



UNIVERSITY OF
LEICESTER

**Identification of the receptor binding
proteins for *Clostridium difficile*
Bacteriophages**

**Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester by**

Ahmed Salim Ali Dowah BSc, MSc.

**Department of Infection, Immunity, and Inflammation
University of Leicester, Leicester, UK**

April 2018

Identification of the receptor binding proteins for *Clostridium difficile* Bacteriophages.

By Ahmed Dowah

Abstract

Clostridium difficile causes severe infectious diarrhoea referred to as *C. difficile* infection (CDI) and due to the organism being naturally resistant to antibiotics, alternative treatments for CDI are urgently required. Phages could provide an alternative source of antimicrobials for this pathogen due to their specificity, minimal disruption of microbiota and ability to self-amplify at the site of infection. However, the therapeutic development of phages will significantly benefit from a full understanding of the *C. difficile* phage infection process. To date no studies have identified the phage receptors binding proteins (RBPs) or the corresponding receptors on the bacterial surface that phages bind or adsorb, to establish infection. In other words, how does the first physical contact between phage receptor binding proteins located in the distal part of the phage tail and the surface of the bacterium occur?

This project aims to identify the receptor binding proteins for two phages of *C. difficile*; phiCDHS1 (siphovirus), which infects CD105LC1 and CDR20291 that belong to the Ribotype 027 hypervirulent strains. In addition, phiCDMH1 (myovirus) that infects CD105HE1 ribotype 076. The approaches employed to identify the RBPs for these phages, were to over-express the four predicted phage tail fiber proteins Gp18, Gp19, Gp21 and Gp22 from CDHS1 phage and Gp29 and Gp30 from CDMH1. Which presumably, one or two of them is involved in the phage host binding. After significant optimisation, the expressed proteins were purified and polyclonal antibodies were generated against them. The antibodies were then used to neutralize phage infection, and

were immunogold labelled to visualise the location of the proteins using TEM. The proteins were also crystallised in order to identify their structure.

It was found that the anti-Gp22 protein was able to block phiCDHS1 infection, indicating that Gp22 is the protein responsible for *C. difficile* recognition. In addition, the anti-Gp29 protein was able to inhibit phiCDMH1 infection, which indicates that the Gp29 is the RBP for this phage. This is the first observation for *C. difficile* bacteriophages.

This finding provides a novel insight into *C. difficile* bacteriophage biology and mechanisms of interaction with their hosts.

Acknowledgments

First, All praise and glory is due to Allah, and Allah's Peace and Blessings be upon His messenger.

I genuinely would like to express my deep gratitude to both of my supervisors Prof Martha Clokie and Prof Russell Wallis for their guidance, patience, attention and support throughout the journey of this research. The door to Prof Clokie and Prof Wallis office was always open whenever I ran into trouble or had a question about my research. They consistently allowed this PhD project to be my own work but, steered me in the right the direction whenever they thought I needed it. I am very grateful, and I am very lucky and honoured to have had you both as my supervisors.

Also, I must thank all people who helped me a lot during this journey, starting with Dr Christopher Furze and Dr Bayan Faraj from lab 218. From phage group thanks to Dr Janet Nale, Dr Katherine Hargreaves, Dr Jinyu Shan, Dr Christopher Turkington, Dr Srwa Rashid, Peter Hickenbotham, Dr Ed Galyov, my PRP committee (Dr Helen O'Hare and Dr Galina Mukamolova) and last but not least special thanks to Dr Anisha Thanki for your help and your kindness in reading my thesis chapters. I almost forget Dr Jaspreet Sahota, Thank you.

I would like to extremely thank all members of phage group for their support and creating a friendly environment to work, thanks to, Dr Ali Abdulkareem, Mohammed, Aisha, Guillermo, Ananthi, Neda, Faizal, Bander, Dr Tolis Panayi, Thekra, Wafaa and Lamiaa. Thanks people for being amazing.

Most importantly, and for sure words cannot express my deep gratitude to my mother and father for their, love, support, prayers, and continues engorgements during this journey and whole my life. Also I would like to thank my grandmother for her encouragement

and continued prayers. My brothers and sisters, thank you so much for your unconditional love and support, I am very fortunate to have you.

Finally, thanks to the Libyan government for funding me and giving me the opportunity to complete my PhD, without this funding, this would not have been able to happen.

Publications

Dowah, A. S. A. & Clokie, M. R. J., 2018. Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. *Biophys Rev.*

Table of contents

Chapter 1	General Introduction	1
1.1	Introduction	2
1.1.1	<i>Clostridium difficile</i>	2
1.1.2	<i>C. difficile</i> pathogenicity	4
1.1.3	<i>C. difficile</i> treatment	5
1.2	Bacteriophage background	8
1.2.1	Application of bacteriophages	10
1.2.2	Phage adsorption	11
1.2.3	Bacteriophage receptors on the surface of the Bacteria	12
1.2.4	Phage receptor binding proteins (RBPs)	16
1.2.5	RBPs for phages of Gram-Negative bacteria	17
1.2.6	RBPs for Siphoviruses of Gram-positive bacteria	18
1.2.7	Genomic architecture and the position of the tail fiber/baseplate sequences	19
1.3	Structural aspect of Siphophages baseplate	22
1.3.1	Distal tail proteins (Dit)	22
1.3.2	Tail associated lysine proteins (Tal)	23
1.3.3	Like -Upper baseplate protein (BppU)	24
1.3.4	Receptor binding protein (RBPs)	24
1.3.5	RBPs shape depend on the bacterial receptors nature	26
1.3.6	Two strategies found within siphoviruses binding	27
1.3.7	RBPs for myoviruses of Gram- positive bacteria	27
1.4	Bacteriophage that infects <i>C. difficile</i>	28

1.5	Project Aims -----	31
Chapter 2 <i>in silico</i> analysis to identify putative functions of the baseplate proteins and RBP candidates of phages CDHS1 and CDMH1 that infect <i>C. difficile</i> -----		
2.1	Introduction -----	35
2.2	Aim of this study -----	36
2.3	Results -----	38
2.3.1	<i>in silico</i> analysis of the putative tail proteins that involved in baseplate structure of CDHS1 phage -----	38
2.3.1.1	Gp18 protein -----	38
2.3.1.2	Gp19 protein -----	40
2.3.1.3	Gp21 protein -----	41
2.3.1.4	Gp22 protein -----	42
2.3.2	<i>in silico</i> analysis of the putative tail proteins of CDMH1 -----	42
2.3.2.1	Gp23 protein -----	43
2.3.2.2	Gp24 protein of CDMH1 -----	44
2.3.2.3	Gp25 and Gp26 proteins of CDMH1 -----	45
2.3.2.4	GP27 and gp28 proteins of CDMH1 -----	46
2.3.2.5	Gp29 and Gp30 proteins of CDMH1 -----	46
2.4	Discussion -----	47
2.4.1	<i>in silico</i> analysis of CDHS1 baseplate proteins -----	47
2.4.1.1	Gp18 protein -----	47
2.4.1.2	The Gp19 protein of the CDHS1 phage -----	48
2.4.1.3	Gp21 protein of CDHS1 -----	49
2.4.1.4	Gp22 protein of CDHS1 -----	49

2.4.2	<i>in silico</i> analysis of the CDMH1 baseplate proteins-----	50
2.4.2.1	Gp23 and Gp24 proteins -----	50
2.4.2.2	Gp25 and Gp26 proteins -----	51
2.4.2.3	Gp27 and Gp28 proteins -----	51
2.4.2.4	Gp29 and Gp30 proteins -----	52
2.5	Summary of Chapter 2 -----	53
Chapter 3	Cloning, expression and purification of the tail proteins for CDHS1 & CDMH1 phages-----	55
3.1	Introduction -----	56
3.2	Aim of this study -----	57
3.3	Method -----	58
3.3.1	Amplification of the phage proteins that were bioinformatically predicted to be tail proteins for CDHS1 and CDMH1 phages-----	58
3.3.1.1	Phage DNA extraction -----	58
3.3.1.2	PCR primers designed-----	59
3.3.1.3	PCR assay -----	59
3.3.2	Preparation of BL21 (DE3) and DH5 α <i>E. coli</i> competent cells-----	61
3.3.3	Heat shock transformation -----	61
3.3.4	Protein expression and purification for the tail proteins Gp18, Gp19, Gp21 and Gp22 of CDHS1 Phage -----	62
3.3.4.1	Expression and purification of Histidine (His) tagged Gp18 -----	62
3.3.4.2	Poly Acrylamide Gel Electrophoresis SDS- PAGE preparation -----	64
3.3.4.3	Gel Filtration chromatography for Gp18 -----	65

3.3.5 Proteins expression of Glutathione S-transferases (GST) tagged Gp19 and Gp22 proteins-----65

3.3.6 Purification of Gp19 and Gp22 proteins -----67

 3.3.6.1 Affinity chromatography purification for Gp19 and Gp22 proteins----67

 3.3.6.2 Gel Filtration Chromatography for Gp19 and Gp22-----67

3.3.7 Cloning and preparation of recombinant Gp 21 proteins -----68

 3.3.7.1 Plasmid extraction and purification -----68

 3.3.7.2 The *gp21* gene amplification and purification -----69

 3.3.7.3 Digestion with restricted Endonucleases -----69

 3.3.7.4 Ligation-----69

 3.3.7.5 Transformation -----70

 3.3.7.6 Screen transformed colony -----70

 3.3.7.7 Restricted Endonucleases -----70

 3.3.7.8 Sequencing-----70

 3.3.7.9 Protein expression and purification for Gp21 protein -----71

3.3.8 Protein expression and purification of the tail protein for CDMH1-----71

 3.3.8.1 Protein expression of Gp29 and Gp30 -----71

 3.3.8.2 GST tagged Gp29 purification -----71

 3.3.8.3 GST tagged Gp30 purification -----72

3.4 Results -----73

3.4.1 Amplification of the putative genes that encode tail proteins for CDHS1 and CDMH1 phages -----73

 3.4.1.1 Amplification of the putative genes *gp18*, *gp19*, and *gp22* of CDHS1 73

 3.4.1.2 Cloning -----74

 3.4.1.3 Cloning *gp21* genes-----74

3.4.1.4	Amplification of the two putative genes that predicted to be encoding tail proteins for CDMH1 -----	75
3.4.2	Protein expression and purification of tail proteins of CDHS1 phage -----	76
3.4.2.1	Expression and purification of His tagged Gp18 -----	76
3.4.2.2	Expression and purification of GST tagged proteins from CDHS1 ----	78
3.4.2.2.1	GST tagged Gp19-----	78
3.4.2.2.2	GST tagged Gp22-----	79
3.4.2.2.3	GST tagged GP21 -----	81
3.4.3	Protein expression and purification of the tail proteins of CDMH1 phage--	83
3.4.3.1	GST tagged Gp29 from CDMH1 -----	83
3.4.3.2	GST tagged Gp30 CDMH1-----	84
3.5	Discussion-----	86
3.5.1	Protein expression and purification for the tail proteins of CDHS1-----	86
3.5.2	Protein expression and purification for tail proteins Gp29 and Gp30 of CDMH1 -----	89
3.6	Conclusion-----	91
Chapter 4	Identifying the Receptor binding proteins for <i>C. difficile</i> phages CDHS1 and CDMH1 -----	92
4.1	Introduction-----	93
4.2	The aims of this Chapter -----	96
4.3	Methods -----	97
4.3.1	Strains and Culturing-----	97
4.3.2	Phage propagations-----	97
4.3.2.1	Plaque Assay-----	97

4.3.2.2	Phage propagation using liquid culture -----	98
4.3.3	Spot test -----	98
4.3.4	Phage purification -----	98
4.3.5	Phage neutralization using antibodies -----	99
4.3.6	Immune gold labelling -----	100
4.4	Results -----	101
4.4.1	Investigation of the role of Gp18, Gp21 and Gp22 in CDHS1 phage adsorption -----	101
4.4.1.1	Gp18 protein -----	101
4.4.1.2	Gp21 protein -----	103
4.4.1.3	Gp22 protein -----	105
4.4.2	Immunogold labelling of the baseplate proteins or tail proteins of CDHS1 using TEM -----	107
4.4.2.1	The location of Gp18 on CDHS1 phage -----	107
4.4.2.2	The location of Gp 21 on CDHS1 phage -----	109
4.4.2.3	The localization of Gp22 on CDHS1 Phage -----	111
4.4.3	Investigation the role of Gp29 and Gp30 in CDMH1 phage adsorption --	113
4.4.3.1	Gp29 protein -----	113
4.4.3.2	Gp30 protein -----	115
4.4.4	Neutralization assay of phage CDHM3 & CDHM6 infection -----	116
4.4.4.1	Neutralization assay of phage CDHM3 infection using anti-Gp29 and anti-Gp30 serum -----	116
4.4.4.2	Neutralization Assay of phage CDHM6 infection using Anti-Gp29 and Anti- Gp30 serum -----	118
4.5	Discussion -----	120

4.5.1	Identification of RBPs of CDHS1 that infect <i>C. difficile</i> -----	121
4.5.2	Immunogold labelling of the baseplate proteins or tail proteins of CDHS1 using transmission electron microscope -----	122
4.5.3	Identification of RBPs of CDMH1 that infect <i>C. difficile</i> -----	124
4.5.4	Determining the ability of anti-Gp29 to neutralize CDHM3 and CDHM6 phages that infect CD105HE1 -----	126
4.6	Conclusion -----	127
Chapter 5	Structural characterization of the protein Gp22 -----	128
5.1	Introduction -----	129
5.2	The Aim -----	134
5.3	Methods -----	135
5.3.1	Expression and purification of Gp22 protein -----	135
5.3.2	Expression of Selenomethionine- Gp22 protein -----	135
5.3.3	Purification of Selenomethionine- Gp22 protein -----	136
5.3.4	Gp22 and Selenomethionine- Gp22 protein crystallization and optimization - 136	
5.3.5	The optimization of the conditions that produced crystals for Gp22 protein -- 136	
5.3.6	Processes to improve the quality of the crystals after crystallization -----	138
5.3.6.1	Dehydration -----	138
5.3.6.2	Micro seeding -----	138
5.4	Results -----	139
5.4.1	The use of 0.8 M Succinic acid at a pH of 7.0 as a crystallization condition 139	

5.4.2	The use of 0.2 M Sodium citrate tribasic dehydrate, and 20 % w/v PEG 3350 as a crystallization condition -----	141
5.4.3	Dehydration -----	143
5.4.4	Micro-seeding -----	145
5.4.5	X- ray diffraction -----	146
5.5	Discussion-----	148
5.6	Conclusion-----	152
Chapter 6	Conclusion and future work -----	153
6.1	Key finding of this study-----	154
6.1.1	Identification of the putative function of the tail proteins -----	154
6.1.2	Production and purification of the tail proteins of phages CDHS1 and CDMH1 -----	155
6.1.3	Determining the RBPs of phages CDHS1 and CDMH1 -----	156
6.1.4	Determining the three-dimensional structure of protein Gp22-----	156
6.2	Future work -----	157
6.2.1	Expression and purification of the tail associated lysin for the two phages used in this project -----	157
6.2.2	Resolving the structure of the RBPs -----	158
6.2.3	Exploiting the RBPs of the two phages as diagnostic tools -----	158
6.2.4	Exploiting RBPs of the two phages to identify the binding ligand on the <i>C. difficile</i> cell wall-----	158
Chapter 7	Appendices-----	160
7.1	Appendix 1: Purification of bacteriophages-----	161
7.2	Appendix 2– Protein Sequences confirmation -----	163

List of Tables

Table 1-1: Examples of Known phages receptors for phages that infect Gram-negative and positive bacteria	14
Table 1-2. A list of some of the phages that target Gram-positive bacteria that have their RBPs have been identified.....	18
Table 3-1:- Outline of the primers that were designed to amplify the genes that encode the tail proteins for CDHS1 and CDMH1 phages	60
Table 3-2 :- Reagents used for the PCR master-mix to amplify the target tail proteins.....	60
Table 3-3 :- 12%&15%SDS–PAGE resolving gel. Sufficient for two 1.0mm mini-gels	65
Table 3-4 :- 5% SDS –PAGE stacking gel. Sufficient for two 1.0mm mini- gels	65
Table 3-5 : -Reagents used for the ligation Reaction.....	70
Table 5-1:- List of the conditions that have been used to produce Gp22 protein Crystals	137
Table 5-2:- X ray data collection of Gp 22 protein.....	146

List of Figures

Figure 1-1: Image representing a Pseudomembranous Colitis	3
Figure 1-2: A proposed model to represent <i>C. difficile</i> infection and the potential phage intervention	7
Figure 1-3 : represent the two-main forms of Bacteriophage life cycle	9
Figure 1-4 : Schematic representation of baseplate genes location of four siphoviruses	20
Figure 1-5: Diagram represent the complete structure of three <i>Lactococcus</i> phages of known structure.....	21
Figure 1-6: Structure of Dit protein of SPP1 phage that infect <i>B. subtilis</i>	23
Figure 1-7: Structure of RBPs of two phages that infect <i>L. lactis</i>	25
Figure 1-8 Diagram of the cell wall structure of <i>C. difficile</i>	29
Figure 1-9: TEM image represent the overall structure for the two phages used.....	30
Figure 2-1: Schematic representation of the baseplate genes for CDHS1 and CDMH1 phages	36
Figure 2-2: in silico structural analysis of the Gp18.....	39
Figure 2-3 : in silico structural analysis of Gp19.....	40
Figure 2-4: in silico structural analysis of Gp21.....	41
Figure 2-5: in silico analysis of gp23 of CDMH1	43
Figure 2-6: in silico analysis of Gp24 from CDMH1	44
Figure 2-7: in silico analysis of Gp25 and Gp26 of CDMH1.....	45
Figure 3-1: Schematic map of PLEICS-01 used to make His tagged Gp18 protein	63
Figure 3-2: Schematic map of PLEICS-02 used to make GST tagged Gp19 andGp22 proteins.....	66
Figure 3-3: Schematic map of PGEX-4T-1 used to make GST tagged Gp21	68

Figure 3-4:- PCR products for the three putative genes encoding tail fiber proteins for phage CDHS1 separated on a 1% agarose gel	73
Figure 3-5: PCR products of CDHS1 gp21 gene separated on a 1% agarose gel	74
Figure 3-6: PCR products for the two putative genes encoding tail fiber proteins for phage CDMH1	75
Figure 3-7 : Purification of the His tagged Gp18 protein	77
Figure 3-8 : SDS-PAGE for the purification of GST tagged Gp19	78
Figure 3-9: Purification of GST tagged Gp22	80
Figure 3-10: Purification of the GST tagged Gp21 protein	82
Figure 3-11: SDS PAGE for CDMH1 GP 29 purification	83
Figure 3-12: purification of GST tagged Gp30 from CDMH1.....	85
Figure 4-1 Investigation of the role of Gp18 protein in phage CDHS1 adsorption.....	102
Figure 4-2 :- Investigation of the role of Gp21 protein in phage CDHS1 adsorption ..	104
Figure 4-3 :- Neutralization of CDHS1 infection with rabbit anti-Gp22	106
Figure 4-4 :- Immunogold labelling of tail protein Gp18.....	108
Figure 4-5 Immunogold labelling of tail protein Gp21	110
Figure 4-6 Immunogold labelling of tail protein Gp22	112
Figure 4-7 Neutralization of CDMH1 infection with rabbit anti-Gp29.....	114
Figure 4-8 Neutralization of CDMH1 infection with rabbit anti-Gp30.....	115
Figure 4-9 Neutralization of CDHM3 infection with rabbit anti-Gp29 & anti-Gp30 ..	117
Figure 4-10 Neutralization of CDHM6 infection with rabbit anti-Gp29 & anti-Gp30	119
Figure 5-1 Schematic diagram of protein crystallization phase	130
Figure 5-2 Schematic diagram represent the X- ray diffraction	131
Figure 5-3: Schematic representation of Bragg's law	132
Figure 5-4: Gp22 crystal image	140

