

IDENTIFICATION OF NOVEL TRANSGLUTAMINASE-LIKE PROTEINS ASSOCIATED
WITH SPORULATION IN *CLOSTRIDIODES DIFFICILE*

TATIANA JOUMSI TAGNE
Doctor of Philosophy

ASTON UNIVERSITY
September 2021

©Tatiana Jousmi Tagne, 2021

Tatiana Jousmi Tagne asserts her moral right to be identified as the author of this thesis.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without appropriate permission and acknowledgement.

Aston University

IDENTIFICATION OF NOVEL TRANSGLUTAMINASE-LIKE PROTEINS ASSOCIATED WITH SPORULATION IN *CLOSTRIDIODES DIFFICILE*

A Thesis submitted by Tatiana Jousmi Tagne

For the degree of Doctor of Philosophy, 2021

Thesis Abstract

Clostridioides difficile is a spore-forming, Gram-positive, obligate anaerobe majorly implicated in nosocomial infections. It causes antibiotic-induced nosocomial diarrhoeas and related healthcare infections in most parts of the world. Dissemination of *C. difficile* infections occurs mainly via surfaces contaminated with spores especially in clinical settings and disease control is significantly hindered by spore formation. Spores are highly infectious, and resistant to heat, alcohol, and standard disinfectants. Transglutaminases are produced during sporulation in *Streptomyces mobaraense* and *Bacillus subtilis* and promote survival and virulence by modulating cellular protein crosslinking which increases resistance of cellular structures. Transglutaminases are enzymes in microbial and mammalian organisms that irreversibly crosslink proteins by forming covalent epsilon (gamma-glutamyl) lysine bonds that are proteolysis-resistant. The detection of novel important genes not explored previously by antibiotics can facilitate the discovery of alternative efficient antimicrobials to surmount existing resistance associated with disease control and treatment. The aim of this study was to identify novel transglutaminase-like genes associated with sporulation in *C. difficile* 630. We detected transglutaminase activity in sporulating cells and spores of CD 630. Also, transglutaminase gene expression was detected during sporulation. Three transglutaminase-like genes were identified in CD 630 and successfully cloned in various vectors and expressed in *E. coli*. No transglutaminase activity was detected in culture supernatants and in purified recombinant proteins. However, transglutaminase-like proteins were purified from sporulating *C. difficile* 630 cells by a two-step strategy to attain a specific activity and purification fold of 4 and 57 respectively. The purified proteins were characterized and found to be calcium-independent like most microbial transglutaminases, significantly inhibited by PMSF, and only inhibited to a small extent by thiol group inhibitors. Mass spectrometry analysis of the purified proteins suggests that they are mainly involved in metabolism in CD 630. We conclude that CD630 produces transglutaminase-like proteins during sporulation whose role necessitates functional analysis involving gene knockouts and the associated phenotypes of spores.

Keywords: transglutaminase, sporulation, *C. difficile*, purification, identification, characterization.

DEDICATION

To my dad, **Dr. Tagne Emmanuel** who passed in the last months of my studies, I dedicate this thesis to you! Thanks for being the endless beam of hope during this particularly long and difficult journey.

'Mon petit Papa' your sudden demise has left me without words and created a void in my heart. More so at a time when I was almost at the end of the journey for which you contributed immensely. Now that I cannot turn back the hands of time, I can only promise to honour you with a cascade of successes in my future endeavours. Rest on Papa!

ACKNOWLEDGEMENTS

Firstly, I thank the commonwealth scholarship commission in the U.K. for the financial support towards my studies and PhD research project. I would like to hugely appreciate the efforts of my supervisor Dr. Russel Collighan for the continual support and guidance, plus the freedom granted to manage my research. Gratitude also goes towards Prof. A.C. Hilton for the support during the initial stages of this work and granting me the opportunity of pursuing my research, and Dr. Tony Worthington for facilitating my research.

I am thankful to all those who provided support, advise and encouragements during this journey particularly Dr. Oluseyi Ayinde, Dr. Affiong Iyire, but also Dr. James Gavin, Dr. Shaun Fell, Fritz de la Rafa. A special thanks goes to current and previous members of labs MB327/MB633 for their suggestions and expertise, including making the past years a rich experience. I also thank Guan Cui for help with qPCR experimental guidance.

Most importantly, I am grateful to my family and friends for the constant love and support, especially my parents for always believing in me through these years. I am thankful for their encouragements and continuous reassurance. This could not have been a reality without them particularly my mum Tagne Henriette, my siblings Dr. Corinne, Romuald, Aubin and friend turned sister Juanita. Utmost gratitude goes to the almighty God for holding my hand during this particularly long and rough journey and making it a success in the end.

TABLE OF CONTENTS

Thesis Summary.....	2
DEDICATION.....	3
ACKNOWLEDGEMENTS.....	4
LIST OF ABBREVIATIONS.....	9
LIST OF TABLES.....	13
LIST OF FIGURES.....	14
CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW.....	16
1.1 Transglutaminases: Overview.....	16
1.1.1 Mammalian Transglutaminases.....	18
1.1.2 Microorganisms exploiting host transglutaminases.....	19
1.1.3 Microbial Transglutaminases.....	20
1.1.4 Evidence for the role of transglutaminases in sporulation.....	23
1.1.5 Microbial transglutaminases and their role in non-spore forming bacteria.....	25
1.1.6 Production of transglutaminase in microbial systems.....	26
1.1.7 Applications of microbial transglutaminase.....	30
1.1.8 Transglutaminase inhibition.....	31
1.2 CLOSTRIDIODES DIFFICILE.....	33
1.2.1 Overview: Epidemiology and Risk factors.....	33
1.3 SPORULATION.....	34
1.3.1 Entry into sporulation.....	35
1.3.2 Regulation of the Sporulation process.....	39
1.4 Spore structure and Function.....	41
1.4.1 Function of Small acid-soluble proteins (SASPs).....	41
1.4.2 Core.....	42
1.4.3 Inner membrane.....	42
1.4.4 Cortex.....	43
1.4.5 Outer membrane.....	44
1.4.6 Coat.....	44
1.4.7 The Exosporium in Spore-forming bacteria.....	45
1.4.8 Role of surface associated proteins.....	46
1.5 PATHOGENESIS OF <i>C. DIFFICILE</i> INFECTIONS.....	47
1.5.1 Bile production and metabolism.....	47
1.5.2 Dissemination of <i>C. difficile</i> infection.....	47
1.5.3 Control of <i>C. difficile</i> infection.....	48
1.6 Diagnosis.....	50
1.7 Treatment.....	50
1.8 GERMINATION.....	51
1.8.1 Role of subtilisin-like serine proteases (Csps) in <i>C. difficile</i> spore germination....	52
1.8.2 Binding of <i>C. difficile</i> CspC germinant receptor to bile salts.....	53
1.8.3 Cortex hydrolysis.....	54
1.9 Rationale.....	56
1.10 Objectives.....	57
Specific objectives:.....	57
CHAPTER 2. MATERIALS AND METHODS.....	58

2.1 Materials	58
2.1.1 Microorganisms and growth conditions	58
2.2 Bioinformatics analyses	59
2.3 Analysis of Transglutaminase transcripts during sporulation in <i>C. difficile</i> 630	59
2.3.1 Total RNA isolation	59
2.3.2 cDNA synthesis and qPCR amplification of transglutaminase genes.....	60
2.4 DNA METHODS	60
2.4.1 Genomic DNA extraction.....	60
2.4.2 Primer Design	61
2.4.3 PCR amplification.....	62
2.4.4 PCR gel purification	63
2.4.5 Agarose gel electrophoresis procedure	63
2.4.6 Quantification of DNA.....	63
2.4.7 Preparation of competent cells.....	64
2.4.8 Transformation of competent cells	64
2.4.9 Cloning of Transglutaminase-like genes from <i>C. difficile</i> 630	64
2.4.10 Sequencing of recombinant plasmids.....	67
2.5 Recombinant transglutaminase Protein Expression and Purification	68
2.5.1 Expression of recombinant proteins	68
2.5.2 Purification of recombinant proteins	68
2.5.3 Dialysis of purified proteins	69
2.6 PROTEIN METHODS	69
2.6.1 Protein concentration assay	69
2.6.2 Protein precipitation with trichloroacetic acid (TCA).....	70
2.6.3 SDS-PAGE gel preparation and protein analysis.....	70
2.6.4 Coomassie staining of acrylamide gels	71
2.6.5 Western blot.....	71
2.6.6 Transglutaminase activity assay.....	72
2.6.7 Protease activity assay	73
2.6.8 Transglutaminase-catalysed protein crosslinking assay	73
2.6.9 Cleavage and removal of GST tags from recombinant proteins.....	74
2.6.10 Digestion of recombinant proteins with Trypsin	74
2.7 Optimization of Recombinant Transglutaminase Protein Expression and Purification	75
2.7.1 Removal of signal peptides by PCR amplification of Recombinant plasmids	75
2.8 Purification and Characterisation of Transglutaminase-like proteins from <i>C. difficile</i> 630 spores	75
2.8.1 <i>C. difficile</i> growth conditions and Spore production	75
2.8.2 Transglutaminase production during sporulation in <i>C. difficile</i>	76
2.8.3 Solubilisation of transglutaminase-like proteins from <i>C. difficile</i> spores	77
2.8.4 Purification of crude enzyme extracts from <i>C. difficile</i> 630 spores	78
2.8.5 Characterization of partially purified <i>C. difficile</i> 630 spore enzymes	79
2.9 Statistical Analysis	80
CHAPTER 3. Gene Expression during sporulation in <i>C. difficile</i> 630 and Bioinformatics analysis of putative transglutaminase-like genes	81
3.1 Introduction	81
3.2 Objectives	81
3.3 RESULTS	82
3.3.1 Bioinformatics	82

3.3.2 <i>C. difficile</i> 630 spores contain transglutaminase activity	88
3.3.3 Putative transglutaminase genes are expressed during sporulation in <i>C. difficile</i> 630.....	88
3.4 DISCUSSION	94
3.4.1 <i>C. difficile</i> 630 spores contain transglutaminase activity which is produced during sporulation	94
3.4.2 Putative transglutaminases are expressed during sporulation in <i>C. difficile</i> 630 ..	94
3.4.3 Characteristics of microbial transglutaminase-like proteins from <i>C. difficile</i> 630 ..	96
3.5 Conclusion	98
CHAPTER 4. Recombinant Expression, Purification and Characterisation of transglutaminase-like proteins from <i>C. difficile</i> 630.....	100
4.1 Introduction	100
4.2 Objectives	100
4.3 RESULTS.....	101
4.3.1 <i>C. difficile</i> 630 genomic DNA extraction	101
4.3.2 PCR amplification of transglutaminase-like genes in <i>Clostridioides difficile</i> 630	102
4.3.3 Cloning Transglutaminase-like genes from <i>C. difficile</i> 630	103
4.3.4 Protein Expression and Purification of GST-tagged transglutaminase-like Proteins	105
4.3.5 Cloning and Recombinant Expression of <i>C. difficile</i> 630 putative transglutaminase proteins in pRSETc.....	115
4.4 Recombinant Expression And Purification Of Mature <i>C. difficile</i> 630 transglutaminase-Like Proteins	125
4.4.1 PCR amplification of mature regions in <i>C. difficile</i> 630 transglutaminase genes	126
4.4.2 Plasmid preparations of recombinant mature <i>C. difficile</i> 630 transglutaminase constructs	127
4.4.3 Removal of signal peptides promotes soluble expression of <i>C. difficile</i> 630 transglutaminases.....	128
4.5 Characterization of recombinant transglutaminase-like proteins from <i>C. difficile</i> 630	130
4.5.1 Characterization of recombinant GST-tagged CD-720 protein.....	130
4.5.2 Cleavage of GST-tags from recombinant mature proteins and column purification	141
4.5.3 Trypsin digestions of recombinant His-tagged proteins	143
4.6 DISCUSSION	144
4.6.1 Production of expression plasmids.....	144
4.6.2 Recombinant <i>C. difficile</i> 630 transglutaminase protein expression in <i>E. coli</i>	145
4.6.3 Effect of temperature on recombinant transglutaminase protein expression	145
4.6.4 Characterization of recombinant proteins	151
4.7 Conclusion	152
CHAPTER 5. Purification and Characterization of transglutaminase from <i>C. difficile</i> 630 spores	154
5.1 Introduction	154
5.2 Transglutaminase activity during <i>C. difficile</i> 630 sporulation on solid and liquid media.....	155
5.3 Solubilisation of transglutaminase from sporulating cells using different temperature and pH conditions.....	157
5.4 High temperatures inactivate crude enzyme activity	161

5.5 Purification of transglutaminase activity from crude extracts of <i>C. difficile</i> 630	161
5.5.1 Purification of transglutaminase by ion-exchange chromatography	162
5.5.2 Purification of active fractions by hydrophobic interaction chromatography	169
5.6 SDS-PAGE analysis of purified transglutaminase from crude extracts of <i>C. difficile</i> 630	171
5.7 Characterization of purified enzymes from <i>C. difficile</i> 630 spores	172
5.7.1 Optimal temperature of purified spore protein activity	173
5.7.2 Thermostability of purified spore enzymes	173
5.7.3 Effect of various inhibitors and metal ions on purified enzyme activity	174
5.7.4 Purified transglutaminase from <i>C. difficile</i> spores does not crosslink casein substrate	176
5.7.5 Mass spectrometry analysis of purified transglutaminase from <i>C. difficile</i> 630 spores	177
5.8 DISCUSSIONS	179
5.8.1 Purification of transglutaminase from sporulating <i>C. difficile</i> 630 cells	182
5.8.2 Characterization of transglutaminase purified from sporulating <i>C. difficile</i> 630	183
5.8.3 Mass spectrometry analysis of purified spore proteins	186
5.9 CONCLUSION	188
CHAPTER 6. GENERAL DISCUSSIONS	190
6.1 Introduction	190
6.2 Gene expression analysis during sporulation in <i>C. difficile</i> 630	191
6.3 Characterization of transglutaminase activity produced during sporulation in <i>C. difficile</i>	192
6.4 Transglutaminase gene expression and mass spectrometry analysis of purified proteins	193
6.5 Recombinant Expression of transglutaminases in <i>C. difficile</i> 630	194
6.6 CONCLUSION AND FUTURE PERSPECTIVES	195
REFERENCES	196
APPENDICES	205
Appendix 1: Bacterial strains and plasmids used in this study	205
Appendix 2: Plasmids used in this study	205
Appendix 3: List of antibodies used in this study	207
Appendix 4: Amplification and melting curves obtained from qRT-PCR reactions of target genes	207
Appendix 5: Detection of transglutaminase activity in vegetative cells during sporulation in <i>C. difficile</i> 630	208
Appendix 6: Effect of washes on transglutaminase activity of lysed cells	208
Appendix 7: Screening of HIC media by purification of active Tgase fractions on selected columns	209
Appendix 8: SDS-PAGE showing reaction of purified protein with BSA substrate	211
Appendix 9: Transglutaminase activity in culture supernatants of recombinant proteins	211

LIST OF ABBREVIATIONS

AAD: Antibiotic associated diarrhoea
AGE: agarose gel electrophoresis
ANOVA: Analysis of variance
BHI: Brain heart infusion
BLAST: basic local alignment search tool
Bp: base pair
BSA: bovine serum albumin
BTC: Biotin cadaverine
Ca-DPA: Calcium dipicolinic acid
CBB: Coomassie brilliant blue
CDI: *Clostridioides difficile* infection
cDNA: complementary DNA
CFU: colony forming unit
CNF1: cytotoxic necrotizing factor
Csp: serine protease
CV: column volume
DBD: DNA binding domain
DMSO: dimethyl sulfoxide
DNA: deoxyribonucleic acid
DNT: dermonecrotizing toxin
dNTP: deoxyribonucleotide triphosphate
DPA: dipicolinic acid
DTT: dithiothreitol
ECM: Extracellular matrix
EDTA: Ethylenediaminetetraacetic acid
ELISA: Enzyme linked immunosorbent assay
FF: fast flow
FnbA: Staphylococcal fibronectin binding protein
FR: flow rate
FS: forespore
FTC: flouroscein casein
FXIII: factor XIII
GCW: germ cell wall
GDH: Glutamate dehydrogenase
gDNA: genomic DNA

GDP: Guanosine diphosphate
Gln: glutamine
GR: germination receptor
GST: Glutathione S-transferase
GTP: Guanosine triphosphate
HCL: hydrochloric acid
HIC: hydrophobic interaction chromatography
HK: histidine kinase
HP: high performance
HRP: Horseradish peroxidase
IAA: iodoacetamide
IB: inclusion bodies
IEX: ion-exchange
IgG: immunoglobulin G
IM: inner membrane
IPTG: Isopropyl β -D-1-thiogalactopyranoside
kb: kilobase
kDa: kilodalton
LB: Luria-Bertani
LPS: Lipopolysaccharide
Lys: lysine
MAL: muramic d-lactam
MC mother cell
MLSA: Multiple sequence alignment
mRNA: messenger RNA
MS: mass spectrometry
MTG: microbial transglutaminase
NAM: N-acetylmuramic acid
NC: nitrocellulose
NCBI: Nation center for biotechnology information
NEM: N-ethylmaleimide
NFW: nuclease free water
NO: nitric oxide
O/N: overnight
OD: optical density
OM: outer membrane
OPD: O-phenylenediamine

PBS: phosphate buffered saline
PCR: Polymerase chain reaction
PDB: protein data base
PG: peptidoglycan
pI: isoelectric point
PLC: phospholipase C
PMSF: phenylmethylsulfonyl fluoride
PMT: *Pasteurella multocida* toxin
PNK: polynucleotide kinase
qPCR: quantitative polymerase chain reaction
qRT-PCR: real-time quantitative reverse transcription PCR
rFnbA: recombinant fibronectin-binding protein A
RFU: relative fluorescence units
RNA: ribonucleic acid
ROS: reactive oxygen species
Rpm: revolutions per minute
RT-PCR: reverse transcription PCR
SASP: Small acid spore protein
SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SDW: sterile distilled water
SEC: size exclusion chromatography
SEM: standard error of the mean
Sig: sigma factor
SipL: SpoIVA interacting protein L
SP: signal peptide
SpoOA: stage 0 sporulation protein A
TAE: Tris-acetate EDTA
TBS: Tris buffered saline
TBST: Tris buffered saline Tween
TCA: trichloroacetic acid
TCS: two-component signal transduction system
TE: Tris-EDTA
TEMED: Tetramethylethylenediamine
TG2: Tissue transglutaminase
Tgase: Transglutaminase
TgpA: Transglutaminase protein A
tRNA: transfer RNA

TY: tryptone yeast media

U.K.: United Kingdom

UV: ultraviolet

X-Gal: 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

Z-Gln-Gly: Z-glutamine-glycine

LIST OF TABLES

Table 1.1. Summary of transglutaminase functions in bacteria.....	23
Table 1.2. Summary of Spores and vegetative cell responses to physical and chemical stressors.	50
Table 2.1. Primers used in this study.	62
Table 2.2. Recipe for SDS separating and stacking gels.....	71
Table 3.1. Summary of the predicted characteristics of transglutaminase-like genes in <i>C. difficile</i> 630.....	87
Table 5.1. Summary of transglutaminase purification from crude spore extracts of <i>C. difficile</i> 630.....	171
Table 5.2. Representative highly abundant proteins identified by mass spectrometry analysis.	177

LIST OF FIGURES

Figure 1.1. General reactions mediated by transglutaminases.....	17
Figure 1.2. Structure of microbial transglutaminase.....	21
Figure 1.3. Spore morphogenesis during sporulation in <i>C. difficile</i>	35
Figure 1.4. Structure of a bacterial spore.....	41
Figure 1.5. Flow chart of spore germination in <i>C. difficile</i>	56
Figure 2.1. Flow diagram of cloning strategy for transglutaminase-like genes in <i>C. difficile</i>	65
Figure 2.2. Flow diagram for Transglutaminase production.....	77
Figure 3.1. Multiple alignment of Tgase-core domains in <i>C. difficile</i> 630.....	83
Figure 3.2. Multiple sequence alignment of conserved Tgase regions in mammalian and microbial organisms.....	84
Figure 3.3. Phylogenetic tree analysis of Tgase-like genes in mammalian and microbial organisms.....	85
Figure 3.4. Prediction of Signal peptides in CD-1062 and CD-1083 Tgase genes.....	86
Figure 3.5. Transglutaminase activity in sporulating <i>C. difficile</i> 630 cells.....	88
Figure 3.6. Representative amplification and melting curves obtained from qRT-PCR reactions.....	90
Figure 3.7. qRT-PCR analysis of Tgase genes and spo0A during sporulation in <i>C. difficile</i> 630.....	92
Figure 3.8. Growth curve of <i>C. difficile</i> 630.....	93
Figure 3.9. Possible spo0A targets upstream transglutaminase-like DNA sequences in <i>C. difficile</i> 630.....	95
Figure 4.1. Genomic DNA extraction from <i>C. difficile</i> 630.....	101
Figure 4.2. PCR amplifications of <i>C. difficile</i> 630 transglutaminase genes.....	102
Figure 4.3. Restriction digestion of recombinant pJET1.2/blunt and pUC19 plasmids.....	103
Figure 4.4. Restriction digestion of recombinant plasmids after sub-cloning in pGEX-3X.....	104
Figure 4.5. Affinity purification scheme of recombinant GST-CD720 proteins from <i>C. difficile</i> 630.....	106
Figure 4.6. Affinity purification scheme of recombinant GST-CD1062 and GST-CD1083 proteins from <i>C. difficile</i> 630.....	107
Figure 4.7. SDS-PAGE of recombinant GST-tagged protein expression at 37°C.....	108
Figure 4.8. SDS-PAGE of recombinant GST-tagged protein expression at 20°C.....	110
Figure 4.9. SDS-PAGE of recombinant GST-tagged proteins expressed at low temperature and IPTG concentration.....	112
Figure 4.10. SDS-PAGE showing recombinant GST-tagged proteins expressed in <i>E. coli</i> Rosetta 2 cells.....	114
Figure 4.11. Restriction digestion of recombinant plasmids after sub-cloning in pRSETc.....	116
Figure 4.12. SDS-PAGE of recombinant His-tagged proteins expressed at 37°C.....	118
Figure 4.13. SDS-PAGE of recombinant His-tagged proteins expressed in <i>E. coli</i> BL21 at 20°C.....	120
Figure 4.14. SDS-PAGE of recombinant His-tagged proteins expressed in <i>E. coli</i> Rosetta 2 at 37°C.....	122
Figure 4.15. SDS-PAGE of recombinant His-tagged proteins expressed in <i>E. coli</i> Rosetta 2 at 20°C.....	123
Figure 4.16. Western blot of recombinant His-tagged proteins expressed in <i>E. coli</i> BL21 and Rosetta 2 cells.....	124
Figure 4.17. Diagram of recombinant pGEX3X plasmid with primers for signal peptide cleavage.....	125
Figure 4.18. PCR amplifications of mature regions in <i>C. difficile</i> 630 transglutaminase genes.....	126
Figure 4.19. Plasmid preparations of recombinant mature <i>C. difficile</i> 630 transglutaminase constructs.....	127
Figure 4.20. SDS-PAGE analysis of recombinant GST-tagged CD-1062m proteins.....	128

Figure 4.21. SDS-PAGE analysis of recombinant GST-tagged CD-1083m proteins..	129
Figure 4.22. SDS-PAGE analysis of GST tag cleavage from recombinant CD-720 proteins..	131
Figure 4.23. Transglutaminase activity assay of recombinant CD-720 proteins.....	132
Figure 4.24. Protease activity assay of recombinant GST-CD720 proteins..	133
Figure 4.25. Protease activity assay of recombinant CD-720 protein with cleaved GST-tags.	134
Figure 4.26. Hide powder azure protease assay for recombinant GST-CD720..	135
Figure 4.27. Hide powder azure protease assay for recombinant GST-CD720..	136
Figure 4.28. SDS-PAGE analysis following trypsin digestions of recombinant GST-CD720 proteins.....	137
Figure 4.29. Transglutaminase activity assay of recombinant GST-CD720 digested with trypsin.....	138
Figure 4.30. SDS-PAGE analysis of recombinant GST-CD720 protein reaction with BSA.	139
Figure 4.31. SDS-PAGE analysis of recombinant GST-CD720 protein reaction with casein.	140
Figure 4.32. Transglutaminase activity in induced lysates and recombinant eluted protein fractions.....	141
Figure 4.33. SDS-PAGE showing tag cleavage from recombinant mature GST-tagged proteins.....	142
Figure 4.34. SDS-PAGE showing trypsin digestions of recombinant His-tagged proteins. P.	143
Figure 5.1. Detection of transglutaminase activity during sporulation in <i>C. difficile</i> ...	156
Figure 5.2. Solubilisation of transglutaminase from spores at different temperatures.	158
Figure 5.3. Change in transglutaminase activity over time at different pH.	159
Figure 5.4. Effect of detergents and reducing agents on transglutaminase solubilisation from spores.....	160
Figure 5.5. Effect of heat on transglutaminase activity of crude extracts from <i>C. difficile</i> 630.	161
Figure 5.6. Purification of crude enzyme extracts from sporulating <i>C. difficile</i> on 20 ml Q-Sepharose FF.....	163
Figure 5.7. Purification of crude enzyme extracts from sporulating <i>C. difficile</i> on 100 ml Q-Sepharose FF.....	164
Figure 5.8. Effect of various salts on transglutaminase activity from <i>C. difficile</i> spores..	166
Figure 5.9. Effect of NaCl on pooled transglutaminase activity and recovery of activity after desalting.....	167
Figure 5.10. Screening of butyl Sepharose columns for purification of active transglutaminase fractions.....	168
Figure 5.11. Purification of active transglutaminase fractions on Butyl Sepharose FF..	170
Figure 5.12. SDS-PAGE analyses of purified active enzyme fractions from Butyl Sepharose columns..	172
Figure 5.13. Effect of temperature on activity of purified transglutaminase from <i>C. difficile</i> 630 spores.....	173
Figure 5.14. Thermostability of purified transglutaminase activity from <i>C. difficile</i> 630 spores..	174
Figure 5.15. Effect of ions and inhibitors on purified transglutaminase activity..	175
Figure 5.16. SDS-PAGE showing reaction of purified protein with casein substrate..	176

CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW

1.1 Transglutaminases: Overview

Transglutaminases are enzymes that mediate three kinds of posttranslational modifications: transamidation, deamidation and crosslinking (Figure 1.1) (Nurminskaya and Belkin, 2012). Transglutaminases (Tgases) form proteolysis-resistant isopeptide bonds through acyl transfer reactions between the γ -carboxamide of a peptide-bound glutamine acting as amine acceptor and the ϵ -amino group of lysine acting as amine donor (Bergamini *et al.*, 2011). Transglutaminases irreversibly crosslink proteins forming insoluble and high molecular weight aggregates. The reaction proceeds by a ping-pong mechanism (Figure 1.1) whereby the active site thiol (SH) in cysteine attacks a glutamine moiety in the side chain of the target protein, forming an acyl enzyme with the release of ammonia (Pasternack *et al.*, 1998). In the next step, the enzyme is displaced by a primary amine forming a γ -carboxamide and in the presence of lysine in a protein, an isopeptide bond is formed resulting from inter/intra molecular crosslinking (Griffin, Casadio and Bergamini, 2002). A ping pong reaction usually involves two substrates which react individually with the enzyme, and formation of an enzyme intermediate complex plus a product upon reaction with the first substrate. The final product is released after reaction with the second substrate and the enzyme returns to the original form (Ulusu, 2015). Isopeptide bonds provide stiffness and resistance to physical, mechanical, chemical and enzymatic breakdown in host cells and tissues (Griffin, Casadio and Bergamini, 2002). Transglutaminases catalyse deamidation reactions in the absence of an acceptor lysine, where water becomes the acyl acceptor producing glutamic acid (Kieliszek and Misiewicz, 2014). Figure 1.1 shows the transglutaminase reaction mechanism.

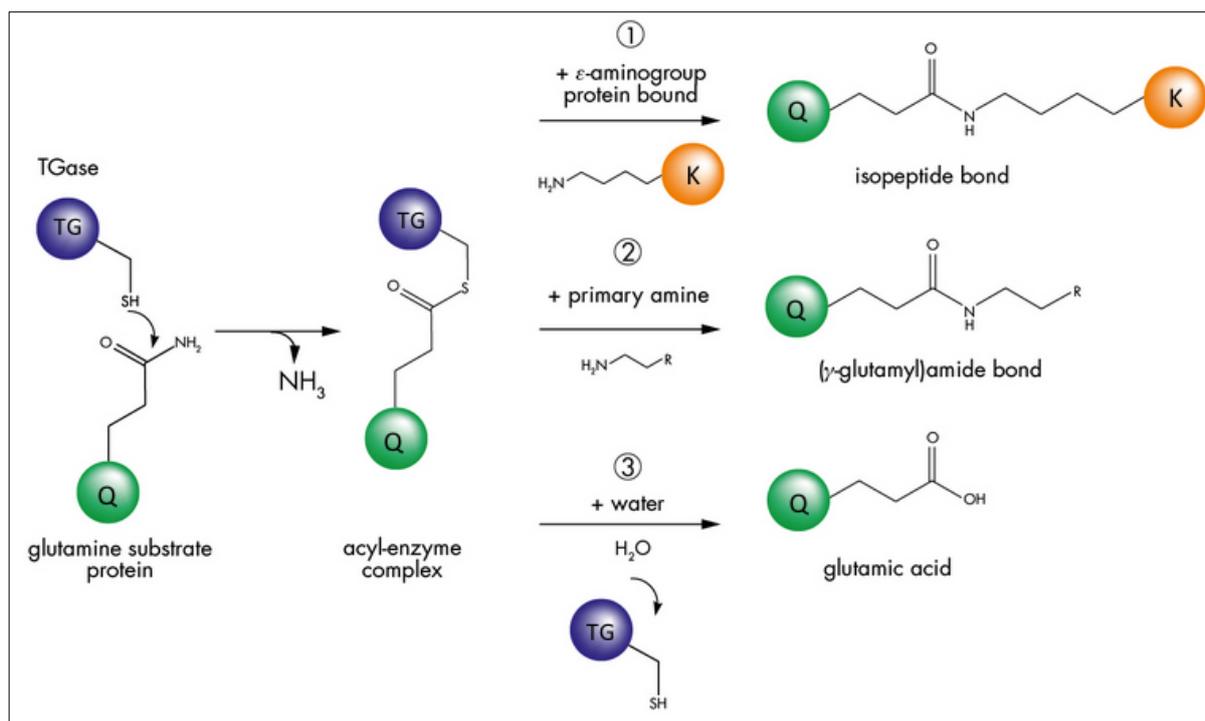


Figure 1.1. General reactions mediated by transglutaminases. The active site thiol group in transglutaminase attacks the γ -carboxamide of glutamine in a protein substrate to form an acyl-enzyme intermediate with the release of ammonia. 1) Crosslinking/Transamidation. The acyl-enzyme reacts with the ϵ -amino group of lysine (acyl acceptor) to form an ϵ - γ -(glutamyl) lysine isopeptide bond. 2) Acylation. Acyl-enzyme intermediates undergo an acyl-transfer to a primary amine forming a (γ -glutamyl)amine bond. 3) Deamidation. In the absence of an amine acceptor, glutamine undergoes a hydrolysis reaction with water serving as acyl acceptor to form glutamic acid. Q and K represent glutamine and lysine amino acids, and TG represents transglutaminase.

Transglutaminases have been reported in plants, animals and microorganisms (Kieliszek and Misiewicz, 2014). The active site of transglutaminases in all domains of life consists of a catalytic triad (Cys His Asp) which differ in structural orientation, indicating an evolutionary relationship. Indeed all characterised transglutaminases contain a conserved cysteine residue at the catalytic site which is indispensable for activity (Javitt *et al.*, 2017). Using computer BLAST analysis programs, protein homologs of eukaryotic transglutaminases were identified in archaea, bacteria, yeast and nematodes. Members of this newly identified protein superfamily possess a conserved catalytic triad of cysteine, histidine, aspartate similar to characterised transglutaminases (Fernandes *et al.*, 2015). Former comparison studies have suggested an evolutionary link between transglutaminases and papain-like thiol proteases, consistent with the localised sequence similarities observed in the proximity of the catalytic sites (Makarova, Aravind and Koonin, 1999). Papain-like thiol proteases have a similar core to animal transglutaminases and belong to the same superfamily as observed from BLAST analysis

(Makarova, Aravind and Koonin, 1999). Similarly, microbial transglutaminases from *Streptomyces mobaraensis* and animal transglutaminases have evolved from the same ancestor (Rickert *et al.*, 2016). Following BLAST analyses, many microbial homologs of eukaryotic transglutaminases were identified as proteases with a conserved catalytic triad of cysteine, histidine, aspartate analogous to that of eukaryotic transglutaminases, though with different structural configuration. This elucidates a convergent evolutionary link implying that animal and microbial transglutaminases originate from ancient proteases (Fernandes *et al.*, 2015).

Transglutaminase-like proteins can be separated into three categories based on their domain constitution. The first category includes secreted proteins as specified by the predicted signal peptide regions. Secreted microbial transglutaminases from pathogenic bacteria such as *Mycobacteria* participate in breakdown of host proteins during their life cycle i.e., exhibit some form of proteolytic activity (Makarova, Aravind and Koonin, 1999). The second category describes proteins that contain only the transglutaminase-like domains and these proteins are usually intracellular enzymes (Suzuki *et al.*, 2000). The third category describes intracellular proteins that contain the transglutaminase-core domain and several other protein-protein interaction regions and therefore would probably operate as components of protein complexes (Makarova, Aravind and Koonin, 1999). Proteins that carry inactive transglutaminase domains probably participate in structural or protein-binding roles like the erythrocyte Band 4.2 proteins (described in 1.1.1) but can also act as negative modulators of active enzymes (Griffin, Casadio and Bergamini, 2002).

1.1.1 Mammalian Transglutaminases

Mammalian transglutaminases belong to the superfamily of cysteine proteases with a conserved catalytic triad. Mammalian transglutaminases are calcium-dependent enzymes that irreversibly crosslink proteins by forming covalent epsilon (gamma-glutamyl) lysine bonds that are highly resistant to breakdown by matrix proteases (Bergamini *et al.*, 2011). Transglutaminase 2 (TG2) is the most abundant and multifunctional member of the animal transglutaminase family which comprises eight functional transglutaminases which perform different roles (Wang and Griffin, 2012). TG1, TG3, and TG5 are mostly expressed on squamous epithelial surfaces (Strop, 2014). TG4 is found in the prostate gland; TG6 is found in the lung, brain and testis; TG7 just like TG2 is ubiquitous but predominantly found in the lung and testis. Band 4.2 is an inactive protein located on erythrocyte membranes and part of the TG2 family (Odi and Coussons, 2014). Tissue transglutaminase (tTG or TG2) is found in

intracellular and extracellular cell compartments particularly in the cytosol, mitochondria, nucleus, plasma membrane, extracellular matrix (ECM) and cell surface (Nurminskaya and Belkin, 2012). Likewise, Factor XIII is located in intracellular and extracellular cell compartments but mostly found in blood plasma (Matsuka *et al.*, 2003).

1.1.1.1 Physiological roles of mammalian transglutaminases

Transglutaminases are found in several mammalian organisms and participate in several physiological and pathological processes (Pasternack *et al.*, 1998). Transglutaminases contribute to many physiological functions including wound healing, blood clotting, skin protection, cell adhesion, migration, cell development, differentiation and apoptosis (Rachel and Pelletier, 2013). In the cytosol, TG2 acts as a G-protein whereby it activates phospholipase C (PLC) during hormonal signal transduction, and this has an effect on cell survival and cell cycle advancement (Wang and Griffin, 2012). Transglutaminase is responsible for extracellular matrix (ECM) deposition and stabilisation via protein crosslinking. Additionally, transglutaminases play a protective role in tissue repair, and provide mechanical and chemical strength to tissues preventing injury (Wang and Griffin, 2012). Transglutaminase 1 is a membrane bound protein that crosslinks surface proteins that line the stratified squamous epithelial cells, thus providing resistance to chemical and mechanical injury or in certain cases dehydration (Iismaa *et al.*, 2009). Under normal physiological conditions transglutaminases are inactive and present at low levels in tissues. Their activation is triggered by perturbations in normal homeostasis (Kim, Jeitner and Steinert, 2002). Transglutaminase is involved in protein crosslinking in blood plasma, serving as a healing mechanism after injury. Factor XIIIa takes part in stabilisation of blood clots following injury via establishment of isopeptide crosslinks amongst fibrin monomers (Matsuka *et al.*, 2003). The established clot is very stiff and highly resistant to enzymatic lysis (Iismaa *et al.*, 2009).

1.1.2 Microorganisms exploiting host transglutaminases

Some bacteria such as pathogenic *Staphylococcus aureus* employ transglutaminase activity to adhere to cell surfaces, proteins and host tissues. In order to colonise its hosts and cause infection, the surface-associated binding proteins (FnbA) expressed in *S. aureus* bind to extracellular matrix (ECM) proteins, fibronectin and fibrinogen which are Factor XIIIa substrates. This suggests that surface-associated Staphylococcal binding protein acts as a bifunctional transglutaminase substrate (Matsuka *et al.*, 2003). Therefore *S. aureus* exploits Factor XIIIa crosslinking activity to bind to its ligands present on host cell surfaces, plasma proteins and ECM. Factor XIIIa mediates covalent crosslinking of Staphylococcal binding

protein to ECM proteins, which plays a key function upon bacterial colonization at the injury site. Staphylococcal binding protein could inhibit the wound healing process upon incorporation at the injury site by obstructing fibrin polymerization. Following colonisation at the injury site, FnbA could promote the spread of infection (Matsuka *et al.*, 2003).

Likewise, some yeast express surface proteins that act as transglutaminase substrates and through crosslinking enable the yeast to invade and infect hosts. *Candida albicans* (*C. albicans*) is an opportunistic pathogen that causes severe disease in susceptible and immunocompromised hosts (Staab *et al.*, 1999). *C. albicans* expresses a hyphae-specific surface protein, Hwp1 that plays a role in the pathogenesis of candidiasis. The Hwp1 protein which contains amino and carboxyl terminal domains is a transglutaminase substrate that interacts with human buccal epithelial cells forming stable complexes typical of transglutaminase-catalysed reactions (Staab *et al.*, 1999). Therefore, in keratinised epithelium the growth of *C. albicans* is possibly mediated by the reaction of Hwp1 proteins with keratinocyte transglutaminases and is impeded by transglutaminase inhibitors. This reaction enables *C. albicans* to invade its hosts and cause disease (Staab *et al.*, 1999).

1.1.3 Microbial Transglutaminases

1.1.3.1 Properties of microbial transglutaminases

Microbial transglutaminases (MTG) have a single subunit and are calcium-independent, in contrast to the multi domain animal transglutaminases which require calcium for activity. Microbial transglutaminases have lower molecular weights e.g. *Bacillus subtilis* Tgl: 29 kDa, *Streptomyces mobaraense* MTG: 38 kDa compared to animal transglutaminases like TG2: 80 kDa (Strop, 2014), with optimum catalytic activity at 40°C, pH 5.5 and are stable within a broad pH (4.5-8) and temperature range (Ho *et al.*, 2000). MTG from *Streptomyces* is unstable at 50°C, with high activity at lower temperatures (45°C) and retains full activity at 0°C. The thermostability of MTG is enhanced in the presence of carbohydrates (Yokoyama, Nio and Kikuchi, 2004; Cui *et al.*, 2007). *B. subtilis* transglutaminase (Tgl) has greater stability than MTG and a wider temperature (30-60°C) and pH range (5-9) of activity, with peak activity at pH 8, 60°C. MTG and Tgl have 10% similarity in the amino acid sequences at the catalytic centre (Liu *et al.*, 2018). Microbial transglutaminase activity is enhanced by Co²⁺, K⁺, Ba⁺ ions and impeded by ions which react with thiol groups of cysteine (Pb²⁺, Zn²⁺, Hg²⁺, Cu²⁺) at the active site (Kieliszek and Misiewicz, 2014). Figure 1.2 illustrates the crystal structure of single domain *Streptomyces* MTG, with the catalytic cysteine residue at the base of the active site cleft.

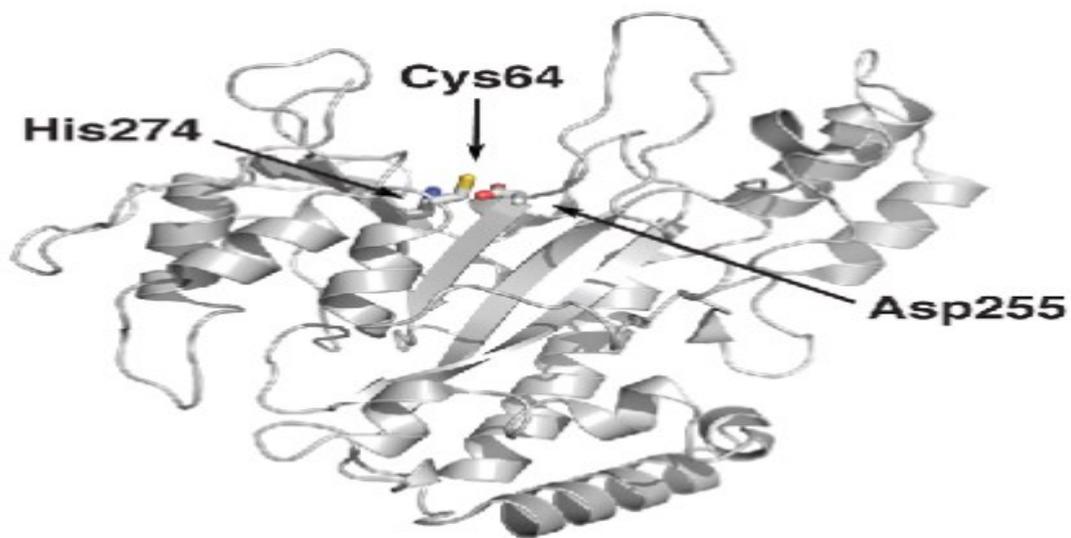


Figure 1.2. Structure of microbial transglutaminase. The enzyme as shown is proteolytically activated and the yellow balls illustrate the cysteine active site. Adapted from (Rachel and Pelletier, 2013).

The single domain structure of MTG differs from the multidomain structure of mammalian transglutaminases highlighting structural differences between mammalian and microbial transglutaminases (Fernandes *et al.*, 2015). The N-terminus in MTG prevents substrate accessibility by shielding the active site, thus proteases are required for activation. Transglutaminase from *S. mobaraense* (MTG) was produced as a zymogen and activated by endogenous and exogenous proteases (Pasternack *et al.*, 1998). MTG and mammalian transglutaminase have the same catalytic triad in the active site that differ in structural orientation (Kashiwagi *et al.*, 2002). From a biotechnology perspective microbial transglutaminases offer several advantages over the mammalian counterparts due to their high thermostability, lesser substrate specificity, low molecular weights, greater stability over a wide pH range (4.5-8), and independence from cofactors for activity (Kieliszek and Misiewicz, 2014). Additionally, MTG is an economical source of transglutaminase with broad substrate specificity and a single domain protein that displays a unique structure (Yokoyama, Nio and Kikuchi, 2004).

1.1.3.2 Functions of microbial transglutaminases

Transglutaminase is implicated in various roles in bacteria which contribute to its protection. *Streptomyces* transglutaminase is involved in the crosslinking of protease inhibitory proteins originating from *Streptomyces* during spore and hyphae development (Strop, 2014). The protein crosslinks could contribute to the establishment of an antibiotic barrier and protection

from host proteases in *Streptomyces*. Additionally *Streptomyces* MTG participates in spore coat morphogenesis, cell division, elongation, germination and cell growth (Kieliszek and Misiewicz, 2014). It is predicted that MTG in *S. mobaraensis* confers physical and chemical toughness to cell structures like the spore coat by modifying cellular proteins through crosslinking (Strop, 2014). Similarly, transglutaminases are secreted during sporulation and crosslink spore coat proteins in *B. subtilis*, which may provide spore resistance against mechanical and proteolytic degradation (Katsunori Kobayashi *et al.*, 1998; Monroe and Setlow, 2006).

Like their mammalian counterparts, microbial transglutaminases are also implicated in the pathogenesis of certain infections and diseases. They are involved in surface modification of proteins in bacterial and yeast biofilms (Kuwana *et al.*, 2006). More recently like TG2, *Streptomyces* MTG has been implicated in the pathogenesis of celiac disease by crosslinking and possibly deamidating gluten and gliadin. Additionally, MTG has been reported to be important in gut microbial survival by exerting protease inhibitory activity and inhibiting phagocytosis and anti-microbial peptides (Aaron and Torsten, 2018). Transglutaminases have been identified in non-spore forming pathogenic bacteria and play several roles thereby promoting their pathogenicity (Matsuka *et al.*, 2003; Milani *et al.*, 2012). A summary of Tgase functions in microbial organisms is described in Table 1.1.

Table 1.1. Summary of transglutaminase functions in bacteria (Katsunori Kobayashi *et al.*, 1998; King *et al.*, 2009; Milani *et al.*, 2012; Kieliszek and Misiewicz, 2014).

organism	Gene or protein	Cellular location	Functions
<i>Bacillus subtilis</i>	Tgl gene	Spore coat	Provides resistance against toxic chemicals and lytic enzymes through crosslinking of spore surface structures
<i>Streptomyces mobaraensis</i>	MTG	Spore coat	Confers resistance to cellular structures and protection from host proteases
<i>Pseudomonas aeruginosa</i>	TgpA protein	Periplasm of PA2873 protein	Essential for viability
<i>Bordetella bronchiseptica</i>	WbmE	periplasm	lipopolysaccharide development

1.1.4 Evidence for the role of transglutaminases in sporulation

In spores and sporulating cells of *B. subtilis*, a Tgase activity from the Tgl gene has been reported in several studies (Katsunori Kobayashi *et al.*, 1998; Katsunori. Kobayashi *et al.*, 1998; Suzuki *et al.*, 2000). Also, ϵ - γ -glutamyl lysine cross-links were found in the spores of *B. subtilis* (Kobayashi *et al.*, 1996). Transglutaminase activity was linked to sporulation in *B. subtilis* since several strains simultaneously demonstrated high activity with high sporulation rates, meanwhile the strains which were unable to sporulate showed no activity (Katsunori Kobayashi *et al.*, 1998). Ensuing, it was proposed that Tgases are bound to spores following their production during sporulation. Many post-transcriptional and post-translational modifications including crosslinking, proteolysis, tyrosine peroxidation and glycosylation regulate the coat assembly process (Plomp *et al.*, 2014). Previous studies revealed the existence of reversible and irreversible crosslinking of coat proteins. A highly insoluble crosslinked fraction was refractory to extraction by a combination of detergents, reducing agents and denaturants, which constituted about 30% of the total spore coat protein (Ragkousi and Setlow, 2004). Accordingly transglutaminase was postulated to be involved in coat protein assembly because it was found attached to spores at the level of coat proteins, and crosslinking could be important in this assembly (Zilhão *et al.*, 2005). Spore coat proteins including Tgases are produced in the mother cell of the developing spore (Monroe and Setlow, 2006).

Transglutaminase activity forms part of the covalent crosslinks found in the spore coats of *B. subtilis* (Monroe and Setlow, 2006). In *Bacillus subtilis*, a spore coat protein (GerQ) was crosslinked by Tgl forming high molecular weight complexes in the late stages of sporulation after lysis of the mother cell. Thus Tgl contributes to GerQ association with the insoluble coat fraction (Ragkousi and Setlow, 2004). Following purification and characterization experiments, its molecular weight was detected to be 23 kDa and cysteine was shown to be important for its crosslinking activity (Katsunori Kobayashi *et al.*, 1998). The GerQ protein contains several lysine residues in its amino terminus and thus serves as a good substrate for Tgl catalysed crosslinking (Monroe and Setlow, 2006). Additionally Tgl is thought to play a role in the insolubilisation of small molecular weight polypeptides in the spore coat (Zilhão *et al.*, 2005). There are no homologs of Tgl in *C. difficile*, and no work has been done so far. It is hypothesized that Tgl-catalysed crosslinking of the coat protein GerQ is more efficient at 60°C, with less activity at lower temperatures (37°C), during which GerQ crosslinking is mediated by the YabG coat protein (Kuwana *et al.*, 2006). Perhaps prior modification of GerQ by YabG is required for Tgase-catalysed crosslinking of GerQ at higher temperatures. Both YabG and Tgl play a role in spore germination by modifying GerQ through crosslinking (Kuwana *et al.*, 2006). Even so, Tgl-mediated crosslinking of GerQ is not crucial for normal functioning of the spore coat and germination. However, the crosslink could play an important role in providing extra stiffness and resistance against degradation (Ragkousi and Setlow, 2004).

During sporulation in *C. difficile* 630, SigK dependent genes were the most greatly expressed. A cysteine rich protein CD1067 was identified as the most abundant protein in purified spores and is also regulated by SigK (Fimlaid *et al.*, 2013). This protein under denaturing conditions forms highly resistant high molecular weight multimers and therefore could contribute to a more rigid spore coat structure. Interestingly, CD1067 protein is coded in a location enriched with majorly SigK regulated spore-protein coding genes and Sig K codes for hypothetical proteins specific to *C. difficile* (Lawley *et al.*, 2009). These genes could code for coat proteins that provide structural integrity and toughness to the spore coat. Therefore, SigK possibly regulates gene expression of Tgase-like genes during the late stages of sporulation. The transcription of a *B. subtilis* Tgase, Tgl is mediated by SigK, and GerE acts as transcriptional regulator of Tgl expression (Zilhão *et al.*, 2005). GerE is a transcription factor produced from a SigK-regulated gene which controls expression of SigK, and thus Tgl. The SigK consensus sequence was found in the promoter region of Tgl, including the putative GerE target sites. Tgl crosslinking of GerQ into high molecular weight products occurs in the late stages of sporulation following lysis of the mother (Katsunori. Kobayashi *et al.*, 1998).

1.1.5 Microbial transglutaminases and their role in non-spore forming bacteria

1.1.5.1 Deamidation as a bacterial virulence mechanism

Pathogenic bacteria have long employed protein toxins as virulence factors to cause disease in their hosts via modification of protein targets and physiological processes. The Dermonecrotic toxin (DNT) family members which include many bacterial toxins utilise deamidation and transamidation to modify target G-proteins in host cells (Schmidt *et al.*, 1998). Dermonecrotic toxin catalyses Rho transglutamination which results to its activation (Washington *et al.*, 2013). Just like transglutaminases, deamidases possess a cysteine-based active site and belong to the same papain-like superfamily. The deamidation of host proteins is irreversible, contributing to their potency and virulence (Washington *et al.*, 2013). More so, the specific residues targeted for modification are important in the functionality of the host proteins. The specificity of the deamidation of host target proteins is high and the downstream effects are detrimental to the host. The deamidation and transamidation of catalytic glutamine triggers GTPase activation. In the pathogenic strains of *Escherichia coli* (*E. coli*), CNF1 (cytotoxic necrotizing factor) has been shown to deamidate specific RhoA GTPase family members to glutamic acid (Schmidt *et al.*, 1998). This causes constitutive activation of host protein GTPases leading to rearrangement of the actin cytoskeleton, thereby promoting entry of pathogenic *E. coli* into uroepithelial cells that trigger urinary tract infections. Also pathogenic bacteria like *Salmonella* modulate the host cytoskeleton to ease entry into their host (Washington *et al.*, 2013). Analogous to CNF1, in *Pasteurella multocida* the PMT toxin (*P. multocida* toxin) deamidates heterotrimeric G-proteins leading to their activation that triggers a sequence of events thereby contributing to its virulence. The action of PMT on the host target is non-specific and indirect via α -subunits of G proteins, but still alters physiological processes in the host (Wilson and Ho, 2011). Similar to *E. coli*, this results to rearrangement of the actin cytoskeleton, development of actin stress fibres and endothelial cell permeability (Washington *et al.*, 2013).

In *Bordetella bronchiseptica*, a deamidase has been identified that targets the O-antigen in the complete lipopolysaccharide (LPS) domain. *Bordetella* is a Gram-negative bacterium associated with respiratory infections in mammals. The WbME protein in *Bordetella* is part of the transglutaminase superfamily with a catalytic triad reminiscent of transglutaminases (King *et al.*, 2009). The WbME protein is located in the periplasm and targets uronamide-rich O chains for deamidation to uronic acid and this event occurs late during LPS biosynthesis. This demonstrates how *Bordetella* uses deamidation as a virulence mechanism to colonise and infect hosts in response to adverse environmental cues (King *et al.*, 2009). This was the first

report of a polysaccharide post-assembly modification by a transglutaminase member. It is thought that O-antigen modification by WbME after polymerization may provide resistance against phagocytosis, and serve as a survival mechanism outside the host to evade desiccation, thus representing an adaptation to the environmental reservoir (King *et al.*, 2009).

Likewise, periplasmic proteins with transglutaminase activity have been found in other microorganisms like *Pseudomonas aeruginosa* (*P. aeruginosa*). A transglutaminase has been identified in *P. aeruginosa* (PAO1 strain), an opportunistic bacterium that causes infection in immunocompromised hosts. *Pseudomonas* can survive in several habitats and hosts including mammals, plants, and nematode hosts. Its ability to survive in varied hosts and ecological niches contributes to its virulence (Milani *et al.*, 2012). Control of *P. aeruginosa* infections has been limited by the high antibiotic resistance associated with the different strains. A shot-gun antisense technique previously used in Gram-negative bacteria was employed to identify the novel PA2873 protein in the PAO1 strain. The results from this study showed that transglutaminase activity was associated with the periplasmic region of the PA2873 protein (Milani *et al.*, 2012). Further investigation revealed that the transglutaminase protein named TgpA performs an important function in the viability and growth of the PAO1 strain. The catalytic triad of the PA2873 protein was well conserved when aligned with characterised transglutaminases in prokaryotes and eukaryotes (Milani *et al.*, 2012).

1.1.6 Production of transglutaminase in microbial systems

Transglutaminase has been produced and detected both extracellularly in *Streptomyces* and intracellularly in *Bacillus*. Mostly, transglutaminases have been obtained by extracellular production using fermentation techniques which can be scaled up (Ando *et al.*, 1989). Several microbiological and molecular techniques have been employed to isolate and produce transglutaminase from microorganisms, to characterise and understand their functions. The first was extraction and purification of crude enzyme extracts obtained from microorganisms (Zhang, Zhu and Chen, 2009). Secondly, several strains have been screened for transglutaminase production and those with high yields have been isolated to optimise their transglutaminase production. Because higher yields were obtained with *S. mobaraense*, it was produced industrially for commercialisation (Yokoyama, Nio and Kikuchi, 2004). Attempts directed towards optimisation of transglutaminase production generated good results using fermentation broth, through which higher yields were obtained with lower production costs. However, the nitrogen sources used in the culture of *Streptomyces* for transglutaminase production contributed to 30% of bioproduction expenses, making the fermentation route expensive (Kieliszek and Misiewicz, 2014).

This finally led to the adoption of recombinant DNA techniques to obtain recombinant strains capable of high yield production (Yokoyama, Nio and Kikuchi, 2004). *Streptomyces mobaraense* was the first organism used for recombinant expression due to the available crystal structure (Kashiwagi *et al.*, 2002). First attempts of recombinant expression in *E. coli* yielded low activity, though the gene product had molecular weight and immunological properties identical to native transglutaminase (Zhang, Zhu and Chen, 2009). Yet another group succeeded in expressing *S. mobaraense* transglutaminase in a *Streptomyces* host as a pro-transglutaminase which resulted to an active, mature protein. From there scientists employed varying hosts and expression mechanisms to obtain increased transglutaminase yields (Zhang, Zhu and Chen, 2009).

1.1.6.1 Cloning and expression of microbial transglutaminases in *E. coli*

Transglutaminases from several microorganisms including *Streptomyces* species and *B. subtilis* have been recombinantly expressed in *E. coli* (Marx, Hertel and Pietzsch, 2007). *E. coli* is a widely used system in recombinant protein production due to the large plethora of strains and expression vectors available. Additionally, *E. coli* is easy to grow and manipulate but also facilitates cloning, plasmid propagation and transformation (Rosano and Ceccarelli, 2014). The potential is further elucidated by its use in the production of about 30% recombinant therapeutic proteins, and recombinant foreign proteins overexpressed in *E. coli* account for more than 45% of the overall proteins made (Javitt *et al.*, 2017). Several trials of recombinant soluble protein overexpression in *E. coli* have been hampered by numerous challenges, including inclusion body (IB) formation coupled with low yields (Zhang, Zhu and Chen, 2009). IB formation has mostly resulted from cytosolic expression of proteins in *E. coli*, and refolding of enzymes from IB is time consuming and not always successful (Kashiwagi *et al.*, 2002). This was observed during overexpression of transglutaminase in *E. coli* optimised with codon usage and signal peptides to direct secretion to the periplasm. IB were formed in the cytosol and yields of protein following refolding were low. In studies which employed different refolding strategies, yields from refolding insoluble IB were as low as 15-20% (Kawai *et al.*, 1997). However, some studies used IB to obtain high yields of chemically renatured mTG protein (Zhang, Zhu and Chen, 2009). After failures of soluble mTG expression in *E. coli*, *Corynebacterium* was explored as an expression system for soluble protein excretion in culture medium (Marx, Hertel and Pietzsch, 2007). Recombinant expression in *E. coli* has been mostly achieved by expressing Tgases as pro-enzymes and active enzymes were obtained by protease activation (Marx, Hertel and Pietzsch, 2007; Liu *et al.*, 2011; Javitt *et al.*, 2017). In other studies *Streptomyces* as host for recombinant expression resulted in higher

yields (Javitt *et al.*, 2017). Therefore, it was proposed that exogenous methods utilised for recombinant transglutaminase production needs to include gene cloning, expression and enzyme activation mechanisms. So far, the best yields for recombinant mature microbial transglutaminase expression have been with *Streptomyces* hosts (Zhang, Zhu and Chen, 2009).

1.1.6.2 Improving soluble transglutaminase protein expression in *E. coli*

Optimisation of transglutaminase protein expression can be done by altering plasmid copy numbers, inductor and inductor concentrations, codon usage, temperature following induction and localization. Secretion to the periplasm in many instances has produced high levels of soluble, properly folded proteins attributed to chaperones and decreased protein concentration (Marx, Hertel and Pietzsch, 2007). However in some cases, low levels of transglutaminase expression were observed despite the inclusion of signal peptides for periplasm secretion (Javitt *et al.*, 2017). The relevance of temperature on growth and post-induction has been demonstrated to impact soluble transglutaminase protein expression levels. Increased quantities of soluble protein were obtained with cultivation at temperatures less than 37°C (Marx, Hertel and Pietzsch, 2007). Also, solubility could be enhanced by decreasing the cultivation temperatures following induction with IPTG. The temperature shift approach has been used to solve the problems associated with refolding of insoluble IB. Prior cultivation at 37°C to generate enough biomass and later induction at lower temperatures (24°C) favoured slow expression, resulting to high amounts (90%) of soluble transglutaminase protein. Similarly, high solubility was achieved with a lactose auto-induction culture at a growth temperature of 28°C (Marx, Hertel and Pietzsch, 2007).

Another expression strategy has adopted a polycistronic gene expression system comprising a pro-domain with the mature protein, both under the control of a T7 promoter and non-covalently linked to each other (Liu *et al.*, 2011). The active transglutaminase protein was obtained by sequential co-expression of the pro-enzyme followed by the mature protein. Direct expression of *Streptomyces* MTG as mature active enzyme resulted to inactive protein or insoluble IB formation, stipulating the role of the pro-peptide in protein folding for expression (Liu *et al.*, 2011). Yet another recombinant protein production approach consisted of a constitutive expression system for production of soluble, active *S. mobaraense* mTG protein in *E. coli*. In these studies the expressed proteins were directed to the periplasm to facilitate correct protein folding, thereby decreasing IB formation and equally reducing proteolytic degradation of expressed products (Javitt *et al.*, 2017). Nonetheless, constitutive expressions can be optimised by including periplasm protease deficient strains to decrease enzyme

degradation (Marx, Hertel and Pietzsch, 2007). These systems simplify the expression scheme of active enzyme, bypassing downstream processes such as proteolytic activation and renaturation from IB (Kieliszek and Misiewicz, 2014). This enables a quicker expression analogous to current industrial scale production which bypass protease activation to produce active mTG. Another benefit is that intracellular enzyme expression decreases the workload associated with large volumes as observed with extracellular production (Javitt *et al.*, 2017). These constitutive protein expression systems illustrate that active transglutaminase accumulation in the periplasm may not be entirely toxic to the host as formerly hypothesised. Given that secretion to the periplasm usually limits production, higher yields can be obtained by co-expression with periplasmic secretory proteins (Javitt *et al.*, 2017).

Streptomyces hosts have been successfully used for recombinant active MTG production because it contains endogenous proteases capable of activating the proenzyme. Co-expression of the protease with the mature enzyme is required for other expression strains lacking activating proteases (Washizu *et al.*, 1994). Heterologous expression of active enzyme in *S. lividans* have been successful, even though further optimisation of production and yield are still required. The use of strong promoters for expression in *Streptomyces* improved yield, signifying that enzyme expression could be optimised with endogenous promoters from various transglutaminases or ameliorated versions. In *S. lividans*, MTG expression could be greatly decreased due to high GC content in the host genome and the presence of rare codons. However, rare codons were also found in the *Streptomyces* transglutaminase gene, implying expression of transglutaminase in *S. lividans* could be enhanced by codon optimisation. Ensuing, using optimised endogenous transglutaminase promoters and codon optimisation, high yields of active *Streptomyces* transglutaminase were obtained following expression in *S. lividans* host. Codon optimisation boosted productivity by 73.6% and lessened the fermentation time frame (Liu *et al.*, 2016).

Genetically modified organisms when used as expression systems did not attain the yields of native *Streptomyces* species (Zhang, Zhu and Chen, 2009). Conversely genetic modification of MTG for recombinant cytoplasmic expression in *E. coli* resulted to soluble and active enzyme, with activity comparable to wild type *S. mobaraense* MTG. Specifically, the high yields and enzyme activity was achieved by introduction of a 3C protease cleavage site along with pro-domain mutants (Rickert *et al.*, 2016). Important residues required for chaperone folding activity and mTG reaction stability were identified after conducting an alanine-scan of the enzyme pro-domain. The identified pro-domain amino acid residues preserved chaperone activity, but weakened mTG-cleaved pro-domain interactivity in a temperature dependent manner (Rickert *et al.*, 2016). This enabled correct mTG folding and kept the enzyme inert

throughout expression at 20°C yet generated full activity when moved to 37°C because of weakened domain interactions. The mTG-bound pro-domain in *E. coli* promoted increased protein expression, excluding toxicity associated with protein crosslinking (Rickert *et al.*, 2016).

1.1.7 Applications of microbial transglutaminase

Microbial transglutaminase protein crosslinking activity influences the physicochemical characteristics of food. More so the protein modifications alter size, conformation, charge and stability of proteins (Fatima and Khare, 2018). Importantly transglutaminase from *Streptomyces* has attracted attention in industry due to its small size, fast rates of reaction and wide substrate specificity (Liu *et al.*, 2016). *S. mobaraense* has been produced by fermentation and commercialised for the food industry. In the food industry, transglutaminase modifications provide lower allergenicity, shelf life stability, enhance flavour, structure, texture, appearance, solubility, water holding capacity and other properties of food (Zhu and Tramper, 2008). Also transglutaminase is a protein modifier that has been used to increase the stability and volume of some food products, as well as prolonging shelf life by forming protein films (Kieliszek and Misiewicz, 2014). Microbial transglutaminases (MTG) have been used in the manufacture of dairy, baking and meat products with greater nutritional value and permitting the use of lower quality starting materials. Transglutaminase activity can alter the properties of foods such as dairy products where structure, gelling and emulsifying capacities, thermostability and shelf-life is increased (Yokoyama, Nio and Kikuchi, 2004). Transglutaminase crosslinking facilitates the incorporation of essential amino acids to obtain foods with greater nutritional content, and can be exploited to obtain foods with decreased fat contents (Kieliszek and Misiewicz, 2014).

Microbial transglutaminase is also exploited in other sectors where it provides novel and diverse applications in the textile, leather, material science and biomedical industries (Zhu and Tramper, 2008). In the textile industry, protein modifications provide greater resistance, toughness and improves leather via crosslinking (Javitt *et al.*, 2017). In pharmaceuticals transglutaminase can be used for conjugation of DNA, peptides, and proteins to manufacture antibody-drug conjugates for therapeutic usage (Strop, 2014). In tissue engineering transglutaminase has been employed to crosslink hydrogels used to deliver drugs (Rickert *et al.*, 2016). Also, MTG can be used for enzyme immobilisation on solid supports e.g. silica gel and ion-exchangers (Marx, Hertel and Pietzsch, 2007).

1.1.8 Transglutaminase inhibition

Given the involvement of transglutaminase transamidation activity (especially TG2) in several disease processes, the development of novel Tgase-specific inhibitors would provide potential therapeutic solutions (Griffin, Casadio and Bergamini, 2002). There exist two main classes of transamidation inhibitors with different mechanisms of enzyme inactivation: reversible inhibitors which could be competitive or non-competitive, and then irreversible inhibitors (Martucciello *et al.*, 2020). Competitive reversible inhibitors which contain aliphatic chains with primary amines incorporated in their structure act as competitive inhibitors of primary amine substrates, thereby blocking crosslinking and in the process are included in the γ -glutamyl moiety of the protein substrate. Reversible non-competitive inhibitors compete for the transglutaminase 2 nucleotide cofactors (GTP/GDP) binding site and thus allosterically inhibit the enzyme (Griffin, 2013). Transglutaminase 2 irreversible inhibitors are the most studied inhibitors. They target cysteine at the enzyme active site and incorporate an important electrophilic warhead capable of reacting covalently with the nucleophilic sulphur contained in cysteine at the active site, thus completely inhibiting the enzyme. It is also established that disulphide compounds like cystamine have affinity for the transglutaminase active site (Griffin, 2013). Furthermore, potent inhibitors (R281/R283) of mammalian transglutaminase activity impede transglutaminase-catalysed modifications which contribute to several human diseases. This could suggest the relevance of such inhibitors in blocking microbial transglutaminase-directed activities, and thus could serve to screen for novel microbial transglutaminases (Martucciello *et al.*, 2020).

1.1.8.1 Effect of inhibitors on microbial transglutaminase activity

Several reagents and chemicals were tested for their effect on inhibition of MTG activity. Reagents like PMSF was shown to be inhibitory to microbial transglutaminase activity in some studies, indicating the presence of serine within the transglutaminase active site in these microorganisms (Nur'amaliyah, Zilda and Mubarik, 2016). In *B. subtilis*, PMSF was mildly inhibitory to transglutaminase activity (Suzuki *et al.*, 2000). On the other hand, PMSF had no effect on microbial transglutaminases from other sources including *S. mobaraense* transglutaminase and MTG-TX from soil, suggesting the lack of serine within the transglutaminase catalytic site (Ando *et al.*, 1989; Jin *et al.*, 2016). In most studies, N-ethylmaleimide (NEM) was shown to inhibit microbial transglutaminase activity again confirming the presence of cysteine at the active site (Ando *et al.*, 1989; Suzuki *et al.*, 2000; Macedo, Sette and Sato, 2011; Jin *et al.*, 2016). Likewise, NEM and certain divalent cations were highly inhibitory to mammalian transglutaminases confirming cysteine at the catalytic

site, meanwhile PMSF was only mildly inhibitory (Kumazawa *et al.*, 1996, 1997). The action of IAA (iodoacetamide acid) on transglutaminase activity was controversial just like PMSF. In some studies IAA completely abolished microbial transglutaminase activity, whereas it increased transglutaminase activity in other microbial sources (Ando *et al.*, 1989; Nur'amaliyah, Zilda and Mubarik, 2016). However, IAA was inhibitory to transglutaminase in mammalian organisms such as Japanese oyster and walleye pollack liver (Kumazawa *et al.*, 1996, 1997).

1.2 CLOSTRIDIODES DIFFICILE

1.2.1 Overview: Epidemiology and Risk factors

Clostridioides difficile (*C. difficile*) is a spore-forming, Gram-positive, obligate anaerobe majorly implicated in nosocomial infections. It causes antibiotic-induced nosocomial diarrhoeas and related healthcare infections in most parts of the world (Worthington and Hilton, 2016). In the U.K, *C. difficile* is the main causative agent of nosocomial diarrhoea (Polage, Solnick and Cohen, 2012). This pathogen is a major threat to healthcare institutions worldwide due to its incidence, constitutively high resistance to antibiotics due to the spore form, and high recurrence in treated patients (Worthington and Hilton, 2016). About 20-30% of all admitted patients under antibiotic treatment get *C. difficile* associated antibiotic-induced diarrhoea (Paredes-Sabja, Shen and Sorg, 2014). In the preceding decade there has been a significant rise in the occurrence of *C. difficile* infections (CDI) and the high frequency of recurrence has imposed a huge expense on treatment for health systems (Fimlaid *et al.*, 2013). About 25% of patients after treatment are prone to a reinfection and succeeding infections render patients more likely to experience more severe forms. The recurrence observed in CDI occurs due to the capacity of the pathogen to form endospores (Putnam *et al.*, 2013). Epidemic *C. difficile* strains including ribotypes 027 and 001 have a greater intrinsic sporulation ability suggesting a role in their dissemination (Underwood *et al.*, 2009).

Importantly, the major risk factors for CDI include hospitalization, contact with infected persons, improper hygiene, old age, antibiotic consumption, and reduced immunity. Antibiotic consumption especially in hospitalized patients increases susceptibility to *C. difficile* infections because they perturb the normal gut flora that prevents colonization of the pathogen (Curry, 2017). This significantly depletes the species responsible for dehydroxylating primary bile salts at the 7 α -hydroxyl position. This results to a rise in the amount of cholate derivatives and a reduction in deoxycholate levels in the caecum, which favours *C. difficile* growth and colonization (Ridlon, Kang and Hylemon, 2006). The degree of colonization is influenced by primary bile salts that are abnormally present in the large bowel, thus providing a harmless germinant for residual spores in the colon. Hence, the normal bacterial flora plays a vital protective function in the metabolism of cholate products to deoxycholate which prevents *C. difficile* growth (Sorg and Sonenshein, 2008). However, CDI can also be found in healthy young adults with no background of antibiotic use (Vohra and Poxton, 2011). Likewise in susceptible individuals (e.g. immunocompromised), destruction or perturbation of the gut flora

creates an ideal environment for CDI to develop (Sorg and Sonenshein, 2008). More so during antibiotic therapy spores enable *C. difficile* to persist in the host thus promoting recurrence after treatment completion (Wheeldon *et al.*, 2008). Furthermore, treatments such as antacids or proton pump inhibitors raise the risk of acquiring *C. difficile* antibiotic-associated diarrhoea because the increased gastric pH promotes the survival and then colonisation of the spores (Jump, Pultz and Donskey, 2007). The bile salt chenodeoxycholate has been demonstrated to block spore germination and growth in *C. difficile*, providing a level of control in gut colonization by C.D (Sorg and Sonenshein, 2009).

