

LONDON
SCHOOL *of*
HYGIENE
& TROPICAL
MEDICINE



**Investigating the roles of the cell wall anchoring
sortase enzyme and sorted proteins in
*Clostridium difficile***

Elizabeth Harden Donahue

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy

University of London

July 2014

Department of Pathogen Molecular Biology

Faculty of Infectious and Tropical Diseases

London School of Hygiene and Tropical Medicine

No funding received

Research group affiliation: Professor Brendan Wren

Declaration

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

Domainex Ltd. (Cambridge, UK) performed the *Leadbuilder* virtual screen using the *PROTOCATS* database to identify potential CD2718 inhibitors. Mark Donahue wrote the Python scripts for the FRET assay data analysis. Dr. Jun Wheeler (National Institute for Biological Standards and Control, Medicine Healthcare Regulatory Agency, Hertfordshire, UK) performed MALDI-TOF mass spectrometry to confirm expression of recombinant proteins. Dr. Len Packman (Protein and Nucleic Acid Chemistry Facility, University of Cambridge, UK) performed MALDI-TOF mass spectrometry to confirm expression of recombinant proteins and to analyse FRET reaction samples for cleavage products. Dr. Johann Peltier expressed and purified the H-CD0183-CwpV protein for use in our assays.

None of the material presented herein has been submitted previously for the purpose of obtaining another degree. This thesis does not exceed 100,000 words as required by the London School of Hygiene and Tropical Medicine.

Elizabeth Harden Donahue

July 2014

Abstract

Clostridium difficile is a Gram-positive, anaerobic bacterium that is the most frequent cause of antibiotic-associated colitis and healthcare-acquired diarrhoea worldwide. In many Gram-positive bacteria, a membrane bound sortase enzyme covalently anchors surface proteins to the cell wall, a process that is essential for virulence. Sortase protein anchoring is mediated by a conserved cell wall sorting signal on the anchored protein, containing the “LPXTG-like” motif. Sequence analysis confirmed that *C. difficile* strain 630 encodes a single sortase, CD2718, but little is known about its function. In this study, we identify seven predicted cell wall proteins with the (S/P)PXTG sorting motif, four of which are conserved across all five *C. difficile* lineages and include potential adhesins and cell wall hydrolases. A FRET-based assay was developed to confirm that recombinant CD2718 catalyses the cleavage of fluorescently labelled peptides containing (S/P)PXTG motifs *in vitro*. Mass spectrometry reveals the cleavage site to be between the threonine and glycine residues of the (S/P)PXTG peptide. Replacement of the predicted catalytic cysteine residue at position 209 with alanine abolishes CD2718 activity, as does addition of the cysteine protease inhibitor MTSET to the reaction. The activity of CD2718 can also be inhibited by several small-molecule inhibitors identified through an *in silico* screen. CD2718-mediated cleavage of a recombinant fusion protein containing the full length predicted sortase substrate CD0183 was also observed. These results demonstrate for the first time that *C. difficile* encodes a single sortase enzyme that recognises (S/P)PXTG sequences. The activity of CD2718 can be inhibited by rationally designed small-molecule inhibitors, and may be an appropriate target for downstream anti-infective therapies against *C. difficile* infection.

Acknowledgements

Firstly, I would like to express my deepest appreciation to Professor Brendan Wren. I will be forever indebted to him for the extraordinary support and kindness he has shown me over the years, along with the time, supervision, expertise and resources he has provided me to complete my studies. Likewise, a big thank you must go to Dr. Lisa Dawson for all of her technical guidance, for constantly engaging in my research and scientific thinking, and for her encouragement throughout the ups and downs over the years. I would like to thank both of them for reading through drafts of this thesis and their insightful comments.

I would like to thank the entire Wren lab for all their support and encouragement; it has been an absolute pleasure to work with all of you. I would like to thank my fellow *C. diff*-ers for their invaluable scientific advice and insightful discussions. Special thanks must also go to Dr. Jon Cuccui for his willingness to share his time (both in the lab and on the golf course), his technical expertise, and his enthusiasm for science.

I would like to thank my family and friends for all of their love and support. Thank you to Sarah and Melissa, for listening to my many rants and for keeping me sane throughout this PhD; Maddie, for all the coffee runs and helping to bring March Madness to England; Meredith, my running buddy and supplier of Australian delicacies; Alice and family, for all the lovely times at Studley End; and Mathieu and Muriel, for sharing their wine, for jass and “nail to nail”. I would like to thank my entire family, Harden and Donahue alike, for their unwavering optimism and continuous support. To my amazing husband, Mark: thank you for all the laughs and always believing in me; I couldn’t have done it without you.

Finally, I dedicate this thesis to my dear grandmother, Evelyn, my first biology teacher.

Table of Contents

Declaration	2
Abstract	3
Acknowledgements	4
List of Figures	9
List of Tables	11
List of Equations.....	12
List of Abbreviations.....	13
1 Introduction.....	15
1.1 Clostridium difficile.....	15
1.1.1 <i>C. difficile</i> infection	16
1.1.2 Risk factors of CDI.....	19
1.1.3 Diagnosis.....	20
1.1.4 Treatment.....	21
1.1.5 Impact of CDI on the hospital environment	23
1.2 Epidemiology of <i>C. difficile</i>	23
1.2.1 <i>C. difficile</i> typing methods	23
1.2.2 The changing epidemiology of <i>C. difficile</i> strains	27
1.2.3 CDI in the community	28
1.3 <i>C. difficile</i> virulence factors	28
1.3.1 Toxins.....	28
1.3.2 Flagella.....	29
1.3.3 Surface layer proteins.....	29
1.3.4 Other surface proteins.....	31
1.3.5 Additional <i>C. difficile</i> virulence factors	31
1.4 Sortases in Gram-positive bacteria.....	32
1.4.1 Sortase transpeptidation	32
1.4.2 Class A sortases – housekeeping sortases.....	34
1.4.3 Class B sortases – iron acquisition	36
1.4.4 Class C sortases – pili assembly	38
1.4.5 Other sortases	39
1.5 Sortases as drug targets	40
1.6 The sortase and its substrates in <i>C. difficile</i>	42
1.7 Aims and objectives.....	43
2 Materials and Methods	44
2.1 Materials	44
2.1.1 Reagents	44
2.1.2 Primers.....	44
2.1.3 Bacterial strains and plasmids used	44
2.1.4 Fluorescence resonance energy transfer (FRET) peptides	45
2.1.5 Sortase inhibitors.....	46
2.1.6 Antibodies used for western blotting	46
2.2 Methods	47
2.2.1 Bacterial growth conditions.....	47
2.2.1.1 Growth kinetics	48
2.2.2 Bioinformatics.....	48
2.2.3 DNA manipulation	48
2.2.3.1 DNA isolation	48

2.2.3.2	Polymerase chain reaction.....	49
2.2.3.3	Cloning	49
2.2.3.4	Colony PCR	49
2.2.3.5	Preparation of electrocompetent <i>E. coli</i>	50
2.2.3.6	Transformation via electroporation.....	50
2.2.4	RNA methods.....	50
2.2.4.1	RNA extraction and purification.....	51
2.2.4.2	Genomic DNA removal.....	51
2.2.4.3	Reverse transcription of total RNA	52
2.2.5	<i>C. difficile</i> mutagenesis	52
2.2.5.1	Intron retargeting.....	52
2.2.5.2	Conjugations	53
2.2.5.3	Mutant selection	53
2.2.5.4	PCR screening mutant clones	54
2.2.5.5	Southern blot	55
2.2.6	Phenotypic tests	55
2.2.6.1	Visualisation of <i>C. difficile</i> colony morphology	55
2.2.6.2	Autoagglutination assay.....	55
2.2.6.3	Bacterial sedimentation assay	56
2.2.6.4	Motility assay	56
2.2.6.5	Biofilm assays.....	56
2.2.7	Protein methods	57
2.2.7.1	SDS-PAGE	57
2.2.7.2	Fluorescent western blots.....	58
2.2.7.3	Protein expression	58
2.2.7.4	Cell lysis.....	59
2.2.7.5	Nickel purification using Ni-NTA Agarose	59
2.2.7.6	Purification using chitin resin.....	60
2.2.7.7	Nickel purification using HPLC	60
2.2.7.8	Gel filtration	60
2.2.7.9	Protein quantification	61
2.2.7.10	Identification of recombinant proteins.....	61
2.2.8	<i>In vitro</i> analysis of enzyme activity.....	61
2.2.8.1	Fluorescence resonance energy transfer (FRET) assays.....	61
2.2.8.2	Data analysis	62
2.2.8.3	Analysis of FRET products	62
2.2.8.4	Enzyme kinetics.....	62
2.2.8.5	Design and testing of sortase inhibitors.....	63
2.2.8.6	Effect of sortase inhibitors on <i>C. difficile</i> culture	64
2.2.8.7	Whole protein cleavage assay.....	65
3	Bioinformatic analysis of the potential <i>C. difficile</i> sortase and identification of putative sortase substrates	66
3.1	Introduction	66
3.1.1	Aims of work described in this Chapter.....	67
3.2	Comparative analysis of CD2718	67
3.3	Distribution of CD2718 in the <i>C. difficile</i> phylogenetic spectrum.....	69
3.4	Bioinformatic prediction of sortase substrates.....	71
3.5	Distribution of predicted sortase substrates among <i>C. difficile</i> strains.....	76
3.6	Analysis of predicted substrate proteins	77
3.6.1	Putative surface protein, CD3246.....	77
3.6.2	Collagen binding proteins.....	78
3.6.3	Cell wall hydrolases	81
3.6.4	5' nucleotidase/phosphoesterase	82
3.7	Discussion.....	84
4	Initial characterisation and purification of the <i>C. difficile</i> sortase, CD2718.....	87

4.1	Introduction	87
4.1.1	Aims of the work described in this Chapter	88
4.2	Transcriptional analysis of CD2718 and CD3146	88
4.2.1	Attempted construction of a defined isogenic knockout in CD2718	90
4.3	Characterisation of CD2718 overexpression in <i>C. difficile</i>	92
4.3.1	Growth dynamics.....	93
4.3.2	Colony morphology	94
4.3.3	Sedimentation of bacterial cells	94
4.3.4	Biofilm formation.....	95
4.4	Expression of recombinant CD2718.....	96
4.4.1	Expression of CD2718 in <i>C. difficile</i>	97
4.4.2	Expression of CD2718 in <i>E. coli</i>	98
4.4.2.1	Expression and purification of full length CD2718.....	98
4.4.2.2	Expressing CD2718 _{ΔN28}	103
4.4.2.3	Expressing CD2718 _{ΔN26}	106
4.4.2.4	Expressing <i>S. aureus</i> SrtB	109
4.5	Discussion.....	112
5	Development of a FRET-based assay to assess relative <i>C. difficile</i> sortase activity	
	115	
5.1	Introduction	115
5.1.1	Aims	115
5.2	Development of a FRET assay for CD2718	116
5.3	Characterisation of CD2718_{ΔN26} activity	123
5.3.1	CD2718 _{ΔN26} cleavage specificity.....	123
5.3.2	Addition of a nucleophile to the FRET reaction.....	125
5.3.3	Cysteine residue is essential for CD2718 _{ΔN26} activity	126
5.3.4	Analysis of FRET reaction products	127
5.3.5	Kinetic measurements of CD2718 _{ΔN26} activity	130
5.3.6	Inhibiting CD2718 _{ΔN26} activity	133
5.3.7	Effect of CD2718 inhibitors on <i>C. difficile</i> growth	139
5.4	Discussion.....	140
6	Analysis of putative sortase substrates in <i>C. difficile</i>	143
6.1	Introduction	143
6.1.1	Aims	143
6.2	Transcriptional analysis of sortase substrates	144
6.3	Construction of sortase substrate mutants in 630Δerm	148
6.4	Initial phenotypic characterisation of mutants.....	152
6.4.1	Growth dynamics.....	152
6.4.2	Colony morphology	153
6.4.3	Cell aggregation	154
6.4.4	Motility	156
6.5	Discussion and future directions	157
7	Development of a sortase cleavage assay using recombinantly expressed sortase substrate proteins	160
7.1	Introduction	160
7.1.1	Aims	161
7.2	Recombinant His-gene-HA substrate proteins.....	162
7.2.1	Expression testing.....	164
7.2.2	Purification of CD3392-s2	165
7.2.3	Initial cleavage assay attempts.....	166
7.3	Recombinant His-CD3392-s2-Strep protein	167
7.3.1	Expression testing and purification	168

7.3.2	Cleavage assay attempts	169
7.4	Expression of truncated substrate protein with C-terminal tags	170
7.4.1	Expression testing of successfully cloned constructs	171
7.4.2	Cleavage assay attempts	176
7.5	Substrate-CwpV fusion proteins.....	178
7.5.1	Expression testing.....	179
7.5.2	Purification	180
7.5.3	Testing H-CD3392-CwpV fusion protein.....	182
7.5.4	Testing the H-CD0183-CwpV fusion protein.....	183
7.6	Discussion.....	185
8	Discussion and conclusions	188
8.1	The <i>C. difficile</i> sortase.....	188
8.2	Sortase inhibitors and CDI	190
8.3	Substrates and the cell wall.....	192
8.4	Final conclusions	192
References		194
A	Materials and Methods tables	214
B	Construction of plasmids	232
B.1	Construction of <i>C. difficile</i> overexpression vectors, pEHD007 and pEHD008	232
B.2	Construction of vectors for sortase expression in <i>E. coli</i> (pEHD009-pEHD015).....	232
B.3	Construction of vectors for sortase substrate expression in <i>E. coli</i> (pEHD016-pEHD033)	
	233	
C	Python scripts for FRET data analysis	235
C.1	Python packages used	235
C.2	Functions used for curve-fit optimisation.....	235
C.3	Script for linear regression.....	235
C.4	Script for least squares fit.....	235
D	Alternative sortase substrate orthologue tables	237
E	Replicates of H-CD0183-CwpV cleavage assay	238

List of Figures

Figure 1.1: <i>C. difficile</i>	15
Figure 1.2: <i>C. difficile</i> pathogenesis	18
Figure 1.3: Phylogenetic tree of <i>C. difficile</i> population structure	26
Figure 1.4: Organisation of the <i>C. difficile</i> cell wall	30
Figure 1.5: Sortase-mediated anchoring of LPXTG surface proteins	33
Figure 1.6: Isd-locus mediated haem-iron uptake across the cell wall of <i>S. aureus</i>	37
Figure 1.7: Assembly of the SpaCAB pili of <i>C. diphtheriae</i> by the class C sortase, CdSrtA	39
Figure 3.1: CD2718 comparative alignment.....	68
Figure 3.2: Location of stop codon in sortase pseudogene, CD3146	69
Figure 3.3: Alignment of CD3146 orthologues	71
Figure 3.4: Structure of the collagen binding protein, Cna, from <i>S. aureus</i>	79
Figure 3.5: Schematic representation of CD0386, CD3392 and CD2831 protein organisation	80
Figure 3.6: Schematic representation of CpbA (CD3145) protein organisation	81
Figure 3.7: Schematic representation of CD0183 and CD2768 protein organisation	82
Figure 3.8: Schematic representation of CD2537 protein organisation.....	83
Figure 4.1: Genomic localisation of CD2718 and CD3146.....	89
Figure 4.2: Extraction of <i>C. difficile</i> 630Δerm RNA	89
Figure 4.3: CD2718 and CD3146 transcription	90
Figure 4.4: Screening potential CD2718 ClosTron mutants	91
Figure 4.5: Schematic diagram of the overexpression construct.....	93
Figure 4.6: Growth analysis of sortase overexpression.....	93
Figure 4.7: Colony morphology of CD2718 overexpression.....	94
Figure 4.8: Sedimentation assay.....	95
Figure 4.9: Biofilm assay.....	96
Figure 4.10: Overexpression of CD2718 in <i>C. difficile</i>	97
Figure 4.11: Comparison of recombinant CD2718 proteins for expression in <i>E. coli</i>	98
Figure 4.12: Schematic diagram of pEHD011 construct	99
Figure 4.13: Expression of CD2718 from BL21(DE3) cells	100
Figure 4.14: Small-scale purification of CD2718.....	101
Figure 4.15: Purification of CD2718 from <i>E. coli</i> BL21(DE3).....	102
Figure 4.16: Schematic diagram of the pEHD012 expression construct.....	103
Figure 4.17: Initial expression testing of pEHD012	104
Figure 4.18: Small-scale purification of CD2718 _{ΔN28}	105
Figure 4.19: Purification of CD2718 _{ΔN28} from <i>E. coli</i> NiCo21(DE3)	106
Figure 4.20: Initial expression testing of pEHD013	107
Figure 4.21: Purification of CD2718 _{ΔN26} from <i>E. coli</i> NiCo21(DE3)	108
Figure 4.22: MALDI fingerprinting analysis of CD2718 _{ΔN26}	109
Figure 4.23: Expression and purification of pEHD015	110
Figure 4.24: Purification of SaSrtB _{ΔN28} from <i>E. coli</i> NiCo21(DE3)	111
Figure 5.1: Absorbance and emission spectra for Dabcyl and Edans	116
Figure 5.2: Effect of downstream residues on FRET peptide cleavage	117
Figure 5.3: Initial FRET experiments	119
Figure 5.4: FRET activity with CD2718 _{ΔN26}	120
Figure 5.5: Increase in fluorescence over time is dependent on CD2718 _{ΔN26}	121
Figure 5.6: Optimising peptide concentration	122
Figure 5.7: Time dependent cleavage of (S/P)PXTG peptides by recombinant CD2718 _{ΔN26}	123
Figure 5.8: CD2718 _{ΔN26} substrate specificity	124
Figure 5.9: Nucleophiles have no effect on CD2718 _{ΔN26} mediated cleavage of FRET peptides.....	126
Figure 5.10: CD2718 _{ΔN26} activity requires a cysteine residue at position 209	127
Figure 5.11: CD2718 _{ΔN26} cleavage efficiency improves with native peptide sequence	128
Figure 5.12: CD2718 _{ΔN26} cleaves FRET peptides between T and G residues	130
Figure 5.13: Edans fluorophore standard curve	131
Figure 5.14: Kinetic parameters of CD2718 _{ΔN26}	132
Figure 5.15: Space-filling model of compound interacting with active site of BaSrtB	133
Figure 5.16: Determining the IC ₅₀ of MTSET.....	134
Figure 5.17: Effect of DMSO concentration on CD2718 _{ΔN26} activity.....	135

Figure 5.18: Structures and IC ₅₀ values of CD2718 _{ΔN26} inhibitors.....	138
Figure 5.19: CD2718 inhibitors have no effect on <i>C. difficile</i> growth.....	139
Figure 6.1: Assessment of cDNA quality	144
Figure 6.2: Transcriptional analysis of <i>CD0183</i> , <i>CD2537</i> , <i>CD2831</i> , and <i>CD3246</i>	146
Figure 6.3: Transcriptional analysis of <i>CD0386</i> and <i>CD3392</i>	147
Figure 6.4: Transcriptional analysis of <i>CD2768</i>	148
Figure 6.5: PCR screen of constructed <i>CD0386</i> mutants	149
Figure 6.6: PCR screen of constructed <i>CD2831</i> mutants	150
Figure 6.7: PCR screen of constructed <i>CD3392</i> mutants	151
Figure 6.8: Southern blot analysis of mutants	152
Figure 6.9: Growth analysis of sortase substrate mutants.....	153
Figure 6.10: Colony morphology of substrate mutants.....	154
Figure 6.11: Autoagglutination of <i>C. difficile</i> mutants	155
Figure 6.12: Sedimentation assay	156
Figure 6.13: Motility testing of <i>C. difficile</i> mutants.....	157
Figure 7.1: Sortase-mediated cleavage of a protein substrate.....	161
Figure 7.2: Sample of <i>C. difficile</i> protein array pilot data.....	162
Figure 7.3: Construct design from pilot array	163
Figure 7.4: Expression of His-HA proteins.....	165
Figure 7.5: Confirmation of <i>CD3392-s2</i> expression	166
Figure 7.6: Cleavage assay with <i>CD3392-s2</i>	167
Figure 7.7: Design of <i>CD3392s2</i> -Strep protein	168
Figure 7.8: Expression of the <i>CD3392s2</i> -Strep protein.....	169
Figure 7.9: Cleavage assay attempts with <i>CD3392s2</i> -Strep.....	170
Figure 7.10: Schematic of new substrate fragment expression constructs	171
Figure 7.11: Expression of <i>CD0183</i> fragments.....	173
Figure 7.12: Expression of <i>CD2768</i> fragments.....	173
Figure 7.13: Expression testing of <i>CD2831</i> fragments	174
Figure 7.14: Expression of <i>CD3392</i> fragments.....	174
Figure 7.15: Confirmation of H- <i>CD2768</i> -Strep expression.....	176
Figure 7.16: Cleavage assay attempts with H- <i>CD2768</i> -Strep protein	177
Figure 7.17: Substrate-CwpV fusion protein design	179
Figure 7.18: Expression testing of CwpV fusion constructs	180
Figure 7.19: Purification of H- <i>CD3392</i> -CwpV protein.....	181
Figure 7.20: Confirmation of CwpV fusion protein identity	182
Figure 7.21: H- <i>CD3392</i> -CwpV cleavage assay attempt	183
Figure 7.22: Cleavage of H- <i>CD0183</i> -CwpV by <i>CD2718</i> _{ΔN26}	184
Figure 8.1: Replicate 2 of H- <i>CD0183</i> -CwpV cleavage assay	238
Figure 8.2: Replicate 3 of H- <i>CD0183</i> -CwpV cleavage assay	239
Figure 8.3: Replicate 4 of H- <i>CD0183</i> -CwpV cleavage assay	240

List of Tables

Table 2.1: Bacterial strains used in this study	45
Table 2.2: FRET peptides used in this study	46
Table 2.3: Antibodies used in this study	47
Table 3.1: Representative <i>C. difficile</i> strains	70
Table 3.2: CD2718 orthologues and % amino acid sequence identity between representative <i>C. difficile</i> strains	71
Table 3.3: (S/P)PXTG and NVQTG containing proteins in <i>C. difficile</i> 630	74
Table 3.4: Distribution and % amino acid identity of predicted CD2718 substrate proteins across <i>C. difficile</i> strains*	76
Table 3.5: Additional (S/P)PXTG containing proteins in <i>C. difficile</i> strains	77
Table 5.1: Kinetic parameters for CD2718 _{ΔN26} compared with those for other sortases	132
Table 5.2: Percent CD2718 _{ΔN26} activity in presence of inhibitors.....	137
Table 7.1: His-HA expression constructs from the Felgner lab	164
Table 7.2: Expected sizes of recombinant truncated substrate proteins.....	172
Table 7.3: Expected sizes of cloned substrate-CwpV fusion proteins.....	179
Table A.1: Oligonucleotides used in this study	214
Table A.2: Plasmids used in this study	217
Table A.3: Sortase inhibitor details	219
Table D.1: Gene ID of predicted sortase substrate orthologues	237

List of Equations

Equation 2.1: Michaelis-Menten equation	63
Equation 2.2: Sigmoidal concentration-response curve.....	64

List of Abbreviations

2,6-Dap	2,6-diaminopimelic acid
aa	amino acid
ADP	dadenosine diphosphate
BCA	bicinchoninic acid
BHIS	brain heart infusion (supplemented with 0.05% L-cysteine)
BHISc	brain heart infusion (supplemented with 0.1% L-cysteine)
bp	base pair
CDI	<i>C. difficile</i> infection
cDNA	complementary DNA
CDT	<i>C. difficile</i> binary toxin
CWP	cell wall protein
CWSS	cell wall sorting signal
Dabcyd or <i>d</i>	4-([4-(dimethylamino)phenyl]-azo)-benzoic acid
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
DTT	dithiothreitol
Edans or <i>e</i>	5-((2-aminoethyl)amino)naphthalene-1-sulfonic acid
EDTA	ethylenediaminetetraacetic acid
EIA	enzyme immunoassays
FRET	fluorescence resonance energy transfer
GDH	glutamate dehydrogenase
GlcNAc	<i>N</i> -acetylglucosamine
HMW	high molecular weight
IPTG	isopropyl β -D-1-thiogalactopyranoside
kDa	kilodaltons
LB	Luria-Bertani
LMW	low molecular weight
MALDI-TOF	Matrix-assisted laser desorption/ionisation time of flight
<i>m</i> -Dap	<i>meso</i> -diaminopimelic acid
MLST	multi-locus sequence typing

MOPS	3-(N-morpholino)propanesulfonic acid
MRSA	methicillin resistant <i>S. aureus</i>
MTSET	2-(trimethylammonium)ethyl]methanethiosulfonate
MurNAc	<i>N</i> -acetylmuramic acid
Ni-NTA	nickel nitrilotriacetic acid
OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
PMC	pseudomembranous colitis
PVDF	polyvinylidene difluoride
REA	restriction endonuclease analysis
RFLP	restriction fragment length polymorphism
RFU	relative fluorescence units
RNA	ribonucleic acid
rRNA	ribosomal RNA
RT	ribotype
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SLP	surface-layer protein
SOE	splicing by overlap extension
SrtA	sortase A
SrtB	sortase B
TcdA	toxin A
TcdB	toxin B
UDP	undecaprenyl phosphate
VRE	vancomycin-resistant <i>Enterococci</i>

1 Introduction

1.1 *Clostridium difficile*

Clostridium difficile is a member of the phylum Firmicutes and class Clostridia. Members of the genus *Clostridium* are described as obligate anaerobes, Gram-positive bacilli capable of producing metabolically dormant endospores in response to nutrient deficiency or stress (Figure 1.1). The genus contains over 150 species, with only a few considered to be pathogenic to animals and humans, including but not limited to *C. botulinum*, *C. tetani*, *C. perfringens*, and *C. difficile* (Woo *et al.*, 2005). These pathogenic species have been well characterised and are noted for their production of exotoxins, which are among some of the most potent toxins found in nature (Johnson, 1999). *C. difficile* vegetative cells are thick rods, typically two to eight µm in length and 0.5 µm in width, with oval, subterminal spores (Figure 1.1A and B) (Smith, 1975). *C. difficile* spores are highly resistant to desiccation and most disinfectants. Spores persist in the environment for months and facilitate spread of the organism in health care settings (Vonberg *et al.*, 2008, Gerding *et al.*, 2008).

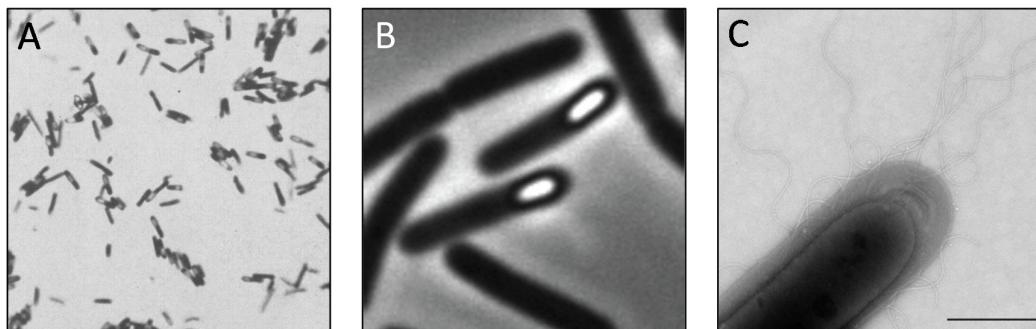


Figure 1.1: *C. difficile*

- A.** Light microscopy image of Gram-stained *C. difficile*. Image from Hall & O'Toole, 1935. **B.** Phase-contrast microscopy image of *C. difficile* cells producing spores. Image from Pereira *et al.*, 2012. **C.** Electron microscopy image of *C. difficile* flagella. Image from Martin *et al.*, 2013.

As an obligate anaerobe, *C. difficile* must be cultured under anaerobic conditions. *C. difficile* is routinely grown on selective media containing the antibiotics cefoxitin and cycloserine to suppress the growth of other organisms from stool (George *et al.*, 1979). *C. difficile* colonies are typically large and circular with rough edges, and fluoresce green-

yellow under ultraviolet light (George *et al.*, 1979). *C. difficile* also produces a characteristic odour resembling fresh horse manure due to its production of various volatile fatty acids and *para*-cresol (Levett, 1984). Diarrheal samples from patients infected with *C. difficile* have been noted to have a similar horse manure smell (Bomers *et al.*, 2014), though the diagnostic power of stool odour from patients with suspected *C. difficile* infection (CDI) remains debatable (Rao *et al.*, 2013). Vegetative cells appear as Gram-positive or Gram-variable bacilli (Cohen *et al.*, 2010), with spores visible depending on the agar used for culturing (George *et al.*, 1979). Most strains of *C. difficile* are motile by peritrichous flagella which may also aid in adherence to colonic mucosal cells (Figure 1.1C) (Twine *et al.*, 2009).

Originally named *Bacillus difficile* because of its fastidious growth requirements, *C. difficile* was first described in 1935 as part of the intestinal microflora of healthy infants. Though it was observed to produce a heat sensitive toxin that was pathogenic in guinea pigs, the bacteria was considered non-pathogenic to humans for almost four decades as newborns had no pathology associated with carriage (Hall & O'Toole, 1935). *C. difficile* was not recognised as a causative agent of antibiotic-associated colitis and pseudomembranous colitis (PMC) until 1978 (Bartlett *et al.*, 1978, Larson *et al.*, 1978). PMC had been a rare condition until the introduction of broad-spectrum antibiotics in the latter half of the 20th century, when it became a relatively common adverse effect of treatment with clindamycin (Tedesco *et al.*, 1974).

1.1.1 *C. difficile* infection

C. difficile is now recognised as one of the most important causes of antibiotic-associated diarrhoea in North America and Europe. CDI is responsible for an estimated 14,000 deaths per year in the USA (Centers for Disease Control and Prevention, <http://www.cdc.gov/HAI>), and an estimated 1500-2000 deaths per year in England and Wales (Office for National Statistics, <http://www.ons.gov.uk>). *C. difficile* is the most common pathogen associated with nosocomial infections in the USA, accounting for an estimated 12% of total healthcare-acquired infections compared to the 10% caused by MRSA (Magill *et al.*, 2014). Infection with *C. difficile* infection results in a broad range of clinical presentations, from asymptomatic colonisation and mild diarrhoea, to more

severe forms including the life-threatening PMC (Kazanowski *et al.*, 2014). *C. difficile* enters the body via the faecal-oral route through the ingestion of metabolically dormant endospores, considered the primary transmission factor of *C. difficile* (Vonberg *et al.*, 2008, Deakin *et al.*, 2012). Spores are excreted in large numbers in the faeces of patients with CDI and readily contaminate hospital environments (Verity *et al.*, 2001). *C. difficile* spores are highly resistant to commonly used hospital disinfectants (Dawson *et al.*, 2011b, Lawley *et al.*, 2010, Vohra & Poxton, 2011), and thus are difficult to eradicate and can remain in the environment for months (Gerding *et al.*, 2008, Fekety *et al.*, 1981). The hands of hospital personnel are also frequently contaminated with *C. difficile* spores and considered a major source of nosocomial transmission (McFarland *et al.*, 1989). Following the disruption of the colonic microflora by antibiotic therapy, ingested *C. difficile* spores are stimulated to germinate in the presence of bile salts and glycine found in the intestines (Sorg & Sonenshein, 2008). The germinated spores then undergo vegetative replication and are able to colonise the intestinal epithelium, produce toxins and cause disease (Figure 1.2A).

The *C. difficile* toxins, toxin A (TcdA) and toxin B (TcdB), are cytotoxic and responsible for the main clinical symptoms of CDI (Voth & Ballard, 2005). Common symptoms include watery diarrhoea, abdominal pain, fever and nausea, with fulminant colitis occurring in 3-8% of patients (Kazanowski *et al.*, 2014). Severe CDI can result in PMC, characterised the formation of summit-shaped pseudomembranes on the colonic mucosa composed of an exudate of inflammatory cells, fibrin, and necrotic debris (Figure 1.2B and C) (Price & Davies, 1977). Complications of severe CDI and PMC can lead to toxic megacolon, colonic dilation and perforation, sepsis, and mortality in 25-40% of cases (Cleary, 1998, Kuijper *et al.*, 2007). Increased toxin production, measured by high toxin levels in faeces, has been linked to greater severity of CDI symptoms, such as the development of PMC, and increased mortality (Akerlund *et al.*, 2006).

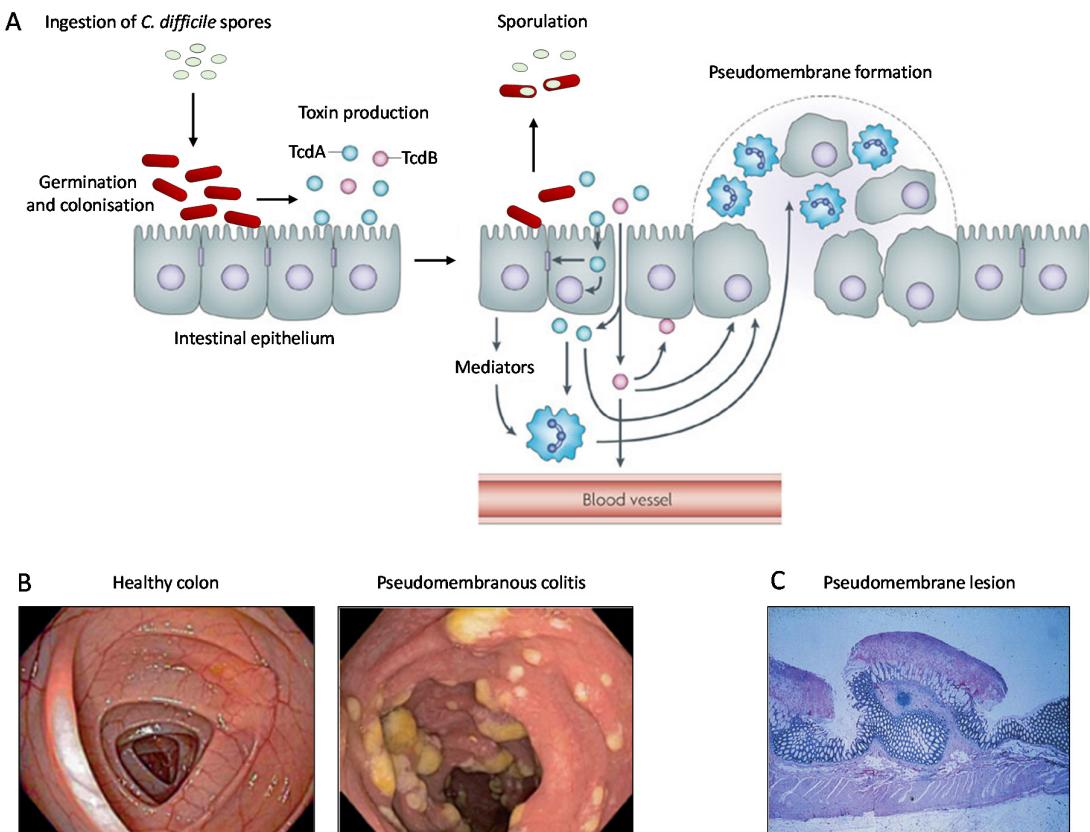


Figure 1.2: *C. difficile* pathogenesis

A. Ingested *C. difficile* spores germinate in the anaerobic environment of the colon, and are able to colonise following disruption of the normal intestinal flora. Toxin production (TcdA and TcdB) by vegetative cells induces the release of various inflammatory mediators from intestinal epithelial cells, phagocytes and mast cells, resulting in inflammation and the accumulation of neutrophils. Figure adapted from Rupnik *et al.* 2009 and Shen, 2012. **B.** Colonoscopy images from a healthy patient (left) and a patient diagnosed with severe CDI with pseudomembranous lesions characteristic of PMC (right). Images from Borody & Khoruts, 2012. **C.** Histology of characteristic PMC lesion. Image from Bartlett, 2008.

Exposure to *C. difficile* can also result in asymptomatic colonisation of the intestinal tract of healthy adults and neonates, with 3-4% adults and up to 66% infants colonised with toxigenic strains (Kato *et al.*, 2001, Ozaki *et al.*, 2004, Kazanowski *et al.*, 2014). In the healthcare setting, colonisation rates can reach much higher. Though the *C. difficile* colonisation rates of patients at admission are similar to that of the general population (Loo *et al.*, 2011, Eyre *et al.*, 2013), it is estimated that approximately 50% of hospital patients become asymptomatic carriers of *C. difficile* following a hospital stay (Lawrence

et al., 2007). Asymptomatic *C. difficile* colonisation with either toxigenic or non-toxigenic strains is associated with a decreased risk of CDI (Johnson *et al.*, 1990, Shim *et al.*, 1998). However, *C. difficile* carriers may serve as a reservoir for infection in healthcare settings, as cases of hospital-acquired CDI have been linked to newly admitted asymptomatic carriers (Clabots *et al.*, 1992, Curry *et al.*, 2013), and carriers often exhibit skin and environmental spore contamination (Riggs *et al.*, 2007).

1.1.2 Risk factors of CDI

The antibiotic-associated nature of PMC was known before it was linked to CDI (Cohen *et al.*, 1973, Tedesco *et al.*, 1974). Broad-spectrum antibiotic usage has since been identified as one of the primary risk factors of CDI, in addition to prolonged hospitalisation, a compromised immune system and advanced age (Bartlett, 2002, Kazanowski *et al.*, 2014). As a result, CDI is problematic in healthcare facilities, especially in long-term care facilities as a high proportion of patients are particularly susceptible to CDI. The native bacterial population of the colon provides a host defence by inhibiting colonisation and overgrowth of *C. difficile* (Borriello & Barclay, 1986, Vollaard & Clasener, 1994). *C. difficile*, resistant to many commonly used antibiotics (Gerding, 2004, Owens *et al.*, 2008), is able to proliferate and cause disease when the intestinal microbiota is disturbed in the presence of antibiotics (Pultz & Donskey, 2005).

C. difficile strains have been shown to be resistant to fluoroquinolones due to mutations in DNA gyrase genes (Dridi *et al.*, 2002, Drudy *et al.*, 2006), and conjugative transposons are frequently involved in resistance to macrolides and tetracycline (Lyra & Rood, 2000, Spigaglia *et al.*, 2005, Farrow *et al.*, 2001, Roberts *et al.*, 2001). The use of broad-spectrum antimicrobials to which *C. difficile* is resistant; such as clindamycin, cephalosporins and fluoroquinolones; has been associated with increased risk of developing CDI (Johnson *et al.*, 1999, Gaynes *et al.*, 2004, Gerding, 2004, Pépin *et al.*, 2005b, Muto *et al.*, 2005, Owens *et al.*, 2008). Epidemics of CDI in hospitals are frequently attributed to resistant *C. difficile* strains and to changing antimicrobial usage patterns (Owens *et al.*, 2008). Several studies have shown that restricting antibiotic usage can significantly reduce nosocomial infections caused by *C. difficile* (Carling *et al.*, 2003, Valiquette *et al.*, 2007).

Advanced age is a major risk factor for acquiring CDI, as the incidence of CDI in geriatric patients is routinely higher in population-based analyses. In a Swedish hospital, CDI incidence was 10-fold greater in patients 60 years of age or older compared to patients under 60 (Karlstrom *et al.*, 1998). In 2003, CDI incidence in Quebec was estimated to be 156.3 cases per 100,000 overall, but that increased to 866.5 and 1681 cases per 100,000 in those older than 65 and 80 years of age, respectively (Pépin *et al.*, 2004). Those over 65 years of age have also been found to be at greater risk of severe, complicated CDI, and experience more recurrent infections (Pépin *et al.*, 2005a). Age itself is not necessarily the reason for the increase in associated risk of CDI, but likely due to the disproportionate number of elderly people with underlying health conditions on antibiotics in hospitals and long term care facilities.

1.1.3 Diagnosis

Most patients with CDI present with foul smelling, watery diarrhoea following hospitalisation or antibiotic use for an underlying condition (Bartlett & Gerding, 2008). When CDI is the suspected cause of antibiotic associated diarrhoea, faecal samples are often tested for the presence of TcdA and TcdB using either a cell cytotoxicity assay or enzyme immunoassays (EIA) (Wilcox, 2012). The cell cytotoxicity assay is highly sensitive (98%) but technically demanding and takes 24-28 hours, while the toxin EIA tests are commercially available, are cheaper and quicker to perform, but are less sensitive and miss about 40% of cases (Bartlett & Gerding, 2008, Crobach *et al.*, 2009). The accuracy of tests that rely on the detection of toxin is complicated by the toxin variable strains, including the widely distributed toxin A-/B+ lineage (Johnson *et al.*, 2003, Lemee *et al.*, 2004). Culturing *C. difficile* from patient stool and PCR-based nucleic acid amplification tests (NAATs) to detect toxin genes can provide valuable epidemiological information in strain typing, but are unable to discriminate between asymptomatic colonisation with *C. difficile* and isolates causing disease (Crobach *et al.*, 2009, Wilcox, 2012). Another commonly used EIA test detects the presence of glutamate dehydrogenase (GDH), a metabolic enzyme produced by *C. difficile*, in faecal samples. This test is highly sensitive and can be used to rule out the presence of *C. difficile* in the gut, but is poorly specific for CDI and is often used in conjunction with toxin tests (Ticehurst *et al.*, 2006, Sharp *et al.*, 2010). There is no single test that provides reliable sensitivity and specificity with rapid

turnaround and at low cost, so CDI diagnosis is often made using a combination of techniques or multi-step algorithms (Crobach *et al.*, 2009, Wilcox, 2012).

1.1.4 Treatment

Treatment of CDI depends largely on the severity of the disease. Initial management of patients with CDI includes cessation of the course of antibiotics that precipitated its onset, or switching to antibiotics that are less associated with CDI such as aminoglycosides, macrolides, tetracyclines, and vancomycin (Kazanowski *et al.*, 2014). Standard treatment of CDI includes the oral antibiotics metronidazole or vancomycin, and palliative care of the diarrheal symptoms including hydration and electrolyte replacement. For mild CDI, no significant difference in initial treatment efficacy has been observed between metronidazole and vancomycin (Zar *et al.*, 2007, Drekonja *et al.*, 2011). Metronidazole is the preferred choice for treatment of mild to moderate cases as it is less expensive and causes fewer side effects than vancomycin and does not contribute to the spread of vancomycin-resistant *Enterococci* (VRE) (Fekety, 1997, Cohen *et al.*, 2010). However, oral vancomycin treatment is recommended for patients with severe disease, as recent reports have suggested that metronidazole may be less effective in severe cases (Zar *et al.*, 2007, Kelly & LaMont, 2008, Cohen *et al.*, 2010). There have been reports of an increase in metronidazole resistance among some clinical isolates (Al-Nassir *et al.*, 2008, Baines *et al.*, 2008, Kuijper & Wilcox, 2008). Up to 5% of CDI cases progress to fulminant colitis, where patients develop toxic megacolon, multi-organ failure, and perforation or necrosis of the intestine (Kazanowski *et al.*, 2014). These require immediate surgical intervention to either remove the entire colon, which carries a 40-50% 30-day mortality rate (Al-Abed *et al.*, 2010, Perera *et al.*, 2010), or perform the less invasive diverting loop ileostomy with colonic lavage, which carries a 20% mortality rate (Neal *et al.*, 2011).

Treating CDI is inherently difficult, as standard antibiotic therapy with metronidazole or vancomycin may exacerbate the problem by continuing to disrupt the gut microflora. Initially, CDI patients may respond well to antibiotics, but their symptoms resume within one month after cessation of treatment (Rupnik *et al.*, 2009). Known as recurrent CDI, this includes relapse of the initial infection due to persistence of the *C. difficile* isolate within the patient, as well as re-infection with another or even the same isolate because the

patient remains susceptible following antibiotic treatment (Wilcox *et al.*, 1998). Approximately one third of patients treated for CDI experience recurrent infection following treatment, and those that relapse have a greater risk of succumbing to the infection (Pépin *et al.*, 2006, Johnson, 2009). Treatment with metronidazole and vancomycin are reported to have similar recurrence rates (Kelly & LaMont, 2008).

The narrow spectrum macrolide antibiotic, fidaxomicin, was recently approved for first-line CDI treatment in the USA and Europe (Johnson & Wilcox, 2012). Fidaxomicin targets *C. difficile* with limited effect on other predominant Gram-positive and Gram-negative members of the commensal microbiota species (Tannock *et al.*, 2010, Louie *et al.*, 2012, Johnson & Wilcox, 2012). This narrow-spectrum of activity may be responsible for the reduced rates of recurrent infection observed following fidaxomicin treatment (7.8% to 19.7%) when compared to vancomycin treatment (25.5% to 35.5%) in a number of double-blind clinical studies (Drekonja *et al.*, 2011, Louie *et al.*, 2011, Cornely *et al.*, 2012, Crook *et al.*, 2012, Johnson & Wilcox, 2012). However, fidaxomicin is considerably more expensive than vancomycin, and its cost-effectiveness as a first-line treatment for CDI has yet to be demonstrated (Drekonja *et al.*, 2011, Johnson & Wilcox, 2012).

CDI is considered a disease that results from an imbalance of the microbiota, known as dysbiosis. Because of the role of the intestinal microbiota in the development of CDI, restoration of the balance of the intestinal flora is important in resolving infection and preventing recurrence. In a number of case studies, CDI resolution has been achieved through faecal transplants, where faecal bacteria from a healthy donor are transferred to the patient (Aas *et al.*, 2003, Persky & Brandt, 2000, Rohlke *et al.*, 2010, Garborg *et al.*, 2010). Faecal transplants restore the intestinal microbial community and are an effective alternative therapy with a reported success rate of resolving *C. difficile* disease in ~90% of infections (van Nood *et al.*, 2009, Gough *et al.*, 2011). A recent randomised, controlled clinical trial provided the first direct evidence that faecal transplants were more effective at curing CDI than vancomycin after a single treatment (van Nood *et al.*, 2013). This phenomenon has also been demonstrated in animal experiments, with faecal transplantation resolving the *C. difficile* supershedding state in the murine infection model, restoring a healthy and diverse microbiota and reducing contagiousness (Lawley *et al.*, 2012). In this same study, the use of a simple culture mixture of six intestinal

bacteria was enough to clear *C. difficile* pathogenesis in mice, highlighting the potential for bacteriotherapy as a treatment for CDI in humans (Lawley *et al.*, 2012).

1.1.5 Impact of CDI on the hospital environment

C. difficile places an enormous burden on the health care system, not just in terms of treatment cost but also due to the high operational costs associated with infection control. Annual costs of CDI are difficult to quantify, and estimates range from \$800 million to \$4.8 billion for the USA, and €3 billion in Europe (Bouza, 2012, Dubberke & Olsen, 2012). Patients who contract CDI while in hospital remain hospitalised for an additional 2-8 days on average, with a treatment cost of up to \$15,000 per episode of CDI in the USA (Bouza, 2012, Dubberke & Olsen, 2012). Patient isolation and the use of dedicated equipment in CDI treatment are recommended but compliance is often challenged by the limited availability of hospital resources (Vonberg *et al.*, 2008). Current recommendations for *C. difficile* control and containment in health care facilities are difficult to maintain and costly. *C. difficile* spores readily contaminate the hospital environment and the hands of health-care workers as well as visitors, and are difficult to eradicate. Sodium hypochlorite is commonly used to reduce environmental contamination of *C. difficile* spores in patients' rooms, but can be corrosive at high concentrations (Gerding *et al.*, 2008). To reduce the spread of spores by health care works, barrier techniques such as the use of disposable gowns and gloves are recommended during patient care (Vonberg *et al.*, 2008). Traditional alcohol based antiseptics and hand gels are ineffective at removing spores from hands, and recent guidelines suggest use of gloves and frequent hand washing with soap and water (Oughton *et al.*, 2009, Cohen *et al.*, 2010).

1.2 Epidemiology of *C. difficile*

1.2.1 *C. difficile* typing methods

Typing methods for discriminating between bacterial isolates of the same species are important tools in epidemiology and infection control. Many molecular typing schemes are used to study the epidemiology of *C. difficile* infection, including restriction

endonuclease analysis (REA) and pulsed-field gel electrophoresis (PFGE), PCR-ribotyping, toxinotyping, and multi-locus sequence typing (MLST).

Both the REA and PFGE methods utilise restriction enzymes to cut the bacterial genome, resulting in various DNA fragments that are separated by gel electrophoresis (Rupnik *et al.*, 2009). In REA, the genome is cut frequently, producing a large number of smaller DNA fragments that are separated in an agarose gel (Clabots *et al.*, 1993). This method is usually highly discriminatory, but produces complex DNA banding patterns that can be difficult to interpret and reproduce. In contrast, the DNA fragments produced by PFGE are much larger than those in REA, and cannot be resolved on an agarose gel (Gal *et al.*, 2005). Instead, the fragments are separated in a polyacrylamide gel with an alternating electrical field that enables the large DNA fragments to migrate through the gel according to their size. The resulting banding patterns, often referred to as pulsovars, are highly discriminatory. However a large proportion of *C. difficile* strains are non-typable by PFGE due to extensive genomic DNA degradation.

PCR-ribotyping, one of the most commonly used *C. difficile* typing techniques, is based on the variation that occurs in the intergenic spacer regions of the 16S and 23S ribosomal RNA genes (Stubbs *et al.*, 1999). Specific primers are used for PCR-mediated amplification of the DNA that encodes these RNA regions. This method generates a few DNA bands as visualised by gel electrophoresis, and the different DNA band patterns are referred to as ribotypes. To date, over 190 PCR-ribotypes have been reported (Stubbs *et al.*, 1999, Kuijper *et al.*, 2007, Rupnik *et al.*, 2009). Toxinotyping is another PCR-based typing method used for detection of sequence variations within the pathogenicity locus that encodes the *C. difficile* toxins, TcdA and TcdB (Rupnik *et al.*, 1998). PCR products are cut with several restriction enzymes to monitor restriction fragment length polymorphisms (RFLPs); the resulting fragments are resolved by gel electrophoresis and toxinotypes are determined by unique banding patterns (Rupnik *et al.*, 1998). To date, over 27 toxinotypes have been reported (Rupnik *et al.*, 2009).

Multi locus sequence typing (MLST) is based on the allele pattern of sequences for seven housekeeping loci (Maiden *et al.*, 1998, Lemee *et al.*, 2004). The seven housekeeping genes are amplified and sequenced, and sequence types (ST) are assigned based on the

allelic profiles identified. MLST was developed for *C. difficile* in 2004 and is a reliable typing method that facilitates comparisons of strains between laboratories. The University of Oxford hosts the *C. difficile* MLST database (<http://pubmlst.org/cdifficile/>), which provides an open-access locus and sequence definitions database available for researchers around the world to type their strains. MLST and whole genome sequencing studies reveal a broad genetic diversity within *C. difficile* strains, and that *C. difficile* forms at least five distinct lineages associated with human infection (He *et al.*, 2010, Dingle *et al.*, 2011, Stabler *et al.*, 2012), with evidence for a sixth lineage recently described (Knetsch *et al.*, 2012) (Figure 1.3). Clade 1 represents a diverse range of ribotypes, while the other 4 major clades are clonal in structure and are associated with the increasing rates of CDI seen in recent years.

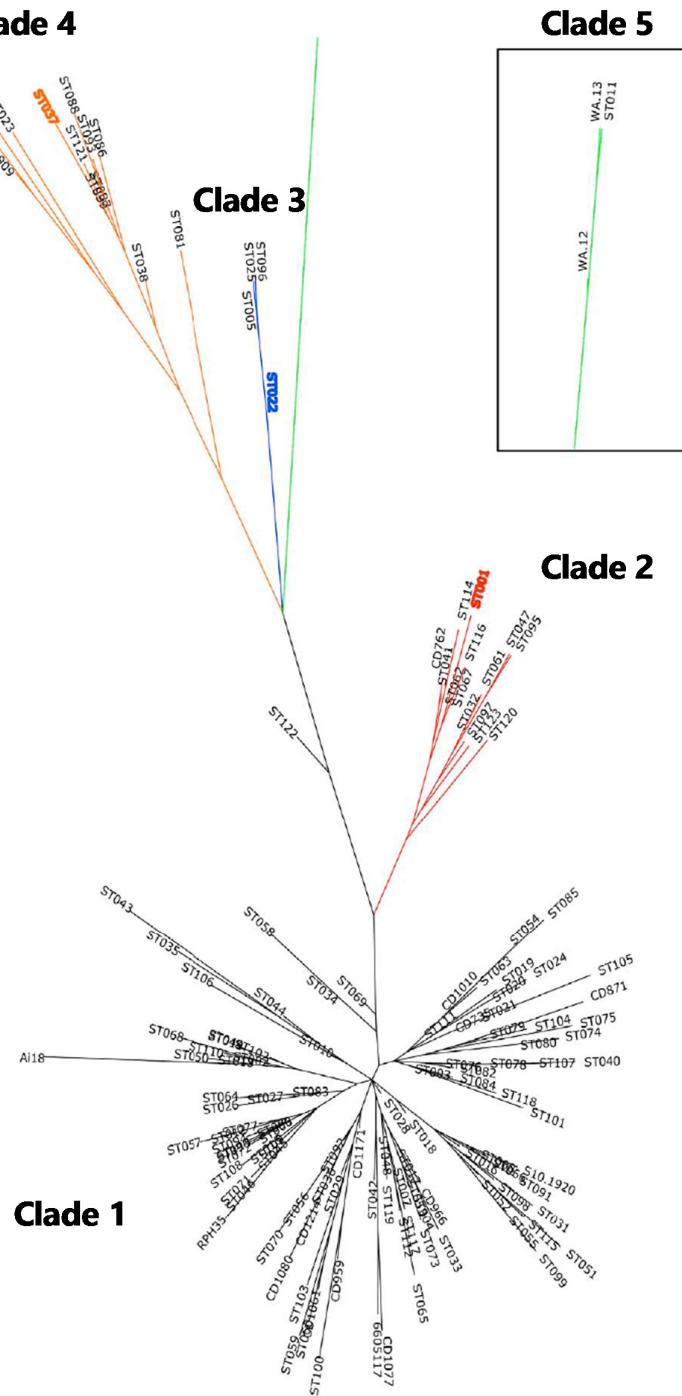


Figure 1.3: Phylogenetic tree of *C. difficile* population structure

This phylogenetic tree, taken from an MLST analysis of 385 diverse *C. difficile* strains, shows the relative evolutionary relatedness of five distinct clonal lineages. Clade 1 is genetically diverse, while Clades 2-5 are clonal in nature. Branch coloring; black = Clade 1 (containing Strain 630), red = Clade 2 (RT027), blue = Clade 3 (RT023), orange = Clade 4 (RT017), green = Clade 5 (RT078). Adapted from Stabler *et al.*, 2012.

1.2.2 The changing epidemiology of *C. difficile* strains

In recent years, the number of reported CDI cases in North American and European hospitals has risen dramatically. Following several hospital outbreaks in Quebec in 2002, the Centre Hospitalier Universitaire de Sherbrooke in Quebec reviewed the charts of its patients with CDI across a 14 year range (1991 to 2003) and noted, not only an increase in CDI incidence, but also an increase in disease severity (Pépin *et al.*, 2004). The authors found that complicated CDI cases, including those resulting in perforated colon or requiring emergency surgery, increased from 7.1% to 18.2% of CDI cases over this time period. Patient death within 30 days of diagnosis also increased from 4.7% of CDI cases seen by the hospital to 13.8% over the same time period (Pépin *et al.*, 2004). Subsequent epidemiological studies revealed this change in the epidemiology of *C. difficile* was attributed largely to *C. difficile* strains of ribotype 027 (RT027), B1 by REA, NAP1 by PFGE, sequence type 1 (ST-1), and toxinotype III (Pépin *et al.*, 2004, Pépin *et al.*, 2005c), representing Clade 2 in Figure 1.3.

Since then, RT027 strains have been reported in hospital outbreaks across the United States and Europe (McDonald *et al.*, 2005, Kuijper *et al.*, 2006a, Kuijper *et al.*, 2006b). The first such outbreak in the UK occurred in 2004 and 2005, at Stoke Mandeville Hospital, which saw two outbreaks attributed to a single RT027 strain, R20291, over that period that resulted in 332 cases of hospital-acquired CDI and 38 deaths (HCC report 2006). Recent RT027 isolates show increased resistance to fluoroquinolones when compared to historical RT027 isolates (McDonald *et al.*, 2005, Razavi *et al.*, 2007). In several of the North American health care facilities that reported RT027 outbreaks, these outbreaks occurred shortly after an increase in the facilities usage of these two antibiotics (McDonald *et al.*, 2005). It is believed that increased use of these broad-spectrum antibiotics during post-surgical prophylaxis has selected for the emergence of this resistant strain (Razavi *et al.*, 2007). In addition to RT027 strains, new groups of virulent *C. difficile* strains have emerged in recent years to cause outbreaks of increased disease severity globally. These include the toxin A-/B+ RT017 strains and RT078 strains (Dawson *et al.*, 2009).

1.2.3 CDI in the community

Though a leading cause of hospital-acquired infection, *C. difficile* also causes infections in the community, and the incidence of community-acquired CDI is increasing. Studies in North America and Europe have reported that up to one third of cases of CDI are community-associated (Karlstrom *et al.*, 1998, Khanna *et al.*, 2012, Leffler & Lamont, 2012, Lessa, 2013). Community acquired CDI cases are less frequently associated with the known risk factors, hospitalisation or antibiotic treatment, and the patients are typically younger in age (Wilcox *et al.*, 2008). *C. difficile* is also recognised as an emerging animal pathogen and has been isolated from a variety of animal species, including horses and livestock (Songer & Anderson, 2006, Hammitt *et al.*, 2008, Keel & Songer, 2007, Rodriguez-Palacios *et al.*, 2006). A retrospective study isolated *C. difficile* from 31 of 278 calves in Canada, with eight ribotypes being represented including RT017 and RT027 (Rodriguez-Palacios *et al.*, 2006). In addition to causing disease in humans, *C. difficile* RT078 is often associated with disease in animals, including pigs and calves (Jhung *et al.*, 2008). This may serve as a reservoir for community-acquired infections, and has implications that *C. difficile* might be potential food borne pathogen.

1.3 *C. difficile* virulence factors

1.3.1 Toxins

Most pathogenic *C. difficile* strains produce two cytotoxins, toxin A (TcdA) and toxin B (TcdB), that glycosylate the Rho and Ras superfamilies of small GTPases (Rupnik *et al.*, 2009). This causes colonic injury through the disruption of the actin cytoskeleton and tight junctions, and the release of cytokines, resulting in inflammation and the destruction of the intestinal epithelium (Voth & Ballard, 2005). Inactivation of the either toxin encoding gene reduces virulence in the hamster model of infection, though the relative importance of each toxin in disease has not yet been determined (Lyras *et al.*, 2009, Kuehne *et al.*, 2011). Non-toxigenic *C. difficile* strains are not usually considered pathogenic (Natarajan *et al.*, 2013). In addition to TcdA and TcdB, approximately 6% of strains also produce an actin modifying ADP-ribosylating binary toxin called *C. difficile* transferase (CDT) (Stubbs *et al.*, 2000, Goncalves *et al.*, 2004, Schwan *et al.*, 2009). CDT is a binary toxin and consists of two components: the enzymatic component, CdtA, and the receptor binding and

translocation component, CdtB (Perelle *et al.*, 1997, Gulke *et al.*, 2001). CDT inhibits actin polymerisation, causes disruptions to the host cell cytoskeleton, and contribute to adherence to epithelial cells (Schwan *et al.*, 2009, Schwan *et al.*, 2014). The role of CDT in *C. difficile* pathogenesis is unclear, but there is evidence that strains which produce binary toxin, irrespective of PCR ribotype, cause more severe clinical symptoms (Bacci *et al.*, 2011).

1.3.2 Flagella

Flagella are important virulence factors in many bacterial species because of their role in motility, cellular adherence, colonisation, and activation of the host innate immune system (Eaton *et al.*, 1996, Barken *et al.*, 2008, Yildiz & Visick, 2009, Kinnebrew *et al.*, 2010, Yoshino *et al.*, 2013). *C. difficile* produce numerous peritrichous flagella, which are distributed uniformly across the bacterial cell surface (Tasteyre *et al.*, 2001, Twine *et al.*, 2009). The major flagella subunit antigen, FliC, is required for the production of flagellar filaments and motility (Tasteyre *et al.*, 2001, Twine *et al.*, 2009). *C. difficile* flagella contribute to adherence within the mouse model; purified flagellin proteins bind to murine mucus, and flagellated strains exhibit a 10-fold increase in adherence to the mouse caecum to non-flagellated strains (Tasteyre *et al.*, 2001). The role of the *C. difficile* flagella in disease remains unclear: despite reacting with patient sera (Pechine *et al.* 2005) and stimulating toll-like receptor five (Yoshino *et al.*, 2013), knockout mutants in both *fliC* and *fliD*, the flagella cap protein, exhibited increased adherence *in vitro* to intestinal Caco-2 cells and increased virulence in the hamster model of infection compared to wildtype (Dingle *et al.*, 2011).

1.3.3 Surface layer proteins

The cell wall of *C. difficile* is typical of Gram-positive bacteria, with a cytoplasmic membrane and a thick peptidoglycan layer. Like many bacteria and archaea, *C. difficile* expresses a surface (S)-layer (Calabi *et al.*, 2001), proteins arranged in a paracrystalline array that completely coat the bacterium and have been shown to be essential for virulence in many pathogens (Sára & Sleytr, 2000). In *C. difficile*, the S-layer is a heterodimeric complex composed of two major S-layer proteins (SLP), a high molecular weight (HMW) SLP and the low molecular weight (LMW) SLP, both of which are derived

from the posttranslational cleavage of the common precursor, SlpA (Calabi *et al.*, 2001) (Figure 1.4). SLPs produced by *C. difficile* may contribute to colonisation of the gut, as they bind strongly to mucosal cells and the underlying lamina propria (Calabi *et al.*, 2002).

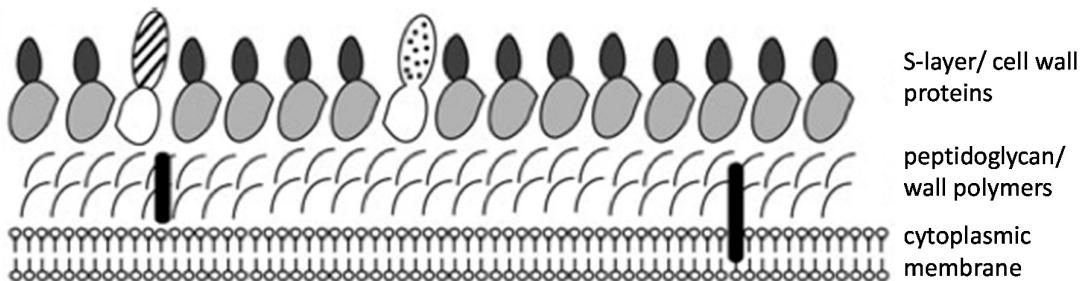


Figure 1.4: Organisation of the *C. difficile* cell wall

The *C. difficile* cell wall is comprised of a thick peptidoglycan layer that is covered by a proteinaceous array called an S-layer. The S-layer is composed of two SLPs that are derived from the cleavage of SlpA: the HMW SLP (light grey), and the LMW SLP (dark grey). Minor CWPs are found within the S-layer framework. Figure from Fagan *et al.*, 2009.

In *C. difficile*, there are a total of 29 cell wall proteins (CWP), including SlpA, that use the cell wall binding 2 (CWB2) domain for non-covalent anchoring to the cell wall (Fagan *et al.*, 2011). Cleavage of the SlpA precursor into HMW SLP and LMW SLP is mediated by the CWP cysteine protease, Cwp84, which has been shown to degrade extracellular matrix proteins, including fibronectin, laminin, and vitronectin (Janoir *et al.*, 2007). *cwp84* knockouts failed to display matured cleaved HMW and LMW SLP, instead the S-layer only contained the full length immature SlpA (Kirby *et al.*, 2009). These changes to the S-layer were observable at the macroscopic level, resulting in significantly altered colony morphology, but had no effect on virulence in the hamster model (Kirby *et al.*, 2009). Another cell wall protein, CwpV, is expressed in a phase variable manner, and post-translationally cleaved into a complex of mature products similarly to SlpA (Emerson *et al.*, 2009). CwpV, the largest of the cell wall proteins, has been shown to promote *C. difficile* aggregation, mediated by a variable C-terminal repeat region (Reynolds *et al.*, 2011). *C. difficile* SLPs and CWPs may play a role in infection, as studies have reported that they are reactive against human antiserum from patients with CDI (Pechine *et al.*, 2005, Wright *et al.*, 2008), and have adhesive properties (Calabi *et al.*, 2002, Waligora *et al.*, 2001).

1.3.4 Other surface proteins

Bacterial cell surface proteins are typically associated with the successful colonisation of a host by pathogenic bacteria and their evasion of host immune responses (Pizarro-Cerda & Cossart, 2006). In addition to the adhesive properties demonstrated by the SLPs and some CWPs (Calabi *et al.*, 2002, Waligora *et al.*, 2001), *C. difficile* expresses several other surface proteins likely to contribute in adherence and colonisation. *C. difficile* expresses a surface exposed fibronectin binding protein, Fbp68 or FbpA, which mediates adherence to epithelial cells (Hennequin *et al.*, 2003, Barketi-Klai *et al.*, 2011). Antibody to Fbp68 can be detected in sera from patients with CDI, indicating that Fbp68 induces a host immune response during infection (Pechine *et al.*, 2005), and fbpA knockout mutants exhibited reduced shedding and colonisation in mice (Barketi-Klai *et al.*, 2011). Another surface exposed protein, CbpA, was recently described to have collagen binding properties (Tulli *et al.*, 2013). In an engineered *Lactococcus lactis* strain expressing CbpA, binding to collagen was significantly increased, as was adherence to extracellular matrix producing cells (Tulli *et al.*, 2013).

Fimbriae, or pili, are short hair-like structures found on many bacteria that are thought to be important in attachment to surfaces and host cells (Pizarro-Cerda & Cossart, 2006). Fimbriae-like structures were first observed on the cell surface of *C. difficile* in 1988 (Borriello *et al.*, 1988), and pilus-like structures have been observed on *C. difficile* in the hamster intestine by electron microscopy (Goulding *et al.*, 2009). However, the role of these structures in colonisation is unclear as no correlation was found between the presence of fimbriae and the relative ability of different strains of *C. difficile* to adhere to hamster gut mucus (Borriello *et al.*, 1988).

1.3.5 Additional *C. difficile* virulence factors

In addition to the toxins and adhesins described above, *C. difficile* encodes other virulence determinants that are hypothesised to contribute toward pathogenesis. *C. difficile* produces a number of hydrolytic and proteolytic enzymes that are thought to play a role in the breakdown of host tissues (Borriello *et al.*, 1990). These enzymes may also facilitate bacterial growth within the gut by releasing essential nutrients (Seddon *et al.*, 1990). *C.*

difficile strains have been observed to produce a capsule (Davies & Borriello, 1990, Baldassarri *et al.*, 1991), which prevented the phagocytosis of *C. difficile* by white blood cells *in vitro* (Dailey *et al.*, 1987). *C. difficile* also produces a phenolic compound, *p*-cresol, that is bacteriostatic and inhibits the growth of other bacteria (Scheline, 1968, Hafiz & Oakley, 1976). It is hypothesised that the production of *p*-cresol may provide *C. difficile* with a competitive advantage over the other gut microflora and contribute to recurrent disease (Dawson *et al.*, 2008, Dawson *et al.*, 2011a). However, not much is known how these factors contribute to infection *in vivo*.

1.4 Sortases in Gram-positive bacteria

Sortases are membrane-bound cysteine transpeptidases that anchor surface proteins to the peptidoglycan cell wall in Gram-positive bacteria. Surface proteins anchored via sortases are often essential virulence factors important in colonisation and invasion, evasion of the host immune system, and nutrient acquisition. The sorting process is mediated by a conserved C-terminal cell wall sorting signal (CWSS) on the anchored protein, comprised of a C-terminal recognition sequence (often LPXTG, where X is any amino acid), followed closely by a hydrophobic transmembrane domain and a positively charged tail (Mazmanian *et al.*, 2001). A conserved active site cysteine residue of the sortase cleaves the CWSS of the polypeptide and covalently attaches the protein to the peptidoglycan (Ton-That *et al.*, 2004). Sortases are classified based on their sequence, and approximately 60% of all sortase proteins can be assigned to six distinct sortase families, A-F (Spirig *et al.*, 2011). Many bacteria encode more than one class of sortase, each catalysing similar transpeptidation reactions, but recognise different CWSS and serve different functions within the cell. Class A enzymes appear to function as housekeeping sortases, as these anchor a variety of proteins to the cell wall. Class B enzymes primarily attach haem-receptors to peptidoglycan, while class C enzymes polymerise pilin subunits. In *Bacillus*, class D sortases anchor proteins involved in sporulation, and in *Corynebacterium diphtheriae*, class E sortases appear to serve a housekeeping function similar to class A enzymes. Little is known about class F enzymes or the sortase enzymes that have yet to be classified.

1.4.1 Sortase transpeptidation

Sortase substrates proteins are translated in the cytoplasm with an N-terminal signal peptide and a C-terminal cell wall sorting signal (CWSS), comprising of an LPXTG-like motif followed by a hydrophobic region and a positively charged tail (Schneewind *et al.*, 1992, Schneewind *et al.*, 1993). All sortases recognise and cleave a specific sorting signal at the C-terminus of their substrate protein(s). This results in the formation of an acyl (thioester) bond between the sortase active site cysteine and a residue in the sorting signal, which is resolved by the nucleophilic attack of a cell wall component (Hendrickx *et al.*, 2011, Schneewind & Missiakas, 2014). The sorting reaction of the *Staphylococcus aureus* class A sortase (SaSrtA) has been studied in most detail (Figure 1.5). SaSrtA cleaves the LPXTG sorting signal between the threonine and glycine residues to form an acyl-enzyme intermediate between the threonine residue of the surface protein and a reactive cysteine in the TLXTC active site of the sortase (Navarre & Schneewind, 1994, Ton-That *et al.*, 1999, Ton-That *et al.*, 2000). Subsequently, the acyl-enzyme intermediate is resolved by nucleophilic attack by the amino group of the pentaglycine cross-bridge within the peptidoglycan precursor lipid-II (Ton-That *et al.*, 1997, Ton-That *et al.*, 1999, Ton-That & Schneewind, 1999, Perry *et al.*, 2002, Ruzin *et al.*, 2002), which in *S. aureus* consists of an UDP-MurNAc-GlcNAc bound to the cell wall pentapeptide [L-Ala – D-Gln – L-Lys – D-Ala – D-Ala] (Schleifer & Kandler, 1972, Dziarski, 2004).

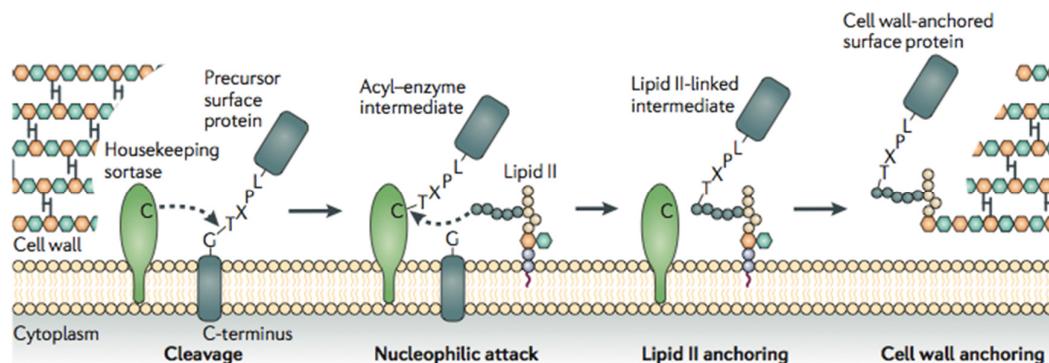


Figure 1.5: Sortase-mediated anchoring of LPXTG surface proteins

Following secretion of sortase substrates by the Sec secretion system, signal peptidases cleave their N-terminal signal peptides resulting in precursor surface proteins. The hydrophobic and charged residues of the CWSS retain the C-terminus of the precursor protein in the cell membrane, allowing the membrane-bound sortase to recognise and interact with the protein. The active-site cysteine of the sortase cleaves the amide bond between threonine and glycine of the LPXTG motif and generates an acyl-enzyme

intermediate. Nucleophilic attack by the pentaglycine cross-bridge of lipid-II (in *S. aureus*) links the C-terminal threonine of the surface protein to the lipid-II peptidoglycan precursor. Penicillin-binding proteins then incorporate this precursor into the peptidoglycan cell wall by catalysing a transpeptidation reaction. Figure from Hendrickx *et al.*, 2011.

The active site cysteine residue is conserved in all sortases to date: substitution of this cysteine (at position 184 in *S. aureus*) with an alanine abolishes sortase catalytic activity *in vitro* and *in vivo* (Ton-That *et al.*, 1999, Ton-That *et al.*, 2000). The crystal structure of sortases and their respective kinetic analyses has led to several different models of their reaction mechanism: it appears the sulphhydryl of the active site cysteine forms a possible catalytic diad with a conserved histidine (Dessen, 2004, Kang *et al.*, 2011), or a possible catalytic triad between these two residues and either an aspartic acid or arginine (Ilangovan *et al.*, 2001, Ton-That *et al.*, 2002, Zong *et al.*, 2004a, Zong *et al.*, 2004b, Zhang *et al.*, 2004, Marraffini *et al.*, 2004). Recently, a reverse protonation mechanism for the catalytic histidine and cysteine residues has been proposed (Frankel *et al.*, 2005, Bentley *et al.*, 2008). This model indicates that one residue must be protonated (histidine) and the other deprotonated (cysteine) for the enzyme to be active, and that at physiological pH a small fraction of the enzyme population (0.06%) in the active reverse-protonated state (Frankel *et al.*, 2005).

1.4.2 Class A sortases – housekeeping sortases

Class A sortases (SrtA), such as the prototypical *Staphylococcus aureus* Sortase A (SaSrtA), are considered housekeeping sortases as they are capable of anchoring many functionally distinct proteins to the cell wall. SaSrtA, which recognises an LPXTG motif, is responsible for anchoring a variety of surface proteins involved in adherence and immune response evasion, and is essential for virulence in animal models (Mazmanian *et al.*, 1999, Mazmanian *et al.*, 2000). SrtA orthologues have been found in the genomes of almost all Gram-positive bacteria (Mazmanian *et al.*, 2000, Pallen *et al.*, 2001, Barnett & Scott, 2002, Bierne *et al.*, 2002, Garandeau *et al.*, 2002, Kharat & Tomasz, 2003, Gaspar *et al.*, 2005). The importance of the C-terminal LPXTG motif, hydrophobic domain, and charged tail was confirmed by Schneewind *et al.* (1992), as deletion of these elements prevented anchoring of protein A (Spa), a surface protein of *S. aureus* that binds to antibodies and

precipitates immunoglobulin, to the cell surface. Mazmanian *et al.* (1999) later discovered that SrtA was responsible for anchoring LPXTG-containing proteins to the cell wall, specifically to the pentaglycine cross-bridge of the peptidoglycan (Schneewind *et al.*, 1995). In a *S. aureus* mutant library screen, this group discovered *srtA* (surface protein sorting A) mutants failed to cleave and anchor Spa. The same group characterised the sortase reaction using purified recombinant SrtA in a fluorescence resonance energy transfer assay (FRET) and identified the essential catalytic cysteine residue (Ton-That *et al.*, 1999).

SrtA is essential for *S. aureus* virulence, as *srtA* mutants are defective in the display of surface proteins, and are severely attenuated in the pathogenesis of murine organ abscesses, infectious arthritis, mastitis, or endocarditis (Mazmanian *et al.*, 2000, Jonsson *et al.*, 2002, Jonsson *et al.*, 2003, Weiss *et al.*, 2004, Cheng *et al.*, 2009, Chen *et al.*, 2013). *S. aureus* strains encode 18-22 LPXTG-containing cell wall proteins that function as microbial surface components recognising adhesive matrix molecules (MSCRAMMS) (Foster & Hook, 1998, Roche *et al.*, 2003, Nandakumar *et al.*, 2005). Besides Spa, SaSrtA substrates include fibrinogen binding proteins (CflA and CflB), fibronectin binding proteins (FnbpA and FnbpB), collagen binding protein, the serine asparate repeat proteins SdrC, SdrD and SdrE, and several uncharacterised staphylococcal surface (Sas) proteins, many of which contribute to staphylococcal infection in mice (Marraffini *et al.*, 2006, Cheng *et al.*, 2009).

Streptococcus pyogenes SrtA (SpSrtA) is a general housekeeping sortase and is responsible for anchoring proteins containing the LPXTG sorting signal to the cell wall (Barnett & Scott, 2002). This includes several key streptococcal virulence proteins, such as the adhesin and main cell wall antigen, M-protein, and the immune evasion factors, G-related α 2-macroglobulin-binding protein (GRAB) and ScpA (Cunningham, 2000, Nobbs *et al.*, 2007). The SpSrtA cell wall anchor is the dialanine peptidoglycan cross-bridge of *S. pyogenes* (Race *et al.*, 2009), which is equivalent to pentaglycine in *S. aureus* (Navarre & Schneewind, 1999).

The genome sequence of *B. anthracis* encodes seven predicted surface proteins with LPXTG sequences (Read *et al.*, 2003). The *B. anthracis* SrtA (BaSrtA) anchors these proteins

to the cell wall by joining the threonine of the LPXTG sequence to the *meso*-diaminopimelic acid within the lipid II peptidoglycan precursor (Gaspar *et al.*, 2005). Many of these proteins remain uncharacterised experimentally, but sequence similarities suggest they include several putative collagen binding proteins (Gaspar *et al.*, 2005). *B. anthracis* *srtA* mutants are unable to grow in macrophages (Zink & Burns, 2005), a presumed early step in the development of inhalation anthrax, but interestingly did not display any virulence defects in mice (Gaspar *et al.*, 2005).

The *L. monocytogenes* genome encodes the highest number of LPXTG proteins among Gram-positive bacteria, with 41 predicted sortase substrates, the most studied of which is the invasion protein, internalin (InlA) (Bierne & Cossart, 2007). InlA contains an LPTTG sequence and is covalently anchored by SrtA to *meso*-diaminopimelic acid residues of the peptidoglycan (Dhar *et al.*, 2000, Garandeau *et al.*, 2002). *L. monocytogenes* *srtA* mutants are defective in the attachment and display of surface proteins, including InlA, are unable to invade cell types for which entry is mediated by InlA, and attenuated in the murine model of infection (Bierne *et al.*, 2002, Garandeau *et al.*, 2002).

