATGTCAGAGTACATGAATAAACGTACTAATGGTAAAGAAACTAGAAATTGGAGCCA AGGTATAGCATCCTATTATGAAAGATTGTATTATCCACATATAAGAGAAATGTTAGAAAAGCAAGGGTTTAAAGTCGAAAAATAA	3
mGM-CSF	GCGCCTACAAGGAGCCCAATAACTGTAACAAGACCTTGGAAACATGTTGAAGCTATTAAGGAAGCACTTAATTTACTAGATGACATGCCTGTAAC TTTAAATGAAGAGGTAGAAGTTGTAAGTAACGAATTTTCATTTAAAAAGTTAACATGTGTACAAACTAGATTAAAAATTTTTGAACAAGGACTTA GAGGTAATTTTACTAAATTGAAAGGAGCTTTAAATATGACTGCTTCTTATTACCAAACATATTGTCCACCTACTCCAGAAACCGATTGCGAAACTC AAGTTACTACTTATGCAGATTTTATAGATTCTCTAAAAACTTTCTTAACAGATATACCCTTTGAATGTAAAAAACCAGGTCAAAAAGATTATAAAG ATGATGATGATAAATAA	5
mIL-4	CATATTCATGGATGTGATAAGAATCATCTAAGGGAAATTATAGGAATACTTAACGAAGTAACTGGAGAGGGTACTCCATGTACAGAAATGGATG TTCCTAATGTATTAACAGCTACAAAAAATACTACAGAATCAGAATTAGTATGTAGAGCAAGTAAAGTATTAAGAATATTTTATTTAAAACACGGTA AAACTCCATGTTTGAAAAAGAATTCTTCAGTTTTAATGGAATTACAAAGACTATTTAGAGCTTTCAGATGTCTTGACTCTTCAATAAGTTGCACTAT GAATGAATCTAAATCAACCAGCTTAAAAGATTTTTTAGAAAGTTTAAAATCCATTATGCAAATGGATTATTCTGATTATAAAGATGATGATGATAA ATAA	5

VII. P_{thl14} promoter/RBS variant for gene expression in *C. sporogenes*

Below is a section from the PhD thesis of Dr Aleksandra Kubiak (pp. 119-122 of the original document) describing the characterisation of the Pthl14 expression signal in *C. sporogenes*. The nucleotide sequence of P_{thl14} is highlighted in red.

4.2.3 Expression study on promoter-RBS variants

In order to investigate the effects of regulation the transcription and translation processes, the study of promoter library and ribosome binding sites were combined and a set of 8 new expression variants was created. All variants consist of experimentally tested and selected promoters (P*pyrE*, P*tRNA*, P*thl* and P*fdx*), RBS (R13 and R14) sequences and the first 269 base pairs of *nmeNTR* gene, as shown in Table 4.4. In order to avoid the addition of restriction sites before the ATG start codon, all 8 sequences were synthesized by the company Biomatic. The sequences made, possessed *Not*I and *Bsm*I restriction sites, which allow their cloning into digested pMTL84153:NmeNTR vector as schematically presented in Figure 4.7.