1.3 SPORULATION

To survive conditions of nutrient depletion, bacterial cells alter their physiology and metabolism on entry into the stationary phase. In the case of *Clostridia* and *Bacilli* they undergo differentiation into a spore formation process during unfavourable conditions and germinate when favourable conditions return (Lawley *et al.*, 2009). Endosporulation is a primitive cell differentiation mechanism in bacteria that results in the formation of very tough dormant structures called endospores. Due to their resistance, spores can accumulate in many different environments for prolonged periods particularly in the gut in metazoan hosts (Pereira *et al.*, 2013). Thus sporulation also serves as an intrinsic mode of transmission and provides bacterial resistance towards microbicidal agents (Rosenbusch *et al.*, 2012). In *Clostridia*, little information is known about the regulatory mechanisms after the exponential phase and the commencement of sporulation (Heap *et al.*, 2007). Spore biology in *Clostridioides difficile* has been limited by the lack of homologs in other spore-formers such as *Bacillus* and some *Clostridium* species. For instance, spore coat proteins in *Bacillus subtilis* (*B. subtilis*) have roughly 25% homologs in *C. difficile* (Fimlaid *et al.*, 2013). In *C. difficile* sporulation is induced by stress factors such as nutrient deprivation, quorum sensing and other unknown causes. Also *C. difficile* is induced to sporulate in the presence of oxygen (Jump, Pultz and Donskey, 2007). The sporulation process serves as a survival mechanism in harsh environmental contexts and spores can persist in these dormancy conditions for several years (Paredes-Sabja, Shen and Sorg, 2014). There are considerable differences in the sporulation regulation program and activity of sigma factors between *C. difficile* and *B. subtilis* species or other *Clostridium* species (Underwood *et al.*, 2009; Paredes-Sabja, Shen and Sorg, 2014). Figure 1.3 summarises the sporulation process in *C. difficile*.

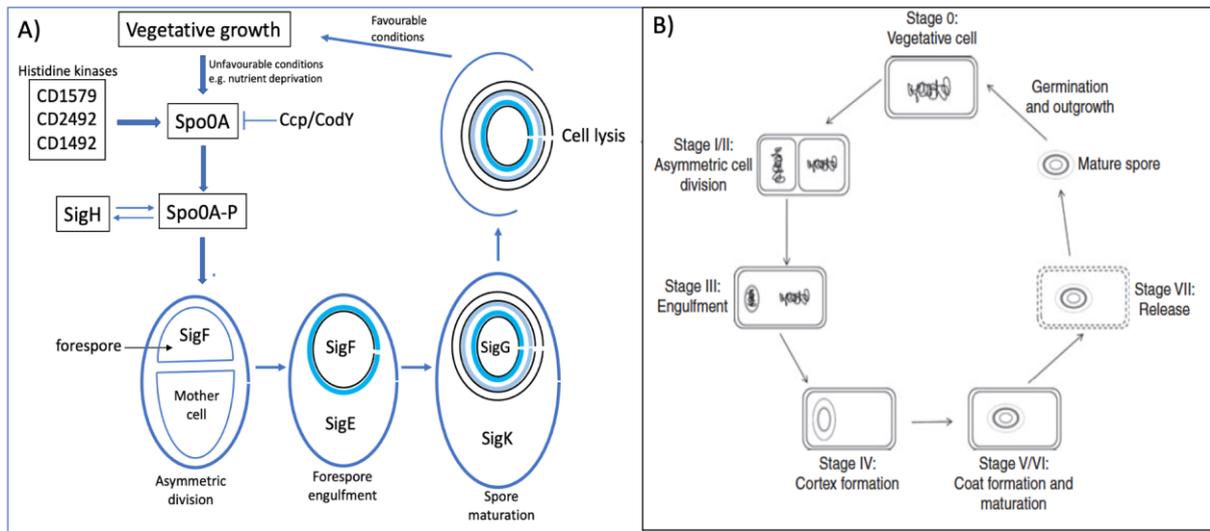


Figure 1.3. Spore morphogenesis during sporulation in *C. difficile*. Adapted from (Leggett et al., 2012). A) Morphogenesis. In response to appropriate signals the histidine kinases CD1579, 1492, 2492 induce Spo0A phosphorylation and thus sporulation initiation in *C. difficile*. This creates a cascade activation of sigma factors in the forespore and mother cell which regulate the sporulation process. SigF and SigE are early sporulation sigma factors activated respectively in forespore (FS) and mother cell (MC), meanwhile SigG and SigK are late sporulation sigma factors. Each transcription factor is activated during a specific event in sporulation as shown on the diagram. B) Sporulation process. The sporulation developmental process comprises a sequence of events that include (I/II) Asymmetric division of vegetative cells which generates a smaller prespore and a larger mother cell compartment, (III) Prespore engulfment by the mother cell; (IV) Peptidoglycan synthesis in the germ cell wall and cortex; (V) formation of spore coat; (VI) maturation of spore coat; (VII) mother cell lysis to release the mature, dormant spore; (0) during favourable conditions spore germinates and cell resumes vegetative growth.

1.3.1 Entry into sporulation

In *C. difficile* the signals that trigger sporulation are unique and species-specific, and Spo0A is a very pleiotropic cellular protein with great effects on overall gene expression (Underwood *et al.*, 2009). Sporulation is triggered under nutritionally unfavourable conditions which induces phosphorylation of Spo0A, the major transcriptional regulator in the initial stages in *C. difficile* sporulation (Paredes-Sabja, Shen and Sorg, 2014). In *C. difficile* several histidine kinases (HK) are involved in Spo0A activation either directly or indirectly in response to environmental signals. Environmental triggers of Spo0A phosphorylation are mostly unknown but could be found within the gastrointestinal tract when the bacteria are metabolically active (Lawley *et al.*, 2009; Rosenbusch *et al.*, 2012). Sensor kinases are in charge of TCS (two-component signal transduction system) activation and relay of phosphoryl groups in *Clostridia* (Underwood *et al.*, 2009; Steiner *et al.*, 2012). Depending on the species, sporulation can involve direct phosphorylation of Spo0A by the histidine kinase or the involvement of a protein that relays

the phosphate moiety directly to Spo0A. In *Clostridioides*, sporulation sensor kinases sense the signals and mediate direct phosphorylation of Spo0A (Paredes-Sabja, Shen and Sorg, 2014). Usually one sensor kinase responds to one particular signal ligand, and therefore having many kinases boosts the signal sensing, allowing many signals to control *spo0A* activation (Underwood *et al.*, 2009). The histidine kinases CD1579, 1492, 2492 are involved in Spo0A phosphorylation and thus sporulation initiation in *C. difficile*, with each responding to different ligands. In *C. difficile*, CD1579 is the only known histidine kinase that undergoes autophosphorylation and relays a phosphoryl group directly to Spo0A (Underwood *et al.*, 2009). The importance of histidine kinases in *C. difficile* sporulation was illustrated by a decrease in *C. difficile* 630 spore formation following inactivation of the histidine kinase CD2492 as compared to the wild type. Meanwhile no major effect was observed on vegetative growth and viability of batch cultures in rich medium which further demonstrated the importance of HK and Spo0A in sporulation (Underwood *et al.*, 2009).

1.3.1.1 Activation of Spo0A by Histidine Kinases

Sensor kinase domain structure consists of an N-terminus input domain and a catalytic C-terminal kinase that contains two subdomains. Histidine kinases sense and respond to cellular and environmental signals that mark termination of vegetative growth (Underwood *et al.*, 2009). This results to autophosphorylation of specified histidine residues at the catalytic centre of HK. The phosphate moiety is transferred to aspartate on Spo0F which is then relayed to Spo0A transcription factor by the Spo0B phosphotransferase itself being temporarily phosphorylated on histidine (Underwood *et al.*, 2009). The phosphorylation of aspartate at active sites of Spo0A enhances its activation and binding to target sequences (0A box) around gene promoters governed by Spo0A regulation leading to gene activation or suppression (Rosenbusch *et al.*, 2012; Steiner *et al.*, 2012). Upon activation by kinases, Spo0A mediates a series of events that leads to formation of a mature spore in a process called sporulation as summarised in Figure 1.3 (Paredes-Sabja, Shen and Sorg, 2014). There is less conservation amongst *B. subtilis* and *Clostridia* around the pathway towards Spo0A activation (Saujet *et al.*, 2013). Spore formation in *C. difficile* 630 was completely absent after a mutation in *spo0A*. Additionally, mutation of *spo0A* in *C. difficile* strain results in decreased persistence and transmission of disease (Rosenbusch *et al.*, 2012; Fimlaid *et al.*, 2013; Pettit *et al.*, 2014). Therefore spore-forming bacteria employ Spo0A to regulate their virulence and survival response that is most important for their pathogenicity (Underwood *et al.*, 2009). Further work is required to decipher the proteins involved in *spo0A* phosphorylation and activation of RNA polymerase sigma factors.

1.3.1.2 Role of Spo0A binding sites in spore formers

Spo0A undergoes phosphorylation dependent activation which disrupts transcriptional-inhibitory interactions enabling the DNA binding domain (DBD) to bind its targets. Indeed Spo0A recognises a specific motif, the '0A' box to which it binds with high affinity. In *C. difficile* Spo0A is involved in transcriptional regulation of a subgroup of early sporulation genes by specific binding to promoter regions (Rosenbusch *et al.*, 2012). There is high conservation of Spo0A DNA binding domain (DBD) between *B. subtilis* and *C. difficile*. Actually a homolog for the *B. subtilis* Spo0A DBD was found in *C. difficile* 630 Δ erm strain, CD1214 following BLAST searches. Thus *C. difficile* Spo0A DBD recognises the same 7 bp (TGTCGAA) 0A box in which Guanine and Cytosine are indispensable for high affinity binding (Heap *et al.*, 2007). Many early sporulation genes in other bacteria are directly regulated by Spo0A through binding to promoter regions illustrating conservation of sporulation. Notably some direct targets of Spo0A in *C. difficile* do not possess an exact equivalent to the consensus sequence or contain variations in the consensus box. These differences may indicate distinct levels of control, mechanism of activation, and promoter organisation (Rosenbusch *et al.*, 2012).

1.3.1.3 Role of Spo0A protein in non-sporulation events

Spo0A also regulates other programs aside sporulation including toxin production, biofilm formation, DNA replication, and genetic transformation (Rosenbusch *et al.*, 2012). In *C. difficile* Spo0A binds directly to SigH promoter which codes for the major transition phase sigma factor and controls processes beyond sporulation (Saujet *et al.*, 2013). Also Spo0A regulates several genes associated with metabolism, gene regulation, transport/binding and cell envelope architecture (Pettit *et al.*, 2014). Consistent with this view, two genes unrelated to sporulation were regulated by Spo0A following identification of *spo0A* recognition sites on their promoters. This observation may be found in other strains elucidating conservation of Spo0A regulation in different *C. difficile* strains (Rosenbusch *et al.*, 2012). It was reported that *spo0A* positively regulates the transcript and protein levels in most genes that contain Spo0A binding consensus. Notably the absence of a Spo0A binding consensus in many genes was reflected in disrupted transcript levels. This indicated indirect regulation or control via a non-Spo0A binding consensus. However, even in the absence of distinct Spo0A binding sites as observed in *C. acetobutylicum*, numerous transcriptional regulators display alteration in transcript quantities and strong *spo0A* directed control (Pettit *et al.*, 2014).

1.3.1.4 Gene expression control by Spo0A and SigH during sporulation

All *Clostridia* possess Spo0A and SigH just like *Bacillus* but do not rely on phosphorelay for Spo0A activation, instead utilise a two-component system (TCS) (Saujet *et al.*, 2013). Sig H plays an important role in the early stages of sporulation in *C. difficile* and *B. subtilis*. This is elucidated by the fact that majority of the traditional genes that act as direct targets of SigH are associated with sporulation or cellular division. SigH mediates indirect effects on other sporulation genes via SigF or *spo0A*. In *C. difficile*, SigH is very important as it regulates 60% of genes expressed after the exponential growth phase. In *C. difficile* 630E, *spo0A* and SigH were expressed during the exponential and early stationary phases following growth in TY medium (Saujet *et al.*, 2013). Furthermore, SigH controlled expression of most sporulation and cell differentiation genes at the beginning of the stationary phase and during the late phases of sporulation. Sporulation and stress response genes were expressed in the early stationary phase. Several other genes were expressed during these growth phases. This demonstrates an adaptation to environmental conditions triggered early at stationary phase when conditions become limiting (Saujet *et al.*, 2013).

Inhibition of SigH or *spo0A* leads to a nonsporulation phenotype. Therefore, SigH is equally as important in *C. difficile* sporulation as Spo0A (Underwood *et al.*, 2009). Analogous to *B. subtilis*, there is mutual regulation of SigH/*spo0A* synthesis in *C. difficile* and both control many sporulation genes. Indeed, *spo0A* positively regulates SigH transcription suggested by the presence of a *spo0A* box upstream the SigA promoter of SigH (Saujet *et al.*, 2013). Again, as seen in *B. subtilis* at the start of the stationary phase SigH is involved in the initiation of *spo0A* transcription. Additionally, in CD630E, SigH controls the production of CD2992 kinase highlighting another stage of regulation of Spo0A activity and further illustrating its role in sporulation. This is not observed with other histidine kinases. Thus in *Bacillus subtilis* (kinA, kinE) and *C. difficile*, SigH controls transcription of *spo0A* and some *spo0A* kinases (Saujet *et al.*, 2013).

In *C. difficile* both *spo0A* and SigH are under the direct control of Spo0A and so there is a possibility of *spo0A* autoregulation (Saujet *et al.*, 2013). This putative direct control by Spo0A may indicate absence of the AbrB homolog in the *C. difficile* genome. In *B. subtilis*, AbrB is a pleiotropic regulator which mediates *spo0A* dependent control of SigH. A steep rise in Spo0A levels was observed during growth towards the stationary phase, in line with a prototype where *spo0A* undergoes positive autoregulation. This suggests both SigH and *spo0A* trigger a rise in Spo0A levels at stationary phase. There is a likelihood of an indirect transcriptional control

on more regulators through Spo0A-mediated effects by AbrB, although it does not exist in *C. difficile* (Rosenbusch *et al.*, 2012).

1.3.2 Regulation of the Sporulation process

The sporulation process which is summarised in Figure 1.3 consist of distinct stages from asymmetric cell division, engulfment, cortex/coat formation and maturation, and finally release of the mature spore (Leggett *et al.*, 2012). There are still gaps in the understanding of cellular and molecular events during sporulation in *C. difficile* (Pereira *et al.*, 2013). There is high conservation of gene expression during sporulation amongst *B. subtilis* and *C. difficile*. In *C. difficile* spore morphogenesis, there are disparities in gene expression and activity. Regulation of gene expression by sigma factors takes place alternately in mother cell and forespore. This is aligned perfectly with cellular morphogenesis (Fimlaid *et al.*, 2013). Four specific RNA polymerases govern the expression of several genes and proteins produced during sporulation (Ragkousi and Setlow, 2004). The RNA polymerase sigma factors are activated in specific compartments in *C. difficile* and *B. subtilis*. The early events of sporulation are regulated by SigF in the forespore and SigE in the mother cell. This is followed by SigG and SigK after completion of engulfment. Also, early events in sporulation such as asymmetric division and engulfment are under SigH control (Saujet *et al.*, 2013).

1.3.2.1 Asymmetric cell division

The mechanism of Spo0A phosphorylation and activation is detailed in Section 1.3.1.1. Phosphorylated Spo0A triggers entry into sporulation and the shift to asymmetric division of the vegetative cell cytoplasm which generates two unequal units, the smaller forespore and the larger compartment called the mother cell, both separated by a polar septum. Each compartment acquires one chromosome copy from the parent (Monroe and Setlow, 2006). Notably SigF and SigG are activated following asymmetric division, with activity being restricted to the forespore, meanwhile SigK and SigE activity is confined in the mother cell (Kuwana *et al.*, 2006). Communication between the two compartments ensures progress of gene expression and sporulation. However, the intercompartmental crosstalk amongst sporulation sigma factors seen in *B. subtilis* is absent in *C. difficile* (Underwood *et al.*, 2009; Paredes-Sabja, Shen and Sorg, 2014).

1.3.2.2 Engulfment and spore maturation

The engulfment of the forespore (FS) is the hallmark of sporulation and is delineated by a double membrane in the mother cell cytoplasm (Pereira *et al.*, 2013). Two RNA polymerases, SigG and SigK are activated following forespore engulfment, and their activities are confined within the mother cell (Kuwana *et al.*, 2006). Following asymmetric division, the mother cell engulfs the prespore and mediates a series of maturation events including peptidoglycan cortex, coat, and exosporium formation (Fimlaid *et al.*, 2013). Assembly of most spore structures takes place within the mother cell except the germ cell wall made within the FS (Pereira *et al.*, 2013). The spore coat is layered on the outer spore surface and certain species have an outermost exosporium layer (Paredes-Sabja, Setlow and Sarker, 2011). Eventually the mature spore is released into the environment following lysis of the mother cell and assumes dormancy, even though there could be a short period of metabolic activity after which the spore resumes to metabolic inactivity or dormancy (Fimlaid *et al.*, 2013). The spores produced though metabolically inactive present high resistance to various insults and stress stimuli (Setlow, 2014). Nevertheless, they respond to nutrients during favourable conditions allowing the spore to germinate and outgrow, and resume vegetative growth (Kuwana *et al.*, 2006; Leggett *et al.*, 2012).

1.3.2.3 Initiation of germination and outgrowth following spore lysis

The bacterial spores produced though metabolically inactive, can resume active metabolism via germination and outgrowth (Kuwana *et al.*, 2006). Three distinct phases namely activation, germination and outgrowth are required in the conversion of a dormant spore to a vegetative cell. Firstly, 'activation' induced by environmental conditions such as pH, heat or chemicals which makes the dormant spore obliged to commence germination by interrupting the dormancy state (Leggett *et al.*, 2012). Spore activation prior to germinant addition starts germination more promptly and completely. However, activation is reversible and does not automatically resolute to germination and outgrowth (Kuwana *et al.*, 2006). Also activated spores preserve majority of the characteristics related to dormancy. Contrarily a spore already destined to germinate cannot reverse to the dormancy state. Spores germinate in response to several stimuli, which include amongst others metabolizable nutrient germinants and nonmetabolizable germinants (Paredes-Sabja, Setlow and Sarker, 2011; Leggett *et al.*, 2012). Outgrowth refers to the developmental processes that happen following germination such as metabolism initiation, macromolecular production, spore expansion, appearance and

maturation of a new cell which marks spore resumption to vegetative cell growth. Hence the germinated spore is converted to a growing cell in the process of outgrowth (Setlow, 2014).

1.4 Spore structure and Function

The unique structure of the spore contributes to its resistance properties. Spores have a multi-layered structure absent in growing cells which consists of the coat surrounded by the outermost exosporium layer found in certain species. The detailed structure of the bacterial spore is shown in Figure 1.4.

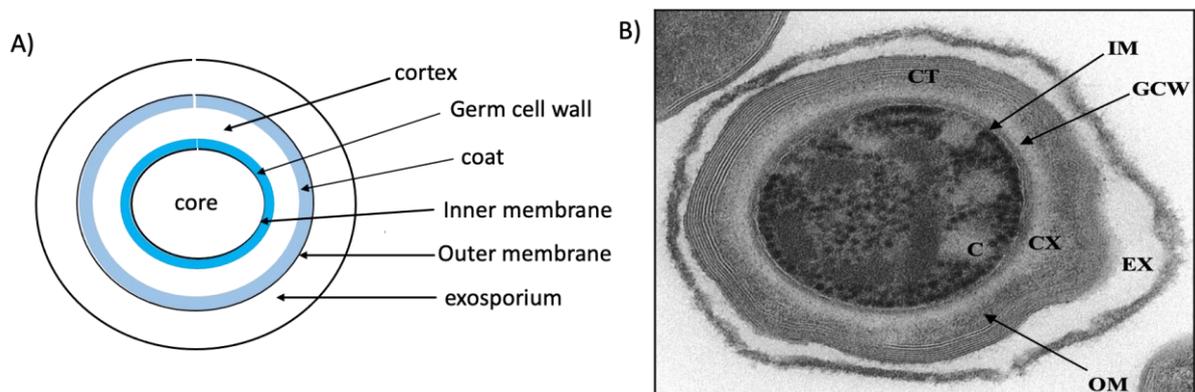


Figure 1.4. Structure of a bacterial spore. A) Structure of a *C. difficile* bacterial spore. B) Electron micrograph of *C. difficile* spore (Lawley *et al.*, 2009). The image shows the ultrastructure of a mature spore consisting of several concentric layers: an outermost exosporium and innermost core surrounded by an IM. Next to the IM is the GCW overlaid by a thick cortex, outer membrane and coat respectively. C: core; CX: cortex; CT: coat, EX: exosporium; IM: inner membrane; OM: outer membrane; GCW: germ cell wall.

1.4.1 Function of Small acid-soluble proteins (SASPs)

Small acid spore proteins (SASPs) are low molecular weight proteins that are highly abundant and only located in spores. These proteins are produced in the developing spore in the late stages of sporulation and constitute an important source of amino acids for the developing spore. Two major types of SASPs exist in *Clostridium* and *Bacillus* species (Leggett *et al.*, 2012). The α/β -type is the only SASP found in *Clostridia*, rich in hydrophobic amino acids and plays an important role in spore protection against chemicals and related agents such as wet heat which target the spore DNA. Thus α/β -SASPs by binding the spore DNA backbone provides a shield against DNA damage (Sanchez-Salas *et al.*, 2011). Likewise, the core is protected from UV damage because of the conformational alterations induced following binding of α/β -SASP proteins (Leggett *et al.*, 2012).

1.4.2 Core

The innermost spore layer is the core which is comprised of genetic elements (DNA, RNA), ribosomes and enzymes. Additionally, the core contains high amounts of Ca-DPA (calcium dipicolinic acid), has low water content and is saturated with DNA α/β small acid soluble proteins (SASPs) which contribute to its resistance characteristics (Paredes-Sabja, Setlow and Sarker, 2011). The core is deprived of water, i.e. highly dehydrated which plays an important role in the spore's metabolic dormancy and resistance characteristics (Leggett *et al.*, 2012). In dormancy conditions during which bacterial spores are metabolically inactive, significant damage to proteins, nucleic acids and ribosomes accumulate (Setlow, 2006). Moreover, spores cannot repair and/or degrade the damaged proteins and nucleic acids. Repair of damage to nucleic acids and proteins occurs during germination and outgrowth wherein spores are reactivated and resume vegetative growth (Setlow, 2006; Leggett *et al.*, 2012). The core which is enriched with Ca-DPA contributes to the resistance features of the spore, mostly directed towards preventing DNA damage (Sanchez-Salas *et al.*, 2011). The spore core water level is the main determinant of wet heat resistance, whereby higher heat resistance is associated with reduced core water levels. In certain *Clostridium* species the link between core water levels and resistance has been shown (Sanchez-Salas *et al.*, 2011). The spore core water levels can be altered using several methods such as varying the sporulation temperatures, as higher temperatures result to reduced core water levels or utilizing strains that cannot synthesize DPA. Spores that cannot produce DPA have considerably higher core water levels, even though this can be decreased to wildtype amounts by addition of DPA to the sporulation medium (Leggett *et al.*, 2012). In addition, core mineralization provides some amount of wet heat resistance given that greater mineralization produces greater wet heat resistance, some of which is contributed by reduced core water levels. Compared to other mineral ions, calcium provides a higher degree of wet heat resistance (Leggett *et al.*, 2012).

1.4.3 Inner membrane

The inner membrane (IM) surrounding the core is highly compressed and contains largely immobile lipids. The IM is a phospholipid layer with extremely low permeability and therefore shields the core DNA from damaging chemicals (Paredes-Sabja, Setlow and Sarker, 2011). In some spore-forming bacteria, the IM harbours germinant receptors (GR) that binds specific germinants during spore germination (Setlow, 2014). Biophysical analysis reveal that the impermeability is conferred by the state of lipids inside the membrane. It was further proposed that a possible constriction of the IM by the cortex could justify the impermeability and lipid immobility. Another suggestion was that decreased lipid mobility could be caused by the low

core water levels that have an effect on the lipid mobility in the IM (Leggett *et al.*, 2012). The permeability of the spore is increased by many oxidizing agents that destroy the inner membrane, allowing toxic chemicals to penetrate and mediate the release of DPA from the core reserves. Thus oxidizing agents target several proteins that lie inside the IM which if altered could affect membrane integrity (Leggett *et al.*, 2012).

1.4.4 Cortex

The cortex plays an important role in keeping the core in a metabolically inactive and dehydrated state, and altering inner membrane permeability thereby contributing to spore resistance (Kevorkian, Shirley and Shen, 2016). The intact cortex maintains spore dehydration and thus important in preserving spore heat resistance (Leggett *et al.*, 2012). During the last sporulation stages, the cortex exerts a large pressure on the spore core thereby contributing to attainment in core dehydration, although recent studies suggests that its static structure preserves core dehydration (Setlow, 2006). The pressure exerted by the cortex on the core may be influenced by the varying levels of crosslinking of the cortex (Leggett *et al.*, 2012). The spore cortex is made of peptidoglycan (PG) which differs from the PG in vegetative cells, since the N-acetylmuramic acid (NAM) residues are deprived of teichoic acids. In cortex PG, peptide chains are absent in about 50% of NAM groups. Rather the NAM residues undergo cyclisation to produce muramic- δ -lactam (MAL). Yet another 25% of NAM residues contain only L-alanine in the side chains (Paredes-Sabja, Setlow and Sarker, 2011). All these NAM modifications prevent peptide crosslinks to be established amongst glycan strands. Crosslinked peptide chains are found in only about 3% of NAM groups. It is predicted that the reduced crosslinks in spore peptidoglycan could enable the spore to achieve and preserve a highly dehydrated core. Even so, later studies have shown that spore dehydration is not altered by the degree of crosslinking (Leggett *et al.*, 2012). The germ cell wall (GCW) which surrounds the IM is made of PG and during outgrowth this structure is converted to the cell wall. The PG cortex surrounds the GCW with a structure that differs from that of growing cells by a few modifications as described above (Paredes-Sabja, Setlow and Sarker, 2011). No role in spore resistance has been ascribed to the GCW. The GCW and cortex peptidoglycan differ in that MAL is absent in the former, which enables specific degradation of the cortex during germination. During germination, cortex lytic enzymes target the MAL substance in the cortex (Setlow, 2014).

1.4.5 Outer membrane

The outer membrane just beneath the spore coat originates from the mother cell and has no role in spore resistance but is important for sporulation. More research is needed to explore the morphology and function, especially in relation to spore resistance and permeability (Paredes-Sabja, Setlow and Sarker, 2011).

1.4.6 Coat

Analysis of *C. difficile* spores reveals a coat structure with distinct laminations. The spore coat is a highly proteinaceous complex structure made of two layers: a thicker outer coat and an inner coat with a lamellar structure as observed under a microscope (Ragkousi and Setlow, 2004). The coat surrounds the cortex and is majorly made of protein, with small amounts of carbohydrates. In fact, about 50-80% of the overall spore proteins are contained in the coat (Leggett *et al.*, 2012). The coat is made of soluble and insoluble fractions, with about 70% of coat proteins being soluble (Monroe and Setlow, 2006). Additionally, the spore coat contains other crosslinks such as isopeptide and dityrosine crosslinks. These crosslinks perhaps contribute to the spore's resistance to mechanical and chemical agents (Katsunori. Kobayashi *et al.*, 1998; Zilhão *et al.*, 2005). The fully assembled coat is important for spore germination and provides resistance to peptidoglycan lytic enzymes, toxic chemicals and mechanical damage (Ragkousi and Setlow, 2004). The coat is highly resistant to several chemical agents and toxic compounds such as hydrogen peroxide, hypochlorite, which would otherwise kill spores more quickly in the absence of the coat. Thus the coat could serve as a detoxifying structure for harmful chemicals prior to entry into the innermost core or inner membrane layers (Klobutcher, Ragkousi and Setlow, 2006; Kuwana *et al.*, 2006). However, nutrient germinants which are small molecules can penetrate through the coat. Therefore the coat serves like a sieve by allowing only germinant molecules to bind to their germinant receptors underneath the coat, and rejecting possibly toxic chemicals (Monroe and Setlow, 2006). Also the coat is permeable to certain low molecular weight toxic chemicals which due to their small sizes can bypass the coat barrier and gain entry into the innermost spore core. Up-to-date no distinct protein has been detected as a component in spore protection (Leggett *et al.*, 2012).

Coat assembly requires an organized mobilization of morphogenetic proteins at the spore surface. Morphogenetic proteins are proteins that direct the assembly of coat proteins (Putnam *et al.*, 2013). Coat formation occurs late during sporulation and is preceded by gene expression and protein assembly which are highly regulated. In sporulating cells, the mother cell is the site of synthesis of most coat proteins and coat structural constituents, regulated by

SigK. However, SigE controls the production of some coat proteins before the prespore is engulfed by the mother cell (Zilhão *et al.*, 2005). GerR and SpoIID transcription factors control the activity of SigE, meanwhile GerE transcription factor regulates activity of SigK (Plomp *et al.*, 2014). Complete coat assembly ensures full spore resistance, although most coat proteins display redundancy in their functions (Monroe and Setlow, 2006). In *C. difficile*, SpoIVA and SpoVM mediate spore morphogenesis, specifically SpoIVA which is important for coat assembly around the forespore analogous to *B. subtilis*. However, contrary to *B. subtilis*, SpoIVA in *C. difficile* is not required for cortex synthesis (Putnam *et al.*, 2013). Additionally, *C. difficile* lacks homologs of SafA, SpoVID and CotE as seen in *B. subtilis*, implying that different proteins possibly mediate their assembly process (Paredes-Sabja, Shen and Sorg, 2014). A new morphogenetic protein, CD3567 has been identified recently in *C. difficile* spores. This protein displayed direct interactions with SpoIVA forming a complex called SipL (SpoIVA interacting protein L) that orchestrates assembly of the coat around the forespore (Touchette *et al.*, 2019). Actually similar to SpoVID in *B. subtilis*, the SipL protein mediates spore coat conversion from the mother cell proximal pole to full casing (Putnam *et al.*, 2013). Further work is required to define the functions of coat proteins, understand the mechanism of coat protein assembly, and to decipher the proteins involved in activation of RNA polymerase sigma factors.

1.4.7 The Exosporium in Spore-forming bacteria

In *C. difficile* the outermost layer of the spore is the exosporium characterised by hair-like projections, and similar to *B. subtilis* lacks the interspace gap that distinguishes it from the spore coat. In *C. difficile* 630 the exosporium is described as an electron dense compact structure (Mora-uribe *et al.*, 2016). The exosporium is a glycoprotein-rich, collagen-like layer produced in the mother cell. In *C. difficile*, the exosporium contains glycoproteins that encode for BclA proteins also produced by the mother cell (Leggett *et al.*, 2012). The BclA proteins are made of three domains, the N-terminal, collagen-like, and C-terminal domains. A role has been found for *bclA1* which is posttranslationally cleaved and in *C. difficile* 630 spores is located entirely on the exosporium layer. The other BclA proteins have undefined roles. During sporulation TOMATO reporter proteins are localised to *C. difficile* spore surfaces by BclA proteins via their N-terminal domain (Paredes-Sabja, Shen and Sorg, 2014). Furthermore, a cysteine-rich protein, CdeC has been identified in the exosporium of *C. difficile* spores and plays an important role in the correct assembly of the exosporium. A mutation in CdeC severely diminished exosporium mass, resulting in incomplete assembly of the coat and exosporium layers (Mora-uribe *et al.*, 2016). Additionally, CdeC was suggested to play an anchoring role at the coat-exosporium interface. Eventually CdeC was found to exert multiple

functions during sporulation which demands further investigation (Paredes-Sabja, Shen and Sorg, 2014). In *C. difficile* the coat and exosporium both facilitate direct spore interactions with the environment such as small molecule nutrient germinants (Paredes-Sabja, Shen and Sorg, 2014). There is still a lot of controversy regarding the exosporium stability in *C. difficile*, which could be dependent on strain type or on proteases utilised during purification. Recent studies show that the exosporium can be removed with proteinase K (Mora-uribe *et al.*, 2016). The exosporium has not been completely characterized but is thought to play a role in spore adherence to surfaces and spore persistence (Leggett *et al.*, 2012). Specifically, the exosporium by conferring hydrophobicity to spores influences spore adherence to static surfaces. This prevents spore attachment to colonic surfaces and germination, thereby inhibiting disease transmission (Mora-uribe *et al.*, 2016).

1.4.8 Role of surface associated proteins

In *C. difficile* strain 630, immunofluorescence analysis enabled identification of proteins localised at the spore surface. These included proteins with catalase activity, CotD and CotG confined to the exosporium meanwhile CotCB had little activity (Paredes-Sabja, Shen and Sorg, 2014). RNA polymerase specific sigma factors control the expression of spore surface associated proteins produced during sporulation. The expression of *bclA1*, *bclA2*, *bclA3*, *CdeC*, *CotCB*, *CotD*, *CotA*, *CotE* genes is under the control of SigK and *CotB* expression is dependent on SigE (Saujet *et al.*, 2013). In *C. difficile*, SigG regulates the expression of *SodA* and *CotG* genes. Therefore, confinement of *SodA* and *CotG* to the spore surface implies that they are incorporated onto the spore surface after synthesis in the forespore, and translocation via the membranes (Kuwana *et al.*, 2006). Also other spore surface detected proteins possessed enzymatic activities including *SodA* (superoxide dismutase) homologous to *B. subtilis* *SodA* and *CotE*, a peroxiredoxin reductase and chitinase (Zilhão *et al.*, 2005). *SodA* had no role in spore ultrastructure and resistance. No structure or functional homology has been found in *C. difficile* *CotE* when compared with *B. subtilis* *CotE*. Additionally, a group of hypothetical proteins (*CotA*, *CotB*, *CotF*) was identified on the spore surface, and eventually it was discovered that these proteins in *C. difficile* 630 are present in an intact exosporium or translocate to the exosporium. Similar to *CdeC*, *CotA* is important in the assembly of the coat and exosporium layers (Paredes-Sabja, Shen and Sorg, 2014).

1.5 PATHOGENESIS OF *C. DIFFICILE* INFECTIONS

1.5.1 Bile production and metabolism

Primary bile is synthesised in the liver predominantly made of cholate and chenodeoxycholate, bound to glycine or taurine and secreted by the gall bladder during digestion. Active reabsorption of bile occurs during transit via the distal ileum after which it is salvaged in the liver (Ridlon, Kang and Hylemon, 2006). In the lower bowel, the normal microbial flora deconjugate the glycine-conjugated cholate (glycocholate) to release cholate and glycine which is enough to initiate spore germination and outgrowth (Ridlon, Kang and Hylemon, 2006). In humans the small intestine is the site for *C.D* spore germination, and taurocholate which is the germinant for *C. difficile* is found in the small intestine (Sorg and Sonenshein, 2009). In a healthy individual, ingestion of *C.D* spores results in germination in the jejunum which contains high levels of nutrients and cholate derivatives (Ridlon, Kang and Hylemon, 2006). Notably spores resist transit via the stomach and duodenum, after which spore germination is triggered by glycine and cholate in the jejunum (Sorg and Sonenshein, 2008, 2009). The primary bile salts and germinated spores that travel through the small intestine to the anaerobic caecum are deconjugated by the normal bacteria flora to deoxycholate, a potent inhibitor of vegetative growth thereby preventing colonization of the host (Ridlon, Kang and Hylemon, 2006). Deoxycholate is the major bile salt in human faeces. However, spores which could not germinate while passing through the upper digestive tract could germinate in the presence of deoxycholate, but the toxicity in the lower bowel prevents spore outgrowth (Sorg and Sonenshein, 2008).

1.5.2 Dissemination of *C. difficile* infection

Dissemination of *C. difficile* infections (CDI) occurs mainly via surfaces contaminated with spores especially in the clinical setting. Spores are the main vectors for disease transmission due to their high resistance and prevalence (Worthington and Hilton, 2016). In susceptible individuals, CDI is caused by spore ingestion which triggers irreversible spore germination upon exposure to nutrient germinants in the small intestine (Sorg and Sonenshein, 2008; Wheeldon *et al.*, 2008). Thus, colonization of susceptible host by *C. difficile* requires the dormant spore to germinate following sensing of specific bile salts in the gut (Sorg and Sonenshein, 2008). This gives rise to metabolically active toxin-producing vegetative cells which trigger considerable epithelial damage and severe inflammatory reactions of the gut. This results to toxin-mediated disease characterized by diarrhoea which can progress to pseudomembranous colitis, toxic megacolon and then death in some cases induced by sepsis (Dembek *et al.*, 2013). Upon establishment of infection, two glucosylating toxins namely

enterotoxin TcdA and cytotoxin TcdB are produced by *C. difficile* spores which generate the symptoms associated with disease (Dembek *et al.*, 2013). Both toxins permeate the cell (intestinal barrier) via endocytosis following binding of the C-terminus to distinct receptors on the plasma membrane. Acidification of the endosome is required for effective toxin translocation and once internalised Rho GTPases (Rho, Ras, Rac) are glucosylated via the catalytic N-terminal domain leading to their inactivation (Di Bella *et al.*, 2016). Because Rho GTPases regulate the actin cytoskeleton, their inactivation disrupts tight junctions and actin cytoskeleton resulting in fluid accumulation, reduced transepithelial resistance and eventually damage of the intestinal lining giving rise to diarrhoeal symptoms associated with CDI. Also, the toxins induce cytotoxic effects that trigger inflammation and pseudomembranous colitis (Kuehne, Cartman and Minton, 2011).

Infected patients frequently excrete vegetative cells in faeces that are metabolically active and thus easy to kill. However, vegetative cells are intolerant to continued oxygen exposure during which they sporulate, making elimination more challenging, thus promoting dissemination and infection. Hence prolonged survival of *C. difficile* bacteria outside the host or anaerobic milieu of the large bowel requires the spore form (Jump, Pultz and Donskey, 2007). Also in infected patients, excretion of spores in faeces explains their persistence in hospital facilities due to the difficulty associated with spore elimination. This remains a major source of infection and reinfection via inadvertent intake of contaminated matter (Wheeldon *et al.*, 2008; Lawley *et al.*, 2009).

1.5.3 Control of *C. difficile* infection

The control of clinically-acquired *C. difficile* related-disease is significantly hindered by spore formation (Wheeldon *et al.*, 2008; Worthington and Hilton, 2016). Induction of the sporulation pathway during CDI culminates in more spore production which explains the persistence and horizontal propagation amongst hospitalized patients. However, during CDI strains which cannot form spores would not survive in the host colon to be disseminated horizontally (Paredes-Sabja, Shen and Sorg, 2014). Due to their metabolic dormancy C.D spores show inherent resistance to antibiotics, host immune system attacks, and upon excretion into the surrounding they show resistance to routinely used disinfectants (Paredes-Sabja, Shen and Sorg, 2014). Additionally, their multi-layered structure provides robustness and resistance to several physical and chemical stimuli, which enables them escape killing following treatment with routine disinfectants (Dembek *et al.*, 2013). Cleaning is hampered by spore resistance to heat, chemicals and radiation, which explains their persistence in the environment for prolonged periods (Dembek *et al.*, 2013). Thus vigorous disinfection strategies are required in

the clinical milieu to limit the spread of infection (Worthington and Hilton, 2016). Consequently, spores are eliminated by using sporicidal agents like bleach since the metabolically inactive spores show resistance to routinely used bleach-free sanitizers. However sporicidal agents that are effective towards spore elimination are often harsh and toxic (Wheeldon *et al.*, 2008).

In the quest for strategies to limit dissemination of infection in the healthcare setting, the use of nutrient germinants on spores could be one means of decreasing spore resistance, thereby making them more susceptible to external stressors. Due to increased sensitivity of germinating spores to antimicrobials in comparison to dormant spores, a 'germinate to exterminate' approach was suggested as a novel mechanism to eliminate *C. difficile* spores (Wheeldon *et al.*, 2008). In this strategy, inducing spores to germinate following treatments with taurocholate-based germinants would decrease resistance associated with dormant spores, thereby making them easier to kill with standard disinfectants. The special germination combination enhances irreversible germination of spores to metabolically active cells (Worthington and Hilton, 2016). Therefore this approach can be incorporated into future infection control programmes thereby circumventing the use of more dangerous sporicidal agents (Worthington and Hilton, 2016). The susceptibility of spores and vegetative cells to several physical and chemical agents is summarised in Table 1.2.

Table 1.2. Summary of Spores and vegetative cell responses to physical and chemical stressors. (Setlow, 2006; Wheeldon *et al.*, 2008; Lawley *et al.*, 2010; Edwards *et al.*, 2016).

parameter	Spores	Vegetative cells
temperature	Resistant (up to 70°C)	sensitive
70% (v/v) ethanol	resistant	Sensitive (as low as 9.5% (v/v))
chloroform	resistant	sensitive
butanol	resistant	sensitive
Hydrogen peroxide 3% (v/v), 10% (v/v)	resistant	sensitive
1% (v/v) sodium hypochlorite (bleach) 1% (w/v) Virkon Strong acid	sensitive	sensitive
pH 5-10	Sensitive to bleach at pH 8-10 Maximal germination with germinants pH 6-8	≥5 (increased survival)

1.6 Diagnosis

Early signs of infection are based on the presence of foul-smelling diarrhoea. Previously, the gold standard for *C. difficile* diagnosis was toxin identification in stool by toxigenic culture methods (Crobach *et al.*, 2009). Eventually, toxin detection evolved to ELISA based methods and PCR assays (Curry, 2017). However, there were a number of drawbacks associated with some of the diagnostic methods including time consumption, labour-intensiveness, false positives, high costs and particularly low sensitivity. Recently a two-step approach has been adopted in which a conserved antigenic Glutamate dehydrogenase (GDH) assay is performed as first line test, and further analysis is pursued 'only' if the GDH test comes positive (Crobach *et al.*, 2009).

1.7 Treatment

Standard treatment involves antibiotic therapy usually metronidazole for mild cases, and vancomycin or fidaxomicin for severe infection. Current antibiotic treatment for CDI include metronidazole, fidaxomicin and vancomycin (Sorg and Sonenshein, 2008). The emergence of hypervirulent strains like ribotype 027 show an increasing resistance to antibiotic treatment

probably due to higher sporulation rates. *C. difficile* is resistant to several frequently used antibiotics against bacterial infections. There has been a rising resistance of *C. difficile* towards metronidazole treatment (Peng *et al.*, 2017). However, recurrences can occur after treatment completion because of the time required for repopulation of the colon and restoration of the natural equilibrium of primary and secondary bile salts (Sorg and Sonenshein, 2008). Other studies have also suggested the importance of faecal microbiota transplant therapy in recurrent cases of infection (Curry, 2017). Also the ability of chenodeoxycholate to block taurocholate induced germination can be explored in the development of novel therapies for prevention of CDI (Sorg and Sonenshein, 2009). The introduction of chenodeoxycholate in treatment regimens of patients taking antibiotics could prevent colonization due to its role in germination suppression. Chenodeoxycholate administration to patients undergoing treatment for CDI with vancomycin or metronidazole could block germination and additional vegetative growth following completion of antibiotic therapy, which serves to minimize chances of relapse (Sorg and Sonenshein, 2008). Furthermore, the introduction of probiotics that contain bacterial species capable of dehydroxylating cholates at the 7 α hydroxyl position or probiotics in the diets of patients under antibiotic treatment could be beneficial in the development of treatment options (Kalakuntla *et al.*, 2019).

1.8 GERMINATION

In several pathogenic spore forming species, germination is an initial and important stage in the advancement of spore pathogenesis. In presence of the appropriate stimuli such as germinants, spores forfeit their dormancy and resistance characteristics thereby causing disease (Paredes-Sabja, Setlow and Sarker, 2011; Setlow, 2014). Thus germination is a potential target in the research of treatments to obstruct disease progress (Paredes-Sabja, Setlow and Sarker, 2011). Traditionally germination is described as a variation in optical density produced by Ca-DPA release and core hydration (Sorg and Sonenshein, 2008, 2009). Germination is marked by a conversion of spores from phase bright to phase dark as described by phase contrast microscopy (Sorg and Sonenshein, 2008). Germination in *C. difficile* spores is a critical step for disease initiation. Spore germination is initiated when germinant receptors (GR) react with species-specific germinants (Dembek *et al.*, 2013). This initiates a cascade of enzymatic reactions that result to Ca-DPA release, core rehydration, cortex breakdown and vegetative cell outgrowth (Sorg and Sonenshein, 2008). The germination pathway in *C. difficile* is less well defined and characterized compared to *B. subtilis*. The germination activation pathway in *C. difficile* differs from that of *Bacillus* or other *Clostridium* species especially with respect to germinant sensing, induction of Ca-DPA release

and activation of cortex hydrolysis. *C. difficile* employs the Csp-SleC cortex hydrolysis pathway which is very well conserved (Kevorkian, Shirley and Shen, 2016).

Bacterial spores possess many germinant receptors with distinct specificities for germinants in different strains and species (Setlow, 2014). This possibly indicates the adjustments of the spore formers to their precise environmental niches. All GR are produced in the developing spore during the late stages of sporulation. Most spore formers would germinate in the presence of specific nutrient germinants usually found in the environments favourable for the growing bacteria from which spores originate (Paredes-Sabja, Setlow and Sarker, 2011). Germination induced by germinant binding to GR can be driven in several ways whereby single germinants or multiple germinants can bind a single GR, or multiple germinants cooperatively bind to multiple GR (Paredes-Sabja, Setlow and Sarker, 2011; Kevorkian, Shirley and Shen, 2016). Most spore-forming bacteria possess GR localised in the inner membrane that recognise particular amino acid germinants but these GR are absent in *C. difficile* (Paredes-Sabja, Setlow and Sarker, 2011; Setlow, 2014). *C. difficile* does not have GR homologs observed in other spore formers. Rather in *C. difficile*, a subtilisin-like pseudoprotease CspC serves as a non-canonical GR that senses bile salts (Paredes-Sabja, Shen and Sorg, 2014).

1.8.1 Role of subtilisin-like serine proteases (Csps) in *C. difficile* spore germination

Pseudoproteases play important roles in the regulation of *C. difficile* spore germination. Subtilisin-like serine proteases, known as Csps possess a conserved catalytic triad composed of histidine, aspartate and serine. In *C. difficile*, CspC is catalytically inactive due to a mutation in two of its active site residues (Kevorkian, Shirley and Shen, 2016). Additionally, contrary to earlier identified subtilisin-like serine proteases, Csps have a stabilising, large central jellyroll domain fold (Setlow, 2014). Essentially, CspB plays an important role in spore germination by utilising its protease activity to cleave pro-SleC to generate mature SleC. In *C. difficile* SleC is a cortex hydrolase important for degrading the peptidoglycan cortex. The importance of CspB was further illustrated by a 20-fold reduction in spore germination and impairment in pro-SleC processing following a mutation in its active site serine (Paredes-Sabja, Shen and Sorg, 2014; Kevorkian, Shirley and Shen, 2016). During spore germination in *C. difficile*, the pseudoproteases CspC and CspA proportionately control CspB protease activity. The regulation of CspB activity by CspC pseudoprotease involves protein-protein interactions. Therefore in *C. difficile*, CspC has a bifunctional role because it is a GR and also controls CspB activity (Kevorkian, Shirley and Shen, 2016). During germination, *C. difficile* codes for only one active CspB protease meanwhile *C. perfringens* employs three Csp proteases

(CspB, CspA, or CspC) to activate SleC (Paredes-Sabja, Shen and Sorg, 2014; Kevorkian, Shirley and Shen, 2016). Similarly in *B. subtilis*, the cortex is degraded by two cortex hydrolases SleB and CwlJ, and CwlJ binds to the Ca-DPA released following germinant sensing in order to be activated (Setlow, 2014).

Furthermore, a CspBA fusion protein in *C. difficile* results from the fusion of CspB to the catalytically inactive CspA domain. Germination efficiency in CD630 is enhanced by the tandem organization of active CspB and inactive CspA pseudoprotease domain and this is not conserved in all CD strains (Kevorkian, Shirley and Shen, 2016). During spore formation, the CspBA fusion protein is subject to interdomain processing which releases CspB that attaches onto spores, and the degree of attachment relies on SleC (Paredes-Sabja et al., 2014). The interdomain processing of CspBA fusion protein is mediated by YabG protease, after which the CspA pseudoprotease released from CspB protease attaches onto mature spores (Kevorkian, Shirley and Shen, 2016). It was also reported that the CspA domain performs a more important role than CspB domain during germination. In mature spores, the CspA pseudoprotease controls the levels of CspC germinant receptor and is essential for maximal spore germination especially when fused to CspB (Kevorkian, Shirley and Shen, 2016). Thus during sporulation direct interaction of CspBA with CspC helps in CspC stabilisation. Additionally, a complex between the three Csp proteins is formed in dormant spores, and in mature spores the complex may incorporate SleC whereby CspC/CspA pseudoproteases are thought to have scaffolding and regulatory roles (Kevorkian, Shirley and Shen, 2016).

1.8.2 Binding of *C. difficile* CspC germinant receptor to bile salts

C. difficile lacks genes that encode for identified GR in *Bacillus* or other *Clostridia*, but still germinates in presence of bile salts and cogerminants (Setlow, 2014). Through an undetermined mechanism, CspC binds to cholate bile salts such as taurocholate which are the main germinants for germination of *C. difficile* spores, and germination with taurocholate is enhanced by the cogerminant glycine (Paredes-Sabja, Shen and Sorg, 2014; Setlow, 2014). An enzymatic signalling cascade involving CspB protease and SleC cortex hydrolase is activated following germinant binding to CspC pseudoprotease (Kevorkian, Shirley and Shen, 2016). A cogerminant is a compound which cannot drive spore germination by itself, but can induce the activity of a germinant or another cogerminant (Ridlon, Kang and Hylemon, 2006; Paredes-Sabja, Setlow and Sarker, 2011). The addition of taurocholate to rich medium triggers rapid spore germination in *C. difficile* which cannot effectively induce germination by itself. Cholates derivatives act as cogerminants essential for *C. difficile* spores which initiate germination at varying rates, with taurocholate being the most effective for rapid induction of

germination (Sorg and Sonenshein, 2008; Wheeldon *et al.*, 2008). Notwithstanding certain *C. difficile* strains can germinate in the absence of taurocholate in rich medium, which highlights the complexity of GR mechanism and specificities (Wheeldon *et al.*, 2008). Glycine and taurocholate have not been used previously for germination of *Bacillus* or *Clostridium* spores highlighting their originality. Previously, reducing agents and lysozyme at high concentrations were used to induce germination in *C. difficile* spores (Sorg and Sonenshein, 2008). The first event following germinant-germinant receptor binding in germination is 'commitment', the point whereby interruption of germination can no longer reverse the germination process (Paredes-Sabja, Setlow and Sarker, 2011; Setlow, 2014).

A point mutation in the active site of CspC triggers spore germination in response to chenodeoxycholate regarded as a strong germination inhibitor (Paredes-Sabja, Shen and Sorg, 2014; Kevorkian, Shirley and Shen, 2016). *C. difficile* spores showed higher affinity towards chenodeoxycholate when compared to taurocholate. Chenodeoxycholate is a bile salt which provides strong competitive inhibition towards taurocholate-induced germination in *C. difficile* spores (Sorg and Sonenshein, 2009). It is hypothesised that *C. difficile* spores specifically recognize the 12 α -hydroxyl side chain in cholate products. The absence of this side chain in chenodeoxycholate or its sparse water solubility perhaps explains its inability to stimulate germination and growth (Sorg and Sonenshein, 2008).

1.8.3 Cortex hydrolysis

The last stage of spore germination is lysis of the peptidoglycan cortex which releases the physical restraints to core enlargement that permits uptake of water into the core (Setlow, 2014). After the constraints on the outer spore layers are cleared, the cell commences a phase of longitudinal growth along with restoration of cell metabolism which is when DNA, RNA and protein synthesis restart (Dembek *et al.*, 2013). Cortex hydrolysis is a prerequisite for activation of Ca-DPA release from *C. difficile* spores (Kevorkian, Shirley and Shen, 2016). Peptidoglycan cortex hydrolysis is mediated by cortex lytic enzymes which specifically recognize muramic- δ -lactam (MAL) in the PG cortex. The cortex in *C. difficile* is degraded by the cortex hydrolase SleC which is found in mature spores as a zymogen (pro-SleC) and subjected to proteolytic activation catalysed by CspB proteases (Paredes-Sabja *et al.*, 2011/14). Cleavage of the inhibitory pro-peptide in pro-SleC releases active SleC that digests the peptidoglycan cortex allowing metabolism and growth to resume in the spore (Kevorkian, Shirley and Shen, 2016). Complete core hydration enables enzyme activity and resumption of

metabolism, macromolecular synthesis and outgrowth of spores (Paredes-Sabja, Setlow and Sarker, 2011).

In *C. difficile*, there is no clear explanation on the role of CspC in Ca-DPA release following germinant sensing, but could be mediated through the cortex hydrolysis pathway or via direct stimulation of DPA channels (Setlow, 2014). It follows that in *C. difficile* spores, cortex degradation is required for activation of Ca-DPA release meanwhile in other spore-formers Ca-DPA is released immediately following germinant binding (Kevorkian, Shirley and Shen, 2016). The precise germination pathway via cortex hydrolysis is not well outlined. More studies are required to understand the mechanisms associated with GR-germinant binding, and activation of the cortex hydrolysis machinery for completion of germination (Paredes-Sabja, Shen and Sorg, 2014). Because *C. difficile* employs catalytically inactive pseudoproteases plus catalytically active CspB proteases in the regulation of spore germination, it is considered more of a *Peptostreptococcaceae* than a *Clostridiaceae*. *C. difficile* has been reclassified under the genus *Clostridioides* hence the name *Clostridioides difficile* (Kevorkian, Shirley and Shen, 2016). The summary of spore germination in *C. difficile* as discussed is shown in Figure 1.5.

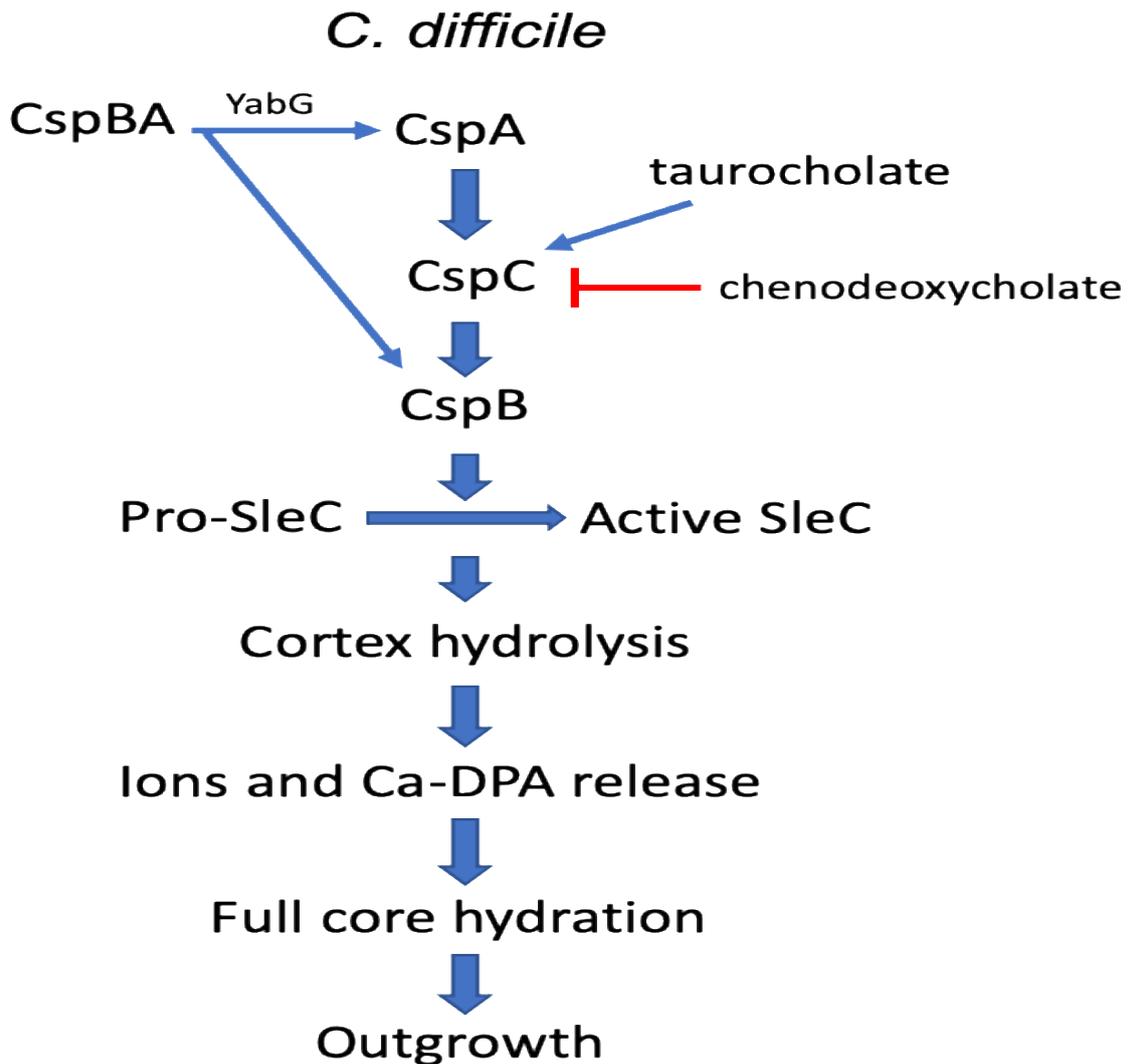


Figure 1.5. Flow chart of spore germination in *C. difficile*. Adapted from (Paredes-Sabja, Shen and Sorg, 2014). Germination is initiated following binding of bile salts (taurocholate) and cogerminants (CspA) to CspC receptor. This leads to CspB activation by an unknown mechanism which proteolytically cleaves pro-SleC. Active SleC degrades the peptidoglycan cortex which activates Ca-DPA release from the core, and uptake of water into the core. Complete core hydration allows resumption of metabolism, macromolecular synthesis and outgrowth of spores. Chenodeoxycholate is an inhibitor of taurocholate-induced spore germination.

1.9 Rationale

The involvement of transglutaminases in sporulation of some bacteria, their ability to increase resistance and promote virulence suggests the need for further investigation on the role of Tgases in sporulation in *C. difficile*. The importance of *C. difficile* spores in the persistence and propagation of infection necessitates further studies to discern the mechanisms associated with spore formation, and host interaction and possibly development of novel

treatments aimed at preventing transmission of *C. difficile* infection. The detection of novel important genes not explored previously by antibiotics can facilitate the discovery of alternative efficient antimicrobials to surmount existing resistance associated with disease control and treatment. Thus this study seeks to identify transglutaminases in *C. difficile* and determine its association with sporulation. The *C. difficile* '630' strain was used in this study because it has been previously involved in an outbreak of pseudomembranous colitis, but also has a fully sequenced genome.

1.10 Objectives

Overall Objective:

To determine whether *C. difficile* 630 contains transglutaminase-like genes and investigate its association with sporulation.

Specific objectives:

- Detection of transglutaminase activity and expression of transcripts during sporulation in *C. difficile* 630.
- To identify putative transglutaminases in *C. difficile* 630 using bioinformatics analysis and detect homology with other microbial transglutaminases and characterised mammalian transglutaminases.
- Recombinant expression, Purification and Characterization of transglutaminase-like proteins produced in *E. coli*.
- Purification and characterisation of transglutaminase-like proteins produced during sporulation in *C. difficile* 630, and identification by proteomic analysis.

CHAPTER 2. MATERIALS AND METHODS

2.1 Materials

All media used for bacteria culture were purchased from Oxoid (Hampshire, United Kingdom). Library efficiency DH5 α competent cells, CloneJET PCR cloning kit, Bovine Factor Xa native protein, TRIzol reagent and the Pierce fluorescent protease assay kit were purchased from ThermoFisher Scientific (Loughborough, United Kingdom). The GenElute bacterial genomic DNA extraction kit, ExtrAvidin peroxidase, cysteamine dihydrochloride, N-ethylmaleimide (NEM) was obtained from Sigma-Aldrich (Dorset, U.K). Wizard SV Gel and PCR Clean-up system and the Wizard Plus SV Minipreps DNA Purification System were purchased from Promega (Southampton, United Kingdom). Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase, Antarctic phosphatase, loading dye and DNA ladder were obtained from New England Biolabs (Hertfordshire, United Kingdom). The cDNA synthesis and SYBR green qPCR kits were purchased from NEB. The SDS-PAGE prestained protein ladder was purchased from Geneflow (Staffordshire, United Kingdom). Yeast, bovine serum albumin (BSA), ethylenediaminetetraacetic acid (EDTA), Phenylmethanesulfonyl fluoride (PMSF), ammonium sulphate, and Dithiothreitol (DTT) were purchased from Melford (Suffolk, United Kingdom). N-(Biotinyl)cadaverine, and Z-Gln-Gly were procured from Zedira (Carmstadt, Germany). All purification columns used in this study were obtained from GE Healthcare (Loughborough, United Kingdom). Instant Blue protein stain was procured from Expedeon (Cambridge, United Kingdom). All other chemicals and reagents were purchased from Melford or Sigma unless otherwise stated.

2.1.1 Microorganisms and growth conditions

All media used in this study were prepared as per manufacturer's instructions. After autoclaving, media intended for anaerobic culture were equilibrated for about 15 hours in the anaerobic chamber prior to use. All bacteria strains and plasmids used in this study are listed in Appendix 1. For generation of plasmids, *E. coli* DH5 α strains were used and *E. coli* BL21 (DE3) and Rosetta 2(DE3) (Novagen) strains were used for bacterial protein expression. *E. coli* cells were grown at 37°C in Luria-Bertani (LB) broth or agar supplemented with carbenicillin (50 μ g/ml) and ampicillin (50-100 μ g/ml) as appropriate. Chloramphenicol (34 μ g/ml) was included in culture media for protein expression with Rosetta cells. Liquid cultures were grown on a shaking incubator (180 rpm) at 37°C. DNA sequencing was done on all plasmid DNA constructs by Eurofins Genomics. *C. difficile* 630 strain used in this study was a kind gift from Dr Sarah Kuehne (University of Birmingham). *C. difficile* cultivation was performed in an anaerobic cabinet (Don Whitley MiniMACS, Bingley, England) at 37°C with a

gas mixture of 5% CO₂, 85% N₂ and 10% H₂. *C. difficile* culture was performed in Brain heart Infusion (BHI) broth or agar supplemented with yeast and L-cysteine with overnight static growth in anaerobic conditions at 37°C. Bacterial strains were stored as glycerol stocks at -80°C, achieved by mixing liquid overnight cultures with 50% (v/v) glycerol to obtain a 15% (v/v) glycerol stock.

2.2 Bioinformatics analyses

The NCBI database was used to search for transglutaminase-like genes in *C. difficile* 630 strain by keyword searches. The protein sequences of putative transglutaminases (Tgases) in *C. difficile* 630 were entered as query in the NCBI BLAST program to search for sequence homologs. The protein sequences of selected bacteria and mammals containing functional Tgases were downloaded from NCBI, and a phylogenetic tree was generated with Clustal Omega MLSA tool using the neighbour joining method. Also, multiple alignments of the selected Tgase sequences were generated with the same tool. Signal peptides within the *C. difficile* 630 Tgase genes was predicted using the SignalP 4.1 server. Predictions of protein localizations and isoelectric points were performed using Softberry's ProtCompB and ExPasy tools respectively.

2.3 Analysis of Transglutaminase transcripts during sporulation in *C. difficile* 630

2.3.1 Total RNA isolation

Total RNA was extracted from *C. difficile* 630 using the Trizol extraction method described in the manufacturers' instructions. Briefly 10 ml of sporulating culture was harvested at different time points (OD₆₀₀= 0.1-1.6) and centrifuged at 16 000 g for 3 min. Cell pellets were resuspended in 1 ml TRIzol reagent and mixed by repeated pipetting. After incubation for 5 minutes at room temperature, chloroform was added to the lysed cell suspensions and shaken vigorously for 10 secs. Samples were further incubated for 15 minutes, and centrifuged at 4°C, 12 000 g for 15 minutes. The uppermost supernatant fraction was mixed with 0.1 vol of 3 M sodium acetate plus 2 vols of 70% (v/v) ethanol and incubated for 1 hour at -20°C. After centrifugation for 15 minutes the RNA pellet was washed twice with 70% (v/v) ethanol by repeated centrifugation. The resulting pellet was air dried and dissolved in 50 µl nuclease-free water (NFW). The concentration and purity of total RNA extractions was determined on a Nanodrop 1000 spectrophotometer. RNA samples were stored at -80°C for further analysis. Using this protocol about 55 µg total RNA was obtained from sporulating cells. RNA extractions were relatively pure with absorbance at 260/280 nm greater than 1.7.

2.3.2 cDNA synthesis and qPCR amplification of transglutaminase genes

For reverse transcription reactions, cDNA was synthesised with the LunaScript RT SuperMix kit following the manufacturers' instructions. About 750 ng of total RNA was converted to cDNA using random hexamers and oligo-dT primers in a 20 µl reaction volume containing 2 µl cDNA, 4 µl of 5X LunaScript RT supermix and 14 µl NFW. The cDNA synthesis was performed at 55°C for 10 min. At the end of the program, a 1 min incubation step at 95°C was included to inactivate reverse transcriptions. The cDNA was immediately stored in the freezer at -20°C for further analysis.

Primer pairs for the qPCR reactions were designed on Primer 3 (NCBI) based on the gene sequences on NCBI. The primers had a GC content of 45-55%, T_m of 62-68°C and amplicon size of 110-217 bases. The primer sequences were analysed through BLAST to ensure specificity. The Roche light cycler 480 was used for real time qPCR amplification of Tgase genes in *C. difficile* 630, sporulation gene (*spo0A*) and reference gene (*rpsJ*). The qPCR amplification was conducted in triplicate in 20 µl reactions containing 2 µl synthesized cDNA, 250 nM gene specific primers, and 10 µl SYBR green qPCR master mix. The qPCR reaction program included a 5 min initial denaturation at 95°C followed by 30 cycles of amplification (95°C for 10 sec, 60°C for 10 sec, 72°C for 10 sec) and a melting curve analysis at the end of the run. In every qPCR run, negative (non-template) controls were included. Primer efficiencies were determined using the standard curves obtained from serially diluted cDNA templates representative of the different time points. Primer specificity for each primer pair was checked by melting curve analysis. Reference genes were used for normalisation of gene expression. The qPCR reactions were performed on three independent biological replicates.

2.4 DNA METHODS

2.4.1 Genomic DNA extraction.

Genomic DNA was extracted from overnight cultures of *C. difficile* 630 using the GenElute Bacterial Genomic DNA kit (Sigma) following the manufacturer's instructions. An overnight culture (1.5 ml) of *C. difficile* 630 in BHI supplemented broth was centrifuged at 16000 g for 5 min. The cell pellet was re-suspended and incubated with 10 mg/ml lysozyme for 30 min at 37°C. Proteinase K (2 mg/ml) and 200 µl lysis solution was added to the cell suspension and incubated for 10 min at 55°C. Ethanol (200 µl) was then added to the cell lysate and vortexed for 10 seconds. Prior to sample loading, the column preparation solution was centrifuged through the binding column at 16000 g for 1 min. Afterwards the lysate was spun through the column at 6500 g for 1 min. Following two washes with wash solution, genomic DNA was

eluted from the column with 200 µl of Tris-EDTA solution. The success of genomic DNA extraction was confirmed after analysis of 10 µl eluate by agarose gel electrophoresis (AGE).

2.4.2 Primer Design

The three oligonucleotide primers used for PCR amplification of the DNA portion encoding 720 bp, 1062 bp, 1083 bp Tgase-like genes in *C. difficile* 630 were designed based on the published gene sequence on NCBI. Two restriction sites were incorporated at the ends of the primer sequences. The forward primers included a *Bam*HI site and ATG start codon immediately before the coding sequence, meanwhile the reverse primers incorporated the stop codons after the coding region followed by the *Eco*RI restriction site. The addition of two bases 'TC' immediately upstream of the ATG start codon in each forward primer was to allow for in-frame read through from GST. All primers used in this study were produced by Sigma as listed in Table 2.1 and prior to use was dissolved to 100 µM in nuclease-free water.

Table 2.1. Primers used in this study.

Primer	Sequence 5'-3'
	Amplification and plasmid screening primers
CDTGAF (CD-720)	GGATCCT CATGAATAAAGAATATTTAG
CDTGAR	GAATTC TCAATTCGTTTTTATC
CDTGBF (CD-1062)	GGATCCT CATGAAGAAAATTAGGAAG
CDTGBR	GAATTC TTATTCATATCGAACATAC
TGCF (CD-1083)	GGATCCT CATGAAAAAATCAGGAAATTAGC
TGCR	GAATTC TTAATTCACATATCTAGCG
pGEXF	GGGCTGGCAAGCCACGTTTGGTG
pGEXR	CCGGGAGCTGCATGTGTCAGAGG
CDTGBIF (CD-1062m)	ATGACAGCCAAAGAGGTAAC
CDTGBIR/ CDTGDIR	GAGGATCCCACGACCTTCG
CDTGDIF (CD-1083m)	ATGACAGCTAAAGATGTTAC
	qPCR primers
<i>rpsJ</i> -F (30S ribosomal protein S10)	GATCACAAGTTTCAGGACCTG
<i>rpsJ</i> -R	GTCCTTAGGTGTTGGATTAGC
Cd720qPCR9-F	TGTTAAGCGCAACACAGGTA
Cd720qPCR9-R	AGCATGAAGCCTACAAGGGA
Cd1062qPCR3-F	GTCCAAAGCTGTCACAAGCC
Cd1062qPCR3-R	GCCATTCTGAACATGCACC
Cd1083qPCR3-F	GGTAGCACAAGCAACAGCAG
Cd1083qPCR3-R	CCTCTGCTGCGTTCCAGTTA
<i>spo0A</i> qPCR-F	CTCAAAGCGCAATAAATCTAGGAGC
<i>spo0A</i> qPCR-R	TTGAGTCTCTTGAACCTGGTCTAGG

restriction enzyme sites are indicated in bold

2.4.3 PCR amplification

The KOD hot start DNA polymerase kit used for PCR amplifications was purchased from Novagen. This polymerase produces blunt-end PCR products for blunt cloning purposes, and thus removes overhanging nucleotides through its exonuclease activity. The PCR amplification of Tgase-like gene(s) of interest was performed using as template genomic DNA from *C. difficile* strain 630. The PCR amplification reaction was carried out in a total volume of 50 µl comprising 5 µl of 10x reaction buffer, 2 µM forward and reverse primers, 2.5 mM MgSO₄, 0.2 mM dNTPs, 1 µl genomic DNA, 1 unit KOD Hot Start DNA polymerase and nuclease-free water. The reaction was conducted in a DNA thermocycler (ABI GeneAmp PCR system 9700) set to the following conditions: a hot start at 95°C for 5 minutes, 10 cycles (0.5 min denaturation at 95°C, 1 min touch-down/annealing step from 66-56°C, 71-61°C, 70-60°C for 1062 bp, 720 bp, 1083 bp Tgases respectively, 1 min extension at 70°C), 30 cycles denaturation (0.5 min

at 95°C, 1 min at 56°C, 61°C, 60°C for 1062 bp, 720 bp, 1083 bp Tgases respectively), extension for 1 min at 70°C and a final extension at 70°C for 5 minutes. The success of the PCR reactions was determined by analysing 20 µl of PCR products by electrophoresis on a 1% (w/v) agarose gel. The 720 bp PCR product from CDTGAF/CDTGAR was named as **A**; the 1062 bp PCR product from CDTGBF/CDTGBR was named as **B**; the 1083 bp PCR product from TGCF/TGCR was named as **C**.