1.4.3 Class B sortases – iron acquisition

Class B sortases have sequences with similarity to Sortase B from *S. aureus* (SaSrtB) (Mazmanian *et al.*, 2002). In *S. aureus*, *srtB* is part of an iron-regulated locus *isd* (iron-responsive surface determinant) responsible for haem-iron transport (Mazmanian *et al.*, 2002). The *S. aureus* *isd* locus is expressed under low iron conditions and, in addition to SaSrtB, encodes the cell wall anchored haem-binding proteins IsdA, IsdB, and IsdC, and a membrane-based haem transport system, IsdD, IsdE, and IsdF (Figure 1.6). SaSrtB cleaves IsdC, the only known substrate of SaSrtB in *S. aureus*, between the threonine and asparagine of its NPQTN motif and covalently attaches it within the peptidoglycan layer in a reaction similar to that of SrtA (Mazmanian *et al.*, 2002). IsdA and IsdB contain an LPXTG motif and are anchored by SrtA, combining with IsdC and other Isd proteins encoded by the operon to form a haem-iron transport channel across the cell wall (Muryoi *et al.*, 2008).

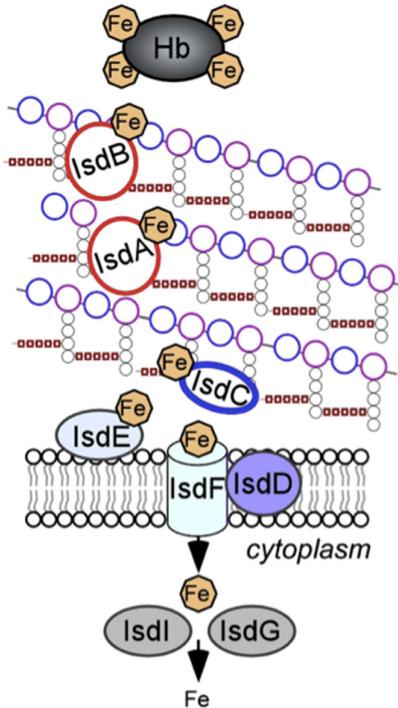


Figure 1.6: Isd-locus mediated haem-iron uptake across the cell wall of *S. aureus*

Haemoglobin binds to IsdB and IsdA, both LPXTG- containing proteins that are anchored to the cell wall by SrtA, and is transferred to the cell wall receptor IsdC, which carries the NPQTN motif and is anchored deeper within the peptidoglycan by SrtB. Haem is transported across the membrane and into the cytoplasm by the IsdDEF membrane transport system, where it is cleaved into free iron by IsdG. Figure from Schneewind & Missiaka, 2014.

B. anthracis and *L. monocytogenes* encode class B sortases that appear to have similar functions in haem-scavenging as SaSrtB (Maresco *et al.*, 2006, Mariscotti *et al.*, 2009). *B. anthracis* has a genomic region with significant sequence identity to the *S. aureus* *isd* operon, including orthologues of IsdA, IsdC, and SrtB (Skaar *et al.*, 2006). SrtB cleaves an NPKTG motif within IsdC between the threonine and glycine residues, with both proteins required for efficient haem-iron scavenging (Maresco *et al.*, 2006). In *L. monocytogenes*, SrtB is responsible for anchoring two surface proteins with sequence similarity to IsdC, SvpA and Lmo2186, to the cell wall (Bierne *et al.*, 2004, Pucciarelli *et al.*, 2005). SvpA and Lmo2186 contain an NAKTN and an NPKSS motif, respectively (Mariscotti *et al.*, 2009), and are encoded by genes clustered in the same iron-regulated operon as *srtB* (Newton *et al.*, 2005), but their role in haem-scavenging has not yet been demonstrated.

The ability to acquire and utilise haem-iron from the host as a source of iron is an important virulence trait of many pathogens. In *S. aureus* and *L. monocytogenes*, SrtB is required for bacterial persistence over time (Mazmanian *et al.*, 2002, Jonsson *et al.*, 2003, Newton *et al.*, 2005). *S. aureus* and *B. anthracis* *srtB* mutants are unable to grow in iron-depleted media that has been supplemented with haem-iron (Mazmanian *et al.*, 2003, Maresso *et al.*, 2006). Although class B sortases share significant sequence identity with SrtB of *S. aureus*, class B members are not limited to anchoring haem-iron binding proteins. In *S. pyogenes*, a class B sortase has recently been shown to assemble pilin subunits, a function thought to be reserved for class C sortases (Kang *et al.*, 2011).

1.4.4 Class C sortases – pili assembly

Class C sortases are involved in the building of pili, structures important in adhesion and biofilm formation. The prototypical SpaA pilus in *Corynebacterium diphtheriae* is assembled in a two-stage process: the class C sortase, CdSrtA, joins the pilin subunits SpaA, SpaB and SpaC to form the pilus (Ton-That & Schneewind, 2003), which is then anchored to the cell wall by a housekeeping class E sortase, CdSrtF (Swaminathan *et al.*, 2007) (Figure 1.7). The pilus tip subunit, SpaC, is cleaved between the threonine and glycine residues of an LPLTG motif, forming a SpaC-sortase complex that is resolved by the nucleophilic attack of a lysine residue from another pilin subunit, SpaA (Ton-That & Schneewind, 2003, Guttilla *et al.*, 2009). SpaA subunits are added to the elongating pilus in this fashion by CdSrtA, forming the majority of the pilus shaft (Ton-That & Schneewind, 2003, Guttilla *et al.*, 2009). SpaB is incorporated into the pilus by a similar mechanism involving a lysine residue, and cleaved at LAFTG by the housekeeping sortase. The addition of SpaB then acts as a switch to terminate pilus polymerisation in favour of cell wall anchoring by the housekeeping sortase (Swaminathan *et al.*, 2007, Mandlik *et al.*, 2008).

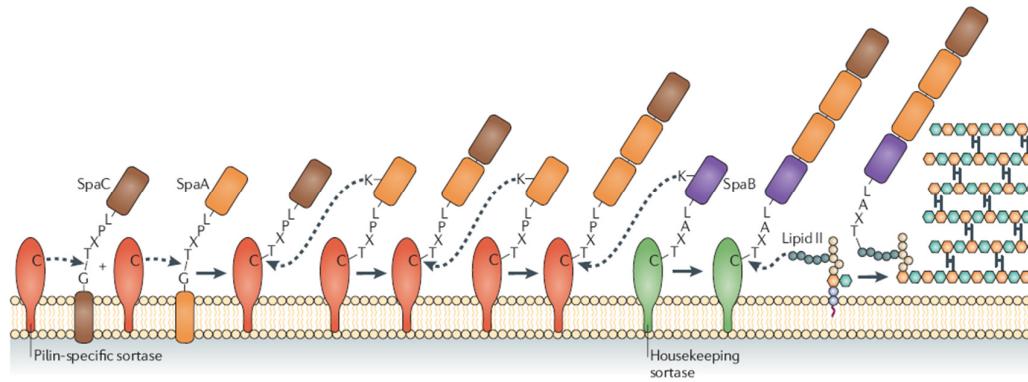


Figure 1.7: Assembly of the SpaCAB pili of *C. diphtheriae* by the class C sortase, CdSrtA

Pilin subunits SpaC and SpaA are typical sortase substrates, containing an N-terminal signal peptide that promotes secretion through the Sec system, and a C-terminal CWSS. The pilin-specific sortase cleaves the sorting signals of SpaC and SpaA at the threonine of the LPXTG motif, generating an acyl–enzyme intermediate. This intermediate is resolved nucleophilic attack of a conserved lysine residue (K) present in an incoming SpaA subunit, resulting in an intermolecular isopeptide bond between SpaC and SpaA, or SpaA with another SpaA subunit. The remainder of the pili is assembled in this manner, until the addition of SpaB. The assembled pilus is then transferred to the cell wall by the housekeeping sortase. Figure from Hendrickx *et al.*, 2011 .

A similar mechanism occurs in pilus assembly in *B. cereus*. The pilus in *B. cereus* contains only two components, the major subunit that forms the shaft, BcpA, and the minor tip pilin, BcpB (Budzik *et al.*, 2007). The class C sortase, BcSrtD, cleaves LPVTG in BcpA and IPNTG in BcpB at the threonine residue. Nucleophilic attack by the side chain lysine in a conserved YPKN motif of BcpA results in an isopeptide bond between BcpB and BcpA or BcpA and BcpA, and catalyses pilus polymerisation (Budzik *et al.*, 2009). Assembly is terminated by the class A sortase, BcSrtA, which attaches the assembled pilus to the diaminopimelic acid moiety within the cell wall (Budzik *et al.*, 2008). In streptococcal species, pilus assembly is more complicated, and requires two or more class C sortases that appear to function redundantly to polymerise pilin subunits (Spirig *et al.*, 2011).

1.4.5 Other sortases

There also exist distinct class D, E, and F sortases, but very little is known about their function (Spirig *et al.*, 2011). The class D sortase has only been characterised in *B.*

anthracis, and is involved in sporulation initiation. Named BaSrtC prior to the established nomenclature, this class D sortase anchors two proteins required for proper spore formation, BasH and BasI, to the cell wall (Marraffini & Schneewind, 2006, Marraffini & Schneewind, 2007). *B. anthracis* *srtC* mutants are less efficient at spore formation under oxygen limiting conditions, though they remain fully pathogenic in mice (Marraffini & Schneewind, 2006). This class D sortase attaches each of its two substrates to a distinct structure within the cell; BasI is attached to the diaminopimelic acid moiety of the peptidoglycan, while BasH is attached exclusively to the forespore. Though *B. anthracis* contains a class A and D sortase enzyme that recognise closely related sorting signals, an LPXTG and an LPNTA sorting signal respectively, the enzymes function non-redundantly suggesting that they may have evolved a high degree of specificity for their respective sorting signals (Marraffini & Schneewind, 2006).

Class E sortases are found in GC rich organisms lacking a class A sortase, and are hypothesised to function as housekeeping sortases in place of SrtA (Ton-That *et al.*, 2004). This is based on the fact that bacteria that encode a class E sortase do not encode a class A sortase, and that genes encoding class E sortases are not adjacent to genes encoding potential substrates (Comfort & Clubb, 2004). Comparative genome analyses suggest that class E enzymes recognise an LAXTG sorting signal, instead of the canonical LPXTG motif processed by class A enzymes (Comfort & Clubb, 2004). The only class E sortase characterised experimentally, CdSrtF from *Corynebacterium diphtheriae*, anchors fully assembled pili to the cell wall (Chang *et al.*, 2011), similarly to many SrtA proteins (Budzik *et al.*, 2007, Nobbs *et al.*, 2007). Class F enzymes are present in *Sreptomyces coelicolor* and other Actinobacteria, but their functions have yet to be determined (Spirig *et al.*, 2011).

1.5 Sortases as drug targets

Numerous *srtA* mutants have been reported to be severely attenuated in various *in vitro* models, despite showing no difference in bacterial viability (Paterson & Mitchell, 2004). Consequently, there is a great deal of interest in the development of sortase inhibitors as a potential strategy against Gram-positive bacterial infections (Mazmanian *et al.*, 2000, Cossart & Jonquieres, 2000). Such compounds would be considered anti-infective agents

that disrupt pathogenesis allowing the host immune system to overcome the bacterium, rather than antibiotics that kill the organism in question. By targeting the infective capability of the organism rather than the organism itself, anti-infectives such as sortase inhibitors would be unlikely to exert a strong selective pressure toward the development of bacterial resistance and would be useful in treating drug-resistant infections. Prior to the discovery of the sortase enzyme, it was observed that sulfhydryl modifying reagents inhibited surface-protein anchoring in *S. aureus* (Ton-That & Schneewind, 1999). When the sortase enzyme, SrtA, was initially characterised in *S. aureus*, it was further reported that methanethiosulfonates and organomercurials inhibit the SaSrtA-mediated cleavage of LPXTG-containing fluorescent peptides (Ton-That *et al.*, 1999). The activity of these two types of compounds is directed at the sulfhydryl group of the essential cysteine residue.

Since this report, different classes of sortase inhibitors have been described, despite having limited or no effect on bacterial cell growth. Several studies have developed substrate-derived sortase inhibitors based on LPXTG sequences, but in which the scissile amide bond between the threonine and glycine residues was replaced chemical groupings known to alkylate the active site thiol group of cysteine proteases (Shaw & Green, 1981, Shaw, 1990). Diazoketone, chloromethyl ketone (Scott *et al.*, 2002), and vinyl sulfone groups (Connolly *et al.*, 2003, Frankel *et al.*, 2004) were shown to be irreversible SaSrtA inhibitors. Vinyl sulfones also inhibited *S. aureus* fibronectin-binding to a magnitude comparable to the behaviour of untreated sortase knockout mutants (Frankel *et al.*, 2004). Screens of small molecule libraries have also identified compounds that irreversibly modify the cysteine sulfhydryl of sortases. Diarylacrylonitriles were identified as SaSrtA inhibitors in a screen of 1000 compounds (Oh *et al.*, 2004), and aryl(B-amino)ethyl ketones were identified as inhibitors of SaSrtA as well as BaSrtA, BaSrtB, and BaSrtC from a screen of over 100,000 compounds (Maresco *et al.*, 2007).

A number of natural alkaloid compounds isolated from plants prevent cleavage of fluorescently labelled peptides by sortases *in vitro*. Extracts from medicinal plants were screened for anti-SaSrtA activity (Kim *et al.*, 2002), resulting in the identification of a glucosylsterol (Kim *et al.*, 2003), berberine chloride (Kim *et al.*, 2004), and kurarinol (Oh *et al.*, 2011) as SaSrtA inhibitors. Some of these compounds also demonstrated activity against SaSrtB, and inhibited *S. aureus* cell adhesion to fibronectin-coated plates (Oh *et*

al., 2006). Isoaptamine, isolated from a marine sponge, is also a potent inhibitor of SaSrtA and reduced *S. aureus* fibronectin-binding (Jang *et al.*, 2007). The spice, turmeric, inhibits the *in vitro* activity of SaSrtA and reduces *S. aureus* adhesion to fibronectin (Park *et al.*, 2005), while the related curcumin not only inhibits the activity of SrtA from *Streptococcus mutans* but also interferes with *in vitro* biofilm formation (Hu *et al.*, 2013b, Hu *et al.*, 2013a). Synthetic analogues of natural products such as these can be refined such that SrtA inhibitory activity is enhanced (Lee *et al.*, 2010b).

Sortase inhibitors have also been demonstrated to reduce virulence in animal models. Vinyl sulfone treatment reduced adhesion and invasion of salmon embryo cells *in vitro* by *Renibacterium salmoninarium*, an invasive bacterial pathogen of fish related to *Micrococcus* (Sudheesh *et al.*, 2007). Similarly, infection and mortality rates were reduced in mice challenged with *S. aureus* when they were treated with diarylacrylonitrile sortase inhibitors (Oh *et al.*, 2010). The ability of sortase inhibitors to affect bacterial adhesion to fibronectin and cells, and reduce virulence in mice supports their potential for further development as therapeutics against Gram-positive infections.

1.6 The sortase and its substrates in *C. difficile*

The first *C. difficile* genome sequence was published in 2006, revealing that strain 630 encodes a single predicted sortase, CD2718, with amino acid similarity to SrtB of *S. aureus* and *B. cereus* (Sebaihia *et al.*, 2006). With regards to the predicted cell wall sorting signal recognised by CD2718 and its potential substrate proteins, the literature is inconsistent and vague. Initially, five substrates containing an LPXTG-like sequon, (S/P)PXTG, were predicted by Pallen *et al.* (2001), while seven substrates, without mention of a sequon, were predicted by Comfort and Clubb (2004). The authors of the 630 genome paper identified 14 potential sortase substrates with an NPQTN motif (Sebaihia *et al.*, 2006), but this manuscript does not elaborate on their role. Recently, the surface associated collagen binding protein, CbpA, from *C. difficile* 630 was characterised and proposed as a substrate of CD2718 with an NVQTG recognition motif (Tulli *et al.*, 2013). However, none of these substrate predictions have been demonstrated experimentally, nor has sortase activity been confirmed for CD2718.

1.7 Aims and objectives

The overall aim of this project is to investigate the function of both the predicted sortase and predicted sortase substrates in *C. difficile*.

The specific aims of this thesis are:

- Identify potential substrates of CD2718
- Characterise the activity of CD2718
 - Purify recombinant CD2718
 - Develop a FRET assay for quantifying CD2718 activity
 - Demonstrate cleavage of sortase substrate(s) *in vitro*
- Investigate the role(s) of the sortase and its predicted substrates in pathogenesis
 - construct and analyse gene inactivation mutants
- Determine the suitability of CD2718 as novel drug target for CDI
 - Design inhibitors of CD2718
 - Determine efficacy of these inhibitors *in vitro* using FRET assay

Chapter 3 of this thesis describes a detailed bioinformatics analysis of the *C. difficile* sortase and of potential substrates identified in this study. Chapters 4 and 5 report on the characterisation of the sortase, and the development and optimisation of a fluorescent based *in vitro* assay to measure the activity of sortase recombinantly expressed in *Escherichia coli*. Additionally, Chapter 5 describes the screening of potential inhibitors of the *C. difficile* sortase. Finally, Chapters 6 and 7 focus on the analysis of the predicted sortase substrates, describing the characterisation of the predicted substrates through mutational analysis in *C. difficile* and recombinant expression in *E. coli*.

2 Materials and Methods

2.1 Materials

2.1.1 Reagents

Reagents were purchased from Sigma-Aldrich Company Ltd. (Dorset, UK) unless otherwise specified. Buffers and media were made up using Milli-Q grade water (Merck Millipore, MA, USA).

2.1.2 Primers

A complete list of oligonucleotides used in this study is described in Table A.1 (in Appendix A). Oligonucleotides were designed using Primer3 Version 0.4.0 software (Untergasser *et al.*, 2007) and checked for secondary structures using NetPrimer software (Premier Biosoft, CA, USA). Oligonucleotides were purchased from Invitrogen (Life Technologies, CA, USA).

2.1.3 Bacterial strains and plasmids used

The bacterial strains and plasmids used in this study are summarised in Table 2.1 and Table A.2 (Appendix A), respectively.

Table 2.1: Bacterial strains used in this study

Strain	Description	Reference
<i>C. difficile</i>		
630	PCR-ribotype 012 isolated in Zurich, Switzerland 1982	Sebaihia <i>et al.</i> , 2006, Monot <i>et al.</i> , 2011
630Δerm	Ery ^S derivative of strain 630, used for ClosTron mutagenesis	Hussain <i>et al.</i> , 2005
R20291	Epidemic PCR-ribotype 027, Stoke Mandeville Hospital outbreak, UK 2005	Stabler <i>et al.</i> , 2009
CD196	Non-epidemic PCR-ribotype 027 isolate, Paris, France 1985	Stabler <i>et al.</i> , 2009
M68	PCR-ribotype 017, Dublin, Ireland 2006	He <i>et al.</i> , 2010
CF5	PCR-ribotype 017, Belgium 1995	He <i>et al.</i> , 2010
CD305	PCR-ribotype 023, London, UK 2008	WTSI (unpublished)
M120	Non-motile, non-epidemic PCR-ribotype 078, UK 2007	He <i>et al.</i> , 2010
630Δerm CD0386::CT	ClosTron knockout in the 630Δerm CD0386 gene	This study
630Δerm CD2831::CT	ClosTron knockout in the 630Δerm CD2831 gene	This study
630Δerm CD3392::CT	ClosTron knockout in the 630Δerm CD3392 gene	This study
<i>E. coli</i>		
CA434	Conjugation donor strain HB101 [<i>thi-1 hsdS20 (r_{B-} m_{B-}) supE44 recAB ara-14 leuB5proA2 lacY1galK rpsL20 (Str^R) xyl-5 mtl-1</i>] carrying plasmid R702 (Kan ^R)	Williams <i>et al.</i> , 1990, Purdy <i>et al.</i> , 2002
One Shot TOP10 Electrocomp	Cloning strain: F ⁻ <i>mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu) 7697 galU galK rpsL (Str^R) endA1 nupG λ</i>	Invitrogen, Life Technologies, CA, USA
Epicurian Coli XL1-Blue Supercompetent cells	Cloning strain: <i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F⁻ proAB lacIqZΔM15 Tn10 (Tet^R)]^c</i>	Agilent Technologies Inc., CA, USA
BL21(DE3)	Protein expression strain: F ⁻ <i>ompT hsdS_B(r_{B-} m_{B-}) gal dcm(DE3)</i>	Invitrogen, Life Technologies, CA, USA
Rosetta(DE3)	Protein expression strain: F ⁻ <i>ompT hsdS_B(r_{B-} m_{B-}) gal dcm(DE3) pRARE (Cam^R)</i>	Novagen, Merck Millipore, MA, USA
NiCo21(DE3)	Protein expression strain: <i>can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamH1o ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnins</i>	New England BioLabs Inc., MA, USA

^S = sensitivity to indicated antimicrobial

^R = resistance to indicated antimicrobial

2.1.4 Fluorescence resonance energy transfer (FRET) peptides

Fluorescently self-quenched peptides were synthesised by Protein Peptide Research Ltd. (Hampshire, UK) and solubilised in DMSO before use. The peptide sequences were based

on the predicted sortase substrate sequences identified in Section 3.4, and are tagged with 5-((2-aminoethyl)amino)naphthalene-1-sulfonic acid (Edans or *e*) as a fluorophore and 4-([4-(dimethylamino)phenyl]-azo)-benzoic acid (Dabcyl or *d*) as a quencher (Matayoshi *et al.*, 1990). All peptides used in this study are listed in Table 2.2.

Table 2.2: FRET peptides used in this study

Peptide sequence*	Description
<i>d</i> -IHSPSTGGG- <i>e</i>	Based on CD0183 sequence
<i>d</i> -IHGSSTPGG- <i>e</i>	Scrambled control for above peptide
<i>d</i> -SDSPKTGGG- <i>e</i>	Based on CD0386, CD3392 sequence
<i>d</i> -SDGSKTPGG- <i>e</i>	Scrambled control for above peptide
<i>d</i> -IHSPQTGGG- <i>e</i>	Based on CD2768 sequence
<i>d</i> -IHGSQTPGG- <i>e</i>	Scrambled control for above peptide
<i>d</i> -PVPPKTGGG- <i>e</i>	Based on CD2831 sequence
<i>d</i> -PVGPKTPGG- <i>e</i>	Scrambled control for above peptide
<i>d</i> -GQNVQTGGG- <i>e</i>	Based on CbpA sequence
<i>d</i> -QALPETGGG- <i>e</i>	SaSrtA peptide
<i>d</i> -NPQTN- <i>e</i>	SaSrtB peptide
<i>d</i> -IHSPSTGKT- <i>e</i>	Based on CD0183 sequence
<i>d</i> -SDSPKTGDN- <i>e</i>	Based on CD0386 sequence
<i>d</i> -IHSPQTGDV- <i>e</i>	Based on CD2768 sequence
<i>d</i> -PVPPKTGDS- <i>e</i>	Based on CD2831 sequence

*Where *d* is Dabcyl (4-([4-(dimethylamino)phenyl]-azo)-benzoyl) and *e* is EDANS ((2-aminoethyl)-amino)naphthalene-1-sulfonyl)

2.1.5 Sortase inhibitors

Potential sortase inhibitors were identified by Domainex Ltd. (Cambridge, UK) and purchased from either ChemBridge Corp. (CA, USA), Enamine Ltd. (Kiev, Ukraine), or Key Organics Ltd. (Cornwall, UK). Inhibitors were solubilised in DMSO before use. The inhibitors tested in this study are detailed in Table A.3 (Appendix A).

2.1.6 Antibodies used for western blotting

The antibodies used for western blotting are detailed in Table 2.3.

Table 2.3: Antibodies used in this study

Primary antibody	Host	Dilution	Source
Anti-His tag	Mouse	1:5000	Abcam PLC, UK
Anti-His tag	Rabbit	1:5000	Abcam PLC, UK
Anti-CBD	Mouse	1:2000	New England BioLabs, USA
Anti-Strep tag II	Mouse	1:100 to 1:500	IBA GmbH, Germany
Anti-HA tag	Rabbit	1:100 to 1:500	Abcam PLC, UK
Anti-CD0183	Mouse	1:2500	N. Fairweather lab
Anti-CwpV type II repeats	Rabbit	1:2500	N. Fairweather lab
Secondary antibody	Host/Conjugate	Dilution	Source
Anti-Mouse	Goat/IRDye 680CW	1:7500	LI-COR Biosciences, USA
Anti-Rabbit	Goat/IRDye 680CW	1:7500	LI-COR Biosciences, USA
Anti-Mouse	Goat/IRDye 800CW	1:7500	LI-COR Biosciences, USA
Anti-Rabbit	Goat/IRDye 800CW	1:7500	LI-COR Biosciences, USA

2.2 Methods

2.2.1 Bacterial growth conditions

Escherichia coli strains were routinely cultured at 37 °C on Luria-Bertani (LB) agar (Novagen, Merck Millipore) or in LB broth (Merck Millipore) with shaking at 170 rpm. Media was supplemented with 100 µg/ml ampicillin, 25 µg/ml chloramphenicol, or 50 µg/ml kanamycin as required.

Clostridium difficile strains were routinely cultured on Brazier's CCEY agar (BioConnections, Knypersley, UK) supplemented with 4% egg yolk (BioConnections) and 1% defibrinated horse blood (TCS Biosciences Ltd., Buckingham, UK), or brain heart infusion agar (Oxoid Ltd., Hampshire, UK) supplemented with 0.05% L-cysteine (BHIS agar). Liquid cultures were grown in brain heart infusion broth (Oxoid Ltd.) supplemented with 0.05% L-cysteine (BHIS broth). Media was supplemented with *C. difficile* antibiotic supplement (250 µg/ml D-cycloserine and 8 µg/ml cefoxitin, BioConnections), 15 µg/ml thiamphenicol, or 20 µg/ml lincomycin as required. *C. difficile* cultures were incubated at 37 °C for 24-48 hours under anaerobic conditions (10% CO₂, 10% H₂ and 80% N₂) in a Whitley MG500 anaerobic workstation (Don Whitley Scientific Ltd., West Yorkshire, UK).

2.2.1.1 Growth kinetics

Growth kinetics of *C. difficile* cultures were assessed in BHIS or YP broth by inoculating single colonies into 10 ml of primary culture, and incubating overnight with shaking at 120 rpm. Secondary cultures were inoculated by a dilution factor of 1/100, and growth dynamics were monitored by OD₆₀₀ readings using a Jenway 6300 spectrophotometer (Scientific Laboratory Supplies Ltd., Wilford, UK). GraphPad Prism 4 (GraphPad Software, Inc., CA, USA) was used to graph growth data.

2.2.2 Bioinformatics

Multiple sequence alignments were created using ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) (Larkin *et al.*, 2007). Putative protein sequences were analysed using the following programs: protein structure predictions were made using the Phyre2 Server (Protein Homology/analogY Recognition Engine V 2.0, <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index> (Kelley & Sternberg, 2009)), conserved protein domains were identified using the Pfam 27.0 families database (<http://pfam.sanger.ac.uk/>) (Punta *et al.*, 2012), transmembrane domains were identified using the TMHMM Server v 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) (Moller *et al.*, 2001), and signal peptide sequences identified using the SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>) (Petersen *et al.*, 2011) or the Phobius web server (<http://phobius.sbc.su.se/>) (Kall *et al.*, 2007). Protein size estimates were made using ExPASy ProtParam Server (<http://web.expasy.org/protparam/>) (Wilkins *et al.*, 1999).

2.2.3 DNA manipulation

2.2.3.1 DNA isolation

Plasmid DNA was isolated from *E. coli* cultures using the QIAprep Spin Miniprep kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. *C. difficile* genomic DNA was isolated from overnight static cultures by cell lysis, phenol/chloroform extraction and ethanol precipitation as previously described (Stabler *et al.*, 2009). Purity and quantification was assessed using a NanoDrop1000 spectrophotometer (NanoDrop Technologies, Inc.). Genomic DNA quality was visualised by agarose gel electrophoresis

with a 1.0% (w/v) agarose gel resolved at 120 volts for 45 minutes and stained with 5 µl of 1% ethidium bromide solution (Fisher Scientific UK Ltd., Loughborough, UK).

2.2.3.2 Polymerase chain reaction

Routine polymerase chain reaction (PCR) amplifications were performed using either MyTaq Red DNA polymerase (Bioline, London, UK) or GoTaq DNA polymerase (Promega Corp., WI, USA) in accordance with the manufacturer's instructions. For cloning purposes, the high-fidelity Phusion DNA polymerase (New England BioLabs Inc., MA, USA) was used following the manufacturer's instructions with genomic DNA from *C. difficile* strain 630 as a template. All PCR amplifications were performed on a DNA Engine Tetrad 2 thermal cycler (Bio-Rad Laboratories Inc., Hertfordshire, UK) and visualised by agarose gel electrophoresis as described in Section 2.2.3.1.

2.2.3.3 Cloning

Details on the construction of the recombinant plasmids listed in Table A.2 (Appendix A) can be found in Appendix B. Generally, restrictions endonucleases and T4 DNA ligase were purchased from either Promega or New England BioLabs, and used in accordance with the manufacturer's instructions. Digested plasmids were treated with 1 µl Antarctic phosphatase (New England BioLabs) for 30 minutes at 37 °C to prevent recircularisation. Site-directed mutagenesis was performed using the QuikChange Site-Directed Mutagenesis kit (Agilent Technologies Inc., CA, USA). PCR products and restriction digest products were routinely purified using the QIAquick PCR Purification kit (Qiagen), or the QIAquick Gel Extraction kit (Qiagen) following agarose gel electrophoresis. Electrocompetent TOP10 *E. coli* cells (Invitrogen) were used for cloning and recombinant plasmid maintenance.

2.2.3.4 Colony PCR

Colony PCR was performed using MyTaq Red DNA polymerase (Bioline) to screen for correct plasmid constructs. An *E. coli* colony served as the DNA template and an additional 5 minute initial denaturing step was incorporated prior to cycling. Plasmid constructs were verified with Sanger sequencing carried out by Source BioScience LifeSciences

(Nottingham, UK). Sequence data was analysed using GENtle v 1.9.4 software (Cologne, Germany).

2.2.3.5 Preparation of electrocompetent *E. coli*

E. coli strains were made electrocompetent to transform with constructed plasmids. Five ml of LB broth was inoculated with a single *E. coli* colony and incubated overnight. Secondary cultures were inoculated with a dilution of 1/100 and cultures were incubated until OD₆₀₀ was between 0.5 and 1.0; after which cells were chilled in 50 ml aliquots on ice for 30 minutes before centrifuging in a chilled rotor at 4000 rpm for 15 minutes. Cell pellets were washed once with 50 ml of cold sterile dH₂O and twice with 4 ml of cold sterile 10% glycerol solution, before resuspending to a final volume of 400 µl in 10% glycerol and separating into 50 µl aliquots for storage at -80 °C.

2.2.3.6 Transformation via electroporation

Plasmid DNA was transferred into *E. coli* cells via electroporation. Ligations were dialysed using a membrane filter disc (Merck Millipore) against nuclease free dH₂O prior to electroporation. Briefly, 50 µl of electrocompetent cells were aliquoted into pre-chilled 0.2 cm gap electroporation cuvettes (Bio-Rad). Plasmid DNA (100-200 ng) was added and cells were electroporated using a Gene Pulser Xcell (Bio-Rad) at 2.5 kV, 25 µF, 200 ohms. Cells were immediately recovered in 500 µl of SOC medium (Invitrogen) and incubated for 1 hour. Transformants were selected for on LB agar supplemented with the appropriate antibiotics.

2.2.4 RNA methods

Ten ml secondary cultures of *C. difficile* 630Δerm in BHIS were inoculated with 1 ml primary cultures (prepared as described in Section 2.2.1.1) and grown to early exponential phase (t=3 hours, OD₆₀₀ = 0.4-0.5), late exponential phase (t=5 hours, OD₆₀₀ = 0.8-0.9), and stationary phase (t=24 hours, OD₆₀₀ = 0.6-0.8). Cultures were incubated with 2 volumes of RNAProtect Bacteria Reagent (Qiagen) for 5 minutes before harvesting via centrifugation at 4000 rpm for 10 minutes at 4 °C. The supernatant was discarded and cell pellets were

processed as follows. For RNA work, nuclease-free water (Ambion, Life Technologies) was regularly used, and pipettes and the workbench were treated with RNaseZap (Ambion).

2.2.4.1 RNA extraction and purification

Total RNA was extracted using the FastRNA Pro Blue Kit (MP Biomedicals LLC., Illkirch, France) according to the manufacturer's protocol. Cell pellets were re-suspended in 2 ml RNAPro Solution (MP Biomedicals) and mechanically disrupted for 40 seconds in a FastPrep-24 Instrument (MP Biomedicals) at speed six. Cell debris and lysing matrix were removed by centrifugation at 13,000 rpm for 15 minutes at 4 °C. The supernatant was transferred to a new tube and washed with 300 µl chloroform (Fisher Scientific) before centrifuging once more. The upper aqueous phase was precipitated overnight at -20 °C in 500 µl 100 % ethanol. Precipitated samples were centrifuged at 13,000 rpm for 15 minutes at 4 °C, and the pellet washed with 500 µl 70 % ethanol before a final centrifugation. The pellet was resuspended in 50 µl nuclease-free water and quantified using the NanoDrop1000 spectrophotometer (NanoDrop Technologies, Inc.).

2.2.4.2 Genomic DNA removal

Genomic DNA was removed from total RNA samples using TURBO DNase (Life Technologies). Four units of TURBO DNase, 80 units RNasin Plus RNase inhibitor (Promega), 5 mM magnesium sulphate (Ambion), and 100 mM sodium acetate (Ambion) were added to 25 µl of eluted total RNA to a final reaction volume of 150 µl and incubated for one hour at 37°C. An additional 4 units of TURBO DNase was added and incubation repeated. DNase-treated RNA samples were cleaned using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA was quantified using the NanoDrop1000 spectrophotometer (NanoDrop Technologies, Inc.). To confirm genomic DNA depletion, PCRs were performed using the treated RNA samples with oligonucleotide pairs specific to the following housekeeping genes: *atpA*, *sigB*, and *16sRNA*. RNA quality was also assessed using a Small RNA Chip (Agilent Technologies, Berkshire, UK) on a 2100 Bioanalyzer (Agilent Technologies).

2.2.4.3 Reverse transcription of total RNA

Equal amounts of RNA were reverse transcribed into complementary DNA (cDNA) for expression analysis as described below. Briefly, one µg random hexamer primers (Invitrogen) and 40 units RNasin Plus RNase inhibitor (Promega) was added to one µg RNA in a final volume of 10 µl, and incubated at 65°C for 10 min followed by room temperature for 30 min. The following first-strand mixture was added for cDNA synthesis: four µl of 5x first-strand buffer (Invitrogen), two µl 0.1 M DTT (Invitrogen), two µl 10 mM dNTP mix (New England BioLabs), and 1.5 µl Superscript II (Invitrogen). The reaction mixture was incubated at 25 °C for 10 minutes, 42 °C for 1 h, and finally 70 °C for 15 minutes. Confirmation of cDNA synthesis was performed using 1 µl of cDNA in a PCR reaction with oligonucleotide pairs to amplify the housekeeping genes. Samples were stored at -80°C.

2.2.5 *C. difficile* mutagenesis

The ClosTron mutagenesis system (Heap *et al.*, 2007, Heap *et al.*, 2010) was used to construct insertional mutants in *C. difficile* 630Δerm. This method is based on a mobile group II intron from the *ltrB* gene of *Lactococcus lactis* (Li.ltrB) modified using a computer algorithm (Perutka *et al.*, 2004) to target the gene of interest. Within the LL.ltrB intron is an erythromycin retrotransposition-activated marker (RAM) that allows for the selection of integration of the intron into the *C. difficile* genome.

2.2.5.1 Intron retargeting

The genes of interest (the sortase *CD2718*, three of the predicted substrates *CD0386*, *CD2831*, and *CD3392*) were analysed by the Intron Targeting and Design Tool (<http://clostron.com/>) to identify insertion sites for the Li.ltrB intron using the Perukta algorithm (Perutka et al 2004). The retargeted introns were synthesised and cloned into the HindIII and BsrGI sites of pMTL007C-E2 by DNA2.0 Inc. (CA, USA) to construct the following plasmids: pEHD002, pEHD003, pEHD004, and pEHD005. The sequences of the retargeted introns in pMTL007C-E2 were confirmed using primers cspfdxF1 and pMTL007-R1.

The plasmid pEHD001 was previously constructed by L. Dawson in the Wren lab. Target sites within *CD2718* were identified, and *Li.ltrB* intron retargeting PCR primers designed using the Perutka algorithm (Perutka *et al.*, 2004) as part of the TargeTron Gene Knockout System kit (Sigma-Aldrich). One-tube SOEing PCR was used to assemble and amplify the 350 bp PCR product containing the modified IBS, EBS1 δ , and EBS2 regions of the *Li.ltrB* intron, in accordance with the kit instructions using primers IBS-213s, EBS1 δ -213s, EBS2-213s, and the EBS Universal. The sequence of the retargeted intron in pMTL007 was confirmed using primers pMTL007-F1 and pMTL007-R1.

2.2.5.2 Conjugations

These constructs were transformed into electrocompetent *E. coli* CA434 (Williams *et al.*, 1990) prior to conjugation into *C. difficile* 630 Δ erm as previously described (Purdy *et al.*, 2002). Briefly, the *E. coli* CA434 donor strain, carrying the pMTL007 or pMTL007C-E2 plasmids, was grown overnight in five ml of LB broth supplemented with chloramphenicol and kanamycin. In parallel, the recipient *C. difficile* 630 Δ erm strain was grown overnight in BHIS broth. One ml of *E. coli* CA434 cells was pelleted at 2000 rpm for one minute, before gently washing with 500 μ l PBS and centrifuging as before. The supernatant was removed and the pellet was taken into the anaerobic chamber, where it was resuspended in 150 μ l of overnight 630 Δ erm culture. The mixture was then plated as 10 μ l drops onto non-selective BHIS agar and incubated anaerobically for 24 hours. Growth was scraped off the plates, and suspended in 400 μ l PBS. Transconjugant *C. difficile* were selected by plating the suspension on BHIS plates supplemented with thiampenicol (to select for *C. difficile* cells carrying the ClosTron plasmid) and *C. difficile* antibiotic supplement (to select against the *E. coli* donor strain), and incubated for three days.

2.2.5.3 Mutant selection

In the first generation ClosTron plasmid, pMTL007, expression of the *Li.ltrB* intron is under the control of an IPTG-inducible promoter from *C. acetobutylicum*. *C. difficile* 630 Δ erm transconjugants colonies carrying the pEHD001 plasmid were suspended in one ml BHIS supplemented with thiampenicol and one mM IPTG (Fisher Scientific), and incubated anaerobically for one hour to induce *Li.ltrB* expression before centrifuging at 8,000 rpm for 1 minute. Cells were washed with 500 μ l PBS before resuspending in 300 μ l non-

selective BHIS, and serially diluting in PBS and spreading onto pre-equilibrated BHIS agar supplemented with lincomycin.

In the second generation ClosTron vector, pMTL007C-E2, the constitutive promoter from the *fdx* gene in *C. sporogenes* has replaced the inducible promoter in pMTL007, so IPTG induction is not necessary. *C. difficile* 630Δerm transconjugant colonies carrying the pEHD002-pEHD005 plasmids were suspended in 500 µl PBS, serially diluted and spread onto pre-equilibrated BHIS agar supplemented with lincomycin.

In both cases, lincomycin supplemented BHIS plates were incubated for three days to allow for growth of ClosTron integrants. Lincomycin resistant colonies were subcultured onto BHIS supplemented with lincomycin twice more to confirm lincomycin resistance. These colonies were then replica-plated onto thiamphenicol, and lincomycin-supplemented BHIS agar. Potential ClosTron mutants were identified as thiamphenicol sensitive (indicating loss of the ClosTron plasmid) but lincomycin resistant (indicating genomic integration of LI.ltrB intron).

2.2.5.4 PCR screening mutant clones

DNA from potential *C. difficile* mutants was isolated by Tris-EDTA boilate as previously described (Griffiths *et al.*, 2010). Briefly, a few colonies of putative mutant clones were emulsified in Tris-EDTA buffer, and heated at 100 °C for 10 minutes. Cell debris was removed by centrifugation at 13,000 rpm for 2 minutes, and the supernatant removed for screening PCRs.

Two PCRs were performed to confirm successful integration of the LI.ltrB intron into the target genes: 1) insertion into the target gene was confirmed using a gene-specific primer near to the predicted insertion site (CD2718-F, CD0386-F, CD2831-F, or CD3392-F), and the intron-specific primer EBS Universal, and 2) splicing of the group I intron from the RAM element and loss of the pMTL007C-E2 plasmid was confirmed using primers RAM-F and RAM-R. A third PCR using primers flanking the insertion site (CD2718-F and CD2718-R) was also performed to screen potential mutants in *CD2718* for a size shift corresponding to integration of the LI.ltrB intron.

2.2.5.5 Southern blot

Genomic DNA was isolated from PCR-positive mutant clones, and single insertion of the Ll.ltrB intron was verified by Southern blot hybridisation using the AlkPhos Direct Labelling and Detection System (GE Healthcare Life Sciences, Buckinghamshire, UK). An intron-specific probe was amplified from pMTL007C-E2 using primers IBS-213s and EBS1δ-213s and labelled in accordance with the manufacturer's instructions. Genomic DNA from mutant clones and wildtype 630Δerm was digested for 8 hours with the following restriction enzymes: BtgI and PciI for *CD0386* and *CD3392* mutants, and BsaBI for *CD2831* mutants. Digests were separated by agarose gel electrophoresis and transferred to a nylon Hybond-N+ membrane (GE Healthcare) according to standard techniques before probe hybridisation at 55 °C overnight, and subsequent detection with CDP-Star (GE Healthcare).

2.2.6 Phenotypic tests

2.2.6.1 Visualisation of *C. difficile* colony morphology

Primary BHIS broth cultures were prepared as described in Section 2.2.1.1. Two µl of overnight primary culture was spotted onto pre-equilibrated BHIS agar and incubated for 72 hours. Plates were removed from the anaerobic chamber and colonies photographed using a Canon 600D SLR. For microscopic visualisation of *C. difficile* colony morphology, plates were visualised using a Zeiss Axioplan 2 upright microscope (Carl Zeiss Ltd., Cambridge, UK). Microscope images were analysed using Volocity Image Analysis 6.3 software (PerkinElmer, MA, USA). Experiments were performed in triplicate.

2.2.6.2 Autoagglutination assay

Autoagglutination was measured as previously described (Stabler *et al.*, 2009). *C. difficile* strains were grown on Brazier's CCEY agar for 1-2 days and then inoculated into pre-reduced PBS to an OD₆₀₀ 1.0 (+/- 0.1). Then two ml of the cell suspension was added to sterile glass tubes and incubated for 24 hours at 37 °C, after which one ml was removed from the tube surface to measure the OD₆₀₀. The results were normalised to the starting OD using the following equation: 100-((Final OD₆₀₀ - Starting OD₆₀₀) × 100) to show the

actual autoagglutination percentage. Strain M120 was used as a positive control as this strain autoagglutinates more than 95%, and strain R20291 was used as a negative control as it exhibits a lower percentage of autoagglutination. Experiments were performed in triplicate. GraphPad Prism 4 (GraphPad Software, Inc., CA, USA) was used to graph data and perform an unpaired two-tailed Student's *t*-tests with a confidence interval of 99% ($p<0.01$).

2.2.6.3 Bacterial sedimentation assay

The propensity of *C. difficile* cells to sediment at the bottom of a liquid culture during overnight growth was assessed using an assay developed in the Wren lab (A. Faulds-Pain, manuscript in preparation). Five ml BHIS in glass vials was inoculated with a single *C. difficile* colony and incubated for 24 hours. Sedimentation of the cultures was visualised using a Canon 600D SLR. Strain M120 sediments strongly in this assay and was used as a positive control. Experiments were performed in triplicate.

2.2.6.4 Motility assay

Relative motility was determined according as described previously (Twine *et al.*, 2009). Briefly, *C. difficile* cultures were grown anaerobically for 1-2 days on Brazier's CCEY agar. BHIS with 0.175% agar was poured into 5 ml glass vials and pre-equilibrated overnight in the anaerobe chamber. Single colonies were stabbed into the top 2-5 mm of the semi-solid agar in and incubated overnight in the anaerobe chamber. The tubes were visualised using a Canon 600D SLR to record relative motility. Strain M120 is non-motile and was used as a negative control. Experiments were performed in triplicate.

2.2.6.5 Biofilm assays

Biofilm formation of *C. difficile* strains was assessed by a crystal violet staining assay as previously described (Dawson *et al.*, 2012). Primary BHIS broth cultures were prepared as described in Section 2.2.1.1. To assess biofilm formation, the primary cultures were used to inoculate tissue culture (TC) flasks and 24-well plates. Duplicate 10 ml BHIS broth (supplemented with 1% cysteine instead of 0.05% cysteine – BHISc) in 25 ml TC flasks (BD Biosciences, MA, USA) were inoculated with a 1/10 dilution of primary culture and

incubated statically for three days prior to being removed from the anaerobic chamber. The biofilm mass was mechanically detached from the bottom of the TC flask by gentle agitation as described in Dawson *et al.*, (2012) prior to photographing with a Canon 600D SLR.

Two ml pre-equilibrated BHISc broth in 24-well plates (Nalge Nunc International, NY, USA) was inoculated with a 1/10 dilution with primary cultures, and incubated statically for three days alongside three media blank controls. To prevent evaporation, plates were sealed with Parafilm (Sigma-Aldrich) after inoculation. After incubation, the plates were removed from the anaerobic chamber. The media was removed by pipetting, after which wells were washed once with 700 µl PBS prior to incubation for 30 minutes at room temperature with 700 µl of filter-sterilised 1% (v/v) crystal violet (TCS Biosciences). Excess crystal violet was removed from the wells, after which wells were washed three times with 700 µl PBS. Images were taken of the 24-well plates using a Canon 600D SLR. The crystal violet was then extracted by adding 700 µl methanol to each well, and incubating for 15 minutes at room temperature. The OD₅₉₅ was measured using an ELx800 Absorbance microplate reader (Biotek, VT, USA). All assays were performed with a minimum of three biological and three technical replicates. GraphPad Prism 4 (GraphPad Software, Inc., CA, USA) was used to graph data and perform an unpaired two-tailed Student's *t*-tests with a confidence interval of 99% (*p*<0.01).

2.2.7 Protein methods

2.2.7.1 SDS-PAGE

Samples were mixed 1:1 with Laemmli sample buffer, boiled for 5 minutes, and then separated on NuPAGE® Novex 10% Bis-Tris gel (Invitrogen) in 1x MOPS SDS running buffer (Invitrogen). Gels were either stained with Coomassie Brilliant Blue R-250 (Thermo Scientific) for visualisation or transferred to nitrocellulose for western blotting according to standard techniques.

2.2.7.2 Fluorescent western blots

SDS-PAGE separated samples were transferred to Hybond-C Extra nitrocellulose (GE Healthcare) using a SemiPhor semi-dry transfer unit (Hoefer Inc., MA, USA). Membranes were rinsed with water before blocking with 2% skimmed milk (Marvel, Hertfordshire, UK) in PBS for one hour. The membrane was incubated with primary antibody (detailed in Table 2.3) in PBS, 2% skimmed milk, 0.1% Tween-20 for one hour before three separate five minute washes with PBS, 0.1% Tween-20, and a single five minute wash with PBS alone. The membrane was then incubated with 1:7500 dilution of appropriate IRDye-conjugated secondary antibodies (detailed in Table 2.3) in PBS, 2% skimmed milk, 0.1% Tween-20, and 0.01% SDS for 45 minutes before washing with PBS, 0.1% Tween-20, and then PBS as before. Due to the light sensitivity of the IRDye-conjugated secondary antibody, incubation with the antibody and the subsequent washes were carried out in the dark. Membranes were visualised using an Odyssey near-infrared imager (LI-COR Biosciences, NE, USA) using the appropriate channels for detection of the IRDye conjugated to the secondary antibody (either 680 nm or 800 nm).

2.2.7.3 Protein expression

For 10 to 200 ml expression cultures, recombinant protein was expressed as follows: to generate primary cultures for protein expression, 10 ml LB broth with appropriate antibiotic selection was inoculated with a single *E. coli* colony and incubated with shaking overnight. *E. coli* cultures for protein expression were inoculated with a 1/20 dilution of overnight primary culture and incubated until the OD₆₀₀ was 0.4-0.6. A volume of culture was removed corresponding to OD₆₀₀ = 1 (ex. two ml of a culture with OD₆₀₀ = 0.5) and pelleted to serve as a non-induced control, prior to inducing the remaining culture with 1 mM IPTG (Fisher Scientific). After inducing for the desired length of time, cells were harvested and stored as cell pellets at -20 °C until processing.

For large-scale two litre expression cultures, recombinant protein was expressed as above with the following differences: 200 ml primary cultures were inoculated either with a single *E. coli* colony or directly from glycerol stocks stored at -80 °C. Two litres LB broth was inoculated with a 1:10 dilution of the primary cultures and induced with 1 mM when

the OD₆₀₀ reached 0.4-0.6. Cells were harvested after five hours of induction and stored as cell pellets at -20 °C until processing.

2.2.7.4 Cell lysis

Frozen cell pellets were thawed on ice and resuspended in 5 ml Lysis buffer [1x Bugbuster (Novagen) diluted in 50 mM NaH₂PO₄, 300 mM NaCl, 20-40 mM imidazole (Fisher Scientific), 1 mM DTT, pH 7.5] per gram of cell pellet. Lysates were sonicated on ice for 2 min (15 sec on/off) at 50 % Vibra Cell high intensity ultrasonic processor (Jencons, Leighton Buzzard, Bedfordshire, UK). One mg/ml lysozyme and 25 U/ml Benzonase Nuclease (Novagen) was added to the cell suspension, and incubated for 20-60 minutes with gentle shaking at room temperature until no longer viscous. Insoluble cell debris was removed by centrifugation at 13000 rpm for 45 minutes at 4 °C, after which the soluble cell lysate was removed and passed through a 0.22 µm filter (Merck Millipore) prior to nickel purification.

2.2.7.5 Nickel purification using Ni-NTA Agarose

Ni-NTA agarose (Qiagen) was pre-equilibrated with Binding buffer [50 mM NaH₂PO₄, 300 mM NaCl, 20-40 mM imidazole (Fisher Scientific), 1 mM DTT, pH 7.5]. An aliquot of Ni-NTA agarose slurry was pelleted at 4000 rpm for 5 minutes. The supernatant was then removed and the Ni-NTA slurry washed three times with an equal volume of Binding buffer. A 1/20 volume of pre-equilibrated Ni-NTA agarose slurry was added to soluble cell lysates and incubated for 1-2 hours with gentle shaking at 4 °C. For volumes less than 2 ml, the cell lysate and Ni-NTA slurry mixture was added to a microcentrifuge spin cup (Thermo Scientific) 500 µl at a time and centrifuged at 4000 rpm for one minute. The flow through was discarded and the collected resin was washed with 5x volume (of cell lysate) of Binding buffer, before eluting with 2x volume of Elution buffer [50 mM NaH₂PO₄, 300 mM NaCl, 500 mM imidazole (Fisher Scientific), 1 mM DTT, pH 7.5]. For volumes greater than 2 ml, the cell lysate and Ni-NTA slurry mixture was added to a 10 ml disposable polypropylene column (Thermo Scientific). Samples were washed with a 5x volume of Binding buffer before eluting recombinant protein by gravity flow using 2x volumes of Elution buffer. Eluted samples were stored at 4 °C on ice until processed.

2.2.7.6 Purification using chitin resin

To remove *E. coli* NiCo21(DE3) contaminant proteins tagged with a chitin binding domain (CBD) (Robichon *et al.*, 2011), eluted fractions obtained after Nickel affinity chromatography were pooled and incubated with one ml chitin beads (New England BioLabs), previously equilibrated with 5 ml buffer C (50 mM NaH₂PO₄, 500 mM NaCl, pH 7.4). After gentle shaking for 30 minutes at 4 °C, the target protein was eluted from the chitin slurry by gravity flow, and the chitin beads washed with five ml buffer C. The chitin eluate was analysed by SDS-PAGE and immunoblotting using an anti-CBD monoclonal antibody to check for removal of the CBD-tagged proteins from the target recombinant protein sample.

2.2.7.7 Nickel purification using HPLC

An AKTApurifier system (GE Healthcare) was used to purify His-tagged proteins from two litre induced *E. coli* cultures. Cell lysates prepared as described in 2.2.7.4 were applied to a one ml HisTrap HP column (GE Healthcare), pre-equilibrated with Binding buffer, using an LKB 2132 Microperpex peristaltic pump (Pharmacia LKB Biotechnology, Sweden) at 30 rpm (one ml/min flow rate). The loaded column was then attached to the AKTA system and washed with 10-20 ml Binding buffer at a rate of one ml/min until the UV_{280nm} baseline stabilised. Recombinant protein was eluted by applying an imidazole gradient (40 – 500 mM) over 25 column volumes. Any peaks observed in the UV_{280nm} trace were analysed by SDS-PAGE and immunoblotting to determine which fractions contained the protein of interest.

2.2.7.8 Gel filtration

Gel filtration was used to further purify the eluted protein obtained from nickel affinity chromatography on the AKTA (Section 2.2.7.7), and to buffer exchange the protein into the appropriate downstream buffer. Fractions containing the recombinant protein (as identified by SDS-PAGE) were pooled and injected onto a HiLoad 16/60 Superdex 200 column (GE Healthcare), pre-equilibrated with two column volumes of FRET buffer (5 mM CaCl₂, 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT). Separation was performed at one ml/min over one column volume, and peaks observed in the UV_{280nm} trace were

analysed by SDS-PAGE and immunoblotting to determine which fractions contained the protein of interest.

2.2.7.9 Protein quantification

Eluted fractions containing recombinant protein were pooled and concentrated using an Amicon Ultra-15 (10 kDa) centrifuge filter unit (Merck Millipore). Concentrated protein samples were quantified using BCA assay (Thermo Scientific, for samples without DTT) or Bradford reagent (Thermo Scientific, for samples with DTT) and analysed by SDS-PAGE.

2.2.7.10 Identification of recombinant proteins

The identity of purified recombinant proteins was confirmed either by N-terminal sequencing (Edman degradation) or MALDI fingerprinting. For Edman degradation, samples were separated by SDS-PAGE and transferred to a Hybond-P PVDF membrane (GE Healthcare). The membrane was washed in water for 10 minutes and then stained in 0.1% Coomassie Blue in 50% methanol for 5 minutes. After destaining (50% methanol, 10% acetic acid) twice for 5 minutes and washing twice in water for 5 minutes, the membrane was allowed to dry thoroughly. N-terminal sequencing by Edman degradation was performed by the PNAC Facility (Department of Biochemistry, University of Cambridge, UK). For MALDI fingerprinting, samples were separated by SDS-PAGE and stained by Coomassie. Protein bands were digested by trypsin and analysed by MALDI-TOF performed by Jun Wheeler (NIBSC, PHE, Hertfordshire, UK), or by the PNAC Facility.

2.2.8 *In vitro* analysis of enzyme activity

2.2.8.1 Fluorescence resonance energy transfer (FRET) assays

Recombinant CD2718 activity was monitored using a FRET (Fluorescence Resonance Energy Transfer) assay in FRET buffer. Fluorescently self-quenched peptides tagged with 5-((2-aminoethyl)amino)naphthalene-1-sulfonic acid (Edans) as a fluorophore and 4-([4-(dimethylamino)phenyl]-azo)-benzoic acid (Dabcyl) as a quencher (Matayoshi *et al.*, 1990), and containing the predicted sorting signals of SrtB were purchased from Protein Peptide Research Ltd and solubilised in DMSO (Table 2.2). The FRET assay was performed in a final volume of 100 µl buffer F containing 10 µM SrtB_{ΔN} and 20 µM fluorogenic peptide

in clear-bottomed, black polystyrene 384-well plates (Nunc). Plates were incubated for 48 hours at 37 °C, during which fluorescence (excitation = 340 nm, emission = 490 nm) was measured using a SpectraMax M3 plate reader (Molecular Devices, LLC., CA, USA). Five mM [2-(trimethylammonium)ethyl]methanethiosulfonate (MTSET), 150 µM hydroxylamine (NH₂OH, Sigma) and 150 µM 2,6-diaminopimelic acid (Sigma) were added to the reactions as indicated. Each experiment was performed in triplicate with a minimum of 3 biological replicates, and the results are presented as the means and the standard error of the data obtained.

2.2.8.2 Data analysis

FRET data was exported from the SoftMax Pro v5.4 software in comma separated value (CSV) format. The data were then imported and compiled in Python 2.7.3 (<http://www.python.org>). A custom Python application was used to analyse the data (script detailed in Appendix C). The mean and standard error of the data was then graphed in Excel.

For curve fits, the following SciPy 0.11.0 packages for Python were used: the Stats package for linear regression (<http://docs.scipy.org/doc/scipy-0.13.0/reference/generated/scipy.stats.linregress.html>) and the Optimisation package (<http://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.leastsq.html#scipy.optimize.leastsq>) for the least squares curve fits to Equation 2.1 and Equation 2.2. The plotting package Matplotlib (<http://matplotlib.org>) was used to graph the data.

2.2.8.3 Analysis of FRET products

MALDI-TOF analysis of FRET reaction samples was performed by the PNAC Facility to determine exact cleavage site within each peptide.

2.2.8.4 Enzyme kinetics

Kinetic data for CD2718_{ΔN26} were obtained by incubating varying concentrations of peptide (8, 10, 20, 40, 80, 160, 200 and 240 µM) with 10 µM CD2718_{ΔN26}. All reactions were performed as described in Section 2.2.8.1, with fluorescence monitored every ten

minutes over a 13 hour period. To correlate fluorescence signal, expressed as Relative Fluorescence Units (RFU), with concentration of product formed, the standard curves of the fluorophore, Edans (Edans acid, Eurogentec, Liège, Belgium), in the absence and presence of an equal concentration of the quencher, Dabcyl (Dabcyl acid, Eurogentec), were collected. The presence of the Dabcyl decreased the fluorescence of the Edans molecule, but this effect was minimal at concentrations below 2 µM. The linear segment of the fluorophore standard curve generated a conversion ratio of 703.9 RFU/ µM Edans.

Initial velocities (V) were determined from the progress curves and plotted against substrate concentration [S]. The data were fitted to a modified version of the Michaelis-Menten equation incorporating substrate inhibition (Equation 2.1), where V_{max} is the apparent maximal enzymatic velocity, K_m is the apparent Michaelis constant, and K_i is the apparent inhibitor dissociation constant for unproductive substrate binding. All data points were collected in triplicate, and the overall assay was run in duplicate.

Equation 2.1: Michaelis-Menten equation

$$V = \frac{V_{max} \cdot [S]}{K_m + [S] + \frac{[S]^2}{K_i}}$$

2.2.8.5 Design and testing of sortase inhibitors

The proprietary *LeadBuilder* virtual screening method (Domainex, Ltd) was used to interrogate a database (*PROTOCOLS*) of 80,000 potential compounds which had been pre-selected as protease inhibitors. The virtual screening protocol used pharmacophoric and docking filters derived from analysis of the BaSrtB crystal structure (with which the CD2718 shows 70% identity and 90% similarity at the active site). Sixty-two compounds identified in this screen as potential inhibitors of CD2718 were obtained from Enamine, ChemBridge, and Key Organics, and solubilised in DMSO (listed in

Table A.3). Selected compounds and MTSET were incubated with 10 µM CD2718_{ΔN26} at a range of concentrations in the FRET-based assay conditions described above, so that final DMSO concentrations were ≤ 3.75%.

IC₅₀ values were determined by non-linear least squares fit to a sigmoidal function (Equation 2.2), where x is the concentration of the inhibitor and y is the relative fluorescence observed after 48 hours. This equation describes concentration-response curves using four parameters: I_{min} is the lowest observed value, I_{max} is the highest observed value, S is the slope, IC₅₀ is the inflection point of the sigmoidal curve and refers to the concentration leading to 50% of I_{max}. All data points were collected in triplicate, and the overall assay was run in triplicate.

Equation 2.2: Sigmoidal concentration-response curve

$$y = I_{min} + \frac{I_{max} - I_{min}}{1 + \left[\frac{x}{IC_{50}} \right]^S}$$

2.2.8.6 Effect of sortase inhibitors on *C. difficile* culture

The positive sortase inhibitors from the FRET assay were tested for inhibitory activity on *C. difficile* growth. Five hundred µM MTSET, and the highest possible concentration of each inhibitor was tested at which the final DMSO concentrations were ≤ 3.75%. 3.75% DMSO was tested to control for the presence of the solvent in the inhibitor sample. Two *C. difficile* strains, 630Δerm and R20291, were tested to assess the potential for strain to stain variation. Ten ml secondary cultures of *C. difficile* in BHIS were inoculated with 1 ml primary cultures (prepared as described in Section 2.2.1.1) and grown until OD₆₀₀ = 0.3. One hundred µl of culture was added to each well of a 24-well plate (Nalge Nunc International, NY, USA) containing 900 µl of BHIS supplemented with the sortase inhibitor at the desired final concentration. Plates were sealed with Parafilm to prevent evaporation, and incubated statically for 24 hours. Following incubation, vegetative cell counts were determined for all cultures. Contents from each well were serially diluted in PBS and plated in triplicate onto BHIS supplemented with 0.1% (w/v) sodium taurocholate (Sigma) to induce spore germination. Bacterial counts were enumerated on plates after

24 hours and calculations were performed to give colony forming units per ml (CFU/ml). Assays were performed with a minimum of two biological replicates.

2.2.8.7 Whole protein cleavage assay

To assess the activity of the recombinant sortase enzyme CD2718_{ΔN26} against a purified protein substrate, a whole protein cleavage assay was developed to mirror the results of the FRET assay. Fifteen µM CD2718_{ΔN26} (purification described in Section 4.4.2.3) was incubated with 20 µM recombinant substrate protein in FRET buffer (5 mM CaCl₂, 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT) for 48 hours at 37 °C, after which samples were analysed by SDS-PAGE and immunoblotting.

3 Bioinformatic analysis of the potential *C. difficile* sortase and identification of putative sortase substrates

3.1 Introduction

Sortases are membrane bound enzymes that covalently anchor virulence-associated surface proteins containing LPXTG-like sequences to the Gram-positive cell wall. This process is mediated by a conserved catalytic cysteine residue found in all sortases to date, which is essential for activity (Ton-That *et al.*, 1999). Sortase substrates contain an N-terminal signal peptide that directs the protein for secretion from the cell, and a conserved C-terminal cell wall sorting signal (CWSS) comprising an LPXTG-like motif, a membrane spanning hydrophobic region and a positively charged tail (Schneewind *et al.*, 1992, Schneewind *et al.*, 1993). The hydrophobic region and charged tail of the substrate retain the protein within the cell membrane following secretion, allowing the membrane-bound sortase to recognise the CWSS of the substrate, and covalently anchor it to the cell wall (Schneewind *et al.*, 1993).

The genome sequence of *C. difficile* strain 630 indicates the presence of a single sortase-like protein, encoded by the gene, *CD2718* (Sebaihia *et al.*, 2006). The putative protein, *CD2718*, is 225 amino acids in length with a predicted size of 26.7 kDa and annotated as a class B sortase based on sequence similarity to Sortase B (SrtB) from *S. aureus* (SaSrtB). The predicted recognition sequence for *CD2718* has been proposed to be (S/P)PXTG (Pallen *et al.*, 2001), and to also include the sequence NVQTG, found in the surface-associated collagen binding protein, CbpA (Tulli *et al.*, 2013). However, this has not been demonstrated experimentally, and little is known about what other proteins may also be substrates of *CD2718*.

3.1.1 Aims of work described in this Chapter

This Chapter reports bioinformatics analyses of the predicted *C. difficile* sortase, CD2718, and its potential substrates to determine whether it is likely to be a functional system.

The aims of this study were to:

- analyse the putative sortase through comparisons with other well-characterised sortases and identify potential substrate proteins within the genome
- determine the distribution of the sortase and predicted substrates across diverse *C. difficile* strains

3.2 Comparative analysis of CD2718

A BLAST analysis (BlastP) of the predicted *C. difficile* 630 CD2718-encoded protein revealed that it shares 32% and 34% amino acid sequence identity with SrtB from *S. aureus* (SaSrtB) and *B. anthracis* (BaSrtB), respectively. Alignment in ClustalW shows a high degree of similarity between CD2718 and both sortases (Figure 3.1A). Three active site residues (His120, Cys184 and Arg197 in Sortase A from *S. aureus* (SaSrtA)) essential for the catalytic activity of all sortases to date are conserved in CD2718 (Ton-That *et al.*, 2002, Liew *et al.*, 2004, Marraffini *et al.*, 2004). A structural prediction analysis of CD2718 was performed using Phyre2 Protein Fold Recognition Server (Kelley & Sternberg, 2009). The Phyre2 Server modelled 197 residues of CD2718 (a coverage of 88%) with 100% confidence by the highest scoring template found in the database, the crystal structure of BaSrtB (Zhang *et al.*, 2004). The resulting alignment suggests a high level of conservation between the predicted secondary structure of CD2718 and the structure of BaSrtB (Figure 3.1B).

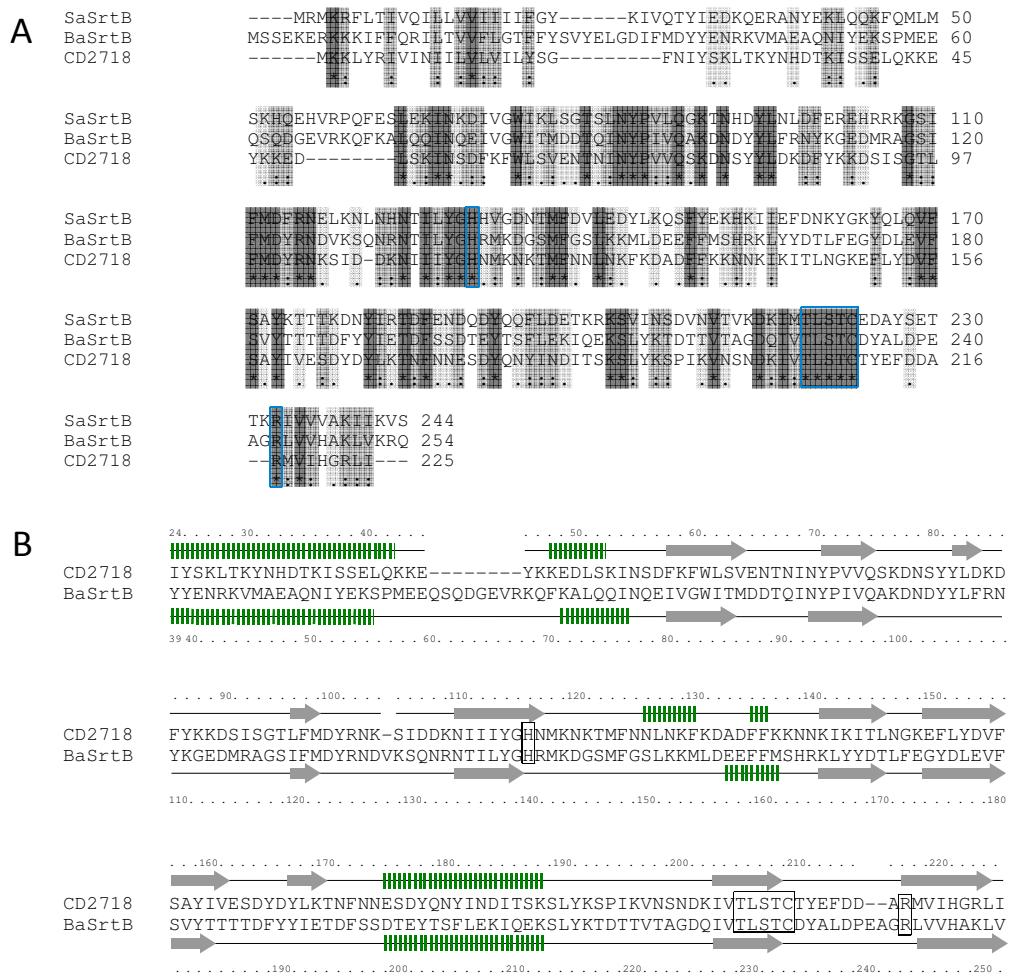


Figure 3.1: CD2718 comparative alignment

A: ClustalW2 multiple alignment of SrtB from *S. aureus* (SaSrtB), *B. anthracis* (BaSrtB), and *C. difficile* (CD2718). CD2718 shares 32% and 34% amino acid sequence identity with SaSrtB and BaSrtB, respectively, as determined by BLAST search. Symbols represent: dark shading and “*”= identical residues; lighter shading and “:” = strongly similar residues; “.” = weakly similar residues. Amino acid positions relative to start are indicated. The sortase active site signature sequence, TLXTC, is boxed, as are the conserved essential histidine and arginine residues. **B:** Structural alignment between the known crystal structure of BaSrtB (Zhang *et al.*, 2004) and the predicted structure of CD2718 using the Phyre2 Protein Fold Recognition Server suggests a high degree of structural conservation. **Top:** CD2718 predicted secondary structure and sequence. **Bottom:** BaSrtB sequence and known structure. Arrows indicate beta sheets, and striped rectangles indicate alpha helices. Amino acid positions relative to start are indicated. The sortase active site signature sequence, TLXTC, is boxed, as are the conserved essential histidine and arginine residues.

CD3146 is a potential second sortase encoded within the genome of *C. difficile* strain 630, but is interrupted by a stop codon (Sebaihia *et al.*, 2006). The in-frame stop codon occurs upstream of the catalytic cysteine residue within the sortase signature sequence, TLXTC (Figure 3.2), supporting the assertion that *CD3146* is a pseudogene. The full open reading frame of *CD3146*, if uninterrupted, shares 44% amino acid sequence identity with CD2718 (and 29% identity with BaSrtB).

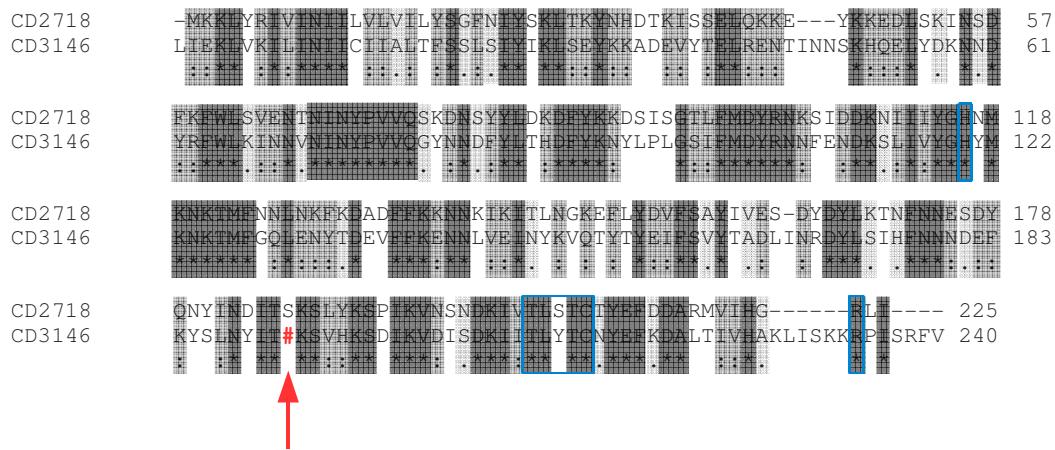


Figure 3.2: Location of stop codon in sortase pseudogene, CD3146

ClustalW2 alignment of the predicted SrtB, CD2718, and the predicted pseudogene, CD3146. CD2718 and CD3146 share 44% amino acid identity. The CDS of CD3146 contains an in-frame stop codon (indicated by the red # and arrow) upstream of the sortase signature sequence TLXTC, and is hypothesised to be a non-functional pseudogene.

3.3 Distribution of CD2718 in the *C. difficile* phylogenetic spectrum

To better understand the prevalence of the predicted sortase within the *C. difficile* species, we carried out a comparative genomic analysis to determine its distribution within selected *C. difficile* isolates. In associated research, I undertook a phylogenetic analysis using multilocus sequence typing (MLST) that revealed a broad genetic diversity within *C. difficile* strains and population structure of at least five distinct clonal lineages that are all associated with human infection (Stabler *et al.*, 2012). This was in line with previous phylogenetic analysis based on whole genome sequencing and MLST (He *et al.*, 2010, Dingle *et al.*, 2011). The five distinct lineages, or clades, are highly correlated with strain PCR-ribotypes, a typing method for *C. difficile* based on the variation that occurs in the intergenic spacer regions of the 16S and 23S ribosomal RNA genes (Stubbs *et al.*,

1999). Clade 1 is formed of a diverse range of ribotypes (RT), including *C. difficile* 630 (RT012), while the other 4 major clades are more clonal in structure: Clade 2 including the hypervirulent RT027 strains, Clade 3 including RT023 strains, Clade 4 including RT017 strains, and Clade 5 including the highly divergent RT078 (see Figure 1.3). Representatives for each of the five distinct clades were chosen for our investigation into the distribution of the predicted sortase gene within *C. difficile* based on the availability of a fully annotated sequence. Details of the strains used in the analysis are summarised in Table 3.1.

Table 3.1: Representative *C. difficile* strains

	Strain	RT	Origin	Reference	Accession number
Clade 1	630	RT012	1982, Zurich, Switzerland	Sebaihia <i>et al.</i> , 2006	AM180355
Clade 2	R20291	RT027	2006, London, UK	Stabler <i>et al.</i> , 2009	NC_013316
	CD196	RT027	1985, Paris, France	Stabler <i>et al.</i> , 2009	NC_013315
Clade 3	CD305	RT023	2008, London, UK	Unpublished, WTSI	
Clade 4	M68	RT017	2006, Dublin, Ireland	He <i>et al.</i> , 2010	FN668375
	CF5	RT017	1995, Belgium	He <i>et al.</i> , 2010	FN665652
Clade 5	M120	RT078	2007, UK	He <i>et al.</i> , 2010	FN665653

BLAST searches of these representative strains from the five major *C. difficile* lineages show that CD2718 is conserved in all *C. difficile* lineages, with each strain encoding an orthologue with at least 94% amino acid sequence identity to CD2718 (Table 3.2). There was only one intact sortase identified in all the lineages with the exception of Clade 3. Interestingly, strain CD305 (RT023) appears to encode for two intact putative sortases. In addition to the CD2718 orthologue, annotated as CD305_2958, a second distinct sortase is found, CD305_0570, with 26% amino acid sequence identity with CD2718, and 32% amino acid sequence identity with BaSrtB. The predicted pseudogene, CD3146, found in *C. difficile* strain 630, is absent in Clade 2 and 3 strains, but is present in Clade 4 and 5 strains. The in-frame stop codon that occurs in CD3146 is conserved in strains M68 and CF5, but the copy present in strain M120 is further truncated at the C-terminus (Figure 3.3).

Table 3.2: CD2718 orthologues and % amino acid sequence identity between representative *C. difficile* strains

CD3146	LIEKLVKILINIICIITALFSSLSIYIKLSEYKKADEVYTELRENTINN-SKHQELEYDKNN	59
CDM68_31481	-LKILVKILINIICIITLIFSSLSAYIKLSEYRKADEVYTELRENTSNNSNSHQQELEYDKNN	59
CF5_31421	-LKILVKILINIICIITLIFSSLSAYIKLSEYRKADEVYTELRENTSNNSNSHQQELEYDKNN	59
CDM120_31201	-LKKLVKILINIICIITLIFSSLSIYIKLSEYKKADEVYTELRENTSNNSNSHQQELEYDKNN	59
	: : *****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	
CD3146	DYRFWLKINNVNINYPVVQGYNDFYLTHDFYKNYLPGLSIFMDYRNNFENDKS LIVYGHY	119
CDM68_31481	DYRFWLKINNTNIDYPVVQGYNDFYLTHNFYKNYLPAGSIFMDYRNNFENDKS LIVYGHY	119
CF5_31421	DYRFWLKINNTNIDYPVVQGYNDFYLTHNFYKNYLPAGSIFMDYRNNFENDKS LIVYGHY	119
CDM120_31201	DYRFWLKINNANIDYPVVQGYNKDFYLTHDFYKNYLPAGSIFMDYRNNFENDKS LIVYGHY	119
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	
CD3146	MKNKTMFGQLENYTDEVFFKENNLVEINYKVQTYYEIFS VYTADL INRDYLSIHFNND E	179
CDM68_31481	MKNKTMFGQLENYTDEALFKENNLVETKYKVQTYYKIFS IYTADL IYRDYLSIHFN-DE	179
CF5_31421	MKNKTMFGQLENYTDEALFKENNLVETKYKVQTYYKIFS IYTADL IYRDYLSIHFN-DE	179
CDM120_31201	MKNKTMFGQLE-----SRNKI#-----	137
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	
CD3146	FKYSLNYIT#KSVHKSDIKVDISDKIITLYTONYEFKDALTIHVAKLISKKRPI SRFV	239
CDM68_31481	FKDSLNYIT#-----	190
CF5_31421	FKDSLNYIT#-----	190
CDM120_31201	-----	

Figure 3.3: Alignment of CD3146 orthologues

ClustalW2 alignment of the predicted pseudogene, CD3146, with orthologues from other *C. difficile* strains. The in-frame stop codon (#) is conserved in strains M68 and CF5, while strain M120 is truncated at the C-terminus. The stop codon occurs before the sortase active site TLXTC (boxed in).

3.4 Bioinformatic prediction of sortase substrates

In addition to an N-terminal signal peptide sequence, sortases substrates contain a highly conserved CWSS at the C-terminus that facilitates their preliminary identification in

genome-wide analyses. The CWSS has been defined by the following three features: an LPXTG-like pattern positioned 17-45 residues from the C-terminus, followed by a membrane spanning region and a positively charged tail (Janulczyk & Rasmussen, 2001, Comfort & Clubb, 2004). A bioinformatics approach was used for the preliminary identification of sortase substrate proteins in *C. difficile* 630 that met these criteria. The *C. difficile* sortase recognition sequence has been predicted to be (S/P)PXTG, where the first residue is either a serine or proline instead of a leucine (Pallen *et al.*, 2001). The sequence NVQTG has also been proposed to be a CD2718 recognition sequence, appearing in the recently characterised collagen binding protein A, CbpA (Tulli *et al.*, 2013). All protein sequences in *C. difficile* 630 were searched for the patterns (S/P)PXTG and NVQTG using the Artemis Genome Browser Tool (Rutherford *et al.*, 2000). Also included in the search were the recognition sequences for SaSrtB and BaSrtB, NPQTN and NPKTG, respectively.

Of the 23 putative protein coding regions that contained the sequences (S/P)PXTG or NVQTG, only 10 contained them within the final 45 residues of the C-terminus (Table 3.3). There were no putative protein coding regions found that contained a C-terminally located NPQTN or NPKTG. To determine the likelihood of these 10 candidate proteins being sortase substrates, the proteins were assessed for the presence of the following: (i) a suitable signal peptide sequence at the N-terminus, (ii) a likely membrane-spanning region following the proposed sortase recognition sequence, and (iii) a charged tail consisting of at least two consecutive basic residues (lysine or arginine) (Schneewind *et al.*, 1993). This was performed using the SignalP 4.1 Server to predict signal peptide sequences (Petersen *et al.*, 2011), and the TMHMM Server v. 2.0 (Moller *et al.*, 2001), in addition to manually identifying hydrophobic residues to predict transmembrane domains (Monera *et al.*, 1995). The results of these searches are found in Table 3.3.