Figure 5-5 : Gp22 protein crystal image represent	142
Figure 5-6 : Gp22 crystal images.....	144
Figure 5-7: Gp 22 Crystals image.....	145
Figure 5-8: diffraction pattern for the native GP22 protein.....	147

Abbreviations

ϕ	Phage symbol
BHI	Brain Heart Infusion
Bp	Base Pair
BppL	Lower Baseplate Protein
BppU	Upper baseplate protein
CDAD	<i>Clostridium difficile</i> -Associated Diarrhoea
CDC	Centres for Disease Control and Prevention
CDI	<i>Clostridium difficile</i> Infection
Cfu	Colony forming units
CIMs	Convective Interactive Medias
Dit	Distal Tail Protein
DTT	1, 4-Dithiothreitol
DHB	Defibrinated Horse Blood
DH ₂ O	Distilled Water
DNA	Deoxyribonucleic Acid
EDTA	Ethylene Diamine Tetra Acetic Acid
ELISA	Enzyme-Linked Immunosorbent Assay
ESCMID	European Society of Clinical Microbiology and Infectious Diseases
FAB	Fastidious Anaerobic Broth
GFP	Green fluorescent protein
GST	Glutathione S-transferases
HHpred	Homology Detection and Structure Prediction
His	Histidine
HMW	High-molecular-weight protein
Hol	Holin
IPTG	Isopropyl β -D-1-thiogalactopyranoside
Kb	Kilo Base

kDa	Kilo Dalton
LB	Luria Broth
LMW	Low-molecular-weight protein
LPS	Lipopolysaccharide
NAP1	North American Pulsed-Field Type 1
OD	Optical Density
OmpC	Outer membrane porin C
ORF	Open Reading Frame
PaLoc	<i>C. difficile</i> Pathogenicity Island
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PFU	Plaque Forming Units
PHYER2	Protein Homology/Analogy Recognition Engine
Ply	Lysin
PNACL	Protein nucleic acid chemistry laboratory
PROTEX	Protein Expression Laboratory
QA	Quaternary Amine
R	Ribotype
RBPs	Receptor Binding Proteins
Rpm	Round per minute
S-layer	Surface layer proteins
SDS	Sodium dodecyl sulphate
Slp A	S-Layer Protein A
SM	Saline Magnesium
Tal	Tail-associated lysin protein
TcdA	<i>C. difficile</i> Toxin A
TcdB	<i>C. difficile</i> Toxin B
TFB1	Transformation Buffer 1

TFB2	Transformation Buffer 2
TEM	Transmission Electron Microscopy
TMP	Tape Measure Protein
TEV	Tobacco Etch Virus nuclear-inclusion-a endopeptidase
UpH ₂ O	Ultrapure water
VAPGHs	Virion-associated peptidoglycan hydrolases
TMP	Tape Measure Protein

Chapter 1 General Introduction

1.1 Introduction

1.1.1 *Clostridium difficile*

Clostridium difficile or *Peptoclostridium difficile* is an anaerobic, Gram-positive, spore-forming bacterium that is responsible for a range of gastrointestinal diseases in humans. Although it was isolated in 1935 by Hall and O'Toole from a stool sample of a healthy child, it wasn't until the late 1970s when it was discovered that it could cause disease in humans (Rodriguez-Palacios *et al.*, 2013). *C. difficile* infection (CDI) is often associated with antibiotic treatment due to the disruption of the host gut microbiota. That allows overgrowth of endogenous *C. difficile* (strains present in the person) or exogenous *C. difficile* (strains acquired from an external source) (Burke *et al.*, 2014, Hargreaves *et al.*, 2014). Although, more than one strain of *C. difficile* can contribute to cause mixed infections, this is believed to occur only in low frequency, with the majority of infections caused by a single strain (Hargreaves *et al.*, 2014).

CDI often occurs via the faecal-oral route, where patients become infected with *C. difficile* during hospitalisation. However, they generally become susceptible to infection after being treated with broad spectrum of antibiotics that disrupt the gut microbiota, leading to *C. difficile* being allowed to grow and proliferation (Vaishnavi, 2015). Patients at hospital are at risk of *C. difficile* infection, due to exposure to spores that contaminate surfaces. Then these spores can germinate and transform into vegetative forms, colonize the large intestine, where they grow and produce toxins (Vaishnavi, 2015).

CDI ranges from mild to severe, and sometimes lethal; the course of *C. difficile* infection starts with mild diarrhoea and abdominal cramps. Severe CDI results in colitis high fever of more than 38.3°C, with signs of peritonitis, leucocytosis, and elevated serum lactate levels, pseudomembranous colitis as shown in Figure (1-1), toxic megacolon, thick colonic wall and ascites. Furthermore, in some cases, colonic perforations and sepsis might occur, and lead ultimately to death. The causes of morbidity and mortality in CDI

ranges from dehydration to gastrointestinal haemorrhage (Martinez *et al.*, 2012, Vaishnavi, 2015).

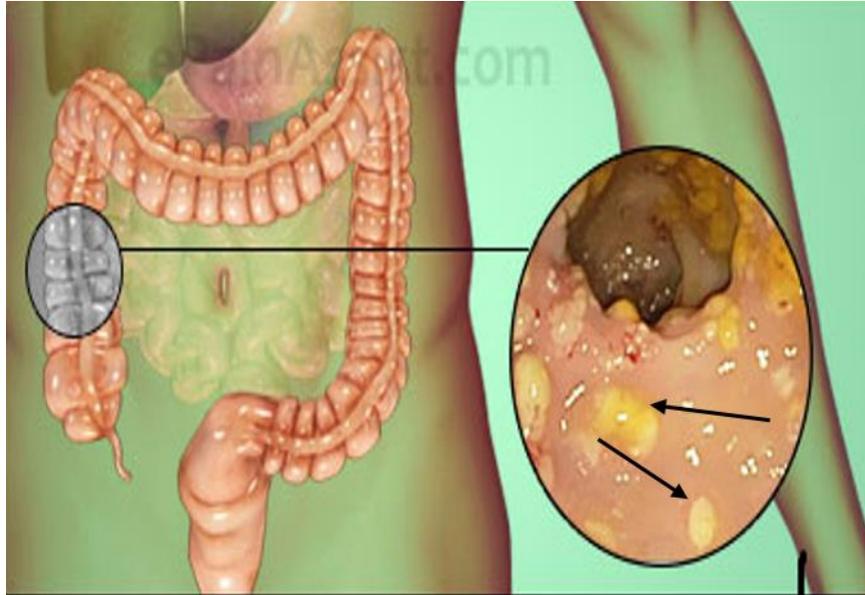


Figure 1-1: Image representing a Pseudomembranous Colitis

This condition resulted from *C. difficile* colonisation. Pseudomembranous Colitis in this image are the raised yellow plaques indicated by the black arrow. This image was obtained from the link below.

<https://www.epainassist.com/abdominal-pain/intestine/pseudomembranous>

CDI is the cause of around 39% of antibiotic-associated diarrhoea in developed countries. It has been reported that around 453,000 infections occur in the United States every year and approximately 172,000 and 18,000 in Europe and England, respectively (Borren *et al.*, 2017, Knight *et al.*, 2015). The geographical distribution of CDI has occurred due to the emergence of diverse virulent ribotypes (strains) such as North American pulsed-field gel electrophoresis type 1 (NAP1) strain, that cause CDI and their resistance to antibiotic courses (Knight *et al.*, 2015, Mónica Oleastro1 *et al.*, 2014, Oleastro *et al.*, 2014). Moreover, these strains have the capability to readily form spores in comparison to the wild types (Ivarsson *et al.*, 2015). *C. difficile* most commonly infects elderly people, however, it has been reported that *C. difficile* may infect young adults, children and

General Introduction

pregnant women, groups who were previously considered to be at low risk from infection (Khanna *et al.*, 2013). One of the key issues that healthcare facilities face during CDI management is the recurrence of infection, as it has been reported that up to 25% of CDI patients experienced a recurrence of the infection within 30 days after an antibiotics course. Recurrent CDI is problematic in terms of treatment, with an impact on long period of hospitalization time and the costs. Patients who immunosuppressed, renal impaired or who are aged 65 years or older are particularly at risk (Aguado *et al.*, 2015).

1.1.2 *C. difficile* pathogenicity

As stated above, *C. difficile* is transmitted through the faecal-oral route, where the bacterium is either at the vegetative or the spore form. The spore form could remain viable outside the host for extended period of time due to its capacity to resist harsh environmental conditions such as temperature and heat. In the gut, spores will germinate, colonize the colon, produce toxins and cause CDI (Khanna *et al.*, 2013). The two main virulence factors that are always associated with the pathogenicity of *C. difficile* are toxin A (TcdA) and toxin B (TcdB), both of which are encoded in the PaLoc Pathogenicity island (Rupnik *et al.*, 2009). Toxin A and B are the main cause for the symptoms experienced by CDI patients. Furthermore, the traditional gold standard to diagnose CDI is a cytotoxic assay that detects toxin B or toxin A depending on the cell line used (Rupnik *et al.*, 2009).

Toxin B and toxin A are large secreted proteins which have four structurally homologous domains. They harbour RHO and RAC glucosyltransferase domains (GTDs) that mediate toxicity by glycosylating and thereby inactivating bacterial RHO and RAC GTPases, that cause damage to normal cytoskeletal architecture and the tight junction of the cell (Hunt *et al.*, 2013). This in turn leads to increased vascular permeability, inflammation and cell death amongst the epithelial cells at the surface of the colon (Pruitt *et al.*, 2012).

General Introduction

In addition, a third toxin is produced by some strains of *C. difficile* known as the binary toxin. Infection with strains producing this toxin are correlated with high rates of death, compared with the non-binary toxin producing strains. It is therefore thought that this toxin may be involved in the pathogenicity of *C. difficile* (Vedantam *et al.*, 2012). Furthermore, many other components may contribute to the *C. difficile* pathogenicity. For example, Surface layer proteins (S-layer) that surrounds the whole *C. difficile*. Unlike the S-layer of other organism that composed of one or two proteins encoded by different genes. The S-layer of *C. difficile* is composed of two major proteins that result from the cleavage of a single Slp A protein, this occurs by Cwp84, a cysteine protease. The two proteins are a high-molecular-weight HMW protein, and a low-molecular-weight (LMW) protein. It has been found that the LMW protein contributes to the adherence of *C. difficile* to human epithelial cells (Hunt *et al.*, 2013, Merrigan *et al.*, 2013). Another, important factor that mediates *C. difficile* attachment to the host and may enhance the colonization is the Flagella, specifically the flagellar FliD protein (Hunt *et al.*, 2013, Tasteyre *et al.*, 2001).

1.1.3 *C. difficile* treatment

Treatment of CDI has become critically problematic, due to this organism being naturally resistant to antibiotics and also due to the increased development of antibiotic resistance by this organism. There are currently three antibiotics, Vancomycin, Metronidazole, and Fidaxomicin that are available as first-line treatments for CDI. Vancomycin and Metronidazole, are the first options used. They can be administered separately or in combination (Nale *et al.*, 2016). In 2014, the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) guidelines suggested that metronidazole should be used as the initial cure for mild-to-moderate CDI, and oral vancomycin is for more severe cases (Crowther *et al.*, 2015). Fidaxomicin is a narrow-spectrum antibiotic with low activity against normal gut microbiota but is efficient against *C. difficile*. Fidaxomicin

General Introduction

has been reported to be as good as vancomycin for initial cure in patients with CDI, and lower recurrence rates have been noticed when CDI patients are treated with Fidaxomicin, compared to Vancomycin (Crowther *et al.*, 2015, Cornely *et al.*, 2012). However, cases of antibiotic resistance have been reported for Vancomycin and Metronidazole, in addition to the recurrence rates associated with CDI treatment with these antibiotics (Crowther *et al.*, 2015). Although Fidaxomicin has a similar effect to Vancomycin, the high cost of the former makes it clinically restricted in terms of its use (Zucca *et al.*, 2013).

Therefore, due to the emergence of *C. difficile* antibiotic resistance, and the lack of discovery of new antibiotics to control or eradicate this organism, scientists have been brought back to the era before discovery of antibiotics. The attempt to develop an innovative approach to overcome CDI has become a major concern in recent years. Many alternative methods have been proposed for CDI treatment, for example vaccine development, faecal transfer, and phage therapy, the latter appearing quite promising for bacterial infection treatment (Zucca *et al.*, 2013). Phages could be an attractive source of antimicrobials agents to treat for bacterial infections due to their specificity, minimal disruption of microbiota, the ability to self-amplify at the site of infection, no or minimal side effects, and the fact that phages may act in a synergic manner with antibiotics (Nale *et al.*, 2016). Figure 1-2 illustrates the stages that occur during CDI and the possible effect of phage therapy.

General Introduction

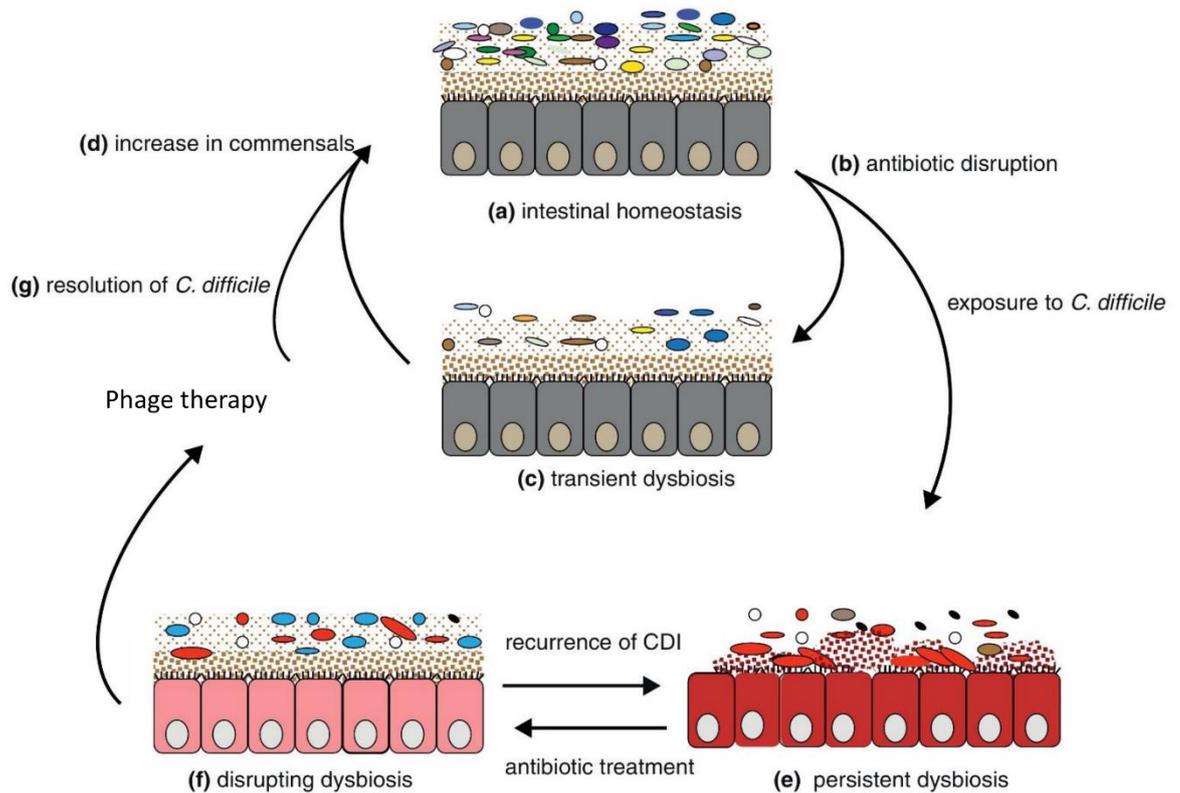


Figure 1-2: A proposed model to represent *C. difficile* infection and the potential phage intervention

(a) Intestinal homeostasis is featured by a stable and diverse gut microbiota and health-associated metabolites. (b–c) Exposure to the antibiotic impairs the equilibrium of the microbiota leading to a reduction in diversity and a lack of colonization resistance. (d) After the antibiotics are stopped, diversity is restored. (e) Microorganisms such as *C. difficile* starts to colonize, subsequently, infection and persistent microbial imbalances occur. (f) After treatment of CDI with antibiotics such as vancomycin, further microbiota disruption occurs potentially leading to recurrent CDI after discontinued use of the antibiotic. (g) Bacteriophages are proposed as an alternative tool to eliminate *C. difficile* and they could have an increase in species diversity of the microbiota (d) and restore intestinal homeostasis. This Figure was modified from (Brandt, 2013)

1.2 Bacteriophage background

Bacteriophages or phages, are bacterial viruses that are thought to be the most abundant and diverse biological entities on earth. They can have either single or double stranded RNA or DNA genomes and these are surrounded by a protein capsid connected to their protein tail, this phage tail binds to the host receptor sites which could be a protein or carbohydrate (Haq *et al.*, 2012). The vast majority of phages that have been morphologically characterised belong to the order *Caudovirales* which have dsDNA genomes. Within this order there are three families that can be distinguished from one another on their overall appearance, and of particular relevance to this project, their tail morphology. Siphoviruses possess long non-contractile tails, but Myoviruses have contractile tails, and Podoviruses have short tails (Spinelli *et al.*, 2014a).

These bacterial viruses rely on their host for their replication, and they have four different life cycles that have been observed. The first is known as the lytic life cycle, where the phages penetrate their bacterial hosts, multiply, and then mature progeny are released after they lyse the bacterial cells. Lytic phages are a source of promising novel therapeutics and indeed have long history of use to treat bacterial infections (Czaplewski *et al.*, 2016). The second life cycle is termed the lysogenic life cycle and here the phage genome integrates into the bacteria genome, often remaining integrated for prolonged periods. However, if the bacteria encounters stressful conditions, such as ultraviolet radiation or antibiotics, the integrated (temperate) phage is induced and enters the lytic cycle (Haq *et al.*, 2012). Figure 1-3 describes the lytic and lysogenic life cycle of bacteriophage. Phages that follow the lysogenic cycle are not desirable for therapeutic purposes, due to their ability to stay silent in the cycle for several generations without affecting the bacterial host. Moreover in some cases, the prophage may encode virulence genes which can be horizontally transferred from one bacteria to another by transduction (Wittebole *et al.*, 2014).