2.4.4 PCR gel purification

Gel purification was performed according to manufacturers' instructions with the SV Gel and PCR Clean-up kit. Briefly, products from three repeats were pooled together and separated by agarose gel electrophoresis (AGE), after which the desired DNA bands were excised from the gel with a scalpel. The gel band for each target gene was weighed and dissolved at 65°C for 15 minutes in a calculated volume of binding buffer corresponding to the gel weight. The dissolved gel band was spun through a silica membrane column for 1 minute at 16000 g. Next the column was washed two times with wash buffer by centrifugation at 16000 g for 1 minute. Finally, the bound DNA was eluted by centrifugation (16000g, 1 min) following incubation with 50 µl nuclease-free water (NFW) at room temperature for 1 minute. The presence of purified products was confirmed by AGE.

2.4.5 Agarose gel electrophoresis procedure

To prepare 1% (w/v) agarose gels, agarose powder was melted in 1X TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) in a microwave. After cooling the gel to 40°C, 2 µl of fluorescent intercalating DNA dye, ethidium bromide (0.3 µg/ml) was added and the gels were cast on horizontal plastic trays with gel combs inserted. Once completely set, gels were immersed in a tank with 1X TAE buffer. Then DNA samples were diluted with 6X loading dye, and after loading samples onto the gel, electrophoresis was conducted at 100V for approximately 50 minutes. The gels were visualized with a UV transilluminator, and images were recorded using Syngene software on the Genebox.

2.4.6 Quantification of DNA

The DNA concentration and purity of samples was measured by absorbance readings at 280/260 nm on a Nanodrop 1000 spectrophotometer (Thermo Scientific).

2.4.7 Preparation of competent cells

An overnight culture of *E. coli* BL21 (DE3) was diluted (1:50) in fresh LB broth and grown at 37°C to log phase ($OD_{600}=0.5$). Cells were harvested by centrifugation at 4000 rpm, 4°C for 10 minutes and the pellet was resuspended in ice-cold 100 mM calcium chloride, then incubated for 20 minutes on ice. Again, cells were centrifuged as above and resuspended in 2 ml of 100 mM ice-cold calcium chloride containing 15% (v/v) glycerol. After incubation for 20 minutes on ice, 200 µl aliquots were prepared, flash-frozen in a dry-ice/ethanol bath and stored at -80°C.

2.4.8 Transformation of competent cells

All ligation reactions, cloning and expression vectors were transformed into Library Efficiency DH5α competent cells following the heat-shock procedure at 42°C. Plasmids were mixed properly with the thawed cells and incubated for 30 minutes on ice, 45 secs at 42°C and then chilled for 2 minutes on ice. Next, LB broth was added, and the cell culture was incubated on a shaker (180 rpm) at 37°C for 1 hour. Following incubation, 100 µl of transformants was plated on pre-warmed LB agar plates containing suitable antibiotics to select for plasmids. Selection of certain plasmids required the addition of X-gal and IPTG to antibiotic supplemented plates for blue/white screening. Recombinant colonies were selected after overnight incubation at 37°C. Transformation was verified by AGE on plasmid DNA preps and restriction enzyme digestion of recombinant plasmids.

2.4.9 Cloning of Transglutaminase-like genes from *C. difficile* 630

The cloning strategy for target transglutaminase-like genes in *C. difficile* 630 is summarised in Figure 2.1, and the method adopted was restriction cloning.

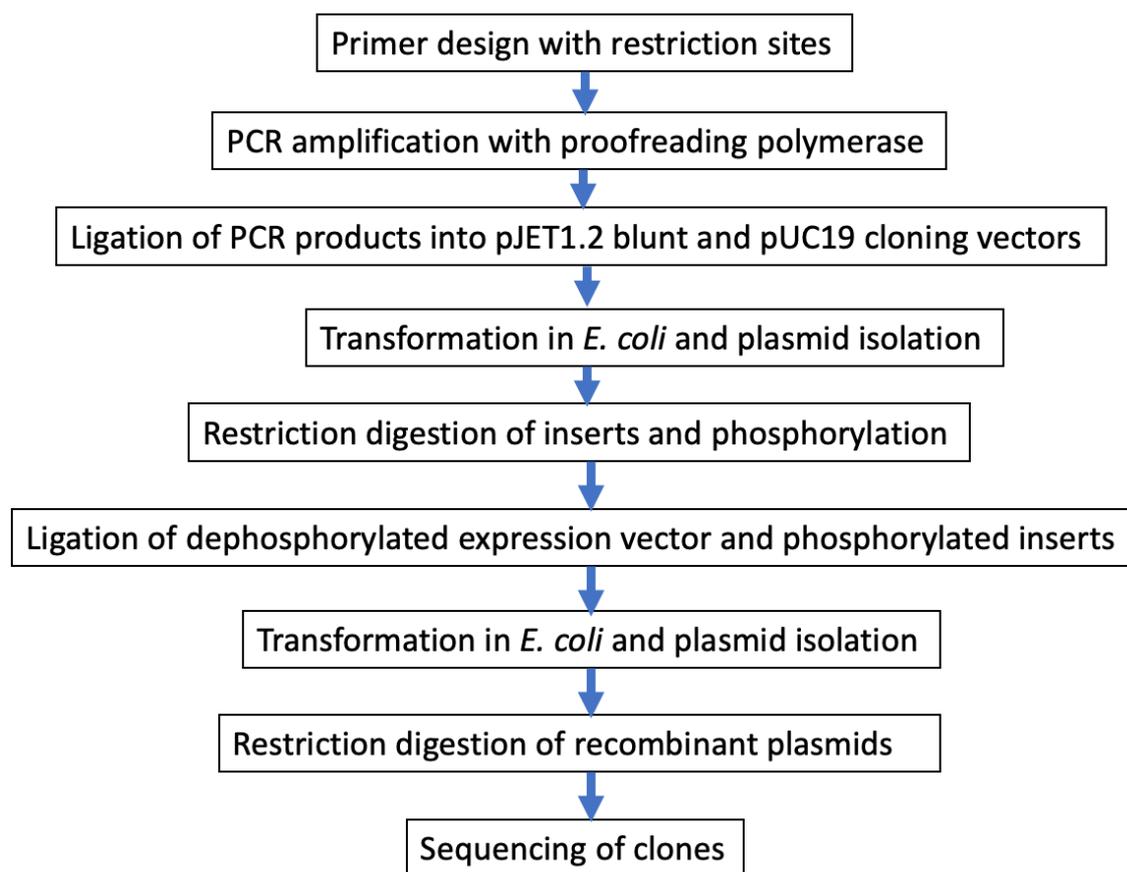


Figure 2.1. Flow diagram of cloning strategy for transglutaminase-like genes in *C. difficile*. The strategy consisted of ligating Tgase blunt PCR products into blunt cloning vectors, followed by restriction digestion of recombinant plasmids and final ligation of inserts into expression vectors. The resulting recombinant expression plasmids were digested to confirm the presence of target inserts and sequenced to select clones containing the correct sequence.

As shown in Figure 2.1, blunt-end cloning was used as the strategy for cloning target genes into cloning vectors. Thus, no overhangs were required for cloning purposes. The KOD polymerase used in this study removes overhangs via its exonuclease activity and cloning vectors have blunt-ended ligation sites within the multiple cloning sites which facilitate blunt-end cloning. The target inserts after restriction digestion from cloning vectors were asymmetrically ligated into expression vectors and recombinant expression plasmids were sequenced.

2.4.9.1 Plasmid isolation

Transformant colonies were seeded in LB broth supplemented with 100 µg/ml ampicillin and grown overnight at 37°C with shaking (180 rpm). Plasmid isolation was carried out following the alkali lysis method as per the manufacturers' instructions with the Wizard Plus SV Minipreps DNA purification kit. Cells (3 ml) were harvested by centrifugation at 10 000 g for 5 minutes. The supernatant was discarded, and excess media blotted on paper towel. The

resultant pellet was resuspended in 250 µl cell resuspension solution. The cell suspension was lysed with 250 µl lysis buffer by incubating for approximately 5 minutes, with gentle tube inversion. To the clarified lysate, 4 µl of alkaline protease was added and the tube was inverted four times during 5 minutes of incubation. The cell lysate was neutralised with 350 µl neutralisation solution and mixed by repeated tube inversion, then centrifuged for 10 minutes at 14 000 g. The clarified supernatant was loaded on a spin column leaving the pellet undisturbed and centrifuged at maximum speed for 1 minute. The flow through was discarded and the column was washed with ethanol-containing wash solution by centrifugation at 14 000 g for 1 minute. The wash step was repeated with 250 µl wash solution, and residual wash solution was removed by centrifugation at maximum speed for 2 minutes. Spin columns were transferred to clean microfuge tubes and plasmid DNA was eluted through centrifugation (16000g, 1 min) following incubation with 100 µl nuclease-free water for 1 minute. The presence of the isolated plasmid DNA was verified after analysis by AGE and plasmid DNA samples were stored at -20°C.

2.4.9.2 Restriction digestion

Restriction digestions were performed for 1-2 hours at appropriate temperatures depending on the enzymes following the kit instructions. The plasmids used allowed for cleavage with *Bam*HI, *Eco*RI and *Sma*I. About 1 µg of plasmid DNA or recombinant DNA was digested with 10 units of restriction enzyme, 1X reaction buffer and nuclease-free water in a final volume of 20 µl. Double digestions were carried out under standard conditions according to manufacturer's guidelines and in compatible buffers. Digestions were verified by AGE.

2.4.9.3 Dephosphorylation of plasmid DNA

Linearized cloning vectors were dephosphorylated for 30 minutes at 37°C in a 20 µl reaction containing 5 units Antarctic phosphatase, 1 µg DNA, 2 µl of 10X Antarctic phosphatase buffer and NFW. This also reduced the occurrence of background relegation of empty vectors. The reaction was heat-inactivated for 2 minutes at 80°C and purified before ligation.

2.4.9.4 Phosphorylation of PCR products for ligation

Blunt ended PCR products were phosphorylated for 30 minutes at 37°C in a 50 µl reaction containing 20 units T4 Polynucleotide kinase (PNK), 300 pmol DNA, 5 µl of 10X T4 PNK buffer, 5 µl of 1 mM ATP and NFW. The reaction was heat-inactivated for 20 minutes at 65°C and purified before ligation.

2.4.9.5 Ligation of DNA

All ligations were performed with T4 DNA ligase following the manufacturer's guidelines. Ligation reactions were performed with linearized dephosphorylated vectors and phosphorylated PCR products purified by agarose gel electrophoresis. Cloning of PCR products was performed with the CloneJET PCR cloning kit following the manufacturer's instructions. For standard cloning 3:1 insert to vector ratio were used. Blunt-end ligation reactions were set up in a final volume of 20 µl containing 1X reaction buffer, 5 units T4 DNA ligase, cloning vector (0.05 pmol ends), 0.15 pmol ends of purified PCR products and nuclease-free water. Cloning controls were blunted at 70°C for 5 minutes in 18 µl reactions containing 1 µl DNA blunting enzyme, 1X reaction buffer, 24 ng control PCR product, and nuclease-free water, followed by ligation with T4 DNA ligase. Plasmid DNA were re-ligated to serve as negative controls. Blunt-end ligation reactions were incubated for 20 minutes at room temperature meanwhile sticky end ligation mixtures were incubated overnight at 16°C. Ligation reactions were heat inactivated at 65°C for 10 minutes prior to transformation. The success of cloning was verified by AGE on plasmid DNA preps and restriction enzyme digestion of recombinant plasmids.

2.4.10 Sequencing of recombinant plasmids

Plasmids containing inserts of correct size were prepared for sequencing according to Eurofins protocols. Briefly, 15 µl of recombinant plasmids were mixed with 2 µl of 10 µM forward (pGEXF) and reverse (pGEXR) primers respectively in separate tubes. Plasmids were analysed by Sanger sequencing on pre-mixed DNA samples as completed by Eurofins Genomics (Germany).

2.5 Recombinant transglutaminase Protein Expression and Purification

2.5.1 Expression of recombinant proteins

Recombinant plasmids were transformed into *E. coli* BL21 (DE3) competent cells and plated on LB agar supplemented with ampicillin. The transformed BL21 cells were cultured overnight at 37°C with agitation (180 rpm) in LB broth containing 100 µg/ml ampicillin. A 1:50 dilution of O/N cultures in fresh LB broth was grown for approximately 2 hours on a shaker (180 rpm) at 37°C to reach an OD₆₀₀ of 0.5. To induce for protein expression, 0.5 mM IPTG was introduced into the culture at log phase and further incubated for 3 hours. Cells were harvested from 500 ml cultures by centrifugation at 10 000 g for 10 minutes at 4°C. Culture supernatants and cell pellets were stored at -80°C. In subsequent optimisation experiments, starter cultures of transformants were sub-cultured in LB broth and grown at 37°C on a shaker to an OD₆₀₀ of 0.3-0.4. Then the cultures were pre-equilibrated at 20°C for up to 1 hour on a shaker (180 rpm) prior to induction at an OD₆₀₀ of 0.5-0.6. After induction with 0.1 mM IPTG at 25°C and overnight incubation, cells were harvested by centrifugation. Furthermore, for codon supplementation, *E. coli* Rosetta 2(DE3) cells were used for protein expression under standard conditions with growth at 37°C and induction at 20°C. Cells were harvested and processed as usual for downstream processing.

2.5.2 Purification of recombinant proteins

All purifications were carried out at 4°C on the AKTA PURE explorer machine (GE Healthcare, United Kingdom). Before purification cell pellets were thawed on ice, resuspended in 20 ml PBS and disrupted by sonication with a micro tip (Bandelin Sonoplus) on ice for 3 minutes in 30 secs bursts at 98% power. After centrifugation (10 000 g, 10 minutes, 4°C) to isolate the cell debris, the supernatant was filtered with a 0.22 µm filter. In subsequent experiments, cell pellets were resuspended in PBS and disrupted using a French press (Thermo IEC) set at 1500 psi for 2 rounds. Supernatants obtained from centrifugation of cell lysates represent the intracellular soluble fraction and the cell debris represent the insoluble fraction. Clarified lysates were diluted in appropriate buffers and loaded on corresponding columns for purification. For GST-tagged proteins the clarified lysate was diluted in PBS, pH 7.4 and loaded on a 5 ml glutathione Sepharose affinity column (GE Healthcare, United Kingdom) pre-equilibrated with the same buffer. After protein binding at a flow rate (FR) of 2 ml/min, the column was washed with 5 column volumes of binding buffer (PBS, pH 7.4) at 5 ml/min to remove impurities. The bound protein was eluted at a flow rate of 0.5 ml/min with elution buffer (50 mM Tris-HCL, 10 mM reduced glutathione, pH 8) and 1 ml fractions were collected.

Meanwhile for His-tagged proteins clarified lysates were prepared with binding buffer (20 mM sodium phosphate, 0.25 M NaCl, 5 mM imidazole, pH 8) for affinity purification on HisTrap HP columns pre-equilibrated with binding buffer. After protein binding at a FR of 1 ml/min, impurities were eliminated by washing at 2 ml/min with the same buffer containing 20-40 mM Imidazole. The target protein was eluted with a linear gradient of 40-250 mM Imidazole in binding buffer and 1 ml fractions were collected at a flow rate of 1 ml/min. After SDS-PAGE analysis on purified protein fractions, insoluble fractions, induced and non-induced cell lysates, fractions containing the target protein band were desalted or dialysed. Fractions were pooled and desalted on a HiTrap Desalting column pre-equilibrated with 50 mM Tris, pH 8. Eluents were collected at a flow rate of 1 ml/min in 1 ml fractions and in the same buffer. Protein assay and transglutaminase activity assays were performed on eluted protein fractions as described previously.

2.5.3 Dialysis of purified proteins

Purified protein solutions were introduced in dialysis bags, immersed in 1 L buffer and allowed on a magnetic stirrer at 4°C. Altogether two to three buffer replacements were made during a 15 hour period. Samples were collected from dialysis bags and tested for proteins levels and enzyme activity. Alternatively, samples containing proteins of interest were pooled and desalted on a desalting column equilibrated with 50 mM Tris, pH 8 on the AKTA PURE system. Samples collected following desalting were tested for enzyme activity and protein content.

2.6 PROTEIN METHODS

2.6.1 Protein concentration assay

The protein concentration of purified proteins was determined using the Bradford DC assay kit (Bradford, 1976), with known concentrations of BSA as protein standards. The Bradford kit used is a detergent compatible assay that involves reduction of Folin-Ciocalteu reagent by protein samples priorly treated with alkaline copper tartrate. The colour produced following this reduction reaction can be measured at an absorbance maximum of 750 nm. A volume of 5 µl protein standards and samples was loaded in triplicate onto 96 well plates. Then 25 µl of alkaline copper tartrate solution and 200 µl of folin reagent were added respectively to the samples and standards. After incubation for 15 minutes at room temperature, the absorbance was measured on a plate reader (Biotek EL808) at 690 nm. Protein concentration in samples was calculated using the equation of the line derived from the BSA standard curve.

2.6.2 Protein precipitation with trichloroacetic acid (TCA)

Culture supernatants were diluted with TCA to a final concentration of 10% (v/v) to precipitate the supernatant proteins. Samples were incubated for 20 minutes on ice followed by centrifugation (5000g, 4°C, 10 min). Pellets were washed twice with ice-cold acetone by repeated centrifugation. Finally, the protein pellets were dried at 50°C for 5 min to eliminate residual acetone prior to sample preparation for further analysis.

2.6.3 SDS-PAGE gel preparation and protein analysis

The SDS polyacrylamide gels for protein separation were made of 10% (v/v) separating gel and 4% (v/v) stacking gel. The SDS gels were 0.75 mm thick and contained a 29:1 acrylamide/bisacrylamide ratio based on conventional protocols (recipe in Table 2.2). The SDS-PAGE gels were made using the Bio-rad Mini gel system. Firstly, the separating gel was poured in the assembled glass plates and isopropanol was layered over the gel to flatten the gel surface. The gel was allowed to set at room temperature for 45-60 minutes. Then isopropanol was washed off the gel surface with distilled water (dH₂O) and excess dH₂O was absorbed with blotting paper. Secondly, the stacking gel was prepared following the recipe in table 2.2 and after pouring into the glass spacers, a 10-well gel comb was introduced into the gel to create wells for sample loading. Finally, after the stacking gel was set, the combs were taken off and the wells flushed with running buffer prior to sample loading.

Samples from protein purifications and cell lysates were prepared for SDS-PAGE analysis as described by Laemmli (Laemmli, 1970). Samples containing 20-50 µg protein were prepared beforehand by mixing in equal proportions with 2x Laemmli buffer. After boiling for 10 minutes, 20 µl samples were loaded into the wells of acrylamide gels immersed in 1X running buffer, pH 8.3. Electrophoresis was conducted using a Bio-Rad Mini gel system in denaturing conditions at 125V. Electrophoresis was stopped when the migration front reached the base of the gel and gels were rinsed with dH₂O prior to staining. The recipe for preparation of acrylamide gels for SDS-PAGE analysis is detailed in Table 2.2.

Table 2.2. Recipe for SDS separating and stacking gels.

Separating gel	10% (v/v) acrylamide gel (2 gels)
30% (v/v) acrylamide	2.5 ml
4x Tris-Cl/ SDS pH 8.8	2.5 ml
Distilled water	5 ml
10% (w/v) Ammonium Persulfate	33.3 μ l
TEMED	6.7 μ l

Stacking gel	10% (v/v) acrylamide gel (2 gels)
30% (v/v) acrylamide	0.65 ml
4x Tris-Cl/ SDS pH 6.8	1.25 ml
Distilled water	3.05 ml
10% (w/v) Ammonium Persulfate	25 μ l
TEMED	5 μ l

2.6.4 Coomassie staining of acrylamide gels

After sample separation by SDS-PAGE, protein bands were visualised by staining. Protein gels were stained overnight with Coomassie brilliant blue R250 (CBB) prepared following the manufacturers' instructions (40% (v/v) methanol, 50% (v/v) water, 10% (v/v) acetic acid). Then the gel was destained with destaining solution (20% (v/v) methanol, 70% (v/v) water, 10% (v/v) acetic acid) for a few hours on a shaker until the gel background was clear. Images were recorded with the Syngene software on a Genebox. In some instances, Instant blue stain was used which is colloidal CBB that does not require a destaining step. Image J software was used to calculate purity and estimate molecular weights of purified proteins based on the SDS-PAGE gels.

2.6.5 Western blot

After protein separation by SDS-PAGE, the proteins were transferred electrophoretically onto nitrocellulose (NC) membranes at 300 A for 70 min using a Bio-rad gel transfer system containing transfer buffer. After transfer, NC membranes were blocked with 5% (w/v) BSA at room temperature for 1 hour. Membranes were incubated with primary anti-His antibody (1:1000, Abcam) in 5% (w/v) BSA, followed by three successive 10 mins washes in Tris buffered saline-Tween (TBST). This was ensued by incubation with HRP-conjugated anti-

mouse IgG (1/3000, R&D) in BSA for 1 hour. Membranes were washed three times with TBS-Tween and rinsed once with TBS. The His-tagged protein bands were detected by chemiluminescence using the ThermoFisher Scientific SuperSignal West Pico PLUS Chemiluminescent substrate. Images were recorded on a Western blot imager. The antibodies used are listed on the table in Appendix 3.

Transfer buffer: 190 mM glycine, 25 mM Tris base, 20% (v/v) methanol

Running buffer: 190 mM glycine, 25 mM Tris base, 0.1% (w/v) SDS

TBS: 0.2 M Tris, 1.5 M NaCl

TBS-Tween: 0.1% (v/v) Tween, 1X TBS

2.6.6 Transglutaminase activity assay

Transglutaminase activity assay as described by Slaughter *et al.*, 1992 was measured by biotin cadaverine (BTC) incorporation into casein. Briefly microtiter plates were coated with 10 mg/ml N,N-Dimethylcasein in 50 mM Tris-Cl, pH 8 and incubated overnight at 4°C. After washing the plate three times with TBS-Tween (pH 7.4), samples (100 µl) were added to the wells in triplicate. The reaction mixture comprised 0.1 mM BTC, 1.1 mM DTT, 50 mM Tris-HCL pH 8, eluted enzyme fractions, cell/spore lysates, flow through, washes, and 1% (w/v) MTG control. Following incubation at 37°C for 1 hour and washing with TBS-Tween, the wells were blocked with blocking buffer (3% (w/v) BSA in 50 mM Tris, pH 7.4) for 30 minutes at 37°C. Biotin-cadaverine incorporation into casein was detected by adding 100 µl ExtrAvidin-Peroxidase (1:2500 dilution with BSA) into wells and incubating for 1 hour at 37°C. After three washes with TBS-Tween and one wash with TBS, the reaction was developed by incubating with 0.4 mg/ml OPD for 30 minutes at room temperature. Finally, colour development with OPD was stopped with 3 M HCL and absorbance was measured on a plate reader (Biotek) at 450 nm. Recombinant microbial transglutaminase (MTG) from *Streptomyces mobaraense* was used as positive control and 50 mM Tris, pH 7.4 as negative control.

2.6.7 Protease activity assay

The protease activity in purified proteins was assessed using two different substrates, a casein substrate, and a collagen-based substrate.

2.6.7.1 Pierce fluorescent protease assay

Proteolytic activity with the Pierce fluorescent protease assay kit was measured using a fluorometer following digestion of FTC-casein by recombinant or purified proteins and trypsin standards. The Pierce fluorescent protease assay kit uses fluorescein labelled casein as substrate for determination of proteolytic activity in protein samples. Fluorescein labelled casein is digested by proteases into smaller fragments producing a change in fluorescence which can be detected by a fluorometer (ThermoFisher scientific).

TPCK treated trypsin eliminates chymotrypsin activity and cleaves proteins at the C-terminus of arginine or lysine. Eluted protein fractions and TPCK trypsin were serially diluted in TBS, pH 7.2 to obtain standards from 0-10 µg/ml and protein fractions from 1-20 µg/ml. Afterwards 100 µl of protein samples and trypsin standards were loaded in triplicate onto 96-well black plates. Then 100 µl (0.01 mg/ml) of working reagent (FTC-casein diluted in TBS) was added to the wells and incubated for 60 minutes at room temperature, during which fluorescence was measured every 5 minutes. The fluorescence emitted was measured by a fluorometer (FLUOstar Omega, BMG Labtech) with excitation/emission filter set to 485/520 nm.

2.6.7.2 Hide Powder Azure assay

Hide powder azure was used as substrate to assess the protease activity of recombinant proteins using trypsin as positive control. Hide powder azure is a collagen substrate stained with Coomassie brilliant blue that produces soluble dye-bound fragments upon digestion by proteases, generating a colour change which can be quantified by absorbance measurements. Recombinant protein samples (10 µg/ml) and trypsin (0-10 µg/ml) were prepared with Hide powder azure (5 mg/ml) in 50 mM Tris, 1 mM CaCl₂, pH 8. After incubating on a shaker (100 rpm) for 0.5-48 hours at 37°C, samples were centrifuged for 10 minutes at 14000 g and the absorbance of supernatants was measured at 595 nm.

2.6.8 Transglutaminase-catalysed protein crosslinking assay

Several candidate substrates were selected for reaction with purified proteins to evaluate evidence of crosslink formation indicating Tgase activity in eluents. BSA (0.35 mg/ml) was

dissolved in 50 mM Tris, pH 8 and 12 µg of recombinant protein was diluted 1:1 with BSA. Simultaneously trypsin (6 µg/ml) and purified MTG (100 µg) from *S. mobaraensis* were prepared with BSA. All samples were incubated at 37°C over a time course (1-5 hours). Fractions were collected every hour and immediately diluted with 2x Laemmli buffer. Also, casein (10 mg/ml) from bovine milk was dissolved in 50 mM Tris, pH 8 and 12 µg of recombinant protein was diluted 1:1 with casein. Simultaneously trypsin (6 µg/ml) and purified MTG (100 µg) were added to casein and samples were incubated at 37°C for 1 hour. SDS-PAGE of reaction products was performed on 10% (v/v) acrylamide gels as described previously. Samples were prepared with 2X Laemmli buffer, boiled and separated by SDS-PAGE. Afterwards gels were stained overnight with Coomassie brilliant blue and visualised on a Genebox after de-staining.

2.6.9 Cleavage and removal of GST tags from recombinant proteins

Recombinant proteins were digested with Factor Xa to cleave the GST tags prior to performing activity assays. A pilot study of GST tag cleavage was carried out by adding Factor Xa to recombinant proteins in buffers with and without calcium, followed by overnight incubation at room temperature or at 37°C for 2 hours. Samples were collected at 2, 5 and 17 hours of incubation at room temperature and analysed by SDS-PAGE. The success of tag cleavage was verified after recovery of GST and cleaved proteins as shown by SDS-PAGE gels. Following successful pilot studies, recombinant proteins were digested with Factor Xa and purified through GSTrap/heparin columns for GST plus Factor Xa removal. Recombinant proteins (1 mg/ml) were treated with Factor Xa (200 µg/ml) in 50 mM Tris, 150 mM NaCl, 1 mM CaCl₂, pH 7.5 for 4 hours at room temperature and purified on GSTrap/heparin columns pre-equilibrated with the same buffer. The cleaved proteins were collected in 1 ml fractions in the same buffer at a FR of 1 ml/min leaving behind the tag bound to the column. The recombinant proteins with cleaved GST tags were then tested for transglutaminase activity.

2.6.10 Digestion of recombinant proteins with Trypsin

Recombinant proteins were digested with trypsin in an attempt to activate the protein through the removal of a probable pro-peptide. Trypsin digestions were performed in a 50 µl reaction containing trypsin (0-10 µg/ml) and 24 µg/ml of purified recombinant proteins. The reactions were incubated for 30 minutes at 37°C. Samples were analysed by SDS-PAGE and visualised after Coomassie blue staining. The digested samples were tested for transglutaminase activity.

2.7 Optimization of Recombinant Transglutaminase Protein Expression and Purification

2.7.1 Removal of signal peptides by PCR amplification of Recombinant plasmids

The first 29 amino acids corresponding to the putative signal peptides in CD-1062 and CD-1083 Tgase-like genes in *C. difficile* 630 were targeted for elimination. To cleave the signal peptides, a PCR reaction was performed under standard conditions described previously with recombinant pGEX-3X plasmids as template and specific primers using the KOD DNA polymerase kit. Primers were designed such that the whole DNA sequence was amplified excluding the signal peptide regions. The success of the PCR reactions was determined by electrophoresis on a 1% (w/v) agarose gel. The PCR product was purified by agarose gel electrophoresis prior to transformation by the heat shock procedure and plasmid DNA isolation as described beforehand. DNA sequencing was performed on isolated plasmids to verify the success of signal peptide cleavage.

2.8 Purification and Characterisation of Transglutaminase-like proteins from *C. difficile* 630 spores

2.8.1 *C. difficile* growth conditions and Spore production

Clostridioides difficile 630 was resuscitated from glycerol stocks by growing on Wilkens Chalgren agar supplemented with 0.1% (w/v) sodium taurocholate for 48 hours at 37°C under anaerobic conditions. One *C. difficile* colony isolated from the solid medium was inoculated into 10 ml BHI broth (supplemented with 0.5% (w/v) yeast, 0.1% (w/v) cysteine), and incubated overnight at 37°C in anaerobic conditions to be used as starter culture. Sporulation was induced following the nutrient starvation procedure in BHIS broth and spores were grown on agar. To make sure the culture was spore-free prior to induction of sporulation, overnight cultures of *C. difficile* 630 were sub-cultured as follows. One percent (v/v) of starter culture was inoculated into 10 ml of BHI broth in a 30 ml universal tube and incubated at 37°C for about 4-5 hours to reach log phase (OD₆₀₀: 0.2-0.5). For spore production, *C. difficile* 630 was cultured on petri dishes containing BHI agar supplemented with 0.5% (w/v) yeast and 0.1% (w/v) L-cysteine. The culture fluid at log phase was spread onto several pre-reduced agar plates and further incubated for 120 hours at 37°C under anaerobic conditions. After 5 days, spores were harvested in PBS and incubated overnight at 4°C to promote the release of residual spores prior to washing. Briefly spores were purified by repeated washing and

centrifugation (10 000 g, 10 min, 4°C) in PBS-Tween (0.1% (v/v)), PBS-Tween (0.01% (v/v)) and final few washes in sterile distilled water. Afterwards the purity of washed spores was assessed by staining following the Schaefer-Fulton method, and purified spore suspensions were stored at 4°C.

To measure spore formation, samples were collected at specific times during 5 days of sporulation in BHIS broth or agar, and vegetative cells were killed by heating for 30 min at 60°C. The principle of this procedure being that solely spores would resist heat treatment implying that the colonies counted on the plate represent the total spores in the population. The samples were then diluted in PBS and spread on BHIS containing 0.1% (w/v) taurocholate to trigger germination and promote spore recovery on agar. After 24 hour incubation at 37°C under anaerobic conditions, colony forming units (CFU) were counted.

2.8.2 Transglutaminase production during sporulation in *C. difficile*

Studies in *B. subtilis* indicate that Tgases were expressed endogenously during sporulation and purified from crude extracts by a two-step purification (Suzuki *et al.*, 2000). This was a good starting point that enabled us to devise a strategy for *C. difficile* 630 enzyme purification from crude Tgase extracts harvested from sporulating cells. Following induction of sporulation by nutrient starvation in BHIS broth, Tgase activity was monitored over 5 days of sporulation in BHIS broth and on agar. At specific time points sporulating cells were collected from agar plates by flushing with distilled water. The resulting pellets obtained by centrifugation of samples from agar and broth cultures at 10 000 g (Beckman Coulter) for 10 min at 4°C was processed for analysis. To establish the time for peak Tgase activity, sporulating cells harvested at specific time points during sporulation were solubilised in 50 mM Tris buffers and tested for Tgase activity. The flow chart in Figure 2.2 summarises the steps required for transglutaminase production from spores up to purification.

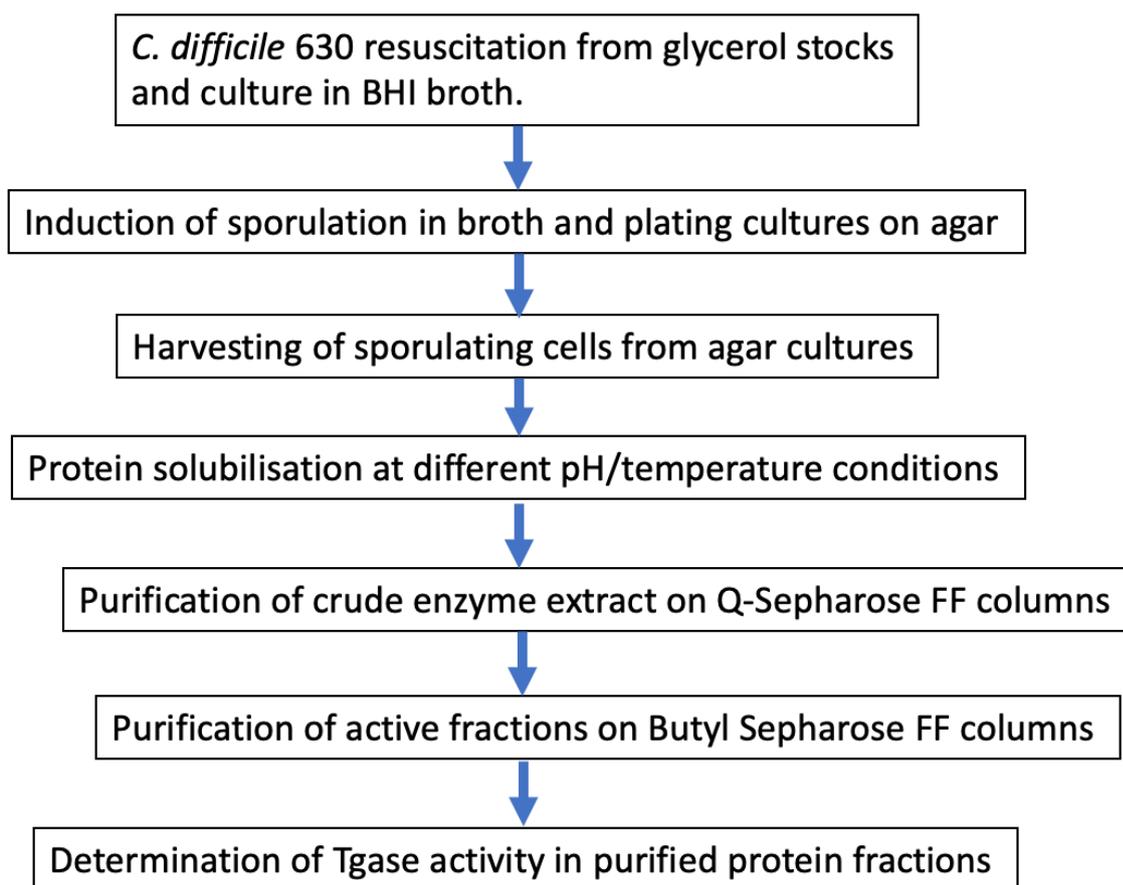


Figure 2.2. Flow diagram for Transglutaminase production. *C. difficile* 630 was resuscitated from glycerol stocks on Wilken Chalgren agar containing 0.1% (w/v) taurocholate and overnight cultures were prepared in BHIS broth. Sporulation was induced from starter cultures in BHIS broth and spores were grown on BHIS agar. After 24 hours, sporulating cells were harvested from agar plates and solubilized in 50 mM Tris buffer, pH 8. The Tgase activity in cell extracts was purified by ion-exchange chromatography and hydrophobic interaction chromatography prior to further characterization studies.

2.8.3 Solubilisation of transglutaminase-like proteins from *C. difficile* spores

To determine the optimum conditions required for efficient solubilisation of Tgase from sporulating cells or spores, different pH and temperature were tested. Spores were incubated at different temperatures ranging from 0-50°C in 50 mM Tris, pH 7.5 or 8 for 2 hours. Equally spores were incubated in 50 mM Tris, with pH ranging from 5-10 for 2 hours at 0°C or 25°C. After protein solubilisation, the cell suspensions were centrifuged and Tgase activity was measured in the supernatants using the standard BTC incorporation assay.

2.8.4 Purification of crude enzyme extracts from *C. difficile* 630 spores

All purifications were carried out at 4°C on the AKTA PURE explorer machine (GE Healthcare, United Kingdom).

2.8.4.1 Ion-exchange chromatography

Sporulating cells were solubilised in 50 mM Tris, pH 8 for 1 hour at room temperature with stirring, and crude extracts were obtained after centrifugation. The supernatants obtained from crude enzyme extractions were put on 100 ml Q-Sepharose Fast Flow columns previously equilibrated with 50 mM Tris, pH 8. During this process the protein was adsorbed onto the column at a FR of 3 ml/min. Impurities were removed by washing extensively at similar flow rates with 1.5 column volume (CV) of binding buffer till baseline absorbance levels. The bound protein was eluted from the column with a linear NaCl gradient from 0 to 2 M in Tris buffer (50 mM, pH 8). Eluents were collected at a flow rate of 3 ml/min in 6 ml fractions in a total volume of 180 ml. After transglutaminase activity assay and SDS-PAGE analysis on the eluted fractions, active fractions were pooled together for further purification.

2.8.4.2 Hydrophobic interaction chromatography

The partially purified Tgase active fractions obtained from Q-Sepharose columns were adjusted to 2 M NaCl in 50 mM Tris (pH 8), with a 5 M NaCl stock. This enzyme solution was applied onto a 5 ml Butyl Sepharose Fast Flow column equilibrated with 2 M NaCl in 50 mM Tris (pH 8). Transglutaminase in active fractions was bound at a flow rate of 1.2 ml/min. The column was extensively washed with 5 CV of NaCl buffer at the same flow rates to remove unabsorbed proteins. The bound protein was eluted with a decreasing linear (100 to 0% (v/v)) gradient starting from 1 M NaCl in 50 mM Tris buffer. Eluents were collected in 1 ml fractions at a flow rate of 1.2 ml/min in a total volume of 60 ml. The eluted fractions were assayed for transglutaminase activity, protein concentration, and SDS-PAGE analysis as previously described. The active Tgase fractions obtained from the purification were pooled for characterization experiments.

2.8.5 Characterization of partially purified *C. difficile* 630 spore enzymes

2.8.5.1 Effect of temperature on activity of purified enzyme

The effect of temperature on activity of purified spore enzymes was determined by testing activity of enzyme solutions following BTC incorporation at different temperatures (0°C, 22°C, 37°C, 45°C). Substrates were added to purified proteins and loaded onto 4 casein-coated plates. Each plate was labelled accordingly and incubated at a specific temperature for 1 hour. After washing with TBS-Tween, plates were blocked with 3% (w/v) BSA at 37°C and all succeeding steps were carried out at 37°C. Transglutaminase activity was then measured under normal assay conditions as described previously. The relative activities were calculated by taking the maximum activity from any given plate initially incubated at a particular temperature as 100%.

2.8.5.2 Thermostability of purified enzyme activity

The thermostability of purified spore protein solutions was investigated by preincubation of samples at different temperatures (0°C, 22°C, 37°C, 45°C) for 15 min followed by immediate cooling on ice. Transglutaminase activity was measured under normal assay conditions after addition of substrates. The relative activities were calculated by taking the maximum activity at any given temperature as 100%.

2.8.5.3 Effect of metal ions and inhibitors on enzyme activity

The effect of inhibitors and metal ions on enzyme activity was assessed by adding different concentration of inhibitors to enzyme solutions and incubating for 30 min on ice prior to Tgase activity assay. The residual activity was determined under established assay conditions at 37°C and calculated as percentage activity relative to activity in the absence of inhibitors taken as 100%. The chemicals tested for inhibition of activity were 2 mM N-ethylmaleimide (NEM), 2 mM cysteamine dihydrochloride, 20 mM ammonium sulphate, 10 mM EDTA, 1 mM calcium chloride, 5 mM PMSF, 10-100 mM DTT. These inhibitor concentrations represent the final concentrations included in the assay.

2.8.5.4 Mass spectrometry analysis of partially purified spore enzymes

Following purification of spore enzyme extracts by ion-exchange and hydrophobic interaction chromatography, eluted proteins were analysed by SDS-PAGE. For protein identification,

purified protein bands were excised from acrylamide gels and sent for mass spectrometry analysis (University of Birmingham mass spectrometry facility).

2.9 Statistical Analysis

Data are presented as mean \pm SEM of three or more repeats and figure sets shown are representative of experiments performed in triplicate. Data were analysed with Microsoft Excel 2019 and Graph Pad Prism 9.1.0. Results were analysed using Student's t-tests and ANOVA tests. All analysis were considered statistically significant only when P values were < 0.05.

CHAPTER 3. Gene Expression during sporulation in *C. difficile* 630 and Bioinformatics analysis of putative transglutaminase-like genes

3.1 Introduction

Up to date no transglutaminase from *C. difficile* 630 has been identified and characterized. Gene transcript analysis have been investigated during sporulation in *C. difficile*, but transglutaminase (Tgase) gene expression has not been studied (Dembek *et al.*, 2013). Equally transglutaminase activity screening during sporulation has not been carried out in *C. difficile* strains as observed in *Bacillus* strains. Transglutaminase activity has been detected in spore formers such as *B. subtilis*, and plays important roles in spore resistance through protein crosslinking at spore surfaces (Fernandes *et al.*, 2015). In *C. difficile* 630, Spo0A is the master regulator for transcription during sporulation (Underwood *et al.*, 2009). Sporulation consists of a sequence of events that ultimately leads to production of metabolically inactive cells from vegetative growth. Spore formation usually occurs during the stationary phase when the cells are stressed due to limiting nutrient availability (Pettit *et al.*, 2014). Sporulation serves as a survival mechanism during harsh environmental contexts and bacteria can remain in this state for extended periods. Bacterial spores are formed by organisms in the Firmicutes phylum, including the popular *Clostridioides* and *Bacillus* genus (Pettit *et al.*, 2014). *C. difficile* is responsible for nosocomial antibiotic associated diarrhoeas and currently a pathogen of public health concern (Worthington and Hilton, 2016). The control of *C. difficile* infections is limited due to the inherent resistance of *C. difficile* spores, prolonged survival in the environment, high recurrence, and transmission rates. Thus the spore formation ability of *C. difficile* enables it to thrive as a germ and cause disease (Lawley *et al.*, 2009). Recently the emergence of hypervirulent strains has been associated with increased incidence of grave disease, higher recurrence, and mortality rates (Pereira *et al.*, 2013). A deeper understanding of *C. difficile* spore biology could serve to establish disinfection regimens, diagnostics and develop specific treatments.

3.2 Objectives

The objectives of this chapter will be:

- Bioinformatics analysis of putative transglutaminase genes in *C. difficile* 630.
- To determine the presence of transglutaminase proteins in *C. difficile* 630 by measuring transglutaminase activity during sporulation.

- To determine the expression of transglutaminase-like genes during sporulation via qRT-PCR and computing the correlation between transglutaminase expression and sporulation.

3.3 RESULTS

The aim of this section is to identify putative Tgases by bioinformatic analysis, detect Tgase activity in *C. difficile* 630, examine the profile of Tgase mRNA transcripts during sporulation to understand its expression, and determine its correlation with sporulation.

3.3.1 Bioinformatics

The aim of this section was to identify putative Tgase genes within the *C. difficile* 630 genome using key word searches and determination of homology with known microbial or mammalian transglutaminases. Using the NCBI database to search for putative transglutaminase-like genes in *C. difficile* 630 by keyword searches, three genes encoding Tgases were identified within the complete genome of *C. difficile* 630 strain. The putative Tgase genes identified had distinct locus tags as shown in Table 3.1. On NCBI the genes were named as 'uncharacterized transglutaminase-like' genes, except the shortest gene named 'transglutaminase-like'. The three identified Tgase-like proteins contain 239, 353 and 360 amino acids respectively with corresponding predicted molecular weights of 27.9 kDa, 38.8 kDa and 39.3 kDa. For simplicity the CD630_25070, CD630_30430 and CD630_11560 Tgase-like genes in *C. difficile* 630 are referred to as CD-720, CD-1062 and CD-1083 proteins respectively. The deduced molecular weights of *C. difficile* 630 Tgases are comparable to the molecular weights of characterized microbial transglutaminases in other spore formers (Ando *et al.*, 1989; Cui *et al.*, 2007). BLAST analysis on NCBI was performed using the amino acid sequences of transglutaminase-like genes as query to identify other Tgase-like genes within *C. difficile* 630, but also homologous sequences in other strains, or bacteria and mammals. From the BLAST analysis using CD-1062 entire DNA coding sequence as query, the CD-1083 Tgase gene was identified with 73% similarity. Also, homology searches using conserved core regions and entire protein sequences of CD-1062 and CD-1083 as query detected homologous regions in CD 630 genome with 28% and 39% similarity respectively. Surprisingly, a homolog was found in *E. coli* with 100% identity when the entire protein sequence of CD-1083 was used as query. No sequence similarity was identified in organisms containing characterised Tgases using *C. difficile* 630 Tgase-like coding sequences as query. A multiple sequence alignment (MLSA) of the Tgase core domains was done to visualise the similarity amongst *C. difficile* 630 Tgase-like genes as reported by BLAST analysis.

```

Clostridium_difficile_630_transglutaminase_720bp      ----ELDDFHKIKSIYEFVQND-ILFGYNASDMLSATQVLNDGMGQCNTKATLLMALLRA      55
Clostridium_difficile_630_transglutaminase_1083bp     KITKNMTTDAKIRVAYNHKNNTTYVNWNAEAGANTAYTLVTKKGACSGMCRSLKALCDA      60
Clostridium_difficile_630_transglutaminase_1062bp     KITKNMTTDAKILVAYNHKNNTTYVDWNAVEGANTAYTLVTKKGACSGMARSMLKALCDA      60
:: ** *:::*: ..:** : .:: .* * * . : ** *

Clostridium_difficile_630_transglutaminase_720bp     VNIPCR LHAFDVTKDFQRGATSKLISLLAPKYILHTWVEVFYQIRWI-ALE      105
Clostridium_difficile_630_transglutaminase_1083bp     MGIECYVHSTT-----NDHQWNLIRFEDGRLYHVD      91
Clostridium_difficile_630_transglutaminase_1062bp     MGIESYYVHSTS-----NDHQWNLIRFGDGRLYHVD      91
:.* . * * : : * : : :

```

Figure 3.1. Multiple alignment of Tgase-core domains in *C. difficile* 630. The figure shows a multiple alignment of the Tgase conserved fold in *C. difficile* 630 putative Tgase genes. The respective Tgase core domains of Tgase-like genes in *C. difficile* 630 were subject to multiple alignment with Clustal Omega MLSA tool. The conserved active site region is highlighted in red (the single-letter code represents amino acid residues, asterisks below the alignment indicate single fully conserved regions, dots indicate conserved regions between groups with similar and less similar properties).

The multiple alignment of Tgase-core domains shows a high degree of similarity between two Tgase-like genes in *C. difficile* 630, CD-1062 and CD-1083 (Figure 3.1). The core domains of CD-1062 and CD-1083 genes have 91% identity. Many conserved regions within the three genes have been identified, some of which correspond to potential catalytic sites (highlighted in red). Notably from the graphics results on NCBI, Tgases from *C. difficile* 630 possess a single domain that consists of a conserved catalytic core of cysteine, histidine, aspartate, and thus are part of the Tgase-like superfamily. To verify the conservation of Tgase core domains in putative *C. difficile* 630 Tgases, sequence alignments with known Tgase-like superfamily members was performed as shown in Figure 3.2 below.

Clostridium_difficile_720	LNDGMGQCNTKATLLMALLRAVNIPCR---LH---AFDVTKDFQRGA---TSKLISLLAP	51
Clostridium_difficile_1062	LVTKKGACSGMARSMKALCDAMGIESY---Y-----VH-----S	31
Clostridium_difficile_1083	LVTKKGACSGMCRSLKALCDAMGIECY---Y-----VH-----S	31
Papain_cysteine_protease	---CGSCWAFSLNFAIESRWYITNSK---FV---KLPEQ--F--EA-----VDLDH	38
Bordetella_bronchiseptica_WbME	LDGGFGICGNHQYLFLELMHRLGLEAR---SV---GF----WY-----	33
P_aeruginosa_PA2873	FDTRSGFCAHYAGAMAFVLRAGIPAR---VV-----A-----VH	32
Tissue-Tgase_TG2	QQVKYGQCVVFAAVACTVLRCLGIPTR---VV---TN-----S-----NKSEM	37
Human_Factor_XIIIA	NPVRYGQCVVFAAGVFNTFLRCLGIPAR---IV---TN-----YF-----DS	35
Fish_Tgase_FTG	--VKYGQCVVFAAVACTVLRCLGIPTR---PI---TN-----FA-----DS	33
Bacillus_subtilis_TgI	-----CAT-----AIVIIY--Y---LALIDTIGED--KFNASFDRIIL--ILLGED	37
Streptomyces_mobaraensis_Tgase_MTG	-----CVG-----VTWVNSGQYPTNRLAFASFDED--RFKNEL-----DA	33
	*	
Clostridium_difficile_720	KYILHTWVE-----D---	61
Clostridium_difficile_1062	TSNDHQWNLIRFDGRL-----YHVD---	52
Clostridium_difficile_1083	TTNDHQWNLIRFDGRL-----YHVD---	52
Papain_cysteine_protease	AINVVGWVDTLDDGTEAQHWILR-----NSW-----	65
Bordetella_bronchiseptica_WbME	SRASHA-----AEVLVD-----KKWRYVDITW	56
P_aeruginosa_PA2873	QFDAHAW-----VEYWQPE-----QGWSVDPTY	56
Tissue-Tgase_TG2	IWNFHCW-----VESW---PDLQPGYEGWQAIIDPT-	64
Human_Factor_XIIIA	VWNYHCW-----NEAWMTRPDLVPGFGGWQAVDSTP	66
Fish_Tgase_FTG	SWNFHCW-----VESWMSREDLPEGNDGWQLDPTP	64
Bacillus_subtilis_TgI	KYFAHGLGILNGK--Q---IIDKLNFRK-KGALQSAY--	69
Streptomyces_mobaraensis_Tgase_MTG	FRPAPGTGLVDMRDR---NIPRSPTSPG-EGFVNFYDYGW	69

Figure 3.2. Multiple sequence alignment of conserved Tgase regions in mammalian and microbial organisms. The conserved Tgase core domain in selected microbial and mammalian organisms were subject to multiple alignment with Clustal Omega MLSA tool. Single fully conserved regions are denoted by an asterisk (*) as shown beneath the cysteine residue at the bottom of the alignment.

Comparative studies of *C. difficile* 630 Tgase core domains with other functional Tgases identified homologous regions in other bacteria and animal Tgases. The catalytic cysteine residue in the active site domains of all Tgases appear to be conserved based on the MLSA in Figure 3.2. The CD-1062 and CD-1083 Tgases in *C. difficile* 630 have an identical active site domain. The Tgase core domain of CD-720 contains 105 amino acids, meanwhile the core domains of CD-1062 and CD-1083 contain 91 amino acids. Notably, in CD-1062 and CD-1083 a serine residue is found next to the putative catalytic cysteine. This observation is not shared with other Tgases analysed. Figure 3.3 shows a phylogenetic tree representing known Tgase superfamily members with *C. difficile* 630 Tgases and their evolutionary relatedness.

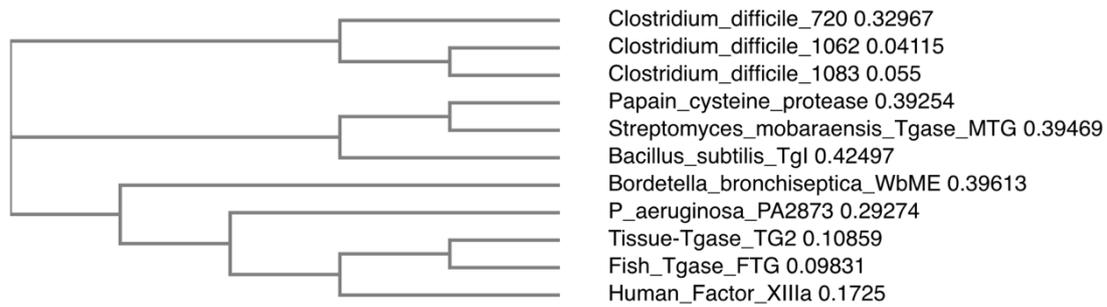


Figure 3.3. Phylogenetic tree analysis of Tgase-like genes in mammalian and microbial organisms. The Tgase core domains in *C. difficile* 630 were compared to that of Tgase superfamily members in microbial and mammalian organisms. The conserved Tgase core sequences were downloaded from NCBI, and the phylogenetic tree was generated with Clustal Omega MLSA tool using the neighbour joining method. The organisms used in the analysis included: *Bacillus subtilis* Tgl, human Factor XIIIa, *Bordetella bronchiseptica* WbME, *Streptomyces mobaraense* MTG, tissue Tgase (TG2), red sea bream liver Tgase (FTG), *Pseudomonas aeruginosa* PA2873, CD-720, CD-1062, and CD-1083 Tgases from *C. difficile* 630, and papain protease. The numbers next to the names are the branch lengths that indicate the evolutionary gap amongst the sequences analysed.

The Tgases used to construct the phylogenetic tree belong to the Tgase-like superfamily of enzymes. The sequence clustering in Figure 3.3 shows that CD-1083 and CD-1062 are closely related as confirmed by BLAST analysis and both share similarity with CD-720. CD-1083 and CD-1062 are orthologues, and probably the product of a gene duplication event. The phylogenetic tree suggests that *C. difficile* Tgase-like genes are different from other Tgases. *C. difficile* Tgase-like genes could be novel Tgases, and perhaps possess a completely different catalytic mechanism. From the analysis MTG is closely related to papain proteases, both of which are related to Tgl. This is consistent with the fact that they contain the NlpC/P60 core reminiscent of ancient cysteine protease fold. More so MTG and cysteine proteases have similar arrangements around the catalytic site (Kashiwagi *et al.*, 2002). Protein relatives of Tgl are found in other spore formers which concurs with the alignment in Figure 3.3 showing Tgl and MTG are related. MTG and Tgl mediate catalysis via a cysteine protease-like mechanism that involves a catalytic dyad (Fernandes *et al.*, 2015). Most of the Tgase genes analysed have descended from an ancestral gene and thus related by convergent evolution. Despite the lack of sequence homology, they have evolved to perform a similar function i.e., protein crosslinking. After bioinformatics analysis of *C. difficile* 630 Tgase-like genes and homology detection, other features specific to *C. difficile* 630 Tgases were investigated on several prediction servers.

3.3.1.1 Transglutaminase-like genes in *C. difficile* 630 contain putative signal peptides

The presence of signal peptides within Tgase genes in *C. difficile* 630 was determined by signal peptide predictions on the SignalP 4.1 server as illustrated in figure 3.4.

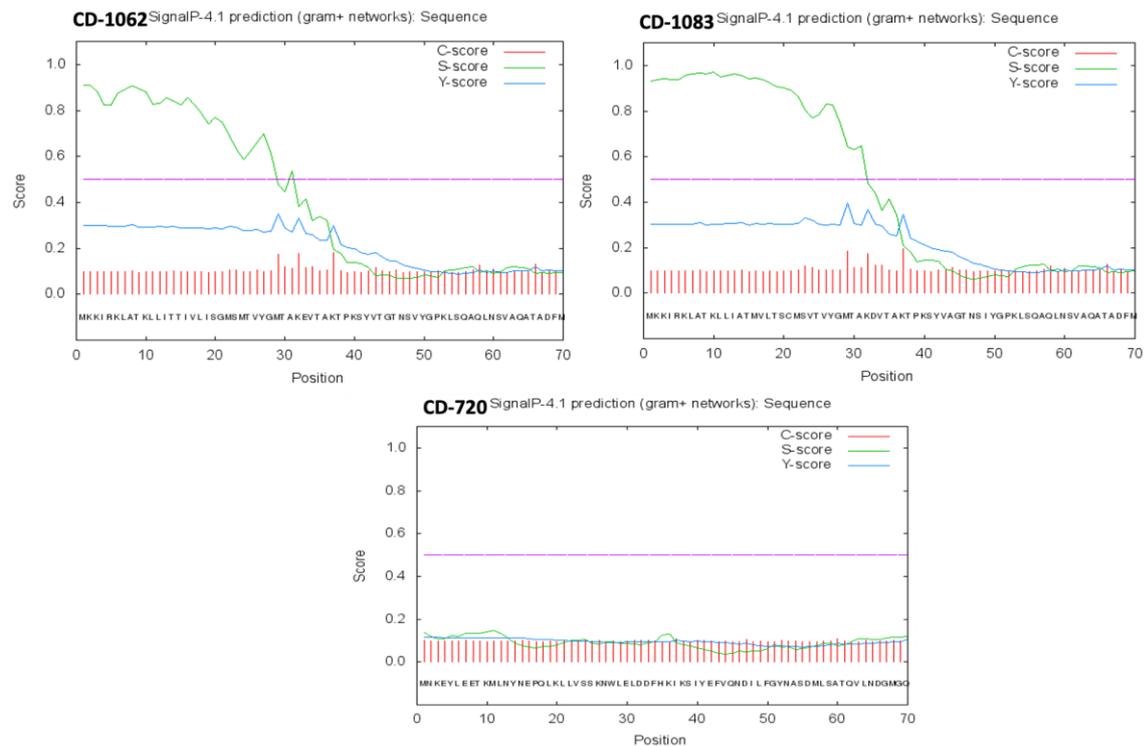


Figure 3.4. Prediction of Signal peptides in CD-1062 and CD-1083 Tgase genes. The protein sequences of Tgase-like genes in *C. difficile* 630 were uploaded into SignalP 4.1 server to predict the presence of signal peptides within the protein. The CD-1062 and CD-1083 Tgase-like genes were predicted to have signal peptides with predicted cleavage sites between amino acids at positions 28 and 29. The CD-720 Tgase-like gene did not contain signal peptides based on the prediction analysis (C-score: raw cleavage site score predicts the cleavage site; S-score: signal peptide score differentiates between positions within and outside the signal peptide site, and mature peptide; Y-score: combined cleavage site score is a better predictor of signal peptide cleavage site than the C-score).

Searches on SignalP 4.1 server predicted the presence of signal peptides in two of three Tgase-like genes in *C. difficile* 630, CD-1062 and CD-1083 as illustrated in Figure 3.4. No signal peptides were found in CD-720 (Figure 3.4). After detection of signal peptides in *C. difficile* 630 Tgases, further investigation on the characteristics of *C. difficile* 630 Tgases was pursued. The molecular weight of mature *C. difficile* 630 Tgases is 36.2 kDa and 37 kDa respectively for CD-1062 and CD-1083. Sequence searches on NCBI-BLAST using the mature regions encoded by *C. difficile* 630 Tgases as query did not detect homologous sequences in other microorganisms.

3.3.1.2 Characteristics of putative microbial transglutaminases from *C. difficile* 630

Analysis of *C. difficile* 630 Tgase gene sequences for predictions of protein localizations was performed using Softberry's ProtCompB and isoelectric points (pI) using ExPASy tool. Rare codons within Tgase gene sequences were searched using the Genscript rare codon analysis tool. The characteristics of putative Tgase-like genes in *C. difficile* 630 as predicted by several tools is summarised in Table 3.1 below.

Table 3.1. Summary of the predicted characteristics of transglutaminase-like genes in *C. difficile* 630.

Locus tag	pI	localization	Length (nucleotide)	Signal peptide	Rare codons	GC content (%)
CD630_25070	8.75	membrane	720	no	yes	29.2
CD630_30430	9.3	secreted	1062	yes	yes	35.91
CD630_11560	9.18	secreted	1083	yes	yes	36.5

The results in Table 3.1 provide a summary of the predicted characteristics of putative transglutaminases in *C. difficile* 630. Both CD-1062 and CD-1083 Tgase proteins are predicted to be secreted and CD-720 is predicted to be membrane-bound. All three Tgase-like genes in *C. difficile* 630 were predicted to contain rare codons. The CD-1062 and CD-1083 proteins have GC contents greater than 30% meanwhile the CD-720 protein has GC content of 29% which is within the acceptable range defined as values below 30%.

3.3.2 C. difficile 630 spores contain transglutaminase activity

To determine if *C. difficile* 630 contains transglutaminase activity, activity was monitored at distinct times during sporulation as shown in Figure 3.5.

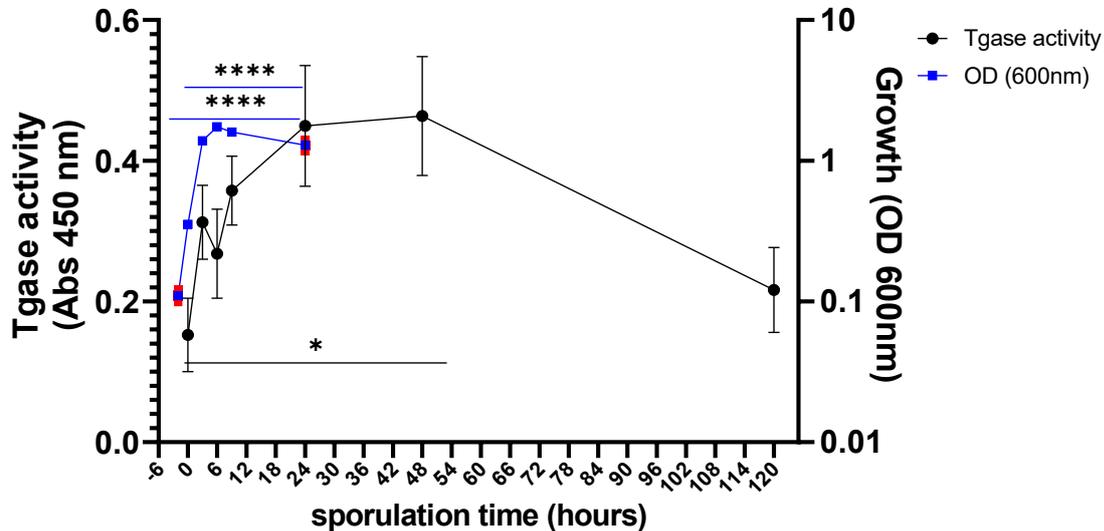


Figure 3.5. Transglutaminase activity in sporulating *C. difficile* 630 cells. Growth was monitored at OD_{600nm} from the point of sporulation induction up to the stationary phase. Sporulation was induced in broth at 0 h and cultures were plated on agar and harvested at different time points to determine Tgase activity. Tgase activity was measured by BTC incorporation into casein as described in the Methods section 2.6.6. Asterisks (*) indicate statistical significance where $p < 0.05$.

Results in Figure 3.5 show that *C. difficile* 630 spores harvested from agar plates at different times during sporulation contain Tgase activity. Growth measurements were not pursued beyond the stationary phase (24 hours) because cells were in the death phase. Tgase activity was detected throughout sporulation but a decrease in activity was observed after 48 h of sporulation. Enzyme activity was detected from the time of sporulation induction and peak Tgase activity occurred at 48 hours. After detection of Tgase activity in *C. difficile* 630 spores, the next step was to determine whether putative Tgase-like genes are expressed during sporulation.

3.3.3 Putative transglutaminase genes are expressed during sporulation in *C. difficile* 630

The quality of RNA is critical in transcript analysis during gene expression. Several methods of RNA extraction have been published and choosing the best method to suit the needs of the research is important (Metcalf, Sharif and Weese, 2010). A protocol that would generate high yields while maintaining RNA integrity was chosen following screening and optimisations of

several kits. Trizol based methods and enzymatic methods using kits were tested. In subsequent experiments, Trizol extraction was adopted due to higher yields compared to enzymatic methods. Very low yields (2-5 μg) were obtained using Promega kits, meanwhile higher yields (10-15 μg) were obtained with NEB kits and the highest yields up to 55 μg resulted from Trizol extractions. Total RNA extraction from 10 ml cultures using the Trizol method is described in Section 2.3.1. All RNA extractions were of good quality with $\text{Abs}_{260/280\text{nm}}$ of approximately 2. Total RNA was extracted from sporulation cultures at different time points corresponding to exponential and early stationary phases (OD_{600} : 0.1 to 1.6) of growth. Primers were designed on Primer 3 based on gene sequences obtained from NCBI. Primers for *spo0A* and *rpsJ* were taken from previously published studies (Pettit *et al.*, 2014). The *rpsJ* reference gene used in this study has served as housekeeping gene in previous studies of gene expression in Gram-positive bacteria, and shown to be stably expressed throughout growth in *C. difficile* (Metcalf, Sharif and Weese, 2010). Hence samples were normalised with *rpsJ* to eliminate experimental variations. The figures below are representative of the amplification and melting curves obtained from typical qPCR reactions in our system.

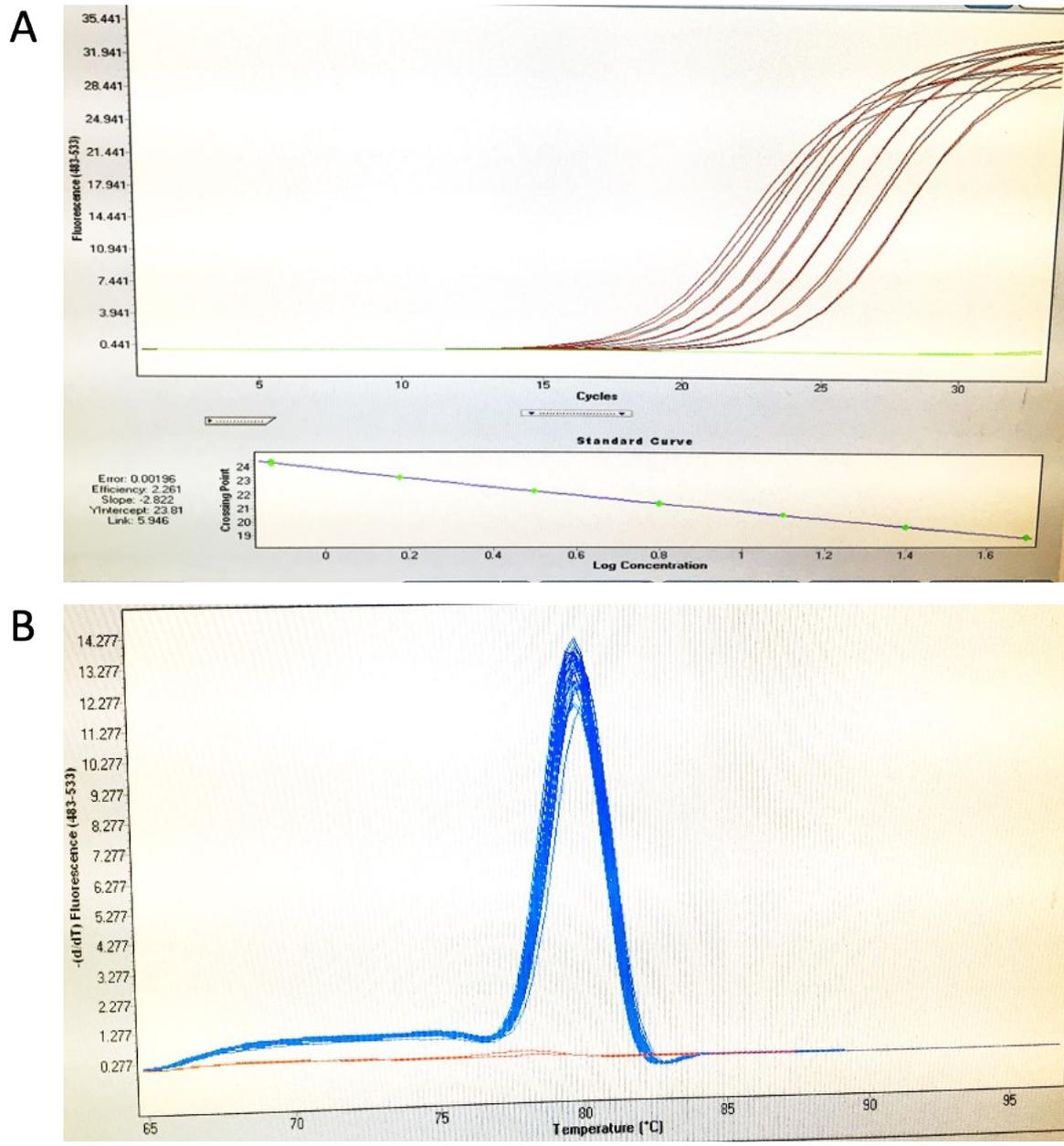


Figure 3.6. Representative amplification and melting curves obtained from qRT-PCR reactions. After qRT-PCR reactions of target and control genes, amplification curves were generated from serial dilutions of cDNA samples. Nuclease-free water was used as negative control. Amplification of target genes was complete after 33 cycles. Melting curves were generated at the end of every run to verify primer specificity. The figure shown is representative of a typical amplification and melting curve obtained under our experimental conditions. (A) amplification curve for *rpsJ* plus standard curve at bottom. The standard curve is a plot of crossing point expressed in number of cycles versus log concentration of standards (B) Melting curve for *rpsJ* gene.

The qRT-PCR reactions were successful under the reaction conditions used. Amplification of target genes was complete after 33 cycles as shown on the amplification curve in Figure 3.6. The efficiency of the PCR reactions was greater than 90% as determined from the slopes of

the standard curves. The qPCR software automatically calculates the efficiency of the amplification reaction, and a perfect amplification reaction produces an efficiency of 2 which corresponds to 100%. The qPCR reactions in our study produced efficiencies ranging from 2.12 to 2.38. The melting curve analysis from all amplification reactions generated a single peak which suggests that there was no non-specific amplification of small products. The gene expression studies were performed using three primer sets corresponding to the Tgase-like genes identified in *C. difficile* 630, and detection of *spo0A* transcripts was used as an indicator for sporulation initiation. Figure 3.7 below represents the relative gene expression of transglutaminases and *spo0A* during sporulation in *C. difficile* 630.