Table 4.4: The promoter sequences used in the promoter library study

Variant name	Sequence (5'-3'); (underlined: *Not*I/ *Bsm*I restriction sites)
pyrE13	GCGGCCGCCCAATAAAACCCAATCTTTTGATTGGGTTTTTATTTTTGTGTAAATAATTAAT ATTTTAAGGTATAGTTGCTTCTATTTTTCAGAAGTATATATGATATAATTATAAAGGAAAT TAATCAGTAAATGCTG**AAAGGAGGT**GAGAAAT:NmeNTRGAATGCA
pyrE14	GCGGCCGCCCAATAAAACCCAATCTTTTGATTGGGTTTTTATTTTTGTGTAAATAATTAAT ATTTTAAGGTATAGTTGCTTCTATTTTTCAGAAGTATATATGATATAATTATAAAGGAAAT TAATCAGTAATAAAAT**AAAGGAGGT**TTATATAT:NmeNTRGAATGCA
tRNA13	GCGGCCGCCCTAATATATTTATATATAATACAATACGTAAAGCAACTTTTAGGGATTTACA CAAAAAGACAATATTAATTATTTCGTTATAATTATCAATAAGATGTAAACTGTTGAAAACC CTTGAAATAAACCATTGACAGTCGAAGATGAGAGTAATATAATAAATCTGCAATGGGAAAT GCTG**AAAGGAGGT**GAGAAAT:NmeNTRGAATGCA
tRNA14	GCGGCCGCCCTAATATATTTATATATAATACAATACGTAAAGCAACTTTTAGGGATTTACA CAAAAAGACAATATTAATTATTTCGTTATAATTATCAATAAGATGTAAACTGTTGAAAACC CTTGAAATAAACCATTGACAGTCGAAGATGAGAGTAATATAATAAATCTGCAATGGGAATA AAAT**AAAGGAGGT**TTATATAT:NmeNTRGAATGCA
thl13	GCGGCCGCGGGGGTTAGCATATGCATAAGTTTAATTTTTTTGTTAAAAAATATTAAACTTT GTGTTTTTTTTTAACAAAATATATTGATAAAAATAATAATAGTGGGTATAATTAAGTTGTTAG AGAAAACGTAAATGCTG**AAAGGAGGT**GAGAAAT:NmeNTRGAATGCA
thl14	GCGGCCGCGGGGGTTAGCATATGCATAAGTTTAATTTTTTTGTTAAAAAATATTAAACTTT GTGTTTTTTTTTAACAAAATATATTGATAAAAATAATAATAGTGGGTATAATTAAGTTGTTA GAGAAAACGTAATAAAAT**AAAGGAGGT**TTATATAT:NmeNTRGAATGCA
fdx13	GCGGCCGCGTGTAGTAGCCTGTGAAATAAGTAAGGAAAAAAAAGAAGTAAGTGTTATATAT GATGATTATTTTGTAGATGTAGATAGGATAATAGAATCCATAGAAAATATAGGTTATACAG TTATATAAAAATTACTTTAAAAATTAATAAAAACATGGTAAAATATAAATCGTATAAAGTT GTGAAATGCTG**AAAGGAGGT**GAGAAAT:NmeNTRGAATGCA
fdx14	GCGGCCGCGTGTAGTAGCCTGTGAAATAAGTAAGGAAAAAAAAGAAGTAAGTGTTATATAT GATGATTATTTTGTAGATGTAGATAGGATAATAGAATCCATAGAAAATATAGGTTATACAG TTATATAAAAATTACTTTAAAAATTAATAAAAACATGGTAAAATATAAATCGTATAAAGTT GTGAATAAAAT**AAAGGAGGT**TTATATAT:NmeNTRGAATGCA

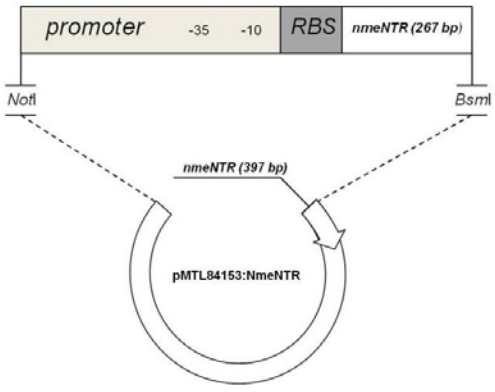


Figure 4.7: Schematic representation of the cloning procedure for the preparation of the promoter/ RBS NmeNTR variants.

pMTL84153:NmeNTR vector was used to prepare the library of synthesized promoter/RBS sequences. *Not*I and *Bsm*I represent the restriction sites used during cloning procedure. All sequences were synthesized containing the first 267 bp of *nmeNTR* gene. After enzymatic digestion pMTL84153:NmeNTR vector contained the remaining 397 bp of *nmeNTR*.