LIFE CYCLE :

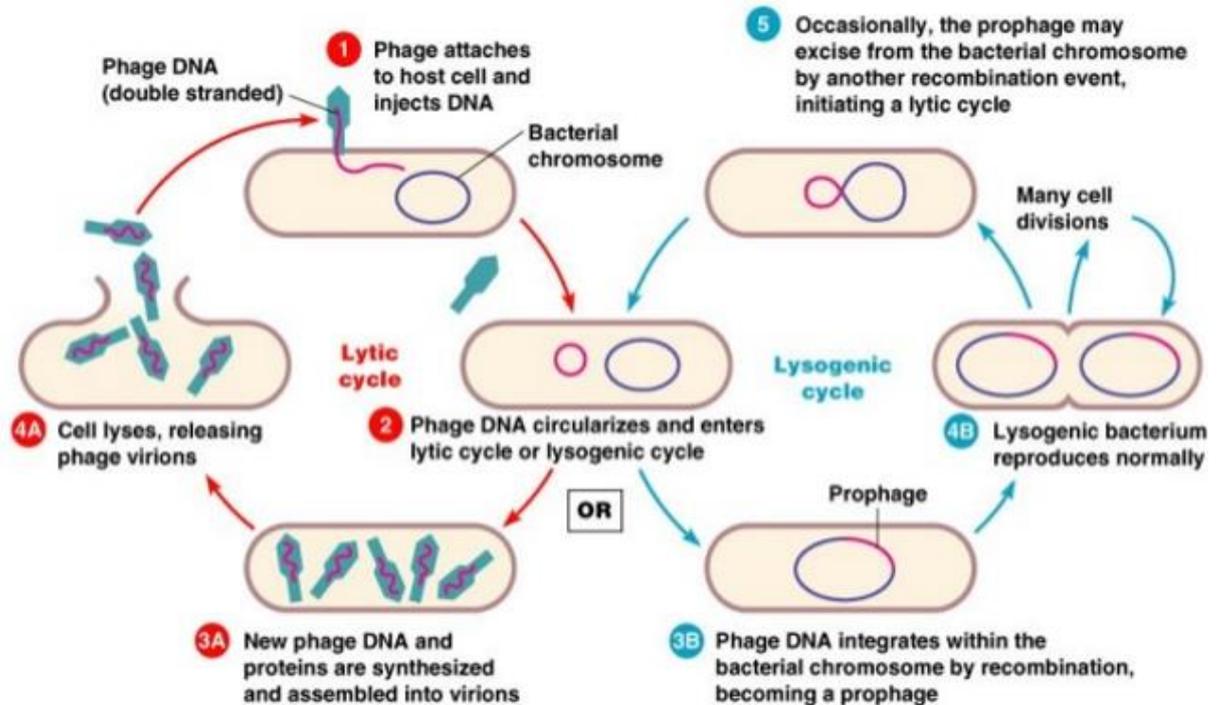


Figure 1-3 : represent the two-main forms of Bacteriophage life cycle

The lytic life cycle (highlighted in red) and Lysogenic life cycle (in blue) that phage undergo during the phage bacterial infection. This Figure taken from the link below.

<http://www.majordifferences.com/2013/03/difference-between-lysogenic-phase-and.html#.WoVN3a5l-Uk>

The third form of life cycle is called pseudo-lysogenic. In this stage, the infecting phage neither integrate with host genome nor lysis the cells, and remains in this pseudo-lysogenic stage until unfavourable physical conditions improve, after which it can resume the lytic or lysogenic pathway (Clokie *et al.*, 2011, Dang *et al.*, 2014). Chronic infection is another form of phage life cycle and it is seen in several archaeal viruses and *E. coli* filamentous phages. In this form of phage life cycle, the phage infects the host for a prolonged period of time and novel particles exit the cell without causing cell lysis as the

progeny passed down to daughter cells asymmetrically post to division (Clokic *et al.*, 2011, Dang *et al.*, 2014).

1.2.1 Application of bacteriophages

The high specificity of phages towards their host enables them to be exploited as tools for different purposes and in different fields, such as agriculture where phages are used to control and prevent plant diseases (Alvarez *et al.*, 2017, Buttimer *et al.*, 2017). Since phages do not cause any harm to mammalian cells, they have been recognised and used in various applications within the food industry. Phages are also used as alternatives to antibiotics in animal health, as biopreservatives in food (Garcia *et al.*, 2008). Several examples have been reported indicating the exploitation of phages throughout the food chain, such as using a cocktail of three phages (e11/2, e4/1c and pp01) on the surface of the beef meat, which caused a significant elimination of the pathogen *E. coli* O157:H7 (O'Flynn *et al.*, 2004). Applying specific phages in the food industry helps in preventing the decay of food products, the spread of pathogenic bacteria, and enhance the safety of animal and plant foods (Sillankorva *et al.*, 2012).

In the medical field, phages therapy offer a promising way to kill pathogens at the site of the infection (Aminov *et al.*, 2017). The idea to utilise phages for therapeutic purposes came into existence 100 years ago, indeed almost immediately after the phages were first discovered. After, phage therapy was quickly recognised, and it has been developed in many countries. A successful phage treatment for various bacterial infections such as dysentery, cholera, and other human infections has been reported (Kutateladze *et al.*, 2008, Jun *et al.*, 2016). In spite of the optimistic early results from phage therapy, it was largely abandoned, and the focus of phages understandably shifted as soon as the antibiotics were discovered (Aminov *et al.*, 2017).

In certain countries such as in Eastern Europe and the former Soviet Union, phage therapy research was continued, and phage preparations were commercialised and are sold in

pharmacies (Pirnay *et al.*, 2011). Conversely, phage therapy development has only recently become the main focus of research in Western Europe and the US, mainly motivated by the increase in antibiotic resistance by bacteria, which has driven the demand to establish a new method for treatment (Aminov *et al.*, 2017, Pirnay *et al.*, 2011). Other limitations in developing the phages in therapeutic purposes, include the lack of knowledge pertaining to phage biology, regulation, and non-standard methods for manufacturing, all of which raise reasonable doubts among physicians and clinicians (Drulis-Kawa *et al.*, 2015).

The high specificity of the phages toward their host makes phages an ideal tool for the use in bacterial diagnosis (Schofield *et al.*, 2012). Phage-based detection assays such as phage amplification, reporter phage, phage-labelling and phage capture elements have been developed as diagnostic tools for bacterial infections caused by *Salmonella*, *E. coli*, *Mycobacterium tuberculosis*, *Yersinia pestis*, *Bacillus anthracis* and *Staphylococcus aureus* (Schofield *et al.*, 2012, Schmelcher *et al.*, 2014). These assays rely on utilising the whole phage. However, phage products such as receptor binding proteins (RBPs) or cell wall binding domains of the endolysin protein may be also used as detection methods to identify the cause of the bacterial infection (Schmelcher *et al.*, 2014). Nevertheless, before utilising phages in any of the proposed applications in both therapeutic and diagnostic, it is important to understand the various aspects of bacteriophage biology mainly phage bacterial host interaction.

1.2.2 Phage adsorption

By definition, the physical phage interaction with bacterial hosts is a *sine qua non* or absolutely essential process for phage infection and is in most cases thought to be highly specific. The infection process starts with phages attaching to one or more components, or receptors on their host bacteria and then penetrating the cell membranes (Rakhuba *et al.*, 2010). This is mediated via interactions between proteins located at the

distal end of the phage, they are the receptor binding proteins (RBPs) of phages (sometimes termed as anti-receptors or adhesins) and ligands (receptors) on the bacterial surface. This highly specific binding is termed adsorption and has two main steps (Mahony *et al.*, 2012).

The first is when phages bind reversibly to a constituent on the bacterial surface. Here, phage-bacterial binding is not complete, and phages can be desorbed from bacteria as evidenced by the presence of viable phages within supernatant (Bertozi Silva *et al.*, 2016). During the second step however, phages irreversibly attach to either the same receptor as in the first step of the adsorption, or to a second receptor. Following this step, the phages penetrate the host cell and insert their DNA. This binding between phages and bacteria is sometimes assisted by enzymatic cleavage that helps phage DNA to be injected (Sao-Jose *et al.*, 2004, Bertozi Silva *et al.*, 2016, Dowah *et al.*, 2018).

1.2.3 Bacteriophage receptors on the surface of the Bacteria

It is possible that all the molecular structures on the surface of the bacterium may act as phage receptors. It is worth noting that the nature of these receptors (protein, polysaccharide, lipopolysaccharides (LPS) and carbohydrate moieties), in addition to their position on the surface of the bacteria, are diverse, as based on both the phage and bacteria (Bertozi Silva *et al.*, 2016). In Gram-negative bacteria, for example, lipopolysaccharides (LPS) are commonly considered to be the phage binding site. In addition to other structures on the surface of Gram -negative bacteria that may act as phage receptors, such as outer membrane proteins, pili and flagella (Sorensen *et al.*, 2011). The T4 phage that infects *E. coli* is one of the well-studied phages which belongs to Myoviridae family. It attaches reversibly to LPSs or with the outer membrane protein porin OmpC on the surface of the host. This attachment additionally leads to irreversible binding of T4 with the outer core region of the *E. coli* LPS. The T7 phage, which infects

General Introduction

E. coli, utilises LPSs as phage binding sites (Sorensen *et al.*, 2011, Gonzalez-Garcia *et al.*, 2015).

For phages that infect Gram-positive bacteria, peptidoglycan is an important phage receptor, as it is a major polymer on bacterial surfaces, along with teichoic acids, that are attached covalently to the peptidoglycan layer (Bertozi Silva *et al.*, 2016). Polysaccharides that are exposed on the surface of the bacteria are also common receptors (Bertozi Silva *et al.*, 2016). In many ways it is surprising that only a small number of phage receptors have been identified for Gram-positive bacteria; this is in part due to their complex outer structure and in part due to the scarcity of research activities on phages that target Gram-positive bacteria in general (Dowah *et al.*, 2018).

The vast majority of such receptors that have been recognised for these type of bacteria are related to teichoic acid or peptidoglycan. It has been reported that out of 30 phages that infect Gram-positive bacteria, 10 use other structures to bind with the host, and among these nine utilise residues of either teichoic acid or peptidoglycan, indicating the importance of these two components as phage receptors (Bertozi Silva *et al.*, 2016).

Table 1-1: Examples of Known phages receptors for phages that infect Gram-negative and positive bacteria

Host	Phage	Phage Family	Name of the receptors on the host	Reference
<i>Escherichia coli</i>	Lambda	Siphoviridae	LamB	(Chatterjee <i>et al.</i> , 2012)
<i>E. coli</i>	T5	Siphoviridae	FhuA	(Mahony <i>et al.</i> , 2012)
<i>Pseudomonas aeruginosa</i>	E79	Myoviridae	Core polysaccharide of LPS	(MEADOW <i>et al.</i> , 1978)
<i>P. aeruginosa</i>	JG004	Myoviridae	LPS	(Garbe <i>et al.</i> , 2011)
<i>Salmonella</i>	ES18	Siphoviridae	Protein FhuA	(Killmann <i>et al.</i> , 2001)
<i>Salmonella</i>	Gifsy-1 Gifsy-2	Siphoviridae	Protein OmpC	(Ho <i>et al.</i> , 2001)
<i>Salmonella</i>	vB SenM- S16	Myoviridae	Protein OmpC	(Marti <i>et al.</i> , 2013)
<i>Streptococcus thermophilus</i>	OBJ	Siphoviridae	Glucosamine/ Ribose	(Quiberoni <i>et al.</i> , 2000)
<i>S. thermophilus</i>	CYM	Siphoviridae	Glucosamine/ Rhamnose	(Quiberoni <i>et al.</i> , 2000)
<i>Lactococcus Lactis</i>	TP901-1	Siphoviridae	Saccharide	(Mahony <i>et al.</i> , 2012)

General Introduction

<i>L. Lactis</i>	Tuc2009	Siphoviridae	Saccharide	(Mahony <i>et al.</i> , 2012)
<i>Bacillus Subtilis</i>	SPP1	Siphoviridae	YueB/	(Sao-Jose <i>et al.</i> , 2004)
<i>Staphylococcus aureus</i>	W	Siphoviridae	N-acetylglucosamine (GlcNAc) glycol peptide on WTA	(Xia <i>et al.</i> , 2011)
<i>S. aureus</i>	φ812 φ K	Myoviridae	Anionic backbone of WTA	(Xia <i>et al.</i> , 2011)

1.2.4 Phage receptor binding proteins (RBPs)

Phage RBPs or tail fiber proteins are key to phage specificity determination, therefore one of the important role of these proteins is being exploited as therapeutic tools to reduce host colonisation, for example, it has been shown that the RBPs (tails pike proteins) of the Podovirus P22 that infect *Salmonella enterica* serovar *Typhimurium*, reduced the colonisation of *S. enterica* in the gut of chicken and the penetration into other internal organs (Simpson *et al.*, 2016, Waseh *et al.*, 2010). RBPs are also promising diagnostic tools for various bacterial infections due to their specific selectivity for their host; for example, RBPs have been used to detect *Campylobacter jejuni* and *Campylobacter coli* (Javed *et al.*, 2013). Therefore, the study and characterization of these proteins is crucial to improving our understanding of the mechanism by which phages adsorb to their bacterial hosts. The stability, the affinity to the carbohydrate, and the specificity to the receptors on the surface of the bacteria make Phage RBPs a potential candidate for the therapeutic and diagnostic applications (Simpson *et al.*, 2016).

These RBPs are structurally diverse between phages due to phage morphology differences and the mechanisms by which they attach to the bacteria. For example, myovirus such as T4 binds to its bacterial host through long- and short-tail fibers, whereas in siphoviruses such as the *L. lactis* phage TP901-1 the RBPs are found within a large complex structure known as a baseplate, or can be found in a tail spike such as in *B. subtilis* phages SPP1 (Simpson *et al.*, 2016, Veessler *et al.*, 2012, Walter *et al.*, 2008).

Phage RBPs have been well studied and characterised for phages that infect Gram-negative bacteria, mainly *E. coli* phages such as T4 and lambda which have been used as examples to enhance our understanding of phage interaction with Gram-negative bacteria (Mahony *et al.*, 2015). However, in comparison, only recently, there has been an increased focus in the study of RBPs for phages that attach to Gram-positive bacteria, as

previously this mechanism was poorly understood (Mahony *et al.*, 2016b, Dowah *et al.*, 2018).

1.2.5 RBPs for phages of Gram-Negative bacteria

The interaction between phages and Gram-negative bacteria has been the main focus of research field for many years, especially for the *E. coli* phages. Phage T4 is one of the most studied and well characterized of *E. coli* phages. As stated above, the T4 phage attaches to *E. coli* cells using two different tail fibers; long tail fibers that T4 uses to adsorb reversibly to the cells, and then short tail fibers that are used to attach irreversibly to the receptors on the surfaces of the cells (Washizaki *et al.*, 2016). The mechanism that T4 applies to attach to *E. coli* is reliant on Gp37 at the tip of the long-tail fiber that binds reversibly to the lipopolysaccharide or the outer membrane protein C based on the *E. coli* strain that T4 attaches to (Bartual *et al.*, 2010). This binding will induce the short tail-fiber Gp12 to attach irreversibly to the lipopolysaccharide, after which a conformational change will occur in the baseplate allowing it to transfer from a hexagonal to a star shape form (Hu *et al.*, 2015). The baseplate of T4 phage consists of a hub and six surrounding wedges. The proteins that form the central hub are Gp5, Gp27 and Gp29, whereas the six surrounding wedges are formed by Gp11, Gp10, Gp7, Gp8, Gp6, Gp53 and Gp25 (Arisaka *et al.*, 2016, Taylor *et al.*, 2016, Kostyuchenko *et al.*, 2003).

Bacteriophage lambda is another phage that infects *E. coli* K-12, and the C-terminal of Gp j is the RBPs that the lambda phage uses to adsorb to its host (Wang *et al.*, 2000). Phage p22, that infects *S. enterica*, uses the tail spike Gp 9 to bind and degrade the O-antigen surface polysaccharide on the surface of the host (Olia *et al.*, 2007). Recently, two phages, JG004 and PaP1, that infect *P. aeruginosa* have had their RBPs identified as ORF84 and ORF69, respectively (van Raaij *et al.*, 2013).

1.2.6 RBPs for Siphoviruses of Gram-positive bacteria

The majority of phages that infect Gram-positive bacteria and have had their RBPs been characterized are belong to the Siphoviridae family. This has largely been driven by the fact that phages from this family are a significant problem in dairy fermentation, as they lysed the lactic acid bacteria, and thus they have been thoroughly studied (Dupont *et al.*, 2004). Table 1-2 presents some of Siphoviruses of Gram-positive bacteria that have their RBPs being characterised.

Table 1-2. A list of some of the phages that target Gram-positive bacteria that have their RBPs have been identified

Phages	Host	RBPs	References
bIL170	<i>L. lactis</i>	ORF20	(Dupont <i>et al.</i> , 2004)
TP901-1	<i>L. lactis</i>	BppL	(Vegge <i>et al.</i> , 2006)
Tuc2009	<i>L. lactis</i>	ORF53	(Vegge <i>et al.</i> , 2006)
SPP1	<i>B. subtilis</i>	Gp21	(Vinga <i>et al.</i> , 2012)
A118	<i>Listeria.</i> <i>monocytogenes</i>	Gp19 &Gp20	(Bielmann <i>et al.</i> , 2015)
P35	<i>L. monocytogenes</i>	Gp16	(Bielmann <i>et al.</i> , 2015)
Phage φ11	<i>S. aureus</i>	Gp45	(Li <i>et al.</i> , 2016a)

RBPs are generally identified using bioinformatics analysis, coincide with antibody studies, for example using polyclonal antibodies raised against overexpressed putative tail fiber proteins, to neutralise phage infection (Li *et al.*, 2016b, Dowah *et al.*, 2018). Molecular approaches are also used, for example, a chimeric phage was produced to identify the RBPs of the *L. lactis* phages TP901-1 and Tuc2009. In this work, the gene

General Introduction

encoding the TP901-1 lower baseplate protein (bppL) was replaced with the gene (orf53) from phage Tuc2009 (Vegge *et al.*, 2006). The results showed that the chimeric TP901-1 phage was able to infect the Tuc2009 host strain efficiently, indicating that the TP901-1 lower baseplate protein (bppL) and (orf53) of phage Tuc2009 are both responsible for the phage attachment to the bacteria (Vegge *et al.*, 2006). To gain further insights into how these phages attach to their host, structural information was determined using X-ray crystallography and morphological information from Transmission Electron microscopy (TEM) (Mahony *et al.*, 2012, Dowah *et al.*, 2018).

1.2.7 Genomic architecture and the position of the tail fiber/baseplate sequences

Based on the genome sequences of siphoviruses that infect the Gram-positive species such as *B. subtilis*, *L. lactis*, *L. monocytogenes*, *S. aureus* and *C. difficile*, it can be seen that genes encoding tail proteins, or proteins that form the baseplate, are located between *tmp* genes encoding the tape measure protein TMP (that determines the length of the phage tail), and the genes encoding the holin and endolysin protein (Li *et al.*, 2016b, Dowah *et al.*, 2018). There are usually four genes located between these anchor points as shown in Figure 1-4 and Figure 1-5 one of these being the *Dit* gene that encodes the distal tail protein (Dit), which has been shown to be fairly conserved within siphoviruses. After *Dit* usually comes the *Tal* gene that encodes Tail-associated lysin protein (Tal), Tal promotes efficient phage penetration and infection by degrading peptidoglycan layer of the bacterial host (Bielmann *et al.*, 2015, Li *et al.*, 2016a, Mahony *et al.*, 2012, Stockdale *et al.*, 2013).

The *Tal* gene architecture has distinct differences within different phages, consistent with the divergent structure of Siphovirus tail tips (Spinelli *et al.*, 2014b). For example, in the *Lactococcus* phage TP901-1 and in most siphoviruses that have been characterized, there are two genes located downstream to Tal gene; one encodes the upper baseplate protein (BppU) (orf48), and the other encodes RBPs. However, in the *B. Subtilis* SPP1 phage,

General Introduction

there is no BppU but instead this phage has a tail spike protein (Bebeacua *et al.*, 2010). The four gene products (Dit, Tal, like BppU protein and RBPs) are discussed in detail below.