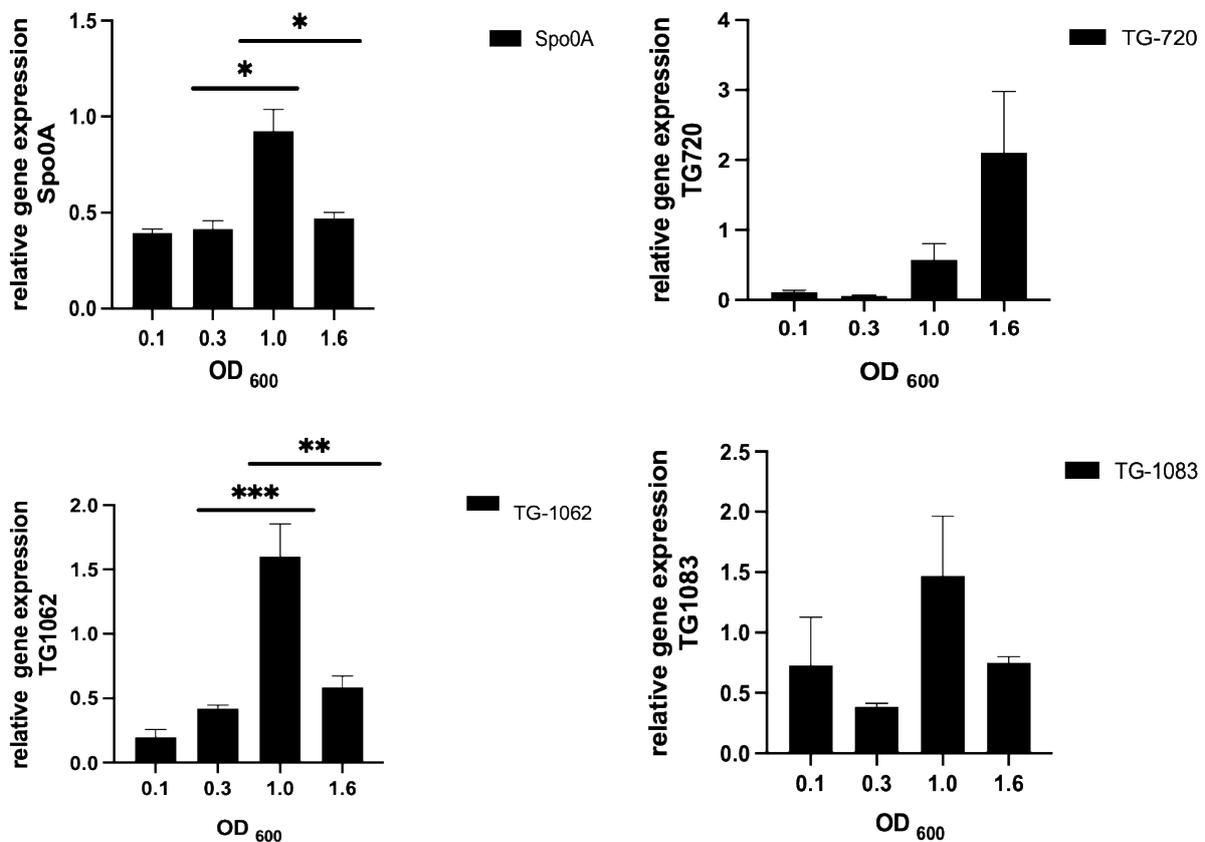


Figure 3.7. qRT-PCR analysis of Tgase genes and *spo0A* during sporulation in *C. difficile* 630. Total RNA was extracted from sporulating cell cultures at specific time points. Reverse transcription of RNA into cDNA was performed and cDNA was used as template in qPCR analysis using gene specific primers for target Tgases, sporulation gene (*spo0A*) and reference gene (*rpsJ*). 33 cycles were sufficient for complete amplification. The *rpsJ* reference gene was used for normalization of gene expression levels. Results are presented as mean \pm SEM of relative gene expression levels representative of 3 biological replicates. Asterisks (*) indicate statistical significance where $p < 0.05$. TG-720, TG-1062, TG-1083 correspond to 720 bp, 1062 bp and 1083 bp transglutaminase-like genes in *C. difficile* 630.

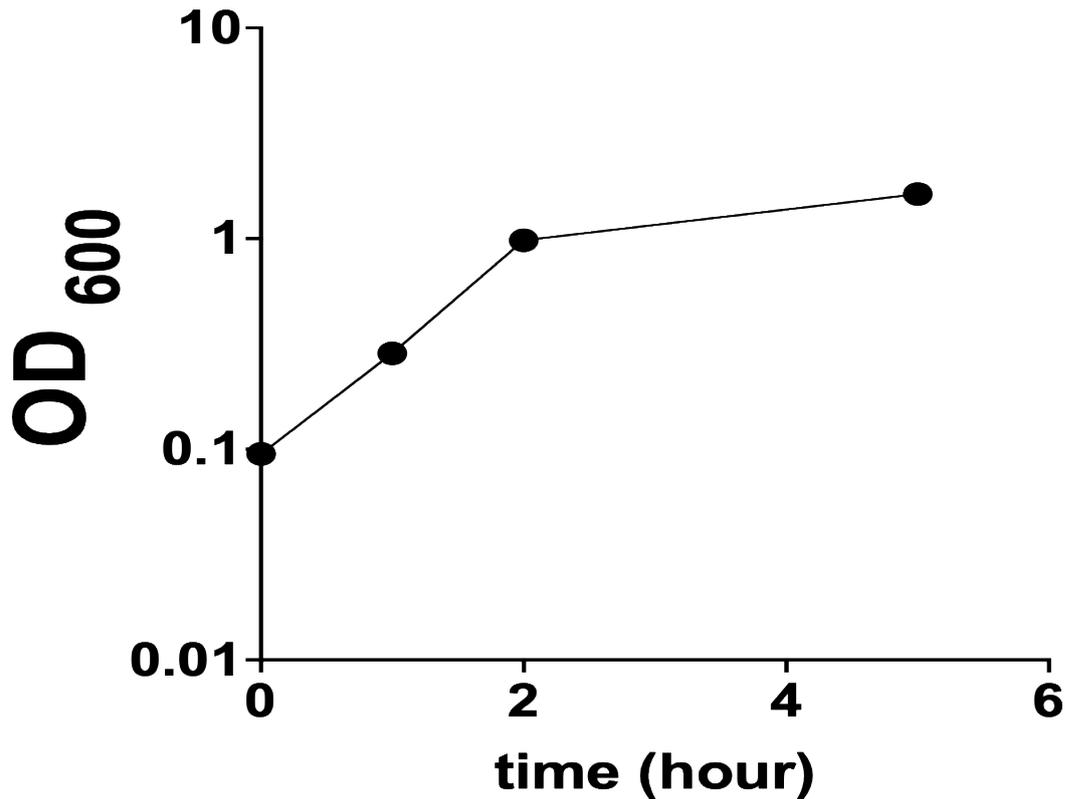


Figure 3.8. Growth curve of *C. difficile* 630. Growth was monitored at 600 nm and sporulation was induced by subculturing starter cultures in BHIS broth and monitoring growth up to stationary phase.

We analysed the Tgase and *spo0A* expression profiles during the exponential and early stationary phase of sporulating *C. difficile* 630 cultures in BHIS broth. Growth was monitored by OD measurements at 600 nm. The expression of *spo0A* upon sporulation initiation was useful in correlating growth (OD) with distinct sporulation stages. In Figure 3.7 and Figure 3.8, OD₆₀₀ of 0.1, 0.3, 1, 1.6 correspond respectively to early, mid, late log phases, and stationary phase. During these growth phases no spores (CFU) were detected when aliquots of broth cultures were heated at 70°C and plated on taurocholate agar plates. Expression of both *spo0A* and Tgases were detected very early at some time during growth that corresponds to the start of sporulation. The results in Figure 3.7 show that transglutaminase (TG-1062, TG-1083) and *spo0A* expression peaked at OD₆₀₀ of 1, corresponding to the late log phase and dropped off afterwards during the stationary phase to constitutive levels. Specifically, expression of TG-1083 increased by three-folds at the late log phase compared to OD₆₀₀ of 0.3 and decreased to baseline levels during the stationary phase. Similarly, TG-1062 expression doubled from early to mid-log phase, then rose significantly by four-folds in the late log phase before decreasing to baseline levels during the stationary phase. The same pattern of gene expression was observed for *spo0A* which increased significantly by two-folds in the

late log phase and levelled off during the stationary phase to constitutive expression levels. In contrast TG-720 was minimally expressed in the early growth stages, but its expression increased by ten-fold towards the end of the exponential phase, with further increases into the stationary phase. Overall Tgases and *spo0A* were expressed throughout sporulation for the period of sampling in our study. To provide evidence that Tgases are associated with sporulation in *C. difficile*, correlation studies between Tgase gene expression and *spo0A* expression were performed. There was no correlation observed between TG-720 ($r=0.02$, $p=0.98$) and TG-1083 ($r=0.94$, $p=0.06$) and *spo0A* expression. Meanwhile there was a statistically significant correlation between TG-1062 Tgase expression and *spo0A* expression ($r=0.99$, $p=0.01$). This evidence supports the possibility that these genes are involved in sporulation.

3.4 DISCUSSION

3.4.1 *C. difficile* 630 spores contain transglutaminase activity which is produced during sporulation

We identified Tgase activity in *C. difficile* 630 spores, and activity was expressed all through sporulation with peak activity coinciding with the stationary phase. A similar observation was made in *B. subtilis* where Tgase activity was found in sporulating cells late during sporulation, and maximal activity coincided with the stationary phase. Moreover the enzyme was reported to be spore-bound (Katsunori Kobayashi *et al.*, 1998). Also, Tgase was detected in culture supernatants of *S. mobaraense* DSMZ and activity increased during growth from 24 h to 95 h (Pasternack *et al.*, 1998). Likewise, Tgase was secreted in cultures of *S. mobaraense* B-3729, *S. paucisporogenes* ATCC-12596 and *S. platensis* during fermentation. Activity increased till 4-5 days for *S. platensis* and *S. mobaraense*, and up to 8 days for *S. paucisporogenes* (Nagy and Szakacs, 2008).

3.4.2 Putative transglutaminases are expressed during sporulation in *C. difficile* 630

We wanted to determine if Spo0A is associated with expression of putative Tgases during sporulation in *C. difficile* 630. Consistent with Tgase gene identification in *C. difficile* 630 genome by keyword searches, we predict expression of Tgases during sporulation following qRT-PCR experiments. Ubiquitous expression of CD-1062/CD-1083 Tgases throughout sporulation mimics *spo0A* expression. In *C. difficile* 630, *spo0A* gene expression was detected all through growth. Expression was perceived just 1 hour after sporulation induction up to the stationary phase, with peaks at the end of exponential phase. This concurs with other studies where *C. difficile spo0A* was detected early during sporulation and expressed all through

growth (Rosenbusch *et al.*, 2012; Pettit *et al.*, 2014). In these studies, *spo0A* expression was perceived 3 hours after sporulation induction and persisted through to the stationary phase which resembles our expression profile (Rosenbusch *et al.*, 2012). Additionally, we observed a significant two-fold increase in *spo0A* expression during the exponential phase, which corroborates the expression pattern in another study where 20-fold increases in *spo0A* expression were observed in mid exponential phases (Rosenbusch *et al.*, 2012). In their study Spo0A levels remained high after 6 hours all through till the stationary phase, which differs from our study where a significant drop to baseline expression levels was observed after the log phase. In the present study *spo0A* expression was maximal in the late log phase. Similarly, a different study reported maximal expression in the late exponential phase where the highest numbers of spores were detected (Pettit *et al.*, 2014). However, the picture is slightly different in another study that reported maximal expression in the stationary phase (Rosenbusch *et al.*, 2012). This is likely due to the heterogeneity in sporulating cells which imparts on gene expression at distinct stages during sporulation (Saujet *et al.*, 2013). A representative sketch of the *spo0A* targets upstream the transcriptional start site of putative Tgase genes in *C. difficile* 630 is shown below.

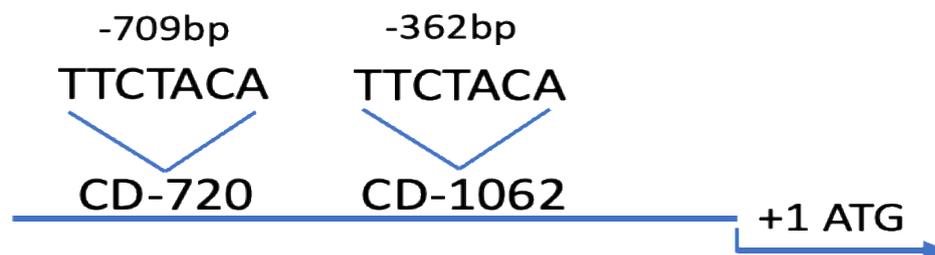


Figure 3.9. Possible *spo0A* targets upstream transglutaminase-like DNA sequences in *C. difficile* 630. The schematics illustrates the putative *spo0A* box found after 709 bp and 362 bp upstream CD-720 and CD-1062 transglutaminase locus respectively and the translational start site.

Following searches for consensus *spo0A* sequences around *C. difficile* 630 Tgase genes, the classical *spo0A* box was found 362 bp upstream CD-1062 and 709 bp upstream CD-720 as illustrated in Figure 3.9. This suggests that CD-1062 is probably directly regulated by Spo0A and therefore a potential candidate sporulation gene. This is in line with the results obtained from correlation analysis of *spo0A* and Tgase gene expression. It shows that CD-1062 gene expression correlated significantly with *spo0A* gene expression ($r=0.99$, $p=0.01$) with comparable expression patterns reinforcing the probability that CD-1062 could be a direct *spo0A* target. Indeed, expression of CD-1062 is upregulated in the transition phase which coincides with upregulation in *spo0A* expression. This agrees with other studies where Spo0A

was upregulated towards the stationary phase and so could suggest the likely association of Tgases with sporulation in CD630 (Pettit *et al.*, 2014). Thus CD-720 which has a different expression pattern from *spo0A* may not be a *spo0A* target. It is also plausible that *C. difficile* Tgase-like genes could be involved in both sporulation and metabolism as is Spo0A given their detection early on during exponential growth. Consistent with this observation Spo0A has been reported to control processes beyond sporulation such as biofilm formation, toxin production and metabolism (Rosenbusch *et al.*, 2012; Pettit *et al.*, 2014).

In another study the sporulation scheme was mainly active during the exponential phase illustrated by a rise in the number of spores during the transition phase. During the stationary phase in the same study no detectable change in the number of spores was recorded (Pettit *et al.*, 2014). This corresponds to our analysis of gene expression for *spo0A* and all Tgase genes except CD-720. In our study, the sporulation program defined by *spo0A* expression was highly active during the late exponential phase. The expression pattern observed is justified by reports in *B. subtilis* where spore formation occurs in the transition growth phase during which gene expression switches to assist spore formation instead of vegetative growth (Edwards and McBride, 2014). In the transition phase, cell growth shifts from exponential to the stationary phase. Indeed, during this growth phase in *B. subtilis*, Spo0A is activated and turns on sporulation-related gene expression thereby acting as the key transcriptional regulator for sporulation (Edwards and McBride, 2014). Therefore, in the present study *spo0A* upregulation towards the late log phase may be responsible for Tgase-like gene expression given its role in transcriptional regulation during sporulation. Other studies used Western blot with Spo0A antibodies or spore enumeration by colony forming units to evaluate changes in Spo0A production during sporulation (Rosenbusch *et al.*, 2012; Pettit *et al.*, 2014). However, CFU counts or Western blots would not have added much information to what was already obtained from the *spo0A* gene expression.

3.4.3 Characteristics of microbial transglutaminase-like proteins from *C. difficile* 630

The three putative Tgases identified in *C. difficile* 630 belong to the Tgase-like superfamily which constitutes a large group of proteins from bacteria, eukaryotes, and archaea that are homologs of animal Tgases (Makarova, Aravind and Koonin, 1999). The molecular weight and pI of bacterial Tgases are different compared to that of mammalian Tgases. Mostly, microbial transglutaminases are smaller in size compared to mammalian Tgases with a few exceptions (Lin *et al.*, 2004). Mammalian Tgases (tTg) have a pI of 4.5 and molecular weight of 80 kDa (Fernandes *et al.*, 2015). The molecular weight of putative *C. difficile* 630 Tgases is lower compared to that of mammalian Tgases, about half that of tissue Tgase and a proportionally higher pI. The results in Table 3.1 show that the pI of CD-720 is very similar to that of

Streptoverticillium S-8112 (Ando *et al.*, 1989). The predicted molecular weight of CD-1062 and CD-1083 is similar to that of most characterized microbial transglutaminases such as *S. hygroscopicus* of 38kDa and *Strv.* S-8112 of 40 kDa (Ando *et al.*, 1989; Cui *et al.*, 2007). Similarly, the predicted molecular weight of CD-720 is almost identical to that of *B. subtilis* Tgl protein. The molecular weights of mature *C. difficile* 630 Tgases are similar to that of mature Tgase from *Strv. ladakanum* which is produced extracellularly (Lin *et al.*, 2004). The CD-720 Tgase was predicted to be membrane-bound analogous to the Tgase protein, TgpA in *P. aeruginosa* PA01 strain which is an integral membrane protein (Milani *et al.*, 2012). The other two *C. difficile* 630 Tgases were predicted to be secreted. Likewise most characterised Tgase proteins from *Streptomyces* species have been produced extracellularly (Kashiwagi *et al.*, 2002). Moreover, both CD-1062 and CD-1083 contain an N-terminal signal peptide sequence with predicted cleavage sites after the first 28 amino acids. The signal peptide prediction ties with the localisation prediction. The presence of signal peptides in Tgases from *S. ladakanum* and *Strv. mobaraense* enabled protein secretion to the extracellular medium.

The catalytic domains of *C. difficile* 630 Tgase-like genes were well conserved as observed from MLSA with characterized Tgases, showing high homology in the region around the catalytic cysteine residue (Figure 3.1). Sequence similarity searches with the entire nucleotide sequence via BLAST analysis did not detect *C. difficile* 630 Tgase homologs in other organisms. Similarly, no homologs were identified using either of the conserved fold domains from Tgase-like superfamily members as query. However homologous sequences were identified in *E. coli* with 100% identity using the CD-1062 entire protein sequence as query. As reported elsewhere, sequence similarity searches did not identify relatives of animal Tgases in other organisms containing the conserved Tgase fold. This was demonstrated by the failure to detect MTG and Tgl as relatives of animal Tgases through sequence similarity (Fernandes *et al.*, 2015). This is surprising given that relatives of animal Tgases possess the NlpC/P60 fold, and truly both proteins contain the NlpC/P60 core. Indeed, the active site domain of all Tgases derive from thiol proteases and thus contain the NlpC/P60 catalytic core. The complete lack of sequence similarity may suggest convergent evolution. Microbial transglutaminases are structurally different from animal Tgases which is coherent with the phylogenetic tree analysis in Figure 3.3 showing unrelatedness. Microbial transglutaminases are single domain proteins meanwhile mammalian Tgases are multidomain proteins and both have different molecular weights. Furthermore, microbial transglutaminases do not use a catalytic triad like their mammalian counterparts (Kashiwagi *et al.*, 2002). Figure 3.3 shows the relatedness between fish Tgase, TG2 and Factor XIIIa. This was conceivable given that all three Tgases operate via a cysteine, histidine, aspartate catalytic triad (Kashiwagi *et al.*, 2002). Additionally, they have some structural features in common including a catalytic

cysteine insulated by a hydrophobic tunnel important in protein crosslinking (Fernandes *et al.*, 2015). The phylogenetic analysis shows that *Streptomyces* MTG is closely related to *Bacillus* Tgl. This is not unusual given that *B. subtilis* Tgl and *Streptomyces* MTG contain the NlpC/P60 core and undergo catalysis via a Cys-Asp catalytic dyad (Kashiwagi *et al.*, 2002; Fernandes *et al.*, 2015). It was reported that *Streptomyces* MTG demonstrates minimal sequence homology with other redundant Tgases, except for *Bacillus* Tgl at the active site cysteine in the catalytic domain. The similarity around the catalytic sites suggests a functional association between MTG and Tgl (Kashiwagi *et al.*, 2002).

No known mammalian or microbial transglutaminase contains a serine residue next to the catalytic cysteine as observed in CD-1062 and CD-1083 from *C. difficile* 630. This could have implications in the catalytic activity given that serine is usually found in the catalytic site of serine proteases (Kashiwagi *et al.*, 2002). However, the serine residue did not appear to be conserved as shown on the alignment. Substitution of active site cysteine with a serine residue has been observed in a highly conserved family of protease homolog from *Plasmodium*. Thus serine if correctly oriented can act as a catalytic nucleophile in cases where the other motifs are conserved (Makarova, Aravind and Koonin, 1999). After detecting Tgase activity in *C. difficile* 630 spores, gene expression studies were pursued to confirm the presence of Tgase transcripts during sporulation which could be suggestive of Tgase association with sporulation. Bioinformatics analysis enabled identification of putative Tgases in *C. difficile* 630 and detecting conserved Tgase domains that aligned well with characterized Tgase-like superfamily members.

3.5 Conclusion

Transglutaminase activity was found in sporulating *C. difficile* 630 cultures. Furthermore, transglutaminase-like transcripts were detected in sporulating cultures. Correlation data showed that CD-1062 expression might be associated with *spo0A* expression, and thus could be involved in sporulation in *C. difficile* 630. In addition, *spo0A* boxes were found upstream the CD-1062 gene. Therefore, Spo0A may be involved in the regulation of Tgase gene expression during sporulation in *C. difficile* 630. Three putative Tgase genes were identified in the complete genome of *C. difficile* 630 and all contained the conserved catalytic domain present in Tgase-like superfamily members. Alignments of *C. difficile* 630 Tgases with conserved Tgase motifs suggests that the catalytic domain around the cysteine residue in *C. difficile* 630 Tgases may represent the typical Tgase conserved catalytic site. The *C. difficile* 630 Tgases had molecular weights similar to characterized microbial transglutaminases and lower compared with human Tgases. Signal peptides were found within CD-1062 and CD-

1083, and both proteins were predicted to be secreted meanwhile CD-720 was predicted to be localised within the membrane. The results shown in this chapter provide evidence of Tgase-like activity and expression in *C. difficile* 630 during sporulation. After identification of novel important genes in *C. difficile* 630, cloning of DNA containing target genes into expression vectors followed by recombinant expression in *E. coli*, purification and characterization studies were pursued.

CHAPTER 4. Recombinant Expression, Purification and Characterisation of transglutaminase-like proteins from *C. difficile* 630

4.1 Introduction

Bacterial hosts have been used as expression system for recombinant production of microbial transglutaminases (Tgases). Microbial transglutaminase (MTG) was cloned and recombinantly expressed as a mature active enzyme in the extracellular medium in *S. lividans* (Washizu *et al.*, 1994). Also, MTG encompassing the entire coding region has been produced as an insoluble and inactive protein in *E. coli*. The mature enzyme was activated following enzymatic cleavage of the fusion partner with Factor Xa (Yokoyama, Nio and Kikuchi, 2004). Also MTG has been recombinantly produced extracellularly as an inactive pro-protein which required proteolytic processing to obtain a mature and active protein (Pasternack *et al.*, 1998). *E. coli* is a widely used system for production of recombinant foreign proteins and has been used to obtain high levels of MTG (Marx, Hertel and Pietzsch, 2007; Javitt *et al.*, 2017). Both *E. coli* BL21 and Rosetta 2 cells were adopted as expression hosts in this study to determine the best strain for optimum growth while preserving enzyme activity. Most expression studies with *E. coli* BL21 have been done at 37°C, which could be the temperature for optimum growth in this strain (Rosano and Ceccarelli, 2014). However, using 37°C for recombinant Tgase expression could affect some properties of the protein. In this chapter, putative Tgases comprising the whole coding region in *C. difficile* 630 were cloned and recombinantly expressed in *E. coli* strains. The recombinant Tgase-like proteins were purified by affinity chromatography and characterised to detect the presence of enzyme activity.

4.2 Objectives

The objectives of this chapter will be:

- To PCR amplify and clone Transglutaminase-like genes from *C. difficile* 630, and recombinantly express using different hosts, temperature conditions and expression systems.
- Optimization of the putative transglutaminase recombinant protein expression.
- Purification and characterization of recombinant putative Tgases from *C. difficile* 630.

4.3 RESULTS

4.3.1 *C. difficile* 630 genomic DNA extraction

After identification of putative transglutaminase-like genes within the genome of *C. difficile* 630, genomic DNA (gDNA) was extracted prior to PCR amplifications and cloning into suitable vectors for recombinant protein expression and purification. *C. difficile* 630 gDNA was extracted from overnight cultures using the GenElute Bacterial Genomic DNA kit following the manufacturers' protocol. Success was verified by agarose gel electrophoresis prior to subsequent experiments.

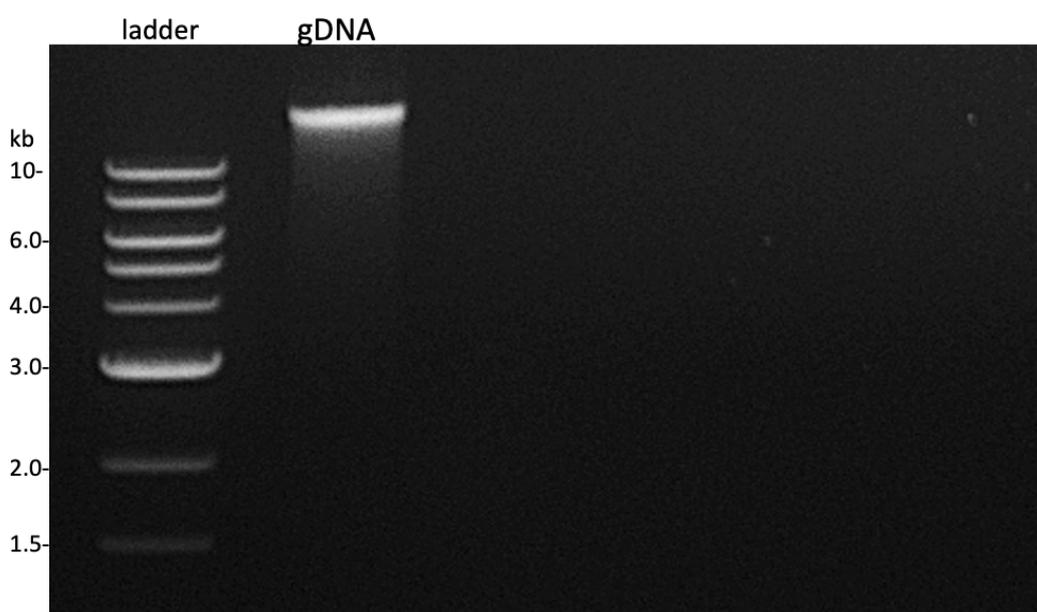


Figure 4.1. Genomic DNA extraction from *C. difficile* 630. Genomic DNA was isolated from an overnight culture of *C. difficile* 630 in BHI supplemented broth, according to the manufacturer's instructions described in Materials and methods. The success of extraction was verified after analysis of isolated DNA by agarose gel electrophoresis.

From Figure 4.1 the gDNA extraction was successful and a yield of 11 μg was obtained from 1.5 ml of culture. The gDNA was relatively pure with absorbance at 260/280nm of 1.9 which is within the acceptable range ($A_{260/280}$: 1.6-1.9).

4.3.2 PCR amplification of transglutaminase-like genes in *Clostridioides difficile* 630

The coding sequence of the three transglutaminase-like genes in *C. difficile* 630 were amplified from gDNA using primers listed in Table 2.1. The PCR amplifications of target genes were analysed by agarose gel electrophoresis as shown in Figure 4.2.

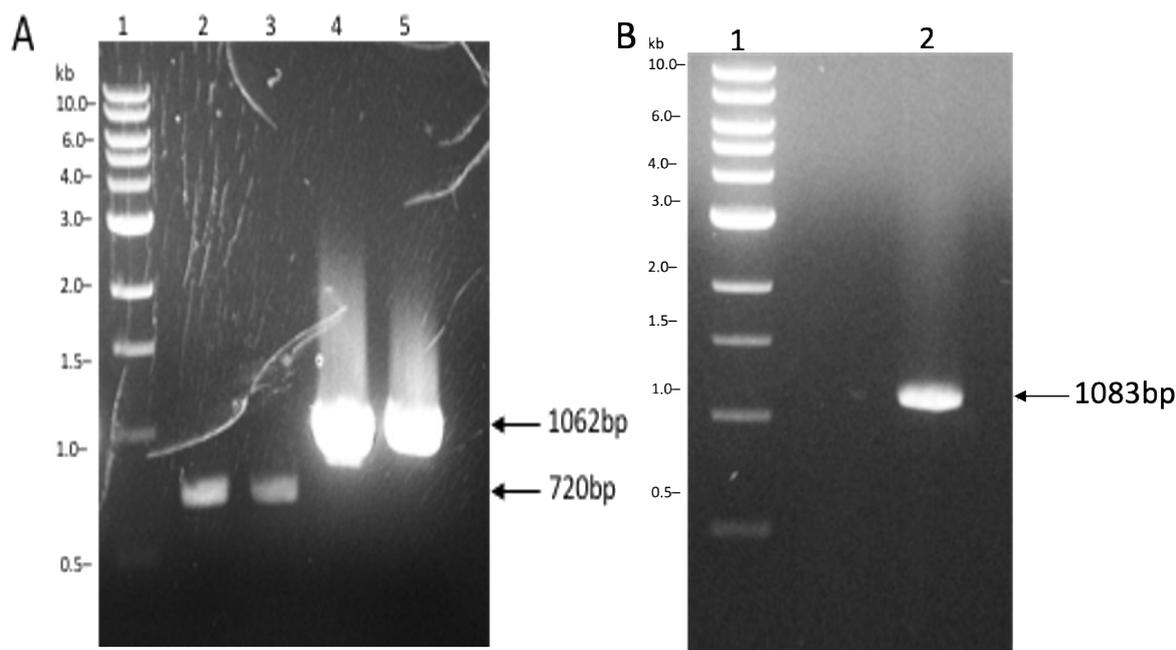


Figure 4.2. PCR amplifications of *C. difficile* 630 transglutaminase genes. For the amplification of transglutaminase-like genes in *C. difficile* 630 from a genomic DNA template, a touch-down PCR with the KOD hot start DNA polymerase kit was performed in a thermocycler as described in Materials and Methods. The success of PCR amplifications was determined by analysis of PCR products on 1% (w/v) agarose gels. lanes 1: DNA ladder; (A) lanes 2-3: 720 bp PCR product, lanes 4-5: 1062 bp PCR product, (B) lane 2: 1083 bp PCR product. Representative images of PCR amplifications.

To amplify the three transglutaminase-like genes identified in *C. difficile* 630, primers complementary to the transglutaminase coding sequence were designed on **Primer 3**, based on the gene sequences identified from National Center for Biotechnology Information (NCBI). The primers successfully amplified the coding region of our target transglutaminase genes from the gDNA template. The primers for CD-720, CD-1062 and CD-1083 transglutaminase-like genes worked efficiently with annealing temperatures set to 66-56°C, 70-60°C and 71-61°C respectively. As shown on the agarose gels in Figure 4.2, PCR products **A**, **B** and **C** contained considerable quantities of amplified DNA of 720 bp, 1062 bp and 1083 bp respectively (the arrows indicate the expected sizes of amplicons). The respective fragments

were excised from agarose gels and purified using a kit (Wizard SV Gel and PCR Clean-up system).

4.3.3 Cloning Transglutaminase-like genes from *C. difficile* 630

4.3.3.1 Cloning of blunt-end PCR products into pJET1.2/blunt and pUC19 cloning vectors

Cloning of transglutaminase inserts into pJET1.2/blunt and pUC19 cloning vectors was verified by agarose gel electrophoresis as shown in Figure 4.3. The pJET1.2/blunt vector is a positive selection system meanwhile the pUC19 vector contains reporter genes based on LacZ complementation. Blunt-ended purified PCR products **A** and **B** were cloned into the 2.97 kb pJET1.2/blunt cloning vector. The pJET1.2 vector includes a positive selection system where the inserted DNA disrupts a lethal gene enabling propagation of plasmids on ampicillin plates. Blunt-ended purified PCR product **C** proved difficult to clone in pJET1.2 and was successfully cloned into the 2.68 kb *Sma*I-digested pUC19 vector. Recombinant colonies from pUC19 ligations were white due to disruption of the α -complementation in LacZ gene by the insert DNA. Between 4 and 7 clones from each transformation was selected for recombinant plasmid isolation. The recombinant plasmids were digested to verify that plasmids contain inserts of correct size with the correct restriction sites present. Restriction digestion of recombinant plasmids confirmed the presence of transglutaminase inserts which were recovered by agarose gel electrophoresis shown in Figure 4.3.

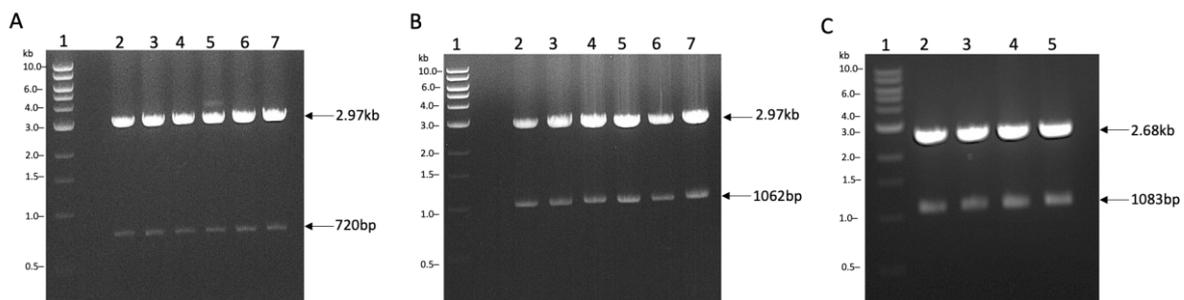


Figure 4.3. Restriction digestion of recombinant pJET1.2/blunt and pUC19 plasmids. The plasmids resulting from cloning PCR products into pJET1.2/blunt and pUC19 cloning vectors were digested with *Bam*HI and *Eco*RI. The digestion reactions were incubated for 2 hours at 37°C and digestion was confirmed following recovery of inserts by AGE. lane 1: DNA ladder, (A) lanes 2-7: CD-720 inserts, (B) lanes 2-7: CD-1062 inserts, (C) lanes 2-5: CD-1083 inserts.

The cloned inserts were released from pUC19/pJET1.2 plasmids by double digestion with *Bam*HI-*Eco*RI. Figure 4.3 illustrates the inserts at their corresponding sizes on the gel, isolated from the larger cloning pJET1.2 (2.97 kb) and pUC19 (2.68 kb) vectors that migrated to a

higher molecular weight. The isolated inserts were sub-cloned in linearized pGEX-3X expression vector. Inserts from several clones were mixed for ligation reactions so as to eliminate chances of error.

4.3.3.2 Sub-cloning of transglutaminase inserts into pGEX-3X expression vector

Cloning of transglutaminase inserts into expression vectors was verified by agarose gel electrophoresis. Isolated inserts were ligated into digested pGEX-3X expression vector and recombinant plasmids were selected on ampicillin plates. Between 5 and 10 clones from each transformation was selected for recombinant plasmid isolation. To confirm the success of cloning, recombinant plasmids were digested to verify the presence of target gene inserts and then sequenced using pGEX-3X F/R universal primers. Restriction digestion of recombinant pGEX-3X plasmids confirmed the presence of transglutaminase inserts which were recovered by agarose gel electrophoresis shown in Figure 4.4.

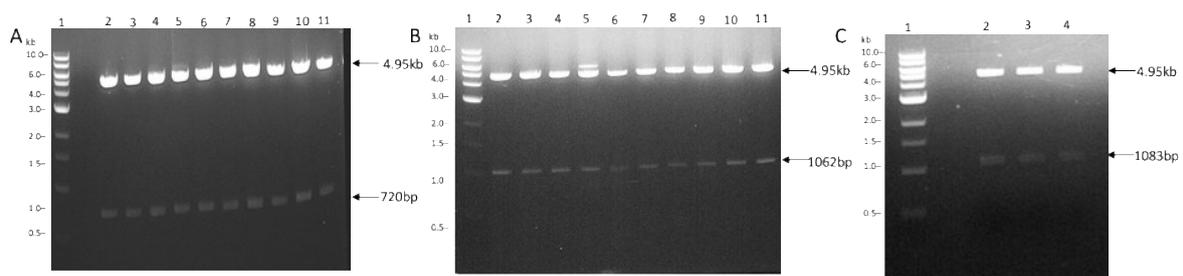


Figure 4.4. Restriction digestion of recombinant plasmids after sub-cloning in pGEX-3X. Recombinant plasmids resulting from ligation of transglutaminase inserts into pGEX-3X expression vector were digested with *Bam*HI and *Eco*RI. The success of cloning was confirmed after recovery of inserts by digestion as shown on agarose gels and the recombinant plasmids were sequenced to verify the integrity of the whole coding sequence. lane 1: DNA ladder, (A) lanes 2-11: CD-720 inserts, (B) lanes 2-11: CD-1062 inserts, (C) lanes 2-4: CD-1083 inserts.

The agarose gel in Figure 4.4 illustrates recovery of cloned inserts following *Bam*HI-*Eco*RI digestion of recombinant expression plasmids. The inserts migrated to their expected sizes which confirmed the success of cloning (expression plasmid construction). The expression constructs were sequenced across the GST tag using pGEX-3X F/R universal primers to verify the integrity of the whole coding sequence prior to recombinant protein production and purification.

4.3.4 Protein Expression and Purification of GST-tagged transglutaminase-like Proteins

About 6 clones from each transformation was subjected to nucleotide sequencing, and clones with 100% match to *C. difficile* 630 target genes were selected for recombinant protein expression. From Sanger sequencing results, target transglutaminase genes were successfully sub-cloned into pGEX-3X expression vector. The cloned transglutaminases were expressed in *E. coli* BL21 cells as GST-tagged proteins under the control of the tac promoter in pGEX-3X expression plasmid that incorporates an N-terminal GST-tag. Briefly plasmids transformed with *E. coli* BL21 cells were grown overnight at 37°C to be used as starter cultures. A 1:50 dilution of overnight cultures was grown at 37°C prior to induction with IPTG at OD₆₀₀ of 0.5. Cells were harvested after 3 hours and processed for purification. The clarified lysates were purified by affinity chromatography on glutathione Sepharose columns. The estimated molecular weights of the recombinant proteins with the incorporated GST-tag were 54 kDa for CD-720, and 65 kDa for both CD-1062 and CD-1083 fusion proteins. Purification of the expressed transglutaminase-like proteins was monitored on an AKTA PURE explorer system as shown in Figure 4.5.

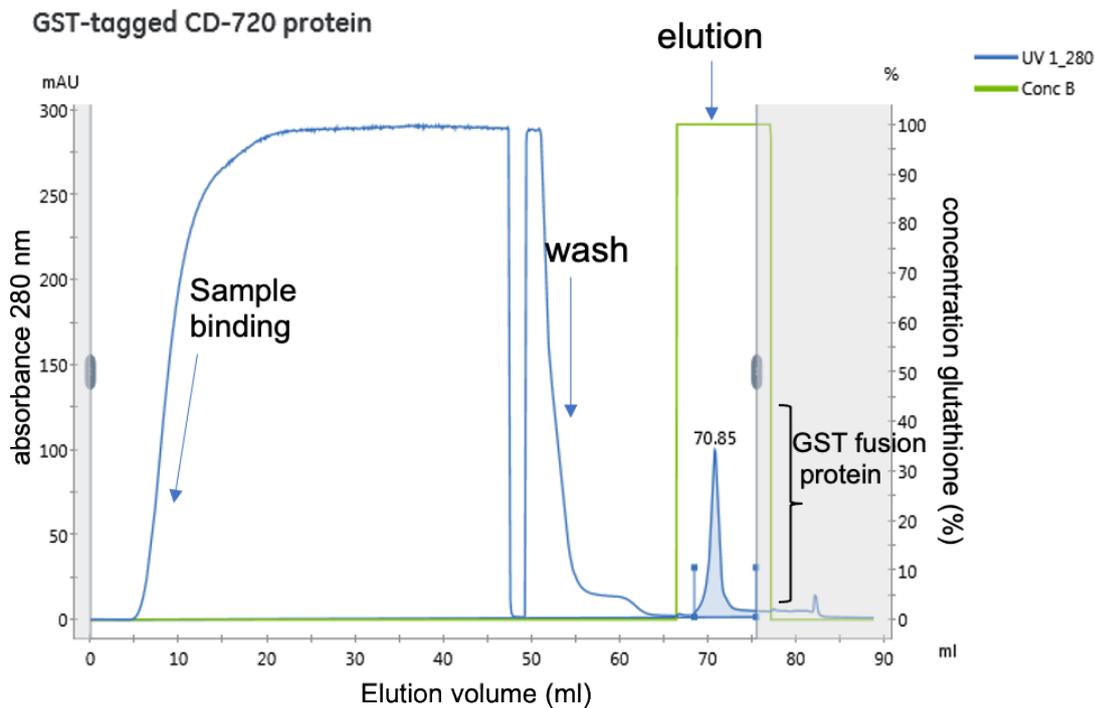


Figure 4.5. Affinity purification scheme of recombinant GST-CD720 proteins from *C. difficile* 630. Expression constructs were induced for protein expression at 37°C and after lysis of harvested cells, clarified lysates were purified on GSTrap HP columns. The purification process was monitored at 280 nm on the AKTA PURE system. The figure shows purification for the GST-CD720 fusion protein (Blue line: protein absorbance, green line: glutathione concentration in % (v/v) from a 10 mM preparation).

The chromatogram in Figure 4.5 shows a representative purification of GST-CD720 fusion protein on Glutathione Sepharose columns monitored on an AKTA PURE system. A prominent elution peak was observed for the recombinant GST-CD720 protein as shown in Figure 4.5 contrary to the very small elution peaks for recombinant GST-CD1062 and GST-CD1083 proteins illustrated in Figure 4.6.

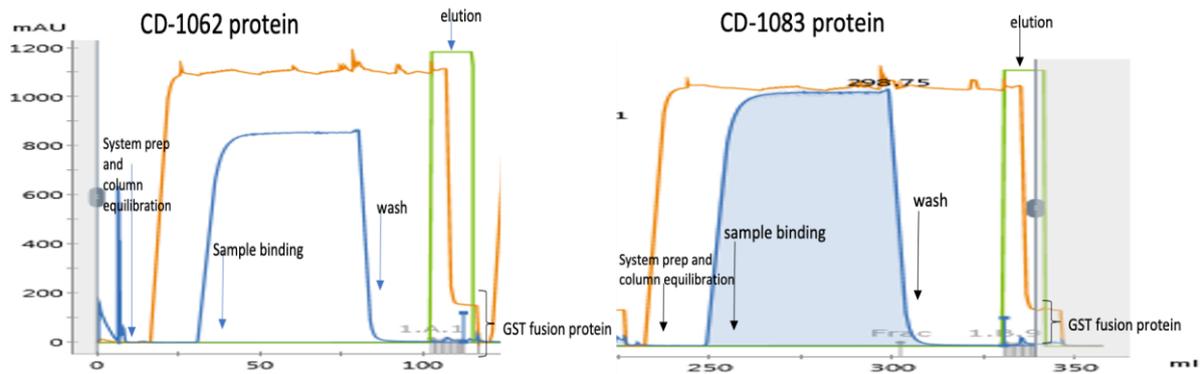


Figure 4.6. Affinity purification scheme of recombinant GST-CD1062 and GST-CD1083 proteins from *C. difficile* 630. Expression constructs were induced for protein expression at 37°C and after lysis of harvested cells, clarified lysates were purified on GSTrap HP columns. The purification process was monitored at 280 nm on the AKTA PURE system. The figure shows purification for the GST-CD1062 and GST-CD1083 fusion proteins. (Blue line: protein absorbance, green line: glutathione concentration in % (v/v) from a 10 mM preparation, orange line: conductivity).

The chromatogram in Figure 4.6 displays very small peaks following elution of recombinant GST-CD1062 and GST-CD1083 fusion proteins.

4.3.4.1 Recombinant expression of GST-tagged proteins in *E. coli* BL21 at 37°C.

After protein expression and affinity purification, cell lysates and eluted fractions were analysed by SDS-PAGE.

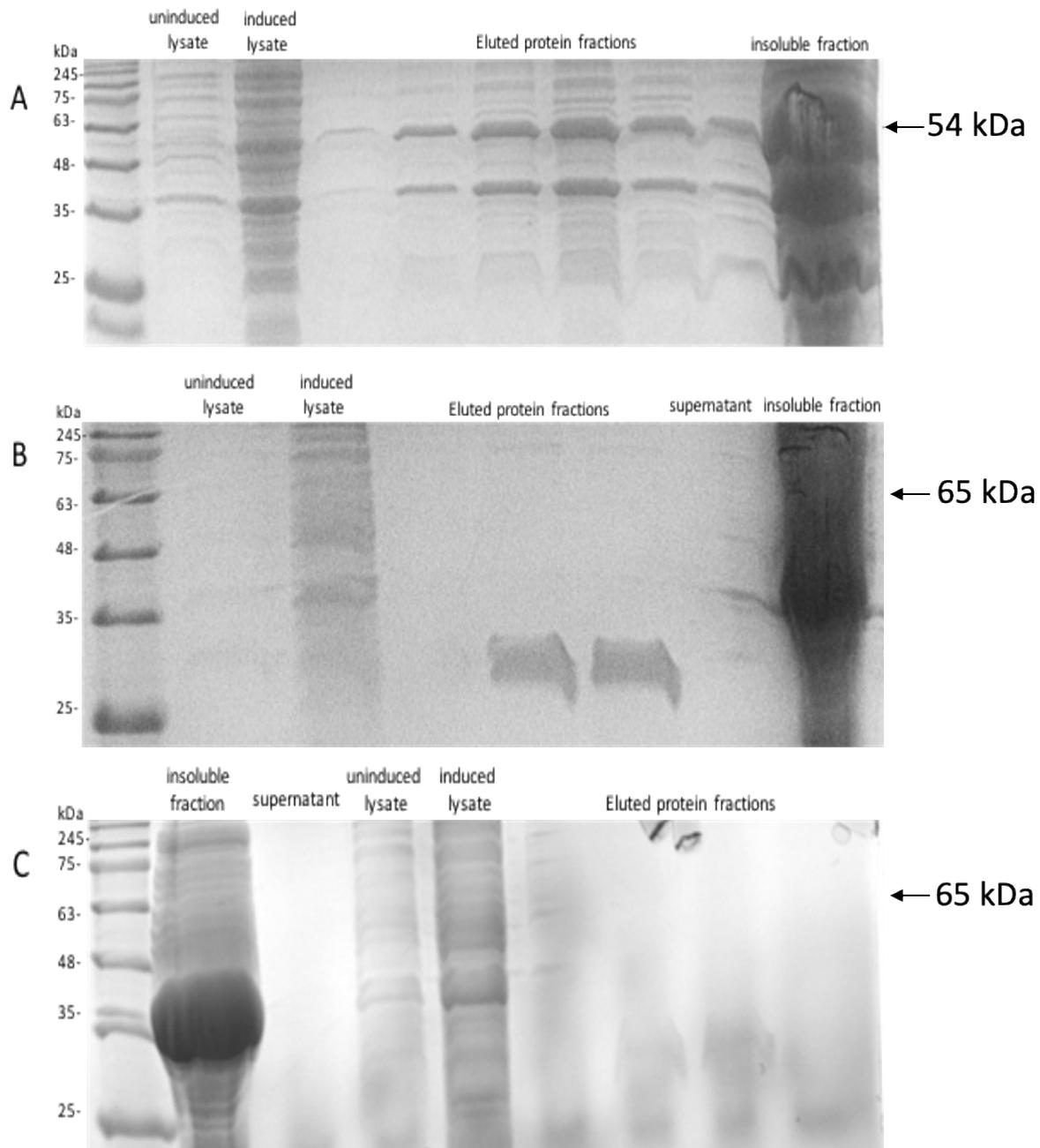


Figure 4.7. SDS-PAGE of recombinant GST-tagged protein expression at 37°C. Recombinant plasmids were induced for protein expression in *E. coli* BL21 as GST-tagged proteins with 0.5 mM IPTG for 3 hours at 37°C. Cells were harvested by centrifugation, lysed and clarified cell lysates were purified by affinity chromatography on GSTrap HP columns. Eluted fractions were analysed by SDS-PAGE and visualised after Coomassie blue staining (A) GST-CD720 (B) GST-CD1062 (C) GST-CD1083.

As shown in Figure 4.7, the protein yield following purification of GST fusion proteins (induced at 37°C) on GSTrap HP columns was non-detectable for GST-CD1062 and GST-CD1083 fusion proteins. Protein bands were observed in the soluble fractions of GST-CD1062 and GST-CD1083 fusion proteins but not at the expected size. There was an apparent non-target

band in the soluble fractions between 26-35 kDa for GST-CD1083 and GST-CD1062. The smallest GST-CD720 protein was detectable in the soluble fraction although an extra band of 35 kDa was present in the eluted fractions. Additionally, major insoluble bands were observed at the target size and at 35 kDa in the GST-CD720 protein gel. The protein purity of GST-CD720 was 40%. Under these expression conditions, GST-CD1062 and GST-CD1083 were majorly expressed as insoluble proteins and some amount of the GST-CD720 protein was insoluble. The insoluble proteins were mostly expressed at 35 kDa in all three induced lysates. Based on these initial results, optimisation was required to improve soluble protein expression.

4.3.4.2 Optimization of Recombinant Protein expression

Different temperatures and IPTG concentrations were evaluated to determine the ideal conditions for optimum protein expression. Cultures were harvested before and after induction and processed for SDS-PAGE analysis to verify transglutaminase protein expression at the target sizes. For optimisation of recombinant expression, cell cultures were scaled up to 500 ml for increased protein production. Cell pellets obtained from induced cultures were resuspended in suitable buffers prior to cell lysis using a French press. Clarified lysates containing recombinant proteins were purified on corresponding columns.

4.3.4.2.1 Effect of temperature on recombinant expression of GST-tagged proteins in *E. coli* BL21

After protein expression at room temperature and affinity purification, cell lysates and eluted fractions were analysed by SDS-PAGE.

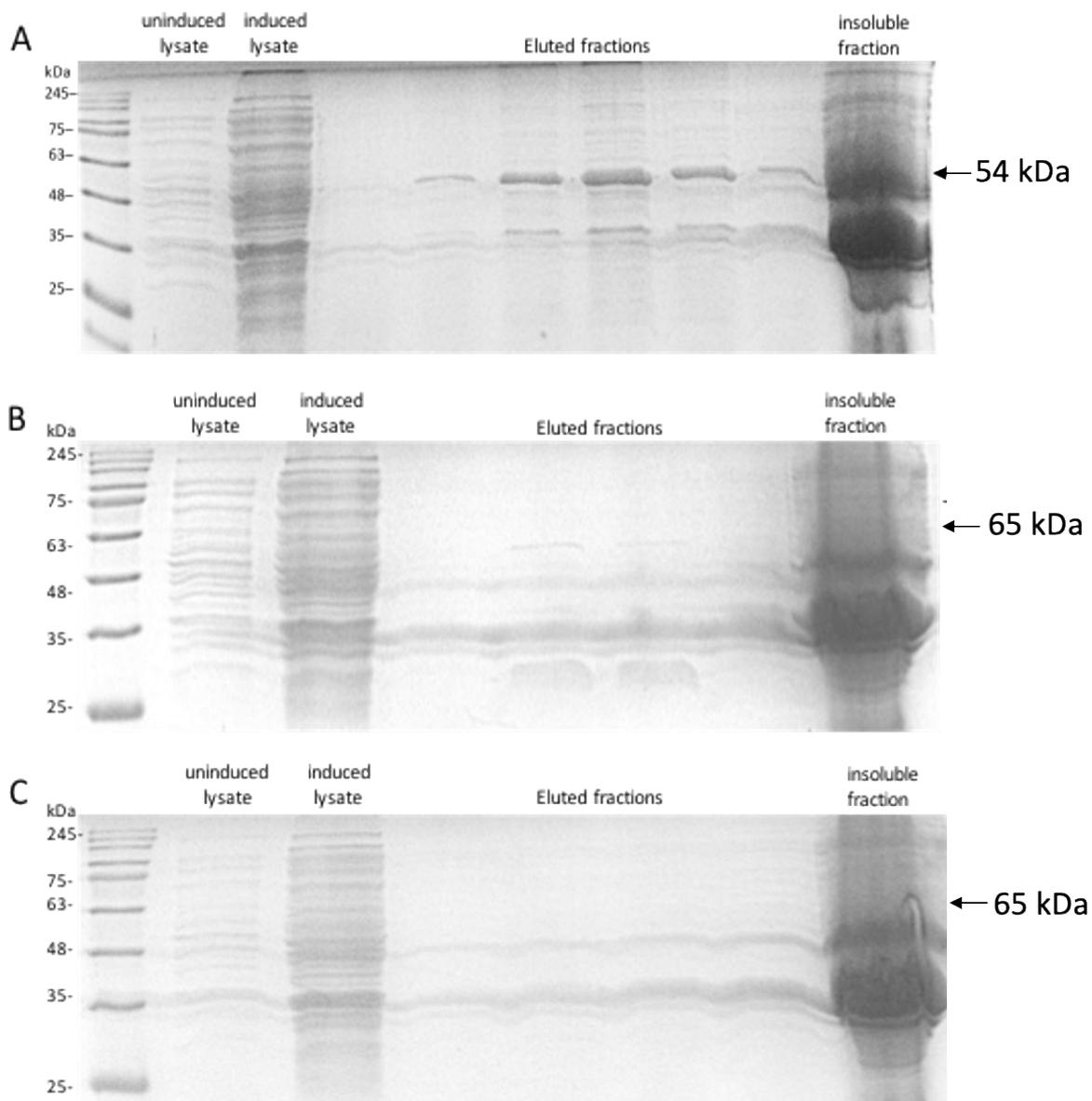


Figure 4.8. SDS-PAGE of recombinant GST-tagged protein expression at 20°C. Recombinant plasmid cultures were induced for protein expression in *E. coli* BL21 with 0.5 mM IPTG at 20°C overnight. Following sonication, clarified cell lysates were purified on GSTrap HP columns. Eluted fractions were analysed by SDS-PAGE and visualised after Coomassie blue staining. (A) GST-CD720 (B) GST-CD1062 (C) GST-CD1083.

Protein yield was optimized by scaling up the cultures from 100 ml to 500 ml, and then inducing at lower temperatures (20°C), which made no difference to the protein expression still undetectable in GST-CD1062 and GST-CD1083 fusion proteins as illustrated in Figure 4.8 (B/C). However, the lower temperatures decreased the appearance of the 35 kDa band in the soluble fractions of GST-CD720 protein. The protein purity of GST-CD720 improved to 53% compared to expression at 37°C. The insoluble bands persisted in all induced lysates and the 35 kDa bands persisted in the soluble fractions but appeared slightly reduced. There seems

to be a faint band slightly beneath the correct size for GST-CD1062 protein. In the GST-CD1062 protein gel, a non-target band around 26 kDa was present in the uninduced lysates, induced lysates and eluted fractions. Further improvements in recombinant expression were required to promote soluble protein production particularly for CD-1062 and CD-1083 transglutaminase-like genes.

Recombinant expression plasmids were transformed into *E. coli* BL21 cells and induced for protein expression as previously with modifications on temperature and IPTG concentrations. Briefly a 1:50 dilution of overnight transformant cultures was performed in fresh LB broth and grown for approximately 2 hours on a shaker (180 rpm) at 37°C to an OD₆₀₀ of 0.3. Next the cells were equilibrated at 20°C for 45 minutes to reach an OD₆₀₀ of 0.5. At this stage to induce for protein expression, 0.1 mM IPTG was introduced into the culture and further incubated overnight at 20°C. Cells were harvested by centrifugation and after lysis, clarified lysates were purified on GSTrap HP columns. Prior to determining protein concentrations and further characterization assays, eluted recombinant protein fractions were dialyzed overnight at 4°C in 50 mM Tris-HCL, pH 8 with few buffer changes or desalted on columns. The next optimisation procedure involved expression at lower temperatures and IPTG concentrations as incentive to promote soluble protein production.

4.3.4.2.2 Effect of temperature and IPTG concentration on recombinant expression of GST-tagged proteins in *E. coli* BL21

After protein induction at room temperature with 0.1 mM IPTG and affinity purification, cell lysates, supernatant fractions, flowthrough, washes and eluted fractions were analysed by SDS-PAGE.

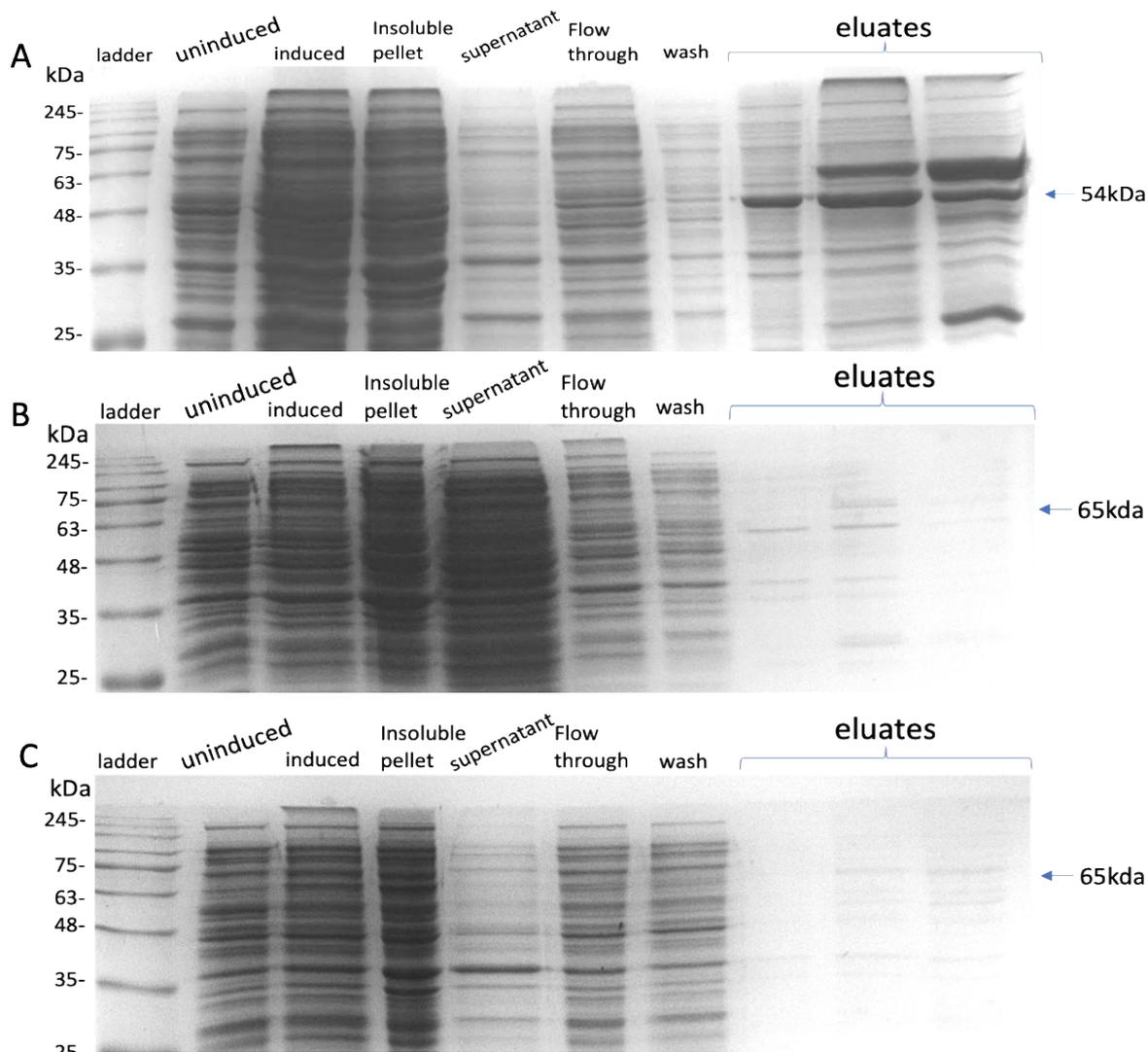


Figure 4.9. SDS-PAGE of recombinant GST-tagged proteins expressed at low temperature and IPTG concentration. Induction of recombinant protein expression in *E. coli* BL21 with 0.1 mM IPTG was performed overnight at 20°C. Cells were harvested by centrifugation, lysed and clarified lysates were purified by affinity chromatography on GStrap HP columns. Eluted fractions were analysed by SDS-PAGE and visualised after Coomassie blue staining. (A) GST-CD720 (B) GST-CD1062 (C) GST-CD1083.

As shown in Figure 4.9, following expression of GST-CD720 at room temperature with lower IPTG, an improvement in solubility plus yield was observed with the protein expressed at the target size. The protein purity of GST-CD720 was 57% which is an improvement compared to expression at 37°C with 0.5 mM IPTG (Figure 4.7). However extra bands can be seen above the target protein and around 26 kDa. The expected size for GST-CD1062 and GST-CD1083 fusion proteins is 64.8 kDa and 65 kDa respectively. For the GST-CD1062 and GST-CD1083 proteins, apparent bands can be seen around the target size in the eluted fractions, even though they are very faint. Additionally, extra bands are present around the target size. In all three induced lysates, non-specific 35 kDa and 26 kDa bands are present in uninduced and

induced lysates, insoluble fractions, cell free culture supernatants and eluted fractions. The non-target bands in the various protein fractions could be explained by the presence of rare codons which can lead to premature termination of translation and production of truncated products. Therefore, protein expression was optimised for codon usage using *E. coli* Rosetta 2 cells.

4.3.4.2.3 Effect of *E. coli* Rosetta 2 cells on recombinant expression of GST-tagged proteins

Rare codons were found in *C. difficile* 630 transglutaminase-like genes after analysis of the gene sequences on the GenScript rare codon analysis tool. To improve the translation efficiency, transglutaminase genes were expressed in *E. coli* Rosetta 2 which incorporates a plasmid containing 7 tRNA genes to complement the deficiency in rare codons that could affect full length protein expression. After protein expression in *E. coli* Rosetta 2 and affinity purification, cell lysates and eluted fractions were analysed by SDS-PAGE (Figure 4.10).

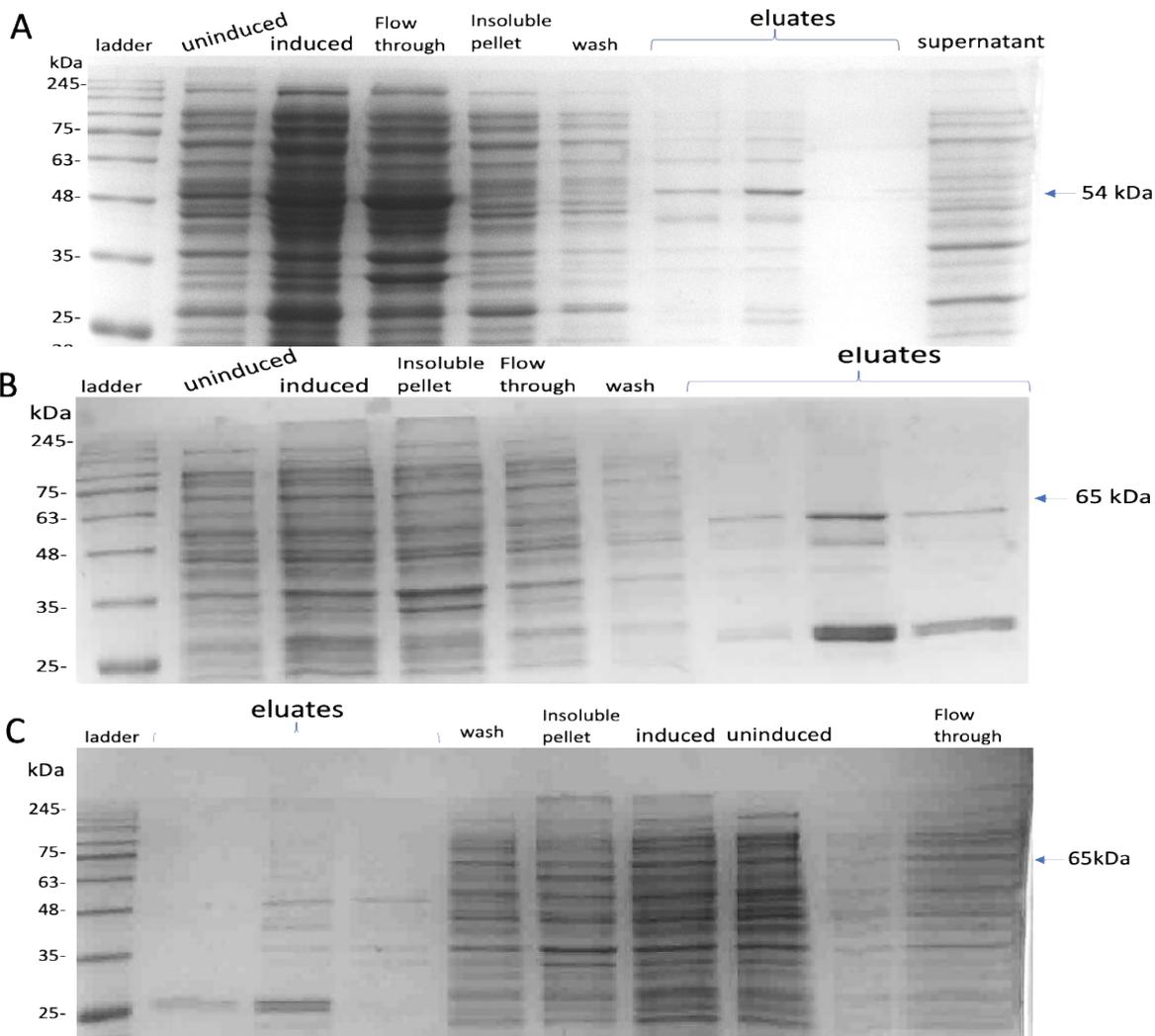


Figure 4.10. SDS-PAGE showing recombinant GST-tagged proteins expressed in *E. coli* Rosetta 2 cells. Induction of recombinant protein expression in *E. coli* Rosetta 2 with 0.1 mM IPTG was performed overnight at 20°C. Cells were harvested by centrifugation, lysed and clarified lysates were purified by affinity chromatography on GSTrap HP columns. Eluted fractions were separated by SDS-PAGE and visualised after Coomassie blue staining. (A) GST-CD720 (B) GST-CD1062 (C) GST-CD1083.

Following expression in Rosetta 2 cells, major bands around 26 kDa were observed in eluted fractions in all recombinant proteins particularly for GST-CD1062 and GST-CD1083. For the GST-CD720 protein, Figure 4.10 shows that expression in Rosetta 2 decreased yields compared to expression in BL21 (Figure 4.9). The target protein band was still present with a few extra bands in the eluted fractions and protein purity was 51%. The protein expression for GST-CD1062 and GST-CD1083 remained unsuccessful. This confirms that the faint band observed in eluted fractions of GST-CD1062 in Figure 4.8 was not the target band. Subsequently smaller peptide tags were used as a strategy to promote proper protein folding and thus soluble expression. To achieve this, transglutaminase inserts were cloned into

pRSETc expression vectors prior to protein expression and purification on nickel charged columns. The pRSETc vector utilises a T7 promoter system that allows expression of proteins with N-terminal 6xHis tags. Additionally, pRSETc contains an enterokinase recognition site for cleavage of the fusion tag after expression.

4.3.5 Cloning and Recombinant Expression of *C. difficile* 630 putative transglutaminase proteins in pRSETc

4.3.5.1 Cloning of transglutaminase genes in pRSETc expression vector

The transglutaminase inserts obtained from double digestions of pGEX3X recombinant constructs with *Bam*HI/*Eco*RI were inserted within the same sites in pRSETc. After transformation of ligation products into DH5 α cells, plasmids were isolated from overnight cultures of transformants. Restriction enzyme digestion of pRSETc recombinant plasmids with *Bam*HI/*Eco*RI were performed to confirm the presence of transglutaminase inserts at the correct sizes as shown on the agarose gel in Figure 4.11.

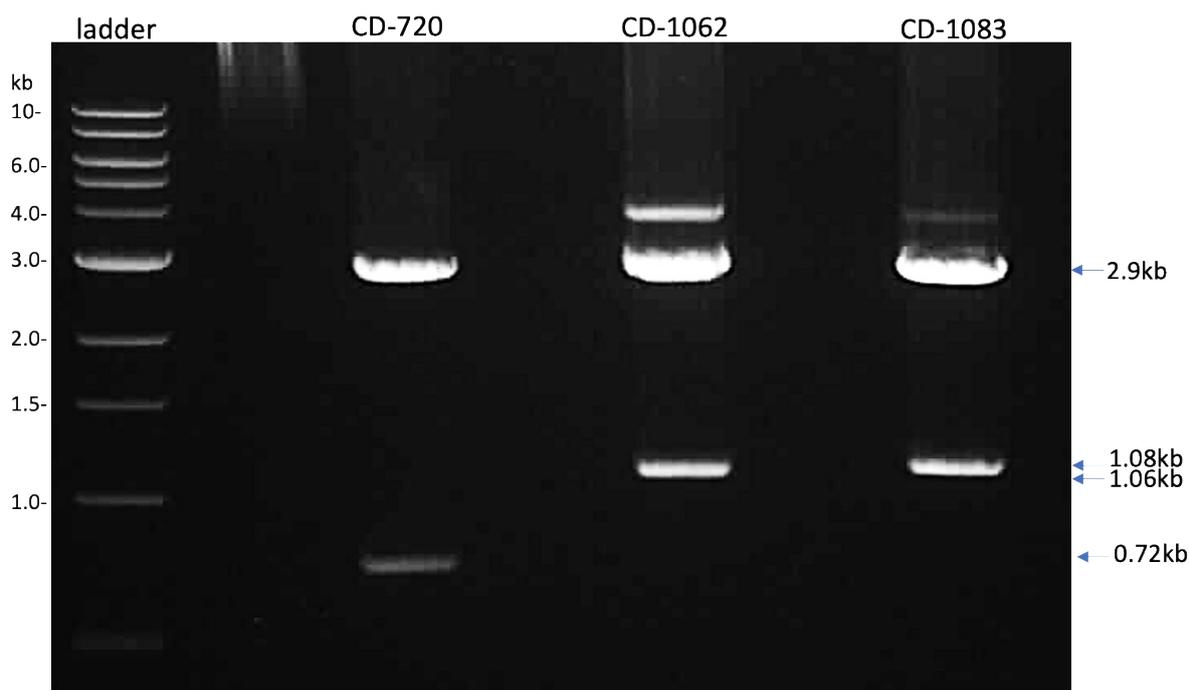


Figure 4.11. Restriction digestion of recombinant plasmids after sub-cloning in pRSETc. The recombinant plasmids resulting from ligation of transglutaminase inserts into pRSETc expression vector were digested with *Bam*HI and *Eco*RI. The success of cloning was confirmed after recovery of inserts following digestion as shown on agarose gels, and the recombinant plasmids were sequenced to verify the integrity of the whole coding sequence. lane L: DNA ladder. The lanes indicate respectively CD-720, CD-1062 and CD-1083 inserts. Top bands: empty plasmids, bottom bands: **transglutaminase** inserts.

The agarose gel in Figure 4.11 illustrates recovery of cloned inserts following *Bam*HI-*Eco*RI digestion of recombinant expression plasmids. The inserts migrated to their expected sizes which confirmed the success of cloning. Some amount of the CD-1062 and CD-1083 clones was partially digested as indicated by the bands above the 2.9 kb plasmid, even though digestion was mostly complete releasing corresponding transglutaminase inserts. Nucleotide sequencing on different clones was performed to select those containing transglutaminase genes in frame with the T7 promoter. Clones with transglutaminase in the correct orientation were transformed into *E. coli* BL21 and Rosetta 2 cells. After protein expression and purification, different fractions were analysed by SDS-PAGE to check for the presence of target proteins at the predicted sizes.