Eight of the 10 proteins satisfied the requirements of a sortase substrate in strain 630 (highlighted in grey in Table 3.3). The annotations of the predicted *C. difficile* sortase substrates suggest a diverse range of surface proteins that include putative cell wall hydrolases, putative adhesins, a collagen binding protein, and a 5' nucleotidase/phosphoesterase. Two of the predicted proteins, CD0386 and CD3392, are 94% identical by amino acid sequence and are located on the similar conjugative

transposons, CTn1 and CTn7, respectively (Brouwer *et al.*, 2012, Brouwer *et al.*, 2011). The newly described *C. difficile* collagen binding protein A, CbpA, is the only protein containing the proposed NVQTG motif (Tulli *et al.*, 2013), and there are four variations of the (S/P)PXTG motif observed: SPKTG, PPKTG, SPSTG and SPQTG. The two excluded proteins of the list of 10 candidates were BclA3 and TopB. BclA3, the exosporium glycoprotein, does not appear to be a potential substrate of CD2718: the PPXTG sequence found at the C-terminus of BclA3 is predicted to be contained within a membrane spanning region, there is no predicted N-terminal signal peptide, and there is only a single charged residue at the C-terminus. Despite the presence of a 4 charged residues at the C-terminus, the DNA topoisomerase III, TopB, was eliminated based on the lack of signal peptide sequence, weakly hydrophobic tail, and its unlikelihood to be located on the bacterial cell surface (Tadesse & Graumann, 2006).

Of the eight *C. difficile* proteins predicted here to be sortase substrates, the putative cell wall hydrolases, CD0183 and CD2768, have weak hydrophobic domains that were not predicted as transmembrane domains by the TMHMM Server. However, they both have hydrophobic amino acids spanning over 15 residues, the predicted minimum required for traversing the membrane (Schneewind *et al.*, 1993, Janulczyk & Rasmussen, 2001): CD0183 has nine hydrophobic residues spanning a region of eighteen residues, and CD2768 has eight hydrophobic residues spanning a region of fifteen residues. A signal peptide sequence was not predicted by the SignalP Server for CD2537, but a membrane spanning segment (residues 7-26) was predicted by the TMHMM Server. As there is considerable overlap between signal peptide sequences and transmembrane domains (Kall *et al.*, 2007), this protein may contain an unrecognised signal sequence. To investigate this, an additional test was performed using the combined membrane topology and signal peptide prediction tool, the Phobius webserver, which predicted a weak signal peptide sequence for CD2537 (score = 0.75).

**Table 3.3: (S/P)PXTG and NVQTG containing proteins in *C. difficile* 630
(Part 1)**

Protein	Function	C-terminus	TMD after LPXTG	Signal Peptide
CD0183	Putative cell wall hydrolase	MIHSPST <u>GKTVSYTSINSYT</u> TARE <u>VAKRL</u>	NP	yes
CD0386	Putative cell surface protein	PSDSP <u>KTGDNNTLYGLLALLTSGAGLAGIFFYKRRKMKK</u> S	yes	yes
CD0551	SleC, Spore cortex-lytic enzyme	270 aa from C-terminus		
CD0660	TcdB, Toxin B	403 aa from C-terminus		
CD0679	Conserved hypothetical protein	180 aa from C-terminus		
CD0684	Putative ATP-dependent peptidase	380 aa from C-terminus		
CD1085	Putative membrane protein	271 aa from C-terminus		
CD1106	TopB, DNA topoisomerase III	KLYSP <u>KTYDGTIALADTGEKYVNYRIELSKK</u> K	NP	NP
CD1647	ABC-type transport system, iron-family permease	237 aa from C-terminus		
CD1657	Glycine dehydrogenase/aminomethyl transferase	501 aa from C-terminus		
CD1704	Thiazole biosynthesis protein	NASSP <u>LTGELGNL</u>	NP	NP
CD1730	Putative ATP-binding protein	118 aa from C-terminus		
CD2094	Putative restriction enzyme	336 aa from C-terminus		
CD2213	Carbonic anhydrase	LIISPE <u>TGKLDVVVNGYEDK</u>	NP	NP
CD2458	Transporter, Major Facilitator Superfamily (MFS)	133 aa from C-terminus		
CD2491	Mannose-6-phosphate isomerase	280 aa from C-terminus		

Bold = (S/P)PXTG and NVQTG sequence

Underlined = hydrophobic residues (Monera et al, J Protein Sci 1995)

Red = positively charged tail (lysine or arginine residues)

NP = not predicted

(Score = 0.75)

b = TMHMM score < 0.85

c = TMHMM score < 0.78

**Table 3.3: (S/P)PXTG and NVQTG containing proteins in *C. difficile* 630
(Part 2)**

Protein	Function	C-terminus	TMD after LPXTG	Signal Peptide
CD2537	Putative 5'-nucleotidase/phosphoesterase	KEK S P K T G <u>D</u> <u>G</u> FNSII <u>I</u> F <u>V</u> S <u>S</u> T <u>L</u> C <u>L</u> N <u>F</u> N <u>Q</u> K E I K D K K S K	yes	yes ^a
CD2768	Putative cell-wall hydrolase	FI H S P Q T G D <u>V</u> <u>V</u> <u>K</u> T <u>S</u> M <u>A</u> P <u>G</u> T <u>N</u> Y A R <u>R</u> L <u>I</u> T <u>A</u> R <u>V</u> <u>L</u> <u>Q</u>	NP	yes
CD2831	Putative adhesin	PPV P P K T G D <u>S</u> <u>T</u> <u>I</u> <u>G</u> E <u>I</u> <u>L</u> <u>V</u> <u>G</u> A <u>V</u> <u>G</u> <u>L</u> <u>V</u> <u>R</u> R <u>N</u> K N T N	yes	yes
CD3034	Transcriptional regulator	218 aa from C-terminus		
CD3145	CbpA, Collagen binding protein A	VGQN <u>V</u> Q T G D <u>Q</u> <u>S</u> N <u>I</u> <u>M</u> <u>L</u> <u>D</u> <u>A</u> <u>M</u> <u>E</u> <u>S</u> <u>F</u> <u>L</u> <u>I</u> <u>K</u> N <u>L</u> <u>T</u> N K I L R K	weak ^b	yes
CD3246	Putative surface protein	IV S P K T G D <u>E</u> <u>T</u> <u>Q</u> <u>L</u> <u>M</u> <u>S</u> <u>Y</u> <u>F</u> <u>I</u> <u>S</u> <u>V</u> <u>I</u> <u>A</u> <u>I</u> <u>C</u> <u>G</u> <u>L</u> <u>A</u> <u>/</u> Q <u>C</u> <u>K</u> I K R N	weak ^c	yes
CD3301	ATP-dependent protease	420 aa from C-terminus		
CD3349	BclA3, Exosporium glycoprotein	T <u>L</u> <u>A</u> <u>P</u> <u>P</u> <u>I</u> <u>G</u> <u>L</u> <u>S</u> <u>V</u> <u>G</u> <u>S</u> <u>I</u> <u>S</u> <u>S</u> <u>G</u> <u>T</u> <u>G</u> <u>I</u> <u>N</u> <u>I</u> <u>A</u> <u>A</u> <u>T</u> <u>A</u> <u>Q</u> <u>T</u> <u>P</u> <u>D</u> R <u>Q</u> <u>Y</u> <u>A</u>	within TMD	NP
CD3392	Putative cell surface protein	PSD S P K T G D <u>S</u> <u>T</u> <u>N</u> <u>L</u> <u>M</u> <u>A</u> <u>F</u> <u>V</u> <u>M</u> <u>L</u> <u>V</u> <u>S</u> <u>G</u> <u>G</u> <u>L</u> <u>A</u> <u>G</u> <u>T</u> <u>L</u> Y K R K M K S	yes	yes
CD3559	FtsH2, Cell division protease	448 aa from C-terminus		

Bold = (S/P)PXTG and NVQTG sequence

Underlined = hydrophobic residues (Monera et al, J Protein Sci 1995)

Red = positively charged tail (lysine or arginine residues)

NP = not predicted

^a = not predicted by SignalP, transmembrane domain predicted by TMHMM (Score = 0.98), and signal peptide predicted by Phobius (Score = 0.75)

^b = TMHMM score < 0.85

^c = TMHMM score < 0.78

3.5 Distribution of predicted sortase substrates among *C. difficile* strains

A comparative genomic analysis was performed to determine the distribution of the eight predicted substrates of CD2718 within the strains chosen as representatives of *C. difficile* genetic diversity in Section 3.3. This analysis revealed that of the eight putative substrates identified in strain 630, four are conserved across all clades: CD0183, CD2537, CD2768, and CD2831 (shaded in grey in Table 3.4). Interestingly, both Clade 2 strains only have a single orthologue of CD0386 and CD3392: all other clades retain orthologues of both highly similar proteins, with the exception of one of the two Clade 4 strains analysed, CF5, which does not encode either. Both Clade 2 strains are missing CbpA and CD3246 orthologues, CD305 (Clade 3) is missing a CbpA orthologue, and M120 (Clade 5) is missing orthologues of CD3392 and CD0386. The genomic co-localisation of CbpA with the pseudogene sortase is retained with the respective orthologues identified in Clade 4 and 5 strains.

Table 3.4: Distribution and % amino acid identity of predicted CD2718 substrate proteins across *C. difficile* strains*

Predicted protein function	Motif	Clade 1		Clade 2		Clade 3		Clade 4		Clade 5	
		630	R20291	CD196	CD305	M68	CF5	M120			
Putative cell wall hydrolase	SPSTG	CD0183	99	99	99	96	96	97			
Putative cell surface protein, collagen binding protein	SPKTG	CD0386 CD3392	83	83	93 83	93 93	x	x			
Putative membrane-associated 5'-nucleotidase/ phosphoesterase	SPKTG	CD2537	98	98	98	97	97	93			
Putative cell-wall hydrolase	SPQTG	CD2768	99	99	99	99	99	94	98		
Putative adhesin, collagen binding protein	PPKTG	CD2831	95	95	95	94	99	92			
CbpA, collagen binding protein	NVQTG	CD3145	x	x	x	88	88	83			
Putative surface protein	SPKTG	CD3246	x	x	87	91	91	84			
Number of predicted substrates		8	5	5	7	8	6	6			

* See Table D.1 for gene ID numbers of predicted substrate orthologues

The genomes of the strains representing Clades 2-5 were screened for additional (S/P)PXTG and NVQTG containing proteins not present in strain 630. Both Clade 2 and clade 4 strains were found to encode an additional SPQTG containing protein, annotated as a putative cell wall hydrolase, with strain CF5 encoding two such proteins (Table 3.5). The R20291_1451, CD196_1476, and CF5_15831 proteins each share 69% amino acid identity with CD0961 from strain 630, while CDM68_32001 and CF5_31931 share 63% amino acid identity with CD1368 from strain 630. Both CD0961 and CD1368 are also

annotated as putative cell wall hydrolases, but lack the SPQTG motif found in the Clade 2 and 4 orthologues. Though no C-terminal transmembrane domains were predicted by the TMHMM or Phobius Servers, all five proteins have very similar C-termini as the two cell wall hydrolases found in strain 630, CD0183 and CD2768, with a similar number of hydrophobic amino acids (Table 3.5). However, no signal peptide was predicted by the SignalP Server, suggesting these five putative proteins are likely not sortase substrates in *C. difficile*.

Table 3.5: Additional (S/P)PXTG containing proteins in *C. difficile* strains

Protein	C-terminus	TMD after LPXTG	Signal peptide	Orthologue in 630 (% identity)
R20291_1451	IIHSPQTGDVVKI E SVNSYRKGKAYTHV RR FI	NP	NP	
CD196_1476	IIHSPQTGDVVK I ESVNSYRKGKAYTHV RR FI	NP	NP	CD0961 (69%)
CF5_15831	MIHSPQTGDVVK I QSVDSYRKGKAYTHV RR FF	NP	NP	
CDM68_32001	IIHSPQTGDVVK I QSVDSYRKGKAYTHV RR YL	NP	NP	
CF5_31931	IIHSPQTGDVVK I QSVDSYRKGKAYTHV RR YL	NP	NP	CD1368 (63%)

NP = not predicted

3.6 Analysis of predicted substrate proteins

To better understand the role these eight predicted sortase substrates may serve in *C. difficile* infection, the proteins were analysed for conserved functional protein domains using the Pfam protein families database (Punta *et al.*, 2012).

3.6.1 Putative surface protein, CD3246

Of the eight predicted substrates, only one had no significant matches to any of the protein family domains in the Pfam database. CD3246, annotated as a putative surface protein, is 727 amino acids in length with a predicted mass of 79.9 kDa. Though not much is known about the putative function of CD3246, translation of the CD3246 ORF was found to be under the control of a c-di-GMP riboswitch (Lee *et al.*, 2010a, Chen *et al.*, 2011), part of the bacterial c-di-GMP second messenger signalling pathway involved in the regulation of many virulence associated physiological properties, including motility and biofilm formation (Hengge, 2009). Riboswitches are *cis*-acting regulatory elements that are commonly involved in bacterial gene expression. When bound to a specific ligand, in this case c-di GMP, riboswitches undergo a conformational change that influence gene expression at either the transcriptional or translational level (Nudler & Mironov, 2004,

Hengge, 2009). Interestingly, CD3246 also contains seven cleavage sites recognised by the recently characterised secreted metalloprotease, CD2830 (Hensbergen *et al.*, 2014). *In vitro* studies demonstrated that native CD2830 secreted by live *C. difficile* cells cleaved endogenously expressed CD3246 (Hensbergen *et al.*, 2014).

3.6.2 Collagen binding proteins

The collagen-binding protein in *S. aureus*, Cna, mediates binding to a variety of collagens and is important for host-tissue adherence and establishment of infection (Foster & Hook, 1998). Cna mutants are considerably less virulent in the mouse than wildtype Cna-expressing *S. aureus* (Patti *et al.*, 1994b, Hienz *et al.*, 1996). Cna is comprised of a non-repetitive A region that mediates attachment to collagen (Symersky *et al.*, 1997), and a repetitive B region that is hypothesised to serve as a stalk that allows display of the functional binding region A (Deivanayagam *et al.*, 2000) (Figure 3.4). There is some variation in the number of B repeats found in the *cna* gene, with versions containing one to four 23 kDa repeats identified (Switalski *et al.*, 1989, Patti *et al.*, 1994a). Despite the differences in the number of B repeats, these Cna proteins do not differ in their collagen binding capacity (Patti *et al.*, 1994a). In the event of loss of the region A binding domain, the B repeats stalk has been predicted to be capable of mediating other adhesive processes or functions in the cell (Deivanayagam *et al.*, 2000).

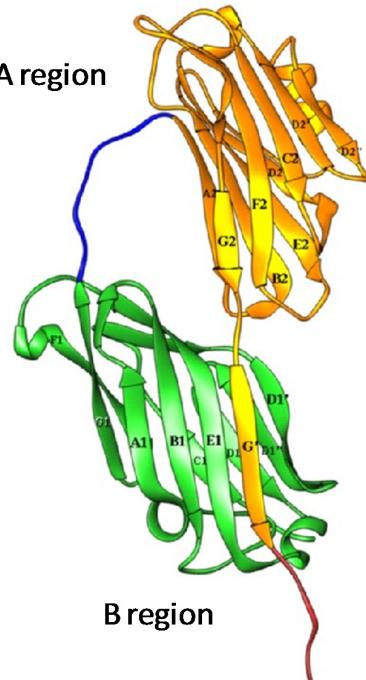


Figure 3.4: Structure of the collagen binding protein, Cna, from *S. aureus*

The A region (in yellow) contains the active collagen binding domain, while the B region repeats (in green) are proposed to aid in the projection of the A region away from the cell surface. Figure from Zong *et al.*, 2005.

CD0386 and CD3392 are both 1014 amino acids in length, with predicted protein sizes of 111.8 kDa, and CD2831 is 972 amino acids in length, with a predicted size of 107.6 kDa. In the Pfam analysis, we find that CD2831 contains the collagen binding region A, and a single B repeat region (Figure 3.5). Both CD0386 and CD3392, which share 94% amino acid similarity, contain three B region repeats but do not retain the collagen binding domain A (Figure 3.4). CD0386 and CD3392 are accessory genes on the highly similar conjugative transposons CTn1 and CTn7, respectively (Brouwer *et al.*, 2012, Brouwer *et al.*, 2011). CD2831 may mediate binding to collagens, while CD0386 and CD3392 may facilitate adherence to an alternative component. Similarly to CD3246, CD2831 contains six cleavage sites recognised by the recently characterised secreted metalloprotease CD2830, located just downstream of CD2831 (Hensbergen *et al.*, 2014), and translation of CD2831 is under the control of a c-di-GMP riboswitch (Lee *et al.*, 2010a, Soutourina *et al.*, 2013).

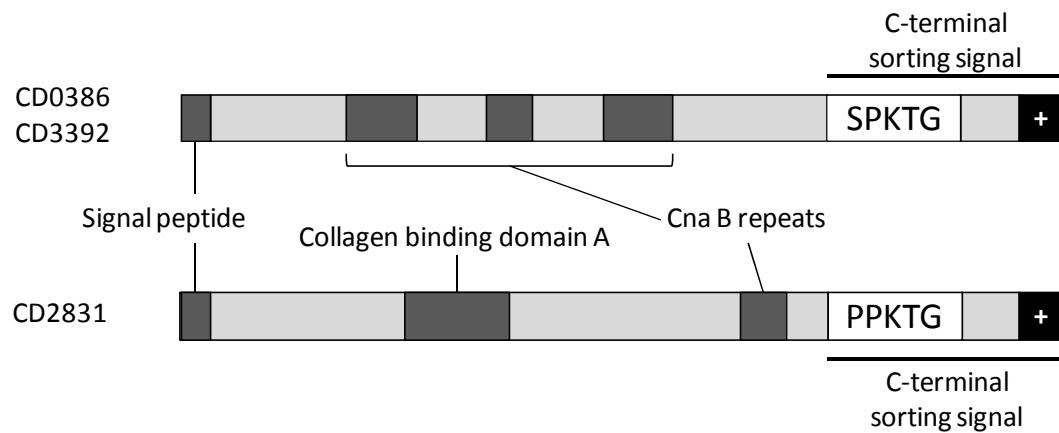


Figure 3.5: Schematic representation of CD0386, CD3392 and CD2831 protein organisation

CD0386 and CD3392 contain three B region repeats and lack the collagen binding A region, while CD2831 contains the collagen binding A region and a single B repeat.

Recently, CD3145 was characterised as collagen binding protein A, CbpA, and shown to mediate attachment to collagen types I and V *in vitro* (Tulli *et al.*, 2013). The *CD3145* gene was originally annotated as a serine-aspartate repeat-containing protein SdrF (Sebaihia *et al.*, 2006, Monot *et al.*, 2011), until its collagen binding properties were demonstrated. CbpA contains three distinct repeat-types (Figure 3.6), one of which resembles calcium-binding sequences found in thrombospondin-like proteins (Tulli *et al.*, 2013), which include a diverse array of large secreted proteins that modulate extracellular matrix structure and eukaryotic cell behavior (Kvansakul *et al.*, 2004). In their study, Tulli and colleagues demonstrated that CbpA is surface localised in *C. difficile*, and that recombinantly expressed CbpA bound to collagens *in vitro*, and to *ex vivo* murine intestinal cells (Tulli *et al.*, 2013). The presence of the NVQTG sequence and the satisfaction of all the other requisites of a sortase substrate prompted the authors to propose that CbpA is anchored to the *C. difficile* cell surface by sortase. The repeats mediating collagen binding in CbpA are functionally different than the domains found in Cna, and are not present in CD0386, CD2831, and CD3392. No domain matches were identified in our Pfam database search.

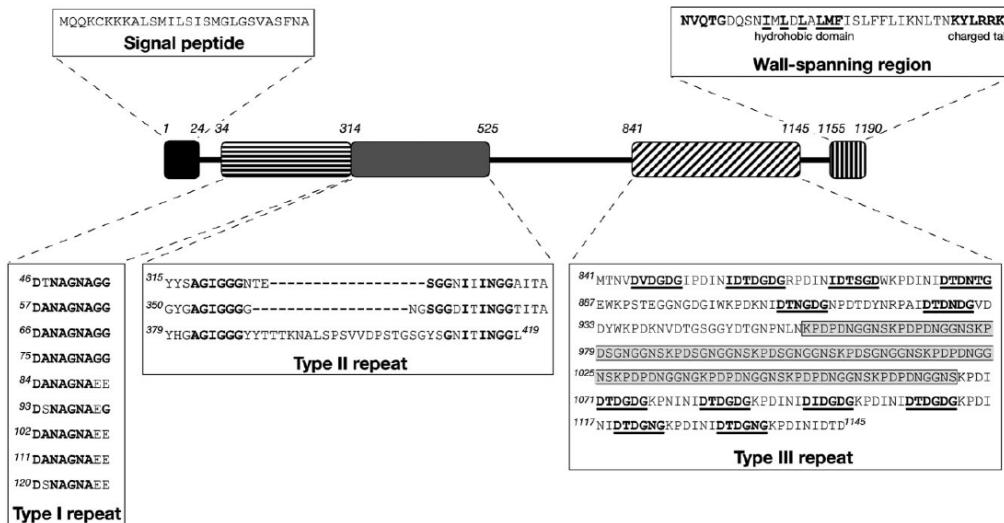


Figure 3.6: Schematic representation of CpbA (CD3145) protein organisation

CpbA contains three distinct repeat types: type I, type II and type III. Type III repeats comprise a central core of consecutive KPD_nDNGGNS motifs flanked by DXDGDG signatures, similar to the calcium-binding sequences found in thrombospondin-like proteins. Figure from Tulli *et al.*, 2013 .

3.6.3 Cell wall hydrolases

Both CD0183 and CD2768 are annotated as putative cell wall hydrolases. They are 340 and 235 amino acid residues long and have a predicted molecular mass of 37.0 and 25.0 kDa, respectively, and share 39% amino acid identity. Translation of CD0183 is under the control of a small non-coding RNA (Soutourina *et al.*, 2013). According to the Pfam database, both proteins contain a src homology-3 (SH3) domain (CD0183 contains two such domains) and an N1pC/P60 domain (Figure 3.7). SH3 domains are found in a variety of intracellular and membrane associated proteins. They are believed to be involved with intracellular signaling pathways by binding proline-rich peptides and mediating protein-protein interactions (Feng *et al.*, 1994). The N1pC/P60 protein family describes a group of peptidoglycan-hydrolysing proteins involved in cell division that are present in a wide range of bacterial species, including the *Escherichia coli* membrane associated lipoprotein N1pC and the *Listeria monocytogenes* protein P60 for which this family is named (Wuenscher *et al.*, 1993, Anantharaman & Aravind, 2003).

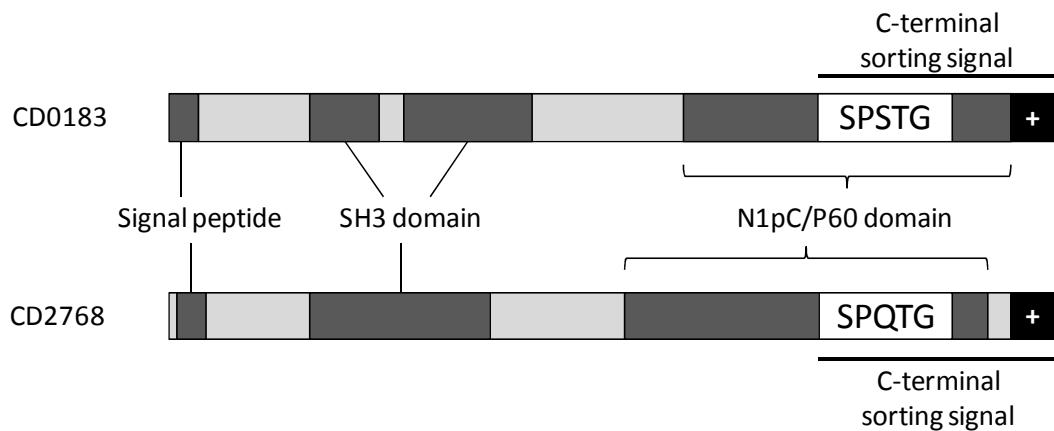


Figure 3.7: Schematic representation of CD0183 and CD2768 protein organisation

Both proteins are predicted cell wall hydrolases that contain Src homology-3 domains (SH3), and also a N1pC/P60 domain, which has been shown to hydrolyse peptidoglycan linkages in many bacteria.

In *L. monocytogenes*, P60 acts as a murein hydrolase required in the last step of cell division (Wuenscher *et al.*, 1993). P60 mutants display cell-wall separation defects and abnormal cell divisions (Pilgrim *et al.*, 2003), exhibit severe defects in attachment to and invasion of several cell types *in vitro*, and cause less severe infection in mice compared to wildtype *L. monocytogenes* (Hess *et al.*, 1996, Faith *et al.*, 2007). In *B. subtilis*, cell-wall separation defects are also observed following mutation of the related LytE autolysin, resulting in the formation of long cell chains (Ishikawa *et al.*, 1998, Margot *et al.*, 1998). Within similar functional domains found in CD0183 and CD2768, these two proteins may be involved in *C. difficile* cell division. In unpublished data described by Soutourina *et al.* (2013), overexpression of CD0183 led to a severe growth deficiency, which the authors proposed was due to enhanced autolytic activity.

3.6.4 5' nucleotidase/phosphoesterase

CD2537, annotated as a putative 5' nucleotidase/phosphoesterase, is 613 amino acids in length with a predicted molecular mass of 68.2 kDa. According to the Pfam database, CD2537 contains both a calcineurin-like phosphoesterase domain and a related 5' nucleotidase domain (Figure 3.8). Phosphoesterases are a diverse category of enzymes that cleave phosphomonoester and phosphodiester bonds found in a large variety of

molecules, including phosphorylated proteins, the backbone structure of DNA and RNA, and in cyclic AMP and cyclic GMP. Hydrolysis of these bonds and dephosphorylation are fundamental biochemical reactions within a broad spectrum of metabolic and regulatory pathways (Koonin, 1994, Aravind & Koonin, 1998). Calcineurin, a member of this enzyme class, is a calcium dependent phosphatase that is involved with regulation of gene expression in eukaryotic cells (Sugiura *et al.*, 2002). One of its principal substrates is the nuclear factor of activated T cells (NFAT), which translocates to the nucleus after dephosphorylation to mediate gene transcription (Im & Rao, 2004).

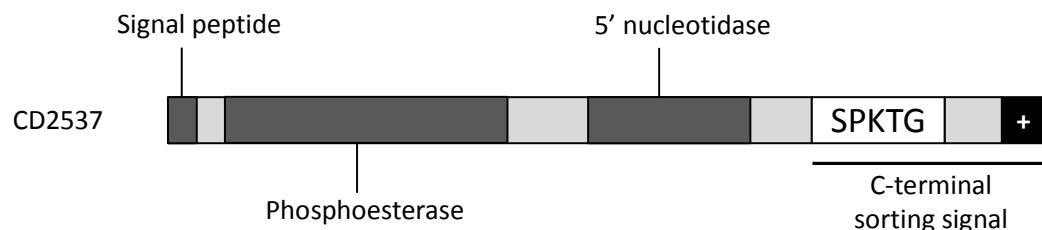


Figure 3.8: Schematic representation of CD2537 protein organisation

CD2537 contains two related domains, a calcineurin-like phosphatase domain and a 5' nucleotidase domain, which have been shown to hydrolyse phosphoester bonds.

In a similar manner, 5' nucleotidases catalyse the hydrolysis of phosphate at the 5' carbon of ribose and deoxyribose within nucleotide molecules, and hydrolyse ATP to adenosine (Zimmermann, 1992). Enzymes hydrolysing 5'-nucleotides display significant differences in the range of substrates hydrolysed as well as in substrate specificity and cellular location. Surface-located 5' nucleotidases have been implicated in cell-cell or cell-matrix interactions, and also in transmembrane signalling (Zimmermann, 1992). The UshA protein from *E. coli* catalyses the degradation of external UDP-glucose to uridine, glucose 1-phosphate and phosphate (Knofel & Strater, 1999), as well as the degradation of NAD to NMN and AMP (Wang *et al.*, 2014). With these functional domains, CD2537 could play a role in signal transduction, though it remains unknown what its potential cellular or extracellular targets might be.

3.7 Discussion

In this chapter, bioinformatics analyses were performed to understand the distribution of sortase-like proteins in the *C. difficile* species, and to identify potential sortase substrate proteins. *C. difficile* 630 encodes a sortase-like protein, CD2718, with similarity to SrtB from *S. aureus* and *Bacillus* species. These analyses have led to the discovery that CD2718 is conserved across all five of the major *C. difficile* lineages, and has led to the identification of eight potential sortase substrates in *C. difficile* 630, based on the predicted recognition patterns of CD2718: (S/P)PXTG and NVQTG.

The existing sortase nomenclature is based on sequence similarity to the six known classes of sortase, A-F (Spirig *et al.*, 2011). Class A sortases are closest in sequence to SrtA from *S. aureus* and known as housekeeping sortases, as they anchor a variety of surface proteins, found throughout the genome, with diverse biological functions to the cell wall. Sortases of class B are most closely related to SrtB from *S. aureus*, and, like SrtB, are typically involved in haem-iron uptake and tend to be expressed in operons with their substrates (Mazmanian *et al.*, 2002, Maresco *et al.*, 2006). The eight potential *C. difficile* sortase substrates identified here comprise a diverse range of predicted surface proteins that include putative cell wall hydrolases, putative adhesins, collagen binding proteins, and a 5' nucleotidase/phosphoesterase, none of which are found in proximity to CD2718. This suggests that the putative class B sortase, CD2178, may serve as a housekeeping sortase in *C. difficile*, a function usually reserved for class A sortases. Several exceptions to these canonical rules have already been described, notably a class B sortase that polymerises pilin subunits in *Streptococcus pyogenes* (Kang *et al.*, 2011), and a class E sortase from *Corynebacteria diphtheriae* that serves a housekeeping function (Chang *et al.*, 2011).

C. difficile strains appear to encode at least one intact sortase with at least 94% amino acid sequence identity with CD2718. Several strains contain a second sortase copy predicted to be a pseudogene, in which an in-frame stop codon occurs prior to the catalytic cysteine residue. The proximity of the predicted pseudogene sortase, CD3146, with the predicted sortase substrate, CbpA, and the conservation of this co-localisation in other *C. difficile* strains suggests that CbpA may have originally been a substrate of this second sortase. Among the strains analysed in this study, CD305 encodes a second intact

sortase B with a similar level of amino acid sequence identity to BaSrtB as CD2718 (32% compared to 36%). It is unknown whether both sortases are functional in strain CD305, and what role either sortase may serve. Many bacteria encode multiple sortase enzymes: typically, bacteria encode one housekeeping sortase and one or more specialised sortases, as is the case with SrtA and SrtB in *S. aureus*. Perhaps the second sortase in strain CD305 has an unknown specialised function. This strain was recently found to contain a putative S-layer glycosylation gene cluster (Dingle *et al.*, 2013), suggesting that CD305 may have an unusual cell surface compared with other *C. difficile* strains.

Of the eight predicted substrates found in strain 630, only four are conserved across *C. difficile* strains, including two putative cell wall hydrolases (CD0183 and CD2768), a putative phosphoesterase (CD2537), and a putative collagen binding protein (CD2831). The substrates that are divergent across all strains include two putative adhesins, so this may demonstrate a redundancy of function for the strains in which they are present. While collagen binding proteins, adhesins, and 5' nucleotidases are commonly sortase-anchored surface proteins (Pallen *et al.*, 2001, Gaspar *et al.*, 2005, Marraffini *et al.*, 2006, Bierne & Cossart, 2007, Nobbs *et al.*, 2009), cell wall hydrolases have not yet been shown to be sortase substrates in any bacteria. This may be a unique to the *C. difficile* sortase, or it may simply be that many LPXTG-containing proteins identified through genome searches, such as those in *B. anthracis* and *L. monocytogenes*, have not yet been characterised and have unknown functions (Pallen *et al.*, 2001, Gaspar *et al.*, 2005, Bierne & Cossart, 2007).

Expression of three of the predicted substrates identified in this study, CD0183, CD2831, and CD3246, appear to be under the regulation of c-di-GMP riboswitches and small noncoding RNAs (Lee *et al.*, 2010a, Chen *et al.*, 2011, Soutourina *et al.*, 2013). Overexpression in *C. difficile* of the small noncoding RNA upstream of the putative hydrolase CD0183 resulted in a similar growth deficiency observed in CD0183 overexpression, which the authors propose may be an example of an RNA-dependent autolysin function in *C. difficile* (Soutourina *et al.*, 2013). Interestingly, the two predicted substrates under the regulation of c-di-GMP riboswitches, CD2831 and CD3246, are also efficiently cleaved by the recently characterised metalloprotease, CD2830 (Hensbergen *et al.*, 2014). The cleavage of adhesins and sortase substrates has previously been observed

in other bacteria, and is thought to be important in regulating the amount of those proteins on the cell surface. In *B. anthracis*, the adhesin BslA, which mediates binding to human endothelial cells, is cleaved by the abundantly secreted protease InhA (Tonry *et al.*, 2012). Expression of BslA and InhA are inversely related, and the level of BslA on the cell surface, and consequent adherence capability of *B. anthracis* vegetative cells, is regulated by InhA (Tonry *et al.*, 2012). A similar phenomenon is observed in *S. aureus*. The SrtA anchored clumping factor B protein (CflB), which mediates staphylococcal adherence to fibrinogen and blood clots, is cleaved by a secreted metalloprotease, aureolysin (McAleese *et al.*, 2001). Surface localisation of CflB varies throughout the growth cycle. McAleese *et al.*, (2001) demonstrate that during stationary phase, in addition to the decrease in CflB expression, CflB on the staphylococcal cell surface is inactivated by aureolysin and shed from the cell wall. *S. aureus* cells in stationary phase are unable to bind fibrinogen in a CflB-dependent manner (McAleese *et al.*, 2001). This suggests a complex regulatory system for the display of these predicted sortase substrates on the *C. difficile* cell surface. The conservation of CD2831 across all *C. difficile* lineages suggests an important function in the *C. difficile* lifestyle.

4 Initial characterisation and purification of the *C. difficile* sortase, CD2718

4.1 Introduction

Analysis of the genome sequence of *C. difficile* strain 630 identified a putative sortase B (SrtB) enzyme, CD2718, though its function remains unknown. Sortase mutants have been used to investigate bacterial protein anchoring and its role during bacterial infections. *S. aureus* sortase A (*srtA*) mutants are defective in anchoring 19 staphylococcal surface proteins bearing an LPXTG motif to the cell wall (Mazmanian *et al.*, 2000), which leads to large defects in virulence in mouse models of disease (Mazmanian *et al.*, 2000, Jonsson *et al.*, 2003, Weiss *et al.*, 2004, Cheng *et al.*, 2009). Mutants in *srtB* are able to cause disease similar to wildtype *S. aureus*, but show reduced bacterial persistence over time during infection (Mazmanian *et al.*, 2002, Jonsson *et al.*, 2003, Weiss *et al.*, 2004).

Inactivating sortases can also have an effect on *in vitro* characteristics. *S. aureus* and *S. epidermidis* *srtA* mutants have an altered colony morphology (Tsompanidou *et al.*, 2012), and *S. mutans* *srtA* mutants exhibit marked differences in cell-surface related properties, including reductions in cell aggregation and hydrophobicity (Lee & Boran, 2003), and impaired biofilm formation (Levesque *et al.*, 2005). Reduction in biofilm production has also been observed in *S. gordonii* *srtA* mutants (Nobbs *et al.*, 2007). Another manner in which to characterise sortase function is through investigating the activity of purified sortase protein in *in vitro* assays, such as the fluorescence resonance energy transfer (FRET) assay. *S. aureus* SrtA recombinantly expressed in and purified from *E. coli* was used to assess its capability to cleave LPXTG motifs (Ton-That *et al.*, 1999). The production of recombinant CD2718 would greatly facilitate such functional characterisation studies.

4.1.1 Aims of the work described in this Chapter

This chapter describes the *in vitro* characterisation and purification of the predicted *C. difficile* sortase enzyme, CD2718, identified through bioinformatic analysis in Chapter 3.

The aims of this study were to:

- Construct an inactivation mutant of CD2718 and characterise its phenotype
- Express and purify recombinant CD2718 for downstream analysis
- Express and purify SaSrtB to serve as control in an *in vitro* assay

4.2 Transcriptional analysis of CD2718 and CD3146

In Chapter 3, two putative sortases were identified in *C. difficile* strain 630; CD2718 appeared intact whereas CD3146 has an in-frame stop codon mutation. Therefore, initial studies were undertaken to determine whether the respective genes were transcribed. Analysis of the genomic localisation of the putative sortase B, *CD2718*, and pseudogene, *CD3146*, revealed that neither gene appeared to be expressed as part of an operon (Figure 3.1).

To determine if *CD2718* and *CD3146* are transcribed when grown under standard laboratory conditions, RT-PCR of wild-type 630Δerm was performed using primer pairs specific to the *C. difficile* *CD2718* and *CD3146* coding sequences. Total RNA was extracted from 10 ml 630Δerm cultures in BHIS media at three time points: early exponential phase ($t=3$ hours, $OD_{600} = 0.4\text{-}0.5$), late exponential phase ($t=5$ hours, $OD_{600} = 0.8\text{-}0.9$), and stationary phase ($t=24$ hours, $OD_{600} = 0.6\text{-}0.8$) (Figure 4.2).

The RT-PCR analysis revealed that the sortase gene, *CD2718*, is transcribed at low levels during both early exponential and stationary phases *in vitro*, and is transcribed highest during late exponential phase (Figure 4.3B, left). The 5' end of the pseudogene *CD3146* open reading frame does appear to be transcribed *in vitro* during early and late exponential phases, but not during stationary phase (Figure 4.3B, right). The gene fragment immediately following the in-frame stop codon is not transcribed, suggesting that *CD3146* is in fact a pseudogene and unlikely to be a functional sortase.

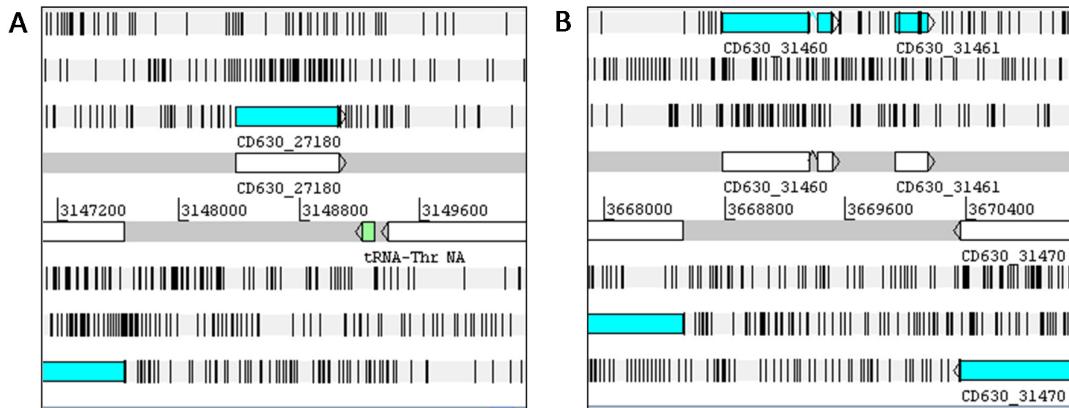


Figure 4.1: Genomic localisation of *CD2718* and *CD3146*

Analysis of the genome of *C. difficile* 630 in the Artemis genome browser (Version 15) revealed that the putative sortase *CD2718* (A.) did not appear to be part of an operon. There is a small gene fragment, *CD31461*, downstream and in-frame with the predicted pseudogene *CD3146* (B.), but this is unlikely to be co-transcribed due to the stop codon interrupting *CD3146*.

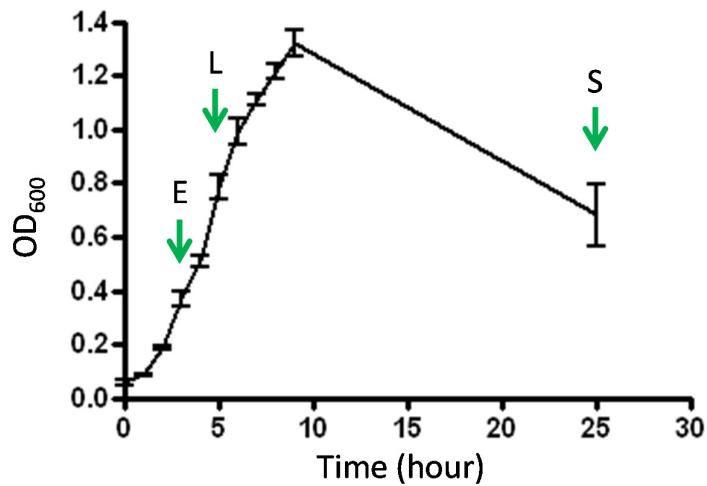


Figure 4.2: Extraction of *C. difficile* 630 Δ erm RNA

Example of 630 Δ erm growth in BHIS media. Green arrows indicate the time points chosen for RNA extraction. E = early exponential phase, L = late exponential phase, S = stationary phase.

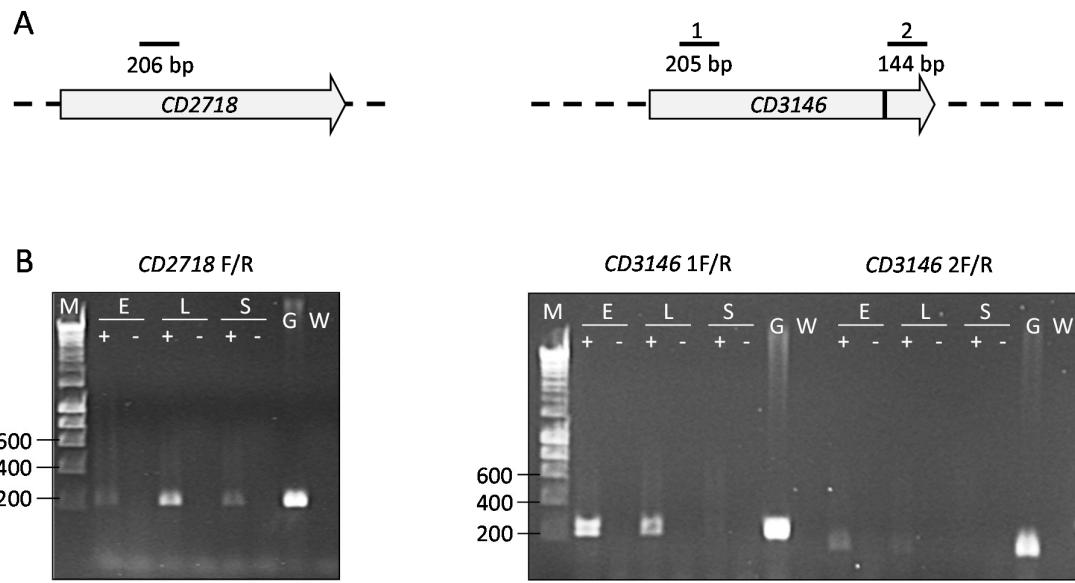


Figure 4.3: CD2718 and CD3146 transcription

A. Schematic diagrams representing the location of the primer pairs used to detect mRNA transcripts, and their expected sizes, from *CD2718* (left) and *CD3146* (right). The stop codon interrupting *CD3146* is indicated with a black line. **B.** PCR reactions were performed with wild-type 630Δerm cDNA that was prepared from cultures grown to early exponential (E), late exponential (L), and stationary phase (S). M = Hyperladder I (Bioline), G = 630Δerm genomic DNA, W = dH₂O. A “+” indicates cDNA reaction with added reverse transcriptase, “-” indicates cDNA reaction without added reverse transcriptase (control for DNA depletion of RNA sample).

4.2.1 Attempted construction of a defined isogenic knockout in CD2718

To assess the role of *CD2718* in *C. difficile*, we attempted to construct an isogenic knockout in *CD2718* using the ClosTron mutagenesis system (Heap *et al.*, 2007, Heap *et al.*, 2010), as described in Section 2.2.5. The mobile group II intron, LI.LtrB, was retargeted to two different positions of the sense strand of *CD2718*; between nucleotide positions 213 and 214 using pMTL007 (resulting in pEHD001), and between nucleotide positions 510 and 511 using pMTL007C-E2 (resulting in pEHD002). Both pEHD001 and pEHD002 were conjugated into *C. difficile* 630Δerm at least four separate times, with limited success. When targeting the 213/214 insertion site, thiamphenicol resistant colonies were obtained following three of the conjugations, suggesting that the pEHD001 plasmid successfully conjugated into *C. difficile*. After induction of the intron using IPTG, however,

integration of the intron into the chromosome was not observed, as there was no growth following lincomycin selection. When targeting the 510/511 insertion site, transconjugant *C. difficile* colonies were isolated following two of the conjugations using the pEHD002, only one of which led to the isolation of lincomycin resistant colonies. When replicated, a total of sixteen colonies were found to be lincomycin resistant (indicating chromosomal integration of the Li.ltrB intron containing the *erm* cassette), but thiamphenicol sensitive (indicating loss of the pEHD002 plasmid). Screening these colonies by PCR analysis revealed that the Li.ltrB intron had indeed integrated into the chromosome of 630Δ erm, but not within the *CD2718* gene (Figure 4.4). This was the only mutagenesis attempt that resulted in clones that could be screened.

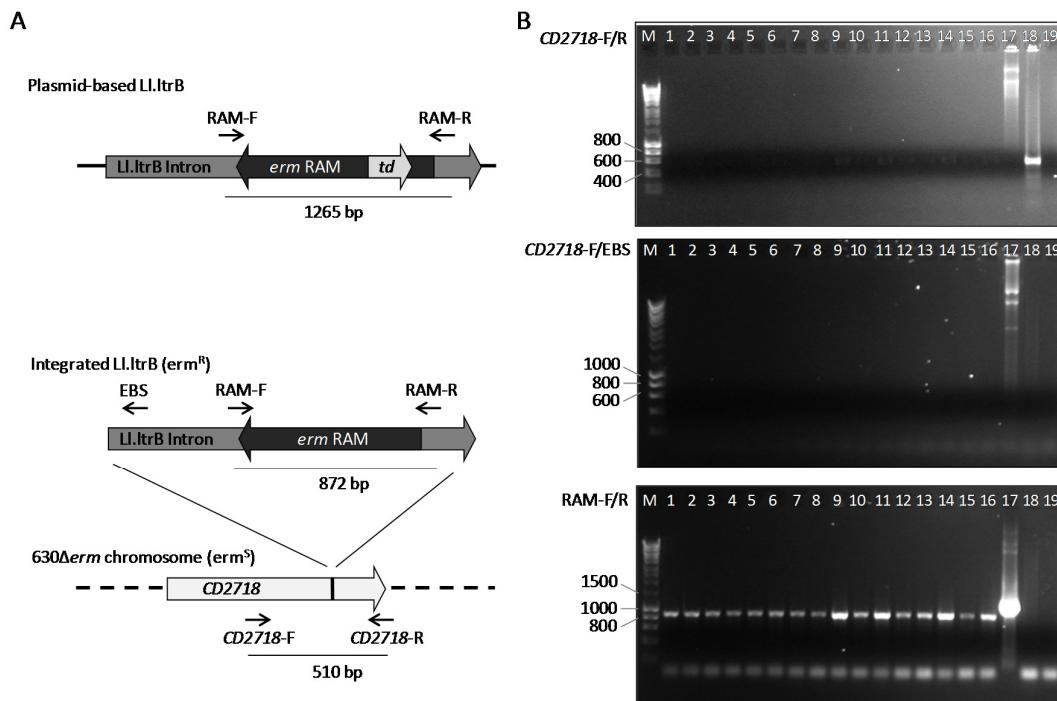


Figure 4.4: Screening potential *CD2718* ClosTron mutants

A. Schematic diagram of the insertional inactivation of *CD2718*. The location of the primers used to screen the potential mutant clones and the expected product sizes are indicated. **B.** Three PCRs were performed to screen sixteen potential 630Δ erm-*CD2718*-510s::CT clones. *CD2718-F/R* = 2.4 kb expected product (WT product of 0.5 kb), *CD2718-F/EBS* = 0.6 kb product, and *RAM-F/R* = 0.9 kb product when the group I intron is spliced out (Heap et al 2007). Despite the *RAM-F/R* PCR revealing that chromosomal integration of the Li.ltrB intron had occurred in all sixteen clones, the *CD2718-F/R* PCRs for all clones resulted in a faint WT size band (located within the dye front). There was no product in the *CD2718-F/EBS* PCRs, indicating that insertion of the

intron occurred at a site other than the intended *CD2718* CDS. M = Hyperladder I (Bioline), 1-16 = potential 630Δ*erm*-*CD2718*-510s::CT clones, 17 = pEHD002 plasmid, 18 = 630Δ*erm*, and 19 = dH2O control.

4.3 Characterisation of *CD2718* overexpression in *C. difficile*

As construction of a *CD2718* knockout remained unsuccessful, the function of *CD2718* was investigated by overexpression in *C. difficile*. Overexpression of SrtA in *S. aureus* was shown to increase the rate of surface protein anchoring *in vivo* (Mazmanian *et al.*, 1999), and this was also observed in SrtB overexpression (Mazmanian *et al.*, 2002). Observing changes that occur on the bacterial cell surface following overexpression of *CD2718* would assist in understanding its potential role in cell surface anchoring.

Overexpression in *C. difficile* was designed as described in Emerson *et al.* (2009). The CDS of *CD2718* was cloned into the *E. coli* – *C. difficile* shuttle vector pMTL960 under control of the moderately expressed Cwp2 promoter with a C-terminal 6xHis-tag to facilitate identification, resulting in plasmid pEHD007 (Figure 4.5). The N-terminal hydrophobic domain of *CD2718* (residues 2-27) was removed from pEHD007 using site-directed mutagenesis, resulting in pEHD008, in an attempt to improve protein solubility. Removing the N-terminal hydrophobic domain of SaSrtA improved protein solubility and yield when expressed in *E. coli*, and had little effect on the activity of SaSrtA observed *in vitro* (Ton-That *et al.*, 1999). Both expression plasmids were confirmed by PCR and sequencing prior to conjugation into *C. difficile* 630Δ*erm*. The empty pMTL960 was also conjugated into 630Δ*erm* to serve as a control for expression and growth in the presence of thiamphenicol. Phenotypic tests were performed to assess potential effects of overexpression of *CD2718* on the cell surface.



Figure 4.5: Schematic diagram of the overexpression construct

The CDS of CD2718 was cloned with a C-terminal 6x-His tag under the control of the moderately expressed *cwp2* promoter on the plasmid pMTL960. The location of the primers used to sequence both pEHD007 and pEHD008 are indicated.

4.3.1 Growth dynamics

Analysis of the growth dynamics of CD2718 overexpression was performed in triplicate in BHIS broth. No differences in growth were found between $630\Delta erm$ carrying either of the overexpression constructs (pEHD007 and pEHD008) or the empty vector (pMTL960) when grown in BHIS broth with thiamphenicol (Figure 4.6).

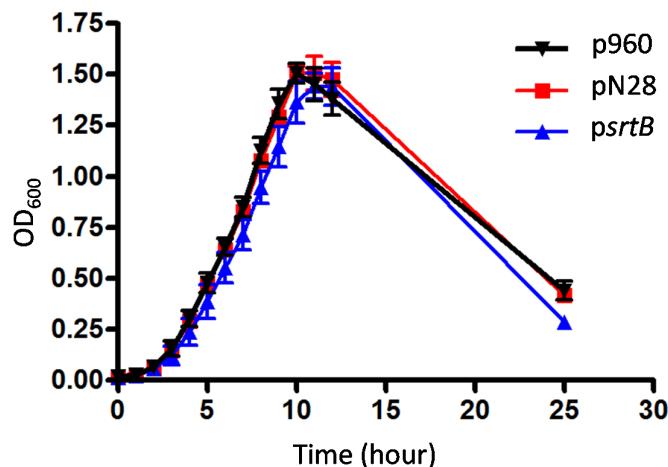


Figure 4.6: Growth analysis of sortase overexpression

Growth dynamics of CD2718 overexpression in BHIS broth. p960 = $630\Delta erm$ carrying the empty vector pMTL960, pN28 = $630\Delta erm$ carrying the pEHD008, psrtB = $630\Delta erm$ carrying the pEHD007 overexpression vector. Cultures were inoculated 1/100 using an overnight primary culture. Cultures were grown in duplicate and data points are from an average of three biological replicates.

4.3.2 Colony morphology

Overexpression of the cell wall protein, CwpV, resulted in an altered colony morphology phenotype (Reynolds *et al.*, 2011). The wildtype and $\Delta cwpV$ colonies were of similar size with ruffled edges, while overexpression of the full length CwpV resulted in smaller colonies with smoother edges. The observed macroscopic differences were visible at the cellular level (63x magnification) and attributed to altered cellular organisation within the bacterial colony. Colony morphology was compared between 630 Δerm (wildtype), 630 Δerm /pMTL960, and the overexpression strain 630 Δerm /pEHD007. All three strains have similar colony shapes with ruffled edges, and no differences were observed under 40x magnification (Figure 4.7).

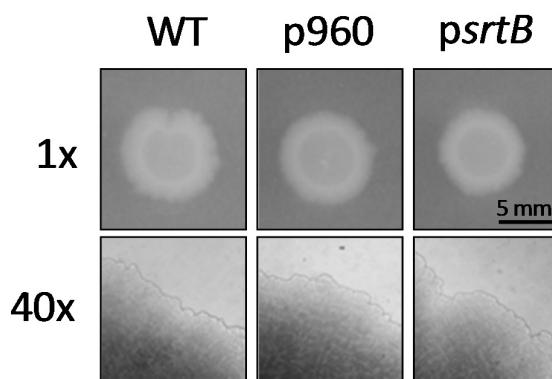


Figure 4.7: Colony morphology of CD2718 overexpression

Two μ l of overnight cultures (WT = 630 Δerm , p960 = 630 Δerm carrying the empty vector pMTL960, psrtB = 630 Δerm carrying the pEHD007 overexpression vector) were spotted onto BHIS plates and incubated for 3 days before removing plates from the anaerobic chamber to observe colony morphology both macroscopically and microscopically (40x magnification). No differences between colonies were observed. Photos were taken with Canon 600D SLR (1x) and a Zeiss Axioplan 2 upright microscope (40x).

4.3.3 Sedimentation of bacterial cells

Aggregation of bacterial cells in a liquid suspension was also observed in overexpression of CwpV, but not in the wildtype or in a *cwpV* mutant (Reynolds *et al.*, 2011). Similar to this aggregation assay, which assesses the propensity of *C. difficile* to aggregate in a highly concentrated suspension (starting OD₆₀₀ of 10), the Wren lab has developed a sedimentation assay that assesses the propensity of *C. difficile* cells to sediment at the bottom of a liquid culture during overnight growth (A. Faulds-Pain, LSHTM, unpublished

manuscript). Sedimentation was compared between *630Δerm* (wildtype), *630Δerm/pMTL960*, and the overexpression strain *630Δerm/pEHD007* (Figure 4.8). The non-motile RT078 strain M120 (Stabler *et al.*, 2009), sediments strongly in this assay, and is used as a positive control. There did not appear to be any differences in the level of sedimentation in strains with a *630Δerm* background.

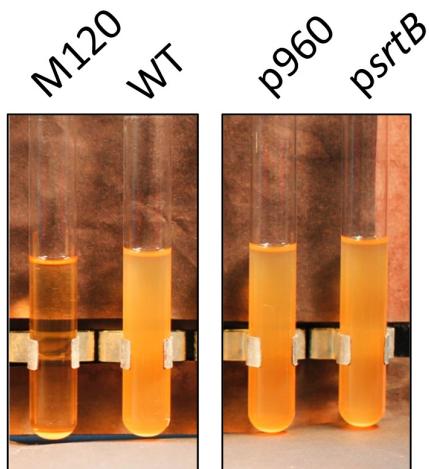


Figure 4.8: Sedimentation assay

Bacterial cell sedimentation following overnight growth in BHIS broth was compared between strains. M120 = a strongly sedimenting RT078 strain that serves as a positive control, WT = *630Δerm*, p960 = *630Δerm* carrying the empty vector pMTL960, psrtB = *630Δerm* carrying the pEHD007 overexpression vector.

4.3.4 Biofilm formation

As inactivating sortase has been shown to reduce biofilm production in some bacteria (Levesque *et al.*, 2005, Nobbs *et al.*, 2007), we investigated the effect overexpression of CD2718 had on *C. difficile* biofilm formation. Two separate biofilm assays were performed as described in Dawson *et al.* (2012). No significant differences were observed in biofilm formation in 10 ml static cultures of *630Δerm* (wildtype), *630Δerm/pMTL960*, and the overexpression strain *630Δerm/pEHD007* (Figure 4.9A). All three strains produced a biofilm attached to the bottom of the tissue culture flask that remained intact following detachment by gentle agitation. The crystal violet biofilm assay similarly revealed no differences in biofilm formation between the CD2718 overexpression strain and the controls (Figure 4.9B and C).

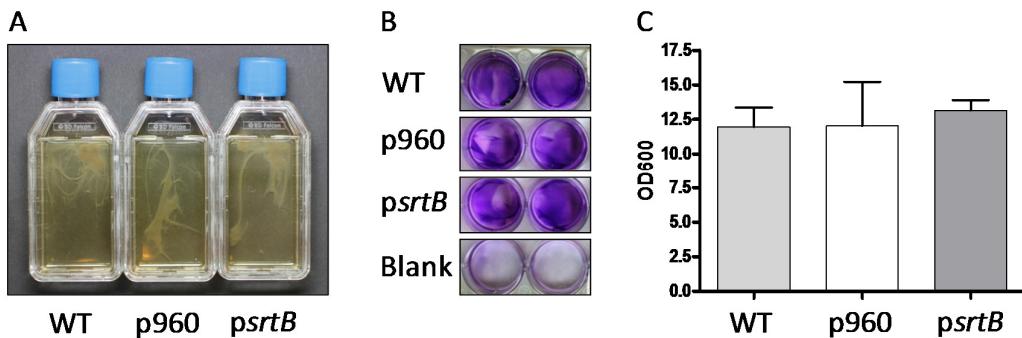


Figure 4.9: Biofilm assay

Three strains tested were WT = $630\Delta erm$, p960 = $630\Delta erm$ carrying the empty vector pMTL960, psrtB = $630\Delta erm$ carrying the pEHD007 overexpression vector. **A.** Ten ml cultures were grown statically for 3 days in tissue culture flasks to visualise biofilm attachment to abiotic surfaces. Biofilms were manually detached from the bottom of the flask by gentle agitation prior to photographing. **B** and **C.** Crystal violet assay: Strains were grown statically in 2 ml BHIS in 24-well plates for 3 days prior to crystal violet staining as described in Dawson *et al.* 2012. Images of the crystal violet stained biofilms were taken prior to methanol extraction (B), after which the OD₅₉₅ of each well was measured (C). Photos were taken with a Canon 600D SLR.

4.4 Expression of recombinant CD2718

Sortase activity is frequently demonstrated by means of an *in vitro* fluorescence resonance energy transfer (FRET) assay. FRET assays allow for the enzymatic cleavage of a fluorescently labelled peptide to be monitored over time, and have been useful in characterising sortases from *S. aureus* (Ton-That *et al.*, 1999, Mazmanian *et al.*, 2002), *B. anthracis* (Gaspar *et al.*, 2005, Marraffini & Schneewind, 2006), and *S. agalactiae* (Necchi *et al.*, 2011). Purification of soluble recombinant CD2718 protein would be a pre-requisite to develop a similar FRET assay to investigate its sortase activity. The following section describes the expression of recombinant CD2718 enzyme in both *C. difficile* and *E. coli*. The testing of the resulting purified proteins in the FRET assay is the subject of Chapter 5.

4.4.1 Expression of CD2718 in *C. difficile*

The CD2718 overexpression strains generated in Section 4.3 were investigated for the possibility of purifying soluble protein for use in an *in vitro* FRET assay. Western blotting was performed on whole cell lysates of *C. difficile* 630 and 630 Δ erm strains carrying the three plasmids (pMTL960, pEHD007, pEHD008) to assess recombinant protein expression levels. Expression of only the full length CD2718, but not the N-terminally truncated CD2718 $_{\Delta N28}$, was observed in whole cell lysates of *C. difficile* cultures (Figure 4.10A). There did not appear to be any difference in the level of CD2718 expression in the two background strains tested, 630 and 630 Δ erm. Following centrifugation, the CD2718 protein sedimented with the membranes (Figure 4.10B). Only a faint band is observed in the lane containing the soluble fraction of 630/pEHD007, indicating that the protein is mostly insoluble and that purification of this protein by nickel affinity chromatography under native conditions was not possible.

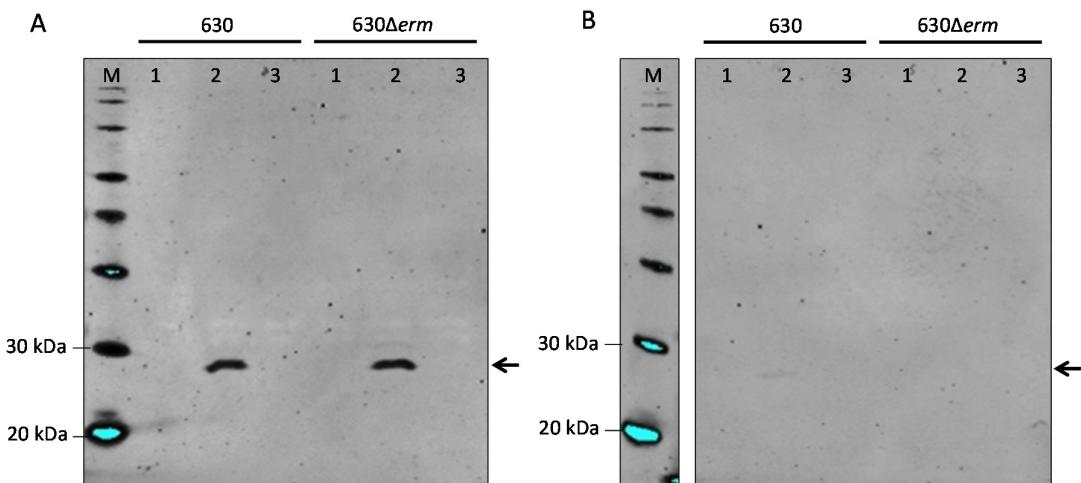


Figure 4.10: Overexpression of CD2718 in *C. difficile*

Anti-His western blots testing the expression and solubility of CD2718 in *C. difficile* strains 630 and 630 Δ erm. **A.** Whole cell lysates of 10 ml *C. difficile* cultures. A 27 kDa band corresponding to the predicted size of the full length CD2718 was observed from the strains carrying the pEHD007 plasmid. **B.** Supernatant of whole cell lysates following centrifugation. The 27 kDa protein observed in Panel A appears to be insoluble as it is only very faintly observed after centrifugation. L = PageRuler™ Plus Prestained Protein Ladder (Thermo Scientific), M = MagicMark™ XP Western Protein Standard (Invitrogen), 1 = pMTL960 empty vector, 2 = pEHD007, and 3 = pEHD008.

4.4.2 Expression of CD2718 in *E. coli*

As overexpression of CD2718 in *C. difficile* did not lead to the expression of soluble protein, constructs were designed for the expression of recombinant CD2718 in *E. coli*. Several different expression constructs were tested – a full length CD2718 (Section 4.4.2.1), as well as two N-terminally truncated CD2718 proteins (Section 4.4.2.2 and 4.4.2.3) – and all proteins were purified both in the presence or absence of 1 mM dithiothreitol (DTT) for testing in a FRET assay (detailed in Chapter 5). DTT is a reducing agent that is often used to stabilise enzymes and prevent oxidation of cysteines (Cleland, 1964), and has been shown to restore activity of sortase that has been inactivated by inhibitors (Ton-That *et al.*, 1999). The sortase B from *S. aureus* (SaSrtB) was also expressed and purified (Section 4.4.2.4) to serve as a control in the FRET assay.

CD2718	M-----KKLYRIVINIILVLVILYSGFNIYSKLTKNHDTKISSELQKKEYKKEDLSKI
60	
CD2718 (N28)	MHHHHHH-----LTKYNHDTKISSELQKKEYKKEDLSKI 40
CD2718 (N26)	MHHHHHH-----SKLTKNHDTKISSELQKKEYKKEDLSKI 42

CD2718	NSDFKFWLSVENTNINYPVVQSKDN SYLDKDFYKKDSISGTLFMDYRNKSIDDKNIIY 120
CD2718 (N28)	NSDFKFWLSVENTNINYPVVQSKDN SYLDKDFYKKDSISGTLFMDYRNKSIDDKNIIY 100
CD2718 (N26)	NSDFKFWLSVENTNINYPVVQSKDN SYLDKDFYKKDSISGTLFMDYRNKSIDDKNIIY 102

CD2718	GHNMKNKTMFNNLNKFKDADFFKKNNKIKITLNGKEFLYDVFSAYIVESDYDYLKTNFNN 180
CD2718 (N28)	GHNMKNKTMFNNLNKFKDADFFKKNNKIKITLNGKEFLYDVFSAYIVESDYDYLKTNFNN 160
CD2718 (N26)	GHNMKNKTMFNNLNKFKDADFFKKNNKIKITLNGKEFLYDVFSAYIVESDYDYLKTNFNN 162

CD2718	ESDYQNYINDITSKSLYKSPIKVNSNDKIVTLSTCTYEFDDARMVIHGRLILEHHHHHH 233
CD2718 (N28)	ESDYQNYINDITSKSLYKSPIKVNSNDKIVTLSTCTYEFDDARMVIHGRLI----- 205
CD2718 (N26)	ESDYQNYINDITSKSLYKSPIKVNSNDKIVTLSTCTYEFDDARMVIHGRLI----- 207

Figure 4.11: Comparison of recombinant CD2718 proteins for expression in *E. coli*

Sequence alignment of the three recombinant CD2718 proteins expressed in *E. coli*. CD2718 contains the full CDS, while CD2718_{N28} and CD2718_{N26} have deletions in the signal peptide. HHHHHH indicates the location of the 6x-His tag, and ----- indicates the signal peptide deletion.

4.4.2.1 Expression and purification of full length CD2718

The expression vector pQE-30 Xa (Qiagen) was initially chosen for expression of CD2718 in *E. coli* as the related pQE-30 was used for SaSrtA expression (Ton-That *et al.*, 2000). The CDS of CD2718 was codon-optimised for expression in *E. coli* due to the usage of rare codons by *C. difficile*. Using the Rare Codon Calculator

(<http://nihserver.mbi.ucla.edu/RACC/>), we determined that CD2718 contains four rare arginine codons (AGA), three rare leucine codons (CTA), and 12 rare isoleucine codons (ATA). The CDS of CD2718 was codon-optimised and cloned into the BamHI-HindIII sites of pQE-30 Xa by Celtek Bioscience LLC to generate plasmid pEHD009. However, there were difficulties in expressing CD2718 from this vector. Resequencing the vector backbone revealed that, though the CDS of *CD2718* was intact and in frame with the N-terminal 6x-His tag on the vector, a 26 bp region had unexpectedly excised from the T5 promoter region.

Another expression vector, pET28a, was chosen for expressing the sortase in *E. coli* under the control of the IPTG-inducible T7 promoter. The codon-optimised CDS of *CD2718* (from pEHD009) was cloned into the Ncol/Xhol sites of pET28a to generate the plasmid pEHD011 (Figure 4.11). The expected recombinant CD2718 is 27.8 kDa with a C-terminal 6xHis tag (encoded on the pET28a vector immediately downstream of the Xhol site).



Figure 4.12: Schematic diagram of pEHD011 construct

Protein expression from pET28a is under the control of the IPTG-inducible T7 promoter, and results in a C-terminally 6xHis-tagged recombinant protein with an expected mass of 27.8 kDa.

CD2718 expression was induced with 1 mM IPTG for 5 hours in 10 ml cultures before harvesting the cells and probing SDS-PAGE separated whole cell lysates with anti-6xHis antibody. A 28 kDa protein that reacted with the anti-6xHis antibody was observed that was not present in the non-induced control (Figure 4.12A). When analysing expression over time in BL21(DE3), the amount of 28 kDa protein observed in whole-cell lysates was roughly proportional to induction duration, with an increase in signal intensity at 5 hours induction over 1 and 3 hours induction (Figure 4.12B). There was not an appreciable increase in production of the 28 kDa for induction times greater than 5 hours.

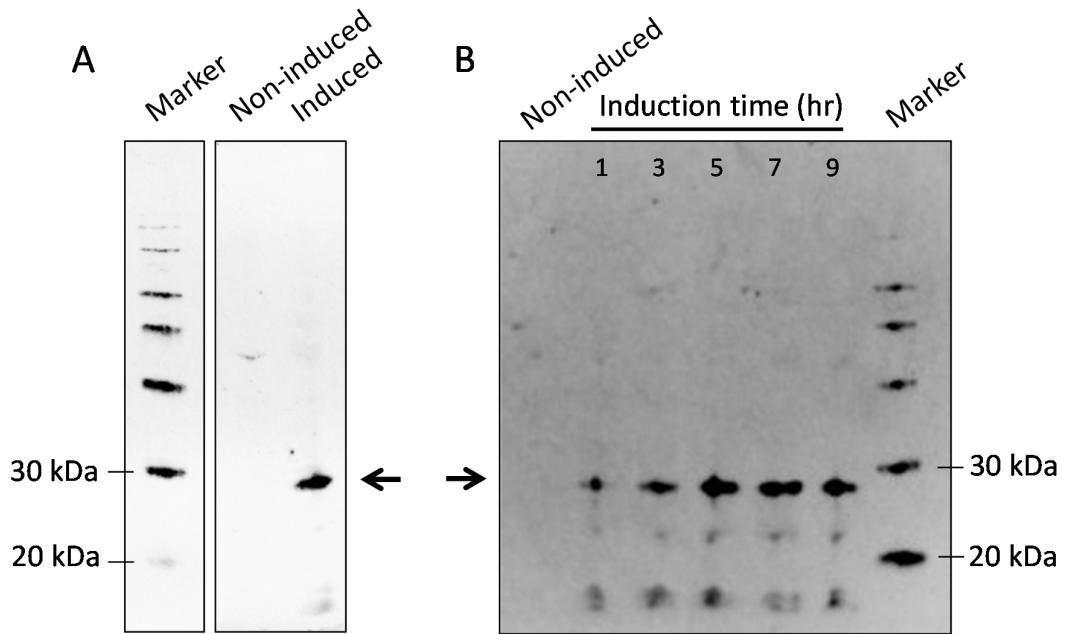


Figure 4.13: Expression of CD2718 from BL21(DE3) cells

Anti-6xHis western blots of OD-matched whole-cell lysates from 10 ml cultures of BL21(DE3)/pEHD011. **A.** A 28 kDa protein was observed after 5 hours of induction with 1 mM IPTG, but not in the non-induced culture. **B.** Time course of expression in BL21(DE3) cells. Hours post-induction are indicated. Marker = MagicMark™ XP Western Protein Standard (Invitrogen).

CD2718 was purified from 10 ml cultures of BL21(DE3)/pEHD011 induced with IPTG for 5 hours. Cell lysates were batch incubated with 1 ml nickel agarose slurry (Qiagen) for 1 hour, before eluting bound protein with 250 mM imidazole. The eluate was separated by SDS-PAGE and transferred to a PVDF membrane for N-terminal sequencing, and the 28 kDa protein was confirmed as CD2718 (Figure 4.13). The 25 kDa band below CD2718 was identified as SlyD, a common *E. coli* contaminant protein following nickel affinity purification. N-terminal sequencing was performed by Mike Weldon (PNAC Facility, University of Cambridge).

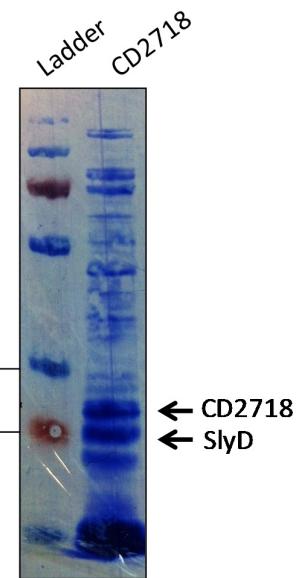


Figure 4.14: Small-scale purification of CD2718

The cell lysate of a 10 ml induced BL21(DE3)/pEHD011 culture was batch incubated with 1 ml nickel agarose slurry (Qiagen) for 1 hour, before eluting bound protein with 250 mM imidazole. The eluate was separated by SDS-PAGE and transferred to a PVDF membrane for N-terminal sequencing. The identity of the 28 kDa protein was confirmed as CD2718. Ladder = PageRuler Plus Prestained Protein Ladder (Fermentas).

Expression of the CD2718 was scaled up for purification using an HPLC AKTApurifier system. Two liters of BL21(DE3)/pEHD011 were induced at 37 °C for 5 hours with 1 mM IPTG. The harvested cells were lysed with detergent and sonication, and the cleared cell lysate applied to a 1 ml HisTrap HP column. Bound protein was eluted from the column over an imidazole gradient, which resulted in two peaks (Figure 4.14A). When elution fractions were separated by SDS-PAGE and stained with Coomassie, the 28 kDa CD2718 was predominantly found in the second more defined peak (Figure 4.14B). The fractions containing the largest amounts of CD2718 were pooled together and applied to a Superdex HiLoad 16/60 200 gel filtration column to improve purity and to remove the high concentration of imidazole (Figure 4.14C). A small peak eluted after 45 ml that corresponded to the void volume of the column, and a larger peak eluted after 90 ml. This peak was analysed by SDS-PAGE and found to contain the 28 kDa CD2718 (Figure 4.14D). Fractions F8-F13 were pooled and concentrated with an Amicon Ultra centrifugal filter column before quantifying the protein content by the BCA or Bradford protein assay

(Pierce). A range of 0.4-0.5 mg recombinant CD2718 was obtained regularly from 2 L cultures of *E. coli*.

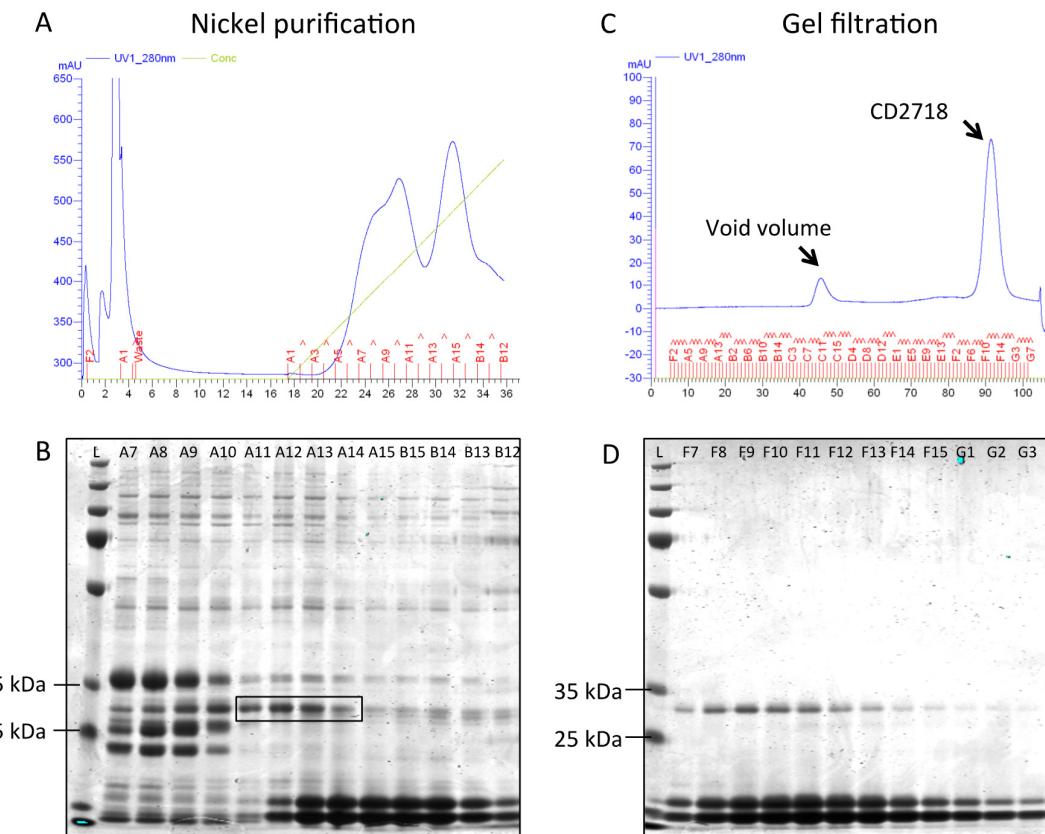


Figure 4.15: Purification of CD2718 from *E. coli* BL21(DE3)

Two litres of BL21(DE3)/pEHD011 were induced with 1 mM IPTG for 5 hours prior to harvesting. The cell pellet was lysed and the lysate applied to a nickel affinity column using an HPLC system. Bound protein was eluted over an imidazole gradient, after which the relevant elution fractions were subjected to gel filtration to improve purity. **A.** HPLC trace of nickel affinity purification. The green line indicates the increasing imidazole concentration, and the blue line corresponds to protein content as measured at 280 nm. Fractions collected by the HPLC fraction collector are indicated as A1-B12. **B.** Fractions from nickel purification were separated by SDS-PAGE and stained by Coomassie. Fractions A11-A14, containing CD2718 (boxed), were pooled together and applied to a gel filtration column. **C.** HPLC trace of gel filtration chromatography of nickel purified sample. **D.** Gel filtration fractions F7-G3 were separated by SDS-PAGE and visualised by Coomassie staining, and found to contain CD2718. L = PageRuler Plus Prestained Protein Ladder (Fermentas).

4.4.2.2 Expressing CD2718_{ΔN28}

Protein yield of recombinant membrane proteins is generally improved following removal of the membrane segment. When the N-terminal transmembrane segment of SaSrtA is replaced with a six histidine tag, the yield of soluble protein is improved without any effect on its activity, either as a deletion of residues 2-25 (Ton-That *et al.*, 1999) or 2-58 (Ton-That *et al.*, 2000, Ilangovan *et al.*, 2001). This was explored to improve yield of CD2718 protein. The N-terminal transmembrane region of CD2718 was predicted by the TMHMM Server to include residues 2-27, so residues 28-225 were amplified from pEHD011 and recloned into the same Ncol-Xhol sites of pET28a to create pEHD012 (Figure 4.15). In this plasmid, the residues 2-27 have been replaced with a six histidine tag (Figure 4.15).