The nucleotide sequences of the assembled, 8 new variants were confirmed by sequencing and thereafter conjugated into *C. sporogenes*. Cell lysates were prepared from 3 individual transformants and menadione assays were conducted to assess the level of nitroreductase expression (Figure 4.8).

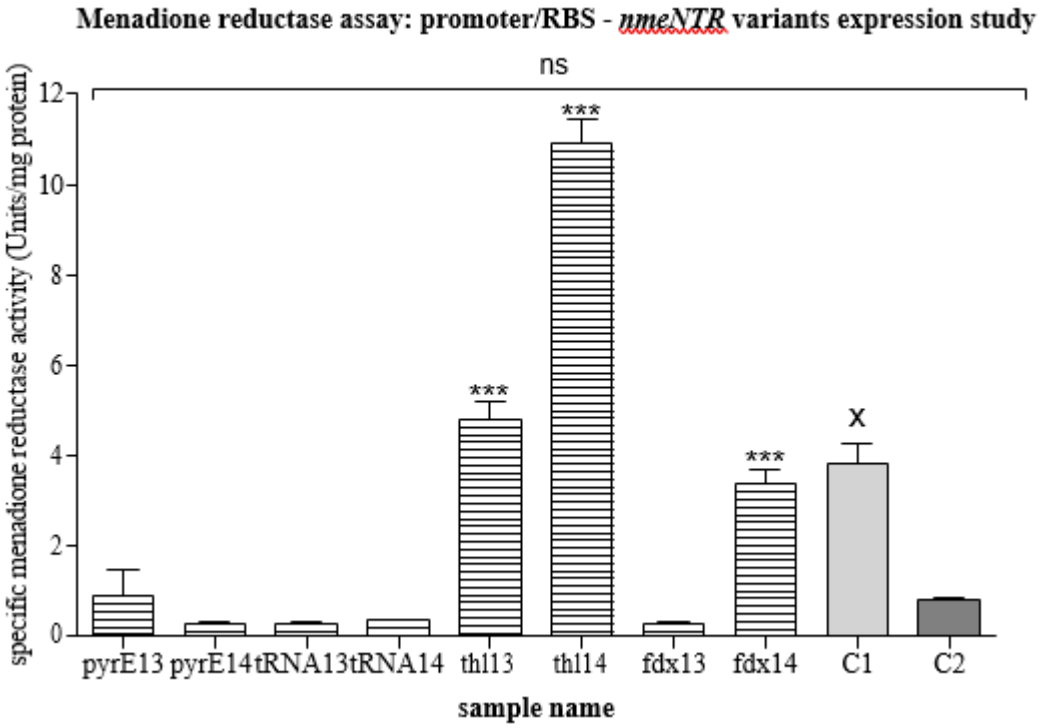


Figure 4.8: Menadione reductase assay on 7th hour soluble fraction lysates of *C. sporogenes* harbouring pMTL84151:NmeNTR vector with different promoter-RBS sequences

(C1) pMTL84153:NmeNTR positive control of the plasmid with native RBS and *fdx* promoter (C2) pMTL84151 negative control (plasmid without *nmeNTR* gene). Soluble fractions of lysates were prepared as described in 2.5.14 and obtained from 7-hour time point from the start of culture incubation. Activities were normalised to total protein concentrations, determined using the Bradford protein assay. Activities are expressed in units per mg. The data represent the average of six independent experiments and error bars indicate the standard errors of the means. Statistical differences between strains with plasmid containing different promoters and control (C2) were calculated using one-way RM ANOVA with Dunnett post tests. With exception of 3 constructs: thl13, thl14 and fdx14 (***, $P < 0.001$) menadione reductase activities were found to be not statistically significant. C1 control was excluded from statistical evaluation (x).