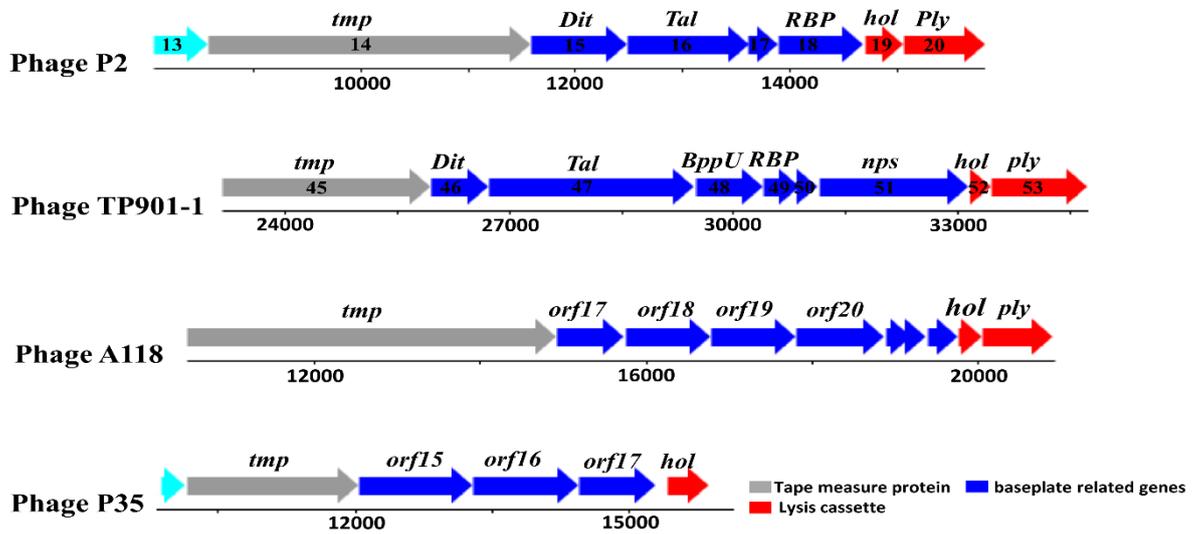


Figure 1-4 : Schematic representation of baseplate genes location of four siphoviruses

P2, TP901-1, A118 and P35 phages that infect *L. lactis* and *L. monocytogenes* respectively. Genes located between tmp and holin and endolysin encoding genes usually encode proteins that involved in baseplate formation. This Figure was taken from (Dowah *et al.*, 2018)

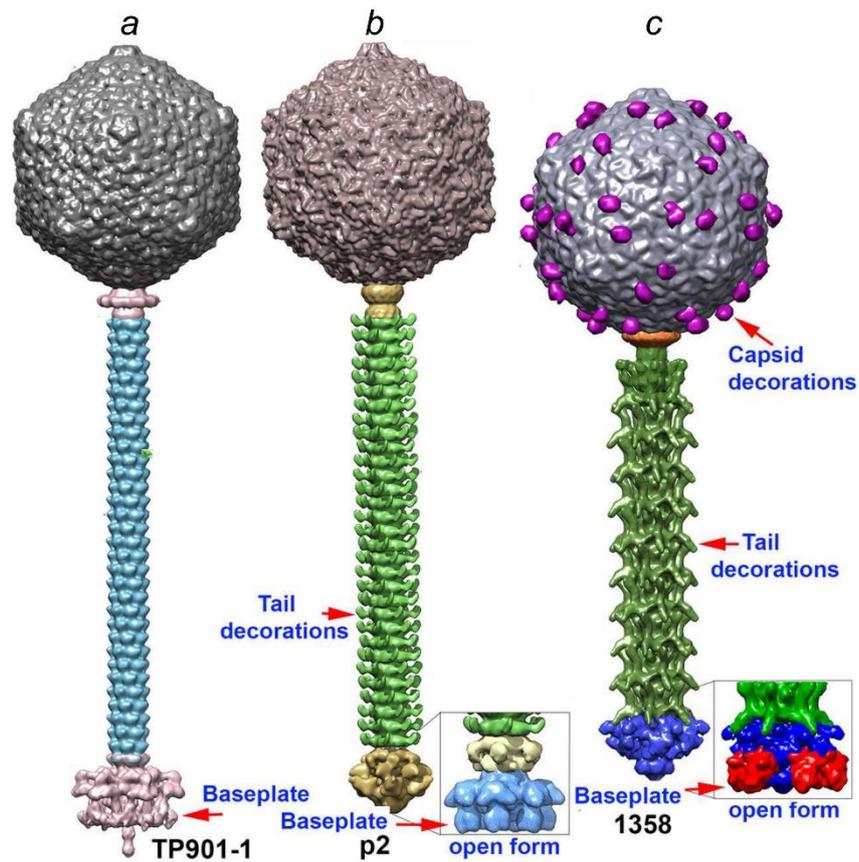


Figure 1-5: Diagram represent the complete structure of three *Lactococcus* phages of known structure

A is the complete structure of TP901-1 phage. **B** is a p2 phage structure and **C** is 1358 phage. The three phages are siphoviruses that infect *L. lactis*. And the diagram represent also a distinct structure of the phages baseplate of the three phages represented. This diagram was adapted from (Spinelli *et al.*, 2014a)

1.3 Structural aspect of Siphophages baseplate

1.3.1 Distal tail proteins (Dit)

The gene encoding the Dit protein is involved in the Siphovirus baseplate structure where it forms the central hub of the phage baseplate. Using HHpred analysis, and data from crystal structures studies, this protein can be identified in several siphoviruses, including the *B. subtilis* phages SPP1, *L. lactis* TP901-1, Tuc2009 and P2 phages, A118 and P35 the *L. monocytogenes* ϕ , and ϕ 11 phage that infects *S. aureus* (Bielmann *et al.*, 2015, Li *et al.*, 2016b, Dowah *et al.*, 2018). These diverse set of phages with a recognizable Dit protein, demonstrate a high level of protein conservation within the siphoviruses (Veesler *et al.*, 2012).

This protein is structurally organized as a hexameric ring that links the tail tube and the adsorption device of the phage. The N-terminal of this protein consists of a β -sandwich, α -helix, and a β -hairpin, whilst the C-terminal is a galectin-like β -sandwich (Veesler *et al.*, 2012, Flayhan *et al.*, 2014). Figure (1-5) shows the structure of Dit protein for *B. subtilis* phage SPP1. Structural comparisons reveal a striking similarity between SPP1 Dit and that of other phages such as the Dit of TP901-1 and P2 phages (Veesler *et al.*, 2012). Furthermore, the N terminal region of Dit has been identified in myoviruses including T4 and Mu, and in siphoviruses such as T5 that infect *E. coli* (Koc *et al.*, 2016). Thus, the conservation of protein of Dit protein extends to both families; the siphoviruses and myoviruses (Dowah *et al.*, 2018).

Dit of SPP1 Phage

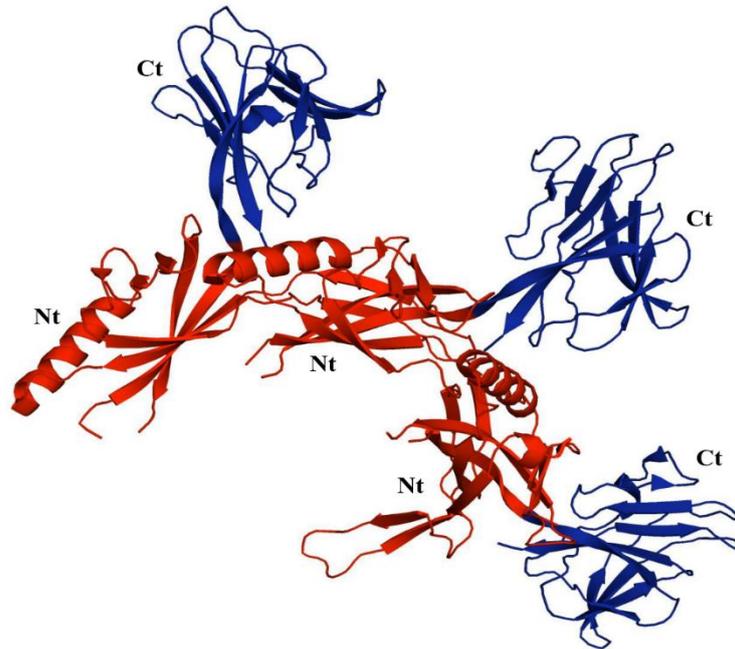


Figure 1-6: Structure of DIT protein of SPP1 phage that infect *B. subtilis*

The N- terminal of this protein highlighted in red, and C terminal was highlighted in blue (Veesler *et al.*, 2010). This picture was drawn for this manuscript using PyMol software <http://www.pymol.org/>

1.3.2 Tail associated lysine proteins (Tal)

The second gene of interest is generally located next to *dit* is *Tal*, which encodes the Tal protein that, can be released via self-cleavage at a consensus site. The Tal protein belongs to a group of proteins known as Virion-associated peptidoglycan hydrolases (VAPGHs). The main function of these proteins or enzymes is peptidoglycan layer degradation, to assist phage infection. The virion-associated muralytic activities may offer a promising tool to treat bacterial pathogens (Drulis-Kawa *et al.*, 2015, Dowah *et al.*, 2018). There are four enzymes included in (VAPGHs) based on their enzymatic activity; lysozymes for example Gp5 from T4 phage, lytic transglycosylases such as Gp16 from T7 phage, protein P7 from PRD1, glucosaminidases and endopeptidases as in protein P5 from phi6 phage and Tal2009 from Tuc2009 phage (Drulis-Kawa *et al.*, 2015, Kenny *et al.*, 2004).

General Introduction

The N-terminal domain of the Tal protein is similar to a group of proteins including ORF16 of *L. lactis* phage p2, Gp21 of phage SPP1, and ORF47 of TP901-1. In addition, the N-terminal of the Tal protein is thought to be involved in phage baseplate structure formation (Sciara *et al.*, 2008).

Interestingly, the N-terminal domain of Tal protein has been found to be conserved in phages that infect Gram-negative bacteria, such as the myoviruses T4 and Mu and siphoviruses such as T5 (Koc *et al.*, 2016). However, the C-terminus of Tal protein has proteolytic activity that is involved in the peptidoglycan cleavage, thus facilitating phage DNA injection into the bacteria. This phenomenon has been noted in Tal C-terminal of *Lactococcus* phages TP901-1 and Tuc2009. Conversely, no such activity has been established for *Lactococcus* phage p2. Instead Tal C terminal of *B. Subtilis* SPP1 phage is found to be the RBPs for this phage and involved in phage attachment with YueB on the surface of the host (Bebeacua *et al.*, 2010, Spinelli *et al.*, 2014b, Dowah *et al.*, 2018).

1.3.3 Like -Upper baseplate protein (BppU)

BppU protein is formed as six asymmetric trimers which link to the Dit central core and the RBPs. Each monomer of BppU contains an N-terminal domain, a linker and two helices joined by a kink and then C-terminal domain (Veesler *et al.*, 2012). HHpred analysis revealed that the homology of BppU protein's N-terminal was found in several phages, such as A118 that infects *L. monocytogenes*, ϕ 11 phage that infects *S. aureus* (Li *et al.*, 2016b, Biemann *et al.*, 2015, Dowah *et al.*, 2018) and in CDHS1 that infects *C. difficile*. Finally, downstream to BppU is the RBPs.

1.3.4 Receptor binding protein (RBPs)

Although the RBPs of various siphoviruses that target different Gram-positive bacteria have been identified, most of the known structures RBPs are from phages that infect *L. lactis*. The first three *Lactococcus* phages that had their RBPs structurally defined were

936, p2, P335 TP901-1 and 936 phage bIL170 (Ricagno *et al.*, 2006). The overall structures of RBPs for these phages are composed of three parts; the shoulder, the neck and the head. The N- terminus (shoulder) of this protein contains an α helix bundle, while the neck is a β prism domain connected to the C- terminus (the head), that possesses a saccharide binding site that is attached to the receptors on the surface of the organism (Sciara *et al.*, 2008). Figure 1-6 A&B shows the structure of RBPs of two phages of *L. lactis* p2 and TP901-1 respectively.

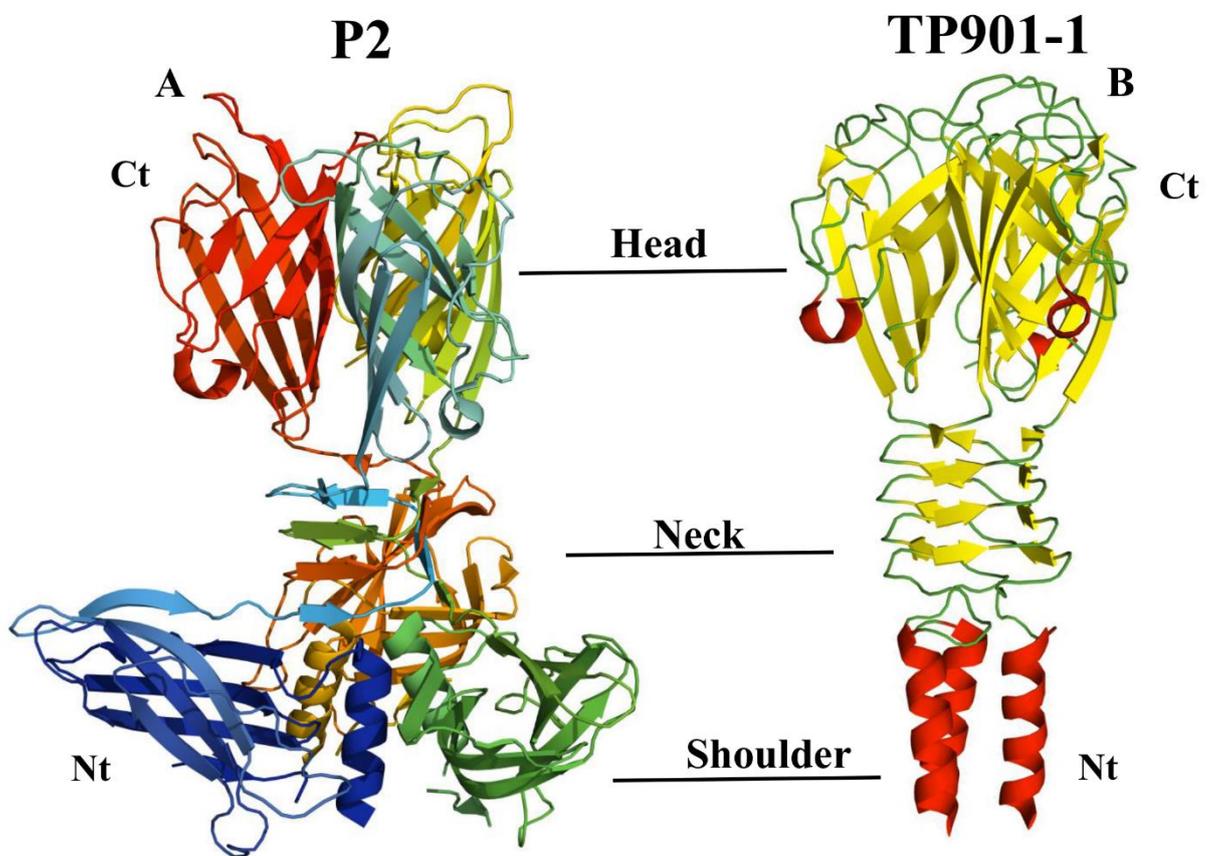


Figure 1-7: Structure of RBPs of two phages that infect *L. lactis*

P2 (A) (Tremblay *et al.*, 2006) and TP901-1 (B) (Spinelli *et al.*, 2006). This Figure was generated using the PyMol software <http://www.pymol.org/>.

General Introduction

The structural comparisons of RBPs between phages reveal that RBPs of phage bIL170 has a high sequence identity of 89 % to 936 phage p2 at the N- terminus domain, however, exhibit differences on the neck and head domains; this explains the difference in host range for these two phages (Ricagno *et al.*, 2006). On the other hand, the RBPs head domain of p335 phage TP901 is 28 % identical to the RBPs head of 936 phage p2, in contrast, no similarity has been noted at the RBPs N- terminal of those phages (Spinelli *et al.*, 2014b). Recently, the RBPs structure for phage ϕ 11 that infects *S. aureus* has been resolved. The structure of this protein also has three regions; the N-terminal domain that consists of triple-helical bundle, the central part, with three β -propeller domains, and the C-terminal “tower” region (Koc *et al.*, 2016). Compared to the RBPs structure of *L. lactis* phages, the RBPs of phage ϕ 11 is distinct from the RBPs of any known phages, apart from the first 30 amino acids being similar to that of the P335 phage and TP901 phage (Dowah *et al.*, 2018, Koc *et al.*, 2016).

1.3.5 RBPs shape depend on the bacterial receptors nature

The information that has been obtained from studying siphoviruses RBPs, particularly the phages that infect Gram-positive bacteria allows for initial prediction of the nature of the host receptor being (protein or carbohydrate) that phages bind to. For example, if a phage binds to host receptor that is protein in nature, the end of the tail fibers will be sharp or spiked in form. The *B. subtilis* SPP1 phage, *B. anthracis* γ phage and c2 - type phages that infect *L. lactis*, are good examples of siphoviruses that have their RBPs as tail spikes that bind to protein receptors on bacterial surfaces (Mahony *et al.*, 2016b). In contrast, the phages that attach to a carbohydrate constituent form a larger baseplate where binding occurs. For example, P2 and TP901-1 that infect *L. lactis*, are known to bind to a carbohydrate on the surface of the host (Mahony *et al.*, 2012, Dowah *et al.*, 2018).

1.3.6 Two strategies found within siphoviruses binding

A comparison of structural studies of these phages has highlighted differences in phage infection mechanisms. *L. lactis* siphoviruses P2 and TP901 have been most extensively studied, both structurally and morphologically (Bebeacua *et al.*, 2013, Veessler *et al.*, 2012). Data obtained from X-ray crystallography and Electron microscopy analysis revealed that they undergo two different mechanisms to attach to their bacterial host. First, the RBPs from P2 phage are oriented towards the capsid. These RBPs then undergo Ca²⁺-mediated conformational changes and effectively infect the bacteria (Spinelli *et al.*, 2014b). Some phages on the other hand, have simple attachment mechanisms, for example phage TP901 has its RBPs displayed in the head domain and they are pointing toward the host, and therefore is ready to infect form without needing to a conformational change or Ca²⁺ for activation (Veessler *et al.*, 2012, Dowah *et al.*, 2018).

1.3.7 RBPs for myoviruses of Gram- positive bacteria

Fewer myovirus RBPs have been identified and characterised for myoviruses of Gram-positive bacteria compared to siphoviruses (Spinelli *et al.*, 2014b). Indeed, the phage A511 that infects *L. monocytogenes* is one of the very few myoviruses whose RBPs has been identified (Habann *et al.*, 2014).