Figure 4.16: Schematic diagram of the pEHD012 expression construct

Residues 28-225 of the codon-optimised CD2718 were amplified from pEHD011 to construct pEHD012. The N-terminal membrane spanning residues 2-27 were replaced with a 6x histidine tag.

The expected size of the purified CD2718_{ΔN28} protein is 24.5 kDa, a similar size observed earlier for the common contaminant following nickel affinity purification in *E. coli*, SlyD (Figure 4.13). In order to avoid confusing expression of SlyD with that of CD2718_{ΔN28}, an engineered BL21(DE3) strain, called NiCo21(DE3), was used for expression of this recombinant protein. Six common *E. coli* contaminant proteins in this expression strain have either been tagged with a chitin-binding domain or have been altered to reduce the number of exposed histidine residues (Robichon *et al.*, 2011). In this strain, a chitin binding domain has been incorporated into SlyD (SlyD-CBD) in order to facilitate removal from the sample following purification. The mutant SlyD-CBD protein migrates at 35 kDa by SDS-PAGE (Robichon *et al.*, 2011). Initial small-scale expression tests of NiCo21(DE3)/pEHD012 identified a 25 kDa protein that reacted with the anti-6xHis tag in the induced culture that was not present in the non-induced control (Figure 4.16A). Whole cell lysates were spun down to separate the soluble cell lysate from the insoluble portion. The 25 kDa protein was found in the soluble fraction, along with a 35 kDa protein that

reacted with both the anti-chitin binding domain (anti-CBD) antibody and the anti-6xHistag antibody and presumed to be the mutant SlyD-CBD protein (Figure 4.16B). A large portion of the SlyD-CBD protein bound to chitin resin and was removed from the soluble cell lysate, but the 25 kDa protein was lost in the processing (Figure 4.16B).

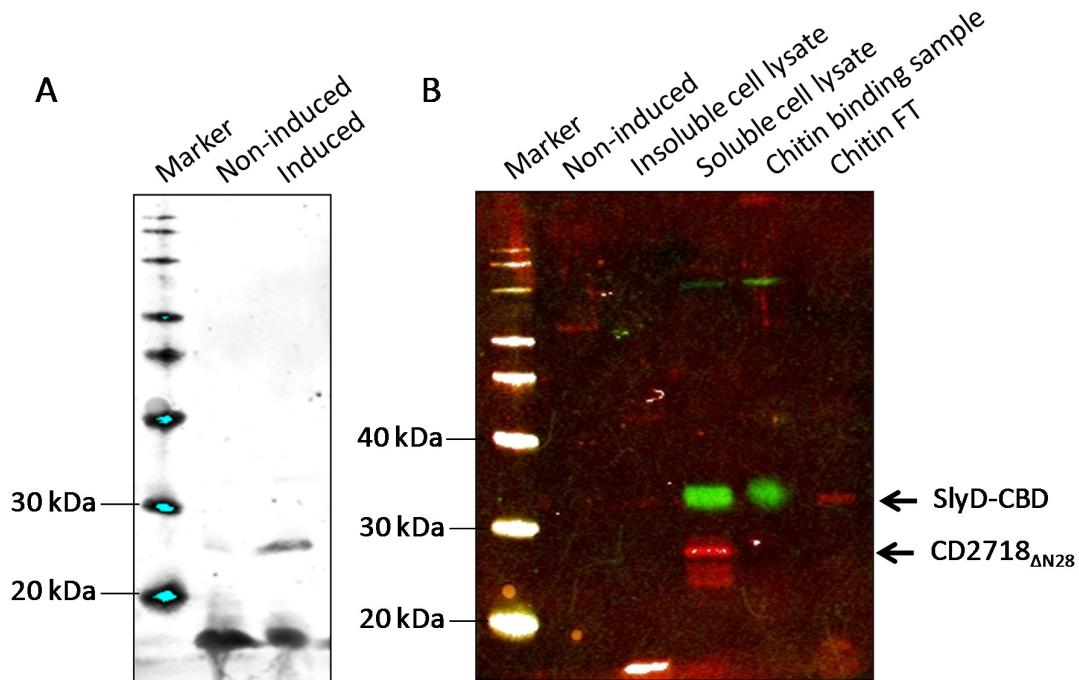


Figure 4.17: Initial expression testing of pEHD012

Anti-6xHis western blots of whole-cell lysates from 10 ml cultures of NiCo21(DE3)/pEHD012. A. A 25 kDa protein was observed after 5 hours of induction with 1 mM IPTG, but not in the non-induced culture. B. Two colour western of NiCo21(DE3)/pEHD012 whole-cell lysates. Centrifugation of whole cell lysates revealed the 25 kDa protein to be soluble, as well as the mutant SlyD-CBD. Incubation of the soluble cell lysate with chitin beads removed most of the SlyD-CBD protein, but the 25 kDa protein was lost in the processing. Green = Anti-CBD antibody, red = anti-6xHistag antibody, Marker = MagicMark™ XP Western Protein Standard (Invitrogen).

CD2718_{ΔN28} was purified from 10 ml cultures of NiCo21(DE3)/pEHD012 induced with 1 mM IPTG for 5 hours. Cell lysates were batch incubated with 1 ml nickel agarose slurry (Qiagen) for 1 hour, before eluting bound protein with 250 mM imidazole. The eluate was separated by SDS-PAGE and transferred to a PVDF membrane for N-terminal sequencing, and the 25 kDa protein was confirmed as CD2718_{ΔN28} (Figure 4.17). N-terminal sequencing was performed by Mike Weldon (PNAC Facility, University of Cambridge).

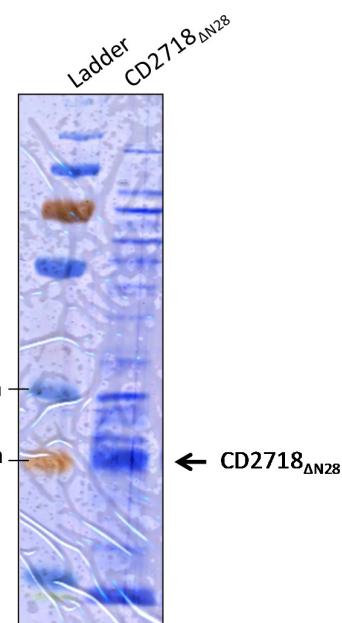


Figure 4.18: Small-scale purification of CD2718_{ΔN28}

The cell lysate of a 10 ml induced NiCo21(DE3)/pEHD012 culture was batch incubated with 1 ml nickel agarose slurry (Qiagen) for 1 hour, before eluting bound protein with 250 mM imidazole. The eluate was separated by SDS-PAGE and transferred to a PVDF membrane for N-terminal sequencing. The identity of the 25 kDa protein was confirmed as CD2718_{ΔN28}. Ladder = PageRuler Plus Prestained Protein Ladder (Fermentas).

Two liters of NiCo21(DE3)/pEHD012 were induced at 37 °C for 5 hours with 1 mM IPTG to scale up for purification using an HPLC AKTApurifier system as described previously for CD2718. The CD2718_{ΔN28} protein eluted from the HisTrap nickel column during fractions A11-A14. Gel filtration of these pooled samples resulted in a large peak after 90 ml that was found to contain CD2718_{ΔN28} (Figure 4.18). Fractions F10-F15 were pooled and concentrated with an Amicon Ultra centrifugal filter column before quantifying the protein content by BCA or Bradford protein assay. A total of 0.5 mg recombinant CD2718_{ΔN28} was obtained from the 2 liter culture. The gel filtration was sufficient to separate the 35 kDa SIYD-CBD from the 25 kDa CD2718_{ΔN28} protein without incubation of the sample with chitin beads.

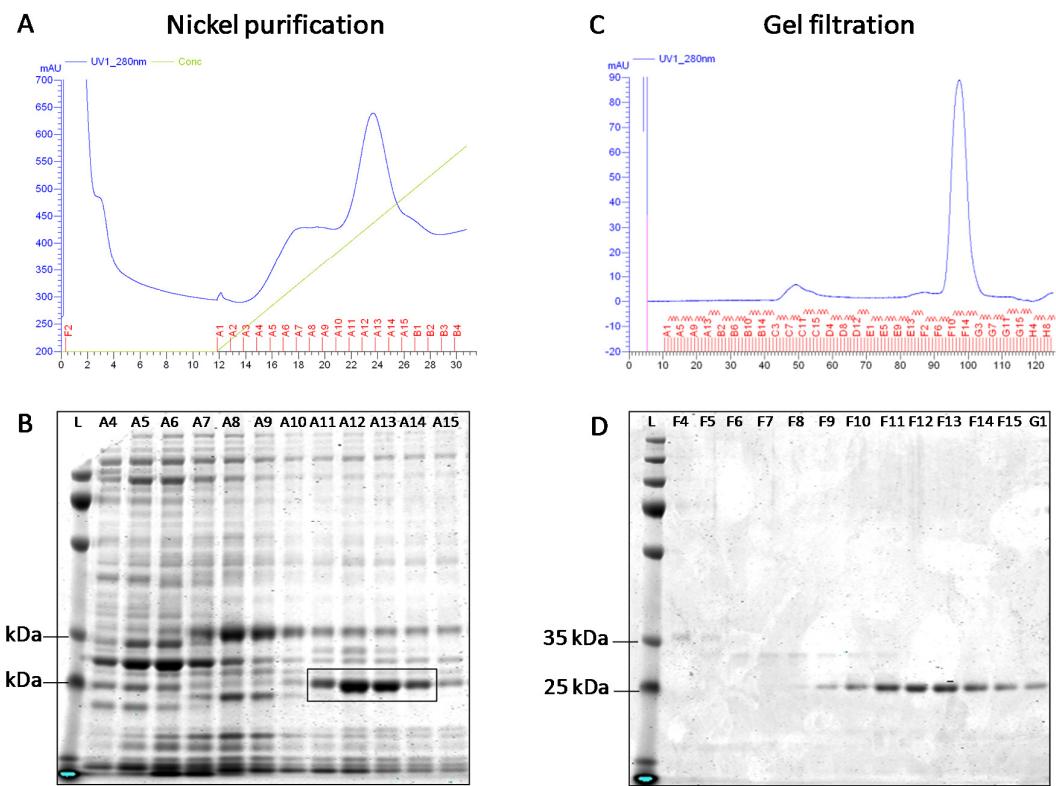


Figure 4.19: Purification of $\text{CD2718}_{\Delta N28}$ from *E. coli* NiCo21(DE3)

Two liters of NiCo21(DE3)/pEHD012 were induced with IPTG for 5 hours prior to harvesting. The cell pellet was lysed and the lysate applied to a nickel affinity column using an HPLC system. Bound protein was eluted over an imidazole gradient, after which the relevant elution fractions were subjected to gel filtration to improve purity. **A.** HPLC trace of nickel affinity purification. The green line indicates the increasing imidazole concentration, and the blue line corresponds to protein content as measured at 280 nm. Fractions collected by the HPLC fraction collector are indicated as A1-B4. **B.** Fractions from nickel purification were separated by SDS-PAGE and stained by Coomassie. Fractions A11-A14, containing $\text{CD2718}_{\Delta N28}$ (boxed), were pooled together and applied to a gel filtration column. **C.** HPLC trace of gel filtration chromatography of nickel purified sample. **D.** Gel filtration fractions F4-G1 were separated by SDS-PAGE and visualised by Coomassie staining, and fractions containing $\text{CD2718}_{\Delta N28}$ (F10-F15) were pooled and concentrated. L = PageRuler Plus Prestained Protein Ladder (Fermentas).

4.4.2.3 Expressing $\text{CD2718}_{\Delta N26}$

Another transmembrane domain prediction server, Phobius, predicted the N-terminal transmembrane region of CD2718 to include residues 2-25. Site-directed mutagenesis was performed to replace residues 26 and 27 (serine and lysine, respectively) to the

pEHD012 plasmid to construct pEHD013. Addition of the six nucleotides was confirmed by sequencing prior to transforming the pEHD013 construct into NiCo21(DE3) cells for expression. The expected size of the recombinant CD2718_{ΔN26} protein is 24.7 kDa, with an N-terminal 6x-Histag similar to pEHD012 (Figure 4.15). Whole cell lysates of 10 ml NiCo21(DE3)/pEHD013 cultures indicated the presence of a 25 kDa band in cultures induced with 1 mM IPTG but not in the non-induced control (Figure 4.19). When induced total cell lysates were centrifuged, this 25 kDa band remained in the soluble portion (Figure 4.19).

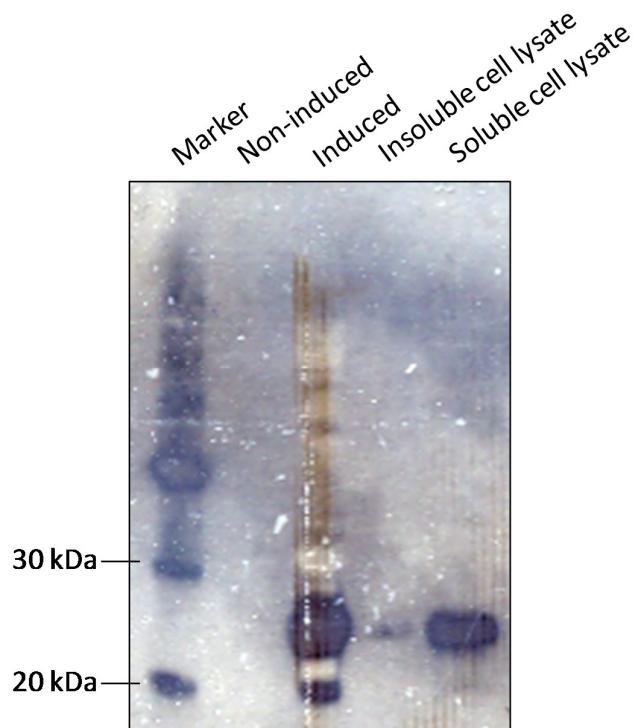


Figure 4.20: Initial expression testing of pEHD013

Anti-6xHis western blots of whole-cell lysates from 10 ml cultures of NiCo21(DE3)/pEHD013. A 25 kDa protein was observed after 5 hours of induction with 1 mM IPTG, but not in the non-induced culture. Centrifugation of whole cell lysates revealed the 25 kDa protein to be soluble. Marker = MagicMark™ XP Western Protein Standard (Invitrogen).

Two liters of NiCo21(DE3)/pEHD013 were induced at 37 °C for 5 hours with IPTG to scale up for purification using an HPLC AKTApurifier system as described previously (Figure 4.20). The CD2718_{ΔN26} protein eluted from the HisTrap nickel column during fractions A13-A15 (Figure 4.20A and B). Gel filtration of these pooled samples resulted in a large peak at 90 ml that was found to contain CD2718_{ΔN26} (Figure 4.20C and D), a peak roughly double

the height of the peaks previously observed following purification of CD2718 and CD2718_{ΔN28}. Fractions F8-F14 were pooled and concentrated with an Amicon Ultra centrifugal filter column before quantifying the protein content by BCA or Bradford protein assay. A total of 1-1.5 mg recombinant CD2718_{ΔN26} was obtained regularly from 2 liter cultures. The gel filtration was sufficient to separate the 35 kDa SlyD-CBD from the 25 kDa CD2718_{ΔN26} protein without incubation of the sample with chitin beads. Identity of the 25 kDa protein purified by HPLC was confirmed as CD2718_{ΔN26} by MALDI fingerprinting (Figure 4.21) performed by Jun Wheeler (NIBSC, HPA).

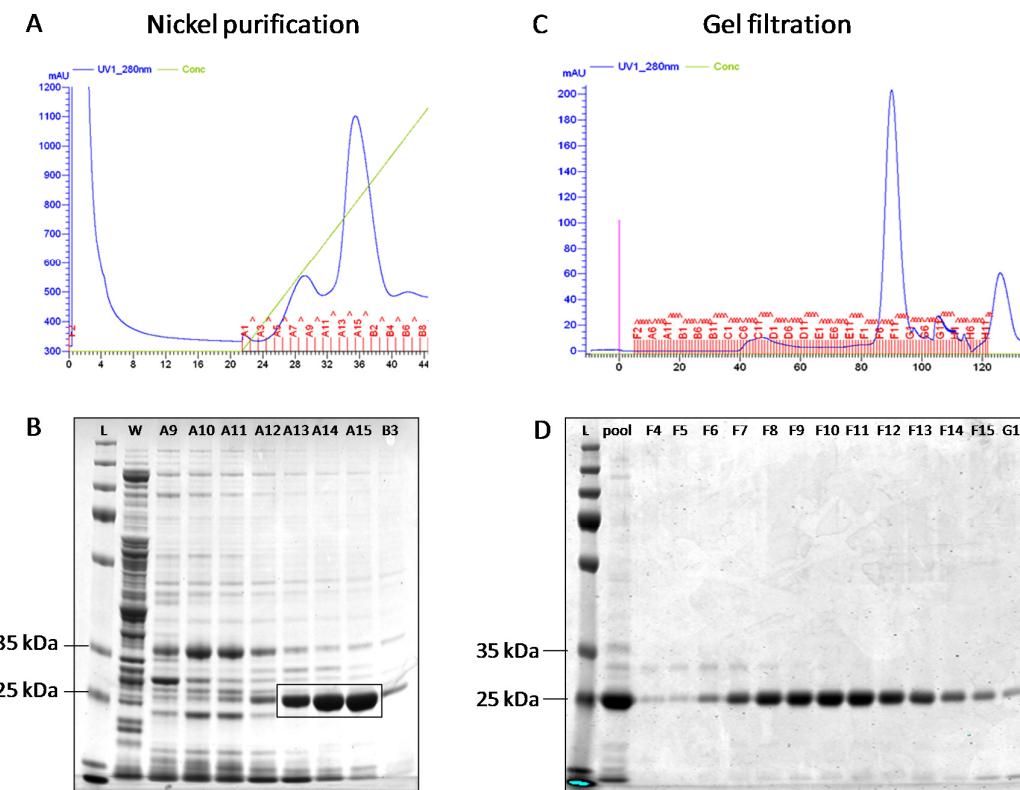


Figure 4.21: Purification of CD2718_{ΔN26} from *E. coli* NiCo21(DE3)

Two liters of NiCo21(DE3)/pEHD013 were induced with IPTG for 5 hours prior to harvesting. The cell pellet was lysed and the lysate applied to a nickel affinity column using an HPLC system. Bound protein was eluted over an imidazole gradient, after which the relevant elution fractions were subjected to gel filtration to improve purity. **A.** HPLC trace of nickel affinity purification. The green line indicates the increasing imidazole concentration, and the blue line corresponds to protein content as measured at 280 nm. Fractions collected by the HPLC fraction collector are indicated as A1-B8. **B.** Fractions from nickel purification were separated by SDS-PAGE and stained by Coomassie. Fractions A13-A15, containing CD2718_{ΔN26} (boxed), were pooled

together and applied to a gel filtration column. **C.** HPLC trace of gel filtration chromatography of nickel purified sample. **D.** Gel filtration fractions were separated by SDS-PAGE and visualised by Coomassie staining, and fractions containing CD2718_{ΔN26} (F8-F14) were pooled and concentrated. L = PageRuler Plus Prestained Protein Ladder (Fermentas).

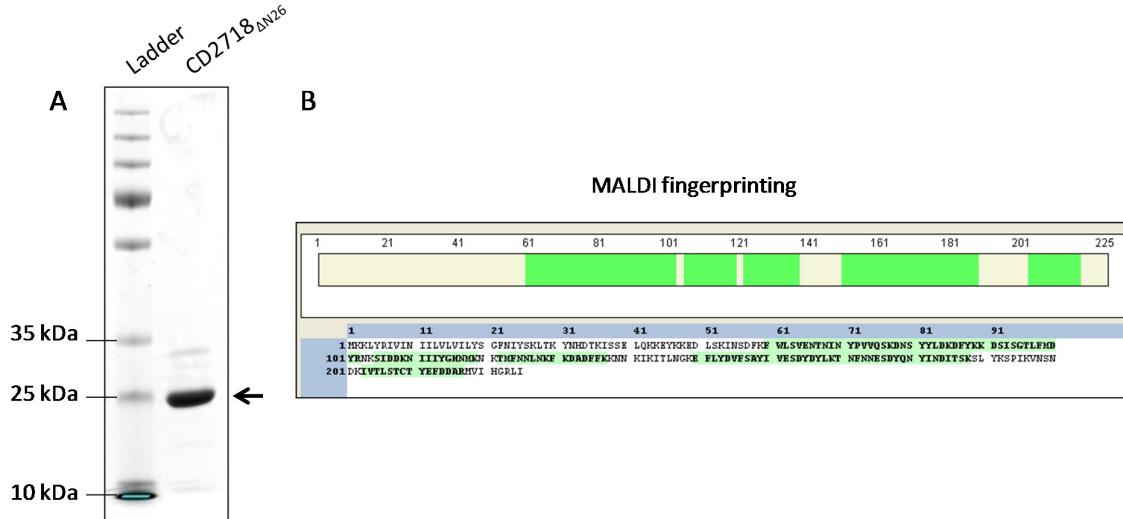


Figure 4.22: MALDI fingerprinting analysis of CD2718_{ΔN26}

A. Gel filtration fractions F7-F14 (Figure 4.20C and D) were pooled and concentrated in an Amicon Ultra centrifugal filter column before quantifying by BCA or Bradford assay, and separating by SDS-PAGE. Ladder = PageRuler Plus Prestained Protein Ladder (Fermentas). **B.** The identity of the 25 kDa protein indicated by an arrow in A was confirmed as CD2718_{ΔN26} by MALDI fingerprinting performed by Jun Wheeler (NIBSC, PHE). Highlighted residues correspond to peptide fragments identified in the sample.

4.4.2.4 Expressing *S. aureus* SrtB

The sortase B from *S. aureus* (SaSrtB) was intended to serve as a control in the FRET assay (Chapter 5). SaSrtB cleaves Dabcyl-NPQTN-Edans peptides (Oh et al 2006). The CDS of SrtB from *S. aureus* Newman was codon optimised for expression in *E. coli* and cloned into the HindIII/BamHI site of pQE-30 Xa by Celtek Bioscience LLC to generate plasmid pEHD010. Residues 26-244 of SaSrtB were cloned into the NcoI/Xhol sites of pET28a to create pEHD015, replacing residues 2-25 with a 6xHistag (Mazmanian et al 2002). The resulting recombinant SaSrtB_{ΔN28} has a predicted size of 27.0 kDa. Induction with IPTG resulted in the appearance of a 28 kDa band that reacted with the anti-6xHistag antibody (Figure 4.22A). This band remained soluble following centrifugation of the total cell lysate, and eluted from nickel agarose resin with 250 mM imidazole (Figure 4.22B). The eluate was

separated by SDS-PAGE and stained with Coomassie (Figure 4.22C), and the 28 kDa protein was confirmed as SaSrtB_{ΔN28} by MALDI fingerprinting performed by Jun Wheeler (Figure 4.22D) (NIBSC, HPA).

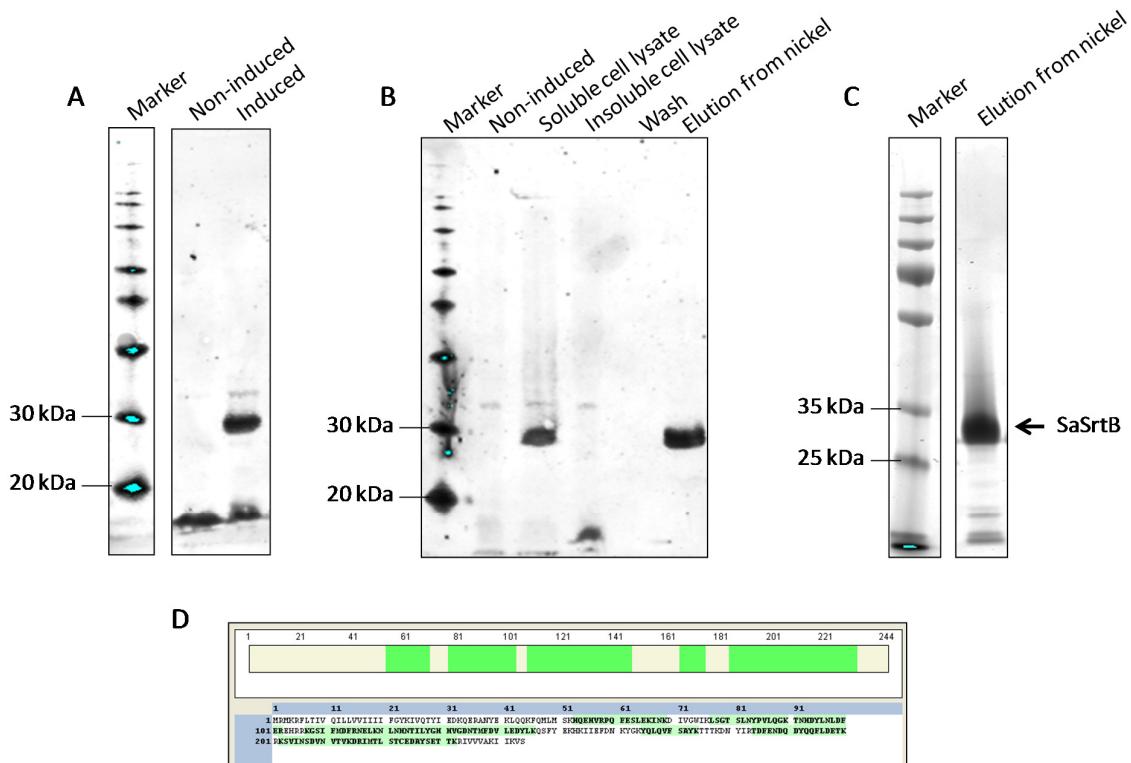


Figure 4.23: Expression and purification of pEHD015

Anti-6xHis western blots of whole-cell lysates from 10 ml cultures of NiCo21(DE3)/pEHD015. **A.** A 28 kDa protein was observed after 5 hours of induction with 1 mM IPTG, but not in the non-induced culture. **B.** Centrifugation of whole cell lysates revealed the 28 kDa protein to be soluble. The protein bound to nickel resin and eluted in 250 mM imidazole. **C.** Eluate from nickel resin incubated with soluble cell lysates were separated by SDS-PAGE, stained with Coomassie. **D.** The resulting 28 kDa band was identified as SaSrtB_{ΔN28} by MALDI fingerprinting. Marker = MagicMark™ XP Western Protein Standard (Invitrogen), Ladder = PageRuler Plus Prestained Protein Ladder (Fermentas). Highlighted residues correspond to peptide fragments identified in the sample.

Expression and purification of SaSrtB_{ΔN28} was scaled up as described previously for CD2718, CD2718_{ΔN28}, and CD2718_{ΔN26}. SaSrtB_{ΔN28} was purified from 2 liters NiCo21(DE3)/pEHD015 by nickel affinity chromatography followed by gel filtration on an HPLC system (Figure 4.23). SaSrtB_{ΔN28} obtained from gel filtration was concentrated and

quantified by BCA or Bradford assay, resulting in 1-1.5 mg of SaSrtB_{ΔN28} from 2 liters of culture.

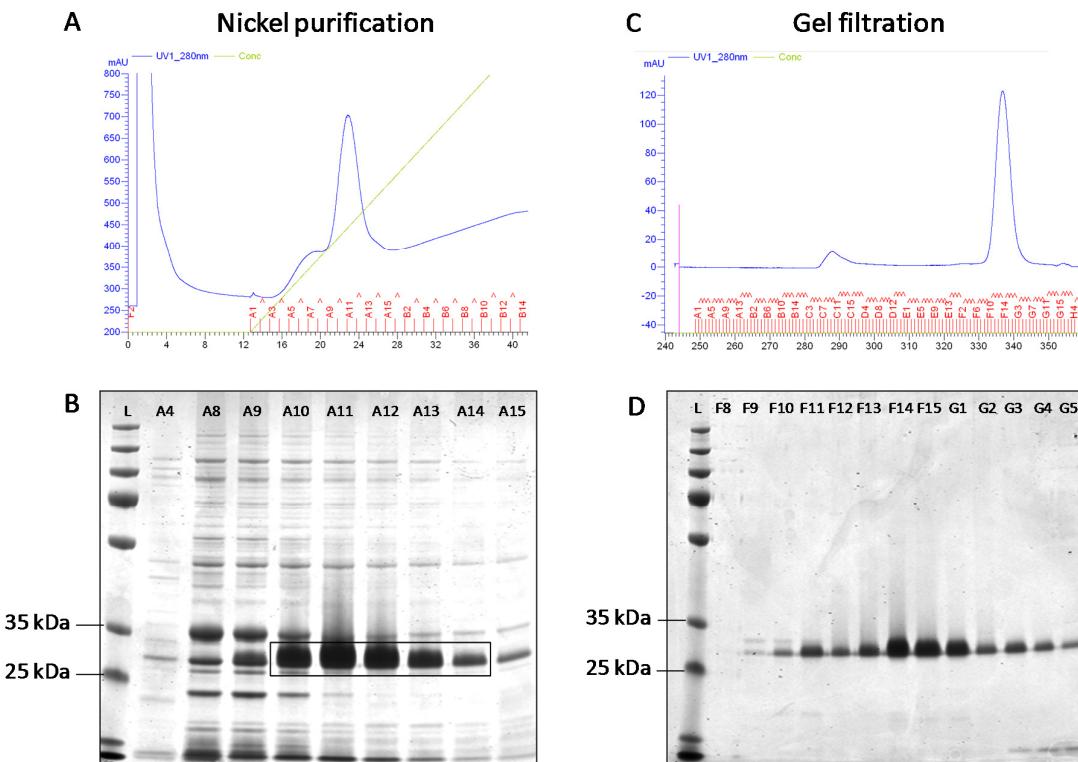


Figure 4.24: Purification of SaSrtB_{ΔN28} from *E. coli* NiCo21(DE3)

Two liters of NiCo21(DE3)/pEHD015 were induced with IPTG for 5 hours prior to harvesting. The cell pellet was lysed and the lysate applied to a nickel affinity column using an HPLC system. Bound protein was eluted over an imidazole gradient, after which the relevant elution fractions were subjected to gel filtration to improve purity. **A.** HPLC trace of nickel affinity purification. The green line indicates the increasing imidazole concentration, and the blue line corresponds to protein content as measured at 280 nm. Fractions collected by the HPLC fraction collector are indicated as A1-B14. **B.** Fractions from nickel purification were separated by SDS-PAGE and stained by Coomassie. Fractions A10-A14, containing SaSrtB_{ΔN28} (boxed), were pooled together and applied to a gel filtration column. **C.** HPLC trace of gel filtration chromatography of nickel purified sample. **D.** Gel filtration fractions F8-G5 were separated by SDS-PAGE and visualised by Coomassie staining, and fractions containing SaSrtB_{ΔN28} (F12-G2) were pooled and concentrated. L = PageRuler Plus Prestained Protein Ladder (Fermentas).

4.5 Discussion

In this chapter, we attempted an initial characterisation of the predicted sortase in *C. difficile*. We sought to construct an inactivation mutant to facilitate functional characterisation of CD2718 using ClosTron technology, but despite repeated attempts, this was unsuccessful. Initially, it was hypothesised the sortase might be essential in *C. difficile* due to the difficulty encountered in mutagenesis. However, this is not the case, as successful construction of a *CD2718* mutant has been performed by others (H. A. Shaw, Imperial College London, personal communication).

As mutagenesis was not possible in our lab, we sought to investigate CD2718 by overexpressing the protein in *C. difficile*. We hypothesised that overexpression of CD2718 could increase the amount of surface protein anchoring and lead to detectable changes in the profile of cell wall associated proteins, as has been observed for *srtA* overexpression in *S. aureus* and *B. subtilis* (Mazmanian *et al.*, 1999, Nguyen *et al.*, 2011). However, we found that overexpression of CD2718 in *C. difficile* did not have an observable phenotype in the simple screening assays tested. *C. difficile* produces biofilms that protect the bacteria from oxidative stress and antibiotics *in vitro* (Dawson *et al.*, 2012, Ethapa *et al.*, 2013). In other bacteria, knocking out sortase and removing the proteins it anchors from the cell surface results in decreased biofilm production (Levesque *et al.*, 2005, Nobbs *et al.*, 2007). We investigated whether overexpression of CD2718 might have the opposite effect and increase biofilm production in *C. difficile*, but found there was no difference. Similarly, we examined colony morphology and cell aggregation as an indication of cell surface alterations, but did not observe any differences due to CD2718 overexpression. It is possible that expression under the *cwp2* promoter, which is only moderately expressed in *C. difficile* (Emerson *et al.*, 2009), is too low to cause any discernable changes to the cell surface. Another promoter with a higher level of expression, or the recently developed Tet-inducible promoter for *C. difficile* (Fagan & Fairweather, 2011) could be better suited to investigate this.

Unfortunately, overexpression of CD2718 in *C. difficile* did not lend itself to the purification of soluble protein for downstream enzymatic assays. The full length CD2718 protein with the N-terminal transmembrane domain intact sedimented with the membranes following centrifugation of *C. difficile* whole cell lysates. In an attempt to

produce soluble CD2718, we removed the hydrophobic region (residues 2-27), but we were unable to detect recombinant expression of this N-terminally truncated CD2718 ($\text{CD2718}_{\Delta\text{N}28}$) in *C. difficile*. In addition to removing the insoluble transmembrane domain, deletion of residues 2-27 would also eliminate the signal peptide sequence from $\text{CD2718}_{\Delta\text{N}28}$. Removal of the signal peptide may result in the mislocation or misfolding of $\text{CD2718}_{\Delta\text{N}28}$, leaving the protein susceptible to degradation by proteases. Proteases such as the Clp proteases in *E. coli* and *B. subtilis* regularly degrade abnormal or misfolded proteins, serving a clean-up function to limit the cellular stress caused by an accumulation of non-functional or abnormal proteins (Gottesman, 1996, Kruger *et al.*, 2000).

As purification of soluble CD2718 was unsuccessful from *C. difficile*, this chapter also describes the successful expression and purification of CD2718, as well as SaSrtB, in *E. coli*. These proteins were purified to enable *in vitro* investigations into the activity of CD2718, described in detail in Chapter 5. Three versions of CD2718 were expressed to test in the FRET assay, including a full length protein and two with different deletions in the N-terminal transmembrane domain. These deletions were designed based on the predictions of two different transmembrane domain prediction servers, TMHMM and Phobius. Both truncations to the N-terminus of CD2718 result in increased protein yield over the expression of the full length CD2718, likely due to improved solubility. Interestingly, the truncation of residues 2-25 ($\text{CD2718}_{\Delta\text{N}26}$) had greater protein yields than that of the 2-27 residue truncation ($\text{CD2718}_{\Delta\text{N}28}$). It is not immediately clear the effect two additional N-terminal residues would have on protein expression. These two additional residues in $\text{CD2718}_{\Delta\text{N}26}$ may perhaps contribute to conformational stability of CD2718, as we have an indication that the $\text{CD2718}_{\Delta\text{N}26}$ protein is more tightly folded than $\text{CD2718}_{\Delta\text{N}28}$ (W. Savory, Domainex Ltd., personal communication).

An engineered strain of *E. coli*, NiCo21(DE3) (Robichon *et al.*, 2011), was used to express the two truncated CD2718 proteins. Initial attempts at CD2718 expression in standard *E. coli* BL21(DE3) cells led to considerable contamination with SlyD, a histidine-rich 24 kDa *E. coli* protein. SlyD is a commonly found *E. coli* contaminant protein in nickel affinity chromatography (Bolanos-Garcia & Davies, 2006). It exhibits a high binding affinity to both nickel and cobalt resins due to its high proportion of histidine residues (10.2%) and elutes from these resins at imidazole concentrations greater than 80 mM, much greater than the

concentrations recommended for washing. Removing SlyD from a protein sample after purification is difficult, especially if the proteins are similar in size which would exclude gel filtration. The NiCo21(DE3) strain expresses a 35 kDa SlyD tagged with a chitin-binding domain, with the purpose of facilitating SlyD removal using chitin resin (Robichon *et al.*, 2011). However, we found that the increase in SlyD size was sufficient to allow easy separation from the two 24 kDa truncated CD2718 proteins by gel filtration.

5 Development of a FRET-based assay to assess relative *C. difficile* sortase activity

5.1 Introduction

In Chapter 4, we describe the expression and purification of recombinant CD2718 protein. The activity of sortases has been thoroughly investigated by means of *in vitro* fluorescence resonance energy transfer (FRET) assays. FRET assays allow for the enzymatic cleavage of an LPXTG-containing peptide to be monitored by utilising a reporter fluorophore and quencher pair bound at either end of the peptide. Fluorescence of the fluorophore within the peptides is blocked when in close proximity to the quencher molecule. When the peptide is cleaved, the quencher is no longer able to absorb the emission of the fluorophore, and a fluorescent signal is observed. FRET-based experiments have contributed greatly to the understanding of the biochemical mechanisms of the sorting reaction. The active site structure of sortases have been investigated by mutating predicted essential residues based on structural analyses in SaSrtA (Ton-That *et al.*, 2002), as well as SaSrtB and BaSrtA (Zhang *et al.*, 2004). Kinetic analyses of SaSrtA (Ton-That *et al.*, 2002, Marraffini *et al.*, 2004, Frankel *et al.*, 2005), SpSrtA (Race *et al.*, 2009), and BaSrtA (Weiner *et al.*, 2010) have allowed insight into reaction mechanisms, and the screening and development of sortase inhibitors (Oh *et al.*, 2006, Maresco *et al.*, 2007, Liew *et al.*, 2004). Such an assay for the putative *C. difficile* sortase would be instrumental in characterisations of its activity, and would allow the screening of potential CD2718 inhibitors.

5.1.1 Aims

This chapter describes the development and testing of a fluorescent based FRET assay to evaluate the activity of the predicted *C. difficile* sortase using the recombinant CD2718 proteins purified in Chapter 4.

The aims of developing this assay were to:

- demonstrate CD2718 exhibits sortase-like activity by cleaving fluorescent peptides containing the predicted (S/P)PXTG and NVQTG motifs (see Chapter 3)

- characterise the activity of CD2718 by measuring enzyme kinetics and substrate specificity
- screen potential CD2718 inhibitors and determine their suitability for downstream drug development

5.2 Development of a FRET assay for CD2718

Fluorescence resonance energy transfer (FRET) peptides were designed based on the variations observed in the predicted (S/P)PXTG sequon found in Chapter 3. The fluorophore and quencher pair used in peptide synthesis (performed by Peptide Protein Research Ltd) was Edans and Dabcyl, respectively. Fluorescence of the Edans fluorophore is blocked when in close proximity to the fluorescent quencher Dabcyl, as the emission spectrum of Edans overlaps entirely with the absorbance spectrum of Dabcyl (Figure 5.1) (Matayoshi *et al.*, 1990). When the peptide is cleaved and the Edans fluorophore (excitation = 340 nm, emission = 490 nm) is separated from Dabcyl, a fluorescent signal is observed.

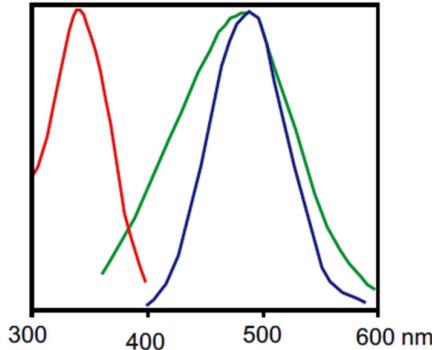


Figure 5.1: Absorbance and emission spectra for Dabcyl and Edans

Edans has a wavelength of maximum absorption of 340 nm (in red) and an emission peak at 490 nm (in blue). The Edans emission peak overlaps entirely with the absorption peak of Dabcyl (in green). Figure adapted from Matayoshi *et al.*, 1990.

Two residues upstream of the pattern were included, and two glycine residues were incorporated downstream, as this has been previously shown to improve sortase cleavage efficiency *in vitro* (Pritz *et al.*, 2007). In the experiment performed by Pritz *et al.* (2007), the cleavage efficiency of all possibilities of LPKTGX peptides was monitored after 7 and

24 hours (Figure 5.2). A glycine residue immediately downstream of the LPKTG motif resulted in the highest peptide turnover, while proline, aspartic acid and tryptophan resulted in the lowest peptide turnover. In the eight possible CD2718 substrates identified in Chapter 3, seven of them have an aspartic acid immediately following the proposed CD2718 recognition sequence (Table 3.3).

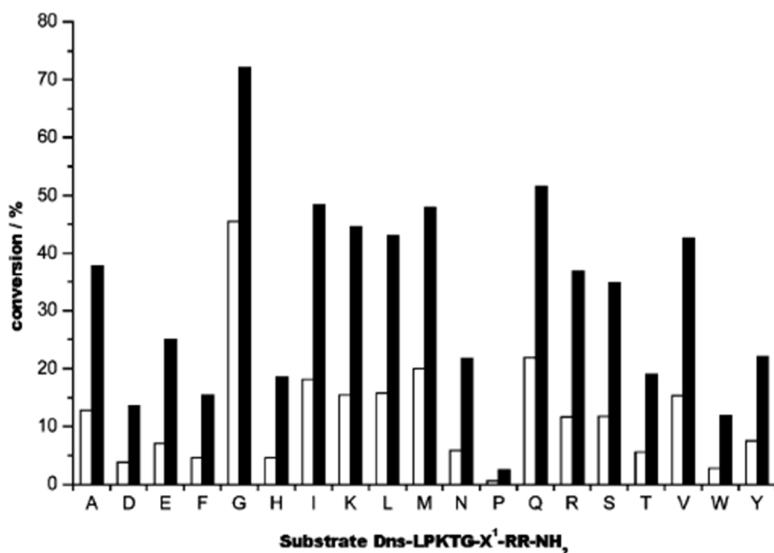


Figure 5.2: Effect of downstream residues on FRET peptide cleavage

In an experiment by Pritz *et al.*, (2007), all twenty possible LPKTGX peptides, where X is any amino acid, were incubated with SaSrtA, and conversion of substrate to product was measured by HPLC fluorescence detection. A glycine (G) residue immediately following the LPKTG motif resulted in the highest peptide cleavage efficiency, while an aspartic acid (D), proline (P), or tryptophan (W) following the motif were among the lowest cleavage efficiencies. White = 7 hours, black = 24 hours. Figure from Pritz *et al.*, 2007.

As the SPKTG motif occurs in half of the predicted *C. difficile* sortase substrates, the Dabcyl-SDSPKTGGG-Edans peptide (*d*-SDSPKTGGG-*e*) was chosen to initially screen for cleavage activity of the recombinant CD2718 proteins purified in Chapter 4 (Section 4.4.2). The peptide *d*-NPQTN-*e*, shown to be cleaved by SaSrtB (Oh *et al.*, 2006), was used to test the activity of the purified SaSrtB. During optimisation of the *C. difficile* sortase FRET assay, the reaction was set up as follows: 4 µM peptide was incubated with 20 µM purified enzyme (with or without 1 mM DTT) and fluorescence measured over a period of 24 hours at 37 °C. The concentrations tested were within ranges previously used in sortase FRET assays: 1-10 µM labelled peptide substrate and 4-10 µM purified sortase enzyme (Ton-That *et al.*, 1999, Ton-That *et al.*, 2000, Mazmanian *et al.*, 2002, Maresso *et al.*, 2007,

Weiner *et al.*, 2010). When the full length CD2718 protein (purification described in Section 4.4.2.1) was incubated with the *d*-SDSPKTGGG-*e* peptide, no change in fluorescence, measured as arbitrary relative fluorescence units (RFU), was observed after 24 hours compared to the peptide alone (Figure 5.3A). Purifying CD2718 in the presence of 1 mM DTT had no effect on fluorescent signal (Figure 5.3A). This assay was repeated with the N-terminally truncated sortase protein, CD2718_{ΔN28}, in which residues 2-27 have been removed (purification described in Section 4.4.2.2). Similar to the full length CD2718, there was no change in fluorescent signal following incubation of the *d*-SDSPKTGGG-*e* peptide with CD2718_{ΔN28} after 24 hours (Figure 5.3B). The addition of DTT to the purification and FRET assay buffers had no effect (Figure 5.3B). Unexpectedly, the SaSrtB_{ΔN28} enzyme (purification described in Section 4.4.2.4) did not appear to be active, as no change in fluorescence was also observed when incubated with *d*-NPQTN-*e* (Figure 5.3C).

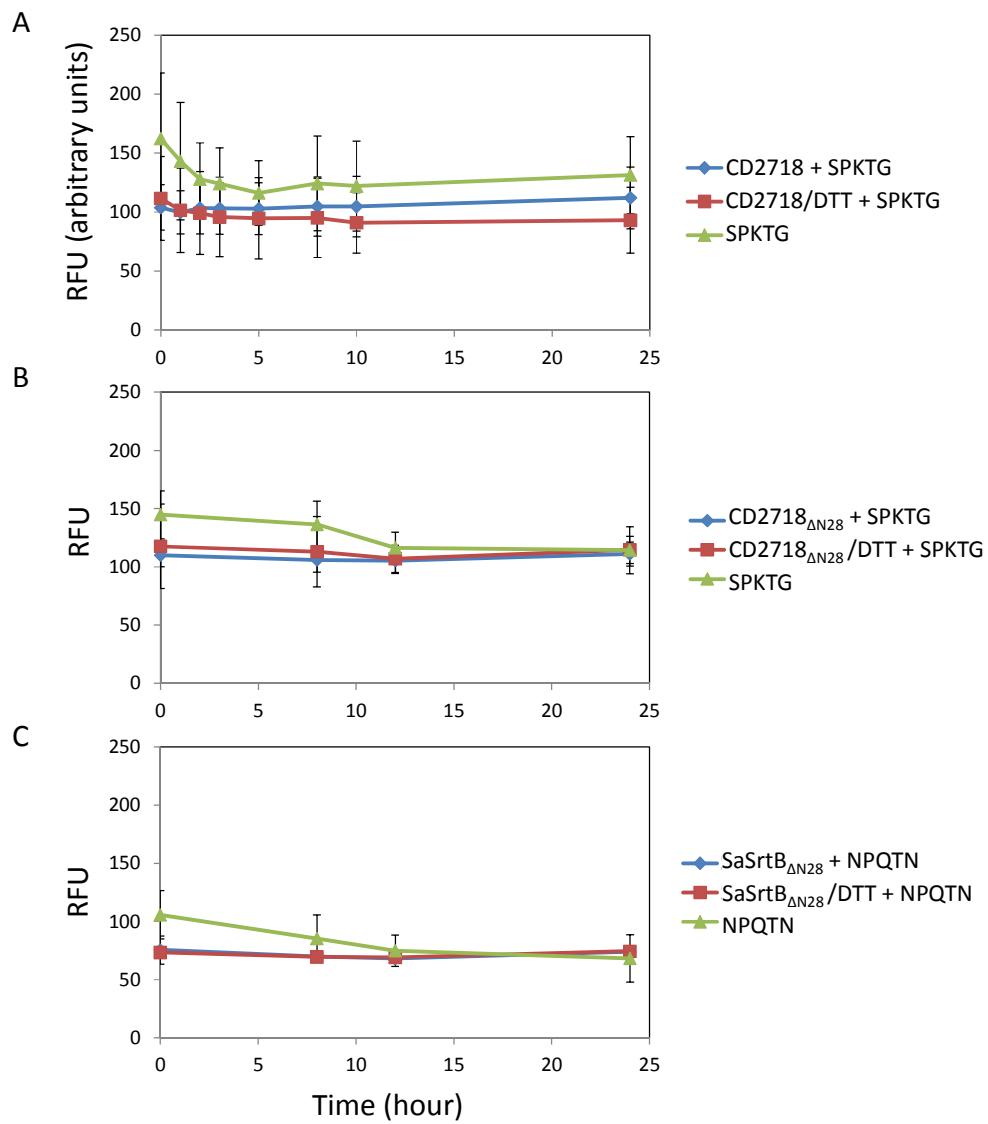


Figure 5.3: Initial FRET experiments

Twenty μM purified sortase was incubated with 4 μM FRET peptide at 37 °C. Fluorescence was measured (excitation at 340 nm, emission at 490 nm) as relative fluorescence units (RFU, arbitrary units) over a period of 24 hours. An increase in fluorescent signal would suggest cleavage of the peptide by sortase, but this was not observed for any of three proteins tested here. **A/B.** CD2718 (**A**) and CD2718_{ΔN28} (**B**) were incubated with *d*-SDSPKTGGG-*e*. **C.** SaSrtB was incubated with *d*-NPQTN-*e*, a peptide previously shown to be cleaved by SaSrtB by Oh *et al.* (2006). “/DTT” indicates the protein was purified in the presence of 1 mM DTT, and that the reaction buffer also contained 1 mM DTT. All data points were collected in triplicate, and the overall assay was run in triplicate. Error bars represent the standard error of the mean.

Another N-terminally truncated sortase protein, CD2718_{ΔN26} (purification described in Section 4.4.2.3), in which only residues 2-25 are removed, was tested in the FRET assay. When CD2718_{ΔN26} was initially tested for activity against the *d*-SDSPKTGGG-e peptide, fluorescence was measured for the first 8 hours, and then not again until after 72 hours. However, at 72 hours, an increase in fluorescence was observed in the three wells containing the CD2718_{ΔN26} purified with 1 mM DTT, suggesting that CD2718_{ΔN26} had cleaved the *d*-SDSPKTGGG-e (Figure 5.4A). When fluorescence was measured again, but this time at regular intervals within the 72 hour period, the same increase in fluorescence was observed over time, indicating cleavage of *d*-SDSPKTGGG-e by the CD2718_{ΔN26} protein purified with DTT (Figure 5.4B).

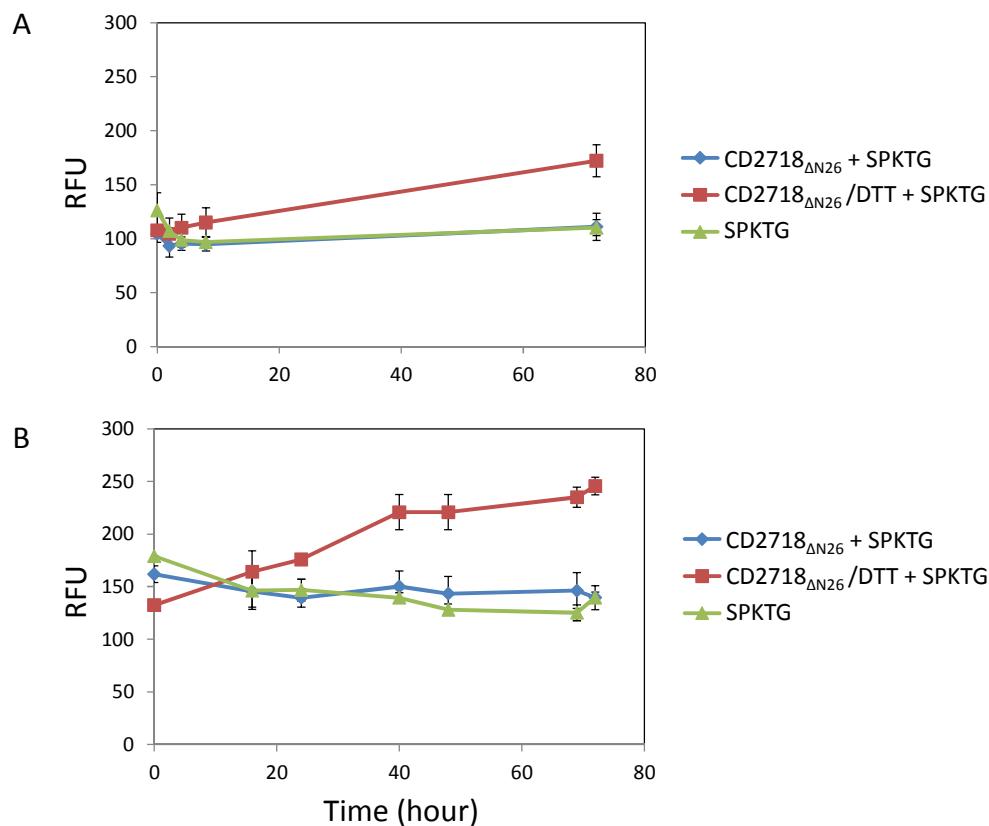


Figure 5.4: FRET activity with CD2718_{ΔN26}

Twenty μM purified CD2718_{ΔN26} (purification detailed in Section 4.4.2.3) was incubated with 4 μM *d*-SDSPKTGGG-e at 37 °C. An increase in fluorescent signal (excitation at 340 nm, emission at 490 nm) over time indicates cleavage of the peptide. **A.** In a single experiment, fluorescence was measured for the first 8 hours, and then after 72 hours. After this time, there was an increase in fluorescence observed in the wells containing

$\text{CD2718}_{\Delta N26}/\text{DTT}$ when compared to the d -SDSPKTGGG-e peptide alone. **B.** Fluorescence was measured over a period of 72 hours, and cleavage of d -SDSPKTGGG-e by the $\text{CD2718}_{\Delta N26}$ protein purified with DTT. “/DTT” indicates the protein was purified in the presence of 1 mM DTT, and that the reaction buffer also contains 1 mM DTT. Error bars represent the standard error of the mean.

To confirm that the increase in fluorescence observed over time was due to the presence of the $\text{CD2718}_{\Delta N26}$ protein, $\text{CD2718}_{\Delta N26}$ and $\text{SaSrtB}_{\Delta N28}$ (both purified with DTT) were incubated with both the d -SDSPKTGGG-e and Dabcyl-NPQTN-Edans peptides. Despite unexpectedly not being active against the d -NPQTN-e peptide, the $\text{SaSrtB}_{\Delta N28}$ protein was still able to be used as a control for the presence of protein in the reaction. Fluorescence was measured over a period of 48 hours, after which, only the $\text{CD2718}_{\Delta N26}$ incubated with the d -SDSPKTGGG-e peptide demonstrated an increase in fluorescent signal (Figure 5.5). $\text{CD2718}_{\Delta N26}$ did not cleave the d -NPQTN-e peptide, and there was no subsequent increase in fluorescence with the d -SDSPKTGGG-e peptide in the presence of another protein (Figure 5.5).

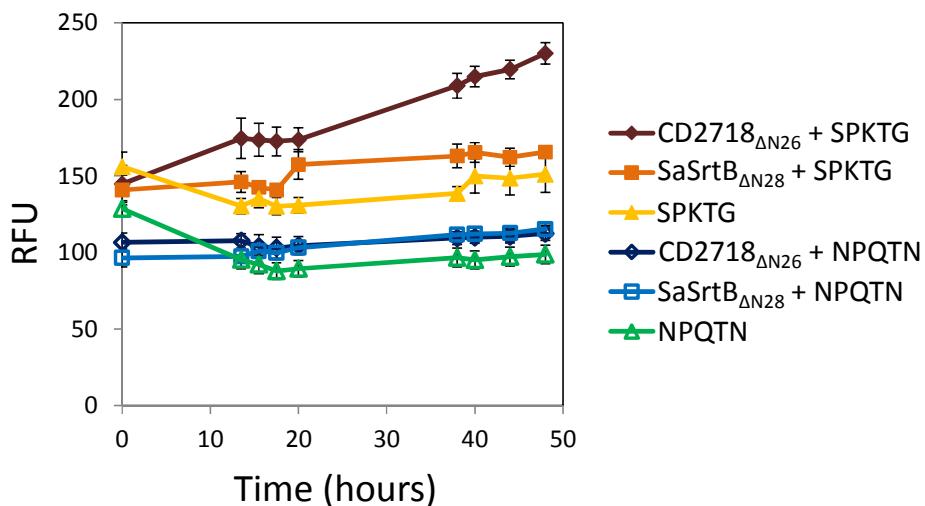


Figure 5.5: Increase in fluorescence over time is dependent on $\text{CD2718}_{\Delta N26}$

Twenty μM $\text{CD2718}_{\Delta N26}$ and SaSrtB were incubated with 4 μM of both the d -SDSPKTGGG-e and d -NPQTN-e peptides. Fluorescence was measured over a period of 48 hours, after which, only the $\text{CD2718}_{\Delta N26}$ incubated with the d -SDSPKTGGG-e peptide demonstrated an increase in fluorescent signal. Error bars represent the standard error of the mean.

Since activity of $\text{CD2718}_{\Delta \text{N}26}$ with the *d*-SDSPKTGGG-e peptide had been demonstrated, we sought to optimise the concentrations of reagents used in the FRET assay going forward. A range of *d*-SDSPKTGGG-e concentrations (2, 4, 8, 10, 20, 40, 80, 160, 200 and 240 μM) were tested with 10 μM $\text{CD2718}_{\Delta \text{N}26}$, and fluorescence measured over a period of time. To control for increased background fluorescence at higher concentrations of *d*-SDSPKTGGG-e, each concentration of peptide was incubated without $\text{CD2718}_{\Delta \text{N}26}$. The increase in fluorescence in the presence of $\text{CD2718}_{\Delta \text{N}26}$ at each peptide concentration was recorded graphically over time (Figure 5.6). This revealed that 10 μM enzyme was sufficient to record an increase in fluorescence after 48 hours, and for *d*-SDSPKTGGG-e concentrations above 40 μM , additional peptide did not increase the amount of cleavage observed. Twenty μM FRET peptide was chosen to carry forward; though the increase in fluorescence was not as high as for 40 μM and above, the signal was detectable and allowed conservation of the amount of peptide used.

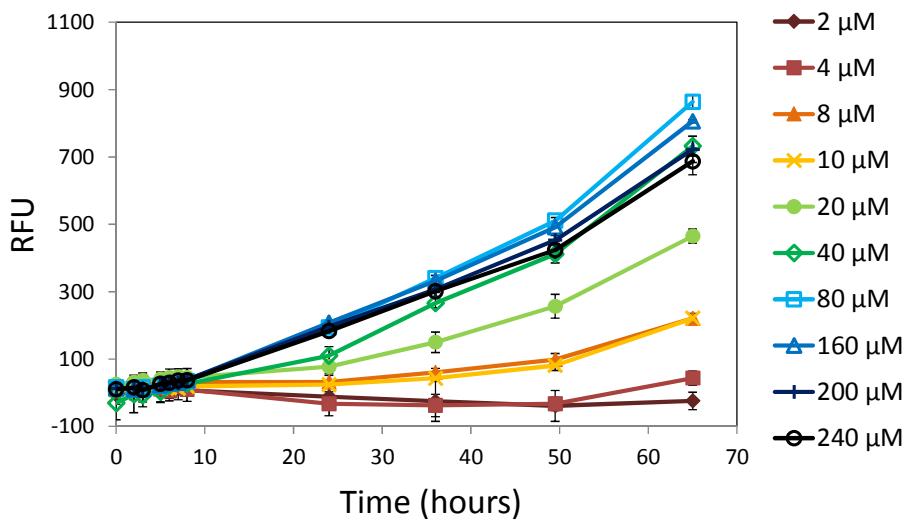


Figure 5.6: Optimising peptide concentration

A range of *d*-SDSPKTGGG-e concentrations (2, 4, 8, 10, 20, 40, 80, 160, 200 and 240 μM) were tested with 10 μM $\text{CD2718}_{\Delta \text{N}26}$, and fluorescence measured over 65 hours. Twenty μM substrate was chosen as a balance between signal intensity and conservation of material.

5.3 Characterisation of CD2718_{ΔN26} activity

5.3.1 CD2718_{ΔN26} cleavage specificity

Using the optimal peptide concentration identified in the previous section, the FRET assay was utilised to examine the substrate specificity of the *C. difficile* sortase. Ten μM purified CD2718_{ΔN26} was incubated with 20 μM FRET peptides containing the (S/P)PXTG motif for 48 hours, during which fluorescence was monitored every hour for the first 8 hours, and at 24, 36, and 48 hours. An increase in fluorescence was observed over time, indicating that cleavage of all four peptides (*d*-SDSPKTGGG-*e*, *d*-PVPPKTGGG-*e*, *d*-IHSPQTGGG-*e*, and *d*-IHSPSTGGG-*e*) occurred in the presence of CD2718_{ΔN26} (Figure 5.7). The greatest cleavage efficiency was observed for the PPKTG and SPQTG containing peptides, while the SPKTG and SPSTG peptides were cleaved to a lesser extent after 48 hours.

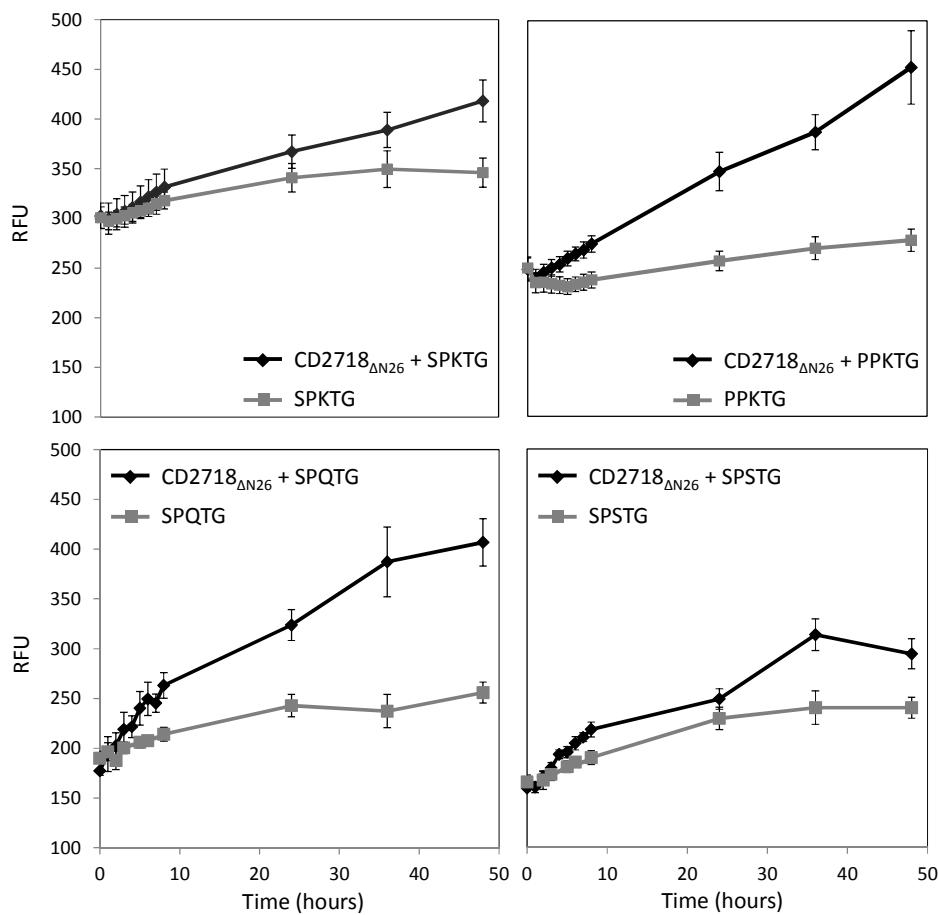


Figure 5.7: Time dependent cleavage of (S/P)PXTG peptides by recombinant CD2718_{ΔN26}

Purified recombinant CD2718_{ΔN26} was incubated with FRET peptides containing the (S/P)PXTG motif and fluorescence measured every hour for the first 8 hours, and also at 24 h, 36 h, and 48 h. An increase in fluorescence

was observed, compared with the peptide incubated alone, indicating that CD2718_{ΔN26}-mediated cleavage of the four peptides occurred over a period of 48 hours. Error bars represent the standard error of the mean.

To quantify the observed cleavage for each peptide, RFU values at 48 hours were standardised by subtracting [(mean RFU_{t=48} of peptide alone)-(standard error of peptide alone)] from each value. The 48 hour timepoint was chosen for this analysis as this was the time at which the level of cleavage observed was significant for all four peptides. These standardised values were graphed side by side as bar graphs to facilitate comparisons despite the differences in background fluorescence of each peptide (Figure 5.8). The bar graphs reflect the varying cleavage efficiencies observed for the different peptides in Figure 5.7. Scrambled peptide sequences GSKTP, GPKTP, GSSTP and GSQTP were tested as controls for the cleavage specificity of the SPKTG, PPKTG, SPQTG, and SPSTG containing peptides. No cleavage of the scrambled peptide sequences was observed, and there was no significant difference observed between the scrambled peptide with and without CD2718_{ΔN26} (Figure 5.8). In addition to the (S/P)PXTG motif, the NVQTG motif from the *C. difficile* collagen binding protein, CbpA, was predicted to be a recognised by CD2718 (Tulli *et al.*, 2013). However, in the FRET assay, CD2718_{ΔN26} failed to cleave the NVQTG peptide (Figure 5.8). The LPETG and NPQTN motifs of SaSrtA and SaSrtB, respectively, were also tested, and were not cleaved by CD2718_{ΔN26} (Figure 5.8).

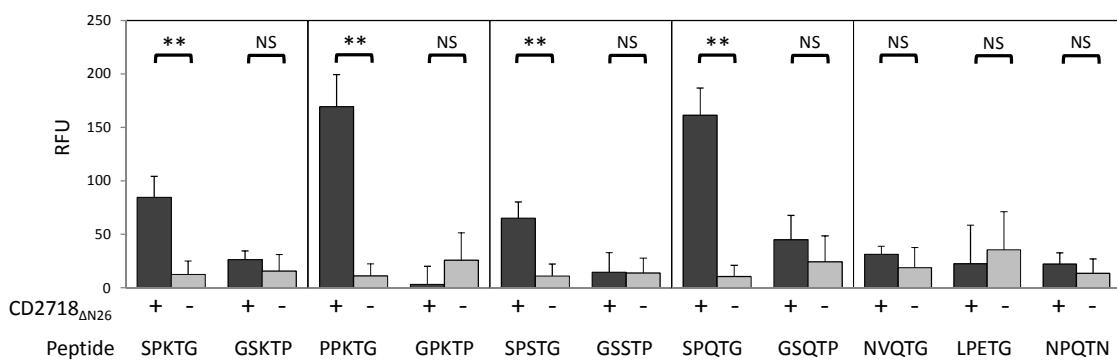


Figure 5.8: CD2718_{ΔN26} substrate specificity

Purified recombinant CD2718_{ΔN26} protein was incubated with a range of peptide sequences to investigate its substrate specificity. The motifs SPKTG, PPKTG, SPSTG and SPQTG were all recognised and cleaved following incubation with CD2718_{ΔN26}. The scrambled peptide sequences GSKTP, GPKTP, GSSTP, and GSQTP serve as controls for the cleavage specificity of CD2718_{ΔN26}. The sequences LPETG and NPQTN, corresponding to the motifs

recognised by *S. aureus* sortase A and B, respectively, do not appear to be substrates for CD2718_{ΔN26}. CD2718_{ΔN26} also failed to cleave the proposed sorting signal NVQTG from newly characterised collagen binding protein, CbpA. Bars indicate the mean, and error bars represent the standard error (** corresponds to $p < 0.01$).

5.3.2 Addition of a nucleophile to the FRET reaction

We sought to test whether the rate of peptide cleavage by CD2718_{ΔN26} could be increased with the addition of a nucleophile. *In vitro*, the *S. aureus* SrtA (SaSrtA) catalyses both peptide hydrolysis and transpeptidation by cleaving the LPXTG motif. In the presence of the nucleophile triglycine ($\text{NH}_2\text{-Gly}_3$), SaSrtA exclusively follows the transpeptidation reaction (Ton-That *et al.*, 2000). The rate of the transpeptidation reaction is at least 2-fold faster than that of the peptide hydrolysis reaction in the absence of the nucleophile (Ton-That *et al.*, 2000). Addition of the strong nucleophile, hydroxylamine (NH_2OH), also increases the rate of peptide cleavage observed by SaSrtA (Ton-That *et al.*, 1999), but not to the same extent as addition of a nucleophile more closely resembling the physiological nucleophile of the transpeptidation reaction, the pentaglycine cross-bridge (Ton-That *et al.*, 2000). This phenomenon is also observed for SrtA in *S. pyogenes* (SpSrtA) – addition of dialanine, the simplest form of cross-bridge between subunits in the peptidoglycan layer of *S. pyogenes* (Navarre & Schneewind, 1999), increases the rate of peptide cleavage and allows transpeptidation to be observed *in vitro* (Race *et al.*, 2009). In *B. anthracis*, peptidoglycan cross-linking occurs directly via *meso*-diaminopimelic acid (*m*-Dap) (Severin *et al.*, 2004), with *m*-Dap serving as the point of covalent attachment of surface proteins by BaSrtA (Budzik *et al.*, 2008). However, the *in vitro* transpeptidation reaction has not been observed for BaSrtA and the addition of *m*-Dap has no effect on the rate of peptide cleavage (Weiner *et al.*, 2010). Peptidoglycan cross-links in *C. difficile* consist of direct connections with 2,6-diaminopimelic acid (2,6-Dap) (Peltier *et al.*, 2011). Addition of NH_2OH or 2,6-Dap in the FRET reaction had no observable effect on the level of CD2718-mediated cleavage (Figure 5.9) indicating that the *in vitro* conditions of this assay may not be optimal for transpeptidation.

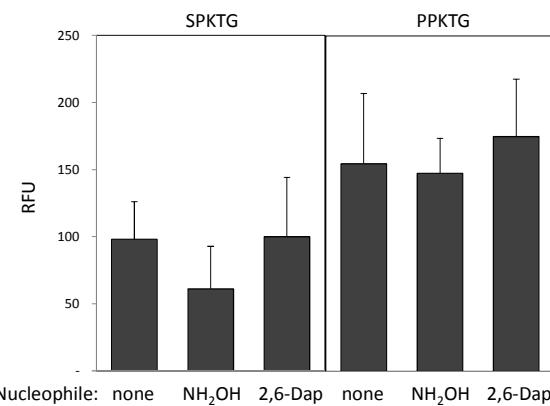


Figure 5.9: Nucleophiles have no effect on CD2718_{ΔN26} mediated cleavage of FRET peptides

The FRET reaction was performed with peptides *d*-SDSPKTGGG-*e* and *d*-PVPPKTGGG-*e* in the presence of either 250 µM hydroxylamine (NH₂OH) or 2,6-diaminopimelic acid (2,6-Dap). Neither had a significant effect on the amount of peptide cleaved over 48 hours.

5.3.3 Cysteine residue is essential for CD2718_{ΔN26} activity

The conserved cysteine residue of sortases is responsible for its catalytic activity (Ton-That *et al.*, 1999, Ton-That *et al.*, 2000, Ton-That *et al.*, 2002, Perry *et al.*, 2002, Ruzin *et al.*, 2002). To determine whether the cleavage of the SPKTG, PPKTG, SPSTG and SPQTG peptides required the only cysteine residue (Cys209) of *C. difficile* CD2718, site-directed mutagenesis was used to replace the cysteine residue at position 209 with an alanine. When the resulting mutant protein CD2718_{ΔN26,C209A} was incubated with the FRET peptides, no fluorescent signal was observed (Figure 5.10).

Another way to test for the importance of the cysteine residue is with the addition of the sulphydryl reagent [2-(trimethylammonium)ethyl]-methanethiosulfonate (MTSET), a cysteine protease inhibitor and known inhibitor of sortases, including SaSrtA and SaSrtB (Ton-That & Schneewind, 1999, Ton-That *et al.*, 2000, Mazmanian *et al.*, 2002). MTSET forms disulfide bonds with sulphhydryl groups, such as the thiol side chain of cysteine (Akabas & Karlin, 1995). Addition of 5 mM MTSET to the FRET reaction abolished cleavage of the SPKTG, PPKTG, SPSTG and SPQTG peptides (Figure 5.10), providing further evidence that CD2718-mediated cleavage requires the cysteine residue at position 209.

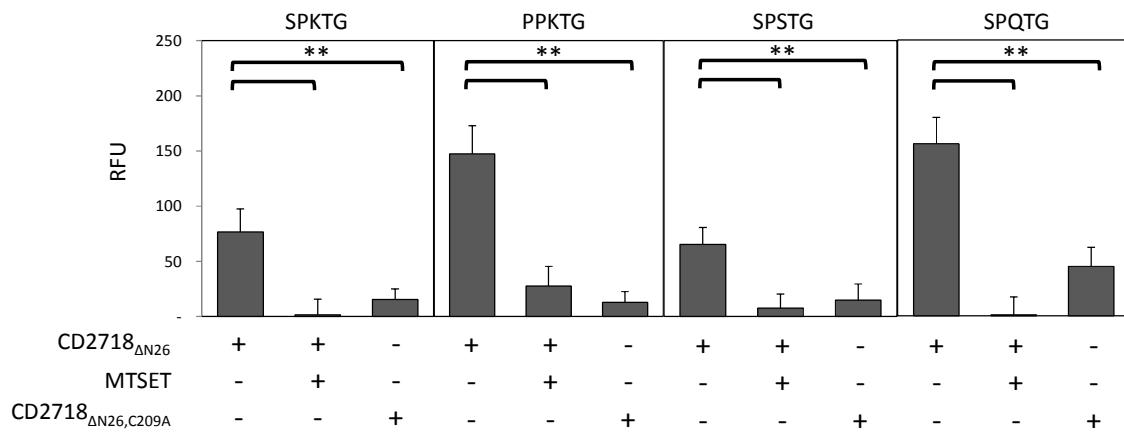


Figure 5.10: CD2718_{ΔN26} activity requires a cysteine residue at position 209

To determine if CD2718_{ΔN26} activity depended on the cysteine residue at position 209, a C209A substitution was made to create CD2718_{ΔN26,C209A}. This enzyme was inactive against the FRET peptides when compared with CD2718_{ΔN26}. Addition at 5 mM of the cysteine protease inhibitor, MTSET, to the reaction also eliminates activity (** corresponds to $p < 0.01$).

5.3.4 Analysis of FRET reaction products

To determine the CD2718_{ΔN26} cleavage site within each peptide, we analysed the reaction products of the FRET reaction by MALDI-TOF mass spectrometry. Sortases recognise and cleave the LPXTG sequence of the sortase substrate between the threonine and glycine residues (Ton-That *et al.*, 2000, Gaspar *et al.*, 2005), and we expected the CD2718-mediated cleavage of the (S/P)PXTG peptides to follow this pattern. Initially, the peptides used in the FRET assay had two glycine residues downstream of the predicted CD2718 recognition sequence because this had previously improved the cleavage efficiency of SaSrtA *in vitro* (Pritz *et al.*, 2007). However, the cleavage of the SPKTG, PPKTG, and SPQTG motifs, but not SPSTG, was enhanced at least two-fold with the addition of the two native amino acids immediately downstream of each sequon (Figure 5.11). Due to the increased rate of cleavage observed for the *d*-SDSPKTGDN-e, *d*-PVPPKTGDS-e, and *d*-IHSPQTGDV-e peptides, these were chosen for the MALDI analysis.

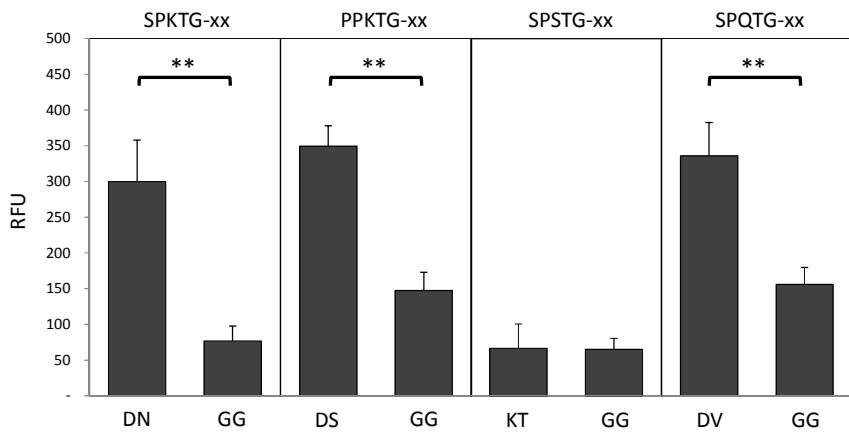
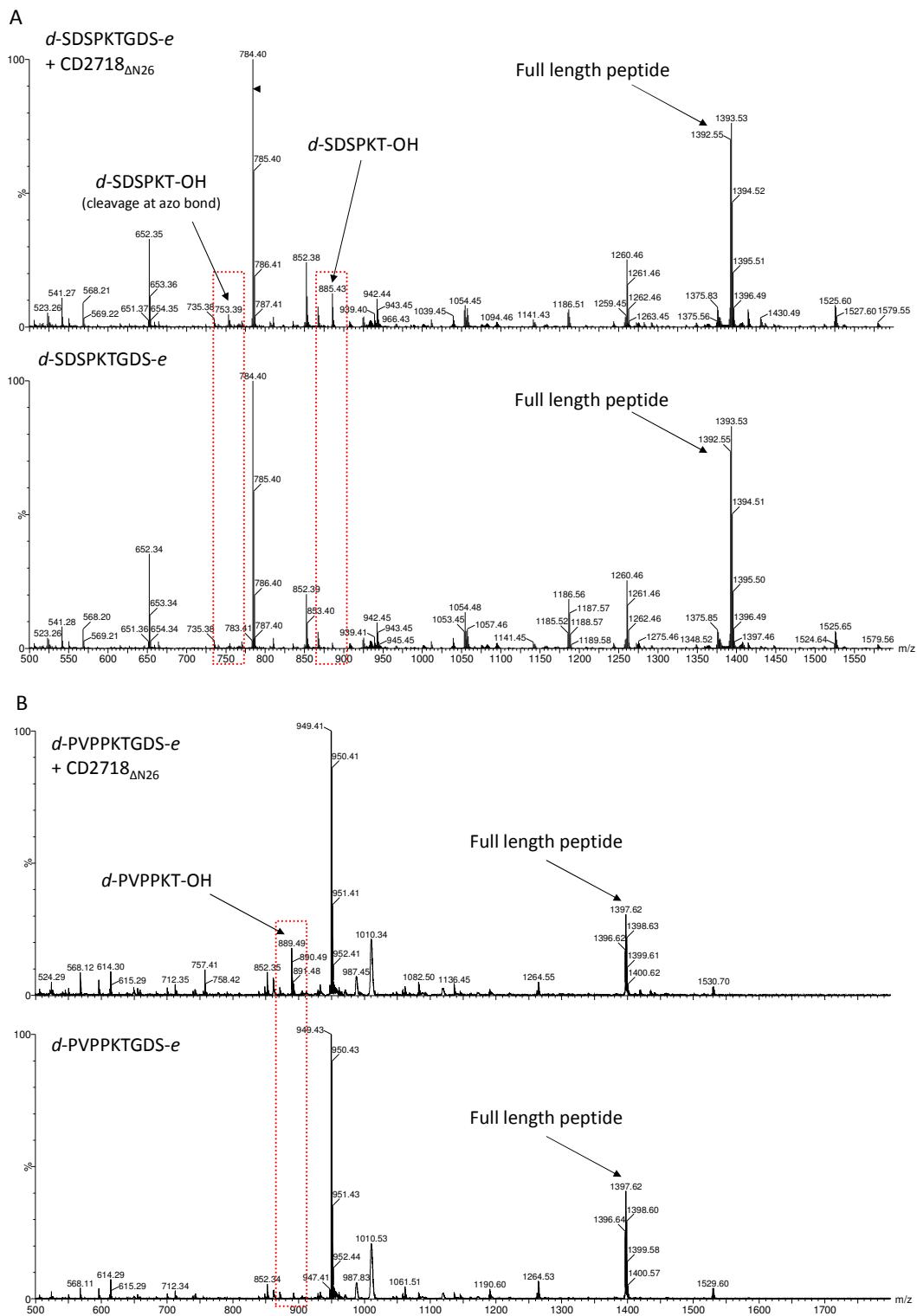


Figure 5.11: CD2718_{ΔN26} cleavage efficiency improves with native peptide sequence

Cleavage of the SPKTG, PPKTG, SPSTG, and SPQTG sequences by CD2718_{ΔN26} was measured using peptides either containing two glycine residues or the two native amino acids immediately following the sequon. For all peptides but SPSTG, the rate of cleavage increased approximately two-fold with the incorporation of the native residues (** corresponds to $p < 0.01$).