A broad range of NmeNTR*synth* expression level was evident. The highest level of expression was obtained for "thl14" variant (approximately 11 Units per mg). This represents a 3-fold increase in expression level compared to the positive control of the plasmid based NmeNTR*synth* with ferredoxin promoter and its native RBS. Other

tested variants do not show a significant improvement or even demonstrate a decrease in enzyme activity. The „thl14" sequence which consists of the P*thl* promoter (derived from *C. acetobutylicum*) and selected variant of ribosome binding site (R14) was selected to be integrated into the *C. sporogenes* genome.

VIII. Protease cleavage site prediction in the mGM-CSF protein sequence

The analysis presented here was carried out according to section 2.6.11.

PeptideCutter

The sequence to investigate:

```
        10         20         30         40         50         60
APTRSPITVT RPWKHVEAIK EALNLLDDMP VTLNEEVEVV SNEFSFKKLT CVQTRLKIFE

        70         80         90        100        110        120
QGLRGNFTKL KGALNMTASY YQTYCPPTPE TDCETQVTTY ADFIDSLKTF LTDIPFECKK

       130
PGQKDYKDDD DK
```

The sequence is 132 amino acids long.

Available enzymes

The enzyme(s) that you have chosen:

- Asp-N endopeptidase
- Asp-N endopeptidase + N-terminal Glu
- Clostripain
- Glutamyl endopeptidase
- LysC
- Staphylococcal peptidase I
- Thermolysin

You have chosen to display all possible cleaving enzymes.

These enzymes cleave the sequence:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N endopeptidase	11	26 27 91 101 104 112 124 127 128 129 130
Asp-N endopeptidase + N-terminal Glu	21	16 20 26 27 34 35 37 42 59 89 91 93 101 104 112 116 124 127 128 129 130
Clostripain	4	4 11 55 64
Glutamyl endopeptidase	10	17 21 35 36 38 43 60 90 94 117
LysC	13	14 20 47 48 57 69 71 108 119 120 124 127 132
Staphylococcal peptidase I	9	17 21 35 38 43 60 90 94 117
Thermolysin	30	6 8 15 18 22 24 25 30 32 39 45 48 51 55 57 58 62 66 69 72 73 75 77 96 100 103 106 109 110 115

At these positions the following enzymes cleave:

- Please note that the size of the peptides are calculated as if **all chosen enzymes were present** during digestion. If you want to obtain the size of the peptides resulting from the cleavage of only one enzyme, please, deselect the others.
- Please be aware of the fact that the present version of the PeptideCutter program does not take into consideration any kind of **modification** neither of the protein sequence nor of modifications evoked by the cleavage. Mass computations are based on average masses of the occurring amino acid residues, and giving peptide masses as [M]. If you want to select different parameters, we recommend to use PeptideMass.