A511 belongs to a group of SPO1 related phages and is strictly lytic with low G+C content, like all phages within this group (Klumpp *et al.*, 2008). When A511 is compared to other SPO1-related phages, such as *Bacillus* phage SPO1, and the *Staphylococcus* phages Twort and ISP, the genes located between the *tmp* gene and Helicase are responsible for encoding the baseplate proteins that contains RBPs. Taking this into account, five gene products (*gp98*, *gp99*, *gp104*, *gp106* and *gp108*) of phage A511 were identified between TMP and Helicase. They are thought to be involved in baseplate formation. Polyclonal antibodies were raised against those proteins and the results showed that two of these proteins (Gp98 and Gp108) are indeed involved in phage

attachment (Habann *et al.*, 2014). Furthermore, the baseplate is a big complex device attached to long and short tail fibers, and during the infection, the baseplate undergoes conformational change and rearranges into a double-ringed shape, accompanied by the contraction of the phage tail (Habann *et al.*, 2014, Dowah *et al.*, 2018).

1.4 Bacteriophages that infect *C. difficile*

While multiple phages that infect *C. difficile* have been isolated (Shan *et al.*, 2012, Nale *et al.*, 2016), the receptors of the phages that infect *C. difficile* have not to date been characterised. However, the S-Layer proteins that surround the entire cell or associated proteins are likely to be the phage receptors for *C. difficile*. S-Layer proteins are defined as two-dimensional (2D) crystalline arrays (Fagan *et al.*, 2014). As stated above *C. difficile*, S-layer proteins are composed of two main proteins, which are derived from S-layer protein A (Slp A). They are the low molecular weight S-layer protein (35 kDa) and high molecular weight S-layer protein (45 kDa). The low molecular weight protein of S-layer is exterior to the high molecular weight S-layer protein and so is more exposed to the external environment. In addition, the low molecular weight S-layer protein differs extensively from one strain to other (Fagan *et al.*, 2014). All these factors indicate this protein to be a potential candidate for phage receptors for *C. difficile*. Other possible structures of the *C. difficile* cell wall that may act as phage binding site are teichoic acids or any of the saccharides on the peptidoglycan layer of Gram-positive bacteria.

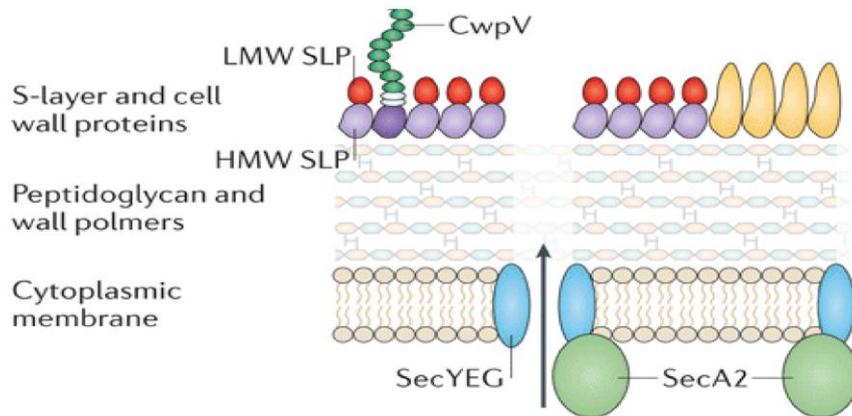


Figure 1-8 Diagram of the cell wall structure of *C. difficile*

The first layer of *C. difficile* cell wall is the surface layer protein(S-layer) that composed of: the HMW SLP (light purple), the LMW SLP (red). Then the peptidoglycan layer and cytoplasmic membrane. SecYEG are a membrane channel and SecA2 is protein translocase subunit. the image was adapted from (Fagan *et al.*, 2014).

To date the receptors binding proteins for phage that infect *C. difficile* have not been identified yet. This is the first study that covers the characterization of RBPs for *C. difficile* bacteriophage. Two phages that were isolated in our lab are involved in this project as shown in Figure (1-7), the first one was PhiCDHS1, and it belongs to the *Siphoviridae* family. The second was PhiCDMH1 which is a myovirus and it infects CD105HE1. Furthermore, the two phages used in this study efficiently infect a wide range of *C. difficile* ribotypes that are clinically of significant importance such as epidemic 027 ribotype (strain R20291 is used in this project and frequently used as example for 027 ribotype) which infected by CDHS1 (Nale *et al.*, 2016).

General Introduction

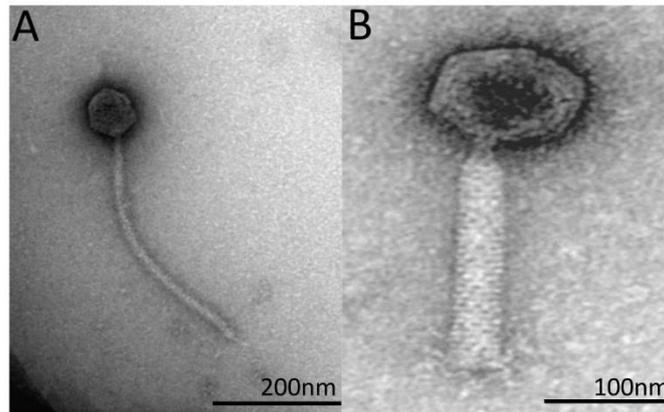


Figure 1-9: TEM image represent the overall structure for the two phages used

A is the siphovirus CDHS1, scale bar 200nm, B is a Myoviruses CDMH1 scale bar 100nm.

1.5 Project Aims

Phage bacterial interaction is one of the key aspects to understanding phage biology, prior to exploiting the phage or its products for any therapeutic or diagnostic applications. Whilst, there has been some extensive studies covering the isolation and characterizations of phages that infect *C. difficile*, the interaction between *C. difficile* and its phages has not been studied yet. Therefore, the overall aim of this PhD project was to identify the receptor binding proteins (RBPs) for two phages (CDHS1& CDMH1) that infect *C. difficile* strains. RBPs is a very important factor for phage specificity, therefore, identification of RBPs for *C. difficile* phages will help us understand the mechanism by which phages attach to *C. difficile*. Furthermore, RBPs can be used in different phage applications such as diagnostic and therapeutic purposes.

The approaches that has been followed to obtain the aim of this project are as below;

- 1- Prediction of the genes encoding tail proteins that involve in the baseplate structure (these proteins will be known as tail proteins throughout the project) for CDHS1& CDMH1 phages.
 - a- Using HHpred software
 - b- Phyer 2 software
- 2- Amplification of the phage genes that bioinformatically predicted to be encoding the tail proteins for the two phages CDHS1& CDMH1.
- 3- Cloning of the phage genes predicted to be encoding the tail proteins for CDHS1& CDMH1 phages.
- 4- Protein expression and purification for the tail proteins for CDHS1& CDMH1 phages.
- 5- Immunogold transmission electron microscopy for the phages CDHS1 & CDMH1 tail proteins.

6- Receptors binding protein for the phages CDHS1 & CDMH1 crystallization trial

Structure of the results chapters

Below is a summary to each chapter: -

Chapter2: *in silico* analysis to identify putative functions of the baseplate proteins and RBP candidates of phages CDHS1 and CDMH1 that infect *C. difficile*

The aim in this chapter is to identify a homology of the proteins Gp18, Gp19, Gp21 and Gp22 from CDHS1 phage and Gp23, Gp24, Gp25, Gp26, Gp27, Gp28, Gp29 and Gp30 from CDMH1 phage that are targeted in this project. This is to determine the putative function of each protein targeted.

Aim of this study will be obtained using

- c- HHpred software
- d- Phyler 2 Software
- e- Blastp search

Chapter 3: - Cloning, expression and purification of the tail proteins for CDHS1&CDMH1 phages

Aim of this study is to clone the genes predicted to be encoding tail proteins for CDHS1& CDMH1 phages. And the resulted construct will be transfer to expression vectors and then overexpressed and purified the protein of interest.

To obtain the aim in this chapter: -

- 1- PCR based assay will be conducted to amplify the desired genes (*gp18*, *gp19*, *gp21* and *gp22*) from phage CDHS1 and *gp29* and *gp30* of phage CDMH1.
- 2- The amplified genes will be sent for cloning in Protein Expression Laboratory (PROTEX) in biochemistry department at University of Leicester.
- 3- The construct resulting from the cloning will be sequenced to confirm the inserts.
- 4- Protein expression assay will be performed on the resultant construct.

- 5- Finally, the overexpressed proteins will be purified, and mass spectrometry will be conducted to confirm the desired proteins.

Chapter4: - Identifying the Receptor binding proteins for CDHS1 and CDMH1

In this chapter, the aim is to determine which of the purified tail proteins of CDHS1 and CDMH1 act as RBPs and therefore involved in phage binding with *C. difficile*.

The approaches that are going to be carried out are

1. The purified Gp18, Gp21 and Gp22 proteins of CDHS1, in addition to Gp29 and Gp30 of CDMH1 will be sent to Eurogentec Company (Brussels, Belgium) to produced poly clonal antibody against these proteins.
2. In the purpose of determination which of these proteins act as RBPs. The produced antibodies will be used in neutralization assay to see which antibody is able to neutralize the phage infection.
3. The antibodies will be used to identify the specific location of these proteins on the phage particle. Using Transmission electron microscopy (TEM).

Chapter 5: - Structural characterisation of Gp22 protein

The aim in this chapter is to determine the protein structure of Gp22 protein, the potential RBP of phage CDHS1: -

To do that

- 1- Gp22 proteins will be cloned, expressed and purified.
- 2- Selenomethionine – Gp22, a modified version of Gp 22 will be expressed and purified
- 3- The purified Gp22 and Selenomethionine – Gp22 will be concentrated and used in sitting-drop vapour diffusion method to crystallize Gp22 protein. For this purpose, different commercial crystallization screens were used such as PACT, JCSG, and Proplex (Molecular Dimensions, UK).

in silico analysis to identify putative functions of the baseplate proteins and RBP candidates of phages CDHS1 and CDMH1 that infect C. difficile

Chapter 2 *in silico* analysis to identify putative functions of the baseplate proteins and RBP candidates of phages CDHS1 and CDMH1 that infect *C. difficile*

2.1 Introduction

Studying and understanding the early steps of phage infection, more specifically the mechanisms by which phages attach to their bacterial hosts, is of fundamental precondition to exploiting phages both as potential antimicrobials and for diagnostics purposes (Habann *et al.*, 2014). Generally, phage attachment to bacteria consists of a multi-step process. Proteins at the end of the phage distal tail known as Receptors binding proteins (RBPs) are generally responsible for the binding process (Bertozzi Silva *et al.*, 2016). RBPs come in the form of tail spikes or tail fiber proteins. They are usually anchored to the big structure of the phage known as the baseplate.

The baseplate structure is very complex and varies with phage morphology. This multiprotein molecular machine is responsible for various functions, such as host recognition, binding and injection of the phage genetic material into the bacterial host (Kostyuchenko *et al.*, 2003). Several different phage proteins are involved in the formation of this highly complex structure, for example in myoviruses such as the *E.coli* phage T4, there are 15 different proteins involved in the T4 phage baseplate structure (Yap *et al.*, 2016). However, in siphoviruses such as the *Lactococcus* phage TP901-1, at least four proteins form its baseplate (Bebeacua *et al.*, 2010).

Amongst the proteins involved in the phage baseplate structure are the RBPs or the tail fiber proteins responsible for phage attachment to bacteria; these are the main focus for this project. Therefore, the proteins involved in the structure formation of the baseplates from the phages (CDMH1 and CDHS1) used in this project will be analyzed. In this chapter, the aim is to determine the putative function of the phage proteins involved in the baseplate structure, and to identify the possible RBPs candidates of the two phages used in this project, namely CDMH1 and CDHS1, using bioinformatics tools.

2.2 Aim of this study

As stated in the introduction (Chapter 1), the genes located between the gene *tmp* (encoding tape measure protein Tmp) and the genes encoding the holin and endolysin are always those that encode the baseplate proteins (Bielmann *et al.*, 2015). Among the genes that located between the *tmp* and genes encoding holin and endolysin are the genes encoding Gp18, Gp19, Gp21 and Gp22 proteins of CDHS1, However for CDMH1 they are Gp23, Gp24, Gp25, Gp26, Gp27, Gp28, Gp29 and Gp30, as shown in Figure 1.10. Those proteins were targeted in this chapter. Therefore the aim in this chapter was to conduct an in silico analysis was carried out to determine the putative functions of those putative proteins involved in the baseplate structure of CDHS1 and CDMH1.

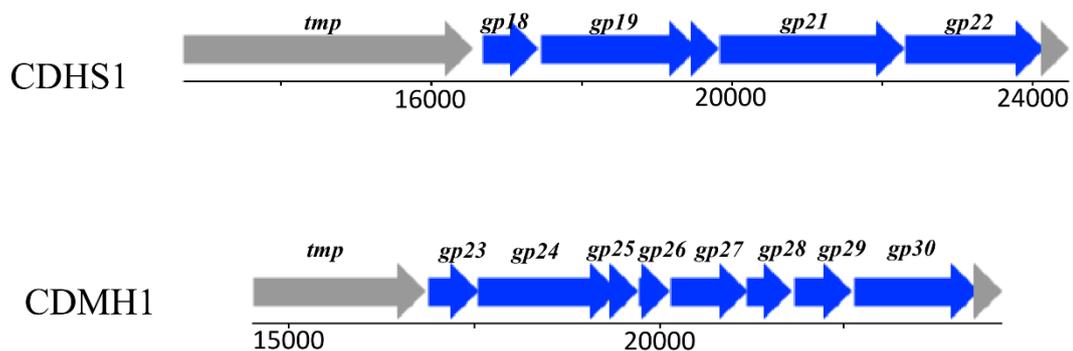


Figure 2-1: Schematic representation of the baseplate genes for CDHS1 and CDMH1 phages

The Genes highlighted with blue in this figure are the genes encoding the proteins that are involved in the baseplate structure of the two phages used (CDHS1 and CDMH1).

in silico analysis to identify putative functions of the baseplate proteins and RBP candidates of phages CDHS1 and CDMH1 that infect C. difficile

The analysis was conducted using the following bioinformatics tools:

- 1- HHpred: - this software is used to detect the protein homology and to predict the structure of the proteins. It is based on the pair wise comparison of profile hidden Markov models (HMMs). It allows to search a wide choice of databases, such as the PDB, SCOP, Pfam, and SMART. This tool was used to identify the homology of the proteins that are targeted in this project. So that assist in identifying the putative function for the proteins targeted.

<https://toolkit.tuebingen.mpg.de/#/tools/hhpred>

- 2- Protein Homology/Analogy Recognition Engine (PHYRE2): - this software is similar to HHpred in terms of function. As it also helps in protein homology prediction. <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>

- 3- Blastp search: this tool is used to compare the sequence of the target proteins with the proteins in the database.

<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>

2.3 Results

2.3.1 *in silico* analysis of the putative tail proteins that involved in baseplate structure of CDHS1 phage

With regards to phage CDHS1, the genes that are located downstream of the *tmp* gene encoding Tmp and upstream the genes encoding the holin are for proteins Gp18, Gp19, Gp21 and Gp22.

2.3.1.1 Gp18 protein

The gene encoding the Gp18 protein is located directly downstream of the gene encoding the tape measure protein (Tmp) of phage CDHS1. HHpred and Phyler 2 analysis of Gp18 protein shows that this protein has a strong structural homology to the Gp19.1 protein from the *B. Subtilis* SPP1 phage as shown in Figure 2-1 and the protein ORF46 of the *L. lactis* phage TP901-1 with 100% probability. The proteins Gp19.1 and ORF46 are known as the distal tail proteins (Dit). Therefore, the Gp18 of CDHS1 is predicted to be the Dit protein for this phage.

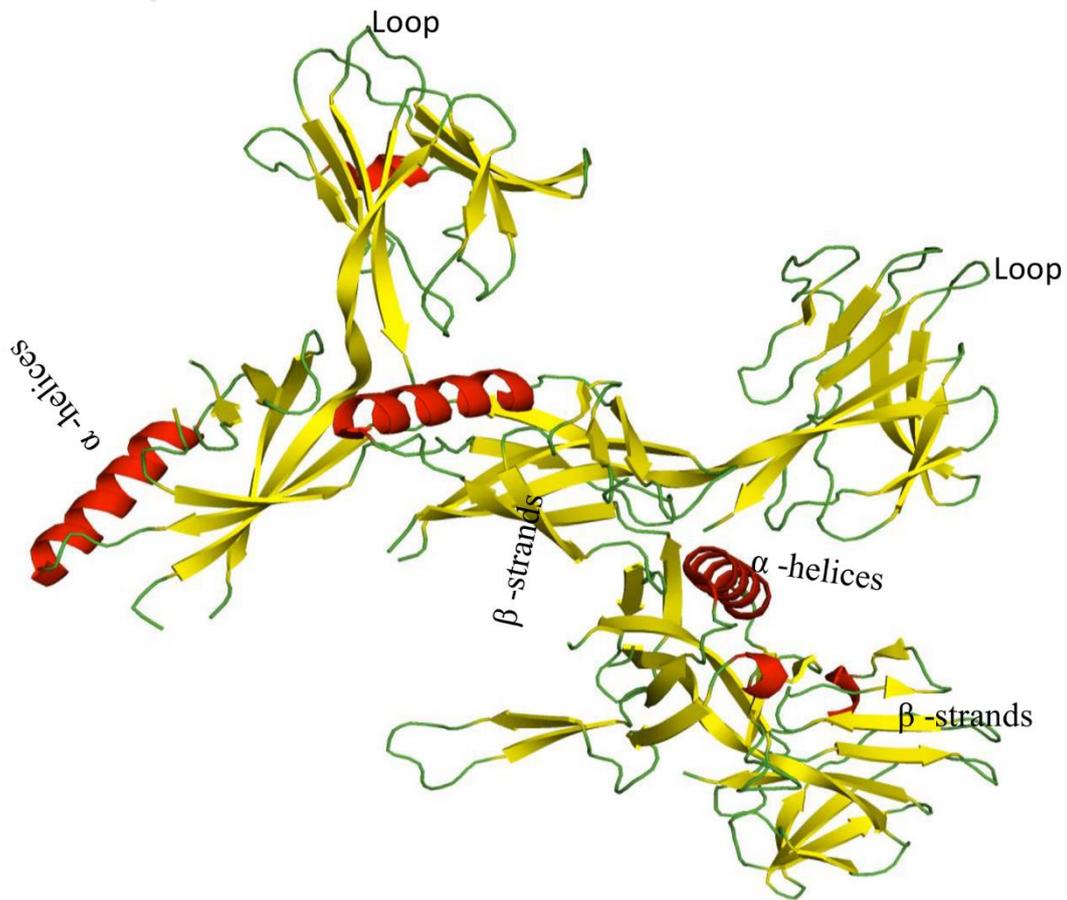


Figure 2-2: *in silico* structural analysis of the Gp18

The Figure represents the PDB entry (2x8k) of the Dit protein structure of Gp 19.1 protein from SPP1 *B. subtilis* phage. The Ribbon structure illustrates α -helices in red, β -strands in yellow and the loops in green. HHpred and phyler2 analysis show that the Gp18 protein has strong structural homology with a probability of 100% to the Dit protein in phages SPP1 of *B. subtilis* and TP901-1 of *L. lactis*.

2.3.1.2 Gp19 protein

The *in-silico* analysis of the Gp19 protein was performed using the same software used for the analysis of Gp18. It revealed that Gp19 has structural homology to the Gp18 protein shown in Figure (2-2) from phage A118 which infects *L. monocytogenes*, also to the protein Gp44 of the Mu phage that infects wide range of enteric bacteria, with a probability of 100% and 97.8% respectively. The proteins Gp44 and Gp18 belong to a group of proteins known as the Tal-like proteins; their action is to degrade the peptidoglycan on the surface of bacteria. In addition to this, a blastp search revealed that the Gp19 protein is a phage tail endopeptidase similar to phages which infect *C. difficile* such as phiCD38-2, phiCD146 and phiCD111 with 99% identity. Gp19 protein is conserved between these phages.