4.3.5.2 Expression of transglutaminase-like proteins as N-terminal 6xHis-tag fusions

The pRSETc vector harbours a T7 promoter which directs the production of cloned transglutaminase genes. It is an IPTG inducible expression system that contains an ampicillin resistance gene for selection, a tag cleavage site, and multiple cloning site containing several restriction sites. The pRSETc construct for recombinant transglutaminase contains an N-terminal 6xHis-tag to enable protein purification by metal affinity chromatography on nickel-charged columns. Expression of transglutaminase-like proteins with His-tags was first performed at 37°C, and eluted fractions were analysed by SDS-PAGE as shown in Figure 4.12.

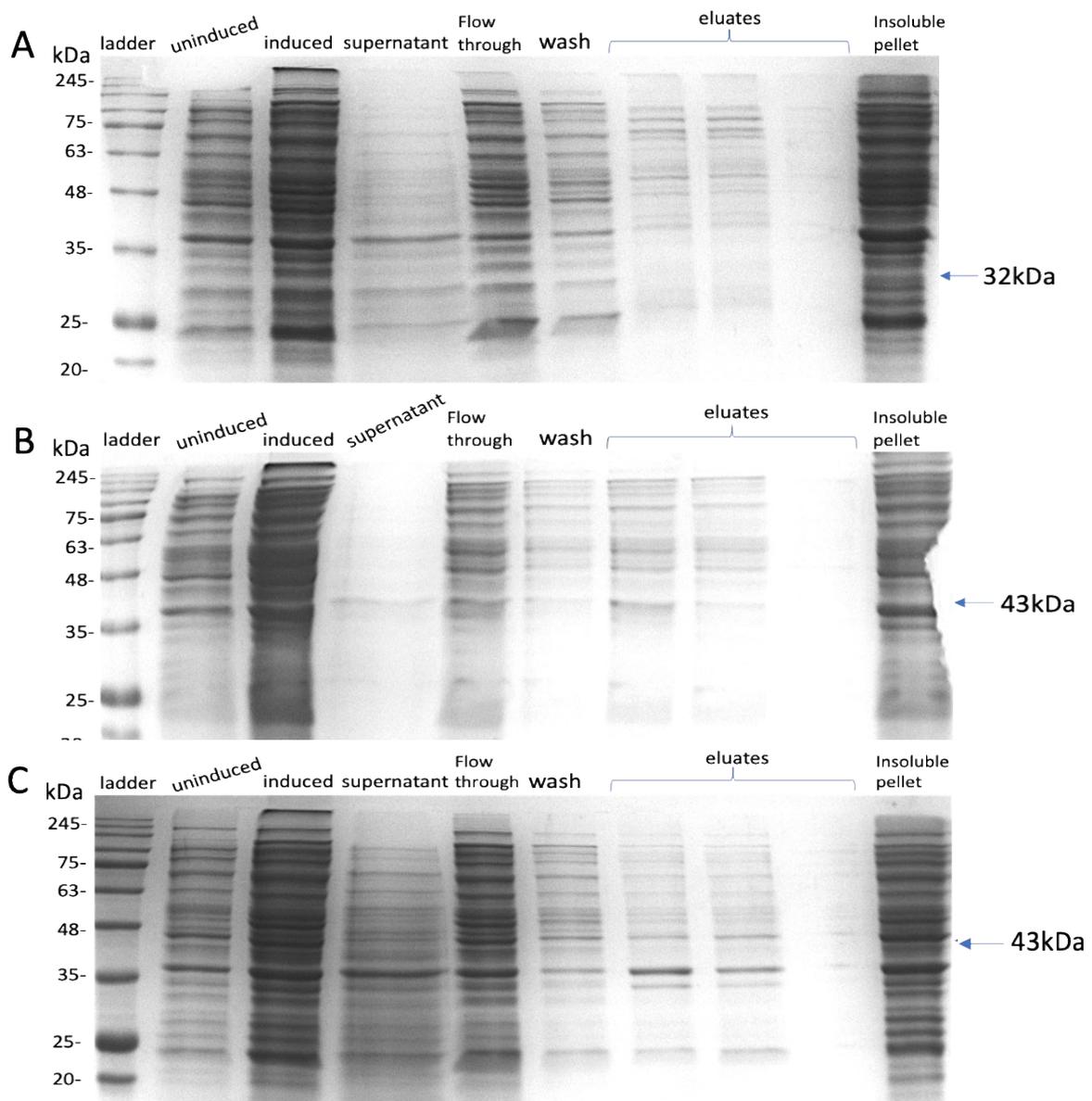


Figure 4.12. SDS-PAGE of recombinant His-tagged proteins expressed at 37°C. Induction of recombinant protein expression in *E. coli* BL21 with 0.1 mM IPTG was performed at 37°C for 3 hours. Cells were harvested by centrifugation, lysed and clarified lysates were purified by affinity chromatography on HisTrap HP columns. Proteins were bound onto columns with 5 mM Imidazole and impurities washed with 40 mM imidazole, followed by gradient elution with 50-250 mM imidazole. Eluted fractions were separated by SDS-PAGE and visualised after Coomassie blue staining. (A) 6xHis-CD720 (B) 6xHis-CD1062 (C) 6xHis-CD1083.

As observed in Figure 4.12 protein expression with His-tags did not seem to be successful for all induced lysates. A band slightly above 35 kDa is observed in all fractions (induced, uninduced lysates, culture supernatants, insoluble fractions) for the three transglutaminase-like proteins. In all expressed proteins, eluted fractions contain many non-target bands. However, for the 6xHis-CD1083 fusion protein there seems to be minimal protein expression indicated by a band around the target size in eluted fractions even though other extra bands

are present. Expression of recombinant proteins was then repeated at lower temperatures (20°C) in attempts to promote solubility of target proteins (Figure 4.13).

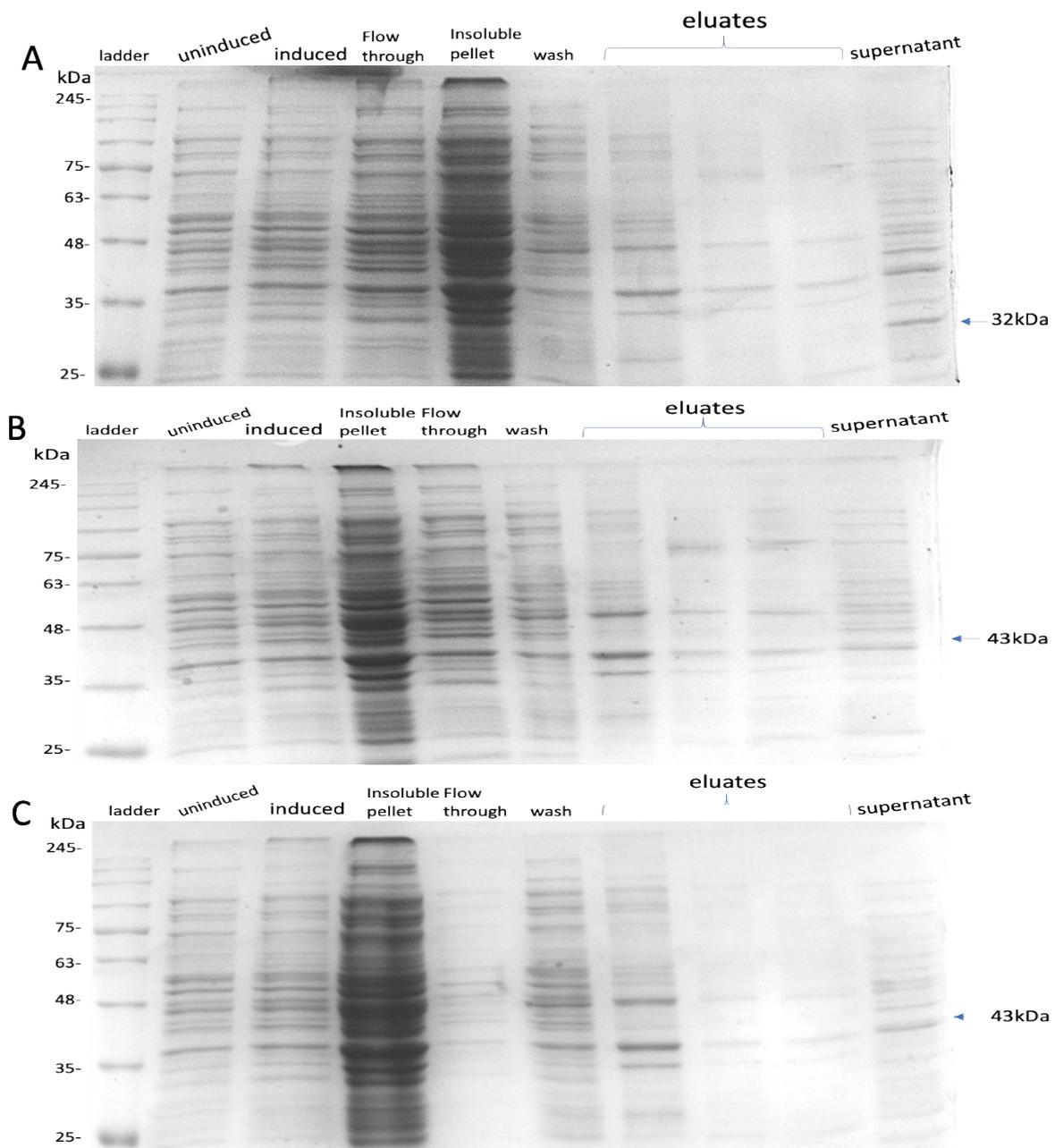


Figure 4.13. SDS-PAGE of recombinant His-tagged proteins expressed in *E. coli* BL21 at 20°C. Induction of recombinant protein expression in *E. coli* BL21 with 0.1 mM IPTG was performed at 20°C overnight. Cells were harvested by centrifugation, lysed and clarified lysates were purified by affinity chromatography on HisTrap HP columns. Proteins were bound onto columns with 5 mM imidazole and impurities washed with 40 mM imidazole, followed by gradient elution with 50-250 mM imidazole. Eluted fractions were analysed by SDS-PAGE and visualised after Coomassie blue staining. (A) 6xHis-CD720 (B) 6xHis-CD1062 (C) 6xHis-CD1083.

From Figure 4.13 no difference in protein expression was observed following induction at lower temperatures in *E. coli* BL21 as target bands were absent in eluted protein fractions.

Recombinant expression was repeated in Rosetta cells to enhance soluble expression and generation of full-length target proteins.

4.3.5.2.1 Recombinant expression of His-tagged proteins in *E. coli* Rosetta

Expression of recombinant transglutaminase-like proteins was first performed at 37°C as shown in Figure 4.14.

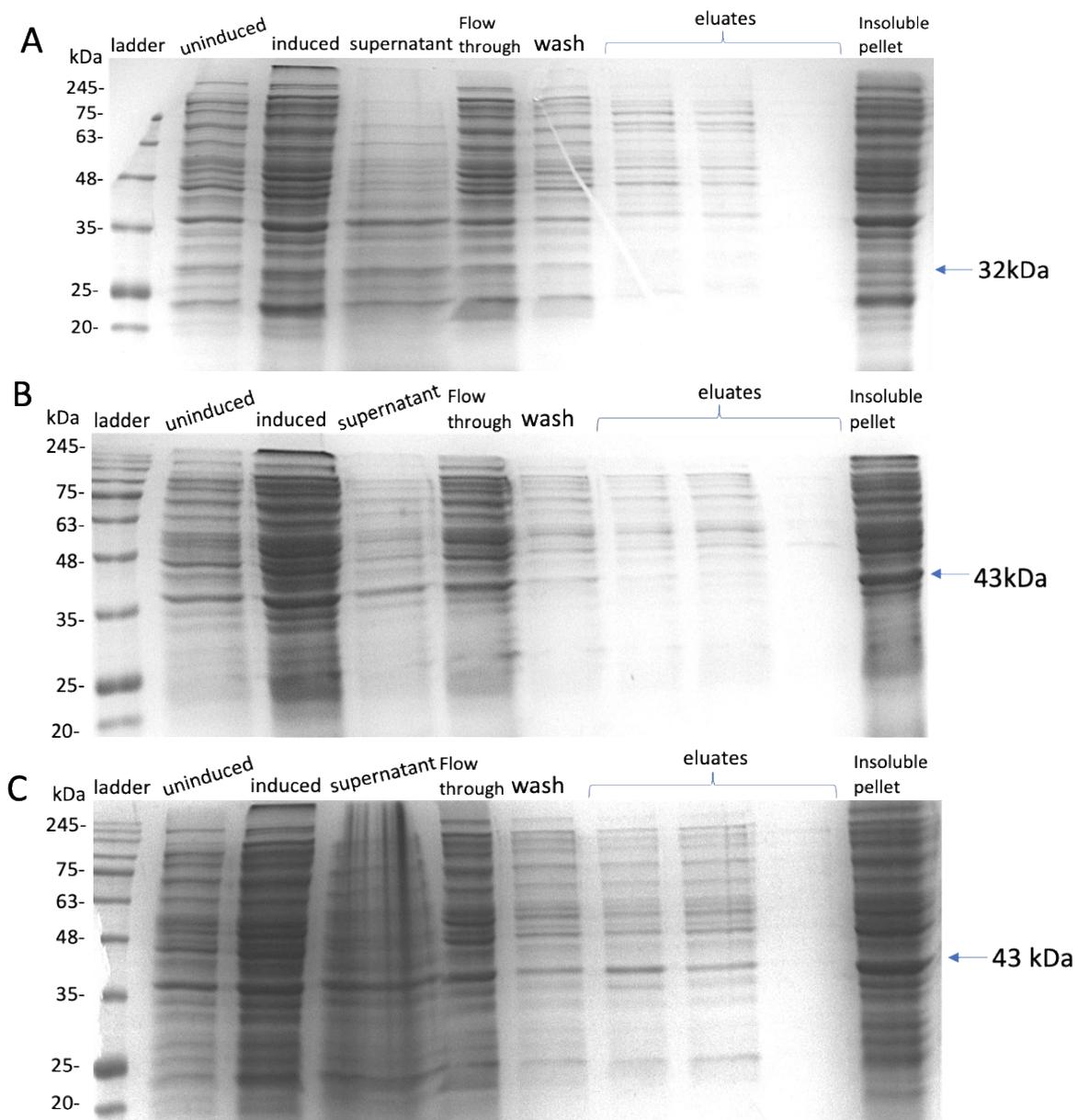


Figure 4.14. SDS-PAGE of recombinant His-tagged proteins expressed in *E. coli* Rosetta 2 at 37°C. Induction of recombinant protein expression in *E. coli* Rosetta 2 with 0.1 mM IPTG was performed at 37°C for 3 hours. Cells were harvested by centrifugation, lysed and clarified lysates were purified by affinity chromatography on HisTrap HP columns. Proteins were bound onto columns with 5 mM imidazole and impurities washed with 40 mM imidazole, followed by gradient elution with 50-250 mM imidazole. Eluted fractions were analysed by SDS-PAGE and visualised after Coomassie blue staining. (A) 6XHis-CD720 (B) 6XHis-CD1062 (C) 6XHis-CD1083.

As shown in Figure 4.14 the same pattern was observed following recombinant protein expression in Rosetta 2 cells. No proteins were observed at the target size except for 6xHis-CD1083, even though a lot of impurities are present in the eluted fractions. Again, as seen

with BL21 expression, a band slightly above 35 kDa is observed in all fractions (induced, uninduced lysates, culture supernatants, insoluble fractions) for the three recombinantly expressed proteins.

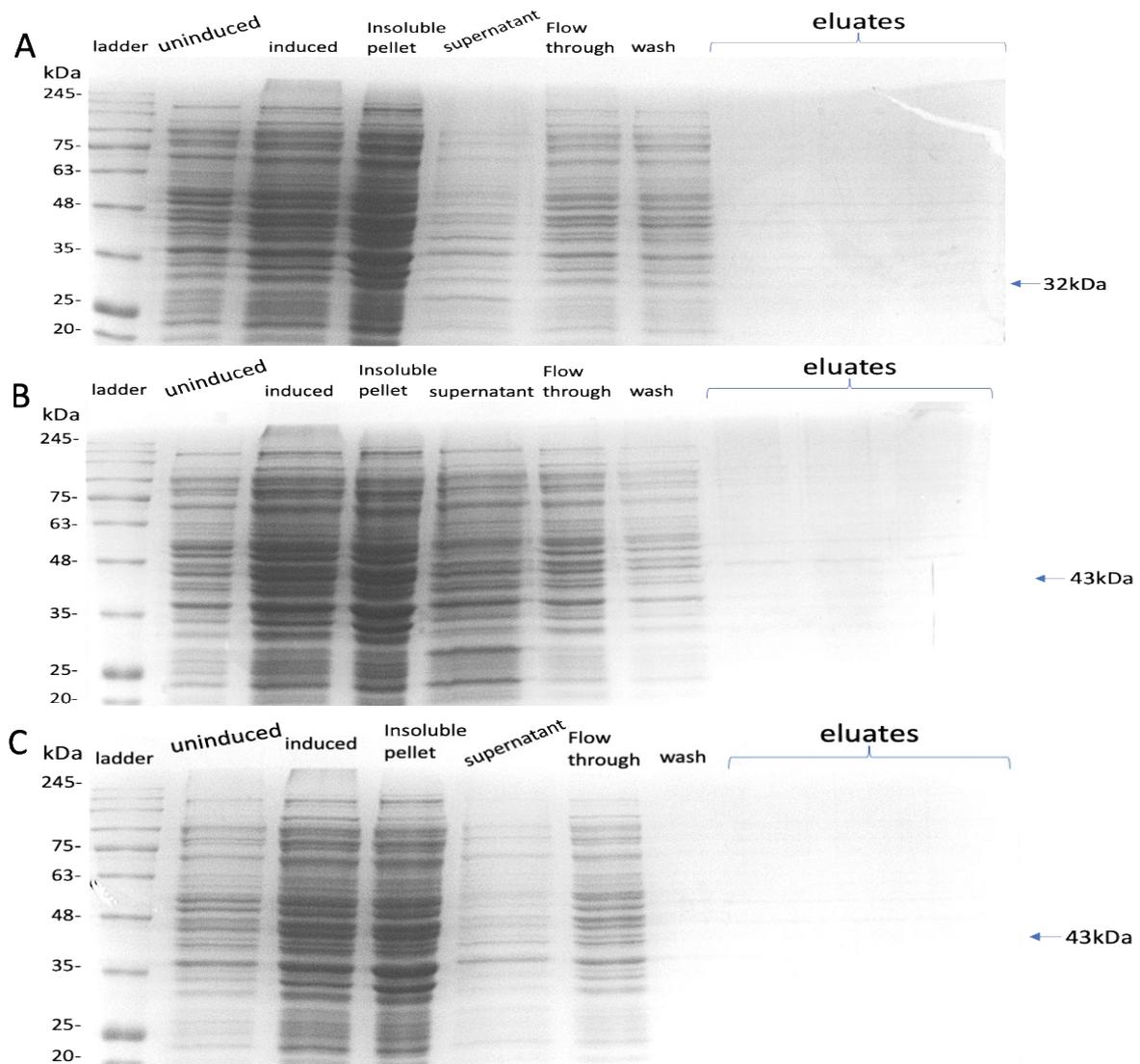


Figure 4.15. SDS-PAGE of recombinant His-tagged proteins expressed in *E. coli* Rosetta 2 at 20°C. Induction of recombinant protein expression in *E. coli* Rosetta 2 with 0.1 mM IPTG was performed at 20°C overnight. Cells were harvested by centrifugation, lysed and clarified lysates were purified by affinity chromatography on HisTrap HP columns. Proteins were bound onto columns with 5 mM imidazole and impurities washed with 40 mM imidazole, followed by gradient elution with 50–250 mM imidazole. Eluted fractions were analysed by SDS-PAGE and visualised after Coomassie blue staining. (A) 6XHis-CD720 (B) 6XHis-CD1062 (C) 6XHis-CD1083.

As shown in Figure 4.15 expression at lower temperatures (20°C) did not make a difference as target His-tagged protein bands were not clearly detectable. Western blot analysis on

purified samples was performed to distinguish His-tagged fusion proteins from truncated proteins and purification impurities.

4.3.5.2.2 Expression with smaller peptide-tags does not improve solubility of CD-1062 and CD-1083 transglutaminase-like proteins

Western blots were performed on all samples following recombinant protein expression and purification on nickel charged columns to confirm the presence of target His-tagged protein bands.

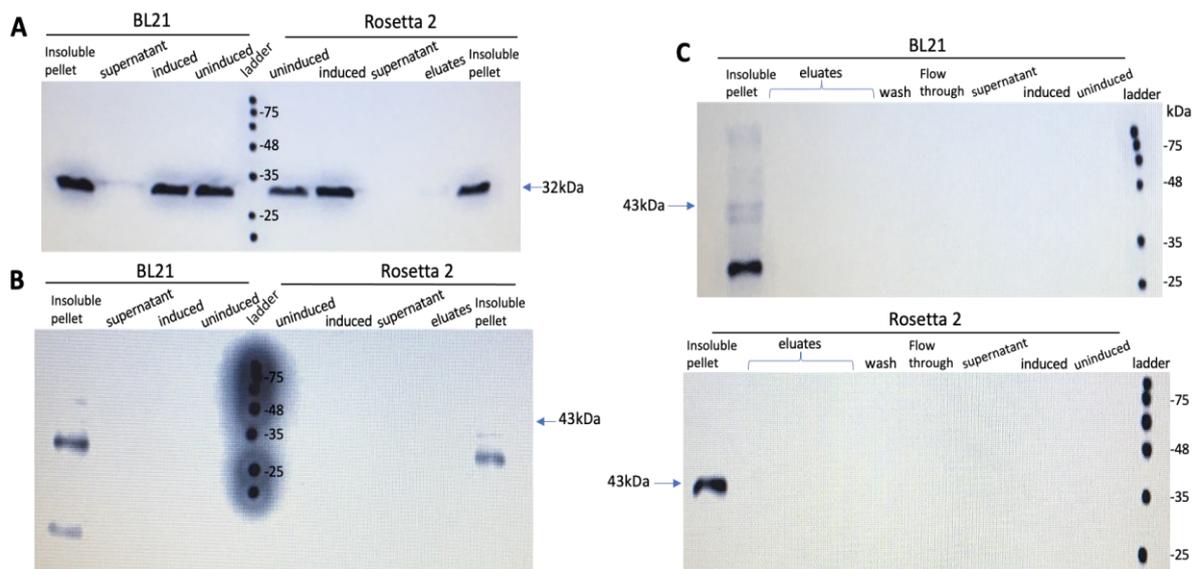


Figure 4.16. Western blot of recombinant His-tagged proteins expressed in *E. coli* BL21 and Rosetta 2 cells. Recombinant expression was induced in *E. coli* BL21 and Rosetta 2 with 0.1 mM IPTG at 20°C overnight and clarified lysates were purified by affinity chromatography on HisTrap HP columns. After sample separation by electrophoresis, protein transfer onto nitrocellulose membranes was completed as described in methods section. Recombinant His-tagged proteins were detected by incubation with mouse anti-His antibody and goat anti-mouse IgG secondary antibody respectively. After washing, the blots were developed with enhanced chemiluminescence reagents and images recorded on a Western blot imager. (Uninduced and induced represent the clarified lysates from uninduced and IPTG-induced cultures, supernatant corresponds to supernatants from induced cultures, and eluates represent the eluted fractions after purification of clarified induced lysates. Insoluble pellet represent inclusion bodies). (A) 6xHis-CD720 (B) 6xHis-CD1062 (C) 6xHis-CD1083.

The Western blot images confirm soluble protein expression only for the 6xHis-CD720 protein as indicated by target bands in induced lysates shown in Figure 4.16. No protein bands could be detected in eluted fractions. The molecular mass of recombinant 6xHis-CD720 matches the predicted size of 32 kDa. Additionally, protein bands could be detected in uninduced lysates, and some amount of protein remained insoluble. This is true using both BL21 and Rosetta cells for CD-720 protein expression. As illustrated in Figure 4.16, the 6xHis-CD1062

and 6xHis-CD1083 proteins were expressed entirely as inclusion bodies indicated by the presence of bands in the insoluble fractions, some of which appear below the target size around 35 kDa. The expression of CD-1062 and CD-1083 as full-length proteins remained futile despite several optimisation strategies. The presence of signal peptides in these two putative transglutaminase-like genes may have prevented soluble intracellular expression (Rosano and Ceccarelli, 2014). Consequently, recombinant expression of GST-CD1062 and GST-CD1083 as mature proteins was implemented as a strategy to promote soluble expression.

4.4 Recombinant Expression And Purification Of Mature *C. difficile* 630 transglutaminase-Like Proteins

Of the three transglutaminase genes in *C. difficile* 630, CD-1062 and CD-1083 were found to contain signal peptides (SP). Recombinant constructs comprising only the mature region of the transglutaminase coding sequence were obtained by PCR and possess an N-terminal GST tag for purification on GSTrap columns. A strategy to express both recombinant proteins in the mature form (i.e., without signal peptides) was implemented which consisted initially of PCR reactions.

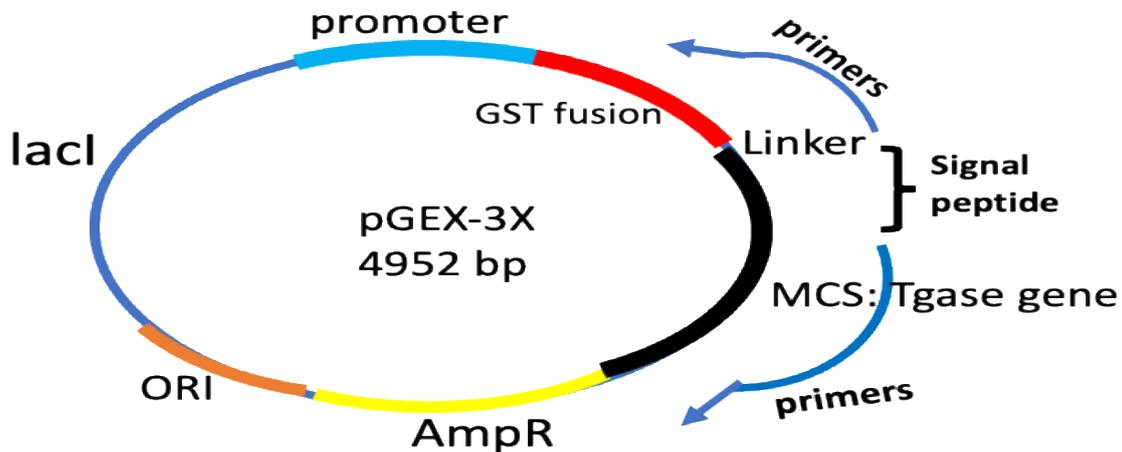


Figure 4.17. Diagram of recombinant pGEX3X plasmid with primers for signal peptide cleavage. The arrows pointing outwards show the direction of PCR amplification away from the N-terminal signal peptide region to produce a GST-fusion product without signal peptides.

The diagram in Figure 4.17 shows the orientation of the primers designed for cleavage of signal peptides from the N-terminus of transglutaminase-like genes. The primers were oriented to amplify in the outwards direction thus by-passing the SP sequence.

4.4.1 PCR amplification of mature regions in *C. difficile* 630 transglutaminase genes

The pGEX3X recombinant constructs for CD-1062 and CD-1083 Tgase-like genes were modified to exclude the SP regions by PCR amplification with specific primers listed on Table 2.1.

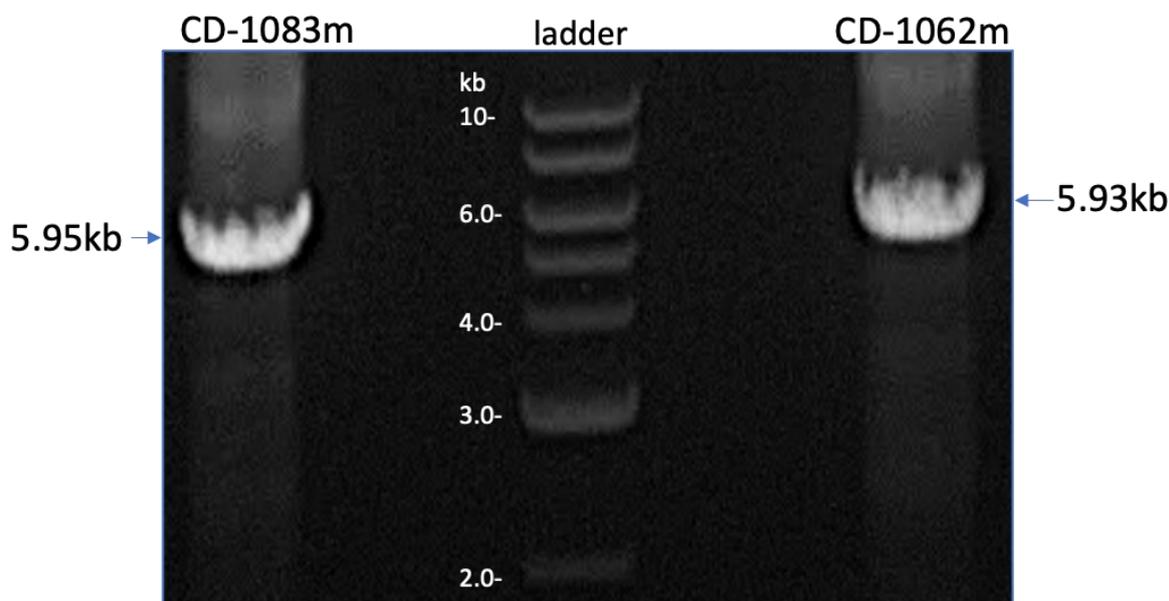


Figure 4.18. PCR amplifications of mature regions in *C. difficile* 630 transglutaminase genes. For amplification of recombinant transglutaminases in pGEX-3X plasmids for the deletion of signal peptide regions, the KOD hot start DNA polymerase kit was used as described in Materials and Methods. Recombinant pGEX-3X plasmids were used as template and success of PCR amplifications was determined by analysis of PCR products on agarose gels. Lane L: DNA ladder, lane A: mature CD-1083 plasmid, lane B: mature CD-1062 plasmid.

Using recombinant pGEX3X plasmids as template, the primer sets successfully amplified the target regions indicated by DNA bands around the expected plasmid size in Figure 4.18. For simplicity the new constructs were renamed GST-CD1062m and GST-CD1083m, where 'm' signifies mature protein. The resulting PCR products were re-ligated and transformed into *E. coli* DH5 α cells. Positive clones were selected on ampicillin plates and plasmids were isolated from overnight cultures of transformants.

4.4.2 Plasmid preparations of recombinant mature *C. difficile* 630 transglutaminase constructs

The success of signal peptide deletion and re-ligation of recombinant plasmids was verified by agarose gel electrophoresis as shown below.

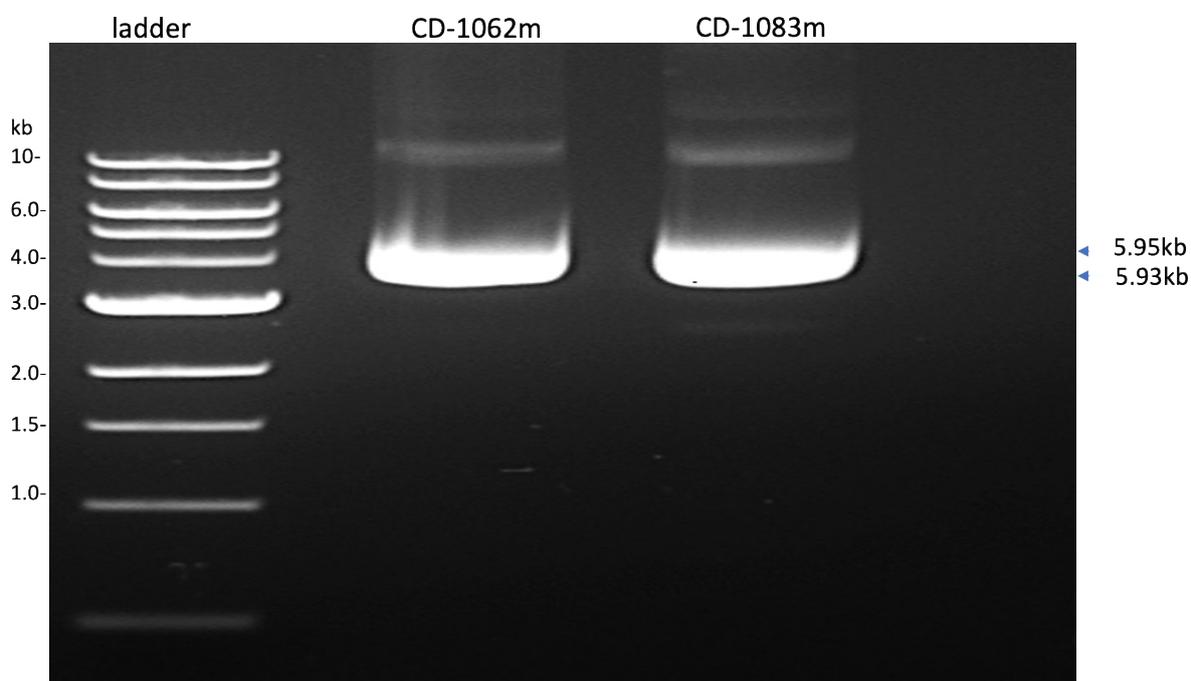


Figure 4.19. Plasmid preparations of recombinant mature *C. difficile* 630 transglutaminase constructs. Recombinant mature CD-1062 and CD-1083 pGEX-3X constructs were transformed into *E. coli* DH5 α cells. The recombinant plasmids were isolated on spin columns from overnight cultures of transformants in LB broth following the alkali lysis method described in Materials and methods. The success of signal peptide cleavage and transformation was verified by AGE on plasmid DNA preps. lane 1: DNA ladder. The lanes indicate respectively recombinant CD-1062m and CD-1083m clones with cleaved signal peptides.

Agarose gel electrophoresis of plasmids isolated from overnight cultures of transformants reveal success of transformation as shown by plasmids at the expected target size in Figure 4.19. Nucleotide sequencing of recombinant plasmids with pGEX-3X F/R universal primers was carried out to verify that the clones contained 100% match to mature putative transglutaminase gene fusions. The new pGEX3X construct still contains a GST tag for purification on glutathione Sepharose columns. The sizes of GST-CD1062m and GST-CD1083m recombinant proteins was 62 kDa and 63 kDa respectively. Recombinant protein expression was performed in *E. coli* BL21 and Rosetta cells, and purified proteins were analysed by SDS-PAGE.

4.4.3 Removal of signal peptides promotes soluble expression of *C. difficile* 630 transglutaminases

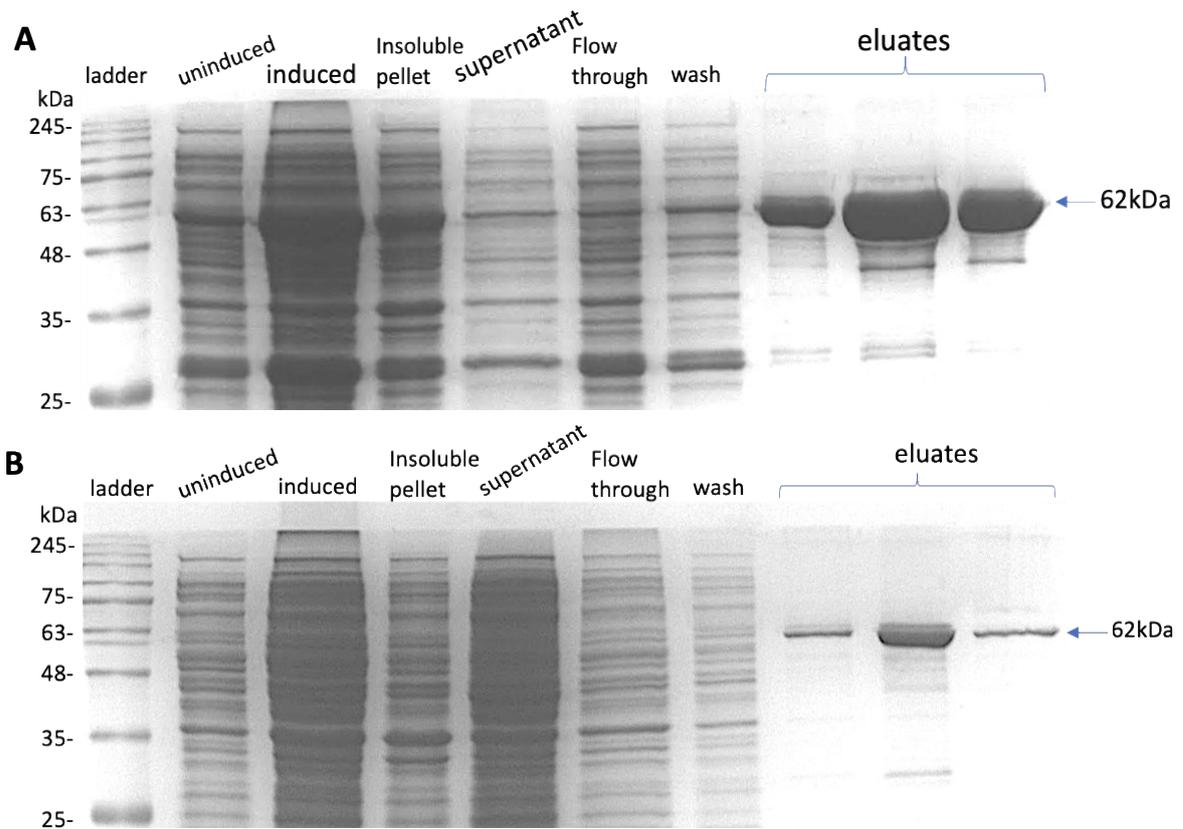


Figure 4.20. SDS-PAGE analysis of recombinant GST-tagged CD-1062m proteins. Recombinant pGEX3X plasmids were transformed into *E. coli* BL21 and Rosetta 2 cells and production of transglutaminase-like proteins was induced with 0.1 mM IPTG overnight at 20°C. Cells were harvested by centrifugation, lysed and clarified lysates were purified by affinity chromatography on GSTrap HP columns. Eluted fractions were analysed by SDS-PAGE and visualised after Coomassie blue staining. A: Expression in BL21, B: expression in Rosetta 2.

The results in Figure 4.20 confirms that the mature sequence of CD-1062 transglutaminase gene within *C. difficile* 630 has been successfully cloned and expressed as soluble GST-tagged proteins in both *E. coli* BL21 and Rosetta 2 cells. The purity for GST-CD1062m proteins after expression in *E. coli* BL21 and Rosetta 2 cells was about 96% and 88% respectively. Similarly, expression of CD-1083m in both *E. coli* strains was followed by purification and SDS-PAGE analysis of eluted fractions as shown in Figure 4.21.

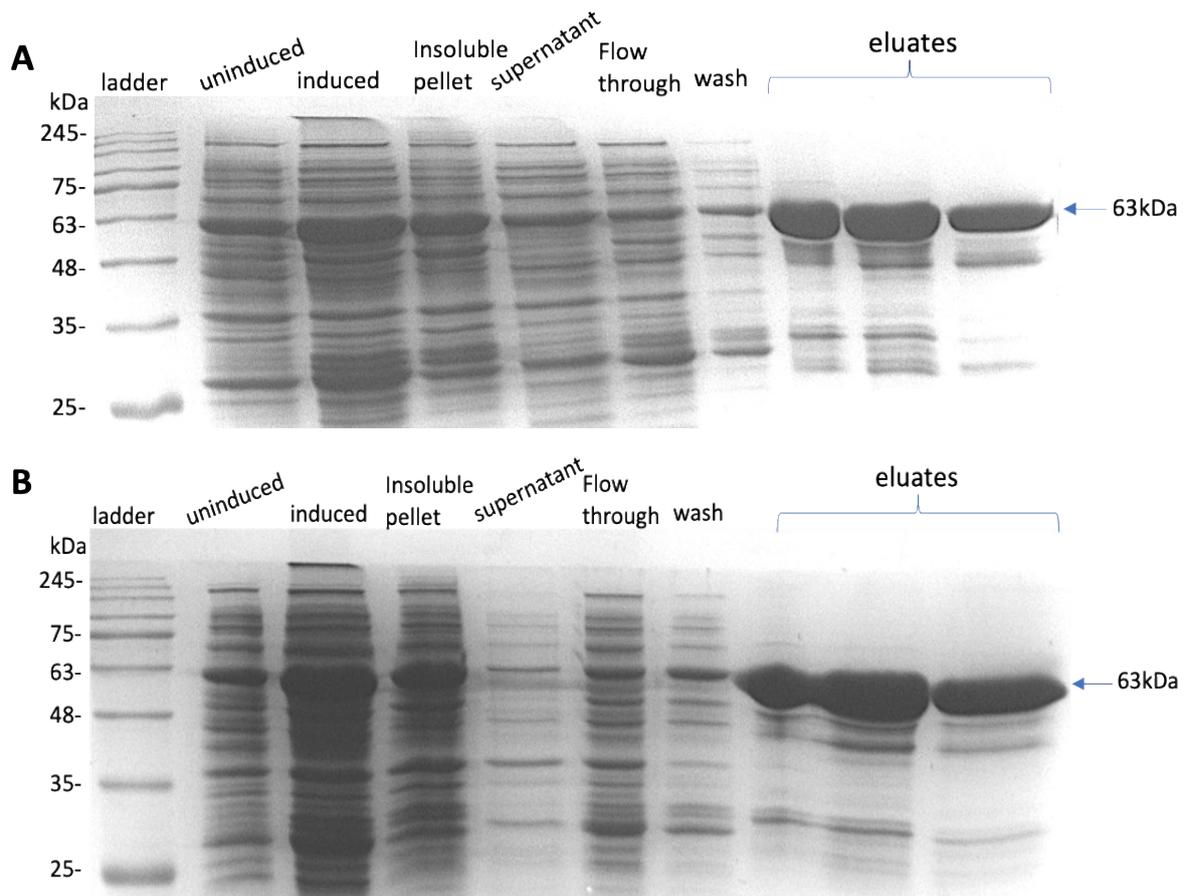


Figure 4.21. SDS-PAGE analysis of recombinant GST-tagged CD-1083m proteins. Recombinant pGEX3X plasmids were transformed into *E. coli* BL21 and Rosetta 2 cells and production of transglutaminase-like proteins was induced with 0.1 mM IPTG overnight at 20°C. Cells were harvested by centrifugation, lysed and clarified lysates were purified by affinity chromatography on GSTrap HP columns. Eluted fractions were analysed by SDS-PAGE and visualised after Coomassie blue staining. A: Expression in BL21, B: expression in Rosetta 2.

Protein gels in Figure 4.21 confirm that the mature sequence of CD-1083 transglutaminase gene within *C. difficile* 630 has been successfully cloned and expressed as soluble GST-tagged proteins in *E. coli* BL21 and Rosetta 2 cells. Similar to CD-1062m expression, the target protein band is also quite prominent in other cell fractions (induced and uninduced lysates, insoluble fractions and culture supernatants). Recombinant expression in Rosetta 2 cells resulted in slightly purer protein compared to expression in BL21 cells. The protein purity for purified GST-CD1083m following expression in *E. coli* BL21 and Rosetta 2 cells was about 70% and 91% respectively. Therefore, similar protein purity was obtained following expression of GST-CD1062m (88%) and GST-CD1083m (91%) in Rosetta 2 cells. After soluble protein expression and purification, recombinant proteins were screened for transglutaminase and protease activities.

4.5 Characterization of recombinant transglutaminase-like proteins from *C. difficile* 630

As a reminder recombinant GST-tagged CD-720 protein was purified in section 4.3.4.1 (Figures 4.7 and 4.8), meanwhile recombinant GST-tagged CD-1062 and CD-1083 were purified in section 4.4.3 (4.20 and 4.21). Characterization studies would begin with recombinant CD-720 which was the first protein to be purified, followed by recombinant CD-1062 and CD-1083. Protein characterization was performed on recombinant fusion proteins to determine transglutaminase and protease activities. Prior to determining protein concentrations and further characterization assays, eluted recombinant protein fractions were dialyzed overnight at 4°C in 50 mM Tris-HCL, pH 8 with few buffer changes or desalted in Tris buffer.

4.5.1 Characterization of recombinant GST-tagged CD-720 protein

Recombinant GST-CD720 protein from *C. difficile* 630 was mostly soluble and protein characterization was performed on eluted fractions after removal of glutathione salt by desalting or dialysis. A transglutaminase activity assay was performed on desalted recombinant protein samples.

4.5.1.1 Recombinant GST-tagged CD-720 protein does not contain transglutaminase activity

The microbial transglutaminase activity assay was performed on recombinant GST-CD720 proteins to determine if it exhibited transglutaminase activity, using recombinant commercial MTG as positive control. No transglutaminase activity was detected in the purified fractions, whole cell lysates, and cell free culture supernatants for which $A_{450\text{nm}}$ was below Tris negative control levels ($A_{450\text{nm}}=0.05$). Therefore recombinant proteins were digested with Factor Xa to cleave the GST-tag prior to measuring activity. This was to verify whether the GST tag was affecting the protein's functionality.

4.5.1.2 Cleavage of fusion partners does not release active recombinant GST-CD720 protein

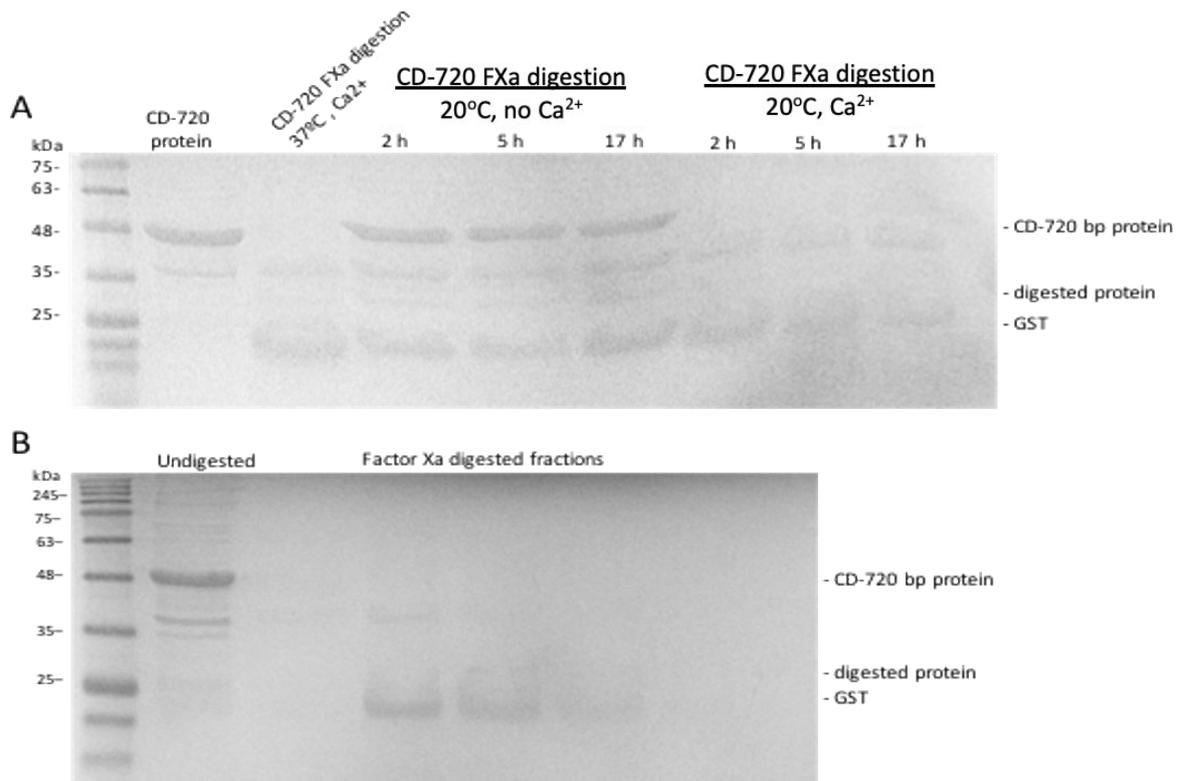


Figure 4.22. SDS-PAGE analysis of GST tag cleavage from recombinant CD-720 proteins. (A): Pilot study of GST tag cleavage from purified recombinant CD-720 proteins following Factor Xa addition in buffers with and without calcium and incubation overnight at room temperature or 37°C for 2 hours. Samples were collected at 2, 5 and 17 hours respectively and analysed by SDS-PAGE. (B): Cleavage of recombinant proteins with Factor Xa in 50 mM Tris, 150 mM NaCl, 1 mM CaCl₂, pH 7.5 for 4 hours at room temperature. Digested samples were separated by SDS-PAGE and visualised by Coomassie blue staining.

Enzymatic cleavage of the GST tag from recombinant CD-720 fusion proteins with Factor Xa (45 kDa) was successful as observed in Figure 4.22, releasing bands corresponding to the GST-tag (26 kDa) and/or cleaved protein (28 kDa). Transglutaminase activity was measured following tag cleavage and on-column purification as shown in Figure 4.23.

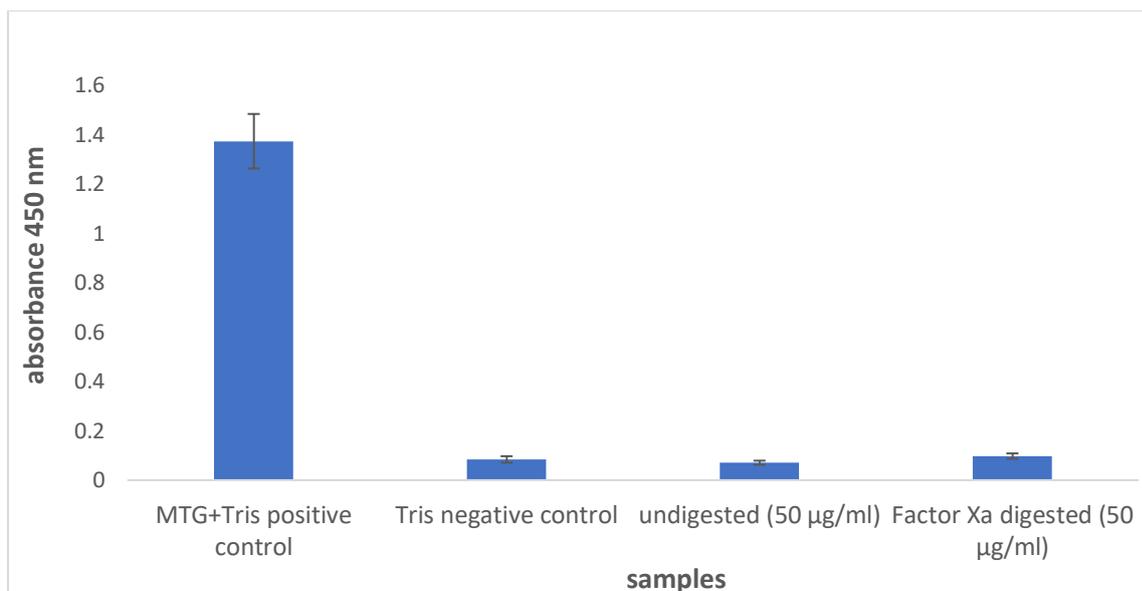


Figure 4.23. Transglutaminase activity assay of recombinant CD-720 proteins. Microbial transglutaminase activity assay as measured by the incorporation of biotin cadaverine (BTC) on to casein coated plates. Recombinant CD-720 proteins were treated with Factor Xa for 4 hours at room temperature and purified on GSTrap/heparin columns. Heparin columns served to remove Factor Xa from the protein digestion preparation. Transglutaminase activity following BTC addition to eluted fractions was measured by the oxidation of OPD by HRP labelled ExtrAvidin, with absorbance max of 450 nm. Recombinant commercial MTG served as positive control. Data are represented as mean \pm SD of two repeats, n=2.

The GST-tag was cleaved off the recombinant CD-720 protein and yet no transglutaminase activity could be detected as shown in Figure 4.23. The protein fractions exhibited activity comparable to negative control levels. Additionally, no differences were observed in transglutaminase activity of Factor Xa-digested and undigested proteins which were both negative. After unsuccessful detection of transglutaminase activity, a protease activity assay was performed on recombinant CD-720 protein given that homologs of microbial transglutaminase are proteases, and both possess a common fold domain.

4.5.1.3 Recombinant GST-tagged CD-720 protein does not contain proteolytic activity

Furthermore, proteolytic activity was measured in recombinant CD-720 protein with the Pierce fluorescence protease assay kit as described in 0 using trypsin standards as positive control.

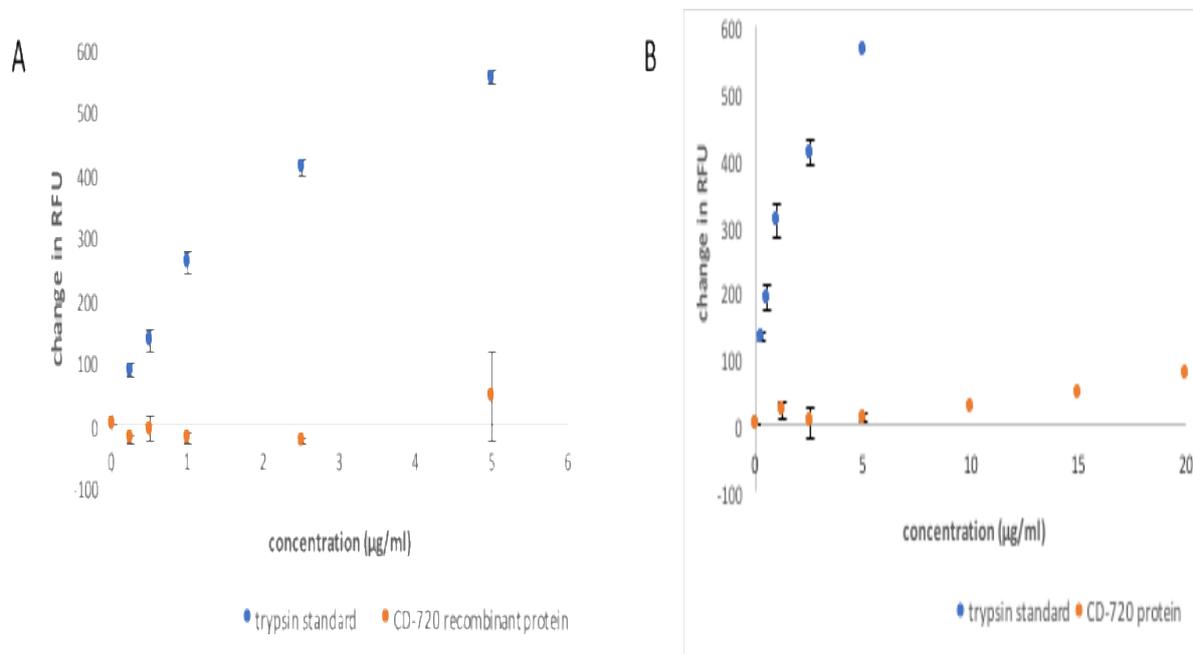


Figure 4.24. Protease activity assay of recombinant GST-CD720 proteins. Protease activity of recombinant GST-CD720 protein fractions (1-20 µg/ml) was verified at room temperature every 5 minutes for 1 hour by measuring digestion of FTC-casein at 485/538 nm, with reference to TPCK trypsin standard (0-10 µg/ml). The graphs represent activity measured at 30 minutes. The RFU measured from samples and controls were adjusted with the blank. Data are represented as mean±SD of two repeats, n=2. (A: samples and standards at 0-5 µg/ml; B: samples at 0-20 µg/ml).

With reference to trypsin positive control, the purified recombinant GST-CD720 protein did not show any proteolytic activity illustrated by the flatness of the curve even with higher concentrations of recombinant protein as shown in Figure 4.24. The protease activity of trypsin standards was dose-dependent and increased as a function of time. At higher trypsin concentrations (>1 µg/ml), most of the substrate was digested and the system was saturated, so even with more enzyme there was little or no change in RFU. A protease activity assay was repeated on purified recombinant proteins without a GST tag as shown in Figure 4.25. This was to verify whether the GST tag was affecting the protein's functionality.

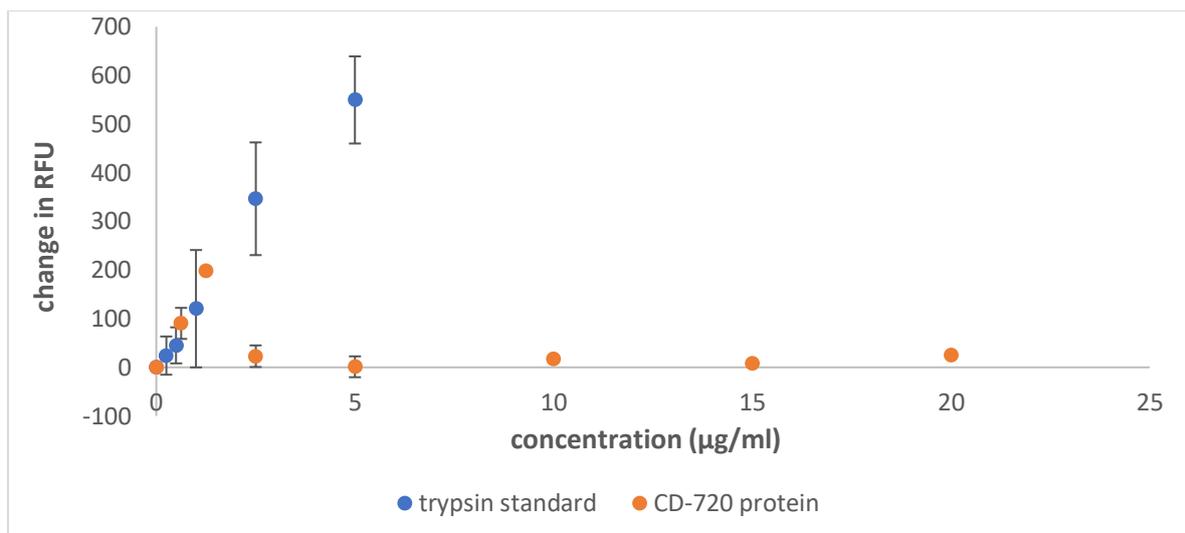


Figure 4.25. Protease activity assay of recombinant CD-720 protein with cleaved GST-tags. Recombinant GST-CD720 protein fractions were digested with Factor Xa for 4 hours at room temperature and purified on GSTrap/heparin columns. Protease activity of eluted fractions prepared at various concentrations (1-20 µg/ml) was verified at room temperature by measuring digestion of FTC-casein at 485/538 nm, with reference to TPCK trypsin standard (0-10 µg/ml). The graph represents activity measured at 30 minutes. Data are represented as mean±SD of two repeats, n=2.

Additionally, no protease activity could be detected following cleavage of GST-tags from recombinant proteins as shown in Figure 4.25. Compared with trypsin standards, no protease activity could be detected in GST-CD720 protein fractions at concentrations up to 20 µg/ml.

Another assay was performed to verify the protease activity using a different substrate, Hide powder azure to confirm the results from the previous protease assay with FTC casein substrate.

4.5.1.3.1 Detection of protease activity with Hide powder azure substrate:

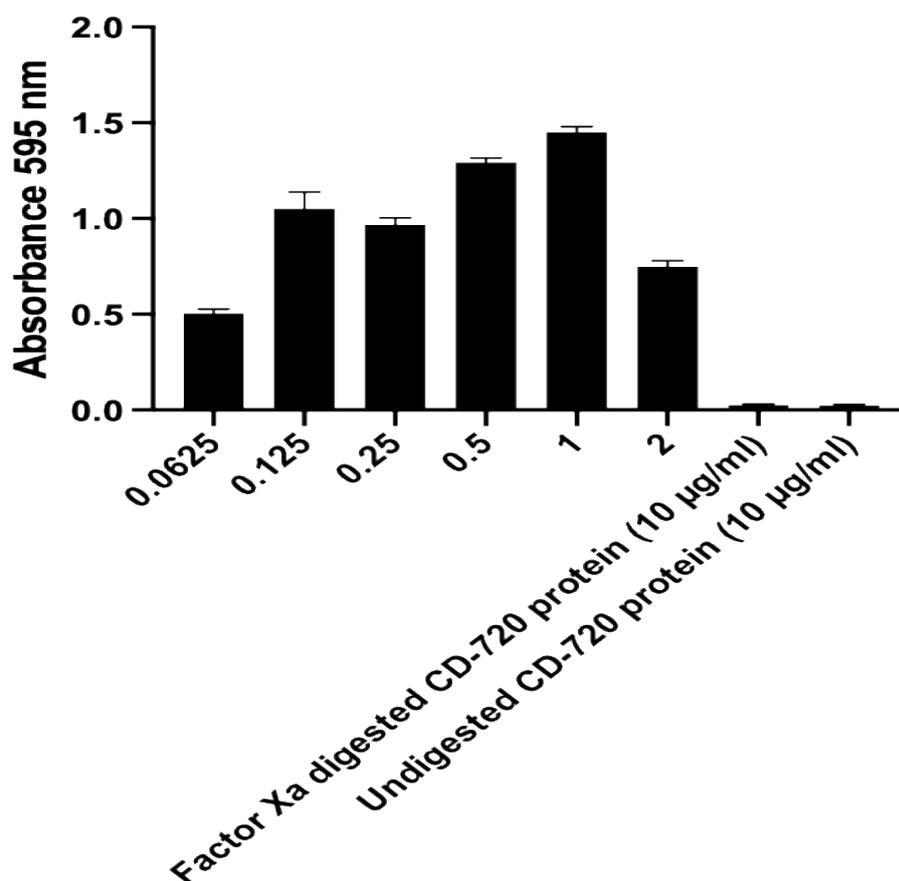


Figure 4.26. Hide powder azure protease assay for recombinant GST-CD720. Factor Xa-digested and undigested GST-CD720 recombinant protein fractions (10 µg/ml) were treated with 5 mg/ml Hide powder azure in 50 mM Tris, 1 mM CaCl₂ pH 8 and incubated at 37°C, with shaking (100 rpm) for 30 minutes. Protease activity was detected by measuring the absorbance of supernatants at 595 nm with reference to TPCK trypsin standards.

As illustrated in Figure 4.26, compared with trypsin positive control no protease activity was detected in undigested and FXa-digested recombinant protein samples using the Hide powder azure assay. The trypsin standard preparations however reached saturation around 1 µg/ml. Recombinant proteins were incubated for longer periods with the Hide powder azure substrate prior to absorbance determination as shown below.

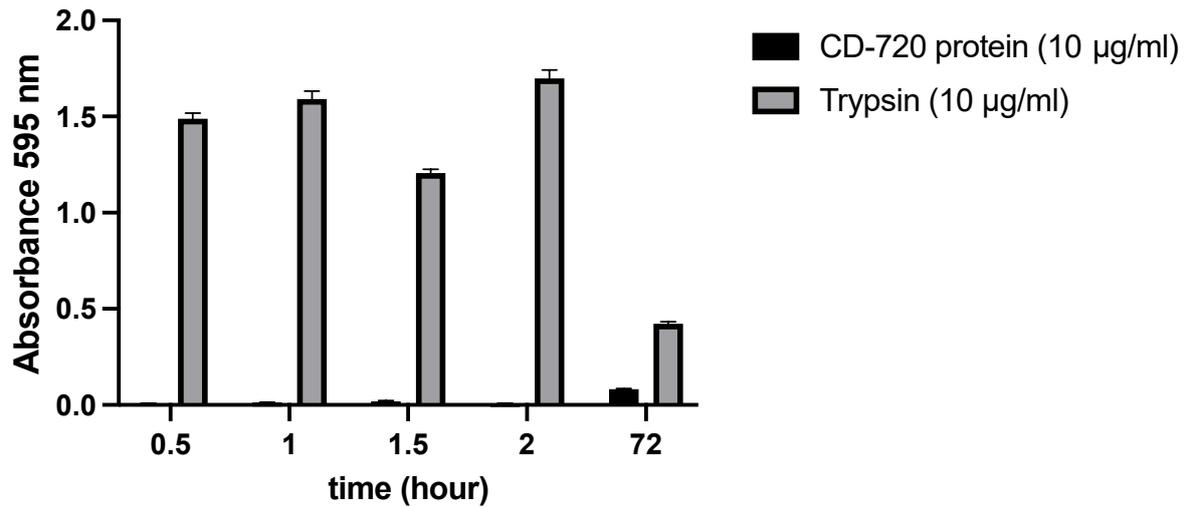


Figure 4.27. Hide powder azure protease assay for recombinant GST-CD720. Factor Xa-digested and undigested recombinant protein fractions (10 µg/ml) were treated with 5 mg/ml Hide powder azure in 50 mM Tris, 1 mM CaCl₂ pH 8 and incubated at 37°C, with shaking (100 rpm) for 72 hours. Protease activity was detected by measuring the absorbance of supernatants at 595 nm with reference to TPCK trypsin standards.

In Figure 4.27, the samples had zero absorbance even after prolonged incubation. The trypsin standard preparations however reached saturation after 30 mins incubation. Next trypsin digestion was performed in attempts to activate the recombinant GST-CD720 protein, given that it could have been expressed as a pro-enzyme that required cleavage of the N-terminal pro-peptide. After digestion with trypsin, samples were analysed by SDS-PAGE prior to assessment of transglutaminase activity.

4.5.1.4 Cleavage of pro-peptides does not activate recombinant GST-CD720 protein

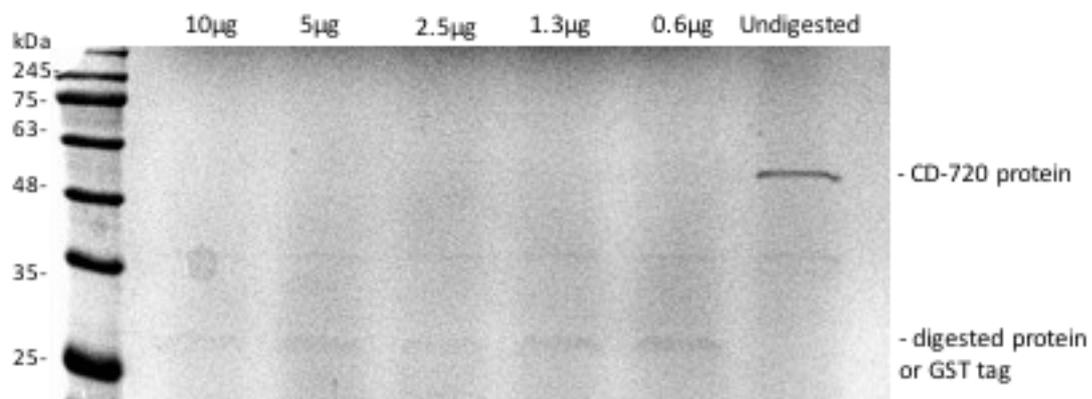


Figure 4.28. SDS-PAGE analysis following trypsin digestions of recombinant GST-CD720 proteins. Purified recombinant proteins (24 µg/ml) were treated with various concentrations of trypsin (0-10 µg/ml) for 30 minutes at 37°C. Samples were analysed by SDS-PAGE and visualised after Coomassie blue staining.

Trypsin digestion of recombinant proteins was successful as observed in Figure 4.28 releasing bands of approximately 26-28 kDa corresponding to the GST-tag (26 kDa) and/or cleaved protein (28 kDa). A transglutaminase activity assay was performed on trypsin-digested samples as shown below.

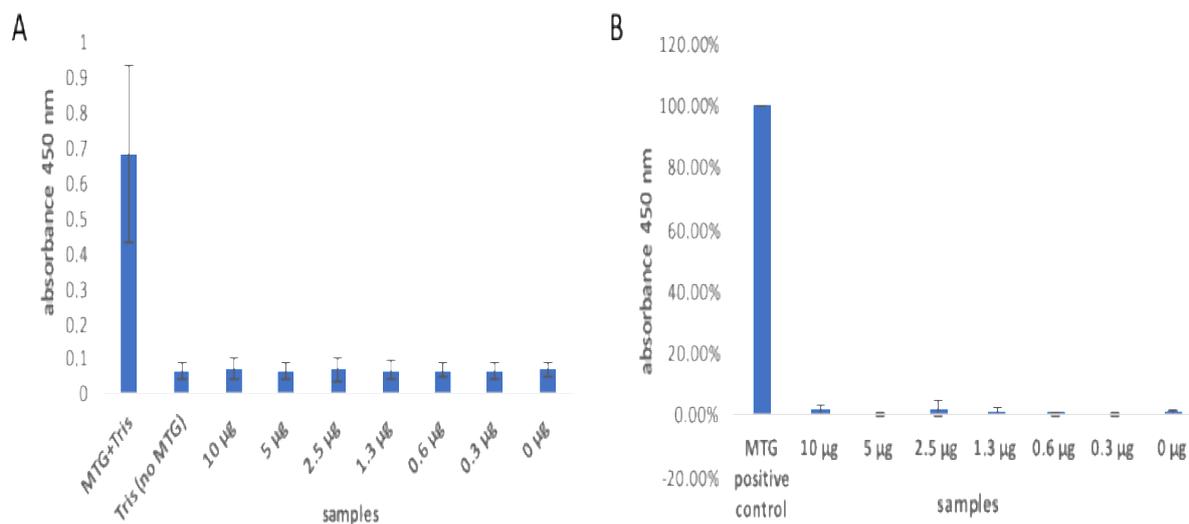


Figure 4.29. Transglutaminase activity assay of recombinant GST-CD720 digested with trypsin. MTG activity assay as measured by the incorporation of biotin cadaverine (BTC) on to casein coated plates. Purified and dialysed recombinant CD-720 proteins (24 µg/ml) were treated with various concentrations of trypsin (0-10 µg/ml) for 30 mins at 37°C. Transglutaminase activity following BTC addition was measured by the oxidation of OPD by HRP labelled ExtrAvidin, with absorbance max of 450 nm. Recombinant commercial MTG served as positive control. (A) activity of purified and trypsinised protein, (B) activity expressed as a percentage relative to the positive control. Data are presented as mean±SD of three repeats, n=3.

The commercially available MTG from *S. mobaraensis* was used as positive control and reference for 100% activity. With reference to the positive control, the purified recombinant GST-CD720 proteins and the trypsin-digested proteins showed no transglutaminase activity, with absorbance comparable to negative control (Tris) levels as shown in Figure 4.29.

Finally, candidate substrates (casein and BSA) were selected for reaction with recombinant GST-CD720 proteins to evaluate evidence of crosslink formation indicating transglutaminase activity using SDS-PAGE analysis. Caseins are established substrates for microbial transglutaminase-catalysed protein crosslinking. Also BSA has been polymerised by microbial transglutaminase (Hsieh and Pan, 2012).