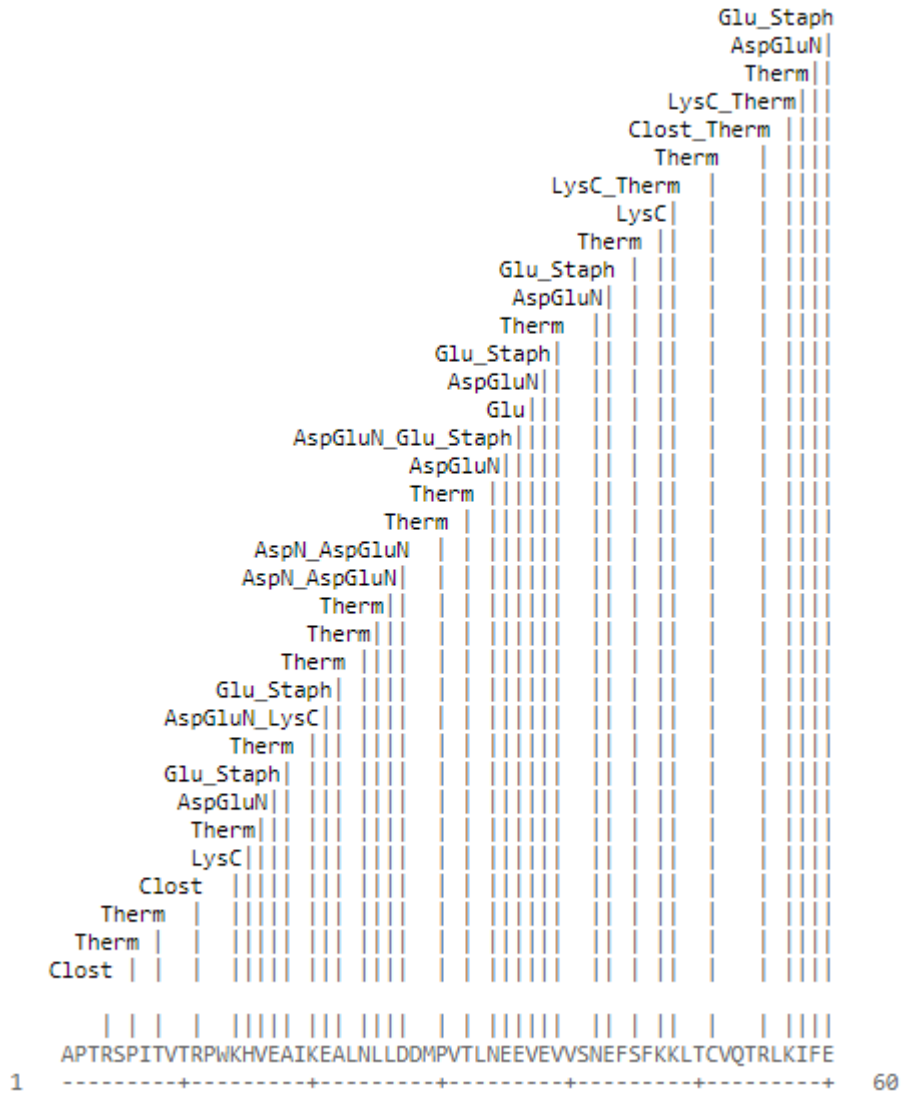
Position of cleavage site	Name of cleaving enzyme(s)	Resulting peptide sequence (see explanations)	Peptide length [aa]	Peptide mass [Da]
4	Clostripain	APTR	4	443.503
6	Thermolysin	SP	2	202.210
8	Thermolysin	IT	2	232.280
11	Clostripain	VTR	3	374.440
14	LysC	PWK	3	429.519
15	Thermolysin	H	1	155.156
16	Asp-N endopeptidase + N-terminal Glu	V	1	117.148
17	Glutamyl endopeptidase Staphylococcal peptidase I	E	1	147.131
18	Thermolysin	A	1	89.094
20	Asp-N endopeptidase + N-terminal Glu LysC	IK	2	259.349
21	Glutamyl endopeptidase Staphylococcal peptidase I	E	1	147.131
22	Thermolysin	A	1	89.094
24	Thermolysin	LN	2	245.278
25	Thermolysin	L	1	131.175
26	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu	L	1	131.175
27	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu	D	1	133.104
30	Thermolysin	DMP	3	361.413
32	Thermolysin	VT	2	218.253
34	Asp-N endopeptidase + N-terminal Glu	LN	2	245.278
35	Asp-N endopeptidase + N-terminal Glu Glutamyl endopeptidase Staphylococcal peptidase I	E	1	147.131
36	Glutamyl endopeptidase	E	1	147.131
37	Asp-N endopeptidase + N-terminal Glu	V	1	117.148
38	Glutamyl endopeptidase Staphylococcal peptidase I	E	1	147.131
39	Thermolysin	V	1	117.148
42	Asp-N endopeptidase + N-terminal Glu	VSN	3	318.330
43	Glutamyl endopeptidase Staphylococcal peptidase I	E	1	147.131
45	Thermolysin	FS	2	252.270
47	LysC	FK	2	293.366
48	LysC Thermolysin	K	1	146.189
51	Thermolysin	LTC	3	335.419
55	Clostripain Thermolysin	VQTR	4	502.571
57	LysC Thermolysin	LK	2	259.349
58	Thermolysin	I	1	131.175