Analysis of FRET reaction products after 48 hours incubation revealed that CD2718_{ΔN26} cleaves the *d*-SDSPKTGDN-*e*, *d*-PVPPKTGDS-*e*, and *d*-IHSPQTGDV-*e* peptides between the threonine and glycine residues (Figure 5.12). MALDI analysis of the *d*-SDSPKTGDN-*e* peptide incubated with CD2718_{ΔN26} results in a peptide fragment with a mass of 885 Da (Figure 5.12A, top), which is absent from the peptide incubated alone (Figure 5.12A, bottom). The 885 Da size matches the expected size of the cleavage product *d*-SDSPKT-OH, where the *d*-SDSPKTGDN-*e* peptide has been cleaved between the threonine and glycine residues. Degradation of this cleavage product was also observed: the 753 Da peptide matches the expected size of the degradation of the azo bond within the Dabcyl molecule of *d*-SDSPKT-OH. MALDI analysis of the *d*-PVPPKTGDS-*e* peptide incubated with CD2718_{ΔN26} results in a peptide with a mass of 889 Da (Figure 5.12B, top) corresponding to the fragment *d*-PVPPKT-OH that is not found in the peptide incubated alone (Figure 5.12B, bottom). In addition to the 933 Da fragment corresponding to the *d*-IHSPQT-OH cleavage product, an 801 Da fragment, indicating cleavage of the azo bond within the Dabcyl molecule of *d*-IHSPQT-OH was also observed (Figure 5.12C, top). Neither of these two fragments were found in the SPQTGKT alone sample (Figure 5.12C, bottom). MALDI

analysis of *d*-IHSPSTGKT-*e* cleavage products was not performed due to the low efficiency of the cleavage reaction with this peptide.



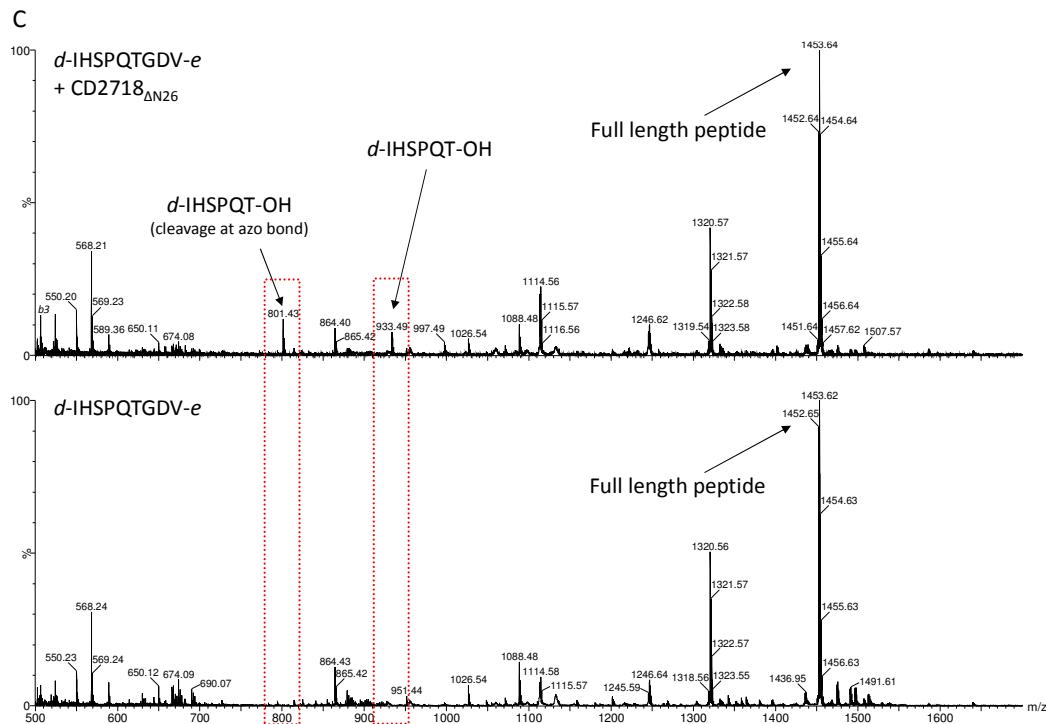


Figure 5.12: CD2718_{ΔN26} cleaves FRET peptides between T and G residues

The *d*-SDSPKTGDN-*e* (**A**), *d*-PVPPKTGDS-*e* (**B**), and *d*-IHSPQTGDV-*e* (**C**) peptides were incubated with CD2718_{ΔN26} for 48 hours, and the resulting reaction products were subjected to MALDI analysis to determine the cleavage site within each peptide. Each peptide incubated for 48 hours without CD2718_{ΔN26} served as controls. Fragments corresponding to the expected cleavage products *d*-SDSPKT-OH, *d*-PVPPKT-OH, and *d*-IHSPQT-OH (boxed in red rectangles) were found in CD2718_{ΔN26} treated samples, but not in the peptide alone sample, indicating that CD2718_{ΔN26} cleaves these three peptides between the threonine and glycine residues. Degradation of the azo bond within the Dabcyl molecule of the *d*-SDSPKT-OH and *d*-IHSPQT-OH cleavage products was also observed.

5.3.5 Kinetic measurements of CD2718_{ΔN26} activity

In order to determine the *in vitro* kinetic parameters of CD2718_{ΔN26} for the SPKTG, PPKTG, SPQTG, and SPSTG motifs, we performed a kinetic analysis of the sortase-catalysed hydrolysis reaction. To correlate fluorescence signal, expressed as relative fluorescence units (RFU), with the concentration of product formed (μM), the standard curves of the fluorophore Edans in the absence and presence of an equal concentration of the quencher Dabcyl were collected (Figure 5.13). The presence of the Dabcyl decreased the fluorescence of the Edans molecule, but this effect was minimal at concentrations below

1 μM . The linear segment of the fluorophore standard curve generated a conversion ratio of 703.9 RFU/ μM Edans (Figure 5.13).

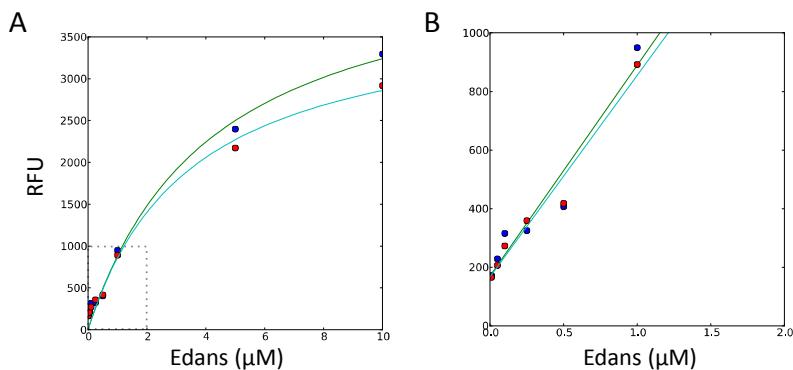


Figure 5.13: Edans fluorophore standard curve

A. A conversion ratio of relative fluorescence units (RFU) to concentration of product formed (μM) was generated by measuring the fluorescence of a range of Edans concentrations (0.01, 0.05, 0.1, 0.5, 1, 5, and 10 μM), both in the presence and absence of an equal concentration of the quencher, Dabcyl. The presence of the quencher did not have an effect on the fluorescence measured at concentrations of Edans below 1 μM . **B.** The slope of a line fitted to the linear segment of the fluorophore standard curve generated a conversion ratio of 703.9 RFU/ μM Edans. Blue = Edans alone, Red = Edans with an equal concentration Dabcyl.

Varying concentrations (8, 10, 20, 40, 80, 160, 200 and 240 μM) of each peptide (*d*-SDSPKTGDN-*e*, *d*-PVPPKTGDS-*e*, *d*-IHSPQTGDV-*e* and *d*-IHSPSTGKT-*e*) were incubated with 10 μM CD2718_{ΔN26} and fluorescence monitored every ten minutes over a 13 hour period (Figure 5.14). The level of cleavage observed for the *d*-IHSPSTGKT-*e* peptide was too low to facilitate kinetic analysis (data not shown). Initial velocities (*V*) were determined from the progress curves and plotted against substrate concentration [*S*]. Figure 5.14A shows the progress curves of the CD2718_{ΔN26}-catalysed hydrolysis reactions at various peptide concentrations of *d*-SDSPKTGDN-*e*, *d*-PVPPKTGDS-*e*, and *d*-IHSPQTGDV-*e*. For each progress curve, the amount of fluorescent product (after conversion from RFU to concentration) was less than 5% of the initial substrate concentration. Within the time period analysed, the progress curves are linear, so the steady state rate (*V*) was determined by fitting the data to a linear function. Figure 5.14B shows *V* plotted against the concentration of the peptide. Nonlinear regression of these data fitted to the modified version of the Michaelis-Menten equation incorporating substrate inhibition resulted in a

K_m of $74.7 \pm 48.2 \mu\text{M}$ for *d*-SDSPKTGDN-*e*, a K_m of $53.3 \pm 25.6 \mu\text{M}$ for *d*-PVPPKTGDS-*e*, and a K_m of $92.2 \pm 76 \mu\text{M}$ for *d*-IHSPQTGDV-*e* (Equation 2.1, Figure 5.14B), values that are not dissimilar to that of other sortases (Table 5.1). With all three peptides, $\text{CD2718}_{\Delta N26}$ appears to be subject to substrate inhibition (Table 5.1) – the reaction turnover rate decreases at the highest substrate concentrations. The phenomenon of substrate inhibition has been observed previously for the *S. pyogenes* SrtA (Race *et al.*, 2009), and is not expected to be physiologically relevant.

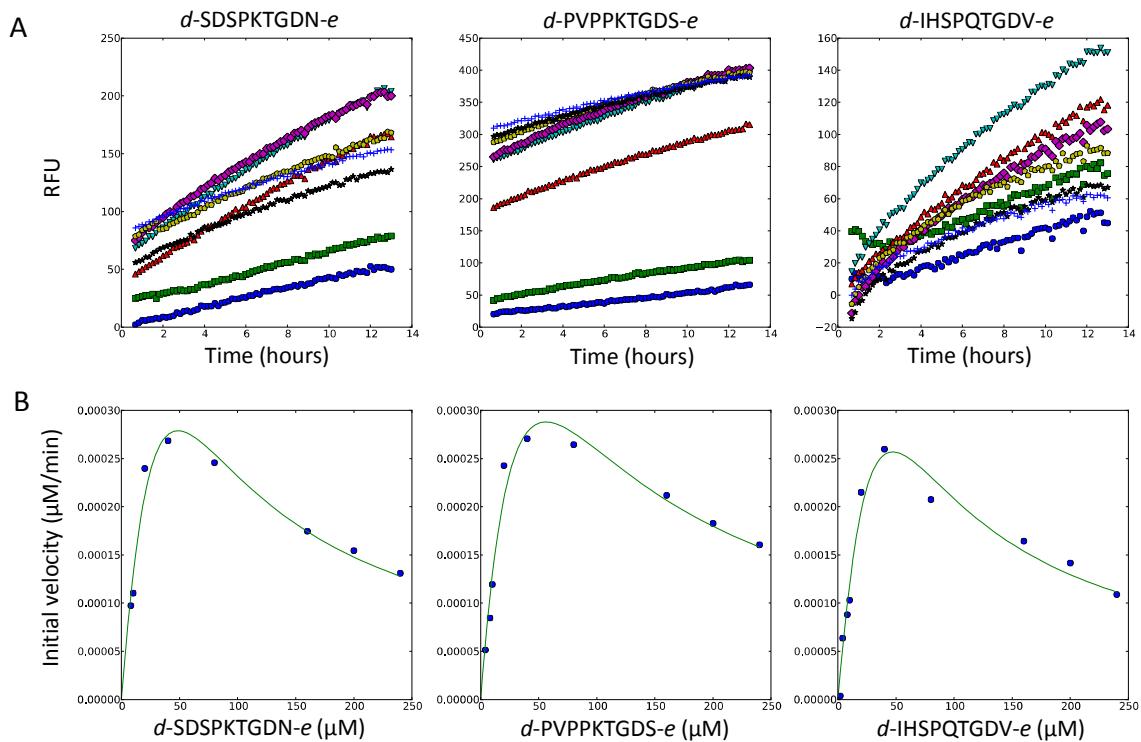


Figure 5.14: Kinetic parameters of $\text{CD2718}_{\Delta N26}$

In order to determine the *in vitro* kinetic parameters of $\text{CD2718}_{\Delta N26}$ for the SPKTGDN, PPKTGDS, and SPQTGDV motifs, we performed a kinetic analysis of the sortase-catalysed hydrolysis reaction. **A.** Progress curves of the $\text{CD2718}_{\Delta N26}$ -catalysed hydrolysis reactions at various concentrations of peptide [8 (●), 10 (■), 20 (▲), 40 (▼), 80 (◆), 160 (◇), 200 (★), and 240 μM (+)]. The steady state rate (V) was determined by fitting the data to a linear function. **B.** Plot of V against the concentration of the peptide [S]. Nonlinear regression of these data fitted to the modified Michaelis-Menten equation resulted in a K_m of $74.7 \pm 48.2 \mu\text{M}$ for *d*-SDSPKTGDN-*e*, a K_m of $53.3 \pm 25.6 \mu\text{M}$ for *d*-PVPPKTGDS-*e*, and a K_m of $92.2 \pm 76 \mu\text{M}$ for *d*-IHSPQTGDV-*e*. $\text{CD2718}_{\Delta N26}$ is subject to substrate inhibition at peptide concentrations $> 24 \mu\text{M}$, which is not expected to be physiologically relevant.

Table 5.1: Kinetic parameters for $\text{CD2718}_{\Delta N26}$ compared with those for other sortases

Sortase	Peptide/Substrate	K_{cat} (min ⁻¹)	K_m (μM)	K_i (substrate, μM)
CD2718 _{ΔN26}	<i>d</i> -PVPPKTGDS- <i>e</i>	$8.3 \times 10^{-4} \pm 3 \times 10^{-4}$	53.3 ± 26	59.3 ± 29
	<i>d</i> -SDSPKTGDN- <i>e</i>	$1.1 \times 10^{-3} \pm 6 \times 10^{-4}$	74.7 ± 48	31.7 ± 20
	<i>d</i> -IHSPQTGDV- <i>e</i>	$1.3 \times 10^{-3} \pm 9 \times 10^{-4}$	92.2 ± 76	24.2 ± 20
BaSrtA _{ΔN26} ^a	<i>abz</i> -LPETG-DNP ^b	$4.0 \times 10^{-4} \pm 1 \times 10^{-5}$	38 ± 4	N/A
SaSrtA _{ΔN26} ^c	<i>abz</i> -LPETG-DNP	$7.3 \times 10^{-3} \pm 6 \times 10^{-4}$	20 ± 4	N/A
SpSrtA _{ΔN81} ^d	<i>abz</i> -LPETGG-DNP	$4.2 \times 10^{-1} \pm 2 \times 10^{-2}$	530 ± 50	16700 ± 2700

^a Weiner *et al.*, 2010

^b Where *abz* is the fluorophore *ortho*-aminobenzoic acid and DNP is the quencher 2,4-dinitrophenyl

^c Huang *et al.*, 2003

^d Race *et al.*, 2009

5.3.6 Inhibiting CD2718_{ΔN26} activity

The published crystal structure of the *Bacillus anthracis* SrtB (Zhang *et al.*, 2004) was used as a template for the selection of potential *C. difficile* CD2718 inhibitors as the predicted structure of CD2718 is highly similar (see Figure 3.1). The proprietary *LeadBuilder* virtual screening method (Domainex Ltd) was used by Domainex, Ltd. to virtually screen a database of 80,000 potential protease inhibitors, *PROTOCATS*, with pharmacophoric and docking filters derived from analysis of the BSrtB crystal structure. These are commercially available compounds which may form reversible transition-state-like complexes with protease enzymes. Compounds in PROTOCATS contain a carbonyl group which is activated to make a fully-reversible complex with the active-site serine/cysteine group by virtue of adjacent moderately electron-withdrawing substituents which are not leaving groups. Examples of these functional groups include α-ketoamides and aryl ketones. Figure 5.15 shows one of the identified compounds docking within the active site structure of BaSrtB.

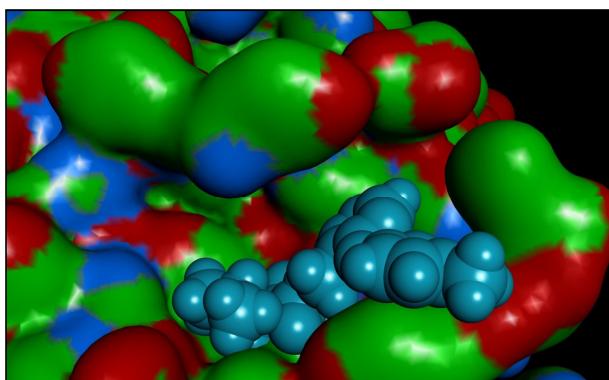


Figure 5.15: Space-filling model of compound interacting with active site of BaSrtB

The proprietary *LeadBuilder* virtual screening method (Domainex Ltd) was used to screen a database of 80,000 potential protease inhibitors,

PROTOCATS, with pharmacophoric and docking filters derived from analysis of the BaSrtB crystal structure (Zhang *et al.*, 2004), which has 70% identity and 90% similarity at the active site with CD2718. This diagram reveals one of the hit compounds fitting into the active site of BaSrtB and interacting with the catalytic cysteine residue.

As described in Section 5.3.3, addition of 5 mM MTSET abolished detectable $CD2718_{\Delta N26}$ activity. In order to identify a suitable concentration range at which to test the hit compounds, we sought to determine the IC_{50} for MTSET. MTSET was tested at a range of concentrations (100, 150, 200, 250, 300, 350, 400, 450, 500, 750, 1000 and 5000 μM) in the FRET assay using the *d*-PVPPKTGDS-*e* peptide (Figure 5.16A). Addition of MTSET reduced $CD2718_{\Delta N26}$ activity to below the limits of detection at concentrations of 500 μM and greater. However, addition of 100 μM MTSET to the reaction had no apparent effect on $CD2718_{\Delta N26}$ (Figure 5.16A). The IC_{50} of MTSET was calculated by nonlinear least squares fit to a four parameter sigmoidal function, and determined to be $286.7 \pm 16.6 \mu M$ (Figure 5.16B).

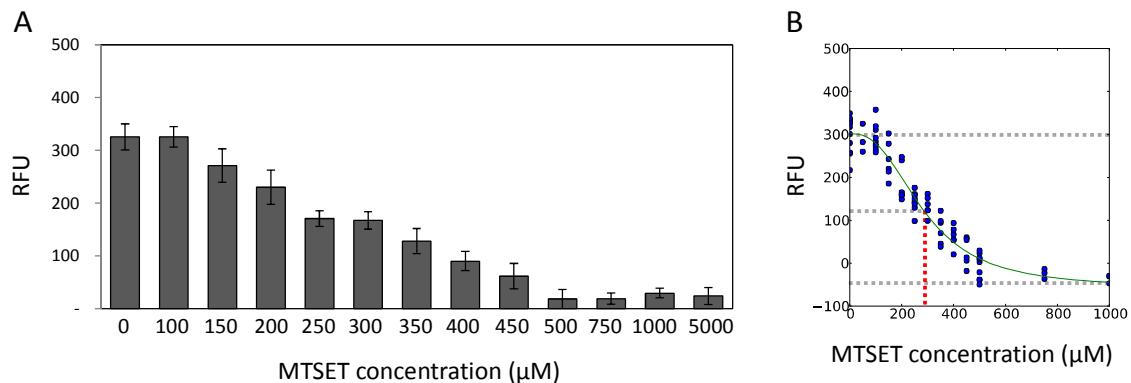


Figure 5.16: Determining the IC_{50} of MTSET

A. The relative activity of $CD2718_{\Delta N26}$ against *d*-PVPPKTGDS-*e* in the presence of a range of MTSET concentrations. **B.** The IC_{50} of MTSET was calculated by fitting a four parameter sigmoidal function to the data in part A. The inflection point of the sigmoidal graph was calculated to be $286.7 \pm 16.6 \mu M$ using SciPy 0.11.0 in Python 2.7.3. Dotted lines have been added to the graph to show the location of the inflection point, and a red dotted line indicates the IC_{50} of 286.7.

Potential CD2718 inhibitors were screened for inhibition of $CD2718_{\Delta N26}$ activity. Sixty-two compounds identified in the *in silico* screen performed by Domainex Ltd were purchased

and solubilised in DMSO. Each compound was tested at 100 µM to allow direct comparisons on their relative inhibitory effect to be made. Due to variations in compound solubility, this was the greatest concentration at which all 62 compounds could be tested in which the final DMSO concentration was at or below 3.75% was tested, which we found to have no effect on CD2718_{ΔN26} activity (Figure 5.17). Where solubility allowed, compounds were also tested at an additional concentration of 200, 250, or 500 µM to see if any observed effect at 100 µM was concentration dependent.

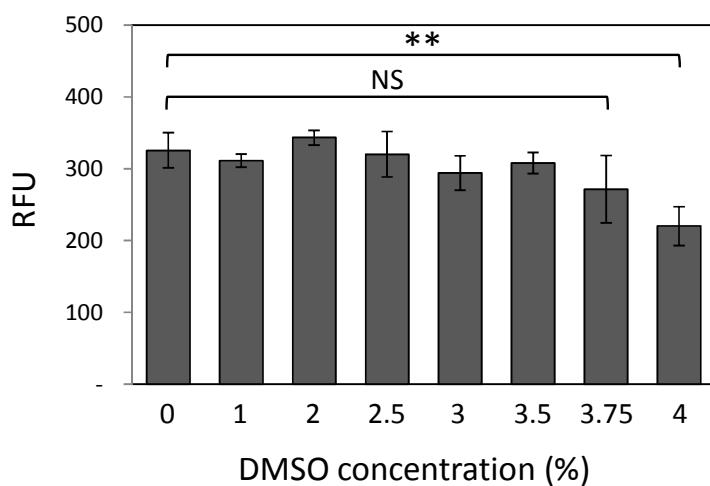


Figure 5.17: Effect of DMSO concentration on CD2718_{ΔN26} activity

Addition of 3.75% had a reduced effect on CD2718_{ΔN26} mediated cleavage of *d*-PVPPKTGDS-*e*, but this was not significant (* corresponds to $p < 0.05$).

A screen of all 62 compounds was performed as follows: inhibitor at each concentration tested was added at the beginning of the FRET reaction containing 10 µM CD2718_{ΔN26} and 20 µM *d*-PVPPKTGDS-*e*. Inhibitor incubated with the peptide alone served as controls for background fluorescence of the compounds, which varied greatly. Relative fluorescence values at 48 hours were standardised as described in Section 5.3.1 to reflect the absolute difference between the wells containing enzyme and the controls. Data points were collected in duplicate, and two independent experiments performed. Inhibitors that caused a reduction in CD2718_{ΔN26} activity of 25-50% at 100 µM were labelled as moderate inhibitors, and those that showed a greater than 50% reduction CD2718_{ΔN26} activity in were labelled as strong inhibitors (Table 5.2). From this screen, the four compounds that showed the greatest levels of inhibition and the least amount of background fluorescence,

LSHTM 13, 40, 50, and 52, were taken forward for further analysis. Compounds that had high background fluorescence, or were sparingly soluble were not chosen to facilitate analysis.

Table 5.2: Percent CD2718_{ΔN26} activity in presence of inhibitors

Inhibitor	Concentration inhibitor (μM)					Inhibitor	Concentration inhibitor (μM)				
	100	200	250	500	Exclude?		100	200	250	500	Exclude?
LSHTM 28	121%	80%			x	LSHTM 31	72%			1%	x
LSHTM 17	107%		72%		x	LSHTM 49	72%		73%		x
LSHTM 9	106%			81%		DMX 9	72%			24%	
LSHTM 5	106%		74%		x	DMX 2	72%			39%	
LSHTM 56	105%			81%	x	DMX 12	71%				x
DMX 26	105%		91%			LSHTM 41	71%		24%	16%	x
DMX 16	104%			80%		LSHTM 51	69%	38%			
LSHTM 47	104%			79%		DMX 17	65%		45%		
DMX 7	103%			89%		LSHTM 11	64%	40%			
LSHTM 12	103%			105%		LSHTM 42	63%	25%			
DMX 28	98%		63%			LSHTM 24	62%	-103%			x
LSHTM 43	97%			86%		LSHTM 30	62%			22%	
LSHTM 16	97%		61%		x	LSHTM 45	62%			34%	
DMX 18	95%			88%		LSHTM 54	61%			66%	
LSHTM 36	95%		40%			LSHTM 29	60%	53%			
LSHTM 7	94%		49%		x	LSHTM 14	60%	6%			
LSHTM 22	93%			110%		LSHTM 18	60%		17%		
LSHTM 21	91%			86%	x	LSHTM 10	58%	61%			
DMX 29	88%		74%			LSHTM 39	58%		55%		
LSHTM 32	88%	81%				LSHTM 46	56%			17%	
LSHTM 35	88%		29%			LSHTM 25	50%				x
DMX 14	85%			49%	x	LSHTM 57	48%	41%			x
DMX 30	83%			40%		LSHTM 40	44%	24%	20%		
LSHTM 19	81%		62%		x	LSHTM 50	37%	8%	6%		
LSHTM 44	78%			42%	x	LSHTM 13	29%	4%			
DMX 1	78%		81%			LSHTM 52	28%	19%			
LSHTM 48	78%			38%		LSHTM 23	19%		126%		x
LSHTM 8	75%			41%	x	LSHTM 53	11%		-74%		x
LSHTM 33	75%		69%			LSHTM 20	6%		-37%		x
LSHTM 38	74%		41%			LSHTM 26	6%		-24%		x
LSHTM 27	74%		74%		x	LSHTM 55	-30%				x

^a Compounds excluded due to high or variable background fluorescence, or limited solubility are marked with an x.

White = compounds that did not significantly reduce sortase activity.

Yellow = compounds with “moderate inhibition,” 25–50% reduction of sortase activity.

Green = compounds with “strong inhibition,” >50% reduction of sortase activity.

In bold are the four inhibitors chosen for further analysis.

These four compounds were then tested at a range of four to five concentrations in order to calculate IC₅₀ values. Data points were collected in triplicate, and the experiment performed in duplicate. The data was fit to a four parameter sigmoidal curve as described for MTSET, and the inflection point calculated as the IC₅₀ values (Figure 5.18). All four compounds, LSHTM 13, 40, 50, and 52, had IC₅₀ values below 100 μM, at 62.4±24.6 μM, 63.1±8.8 μM, 60.1±4.7 μM, and 44.1±6.9 μM, respectively (Figure 5.18). This suggests that they are more potent inhibitors of CD2718_{ΔN26} than MTSET, which had no apparent effect on CD2718_{ΔN26} activity for concentrations at or below 100 μM (Figure 5.16).

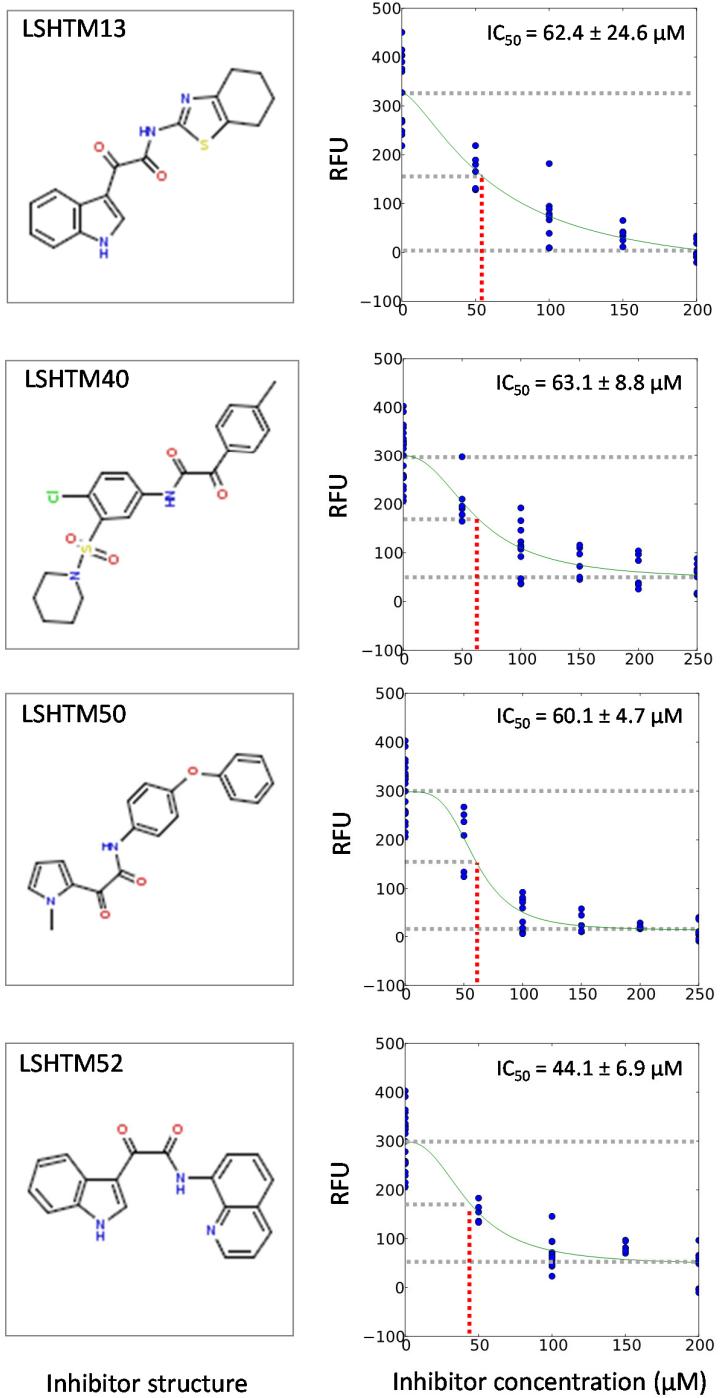


Figure 5.18: Structures and IC_{50} values of $CD2718_{\Delta N26}$ inhibitors

Four hits from the virtual screen were tested in the FRET assay at varying concentrations to screen for inhibition of $CD2718_{\Delta N26}$ mediated cleavage of *d*-PVPPKTGDS-e. The IC_{50} of each compound was calculated as the inflection point of a four parameter sigmoidal function fit to the data. Dotted lines have been added to the graph to show the location of the inflection point, and a red dotted line indicates the IC_{50} . The structures of each compound are depicted to the left of the corresponding data plot.

5.3.7 Effect of CD2718 inhibitors on *C. difficile* growth

The four hit compounds identified with inhibitory activity against CD2718 were assessed for their effect on *C. difficile* growth. The highest concentration of inhibitor such that the final DMSO concentration was at or below 3.75%, as this concentration of DMSO was found to have no effect on *C. difficile* growth (Figure 5.19). Differences in solubility between the compounds meant that LSHTM13 and LSHTM 52 were tested at 200 µM, and LSHTM40 and LSHTM50 were tested at 250 µM. Colony forming units per ml of culture (CFU/ml) were calculated to check for sub-inhibitory effects on growth. However, there were no differences in *C. difficile* growth in the presence of any of the four inhibitors for either strain tested, 630Δerm or R20191 (Figure 5.19). Similarly, 500 µM MTSET had no effect on *C. difficile* growth.

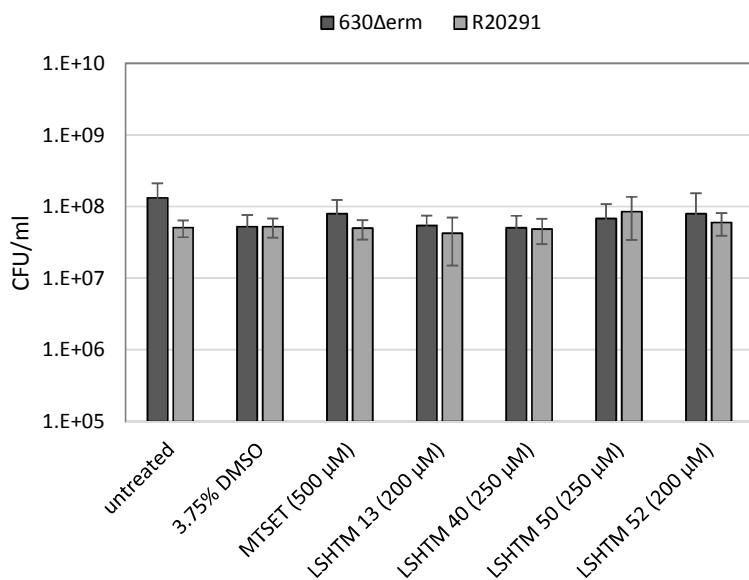


Figure 5.19: CD2718 inhibitors have no effect on *C. difficile* growth

Colony forming unit (CFU) counts from *C. difficile* 630Δerm or R20191 cultures grown overnight in BHIS broth with either 500 µM MTSET, 200 µM LSHTM13, 250 µM LSHTM40, 250 µM LSHTM50, or 200 µM LSHTM52 were compared to those from untreated cultures. No differences in CFU per ml were observed between treated and untreated cultures in either strain. 3.75% DMSO served as a control for the presence of DMSO in the four inhibitors.

5.4 Discussion

This chapter is the first report that demonstrates the putative *C. difficile* sortase B, CD2718, exhibits sortase-like activity. Recombinant *C. difficile* CD2718_{ΔN26} recognises and cleaves an (S/P)PXTG motif between the threonine and glycine residues, and this cleavage is dependent on the single cysteine residue at position 209. This chapter also reports kinetic data for CD2718_{ΔN26}, and demonstrates inhibition of CD2718_{ΔN26} by rationally designed small-molecule inhibitors.

In this chapter, we successfully developed a working FRET assay to analyse CD2718 activity. Chapter 4 described the expression and purification of three different recombinant CD2718 proteins: a full length CD2718, and two N-terminally truncated CD2718 proteins lacking the transmembrane domain (CD2718_{ΔN26} and CD2718_{ΔN26}). These proteins were initially tested for activity against an SPKTG-containing peptide in the presence and absence of DTT. We found that the CD2718_{ΔN26} protein was only active in the presence of DTT, a reducing agent that is often used to stabilise enzymes and prevent oxidation of cysteines (Cleland, 1964), and has been shown to restore activity of sortase that has been inactivated by inhibitors (Ton-That *et al.*, 1999). The cleavage of SPKTG, PPKTG, SPQTG, and SPSTG containing peptides by the recombinant CD2718_{ΔN26} validated the prediction of sortase substrates performed in Chapter 3. This study revealed that CD2718 cleaves these motifs with varying levels of efficiency, cleaving the sequences PPKTG and SPQTG with the greatest efficiency, and SPKTG and SPSTG to a lesser degree. Apparent preferential cleavage efficiency of certain substrate sequences over others has been observed in other sortases. For example, in *B. anthracis*, BaSrtA cleaves LPXTG peptides more readily than a peptide containing the sequence LPNTA (Gaspar *et al.*, 2005). It is unknown what biological role this preference may serve, perhaps in regulating the amount of each protein anchored to the cell wall. Though the SPKTG, PPKTG, SPQTG, and SPSTG sequences have been recognised and cleaved by recombinant CD2718, it still remains to be seen whether the seven proteins containing these sequences identified in Chapter 3 are covalently anchored to the *C. difficile* cell surface.

In this assay, CD2718_{ΔN26} did not cleave the NVQTG motif found in the recently characterised collagen binding protein, CbpA (Tulli *et al.*, 2013). This suggests that the surface localised CbpA may not be a substrate of CD2718, and may be attached to the *C.*

difficile cell surface by another means. Besides the covalent attachment of surface proteins to peptidoglycan by sortases, there exist other major mechanisms of surface protein attachment in Gram-positive organisms. Some bacterial cell surface proteins bind to the cell wall through non-covalent interactions. For example, the surface associated pneumococcal autolysin LytA and the *L. monocytogenes* invasin InlB bind to choline-substituted teichoic acids (TA), and lipoteichoic acids (LTA), respectively (Jonquieres *et al.*, 1999, Fernandez-Tornero *et al.*, 2001). Other Gram-positive surface proteins are anchored to the bacterial membrane by a C-terminal membrane spanning region followed by a positively charged tail thought to serve as a stop-transfer signal, including the *L. monocytogenes* actin assembly inducing protein, ActA (Kocks *et al.*, 1992). The C-terminal membrane anchor of proteins like ActA very closely resembles the C-terminus of sortase substrates and that of CbpA.

The kinetic data reported here for CD2718_{ΔN26} suggests that the CD2718 hydrolysis reaction observed is inefficient and slow. Low catalytic efficiency is observed for most sortases *in vitro*, and is attributed to the fluorogenic peptides used being poor mimics of *in vivo* substrates, and to a small proportion of the enzyme being in a catalytically competent state (Frankel *et al.*, 2005, Race *et al.*, 2009). We found that cleavage efficiency of CD2718_{ΔN26} increased approximately two-fold when the peptide sequence more closely matched the sequence of the predicted protein substrate. *In vivo* the sorting motifs are part of a larger protein, and the transpeptidation substrates are part of a cell wall precursor or mature peptidoglycan (Perry *et al.*, 2002, Ruzin *et al.*, 2002, Marraffini & Schneewind, 2005). The kinetic and cleavage data reported here for CD2718_{ΔN26} follows this trend. Addition of the predicted transpeptidation substrate based on the peptidoglycan structure of *C. difficile*, 2,6-diaminopimelic acid (Peltier *et al.*, 2011), had no observable effect on the reaction rate, indicating that the *in vitro* conditions of this assay may not be optimal for transpeptidation. The transpeptidation reaction has been observed in FRET assays for sortases from bacteria with a Lys-type peptidoglycan, where cross-linking occurs through a peptide bridge (Schleifer & Kandler, 1972, Dziarski, 2004) such as *S. aureus* and *Streptococcus* species (Ton-That *et al.*, 2000, Race *et al.*, 2009, Necchi *et al.*, 2011), but not for bacteria with Dap-type peptidoglycan such as *Bacillus* with direct cross-linkages through *m*-diaminopimelic acid (Weiner *et al.*, 2010). When

transpeptidation is observed *in vitro*, the cleavage efficiency of sortase increases, which may contribute to the low efficiency observed for CD2718_{ΔN26}.

Small-molecule inhibitors with activity against SrtA and SrtB have been reported that prevent cleavage of fluorescently labelled peptide compounds by sortase *in vitro*, and also inhibit cell adhesion to fibronectin (Oh *et al.*, 2006). Inhibitors tested against SrtA, SrtB and SrtC in *B. anthracis* irreversibly modified the active cysteine residue and possessed antimicrobial activity against Gram-positive, but not Gram-negative bacteria (Marešo *et al.*, 2007). When mice challenged with *S. aureus* were treated with sortase inhibitor compounds, infection rates and mortality were reduced, supporting their suitability for further development as therapeutics (Oh *et al.*, 2010). Due to the link between disruption of the intestinal microflora by antibiotics and *C. difficile* infection (CDI), and the complications associated with recurrent infection, developing a *C. difficile* selective antimicrobial is a current clinical imperative. Inhibiting the *C. difficile* sortase could be a strategy to target *C. difficile* but limit the effect on the mostly Gram-negative gut population lacking sortases. The use of *in silico* models such as the LeadBuilder method employed by this study to screen databases of small-molecule inhibitors for further analysis has been validated, as 38 of the 62 compounds identified had at least a moderate inhibitory effect on CD2718_{ΔN26} activity. Further analysis of the structural similarities between these 38 compounds could lead to a refinement of inhibitor design, and increased potency *in vitro*.

6 Analysis of putative sortase substrates in *C. difficile*

6.1 Introduction

Sortase substrates are often adhesins that bind to extracellular matrix binding proteins or pili components and frequently contribute toward colonisation and pathogenesis. Virulence defects in sortase substrate mutants have been well described (Marraffini *et al.*, 2006, Bierne & Cossart, 2007, Nobbs *et al.*, 2009). Inactivating sortase substrates can also have an effect on *in vitro* characteristics, as mutants in *S. aureus* sortase substrates *FnbpA* (fibronectin binding protein A), *FnbpB* (fibronectin binding protein B), *ClfA* (clumping factor A), and *ClfB* (clumping factor B) exhibit an altered colony morphology and cell to cell interactions (Tsompanidou *et al.*, 2012). In Chapter 3, our bioinformatics analysis identified eight predicted sortase substrates in *C. difficile* strain 630; CD0183, CD0386, CD2537, CD2768, CD2831, CD3145 (*CpbA*), CD3246, and CD3392 (Table 3.3). These putative proteins met our criteria for a sortase substrate: an N-terminal signal peptide, and a C-terminal CWSS comprising the predicted sortase recognition sequence (S/P)PXTG or NVQTG, followed by a hydrophobic domain and positively charged tail. Seven of the eight predicted substrates contain an (S/P)PXTG motif, while only *CpbA* contains the NVQTG motif. The *in vitro* FRET assay developed in Chapter 5 demonstrated that the predicted *C. difficile* sortase CD2718 recognises and cleaves peptides containing (S/P)PXTG motifs, but did not cleave the NVQTG motif. This suggests that the *CpbA* protein may not be a substrate of CD2718, and attached to the cell surface by another mechanism. The other seven predicted sortase substrates proteins; which encode putative adhesins, cell wall hydrolases, and a 5' nucleotidase; have not yet been characterised and their role in colonisation and infection remains unknown.

6.1.1 Aims

This chapter describes the preliminary *in vitro* characterisation of some of the predicted *C. difficile* sortase substrates identified through bioinformatic analysis in Chapter 3.

The aims of this study were to:

- Analyse *in vitro* transcription of predicted sortase substrates
- Construct and characterise inactivation mutants of *CD0386*, *CD2831*, and *CD3392*

6.2 Transcriptional analysis of sortase substrates

We sought to characterise these proteins to better understand the function of the seven predicted sortase substrates in *C. difficile*. Therefore, initial studies were undertaken to confirm that the respective genes were transcribed under standard laboratory conditions. As described in Chapter 4, RNA was extracted from 10 ml 630Δerm cultures in BHIS media at three time points: early exponential phase ($t=3$ hours, $OD_{600} = 0.4\text{-}0.5$), late exponential phase ($t=5$ hours, $OD_{600} = 0.8\text{-}0.9$), and stationary phase ($t=24$ hours, $OD_{600} = 0.6\text{-}0.8$) (Figure 4.2). RNA was treated with DNase to remove contaminating genomic DNA prior to cDNA synthesis. To assess cDNA sample quality and depletion of genomic DNA, a PCR screen was performed with primers amplifying the *16s rRNA* and the housekeeping gene *atpA* (Figure 6.1). Bands were successfully amplified from the cDNA with both primer sets, and there was no amplification from the reverse transcriptase negative samples, indicating there was no contaminating genomic DNA in the samples.

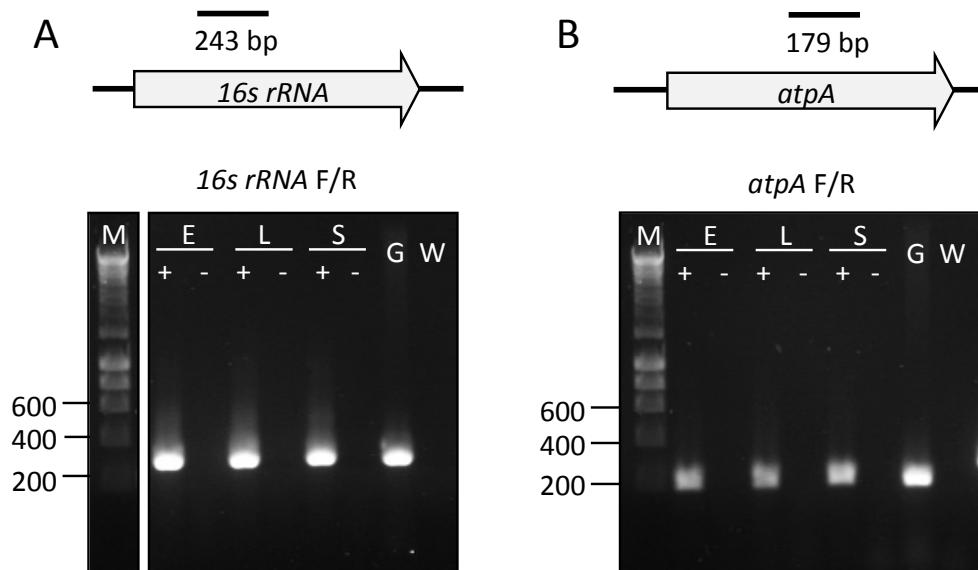


Figure 6.1: Assessment of cDNA quality

PCR screens were performed to amplify the *16s rRNA* (A) and the housekeeping gene *atpA* (B) to confirm cDNA synthesis, and DNA depletion in the RNA samples. **Top:** Schematic diagrams representing the location of the primer pairs used to detect mRNA transcripts, and their expected sizes. **Bottom:** PCR reactions were performed with wild-type 630Δerm cDNA that was prepared from cultures grown to early exponential (E), late exponential (L), and stationary phase (S). M = Hyperladder I (Bioline), G = 630Δerm genomic DNA, W = dH₂O. A "+" indicates cDNA reaction with added reverse transcriptase, "-" indicates cDNA reaction without added reverse transcriptase (control for DNA depletion of RNA sample).

All seven predicted sortase substrate coding sequences in *C. difficile* 630 appear to be transcribed *in vitro* (Figure 6.2-Figure 6.4). *CD0183*, *CD3246*, and *CD2768* are highly transcribed at all three time points (Figure 6.2A, Figure 6.2D, Figure 6.4); *CD2537* is transcribed during early and late exponential phase, but not during stationary phase (Figure 6.2C); *CD2831* is moderately transcribed during all three growth phases, but transcribed at the highest level during early exponential phase (Figure 6.2B); and *CD0386* and *CD3392* are transcribed at very low levels throughout the time course (Figure 6.3). Analysis of the genome localisation of the seven predicted genes revealed only three with in-frame coding sequences either immediately up- or downstream: *CD0386*, *CD2768*, and *CD3392* (Figure 6.3, Figure 6.4). To ascertain whether these neighboring coding sequences were co-transcribed with the predicted sortase substrate genes, PCR screens were designed to amplify products spanning the adjacent ORFs (Figure 6.3, Figure 6.4). Both putative adhesins, *CD0386* and *CD3392*, have small predicted ORFs (*CD0385A* and *CD3391A*, 186 bp and 195 bp in length, respectively) just 12 bp downstream of their stop codons. The putative cell wall hydrolase *CD2768* has a large 1911 bp predicted ORF (*CD2769*) 198 bp upstream of its start codon. *CD0385A* and *CD3391A* are annotated as hypothetical transposon ORFs, while *CD2769* encodes a predicted polysaccharide biosynthesis protein. The RT-PCR analysis revealed a band amplified from each of the three oligonucleotide pairs (Figure 6.3, Figure 6.4). The intergenic regions between *CD0386* and *CD0385A*, and between *CD3392* and *CD3391A*, are transcribed at a much greater level than the corresponding sortase substrate gene (Figure 6.3). Due to the relative difference in their transcription levels observed in this semi-quantitative RT-PCR, it is unlikely these co-localised gene pairs are co-transcribed. The intergenic region between *CD2768* and *CD2769* is transcribed to a similar level as *CD2768* (Figure 6.4), suggesting both genes may possibly form a single transcriptional unit.

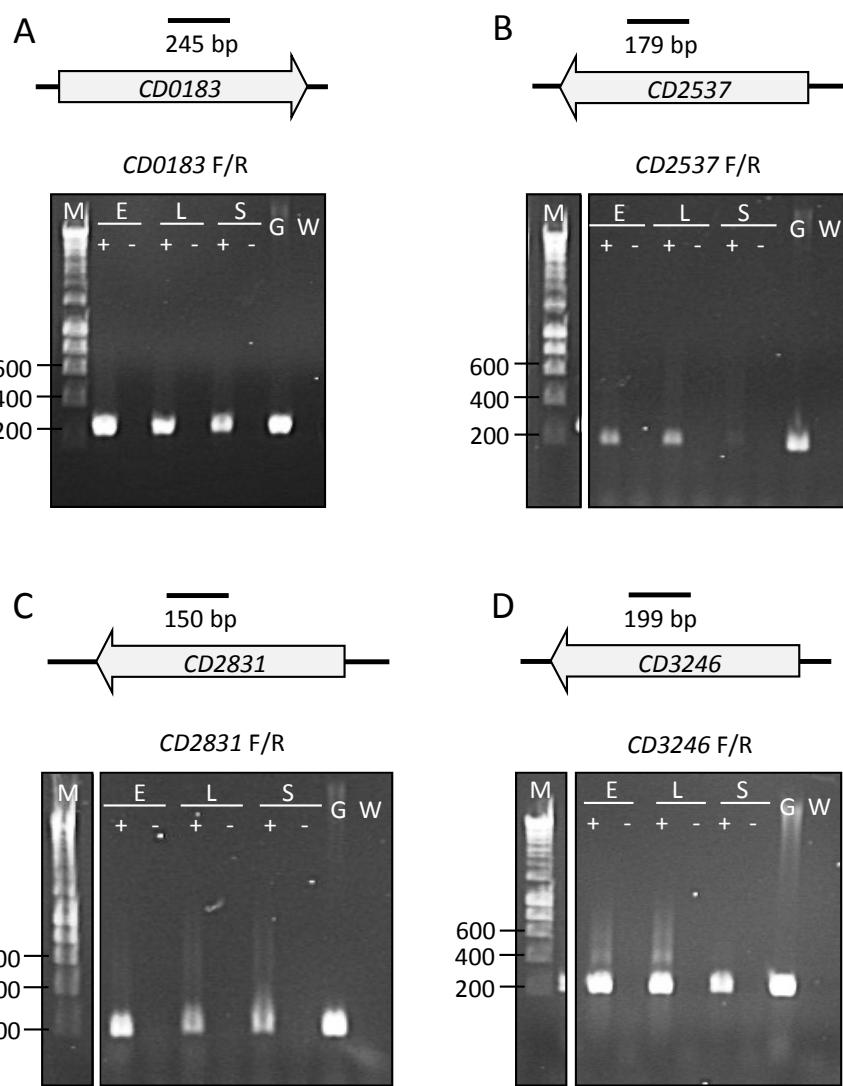


Figure 6.2: Transcriptional analysis of *CD0183*, *CD2537*, *CD2831*, and *CD3246*

Top: Schematic diagrams representing the location of the primer pairs used to detect mRNA transcripts, and their expected sizes for *CD0183* (A), *CD2537* (B), *CD2831* (C), and *CD3246* (D). **Bottom:** PCR screens were performed using wild-type 630Δerm cDNA that was prepared from cultures grown to early exponential (E), late exponential (L), and stationary phase (S). M = Hyperladder I (Bioline), G = 630Δerm genomic DNA, W = dH₂O. A "+" indicates cDNA reaction with added reverse transcriptase, "-" indicates cDNA reaction without added reverse transcriptase (control for DNA depletion of RNA sample).

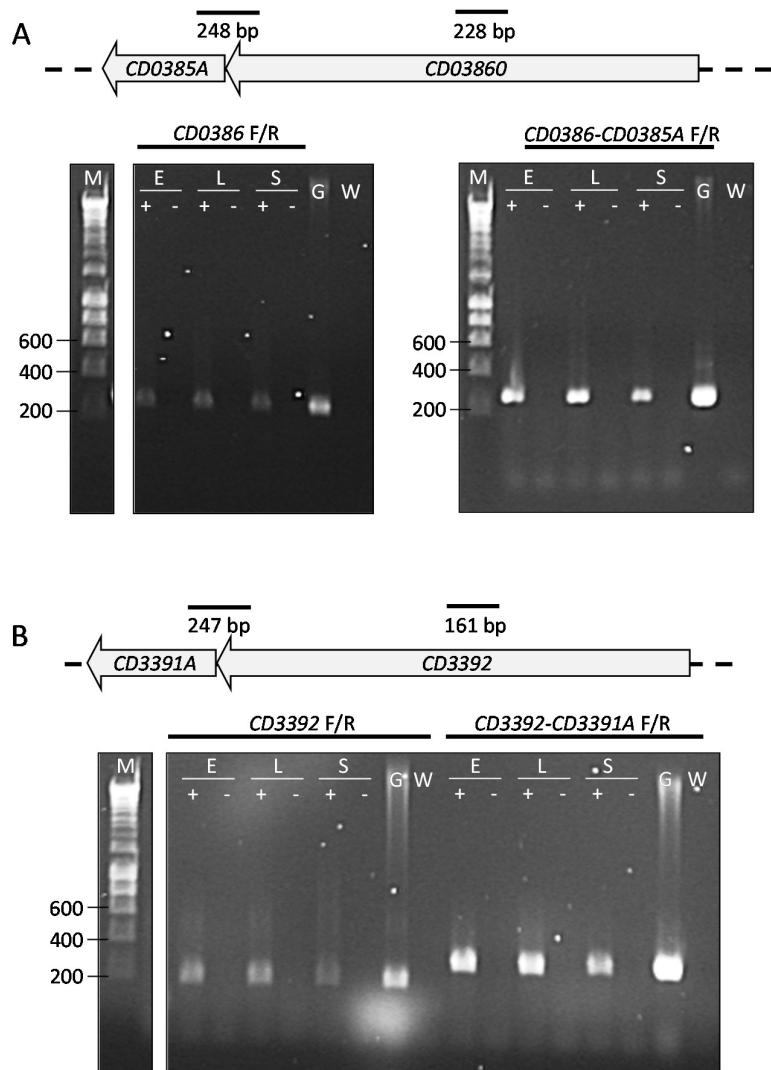


Figure 6.3: Transcriptional analysis of *CD0386* and *CD3392*

Top: Schematic diagrams representing the location of the primer pairs used to detect mRNA transcripts, and their expected sizes for *CD0386* and the intergenic region with the short downstream ORF *CD0385A* (A), and *CD3392* and the intergenic region with the short downstream ORF *CD3391A* (B).

Bottom: PCR screens were performed using wild-type 630Δerm cDNA that was prepared from cultures grown to early exponential (E), late exponential (L), and stationary phase (S). M = Hyperladder I (Bioline), G = 630Δerm genomic DNA, W = dH₂O. A “+” indicates cDNA reaction with added reverse transcriptase, “-” indicates cDNA reaction without added reverse transcriptase (control for DNA depletion of RNA sample).

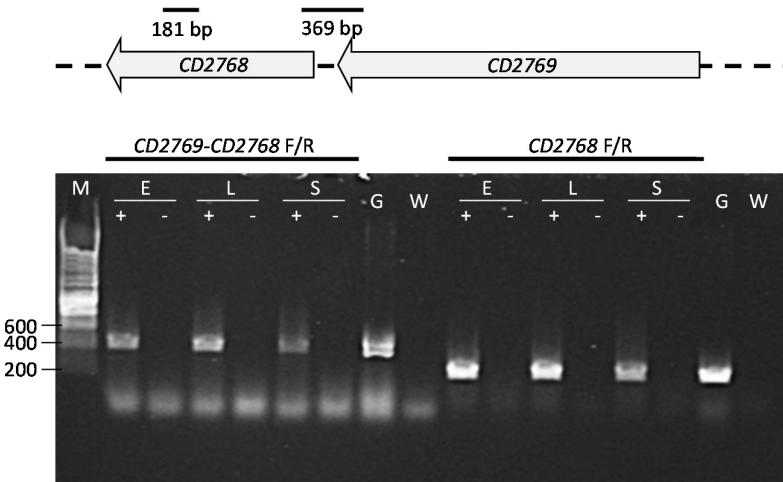


Figure 6.4: Transcriptional analysis of *CD2768*

Top: Schematic diagrams representing the location of the primer pairs used to detect mRNA transcripts, and their expected sizes for *CD2768* and the intergenic region with the upstream ORF *CD2769*. **Bottom:** PCR screens were performed using wild-type 630 Δ erm cDNA that was prepared from cultures grown to early exponential (E), late exponential (L), and stationary phase (S). M = Hyperladder I (Bioline), G = 630 Δ erm genomic DNA, W = dH₂O. A “+” indicates cDNA reaction with added reverse transcriptase, “-” indicates cDNA reaction without added reverse transcriptase (control for DNA depletion of RNA sample).

6.3 Construction of sortase substrate mutants in 630 Δ erm

In order to investigate the potential functional role(s) the sortase substrates, we sought to construct inactivation mutants using the ClosTron mutagenesis system (Heap *et al.*, 2007, Heap *et al.*, 2010). Three of the predicted substrates were chosen for initial mutagenesis studies: the putative collagen-binding proteins CD0386, CD2831, and CD3392. The bioinformatic analyses performed in Chapter 3 revealed that the three proteins CD0386, CD2831, and CD3392 are annotated as putative collagen-binding proteins because they contain domains from the *S. aureus* collagen-binding protein Cna. CD2831 retains both the collagen-binding A region of Cna and a single B-repeat region that may serve as a stalk for displaying the functional A region. The highly similar CD0386 and CD3392 proteins both contain three B-repeat regions, and we predict they may facilitate adherence to the host intestinal epithelium.

To construct ClosTron mutants in these three genes, the mobile group II intron, Li.ItrB, on plasmid pMTL007C-E2 was re-targeted to insert in the sense orientation between nucleotide positions 32 and 33 for *CD0386*, 540 and 541 for *CD2831*, and 840 and 841 for *CD3392*. Mutants were generated in background strain 630Δ erm (Hussain *et al.*, 2005). Potential mutants were identified as lincomycin resistant but thiamphenicol sensitive colonies, indicating chromosomal integration of the intron and loss of the pMTL007C-E2 plasmid, respectively. DNA boilates prepared from potential mutants were screened by PCRs using a gene-specific and the intron-specific EBS Universal primer to confirm insertion of the intron into the target gene, and primers amplifying the retrotransposition-activated marker (RAM) (Figure 6.5-Figure 6.7). Insertion of the Li.ItrB intron into the target genes was verified, and a 900 bp fragment was amplified with the RAM primers from all colonies screened, confirming splicing of the group I intron (*td*) from the RAM during insertion and loss of the 1265 bp plasmid based copy of the RAM (Figure 6.5-Figure 6.7).

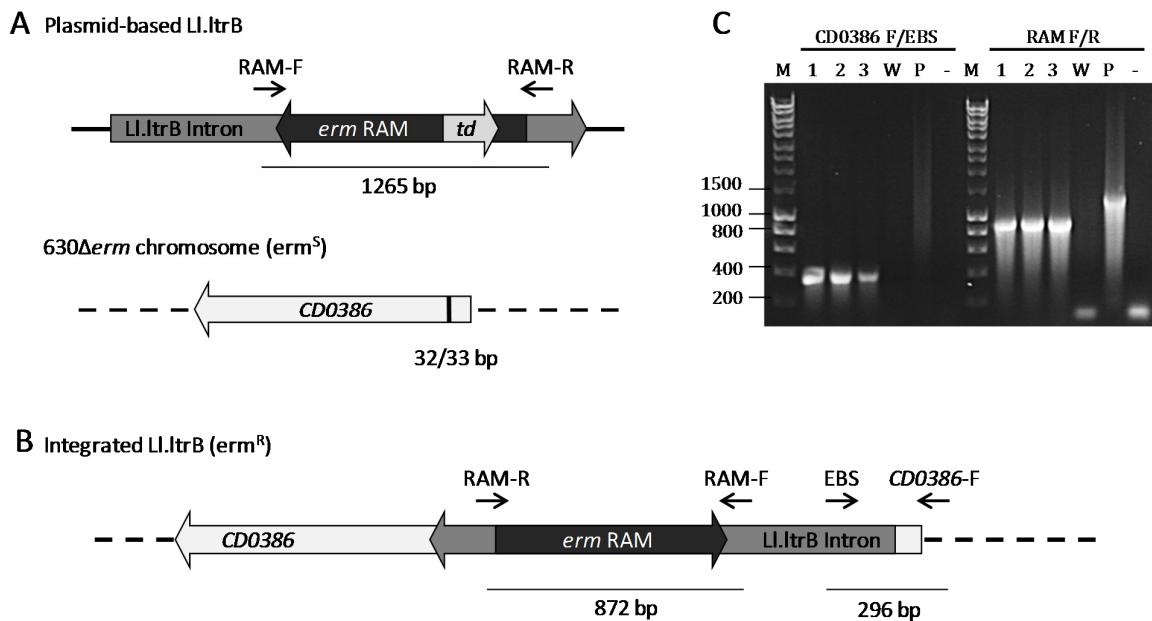


Figure 6.5: PCR screen of constructed *CD0386* mutants

A/B. Schematic diagram of the insertional inactivation of *CD0386*. The location of the primers used to screen the potential mutant clones and the expected product sizes are indicated. **C.** Three potential *CD0386* mutant clones were verified by PCR. The primer pair used for each PCR is indicated above the corresponding lanes. **Left:** Amplification across the gene-intron junctions of each mutant using a gene-specific (*CD0386*-F) and intron-

specific primer (EBS) confirmed insertion of the Li.ItrB intron into *CD0386*. **Right:** Amplification of the retrotransposition-activated marker (RAM) from each mutant confirmed chromosomal integration of RAM (splicing of *td* group I intron) and loss of pMTL007C-E2 plasmid. M: HyperLadder I (Bioline – sizes indicated in base pairs), Lanes 1-3: three mutant clones, W: wildtype 630Δerm, P: plasmid pMTL007C-E2, “–” : dH₂O

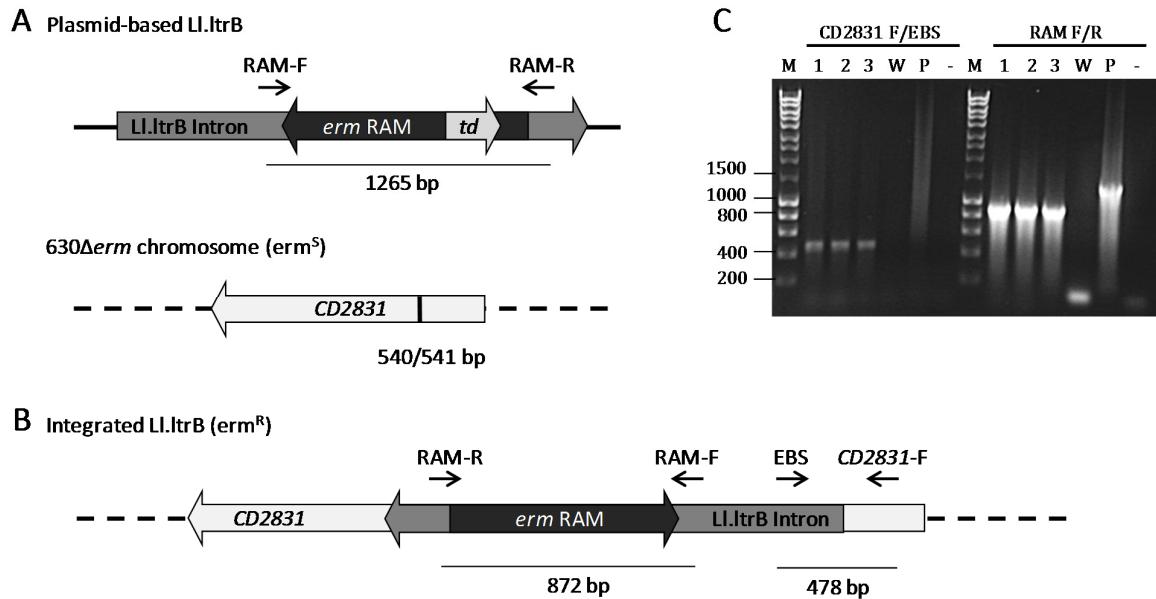


Figure 6.6: PCR screen of constructed *CD2831* mutants

A/B. Schematic diagram of the insertional inactivation of *CD2831*. The location of the primers used to screen the potential mutant clones and the expected product sizes are indicated. **C.** Three potential *CD2831* mutant clones were verified by PCR. The primer pair used for each PCR is indicated above the corresponding lanes. **Left:** Amplification across the gene-intron junctions of each mutant using a gene-specific (*CD2831*-F) and intron-specific primer (EBS) confirmed insertion of the Li.ItrB intron into *CD2831*. **Right:** Amplification of the retrotransposition-activated marker (RAM) from each mutant confirmed chromosomal integration of RAM (splicing of *td* group I intron) and loss of pMTL007C-E2 plasmid. M: HyperLadder I (Bioline – sizes indicated in base pairs), Lanes 1-3: three mutant clones, W: wildtype 630Δerm, P: plasmid pMTL007C-E2, “–” : dH₂O

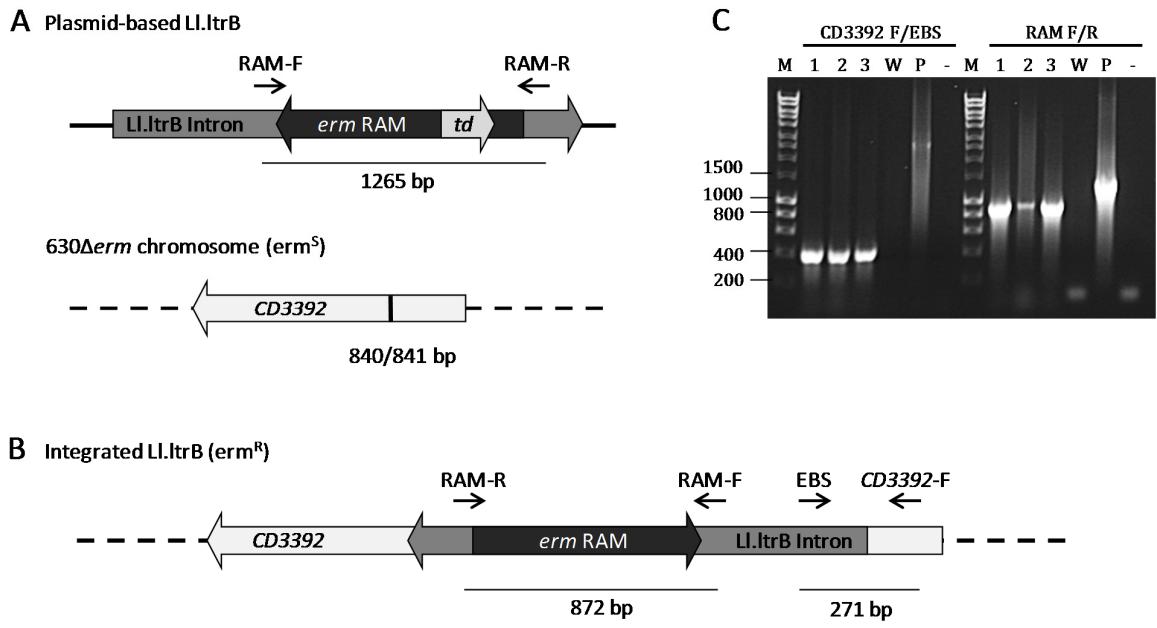
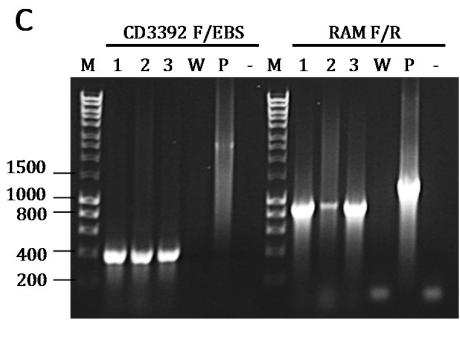


Figure 6.7: PCR screen of constructed *CD3392* mutants

A/B. Schematic diagram of the insertional inactivation of *CD3392*. The location of the primers used to screen the potential mutant clones and the expected product sizes are indicated. **C.** Three potential *CD3392* mutant clones were verified by PCR. The primer pair used for each PCR is indicated above the corresponding lanes. **Left:** Amplification across the gene-intron junctions of each mutant using a gene-specific (*CD3392*-F) and intron-specific primer (EBS) confirmed insertion of the LI.ItrB intron into *CD3392*. **Right:** Amplification of the retrotransposition-activated marker (RAM) from each mutant confirmed chromosomal integration of RAM (splicing of *td* group I intron) and loss of pMTL007C-E2 plasmid. M: HyperLadder I (Bioline – sizes indicated in base pairs), Lanes 1-3: three mutant clones, W: wildtype 630Δerm, P: plasmid pMTL007C-E2, “–” : dH₂O

Single integration of the LI.ItrB intron into the chromosome of each mutant was confirmed by Southern blot analysis using an intron-specific probe (Figure 6.8). As the genes *CD0386* and *CD3392* are highly similar, the corresponding mutants were differentiated by bands of differing size on the Southern blot. Using the same restriction enzymes (BtgI and PciI) for both Southern blots, the expected size for the *CD0386* mutant was 4 kb, while the expected size for the *CD3392* mutant was 3.5 kb. The expected size of the *CD2831* mutant following BsaBI digestion was 3.8 kb. Southern blots confirmed the single insertion of the LI.ItrB intron into three clones of each gene target, and the successful construction of three *C. difficile* sortase substrate mutants (Figure 6.8).



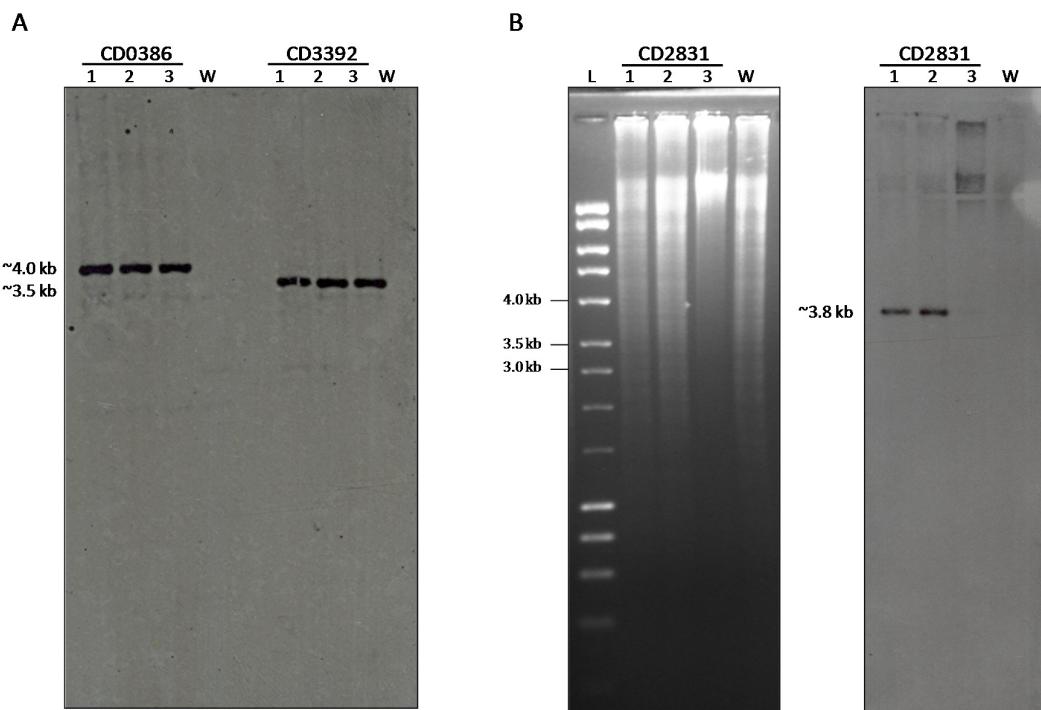


Figure 6.8: Southern blot analysis of mutants

Southern blot using a probe specific to the inserted *Ll.ltrB* intron. **A.** *BtgI* and *PciI* digests were performed on DNA from three clones each of potential *CD0386* and *CD3392* mutants, and the parent strain *630Δerm*. **B.** *BsaBI* digests were performed on three clones of *CD2831* mutants and parent strain *630Δerm*. The digest for clone three was incomplete (left), leading to a faint 3.8 kb band on the Southern blot (right). L = HyperLadder I (Bioline), Lanes 1-3: Three mutant clones, W: wildtype *630Δerm*.

6.4 Initial phenotypic characterisation of mutants

Manipulation of the cell wall protein *CwpV* resulted in visible changes to colony morphology and in cell aggregation (Reynolds *et al.*, 2011), so we wondered if the same could be observed with the potential sortase substrates. Here we describe the preliminary phenotypic tests that were performed on the three sortase substrate mutants to determine whether there was an effect on *C. difficile* growth or cell surface properties.

6.4.1 Growth dynamics

Analysis of the growth dynamics of the three mutants was performed in triplicate in both BHIS broth (Figure 6.9). The three mutants showed no difference in growth compared to the wildtype strain *630Δerm* (Figure 6.9).

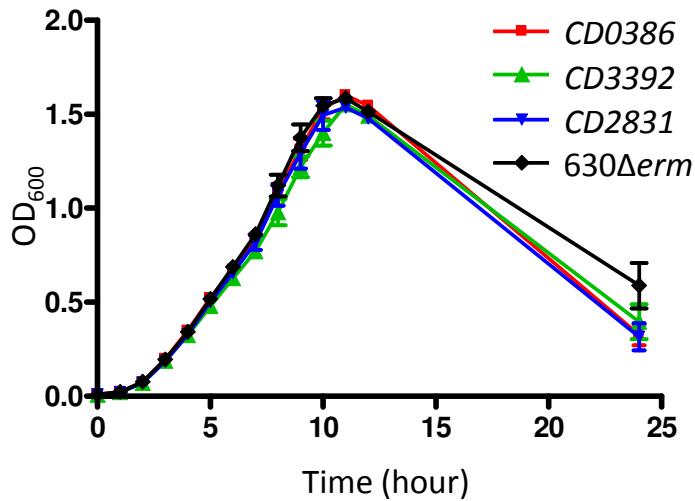


Figure 6.9: Growth analysis of sortase substrate mutants

Growth dynamics of three sortase substrate mutants; *CD0386*, *CD3392*, and *CD2831*; compared to wildtype *630Δerm* in BHIS broth. Cultures were inoculated 1/100 using an overnight primary culture. Cultures were grown in duplicate and data points are from an average of three biological replicates.

6.4.2 Colony morphology

Similarly to Section 4.3.2 with the sortase overexpression strain, the colony morphology of the three sortase substrate mutants were investigated. Colony morphology was compared between *630Δerm* (wildtype), and the *CD0386*, *CD2831*, and *CD3392* mutants: all four strains have similar colony shapes with ruffled edges, and no apparent differences were observed under 40x magnification (Figure 6.10).

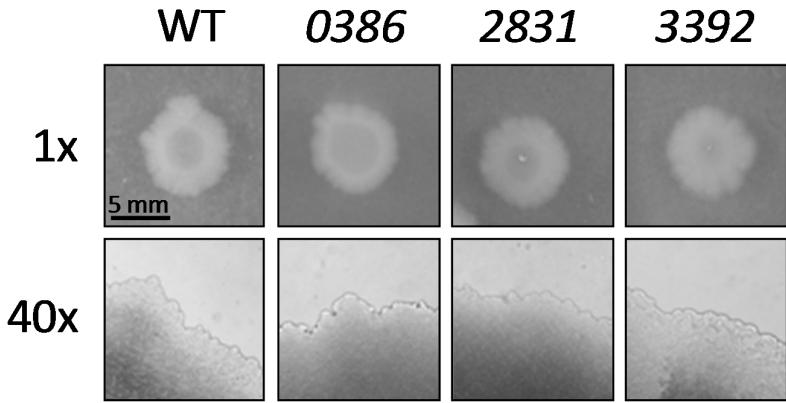


Figure 6.10: Colony morphology of substrate mutants

Two μ l of overnight cultures (WT = $630\Delta erm$, and $CD0386$, $CD2831$, and $CD3392$ mutants) were spotted onto BHIS plates and incubated for 3 days before removing plates from the anaerobic chamber to observe colony morphology both macroscopically and microscopically (40x magnification). No apparent differences between colonies were observed. Photos were taken with Canon 600D SLR (1x) and a Zeiss Axioplan 2 upright microscope (40x).

6.4.3 Cell aggregation

We hypothesized that it was possible that these predicted adhesins were involved in cell-cell interactions, so we assessed the $CD0386$, $CD2831$, and $CD3392$ mutants for their autoagglutination and sedimentation phenotypes. In both of these assays, the non-motile strain M120 was used as a positive control. Autoagglutination activity is a marker of virulence in many bacteria (Chiang *et al.*, 1995, Roggenkamp *et al.*, 1995, Howard *et al.*, 2009). The ability to autoagglutinate has been shown to affect adherence of the intestinal pathogen *Campylobacter jejuni* to intestinal epithelial cells (Misawa & Blaser, 2000, Howard *et al.*, 2009). The multi-faceted virulence factor YadA of *Yersinia pestis* contributes to both collagen-binding, and autoagglutination in static cultures (Tahir *et al.*, 2000). Autoagglutination assays were performed as previously described for *C. difficile* (Stabler *et al.*, 2009) to determine whether mutations in the predicted adhesions $CD0386$, $CD2831$, and $CD3392$ had on this cell surface associated property; however, none of the mutants showed any differences when compared to the $630\Delta erm$ wildtype (Figure 6.11). The $CD0386$, $CD2831$, and $CD3392$ mutants were assessed for their relative sedimentation following overnight growth in a liquid culture using the cell sedimentation assay (A. Faulds-Pain, LSHTM, unpublished manuscript) performed on the CD2718 overexpressing

strain in Section 4.3.3. Two clones of each mutant were tested, but there did not appear to be any differences in the level of sedimentation in strains with a 630Δerm background (Figure 6.12).