4.5.1.5 Recombinant GST-CD720 protein does not crosslink casein and BSA solutions

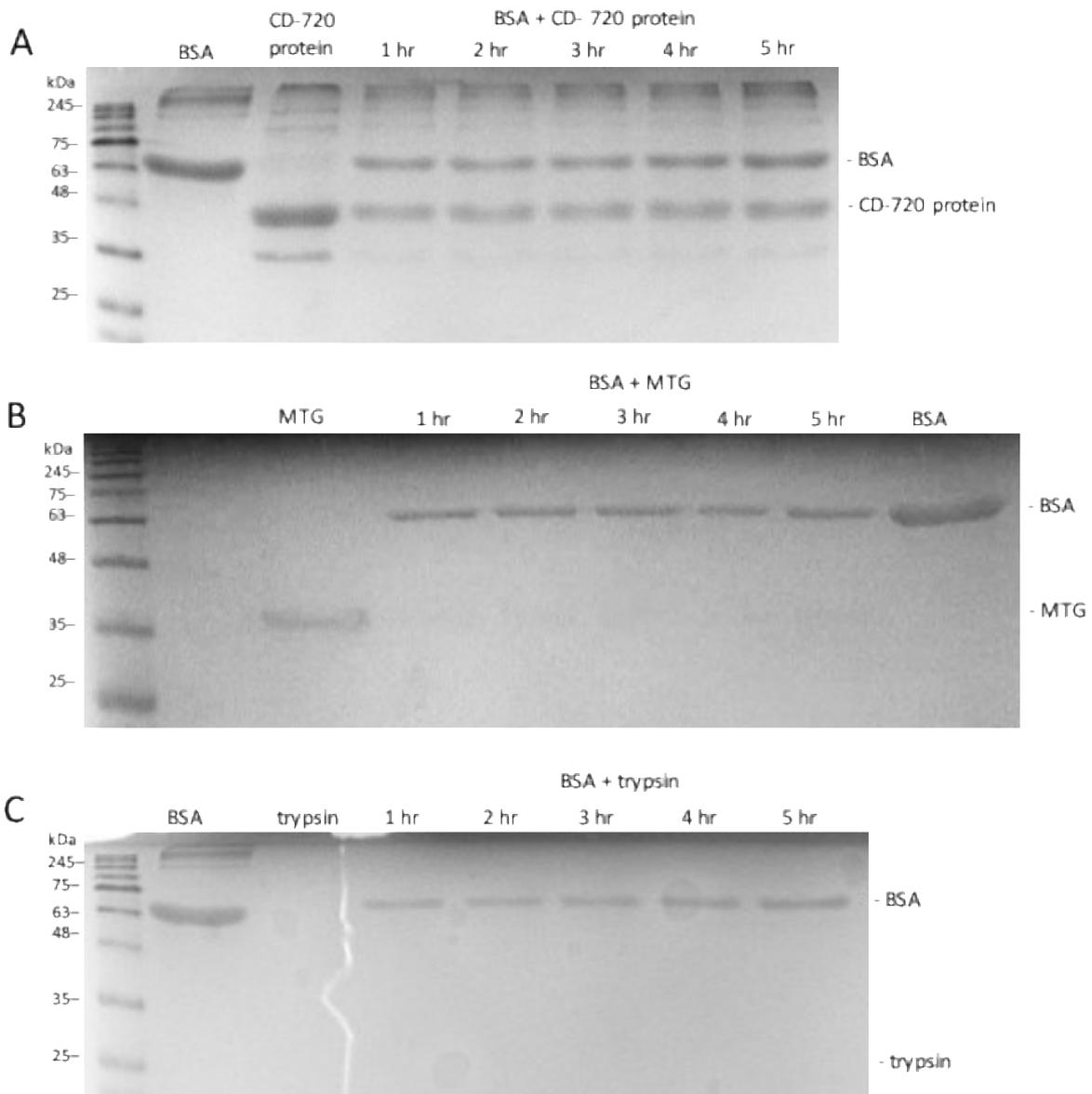


Figure 4.30. SDS-PAGE analysis of recombinant GST-CD720 protein reaction with BSA. BSA (0.35 mg/ml) was treated with purified recombinant GST-CD720 protein (12 μ g), MTG (100 μ g) and trypsin (6 μ g/ml) in 50 mM Tris, pH 8 at 37°C for 1, 2, 3, 4, and 5 hours respectively as shown in lanes 4-8 in A, B and C. Samples were analysed by SDS-PAGE and gels were visualised after Coomassie blue staining. A: BSA treated with recombinant proteins, B: BSA treated with MTG, C: BSA treated with trypsin.

From the SDS-PAGE analysis in Figure 30, there was no difference in banding patterns following incubation of BSA with recombinant GST-CD720 protein, MTG and trypsin illustrated by the two distinct bands in the enzyme-treated BSA lanes. Therefore BSA was not a substrate

for recombinant GST-CD720 protein, trypsin and MTG. Casein was then tested to assess crosslinking activity in recombinant proteins as shown in Figure 4.31 below.

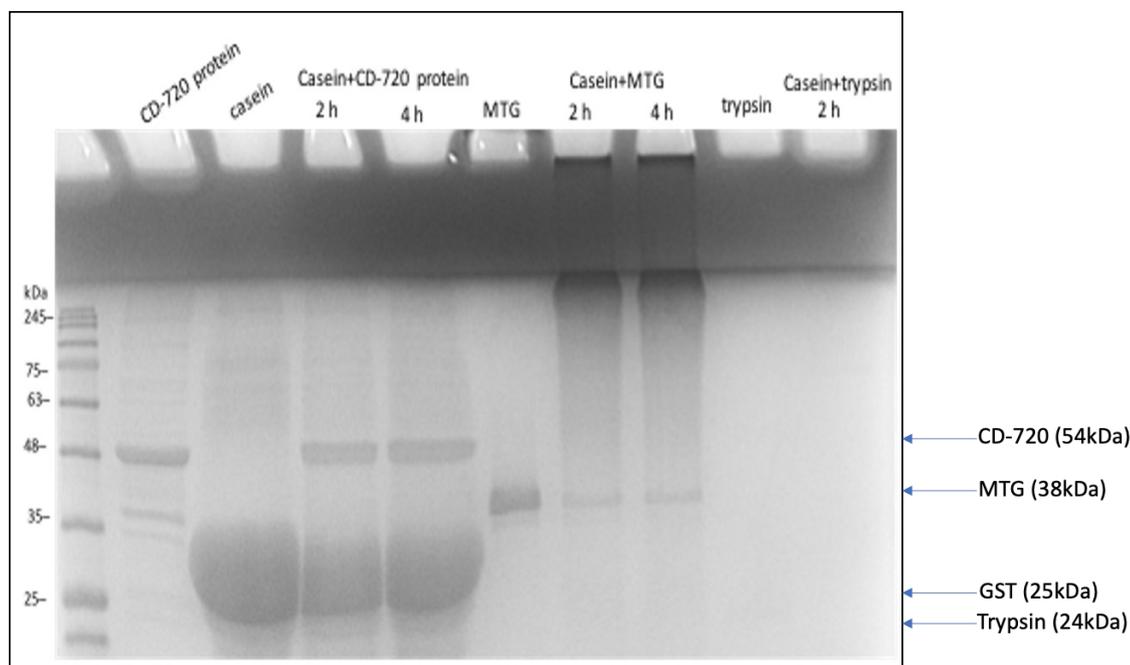


Figure 4.31. SDS-PAGE analysis of recombinant GST-CD720 protein reaction with casein. Casein (10 mg/ml) was treated with purified recombinant GST-CD720 protein (12 μ g), MTG (100 μ g) and trypsin (6 μ g/ml) in 50 mM Tris, pH 8 at 37°C for 4 hours. Samples were analysed by SDS-PAGE after 2 and 4 hours incubation, and gels were visualised after Coomassie blue staining.

As shown in Figure 4.31, no difference in protein bands was observed following incubation of casein with our recombinant GST-CD720 protein illustrated by the two distinct bands in the enzyme-treated casein lanes. Therefore, casein was not a substrate for recombinant GST-CD720 protein. However, in the MTG-treated casein lane, higher molecular mass proteins were observed in the wells and at the top of the gel, accompanied with the disappearance of MTG. There were no visible bands in trypsin-treated casein lanes and in the trypsin only lanes. Also, characterization of mature soluble recombinant GST-tagged proteins (CD-1062m, CD-1083m) which included transglutaminase and protease activity assays on desalted protein fractions remained unsuccessful. Similarly, no transglutaminase or protease activity was detected in full-length recombinant His-tagged proteins. The following figures confirm absence of enzyme activity in induced lysates and eluted fractions.

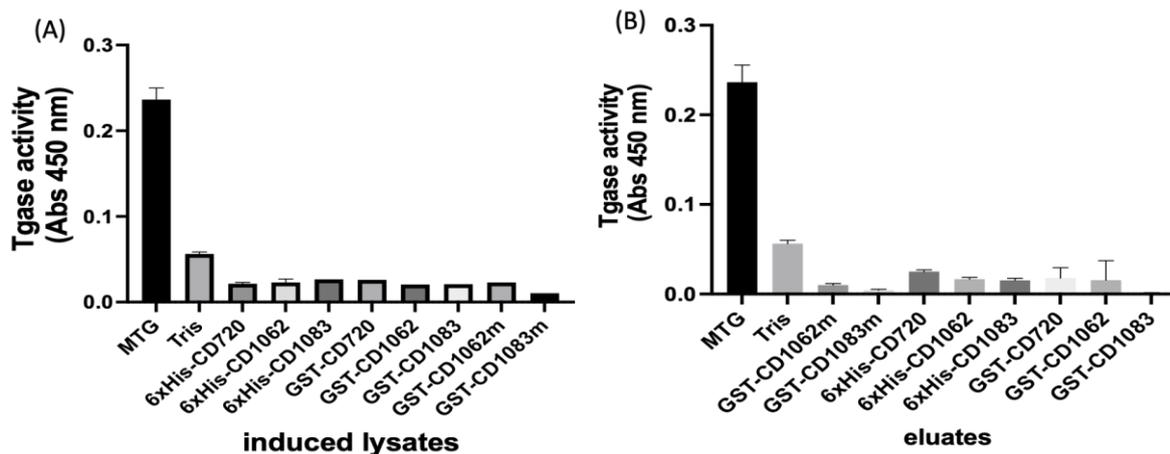


Figure 4.32. Transglutaminase activity in induced lysates and recombinant eluted protein fractions. Transglutaminase activity was measured at 450 nm following biotin cadaverine addition to recombinant GST-tagged and His-tagged full-length and mature proteins as described in Methods section. Enzyme activity was compared to MTG and Tris controls. (A) Activity in induced lysates (B) Activity in recombinant protein fractions.

The results in Figure 4.32 shows that transglutaminase activity was absent in induced lysates and eluted fractions for full-length His-tagged and mature protein expression. Activity was equally absent in cell free supernatants of induced cultures (Appendix 9). Recombinant mature GST-tagged proteins were then digested with Factor Xa to cleave the GST-tag prior to measuring activity. This was to verify whether the GST tag was affecting the protein's functionality.

4.5.2 Cleavage of GST-tags from recombinant mature proteins and column purification

The GST-tags were cleaved from recombinant proteins by enzymatic digestion with Factor Xa. The proteins were purified on GSTrap and heparin columns to eliminate the free GST tag and Factor Xa. The digested samples and eluted fractions were analysed by SDS-PAGE prior to assessment of transglutaminase activity.

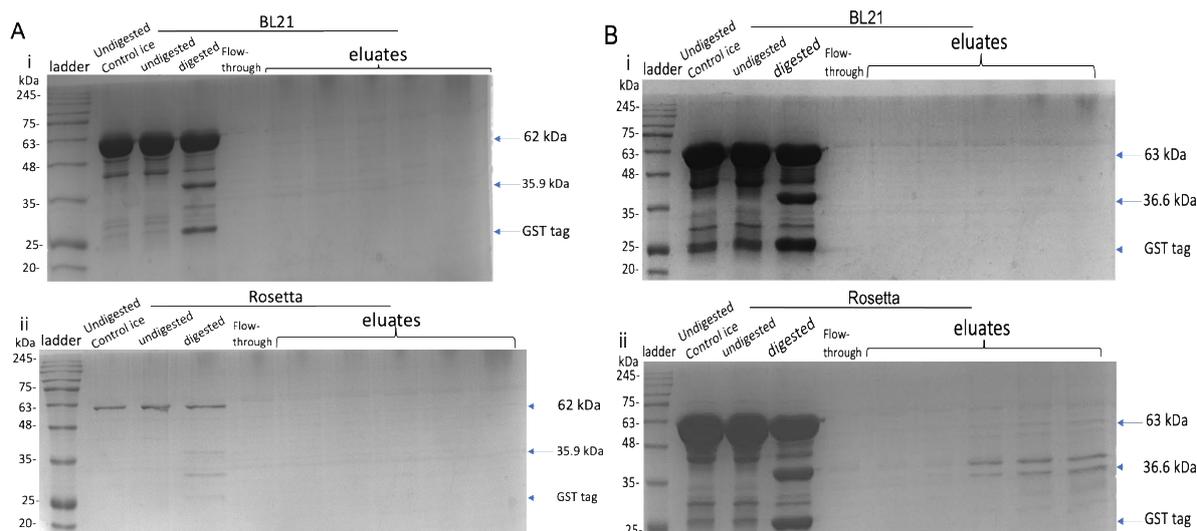


Figure 4.33. SDS-PAGE showing tag cleavage from recombinant mature GST-tagged proteins. Recombinant mature proteins produced in both *E. coli* BL21 and Rosetta 2 cells were desalted and digested with Factor Xa in 50 mM Tris, 150 mM NaCl, 1 mM CaCl₂, pH 7.5 for 2 hours at room temperature prior to purification on heparin plus GSTrap columns. The gels show the undigested and digested recombinant proteins plus eluted protein fractions. (Ai) CD-1062m protein produced in BL21 (Aii) CD-1062m protein produced in Rosetta (Bi) CD-1083m protein produced in BL21 (Bii) CD-1083m protein produced in Rosetta.

The protein gels in Figure 4.33 show that GST tag cleavage from recombinant mature proteins was successful (though incomplete), releasing protein bands corresponding to the target protein sizes plus the free GST tag (26 kDa). For the CD-1062m recombinant protein produced in BL21 and Rosetta cells, no visible bands were detected in eluted fractions after purification on heparin plus GSTrap columns as illustrated on the SDS-PAGE gels (Figure 4.33A). Conversely, for the CD-1083m recombinant protein produced in Rosetta cells visible protein bands were observed at the target size in eluted fractions (Figure 4.33Bii). Meanwhile no protein bands were observed in eluted fractions of CD-1083m proteins produced in BL21 after purification on heparin plus GSTrap columns (Figure 4.33Bi). Importantly, no enzyme activity was detected in eluted fractions following GST-tag cleavage from recombinant mature CD-1062/CD-1083 proteins produced in BL21 and Rosetta cells (Figure 4.32).

On the other hand, we assumed recombinant full-length His-tagged proteins could have been produced as zymogens that require proteolytic activation. Previous studies on *Streptomyces* transglutaminase expression have reported the activation of recombinant MTG following digestion with proteases (Pasternack *et al.*, 1998). Therefore, recombinant His-tagged proteins were digested with trypsin for possible pro-peptide cleavage and analysed by SDS-PAGE prior to testing for enzyme activity.

4.5.3 Trypsin digestions of recombinant His-tagged proteins

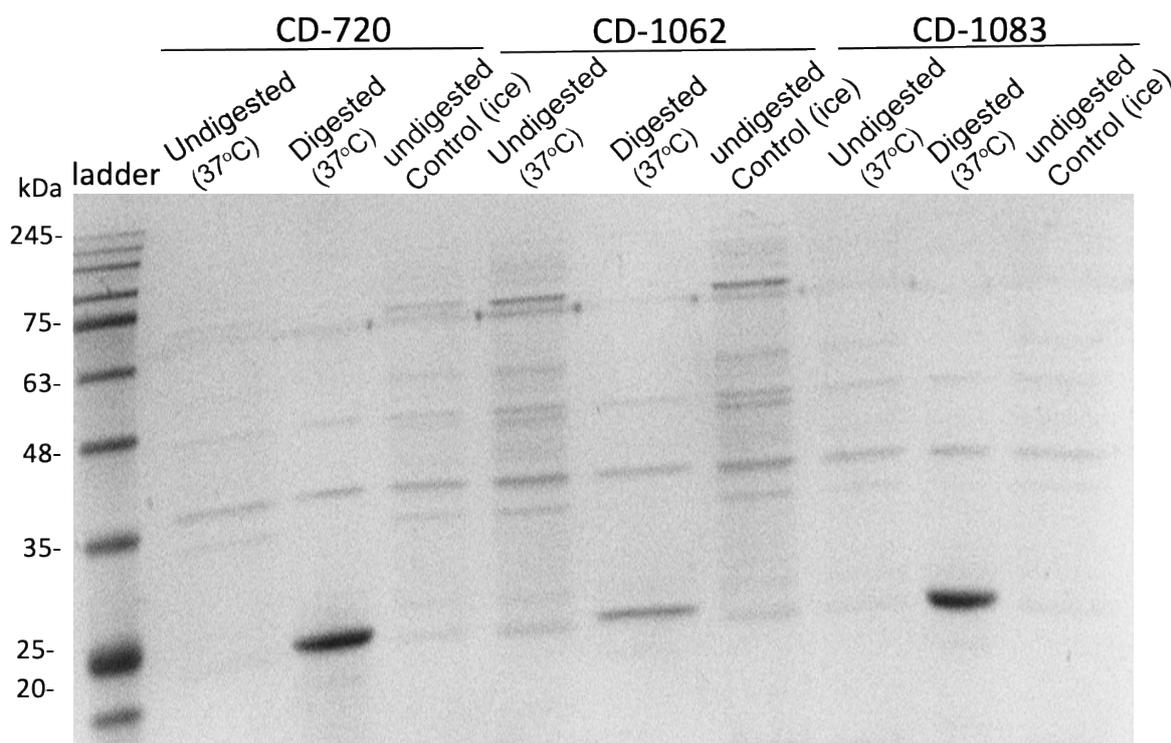


Figure 4.34. SDS-PAGE showing trypsin digestions of recombinant His-tagged proteins. Purified recombinant transglutaminase-like proteins were desalted and digested with trypsin at 8:1 ratio of enzyme to trypsin for 45 minutes at 37°C in TBS buffer. The gel shows the undigested and digested recombinant proteins at 37°C and the undigested controls kept on ice. Trypsin has a molecular weight of 24 kDa.

The gel in Figure 4.34 shows that trypsin digestions of recombinant 6xHis-tagged proteins were successful releasing bands around 25 kDa probably corresponding to trypsin. However, no transglutaminase activity was detected in digested samples (data not shown). Altogether following soluble protein expression, no enzyme activity was detected in recombinant full-length and mature proteins.

4.6 DISCUSSION

In order to determine whether the putative genes identified in Chapter 3 encode transglutaminases, it was necessary to clone, express and purify these proteins. Recombinant protein expression in bacteria is a commonly used method for large scale production of target proteins. Specifically, *E. coli* is widely used for recombinant expression of proteins even though some challenges can be encountered with certain proteins (Tegel, Ottosson and Hober, 2011). Several procedures have been adopted to recombinantly express transglutaminase in *E. coli*. Despite the difficulties encountered, MTG from *S. mobaraensis* was expressed as a soluble protein with high yields by employing several strategies including constitutive expression of the transglutaminase protein using a bi-cistronic vector for co-expression of the pro-enzyme and its activating protease, lactose auto-induction, and a temperature shift approach with growth at 37°C and induction at lower temperatures (Rachel and Pelletier, 2013). In this study target genes were PCR amplified and cloned in expression vectors, followed by recombinant expression in *E. coli* prior to purification and characterization. Several strategies were adopted for intracellular protein expression including different expression systems, *E. coli* strains, temperatures and IPTG concentrations.

4.6.1 Production of expression plasmids

Optimisation of PCR amplifications of the entire coding sequence of transglutaminase genes in *C. difficile* 630 involved the use of touch-down PCR protocols comprising specific temperature ranges for each transglutaminase gene. The experiments were designed to produce blunt-ended PCR products for blunt-end ligations. In the first cloning attempts, PCR product **C** (CD-1083) was cloned within LacZ in pUC19 vector after several failed attempts with pJET1.2 cloning vector. Selection of recombinants with the pJET1.2 vector was based on positive selection, meanwhile pUC19 vector utilises the blue/white screening method. Similar to previous findings, *Sma*I addition to pUC19/PCR product **C** ligation reactions increased the ratio of recombinants and effectively prevented re-ligation of the non-dephosphorylated pUC19 vector. A previous study reported an 80% increase in recombinants in presence of *Sma*I (Ma, Wang and Wang, 2013). Additionally, this ligation is cheap, simple, and fast as the phosphorylation of PCR products and dephosphorylation of cloning vectors can be bypassed. The respective inserts were sub-cloned into pGEX3X expression vector and recombinant plasmids were prepared with pGEX3X sequencing primers. The pre-mixed plasmid preparations were sent for Sanger sequencing. After sequencing, recombinant plasmids were analysed to select constructs with the correct sequence and orientation prior to recombinant expression experiments.

4.6.2 Recombinant *C. difficile* 630 transglutaminase protein expression in *E. coli*

The initial trials to recombinantly express transglutaminases from *C. difficile* 630 as full-length proteins in *E. coli* BL21 at 37°C were abortive for two out of three induced proteins. The expressed proteins formed inclusion bodies under the expression conditions used, except GST-CD720 which was present in the soluble fraction though a lot was still insoluble. The GST-tag at the N-terminus of the cloned transglutaminases enabled affinity purification of the clarified lysates after sonication. In Figure 4.5 the chromatogram illustrates a single, Gaussian peak that eluted following purification of the GST-CD720 protein which corresponds to the 54 kDa band on the SDS-PAGE analysis. Altogether, these initial findings correlate with several other studies in which cytoplasmic expression of transglutaminases at 37°C in *E. coli* resulted in inclusion bodies (Zhang, Zhu and Chen, 2009). The pGEX-3X expression vector used could have a leaky *tac* promoter enabling high basal expression prior to induction although it is expected that repression by *lacI*Q would prevent this. However, this was controlled by adding 2% (w/v) glucose to the culture medium. To solve the problem of inclusion bodies, protein induction was performed at lower temperatures to promote soluble expression.

4.6.3 Effect of temperature on recombinant transglutaminase protein expression

Several publications have emphasized the importance of lowering the temperatures post-induction as a strategy for increased solubility and high protein expression (Marx et al., 2007; Rosano and Ceccarelli, 2014). Additionally, expression temperatures were shown to influence enzyme activities of recombinant proteins (Fazaeli *et al.*, 2019). The temperature switch approach incorporates fast growth rates with lower expression, curtailing inclusion bodies (IB) formation resulting to high levels of soluble protein production. Marx and collaborators employed different parameters to optimize transglutaminase protein expression including temperature. The decreased growth temperature following induction with IPTG produced 30% of soluble protein with periplasm localization (Marx, Hertel and Pietzsch, 2007). Our first optimization approach was to induce at lower temperatures (20°C) to generate more soluble proteins, increase yield and curb formation of inclusion bodies. Additionally, cell cultures were scaled up to 500 ml for increased protein yields. Cell cultures were grown at 37°C and protein expression was induced at 20°C after prior equilibration at the same temperature. Cell pellets obtained from induced cultures were resuspended in suitable buffers prior to cell lysis using a French press and subsequent purification. Indeed, induction at 20°C overnight led to more soluble expression when compared with induction at 37°C for the GST-CD720 protein. This

can be explained by the high metabolic burden impacted on the cell during heterologous production at 37°C, which may overwhelm the folding system resulting in IB formation. Moreover, hydrophobic interactions mainly responsible for IB formation are favoured at high temperatures (Fazaeli *et al.*, 2019). Thus, soluble protein expression and reduced IB formation is favoured at low temperatures. Additionally low temperatures improved protein purity of GST-CD720 and could indicate that lower temperatures possibly favoured proper folding or less proteolysis. By contrast lowering induction temperatures produced no difference in soluble protein expression for GST-CD1062 and GST-CD1083 proteins which remained low or absent. This is illustrated by the persistently high levels of protein accumulated in the insoluble fractions (Figure 4.8 to Figure 4.10). Additionally, a proportion of GST-CD720 protein remained insoluble even with adoption of low temperatures for protein induction.

The composition of culture medium has been shown to influence expression of recombinant proteins in several studies probably by modifying metabolism in hosts (Fazaeli *et al.*, 2019). We did not investigate the effect of culture medium on recombinant expression of target genes. This is because protein yields were not the major concern but soluble expression particularly for GST-CD1062 and GST-CD1083 proteins. It is reported that low cell growth rates observed with protein induction at log phase generates about 50% of overall proteins expressed (Fazaeli *et al.*, 2019). Induction of recombinant transglutaminase protein expression was mostly performed during the early exponential phase, as observed in several other studies. It is true that inducing at higher OD₆₀₀ or stationary phase could increase biomass accumulation (Fazaeli *et al.*, 2019). However, this was not our main interest. Furthermore, the rate of protein synthesis can influence the yield of soluble proteins. It was published that reduced rates of protein synthesis promotes increased production of soluble and properly folded proteins (Tegel, Ottosson and Hober, 2011). Henceforth, we reduced IPTG concentrations from 0.5 mM to 0.1 mM for recombinant expression which concurs with several studies, and yet no significant change in solubility or yield was noticed (Fazaeli *et al.*, 2019).

So far, we have not succeeded with the expression of GST-CD1062 and GST-CD1083 as full-length recombinant proteins in *E. coli* BL21. Despite the adoption of low temperatures post-induction coupled with reduced IPTG levels, the recombinant proteins remained insoluble. Notably, for the GST-CD720 protein where soluble expression has been obtained, extra bands were observed in soluble and insoluble fractions. Particularly, a 26 kDa band suggestive of GST-tags was present in soluble fractions of all expressed proteins. This could result from proteolytic degradation of expressed products by specific proteases that target the GST-tag cleavage sites. However, this should have been prevented by the protease-deficient expression strains and the inclusion of protease inhibitors prior to cell lysis and purification.

Non-target bands were also observed in insoluble fractions of all induced lysates. These bands could represent truncation products arising from premature termination of protein translation. Consequently, the transglutaminase gene sequences were analysed for rare codons given their impact on translation. Indeed, rare codons were found in CD-1062 and CD-1083 transglutaminase-like genes after analysis of *C. difficile* 630 transglutaminase gene sequences with the GenScript rare codon analysis tool. This may explain the presence of non-target bands in the soluble and insoluble fractions of both recombinant proteins, perhaps contributing to failure in recombinant protein expression. The rare codons in the three putative transglutaminase genes were likely to produce truncated GST-tagged products of 35 kDa in GST-CD720 and 39 kDa in GST-CD1062 and GST-CD1083. This could possibly justify the persistent bands between 35-48 kDa in the expressed proteins in supernatants, lysates and eluted fractions (Figure 4.7 to Figure 4.10). Hence optimisation of codon usage was required to improve soluble expression of transglutaminase proteins as full-length products.

The most frequently encountered problem in heterologous protein expression is rare codons which may account for translational errors and decreased levels of recombinant protein production (Fazaeli *et al.*, 2019). Two approaches can be used to resolve this problem. The first is by codon optimisation involving substitution of rare codons to generate a new optimised gene (Fazaeli *et al.*, 2019). However, in certain cases this can result in reduced protein yields and mRNA stability. The second strategy involves the use of engineered strains carrying rare tRNA for heterologous protein expression (Liu *et al.*, 2016). For instance, the *E. coli* Rosetta 2 strain contains a pRARE plasmid supplying tRNA genes to compensate the deficiency in rare codons. Codon and promoter optimization with endogenous transglutaminase promoters produced high yields of recombinant active transglutaminases in *S. lividans* (Liu *et al.*, 2016). Thus the next plausible optimization strategy was addition of rare tRNA genes through plasmid transformation into *E. coli* Rosetta 2 cells, enriched with up to seven rare codons in *E. coli*. The use of *E. coli* Rosetta 2 cells did not improve recombinant expression of our target genes. Major non-target bands were still observed in both soluble and insoluble fractions particularly for GST-CD1062 and GST-CD1083 proteins. Marx and collaborators employed different parameters to optimize transglutaminase protein expression including temperature, localization, and addition of rare t-RNA codons. Similar to our study, in their optimizations neither did the provision of rare tRNA with *E. coli* Rosetta cells nor introduction of a signal leader sequence for periplasm localization increase the solubility of the expressed protein (Marx, Hertel and Pietzsch, 2007). In our system, no improvement in protein expression was observed with optimisation of codon usage via coexpression using Rosetta cells. Accordingly, the problem of translation efficiency was ruled out. As a result, expression of recombinant

proteins with smaller peptide tags in a different expression system was employed to improve proper folding and soluble protein production.

Other factors such as bacterial strains, selected promoters and target protein solubility must be optimised for high yield soluble protein production. Tegel and collaborators demonstrated a positive correlation between protein solubility and expression strains, achieving higher amounts of soluble protein with specific strains (Tegel, Ottosson and Hober, 2011). The *E. coli* 'DE3' strains are ideal for production of recombinant proteins in expression plasmids incorporating T7 promoters. This is because these strains comprise a T7 polymerase gene within the chromosome responsible for production of RNA polymerase for the T7 promoter (Fazaeli *et al.*, 2019). Solubility of target proteins can affect its properties and several factors influence solubility such as hydrophobicity and protein folding (Tegel, Ottosson and Hober, 2011). It is possible that other factors could have contributed to the absence of target bands for GST-CD1062 and GST-CD1083 proteins despite optimisations with Rosetta 2 cells. Our protein expression may have been complicated by the larger sizes of transglutaminase proteins due to the GST-tag, perhaps resulting to incorrect folding which could explain the persistent IB formation. However, GST tags assist protein folding and so soluble expression was envisaged, though this was not achieved (Rosano and Ceccarelli, 2014). Consequently, the next improvement strategy was to use a smaller His-tag for recombinant expression to facilitate protein folding and thus soluble expression. For this purpose, the target inserts were sub-cloned into pRSETc expression vector to enable recombinant expression of transglutaminases as His-tagged proteins. The pRSETc expression system containing a T7 promoter did not favour soluble expression of our target genes in both *E. coli* BL21 and Rosetta 2 cells. Surprisingly, no His-tagged CD-1062 or CD-1083 recombinant protein was expressed at the target size and several non-target bands were present in both soluble and insoluble fractions (Figure 4.12 to Figure 4.15). Western blot images confirm soluble expression of GST-CD720 protein (Figure 4.16), even though target protein bands were found in insoluble fractions and uninduced cell lysates highlighting evidence of leaky expression.

Another crucial factor in promoter choice is the basal levels of transcription. Heterologous protein expression in T7 promoters can lead to basal expression of RNA polymerase in DE3 cells even without IPTG inducers (Tegel, Ottosson and Hober, 2011). Similar amounts of 6xHis-CD720 protein expression were observed in induced and uninduced lysates suggesting probable basal expression. Tightly controlled promoters have negligible basal transcription levels, which is very crucial if the target protein is toxic or harmful to the host. In weak promoters, slower rates of protein synthesis due to lower mRNA give extra time for proper protein folding leading to soluble protein formation (Tegel, Ottosson and Hober, 2011). Asides

leaky expression, it is plausible that the expression vector exerted a metabolic burden on the cell affecting transcription and translation systems involved in target protein production. Another study in addition to leaky expression highlighted decreased cell density upon induction which could be explained by the metabolic burden that affects protein transcription and translation systems (Gomes, Monteiro and Mergulhão, 2020). We did not notice changes in cell growth upon induction and perhaps the impact of metabolic burden during expression was negligible. Promoter strength is predicted by transcription induction, which is majorly influenced by promoter affinity for RNA polymerase. The selective T7 RNA polymerase system engenders efficient transcription and elongation that is about five times quicker compared to *E. coli* RNA polymerase (Tegel, Ottosson and Hober, 2011). Thus, T7 promoter systems are stronger than *E. coli* promoters, and additionally are very tightly controlled. The T7 promoter was recommended for increased levels of soluble protein production. The use of stronger promoters in pRSETc did not benefit protein expression of CD-1062 and CD-1083 which remained trivial. It is plausible that other factors not investigated influenced translation efficiency such as shine-Dalgarno (SD) sequences that play a role in translation initiation. Accordingly, it is expected that the SD sequence length in T7 systems will favour more efficient translation initiation (Tegel, Ottosson and Hober, 2011). The effect of SD sequence in promoting translation seemed to be minimal in our system as indicated by the absence of target 6xHis-CD1062 and 6xHis-CD1083 proteins following expression.

The nature of the target genes may impact significantly on recombinant protein yields with different expression hosts (Fazaeli *et al.*, 2019). The CD-1062 and CD-1083 genes have GC contents greater than 30% which can influence protein transcription and translation (Table 3.1). This could explain the absence of target protein bands following protein induction in different expression systems. Additionally, predictions of protein localisation suggested both CD-1062 and CD-1083 proteins to be secreted, even though no target protein bands were identified in the culture supernatants. The expression of CD-1062 and CD-1083 transglutaminase-like proteins remained problematic given their accumulation in inclusion bodies. Nevertheless, proteins expressed in inclusion bodies can be denatured and refolded to yield soluble protein, but these approaches are time-consuming and the yields of active protein following refolding could be low (Marx, Hertel and Pietzsch, 2007).

Given the role of the pro-peptide in correct transglutaminase protein folding, direct protein expression may have generated high amounts of insoluble IB in line with other studies that highlighted the importance of expressing transglutaminase as pro-enzymes (Fernandes *et al.*, 2015). Additionally the pro-domain is thought to play a role in growth, thermal stability, inhibition of activity, and solubility of the induced protein (Pasternack *et al.*, 1998). Moreover,

our recombinant proteins expressed without the pro-domain could have been toxic to the host. Several authors advocated that transglutaminases when translated in the active form is detrimental to the host, with adverse effects on cell growth and thence cell death caused by protein crosslinking in the cytoplasm (Pasternack *et al.*, 1998; Marx, Hertel and Pietzsch, 2007; Rickert *et al.*, 2016). However, this was checked by monitoring the growth rates of cultures before and after protein induction. No difference in the growth pattern of cultures was observed even after induction. This minimises toxicity as a potential cause of reduced expression. Soluble expression of CD-1062 and CD-1083 proteins as full-length products remained unsuccessful despite various optimisation approaches. Therefore, the proteins were expressed in the mature form to enhance soluble protein production.

After all failed attempts towards soluble expression of CD-1062 and CD-1083 proteins, we wanted to determine the effects of recombinant protein expression without signal peptides. Analysis of the three *C. difficile* 630 transglutaminase protein sequences on the SignalP 4.1 server predicted signal peptides in CD-1062 and CD-1083 transglutaminase-like genes. Signal peptides are short length peptides containing approx. 20-30 amino acids and three domains with specific functions; the N-terminus region, a hydrophobic and a C-terminus region (Freudl, 2018). The two predicted transglutaminases were the biggest with respect to size and the main transglutaminase proteins expressed in highly insoluble forms. It is expected that the signal peptide would decrease problems associated with folding, toxicity, secretion and the two proteins would have been produced in immature forms that could enhance solubility and expression (Rosano and Ceccarelli, 2014).

The mature region of transglutaminase-like genes in *C. difficile* 630 were successfully amplified using specific primers that enabled the cleavage of signal peptides from pGEX3X recombinant constructs. Several optimisations of PCR amplifications were performed including addition of DMSO to eliminate primer dimers. The respective recombinant plasmids were re-ligated and purified by agarose gel electrophoresis. After transformation into BL21 cells, recombinant plasmids were selected on ampicillin plates. Plasmid isolation from transformant cultures followed by nucleotide sequencing with specific primers confirmed that the mature transglutaminase sequences were in-frame with the GST-tag. Finally, the mature sequence of CD-1062 & CD-1083 transglutaminase-like genes within *C. difficile* 630 were successfully expressed as soluble GST-tagged proteins in *E. coli* BL21 and Rosetta 2 cells. Major protein bands were seen at the target size for the two recombinantly expressed transglutaminase proteins (Figure 20-Figure 21). It is plausible that the hydrophobicity within signal peptides (SP) in target genes might have contributed to accumulation in IB during full-length protein expression, especially given that SP did not direct protein secretion in culture

media. After succeeding with soluble protein expression, the recombinant proteins were tested for enzyme activity.

4.6.4 Characterization of recombinant proteins

Recombinant proteins were produced with GST tags as full-length GST-CD720 protein and mature GST-CD1062 and GST-CD1083 proteins. After soluble protein expression, characterization assays were performed to determine protein activity. This involved measuring transglutaminase activity in recombinant proteins. Surprisingly, no transglutaminase activity was detected in recombinant GST-CD720 proteins including clarified lysates, culture supernatants, and eluted fractions (Figure 4.32). Given that fusion partners could affect protein functionality, GST-tags were cleaved and transglutaminase activity was measured once more. Again, no activity was detected in eluted fractions following cleavage of GST-tags from recombinant proteins. There was a possibility that the GST-CD720 recombinant protein may have been produced as a pro-peptide rendering it inactive. Microbial transglutaminases including *Stv. mobaraense* have been produced as zymogens that require proteolytic activation (Pasternack *et al.*, 1998). Consequently, the GST-CD720 protein was digested with trypsin to cleave the inhibitory propeptide and release the active protein. After measuring transglutaminase activity in digested samples no activity was found. Alternatively recombinant proteins could have possessed proteolytic activity given that microbial transglutaminases have evolved from ancient proteases (Makarova, Aravind and Koonin, 1999). No protease activity was detected in recombinant proteins after employing different substrate-based assays; a fluorescent protease assay with FTC-casein substrate and a colorimetric protease assay with a modified collagen substrate (Hide powder azure). Finally, a crosslink assay was performed on recombinant proteins using different substrates to identify protein crosslinks as evidence of transglutaminase activity. No protein crosslinks were identified after incubation with casein and BSA substrate solutions (Figure 4.30). Meanwhile, casein was crosslinked by *Streptomyces* MTG forming higher molecular weight products at the top of the gel illustrating evidence of protein activity (Figure 4.31). In our system, BSA was not crosslinked by MTG control. The fact that BSA acts as glutamine or lysine type substrate and casein as glutamine-lysine substrate may have influenced reactivity with active sites (Hsieh and Pan, 2012).

Furthermore, recombinant proteins were produced as full-length products with 6xHis-tags. After measuring transglutaminase activity in recombinant 6xHis-CD720, no activity was detected in the eluted protein fractions, clarified lysates (Figure 4.32), and culture supernatants (appendix 9). Similarly, for 6xHis-CD1062 and 6xHis-CD1083 constructs expressed entirely as insoluble proteins, no activity was detected in clarified lysates, eluted

protein fractions (Figure 4.32), and culture supernatants (appendix 9). Equally no proteolytic activity was detected in eluted fractions of recombinant 6xHis-CD720 protein. Trypsin digestions of recombinant His-tagged CD-720 protein to cleave a probable pro-peptide and release the active protein could not reveal transglutaminase activity. Therefore, the type of protein tag did not influence transglutaminase activity given that no activity was detected in recombinant His-tagged or GST-tagged proteins. Subsequently recombinant proteins were produced as mature CD-1062 and CD-1083 proteins with GST-tags. No transglutaminase activity was detected in purified mature proteins, clarified lysates (Figure 4.32), and culture supernatants (appendix 9). Then GST-tags were cleaved from recombinant mature proteins with complete digestion even though no protein was recovered following purification except for GST-CD1083 protein expressed in Rosetta 2 cells (Figure 4.33B). As observed for full-length protein expression, no transglutaminase activity was found after GST-tag cleavage from mature proteins. Soluble expression of full-length and mature transglutaminase-like proteins did not yield active proteins. This has been shown in other studies in which recombinant transglutaminase expression in *E. coli* did not result in active products (Marx, Hertel and Pietzsch, 2007).

4.7 Conclusion

The *C. difficile* 630 putative transglutaminases were recombinantly produced in *E. coli* as GST-tagged and His-tagged proteins in the respective expression plasmids. For full-length protein expression in both pGEX3X and pRSETc systems, CD-720 was more soluble than CD-1062/CD-1083 which remained almost exclusively insoluble. This can be explained by differences in structural properties and folding rate of translated proteins or related to the inherent solubility of each protein (Tegel, Ottosson and Hober, 2011). Accordingly, CD1062 and CD-1083 which behaved alike during recombinant expression and have not been produced as soluble full-length proteins are 73% identical based on bioinformatics analysis. Even though a combination of strong promoters with Rosetta cells have been suggested for high yield protein expression, it did not work in our system particularly for full-length expression of CD-1062 or CD-1083. As a result, CD-1062 and CD-1083 genes within *C. difficile* 630 were successfully expressed as proteins in the mature form in *E. coli* BL21 and Rosetta 2. Even with soluble protein expression of transglutaminase-like genes from CD630, no enzyme activity was detected in purified proteins. We conclude that transglutaminase-like genes from *C. difficile* 630 were produced recombinantly as inactive proteins in *E. coli* hosts, or that recombinant proteins may not be transglutaminases or have a particular substrate which was not investigated in this study. To identify putative transglutaminases associated with

sporulation in *C. difficile* 630, transglutaminase activity was purified from sporulating *C. difficile* 630 cells prior to protein characterization assays.

CHAPTER 5. Purification and Characterization of transglutaminase from *C. difficile* 630 spores

5.1 Introduction

In *B. subtilis* there exist spore preparation methods that are reproducible, although very few of these methods have been proposed to study *C. difficile* sporulation. In *C. difficile* there is a large variation in the methodologies used in different studies related to the growth medium used (Burn and Minton, 2011). Up to date, the most commonly utilised culture medium for *C. difficile* sporulation is BHI broth containing 0.5% (w/v) yeast and 0.1% (w/v) L-cysteine (Sorg and Sonenshein, 2008, 2009; Underwood *et al.*, 2009; Burns, Heap and Minton, 2010). In some studies, Wilson's broth is used as the culture medium for optimal induction of sporulation (Lawley *et al.*, 2009; Pereira *et al.*, 2013). The first isolation and purification of microbial transglutaminase (MTG) was from culture supernatants of *Strv. mobaraense* (Ando *et al.*, 1989). From then, microbial transglutaminase has been detected extracellularly in several spore formers (Cui *et al.*, 2007; Macedo, Sette and Sato, 2011; Nur'amaliyah, Zilda and Mubarik, 2016). More recently, transglutaminase has been produced and detected intracellularly in sporulating *B. subtilis* cells (Suzuki *et al.*, 2000). In *Bacillus* species a link has been established between transglutaminase activity and sporulation. The correlation between transglutaminase activity and sporulation was indicated by the presence of transglutaminase activity only in strains that sporulated efficiently (Kobayashi *et al.*, 1996; Katsunori Kobayashi *et al.*, 1998).

Several methods have been employed for purification of microbial transglutaminase with varying yields. This includes ion-exchange chromatography (IEX), hydrophobic interaction chromatography (HIC), ammonium sulphate precipitation, ultrafiltration, affinity chromatography and size exclusion chromatography (SEC) (Suzuki *et al.*, 2000; Cui *et al.*, 2007; Macedo, Sette and Sato, 2011; Nur'amaliyah, Zilda and Mubarik, 2016). In *B. subtilis*, transglutaminase was purified from spores/sporulating cells by a two-step procedure including ammonium sulphate precipitation followed by HIC (Suzuki *et al.*, 2000). No purification of transglutaminase from *C. difficile* has been described to date. Therefore *C. difficile* 630 spore transglutaminase purification was guided by reports on intracellular enzyme production during sporulation in *B. subtilis*. The first part of this chapter involves identification of the best culture method for high yield transglutaminase production during sporulation in *C. difficile* 630. Sporulation in liquid and solid media have been assessed to identify the most suitable for large scale transglutaminase production during sporulation. Also, the solubilisation of transglutaminase enzyme from spores has been evaluated at different pH and temperature conditions to identify the best combo for optimum enzyme extraction. The last part of this

chapter was set out to purify transglutaminase activity from spores, involving optimisation of purification methods to detect relatively pure activity. Biochemical characterization assays and mass spectrometry analysis were performed after obtention of purified enzymes. In summary the objectives of this chapter are:

- To identify the best culture method for high yield transglutaminase production during sporulation in *C. difficile* 630.
- To identify the best pH and temperature conditions for maximal enzyme extraction from spores.
- To purify transglutaminase activity from spores and characterize the purified protein.

5.2 Transglutaminase activity during *C. difficile* 630 sporulation on solid and liquid media

The first objective was to identify a suitable method and time interval for optimal transglutaminase production during sporulation. In this study BHI media was used to generate spores in broth and agar for protein purification purposes. BHI broth containing yeast and cysteine was used to induce sporulation by nutrient starvation. Starter cultures were sub-cultured twice in fresh BHI broth and grown till log phase before plating out on agar. Following sporulation induction in broth and agar, transglutaminase activity was monitored over 5 days of sporulation at 37°C under anaerobic conditions as shown in the figure below.

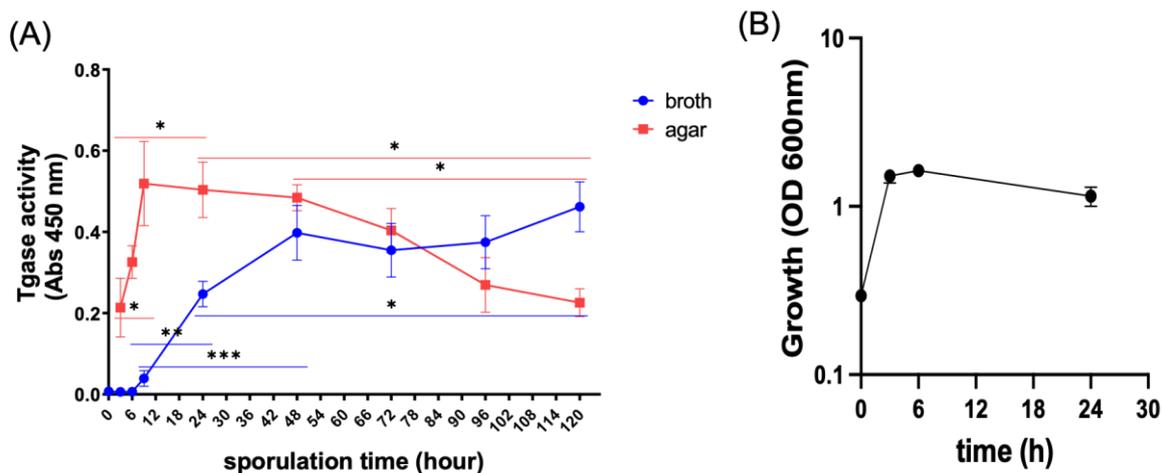


Figure 5.1. Detection of transglutaminase activity during sporulation in *C. difficile*. Sporulation was induced in BHI broth by subculturing an overnight culture of *C. difficile* as previously described, and plating cultures at early exponential phase on several pre-reduced BHI agar plates. Sporulating cells were harvested from broth cultures and agar plates at different time intervals over 5 days of sporulation in anaerobic conditions. Transglutaminase activity was measured in culture supernatants and cell extracts after protein solubilisation from sporulating cells. Results are presented as mean±SEM of three independent repeats. (A) Transglutaminase activity in broth supernatants and cell extracts from agar plates (B) Growth curve. Growth was monitored at OD₆₀₀ in BHI broth prior to plating on agar plates at log phase. Asterisks (*) indicate statistical significance where $p < 0.05$.

The results in Figure 5.1 show the transglutaminase activity profile over time in broth and agar. Previous reports established sporulation in *C. difficile* to be complete within 5 days (Burns, Heap and Minton, 2010). Transglutaminase activity was measured in broth supernatants and agar cultures at different time points during the sporulation course. Sporulating cells collected from agar plates showed transglutaminase activity throughout sporulation. Transglutaminase activity from solid media was higher for the most part of sporulation compared to broth supernatants, except after 120 hours where activity was significantly ($p=0.007$) higher in supernatants. Transglutaminase activity in culture supernatants was only detected after 9 hours and peaking at 120 hours ($p < 0.001$) of sporulation, with significant increase in activity at 24 hours ($p < 0.006$) and 48 hours ($p < 0.001$). Contrarily, transglutaminase activity was observed earlier, from 3 hours of inducing sporulation and peaked at 9 hours ($p=0.046$) when assessed in agar plates. Also, there was a significantly higher activity from agar at 24 hours of sporulation compared to broth cultures ($p=0.0065$). This implies that more time was required to attain maximal enzyme activity levels in broth compared to agar. Therefore, enzyme production on agar medium was faster and agar plates are easier to handle in the anaerobic cabinet allowing for more plates to be processed in a shorter time. Therefore, for subsequent protein purification experiments sporulating cells were harvested from agar plates at 24 hours

of sporulation which was also a convenient time for harvesting and downstream processing. The growth of *C. difficile* in broth cultures was monitored by OD measurements at 600 nm up to the stationary growth phase. The growth curve in Figure 5.1B illustrates that transglutaminase activity on solid media was detected at a time that corresponds to stationary growth phase in liquid culture. Also, maximum enzyme production in liquid culture supernatants occurred after termination of the stationary phase. Following identification of the ideal media and optimal time for high yield transglutaminase production, the next step was to determine the conditions for optimal enzyme extraction.

5.3 Solubilisation of transglutaminase from sporulating cells using different temperature and pH conditions

To investigate the conditions for optimum enzyme solubilisation from sporulating cells, different pH and temperatures were used. Initially spores were incubated at different temperatures ranging from 0°C to 50°C in 50 mM Tris, pH 7.5 for 2 hours. At specific time points during solubilisation, transglutaminase activity was measured in supernatants of spore preparations. Figure 5.2 below shows the time course for enzyme solubilisation at different temperatures.

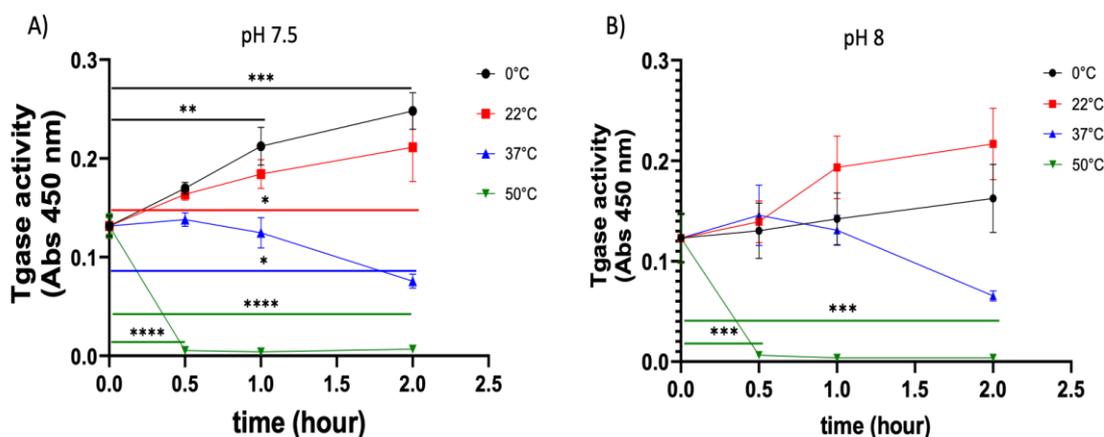


Figure 5.2. Solubilisation of transglutaminase from spores at different temperatures. Sporulating cells were resuspended in 50 mM Tris, pH 7.5 or pH 8 and incubated at different temperatures (0°C, 20°C, 37°C, 50°C) for up to 2 hours. Every 30 minutes cell suspensions were centrifuged for 5 minutes and transglutaminase activity was measured in the supernatants. At time 0 h, spores were resuspended in buffers and supernatants were collected for activity assay. Data are presented as mean±SEM of three independent experiments. (A) solubilisation at pH 7.5 (B) solubilisation at pH 8. Asterisks (*) indicate statistical significance where $p < 0.05$.

Spore enzyme solubilisations were performed in 50 mM Tris pH 7.5 buffer at different temperatures. Figure 5.2A shows that 0°C seemed to have been the best temperature for maximal enzyme solubilisation. The extent of enzyme solubilisation decreased with increasing temperatures and complete enzyme inactivation was observed at 50°C, accompanied with a significant decrease in activity within 30 minutes. At pH 7.5 the enzyme activity was significantly increased from 0-2 hours between 0°C to 22°C, and significantly reduced at 37°C after 2 hours. Solubilisation experiments were performed at pH 8 using various temperatures to confirm the temperature for optimal solubilisation at this pH prior to further purification. At pH 8 enzyme activity at 22°C was greater than activity at other temperatures tested during solubilisation for 1-2 hours (Figure 5.2B). These results suggest that at pH 7.5, 0°C seem to be the best temperature for enzyme solubilisation and 22°C is the best at pH 8.

Also, the pH range for optimal enzyme solubilisation from sporulating cells was investigated. Spore enzyme solubilisations were performed at 0°C in different pH buffers, given that maximal extraction was obtained between 0-22°C using Tris pH 7.5. The graph below shows the time course for enzyme solubilisation at different pH.

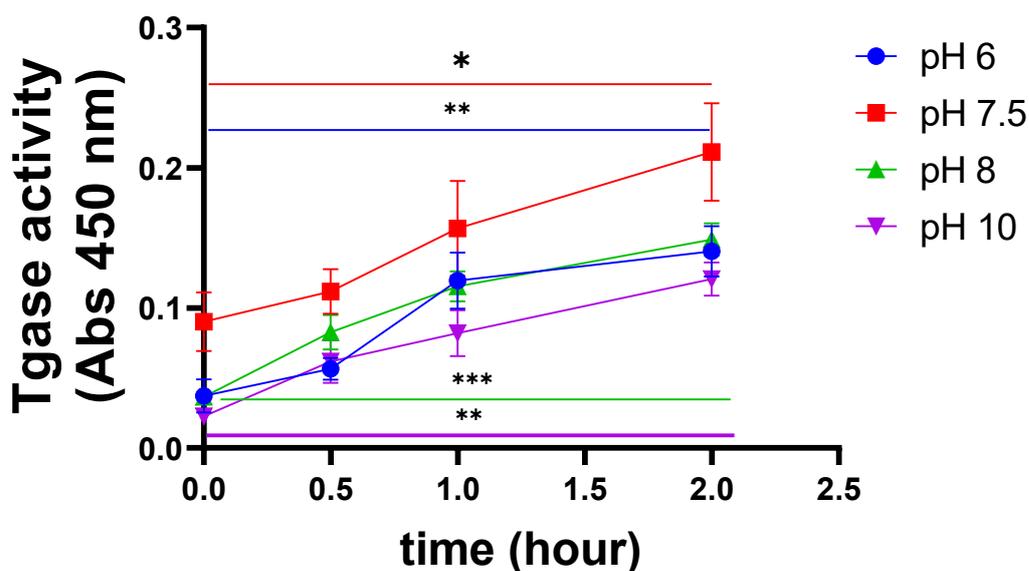


Figure 5.3. Change in transglutaminase activity over time at different pH. *C. difficile* spores were resuspended in 50 mM Tris at different pH and incubated for 2 hours on ice. At different time points, samples were harvested, and transglutaminase activity was measured by the BTC incorporation assay. Results are presented as mean \pm SEM of three independent repeats. Asterisks (*) indicate statistical significance where $p < 0.05$.

Spore enzyme solubilisations were performed at 0°C in Tris buffers ranging from pH 6-10. After enzyme extractions at the various pH, transglutaminase activity assay was performed without pH adjustments and results are plotted as shown in Figure 5.3. The enzyme activity increased significantly after solubilisation for 2 hours at all pH tested. From the gradients obtained for each pH preparation, the rate of enzyme extraction was similar particularly for pH 6, pH 8, pH 10, and pH 7.5 appeared to have a slightly higher rate compared to other pH. Therefore, pH 7.5 was likely the most efficient pH for enzyme extraction given its steeper gradient, implying higher activity was achieved quicker compared to other pH tested. Even so pH 8 was suitable for efficient extraction and was adopted for solubilisation prior to protein purification. Lastly the effects of reducing agents and detergents on enzyme solubilisation from sporulating cells was investigated.

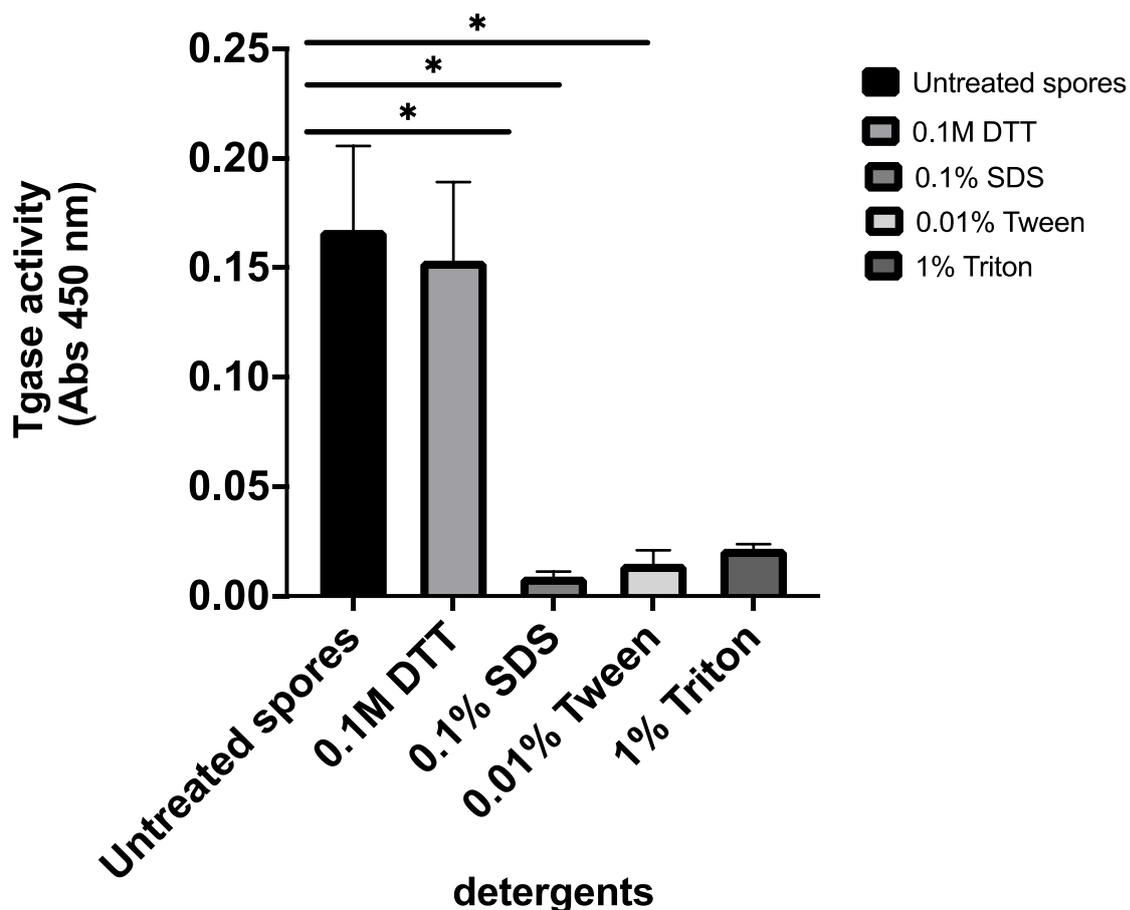


Figure 5.4. Effect of detergents and reducing agents on transglutaminase solubilisation from spores. Sporulating cells were resuspended in Tween 20 (0.01% v/v), Triton (1% v/v), SDS (0.1% w/v), and 0.1 M DTT at pH 8. After incubation for 1 hour at room temperature, transglutaminase activity was measured in cell extracts by the BTC incorporation assay. Results are presented as mean±SEM of three independent repeats. Asterisks (*) indicate statistical significance where $p < 0.05$.

As shown in Figure 5.4 the detergents Tween 20 (0.01% v/v), Triton (1% v/v), and SDS (0.1% w/v) were not suitable for extraction of enzyme activity from sporulating cells. No transglutaminase activity was detected after solubilisation for 1 hour in alkaline conditions. Also compared to untreated spores, no change in enzyme activity was obtained following solubilisations with 0.1 M DTT for 1 hour at pH 8. Based on spore solubilisation assays, the optimum time and conditions for maximal enzyme extraction were determined (Figure 5.2). After successful enzyme solubilisation, crude enzyme extracts were heated at 50°C to evaluate protein inactivation as evidence of true enzyme activity.

5.4 High temperatures inactivate crude enzyme activity

Crude extracts were heated at 50°C and residual activity was measured and plotted on the bar chart below.

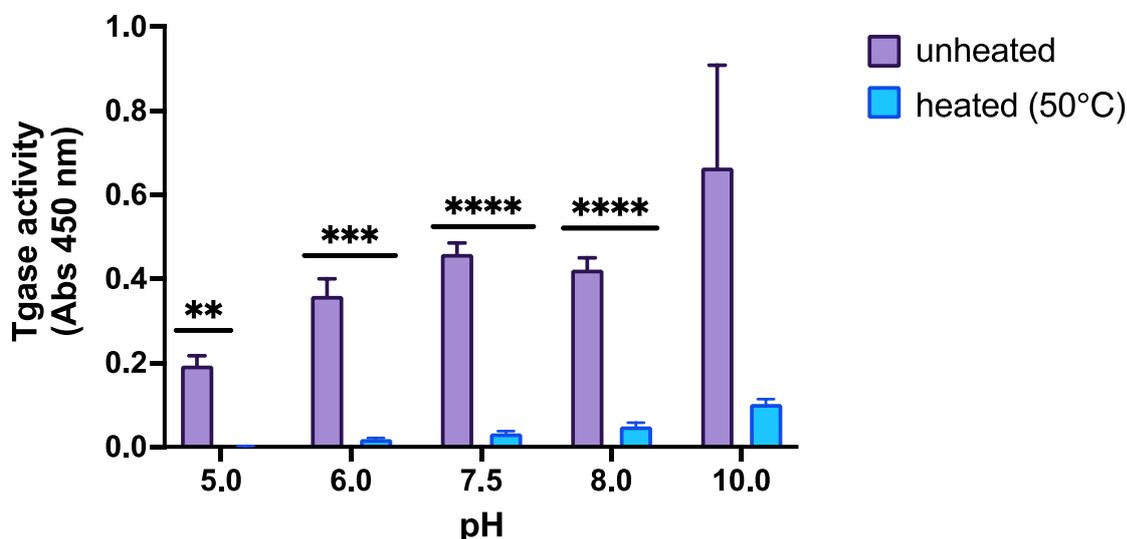


Figure 5.5. Effect of heat on transglutaminase activity of crude extracts from *C. difficile* 630. Crude extracts obtained from solubilisation of sporulating cells at various pH were heated at 50°C for 30 min. Transglutaminase activity in the crude extracts was detected at 450 nm using the BTC incorporation assay as previously described in the Methods section. Results are presented as mean±SEM of three independent repeats. Asterisks (*) indicate statistical significance where $p < 0.05$.

The bar chart in Figure 5.5 indicates that transglutaminase activity was inactivated by heat treatments at 50°C for 30 min. The abolition of activity was significant in crude extracts obtained from pH 5-8. Heat denaturation at high temperatures indicates evidence of true enzyme activity. After identification of the conditions for optimal enzyme solubilisation from spores and assessment of the authenticity of enzyme activity, a protein purification strategy was devised.

5.5 Purification of transglutaminase activity from crude extracts of *C. difficile* 630

In this study enzyme purification was conducted at 4°C. Our purification scheme consisted respectively of IEX, HIC and SEC. For IEX chromatography, comparisons were made between anion exchange and cation exchange columns to determine the most suitable column for initial protein purification. For this purpose, the enzyme was solubilised from sporulating cells in 50 mM Tris at pH 6 and pH 8. Protein purification was performed at pH 6 on RESOURCE S column and pH 8 on RESOURCE Q, corresponding respectively to cation exchange and anion

exchange columns. No transglutaminase activity was bound to RESOURCE S columns at the pH tested, and all the activity was found in the flow-through. On the other hand, transglutaminase activity was detected in eluted fractions from RESOURCE Q column, with no activity found in flow-through and washes. Consequently ion-exchange purification was performed at pH 8 on anion-exchange columns using Q-Sepharose resins. The ideal temperature for enzyme solubilisation at pH 8 was 25°C. Thus enzyme solubilisation for protein purification were performed at 25°C.

5.5.1 Purification of transglutaminase by ion-exchange chromatography

Transglutaminase protein solubilised from sporulating cells for 2 h at 25°C in 50 mM Tris, pH 8 was applied to anion-exchange columns. Initially crude extracts were purified on 20 ml Q-Sepharose FF columns and eluted fractions were tested for enzyme activity as shown in Figure 5.6.

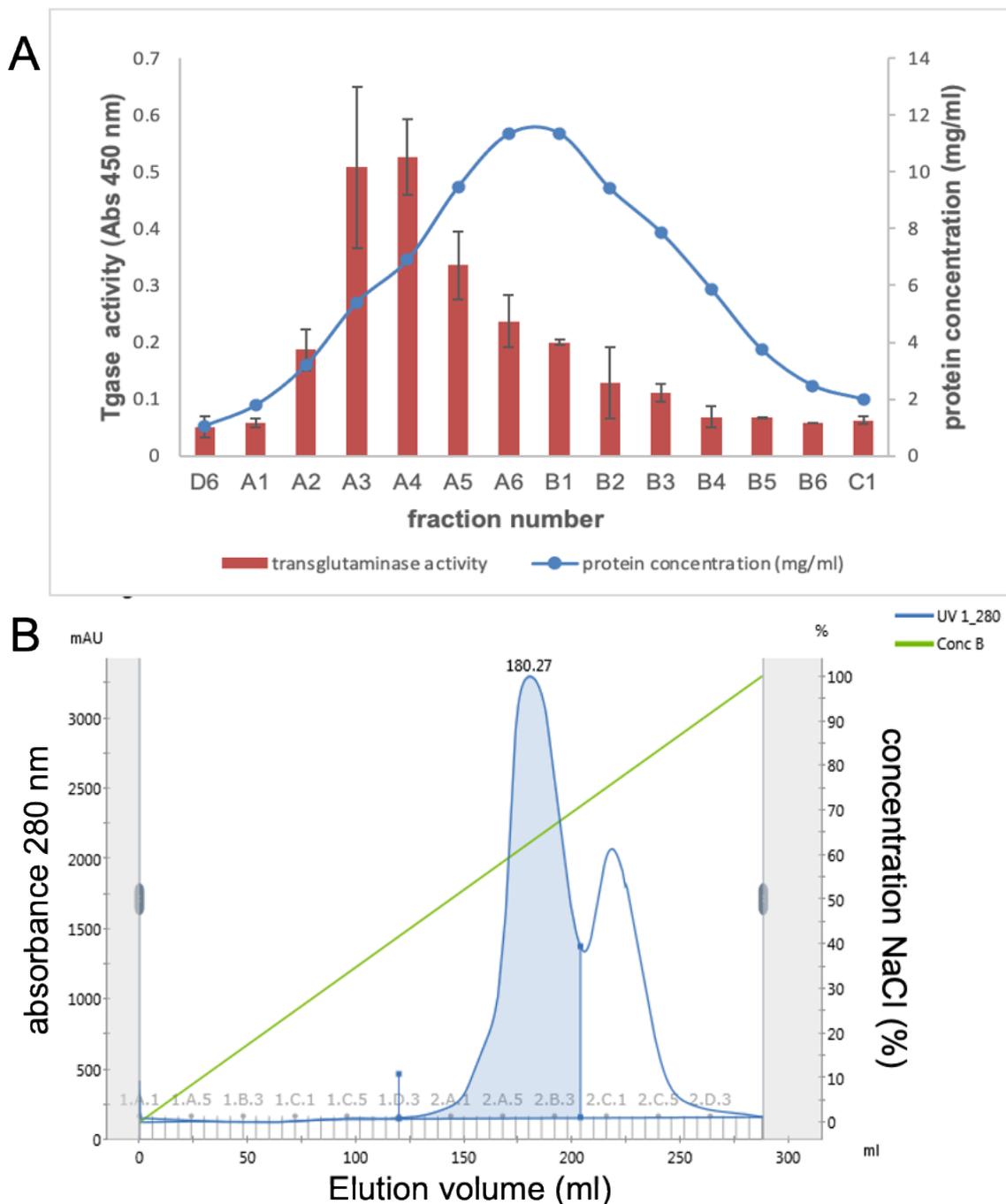


Figure 5.6. Purification of crude enzyme extracts from sporulating *C. difficile* on 20 ml Q-Sepharose FF. Crude extracts obtained following solubilisation of sporulating cells at 25°C for 2 h in 50 mM Tris, pH 8 were purified on Q-Sepharose FF columns. Anion exchange purification was performed on an AKTA system, and protein absorbance was monitored at 280 nm. The bound transglutaminase protein was eluted with an increasing linear NaCl gradient (0-2 M) in Tris buffer. Transglutaminase activity in the eluted fractions was detected at 450 nm using the BTC incorporation assay as previously described. (A): Transglutaminase activity of eluted fractions. Protein concentration is plotted on the right y-axis and transglutaminase activity is plotted on the left y-axis. (B) AKTA purification profile (green line: increasing NaCl gradient, blue line: protein absorbance detected at 280 nm).

From Figure 5.6B, the bound protein was eluted with a linear NaCl gradient, and the peak transglutaminase activity (fractions A2-B3) was found in the first elution peak corresponding to NaCl concentration of 50-65 % (v/v) of a 2 M stock solution. From these results the elution profile was optimised in subsequent scaled-up experiments. In later optimisation of enzyme purification, agar cultures were scaled up for increased protein production and 100 ml Q-Sepharose columns were used in place of 20 ml Q-Sepharose columns to accommodate the increased protein load. Crude extracts were loaded onto Q-Sepharose columns, and the bound protein was eluted with an increasing NaCl gradient. The fractions collected were assessed for transglutaminase activity and protein levels. The purification course on the AKTA is shown in Figure 5.7 alongside the activity profile across eluted fractions.