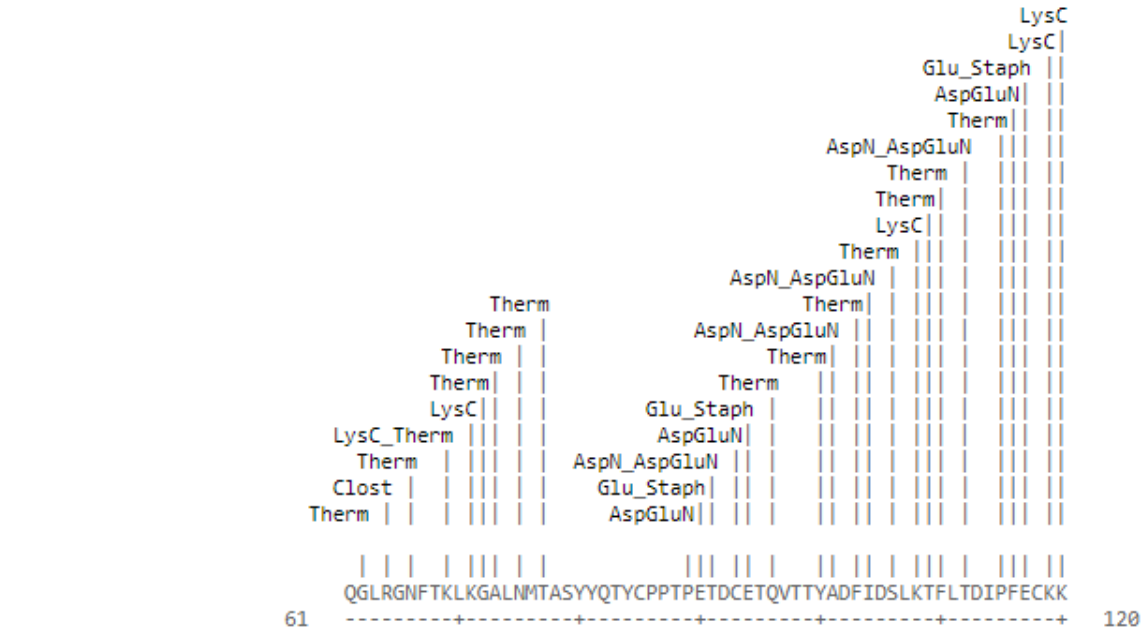
59	Asp-N endopeptidase + N-terminal Glu	F	1	165.192
60	Glutamyl endopeptidase Staphylococcal peptidase I	E	1	147.131
62	Thermolysin	QG	2	203.198
64	Clostripain	LR	2	287.362
66	Thermolysin	GN	2	189.171
69	LysC Thermolysin	FTK	3	394.471
71	LysC	LK	2	259.349
72	Thermolysin	G	1	75.067
73	Thermolysin	A	1	89.094
75	Thermolysin	LN	2	245.278
77	Thermolysin	MT	2	250.313
89	Asp-N endopeptidase + N-terminal Glu	ASYYQTYCPPTP	12	1390.530
90	Glutamyl endopeptidase Staphylococcal peptidase I	E	1	147.131
91	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu	T	1	119.120
93	Asp-N endopeptidase + N-terminal Glu	DC	2	236.243
94	Glutamyl endopeptidase Staphylococcal peptidase I	E	1	147.131
96	Thermolysin	TQ	2	247.251
100	Thermolysin	VTTY	4	482.534
101	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu	A	1	89.094
103	Thermolysin	DF	2	280.280
104	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu	I	1	131.175
106	Thermolysin	DS	2	220.182
108	LysC	LK	2	259.349
109	Thermolysin	T	1	119.120
110	Thermolysin	F	1	165.192
112	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu	LT	2	232.280
115	Thermolysin	DIP	3	343.380
116	Asp-N endopeptidase + N-terminal Glu	F	1	165.192
117	Glutamyl endopeptidase Staphylococcal peptidase I	E	1	147.131
119	LysC	CK	2	249.328
120	LysC	K	1	146.189
124	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu LysC	PGQK	4	428.489
127	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu LysC	DYK	3	424.454