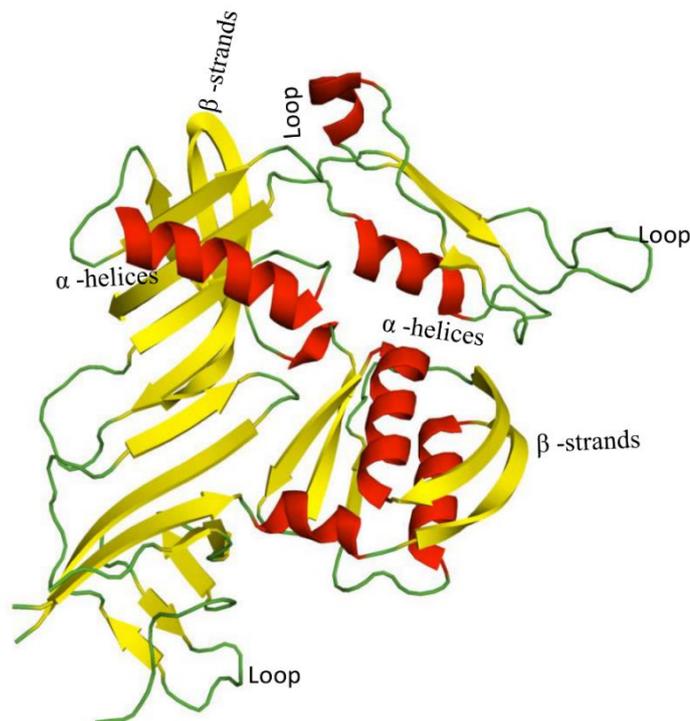


Figure 2-3 : *in silico* structural analysis of Gp19

This Figure represents the PDB entry (3GS9) of the Tal-like protein Gp18 structure of the *L. monocytogenes* phage. Gp19 protein from CDHS1 has strong homology to Gp18 of the *L. monocytogenes* phage with a probability of 100%. The ribbon structure illustrates α -helices in red, β -strands in yellow and the loops in green.

2.3.1.3 Gp21 protein

The analysis of the Gp21 protein revealed that the N-terminal of Gp21 protein harbours strong structural homology with a probability of 99% to the N-terminal of the ORF48 (shown in Figure (2-3)) the upper baseplate protein (BppU) of the TP901-1 phage that infects *L. lactis*. The result also shows that the C. terminal of Gp21 protein has no homology in the database. The Blastp search of this protein identified the Gp21 protein to be the tail fiber protein of many different phages that infect *C. difficile*, such as phage phiCD38-2 with 99% identity and phage phiCD146 with 97% identity. A small level of variability was found at the N- terminal between these phages.

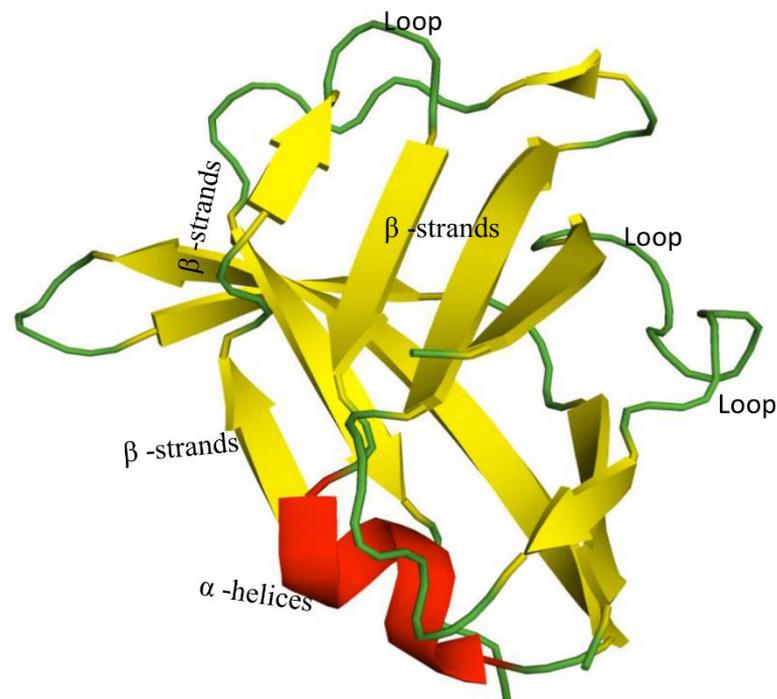


Figure 2-4: *in silico* structural analysis of Gp21

Figure represents the PDB entry (3UH8) for the structure of the N-terminal domain of protein ORF48 of phage TP901-1. The Gp21 protein has strong homology to the ORF48 N-terminal domain of phage TP901-1.

2.3.1.4 Gp22 protein

The *in-silico* analysis of the Gp22 protein shows no homology with any of the other proteins in the database. However, the blast search revealed that Gp22 is a putative tail fiber protein as found in many phages such as phiCD111, phiCD38-2 and phiCD146 that infect *C. difficile*. The blastp search also shows that this Gp22 protein is conserved in CDHS1, phiCD38-2 and phiCD146. However, for phiCD111, Gp22 is conserved at the N-terminal end only. On the other hand, at the C-terminal domain, Gp22 is variable between the two phages, CDHS1 and CD111. This variability at the C-terminal end of this protein was seen in several hits obtained from the analysis.

2.3.2 *in silico* analysis of the putative tail proteins of CDMH1

The tail proteins of the CDMH1 phage were analysed in a similar manner to the analysis carried out for the CDHS1 proteins. Eight proteins (Gp23, Gp24, Gp25, Gp26, Gp27, Gp28, Gp29 and Gp30) were targeted. These proteins are involved in the baseplate formation of phage CDMH1; more importantly, one or two of these proteins may represent the RBPs for this phage.

2.3.2.1 Gp23 protein

The HHpred analysis of CDMH1 Gp23 revealed that this protein has structural homology to the LysM domain presented in Figure (2-4) from the rice blast fungus *Magnaporthe oryzae* protein with a probability of 96%. The blastp search shows that Gp23 also has peptidoglycan binding domain proteins at the C-terminal similar to that of the LysM domain.

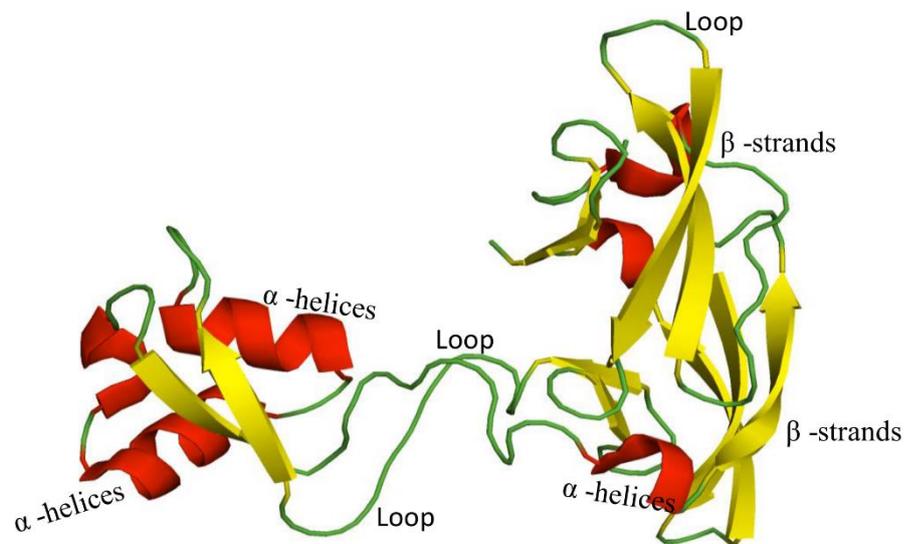


Figure 2-5: *in silico* analysis of gp23 of CDMH1

The Figure shows the BDP entry (2L9Y) of the LysM domain from the rice blast fungus *Magnaporthe oryzae* protein structure. The Gp23 protein of CDMH1 harbours strong homology to the LysM domain from the protein mentioned above, with a probability of 96%.

2.3.2.2 Gp24 protein of CDMH1

The HHpred analysis of the Gp24 protein showed that the N-terminus has a strong structural homology to the 43 kDa tail protein of the *Shewanella oneidensis* prophage MuSo2, with a probability of 99.82%. Moreover, Gp24 has strong homology to Gp44 shown in Figure (2-5) of the phage Mu that infects Enterobacteria (with a probability of 99.82%). Finally, the Gp24 protein is a homologues to the tail-associated lysozyme (Gp27) from the T4 phage with a probability of 83.72%. Moreover, the blastp search revealed the fact that the Gp24 protein has a conserved cell wall hydrolase at the C-terminal of the protein.

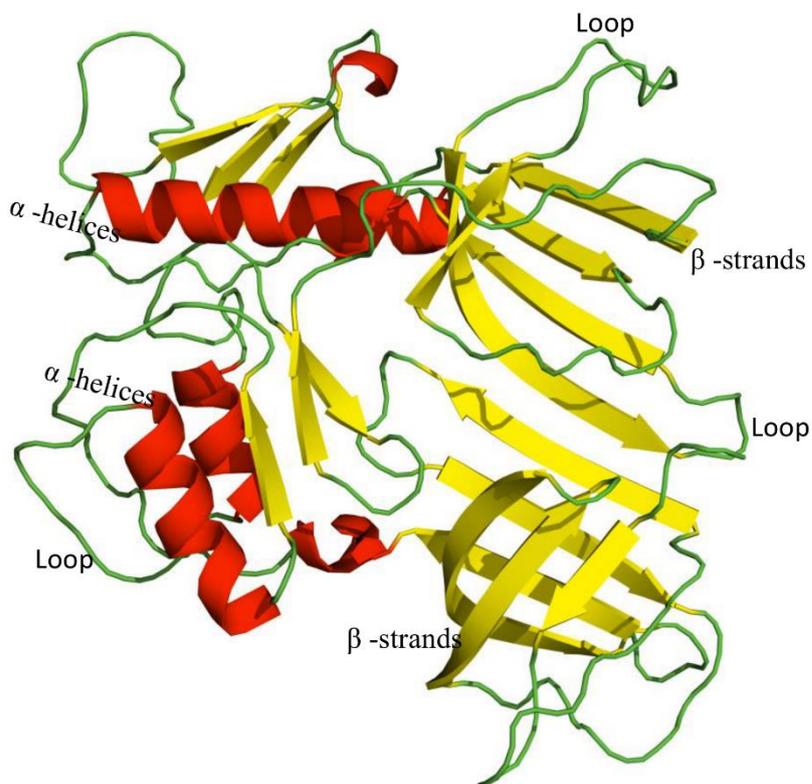


Figure 2-6: *in silico* analysis of Gp24 from CDMH1

The Figure illustrates the BDP entry (1WRU) of Gp44 from phage Mu that infects Enterobacteria. Gp24 protein of CDMH1 harbours strong homology to the protein represented in the Figure, with a probability of 99%.

2.3.2.3 Gp25 and Gp26 proteins of CDMH1

The analysis of both proteins shows that Gp25 has homology to the baseplate protein Gp41 shown in Figure (2-6A) of the phage SN that infects *Pseudomonas*. However, the HHpred analysis and blast search for Gp26 revealed that this protein is homologous to the baseplate wedge protein Gp25 presented in Figure (2-6B) of the T4 phage with a probability of 99.3%. The Gp25 protein acts as the tail initiation sheath. In addition, this protein has lysozyme activity.

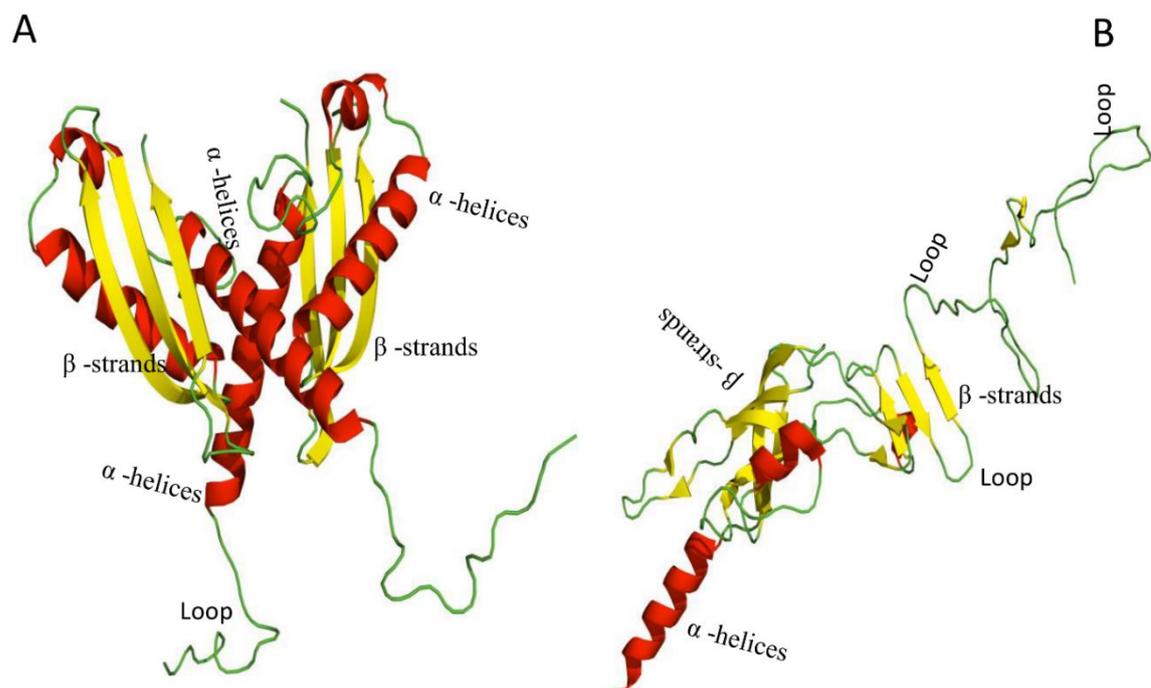


Figure 2-7: *in silico* analysis of Gp25 and Gp26 of CDMH1

Proteins Gp25 and Gp26 have homology to Gp41 of the SN phage that infects *Pseudomonas* represented in Figure A (BDP entry 4RU3) and to Gp5 protein of the T4 phage represented in Figure B (BDP entry 5IW9).

2.3.2.4 GP27 and gp28 proteins of CDMH1

in silico analysis of Gp27 and Gp28 revealed that the two proteins are homologous to Gp6 protein and Gp7 of T4 phages with a probability of 100% and 80% respectively. The Gp6 and Gp7 proteins are components of the T4 baseplate.

2.3.2.5 Gp29 and Gp30 proteins of CDMH1

The analysis for Gp29 and Gp30 show that they have no homology to any of the known phage proteins in the database.

2.4 Discussion

The purpose of this project was to identify the RBPs for the two phages CDHS-1 and CDMH1 that infect different strains (CD105LC1 and CD105HE1 respectively) of *C. difficile*. To obtain this aim, this chapter describes the predictions of the putative function for the proteins involved in the baseplate formation, and the possible candidates for the RBPs of the two phages targeted in this study (CDHS-1 and CDMH1).

2.4.1 *in silico* analysis of CDHS1 baseplate proteins

In the case of the Siphovirus CDHS1 phage, it was shown that the genes encoding proteins involved in the phage baseplate formation are located between Tmp protein and holin proteins (Bielmann *et al.*, 2015). Taking this fact into account, the four proteins of phage CDHS1, Gp18, Gp19, Gp21 and Gp22, were selected as candidates for analysis.

2.4.1.1 Gp18 protein

Gp18 protein from phage CDHS1 was found to have a strong structural homology to the Gp19.1 protein of the *B. Subtilis* SPP1 phage and ORF46 of *L. lactis* phage TP901-1. The two proteins Gp19.1 and ORF46 from the two phages mentioned above are known as distal tail proteins (Dit). The Dit proteins are also found in several siphoviruses that have been characterized such as the *L. lactis* Tuc2009 and P2 phages, phages A118 and P35 that infect *L. monocytogenes*, and the $\phi 11$ phage that infects *S. aureus*. This indicates that the Dit protein is conserved between these siphoviruses (Bielmann *et al.*, 2015, Li *et al.*, 2016b).

Dit proteins from siphoviruses that have been studied represent the central hub of the baseplate (Veesler *et al.*, 2012). Also Dit protein is found to be conserved between these siphoviruses as well as the architecture of the gene encoding this protein (Bielmann *et al.*, 2015). It can then be concluded that phage CDHS1's Gp18 most likely is the Dit for this phage. Therefore, Gp18 protein most likely is responsible for forming of the central hub of the phage CDHS1 baseplate. However, it is worth noting that the amino acid

similarity between Gp18 from CDHS1 and the Dit proteins from other phages is very low.

2.4.1.2 The Gp19 protein of the CDHS1 phage

The Gp19 protein has a strong homology to the Gp18 of A118 phage that infects *L. monocytogenes*, and to Gp44 of the Mu phage that infects Enterobacteria; these two proteins belong to a group of proteins known as Tail associated lysin (Tal) like proteins. The main function of this group of proteins is to help phages inject their genetic material into the bacteria by degrading the peptidoglycan layer (Bielmann *et al.*, 2015). The location of the gene encoding Tal protein in other siphoviruses is located directly downstream to the gene encoding Dit protein. Moreover, it has been reported that the Tal gene location is conserved between several siphoviruses (Bielmann *et al.*, 2015). Given all the above, it is possible that CDHS1's Gp19 belongs to the Tal-like protein group too, and therefore represents the Tal protein for this phage. As mentioned above, Tal proteins degrade the peptidoglycan and assist the phage in infecting bacteria. This opens up the promising possibility of using Tal-like proteins as a tool to treat bacteria (Drulis-Kawa *et al.*, 2015).

The phage enzymes that have potential as therapeutic tools can be classed into two groups of enzymes, hydrolases and lyases, based on their action of carbohydrate degradation. The first category is the one which is of interest here. This group is divided into four protein groups; lysozymes, lytic transglycosylases, glucosaminidases and endopeptidases (Drulis-Kawa *et al.*, 2015). A blastp search for the Gp19 of CDHS1 shows that this protein harbors an endopeptidase domain, which led to the conclusion that Gp19 may belong to a group of endopeptidase proteins (Tal like proteins). Moreover, the protein Tal2009 from the Tuc2009 phage that infects *L. lactis* also belongs to the group of endopeptidase proteins (Drulis-Kawa *et al.*, 2015). The N-terminal of the Tal protein, the

C-terminal of TMP and the Dit protein all together form the so-called initiator complex, which represents the starting point of the phage tail formation (Mahony *et al.*, 2016a).

2.4.1.3 Gp21 protein of CDHS1

Gp21 has strong structural similarity to the N-terminal of the ORF48 protein of phage TP901-1 of *L. lactis*; this protein is known as the Upper baseplate protein (BppU) for the TP901-1 phage. A homologue of BppU has been found in several siphoviruses, such as phage A118 that infects *L. monocytogenes*, and the ϕ 11 phage that infects *S. aureus* (Bielmann *et al.*, 2015, Li *et al.*, 2016b), which again indicates that the level of conservation extended to this protein as stated above in regarding to Dit and Tmp proteins. It has been reported that in phage TP901-1, 18 copies of BppU are arranged around the centre of the Dit protein and connect directly with the receptor binding proteins (RBP) of phage TP901-1 and is required for the RBP fixation step (Sciara *et al.*, 2008, Veessler *et al.*, 2012).