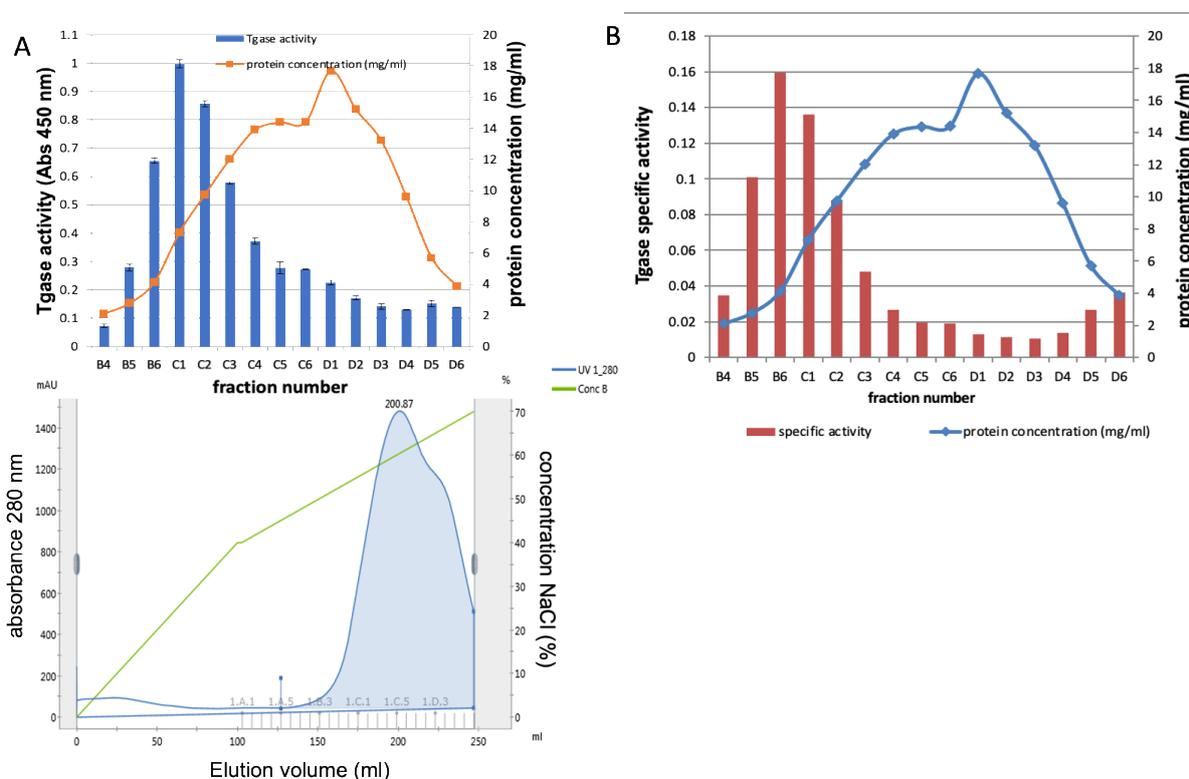


Figure 5.7. Purification of crude enzyme extracts from sporulating *C. difficile* on 100 ml Q-Sepharose FF. Crude extracts obtained following solubilisation of sporulating cells at 25°C for 2 h in 50 mM Tris, pH 8 were purified on Q-Sepharose FF columns. Anion exchange purification was performed on an AKTA system, and protein absorbance was monitored at 280 nm. The bound transglutaminase protein was eluted with an increasing linear NaCl gradient (0-2 M) in Tris buffer. Transglutaminase activity in the eluted fractions was detected at 450 nm using the BTC incorporation assay as previously described. (A) AKTA purification profile and corresponding **transglutaminase activity** across peak (green line: increasing NaCl gradient, blue line: protein absorbance detected at 280 nm). (B): **Specific activity** of eluted fractions calculated as activity per mg protein. Protein concentration is plotted on the right y-axis and transglutaminase specific activity is plotted on the left y-axis.

The chromatogram in Figure 5.7A shows the AKTA purification profile of crude extracts on Q-Sepharose columns, and the corresponding activity obtained from eluted fractions. Purification as monitored on AKTA systems reveal that transglutaminase activity was recovered in the front of the elution peak corresponding to 50-57% (v/v) of a 2 M NaCl elution gradient. The active fractions are contained in fractions B4-C3 as illustrated by the activity peak. Next the effect of various salt preparations on active fractions was evaluated to determine the most appropriate salts for HIC purification.

5.5.1.1 Effect of various salts on partially purified enzyme activity

Pooled active fractions from IEX purification were prepared with several salts and transglutaminase activity in each salt preparation was measured and compared with untreated samples. This experiment was to evaluate the most suitable salts for purification on selected HIC resins. The bar chart in Figure 5.8 shows the residual activity measured after testing various salts on partially purified enzyme activity.

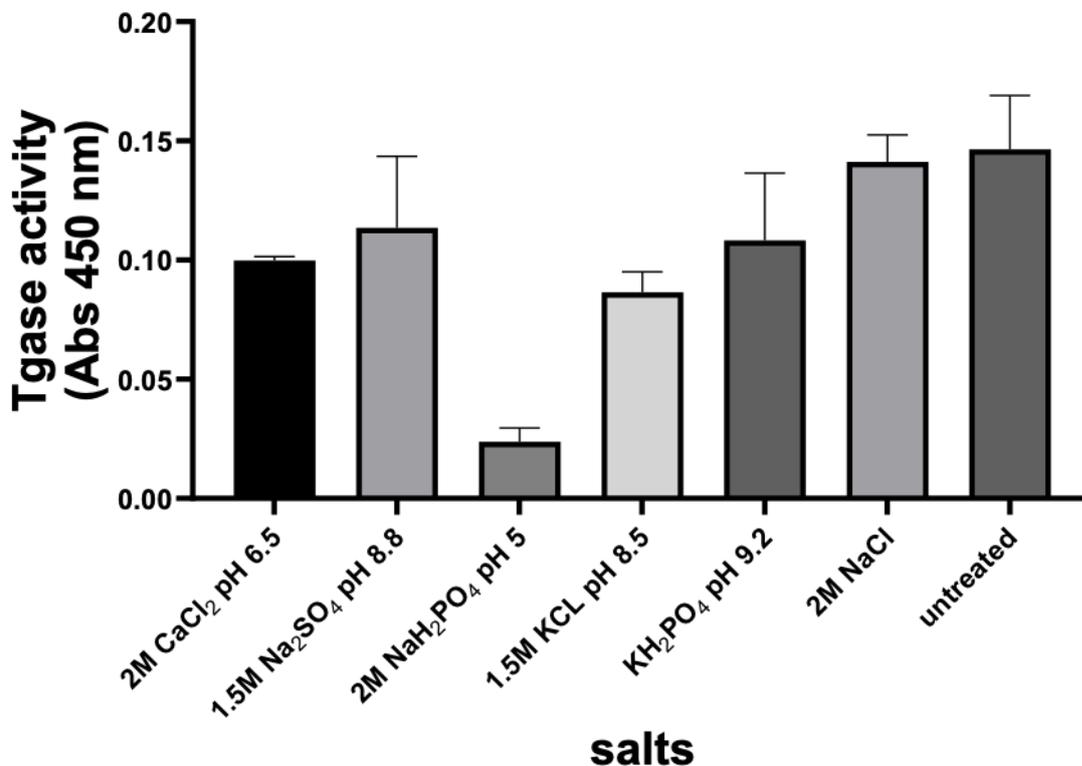


Figure 5.8. Effect of various salts on transglutaminase activity from *C. difficile* spores. After purification on Q-Sepharose columns, pooled active fractions were prepared in salt solutions at different concentrations. After incubation for 1 hour residual activity was measured by the BTC incorporation assay as described previously. Results are presented as mean \pm SEM of two independent repeats. There was a statistically significant difference between the untreated and NaH₂PO₄ samples ($p=0.01$).

The results in Figure 5.8 show that all the salts tested inhibited enzyme activity at the concentrations used. Furthermore, the effects of desalting and lower NaCl concentrations on transglutaminase activity were evaluated.

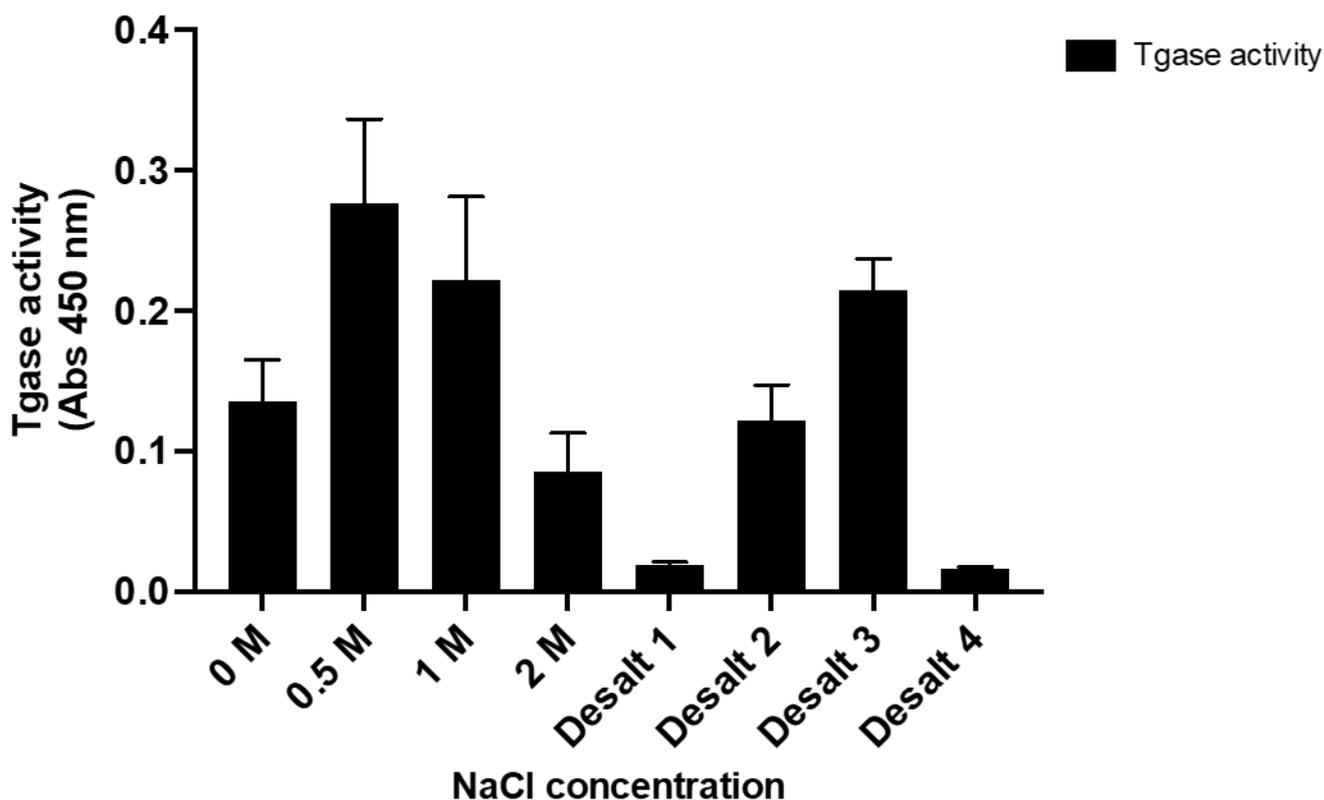


Figure 5.9. Effect of NaCl on pooled transglutaminase activity and recovery of activity after desalting. Pooled active fractions from IEX purification were incubated with different concentrations (0-2 M) of NaCl at 25°C for 1 hour and residual activity was measured using the BTC incorporation assay as described in Materials and methods. Furthermore, fractions treated with 2 M NaCl were desalted in Tris buffer using HiTrap desalting columns, and transglutaminase activity was measured in the eluted fractions. Results are presented as mean±SEM of three independent repeats. There was no statistical significant difference between the different samples.

As shown in Figure 5.9 NaCl was added to active fractions at different concentrations and activity increased with increasing salt concentrations up to 1 M. The enzyme activity at 1 M was higher than at 0 M though reduced compared to 0.5 M. However, a reduction in enzyme activity was observed at 2 M NaCl compared to untreated samples. Interestingly the reduction in activity was recoverable after desalting as seen in desalted fractions 2 and 3. Henceforth NaCl was chosen for purification on HIC columns given its availability in the laboratory, and importantly the recoverability of activity following inhibition. The initial HIC purification experiments consisted of pilot studies involving screening different hydrophobic resins to identify the most suitable for our study. As shown in Figure 5.10, butyl Sepharose was the most suitable because it generated good separation and yield activity compared to other resins (Appendix 7).

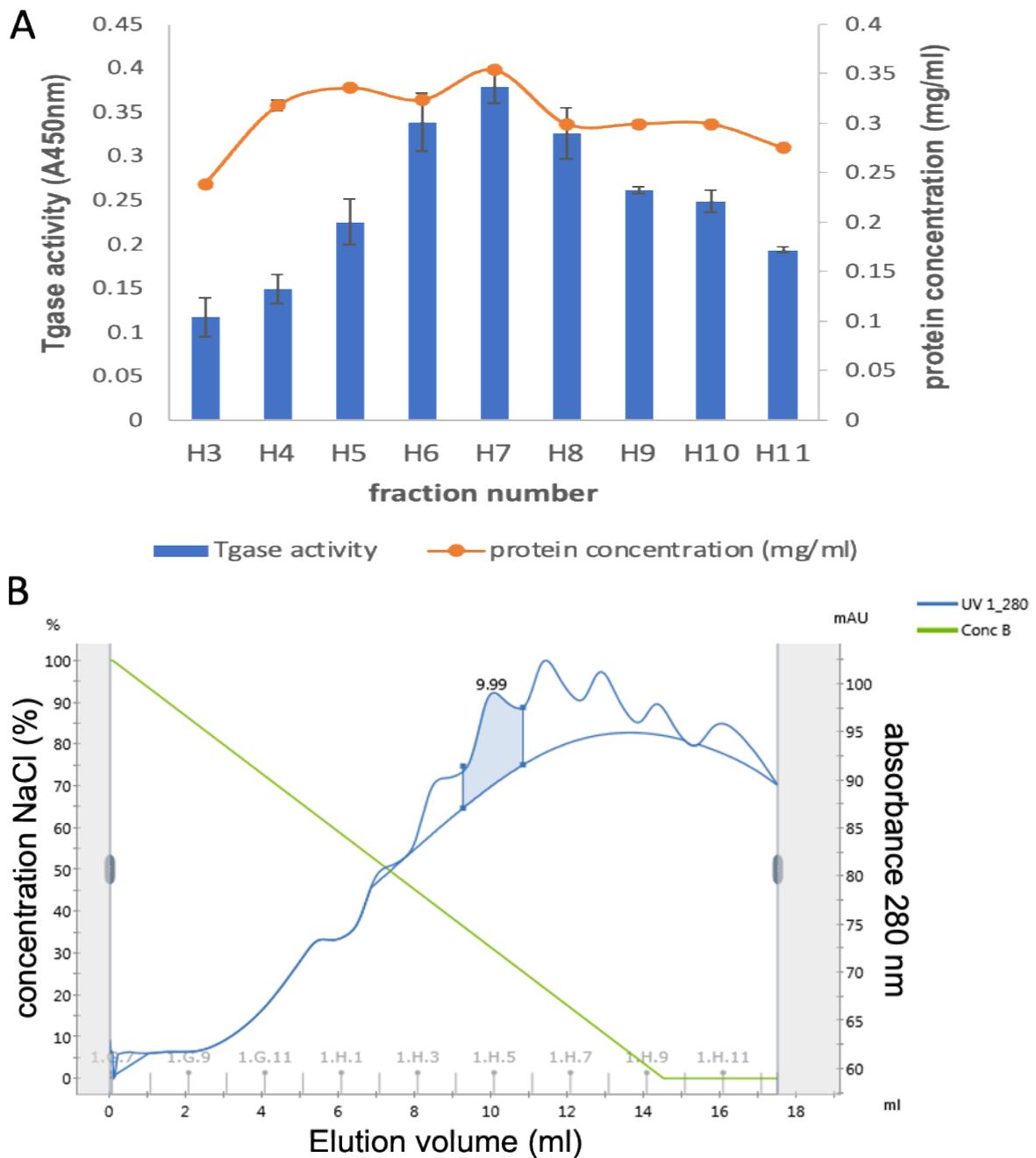


Figure 5.10. Screening of butyl Sepharose columns for purification of active transglutaminase fractions. Active transglutaminase fractions obtained from Q-Sepharose columns were pooled, adjusted to 2 M NaCl and purified on butyl Sepharose resins. Hydrophobic interaction purification was performed on an AKTA system and protein absorbance was monitored at 280 nm. The bound transglutaminase protein was eluted with a decreasing linear (100-0%) NaCl gradient from 2 M in Tris buffer. (green line: decreasing NaCl gradient, blue line: protein absorbance detected at 280 nm, red line: conductivity) (A) Transglutaminase activity of eluted fractions. Protein concentration is plotted on the right y-axis and transglutaminase activity on the left y-axis. (B) AKTA purification profile (green line: decreasing NaCl gradient, blue line: protein absorbance detected at 280 nm).

From Figure 5.10A a distinct peak of activity was obtained from butyl Sepharose resins and thus Butyl Sepharose was adopted as the resin for HIC purification. The resolution was low following purification on other screened resins (octyl Sepharose, phenyl-Sepharose high sub, low sub, high performance) shown in Appendix 7. No protein bound to Butyl-S Sepharose columns, and all the activity was found in the flow-through.

5.5.2 Purification of active fractions by hydrophobic interaction chromatography

The active fractions from Q-Sepharose columns were pooled and adjusted to 2 M NaCl prior to loading onto Butyl Sepharose columns. The purification was monitored on the AKTA PURE and eluted fractions were tested for transglutaminase activity as shown in Figure 5.11.

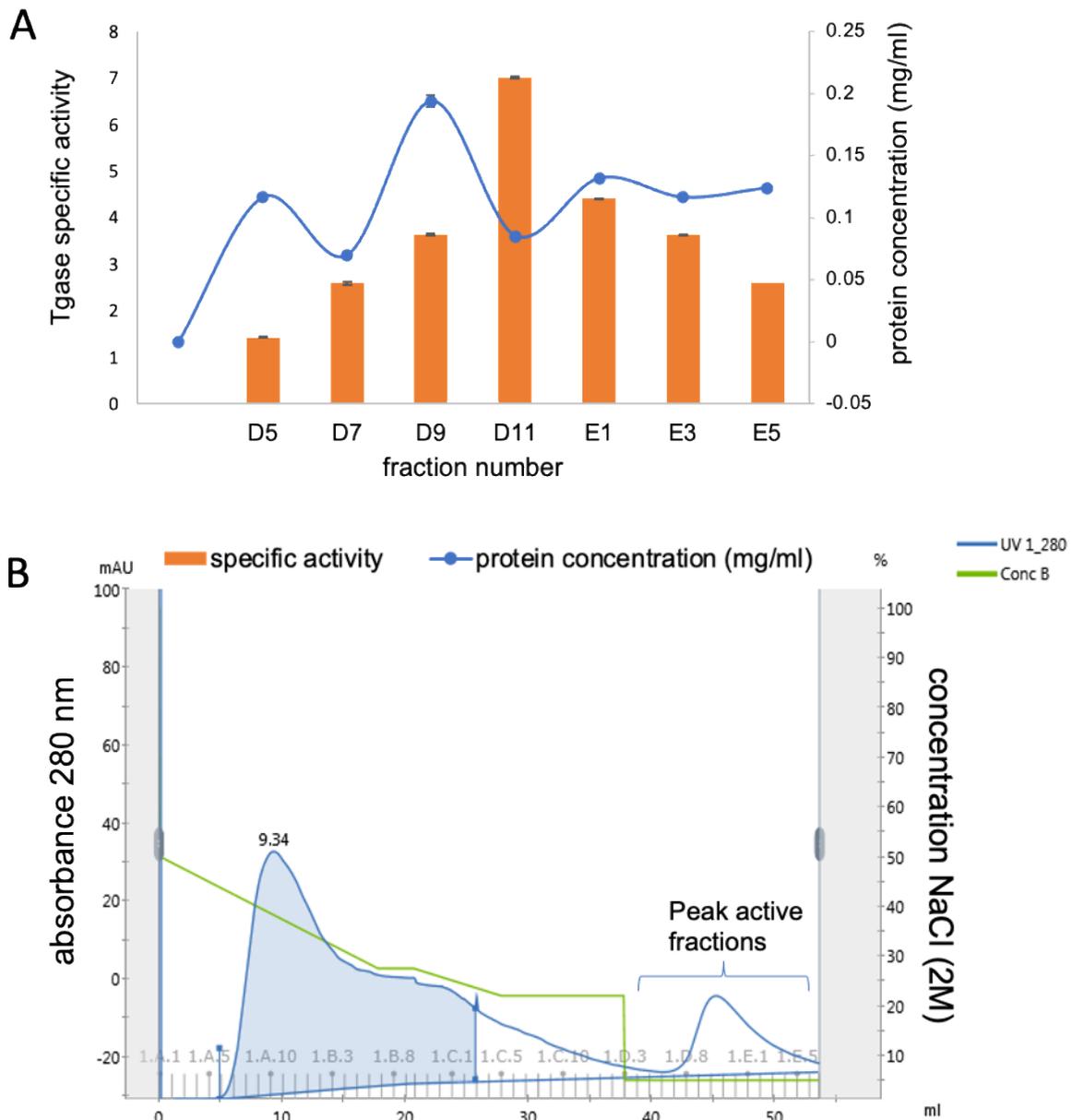


Figure 5.11. Purification of active transglutaminase fractions on Butyl Sepharose FF. Active transglutaminase fractions obtained from anion exchange purification were pooled, adjusted to 2 M NaCl and loaded on Butyl Sepharose columns. Hydrophobic interaction purification was performed on the AKTA system and protein absorbance was monitored at 280 nm. The bound transglutaminase protein was eluted with a decreasing linear (100-0%) NaCl gradient from 1 M in Tris buffer. Transglutaminase activity in eluted fractions was detected at 450 nm using the BTC incorporation assay as previously described. (A) AKTA purification profile (green line: decreasing NaCl gradient, blue line: protein absorbance detected at 280 nm). (B): Specific activity of eluted fractions calculated as activity per mg protein. Protein concentration is plotted on the right y-axis and transglutaminase specific activity on the left y-axis.

Purification of pooled active fractions from Q-Sepharose columns on Butyl Sepharose FF was performed and the bound protein was eluted with a decreasing NaCl gradient in 50 mM Tris. Following purification on Butyl Sepharose, transglutaminase activity eluted in a single peak from 100 mM NaCl corresponding to the active fraction peak as observed on the AKTA graph (figure 5.11). Following successful purification, the entire process and yield obtained at each stage was calculated as shown on Table 5.1.

Table 5.1. Summary of transglutaminase purification from crude spore extracts of *C. difficile* 630.

Procedure	Protein concentration (mg)	Total activity	Specific activity Abs/mg/h	Yield activity (%)	Fold purification
Crude enzyme	498.6	34.25	0.07	100	1
Q-Sepharose	70.8	16.38	0.23	47.8	3.3
Butyl Sepharose	1	4	4	11.7	57.1

The results in Table 5.1 summarise the purification scheme of transglutaminase from sporulating *C. difficile* 630 cells. The transglutaminase enzyme was successfully released in the soluble fraction after incubation of sporulating cells in 50 mM Tris, pH 8 at 25°C. The crude enzyme extract (125 ml) obtained from 12 g wet spores was loaded onto Q-Sepharose FF columns. No transglutaminase activity was identified in the flowthrough implying all the enzyme was bound to the resin. The final step purification on butyl Sepharose was successful and led to an increase in specific activity by 17 times. Overall microbial transglutaminase from solubilised spores was partially purified by 57-fold to obtain a yield activity of 12% corresponding to two major bands on protein gels as observed below. After purification samples were analysed by SDS-PAGE to view the profile of purified proteins as shown in Figure 5.12.

5.6 SDS-PAGE analysis of purified transglutaminase from crude extracts of *C. difficile* 630

The SDS-PAGE gel in Figure 5.12 shows the profile of extracted proteins and eluted fractions following purification of crude extracts on Q-Sepharose and Butyl Sepharose columns.

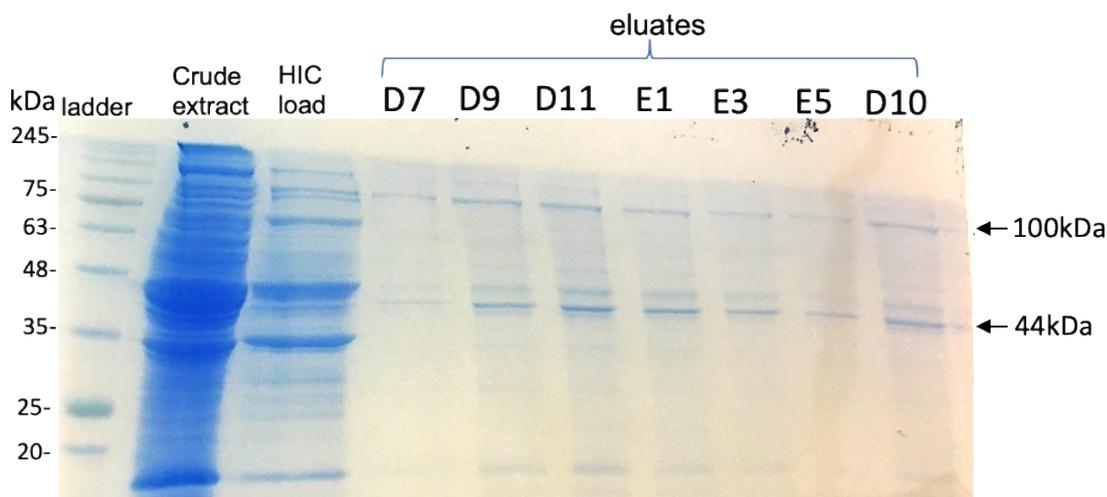


Figure 5.12. SDS-PAGE analyses of purified active enzyme fractions from Butyl Sepharose columns. Active transglutaminase fractions obtained from Q-Sepharose columns were pooled, adjusted to 2 M NaCl and purified on Butyl Sepharose columns. Eluted fractions were analysed by SDS-PAGE on 10% (v/v) gels and protein bands were visualised on a Genebox after Coomassie blue staining. (W: wash, FT: flow-through, HIC: hydrophobic interaction chromatography).

Eluted fractions were analysed by SDS-PAGE and transglutaminases from sporulating *C. difficile* 630 cells were partially purified into two major bands with a purity of 38%. Purity was calculated as the sum of the area under the peaks of target protein bands divided by the total area of the peaks in the lane expressed as a percentage. Multiple bands were obtained from Q-Sepharose purifications, but 2 major bands were identified following purification on Butyl Sepharose FF as observed in eluted fractions (Figure 5.12). There was relatively little purification after loading on Q-Sepharose columns as indicated by low specific activity and impurities in the eluted fractions on protein gels. Based on the protein gels, the molecular weight of the partially purified enzyme is approximately 100 kDa and/or 44 kDa.

5.7 Characterization of purified enzymes from *C. difficile* 630 spores

C. difficile spore enzyme extracts were partially purified on Q-Sepharose and Butyl Sepharose columns respectively. The purified endogenous Tgase-like proteins were characterized by determining optimal activity temperatures, thermostability and effect of inhibitors on enzyme activity.

5.7.1 Optimal temperature of purified spore protein activity

To determine the optimal activity temperature of purified enzymes, transglutaminase activity was measured following biotin cadaverine (BTC) incorporation at different temperatures (0-45°C) for 1 hour as described in the Methods section. The graphs in Figure 5.13 illustrate the temperature profile of purified enzyme activity.

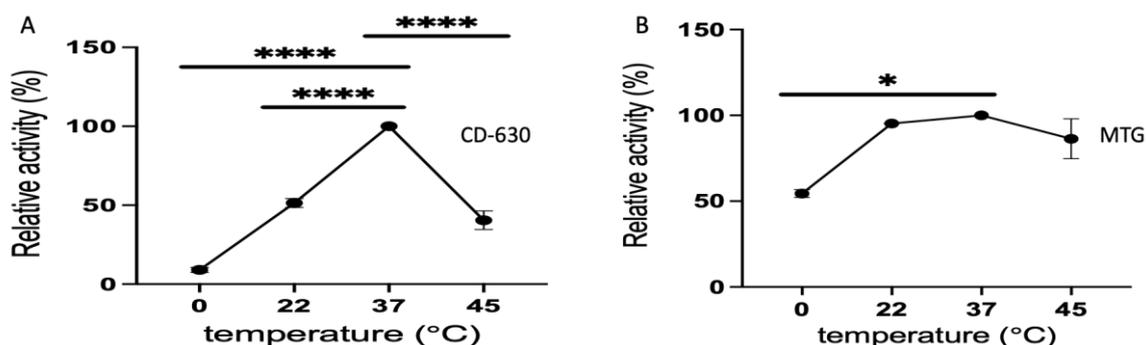


Figure 5.13. Effect of temperature on activity of purified transglutaminase from *C. difficile* 630 spores. Purified enzyme solutions were incubated in casein plates at different temperatures (0°C, 22°C, 37°C, 45°C) for 60 min after addition of substrates. Transglutaminase activity was measured under normal assay conditions. The activity obtained at each temperature is calculated as a percentage of the maximal activity detected at 37°C. Results are presented as mean±SEM of three independent repeats. (A) activity of purified transglutaminase (B) *Streptomyces* MTG activity. Asterisks (*) indicate statistical significance where $p < 0.05$.

The graph in Figure 5.13 illustrates that purified enzyme activity increased steadily with increasing temperatures till 37°C after which the activity dropped rapidly. Thus, the optimal temperature for spore protein activity was around 37°C meanwhile commercial MTG displayed optimal activity between 22-45°C. There was significant reduction in purified enzyme activity at 45°C leading to 40% activity retention. However, there was almost no activity detected at 0°C which suggests that extremely low temperatures do not favour BTC incorporation. Equally a significantly lower MTG activity was observed at 0°C. After identification of optimal activity temperatures for purified spore transglutaminase, the thermostability of purified spore proteins was evaluated.

5.7.2 Thermostability of purified spore enzymes

The thermostability of purified *C. difficile* 630 transglutaminase was evaluated between 0-45°C and activity was determined after 15 min preincubation at these temperatures. The graphs in Figure 5.14 illustrate the temperature profile of residual enzyme activity after preincubation.

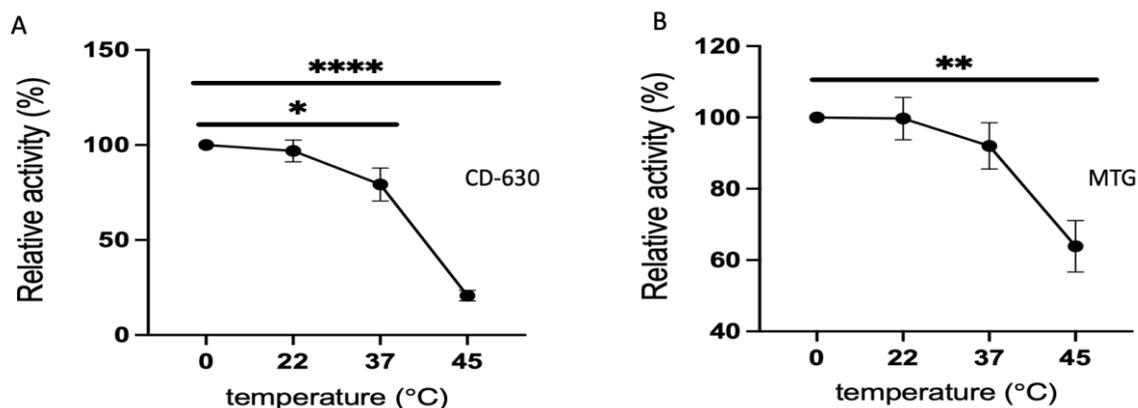


Figure 5.14. Thermostability of purified transglutaminase activity from *C. difficile* 630 spores. Purified transglutaminase in 50 mM Tris pH 8 was incubated at different temperatures (0°C, 22°C, 37°C, 45°C) for 15 min, then chilled on ice. Next transglutaminase activity was measured under normal assay conditions after addition of substrates. The relative activities were calculated by taking the maximum activity at any given temperature as 100%. Results are presented as mean±SEM of three independent repeats. (A) activity of purified transglutaminase (B) *Streptomyces* MTG activity. Asterisks (*) indicate statistical significance where $p < 0.05$.

From Figure 5.14 the spore transglutaminase activity was thermostable between 0°C to 37°C as greater than 80% activity was preserved at temperatures below 37°C. However significant loss in activity was recorded at 45°C accounting for more than 80% of initial activity. This differs from commercial MTG that was thermostable from 0-45°C and retained over 50% activity at 45°C. Following determination of enzyme thermostability, further characterization involved assessing the effects of specific inhibitors on enzyme activity.

5.7.3 Effect of various inhibitors and metal ions on purified enzyme activity

A variety of enzyme inhibitors were tested for their effects on activity of purified enzymes in comparison with *Streptomyces* MTG control. The bar charts in Figure 5.15 illustrate the residual enzyme activity following preincubation with several enzyme inhibitors.

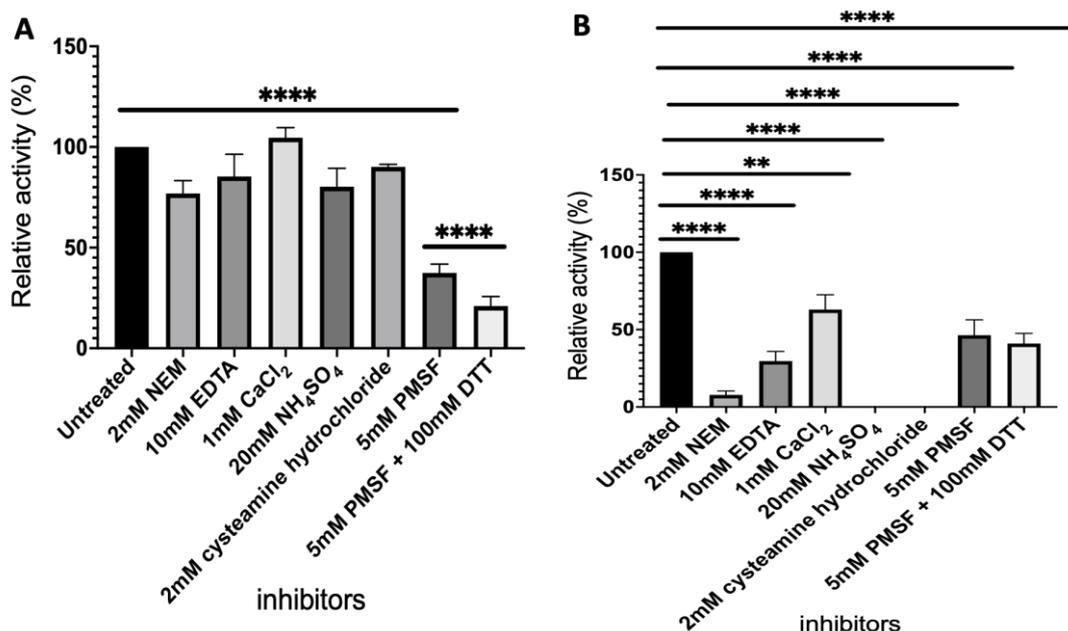


Figure 5.15. Effect of ions and inhibitors on purified transglutaminase activity. Purified transglutaminase protein was incubated with metal ions and inhibitors at different concentrations for 30 minutes on ice, and transglutaminase activity was measured by the BTC incorporation assay. The residual activity was calculated as a percentage considering activity in untreated samples as 100%. Results are presented as mean±SEM of three independent repeats. (A) activity of purified transglutaminase (B) *Streptomyces* MTG activity. Asterisks (*) indicate statistical significance where p<0.05.

The remaining activity in purified spore enzyme was measured after treatment with several potential inhibitors and metal ions at specific concentrations. Concentrations ranging from 2-20 mM were used and activity was determined following 30 min preincubation on ice. Transglutaminase activity from purified spores was considerably inhibited by 5 mM PMSF leading to more than 50% reduction in activity. Simultaneously, 0.1 M DTT was added to PMSF-treated samples to evaluate recovery of activity. Addition of DTT to PMSF-treated samples led to significant reduction in activity when compared with PMSF-treated samples. Cysteine modification by PMSF is reversible by DTT whereas serine modification is irreversible (Nur'amaliyah, Zilda and Mubarik, 2016). *Streptomyces* MTG in our experiment was significantly inhibited by 20 mM NH₄SO₄, 2 mM cysteamine dihydrochloride, 2 mM NEM and 10 mM EDTA. Actually, MTG activity was completely abolished by NH₄SO₄ and cysteamine dihydrochloride but moderately inhibited by PMSF and 1 mM CaCl₂. Similar to observations in spore transglutaminase activity DTT treatment further inhibited MTG activity. Finally, candidate substrates (casein and BSA) were selected for reaction with purified proteins to evaluate evidence of crosslink formation indicating transglutaminase activity using SDS-PAGE analysis.

5.7.4 Purified transglutaminase from *C. difficile* spores does not crosslink casein substrate

The SDS-PAGE gel in Figure 5.16 shows the profile of crosslinked proteins after incubation of MTG, trypsin and purified transglutaminase with casein substrate.

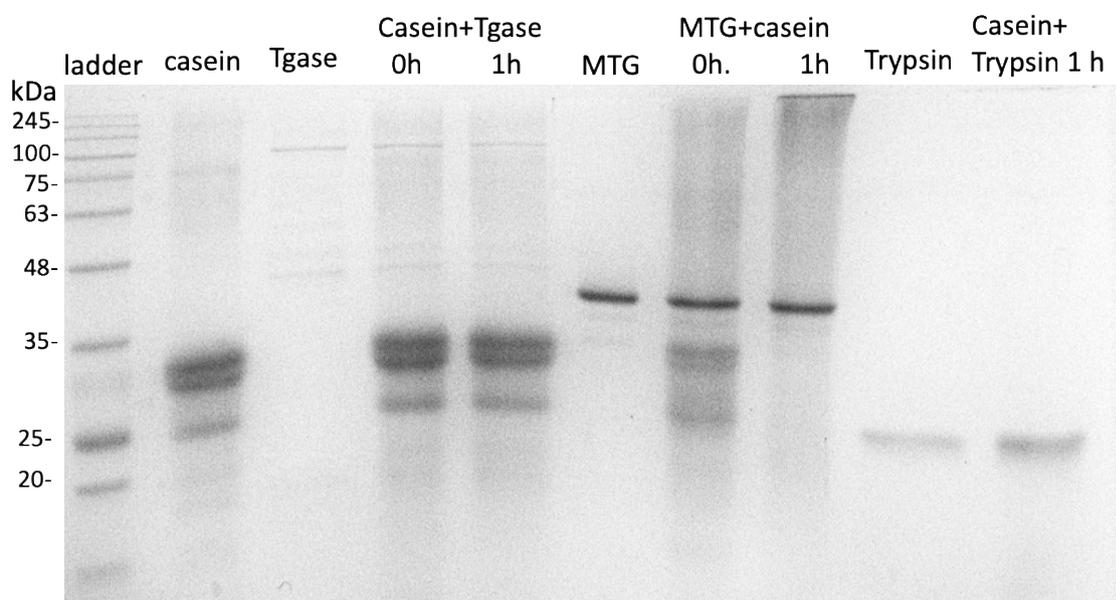


Figure 5.16. SDS-PAGE showing reaction of purified protein with casein substrate. Casein (0.5 mg/ml) was prepared with purified spore protein (0.08 mg/ml) in 50 mM Tris, pH 8 and incubated for 1 hour at 37°C. MTG (100 µg) and trypsin (0.4 mg/ml) were used as positive controls for transglutaminase and protease activity respectively. Following incubations, samples were analysed by SDS-PAGE and gels were visualised after Coomassie blue staining.

From the gel in Figure 5.16 no difference in the casein band is observed in the purified enzyme treated lanes. Indeed, no crosslinks were found following reaction of casein substrate with purified enzymes. By contrast casein was crosslinked by MTG after 1 hour as illustrated by the high molecular mass products at the top of the lane. We can also notice disappearance of the casein band in the MTG treated lanes confirming crosslinking by MTG. On the other hand, casein was digested by trypsin as shown by the absence of casein bands in trypsin-treated lanes. Therefore, our purified enzyme showed neither transglutaminase activity nor proteolytic activity.

5.7.5 Mass spectrometry analysis of purified transglutaminase from *C. difficile* 630 spores

The partially purified spore enzyme was separated as two bands on SDS-PAGE gels corresponding to approximately 44 kDa and/or 100 kDa proteins. The two protein bands were excised from gels and sent for mass spectrometry analysis for protein identification.

From the sequences obtained from mass spectrometry analysis, protein data bases (PDB) were searched against *C. difficile* 630 complete genome to identify possible matches. Matches were found but target proteins contained few modifications. Table 5.2 shows the most abundant matching peptides identified following mass spectrometry analysis of purified protein bands.

Table 5.2. Representative highly abundant proteins identified by mass spectrometry analysis.

Matching peptide identified	abundance (%)	
	44 kDa	100 kDa
Glycolysis: Glyceraldehyde-3-phosphate dehydrogenase	18	8
Butyrate metabolism:		
Acyl-CoA dehydrogenase	29	20
Electron transfer flavoprotein. Alpha subunit	20	14
butyryl-CoA dehydrogenase	23	17
Pyruvate ferredoxin oxidoreductase	26	53
Aldehyde alcohol dehydrogenase	31	64
Putative oxidoreductase	3	48
Amino acid group metabolism: Putative aminotransferase	23	18

From each band analysed, two putative *C. difficile* 630 transglutaminases were found but with low coverage using CD-720, CD-1062 and CD-1083 as matches. In the 44 kDa band both CD-720 and CD-1083 transglutaminases were identified with respective coverage of 11% and 12%. Meanwhile in the 100 kDa band, CD-720 and CD-1083 transglutaminases were identified with respective coverage of 8% and 6%. Lower coverage of target putative transglutaminases was found in the 100 kDa protein band. Importantly several modifications were observed in the purified target proteins. Notably in the 44 kDa band, the CD-1083 protein had modifications in three regions including the putative transglutaminase active site. The active site was modified by phosphorylation plus carbamidomethylation and the other two regions were modified by carbamidomethylation, phosphorylation and deamidations. Equally the CD-1083 protein from the 44 kDa band presented a modification by phosphorylation in just

one region of the complete sequence. Similarly, the CD-1083 and CD-720 proteins identified in the 100 kDa band contained major modifications. The N-terminus of the CD-720 protein was modified by phosphorylation and deamidation. Meanwhile a region within the N-terminus in CD-1083 was modified by carbamidomethylation, phosphorylation and oxidations.

In addition to the very low abundance putative target transglutaminases, other proteins were found in the purified protein bands with high abundance and grouped according to their functions in *C. difficile* as shown in Table 5.2. A putative aminotransferase was found with high abundance and three other aminotransferases with low abundance. Dehydrogenases were another class of highly represented proteins amongst which acetyl-CoA dehydrogenase, aldehyde alcohol dehydrogenase and butyryl dehydrogenase which were highly abundant. Interestingly some of the dehydrogenases have sizes corresponding to the size of our target proteins. Also, oxidoreductases were highly expressed.

5.8 DISCUSSIONS

In *Bacillus* species a link has been established between transglutaminase activity and sporulation, elaborated by the presence of transglutaminase activity only in strains that sporulated efficiently (Katsunori Kobayashi *et al.*, 1998). We studied the expression of transglutaminase activity during sporulation in *C. difficile* 630. Transglutaminase activity was present in sporulating cells and culture supernatants but expressed at different times during sporulation. In our study transglutaminase activity was first detected in sporulating cells grown on agar at a time corresponding to the end of the exponential phase in liquid culture and increased till 72 hours of sporulation. This corroborates similar studies in sporulating *B. subtilis* AJ1307 where transglutaminase activity was detected after termination of exponential growth and increased continuously to reach a maximum after 5 hours (Suzuki *et al.*, 2000). Meanwhile in sporulating *B. subtilis* AJ12866, transglutaminase activity was first detected 5 hours after the end of exponential growth with further increases onwards (Katsunori Kobayashi *et al.*, 1998). This timeline corresponds to that of transglutaminase production from culture supernatants in our study where activity was first detected around 6 hours post-exponential phase. Maximal transglutaminase activity was detected in agar cultures at 9 hours of sporulation and after 48 hours of sporulation in culture supernatants. These observations match with previous studies in *C. difficile* that reported maximal sporulation in BHI or BHIS broth within 24-48 hours of sporulation (Dembek *et al.*, 2013). This suggests maximum transglutaminase activity possibly ties with maximum sporulation as reported in other studies. Culture on solid media has been shown to generate higher activity compared to liquid media contributed by desiccation associated with agar culture. Another advantage is reduced lysis observed with cultures on solid media and more homogeneity in the sporulation process in comparison to sporulation in liquid media (Pereira *et al.*, 2013; Putnam *et al.*, 2013). Agar plates are easier to handle in the anaerobic cabinet in addition to their faster activity generation, and thus more protein can be obtained during the same time required for maximal enzyme production in broth. Consistent with these reports and judging by our results the most suitable method for high yield transglutaminase production was on agar.

Additionally, transglutaminase activity was found in vegetative cells of *C. difficile* 630 (appendix 5). This differs from studies in *B. subtilis* where no transglutaminase activity was found in vegetative cells, culture supernatants or washed unlysed cells during sporulation. In their study, detection of transglutaminase activity was performed on lysed cells (Katsunori Kobayashi *et al.*, 1998). Similarly in our study transglutaminase activity was found in lysed cells, and washed unlysed cells (appendix 6), although determination of activity was routinely done on unwashed unlysed cells. After determining the period for maximal transglutaminase

production during sporulation, we sought to investigate the conditions for optimum enzyme solubilisation.

Protein solubility can be influenced by extrinsic factors such as temperature, pH, ionic strength, solvents, and improved solubility can be attained by alterations in these extrinsic factors (Kramer *et al.*, 2012). In the present study pH 10 seemed to be the most efficient for transglutaminase extraction when compared to other pH, even though similar gradients were obtained for pH 6 and pH 8 (Figure 5.3). This could suggest the importance of pH in promoting protein solubility. Even so, variations in extrinsic factors may not be sufficient to increase protein solubility (Kramer *et al.*, 2012). For instance, high temperatures disrupt hydrogen bonds in the protein and establishment of new hydrogen bonds during this process with further heating breaks the protein helical structure (Kramer *et al.*, 2012). Indeed transglutaminase activity was completely abolished after incubation of crude enzyme extracts at 50°C for 30 min (Figure 5.5). Activity inhibition at high temperatures was indicative of a real enzyme. On the other hand, intrinsic factors mainly determined by surface expressed amino acids can also play a role in protein solubility (Kramer *et al.*, 2012). High levels of salt negatively affect protein structure and thus protein functionality. The high amount of salt affects hydrogen bonding which then influence protein solubility, stability, and binding (Sinha and Khare, 2014). Salt bridges are the major stabilisers of ionic interactions amongst surface exposed charged amino acids in proteins. Disruptions in salt bridges can be caused by the pH of buffer solutions (Sinha and Khare, 2014). From our results higher salt concentrations resulted in gradual loss of activity. Protein activity was greatest with dilute salt concentrations and decreased at higher concentrations (Figure 5.9). This is because with dilute salt preparations, increased surface hydration is achieved following interactions of ionic charges in salts with other charges inside the protein (Sinha and Khare, 2014). Meanwhile at higher salt concentrations, both proteins and salt ions compete for hydration and due to higher salt affinity, proteins are stripped of water resulting in denaturation (Sinha and Khare, 2014). Likewise, detergents and surfactants (Tween, SDS, Triton) in alkaline buffers used for protein solubilisation led to complete loss of activity (Figure 5.4). This suggests that protein denaturation by these agents resulted in abolition of activity. It is well known that SDS is a protein denaturant and denatures proteins by disrupting the tertiary structure. Therefore, the effect of SDS was not surprising. Accordingly physical and chemical factors alter structural properties of proteins which could denature the protein resulting in loss of biological activity (Sinha and Khare, 2014).

Transglutaminase activity was processed from agar plates instead of culture supernatants due to the ease of handling agar plates. Broth cultures usually involve large volumes which may be difficult to handle within the anaerobic cabinet. Additionally, several steps are required to

process the broth to obtain solubilised protein for purification. These may involve centrifugation, protein precipitation, and dialysis. With each of these steps there are possibilities of protein loss and activity depletion. Importantly less time was required to obtain maximal activity with agar cultures. Consequently, agar plates were sampled for transglutaminase activity for purification experiments. During sporulation, maximum transglutaminase activity from agar cultures occurred between 9-48 h (Figure 5.1). Thus, sporulating cells were collected at 24 hours for enzyme solubilisation. From our results, higher temperatures used during enzyme solubilisation resulted in gradual loss of activity (Figure 5.2). Enzyme solubilisation decreased with increasing temperatures and complete inactivation was obtained at 50°C. This contrasts with studies in *B. subtilis* where more enzyme was solubilised with increasing temperatures, and maximum solubilisation was observed at 50°C, though accompanied with enzyme inactivation (Suzuki *et al.*, 2000). Indeed, protein denaturation is enhanced at higher temperatures and solubility decreases at higher temperatures contributed by protein denaturation (Veide Vilg and Undeland, 2017).

In our study mild acidic or alkaline conditions were sufficient to solubilise spore proteins. The pH range used for spore solubilisation seemed efficient because the gradients followed the same trend (figure 5.3). It is true that pH 7.5 seemed to be the most effective for transglutaminase solubilisation from spores at low temperatures (0-22°C) (Figure 5.2). This contrasts with *Bacillus* in which no solubilisation was observed using these conditions (Suzuki *et al.*, 2000). Maximal solubilisation at pH 7.5-8 occurred between 0°C and 22°C, which suggests that lower temperatures favoured enzyme solubilisation that also increased with time. Therefore, it is probable that temperature and pH play a synergistic role on enzyme solubilisation. The scenario is different in *B. subtilis* where maximal solubilisation was achieved at higher temperatures (37°C or 50°C) and pH values (Suzuki *et al.*, 2000). The exact mechanism of enzyme solubilisation is not known and needs to be researched further. It is suggested that protein solubilisation could be increased with higher volumes of extraction buffer. Veide Vilg and Undeland, (2017) published that an increase in extraction liquid leads to increased yields due to greater solubility achieved with more volume. Additionally, more dilution causes retention of insoluble proteins in the pellet (Veide Vilg and Undeland, 2017). We did not evaluate the effect of increased buffer volumes on protein solubilisation. For purification experiments, pH 8 was selected given that pilot purification studies on IEX columns were done at pH 8. Accordingly, prior to purifications, enzyme solubilisation was performed at pH 8, 22°C based on the results obtained from solubilisations at different temperatures.

5.8.1 Purification of transglutaminase from sporulating *C. difficile* 630 cells

Sporulating cells were harvested when transglutaminase activity was maximal and transglutaminase was solubilised at optimal pH/temperature (pH 8/25°C) prior to purification. Standard purification methods combine IEX, HIC and SEC in a two-step or three-step strategy. Mostly microbial transglutaminase purifications from fermentation culture supernatants have been carried out in two-step procedures involving cation-exchange plus gel filtration (Jin *et al.*, 2016). The initial purification by anion-exchange chromatography on Q-Sepharose generated a yield activity of 48% and 3-folds purification. This is comparable to the first step purification of crude enzymes from soil *Streptomyces* on Sephadex columns that resulted in 48% activity recovery plus 2-folds purification (Macedo, Sette and Sato, 2011). Similar to our study, cation exchange purification of MTG from *S. mobaraensis* produced a yield activity of 49% accompanied with 5 folds increase in specific activity (Jin *et al.*, 2016). Our final purification step on Butyl Sepharose was the most efficient. The protein was eluted with 100 mM NaCl, and an increase in specific activity by 17 times compared to IEX chromatography was achieved (Table 5.1). In line with our results, Nur'amaliyah and colleagues also reported an increase in specific activity by 11 times following a final purification on gel filtration columns compared with prior purification on Q-Sepharose (Nur'amaliyah, Zilda and Mubarik, 2016). Though our final purification generated lower yields of activity, the enzyme was purified when compared to crude extracts as shown on SDS-PAGE gels (Figure 5.12). Past studies have included SEC in the purification scheme. However, further purification of active fractions from Butyl Sepharose columns onto Superdex columns were unsuccessful. The protein yields from Butyl Sepharose active fractions were low, and activity was lost after protein concentration by ultrafiltration. Additionally, loading Superdex columns with low concentration enzyme solutions resulted in absence of activity and protein bands following purification. Two protein bands resulted from the final purification on Butyl Sepharose corresponding to approximately 44 kDa and 100 kDa proteins. It is surprising that two bands resulted from purified transglutaminase given that a single peak of activity was obtained corresponding to the protein elution peak on the AKTA (Figure 5.11). In most studies, purified microbial transglutaminases were separated as single bands following SDS-PAGE (Ando *et al.*, 1989; Cui *et al.*, 2007; Jin *et al.*, 2016; Nur'amaliyah, Zilda and Mubarik, 2016). The purified 44 kDa band was smaller than mammalian transglutaminases (Cui *et al.*, 2007) but bigger than *B. subtilis* Tgl (29 kDa), MTG from *S. mobaraensis* (37 kDa), and *S. hygroscopicus* (38 kDa). Meanwhile the purified 100 kDa band was bigger than purified transglutaminases from most microbial and mammalian sources (Ando *et al.*, 1989; Suzuki *et al.*, 2000; Cui *et al.*, 2007; Zhang *et al.*, 2012).

5.8.2 Characterization of transglutaminase purified from sporulating *C. difficile* 630

The purified proteins had 38% purity, with molecular weights of 44 kDa and/or 100 kDa. Protease activity assays using FTC-casein as substrate and trypsin control showed that enzymes were not proteases from the specific activity obtained (11 RLU/min/ μ g), which was less than that of trypsin standard (455 RLU/min/ μ g). The transglutaminase activity in purified proteins was measured through colorimetric assays using *Streptomyces* MTG as reference for 100% activity and Tris negative control. Purified eluted protein fractions tested positive for transglutaminase activity. Actually, 0.1 mg/ml of purified *C. difficile* 630 transglutaminase exhibited activity that was higher than Tris control, corresponding to *Streptomyces* MTG activity of 0.125<MTG<0.25 μ g/ml. Therefore, the purified protein exhibited more of transglutaminase-like activity than protease activity. For initial characterization assays, transglutaminase activity was measured following incubation with biotin-cadaverine (BTC) substrate at various temperatures to identify the optimum temperature for maximal enzyme activity.

The effect of temperature on purified protein activity was determined by measuring activity following BTC incorporation for 1 hour at temperatures ranging from 0°C to 45°C and pH 8. The enzyme activity of purified protein increased steadily with increasing temperatures till it reached 37°C, after which the activity dropped rapidly (Figure 5.13). The optimal temperature of purified enzyme activity was 37°C meanwhile commercial MTG had peak activity between 22-45°C. Moderate spore enzyme activity was detected at 45°C. Similarly, the optimal enzyme activity for soil *Streptomyces* MTG was between 35-40°C using the hydroxylamine hydroxamate assay (Macedo, Sette and Sato, 2011). This is closer to our spore transglutaminase peak activity temperature and equally the commercial MTG used in this study. The optimal temperature for activity of purified *C. difficile* 630 transglutaminase is not very different to that for *Strv. ladakanum* transglutaminase with optimal activity at 40°C and *S. hygrosopicus* transglutaminase which displayed optimal activity in the range of 37-45°C at pH 6-7 using the hydroxamate assay (Ho *et al.*, 2000; Kashiwagi *et al.*, 2002; Cui *et al.*, 2007). Conversely some microbial transglutaminases had much higher optimal activity temperatures around 60°C and 55°C respectively for *B. subtilis* and *S. mobaraense* S-8112, DSM 40587 (Ando *et al.*, 1989; Suzuki *et al.*, 2000; Zhang *et al.*, 2012). Equally purified enzyme from *Streptomyces* sp TTA 02 SDS 114 had optimal enzyme activity at 50°C which is higher than that of *C. difficile* 630 spore transglutaminase and commercial transglutaminase used in this study. We did not determine enzyme activity above 45°C. It should be noted that transglutaminase activity in the current study was evaluated using the BTC incorporation

assay and incubation times of 1 hour with samples at pH 8. In most other studies, hydroxamate assays in the acidic pH range and reduced incubation times were used. After determination of the optimal temperature for enzyme activity, the thermostability of purified enzyme was investigated.

The thermostability of purified *C. difficile* 630 transglutaminase was evaluated between 0-45°C and transglutaminase activity was measured under standard conditions after 15 min preincubation at these temperatures. The purified *C. difficile* 630 spore transglutaminase was thermostable around 0-37°C, with greater than 80% activity being preserved in this interval (Figure 5.14). A similar observation was reported in MTG from *S. mobaraensis* DSM 40587 that retained over 90% activity at temperatures less than 40°C (Zhang *et al.*, 2012). The activity of purified MTG from *S. ladakanum* was maintained by greater than 60% at 37°C, which is slightly lower compared to our *C. difficile* 630 transglutaminase (Ho *et al.*, 2000; Kashiwagi *et al.*, 2002). Indeed, purified *C. difficile* 630 spore transglutaminase has lower thermostability compared to most known microbial transglutaminases (Ho *et al.*, 2000; Cui *et al.*, 2007; Macedo, Sette and Sato, 2011; Nur'amaliyah, Zilda and Mubarik, 2016). Additionally, the enzyme activity was almost abolished at 45°C with more than 80% reduction in activity. By contrast commercial MTG retained nearly 70% activity at 45°C. Analogous to the commercial MTG in this study, the transglutaminase from *S. ladakanum* was thermostable from 22-45°C and preserved about 70% activity at 45°C, meanwhile activity was almost abolished at temperatures above 45°C (Ho *et al.*, 2000). We did not investigate the thermostability at temperatures over 45°C. Like our commercial MTG, after 30 min preincubation MTG from soil *Streptomyces* sp was very thermostable between 4-45°C with about 80% activity retention but completely lost activity at higher temperatures around 60°C (Macedo, Sette and Sato, 2011). A similar pattern of enzyme thermostability was observed for microbial MTG from *S. hygrosopicus* where higher temperatures (60°C) prompted enzyme inactivation and temperatures less than 50°C maintained almost complete enzyme activity (Cui *et al.*, 2007). Also, the thermostability of *Streptomyces* sp TTA 02 SDS 14 was about the same with commercial MTG in this study though the former protein was thermostable for up to 2 hours (Nur'amaliyah, Zilda and Mubarik, 2016). This differs from the present study where enzyme thermostability was evaluated after 15 min preincubation. After determination of thermal stability of purified enzyme activity, the effect of selected inhibitors on enzyme activity was evaluated.

The residual activity of purified spore enzyme was measured after incubation with several inhibitors and metal ions at specific concentrations (Figure 5.15). The enzyme activity purified

from *C. difficile* 630 spores was weakly inhibited by ammonium sulphate (NH_4SO_4), cysteamine, NEM, EDTA, and unaffected by calcium chloride (CaCl_2). Meanwhile significant inhibition was observed with PMSF, and considerably higher inhibition following addition of 100 mM DTT. On the contrary commercial MTG in our experiments was greatly inhibited by NH_4SO_4 , cysteamine, NEM and EDTA, but moderately inhibited by PMSF and CaCl_2 . Notably all the MTG activity was lost within 30 min of incubation on ice following treatment with NH_4SO_4 and cysteamine. As observed in purified spore transglutaminase, DTT treatment further inhibited MTG activity. In other studies, *B. subtilis* Tgl activity was highly inhibited following treatment with 1 mM NEM, and completely abolished with 2 mM cysteamine and 10 mM NH_4SO_4 . This corresponds to the inhibition exhibited by NEM, NH_4SO_4 , and cysteamine on our commercial *Streptomyces* MTG. The high activity inhibition by NEM ties with other studies of microbial transglutaminases indicating the participation of thiol groups in the catalytic reaction. For instance, MTG-TX from *S. mobaraensis* was highly inhibited by NEM but unaffected by Ca^{2+} , Na^+ and PMSF. However, this novel MTG-TX differs from most microbial transglutaminases inhibited by PMSF (Nur'amaliyah, Zilda and Mubarik, 2016). Therefore, purified spore transglutaminase may not contain thiol groups at the active site based on the little-to-no inhibition with NEM. These results show that the participation of sulfhydryl groups in the activity of purified *C. difficile* 630 transglutaminase was almost non-existent in contrast to other microbial transglutaminases or even the commercial transglutaminase control. Notwithstanding our purified enzyme demonstrates some features of characterised microbial transglutaminases such as calcium-independence for activity, and inhibition by PMSF. The catalytic site of most characterised microbial transglutaminases contains cysteine residues suggested by significant inhibition with NEM (Nur'amaliyah, Zilda and Mubarik, 2016). Our experimental conditions for transglutaminase activity cannot account for the absence of significant inhibition with NEM, and consequently our purified protein is likely not a transglutaminase.

Little inhibition was detected following addition of EDTA to purified *C. difficile* 630 spore enzyme. This corroborates studies with transglutaminase purified from sporulating cells of *B. subtilis* in which EDTA was not inhibitory and calcium did not increase activity. This suggests that purified spore enzymes are not metalloproteases given the little inhibition by EDTA and the unaltered enzyme activity in presence of calcium ions. The fact that the purified enzyme does not require calcium for activity is an established feature of microbial transglutaminases as published in several studies (Macedo, Sette and Sato, 2011). Likewise, transglutaminase activity purified from culture filtrates of *Strv.* S-8112 was unaffected by EDTA and CaCl_2 , highly inhibited by NEM with little or no activity inhibition from NaCl (Ando *et al.*, 1989). In the same line, *S. mobaraensis* DSM 40587 was unaffected by calcium (Zhang *et al.*, 2012). Meanwhile

recombinant commercial *Streptomyces* MTG was significantly inhibited by calcium and EDTA under the same reaction conditions. This somewhat corresponds to observations in *B. subtilis* Tgl which was moderately inhibited by calcium (Suzuki *et al.*, 2000). Generally, transglutaminases from microorganisms are not affected by calcium and thus not inhibited by metal chelators like EDTA (Nur'amaliyah, Zilda and Mubarik, 2016). The high reduction in MTG activity with 10 mM EDTA may indicate the loss of cofactors essential for transglutaminase activity. Additionally, reduction in MTG activity with CaCl₂, NH₄SO₄, and cysteamine could evoke competitive inhibition by Ca²⁺ ions, NH₄⁺ ions, and cysteamine with BTC for incorporation into N, N-dimethyl casein plates. This is in line with studies in *B. subtilis* Tgl where enzyme activity was reduced following competitive inhibition by NH₄⁺ ions and cysteamine (Suzuki *et al.*, 2000). Metal ions could trigger conformational alterations on the enzyme that can enhance or diminish activity based on the amino acid moieties present at the active sites (Kieliszek and Misiewicz, 2014). After protein characterization with inhibitors, purified enzymes were treated with substrate solutions to identify crosslinks as evidence of pure transglutaminase activity.

Caseins are known substrates of MTG which they crosslink to form high molecular weight polymers (Hsieh and Pan, 2012). No protein crosslinks were identified after incubation of casein substrate and BSA solution with purified *C. difficile* 630 spore enzyme. Our commercial MTG was able to crosslink casein but not BSA forming high molecular weight products on the gel (appendix 8). Similarly casein was crosslinked by *S. hygroscopicus* MTG and the amount of casein reduced with time, forming high molecular mass polymers that accumulated at the top of the gels (Cui *et al.*, 2007). Meanwhile microbial transglutaminase purified from sporulating cells of *B. subtilis* AJ12866 could crosslink BSA substrate as shown by the high molecular mass products on the SDS-PAGE separation gels (Katsunori Kobayashi *et al.*, 1998; Suzuki *et al.*, 2000). Likewise purified transglutaminase from *B. subtilis* AJ1307 crosslinked protein substrate solutions of α -casein and BSA indicated by gel formation at 37°C (Suzuki *et al.*, 2000). Therefore, it is likely that our purified spore enzyme may not be a pure transglutaminase. Another reason for the absence of crosslinks is that casein may not be the right substrate for the protein, and the specificity of the reaction requires the right substrate.

5.8.3 Mass spectrometry analysis of purified spore proteins

A non-exhaustive list of the proteins identified by mass spectrometry (MS) was obtained from service providers and Table 5.2 reports only those with high abundance. We show only a few highly expressed proteins in each metabolism group. Following MS analysis, 111 and 209 proteins were identified respectively from the purified 44 kDa and 100 kDa protein bands.

Protein identification was based on detection of at least two peptides with a confidence of 1% false discovery rate. The major proteins purified as identified by mass spectrometry are metabolic proteins. These include proteins involved in butyrate metabolism, energy metabolism (glycolysis), metabolism of amino groups and oxidoreductases. Another author reported high levels of oxidoreductases in transcripts of dormant *C. difficile* spores (Dembek *et al.*, 2013). Also, ABC type transport proteins were found mostly in the 44 kDa band and these are usually associated with the cell surface. These results are not surprising given that metabolic proteins especially those involved in energy and amino acid metabolism were highly represented following proteomic analysis of *C. difficile* 630 in another study (Lawley *et al.*, 2009). We purified proteins mainly involved in energy metabolism which could explain the non-specific transglutaminase-like activity obtained following characterisation assays. The time of harvest may have influenced the class of proteins purified. The major proteins identified by MS suggests that during the period of harvest the cell was geared towards energy production.

There are several methods for determining transglutaminase activity each with a distinct principle. The first involves detecting the ammonia liberated following the formation of an acyl enzyme between transglutaminase enzyme and glutamine substrate (Katona *et al.*, 2012). The second type involves measuring transglutaminase catalysed amine incorporation into glutamine containing substrate proteins. This involves a covalent association between amine donor substrates (acyl acceptor) and glutamine side chains of proteins acting as acyl donors, and this reaction is catalysed by transglutaminase (Lilley *et al.*, 1998). Substrate proteins are usually covalently immobilised to plastic surfaces such as 96 well plates and repeated washing steps eliminates unbound amine substrates. In this study, N, N-dimethyl casein was used as the glutamine substrate and biotin-cadaverine (BTC) the amine substrate. An enzymatic method involving avidin conjugated to peroxidase was used to detect the incorporated BTC. The amine incorporation assays are very sensitive though time-consuming in comparison to the ammonia monitoring assays which are faster but less sensitive (Katona *et al.*, 2012). Also, microbial transglutaminase activity can be determined using the hydroxamate assay. In this assay transglutaminase catalyses the incorporation of hydroxylamine (acyl acceptor) into Z-glutamine-glycine (acyl donor) forming a Z-glutamyl(hydroxamate)-glycine complex that produces a colour with iron III chloride which can be measured at 525 nm (Ando *et al.*, 1989). In the present study, the BTC incorporation assay due to its higher sensitivity was chosen over the hydroxamate assay which is widely used for microbial transglutaminase activity detection. Also, the hydroxamate assay requires higher amounts of substrates and reagents compared to the BTC incorporation assay which requires small volumes. The last method involves detection of protein crosslinks between glutamine and lysine substrates which can be

visualised by SDS-PAGE analysis (Hsieh and Pan, 2012). We did not detect protein crosslinks after reactions of purified proteins with glutamine and lysine substrates.

It is plausible that off-target reactions from contaminating proteins such as aminotransferases could possibly mediate BTC incorporation into dimethyl casein plates. Similar observation has been reported from Ca^{2+} -independent transglutaminase activity reactions in plant extracts that suffered interference from diamine oxidase which also mediates incorporation reactions (Lilley *et al.*, 1998). Aminotransferases are enzymes that transfer amino groups originating from amino acids to α -keto acids to produce amino acid products. Thus, they are vital in metabolism of amino acids and require pyridoxal 5' phosphate cofactors to mediate catalytic reactions (Shin, Yun and Park, 2018). Aminotransferases could catalyse incorporation of BTC into dimethyl casein given its role in amino group transfer. More so putative pyridoxal 5' phosphate-dependent transferase was found in the 44 kDa purified bands. Though the cofactor was found with low abundance (5 peptides), it may have just been enough to mediate catalysis (Shin, Yun and Park, 2018). The determination of ϵ - γ -(glutamyl) lysine isopeptide bonds is a definite proof of transglutaminase activity in any system (Lilley *et al.*, 1998). Indeed, to validate the authenticity of transglutaminase activity we needed to detect isopeptide crosslinks. However, we already know that the purified transglutaminase-like activity was non-specific from the inhibition assays and crosslinking assays with casein. The non-specificity of the transglutaminase reaction could arise from the few non-specific bands in the eluted peak fractions, even though the pooled fraction had 2 major bands. Another explanation may be that the specific substrates for our novel enzymes have not been tested.

5.9 CONCLUSION

We set out to purify and characterise transglutaminase from sporulating *C. difficile* 630 cells. This required prior detection of high yield enzyme production during sporulation. Transglutaminase was produced during sporulation in *C. difficile* 630 and transglutaminase activity differed by culture method. Assessment of transglutaminase production in liquid and solid media demonstrated that activity was detected on agar earlier during sporulation and increased to attain high levels at a time coinciding with the stationary phase in liquid cultures. By contrast transglutaminase activity in broth was detected later during sporulation towards the end of the stationary phase and increased to maximum levels by the end of the sporulation course. The best temperature/pH combo for maximal transglutaminase solubilisation was obtained at pH7.5/0°C or pH8/25°C. High yield enzyme extraction at pH8/25°C was attained prior to purification at pH 8 on appropriate columns. We successfully purified transglutaminase from sporulating cells on Q-Sepharose and Butyl Sepharose columns respectively. A partially

purified enzyme was obtained which was separated into two major bands after SDS-PAGE analysis. It is important to note that up to date, only the transglutaminase from *S. mobaraense* has been commercialised. Biochemical characterization of purified spore protein showed that it did not require calcium for activity like most microbial transglutaminases and unlike mammalian transglutaminases. The purified enzyme had optimal temperature of activity at 37°C following the BTC incorporation assay. Notably extremely low temperatures (0°C) did not favour enzyme activity meanwhile high temperatures prompted inactivation. *Streptomyces* MTG had a wider range of enzyme activity with at least 50% recovery at 0°C. Furthermore, the purified enzyme was thermostable between 0-37°C with significant activity depletion at 45°C. Analogous to purified *C. difficile* 630 transglutaminase, some microbial transglutaminases retained over 80% of activity for a period of 30 min at temperatures below 40°C (Cui *et al.*, 2007; Macedo, Sette and Sato, 2011; Zhang *et al.*, 2012). The thermostability of purified enzyme was lower than most characterised microbial transglutaminases that were thermostable at higher temperatures and for longer incubation times. Also purified enzyme activity was significantly inhibited by PMSF, and addition of DTT led to more inhibition. The purified enzyme was unaffected by EDTA and uninhibited by NEM or cysteamine suggesting it does not contain thiol groups in the active site, and probably not a metalloprotease. This was complemented by the MS analysis which detected target *C. difficile* 630 transglutaminase-like genes with very low coverage and containing few modifications. Mainly enzymes involved in metabolism in *C. difficile* 630 were identified from purified protein bands. This implies that the purified enzyme was not a transglutaminase as also indicated by activity persistence following inhibition studies with cysteine inhibitors. More so casein was not crosslinked by the purified enzyme as observed for MTG confirming we may be dealing with an entirely different protein other than a transglutaminase. There is also a possibility that the purified enzyme could be a transglutaminase-like protein for which the specific substrates have not been identified. We conclude that transglutaminase-like activity was detected during sporulation in *C. difficile* 630, although purified proteins did not display key characteristics of true transglutaminases or proteases. Mass spectrometry analysis corroborated the characterisation assays confirming the absence of Tgase in purified enzymes.

CHAPTER 6. GENERAL DISCUSSIONS

6.1 Introduction

C. difficile is a Gram-positive, anaerobic enteropathogen that has evolved over the past years and remains a threat to the healthcare sector and an important economic burden (Underwood *et al.*, 2009). More virulent strains continue to spread widely and are usually associated with higher incidence and disease severity. Spores are the main vectors for disease transmission due to their high resistance and prevalence (Worthington and Hilton, 2016). Compared to the model Gram-positive *B. subtilis* in which sporulation/germination pathways have been established, studies in *C. difficile* have been retarded due to problems associated with bacterial culture plus few available resources for molecular characterisation. Recent developments have identified genetic tools to study spore biology and physiology (Dembek *et al.*, 2013). As a result, there has been some advancement in sporulation studies in *C. difficile* using molecular and microbial approaches. Some studies have combined several approaches including transcriptome, proteome, genome, and microbiology to foster understanding of *C. difficile* sporulation and germination (Pereira *et al.*, 2013; Saujet *et al.*, 2013). One such study has demonstrated the role of Spo0A in regulation of sporulation genes and non-sporulation functions such as virulence and metabolism (Pettit *et al.*, 2014). Further exploration of major players in *C. difficile* sporulation can underlie the discovery of alternative routes for establishing novel diagnostics, disinfection, and treatment regimes. Transglutaminases (Tgases) are enzymes found during sporulation in microorganisms and provide resistance to cellular structures via crosslinking (Katsunori Kobayashi *et al.*, 1998). Therefore, targeting sporulation via Tgase could decrease persistence of infectious spores and thus disease transmission. The selected strain for this study, *C. difficile* 630 was involved in an outbreak and its complete genome sequence is available (Dembek *et al.*, 2013). Additionally, predefined culture conditions for *C. difficile* 630 have been published (Pereira *et al.*, 2013; Putnam *et al.*, 2013).