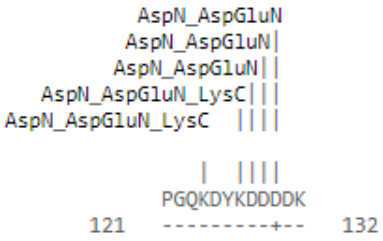
128	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu	D	1	133.104
129	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu	D	1	133.104
130	Asp-N endopeptidase Asp-N endopeptidase + N-terminal Glu	D	1	133.104
132	**end of sequence**	DK	2	261.278

These are the cleavage sites of the chosen enzymes and chemicals mapped onto the entered protein sequence:

- You have chosen a block size of **60** for the map.
- Please note that the cleavage occurs at the **right side** (C-terminal direction) of the marked amino acid.
- You have the possibility to display the results of a single enzyme by **mouseclicking** on the respective enzyme name in the map.




LysC
LysC
Glu_Staph
AspGluN
Therm
AspN_AspGluN
Therm
Therm
LysC
Therm
AspN_AspGluN
Therm
AspN_AspGluN
Therm
Therm
Glu_Staph
AspGluN
AspN_AspGluN
Glu_Staph
AspGluN
LysC_Therm
Therm
Clost
Therm
QGLRGNFTKLKGALNMTASYYQTYCPPTPETDCETQVTTYADFIDSLKTFLTDIPFECKK
61
120


AspN_AspGluN
AspN_AspGluN
AspN_AspGluN
AspN_AspGluN_LysC
AspN_AspGluN_LysC
PGQKDYKDDDDK
121
132

Appendix Bibliography

1. Purdy, D., O'Keeffe, T.A., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., and Minton, N.P., *Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier.* Mol Microbiol, 2002. **46**(2): p. 439-52.
2. Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H., and Wright, F., *Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity.* Nucleic Acids Res, 1988. **16**(17): p. 8207-11.
3. Sharp, P.M., Bailes, E., Grocock, R.J., Peden, J.F., and Sockett, R.E., *Variation in the strength of selected codon usage bias among bacteria.* Nucleic Acids Res, 2005. **33**(4): p. 1141-53.
4. Brüggemann, H., Bäumer, S., Fricke, W.F., Wiezer, A., Liesegang, H., Decker, I., Herzberg, C., Martínez-Arias, R., Merkl, R., Henne, A., and Gottschalk, G., *The genome sequence of Clostridium tetani, the causative agent of tetanus disease.* Proc Natl Acad Sci U S A, 2003. **100**(3): p. 1316-1321.
5. Nolling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I., Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M.J., Bennett, G.N., Koonin, E.V., and Smith, D.R., *Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.* J Bacteriol, 2001. **183**(16): p. 4823-38.
6. Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S., and Hayashi, H., *Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater.* Proc Natl Acad Sci U S A, 2002. **99**(2): p. 996-1001.
7. Villalobos, A., Ness, J.E., Gustafsson, C., Minshull, J., and Govindarajan, S., *Gene Designer: a synthetic biology tool for constructing artificial DNA segments.* BMC Bioinformatics, 2006. **7**: p. 285.