2.4.1.4 Gp22 protein of CDHS1

The *in silico* analysis for this protein revealed that Gp22 has no homology to other proteins in the database. The Gp22 protein is of special interest as it might be the putative RBP for this phage given its location, which is at the end of the phage tail. Additionally, RBPs for phages are the least conserved proteins in the phage baseplate (Habann *et al.*, 2014). Gp22 has no homology to any of the proteins of known function. Moreover, RBPs for related phages are usually present a conserved manner at the N-terminal level. Taking into account the results from the *in silico* analysis and the putative function of the other proteins of phage CDHS1, these being Gp18 (Dit), Gp19 (Tal) and Gp21 (BppU), in addition to this, the Blastp search of Gp22 protein shows that this protein has a highly variable region at C-terminus. All this renders the Gp22 protein as a possible RBP for the CDHS1 phage. This will be further discussed and confirmed in chapter 4.

2.4.2 in silico analysis of the CDMH1 baseplate proteins

Eight proteins were targeted to be analysed as possible CDMH1 baseplate proteins. These are Gp23, Gp24, Gp25, Gp26, Gp27, Gp28, Gp29 and Gp30, and are located between the Tmp protein and the holin and endolysin too. The *in-silico* analysis for these proteins will be discussed in detail below.

2.4.2.1 Gp23 and Gp24 proteins

The analysis of Gp23 revealed that it contains a LysM peptidoglycan-binding domain at the C-terminal. This indicates that this protein may bind to the peptidoglycan of *C.difficile*; however, this has to be confirmed experimentally. A LysM like domain has been found in many other phages; for instance, it has been shown that this domain was found in almost half of the myoviruses that have like T4 phage baseplates (Maxwell *et al.*, 2014). Well-studied phage proteins containing LysM like domain include Gp X from phage P2 and Gp53 of T4 phages that infect *E. coli*. Both proteins are located between the tail tube and the baseplate junction. The Gp X of P2 and Gp53 of T4 are involved in phage baseplate structure (Yap *et al.*, 2016, Maxwell *et al.*, 2014). It is worth mentioning that Gp23 has no amino acid sequence similarity with these proteins except at the LysM like domain.

The N-terminus of the Gp24 protein has structural homology to the 43 kDa tail protein prophage MuSo2 from *Shewanella oneidensis*, and Gp44 of phage Mu that infects Enterobacteria. The 43 KDa protein and Gp44 are known structural orthologues of the Gp27 from T4 phage (Habann *et al.*, 2014). Results also show that Gp24 has structural homology to the T4 phage Gp27 protein, with a probability of 83.72%. Moreover, Gp27 of T4 is known to be involved in phage baseplate formation. More specifically, Gp27 forms the central hub of the T4 phage baseplate by making a complex with another baseplate protein, Gp5 in the T4 phage (Leiman *et al.*, 2010).

Interestingly, the Gp24 of CDMH1 has a conserved cell wall hydrolase domain at the C-terminus of the protein. The activity of this domain is to degrade the peptidoglycan layer. Since phages in general require access to the cell wall of the bacteria so that they can inject the genetic material, a protein with cell wall degradation activity is required (Habann *et al.*, 2014). Moreover, cell wall hydrolase domains have been reported to be found in many baseplate proteins of different phages such as Gp98 protein of the *L. monocytogenes* phage A511, ORF53 of the phage K that infects *S. aureus*. The ORF53 has been shown to be potent against *S. aureus* (Habann *et al.*, 2014, Paul *et al.*, 2011). However, no similarity has been found between Gp24, Gp98 and ORF53 in term of amino acid sequence. As result of the all above, it can be deduced that Gp24 from CDMH1 could be the protein assisting the phage to inject its genome into *C.difficile*.

2.4.2.2 Gp25 and Gp26 proteins

The Gp25 protein has structural homology to Gp41 from the *Pseudomonas* phage SN, and although this protein is known as a putative phage baseplate protein, its function is unknown. However, Gp26 from CDMH1 shows strong homology to Gp25 of the T4 phage. Gp25 proteins of T4 phage is highly conserved in the phages that are structurally like T4. Homology of the T4 Gp25 protein was found with the A511 phage that infects *L. monocytogenes*. This protein harbours a LysM motif, which was found in several enzymes that have the function of degrading the cell wall (Leiman *et al.*, 2010, Habann *et al.*, 2014). The role of Gp25 like proteins is the initiation of sheath assembly (Taylor *et al.*, 2016).

2.4.2.3 Gp27 and Gp28 proteins

The Gp27 protein has a strong homology to Gp6 of the T4 phage. Gp6 protein is one of the phage T4 baseplate proteins that have 12 copies and forms a continuous ring. This ring made up most of phage T4's inner baseplate (Taylor *et al.*, 2016). The role of this

ring is to maintain the baseplate integrity during the conformational changes that occurs when the phage infects the host (Aksyuk *et al.*, 2009). Structural homology of Gp6 was found in the A115 phage (Gp103) that infects *L. monocytogenes* (Habann *et al.*, 2014). In addition to this, the blast search of Gp27 of CDMH1 shows strong homology to Gp J of the P2 phage that infects *E.coli*; this protein is involved in the phage baseplate wedge formation (Christie *et al.*, 2016).

In the case of Gp28 of CDMH1, this protein has homology to Gp7 from the T4 phage wedge baseplate; Gp7 is involved in phage wedge baseplate assembly (Taylor *et al.*, 2016). This protein's specific role is to initiate a complex with Gp10 proteins to start forming the wedge, followed by the sequential binding of Gp8 and Gp6. In addition to this, Gp7 acts as a cornerstone for the attachment of other proteins (Yap *et al.*, 2016).

2.4.2.4 Gp29 and Gp30 proteins

These two proteins were of most interest as possible candidates for acting as the CDMH1 RBPs. These two proteins show no homology to any other phage proteins from the database; it has been stated that RBPs from phages are the least conserved proteins in phage particles, therefore this variability reinforces the chance that these two proteins are in fact RBPs (Habann *et al.*, 2014). Gp29 and Gp30 proteins harbour a variable region at the C-terminal. A similar phenomenon occurs where the RBPs from related phages have conservation at the N-terminal region, however, variability at the C-terminal region (Habann *et al.*, 2014). Therefore, it was of interest to investigate whether these two proteins are RBPs for the phage CDMH1.

2.5 Summary of Chapter 2

In this chapter, four proteins from the phage CDHS1 were bioinformatically analysed, and it was concluded that:

- Three of these proteins show strong structural homology to three well-characterised proteins belonging to other siphoviruses; these include:
 - a) Gp18 from CDHS1 is most likely the distal tail protein (Dit) for this phage. This is because it has strong structural homology to Dit from other siphophages such as the TP901-1, Tuc2009 and P2 phages of *L. lactis*, phages A118 and P35 that infect *L. monocytogenes*, and the $\phi 11$ phage that infects *S. aureus*.
 - b) The location of the gene encoding Gp19 of CDHS1 and the *in-silico* analysis of this protein suggests that Gp19 could be the Tail associated lysin (Tal) protein for this phage.
 - c) The N-terminal of Gp21 of CDHS1 has strong homology to the N-terminal of the ORF48, the upper baseplate of phage TP901-1 of *L. lactis*.
 - d) Gp22 from CDHS1 is most likely to be the RBPs for this phage, as this protein shows no homology to any other phage proteins. Gp22 also has a variable region at the C- terminal of this protein.

This chapter also covered the *in-silico* analysis of the putative proteins involved in the phage CDMH1 baseplate formation.

- Eight proteins were analysed (Gp23, Gp24, Gp25, Gp26, Gp27, Gp28, Gp29 and Gp30). The majority of these proteins show structural homology to proteins from the T4 phage, leading to the conclusion that this phage may belong to the group of phages that have structures like the T4 phage. The main findings from the phage CDMH1 analysis are:

in silico analysis to identify putative functions of the baseplate proteins and RBP candidates of phages CDHS1 and CDMH1 that infect *C. difficile*

- a) Gp24 from phage CDMH1 harbours a cell wall hydrolases domain at the C-terminal of the protein that binds to peptidoglycan. In addition to this, this protein has strong homology to the tail associated lysosome of the T4 phage (Gp27), thereby suggesting that Gp24 may bind to the cell wall of *C. difficile*.
- b) As they have no homology to any other phage proteins, Gp29 and Gp30 are the proteins of most interest one of them most likely to be RBPs for this phage.

Chapter 3 Cloning, expression and
purification of the tail proteins for CDHS1 &
CDMH1 phage

3.1 Introduction

The phage bacterial interaction is dictated by the high specificity of the phage receptor binding protein (RBP). The RBPs may attach to either a protein, carbohydrate, or teichoic acid moiety on the surface of the bacterial host. When the RBPs recognizes a receptor that is a conserved protein on the cell surface of a bacterial species, the phage may be able to infect different strains of the same species. Whereas, if the phage recognizes a specific yet variable molecule such as an oligosaccharide on the bacterial host, this means that, such phage will have more strain-specific interactions and therefore a much narrower host range (Mahony *et al.*, 2016a).

Much work has been conducted to identify the RBPs of phages that infect Gram-negative bacteria. However, recently there has been an increased focus in studying the RBPs of phage that infect Gram-positive bacteria (Mahony *et al.*, 2016a). Such studies are important, because characterisation of the RBPs of phages will enhance our understanding of the mechanism by which phage attach to their host. Moreover, this knowledge undoubtedly will have implications on our understanding of host range specificity.

In chapter 2, I described the Bioinformatic tools that were used to predict the function of the tail proteins for two phages that infect *C. difficile*; CDHS1 and CDMH1. Four genes *gp18*, *gp19*, *gp21* and *gp22* of CDHS1 were targeted to be encoding tail proteins, whereas for CDMH1, eight genes (*gp23*, *gp24*, *gp25*, *gp26*, *gp27*, *gp28*, *gp29* and *gp30*) have been predicted to be encoding tail proteins. However two genes *gp29* and *gp30* were chosen as possible candidate as the RBPs of this phage, due to the lack of protein homology to the proteins that were encoded by these two genes. In this chapter the process of amplification, cloning, and subsequent production and purification of these predicted tail proteins is detailed.

3.2 Aim of this study

The aim of this study was to over express and purify the tail proteins of CDHS1 encoded by *gp18*, *gp19*, *gp21* and *gp22* and the tail proteins of CDMH1 encoded by *gp29* and *gp30* respectively.

To obtain this purpose several approaches were carried out:

- 1- PCR was conducted to amplify the desired genes: *gp18*, *gp19*, *gp21* and *gp22* for phage CDHS1 and *gp29* and *gp30* of phage CDMH1.
- 2- The amplified genes were sent for cloning by the Protein Expression Laboratory (PROTEX) in the biochemistry department at University of Leicester.
- 3- The cloned genes were sequenced to confirm the inserts.
- 4- Proteins were expressed in *E. coli*.
- 5- Finally, the overexpressed proteins were purified, and mass spectrometry was conducted to confirm their identities.

3.3 Method

3.3.1 Amplification of the phage proteins that were bioinformatically predicted to be tail proteins for CDHS1 and CDMH1 phages

The *gp18*, *gp19*, *gp21* and *gp22* were predicted to be encoding tail proteins in phage CDHS1. Whereas eight genes were predicted to be encoding tail proteins in CDMH1. Two of these genes (*gp29* and *gp30*) were used in this project. Genes were amplified and cloned using the methods outlined below.

3.3.1.1 Phage DNA extraction

To achieve phage DNA extraction, 10 µl of DNase (30 mg/ml) (BBI enzymes) and 2 µl RNase (100 mg/ml) (BBI enzymes) was added to 1 ml of 10¹⁰ of phage stock and incubated overnight at 37°C. After incubation, the mixture was first mixed gently with 12.5 µl 1 M MgCl₂, thereafter vigorously mixed with 40 µl of 0.5 M EDTA, 5 µL of proteinase K (10 mg/ml) and 50 µL of 10% SDS. This whole mixture was then incubated for 1 hour at 55°C, during the incubation period the sample was vortexed twice. Thereafter, the sample was transferred in to 2 ml centrifuge tubes, each tube receives 500 µl of the sample and equal amounts of phenol: chloroform: isoamyl-alcohol (25:24:1) was added. The tube was mixed well and centrifuged at 4 °C for 15 minutes at 20.000-x g. Phage DNA is in the top aqueous layer after centrifugation; this layer was removed and placed into a new centrifuge tube. Phenol: chloroform: isoamyl-alcohol (25:24:1) was added once again and the centrifugation step was repeated.

In order to precipitate the DNA, isopropanol was added to the sample in a volume double to the volume of the sample, in addition to this, 1/10 volume 3 M sodium acetate was added too. Incubation for 24 hours was performed at 4 °C, post incubation the sample was centrifuged at 20.000 x g for 20 minutes. After which, the supernatant was carefully removed ensuring that the pellet was not disrupted. Then 500 µl of 70% ethanol was added to wash the pellet without dissolving it, and then centrifugation process was

performed for 10 minutes at full speed. Finally, the pellet was left to dry for five minutes, then the pellet was dissolved in 50 µl of 5mM Tris-HCL pH 8-8.5 and stored at -20 °C until used.

3.3.1.2 PCR primers designed

Sets of primers were used to amplify the genes of interest as shown in Table (3-1). They were designed manually with restriction endonuclease sites included according to the PROTEX guidelines to enable cloning. All primers were purchased from IDT Integrated DNA Technologies, UK and the working stocks of the primers used were at a concentration of 10 mM.

3.3.1.3 PCR assay

After obtaining the DNA template, PCR was performed to amplify the genes encoding the putative tail-fiber proteins for phage CDHS1. Three sets of primers were designed and used as shown in Table (3-1). The master mix used for the PCR reactions is shown in table (3-2). The PCR cycles were as follows: initial denaturation step at 98°C for 30 seconds, 30 cycles at 98°C for 10 seconds, annealing at 65°C for 20 seconds, extension at 72°C for 1 minute, and the final extension at 72°C for 2 minutes. The PCR products were then prepared for gel electrophoresis by using 5 µl of 6 x loading dye. Samples were subsequently loaded on to a prepared 1% agarose gel. The gel was run at 100 V for 1 hour. Then the bands corresponding to the size of the genes were excised from the gel. The ISOLATE II PCR and Gel Kit (Bio line) was used to clean up the PCR product, following the manufactures' protocol. Finally, purified PCR products were sent to PROTEX for cloning.

Table 3-1:- Outline of the primers that were designed to amplify the genes that encode the tail proteins for CDHS1 and CDMH1 phages

Name of the primer	Forward primer	Revers primer
Gp18	TACTTCCAATCCATGagagga gggcataaagctatgt	TATCCACCTTTACTGTCAAttataaaatattccatct aggatt
Gp19	TACTTCCAATCCATGaccatggt agagaggaagaaag	TATCCACCTTTACTGTCAAttaagcataaacatagt attgtact
Gp21	gcgcggatccGTGATAAATTTG AGAGATAG	accgggaattcTTAACTCACCTCTTCTTTTAT TTC
Gp22	TACTTCCAATCCATGagttggg cggagacatacaaag	TATCCACCTTTACTGTCAAttaaattgcttgatacat tgcgtaa
Gp29m	TACTTCCAATCCTTGGCTATAGAT AAAAGTTAT	TATCCACCTTTACTGTCACTATATAGGTAACATAT CAT
Gp30m	TACTTCCAATCCATGACTACTGAA TGGAATTTAAT	TATCCACCTTTACTGTCATTAATAAATTTTAATAG CAC

Table 3-2 :- Reagents used for the PCR master-mix to amplify the target tail proteins

PCR Reagent	Volume Added per Sample (µl)
Reaction Buffer (5XQ5)	5
(10µM) Reverse Primer	1.25
(10µM)Forward Primer	1.25
10mM dNTPs	0.5
Q5 High –Fidelity DNA Polymerase	0.25
DNA template(sample)	1.0
Ultra-pure Water	15.75
Total Volume	25

3.3.2 Preparation of BL21 (DE3) and DH5 α *E. coli* competent cells

DH5 α and BL21 (DE3) *E. coli* competent cells were prepared for the cloning process and protein expression using the rubidium chloride method. In brief, a single colony from a Luria Broth (LB) agar plate was inoculated into 5 ml LB broth and kept in a 37 °C shaking incubator overnight. Then 2 ml of starter culture was inoculated into 200 ml of LB and grown at 37 °C with shaking at 200 rpm until the culture reached an OD₆₀₀ of 0.3. Cells were then chilled on ice for 15 minutes. Thereafter the cells were centrifuged at 4500 x g for 15 minutes at 4 °C. After centrifugation, the pellet was re-suspended in 33 ml of pre-chilled Transformation Buffer 1 (TFB1) (1.2 ml of 1 M potassium acetate, 0.06 g CaCl₂, 0.48 g RbCl, 0.4 g MnCl, and 6 ml glycerol made up to 40 ml in H₂O, at a pH of 5.8) at 1/3 volume of the original culture. Then cells were incubated on ice for one hour. Cells were pelleted and resuspended in 12.5 ml of Transformation Buffer 2 (TFB2) (0.5 M MOPS, 0.22 g CaCl₂, 0.02 g RbCl, and 3 ml of glycerol made up to 20 ml in H₂O at pH 6.8). The cells were then incubated on ice for 15 minutes. Finally, the competent cells suspension was dispensed into 100 μ l aliquots, snap-frozen using liquid nitrogen and then stored at -80 °C.

3.3.3 Heat shock transformation

To do transformation 1 μ l of the plasmid (~100 ng) was transformed into the competent cells (BL-21 DE3 *E. coli*), then the cells were incubated for 1 hour on ice, after which the cells were heated for 1 minute at 42 °C, then 300 μ l of LB broth was added to the cells and incubated with shaking at 37 °C for 1 hour. Finally the cells were plated on LB agar supplemented with 0.15 mg/ml Carabancillin and incubated overnight at 37 °C.

3.3.4 Protein expression and purification for the tail proteins Gp18, Gp19, Gp21 and Gp22 of CDHS1 Phage

To overexpress the tail proteins of phage CDHS1, the PCR amplified genes (*gp18*, *gp19*, and *gp22*) were cloned into a pET-based expression plasmids using the cloning service (PROTEX) at the Biochemistry Department at the University of Leicester. The sequences of the resulting constructs were confirmed and the plasmids were transformed in to *E. coli* BL21 (DE3) prepared as in section 3.3.2 for expression. Due to difficulties encountered in cloning the gene encoding Gp21 protein by the PROTEX service, cloning was carried out in our lab using the expression plasmid pGEX-4T-1, as described in more detail below.

3.3.4.1 Expression and purification of Histidine (His) tagged Gp18

The Gp18 protein was produced with an N-terminal 6x histidine tag from expression via the plasmid pLEICS01 as shown in Figure 3-1.

To express the Gp18 protein, cells were grown in 1 L LB culture with shaking at 37 °C until the optical density OD₆₀₀ of the cells was 0.5, then cells were induced using Isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.1 mM and cells were then incubated at 17 °C overnight. The resulting culture was centrifuged at 5000 x g for 20 minutes at 4 °C, the supernatant was removed, and the pellet was stored at -80 °C.