This chapter would discuss the major findings of this study from a broader perspective, provide some explanations to questions that arose and point to future directions in view of providing answers for Tgase studies during sporulation in *C. difficile*. The study was divided in two major parts. The first step was to identify putative Tgases within *C. difficile* 630 genome, search for the conserved Tgase fold and determine homology with characterised microbial transglutaminases. Secondly, we detected Tgase gene expression and activity during sporulation prior to subsequent characterisation. Two main approaches were used to isolate and massively produce Tgases from *C. difficile* 630 for purification and characterisation studies. This included recombinant expression in *E. coli* and purification of crude extracts

harvested from sporulating cultures. Using transcriptomics and proteomics techniques, we provide preliminary information on target Tgase genes and potential involvement in sporulation in *C. difficile* 630. For better understanding of the role of Tgases in sporulation, gene knockout studies are required to identify the phenotypes associated.

6.2 Gene expression analysis during sporulation in *C. difficile* 630

During sporulation in *C. difficile* 630, Tgase activity was detected at different time points on solid and liquid media. Following, three putative Tgase-like genes were found within the complete genome of *C. difficile* 630 and all possess the Tgase core domain that is conserved in Tgase superfamily members. Sequence alignment with Tgase superfamily members confirmed conservation of catalytic cysteine in the Tgase fold of *C. difficile* 630. Sequence similarity searches did not identify homologs of *C. difficile* Tgase-like genes in other Tgase superfamily members, even though they contain Tgase activity plus the conserved cysteine at the active site. This observation has been reported elsewhere, as such homology searches may not be a good predictor of evolutionary links between genes. Transglutaminase transcripts were detected in sporulating cultures of *C. difficile* 630, and CD-1062 expression correlated with *spo0A* expression. This initial result provides an insight that there could be a possible involvement of Tgases during sporulation in *C. difficile* 630. The similarity between CD-1062 and CD-1083 genes illustrated by BLAST analysis plus the comparability during gene expression could suggest that they are regulated in the same manner. The absence of *spo0A* boxes upstream CD-1083 may evoke an indirect regulation by *spo0A* or that they are not regulated by Spo0A. Indirect regulation has been hypothesised for proteins under *spo0A* regulation but lacking *spo0A* boxes in the region upstream the target gene (Pettit *et al.*, 2014). At least for CD-1062 expression that correlates with *spo0A* expression, there may exist a direct regulation by Spo0A suggesting a likely association between Tgase genes and sporulation. We did not determine Tgase protein production during gene expression studies in sporulating cultures. Consequently, we do not have data to suggest possibility of a correlation between transcript expression and protein expression. However, from later studies of sporulation in *C. difficile* (appendix 5), Tgase protein was produced very early in vegetative cells from the time of sporulation induction, with increases to reach peak levels at 6 hours of sporulation corresponding to the late exponential phase. This somewhat matches the gene expression profile in that Tgase was expressed very early in sporulating cells with peaks towards the end of the exponential phase, and a drop in the stationary phase. Nonetheless a study of both gene and protein expression at specific times during sporulation is required to enable correlation assessments between mRNA and protein. In summary the timing of

transglutaminase expression in *C. difficile* 630 with respect to sporulation was investigated and expression seems to correlate with the time one would expect sporulation to occur.

6.3 Characterization of transglutaminase activity produced during sporulation in *C. difficile*

Transglutaminase activity differs by culture method as higher activity was obtained from Tgase production on agar cultures. Peak Tgase production occurred between 9-24 hours of sporulation and coincided with the stationary/death growth phase. This could predict a likely association between Tgase production and spores, considering that maximal spore yields in *C. difficile* systems occurs in this same interval. This is foreseeable given that spores are produced when nutrient conditions become limiting usually in the stationary phase and serves as a survival mechanism. Perhaps the Tgase gene is a survival gene associated with sporulation. This hypothesis requires further investigation. Furthermore, low temperatures (0-25°C) and pH favoured optimal enzyme solubilisation from sporulating cells in our system. Tgase was solubilised from spores and purified by IEX and HIC. The partially purified enzyme was separated by SDS-PAGE into two major bands of approximately 44 and 100 kDa. In this study, commercial MTG seemed to be more active at higher temperatures compared to purified *C. difficile* 630 Tgase. The optimal temperature of activity of purified *C. difficile* 630 Tgase is lower than most microbial transglutaminases that have maximal enzyme activity around 55-60°C. A possible reason for this divergence could be associated with the thermostability of the *C. difficile* 630 Tgase spore enzyme. Indeed, the purified enzyme had lower thermostability compared to several microbial transglutaminases that are thermostable at higher temperatures and for longer periods.

Following, purified enzymes were significantly inhibited by PMSF, and DTT addition to inhibited samples did not restore activity. Rather there was more inhibition with DTT suggesting it is not a cysteine protease. To test the dependence on calcium for activity essential for eukaryotes, EDTA was included in purified enzymes and MTG control. Calcium chloride was also added to samples to evaluate its effect on activity. Enzyme activity was unaffected by calcium or EDTA which suggests that *C. difficile* 630 Tgase did not require calcium for activity analogous to most microbial transglutaminases. Additionally, no protease activity was found in purified proteins and EDTA was not inhibitory eliminating the possibility of a metalloprotease. Surprisingly no inhibition was observed in purified samples following treatments with thiol group inhibitors (NEM, cysteamine dihydrochloride) which could imply that cysteine is not present in catalytic sites. *Streptomyces* MTG in this study like most other microbial transglutaminases was significantly inhibited by thiol group modifiers. Also, MTG

was inhibited by PMSF, and DTT did not restore enzyme activity which is a 'proof of principle' that MTG does not contain protease activity as already known. The *C. difficile* 630 putative Tgases possibly contain other domains in addition to the predicted Tgase core domain. The persistence of Tgase activity in response to cysteine inhibitors could suggest the possibility of other domains with potential implications in protein functionality or that its specific substrates were not tested during activity assays.

6.4 Transglutaminase gene expression and mass spectrometry analysis of purified proteins

From gene expression analysis of Tgase transcripts, CD-720 was mostly produced during the post-exponential phase and possibly long after the stationary phase. Perhaps CD-720 could have been expressed in transcripts after 24 hours of sporulation corresponding to the time of harvest for protein purification. A more elaborate analysis would be to detect both gene expression and protein activity at the time of harvest. Therefore, our gene expression data cannot be complemented by proteomic analysis of purified proteins harvested from *C. difficile* sporulation cultures. Proteomic analysis data show very low coverage for target *C. difficile* 630 Tgase genes and high abundance for other non-target proteins. Presumption that transcript expression correctly predicts protein expression was not obvious in our system, considering that no Tgases were identified by MS analysis of purified proteins as predicted by gene expression. However, we did not measure Tgase gene expression at the time of harvest prior to purification. Other studies in *C. difficile* have reported disagreement between mRNA and protein expression (Pettit *et al.*, 2014). Altogether these results contribute to our understanding of Tgase expression during sporulation providing a foundation for further characterisation of their role in sporulation and germination.

Predicted Tgase fold conservation does not match MS analysis of purified endogenous spore proteins. Consistent with the MS results, no inhibition of enzyme activity was obtained with known inhibitors of thiol groups. Also sequence similarity searches of target Tgases did not reveal the newly identified proteins as matches. The fact that purified enzymes tested positive for Tgase activity suggests that other purified proteins could have catalysed BTC incorporation, thereby giving false positive Tgase activity. More so, protein crosslinking with casein, a known Tgase substrate was negative. The main proteins purified are involved in energy metabolism in *C. difficile* 630. During sporulation, these proteins maybe accumulated to serve the spore during conversion to vegetative growth when conditions become more favourable. Genes involved in metabolism are under the control of Spo0A. Therefore, it is likely that Spo0A was still active at the time of culture sampling for purification. Our study

reported expression of *spo0A* during the exponential and stationary growth phase, though no measurements were pursued beyond the early stationary phase. Spo0A is involved in regulation of butyrate synthesis elucidating a role with *C. difficile* metabolism. Indeed the most abundant proteins from mass spectrometry analysis are those involved in butyrate metabolism confirming that spore purified proteins are involved in metabolism and possibly regulated by Spo0A. Spo0A directs transport of nutrients and energy metabolism dependent on nutrient availability on mucosal surfaces in the intestinal epithelium (Pettit *et al.*, 2014).

6.5 Recombinant Expression of transglutaminases in *C. difficile* 630

The nature of the target genes can influence expression given that difficulties in full-length expression was seen only in CD-1062 and CD-1083 genes. Meanwhile CD-720 was successfully expressed as a soluble full-length protein with GST and His tags. This matches bioinformatics analysis that highlighted high similarity between CD-1062 and CD-1083 and thus their comparable expression was not surprising. Both CD-1062 and CD-1083 were predicted to be secreted, supported by the presence of signal peptides though target proteins were not released in culture supernatants during expression. Signal peptides (SP) decrease contamination of target proteins through protection from proteolytic breakdown by host proteases; improve biological activity of targets via proper disulphide bond formation that enables protein expression in the proper conformation; facilitates downstream processes and purification (Freudl, 2018). Another benefit is that SP prevent cytosolic accumulation of target proteins into IB and protect from toxicity on the host (Freudl, 2018). Signal peptides did not preserve enzyme activity as would be expected since no enzyme activity was found in recombinant proteins. Also eluted fractions contained several extra bands suggestive of proteolytic degradation despite the presence of SP that are protective. Meanwhile the mature sequence of Tgase genes within *C. difficile* 630 were successfully cloned and expressed as soluble proteins. The biggest Tgase genes CD-1062, CD-1083 were expressed without signal peptides in both *E. coli* Rosetta 2 and BL21 strains. Yet we detected no Tgase activity following intracellular soluble protein expression.

Conversely, microbial transglutaminase from a Gram-positive, spore-forming bacteria has been successively cloned and expressed intracellularly in *E. coli* as an active protein (Washizu *et al.*, 1994; Kobayashi *et al.*, 1996; Katsunori. Kobayashi *et al.*, 1998). It is intriguing that we found no Tgase or protease activity in recombinant proteins. The putative CD-1062/CD-1083 active sites contain a serine residue next to the catalytic cysteine. It is possible that we may have purified serine proteases instead or an entirely different protein. We conclude that

recombinant proteins were expressed as inactive proteins or require specific substrates to detect activity.

6.6 CONCLUSION AND FUTURE PERSPECTIVES

The results reported in this thesis illustrate an effort to identify novel genes associated with sporulation, characterise them and understand their function in sporulation. Taken together the results obtained from our experiments, *C. difficile* 630 contains three putative Tgases, each with a conserved Tgase core. The three Tgase genes were successfully expressed as soluble proteins in *E. coli* even though enzyme activity was not detected. Furthermore, Tgase transcripts were found in sporulating cultures, and CD-1062 gene expression correlated significantly with *spo0A* expression suggesting a potential association of Tgases with sporulation. Moreover, the *spo0A* consensus was found upstream the CD-1062 gene. Tgase activity was detected during sporulation and Tgase was partially purified by IEX and HIC chromatography into two major bands. However, the purified enzyme activity did not exhibit pure Tgase activity after biochemical characterization assays, and mass spectrometry analysis identified other proteins with high abundance and very low coverage for target Tgases. We did not reach the final objective which was to determine the role of Tgases during sporulation and germination, and therefore requires further investigation. The role of Tgases in *C. difficile* sporulation can be verified via the creation of mutations or knock outs within the Tgase gene using the Clostron system. Subsequently phenotype studies that involve measuring changes in spore structure and stability, sporulation efficiency, spore resistance, germination, growth, viability, and morphology can be pursued to determine if a phenotype is associated. All mutants and knockouts need to be complemented and reassessed to evaluate if the observed effects are Tgase specific. Furthermore RT-qPCR analysis to determine Tgase transcript and measuring activity levels during sporulation to investigate the association between Tgase protein and Tgase-like gene. The localisation of Tgase within *C. difficile* spores could be determined using immunofluorescence staining with specific antibodies. Prior to localisation studies, antibodies need to be raised against *C. difficile* 630 Tgase to generate a stock of Tgase specific antibodies. Given the latest advancements in *Clostridial* research with the availability of varied molecular and biochemical methods coupled to the findings from this study, further research would clarify the role of Tgases in sporulation and germination in *C. difficile*. This might facilitate development of novel disinfection and treatment regimes.

REFERENCES

- Aaron, L. and Torsten, M. (2018) 'Microbial transglutaminase: A new potential player in celiac disease', *Clinical Immunology*, (xxxx), pp. 1–7. doi: 10.1016/j.clim.2018.12.008.
- Ando, H. *et al.* (1989) 'Purification and characteristics of a novel transglutaminase derived from microorganisms', *Agricultural and Biological Chemistry*, 53(10), pp. 2613–2617. doi: 10.1080/00021369.1989.10869735.
- Di Bella, S. diarrhea uk. pdfan. *et al.* (2016) 'Clostridium difficile toxins A and B: Insights into pathogenic properties and extraintestinal effects', *Toxins*, 8(5), pp. 1–25. doi: 10.3390/toxins8050134.
- Bergamini, C. M. *et al.* (2011) 'Structure and regulation of type 2 transglutaminase in relation to its physiological functions and pathological roles.', *Advances in enzymology and related areas of molecular biology*, 78, pp. 1–46. doi: 10.1002/9781118105771.ch1.
- Burns, D. A., Heap, J. T. and Minton, N. P. (2010) 'The diverse sporulation characteristics of Clostridium difficile clinical isolates are not associated with type', *Anaerobe*, 16(6), pp. 618–622. doi: 10.1016/j.anaerobe.2010.10.001.
- Crobach, M. J. T. *et al.* (2009) 'European Society of Clinical Microbiology and Infectious Diseases (ESCMID): Data review and recommendations for diagnosing Clostridium difficile -infection (CDI)', *Clinical Microbiology and Infection*, 15(12), pp. 1053–1066. doi: 10.1111/j.1469-0691.2009.03098.x.
- Cui, L. *et al.* (2007) 'Purification and characterization of transglutaminase from a newly isolated Streptomyces hygrosopicus', *Food Chemistry*, 105(2), pp. 612–618. doi: 10.1016/j.foodchem.2007.04.020.
- Curry, S. R. (2017) 'Clostridium difficile', 30(1), pp. 329–342. doi: 10.1016/j.cll.2010.04.001.Clostridium.
- Dembek, M. *et al.* (2013) 'Transcriptional Analysis of Temporal Gene Expression in Germinating Clostridium difficile 630 Endospores', *PLoS ONE*, 8(5). doi: 10.1371/journal.pone.0064011.
- Edwards, A. N. *et al.* (2016) 'Chemical and stress resistances of clostridium difficile spores and vegetative cells', *Frontiers in Microbiology*, 7(OCT), pp. 1–13. doi: 10.3389/fmicb.2016.01698.
- Edwards, A. N. and McBride, S. M. (2014) 'Initiation of sporulation in Clostridium difficile: A twist on the classic model', *FEMS Microbiology Letters*, 358(2), pp. 110–118. doi: 10.1111/1574-6968.12499.
- Fatima, S. W. and Khare, S. K. (2018) 'Current insight and futuristic vistas of microbial transglutaminase in nutraceutical industry', *Microbiological Research*, 215(June), pp. 7–14. doi: 10.1016/j.micres.2018.06.001.

- Fazaeli, A. *et al.* (2019) 'Expression optimization, purification, and functional characterization of cholesterol oxidase from *Chromobacterium* sp. DS1', *PLoS ONE*, 14(2), pp. 1–15. doi: 10.1371/journal.pone.0212217.
- Fernandes, C. G. *et al.* (2015) 'Structural and Functional Characterization of an Ancient Bacterial Transglutaminase Sheds Light on the Minimal Requirements for Protein Cross-Linking', *Biochemistry*, 54(37), pp. 5723–5734. doi: 10.1021/acs.biochem.5b00661.
- Fimlaid, K. A. *et al.* (2013) 'Global Analysis of the Sporulation Pathway of *Clostridium difficile*', *PLOS Genetics*, 9(8). doi: 10.1371/journal.pgen.1003660.
- Freudl, R. (2018) 'Signal peptides for recombinant protein secretion in bacterial expression systems', *Microbial Cell Factories*, 17(1), pp. 1–10. doi: 10.1186/s12934-018-0901-3.
- Gomes, L., Monteiro, G. and Mergulhão, F. (2020) 'The impact of IPTG induction on plasmid stability and heterologous protein expression by *Escherichia coli* biofilms', *International Journal of Molecular Sciences*, 21(2). doi: 10.3390/ijms21020576.
- Griffin, E. B. R. J. C. M. (2013) 'Recent advances in the development of tissue transglutaminase', pp. 119–127. doi: 10.1007/s00726-011-1188-4.
- Griffin, M., Casadio, R. and Bergamini, C. M. (2002) 'Transglutaminases: Nature's biological glues', *Biochem. J.*, 368, pp. 377–396.
- Heap, J. T. *et al.* (2007) 'The Clostron: A universal gene knock-out system for the genus *Clostridium*', *Journal of Microbiological Methods*, 70(3), pp. 452–464. doi: 10.1016/j.mimet.2007.05.021.
- Ho, M. L. *et al.* (2000) 'Technical approach to simplify the purification method and characterization of microbial transglutaminase produced from *Streptovorticillium ladakanum*', *Journal of Food Science*, 65(1), pp. 76–80. doi: 10.1111/j.1365-2621.2000.tb15959.x.
- Hsieh, J. F. and Pan, P. H. (2012) 'Proteomic profiling of microbial transglutaminase-induced polymerization of milk proteins', *Journal of Dairy Science*, 95(2), pp. 580–589. doi: 10.3168/jds.2011-4773.
- lismaa, S. E. *et al.* (2009) 'Transglutaminases and disease: Lessons from genetically engineered mouse models and inherited disorders', *Physiological Reviews*, 89(3), pp. 991–1023. doi: 10.1152/physrev.00044.2008.
- Javitt, G. *et al.* (2017) 'Constitutive expression of active microbial transglutaminase in *Escherichia coli* and comparative characterization to a known variant', *BMC Biotechnology*, 17(1), pp. 1–10. doi: 10.1186/s12896-017-0339-4.
- Jin, M. *et al.* (2016) 'Purification and characterization of a high-salt-resistant microbial transglutaminase from *Streptomyces mobaraensis*', *Journal of Molecular Catalysis B: Enzymatic*, 133, pp. 6–11. doi: 10.1016/j.molcatb.2016.07.005.
- Jump, R. L. P., Pultz, M. J. and Donskey, C. J. (2007) 'Vegetative *Clostridium difficile* survives in room air on moist surfaces and in gastric contents with reduced acidity: A potential

mechanism to explain the association between proton pump inhibitors and *C. difficile*-associated diarrhea?', *Antimicrobial Agents and Chemotherapy*, 51(8), pp. 2883–2887. doi: 10.1128/AAC.01443-06.

Kalakuntla, A. S. *et al.* (2019) 'Probiotics and *Clostridium Difficile*: A Review of Dysbiosis and the Rehabilitation of Gut Microbiota', *Cureus*, 11(7). doi: 10.7759/cureus.5063.

Kashiwagi, T. *et al.* (2002) 'Crystal structure of microbial transglutaminase from *Streptovercillium mobaraense*', *Journal of Biological Chemistry*, 277(46), pp. 44252–44260. doi: 10.1074/jbc.M203933200.

Katona, É. *et al.* (2012) 'Measurement of factor XIII activity in plasma', *Clinical Chemistry and Laboratory Medicine*, 50(7), pp. 1191–1202. doi: 10.1515/cclm-2011-0730.

Kevorkian, Y., Shirley, D. J. and Shen, A. (2016) 'Biochimie Regulation of *Clostridium difficile* spore germination by the CspA pseudoprotease domain', *Biochimie*, 122, pp. 243–254. doi: 10.1016/j.biochi.2015.07.023.

Kieliszek, M. and Misiewicz, A. (2014) 'Microbial transglutaminase and its application in the food industry . A review', *Folia Microbiol*, 59, pp. 241–250. doi: 10.1007/s12223-013-0287-x.

Kim, S. Y., Jeitner, T. M. and Steinert, P. M. (2002) 'Transglutaminases in disease', *Neurochemistry International*, 40(1), pp. 85–103. doi: 10.1016/S0197-0186(01)00064-X.

King, J. D. *et al.* (2009) 'Post-assembly Modification of *Bordetella bronchiseptica* O Polysaccharide by a Novel Periplasmic Enzyme Encoded', 284(3), pp. 1474–1483. doi: 10.1074/jbc.M807729200.

Klobutcher, L. A., Ragkousi, K. and Setlow, P. (2006) 'The *Bacillus subtilis* spore coat provides "eat resistance" during phagocytic predation by the protozoan *Tetrahymena thermophila*', *PNAS*, 103(1), pp. 165–170.

Kobayashi, K. *et al.* (1996) 'ε-(γ-Glutamyl)lysine cross-links of spore coat proteins and transglutaminase activity in *Bacillus subtilis*', *FEMS Microbiology Letters*, 144(2–3), pp. 157–160. doi: 10.1016/0378-1097(96)00353-9.

Kobayashi, Katsunori. *et al.* (1998) 'Molecular cloning of the transglutaminase gene from *Bacillus subtilis* and its expression in *Escherichia coli*', *Biosci Biotechnol Biochem*, 62(6), pp. 1109–1114. doi: 10.1271/bbb.62.1109.

Kobayashi, Katsunori *et al.* (1998) 'Transglutaminase in sporulating cells of *Bacillus subtilis*', *Journal of General and Applied Microbiology*, 44(1), pp. 85–91. doi: 10.2323/jgam.44.85.

Kramer, R. M. *et al.* (2012) 'Toward a molecular understanding of protein solubility: Increased negative surface charge correlates with increased solubility', *Biophysical Journal*, 102(8), pp. 1907–1915. doi: 10.1016/j.bpj.2012.01.060.

Kuehne, S. A., Cartman, S. T. and Minton, N. P. (2011) 'Both, toxin A and toxin B, are important in *Clostridium difficile* infection', *Gut Microbes*, 2(4). doi: 10.4161/gmic.2.4.16109.

Kumazawa, Y. *et al.* (1996) 'Purification and Characterization of Transglutaminase from

- Walleye Pollack Liver', *Fisheries Science*, 62(6), pp. 959–964. doi: 10.2331/fishsci.62.959.
- Kumazawa, Y. *et al.* (1997) 'Purification and Characterization of Transglutaminase from Japanese Oyster (*Crassostrea gigas*)', *Journal of Agricultural and Food Chemistry*, 45(3), pp. 604–610. doi: 10.1021/jf9604596.
- Kuwana, R. *et al.* (2006) 'Modification of GerQ reveals a functional relationship between Tgl and YabG in the coat of *Bacillus subtilis* spores', *Journal of Biochemistry*, 139(5), pp. 887–901. doi: 10.1093/jb/mvj096.
- Lawley, T. D. *et al.* (2009) 'Proteomic and genomic characterization of highly infectious *Clostridium difficile* 630 spores', *Journal of Bacteriology*, 191(17), pp. 5377–5386. doi: 10.1128/JB.00597-09.
- Lawley, T. D. *et al.* (2010) 'Use of purified *clostridium difficile* spores to facilitate evaluation of health care disinfection regimens', *Applied and Environmental Microbiology*, 76(20), pp. 6895–6900. doi: 10.1128/AEM.00718-10.
- Leggett, M. J. *et al.* (2012) 'Bacterial spore structures and their protective role in biocide resistance', *Journal of Applied Microbiology*, 113, pp. 485–498. doi: 10.1111/j.1365-2672.2012.05336.x.
- Lilley, G. R. *et al.* (1998) 'Detection of Ca²⁺-dependent transglutaminase activity in root and leaf tissue of monocotyledonous and dicotyledonous plants', *Plant Physiology*, 117(3), pp. 1115–1123. doi: 10.1104/pp.117.3.1115.
- Lin, Y. S. *et al.* (2004) 'Cloning and expression of the transglutaminase gene from *Streptovercillium ladakanum* in *Streptomyces lividans*', *Process Biochemistry*, 39(5), pp. 591–598. doi: 10.1016/S0032-9592(03)00134-1.
- Liu, S. *et al.* (2011) 'The order of expression is a key factor in the production of active transglutaminase in *Escherichia coli* by co-expression with its pro-peptide', *Microbial Cell Factories*, 10, pp. 1–7. doi: 10.1186/1475-2859-10-112.
- Liu, S. *et al.* (2016) 'Improving the active expression of transglutaminase in *Streptomyces lividans* by promoter engineering and codon optimization', *BMC Biotechnology*, 16(1), pp. 1–9. doi: 10.1186/s12896-016-0304-7.
- Liu, Y. *et al.* (2018) 'Characterization of transglutaminase from: *Bacillus subtilis* and its cross-linking function with a bovine serum albumin model', *Food and Function*, 9(11), pp. 5560–5568. doi: 10.1039/c8fo01503a.
- Ma, J., Wang, Y.-G. and Wang, Y.-J. (2013) 'A simple, fast and efficient method for cloning blunt DNA fragments', *African Journal of Biotechnology*, 12(26), pp. 4094–4097. doi: 10.5897/AJB12.691.
- Macedo, J. A., Sette, L. D. and Sato, H. H. (2011) 'Purification and characterization of a new transglutaminase from *streptomyces* sp. isolated in brazilian soil', *Journal of Food Biochemistry*, 35(4), pp. 1361–1372. doi: 10.1111/j.1745-4514.2010.00456.x.

- Makarova, K. S., Aravind, L. and Koonin, E. V (1999) 'A superfamily of archaeal , bacterial , and eukaryotic proteins homologous to animal transglutaminases', *Protein Science*, 8, pp. 1714–1719.
- Martucciello, S. *et al.* (2020) 'Interplay between type 2 transglutaminase (Tg2), gliadin peptide 31-43 and anti-tg2 antibodies in celiac disease', *International Journal of Molecular Sciences*, 21(10). doi: 10.3390/ijms21103673.
- Marx, C. K., Hertel, T. C. and Pietzsch, M. (2007) 'Soluble expression of a pro-transglutaminase from *Streptomyces mobaraensis* in *Escherichia coli*', *Enzyme and Microbial Technology*, 40(6), pp. 1543–1550. doi: 10.1016/j.enzmictec.2006.10.036.
- Matsuka, Y. V. *et al.* (2003) 'Staphylococcus aureus Fibronectin-Binding Protein Serves as a Substrate for Coagulation Factor XIIIa: Evidence for Factor XIIIa-Catalyzed Covalent Cross-Linking to Fibronectin and Fibrin', *Biochemistry*, 42(49), pp. 14643–14652. doi: 10.1021/bi035239h.
- Metcalf, D., Sharif, S. and Weese, J. S. (2010) 'Evaluation of candidate reference genes in *Clostridium difficile* for gene expression normalization', *Anaerobe*, 16(4), pp. 439–443. doi: 10.1016/j.anaerobe.2010.06.007.
- Milani, A. *et al.* (2012) 'TgpA , a Protein with a Eukaryotic-Like Transglutaminase Domain , Plays a Critical Role in the Viability of *Pseudomonas aeruginosa*', *PLoS ONE*, 7(11). doi: 10.1371/journal.pone.0050323.
- Monroe, A. and Setlow, P. (2006) 'Localization of the Transglutaminase Cross-Linking Sites in the *Bacillus subtilis* Spore Coat Protein GerQ Localization of the Transglutaminase Cross-Linking Sites in the *Bacillus subtilis* Spore Coat Protein GerQ □', *Journal of Bacteriology*, 188(21), pp. 7609–7616. doi: 10.1128/JB.01116-06.
- Mora-uribe, P. *et al.* (2016) 'Characterization of the Adherence of *Clostridium difficile* Spores : The Integrity of the Outermost Layer Affects Adherence Properties of Spores of the Epidemic Strain R20291 to Components of the Intestinal Mucosa', *Frontiers in Cellular and Infection Microbiology*, 6(99), pp. 1–16. doi: 10.3389/fcimb.2016.00099.
- Nagy, V. and Szakacs, G. (2008) 'Production of transglutaminase by *Streptomyces* isolates in solid-state fermentation', 47, pp. 122–127. doi: 10.1111/j.1472-765X.2008.02395.x.
- Nur'amaliyah, L. S., Zilda, D. S. and Mubarik, N. R. (2016) 'Purification and Characterization of Transglutaminase from *Streptomyces* sp. TTA 02 SDS 14', *Squalen Bulletin of Marine and Fisheries Postharvest and Biotechnology*, 11(3), pp. 75–83. doi: 10.15578/squalen.v11i3.234.
- Nurminskaya, M. V. and Belkin, A. M. (2012) *Cellular Functions of Tissue Transglutaminase, International Review of Cell and Molecular Biology*. doi: 10.1016/B978-0-12-394305-7.00001-X.
- Odi, B. O. and Coussons, P. (2014) 'Biological functionalities of transglutaminase 2 and the possibility of its compensation by other members of the transglutaminase family', *The*

- Scientific World Journal*, 2014(Table 1), pp. 7–9. doi: 10.1155/2014/714561.
- Paredes-Sabja, D., Setlow, P. and Sarker, M. R. (2011) 'Germination of spores of Bacillales and Clostridiales species: Mechanisms and proteins involved', *Trends in Microbiology*, 19(2), pp. 85–94. doi: 10.1016/j.tim.2010.10.004.
- Paredes-Sabja, D., Shen, A. and Sorg, J. (2014) 'Clostridium difficile spore biology: sporulation, germination and spore structural proteins', *Trends Microbiol*, 22(7), pp. 406–416. doi: 10.1016/j.tim.2014.04.003.Clostridium.
- Pasternack, R. *et al.* (1998) 'Bacterial pro-transglutaminase from Streptovercillium mobaraense - Purification, characterisation and sequence of the zymogen', *European Journal of Biochemistry*, 257(3), pp. 570–576. doi: 10.1046/j.1432-1327.1998.2570570.x.
- Peng, Z. *et al.* (2017) 'Antibiotic resistance and toxin production of Clostridium difficile isolates from the hospitalized patients in a large hospital in Florida', *Frontiers in Microbiology*, 8(DEC), pp. 1–8. doi: 10.3389/fmicb.2017.02584.
- Pereira, F. C. *et al.* (2013) 'The Spore Differentiation Pathway in the Enteric Pathogen Clostridium difficile', *PLoS Genetics*, 9(10). doi: 10.1371/journal.pgen.1003782.
- Pettit, L. J. *et al.* (2014) 'Functional genomics reveals that Clostridium difficile Spo0A coordinates sporulation, virulence and metabolism', *BMC Genomics*, 15(1), pp. 1–15. doi: 10.1186/1471-2164-15-160.
- Plomp, M. *et al.* (2014) 'Architecture and Assembly of the Bacillus subtilis Spore Coat', *PLoS ONE*, 9(9). doi: 10.1371/journal.pone.0108560.
- Polage, C. R., Solnick, J. V. and Cohen, S. H. (2012) 'Nosocomial diarrhea: Evaluation and treatment of causes other than clostridium difficile', *Clinical Infectious Diseases*, 55(7), pp. 982–989. doi: 10.1093/cid/cis551.
- Putnam, E. E. *et al.* (2013) 'SpoIVA and SipL Are Clostridium difficile Spore Morphogenetic Proteins', *Journal of Bacteriology*, 195(6), pp. 1214–1225. doi: 10.1128/JB.02181-12.
- Rachel, N. M. and Pelletier, J. N. (2013) 'Biotechnological applications of transglutaminases', *Biomolecules*, 3(4), pp. 870–888. doi: 10.3390/biom3040870.
- Ragkousi, K. and Setlow, P. (2004) 'Transglutaminase-mediated cross-linking of GerQ in the coats of Bacillus subtilis spores', *Journal of Bacteriology*, 186(17), pp. 5567–5575. doi: 10.1128/JB.186.17.5567-5575.2004.
- Rickert, M. *et al.* (2016) 'Production of soluble and active microbial transglutaminase in Escherichia coli for site-specific antibody drug conjugation', *Protein Science*, 25, pp. 442–455. doi: 10.1002/pro.2833.
- Ridlon, J. M., Kang, D. and Hylemon, P. B. (2006) 'review Bile salt biotransformations by human intestinal bacteria', *Journal of Lipid Research*, 47, pp. 241–259. doi: 10.1194/jlr.R500013-JLR200.
- Rosano, G. L. and Ceccarelli, E. A. (2014) 'Recombinant protein expression in Escherichia

coli: Advances and challenges', *Frontiers in Microbiology*, 5(APR), pp. 1–17. doi: 10.3389/fmicb.2014.00172.

Rosenbusch, K. E. *et al.* (2012) 'C. difficile 630 Δ erm Spo0A Regulates Sporulation, but Does Not Contribute to Toxin Production, by Direct High-Affinity Binding to Target DNA', *PLoS ONE*, 7(10). doi: 10.1371/journal.pone.0048608.

Sanchez-Salas, J. L. *et al.* (2011) 'Maturation of released spores is necessary for acquisition of full spore heat resistance during *Bacillus subtilis* sporulation', *Applied and Environmental Microbiology*, 77(19), pp. 6746–6754. doi: 10.1128/AEM.05031-11.

Saujet, L. *et al.* (2013) 'Genome-Wide Analysis of Cell Type-Specific Gene Transcription during Spore Formation in *Clostridium difficile*', *PLOS Genetics*, 9(10). doi: 10.1371/journal.pgen.1003756.

Schmidt, G. *et al.* (1998) 'The Rho-deamidating cytotoxic necrotizing factor 1 from *Escherichia coli* possesses transglutaminase activity: Cysteine 866 and histidine 881 are essential for enzyme activity', *Journal of Biological Chemistry*, 273(22), pp. 13669–13674. doi: 10.1074/jbc.273.22.13669.

Setlow, P. (2006) 'Spores of *Bacillus subtilis*: Their resistance to and killing by radiation, heat and chemicals', *Journal of Applied Microbiology*, 101(3), pp. 514–525. doi: 10.1111/j.1365-2672.2005.02736.x.

Setlow, P. (2014) 'Germination of Spores of *Bacillus* Species: What We Know and Do Not Know', *Journal of Bacteriology*, 196(7), pp. 1297–1305. doi: 10.1128/JB.01455-13.

Shin, Y. C., Yun, H. and Park, H. H. (2018) 'Structural dynamics of the transaminase active site revealed by the crystal structure of a co-factor free omega-transaminase from *Vibrio fluvialis* JS17', *Scientific Reports*, 8(1), pp. 1–9. doi: 10.1038/s41598-018-29846-0.

Sinha, R. and Khare, S. K. (2014) 'Protective role of salt in catalysis and maintaining structure of halophilic proteins against denaturation', *Frontiers in Microbiology*, 5(APR), pp. 1–6. doi: 10.3389/fmicb.2014.00165.

Sorg, J. A. and Sonenshein, A. L. (2008) 'Bile Salts and Glycine as Cogermnants for *Clostridium difficile* Spores □', *Journal of Bacteriology*, 190(7), pp. 2505–2512. doi: 10.1128/JB.01765-07.

Sorg, J. A. and Sonenshein, A. L. (2009) 'Chenodeoxycholate Is an Inhibitor of *Clostridium difficile* Spore Germination □', *Journal of Bacteriology*, 191(3), pp. 1115–1117. doi: 10.1128/JB.01260-08.

Staab, J. F. *et al.* (1999) 'Adhesive and mammalian transglutaminase substrate properties of *Candida albicans* Hwp1', *Science*, 283(5407), pp. 1535–1538. doi: 10.1126/science.283.5407.1538.

Steiner, E. *et al.* (2012) 'Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*', *Mol Microbiol*, 80(3), pp.

- 641–654. doi: 10.1111/j.1365-2958.2011.07608.x.Multiple.
- Strop, P. (2014) 'Versatility of Microbial Transglutaminase', *Bioconjugate Chemistry*, (25), pp. 855–862. doi: 10.1021/bc500099v.
- Suzuki, S. *et al.* (2000) 'Purification and Characterization of Novel Transglutaminase from *Bacillus subtilis* Spores', *Biosci Biotechnol Biochem*, 64(11), pp. 2344–2351. doi: 10.1271/bbb.64.2344.
- Tegel, H., Ottosson, J. and Hober, S. (2011) 'Enhancing the protein production levels in *Escherichia coli* with a strong promoter', *FEBS Journal*, 278(5), pp. 729–739. doi: 10.1111/j.1742-4658.2010.07991.x.
- Touchette, M. H. *et al.* (2019) 'SpoIVA-SipL complex formation is essential for *Clostridioides difficile* spore assembly', *Journal of Bacteriology*, 201(8), pp. 1–16. doi: 10.1128/JB.00042-19.
- Ulusu, N. N. (2015) 'Evolution of Enzyme Kinetic Mechanisms', *Journal of Molecular Evolution*, 80(5–6), pp. 251–257. doi: 10.1007/s00239-015-9681-0.
- Underwood, S. *et al.* (2009) 'Characterization of the sporulation initiation pathway of *Clostridium difficile* and its role in toxin production', *Journal of Bacteriology*, 191(23), pp. 7296–7305. doi: 10.1128/JB.00882-09.
- Veide Vilg, J. and Undeland, I. (2017) 'pH-driven solubilization and isoelectric precipitation of proteins from the brown seaweed *Saccharina latissima*—effects of osmotic shock, water volume and temperature', *Journal of Applied Phycology*, 29(1), pp. 585–593. doi: 10.1007/s10811-016-0957-6.
- Vohra, P. and Poxton, I. R. (2011) 'Comparison of toxin and spore production in clinically relevant strains of *Clostridium difficile*', *Microbiology*, 157(5), pp. 1343–1353. doi: 10.1099/mic.0.046243-0.
- Wang, Z. and Griffin, M. (2012) 'TG2, a novel extracellular protein with multiple functions', *Amino Acids*, 42(2–3), pp. 939–949. doi: 10.1007/s00726-011-1008-x.
- Washington, E. J. *et al.* (2013) 'What a Difference a Dalton Makes : Bacterial Virulence Factors Modulate Eukaryotic Host Cell Signaling Systems via Deamidation'. doi: 10.1128/MMBR.00013-13.
- Washizu, K. *et al.* (1994) 'Molecular cloning of the gene for microbial transglutaminase from *Streptomyces lividans* and its expression in *Streptomyces lividans*', *Bioscience, Biotechnology and Biochemistry*, 58(1), pp. 82–87. doi: 10.1080/bbb.58.82.
- Wheeldon, L. J. *et al.* (2008) 'Physical and chemical factors influencing the germination of *Clostridium difficile* spores', *Journal of Applied Microbiology*, 105(6), pp. 2223–2230. doi: 10.1111/j.1365-2672.2008.03965.x.
- Wilson, B. A. and Ho, M. (2011) 'Cellular and molecular action of the mitogenic protein-deamidating toxin from *Pasteurella multocida*', 278, pp. 4616–4632. doi: 10.1111/j.1742-4658.2011.08158.x.

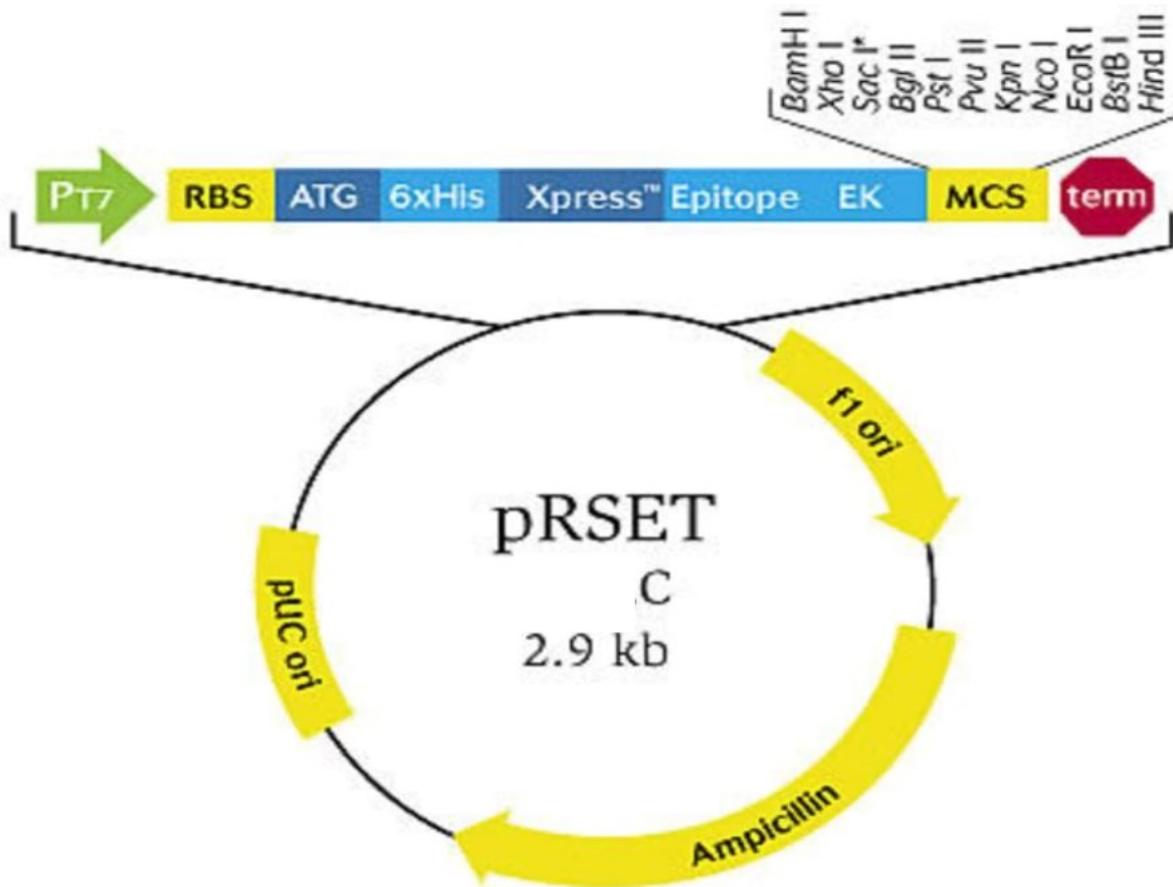
- Worthington, T. and Hilton, A. C. (2016) 'Sporulation of *Clostridium difficile* in Aerobic conditions is Significantly Protracted when Exposed to Sodium Taurocholate', *Journal of Medical Microbiology & Diagnosis*, 05(02), pp. 2–5. doi: 10.4172/2161-0703.1000230.
- Yokoyama, K., Nio, N. and Kikuchi, Y. (2004) 'Properties and applications of microbial transglutaminase', *Applied Microbiology and Biotechnology*, 64(4), pp. 447–454. doi: 10.1007/s00253-003-1539-5.
- Zhang, D., Zhu, Y. and Chen, J. (2009) 'Microbial transglutaminase production: Understanding the mechanism', *Biotechnology and Genetic Engineering Reviews*, 26(1), pp. 205–222. doi: 10.5661/bger-26-205.
- Zhang, L. *et al.* (2012) 'Enzymatic characterization of transglutaminase from *Streptomyces mobaraensis* DSM 40587 in high salt and effect of enzymatic cross-linking of yak milk proteins on functional properties of stirred yogurt', *Journal of Dairy Science*, 95(7), pp. 3559–3568. doi: 10.3168/jds.2011-5125.
- Zhu, Y. and Tramper, J. (2008) 'Novel applications for microbial transglutaminase beyond food processing', (August). doi: 10.1016/j.tibtech.2008.06.006.
- Zilhão, R. *et al.* (2005) 'Assembly and Function of a Spore Coat-Associated Transglutaminase of *Bacillus subtilis* Assembly and Function of a Spore Coat-Associated Transglutaminase of *Bacillus subtilis*', *Journal of Bacteriology*, 187(22), pp. 7753–7764. doi: 10.1128/JB.187.22.7753.

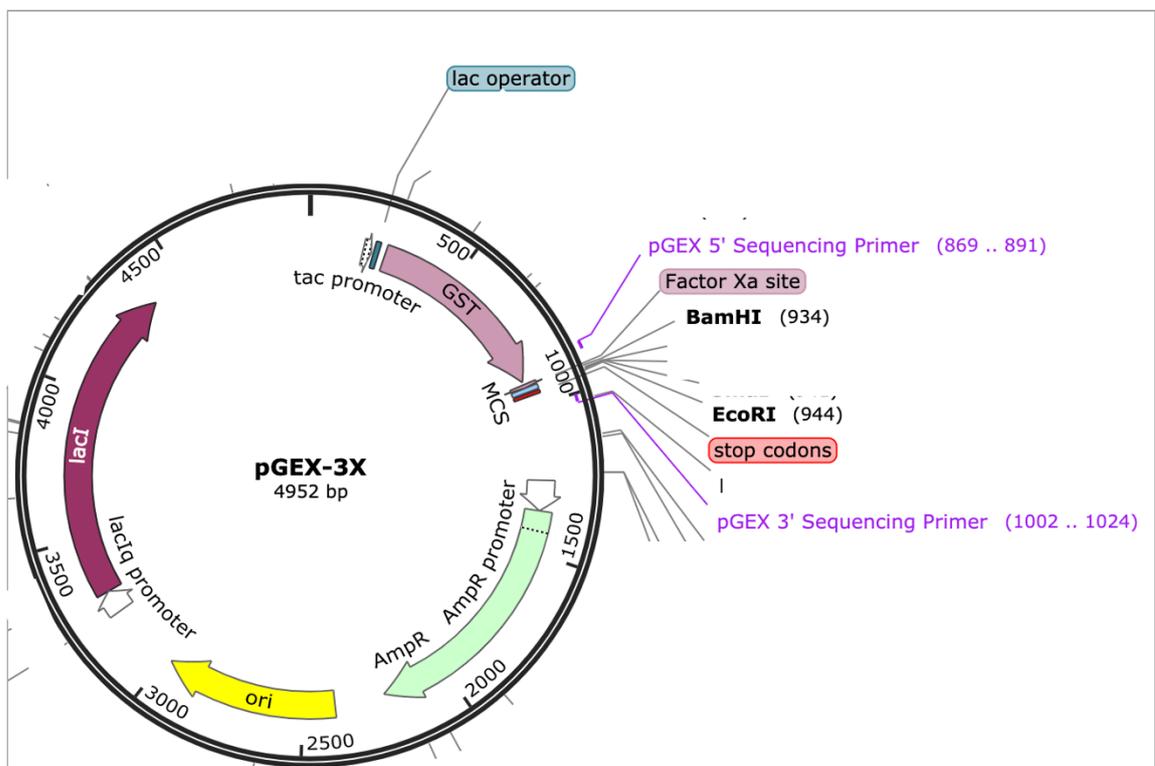
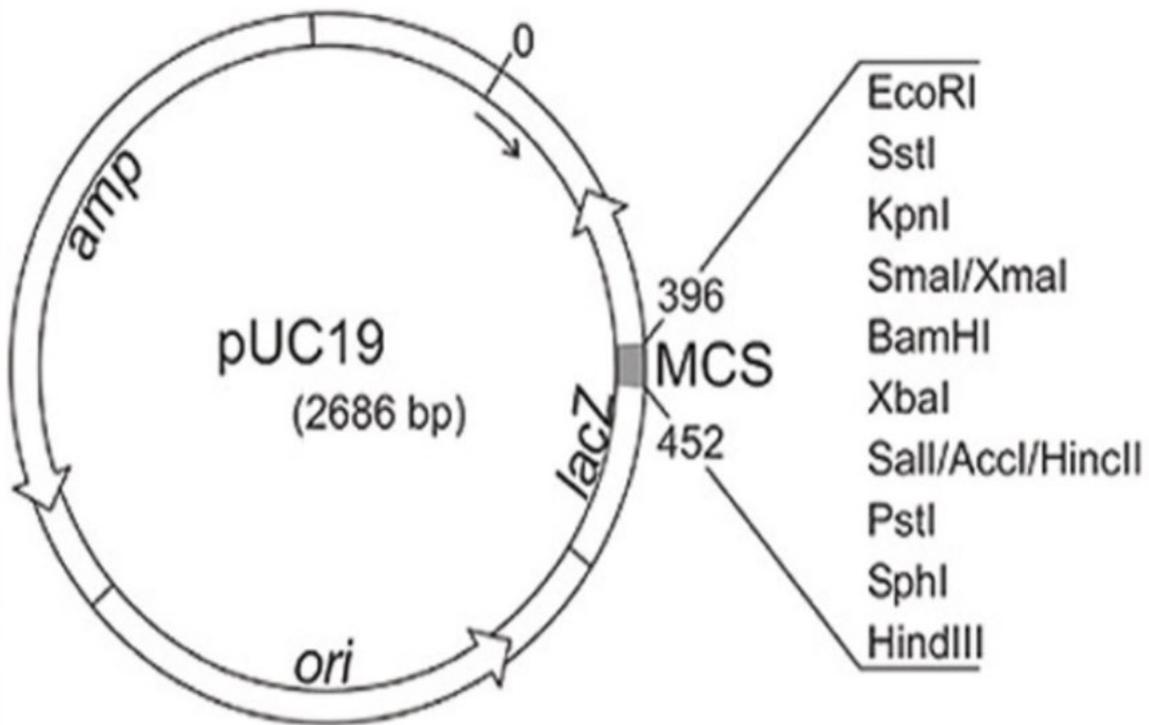
APPENDICES

Appendix 1: Bacterial strains and plasmids used in this study

Bacterial strain	characteristics	source
<i>C. difficile</i> 630	Pathogenic strain	Dr Sarah Kuehne
<i>E. coli</i> Rosetta 2	Competent cell for recombinant expression	Novagen
<i>E. coli</i> BL21 (DE3)	Competent cell for recombinant expression	Novagen
<i>E. coli</i> DH5 α	Competent cell for DNA cloning	Thermofisher
Plasmids		
pUC19	Cloning plasmid	Thermofisher
pGEX3X	Plasmid for construction of recombinant GST-tagged proteins	GE Healthcare
pJET1.2/blunt	Cloning plasmid	Thermofisher
pRSETc	Plasmid for construction of His-tagged recombinant proteins	Thermofisher

Appendix 2: Plasmids used in this study

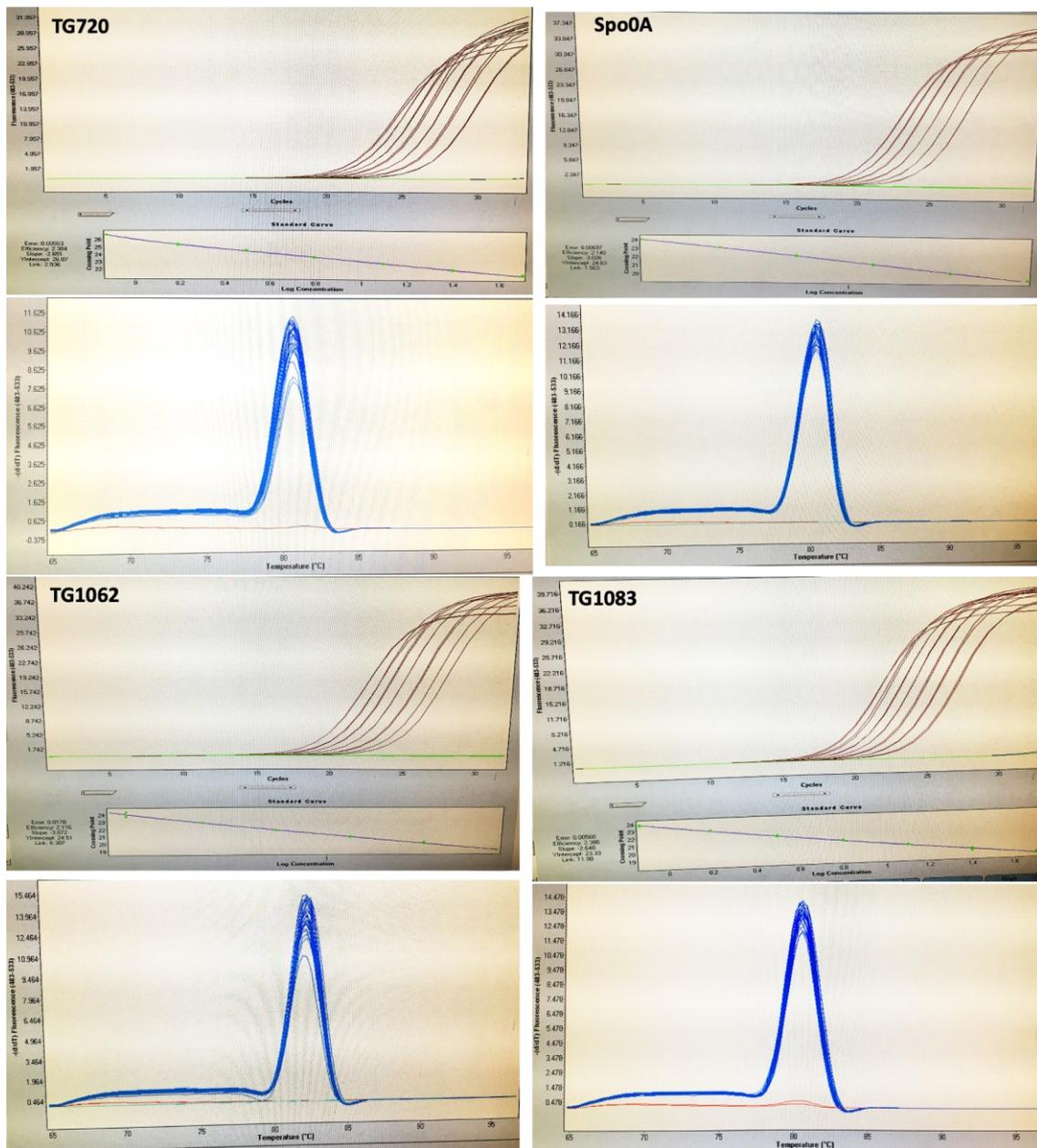


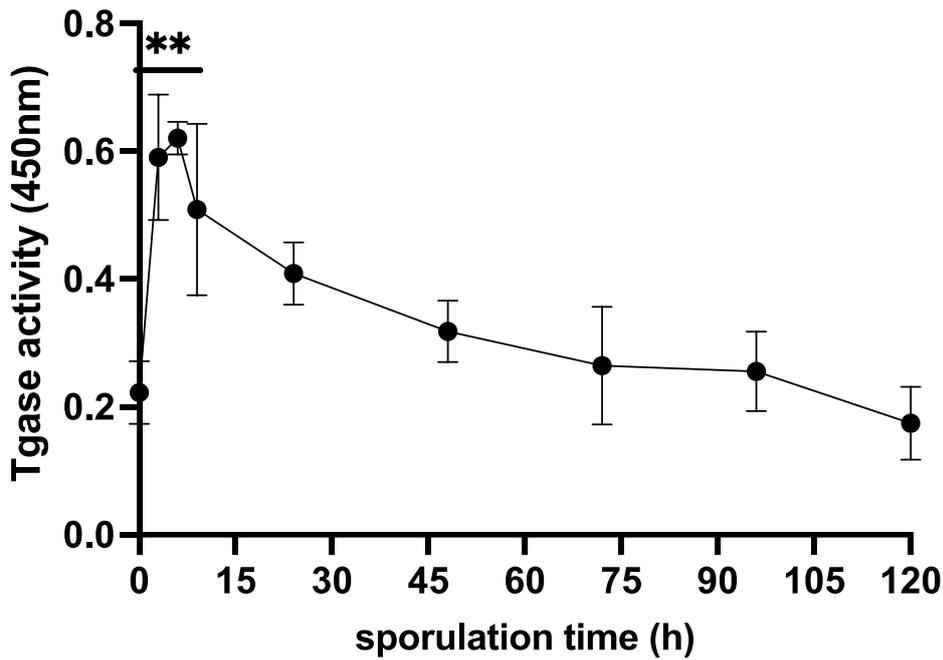


Appendix 3: List of antibodies used in this study

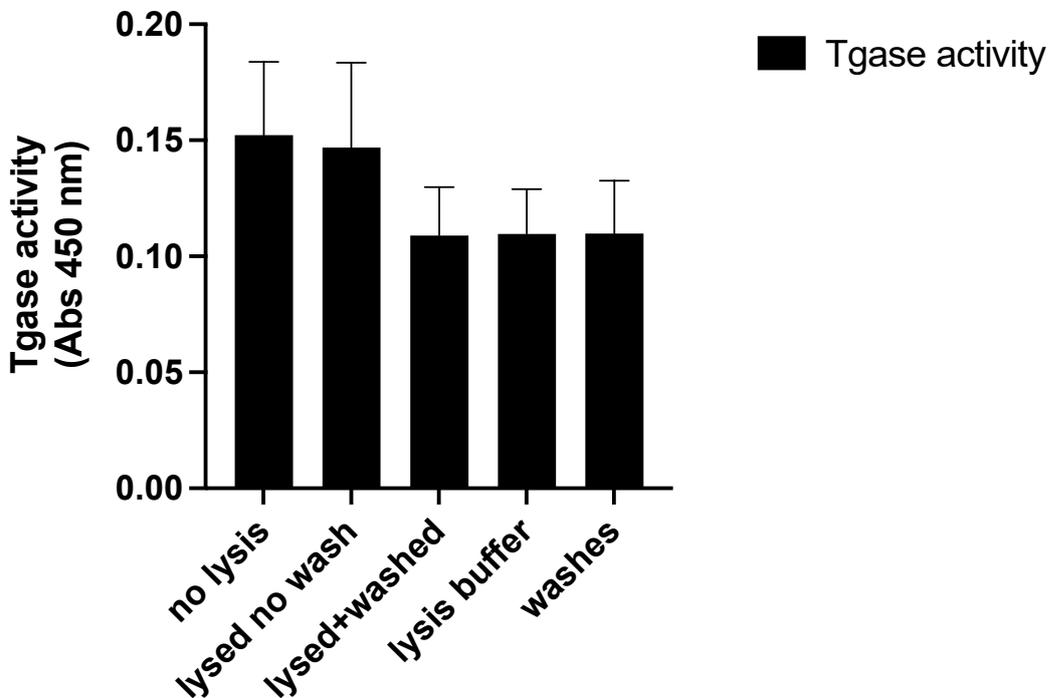
antibody	species	Dilution	manufacturer
ExtrAvidin peroxidase	avian	1:2000-1:5000	Sigma
Anti-6x His tag	mouse	1:1000	Abcam
Anti-mouse IgG	goat	1:1000-1:3000	R&D

Appendix 4: Amplification and melting curves obtained from qRT-PCR reactions of target genes.





Appendix 5: Detection of transglutaminase activity in vegetative cells during sporulation in *C. difficile* 630

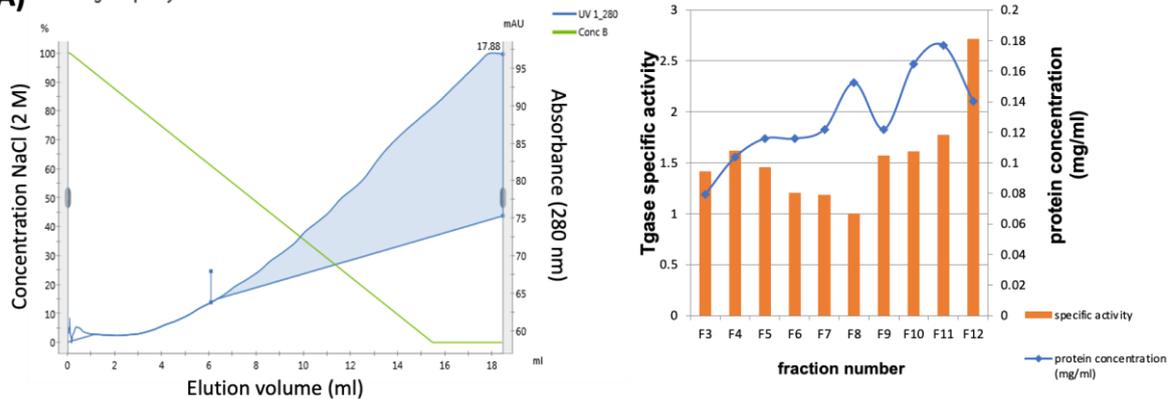


Appendix 6: Effect of washes on transglutaminase activity of lysed cells

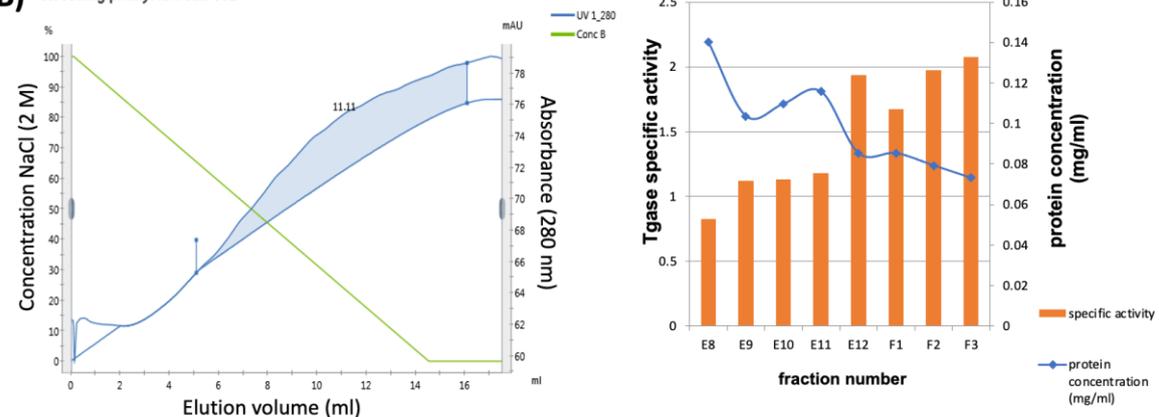
Appendix 7: Screening of HIC media by purification of active Tgase fractions on selected columns.

Active Tgase fractions obtained from Q-Sepharose columns were pooled, adjusted to 2 M NaCl and purified on several HIC columns. Hydrophobic interaction purification was performed on an AKTA system and protein absorbance was monitored at 280 nm. The bound Tgase protein was eluted with a decreasing linear (100-0%) NaCl gradient from 2 M in Tris buffer. The displayed images represent purification on **(A)** Phenyl Sepharose HP **(B)** Phenyl Sepharose FF low sub **(C)** Phenyl Sepharose FF high-sub **(D)** Butyl Sepharose FF **(E)** Octyl Sepharose FF (green line: decreasing NaCl gradient, blue line: protein absorbance detected at 280 nm, red line: conductivity). Purification graphs are on the left side and Tgase activity profile on the right side.

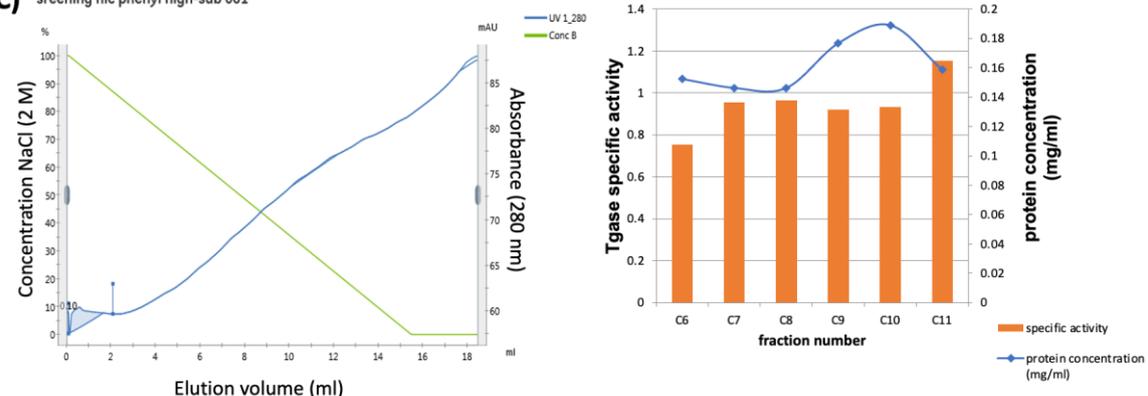
(A) screening HIC phenyl HP 002



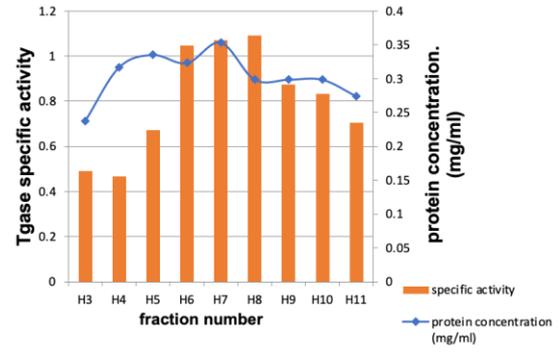
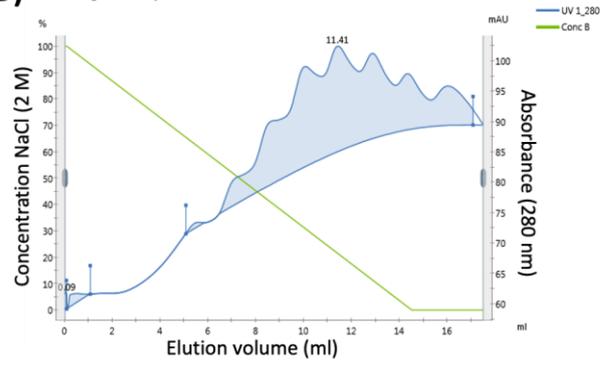
(B) screening phenyl low sub 002



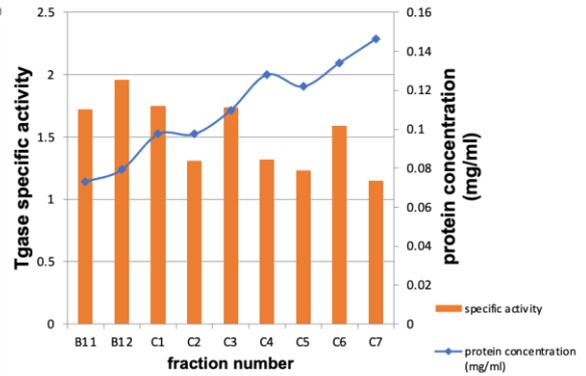
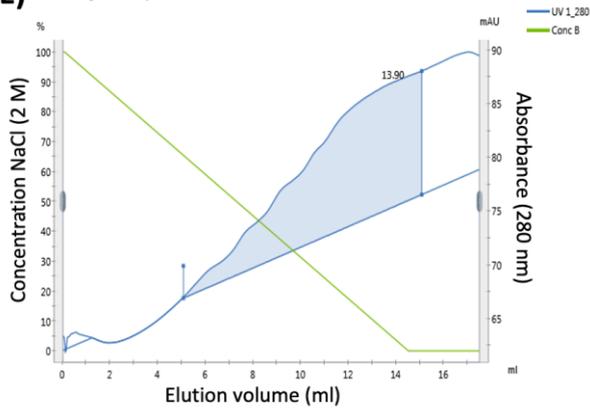
(C) screening hic phenyl high-sub 001

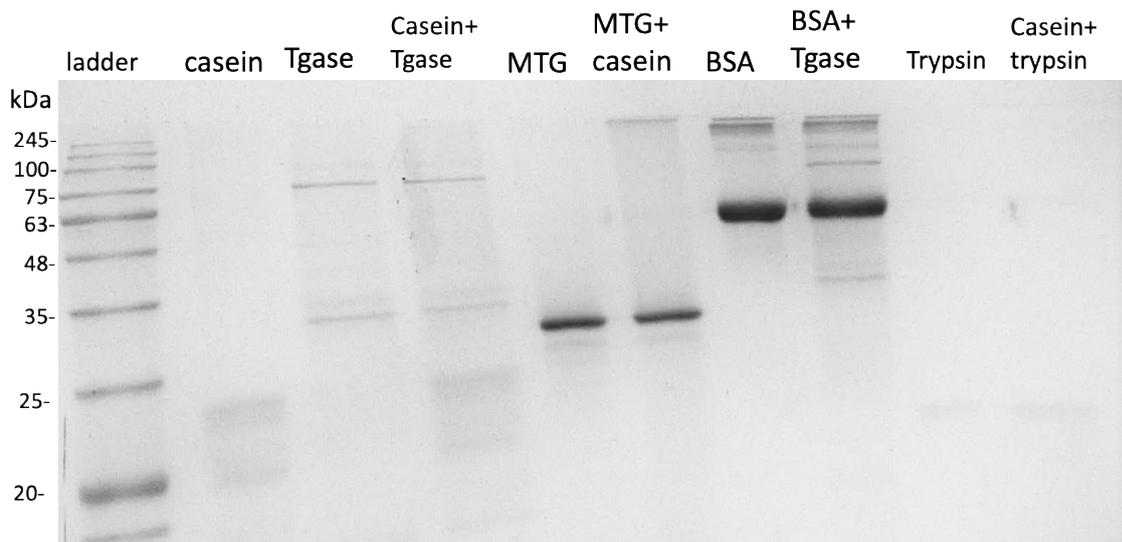


(D) screening HIC butyl ff 1ml 002

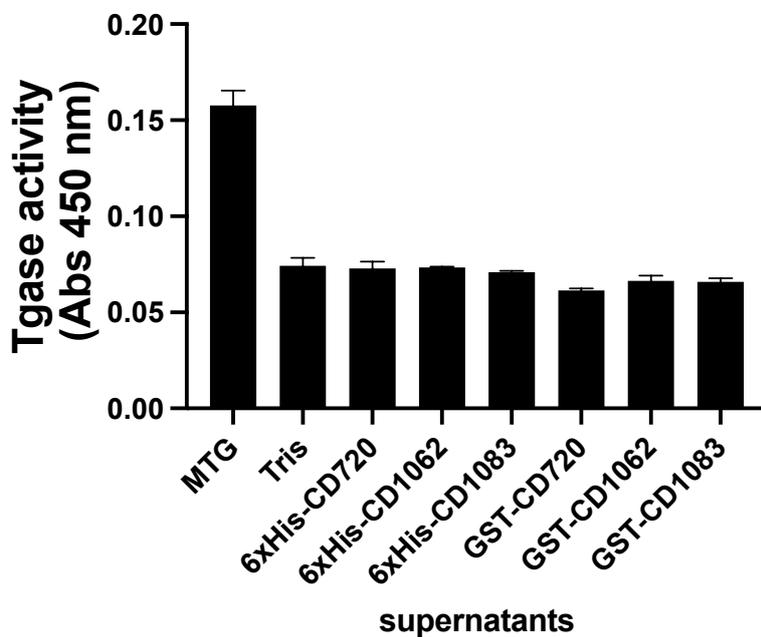


(E) screening HIC octyl ff 002





Appendix 8: SDS-PAGE showing reaction of purified protein with BSA substrate



Appendix 9: Transglutaminase activity in culture supernatants of recombinant proteins