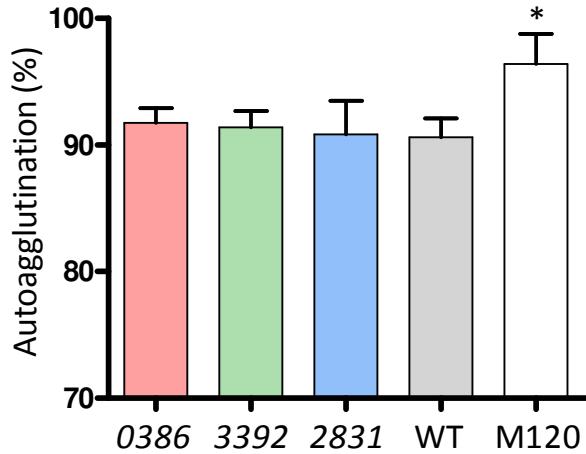


Figure 6.11: Autoagglutination of *C. difficile* mutants

C. difficile strains were grown on BHI plates for 1 to 2 days, and then resuspended to an OD₆₀₀ of approximately 1.0 in PBS. These were incubated for 24 hours, after which the OD₆₀₀ of the top one ml of bacterial suspension was measured. The percentage of autoagglutination was normalised to the starting OD₆₀₀, $((\text{Starting OD}_{600} - \text{Final OD}_{600})/\text{Final OD}_{600}) \times 100$. The bars indicate the percentage of cells autoagglutinating and the error bars indicate the standard error (* corresponds to $p < 0.05$). WT = 630Δerm; M120, a non-motile strain that autoagglutinates to a higher degree than strain 630 and used as a positive control (Stabler *et al.*, 2009).

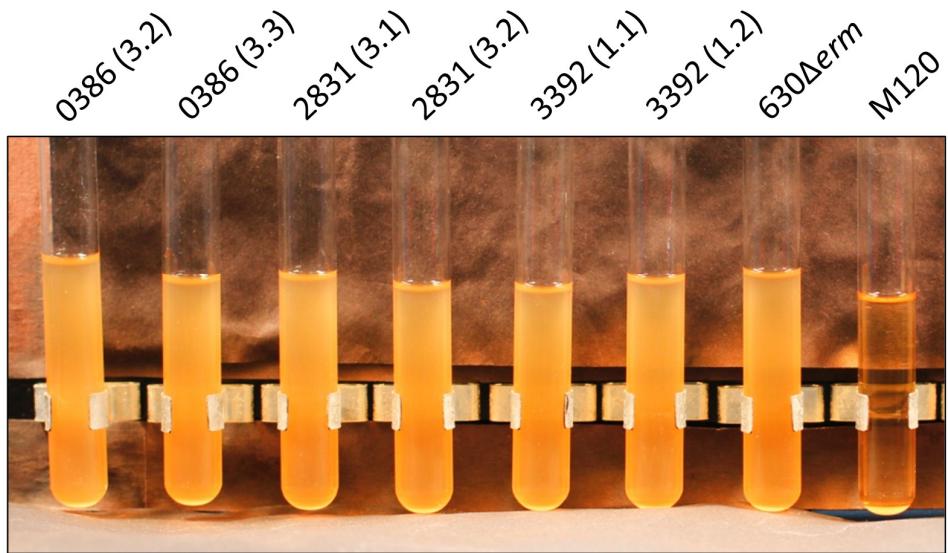


Figure 6.12: Sedimentation assay

Bacterial cell sedimentation following overnight growth in BHIS broth was compared between mutant strains and wildtype *630Δerm*. Two clones of each mutant were tested. The non-motile M120 strain, sediments strongly in this assay, and is used as a positive control.

6.4.4 Motility

Fibronectin binding protein mutants in *S. aureus* exhibit enhanced spreading motility on plates (Tsompanidou *et al.*, 2012). Plate-based motility assays for *C. difficile* have been described (Baban *et al.*, 2013, Purcell *et al.*, 2012), but reproduction of these assays has been unsuccessful in our lab. Instead a stab motility assay using a motility agar tube (BHI + 0.175% agar) (Tasteyre *et al.*, 2001, Twine *et al.*, 2009) was used to screen for differences in relative motility in the *CD0386*, *CD2831*, and *CD3392* mutants. The non-motile strain M120 served as a negative control. There were no apparent differences in relative motility observed between the stab inoculated cultures of the mutants, *630Δerm* wildtype, or strain 630 (Figure 6.13).

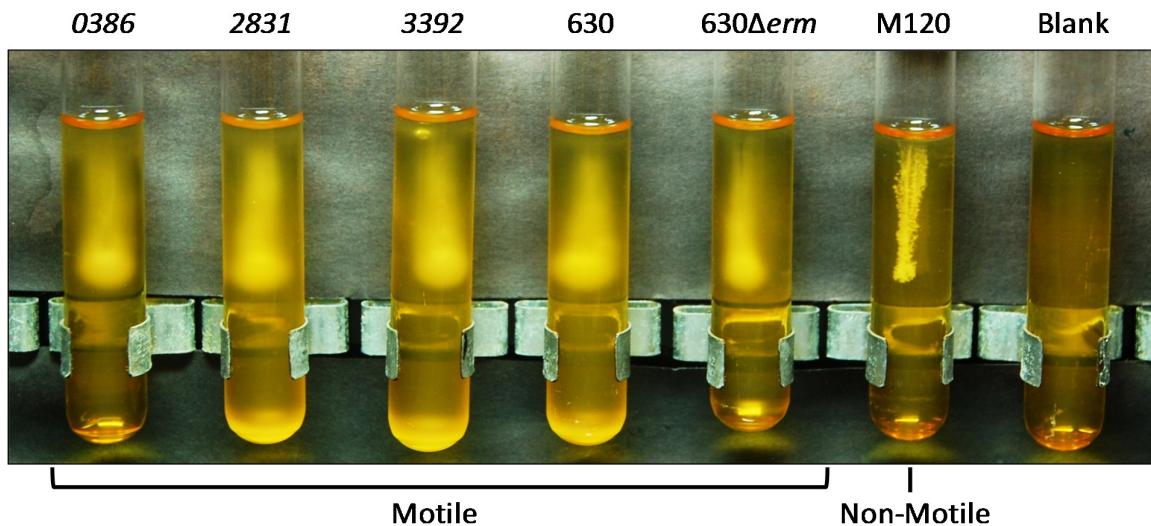


Figure 6.13: Motility testing of *C. difficile* mutants

Motility is assessed by stab inoculating semi-solid BHIS media (0.175% agar).

Motile strains exhibit diffuse growth in the media from the stab, while non-motile strains grow only within the stab. M120, a non-motile strain, served as a negative control (Stabler *et al.*, 2009). There were no apparent differences in motility between the three mutants and parent $630\Delta\text{erm}$ strain.

6.5 Discussion and future directions

This chapter describes the initial attempts to characterise three of the predicted *C. difficile* sortase substrates *in vitro*. Semi-quantitative RT-PCR transcriptional analysis performed in this study confirmed that all seven predicted sortase substrates are transcribed *in vitro*, suggesting that they are functional proteins in *C. difficile*. There were differences observed between the levels of transcription of the seven predicted substrates, indicating potential differences in their relative amounts on the cell surface. Transcription of the highly similar putative adhesion encoding *CD0386* and *CD3392* genes was the lowest of the group. All of the sortase substrates appear to be expressed as single transcriptional units, with the exception of the putative 5' nucleotidase encoding *CD2768*, which may be co-transcribed with an upstream putative polysaccharide biosynthesis protein. However, it is unclear why these two functionally distinct proteins might be co-transcribed.

Three sortase substrate mutants were successfully constructed using the ClosTron method (Heap *et al.*, 2007, Heap *et al.*, 2010), and their genotypes confirmed by PCR and

Southern blot. We investigated their phenotypes in several preliminary *in vitro* tests, but these mutations do not appear to have an effect on the cell surface properties thus far investigated. Further investigation is required to elucidate the function of these three proteins, as well as the other putative sortase substrates. Firstly, as predicted sortase substrates, these proteins are expected to be localised on the cell surface. There is some evidence that CD0386, CD2831, and CD3392 may be surface exposed, as convalescent sera from CDI patients contained antibodies against these proteins (see Section 7.2), but this has not been experimentally demonstrated. Antisera against each of the substrates would need to be generated prior to cell fractionation studies that would confirm their surface localisation and association with the cell wall, and support the hypothesis that they are traditional sortase substrates.

CD0386, CD2831, and CD3392 are predicted to function as adhesins, so their binding capabilities to collagens and other extracellular matrix proteins should be investigated. Recombinant CbpA was shown by ELISA to bind to immobilised type I and V collagens, and to these collagens on the human fibroblast cell line, IMR-90 (Tulli *et al.*, 2013). Similarly, immuno dot-blotting and fibronectin-binding ELISAs were used to investigate the adherent properties of the *C. difficile* fibronectin binding protein, Fbp68 (Hennequin *et al.*, 2003). Hennequin *et al.* (2003) also demonstrated binding of *C. difficile* cells to fibronectin was abolished by pre-incubation of the cells with anti-Fbp68 antibodies. The study by Tulli *et al.*, (2013) attempted to investigate the role of CbpA on *C. difficile* binding to collagen, but found no differences in binding between wildtype *C. difficile* and *cbpA* mutants. The authors suggest that collagen binding in *C. difficile* appears redundant, and that it is likely many proteins contribute towards *C. difficile* binding to the extracellular matrix. Because of this, it may be difficult to observe a phenotype in adherence assays for the three mutants generated in this chapter. So, it might be ideal to perform adherence assays like the ones described above using recombinantly expressed CD0386, CD2831, and CD3392 proteins to investigate whether they facilitate binding to the extracellular matrix.

It would also be interesting to determine whether the mutations in *CD0386*, *CD2831*, and *CD3392* have any influence on *C. difficile* colonisation in mice, but we foresee potential complications in such studies. In *S. aureus*, sortase A (*srtA*) mutants that are deficient in

the anchoring of 19 different surface proteins show severe defects in pathogenesis in animal models (Mazmanian *et al.*, 2000, Jonsson *et al.*, 2003, Weiss *et al.*, 2004, Cheng *et al.*, 2009). However, few individual sortase substrate mutants are attenuated in virulence to the same degree as *srtA* mutants (Cheng *et al.*, 2009), suggesting sortase-anchored proteins may function synergistically in causing infection. The effect of each individual mutation of the *C. difficile* sortase substrates might be too small to observe a phenotype in mice. In addition to this, a recent study demonstrates that two predicted *C. difficile* sortase substrates, CD2831 and CD3246, are readily cleaved by the secreted metalloprotease, CD2830 (Hensbergen *et al.*, 2014). Expression of the predicted sortase substrates and this metalloprotease appears inversely regulated by c-di-GMP (Soutourina *et al.*, 2013, Hensbergen *et al.*, 2014). This may make it difficult to observe differences between wildtype *C. difficile* and CD2831 mutants, as the wildtype cells may not have intact CD2831 protein on the cell surface.

7 Development of a sortase cleavage assay using recombinantly expressed sortase substrate proteins

7.1 Introduction

In Chapter 5, we described the cleavage of short peptides containing (S/P)PXTG sequences by recombinantly expressed CD2718 in a FRET-based assay. This assay demonstrated that the activity of recombinant CD2718 closely resembles that of previously described sortases, as CD2718-mediated cleavage occurred between the same residues (the threonine and glycine) and was dependent on a single cysteine residue. This is an important step in understanding the function CD2718 has in *C. difficile*. However, it remains unknown whether CD2718 cleaves the predicted substrate proteins containing these sequences. In other bacteria, this is repeatedly demonstrated by comparing the localisation of sortase substrates in cell fractions within wildtype and sortase mutants. In sortase mutants, sortase substrates are either released into the culture supernatant or retained in the membrane, instead of found in the cell wall fraction like in wildtype cells (Mazmanian *et al.*, 2000, Barnett & Scott, 2002, Bierne *et al.*, 2002, Kharat & Tomasz, 2003, Gaspar *et al.*, 2005, Marraffini & Schneewind, 2006, Marraffini & Schneewind, 2007, Swaminathan *et al.*, 2007, Necchi *et al.*, 2011). The CD2718 mutant construction attempts in Chapter 3 were repeatedly unsuccessful, so this type of experiment in *C. difficile* was not possible for us.

Instead, we sought to ascertain whether CD2718 cleaved the whole protein substrates through the development of a visual SDS-PAGE based assay, similar to that depicted in Figure 7.1 from Popp *et al.* (2009). In their study, Popp *et al.* (2009) demonstrated that introduction of an LPETG sequence into a flexible loop of a protein (in this case, human ubiquitin) resulted in the observable cleavage of that protein over time by the *S. aureus* SrtA. They were able to observe the time-dependent accumulation of both the N-terminal and C-terminal cleavage products of the LPETG containing protein, as well as traces of the sortase-substrate intermediate. Kang *et al.* (2011) also utilise an *in vitro* western-based assay to demonstrate that a purified SrtB from *S. pyogenes* polymerises purified pilin subunits. A similar assay that demonstrates that the predicted *C. difficile* sortase substrate

proteins identified in Chapter 3 are cleaved by CD2718 would greatly facilitate our understanding of CD2718 activity.

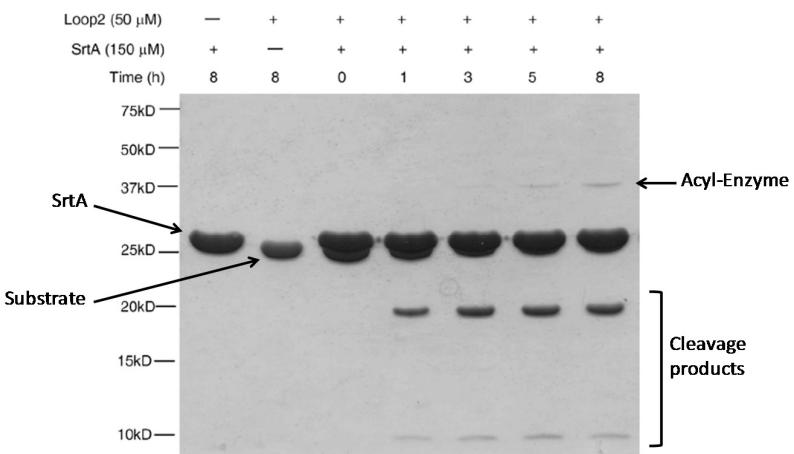


Figure 7.1: Sortase-mediated cleavage of a protein substrate

In a study by Popp *et al.* (2009), an LPETG sequence was introduced to an exposed loop of recombinant ubiquitin hydrolase. The modified protein was then incubated with SaSrtA for the indicated length of time and the reactions separated by SDS-PAGE and stained with Coomassie. Visible is the appearance of both cleavage products over time, as well as traces of the transient sortase-substrae (acyl-enzyme) intermediate. Figure modified from Popp *et al.*, 2009.

7.1.1 Aims

This chapter describes the development of a whole protein cleavage assay to mirror the CD2718 cleavage results obtained in Chapter 5 using recombinantly expressed proteins as substrates in place of the FRET peptides.

The aims of this study were to:

- Express predicted sortase substrates in *E. coli* with N- and C-terminal tags
- Test with the active recombinant CD2718_{ΔN26} and demonstrate cleavage by SDS-PAGE or western blotting

7.2 Recombinant His-gene-HA substrate proteins

We sought to recombinantly express the sortase substrates predicted in Chapter 3 in *E. coli* with both N-terminal and C-terminal tags to facilitate the detection of sortase-mediated cleavage *in vitro*. To begin, we first tested the expression of double-tagged sortase substrate constructs obtained in collaboration with the Felgner lab at the University of California, Irvine.

The Felgner lab has pioneered a high throughput cloning and protein microarray approach that is useful for profiling the immunoreactivity of bacterial proteins on a large scale (Eyles *et al.*, 2007, Felgner *et al.*, 2009, Kunnath-Velayudhan *et al.*, 2010, Molina *et al.*, 2010, Vigil *et al.*, 2010, Liang *et al.*, 2013). The Felgner lab, in collaboration with the Wren lab, piloted a *C. difficile* protein array for 80 subjects (58 CDI cases and 22 control patients) using a panel of 36 protein targets (18 from strain 630 and 18 from R20291) that are predicted to be important for *C. difficile* colonisation, survival or infection were selected. These *C. difficile* proteins were expressed in a cell-free *in vitro* transcription/translation reaction and printed directly onto microarray slides. When probed with serum from hospitalised patients with or without CDI, four antigens were identified as being reactive in patients with CDI (Figure 7.2) that included predicted substrates: CD3392, CDR20291_2722 (the R20291 orthologue of CD2831) and CDR20291_3453 (the R20291 orthologue of CD0386 and CD3392).

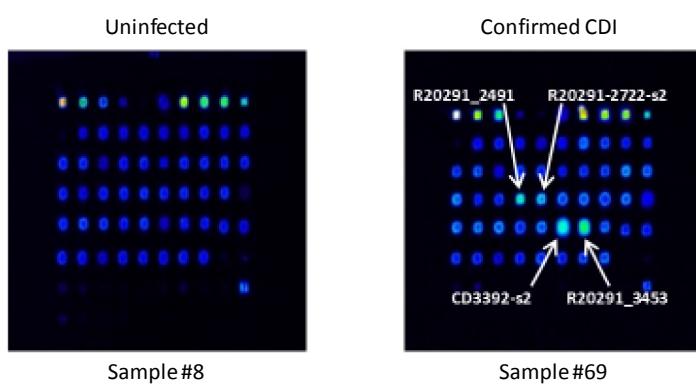


Figure 7.2: Sample of *C. difficile* protein array pilot data

A panel of 36 predicted *C. difficile* virulence protein targets (18 from strain 630 and 18 from strain R20291) were expressed and printed directly onto the microarray slides, and then probed with baseline serum specimens from hospitalised patients with or without CDI confirmed by both ELISA toxin test

and microbiology culture. Four of the antigens reacted with CDI patient sera, and included the predicted substrates CD3392, CDR20291_2722 (R20291 orthologue of CD2831) and CDR20291_3453 (R20291 orthologue of CD0386 and CD3392). Figure from the Felgner lab (University of California, Irvine).

We were able to obtain the protein expression constructs used in the pilot study for the works described in this chapter. The expression constructs from the Felgner lab incorporated a 10x histidine tag at the N-terminal end and an HA tag at the C-terminal end, with expression of the protein under the control of the T7 promoter (Figure 7.3, Table 7.1). Due to the length of the predicted substrate proteins (from 900 to over 1000 amino acid residues long), the gene fragment corresponding to the C-terminal half of these proteins were cloned separately by the Felgner lab, and indicated as “s2” fragments. These constructs were considered suitable for our assay as they still contain the C-terminal sortase cleavage site.



Figure 7.3: Construct design from pilot array

Schematic of recombinant substrate proteins expressed from constructs obtained from the Felgner lab. In these constructs, *C. difficile* gene fragments are cloned under the control of the T7 promoter with an N-terminal 10x histidine tag and a C-terminal HA tag. Included are the expected sizes of the CD3392-s2 protein and the predicted cleavage products.

Table 7.1: His-HA expression constructs from the Felgner lab

Construct name	Cloning details	Expected protein size (kDa)
pUCIrv7	CD2718, residues 1-225 (full length)	31
pUCIrv17	CDR20291_3453-s2, CD0386/CD3392 orthologue in R20291, residues 435-901	56
pUCIrv20	CDR20291_2722-s2, CD2831 orthologue in R20291, residues 486-984	59
pUCIrv26	CD2831-s2, residues 434-972	64
pUCIrv27	CD0386-s2, residues 451-1014	67
pUCIrv28	CD3392-s2, residues 451-1014	67

7.2.1 Expression testing

We chose the Rosetta(DE3) strain (Novagen) for expression of the *C. difficile* proteins in *E. coli*. Rosetta(DE3) cells are a derivative of the BL21(DE3) expression strain that carry a plasmid encoding six tRNAs rare in *E. coli*. As described in Section 4.4.1, rare codons in *E. coli* are common in *C. difficile* proteins, and the use Rosetta(DE3) cells was hypothesised to improve recombinant expression.

Expression of these constructs in Rosetta(DE3) cells was induced with 1 mM IPTG for 5 hours in 10 ml cultures before harvesting the cells and probing SDS-PAGE separated whole cell lysates with anti-His and anti-HA antibodies (Figure 7.4). Expression from the pUCIrv7 construct listed in Table 7.1 was initially tested for potential use in the FRET assay (Figure 7.4, Lane 1), but as the purified CD2718_{ΔN26} protein proved successful, expression of this sortase construct was not pursued further.

A 50-55 kDa band that reacted with the anti-His antibody was observed in the whole cell lysates of the Rosetta(DE3) cells carrying the pUCIrv17 (R20291_3453-s2), pUCIrv20 (R20291_2722-s2), and pUCIrv26 (CD2831-s2) constructs (Figure 7.4A, Lanes 2-4). However, these bands did not react with the anti-HA antibody (Figure 7.4B, Lanes 2-4).

No band strongly reacted with either antibody in the lysates corresponding to the pUCIrv27 construct, indicating that the CD0386-s2 protein was not expressed under these conditions (Figure 7.4, Lane 5). A 60 kDa protein was observed in the whole cell lysate of the induced Rosetta(DE3)/pUCIrv28 culture; this band reacted strongly with the anti-His antibody and also reacted with the anti-HA antibody (Figure 7.4, Lane 6). This band was presumed to be the successfully expressed CD3392-s2 protein (predicted size of 67 kDa), so this construct was carried forward for protein purification.

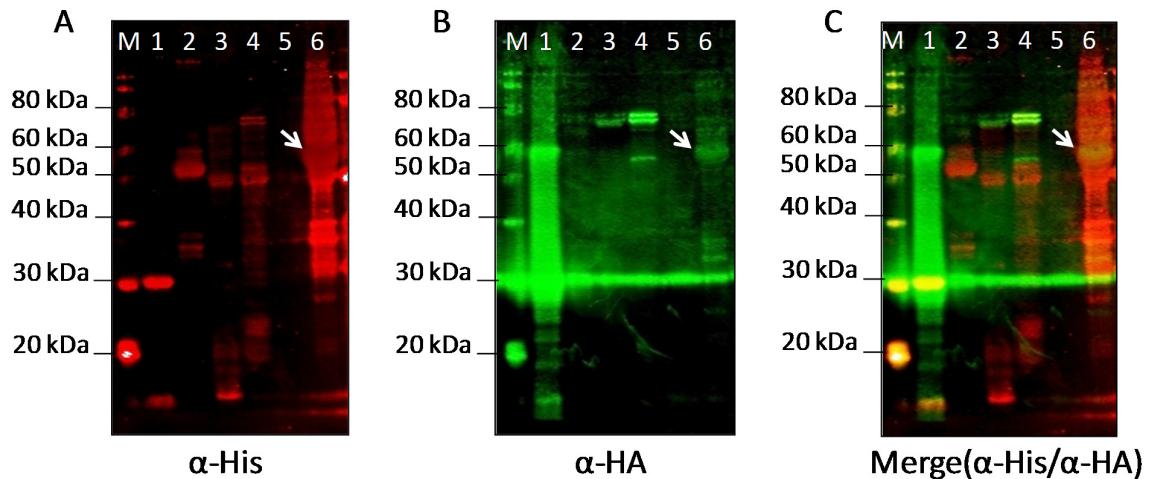


Figure 7.4: Expression of His-HA proteins

Expression from the constructs listed in Table 7.1 was induced in *E. coli* with 1 mM IPTG in 10 ml Rosetta(DE3) cells at 37 °C for 5 hours. Two colour anti-His and anti-HA western blots of whole cell lysates of harvested cells. The individual channels, and the merge of the two are displayed: red corresponds to the anti-His channel (**A**), green corresponds to the anti-HA channel (**B**), and yellow on the merge image indicates where both antibodies bound (**C**). 1=CD2718, 2=R20291_3453-s2, 3=R20291_2722-s2, 4=CD2831-s2, 5=CD0386-s2, 6=CD3392-s2, M= MagicMark™ XP Western Protein Standard (Invitrogen), L= PageRuler Plus Prestained Protein Ladder (Fermentas). An arrow indicates a 60 kDa protein that reacts with both the His and HA antibodies and is predicted to be the CD3392-s2 protein.

7.2.2 Purification of CD3392-s2

Cell lysates of 200 ml induced Rosetta(DE3)/pUCIrv28 cultures were batch incubated with 1 ml nickel agarose slurry for 1 hour, before eluting bound protein with 250 mM imidazole.

The eluate was separated by SDS-PAGE and stained with Coomassie (Figure 7.5A). The 60 kDa band that reacted with both the anti-His and anti-HA antibodies (Figure 7.5B) was revealed to be the CD3392-s2 protein by MALDI fingerprinting (Figure 7.5C).

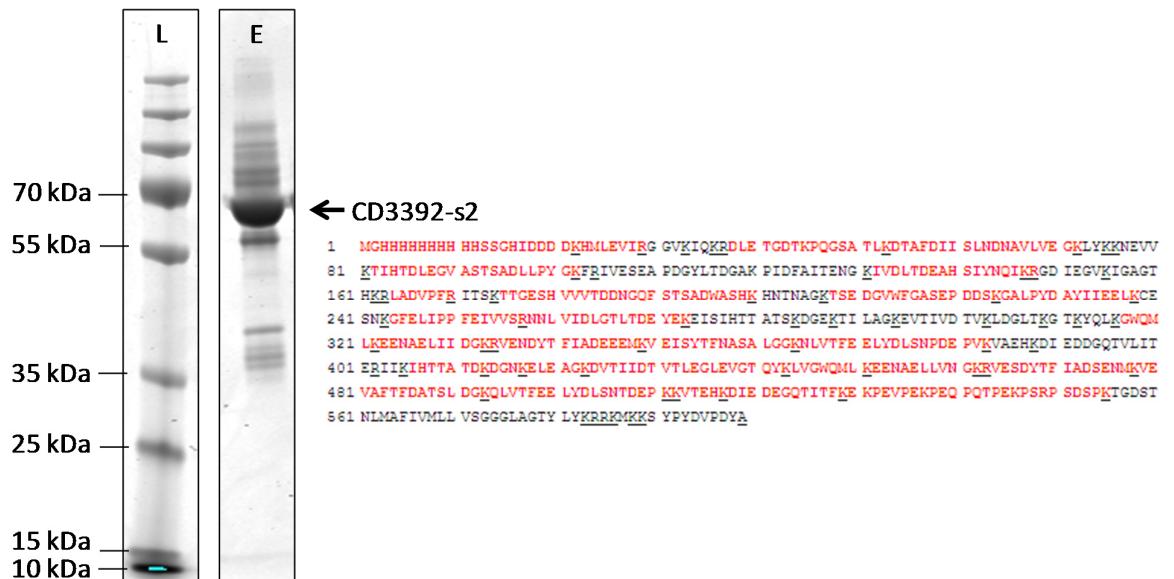


Figure 7.5: Confirmation of CD3392-s2 expression

A 60 kDa protein was purified with Ni-NTA agarose from 200 ml cultures of Rosetta(DE3)/pUCIrv28. Bound protein was eluted with 250 mM imidazole, separated by SDS-PAGE and the eluate stained with Coomassie (E). MALDI fingerprinting of the 60 kDa band indicated by an arrow confirmed this to be the CD3392-s2 protein. The sequence of CD3392-s2 is depicted, with the peptides in red identified during MALDI analysis. L = PageRuler Plus Prestained Protein Ladder (Fermentas).

7.2.3 Initial cleavage assay attempts

The CD3392-s2 protein was concentrated and buffer-exchanged into FRET buffer using an Amicon Ultra centrifugal filter column, before total protein content was quantified using the Bradford protein assay. Fifteen μ M of active *C. difficile* sortase enzyme CD2718 $_{\Delta N26}$ was incubated with 20 μ M CD3392-s2 at 37 °C for 48 hours before separating reaction products by SDS-PAGE (Figure 7.6). Each protein was incubated on its own to serve as a control for protein degradation products over the incubation period. The Coomassie stained gel did not reveal any evidence of cleavage; neither did the anti-His/anti-HA two colour western blot, though this is likely due to weak signal from both antibodies (Figure 7.6).

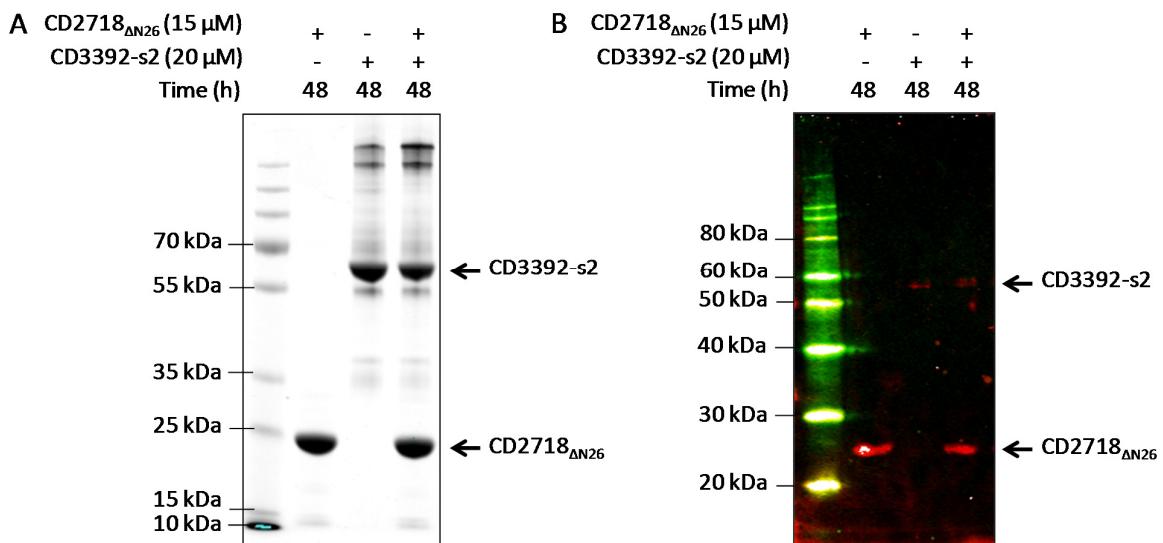


Figure 7.6: Cleavage assay with CD3392-s2

Twenty µM CD3392-s2 was incubated with 15 µM CD2718_{ΔN26} protein in FRET buffer at 37 °C for 48 hours, after which the reaction products were analysed by SDS-PAGE. **A.** Coomassie stained gel. **B.** Two colour anti-His (red) and anti-HA (green) western blot.

7.3 Recombinant His-CD3392-s2-Strep protein

In an attempt to improve the signal detection of the CD3392-s2 protein, alternative C-terminal tags were considered. The detection of the CD3392-s2 protein with the anti-HA antibody required a large amount of antibody to be used (1:100 dilution) and the signal was inconsistent in intensity. This and the relative impurity of the nickel-purified CD3392-s2 protein complicated the SDS-PAGE based detection of CD2718-mediated cleavage. To address both of these issues, a Strep-tag II was considered as a replacement for the HA tag. In addition to simply being another tag to use, the Strep tag has been successfully used for the purification of *C. difficile* proteins in *E. coli* (Dembek *et al.*, 2012), and Strep-tagged proteins can be purified by affinity chromatography using StrepTactin resin. The CD3392-s2 fragment from the pUCIrv28 construct was cloned into the Ncol/Xhol sites of pET28a to generate pEHD016. The resulting recombinant CD3392s2-Strep protein has an N-terminal 6xHis tag and a C-terminal Strep tag in place of the HA tag from CD3392-s2, and a predicted size of 64 kDa (Figure 7.7).

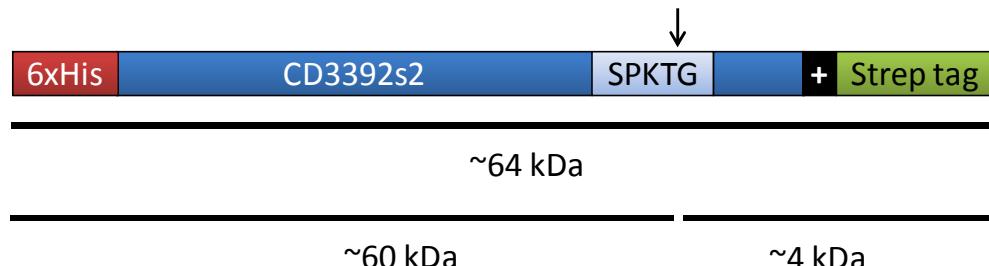
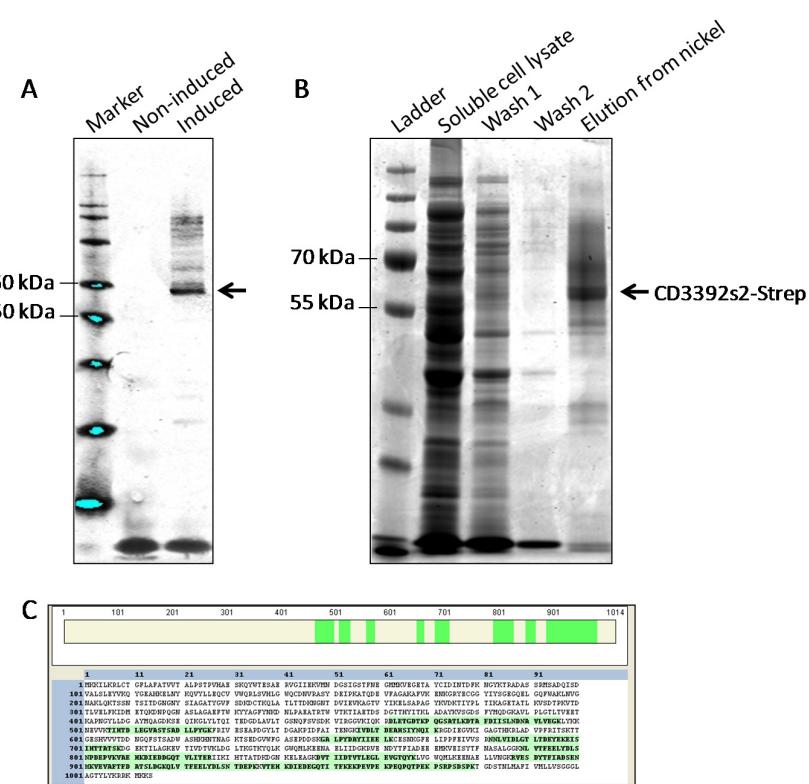


Figure 7.7: Design of CD3392s2-Strep protein

The CD3392-s2 fragment was amplified from the Felgner lab construct and cloned into pET28a to generate the plasmid pEHD016. A C-terminal strep tag was introduced to replace the C-terminal HA tag, resulting in a recombinant protein predicted to be 64 kDa.

7.3.1 Expression testing and purification

Initial expression tests of Rosetta(DE3)/pEHD016 identified a 60 kDa protein that reacted with the anti-His antibody in the induced culture that was absent in the non-induced control (Figure 7.8A). Whole cell lysates of 200 ml cultures were spun down to separate the soluble cell lysate, and incubated with 1 ml nickel agarose, before eluting bound protein with 250 mM imidazole (Figure 7.8B). The 60 kDa protein was visible within a smear in the eluate from nickel resin (Figure 7.8B), and this band confirmed as CD3392s2-Strep by MALDI fingerprinting (Figure 7.8C).



The yield of CD3392s2-Strep protein from a 200 ml culture of Rosetta(DE3)/pEHD016 after concentrated was quite low, and limited the concentration that could be tested in the cleavage assay. This made detecting any evidence of cleavage difficult in either the Coomassie stained gel or the anti-His/anti-Strep two colour western blot (Figure 7.9). Neither antibody worked particularly well on the limited protein sample.

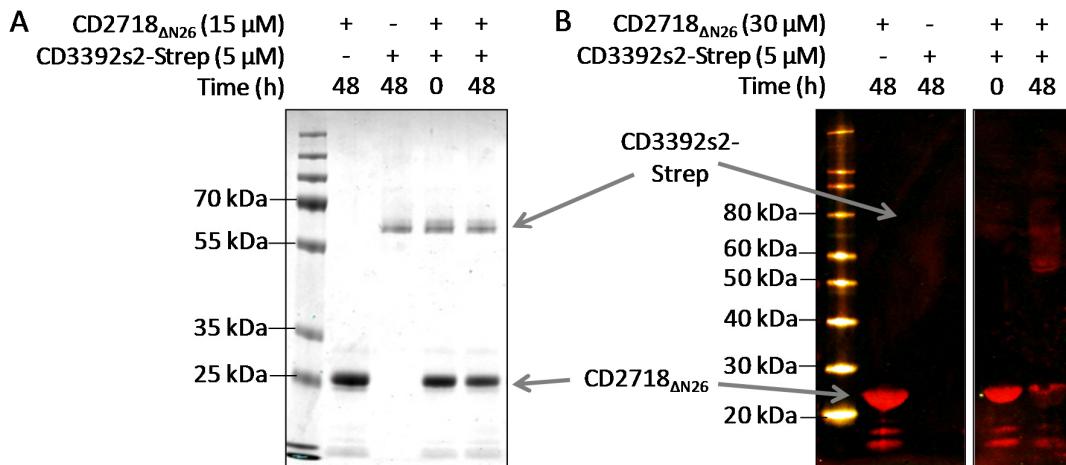


Figure 7.9: Cleavage assay attempts with CD3392s2-Strep

Twenty μM CD3392s2-Strep was incubated with $15 \mu\text{M}$ CD2718 $_{\Delta N26}$ protein in FRET buffer at 37°C for 48 hours, after which the reaction products were analysed by SDS-PAGE. **A.** Coomassie stained gel. **B.** Two colour anti-His (red) and anti-Strep (green) western blot.

7.4 Expression of truncated substrate protein with C-terminal tags

Due to the difficulty in detecting a size shift of the two 60+ kDa recombinant CD3392 proteins described above, further truncated sortase substrate expression constructs were designed. These were designed such that the full length protein would be 20-35 kDa, and that cleavage by CD2718 $_{\Delta N26}$ between the threonine and glycine residues would result in a 3-5 kDa size shift, and with a variety of C-terminal tags (Figure 7.10). A 3-5 kDa size shift in the 20-35 kDa size range was predicted to be detectable based on the differential migration patterns observed for the 28 kDa CD2718 and the 25 kDa CD2718 $_{\Delta N26}$ expressed in Chapter 4 (Figure 4.17 and Figure 4.20, respectively).

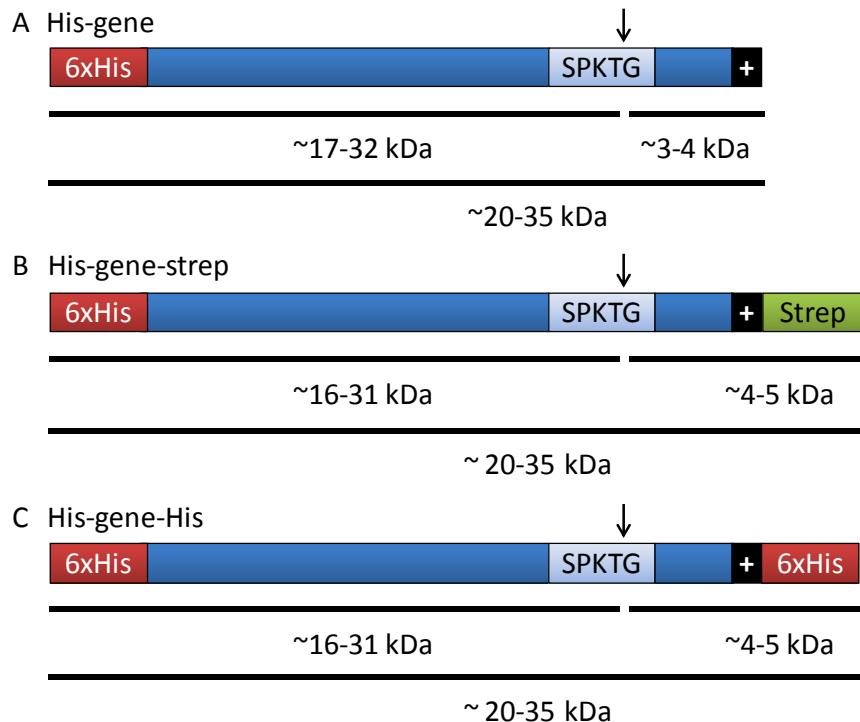


Figure 7.10: Schematic of new substrate fragment expression constructs

Truncated sortase substrate expression constructs were designed such that a size shift resulting from CD2718-mediated cleavage would be detectable by either Coomassie staining or western blotting. The full length proteins were designed to be 20-35 kDa in length, with a size shift of 3-5 kDa expected following cleavage. The cloned constructs included substrates with **A.** an N-terminal 6xHis tag only, **B.** an N-terminal 6xHis tag and a C-terminal Strep tag, and **C.** both an N- and C-terminal His tag.

7.4.1 Expression testing of successfully cloned constructs

Truncated sortase substrate gene products were cloned into the Ncol/Xhol sites of pET28a. Successfully cloned constructs and their respective full length and cleaved product sizes are detailed in Table 7.2.

Table 7.2: Expected sizes of recombinant truncated substrate proteins

Construct	Protein	Residues of substrate cloned	Expected size (kDa)	Cleavage product sizes (kDa)	
				N-terminal	C-terminal
pEHD017	H-CD0183-Strep	23-340	36.6	32.8	3.8
pEHD018	H-CD0183-H		36.6	32.8	3.8
pEHD019	H-CD2768	36-235	21.3	18.3	3
pEHD020	H-CD2768-Strep		22.3	18.3	4
pEHD021	H-CD2768-H		22.3	18.3	4
pEHD022	H-CD2831	811-972	18.4	15.2	3.2
pEHD023	H-CD2831-Strep		19.4	15.2	4.2
pEHD024	H-CD3392	841-1014	20.4	16.7	3.8
pEHD025	H-CD3392-Strep		21.4	16.7	4.8
pEHD026	H-CD3392-H		21.4	16.7	4.8

All pET28a expression constructs were transformed into NiCo21(DE3) except for the pEHD017 and pEHD018 constructs, which were transformed into Rosetta(DE3). The NiCo21(DE3) (Robichon *et al.*, 2011) expression strain was chosen for the ~20 kDa constructs to avoid confusion of their expression with that of the 24 kDa *E. coli* contaminant protein SlyD, as described in detail in Section 4.4.2. Expression was induced in 10 ml cultures before harvesting the cells and probing SDS-PAGE separated whole cell lysates with anti-His or anti-Strep antibody (Figure 7.11-Figure 7.14). The anti-His antibody detected expression of a 40 kDa protein for both the pEHD017 (H-CD0183-Strep) and pEHD018 (H-CD0183-H) expression constructs, but this protein was not detected with the anti-Strep antibody (Figure 7.11). Expression from the three CD2768 constructs (pEHD019, pEHD020, and pEHD021) was leaky, as a 25 kDa band that reacted with the anti-His antibody was observed in the induced and non-induced cultures (Figure 7.12A). This 25 kDa band also reacted with the anti-Strep antibody for the pEHD020 (H-CD2768-Strep) construct (Figure 7.12B). Neither of the two CD2831 constructs (pEHD022 and pEHD023) appeared to express under these conditions (Figure 7.13). A 28 kDa band that reacted with the anti-His antibody was observed for all three CD3392 constructs

(pEHD024, pEHD025, and pEHD026), and this band also faintly reacted with the anti-Strep antibody for the pEHD025 (H-CD3392-Strep) construct (Figure 7.14).

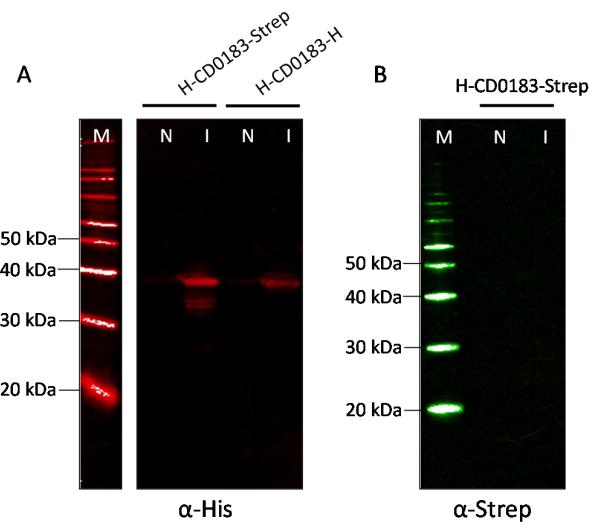


Figure 7.11: Expression of CD0183 fragments

Western blots of whole-cell lysates from 10 ml cultures of Rosetta(DE3)/pEHD017 (H-CD0183-Strep) and Rosetta(DE3)/pEHD018 (H-CD0183-H). **A.** Anti-His western blot revealed a 40 kDa protein in the induced (I), but not the non-induced cultures (N) for both expression constructs. **B.** Anti-Strep western blot of Rosetta(DE3)/pEHD018 did not detect the presence of any strong 40 kDa bands.

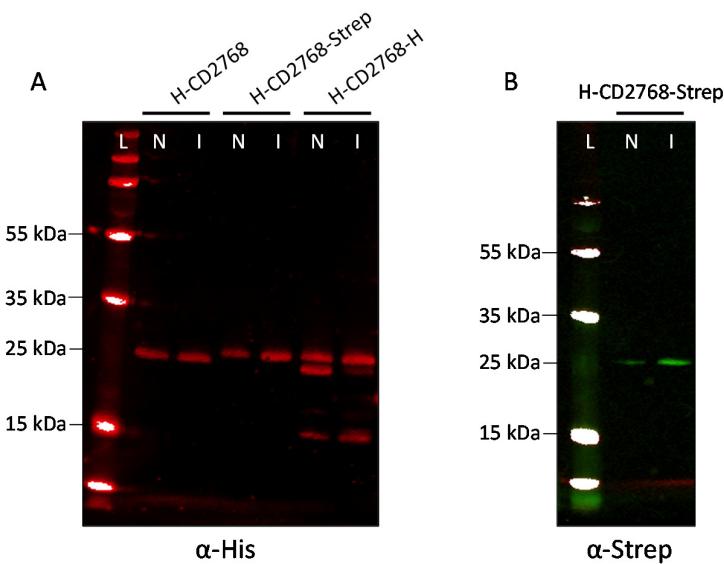


Figure 7.12: Expression of CD2768 fragments

Western blots of whole-cell lysates from 10 ml cultures of Rosetta(DE3)/pEHD019 (H-CD2768), Rosetta(DE3)/pEHD020 (H-CD2768-Strep) and Rosetta(DE3)/pEHD021 (H-CD2768-H). **A.** Anti-His western blot revealed a 25 kDa protein in the induced (I), as well as in the non-induced cultures (N) for all three expression constructs. **B.** Anti-Strep western blot of

Rosetta(DE3)/pEHD020 lysates revealed this 25 kDa band also reacted with the anti-Strep antibody.

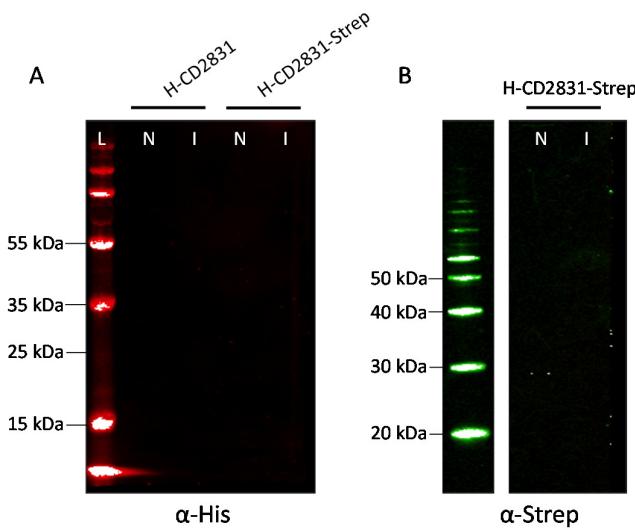


Figure 7.13: Expression testing of CD2831 fragments

Western blots of whole-cell lysates from 10 ml cultures of Rosetta(DE3)/pEHD022 (H-CD2831) and Rosetta(DE3)/pEHD023 (H-CD2831-Strep). Neither the anti-His antibody (**A**) nor the anti-strep antibody (**B**) detected any bands in the whole cell lysates of the induced (I) or non-induced (N) cultures for either expression construct.

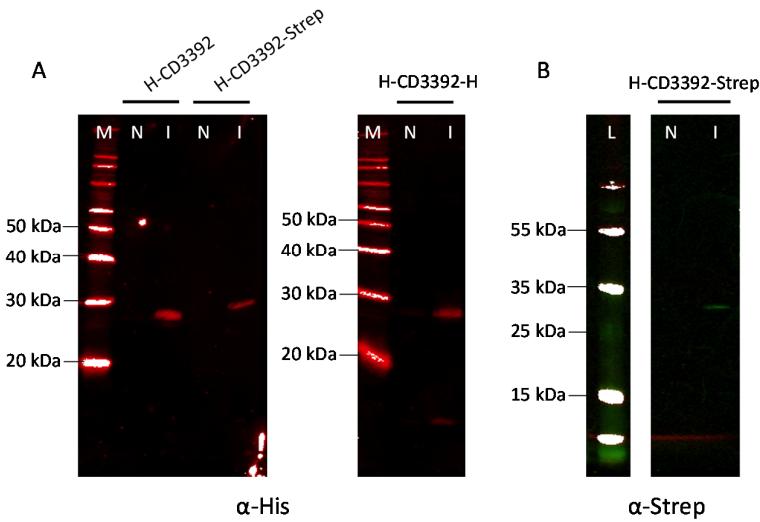


Figure 7.14: Expression of CD3392 fragments

Western blots of whole-cell lysates from 10 ml cultures of Rosetta(DE3)/pEHD024 (H-CD3392), Rosetta(DE3)/pEHD025 (H-CD3392-Strep) and Rosetta(DE3)/pEHD026 (H-CD3392-H). **A** Anti-His western blot revealed a 28 kDa protein in the induced (I), but not the non-induced cultures (N) for all three expression constructs. **B**. Anti-Strep western blot of

Rosetta(DE3)/pEHD025 lysates revealed this 28 kDa band also reacted with the anti-Strep antibody.

Many of the expressed proteins migrated on SDS-PAGE at slightly different sizes than expected based on their predicted sizes listed in Table 7.2. For example, the CD3392 constructs migrated at twice their expected size, at around 40 kDa instead of 21 kDa; and instead of being closer to 20 kDa, the three CD2768 constructs migrated at 25 kDa, the same size as the CD2718_{ΔN26} protein.

While induction of expression was observed for all but two of the constructs tested, the H-CD2768-Strep protein was carried forward for further analysis. Despite the leaky expression and the protein running at the same size as the CD2718_{ΔN26} enzyme, it was hypothesised that a 3 or 4 kDa shift in size would be more easily distinguishable within that size range, rather than at the 40 kDa size range for the CD3392 constructs. The H-CD2768-Strep protein also reacted the best with both antibodies. The CD2831 constructs did not express under these conditions, so they weren't carried on. While the two CD0183 constructs appeared to express well, the SPSTG sequence within this protein was the least efficiently cleaved peptide in the FRET assay (Section 5.3.1).

The H-CD2768-Strep protein was purified using Ni-NTA agarose from 200 ml of Rosetta(DE3)/pEHD020 cultures as described previously, and the resulting eluted protein separated by SDS-PAGE and stained with Coomassie (Figure 7.15). The resulting 25 kDa band was confirmed as H-CD2768-Strep by MALDI fingerprinting (Figure 7.15).



Figure 7.15: Confirmation of H-CD2768-Strep expression

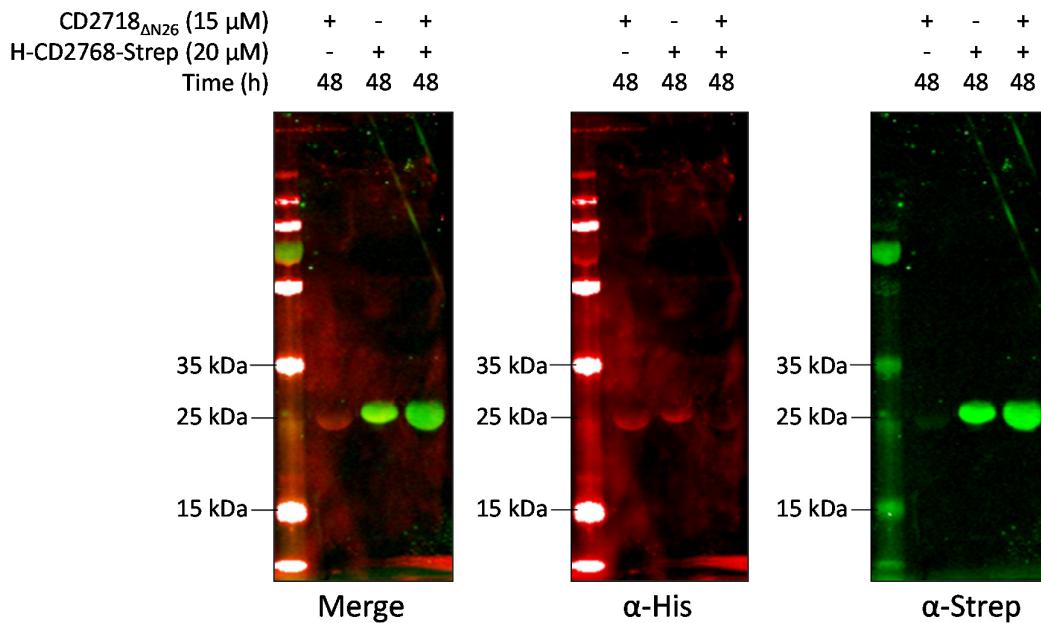
Cell lysates of 200 ml induced Rosetta(DE3)/pEHD020 cultures were batch incubated with 1 ml nickel agarose slurry for 1 hour, before eluting bound protein with 250 mM imidazole. The eluate was separated by SDS-PAGE and stained with Coomassie. C. The 25 kDa band (indicated with an arrow) was revealed to be the H-CD2768-Strep protein by MALDI fingerprinting. The sequence of the H-CD2768-Strep protein is depicted, with the peptides in red identified during MALDI-TOF analysis.

7.4.2 Cleavage assay attempts

The H-CD2768-Strep protein was concentrated and buffer-exchanged into FRET buffer using an Amicon Ultra centrifugal filter column, before total protein content was quantified using the Bradford protein assay. As before, 15 µM of active *C. difficile* sortase enzyme CD2718_{ΔN26} was incubated with 20 µM H-CD2768-Strep at 37 °C for 48 hours before separating reaction products by SDS-PAGE (Figure 7.16). Anti-Strep antibody concentrations were increased (from 1:1000 to 1:100) to improve Strep-tag detection. When anti-His/anti-Strep two colour western blots were performed, the Strep antibody signal at the new concentration was much stronger, but appeared to interfere with the His antibody binding (Figure 7.16A). Separate anti-His and anti-Strep western blots were performed to see whether a size shift in the H-CD2768-Strep protein was apparent and to improve the His signal. There did not appear to be any shift in size of the protein in the His channel when incubated with CD2718_{ΔN26}, and there was no visible difference in the

Strep channel signal intensity between the CD2718_{ΔN26} treated and untreated samples (Figure 7.16B).

A



B

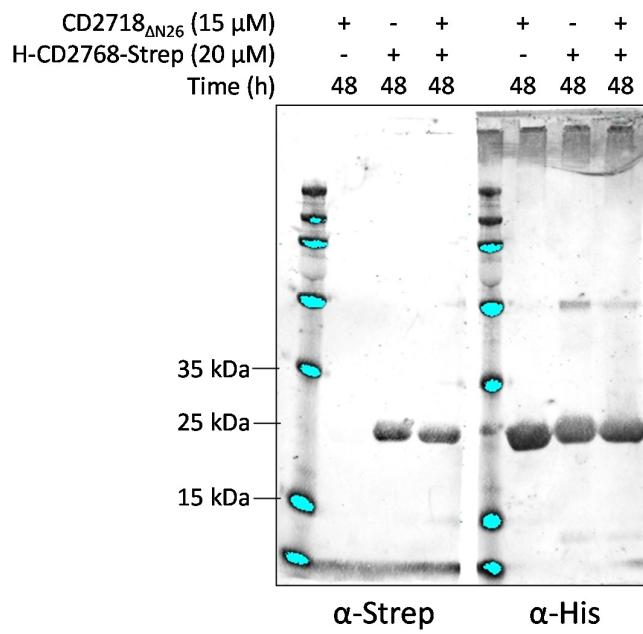


Figure 7.16: Cleavage assay attempts with H-CD2768-Strep protein

Twenty µM H-CD2768-Strep was incubated with 15 µM CD2718_{ΔN26} protein in FRET buffer at 37 °C for 48 hours, after which the reaction products were analysed by SDS-PAGE. A. Two colour anti-His (red) and anti-HA (green)

western blot. **B.** The same samples performed as individual blots using the antibodies indicated.

7.5 Substrate-CwpV fusion proteins

In order to improve the detection of cleavage events *in vitro*, we sought to increase the size of the cleaved C-terminal fragment such that both cleavage products would be an appropriate size to visualise by SDS-PAGE separation. To do this, we used an CD0183-CwpV fusion construct obtained from the Fairweather lab at Imperial College London. CwpV is a cell wall protein which contains C-terminal repeat domains. Five antigenically distinct CwpV repeat types are known: *C. difficile* 630 expresses CwpV with Type I repeats, while the ribotype 027 *C. difficile* strain R20352 expresses CwpV with Type II repeats (Reynolds *et al.*, 2011). Previously, two Type II CwpV repeats were cloned by the Fairweather lab at the C-terminus of the full length CD0183 for overexpression in *C. difficile*. They also cloned this construct into a pACYC vector to create the plasmid pJPK006 for expression of this fusion protein in *E. coli*, resulting in a 52.9 kDa H-CD0183-CwpV protein with an N-terminal His tag.

Highly purified H-CD0183-CwpV produced by the Fairweather lab was provided for our use in developing the whole protein cleavage assay. If CD2718-mediated cleavage of this fusion protein were to occur, a 33.4 kDa N-terminal fragment would be detectable by anti-His and anti-CD0183 antibodies, and a 19.5 kDa C-terminal fragment would be detectable with anti-CwpV antibody (Figure 7.17A). In addition to the H-CD0183-CwpV protein provided, the pJPK006 plasmid encoding the H-CD0183-CwpV protein was used as a template for the cloning of four other substrate-CwpV fusion proteins, listed in Table 7.3. Fifteen kDa of the C-terminal portion of the CD2537, CD2768, CD2831, and CD3392 substrates were cloned in front of the type II CwpV repeats, such that sortase-mediated cleavage would result in a 15 kDa His-reacting fragment in addition to the 20 kDa CwpV reacting fragment (Figure 7.17B).

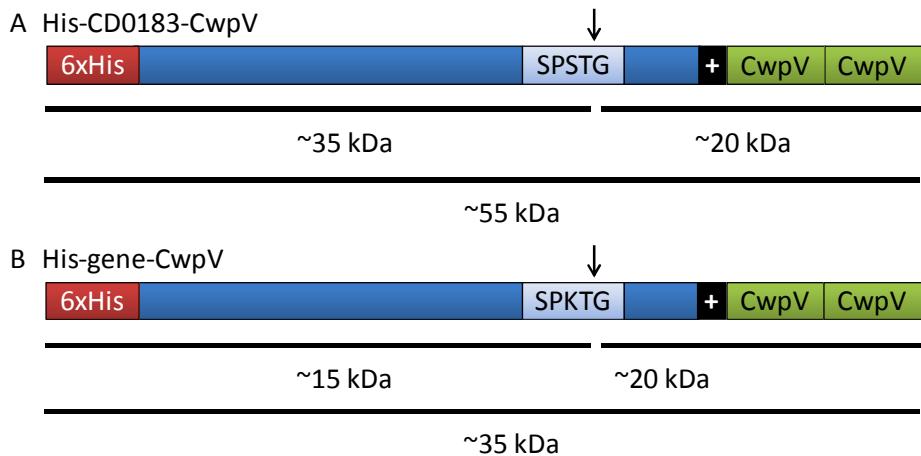


Figure 7.17: Substrate-CwpV fusion protein design

A. Schematic of fusion protein containing CD0183 in its entirety with two CwpV Type II repeats cloned to its C-terminus. **B.** Schematic of other sortase substrate-CwpV fusion proteins. The sizes of the predicted cleavage products are indicated below each protein.

Table 7.3: Expected sizes of cloned substrate-CwpV fusion proteins

Construct	Protein	Residues	Expected	size	Cleavage product sizes (kDa)	
		cloned	(kDa)		N-terminal	C-terminal
pEHD028	H-2537-					
	CwpV	450-613	36.1		15.3	20.8
pEHD030	H-2768-					
	CwpV	123-235	34.2		14.5	19.7
pEHD031	H-2831-					
	CwpV	811-972	35.2		15.2	20.0
pEHD033	H-3392-	862	-			
	CwpV	1015	37.3		16.8	20.5

7.5.1 Expression testing

The constructs listed in Table 7.3 were transformed into Rosetta(DE3) *E. coli*, and expression was induced in 10 ml cultures before harvesting the cells and probing SDS-

PAGE separated whole cell lysates with anti-His and anti-CwpV antibody (Figure 7.18). Of the four constructs tested, only one produced a protein band that reacted with both antibodies: the induced cell lysate of pEHD033 (H-CD3392-CwpV) had a 50 kDa band that reacted strongly with anti-CwpV (Figure 7.18A) and weakly with anti-His (Figure 7.18B). A 40 kDa and a 20 kDa band that reacted with anti-CwpV was observed for the induced cell lysates of pEHD028 (H-CD2537-CwpV) and pEHD031 (H-CD2831-CwpV), respectively, but these bands did not react with the anti-His antibody. The lysate of the pEHD030 construct (H-CD2768-CwpV) had several bands that weakly reacted with the anti-CwpV antibody, but none of these co-reacted with the anti-His antibody.

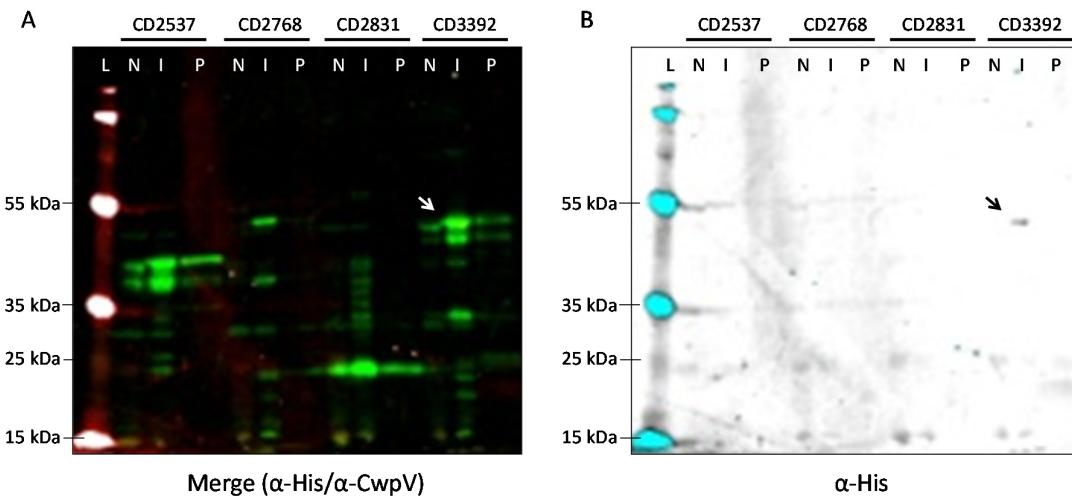


Figure 7.18: Expression testing of CwpV fusion constructs

Western blots of whole-cell lysates from 10 ml cultures of Rosetta(DE3) cells carrying the CwpV fusion protein constructs listed in Table 7.3. N = whole cell lysate of non-induced cultures, I = whole cell lysate of induced cultures, P = insoluble cell lysate. A. Two colour anti-His (red) and anti-CwpV (green) western. B. Anti-His channel alone shown in black and white. An arrow indicates a 50 kDa protein predicted to be H-CD3392-CwpV.

7.5.2 Purification

The H-CD3392-CwpV protein was purified using Ni-NTA agarose from 200 ml of Rosetta(DE3)/pEHD033 cultures as described previously, and the resulting eluted protein separated by SDS-PAGE (Figure 7.19). The same 50 kDa band observed in Figure 7.18 bound reacted with both the anti-His and anti-CwpV antibodies (Figure 7.19), and was

confirmed as H-CD3392-CwpV by MALDI fingerprinting (Figure 7.20). The identity of the purified 55 kDa H-CD0183-CwpV protein kindly provided by the Fairweather lab was also confirmed by MALDI fingerprinting (Figure 7.20).

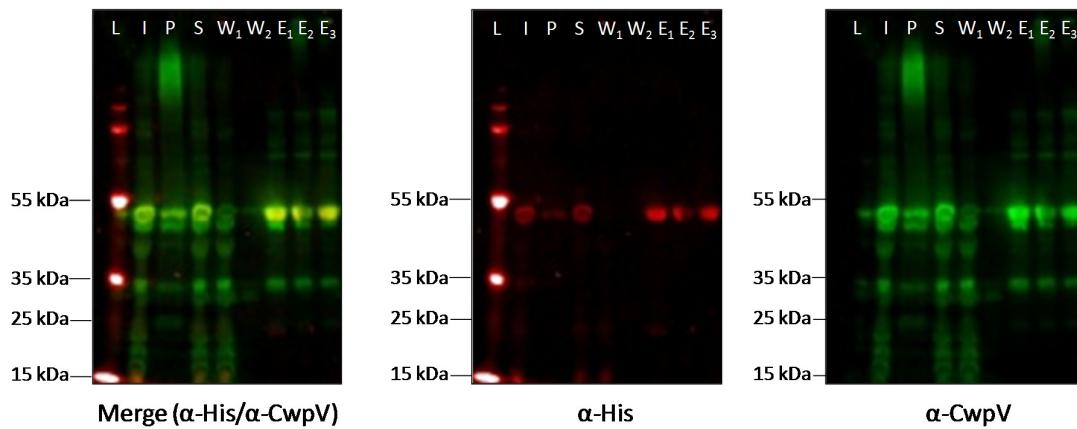


Figure 7.19: Purification of H-CD3392-CwpV protein

Two colour anti-His and anti-CwpV western blots of expression from 200 ml induced Rosetta(DE3)/pEHD033. The individual channels, and the merge of the two are displayed: red corresponds to the anti-His channel (**middle**), green corresponds to the anti-CwpV channel (**right**), and yellow on the merge image indicates where both antibodies bound (**left**). The 50 kDa predicted H-CD3392-CwpV protein reacted with both antibodies. L= ladder, I = induced whole cell lysate, P = insoluble pellet, S = soluble cell lysate applied to nickel column, W = nickel column washes, E = elution from nickel with 250 mM imidazole.

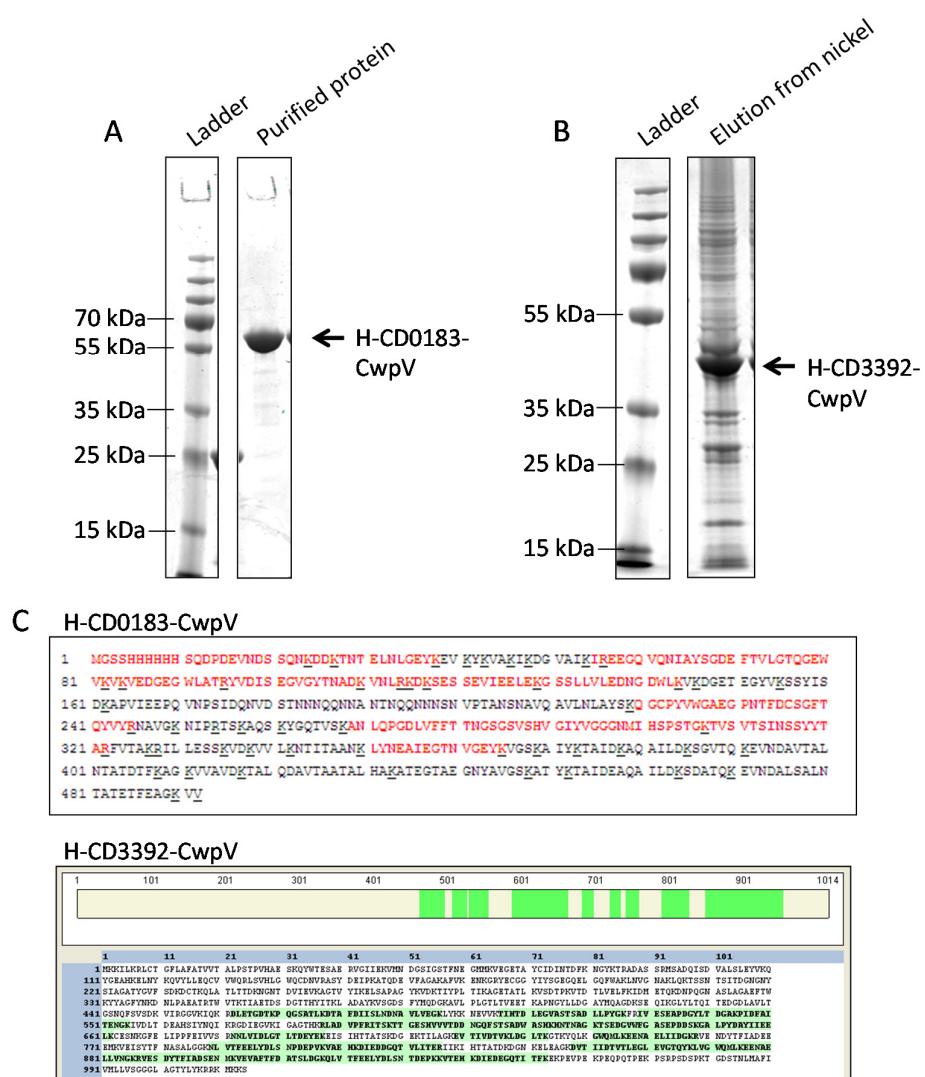


Figure 7.20: Confirmation of CwpV fusion protein identity

Purified substrate-CwpV fusion proteins were separated by SDS-PAGE and stained with Coomassie. The 55 kDa protein indicated with an arrow in **A**, and the 40 kDa protein indicated with an arrow in **B**, were confirmed as H-CD0183-CwpV and H-CD3392-CwpV, respectively, by MALDI fingerprinting (**C**).

7.5.3 Testing H-CD3392-CwpV fusion protein

As before, 15 μ M of CD2718 $_{\Delta N26}$ was incubated with 20 μ M H-CD3392-CwpV in FRET buffer at 37 °C for 48 hours before separating reaction products by SDS-PAGE (Figure 7.21). Because the H-CD3392-CwpV protein migrates closer to 50 kDa than its expected size of 37 kDa, it was unknown what sizes any cleavage products might be. However, in either the anti-His or the anti-CwpV channel of the two colour western, there were no difference

in the banding patterns between the H-CD3392-CwpV incubated with or without CD2718_{ΔN26} (Figure 7.21). There was a lot of non-specific binding of the anti-CwpV antibody to the contaminant proteins in the H-CD3392-CwpV sample. This may have contributed to the weak binding of the anti-His antibody, which hindered the detection of any potential N-terminal cleavage products.

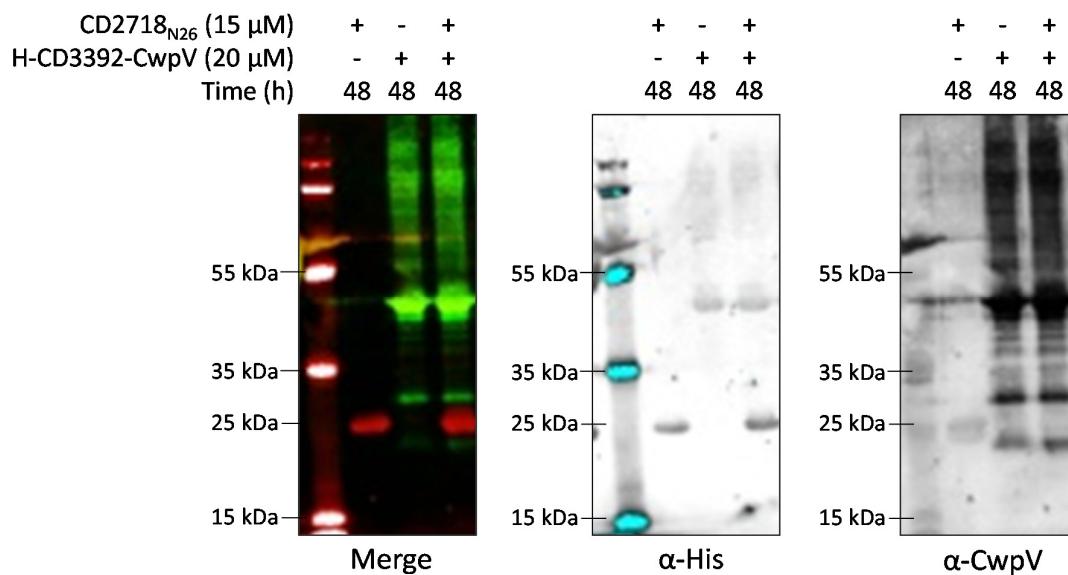


Figure 7.21: H-CD3392-CwpV cleavage assay attempt

Twenty µM H-CD3392-CwpV was incubated with 15 µM CD2718_{ΔN26} protein in FRET buffer at 37 °C for 48 hours, after which the reaction products were analysed by SDS-PAGE. **Left:** Two colour anti-His (red) and anti-CwpV (green) western blot. **Middle:** His channel alone. **Right:** CwpV channel alone.

7.5.4 Testing the H-CD0183-CwpV fusion protein

As the purity of the H-CD3392-CwpV protein sample impeded potential cleavage product detection, it was thought that better success could be obtained with testing the H-CD0183-CwpV fusion protein kindly provided by the Fairweather lab. This protein appeared relatively pure by Coomassie staining (Figure 7.20A) compared to the other recombinant substrate proteins purified in this chapter. Despite the protein containing the weakly cleaved SPSTG sequon, the improved purity was predicted to reduce the non-specific antibody binding that was observed with the other purified substrates and facilitate detection of any cleavage. In addition to the H-CD0183-CwpV protein, the

Fairweather lab kindly provided anti-CD0183 antibody to help corroborate any cleavage products that react with the anti-His.

Fifteen μM $\text{CD2718}_{\Delta\text{N26}}$ and 20 μM H-CD0183-CwpV were incubated together at 37 °C for 48 hours, after which they were analysed by SDS-PAGE and immunoblotting with anti-His, anti-0183, and anti-CwpV antibodies (Figure 7.22). $\text{CD2718}_{\Delta\text{N26}}$ and H-CD0183-CwpV were incubated alone as controls for protein degradation over the experiment time period. $\text{CD2718}_{\Delta\text{N26}}$ and H-CD0183-CwpV together at $t=0$ served as a control for cleavage. Two potential cleavage products, a band at 40-50 kDa and another at 20 kDa, were identified using the anti-His, the anti-CD0183, and the anti-CwpV antibodies (Figure 7.21). The 40-50 kDa band reacted with both the anti-His and anti-CD0183 antibodies, and the 30 kDa band reacted with the anti-CwpV antibody. Though there was non-specific binding of all antibodies in the samples, these bands (indicated with arrows in Figure 7.22) were not observed in the control lanes, and the observed sizes of the cleavage products matched the expected size for both the N-terminal and C-terminal cleavage product of the fusion protein (Figure 7.17A). These faint cleavage products were consistent in size and reproduced in three further independent biological replicates (Appendix E).

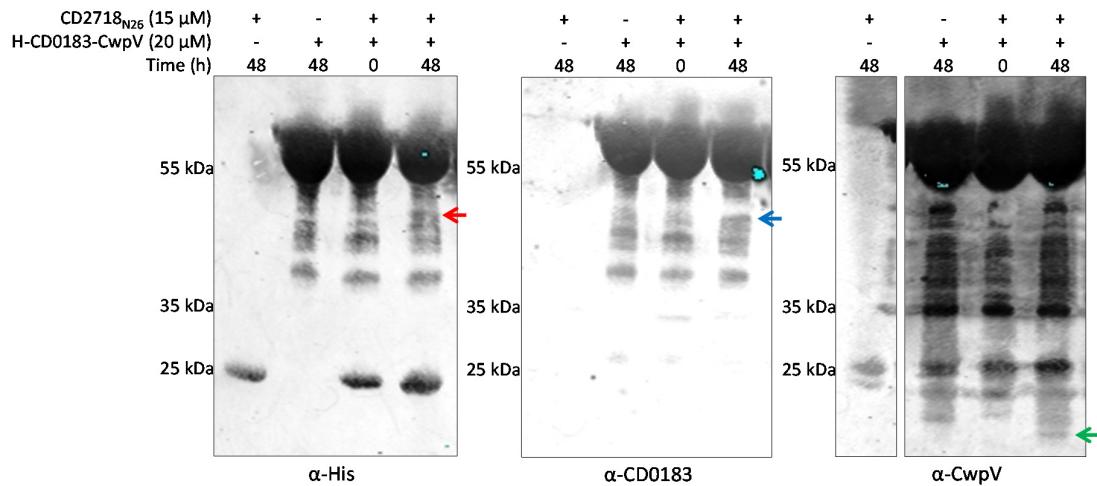


Figure 7.22: Cleavage of H-CD0183-CwpV by $\text{CD2718}_{\Delta\text{N26}}$

Western blots of H-CD0183-CwpV following incubation at 37 °C with $\text{CD2718}_{\Delta\text{N26}}$. The reaction at $t=0$, and the proteins incubated over time served as controls. **Left:** Anti-His western blot showing faint cleavage of H-CD0183-CwpV by $\text{CD2718}_{\Delta\text{N26}}$. A band of approximately 40-50 kDa can be seen in the test well (indicated with a red arrow), but not in either of the controls. **Middle:** Anti-0183 western blot detected the same 40-50 kDa band

(indicated with a blue arrow) observed in the anti-His blot. This band was absent in the control lanes. **Right:** Anti-CwpV western blot showing faint cleavage of H-CD0183-CwpV by CD2718_{ΔN26}. A band of approximately 20 kDa can be seen in the test well (indicated with a green arrow), but not in any of the controls.

7.6 Discussion

In this chapter, we sought to develop an *in vitro* whole protein cleavage assay for the *C. difficile* sortase. In Chapter 5, we demonstrated that the recombinantly expressed sortase, CD2718_{ΔN26}, cleaves fluorescently labelled peptides containing (S/P)PXTG sequences. This sequon is found within seven predicted sortase substrate proteins in *C. difficile* strain 630, but it remained unknown whether these proteins are cleaved by the sortase at these sites. It was also hypothesised that the cleavage efficiency of CD2718 would be markedly improved with a protein substrate rather than a short peptide substrate. The cleavage rates of the FRET peptides by CD2718 improved at least two-fold when the two native residues downstream of the several of the (S/P)PXTG sequences were included, indicating that the region surrounding this conserved five residue sequon may contribute to sortase substrate recognition.

A variety of recombinant sortase substrate proteins were expressed to determine whether they are cleaved by CD2718_{ΔN26} *in vitro*. These proteins were cloned such that they were tagged at both the N- and C-terminus to facilitate the detection of cleavage by SDS-PAGE. Strep and HA tags were tested at the C-terminus, but these tags proved to be inconsistent in their use. Both tags weakly bound their respective antibodies, requiring the use of high concentrations of antisera. Despite using up to a 1:100 dilution for the anti-HA antibody, binding was neither consistent nor sensitive enough to allow for visualisation of a cleavage event. When the anti-Strep antibody was used at a 1:100 dilution, detection of the Strep tag was improved, but the binding efficiency of the anti-His antibody was greatly reduced.

Initially, the truncated recombinant sortase substrates utilised their native C-termini, resulting in predicted C-terminal cleavage products that were only three or four kDa in length. This proved to make observation of any cleavage difficult, so recombinant sortase

substrate fusion proteins were designed incorporating two C-terminal CwpV Type II repeats. The sizes of their expected cleavage products following CD2718-mediated cleavage between the threonine and glycine residues would result in two easily distinguishable large protein fragments. However, it was difficult to detect expression of the substrate-CwpV fusion proteins. Only the H-CD0183-CwpV fusion protein, provided by the Fairweather lab, provided evidence suggesting a low level of cleavage by CD2718_{ΔN26}. Following a cleavage event, the N-terminal fragment of H-CD0183-CwpV was predicted to be approximately 35 kDa in size and to react with both the anti-His and anti-CD0183 antibodies. Though slightly larger than 35 kDa, a faint but reproducible 40-50 kDa band was observed in western blots with these two antibodies. A band corresponding to the predicted 20 kDa C-terminal cleavage product of H-CD0183-CwpV was also observed in multiple replicates of the cleavage assay using the anti-CwpV antibody.

These results are promising and suggest that H-CD0183-CwpV may be cleaved to some extent by CD2718_{ΔN26}. However, the efficiency of this reaction would need to be improved in order to verify that cleavage of H-CD0183-CwpV is occurring. As reported here, the predicted cleavage products are only faintly detectable by western blotting. Ideally, both cleavage products should be visible by Coomassie staining to facilitate identification of both bands by mass spectrometry, or by analysing any suspected C-terminal cleavage product by N-terminal sequencing. As demonstrated in Chapter 5, the sortase recognition sequence found in CD0183, SPSTG, is cleaved with the lowest efficiency in the FRET assay, so it may be unsurprising that only a small amount of the H-CD0183-CwpV protein appears to be cleaved in this chapter. It is possible that the other (S/P)PXTG sortase sequons would be cleaved more readily in CwpV-fusion proteins than SPSTG, similar to what is observed using the FRET peptides. We attempted to demonstrate this with an H-CD3392-CwpV fusion, but unfortunately were unable to detect any cleavage by CD2718_{ΔN26}. We could not test others as expression of the other CwpV fusion proteins was unsuccessful. To potentially improve the detection of H-CD0183-CwpV cleavage, the SPSTG sequon could be modified to one of the other (S/P)PXTG sequences by site- directed mutagenesis.

In Chapter 5, we observed that the *in vitro* activity level of CD2718_{ΔN26} is low. In other bacteria, both the sortase and its substrates are membrane-bound at the start of the sorting reaction, so this may be the optimal arrangement for sortase recognition of its

substrates. In addition, all of the recombinant substrate proteins expressed in this chapter are truncated versions of the full length protein; it is possible they are folding incorrectly such that this hinders their recognition by CD2718. Due to these potential barriers to enzyme efficiency *in vitro*, the cleavage assay attempted in this chapter may not be the ideal approach to demonstrate CD2718-mediated cleavage of the predicted substrate proteins.

8 Discussion and conclusions

C. difficile is a widespread and costly public health problem. It is uniquely adapted to thrive in healthcare settings: its spores facilitate persistence in the environment and transmission between hospital patients, many of whom are elderly and on antibiotics that render them particularly susceptible to the disease. CDI places huge burdens on healthcare systems due to the high cost of infection control measures, and the difficulties and lengthy hospital stays associated with treating recurrent infection. CDI is invariably associated with the disruption of the normal intestinal microflora by the administration of broad spectrum antibiotics, thus there is a pressing need to develop therapies that selectively target *C. difficile* while leaving the intestinal microflora intact.

We hypothesised that the predicted *C. difficile* sortase enzyme, CD2718, could be such a target for antimicrobial development. Sortases are often essential for virulence in Gram-positive organisms; inactivation of sortases reduces colonisation in mice (Mazmanian *et al.*, 2000, Bierne *et al.*, 2002, Weiss *et al.*, 2004, Cheng *et al.*, 2009), and decreases adhesion and invasion *in vitro* (Mazmanian *et al.*, 2000, Bolken *et al.*, 2001, Garandeau *et al.*, 2002, Kharat & Tomasz, 2003, Mandlik *et al.*, 2007). Sortases and their substrates are considered promising targets for the development of new anti-infective compounds for many Gram-positive organisms (Garandeau *et al.*, 2002, Jonsson *et al.*, 2003, Kharat & Tomasz, 2003). In this thesis, we hoped to investigate the function of the *C. difficile* sortase to determine whether inhibiting CD2718 could prove to be an anti-infective strategy for targeting CDI.

8.1 The *C. difficile* sortase

In Chapter 3, investigation of the *C. difficile* 630 genome identified a single intact sortase-like enzyme, CD2718, and seven predicted sortase substrate proteins with an (S/P)PXTG motif. The predicted *C. difficile* sortase substrate proteins appear typically of sortase-anchored proteins, and include potential adhesins, a 5' nucleotidase, and cell wall hydrolases, most of which are conserved across *C. difficile* strains. RT-PCR confirmed that all seven proteins and CD2718 are transcribed during *in vitro* growth, suggesting they may be functional proteins in *C. difficile*. Unfortunately, in Chapter 4 we were unable to generate a CD2718 mutant, and overexpression of CD2718 in *C. difficile* had no apparent phenotype in the assays we performed. Because of this, we focused our attention on the

in vitro characterisation of recombinantly expressed CD2718_{ΔN26} protein. In Chapter 5, we developed a FRET-based assay to assess recombinant CD2718_{ΔN26} protein for enzymatic activity against fluorescently labelled peptides. We demonstrate that CD2718_{ΔN26} exhibits sortase-like activity by recognising and cleaving the (S/P)PXTG motif between the threonine and glycine residues, and that this cleavage is dependent on a single cysteine residue at position 209.

Based on our kinetic analysis of the enzyme (Section 5.3.5), the *in vitro* hydrolytic activity of the *C. difficile* sortase is low compared to that of the SrtA from *S. aureus* and *S. pyogenes* (Huang *et al.*, 2003, Race *et al.*, 2009), but similar to that of SrtA from *B. anthracis* (Weiner *et al.*, 2010). In the proposed reverse protonation mechanism of sortase, only a very small percentage (0.06%) of the SaSrtA enzyme is enzymatically active reverse-protonated state, whereby the residue with the lower pK_a is protonated (histidine) and the residue with the higher pK_a is deprotonated (cysteine) (Frankel *et al.*, 2005). Weiner *et al.* (2010) hypothesised that an even smaller percentage of the BaSrtA enzyme was in its active charged state as compared with SaSrtA due to differences in the pK_a of their respective catalytically essential histidine residues. The low *in vitro* activity level of CD2718 may be due to differences in the percentage of catalytically active enzyme present, but this cannot be confirmed without a solved crystal structure and similar pK_a measurements for residues in the CD2718_{ΔN26} active site. Such experiments would also contribute to an understanding of the mechanism of CD2718 activity, and allow comparisons of catalytic mechanisms to other sortases.

In addition to our kinetic analysis of CD2718_{ΔN26} activity, we also attempted to investigate whether CD2718_{ΔN26} was capable of catalysing a transpeptidation reaction *in vitro* (Section 5.3.2). With SaSrtA and LPETG peptides, hydroxylamine serves as a nucleophile *in vitro*, yielding LPET-hydroxamate products, indicating that hydroxylamine attacks and resolves the acyl-enzyme intermediate (Mazmanian *et al.*, 1999, Ton-That *et al.*, 2000). Addition of peptidoglycan surrogates which mimic the pentaglycine cross-bridge, including chains of one to five glycine residues, increased the rate of sortase cleavage of LPETG peptides and resulted in LPET-Gly_N (where N = 1-5) reaction products (Mazmanian *et al.*, 1999, Ton-That *et al.*, 1999, Ton-That *et al.*, 2000). Addition of the predicted transpeptidation substrate in the *C. difficile* cell wall, based on the predominant

peptidoglycan cross-bridge structure of 2,6-diaminopimelic acid (2,6-Dap) (Peltier *et al.*, 2011), had no observable effect on reaction rate in our FRET assay. This suggests that CD2718_{ΔN26} may not catalyse transpeptidation *in vitro*, similar to what has been observed for the BaSrtA, though this was not confirmed by reverse-phase high-performance liquid chromatography analysis of the FRET reaction products. Transpeptidation in the presence of hydroxylamine or *m*-Dap is not observed for BaSrtA *in vitro*, suggesting that BaSrtA requires additional protein components or larger portions of the lipid II molecule bearing *m*-Dap to be completely active (Weiner *et al.*, 2010). The same may be true for CD2718_{ΔN26} and it is likely that conditions of this assay are not optimal for transpeptidation to occur.

In addition to the cleaved peptides in the FRET assay, this study also provides evidence that recombinant CD2718_{ΔN26} cleaves whole proteins containing an (S/P)PXTG motif *in vitro* (Section 7.5.4). For this assay we used a recombinant fusion protein of the predicted sortase substrates CD0183 with two CwpV type II repeats at the C-terminus (kindly provided by the Fairweather lab), and observed reproducible bands by western blotting suggestive of cleavage at the SPSTG site. We utilised several different recombinant substrates in this assay, but the CD0183 fusion protein was the only one that resulted in reproducible bands that matched the sizes of the expected products following CD2718-mediated cleavage. However, due to the low level of potential cleavage observed we were unable to confirm this experimentally by LC-MS or Edams degradation of the C-terminal cleavage product. Prior to improving the cleavage efficiency of the assay, one way to test whether the appearance of these bands is dependent on CD2718-mediated cleavage would be to repeat the assay in the presence of MTSET and the other inhibitors of CD2718, as well as with the mutant CD2718_{ΔN26,C209A} enzyme. Unfortunately we were unable to perform these additional tests due to limited recombinant CD0183 quantities. Additionally, the SPSTG site within the CD0183 fusion protein could be altered to confirm that this protein is being cleaved at the SPSTG site.

8.2 Sortase inhibitors and CDI

In this study, we describe the identification of small-molecule inhibitors of CD2718_{ΔN26} using the proprietary *Leadbuilder* virtual screening method developed by Domainex, Ltd (Section 5.3.6). All 62 of compounds selected for screening contain α-ketoamide groups, which are known as reversible inhibitors of both cysteine and serine proteases from

diverse pathogens, including hepatitis C virus (Arasappan *et al.*, 2005), *Trypanosoma cruzi* (Choe *et al.*, 2005) and *Plasmodium falciparum* (Lee *et al.*, 2003). We found that more than half of these compounds reduced the amount of the *d*-PVPPKTGDS-e peptide cleaved by CD2718_{ΔN26} by at least 25% at a concentration of 100 μM, and that ten compounds reduced the amount of cleavage by greater than 50% at this concentration (Table 5.2). LSHTM 52, the compound with the lowest IC₅₀ value (Figure 5.18), as well as seven of the other top ten compounds including LSHTM 13, contains an indole pyruvamide, in which an indole group is adjacent to the α-ketoamide. Further analysis of structural similarities such as these between the hit compounds will allow a better understanding of what functional groups adjacent to the α-ketoamide contribute to inhibition of CD2718_{ΔN26} cleavage *in vitro*, and should lead to a refinement of inhibitor design and improved solubility of future compounds. Solving the crystal structure of CD2718_{ΔN26} interacting with one of the hit compounds would also greatly facilitate inhibitor optimisation, as it would reveal how the molecule is binding the active site.

The results presented in this study represent the initial steps in CD2718 inhibitor design, but more investigation is required to fully understand the mechanism of inhibition. The compounds identified in this study are predicted to be reversible inhibitors by nature of their α-ketoamide grouping, though this was not examined due to time and resource constraints, and the lack of crystal structure for CD2718. Reversible inhibition can be described quantitatively in terms of the inhibitor's binding to the enzyme and to the enzyme-substrate complex, but such calculations require biochemical analysis of these interactions beyond the capabilities of our lab. We also tested whether these compounds had any effect on *C. difficile* growth, and found this was not the case. Several inhibitors of SaSrtA have been described that also inhibit *S. aureus* growth to some extent *in vitro*, including several naturally derived alkaloid compounds (Kim *et al.*, 2004, Jang *et al.*, 2007). However, inhibition of bacterial growth is not considered vital in the development of sortase-based drug therapies. For both *Staphylococcus* and *Bacillus*, sortase inhibitors without bactericidal activity are considered promising for development as anti-infectives. This includes several diarylacrylonitrile inhibitors that are effective at reducing mortality in *S. aureus* challenged mice (Oh *et al.*, 2010), despite having no effect on staphylococcal growth (Oh *et al.*, 2006).

8.3 Substrates and the cell wall

Prior to this study, the *C. difficile* sortase recognition sequence was predicted to be (S/P)PXTG and also include NVQTG (Pallen *et al.*, 2001, Tulli *et al.*, 2013). Our analysis of the *C. difficile* 630 genome revealed eight putative surface proteins with these sequences that met the criteria of sortase substrates; seven containing the (S/P)PXTG motif, and only one containing NVQTG, the collagen binding protein CbpA. Our *in vitro* FRET assay developed in Chapter 5 revealed that all four variations of the (S/P)PXTG motif observed (SPKTG, SPQTG, SPSTG, and PPKTG) were cleaved by CD2718_{ΔN26}, but that NVQTG was not. In addition, the protein cleavage assay described in Chapter 7 provides evidence that CD2718 cleaves CD0183 protein *in vitro*. While this does not prove that the (S/P)PXTG containing proteins are true sortase substrates, this suggests that the (S/P)PXTG containing proteins are recognised and processed by CD2718, and that the NVQTG containing CbpA is not.

Further study is needed to demonstrate that these proteins are substrates and that they are indeed covalently anchored to the *C. difficile* cell wall by CD2718. As mentioned earlier, cell localisation studies using antisera generated against the putative substrates would confirm whether they are on the cell surface. Tagging the putative substrates in *C. difficile* would allow identification of their cell wall anchor structure, which is predicted to be the diaminopimelic acid crossbridge in *C. difficile* peptidoglycan (Peltier *et al.*, 2011). In both *S. aureus* and *B. anthracis*, sortase anchored proteins were tagged with a maltose binding protein, prior to digestion of the peptidoglycan and analysis of the resulting fragments by mass spectrometry (Schneewind *et al.*, 1995, Budzik *et al.*, 2008). Such experiments would determine whether these proteins are covalently attached to the cell wall, and support the hypothesis that CD2718 is a functional sortase in *C. difficile*.

8.4 Final conclusions

In conclusion, we demonstrate that the *C. difficile* predicted class B sortase, CD2718, has sortase-like activity *in vitro*. We predicted *C. difficile* to encode seven putative sortase substrates, and have confirmed the CD2718 recognition sequence as (S/P)PXTG through CD2718-mediated cleavage (S/P)PXTG containing peptides. We also have evidence that CD2718 cleaves a recombinant fusion protein containing the predicted *C. difficile* sortase substrate, CD0183. The use of *in silico* approaches such as the *LeadBuilder* method

employed by this study to screen databases of putative small-molecule inhibitors for further analysis has been validated. CD2718 inhibitors such as the ones identified in this study provide a significant step in successfully identifying inhibitor compounds, which can be further refined to enhance their efficacy. We believe that this work is an important first step in the characterisation of the *C. difficile* sortase enzyme, and could contribute towards the development of novel selective therapeutics against CDI.

References

- Aas, J., C.E. Gessert & J.S. Bakken, (2003) Recurrent *Clostridium difficile* colitis: case series involving 18 patients treated with donor stool administered via a nasogastric tube. *Clin. Infect. Dis.* **36**: 580-585.
- Akabas, M.H. & A. Karlin, (1995) Identification of acetylcholine receptor channel-lining residues in the M1 segment of the alpha-subunit. *Biochemistry (Mosc.)*. **34**: 12496-12500.
- Akerlund, T., B. Svenungsson, A. Lagergren & L.G. Burman, (2006) Correlation of disease severity with fecal toxin levels in patients with *Clostridium difficile*-associated diarrhea and distribution of PCR ribotypes and toxin yields *in vitro* of corresponding isolates. *J. Clin. Microbiol.* **44**: 353-358.
- Al-Abed, Y.A., E.A. Gray & N.D. Rothnie, (2010) Outcomes of emergency colectomy for fulminant *Clostridium difficile* colitis. *The surgeon : journal of the Royal Colleges of Surgeons of Edinburgh and Ireland* **8**: 330-333.
- Al-Nassir, W.N., A.K. Sethi, M.M. Nerandzic, G.S. Bobulsky, R.L. Jump & C.J. Donskey, (2008) Comparison of clinical and microbiological response to treatment of *Clostridium difficile*-associated disease with metronidazole and vancomycin. *Clin. Infect. Dis.* **47**: 56-62.
- Anantharaman, V. & L. Aravind, (2003) Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome biology* **4**: R11.
- Arasappan, A., F.G. Njoroge, T.Y. Chan, F. Bennett, S.L. Bogen, K. Chen, H. Gu, L. Hong, E. Jao, Y.T. Liu, R.G. Lovey, T. Parekh, R.E. Pike, P. Pinto, B. Santhanam, S. Venkatraman, H. Vaccaro, H. Wang, X. Yang, Z. Zhu, B. McKittrick, A.K. Saksena, V. Girijavallabhan, J. Pichardo, N. Butkiewicz, R. Ingram, B. Malcolm, A. Prongay, N. Yao, B. Marten, V. Madison, S. Kemp, O. Levy, M. Lim-Wilby, S. Tamura & A.K. Ganguly, (2005) Hepatitis C virus NS3-4A serine protease inhibitors: SAR of P'2 moiety with improved potency. *Bioorg. Med. Chem. Lett.* **15**: 4180-4184.
- Aravind, L. & E.V. Koonin, (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res* **26**: 3746-3752.
- Baban, S.T., S.A. Kuehne, A. Barketi-Klai, S.T. Cartman, M.L. Kelly, K.R. Hardie, I. Kansau, A. Collignon & N.P. Minton, (2013) The role of flagella in *Clostridium difficile* pathogenesis: comparison between a non-epidemic and an epidemic strain. *PLoS One* **8**: e73026.
- Bacci, S., K. Molbak, M.K. Kjeldsen & K.E. Olsen, (2011) Binary toxin and death after *Clostridium difficile* infection. *Emerg. Infect. Dis.* **17**: 976-982.
- Baines, S.D., R. O'Connor, J. Freeman, W.N. Fawley, C. Harmanus, P. Mastrantonio, E.J. Kuijper & M.H. Wilcox, (2008) Emergence of reduced susceptibility to metronidazole in *Clostridium difficile*. *J. Antimicrob. Chemother.* **62**: 1046-1052.
- Baldassarri, L., G. Donelli, M. Cerquetti & P. Mastrantonio, (1991) Capsule-like structures in *Clostridium difficile* strains. *Microbiologica* **14**: 295-300.
- Barken, K.B., S.J. Pamp, L. Yang, M. Gjermansen, J.J. Bertrand, M. Klausen, M. Givskov, C.B. Whitchurch, J.N. Engel & T. Tolker-Nielsen, (2008) Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature multicellular structures in *Pseudomonas aeruginosa* biofilms. *Environmental microbiology* **10**: 2331-2343.
- Barketi-Klai, A., S. Hoys, S. Lambert-Bordes, A. Collignon & I. Kansau, (2011) Role of fibronectin-binding protein A in *Clostridium difficile* intestinal colonization. *J. Med. Microbiol.* **60**: 1155-1161.
- Barnett, T.C. & J.R. Scott, (2002) Differential recognition of surface proteins in *Streptococcus pyogenes* by two sortase gene homologs. *J. Bacteriol.* **184**: 2181-2191.
- Bartlett, J.G., (2002) *Clostridium difficile*-associated Enteric Disease. *Current infectious disease reports* **4**: 477-483.
- Bartlett, J.G., (2008) Historical perspectives on studies of *Clostridium difficile* and *C. difficile* infection. *Clin. Infect. Dis.* **46 Suppl 1**: S4-11.

- Bartlett, J.G., T.W. Chang, M. Gurwith, S.L. Gorbach & A.B. Onderdonk, (1978) Antibiotic-associated pseudomembranous colitis due to toxin-producing Clostridia. *N. Engl. J. Med.* **298**: 531-534.
- Bartlett, J.G. & D.N. Gerdling, (2008) Clinical recognition and diagnosis of *Clostridium difficile* infection. *Clin. Infect. Dis.* **46 Suppl 1**: S12-18.
- Bentley, M.L., E.C. Lamb & D.G. McCafferty, (2008) Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from *Staphylococcus aureus*. *J. Biol. Chem.* **283**: 14762-14771.
- Bierne, H. & P. Cossart, (2007) *Listeria monocytogenes* surface proteins: from genome predictions to function. *Microbiol. Mol. Biol. Rev.* **71**: 377-397.
- Bierne, H., C. Garandeau, M.G. Pucciarelli, C. Sabet, S. Newton, F. Garcia-del Portillo, P. Cossart & A. Charbit, (2004) Sortase B, a new class of sortase in *Listeria monocytogenes*. *J. Bacteriol.* **186**: 1972-1982.
- Bierne, H., S.K. Mazmanian, M. Trost, M.G. Pucciarelli, G. Liu, P. Dehoux, L. Jansch, F. Garcia-del Portillo, O. Schneewind & P. Cossart, (2002) Inactivation of the *srtA* gene in *Listeria monocytogenes* inhibits anchoring of surface proteins and affects virulence. *Mol. Microbiol.* **43**: 869-881.
- Bolanos-Garcia, V.M. & O.R. Davies, (2006) Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochim. Biophys. Acta* **1760**: 1304-1313.
- Bolken, T.C., C.A. Franke, K.F. Jones, G.O. Zeller, C.H. Jones, E.K. Dutton & D.E. Hruby, (2001) Inactivation of the *srtA* gene in *Streptococcus gordonii* inhibits cell wall anchoring of surface proteins and decreases *in vitro* and *in vivo* adhesion. *Infect. Immun.* **69**: 75-80.
- Bomers, M.K., M.A. van Agtmael, H. Luik, C.M. Vandenbroucke-Grauls & Y.M. Smulders, (2014) A detection dog to identify patients with *Clostridium difficile* infection during a hospital outbreak. *J. Infect.*
- Borody, T.J. & A. Khoruts, (2012) Fecal microbiota transplantation and emerging applications. *Nature reviews. Gastroenterology & hepatology* **9**: 88-96.
- Borriello, S.P. & F.E. Barclay, (1986) An *in vitro* model of colonisation resistance to *Clostridium difficile* infection. *J. Med. Microbiol.* **21**: 299-309.
- Borriello, S.P., H.A. Davies & F.E. Barclay, (1988) Detection of fimbriae amongst strains of *Clostridium difficile*. *FEMS Microbiol. Lett.* **49**: 65-67.
- Borriello, S.P., H.A. Davies, S. Kamiya, P.J. Reed & S. Seddon, (1990) Virulence factors of *Clostridium difficile*. *Rev. Infect. Dis.* **12**: S185-S191.
- Bouza, E., (2012) Consequences of *Clostridium difficile* infection: understanding the healthcare burden. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **18 Suppl 6**: 5-12.
- Brouwer, M.S., A.P. Roberts, P. Mullany & E. Allan, (2012) *In silico* analysis of sequenced strains of *Clostridium difficile* reveals a related set of conjugative transposons carrying a variety of accessory genes. *Mobile genetic elements* **2**: 8-12.
- Brouwer, M.S., P.J. Warburton, A.P. Roberts, P. Mullany & E. Allan, (2011) Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of *Clostridium difficile*. *PLoS One* **6**: e23014.
- Budzik, J.M., L.A. Marraffini & O. Schneewind, (2007) Assembly of pili on the surface of *Bacillus cereus* vegetative cells. *Mol. Microbiol.* **66**: 495-510.
- Budzik, J.M., S.Y. Oh & O. Schneewind, (2008) Cell wall anchor structure of BcpA pili in *Bacillus anthracis*. *J. Biol. Chem.* **283**: 36676-36686.
- Budzik, J.M., S.Y. Oh & O. Schneewind, (2009) Sortase D forms the covalent bond that links BcpB to the tip of *Bacillus cereus* pili. *J. Biol. Chem.* **284**: 12989-12997.
- Calabi, E., F. Calabi, A.D. Phillips & N.F. Fairweather, (2002) Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues. *Infect. Immun.* **70**: 5770-5778.

- Calabi, E., S. Ward, B. Wren, T. Paxton, M. Panico, H. Morris, A. Dell, G. Dougan & N. Fairweather, (2001) Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Mol. Microbiol.* **40**: 1187-1199.
- Carling, P., T. Fung, A. Killion, N. Terrin & M. Barza, (2003) Favorable impact of a multidisciplinary antibiotic management program conducted during 7 years. *Infect. Control Hosp. Epidemiol.* **24**: 699-706.
- Chang, C., A. Mandlik, A. Das & H. Ton-That, (2011) Cell surface display of minor pilin adhesins in the form of a simple heterodimeric assembly in *Corynebacterium diphtheriae*. *Mol. Microbiol.* **79**: 1236-1247.
- Chen, A.G., N. Sudarsan & R.R. Breaker, (2011) Mechanism for gene control by a natural allosteric group I ribozyme. *RNA* **17**: 1967-1972.
- Chen, F., B. Liu, D. Wang, L. Wang, X. Deng, C. Bi, Y. Xiong, Q. Wu, Y. Cui, Y. Zhang, X. Li, Y. Wang, B. Liu & Y. Cao, (2013) Role of sortase A in the pathogenesis of *Staphylococcus aureus*-induced mastitis in mice. *FEMS Microbiol. Lett.*
- Cheng, A.G., H.K. Kim, M.L. Burts, T. Krausz, O. Schneewind & D.M. Missiakas, (2009) Genetic requirements for *Staphylococcus aureus* abscess formation and persistence in host tissues. *FASEB J.* **23**: 3393-3404.
- Chiang, S.L., R.K. Taylor, M. Koomey & J.J. Mekalanos, (1995) Single amino acid substitutions in the N-terminus of *Vibrio cholerae* TcpA affect colonization, autoagglutination, and serum resistance. *Mol. Microbiol.* **17**: 1133-1142.
- Choe, Y., L.S. Brinen, M.S. Price, J.C. Engel, M. Lange, C. Grisostomi, S.G. Weston, P.V. Pallai, H. Cheng, L.W. Hardy, D.S. Hartsough, M. McMakin, R.F. Tilton, C.M. Baldino & C.S. Craik, (2005) Development of alpha-keto-based inhibitors of cruzain, a cysteine protease implicated in Chagas disease. *Bioorg. Med. Chem.* **13**: 2141-2156.
- Clabots, C.R., S. Johnson, K.M. Bettin, P.A. Mathie, M.E. Mulligan, D.R. Schaberg, L.R. Peterson & D.N. Gerding, (1993) Development of a rapid and efficient restriction endonuclease analysis typing system for *Clostridium difficile* and correlation with other typing systems. *J. Clin. Microbiol.* **31**: 1870-1875.
- Clabots, C.R., S. Johnson, M.M. Olson, L.R. Peterson & D.N. Gerding, (1992) Acquisition of *Clostridium difficile* by hospitalized patients: evidence for colonized new admissions as a source of infection. *J. Infect. Dis.* **166**: 561-567.
- Cleary, R.K., (1998) *Clostridium difficile*-associated diarrhea and colitis - Clinical manifestations, diagnosis and treatment. *Dis. Colon Rectum* **41**: 1435-1449.
- Cleland, W.W., (1964) Dithiothreitol, a New Protective Reagent for Sh Groups. *Biochemistry (Mosc).* **3**: 480-482.
- Cohen, L.E., C.J. McNeill & R.F. Wells, (1973) Clindamycin-associated colitis. *JAMA* **223**: 1379-1380.
- Cohen, S.H., D.N. Gerding, S. Johnson, C.P. Kelly, V.G. Loo, L.C. McDonald, J. Pepin, M.H. Wilcox, A. Society for Healthcare Epidemiology of & A. Infectious Diseases Society of, (2010) Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA). *Infect. Control Hosp. Epidemiol.* **31**: 431-455.
- Comfort, D. & R.T. Clubb, (2004) A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. *Infect. Immun.* **72**: 2710-2722.
- Connolly, K.M., B.T. Smith, R. Pilpa, U. Ilangovalan, M.E. Jung & R.T. Clubb, (2003) Sortase from *Staphylococcus aureus* does not contain a thiolate-imidazolium ion pair in its active site. *J. Biol. Chem.* **278**: 34061-34065.
- Cornely, O.A., M.A. Miller, T.J. Louie, D.W. Crook & S.L. Gorbach, (2012) Treatment of first recurrence of *Clostridium difficile* infection: fidaxomicin versus vancomycin. *Clin. Infect. Dis.* **55 Suppl 2**: S154-161.
- Cossart, P. & R. Jonquieres, (2000) Sortase, a universal target for therapeutic agents against gram-positive bacteria? *Proc. Natl. Acad. Sci. U. S. A.* **97**: 5013-5015.

- Crobach, M.J., O.M. Dekkers, M.H. Wilcox & E.J. Kuijper, (2009) European Society of Clinical Microbiology and Infectious Diseases (ESCMID): data review and recommendations for diagnosing *Clostridium difficile*-infection (CDI). *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **15**: 1053-1066.
- Crook, D.W., A.S. Walker, Y. Kean, K. Weiss, O.A. Cornely, M.A. Miller, R. Esposito, T.J. Louie, N.E. Stoesser, B.C. Young, B.J. Angus, S.L. Gorbach, T.E. Peto & T. Study, (2012) Fidaxomicin versus vancomycin for *Clostridium difficile* infection: meta-analysis of pivotal randomized controlled trials. *Clin. Infect. Dis.* **55 Suppl 2**: S93-103.
- Cunningham, M.W., (2000) Pathogenesis of group A streptococcal infections. *Clin. Microbiol. Rev.* **13**: 470-511.
- Curry, S.R., C.A. Muto, J.L. Schlackman, A.W. Pasculle, K.A. Shutt, J.W. Marsh & L.H. Harrison, (2013) Use of multilocus variable number of tandem repeats analysis genotyping to determine the role of asymptomatic carriers in *Clostridium difficile* transmission. *Clin. Infect. Dis.* **57**: 1094-1102.
- Dailey, D.C., A. Kaiser & R.H. Schloemer, (1987) Factors influencing the phagocytosis of *Clostridium difficile* by human polymorphonuclear leukocytes. *Infect. Immun.* **55**: 1541-1546.
- Davies, H.A. & S.P. Borriello, (1990) Detection of capsule in strains of *Clostridium difficile* of varying virulence and toxigenicity. *Microb. Pathog.* **9**: 141-146.
- Dawson, L.F., E.H. Donahue, S.T. Cartman, R.H. Barton, J. Bundy, R. Mc너ney, N.P. Minton & B.W. Wren, (2011a) The analysis of para-cresol production and tolerance in *Clostridium difficile* 027 and 012 strains. *BMC microbiology* **11**: 86.
- Dawson, L.F., R.A. Stabler & B.W. Wren, (2008) Assessing the role of p-cresol tolerance in *Clostridium difficile*. *J. Med. Microbiol.* **57**: 745-749.
- Dawson, L.F., E. Valiente, E.H. Donahue, G. Birchenough & B.W. Wren, (2011b) Hypervirulent *Clostridium difficile* PCR-ribotypes exhibit resistance to widely used disinfectants. *PLoS One* **6**: e25754.
- Dawson, L.F., E. Valiente, A. Faulds-Pain, E.H. Donahue & B.W. Wren, (2012) Characterisation of *Clostridium difficile* biofilm formation, a role for Spo0A. *PLoS One* **7**: e50527.
- Dawson, L.F., E. Valiente & B.W. Wren, (2009) *Clostridium difficile* - A continually evolving and problematic pathogen. *Infection, Genetics and Evolution* **9**: 1410-1417.
- Deakin, L.J., S. Clare, R.P. Fagan, L.F. Dawson, D.J. Pickard, M.R. West, B.W. Wren, N.F. Fairweather, G. Dougan & T.D. Lawley, (2012) The *Clostridium difficile* spo0A gene is a persistence and transmission factor. *Infect. Immun.* **80**: 2704-2711.
- Deivanayagam, C.C., R.L. Rich, M. Carson, R.T. Owens, S. Danthuluri, T. Bice, M. Hook & S.V. Narayana, (2000) Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* **8**: 67-78.
- Dembek, M., C.B. Reynolds & N.F. Fairweather, (2012) *Clostridium difficile* cell wall protein CwpV undergoes enzyme-independent intramolecular autoproteolysis. *J. Biol. Chem.* **287**: 1538-1544.
- Dessen, A., (2004) A new catalytic dyad regulates anchoring of molecules to the Gram-positive cell wall by sortases. *Structure* **12**: 6-7.
- Dhar, G., K.F. Faull & O. Schneewind, (2000) Anchor structure of cell wall surface proteins in *Listeria monocytogenes*. *Biochemistry (Mosc)*. **39**: 3725-3733.
- Dingle, K.E., X. Didelot, M.A. Ansari, D.W. Eyre, A. Vaughan, D. Griffiths, C.L. Ip, E.M. Batty, T. Golubchik, R. Bowden, K.A. Jolley, D.W. Hood, W.N. Fawley, A.S. Walker, T.E. Peto, M.H. Wilcox & D.W. Crook, (2013) Recombinational switching of the *Clostridium difficile* S-layer and a novel glycosylation gene cluster revealed by large-scale whole-genome sequencing. *J. Infect. Dis.* **207**: 675-686.
- Dingle, K.E., D. Griffiths, X. Didelot, J. Evans, A. Vaughan, M. Kachrimanidou, N. Stoesser, K.A. Jolley, T. Golubchik, R.M. Harding, T.E. Peto, W. Fawley, A.S. Walker, M. Wilcox & D.W.

- Crook, (2011) Clinical *Clostridium difficile*: clonality and pathogenicity locus diversity. *PLoS One* **6**: e19993.
- Drekonja, D.M., M. Butler, R. MacDonald, D. Bliss, G.A. Filice, T.S. Rector & T.J. Wilt, (2011) Comparative effectiveness of *Clostridium difficile* treatments: a systematic review. *Ann. Intern. Med.* **155**: 839-847.
- Dridi, L., J. Tankovic, B. Burghoffer, F. Barbut & J.C. Petit, (2002) *gyrA* and *gyrB* mutations are implicated in cross-resistance to Ciprofloxacin and moxifloxacin in *Clostridium difficile*. *Antimicrob. Agents Chemother.* **46**: 3418-3421.
- Drudy, D., T. Quinn, R. O'Mahony, L. Kyne, P. O'Gaora & S. Fanning, (2006) High-level resistance to moxifloxacin and gatifloxacin associated with a novel mutation in *gyrB* in toxin-A-negative, toxin-B-positive *Clostridium difficile*. *J. Antimicrob. Chemother.* **58**: 1264-1267.
- Dubberke, E.R. & M.A. Olsen, (2012) Burden of *Clostridium difficile* on the healthcare system. *Clin. Infect. Dis.* **55 Suppl 2**: S88-92.
- Dziarski, R., (2004) Peptidoglycan recognition proteins (PGRPs). *Mol. Immunol.* **40**: 877-886.
- Eaton, K.A., S. Suerbaum, C. Josenhans & S. Krakowka, (1996) Colonization of gnotobiotic piglets by *Helicobacter pylori* deficient in two flagellin genes. *Infect. Immun.* **64**: 2445-2448.
- Emerson, J.E., C.B. Reynolds, R.P. Fagan, H.A. Shaw, D. Goulding & N.F. Fairweather, (2009) A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. *Mol. Microbiol.* **74**: 541-556.
- Ethapa, T., R. Leuzzi, Y.K. Ng, S.T. Baban, R. Adamo, S.A. Kuehne, M. Scarselli, N.P. Minton, D. Serruto & M. Unnikrishnan, (2013) Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*. *J. Bacteriol.* **195**: 545-555.
- Eyles, J.E., B. Unal, M.G. Hartley, S.L. Newstead, H. Flick-Smith, J.L. Prior, P.C. Oyston, A. Randall, Y. Mu, S. Hirst, D.M. Molina, D.H. Davies, T. Milne, K.F. Griffin, P. Baldi, R.W. Titball & P.L. Felgner, (2007) Immunodominant *Francisella tularensis* antigens identified using proteome microarray. *Proteomics* **7**: 2172-2183.
- Eyre, D.W., D. Griffiths, A. Vaughan, T. Golubchik, M. Acharya, L. O'Connor, D.W. Crook, A.S. Walker & T.E. Peto, (2013) Asymptomatic *Clostridium difficile* colonisation and onward transmission. *PLoS One* **8**: e78445.
- Fagan, R.P., D. Albesa-Jove, O. Qazi, D.I. Svergun, K.A. Brown & N.F. Fairweather, (2009) Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. *Mol. Microbiol.* **71**: 1308-1322.
- Fagan, R.P. & N.F. Fairweather, (2011) *Clostridium difficile* has two parallel and essential Sec secretion systems. *J. Biol. Chem.* **286**: 27483-27493.
- Fagan, R.P., C. Janoir, A. Collignon, P. Mastrantonio, I.R. Poxton & N.F. Fairweather, (2011) A proposed nomenclature for cell wall proteins of *Clostridium difficile*. *J. Med. Microbiol.* **60**: 1225-1228.
- Faith, N.G., S. Kathariou, B.L. Neudeck, J.B. Luchansky & C.J. Czuprynski, (2007) A P60 mutant of *Listeria monocytogenes* is impaired in its ability to cause infection in intragastrically inoculated mice. *Microb. Pathog.* **42**: 237-241.
- Farrow, K.A., D. Lytras & J.I. Rood, (2001) Genomic analysis of the erythromycin resistance element Tn5398 from *Clostridium difficile*. *Microbiology* **147**: 2717-2728.
- Fekety, R., (1997) Guidelines for the diagnosis and management of *Clostridium difficile*-associated diarrhea and colitis. American College of Gastroenterology, Practice Parameters Committee. *Am. J. Gastroenterol.* **92**: 739-750.
- Fekety, R., K.H. Kim, D. Brown, D.H. Batts, M. Cudmore & J. Silva, Jr., (1981) Epidemiology of antibiotic-associated colitis; isolation of *Clostridium difficile* from the hospital environment. *Am. J. Med.* **70**: 906-908.
- Felgner, P.L., M.A. Kayala, A. Vigil, C. Burk, R. Nakajima-Sasaki, J. Pablo, D.M. Molina, S. Hirst, J.S. Chew, D. Wang, G. Tan, M. Duffield, R. Yang, J. Neel, N. Chantratita, G. Bancroft, G. Lertmemongkolchai, D.H. Davies, P. Baldi, S. Peacock & R.W. Titball, (2009) A *Burkholderia pseudomallei* protein microarray reveals serodiagnostic and cross-reactive antigens. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 13499-13504.

- Feng, S., J.K. Chen, H. Yu, J.A. Simon & S.L. Schreiber, (1994) Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science* **266**: 1241-1247.
- Fernandez-Tornero, C., R. Lopez, E. Garcia, G. Gimenez-Gallego & A. Romero, (2001) A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA. *Nat. Struct. Biol.* **8**: 1020-1024.
- Foster, T.J. & M. Hook, (1998) Surface protein adhesins of *Staphylococcus aureus*. *Trends Microbiol.* **6**: 484-488.
- Frankel, B.A., M. Bentley, R.G. Kruger & D.G. McCafferty, (2004) Vinyl sulfones: inhibitors of SrtA, a transpeptidase required for cell wall protein anchoring and virulence in *Staphylococcus aureus*. *J. Am. Chem. Soc.* **126**: 3404-3405.
- Frankel, B.A., R.G. Kruger, D.E. Robinson, N.L. Kelleher & D.G. McCafferty, (2005) *Staphylococcus aureus* sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry (Mosc.)* **44**: 11188-11200.
- Gal, M., G. Northey & J.S. Brazier, (2005) A modified pulsed-field gel electrophoresis (PFGE) protocol for subtyping previously non-PFGE typeable isolates of *Clostridium difficile* polymerase chain reaction ribotype 001. *J. Hosp. Infect.* **61**: 231-236.
- Garandeau, C., H. Reglier-Poupet, I. Dubail, J.L. Beretti, P. Berche & A. Charbit, (2002) The sortase SrtA of *Listeria monocytogenes* is involved in processing of internalin and in virulence. *Infect. Immun.* **70**: 1382-1390.
- Garborg, K., B. Waagsbo, A. Stalleemo, J. Matre & A. Sundoy, (2010) Results of faecal donor instillation therapy for recurrent *Clostridium difficile*-associated diarrhoea. *Scand. J. Infect. Dis.* **42**: 857-861.
- Gaspar, A.H., L.A. Marraffini, E.M. Glass, K.L. Debord, H. Ton-That & O. Schneewind, (2005) *Bacillus anthracis* sortase A (SrtA) anchors LPXTG motif-containing surface proteins to the cell wall envelope. *J. Bacteriol.* **187**: 4646-4655.
- Gaynes, R., D. Rimland, E. Killum, H.K. Lowery, T.M. Johnson, 2nd, G. Killgore & F.C. Tenover, (2004) Outbreak of *Clostridium difficile* infection in a long-term care facility: association with gatifloxacin use. *Clin. Infect. Dis.* **38**: 640-645.
- George, W.L., V.L. Sutter, D. Citron & S.M. Finegold, (1979) Selective and differential medium for isolation of *Clostridium difficile*. *J. Clin. Microbiol.* **9**: 214-219.
- Gerding, D.N., (2004) Clindamycin, cephalosporins, fluoroquinolones, and *Clostridium difficile*-associated diarrhea: this is an antimicrobial resistance problem. *Clin. Infect. Dis.* **38**: 646-648.
- Gerding, D.N., C.A. Muto & R.C. Owens, Jr., (2008) Measures to control and prevent *Clostridium difficile* infection. *Clin. Infect. Dis.* **46 Suppl 1**: S43-49.
- Goncalves, C., D. Decre, F. Barbut, B. Burghoffer & J.C. Petit, (2004) Prevalence and characterization of a binary toxin (actin-specific ADP-ribosyltransferase) from *Clostridium difficile*. *J. Clin. Microbiol.* **42**: 1933-1939.
- Gottesman, S., (1996) Proteases and their targets in *Escherichia coli*. *Annu. Rev. Genet.* **30**: 465-506.
- Gough, E., H. Shaikh & A.R. Manges, (2011) Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent *Clostridium difficile* infection. *Clin. Infect. Dis.* **53**: 994-1002.
- Goulding, D., H. Thompson, J. Emerson, N.F. Fairweather, G. Dougan & G.R. Douce, (2009) Distinctive profiles of infection and pathology in hamsters infected with *Clostridium difficile* strains 630 and B1. *Infect. Immun.* **77**: 5478-5485.
- Griffiths, D., W. Fawley, M. Kachrimanidou, R. Bowden, D.W. Crook, R. Fung, T. Golubchik, R.M. Harding, K.J. Jeffery, K.A. Jolley, R. Kirton, T.E. Peto, G. Rees, N. Stoesser, A. Vaughan, A.S. Walker, B.C. Young, M. Wilcox & K.E. Dingle, (2010) Multilocus sequence typing of *Clostridium difficile*. *J. Clin. Microbiol.* **48**: 770-778.

- Gulke, I., G. Pfeifer, J. Liese, M. Fritz, F. Hofmann, K. Aktories & H. Barth, (2001) Characterization of the enzymatic component of the ADP-ribosyltransferase toxin CDTa from *Clostridium difficile*. *Infect. Immun.* **69**: 6004-6011.
- Guttila, I.K., A.H. Gaspar, A. Swierczynski, A. Swaminathan, P. Dwivedi, A. Das & H. Ton-That, (2009) Acyl enzyme intermediates in sortase-catalyzed pilus morphogenesis in gram-positive bacteria. *J. Bacteriol.* **191**: 5603-5612.
- Hafiz, S. & C.L. Oakley, (1976) *Clostridium difficile*: isolation and characteristics. *J. Med. Microbiol.* **9**: 129-136.
- Hall, I.C. & E. O'Toole, (1935) Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, *Bacillus difficilis*. *Am. J. Dis. Child.* **49**: 390-402.
- Hammitt, M.C., D.M. Bueschel, M.K. Keel, R.D. Glock, P. Cuneo, D.W. DeYoung, C. Reggiardo, H.T. Trinh & J.G. Songer, (2008) A possible role for *Clostridium difficile* in the etiology of calf enteritis. *Vet. Microbiol.* **127**: 343-352.
- He, M., M. Sebaihia, T.D. Lawley, R.A. Stabler, L.F. Dawson, M.J. Martin, K.E. Holt, H.M. Seth-Smith, M.A. Quail, R. Rance, K. Brooks, C. Churcher, D. Harris, S.D. Bentley, C. Burrows, L. Clark, C. Corton, V. Murray, G. Rose, S. Thurston, A. van Tonder, D. Walker, B.W. Wren, G. Dougan & J. Parkhill, (2010) Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 7527-7532.
- Heap, J.T., S.A. Kuehne, M. Ehsaan, S.T. Cartman, C.M. Cooksley, J.C. Scott & N.P. Minton, (2010) The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *J. Microbiol. Methods* **80**: 49-55.
- Heap, J.T., O.J. Pennington, S.T. Cartman, G.P. Carter & N.P. Minton, (2007) The Clos Tron: A universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods* **70**: 452-464.
- Hendrickx, A.P., J.M. Budzik, S.Y. Oh & O. Schneewind, (2011) Architects at the bacterial surface - sortases and the assembly of pili with isopeptide bonds. *Nature reviews. Microbiology* **9**: 166-176.
- Hengge, R., (2009) Principles of c-di-GMP signalling in bacteria. *Nature reviews. Microbiology* **7**: 263-273.
- Hennequin, C., C. Janoir, M.C. Barc, A. Collignon & T. Karjalainen, (2003) Identification and characterization of a fibronectin-binding protein from *Clostridium difficile*. *Microbiology* **149**: 2779-2787.
- Hensbergen, P.J., O.I. Klychnikov, D. Bakker, V.J. van Winden, N. Ras, A.C. Kemp, R.A. Cordfunke, I. Dragan, A.M. Deelder, E.J. Kuijper, J. Corver, J.W. Drijfhout & H.C. van Leeuwen, (2014) A novel secreted metalloprotease (CD2830) from *Clostridium difficile* cleaves specific proline sequences in LPXTG cell surface proteins. *Molecular & cellular proteomics : MCP* **13**: 1231-1244.
- Hess, J., A. Dreher, I. Gentschev, W. Goebel, C. Ladel, D. Miko & S.H. Kaufmann, (1996) Protein p60 participates in intestinal host invasion by *Listeria monocytogenes*. *Zentralblatt fur Bakteriologie : international journal of medical microbiology* **284**: 263-272.
- Hienz, S.A., T. Schennings, A. Heimdal & J.I. Flock, (1996) Collagen binding of *Staphylococcus aureus* is a virulence factor in experimental endocarditis. *J. Infect. Dis.* **174**: 83-88.
- Howard, S.L., A. Jagannathan, E.C. Soo, J.P. Hui, A.J. Aubry, I. Ahmed, A. Karlyshev, J.F. Kelly, M.A. Jones, M.P. Stevens, S.M. Logan & B.W. Wren, (2009) *Campylobacter jejuni* glycosylation island important in cell charge, legionaminic acid biosynthesis, and colonization of chickens. *Infect. Immun.* **77**: 2544-2556.
- Hu, P., P. Huang & M.W. Chen, (2013a) Curcumin reduces *Streptococcus mutans* biofilm formation by inhibiting sortase A activity. *Arch. Oral Biol.* **58**: 1343-1348.
- Hu, P., P. Huang & W.M. Chen, (2013b) Curcumin inhibits the Sortase A activity of the *Streptococcus mutans* UA159. *Appl. Biochem. Biotechnol.* **171**: 396-402.
- Huang, X., A. Aulabaugh, W. Ding, B. Kapoor, L. Alksne, K. Tabei & G. Ellestad, (2003) Kinetic mechanism of *Staphylococcus aureus* sortase SrtA. *Biochemistry (Mosc).* **42**: 11307-11315.

- Hussain, H.A., A.P. Roberts & P. Mullany, (2005) Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Δerm) and demonstration that the conjugative transposon Tn916ΔE enters the genome of this strain at multiple sites. *J. Med. Microbiol.* **54**: 137-141.
- Ilangovan, U., J. Iwahara, H. Ton-That, O. Schneewind & R.T. Clubb, (2001) Assignment of the 1H, 13C and 15N signals of Sortase. *J. Biomol. NMR* **19**: 379-380.
- Im, S.H. & A. Rao, (2004) Activation and deactivation of gene expression by Ca2+/calcineurin-NFAT-mediated signaling. *Mol. Cells* **18**: 1-9.
- Ishikawa, S., Y. Hara, R. Ohnishi & J. Sekiguchi, (1998) Regulation of a new cell wall hydrolase gene, *cwlF*, which affects cell separation in *Bacillus subtilis*. *J. Bacteriol.* **180**: 2549-2555.
- Jang, K.H., S.C. Chung, J. Shin, S.H. Lee, T.I. Kim, H.S. Lee & K.B. Oh, (2007) Aaptamines as sortase A inhibitors from the tropical sponge *Aaptos aaptos*. *Bioorg. Med. Chem. Lett.* **17**: 5366-5369.
- Janoir, C., S. Pechine, C. Grosdidier & A. Collignon, (2007) Cwp84, a surface-associated protein of *Clostridium difficile*, is a cysteine protease with degrading activity on extracellular matrix proteins. *J. Bacteriol.* **189**: 7174-7180.
- Janulczyk, R. & M. Rasmussen, (2001) Improved pattern for genome-based screening identifies novel cell wall-attached proteins in gram-positive bacteria. *Infect. Immun.* **69**: 4019-4026.
- Jhung, M.A., A.D. Thompson, G.E. Killgore, W.E. Zukowski, G. Songer, M. Warny, S. Johnson, D.N. Gerding, L.C. McDonald & B.M. Limbago, (2008) Toxinotype V *Clostridium difficile* in humans and food animals. *Emerg. Infect. Dis.* **14**: 1039-1045.
- Johnson, A.P. & M.H. Wilcox, (2012) Fidaxomicin: a new option for the treatment of *Clostridium difficile* infection. *J. Antimicrob. Chemother.* **67**: 2788-2792.
- Johnson, E.A., (1999) Clostridial toxins as therapeutic agents: benefits of nature's most toxic proteins. *Annu. Rev. Microbiol.* **53**: 551-575.
- Johnson, S., (2009) Recurrent *Clostridium difficile* infection: a review of risk factors, treatments, and outcomes. *J. Infect.* **58**: 403-410.
- Johnson, S., C.R. Clabots, F.V. Linn, M.M. Olson, L.R. Peterson & D.N. Gerding, (1990) Nosocomial *Clostridium difficile* colonisation and disease. *Lancet* **336**: 97-100.
- Johnson, S., S.P. Sambol, J.S. Brazier, M. Delmee, V. Avesani, M.M. Merrigan & D.N. Gerding, (2003) International typing study of toxin A-negative, toxin B-positive *Clostridium difficile* variants. *J. Clin. Microbiol.* **41**: 1543-1547.
- Johnson, S., M.H. Samore, K.A. Farrow, G.E. Killgore, F.C. Tenover, D. Lyras, J.I. Rood, P. DeGirolami, A.L. Baltch, M.E. Rafferty, S.M. Pear & D.N. Gerding, (1999) Epidemics of diarrhea caused by a clindamycin-resistant strain of *Clostridium difficile* in four hospitals. *N. Engl. J. Med.* **341**: 1645-1651.
- Jonquieres, R., H. Bierne, F. Fiedler, P. Gounon & P. Cossart, (1999) Interaction between the protein InlB of *Listeria monocytogenes* and lipoteichoic acid: a novel mechanism of protein association at the surface of gram-positive bacteria. *Mol. Microbiol.* **34**: 902-914.
- Jonsson, I.M., S.K. Mazmanian, O. Schneewind, T. Bremell & A. Tarkowski, (2003) The role of *Staphylococcus aureus* sortase A and sortase B in murine arthritis. *Microbes and Infection* **5**: 775-780.
- Jonsson, I.M., S.K. Mazmanian, O. Schneewind, M. Verdrengt, T. Bremell & A. Tarkowski, (2002) On the role of *Staphylococcus aureus* sortase and sortase-catalyzed surface protein anchoring in murine septic arthritis. *J. Infect. Dis.* **185**: 1417-1424.
- Kall, L., A. Krogh & E.L. Sonnhammer, (2007) Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res.* **35**: W429-432.
- Kang, H.J., F. Coulibaly, T. Proft & E.N. Baker, (2011) Crystal structure of Spy0129, a *Streptococcus pyogenes* class B sortase involved in pilus assembly. *PLoS One* **6**: e15969.
- Karlstrom, O., B. Fryklund, K. Tullus & L.G. Burman, (1998) A prospective nationwide study of *Clostridium difficile*-associated diarrhea in Sweden. The Swedish *C. difficile* Study Group. *Clin. Infect. Dis.* **26**: 141-145.

- Kato, H., H. Kita, T. Karasawa, T. Maegawa, Y. Koino, H. Takakuwa, T. Saikai, K. Kobayashi, T. Yamagishi & S. Nakamura, (2001) Colonisation and transmission of *Clostridium difficile* in healthy individuals examined by PCR ribotyping and pulsed-field gel electrophoresis. *J. Med. Microbiol.* **50**: 720-727.
- Kazanowski, M., S. Smolarek, F. Kinnarney & Z. Grzebieniak, (2014) *Clostridium difficile*: epidemiology, diagnostic and therapeutic possibilities-a systematic review. *Techniques in coloproctology* **18**: 223-232.
- Keel, M.K. & J.G. Songer, (2007) The distribution and density of *Clostridium difficile* toxin receptors on the intestinal mucosa of neonatal pigs. *Vet. Pathol.* **44**: 814-822.
- Kelley, L.A. & M.J. Sternberg, (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**: 363-371.
- Kelly, C.P. & J.T. LaMont, (2008) *Clostridium difficile*--more difficult than ever. *N. Engl. J. Med.* **359**: 1932-1940.
- Khanna, S., D.S. Pardi, S.L. Aronson, P.P. Kammer, R. Orenstein, J.L. St Sauver, W.S. Harmsen & A.R. Zinsmeister, (2012) The epidemiology of community-acquired *Clostridium difficile* infection: a population-based study. *Am. J. Gastroenterol.* **107**: 89-95.
- Kharat, A.S. & A. Tomasz, (2003) Inactivation of the *srtA* gene affects localization of surface proteins and decreases adhesion of *Streptococcus pneumoniae* to human pharyngeal cells *in vitro*. *Infect. Immun.* **71**: 2758-2765.
- Kim, S.H., D.S. Shin, M.N. Oh, S.C. Chung, J.S. Lee, I.M. Chang & K.B. Oh, (2003) Inhibition of sortase, a bacterial surface protein anchoring transpeptidase, by beta-sitosterol-3-O-glucopyranoside from *Fritillaria verticillata*. *Biosci. Biotechnol. Biochem.* **67**: 2477-2479.
- Kim, S.H., D.S. Shin, M.N. Oh, S.C. Chung, J.S. Lee & K.B. Oh, (2004) Inhibition of the bacterial surface protein anchoring transpeptidase sortase by isoquinoline alkaloids. *Biosci. Biotechnol. Biochem.* **68**: 421-424.
- Kim, S.W., I.M. Chang & K.B. Oh, (2002) Inhibition of the bacterial surface protein anchoring transpeptidase sortase by medicinal plants. *Biosci. Biotechnol. Biochem.* **66**: 2751-2754.
- Kinnebrew, M.A., C. Ubeda, L.A. Zenewicz, N. Smith, R.A. Flavell & E.G. Pamer, (2010) Bacterial flagellin stimulates Toll-like receptor 5-dependent defense against vancomycin-resistant *Enterococcus* infection. *J. Infect. Dis.* **201**: 534-543.
- Kirby, J., H. Ahern, A. Roberts, V. Kumar, Z. Freeman, K. Acharya & C. Shone, (2009) Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of *Clostridium difficile*. *J. Biol. Chem.* **284**: 34666-34673.
- Knetsch, C.W., E.M. Terveer, C. Lauber, A.E. Gorbalenya, C. Harmanus, E.J. Kuijper, J. Corver & H.C. van Leeuwen, (2012) Comparative analysis of an expanded *Clostridium difficile* reference strain collection reveals genetic diversity and evolution through six lineages. *Infect Genet Evol* **12**: 1577-1585.
- Knofel, T. & N. Strater, (1999) X-ray structure of the *Escherichia coli* periplasmic 5'-nucleotidase containing a dimetal catalytic site. *Nat. Struct. Biol.* **6**: 448-453.
- Kocks, C., E. Gouin, M. Tabouret, P. Berche, H. Ohayon & P. Cossart, (1992) *L. monocytogenes*-induced actin assembly requires the *actA* gene product, a surface protein. *Cell* **68**: 521-531.
- Koonin, E.V., (1994) Conserved sequence pattern in a wide variety of phosphoesterases. *Protein Sci.* **3**: 356-358.
- Kruger, E., E. Witt, S. Ohlmeier, R. Hanschke & M. Hecker, (2000) The *clp* proteases of *Bacillus subtilis* are directly involved in degradation of misfolded proteins. *J. Bacteriol.* **182**: 3259-3265.
- Kuehne, S.A., S.T. Cartman & N.P. Minton, (2011) Both, toxin A and toxin B, are important in *Clostridium difficile* infection. *Gut Microbes* **2**: 252-255.
- Kuijper, E.J., B. Coignard & P. Tull, (2006a) Emergence of *Clostridium difficile*-associated disease in North America and Europe. *Clinical Microbiology and Infection* **12**: 2-18.

- Kuijper, E.J., R.J. van den Berg, S. Debast, C.E. Visser, D. Veenendaal, A. Troelstra, T. van der Kooi, S. van den Hof & D.W. Notermans, (2006b) *Clostridium difficile* ribotype 027, toxinotype III, the Netherlands. *Emerg. Infect. Dis.* **12**: 827-830.
- Kuijper, E.J., J.T. van Dissel & M.H. Wilcox, (2007) *Clostridium difficile*: changing epidemiology and new treatment options. *Current opinion in infectious diseases* **20**: 376-383.
- Kuijper, E.J. & M.H. Wilcox, (2008) Decreased effectiveness of metronidazole for the treatment of *Clostridium difficile* infection? *Clin. Infect. Dis.* **47**: 63-65.
- Kunnath-Velayudhan, S., H. Salamon, H.Y. Wang, A.L. Davidow, D.M. Molina, V.T. Huynh, D.M. Cirillo, G. Michel, E.A. Talbot, M.D. Perkins, P.L. Felgner, X. Liang & M.L. Gennaro, (2010) Dynamic antibody responses to the *Mycobacterium tuberculosis* proteome. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 14703-14708.
- Kvansakul, M., J.C. Adams & E. Hohenester, (2004) Structure of a thrombospondin C-terminal fragment reveals a novel calcium core in the type 3 repeats. *EMBO J.* **23**: 1223-1233.
- Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGgettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson & D.G. Higgins, (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.
- Larson, H.E., A.B. Price, P. Honour & S.P. Borriello, (1978) *Clostridium difficile* and the aetiology of pseudomembranous colitis. *Lancet* **1**: 1063-1066.
- Lawley, T.D., S. Clare, L.J. Deakin, D. Goulding, J.L. Yen, C. Raisen, C. Brandt, J. Lovell, F. Cooke, T.G. Clark & G. Dougan, (2010) Use of purified *Clostridium difficile* spores to facilitate evaluation of health care disinfection regimens. *Appl. Environ. Microbiol.* **76**: 6895-6900.
- Lawley, T.D., S. Clare, A.W. Walker, M.D. Stares, T.R. Connor, C. Raisen, D. Goulding, R. Rad, F. Schreiber, C. Brandt, L.J. Deakin, D.J. Pickard, S.H. Duncan, H.J. Flint, T.G. Clark, J. Parkhill & G. Dougan, (2012) Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing *Clostridium difficile* disease in mice. *PLoS Pathog.* **8**: e1002995.
- Lawrence, S.J., E.R. Dubberke, S. Johnson & D.N. Gerding, (2007) *Clostridium difficile*-associated disease treatment response depends on definition of cure. *Clin. Infect. Dis.* **45**: 1648; author reply 1649-1651.
- Lee, B.J., A. Singh, P. Chiang, S.J. Kemp, E.A. Goldman, M.I. Weinhouse, G.P. Vlasuk & P.J. Rosenthal, (2003) Antimalarial activities of novel synthetic cysteine protease inhibitors. *Antimicrob. Agents Chemother.* **47**: 3810-3814.
- Lee, E.R., J.L. Baker, Z. Weinberg, N. Sudarsan & R.R. Breaker, (2010a) An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* **329**: 845-848.
- Lee, S.F. & T.L. Boran, (2003) Roles of sortase in surface expression of the major protein adhesin P1, saliva-induced aggregation and adherence, and cariogenicity of *Streptococcus mutans*. *Infect. Immun.* **71**: 676-681.
- Lee, Y.J., Y.R. Han, W. Park, S.H. Nam, K.B. Oh & H.S. Lee, (2010b) Synthetic analogs of indole-containing natural products as inhibitors of sortase A and isocitrate lyase. *Bioorg. Med. Chem. Lett.* **20**: 6882-6885.
- Leffler, D.A. & J.T. Lamont, (2012) Editorial: not so nosocomial anymore: the growing threat of community-acquired *Clostridium difficile*. *Am. J. Gastroenterol.* **107**: 96-98.
- Lemee, L., A. Dhalluin, M. Pestel-Caron, J.F. Lemeland & J.L. Pons, (2004) Multilocus sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types. *J. Clin. Microbiol.* **42**: 2609-2617.
- Lessa, F.C., (2013) Community-associated *Clostridium difficile* infection: how real is it? *Anaerobe* **24**: 121-123.
- Levesque, C.M., E. Voronejskaia, Y.C. Huang, R.W. Mair, R.P. Ellen & D.G. Cvitkovitch, (2005) Involvement of sortase anchoring of cell wall proteins in biofilm formation by *Streptococcus mutans*. *Infect. Immun.* **73**: 3773-3777.
- Levett, P.N., (1984) Detection of *Clostridium difficile* in faeces by direct gas liquid chromatography. *J. Clin. Pathol.* **37**: 117-119.

- Liang, L., S. Juarez, T.V. Nga, S. Dunstan, R. Nakajima-Sasaki, D.H. Davies, S. McSorley, S. Baker & P.L. Felgner, (2013) Immune profiling with a *Salmonella Typhi* antigen microarray identifies new diagnostic biomarkers of human typhoid. *Scientific reports* **3**: 1043.
- Liew, C.K., B.T. Smith, R. Pilpa, N. Suree, U. Ilangovan, K.M. Connolly, M.E. Jung & R.T. Clubb, (2004) Localization and mutagenesis of the sorting signal binding site on sortase A from *Staphylococcus aureus*. *FEBS Lett.* **571**: 221-226.
- Loo, V.G., A.M. Bourgault, L. Poirier, F. Lamothe, S. Michaud, N. Turgeon, B. Toye, A. Beaudoin, E.H. Frost, R. Gilca, P. Brassard, N. Dendukuri, C. Beliveau, M. Oughton, I. Brukner & A. Dascal, (2011) Host and pathogen factors for *Clostridium difficile* infection and colonization. *N. Engl. J. Med.* **365**: 1693-1703.
- Louie, T.J., K. Cannon, B. Byrne, J. Emery, L. Ward, M. Eyben & W. Krulicki, (2012) Fidaxomicin preserves the intestinal microbiome during and after treatment of *Clostridium difficile* infection (CDI) and reduces both toxin reexpression and recurrence of CDI. *Clin. Infect. Dis.* **55 Suppl 2**: S132-142.
- Louie, T.J., M.A. Miller, K.M. Mullane, K. Weiss, A. Lentnek, Y. Golan, S. Gorbach, P. Sears, Y.K. Shue & O.P.T.C.S. Group, (2011) Fidaxomicin versus vancomycin for *Clostridium difficile* infection. *N. Engl. J. Med.* **364**: 422-431.
- Lytras, D., J.R. O'Connor, P.M. Howarth, S.P. Sambol, G.P. Carter, T. Phumoonna, R. Poon, V. Adams, G. Vedantam, S. Johnson, D.N. Gerding & J.I. Rood, (2009) Toxin B is essential for virulence of *Clostridium difficile*. *Nature* **458**: 1176-1179.
- Lytras, D. & J.I. Rood, (2000) Transposition of Tn4451 and Tn4453 involves a circular intermediate that forms a promoter for the large resolvase, TnpX. *Mol. Microbiol.* **38**: 588-601.
- Magill, S.S., J.R. Edwards, W. Bamberg, Z.G. Beldavs, G. Dumyati, M.A. Kainer, R. Lynfield, M. Maloney, L. McAllister-Holod, J. Nadle, S.M. Ray, D.L. Thompson, L.E. Wilson, S.K. Fridkin, I. Emerging Infections Program Healthcare-Associated & T. Antimicrobial Use Prevalence Survey, (2014) Multistate point-prevalence survey of health care-associated infections. *N. Engl. J. Med.* **370**: 1198-1208.
- Maiden, M.C., J.A. Bygraves, E. Feil, G. Morelli, J.E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D.A. Caugant, I.M. Feavers, M. Achtman & B.G. Spratt, (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 3140-3145.
- Mandlik, A., A. Das & H. Ton-That, (2008) The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **105**: 14147-14152.
- Mandlik, A., A. Swierczynski, A. Das & H. Ton-That, (2007) *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. *Mol. Microbiol.* **64**: 111-124.
- Marecko, A.W., T.J. Chapa & O. Schneewind, (2006) Surface protein IsdC and Sortase B are required for heme-iron scavenging of *Bacillus anthracis*. *J. Bacteriol.* **188**: 8145-8152.
- Marecko, A.W., R. Wu, J.W. Kern, R. Zhang, D. Janik, D.M. Missiakas, M.E. Duban, A. Joachimiak & O. Schneewind, (2007) Activation of inhibitors by sortase triggers irreversible modification of the active site. *J. Biol. Chem.* **282**: 23129-23139.
- Margot, P., M. Wahlen, A. Gholamhosseini, P. Piggot & D. Karamata, (1998) The *lytE* gene of *Bacillus subtilis* 168 encodes a cell wall hydrolase. *J. Bacteriol.* **180**: 749-752.
- Mariscotti, J.F., F. Garcia-del Portillo & M.G. Pucciarelli, (2009) The *Listeria monocytogenes* sortase-B recognizes varied amino acids at position 2 of the sorting motif. *J. Biol. Chem.* **284**: 6140-6146.
- Marraffini, L.A., A.C. Dedent & O. Schneewind, (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **70**: 192-221.
- Marraffini, L.A. & O. Schneewind, (2005) Anchor structure of staphylococcal surface proteins. V. Anchor structure of the sortase B substrate IsdC. *J. Biol. Chem.* **280**: 16263-16271.
- Marraffini, L.A. & O. Schneewind, (2006) Targeting proteins to the cell wall of sporulating *Bacillus anthracis*. *Mol. Microbiol.* **62**: 1402-1417.

- Marraffini, L.A. & O. Schneewind, (2007) Sortase C-mediated anchoring of BasI to the cell wall envelope of *Bacillus anthracis*. *J. Bacteriol.* **189**: 6425-6436.
- Marraffini, L.A., H. Ton-That, Y. Zong, S.V. Narayana & O. Schneewind, (2004) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. A conserved arginine residue is required for efficient catalysis of sortase A. *J. Biol. Chem.* **279**: 37763-37770.
- Martin, M.J., S. Clare, D. Goulding, A. Faulds-Pain, L. Barquist, H.P. Browne, L. Pettit, G. Dougan, T.D. Lawley & B.W. Wren, (2013) The agr locus regulates virulence and colonization genes in *Clostridium difficile* 027. *J. Bacteriol.* **195**: 3672-3681.
- Matayoshi, E.D., G.T. Wang, G.A. Krafft & J. Erickson, (1990) Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer. *Science* **247**: 954-958.
- Mazmanian, S.K., G. Liu, E.R. Jensen, E. Lenoy & O. Schneewind, (2000) *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proc. Natl. Acad. Sci. U. S. A.* **97**: 5510-5515.
- Mazmanian, S.K., G. Liu, H. Ton-That & O. Schneewind, (1999) *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**: 760-763.
- Mazmanian, S.K., E.P. Skaar, A.H. Gaspar, M. Humayun, P. Gornicki, J. Jelenska, A. Joachmiak, D.M. Missiakas & O. Schneewind, (2003) Passage of heme-iron across the envelope of *Staphylococcus aureus*. *Science* **299**: 906-909.
- Mazmanian, S.K., H. Ton-That & O. Schneewind, (2001) Sortase-catalysed anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. *Mol. Microbiol.* **40**: 1049-1057.
- Mazmanian, S.K., H. Ton-That, K. Su & O. Schneewind, (2002) An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 2293-2298.
- McAleese, F.M., E.J. Walsh, M. Sieprawska, J. Potempa & T.J. Foster, (2001) Loss of clumping factor B fibrinogen binding activity by *Staphylococcus aureus* involves cessation of transcription, shedding and cleavage by metalloprotease. *J. Biol. Chem.* **276**: 29969-29978.
- McDonald, L.C., G.E. Killgore, A. Thompson, R.C. Owens, S.V. Kazakova, S.P. Sambol, S. Johnson & D.N. Gerding, (2005) An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N. Engl. J. Med.* : 2433-2441.
- McFarland, L.V., M.E. Mulligan, R.Y.Y. Kwok & W.E. Stamm, (1989) Nosocomial acquisition of *Clostridium difficile* infection. *N. Engl. J. Med.* **320**: 204-210.
- Misawa, N. & M.J. Blaser, (2000) Detection and characterization of autoagglutination activity by *Campylobacter jejuni*. *Infect. Immun.* **68**: 6168-6175.
- Molina, D.M., S. Pal, M.A. Kayala, A. Teng, P.J. Kim, P. Baldi, P.L. Felgner, X. Liang & L.M. de la Maza, (2010) Identification of immunodominant antigens of *Chlamydia trachomatis* using proteome microarrays. *Vaccine* **28**: 3014-3024.
- Moller, S., M.D. Croning & R. Apweiler, (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**: 646-653.
- Monera, O.D., T.J. Sereda, N.E. Zhou, C.M. Kay & R.S. Hodges, (1995) Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *Journal of peptide science : an official publication of the European Peptide Society* **1**: 319-329.
- Monot, M., C. Boursaux-Eude, M. Thibonnier, D. Vallenet, I. Moszer, C. Medigue, I. Martin-Verstraete & B. Dupuy, (2011) Reannotation of the genome sequence of *Clostridium difficile* strain 630. *J. Med. Microbiol.* **60**: 1193-1199.
- Muryoi, N., M.T. Tiedemann, M. Pluym, J. Cheung, D.E. Heinrichs & M.J. Stillman, (2008) Demonstration of the iron-regulated surface determinant (Isd) heme transfer pathway in *Staphylococcus aureus*. *J. Biol. Chem.* **283**: 28125-28136.
- Muto, C.A., M. Pokrywka, K. Shutt, A.B. Mendelsohn, K. Nouri, K. Posey, T. Roberts, K. Croyle, S. Krystofiak, S. Patel-Brown, A.W. Paschulle, D.L. Paterson, M. Saul & L.H. Harrison, (2005) A large outbreak of *Clostridium difficile*-associated disease with an unexpected

- proportion of deaths and colectomies at a teaching hospital following increased fluoroquinolone use. *Infect. Control Hosp. Epidemiol.* **26**: 273-280.
- Nandakumar, R., M.P. Nandakumar, M.R. Marten & J.M. Ross, (2005) Proteome analysis of membrane and cell wall associated proteins from *Staphylococcus aureus*. *Journal of proteome research* **4**: 250-257.
- Natarajan, M., S.T. Walk, V.B. Young & D.M. Aronoff, (2013) A clinical and epidemiological review of non-toxigenic *Clostridium difficile*. *Anaerobe* **22**: 1-5.
- Navarre, W.W. & O. Schneewind, (1994) Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. *Mol. Microbiol.* **14**: 115-121.
- Navarre, W.W. & O. Schneewind, (1999) Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol. Mol. Biol. Rev.* **63**: 174-229.
- Neal, M.D., J.C. Alverdy, D.E. Hall, R.L. Simmons & B.S. Zuckerbraun, (2011) Diverting loop ileostomy and colonic lavage: an alternative to total abdominal colectomy for the treatment of severe, complicated *Clostridium difficile* associated disease. *Ann. Surg.* **254**: 423-427; discussion 427-429.
- Necchi, F., V. Nardi-Dei, M. Biagini, M. Assfalg, A. Nuccitelli, R. Cozzi, N. Norais, J.L. Telford, C.D. Rinaudo, G. Grandi & D. Maione, (2011) Sortase A substrate specificity in GBS pilus 2a cell wall anchoring. *PLoS One* **6**: e25300.
- Newton, S.M., P.E. Klebba, C. Raynaud, Y. Shao, X. Jiang, I. Dubail, C. Archer, C. Frehel & A. Charbit, (2005) The *svpA-srtB* locus of *Listeria monocytogenes*: *fur*-mediated iron regulation and effect on virulence. *Mol. Microbiol.* **55**: 927-940.
- Nguyen, H.D., T.T. Phan & W. Schumann, (2011) Analysis and application of *Bacillus subtilis* sortases to anchor recombinant proteins on the cell wall. *AMB Express* **1**: 22.
- Nobbs, A.H., R.J. Lamont & H.F. Jenkinson, (2009) *Streptococcus* adherence and colonization. *Microbiol. Mol. Biol. Rev.* **73**: 407-450, Table of Contents.
- Nobbs, A.H., R.M. Vajna, J.R. Johnson, Y. Zhang, S.L. Erlandsen, M.W. Oli, J. Kreth, L.J. Brady & M.C. Herzberg, (2007) Consequences of a sortase A mutation in *Streptococcus gordonii*. *Microbiology* **153**: 4088-4097.
- Nudler, E. & A.S. Mironov, (2004) The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* **29**: 11-17.
- Oh, I., W.Y. Yang, S.C. Chung, T.Y. Kim, K.B. Oh & J. Shin, (2011) *In vitro* sortase A inhibitory and antimicrobial activity of flavonoids isolated from the roots of *Sophora flavescens*. *Arch. Pharm. Res.* **34**: 217-222.
- Oh, K.-B., K.-W. Nam, H. Ahn, J. Shin, S. Kim & W. Mar, (2010) Therapeutic effect of (Z)-3-(2,5-dimethoxyphenyl)-2-(4-methoxyphenyl) acrylonitrile (DMMA) against *Staphylococcus aureus* infection in a murine model. *Biochem. Biophys. Res. Commun.* **396**: 440-444.
- Oh, K.B., S.H. Kim, J. Lee, W.J. Cho, T. Lee & S. Kim, (2004) Discovery of diarylacrylonitriles as a novel series of small molecule sortase A inhibitors. *J. Med. Chem.* **47**: 2418-2421.
- Oh, K.B., M.N. Oh, J.G. Kim, D.S. Shin & J. Shin, (2006) Inhibition of sortase-mediated *Staphylococcus aureus* adhesion to fibronectin via fibronectin-binding protein by sortase inhibitors. *Appl. Environ. Microbiol.* **70**: 102-106.
- Oughton, M.T., V.G. Loo, N. Dendukuri, S. Fenn & M.D. Libman, (2009) Hand hygiene with soap and water is superior to alcohol rub and antiseptic wipes for removal of *Clostridium difficile*. *Infect. Control Hosp. Epidemiol.* **30**: 939-944.
- Owens, R.C., Jr., C.J. Donskey, R.P. Gaynes, V.G. Loo & C.A. Muto, (2008) Antimicrobial-associated risk factors for *Clostridium difficile* infection. *Clin. Infect. Dis.* **46 Suppl 1**: S19-31.
- Ozaki, E., H. Kato, H. Kita, T. Karasawa, T. Maegawa, Y. Koino, K. Matsumoto, T. Takada, K. Nomoto, R. Tanaka & S. Nakamura, (2004) *Clostridium difficile* colonization in healthy adults: transient colonization and correlation with enterococcal colonization. *J. Med. Microbiol.* **53**: 167-172.

- Pallen, M.J., A.C. Lam, M. Antonio & K. Dunbar, (2001) An embarrassment of sortases - a richness of substrates? *Trends Microbiol.* **9**: 97-102.
- Park, B.S., J.G. Kim, M.R. Kim, S.E. Lee, G.R. Takeoka, K.B. Oh & J.H. Kim, (2005) *Curcuma longa* L. constituents inhibit sortase A and *Staphylococcus aureus* cell adhesion to fibronectin. *Journal of Agricultural and Food Chemistry* **53**: 9005-9009.
- Paterson, G.K. & T.J. Mitchell, (2004) The biology of Gram-positive sortase enzymes. *Trends Microbiol.* **12**: 89-95.
- Patti, J.M., B.L. Allen, M.J. McGavin & M. Hook, (1994a) MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu. Rev. Microbiol.* **48**: 585-617.
- Patti, J.M., T. Bremell, D. Krajewska-Pietrasik, A. Abdelnour, A. Tarkowski, C. Ryden & M. Hook, (1994b) The *Staphylococcus aureus* collagen adhesin is a virulence determinant in experimental septic arthritis. *Infect. Immun.* **62**: 152-161.
- Pechine, S., A. Gleizes, C. Janoir, R. Gorges-Kergot, M.C. Barc, M. Delmee & A. Collignon, (2005) Immunological properties of surface proteins of *Clostridium difficile*. *J. Med. Microbiol.* **54**: 193-196.
- Peltier, J., P. Courtin, I. El Meouche, L. Lemee, M.P. Chapot-Chartier & J.L. Pons, (2011) *Clostridium difficile* has an original peptidoglycan structure with a high level of N-acetylglucosamine deacetylation and mainly 3-3 cross-links. *J. Biol. Chem.* **286**: 29053-29062.
- Pépin, J., M.E. Alary, L. Valiquette, E. Raiche, J. Ruel, K. Fulop, D. Godin & C. Bourassa, (2005a) Increasing risk of relapse after treatment of *Clostridium difficile* colitis in Quebec, Canada. *Clin. Infect. Dis.* **40**: 1591-1597.
- Pépin, J., S. Routhier, S. Gagnon & I. Brazeau, (2006) Management and outcomes of a first recurrence of *Clostridium difficile*-associated disease in Quebec, Canada. *Clin. Infect. Dis.* **42**: 758-764.
- Pépin, J., N. Saheb, M.A. Coulombe, M.E. Alary, M.P. Corriveau, S. Authier, M. Leblanc, G. Rivard, M. Bettez, V. Primeau, M. Nguyen, C.E. Jacob & L. Lanthier, (2005b) Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*-associated diarrhea: A cohort study during an epidemic in Quebec. *Clin. Infect. Dis.* **41**: 1254-1260.
- Pépin, J., L. Valiquette, M.E. Alary, P. Villemure, A. Pelletier, K. Forget, K. Pépin & D. Chouinard, (2004) *Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. *Can. Med. Assoc. J.* **171**: 466-472.
- Pépin, J., L. Valiquette & B. Cossette, (2005c) Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *CMAJ* **173**: 1037-1042.
- Pereira, F.C., L. Saujet, A.R. Tome, M. Serrano, M. Monot, E. Couture-Tosi, I. Martin-Verstraete, B. Dupuy & A.O. Henriques, (2013) The spore differentiation pathway in the enteric pathogen *Clostridium difficile*. *PLoS genetics* **9**: e1003782.
- Perelle, S., M. Gibert, P. Bourlioux, G. Corthier & M.R. Popoff, (1997) Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infect. Immun.* **65**: 1402-1407.
- Perera, A.D., R.P. Akbari, M.S. Cowher, T.E. Read, J.T. McCormick, D.S. Medich, J.P. Celebrazze, Jr., S.J. Beck, P.E. Fischer & P.F. Caushaj, (2010) Colectomy for fulminant *Clostridium difficile* colitis: predictors of mortality. *Am. Surg.* **76**: 418-421.
- Perry, A.M., H. Ton-That, S.K. Mazmanian & O. Schneewind, (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an *in vivo* peptidoglycan substrate for sortase-catalyzed surface protein anchoring. *J. Biol. Chem.* **277**: 16241-16248.
- Persky, S.E. & L.J. Brandt, (2000) Treatment of recurrent *Clostridium difficile*-associated diarrhea by administration of donated stool directly through a colonoscope. *Am. J. Gastroenterol.* **95**: 3283-3285.
- Perutka, J., W.J. Wang, D. Goerlitz & A.M. Lambowitz, (2004) Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J. Mol. Biol.* **336**: 421-439.

- Petersen, T.N., S. Brunak, G. von Heijne & H. Nielsen, (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785-786.
- Pilgrim, S., A. Kolb-Maurer, I. Gentschev, W. Goebel & M. Kuhn, (2003) Deletion of the gene encoding p60 in *Listeria monocytogenes* leads to abnormal cell division and loss of actin-based motility. *Infect. Immun.* **71**: 3473-3484.
- Pizarro-Cerdá, J. & P. Cossart, (2006) Bacterial adhesion and entry into host cells. *Cell* **124**: 715-727.
- Popp, M.W., J.M. Antos & H.L. Ploegh, (2009) Site-specific protein labeling via sortase-mediated transpeptidation. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.] Chapter 15*: Unit 15 13.
- Price, A.B. & D.R. Davies, (1977) Pseudomembranous colitis. *J. Clin. Pathol.* **30**: 1-12.
- Pritz, S., Y. Wolf, O. Kraetke, J. Klose, M. Bienert & M. Beyermann, (2007) Synthesis of biologically active peptide nucleic acid-peptide conjugates by sortase-mediated ligation. *J. Org. Chem.* **72**: 3909-3912.
- Pucciarelli, M.G., E. Calvo, C. Sabet, H. Bierne, P. Cossart & F. Garcia-del Portillo, (2005) Identification of substrates of the *Listeria monocytogenes* sortases A and B by a non-gel proteomic analysis. *Proteomics* **5**: 4808-4817.
- Pultz, N.J. & C.J. Donskey, (2005) Effect of antibiotic treatment on growth of and toxin production by *Clostridium difficile* in the cecal contents of mice. *Antimicrob. Agents Chemother.* **49**: 3529-3532.
- Punta, M., P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman & R.D. Finn, (2012) The Pfam protein families database. *Nucleic Acids Res* **40**: D290-301.
- Purcell, E.B., R.W. McKee, S.M. McBride, C.M. Waters & R. Tamayo, (2012) Cyclic diguanylate inversely regulates motility and aggregation in *Clostridium difficile*. *J. Bacteriol.* **194**: 3307-3316.
- Purdy, D., T.A.T. O'Keeffe, M. Elmore, M. Herbert, A. McLeod, M. Bokori-Brown, A. Ostrowski & N.P. Minton, (2002) Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol. Microbiol.* **46**: 439-452.
- Race, P.R., M.L. Bentley, J.A. Melvin, A. Crow, R.K. Hughes, W.D. Smith, R.B. Sessions, M.A. Kehoe, D.G. McCafferty & M.J. Banfield, (2009) Crystal structure of *Streptococcus pyogenes* sortase A: implications for sortase mechanism. *J. Biol. Chem.* **284**: 6924-6933.
- Rao, K., D. Berland, C. Young, S.T. Walk & D.W. Newton, (2013) The nose knows not: poor predictive value of stool sample odor for detection of *Clostridium difficile*. *Clin. Infect. Dis.* **56**: 615-616.
- Razavi, B., A. Apisarnthanarak & L.M. Mundy, (2007) *Clostridium difficile*: Emergence of hypervirulence and fluoroquinolone resistance. *Infection* **35**: 300-307.
- Read, T.D., S.N. Peterson, N. Tourasse, L.W. Baillie, I.T. Paulsen, K.E. Nelson, H. Tettelin, D.E. Fouts, J.A. Eisen, S.R. Gill, E.K. Holtzapfel, O.A. Okstad, E. Helgason, J. Rilstone, M. Wu, J.F. Kolonay, M.J. Beanan, R.J. Dodson, L.M. Brinkac, M. Gwinn, R.T. DeBoy, R. Madpu, S.C. Daugherty, A.S. Durkin, D.H. Haft, W.C. Nelson, J.D. Peterson, M. Pop, H.M. Khouri, D. Radune, J.L. Benton, Y. Mahamoud, L. Jiang, I.R. Hance, J.F. Weidman, K.J. Berry, R.D. Plaut, A.M. Wolf, K.L. Watkins, W.C. Nierman, A. Hazen, R. Cline, C. Redmond, J.E. Thwaite, O. White, S.L. Salzberg, B. Thomason, A.M. Friedlander, T.M. Koehler, P.C. Hanna, A.B. Kolsto & C.M. Fraser, (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **423**: 81-86.
- Reynolds, C.B., J.E. Emerson, L. de la Riva, R.P. Fagan & N.F. Fairweather, (2011) The *Clostridium difficile* cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. *PLoS Pathogens* **7**: e1002024.
- Riggs, M.M., A.K. Sethi, T.F. Zabarsky, E.C. Eckstein, R.L. Jump & C.J. Donskey, (2007) Asymptomatic carriers are a potential source for transmission of epidemic and

- nonepidemic *Clostridium difficile* strains among long-term care facility residents. *Clin. Infect. Dis.* **45**: 992-998.
- Roberts, A.P., P.A. Johanesen, D. Lyras, P. Mullany & J.I. Rood, (2001) Comparison of Tn5397 from *Clostridium difficile*, Tn916 from *Enterococcus faecalis* and the CW459tet(M) element from *Clostridium perfringens* shows that they have similar conjugation regions but different insertion and excision modules. *Microbiology* **147**: 1243-1251.
- Robichon, C., J. Luo, T.B. Causey, J.S. Benner & J.C. Samuelson, (2011) Engineering *Escherichia coli* BL21(DE3) derivative strains to minimize *E. coli* protein contamination after purification by immobilized metal affinity chromatography. *Appl. Environ. Microbiol.* **77**: 4634-4646.
- Roche, F.M., R. Massey, S.J. Peacock, N.P.J. Day, L. Visai, P. Speziale, A. Lam, M. Pallen & T.J. Foster, (2003) Characterization of novel LPXTG-containing proteins of *Staphylococcus aureus* identified from genome sequences. *Microbiology* **149**: 643-654.
- Rodriguez-Palacios, A., H.R. Stampfli, T. Duffield, A.S. Peregrine, L.A. Trotz-Williams, L.G. Arroyo, J.S. Brazier & J.S. Weese, (2006) *Clostridium difficile* PCR ribotypes in calves, Canada. *Emerg. Infect. Dis.* **12**: 1730-1736.
- Roggemann, A., H.R. Neuberger, A. Flugel, T. Schmoll & J. Heesemann, (1995) Substitution of two histidine residues in YadA protein of *Yersinia enterocolitica* abrogates collagen binding, cell adherence and mouse virulence. *Mol. Microbiol.* **16**: 1207-1219.
- Rohlke, F., C.M. Surawicz & N. Stollman, (2010) Fecal flora reconstitution for recurrent *Clostridium difficile* infection: results and methodology. *J. Clin. Gastroenterol.* **44**: 567-570.
- Rupnik, M., V. Avesani, M. Janc, C. von Eichel-Streiber & M. Delmee, (1998) A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. *J. Clin. Microbiol.* **36**: 2240-2247.
- Rupnik, M., M.H. Wilcox & D.N. Gerding, (2009) *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nature Reviews Microbiology* **7**: 526-536.
- Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.A. Rajandream & B. Barrell, (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.
- Ruzin, A., A. Severin, F. Ritacco, K. Tabei, G. Singh, P.A. Bradford, M.M. Siegel, S.J. Projan & D.M. Shlaes, (2002) Further evidence that a cell wall precursor [C(55)-MurNAc-(peptide)-GlcNAc] serves as an acceptor in a sorting reaction. *J. Bacteriol.* **184**: 2141-2147.
- Sára, M. & U.B. Sleytr, (2000) S-layer proteins. *J. Bacteriol.* **182**: 859-868.
- Scheline, R.R., (1968) Metabolism of phenolic acids by the rat intestinal microflora. *Acta Pharmacol. Toxicol. (Copenh.)* **26**: 189-205.
- Schleifer, K.H. & O. Kandler, (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* **36**: 407-477.
- Schneewind, O., A. Fowler & K.F. Faull, (1995) Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science* **268**: 103-106.
- Schneewind, O., D. Mihaylova-Petkov & P. Model, (1993) Cell wall sorting signals in surface proteins of gram-positive bacteria. *EMBO J.* **12**: 4803-4811.
- Schneewind, O. & D. Missiakas, (2014) Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. *Biochim. Biophys. Acta* **1843**: 1687-1697.
- Schneewind, O., P. Model & V.A. Fischetti, (1992) Sorting of Protein A to the Staphylococcal cell-wall. *Cell* **70**: 267-281.
- Schwan, C., A.S. Kruppke, T. Nolke, L. Schumacher, F. Koch-Nolte, M. Kudryashev, H. Stahlberg & K. Aktories, (2014) *Clostridium difficile* toxin CDT hijacks microtubule organization and reroutes vesicle traffic to increase pathogen adherence. *Proc. Natl. Acad. Sci. U. S. A.* **111**: 2313-2318.
- Schwan, C., B. Stecher, T. Tzivelekidis, M. van Ham, M. Rohde, W.D. Hardt, J. Weiland & K. Aktories, (2009) *Clostridium difficile* toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria. *PLoS Pathog* **5**: e1000626.

- Scott, C.J., A. McDowell, S.L. Martin, J.F. Lynas, K. Vandenbroeck & B. Walker, (2002) Irreversible inhibition of the bacterial cysteine protease-transpeptidase sortase (SrtA) by substrate-derived affinity labels. *Biochem. J.* **366**: 953-958.
- Sebaihia, M., B.W. Wren, P. Mullany, N.F. Fairweather, N. Minton, R. Stabler, N.R. Thomson, A.P. Roberts, A.M. Cerdeno-Tarraga, H. Wang, M.T. Holden, A. Wright, C. Churcher, M.A. Quail, S. Baker, N. Bason, K. Brooks, T. Chillingworth, A. Cronin, P. Davis, L. Dowd, A. Fraser, T. Feltwell, Z. Hance, S. Holroyd, K. Jagels, S. Moule, K. Mungall, C. Price, E. Rabbinowitsch, S. Sharp, M. Simmonds, K. Stevens, L. Unwin, S. Whithead, B. Dupuy, G. Dougan, B. Barrell & J. Parkhill, (2006) The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* **38**: 779-786.
- Seddon, S.V., I. Hemingway & S.P. Borriello, (1990) Hydrolytic enzyme production by *Clostridium difficile* and its relationship to toxin production and virulence in the hamster model. *J. Med. Microbiol.* **31**: 169-174.
- Severin, A., K. Tabei & A. Tomasz, (2004) The structure of the cell wall peptidoglycan of *Bacillus cereus* RSVF1, a strain closely related to *Bacillus anthracis*. *Microbial drug resistance* **10**: 77-82.
- Sharp, S.E., L.O. Ruden, J.C. Pohl, P.A. Hatcher, L.M. Jayne & W.M. Ivie, (2010) Evaluation of the C.Diff Quik Chek Complete Assay, a new glutamate dehydrogenase and A/B toxin combination lateral flow assay for use in rapid, simple diagnosis of *Clostridium difficile* disease. *J. Clin. Microbiol.* **48**: 2082-2086.
- Shaw, E., (1990) Cysteinyl proteinases and their selective inactivation. *Adv. Enzymol. Relat. Areas Mol. Biol.* **63**: 271-347.
- Shaw, E. & G.D. Green, (1981) Inactivation of thiol proteases with peptidyl diazomethyl ketones. *Methods Enzymol.* **80 Pt C**: 820-826.
- Shen, A., (2012) *Clostridium difficile* toxins: mediators of inflammation. *Journal of innate immunity* **4**: 149-158.
- Shim, J.K., S. Johnson, M.H. Samore, D.Z. Bliss & D.N. Gerding, (1998) Primary symptomless colonisation by *Clostridium difficile* and decreased risk of subsequent diarrhoea. *Lancet* **351**: 633-636.
- Skaar, E.P., A.H. Gaspar & O. Schneewind, (2006) *Bacillus anthracis* IsdG, a heme-degrading monooxygenase. *J. Bacteriol.* **188**: 1071-1080.
- Smith, L.D., (1975) *The pathogenic anaerobic bacteria*. Charles C. Thomas, Springfield, Ill.
- Songer, J.G. & M.A. Anderson, (2006) *Clostridium difficile*: an important pathogen of food animals. *Anaerobe* **12**: 1-4.
- Sorg, J.A. & A.L. Sonenshein, (2008) Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *J. Bacteriol.* **190**: 2505-2512.
- Soutourina, O.A., M. Monot, P. Boudry, L. Saujet, C. Pichon, O. Sismeiro, E. Semenova, K. Severinov, C. Le Bouguenec, J.Y. Coppee, B. Dupuy & I. Martin-Verstraete, (2013) Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS genetics* **9**: e1003493.
- Spigaglia, P., V. Carucci, F. Barbanti & P. Mastrantonio, (2005) ErmB determinants and Tn916-Like elements in clinical isolates of *Clostridium difficile*. *Antimicrob. Agents Chemother.* **49**: 2550-2553.
- Spirig, T., E.M. Weiner & R.T. Clubb, (2011) Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* **82**: 1044-1059.
- Stabler, R.A., L.F. Dawson, E. Valiente, M.D. Cairns, M.J. Martin, E.H. Donahue, T.V. Riley, J.G. Songer, E.J. Kuijper, K.E. Dingle & B.W. Wren, (2012) Macro and micro diversity of *Clostridium difficile* isolates from diverse sources and geographical locations. *PLoS One* **7**: e31559.
- Stabler, R.A., M. He, L. Dawson, M. Martin, E. Valiente, C. Corton, T.D. Lawley, M. Sebaihia, M.A. Quail, G. Rose, D.N. Gerding, M. Gibert, M.R. Popoff, J. Parkhill, G. Dougan & B.W. Wren, (2009) Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains

- provides insight into the evolution of a hypervirulent bacterium. *Genome biology* **10**: R102.
- Stubbs, S., M. Rupnik, M. Gibert, J. Brazier, B. Duerden & M. Popoff, (2000) Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of *Clostridium difficile*. *FEMS Microbiol. Lett.* **186**: 307-312.
- Stubbs, S.L.J., J.S. Brazier, G.L. O'Neill & B.I. Duerden, (1999) PCR targeted to the 16S-23S rRNA gene intergenic spacer region of *Clostridium difficile* and construction of a library consisting of 116 different PCR ribotypes. *J. Clin. Microbiol.* **37**: 461-463.
- Sudheesh, P.S., S. Crane, K.D. Cain & M.S. Strom, (2007) Sortase inhibitor phenyl vinyl sulfone inhibits *Renibacterium salmoninarum* adherence and invasion of host cells. *Dis. Aquat. Organ.* **78**: 115-127.
- Sugiura, R., S.O. Sio, H. Shunthoh & T. Kuno, (2002) Calcineurin phosphatase in signal transduction: lessons from fission yeast. *Genes Cells* **7**: 619-627.
- Swaminathan, A., A. Mandlik, A. Swierczynski, A. Gaspar, A. Das & H. Ton-That, (2007) Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in *Corynebacterium diphtheriae*. *Mol. Microbiol.* **66**: 961-974.
- Switalski, L.M., P. Speziale & M. Hook, (1989) Isolation and characterization of a putative collagen receptor from *Staphylococcus aureus* strain Cowan 1. *J. Biol. Chem.* **264**: 21080-21086.
- Symersky, J., J.M. Patti, M. Carson, K. House-Pompeo, M. Teale, D. Moore, L. Jin, A. Schneider, L.J. DeLucas, M. Hook & S.V. Narayana, (1997) Structure of the collagen-binding domain from a *Staphylococcus aureus* adhesin. *Nat. Struct. Biol.* **4**: 833-838.
- Tadesse, S. & P.L. Graumann, (2006) Differential and dynamic localization of topoisomerases in *Bacillus subtilis*. *J. Bacteriol.* **188**: 3002-3011.
- Tahir, Y.E., P. Kuusela & M. Skurnik, (2000) Functional mapping of the *Yersinia enterocolitica* adhesin YadA. Identification Of eight NSVAIG - S motifs in the amino-terminal half of the protein involved in collagen binding. *Mol. Microbiol.* **37**: 192-206.
- Tannock, G.W., K. Munro, C. Taylor, B. Lawley, W. Young, B. Byrne, J. Emery & T. Louie, (2010) A new macrocyclic antibiotic, fidaxomicin (OPT-80), causes less alteration to the bowel microbiota of *Clostridium difficile*-infected patients than does vancomycin. *Microbiology* **156**: 3354-3359.
- Tasteyre, A., M.C. Barc, A. Collignon, H. Boureau & T. Karjalainen, (2001) Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect. Immun.* **69**: 7937-7940.
- Tedesco, F.J., R.W. Barton & D.H. Alpers, (1974) Clindamycin-associated colitis - prospective study. *Ann. Intern. Med.* **81**: 429-433.
- Ticehurst, J.R., D.Z. Aird, L.M. Dam, A.P. Borek, J.T. Hargrove & K.C. Carroll, (2006) Effective detection of toxigenic *Clostridium difficile* by a two-step algorithm including tests for antigen and cytotoxin. *J. Clin. Microbiol.* **44**: 1145-1149.
- Ton-That, H., K.F. Faull & O. Schneewind, (1997) Anchor structure of staphylococcal surface proteins. A branched peptide that links the carboxyl terminus of proteins to the cell wall. *J. Biol. Chem.* **272**: 22285-22292.
- Ton-That, H., G. Liu, S.K. Mazmanian, K.F. Faull & O. Schneewind, (1999) Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc. Natl. Acad. Sci. U. S. A.* **96**: 12424-12429.
- Ton-That, H., L.A. Marraffini & O. Schneewind, (2004) Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochim. Biophys. Acta* **1694**: 269-278.
- Ton-That, H., S.K. Mazmanian, L. Alksne & O. Schneewind, (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *J. Biol. Chem.* **277**: 7447-7452.
- Ton-That, H., S.K. Mazmanian, K.F. Faull & O. Schneewind, (2000) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Sortase catalyzed *in vitro* transpeptidation

- reaction using LPXTG peptide and NH(2)-Gly(3) substrates. *J. Biol. Chem.* **275**: 9876-9881.
- Ton-That, H. & O. Schneewind, (1999) Anchor structure of staphylococcal surface proteins. IV. Inhibitors of the cell wall sorting reaction. *J. Biol. Chem.* **274**: 24316-24320.
- Ton-That, H. & O. Schneewind, (2003) Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol. Microbiol.* **50**: 1429-1438.
- Tonyry, J.H., B.A. McNichol, N. Ramarao, D.S. Chertow, K.S. Kim, S. Stibitz, O. Schneewind, F. Kashanchi, C.L. Bailey, S. Popov & M.C. Chung, (2012) *Bacillus anthracis* protease InhA regulates BslA-mediated adhesion in human endothelial cells. *Cell Microbiol* **14**: 1219-1230.
- Tsompanidou, E., E.L. Denham, M.J. Sibbald, X.M. Yang, J. Seinen, A.W. Friedrich, G. Buist & J.M. van Dijl, (2012) The sortase A substrates FnbpA, FnbpB, ClfA and ClfB antagonize colony spreading of *Staphylococcus aureus*. *PLoS One* **7**: e44646.
- Tulli, L., S. Marchi, R. Petracca, H.A. Shaw, N.F. Fairweather, M. Scarselli, M. Soriani & R. Leuzzi, (2013) CbpA: a novel surface exposed adhesin of *Clostridium difficile* targeting human collagen. *Cell Microbiol* **15**: 1674-1687.
- Twine, S.M., C.W. Reid, A. Aubry, D.R. McMullin, K.M. Fulton, J. Austin & S.M. Logan, (2009) Motility and flagellar glycosylation in *Clostridium difficile*. *J. Bacteriol.* **191**: 7050-7062.
- Untergasser, A., H. Nijveen, X. Rao, T. Bisseling, R. Geurts & J.A. Leunissen, (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**: W71-74.
- Valiquette, L., B. Cossette, M.P. Garant, H. Diab & J. Pepin, (2007) Impact of a reduction in the use of high-risk antibiotics on the course of an epidemic of *Clostridium difficile*-associated disease caused by the hypervirulent NAP1/027 strain. *Clin. Infect. Dis.* **45 Suppl 2**: S112-121.
- van Nood, E., P. Speelman, E.J. Kuijper & J.J. Keller, (2009) Struggling with recurrent *Clostridium difficile* infections: is donor faeces the solution? *Euro surveillance : bulletin European sur les maladies transmissibles = European communicable disease bulletin* **14**.
- van Nood, E., A. Vrieze, M. Nieuwdorp, S. Fuentes, E.G. Zoetendal, W.M. de Vos, C.E. Visser, E.J. Kuijper, J.F. Bartelsman, J.G. Tijssen, P. Speelman, M.G. Dijkgraaf & J.J. Keller, (2013) Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N. Engl. J. Med.* **368**: 407-415.
- Verity, P., M.H. Wilcox, W. Fawley & P. Parnell, (2001) Prospective evaluation of environmental contamination by *Clostridium difficile* in isolation side rooms. *J. Hosp. Infect.* **49**: 204-209.
- Vigil, A., R. Ortega, R. Nakajima-Sasaki, J. Pablo, D.M. Molina, C.C. Chao, H.W. Chen, W.M. Ching & P.L. Felgner, (2010) Genome-wide profiling of humoral immune response to *Coxiella burnetii* infection by protein microarray. *Proteomics* **10**: 2259-2269.
- Vohra, P. & I.R. Poxton, (2011) Efficacy of decontaminants and disinfectants against *Clostridium difficile*. *J. Med. Microbiol.* **60**: 1218-1224.
- Vollaard, E.J. & H.A. Clasener, (1994) Colonization resistance. *Antimicrob. Agents Chemother.* **38**: 409-414.
- Vonberg, R.P., E.J. Kuijper, M.H. Wilcox, F. Barbut, P. Tull, P. Gastmeier, P.J. van den Broek, A. Colville, B. Coignard, T. Daha, S. Debast, B.I. Duerden, S. van den Hof, T. van der Kooi, H.J.H. Maarleveld, E. Nagy, D.W. Notermans, J. O'Driscoll, B. Patel, S. Stone & C. Wiuff, (2008) Infection control measures to limit the spread of *Clostridium difficile*. *Clinical Microbiology and Infection* **14**: 2-20.
- Voth, D.E. & J.D. Ballard, (2005) *Clostridium difficile* toxins: mechanism of action and role in disease. *Clin. Microbiol. Rev.* **18**: 247-263.
- Waligora, A.J., C. Hennequin, P. Mullany, P. Bourlioux, A. Collignon & T. Karjalainen, (2001) Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. *Infect. Immun.* **69**: 2144-2153.

- Wang, L., Y.J. Zhou, D. Ji, X. Lin, Y. Liu, Y. Zhang, W. Liu & Z.K. Zhao, (2014) Identification of UshA as a major enzyme for NAD degradation in *Escherichia coli*. *Enzyme Microb. Technol.* **58-59**: 75-79.
- Weiner, E.M., S. Robson, M. Marohn & R.T. Clubb, (2010) The Sortase A enzyme that attaches proteins to the cell wall of *Bacillus anthracis* contains an unusual active site architecture. *J. Biol. Chem.* **285**: 23433-23443.
- Weiss, W.J., E. Lenoy, T. Murphy, L. Tardio, P. Burgio, S.J. Projan, O. Schneewind & L. Alksne, (2004) Effect of *srtA* and *srtB* gene expression on the virulence of *Staphylococcus aureus* in animal models of infection. *J. Antimicrob. Chemother.* **53**: 480-486.
- Wilcox, M.H., (2012) Overcoming barriers to effective recognition and diagnosis of *Clostridium difficile* infection. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **18 Suppl 6**: 13-20.
- Wilcox, M.H., W.N. Fawley, C.D. Settle & A. Davidson, (1998) Recurrence of symptoms in *Clostridium difficile* infection--relapse or reinfection? *J. Hosp. Infect.* **38**: 93-100.
- Wilcox, M.H., L. Mooney, R. Bendall, C.D. Settle & W.N. Fawley, (2008) A case-control study of community-associated *Clostridium difficile* infection. *J. Antimicrob. Chemother.* **62**: 388-396.
- Wilkins, M.R., E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel & D.F. Hochstrasser, (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**: 531-552.
- Williams, D.R., D.I. Young & M. Young, (1990) Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*. *J. Gen. Microbiol.* **136**: 819-826.
- Woo, P.C., S.K. Lau, K.M. Chan, A.M. Fung, B.S. Tang & K.Y. Yuen, (2005) *Clostridium* bacteraemia characterised by 16S ribosomal RNA gene sequencing. *J. Clin. Pathol.* **58**: 301-307.
- Wright, A., D. Drudy, L. Kyne, K. Brown & N.F. Fairweather, (2008) Immunoreactive cell wall proteins of *Clostridium difficile* identified by human sera. *J. Med. Microbiol.* **57**: 750-756.
- Wuenscher, M.D., S. Kohler, A. Bubert, U. Gerike & W. Goebel, (1993) The *iap* gene of *Listeria monocytogenes* is essential for cell viability, and its gene product, p60, has bacteriolytic activity. *J. Bacteriol.* **175**: 3491-3501.
- Yildiz, F.H. & K.L. Visick, (2009) *Vibrio* biofilms: so much the same yet so different. *Trends Microbiol.* **17**: 109-118.
- Yoshino, Y., T. Kitazawa, M. Ikeda, K. Tatsuno, S. Yanagimoto, S. Okugawa, H. Yotsuyanagi & Y. Ota, (2013) *Clostridium difficile* flagellin stimulates toll-like receptor 5, and toxin B promotes flagellin-induced chemokine production via TLR5. *Life Sci.* **92**: 211-217.
- Zar, F.A., S.R. Bakkanagari, K.M. Moorthi & M.B. Davis, (2007) A comparison of vancomycin and metronidazole for the treatment of *Clostridium difficile*-associated diarrhea, stratified by disease severity. *Clin. Infect. Dis.* **45**: 302-307.
- Zhang, R., R. Wu, G. Joachimiak, S.K. Mazmanian, D.M. Missiakas, P. Gornicki, O. Schneewind & A. Joachimiak, (2004) Structures of sortase B from *Staphylococcus aureus* and *Bacillus anthracis* reveal catalytic amino acid triad in the active site. *Structure* **12**: 1147-1156.
- Zimmermann, H., (1992) 5'-Nucleotidase: molecular structure and functional aspects. *Biochem. J.* **285 (Pt 2)**: 345-365.
- Zink, S.D. & D.L. Burns, (2005) Importance of *srtA* and *srtB* for growth of *Bacillus anthracis* in macrophages. *Infect. Immun.* **73**: 5222-5228.
- Zong, Y., T.W. Bice, H. Ton-That, O. Schneewind & S.V. Narayana, (2004a) Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. *J. Biol. Chem.* **279**: 31383-31389.
- Zong, Y., S.K. Mazmanian, O. Schneewind & S.V. Narayana, (2004b) The structure of sortase B, a cysteine transpeptidase that tethers surface protein to the *Staphylococcus aureus* cell wall. *Structure* **12**: 105-112.
- Zong, Y., Y. Xu, X. Liang, D.R. Keene, A. Hook, S. Gurusiddappa, M. Hook & S.V. Narayana, (2005) A 'Collagen Hug' model for *Staphylococcus aureus* CNA binding to collagen. *EMBO J.* **24**: 4224-4236.

A Materials and Methods tables

Table A.1: Oligonucleotides used in this study

Name	Oligonucleotide sequence
<u>C. difficile mutagenesis</u>	
CD2718 F	TTGTATTCTAACAGAAAAGCCAA
CD2718 R	TTAAATCAATCTACCATGA
IBS-213s	AAAAAAGCTTATAATTATCCTAAATATCAATTATGTGCGCCCAGATAGGGTG
EBS16-213s	CAGATTGTACAAATGTGGTGATAACAGATAAGTCATTATCCTAACTTACCTTCTTG
EBS2-213s	TGAACGCAAGTTCTAATTCGGTTATTCCGATAGAGGAAAGTGTCT
EBS Universal	CGAAATTAGAAACTTGC GTT CAG TAAAC
RAM-F	ACGCGTTATATTGATAAAAATAATAATAGTGGG
RAM-R	ACGCGT GCG ACT CAT AGA ATT ATTCCTCCG
CD0386-F	ATGGAACGCCTAAATGACG
CD2831-F	CCAGAGCAAAC TCCAGAAGG
CD3392-F	GGTGCAACCTACGGTGTCTT
cspfdxF1	GATGTAGATAGGATAATAGAATCCATAGAAAATATAGG
pMTL007-F1	TTAAGGAGGTGTATTCATATGACCATGATTACG
pMTL007-R1	AGGGTATCCCCAGTTAGTGTAAAGTCTTGG
<u>Sequencing primers</u>	
M13-F	TGTAAAACGACGCCAGT
M13-R	CAGGAAACAGCTATGACC
T7-F	AATACGACTCACTATAGGG
T7-R	GCTAGTTATTGCTCAGCGG
pQE30Xa-F	CGGATAACAATTCACACAG
pQE30-F	CCCGAAAAGTGCCACCTG
pQE30-R	GTTCTGAGGTCAATTACTGG
pMTL960-F	TTGACTTTAACGCTACGAATACCATA
pMTL960-R	CCAAGGAGCTAAAGAGGTCCC
pXi-R	CCCCAAGGGGTTATGCTAGT
<u>Sortase expression constructs*</u>	
pET_3	GATATT <u>CCATGG</u> ATGAAGAAA <u>CTGTACCGTATCGTTATC</u>
pET_4	GATGAG <u>CTCGAGG</u> ATCAGACGACC <u>GTGGATAAC</u> C
pET_15	GATATA <u>CCATGG</u> AT <u>GCACCA</u> CCAC <u>ACCCAC</u> CTGACCAAA <u>ATACAACCACGACAC</u>
pET_16	GATGAG <u>CTCGAG</u> TTAGATCAGACGACC <u>GTGGATAAC</u>
pET_17	GATATA <u>CCATGG</u> AT <u>GCACCA</u> CCAC <u>ACCCAC</u> ACT <u>CTAA</u> CTGACCAAA <u>ATACAACCACGACAC</u>
C209A	TCGTTACCTGTCTACCGCCACCTACGAATT <u>CGACG</u>
C209A antisense	CGTCGAATT <u>CGTAGG</u> GGCGGTAGACAGGGTAACGA
pET_18	GATATA <u>CCATGG</u> AT <u>GCACCA</u> CCAC <u>ACCCAC</u> AGAC <u>CTACATCGAAGACAAAGCAG</u>
pET_19	GATGAG <u>CTCGAG</u> TTAGAAA <u>ACTTTGATAATCTTGCCAC</u>
pOE_1	<u>CGCGAATT</u> CGATATAAA <u>ACTAC</u> TTAAC <u>CTCGAGGTGATATATTGAA</u>
pOE_2	<u>GCCGGATC</u> TTAG <u>GTGGTGGTGGTGGTGGT</u> GCTCAAGAAT <u>CAATCTACCATGAATTACCAATTCTGGC</u>
Eliz_1	CTACATTAA <u>CTCGAGGTGATATATTGCTTACCAAA</u> UACAAT <u>CTGATACTAA</u>
Eliz_2	TTAGTAT <u>CATGATTGATTTGGTAA</u> GAAT <u>ATATCACCTCGAGTTAAATGTAG</u>
<u>Substrate expression constructs*</u>	
pET-3392 1	CATATA <u>CCATGG</u> AT <u>GCACCA</u> CCAC <u>ACCCAC</u> CGTT <u>ATCCGTGGCGGTGTGAA</u>

pET-3392 3	GATGAG <u>CTCGAG</u> TACTTCTGA <u>ACTGAGGGTGTGACC</u> ATGATTCTCATTTCACGGCGTT
pET_18B	GATA <u>TACCATGG</u> ATGC <u>ACCA</u> CC <u>ACCA</u> CC <u>ACCA</u> CG <u>CAG</u> AG <u>CCA</u> ACT <u>A</u> ATT <u>C</u> ACC
pET_19A	GATA <u>ACTCGAG</u> TATT <u>GTA</u> AAA <u>ACAC</u> GT <u>GT</u> AG <u>CG</u> AG
pET_19B	GATA <u>ACTCGAG</u> TACTTCT <u>CGA</u> <u>ACTGAGGGTGTGACC</u> ATT <u>GTA</u> AA <u>ACAC</u> GT <u>GT</u> AG <u>CG</u> TTA
pET_19C	GATA <u>ACTCGAG</u> TT <u>GTA</u> AAA <u>ACAC</u> GT <u>GT</u> AG <u>CG</u> AG <u>TTA</u>
pET_20	GATA <u>ACCATGG</u> AT <u>GC</u> <u>ACCA</u> CC <u>ACCA</u> CC <u>ACCA</u> AT <u>GC</u> AG <u>ATG</u> AA <u>GT</u> AA <u>ATG</u> ATC
pET_21A	GATA <u>ACTCGAG</u> TT <u>ATA</u> AT <u>ATT</u> CT <u>TTT</u> G <u>CT</u> <u>GT</u> <u>TA</u> ACA
pET_21B	GATA <u>ACTCGAG</u> TACTTCT <u>CGA</u> <u>ACTGAGGGTGTGACC</u> ATA <u>AT</u> ATT <u>CTT</u> TT <u>GCT</u> <u>GT</u> <u>TA</u> ACAA <u>AT</u>
pET_21C	GATA <u>ACTCGAG</u> TA <u>AT</u> ATT <u>CTT</u> TT <u>GCT</u> <u>GT</u> <u>TA</u> ACAA <u>AT</u>
pET_22	GATA <u>ACCATGG</u> AT <u>GC</u> <u>ACCA</u> CC <u>ACCA</u> CC <u>ACCA</u> CA <u>AA</u> AG <u>AT</u> TT <u>GGT</u> AG <u>AA</u> AT <u>AGT</u> GG
pET_23A	GATA <u>ACTCGAG</u> CT <u>TA</u> ATT <u>GT</u> ATT <u>TTT</u> ATT <u>CT</u> <u>CT</u> <u>TA</u> AT <u>AC</u>
pET_23B	GATA <u>ACTCGAG</u> CT <u>ACT</u> CT <u>CGA</u> <u>ACTGAGGGTGTGACC</u> AA <u>TTG</u> TA <u>TTT</u> ATT <u>CT</u> <u>CT</u> <u>TA</u> AT <u>AC</u> G
pET_23C	GATA <u>ACTCGAG</u> ATT <u>GT</u> ATT <u>TTT</u> ATT <u>CT</u> <u>CT</u> <u>TA</u> AT <u>AC</u> G
pET_25A	GATA <u>ACTCGAG</u> TT <u>GT</u> ATT <u>CT</u> <u>CT</u> <u>TA</u> AT <u>AC</u> G
pET_25B	GATA <u>ACTCGAG</u> TT <u>ACT</u> CT <u>CGA</u> <u>ACTGAGGGTGTGACC</u> AT <u>G</u> ATT <u>CT</u> <u>CT</u> <u>TA</u> AT <u>AC</u> GG
pET_25C	GATA <u>ACTCGAG</u> GT <u>ATT</u> CT <u>CT</u> <u>TA</u> AT <u>AC</u> GG
pET_26	GATA <u>ACCATGG</u> AT <u>GC</u> <u>ACCA</u> CC <u>ACCA</u> CC <u>ACCA</u> CA <u>ATG</u> GG <u>AT</u> AA <u>AAA</u> AA <u>AAA</u> AGT
pET_29	GATA <u>ACCATGG</u> AT <u>GC</u> <u>ACCA</u> CC <u>ACCA</u> CC <u>ACCA</u> CA <u>AAA</u> AG <u>TATG</u> AG <u>TTG</u> CA <u>ACG</u> AA
pET_30	GATA <u>ACTCGAG</u> TT <u>TTG</u> ACT <u>TTT</u> AT <u>CT</u> <u>CT</u> <u>TA</u> ACT <u>CT</u>
pET_33	GATA <u>ACCATGG</u> AT <u>GC</u> <u>ACCA</u> CC <u>ACCA</u> CC <u>ACCA</u> CA <u>GG</u> AA <u>ACAC</u> AG <u>TACAA</u> ACT <u>GT</u> GG
pET_34	GATA <u>ACTCGAG</u> GT <u>ATT</u> CT <u>CT</u> <u>TA</u> AT <u>AC</u> GG
cwpV-R	TCTGCTGTTCTCTGTTG
RT-PCR	
0183_F	AAAGATGGGGAAACAGAAAGGA
0183_R	CCCCAAACATATGGACAACC
0386_2F	GGAGATATTGAGGGCGTGAA
0386_2R	TTTGCTGTCGTCGGTCAG
0386-inter_F	GACCTAGCGACAGTCCAAA
0386-inter_R	GCACCCTTGACTATCCGTGT
2537_2F	CTTTTGCAAGTGGGAGGAA
2537_2R	TGGCTTTGAAAAACAAGG
2718_1F	AAATTTGGCTTCTGTTGAGA
2718_1R	AGGTTGTTAACATAGTTTATTTTC
2768-inter_1F	TTATGAGGGATTAGTGCCAACT
2768-inter_1R	TTGCTCCTGCTTCTACTGTGA
2768_F	TAAATGGGAGCAACTGGAC
2768_R	TGAACCACCTGAACCAAAGA
CD2831-int3 F	AAAAGAAGGGAACCTGTGC
CD2831-int3 R	GCTTCAAAAGCATAACCTTGC
3145_F	ATAGGTGGAGGCGGTAAATGG
3145_R	ACTACCATGTGCACCTCCAT
3146_1F	GCATTATTGCGCTAACCTCTAG
3146_1R	GTTATATCCTGTACTACAGGATAG
3146_2F	ATCAGTACACAAATCTGATATAAAGG
3146_2R	TACACAAACCTACTTATAGGTCTC
3246_1F	AGCAAGCTAAAAGACGACA
3246_1R	TCCATCAAATTAGAACCTGCA
CD3392-int2 F	TATCCATACCACAGCGACCA

CD3392-int2_R	TGATTAGCTGGCGTTTCT
atpA_F	AAAAGGTGGGGTAGCAGCTA
atpA_R	CGCCTGTTGGGTCTTT
16srRNA_F	GGCAGCAGTGGGAATATTG
16srRNA_R	CCGTAGCCTTCACTCCTGA
sigB_F	TTCTCCAGCTTGCAACACC
sigB_R	AAGAAGCCTCATAGCCTCT

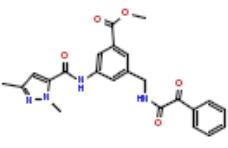
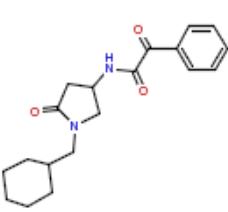
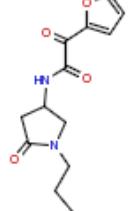
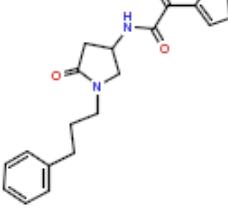
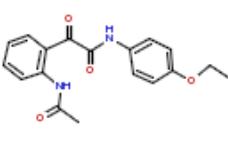
*Underlined regions are restriction sites introduced by primers: CCATGG, Ncol; CTCGAG, Xhol; GAATTC, HindIII; and GGATCC, BamHI. Bolded regions are epitope tags introduced by primers: **CACCACCA****CCAC****CCAC****CCAC** (5'-3') and **GTGGTGGTGGTGGTGGT** (3'-5'), 6xHis tag; and **CTTCTCGAACTGAGGGTGTGACCA** (3'-5'), Strep tag.

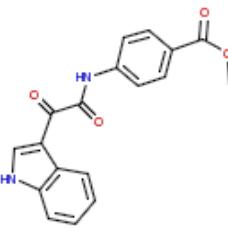
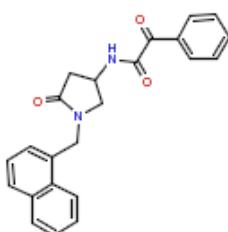
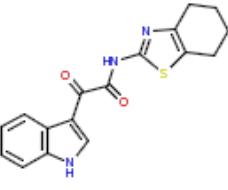
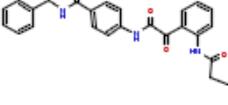
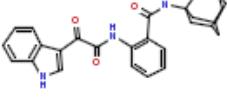
Table A.2: Plasmids used in this study

Name	Details	Description	Reference
<u><i>C. difficile</i> mutagenesis</u>			
pMTL007		ClosTron plasmid, IPTG-inducible RAM expression	Heap <i>et al.</i> , 2007
pEHD001	pMTL007::Cdi- <i>CD2718</i> -213s	ClosTron plasmid targeted to insert between bp 213/214 of <i>CD2718</i>	L. Dawson, LSHTM
pMTL007C-E2		Second generation ClosTron plasmid, constitutive RAM expression	Heap <i>et al.</i> , 2010
pEHD002	pMTL007C-E2::Cdi- <i>CD2718</i> -510s	ClosTron plasmid targeted to insert between bp 510/511 of <i>CD2718</i>	DNA2.0, CA, USA
pEHD003	pMTL007C-E2::Cdi- <i>CD0386</i> -32s	ClosTron plasmid targeted to insert between bp 32/33 of <i>CD0386</i>	DNA2.0, CA, USA
pEHD004	pMTL007C-E2::Cdi- <i>CD2831</i> -540s	ClosTron plasmid targeted to insert between bp 540/541 of <i>CD2831</i>	DNA2.0, CA, USA
pEHD005	pMTL007C-E2::Cdi- <i>CD3392</i> -840s	ClosTron plasmid targeted to insert between bp 840/841 of <i>CD3392</i>	DNA2.0, CA, USA
<u><i>C. difficile</i> overexpression</u>			
pMTL960		<i>E. coli/C. difficile</i> shuttle vector	Purdy <i>et al.</i> , 2002
pCBR023		pUC19 backbone containing <i>cwp2</i> promoter	Emerson <i>et al.</i> , 2009
pEHD006	pCBR023- <i>CD2718</i>	<i>CD2718</i> under the control of the <i>cwp2</i> promoter	This work
pEHD007	pMTL960-pcwp2- <i>CD2718</i>	<i>CD2718</i> under the control of the <i>cwp2</i> promoter	This work
pEHD008	pMTL960-pcwp2- <i>CD2718</i> _{ΔN28}	Residues 28-225 of <i>CD2718</i> under the control of the <i>cwp2</i> promoter	This work
<u><i>E. coli</i> sortase expression</u>			
pEHD009	pQE30Xa- <i>CD2718</i>	Codon optimised <i>CD2718</i> , C-terminal 6xHis tag	Celtek Bioscience LLC, TN, USA
pEHD010	pQE30Xa- <i>SaSrtB</i>	Codon optimised <i>SaSrtB</i> , C-terminal 6xHis tag	Celtek Bioscience LLC, TN, USA
pET28a		Commercial protein expression vector	Novagen, Merck Millipore, MA, USA
pEHD011	pET28a- <i>CD2718</i>	Codon optimised <i>CD2718</i>	This work
pEHD012	pET28a- <i>CD2718</i> _{ΔN28}	<i>CD2718</i> with residues 2-27 replaced with a 6xHis tag	This work
pEHD013	pET28a- <i>CD2718</i> _{ΔN26}	<i>CD2718</i> with residues 2-25 replaced with a 6xHis tag	This work
pEHD014	pET28a- <i>CD2718</i> _{ΔN26} , C209A	Same as above, with C209A substitution	This work
pEHD015	pET28a- <i>SaSrtB</i> _{ΔN26}	<i>SaSrtB</i> with residues 26-244 replaces with 6xHis tag	This work
<u><i>E. coli</i> substrate expression</u>			
pUCIrv7	pXi- <i>CD2718</i>	<i>CD2718</i> with an N-terminal 10xHis tag, and a C-terminal HA tag	P. Felgner, Univeristy of California, Irvine
pUCIrv17	pXi- <i>R20291_3453</i> -s2	Residues 435-901 of <i>R20291_3453</i> with above tags	P. Felgner, Univeristy of California, Irvine
pUCIrv20	pXi- <i>R20291_2722</i> -s2	Residues 486-984 of <i>R20291_2722</i> with above tags	P. Felgner, Univeristy of California, Irvine
pUCIrv26	pXi- <i>CD2831</i> -s2	Residues 434-972 of <i>CD2831</i> with above tags	P. Felgner, Univeristy of California, Irvine
pUCIrv27	pXi- <i>CD0386</i> -s2	Residues 451-1014 of <i>CD0386</i> with above tags	P. Felgner, Univeristy of California, Irvine

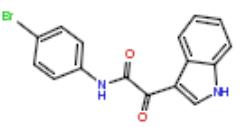
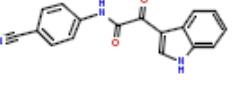
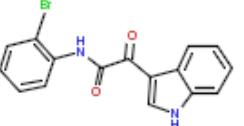
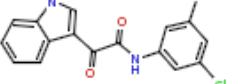
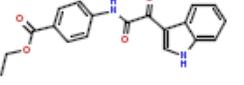
pUCIrv28	pXi-CD3392-s2	Residues 451-1014 of <i>CD3392</i> with above tags	P. Felgner, Univeristy of California, Irvine
pEHD016	pET28a-CD3392s2-Strep	Above <i>CD3392</i> fragment with N-terminal 6xHis tag, and a C-terminal Strep tag	This work
pEHD017	pET28a-CD0183-Strep	Residues 23-340 of <i>CD0183</i> with N-terminal 6xHis tag, and C-terminal Strep tag	This work
pEHD018	pET28a-CD0183-His	Residues 23-340 of <i>CD0183</i> with N- and C-terminal 6xHis tags	This work
pEHD019	pET28a-CD2768	Residues 36-235 of <i>CD2768</i> with N-terminal 6xHis tag	This work
pEHD020	pET28a-CD2768-Strep	Residues 36-235 of <i>CD2768</i> with N-terminal 6xHis tag, and C-terminal Strep tag	This work
pEHD021	pET28a-CD2768-His	Residues 36-235 of <i>CD2768</i> with N- and C-terminal 6xHis tags	This work
pEHD022	pET28a-CD2831	Residues 811-972 of <i>CD2831</i> with N-terminal 6xHis tag	This work
pEHD023	pET28a-CD2831-Strep	Residues 811-972 of <i>CD2831</i> with N-terminal 6xHis tag, and C-terminal Strep tag	This work
pEHD024	pET28a-CD3392	Residues 841-1014 of <i>CD3392</i> with N-terminal 6xHis tag	This work
pEHD025	pET28a-CD3392-Strep	Residues 841-1014 of <i>CD3392</i> with N-terminal 6xHis tag, and C-terminal Strep tag	This work
pEHD026	pET28a-CD3392-His	Residues 841-1014 of <i>CD3392</i> with N- and C-terminal 6xHis tags	This work
pJKP006	pET28a-CD0183-CwpV	Residues 24-340 of <i>CD0183</i> with an N-terminal 6xHis tag, and two C-terminal CwpV type II repeats	J. Peltier, Imperial College London, UK
pEHD027	pET28a-CD2537	Residues 450-613 of <i>CD2537</i> with an N-terminal 6xHis tag	This work
pEHD028	pET28a-CD2537-CwpV	Residues 450-613 of <i>CD2537</i> with an N-terminal 6xHis tag, and two C-terminal CwpV type II repeats	This work
pEHD029	pET28a-CD2768	Residues 123-235 of <i>CD2768</i> with an N-terminal 6xHis tag	This work
pEHD030	pET28a-CD2768-CwpV	Residues 123-235 of <i>CD2768</i> with an N-terminal 6xHis tag, and two C-terminal CwpV type II repeats	This work
pEHD031	pET28a-CD2831-CwpV	Residues 811-972 of <i>CD2831</i> with an N-terminal 6xHis tag, and two C-terminal CwpV type II repeats	This work
pEHD032	pET28a-CD3392	Residues 862-1014 of <i>CD3392</i> with an N-terminal 6xHis tag	This work
pEHD033	pET28a-CD3392-CwpV	Residues 862-1014 of <i>CD3392</i> with an N-terminal 6xHis tag, and two C-terminal CwpV type II repeats	This work

Table A.3: Sortase inhibitor details

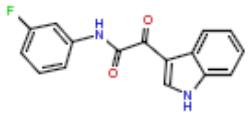
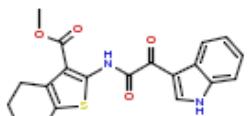
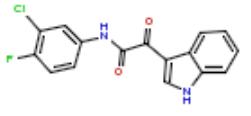
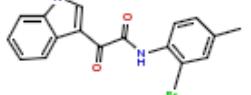
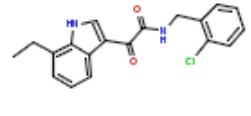
Inhibitor Name	Structure	Chemical formula	Molecular weight	Source
LSHTM-0005		C ₂₃ H ₂₂ N ₄ O ₅	434.4	ChemBridge Corp., CA, USA
LSHTM-0007		C ₁₉ H ₂₄ N ₂ O ₃	328.4	ChemBridge Corp., CA, USA
LSHTM-0008		C ₁₃ H ₁₆ N ₂ O ₄	264.3	ChemBridge Corp., CA, USA
LSHTM-0009		C ₁₉ H ₂₀ N ₂ O ₄	340.4	ChemBridge Corp., CA, USA
LSHTM-0010		C ₁₈ H ₁₈ N ₂ O ₄	326.3	ChemBridge Corp., CA, USA

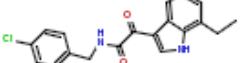
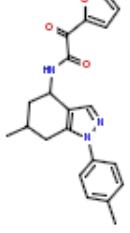
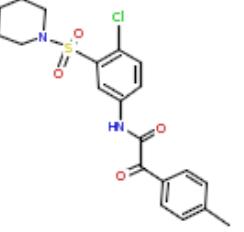
LSHTM-0011		C ₁₈ H ₁₄ N ₂ O ₄	322.3	ChemBridge Corp., CA, USA
LSHTM-0012		C ₂₃ H ₂₀ N ₂ O ₃	372.4	ChemBridge Corp., CA, USA
LSHTM-0013		C ₁₇ H ₁₅ N ₃ O ₂ S	325.4	ChemBridge Corp., CA, USA
LSHTM-0014		C ₂₅ H ₂₃ N ₃ O ₄	429.5	ChemBridge Corp., CA, USA
LSHTM-0015		C ₂₇ H ₂₇ N ₃ O ₃	441.5	ChemBridge Corp., CA, USA

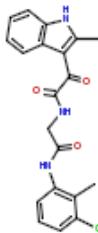
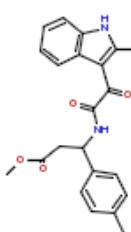
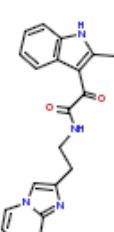
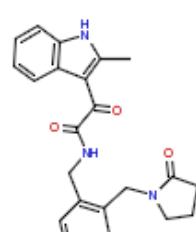
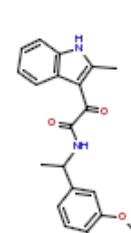
LSHTM-0016		C ₂₂ H ₁₇ ClN ₂ O ₄	408.8	ChemBridge Corp., CA, USA
LSHTM-0017		C ₂₁ H ₂₂ N ₂ O ₃	350.4	ChemBridge Corp., CA, USA
LSHTM-0018		C ₂₀ H ₁₉ N ₃ O ₃	349.4	ChemBridge Corp., CA, USA
LSHTM-0019		C ₁₇ H ₁₂ ClFN ₂ O ₂	330.7	ChemBridge Corp., CA, USA
LSHTM-0020		C ₁₈ H ₁₄ N ₂ O ₄	322.3	ChemBridge Corp., CA, USA

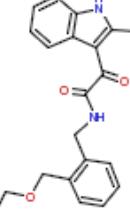
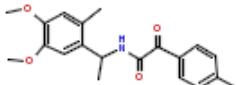
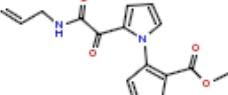
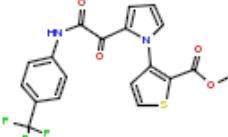
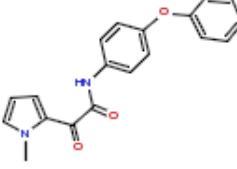
LSHTM-0021		C ₁₆ H ₁₁ Br N ₂ O ₂	343.2	ChemBridge Corp., CA, USA
LSHTM-0022		C ₁₇ H ₁₁ N ₃ O ₂	289.3	ChemBridge Corp., CA, USA
LSHTM-0023		C ₁₆ H ₁₁ Br N ₂ O ₂	343.2	ChemBridge Corp., CA, USA
LSHTM-0024		C ₁₆ H ₁₀ Cl ₂ N ₂ O ₂	333.2	ChemBridge Corp., CA, USA
LSHTM-0025		C ₁₉ H ₁₆ N ₂ O ₄	336.3	ChemBridge Corp., CA, USA

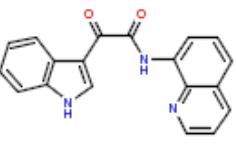
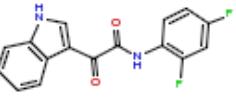
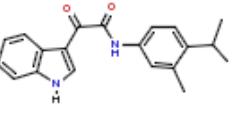
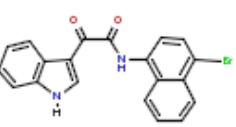
LSHTM-0026		C ₁₈ H ₁₆ N ₂ O ₂	292.3	ChemBridge Corp., CA, USA
LSHTM-0027		C ₁₇ H ₁₁ N ₃ O ₂ S	321.4	ChemBridge Corp., CA, USA
LSHTM-0028		C ₁₇ H ₁₃ Cl N ₂ O ₂	312.8	ChemBridge Corp., CA, USA
LSHTM-0029		C ₁₉ H ₁₆ N ₂ O ₄	336.3	ChemBridge Corp., CA, USA
LSHTM-0030		C ₁₈ H ₁₇ N ₃ O ₂	307.3	ChemBridge Corp., CA, USA

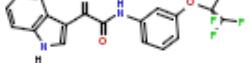
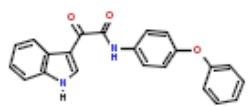
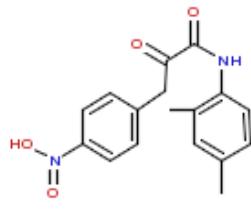
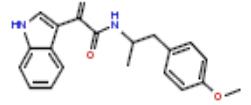
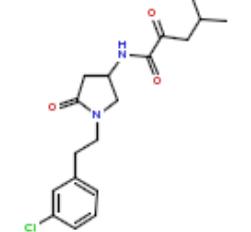
LSHTM-0031		C ₁₆ H ₁₁ FN ₂ O ₂	282.3	ChemBridge Corp., CA, USA
LSHTM-0032		C ₂₀ H ₁₈ N ₂ O ₄ S	382.4	ChemBridge Corp., CA, USA
LSHTM-0033		C ₁₆ H ₁₀ ClFN ₂ O ₂	316.7	ChemBridge Corp., CA, USA
LSHTM-0034		C ₁₇ H ₁₃ BrN ₂ O ₂	357.2	ChemBridge Corp., CA, USA
LSHTM-0035		C ₁₉ H ₁₇ ClN ₂ O ₂	340.8	ChemBridge Corp., CA, USA

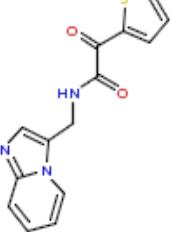
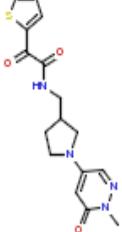
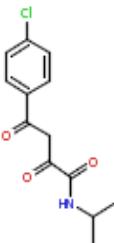
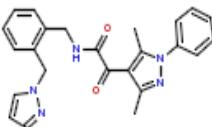
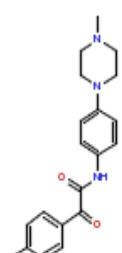
LSHTM-0036		C ₁₉ H ₁₇ ClN ₂ O ₂	340.8	ChemBridge Corp., CA, USA
LSHTM-0037		C ₁₈ H ₁₆ N ₂ O ₂	292.3	ChemBridge Corp., CA, USA
LSHTM-0038		C ₁₉ H ₁₈ N ₂ O ₂	306.4	ChemBridge Corp., CA, USA
LSHTM-0039		C ₂₂ H ₂₃ N ₃ O ₃	377.4	ChemBridge Corp., CA, USA
LSHTM-0040		C ₂₀ H ₂₁ ClN ₂ O ₄ S	420.9	Enamine Ltd., Kiev, Ukraine

LSHTM-0041		C ₂₀ H ₁₈ ClN ₃ O ₃	383.8	Enamine Ltd., Kiev, Ukraine
LSHTM-0042		C ₂₂ H ₂₂ N ₂ O ₄	378.4	Enamine Ltd., Kiev, Ukraine
LSHTM-0043		C ₂₀ H ₁₈ N ₄ O ₂	346.4	Enamine Ltd., Kiev, Ukraine
LSHTM-0044		C ₂₃ H ₂₃ N ₃ O ₃	389.4	Enamine Ltd., Kiev, Ukraine
LSHTM-0045		C ₂₂ H ₂₂ N ₂ O ₄	378.4	Enamine Ltd., Kiev, Ukraine

LSHTM-0046		C ₂₁ H ₂₂ N ₂ O ₃	350.4	Enamine Ltd., Kiev, Ukraine
LSHTM-0047		C ₂₀ H ₂₃ NO ₄	341.4	Enamine Ltd., Kiev, Ukraine
LSHTM-0048		C ₁₅ H ₁₄ N ₂ O ₄ S	318.3	Key Organics Ltd., Cornwall, UK
LSHTM-0049		C ₁₉ H ₁₃ F ₃ N ₂ O ₄ S	422.4	Key Organics Ltd., Cornwall, UK
LSHTM-0050		C ₁₉ H ₁₆ N ₂ O ₃	320.3	Key Organics Ltd., Cornwall, UK

LSHTM-0051		C ₁₃ H ₁₀ F ₂ N ₂ O ₂	264.2	Key Organics Ltd., Cornwall, UK
LSHTM-0052		C ₁₉ H ₁₃ N ₃ O ₂	315.3	Key Organics Ltd., Cornwall, UK
LSHTM-0053		C ₁₆ H ₁₀ F ₂ N ₂ O ₂	300.3	Key Organics Ltd., Cornwall, UK
LSHTM-0054		C ₂₀ H ₂₀ N ₂ O ₂	320.4	Key Organics Ltd., Cornwall, UK
LSHTM-0055		C ₂₀ H ₁₃ BrN ₂ O ₂	393.2	Key Organics Ltd., Cornwall, UK

LSHTM-0056		C ₁₈ H ₁₂ F ₄ N ₂ O ₃	380.3	Key Organics Ltd., Cornwall, UK
LSHTM-0057		C ₂₂ H ₁₆ N ₂ O ₃	356.4	Key Organics Ltd., Cornwall, UK
DMX-0001		C ₁₇ H ₁₆ N ₂ O ₄	314.3	ChemBridge Corp., CA, USA
DMX-0002		C ₂₀ H ₂₀ N ₂ O ₃	336.4	ChemBridge Corp., CA, USA
DMX-0007		C ₁₈ H ₂₃ ClN ₂ O ₃	350.8	ChemBridge Corp., CA, USA

DMX-0009		C ₁₄ H ₁₁ N ₃ O ₂ S	285.3	ChemBridge Corp., CA, USA
DMX-0012		C ₁₆ H ₁₈ N ₄ O ₃ S	346.4	ChemBridge Corp., CA, USA
DMX-0014		C ₁₃ H ₁₄ CINO ₃	267.7	Enamine Ltd., Kiev, Ukraine
DMX-0016		C ₂₄ H ₂₃ N ₅ O ₂	413.5	Enamine Ltd., Kiev, Ukraine
DMX-0017		C ₂₀ H ₂₃ N ₃ O ₂	337.4	Enamine Ltd., Kiev, Ukraine

DMX-0018		C ₂₁ H ₂₅ N ₅ O ₃	395.5	Enamine Ltd., Kiev, Ukraine
DMX-0026		C ₂₃ H ₂₉ N ₃ O ₅	427.5	ChemBridge Corp., CA, USA
DMX-0028		C ₁₂ H ₁₂ N ₂ O ₃ S	264.3	ChemBridge Corp., CA, USA
DMX-0029		C ₂₂ H ₂₃ N ₃ O ₄	393.4	Enamine Ltd., Kiev, Ukraine
DMX-0030		C ₁₇ H ₂₀ N ₄ O ₃	328.4	Enamine Ltd., Kiev, Ukraine

B Construction of plasmids

All plasmid constructs were verified with Sanger sequencing carried out by Source BioScience LifeSciences (Nottingham, UK).

B.1 Construction of *C. difficile* overexpression vectors, pEHD007 and pEHD008

CD2718 was amplified from 630 genomic DNA using primers pOE-1 and pOE-2, and cloned into the EcoRI and BamHI sites of pCBR023 (kindly provided by Neil Fairweather, Imperial College London (Emerson *et al.*, 2009)), a pUC19-based vector containing the *cwp2* promoter which is moderately expressed in *C. difficile*, to create pEHD006. The *pcwp2-CD2718* fragment was excised from this vector using Acc65I and BamHI and cloned into the corresponding sites of the *E. coli* – *C. difficile* shuttle vector pMTL960 (kindly provided by N. Fairweather (Purdy *et al.*, 2002)), to create pEHD007. Site-directed mutagenesis primers Eliz-1 and Eliz-2 were used to delete residues 2-27 of CD2718 from pEHD007, using the QuikChange Site-Directed Mutagenesis kit (Agilent Technologies Inc., CA, USA) in accordance with the manufacturer's instructions, to create plasmid pEHD008. The insert in pCBR023 was sequenced using primers M13-F and M13-R, and the inserts in pMTL960 were sequenced using primers pMTL960-F and pMTL960-R.

B.2 Construction of vectors for sortase expression in *E. coli* (pEHD009-pEHD015)

The coding sequences for CD2718 and SrtB from *S. aureus* (Newman strain) were codon-optimised for expression in *E. coli* and cloned into the BamHI/HindIII sites of pQE30Xa by Celtek Bioscience, LLC (Nashville, TN, USA) to create plasmids pEHD009 and pEHD010, respectively. The cloned inserts were sequenced using the primers pQE30Xa-F and pQE30-R, and the promoter region of the plasmid was sequenced using pQE30-F.

CD2718 was amplified from pEHD009 using primers pET-3 and pET-4, and cloned into the Ncol and Xhol sites of pET28a (Novagen), a commercial expression vector with a C-terminal 6xHis tag (kindly provided by N. Fairweather), to construct pEHD011. Cloned inserts in pET28a were confirmed by sequencing using primers T7-F and T7-R. *CD2718_{ΔN28}* was amplified from pEHD011 using the primers pET-15 and pET-16 to replace the N-

terminal transmembrane domain (residues 2–27) with a 6xHis tag, and the product cloned into the Ncol and Xhol sites of pET28a to construct pEHD012. *CD2718_{ΔN26}* was amplified from pEHD011 using the primers pET-17 and pET-16 to replace residues 2–25 with a 6xHis tag, and the product cloned into the Ncol and Xhol sites of pET28a to construct pEHD013. Site-directed mutagenesis primers C209A and C209A_antisense were used to replace the cysteine residue at position 209 in *CD2718_{ΔN26}* with an alanine, using the QuikChange Site-Directed Mutagenesis kit (Agilent Technologies Inc., CA, USA) in accordance with the manufacturer's instructions, to create plasmid pEHD014. *SaSrtB_{ΔN26}* was amplified from pEHD010 using the primers pET-18 and pET-19 to replace residues 2–25 with a 6xHis tag, and the product cloned into the Ncol and Xhol sites of pET28a to construct pEHD015.

B.3 Construction of vectors for sortase substrate expression in *E. coli* (pEHD016-pEHD033)

Residues 451-1014 of *CD3392* were amplified from pUCIrv28 using primers pET-3392-1 and pET-3392-3, and cloned into the Ncol and Xhol sites of pET28a to construct pEHD016. Residues 23-340 of *CD0183* were amplified from *C. difficile* 630 genomic DNA using primers pET-18B and pET-19B, and pET-18B and pET-19C, and the resulting fragments cloned into the Ncol and Xhol sites of pET28a to construct pEHD017 and pEHD018, respectively. Residues 36-235 of *CD2768* were amplified from *C. difficile* 630 genomic DNA using primers pET-20 and pET-21A, pET-20 and pET-21B, and pET-20 and pET-21C, and the resulting fragments cloned into the Ncol and Xhol sites of pET28a to construct pEHD019, pEHD020, and pEHD021, respectively. Residues 811-972 of *CD2831* were amplified from *C. difficile* 630 genomic DNA using primers pET-22 and pET-23A, and pET-22 and pET-23B, and the resulting fragments cloned into the Ncol and Xhol sites of pET28a to construct pEHD022 and pEHD023, respectively. Residues 841-1014 of *CD3392* were amplified from *C. difficile* 630 genomic DNA using primers pET-24 and pET-25A, pET-24 and pET-25B, and pET-24 and pET-25C, and the resulting fragments cloned into the Ncol and Xhol sites of pET28a to construct pEHD024, pEHD025, and pEHD026, respectively.

Residues 450-613 of *CD2537* were amplified from *C. difficile* 630 genomic DNA using primers pET-29 and pET-30, and cloned into the Ncol and Xhol sites of pET28a to construct pEHD027. Residues 123-235 of *CD2768* were amplified from *C. difficile* 630 genomic DNA

using primers pET-26 and pET-19C, and cloned into the Ncol and Xhol sites of pET28a to construct pEHD029. Residues 862-1014 of *CD3392* were amplified from *C. difficile* 630 genomic DNA using primers pET-33 and pET-34, and cloned into the Ncol and Xhol sites of pET28a to construct pEHD032. A 706 bp Xhol-Xhol fragment containing two C-terminal CwpV type II repeats was excised from pJKP006, and cloned into the Xhol sites in pEHD027, pEHD029, pEHD023, and pEHD032, to construct the CwpV fusion protein constructs pEHD028, pEHD030, pEHD031, and pEHD033, respectively. The directionality of the cloned 706 bp Xhol-Xhol fragment was screened by colony PCR using primers T7-F and cwpV-R prior to sequencing.

C Python scripts for FRET data analysis

C.1 Python packages used

```
from scipy import stats
import scipy.optimize as optimize
import numpy as np
```

C.2 Functions used for curve-fit optimisation

```
def kinetic_function(params,S):
    k_cat, k_m = params
    v = (k_cat * S) / (k_m + S)
    return v

def kinetic_w_substrate_inhibition_function(params,S):
    k_cat, k_m, k_i = params
    v = (k_cat * S) / (k_m + S + (S*S)/(k_i))
    return v

def ic50_function(params,x):
    y_max, y_min, x50, hill = params
    y = y_min + (y_max-y_min)/(1.0 + ((x50/x)**(hill)))
    return y

def kinetic_residuals(params,S,V_0):
    return V_0 - kinetic_function(params,S)

def kinetic_w_substrate_inhibition_residuals(params,S,V_0):
    return V_0 - kinetic_w_substrate_inhibition_function(params,S)

def ic50_residuals(params,x,y0):
    return y0 - ic50_function(params,x)

def estim_errors(func, p, args, covar, data, initial_vals):
    residual = func(p, *args)
    reduced_chi_square =
        (residual**2).sum() / (len(data)-len(initial_vals))
    covar = covar * reduced_chi_square
    std_error = np.array(
        [np.sqrt(covar[i, i]) for i in range(len(p))])
    return std_error
```

C.3 Script for linear regression

```
slope, intercept, r_value, p_value, std_err = stats.linregress(x_data,y_data)
```

C.4 Script for least squares fit

```

p_guess=(0.0784/60.0,57.712, 60.1498)
(p,cov,infodict,mesg,ier)=optimize.leastsq(
    kinetic_w_substrate_inhibition_residuals,
    p_guess,
    args=(np.array(x_data),np.array(y_data)),
    full_output=1)
std_err = estim_errors(
    kinetic_w_substrate_inhibition_residuals,
    p,
    (np.array(x_data),np.array(y_data)),
    cov,
    np.array(y_data),
    p_guess)

p_guess=(0.0,300.0, 250.0, 5.0)
(p,cov,infodict,mesg,ier)=optimize.leastsq(
    ic50_residuals,
    p_guess,
    args=(np.array(x_data),np.array(y_data)),
    full_output=1)
std_err = estim_errors(
    ic50_residuals,
    p,
    (np.array(x_data),np.array(y_data)),
    cov,
    np.array(y_data),
    p_guess)

```

D Alternative sortase substrate orthologue tables

Table D.1: Gene ID of predicted sortase substrate orthologues

Function	Motif	Clade 1	Clade 2	Clade 3	Clade 4	Clade 5
		630	R20291	CD196	CD305	M68
			R20291_0184	CD196_0196	CD305_0316	CF5
Putative cell wall hydrolase	SPSTG	CD0183	R20291_0184	CD196_0196	CD305_0316	CDM68_01881
Putative cell surface protein, collagen binding protein	SPKTG	CD3392	R20291_3453	CD196_3407	CD305_0522	CDM68_03371
putative membrane-associated 5'-nucleotidase/phosphoesterase	SPKTG	CD2537	R20291_2424	CD196_2377	CD305_2758	CDM68_25791
putative cell wall hydrolase adhesin, collagen binding protein	SPQTG	CD2768	R20291_2656	CD196_2609	CD305_3010	CDM68_28181
CbpA, collagen binding protein	PPKTG	CD2831	R20291_2722	CD196_2675	CD305_3096	CDM68_28841
putative surface protein	NVQTG	CD3145	x	x	x	CDM68_31471
number of predicted substrates		8	5	5	7	8
						6

E Replicates of H-CD0183-CwpV cleavage assay

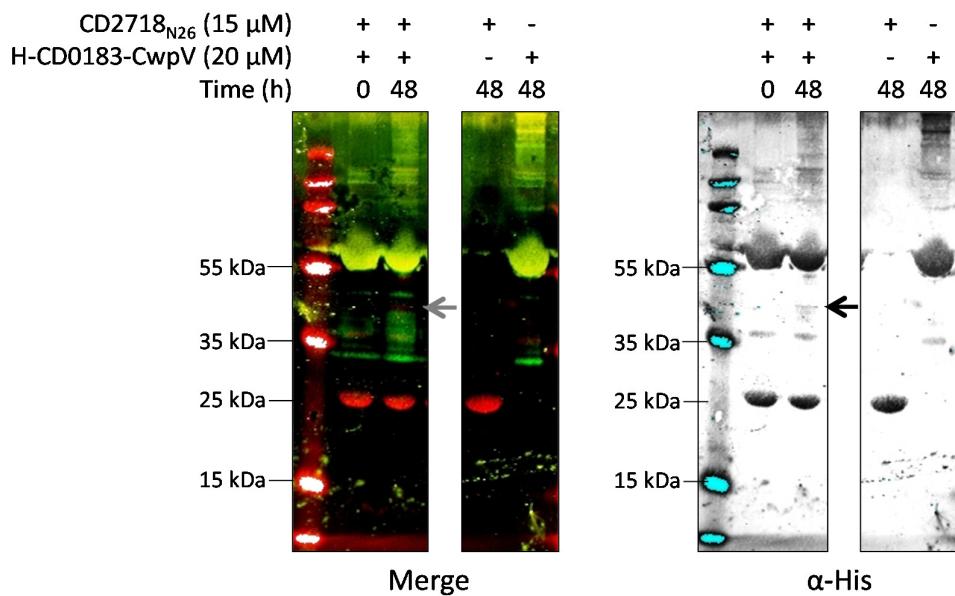


Figure 8.1: Replicate 2 of H-CD0183-CwpV cleavage assay

Western blots of H-CD0183-CwpV following incubation at 37 °C with CD2718_{ΔN26}. The reaction at $t=0$ and the proteins incubated individually over time served as controls. **Left:** Two colour anti-His (red) and anti-CwpV (green) western blot showing faint cleavage of H-CD0183-CwpV by CD2718_{ΔN26}. A band of approximately 40-50 kDa reacting with the anti-His antibody can be seen in the test well (indicated with an arrow), but not in either of the controls. **Right:** Black and white image of the His channel alone. The faint 40-50 kDa band observed in the His channel is more apparent on a light background. No 20 kDa band corresponding to the C-terminal cleavage product was observed in the CwpV channel during this replicate.

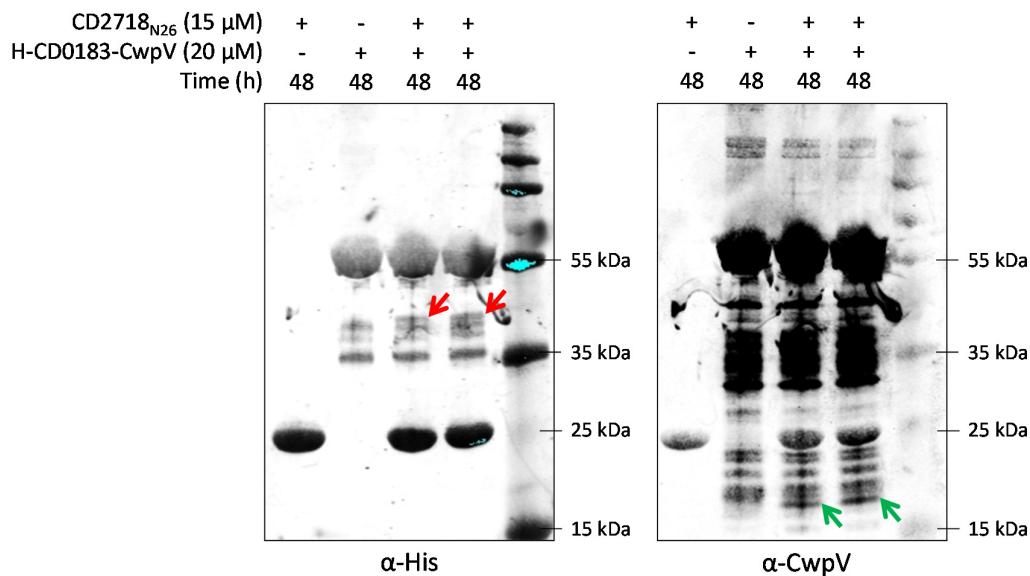


Figure 8.2: Replicate 3 of H-CD0183-CwpV cleavage assay

Western blots of H-CD0183-CwpV following incubation at 37 °C with CD2718_{N26} performed in duplicate. The proteins incubated individually over time served as controls. **Left:** Anti-His western blot showing faint cleavage of H-CD0183-CwpV by CD2718_{N26}. A band of approximately 40-50 kDa (corresponding to the N-terminal cleavage product) can be seen in both test wells (indicated with arrows), but not in either of the controls. **Right:** Anti-CwpV western blot showing faint cleavage of H-CD0183-CwpV by CD2718_{N26}. A band of approximately 20 kDa (corresponding to the C-terminal cleavage product) can be seen in the test wells (indicated with arrows), but not in any of the controls.

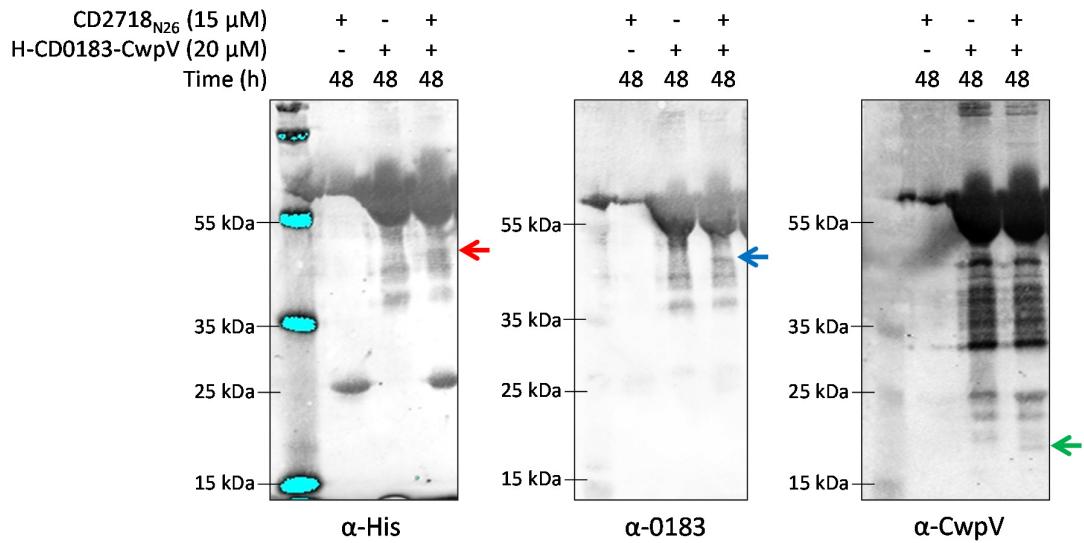


Figure 8.3: Replicate 4 of H-CD0183-CwpV cleavage assay

Western blots of H-CD0183-CwpV following incubation at 37 °C with CD2718_{ΔN26}. The proteins incubated individually over time served as controls. **Left:** Anti-His western blot showing faint cleavage of H-CD0183-CwpV by CD2718_{ΔN26}. A band of approximately 40-50 kDa (corresponding to the predicted N-terminal cleavage product) can be seen in the test well (indicated with an arrow), but not in either of the controls. **Middle:** Anti-CD0183 western blot detected the same 40-50 kDa band (indicated with an arrow) observed in the anti-His blot. This band was not present in the control lanes. **Right:** Anti-CwpV western blot showing faint cleavage of H-CD0183-CwpV by CD2718_{ΔN26}. A band of approximately 20 kDa (corresponding to the C-terminal cleavage product) can be seen in the test well (indicated with an arrow), but not in either of the controls.