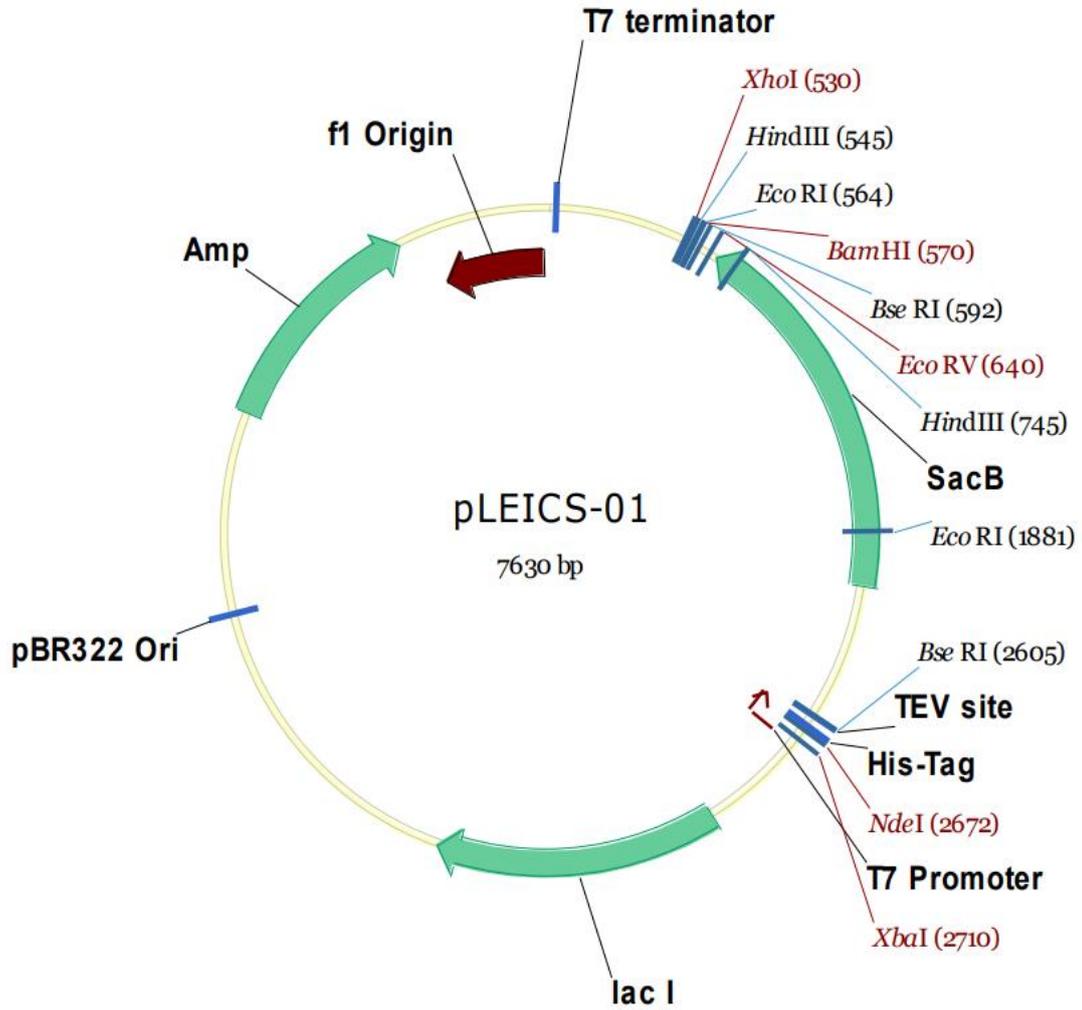


Figure 3-1: Schematic map of PLEICS-01 used to make His tagged Gp18 protein

Cloning, expression and purification of the tail proteins for CDHS1 & CDMH1 phage

The pellet was re-suspended in 50 ml of 25 mM Tris-HCL pH 7.5, 150 mM NaCl (binding buffer), 5 mM of MgCl₂, 50 µl of DNase (BBI Solution) and 1 protease inhibitor tablet (Roche). After the cell pellet was completely suspended, the pellet was subject to sonication, this was carried out using eight pluses, with each pulse lasting for 20 seconds with a one-minute rest in between. Thereafter the mixture was centrifuged at 20,000 x g for 20 minutes at 4°C, upon completion, the protein was purified from the supernatant.

The purification process was carried out at 4°C, where a 1 ml Nickle column was prepared. The Nickle binds to the 6x histidine tagged on the protein of interest. The column was pre-equilibrated using 5 x binding buffer, and then the supernatant of the cell lysate was diluted 1:1 using 25 mM Tris-HCL pH 7.5, 150 mM NaCl and 20 mM imidazole. After which the diluted supernatant was allowed to pass through the column. Then two washing steps were conducted; the first was using 15 ml of 25 mM Tris-HCL pH 7.5 and 1M NaCl, and the second washing step was with 25 mM Tris-HCL pH 7.5, 1 M NaCl and 25 mM Imidazole. The Gp18 protein was eluted using a step gradient of 50 mM, 100 mM, 150 mM, and 250 mM of Imidazole. Collected fractions were screened on a 12% SDS Gel.

3.3.4.2 Poly Acrylamide Gel Electrophoresis SDS- PAGE preparation

Gel casts and electrophoresis equipment (Bio-Rad) were used. SDS-gels were prepared based on the tables 3-3 & 3-4 below. 20 µl of protein samples were mixed with 5 µl of 5x SDS loading dye (250 mM Tris-HCL pH 6.8, 50% v/v glycerol, 10% w/v SDS, 500 mM DTT, 0.25% w/v bromophenol blue) and boiled for 5 minutes at 100 °C before being loaded onto the gel. Gels were immersed in 1x running buffer (25 mM Tris-HCL, 192 mM glycine, 0.1% w/v SDS) at 170V for one hour, after which the electrophoresis was stopped, and the gels were stained using EZ blue stain typically for 1 hour or left overnight to stain . The gels were destained via washing the gels with distilled water with

gently shaking until bands were clear. Then gel images were taken. Tables 3-3 and 3-4 below show the components of the SDS – PAGE gel.

Table 3-4 :- 12%&15%SDS–PAGE resolving gel. Sufficient for two 1.0mm mini-gels

	12%	15%
30% Acrylamide/bis-acrylamide (ml)	4	5
1.5 M Tris-HCL, pH 8.8 (ml)	2.5	2.5
H ₂ O (ml)	3.4	2.4
10% SDS (μl)	100	100
10% APS (μl)	50	50
TEMED (μl)	10	10

Table 3-3 :- 5% SDS –PAGE stacking gel. Sufficient for two 1.0mm mini- gels

	5%
30% Acrylamide/ bis- acrylamide (ml)	833 μl
0.5 M Tris-HCL, pH6.8 (ml)	1.25 ml
H ₂ O	2.87 ml
10% SDS	50 μl
10% APS	25 μl
TEMED	5 μl

3.3.4.3 Gel Filtration chromatography for Gp18

For additional purification of Gp18 protein, the fractions that were obtained from the affinity chromatography purification were pooled together based on the SDS-PAGE gel profile. Then gel filtration chromatography was used in order to obtain highly pure Gp18 protein from those fractions. This method of purification separates samples on the basis of their molecular weight. Gel filtration was performed using a Superdex 75 16/60 column on an AKTA purifier machine. The column was first pre-equilibrated using freshly filtered buffer (120 ml of 20 mM Tris-HCL pH 7.5 and 20 mM NaCl pH 7.5) at a flow rate of 0.5 ml/min. Then, the Ni-purified samples were pooled together, injected in to the AKTA purifier machine and ran at a flow rate of 1 ml/min, 1.5 ml fractions were collected. The fractions that contained the protein of interest were subjected to SDS-PAGE. These fractions were pooled and then concentrated using Amicon Ultra-4 10K Centrifugal Filter Device (Millipore, UK), the concentration of the fractions was determined using a Nano drop spectrophotometer.

3.3.5 Proteins expression of Glutathione S-transferases (GST) tagged Gp19 and Gp22 proteins

The genes encoding Gp19 and Gp22 proteins were fused with N- terminal GST tags by cloning into pLEICS02 as shown in Figure 3-2. Then the resulting constructs were

confirmed by sequencing and expressed using the strategy described above in section 3.2.4.1.

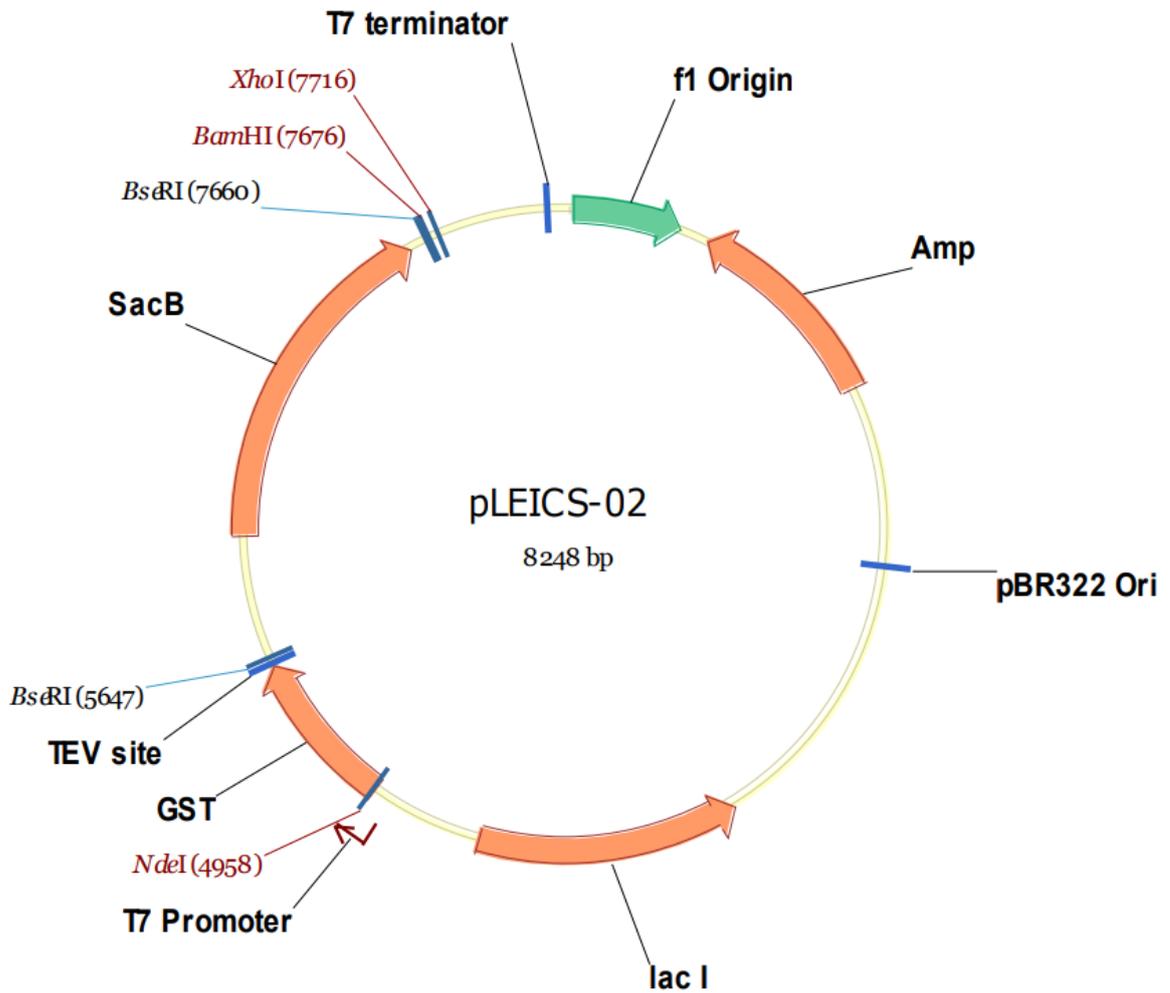


Figure 3-2: Schematic map of PLEICS-02 used to make GST tagged Gp19 and Gp22 proteins

3.3.6 Purification of Gp19 and Gp22 proteins

3.3.6.1 Affinity chromatography purification for Gp19 and Gp22 proteins

To purify Gp19 and Gp22, cell pellets were lysed by sonication (eight pluses lasting 20 seconds with a one minute rest in between pulses) in 50 ml of phosphate buffer solution PBS (binding buffer), 5 mM of MgCl₂, and 50 µl of DNase (BBI Solution) and 1 tablet of protease inhibitor (Roche). Thereafter the mixture was centrifuge at 20,000 x g for 20 minutes at 4°C, after which the proteins were purified from the supernatant.

Affinity chromatography was conducted using 1 ml of glutathione resin (glutathione has the affinity to bind to the GST tag). The prepared column was pre-equilibrated using 5 ml of distilled water and 10 ml of PBS. Then the supernatants of the cell lysates were passed through the column and the proteins were eluted using 50 mM Tris-HCL pH 7.5, 5 mM 1, 4-dithiothreitol (DTT) (Apollo Scientific Limited) and 10 mM reduced glutathione (Duchefa Biochemie). Fractions were screened on a 12% SDS gel prepared as in section 3.2.4.2. After screening, the fractions that contained the protein of interest were digested with TEV protease (Protex) overnight in order to remove the GST tag. Thereafter, gel filtration was conducted to remove the tag along with any impurities or aggregates.

3.3.6.2 Gel Filtration Chromatography for Gp19 and Gp22

Gel filtration was performed as described in 3.2.4.3. However, instead of using Superdex 75 16/60 column, a Superdex 200 16/60 was used to purify Gp19 and Gp22 at a flow rate 1 ml/min, with 1.5 ml fractions were collected. After purification, the fractions containing pure protein based on SDS gels, were pooled, and concentrated using an Amicon Ultra-4 10K Centrifugal Filter Device (Millipore, UK). The final concentration was determining using the Nano drop spectrophotometer.

3.3.7 Cloning and preparation of recombinant Gp 21 proteins

3.3.7.1 Plasmid extraction and purification

PGEX-4T-1 plasmid was used to make Gp21 recombinant protein, the plasmid was extracted by grown *E.coli* (DH5 α) strain in LB media supplemented with Carbenicillin at concentration 0.15 mg/ml. The cells were incubated in 37 °C shaking incubator overnight. Thereafter, the plasmid DNA was extracted using GenElute™ Plasmid Miniprep Kit (Sigma Aldrich, UK), the extraction was conducted according to the manufacturer's instructions. Then plasmid concentration was measured using the Nano drop (Thermo Scientific, USA).

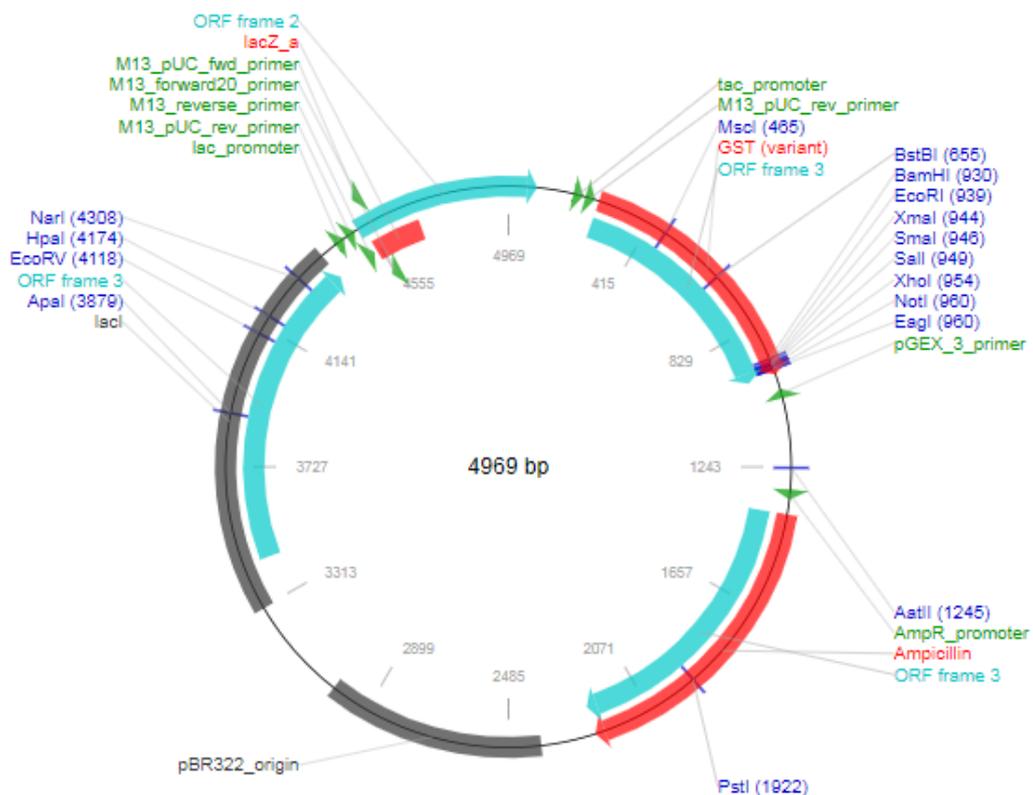


Figure 3-3: Schematic map of PGEX-4T-1 used to make GST tagged Gp21

3.3.7.2 The *gp21* gene amplification and purification

The Gp21 encoding gene was amplified from CDHS1 using the primer gp21F and gp21R Table 3-1 that incorporated a restriction site at the 5' end, the protocol outlined in section 3.2.1.3 was followed with slight a modification in the PCR cycles, the annealing temperature in this reaction was 57.5 °C. After the PCR was complete, the PCR product was purified using The ISOLATE II PCR and Gel Kit (Bio line).and ran on a 1% Agarose gel.

3.3.7.3 Digestion with restricted Endonucleases

After the *gp21* gene been successfully amplified, the *gp21* gene and PGEX-4T-1 plasmid were digested with the restriction endonucleases (BamH1) and (EcoR1). The digestion process was performed according to the manufacturer's instructions (New England Bio labs, UK).

3.3.7.4 Ligation

The *gp21* gene and PGEX-4T-1 plasmid were ligated to each other using a T4 DNA Ligase kit (New England Bio labs, UK). The ligation reaction was conducted at a molecular ratio of 3:1 of the insert to the vector. The reaction consisted of 10 X buffer, T4 DNA ligase (5-10 units for 1 µg of DNA), distilled water, the insert and vector. The reaction mixture had a final volume of 10 µl, and was left to incubate at 16 °C overnight. The PGEX-4T-1 plasmid was used as control. Table 3-5 shows the ligation reaction.

Table 3-5 : -Reagents used for the ligation Reaction

Ligation Reagent	Volume Added per Sample (ul)
T4 DNA ligase	1
10X buffer	1
D.W H ₂ O	2.4
PGEX-4T-1 plasmid	1
Insert	3.4

3.3.7.5 Transformation

To do that 2 µl of the ligation mix and PGEX-4T-1 plasmid were transformed in to DH5α

E. Coli Competent cells using the process mentioned in 3.3.2.

3.3.7.6 Screen transformed colony

Colonies were screened to confirm the insert using, restriction endonucleases digestion and sequencing.

3.3.7.7 Restricted Endonucleases

The predicted positive colonies were inoculated in 5 ml of LB broth supplemented with the relevant appropriate antibiotic incubated overnight in a 37 °C shaking incubator. After which the plasmid was isolated, and digested with the restriction endonucleases BamHI and EcoRI (section 3.2.7.3), and viewed on a 1% agarose gel.

3.3.7.8 Sequencing

The *Gp21* cloned in to PGEX-4T plasmid was sent for sequencing to GATC Biotech. The obtained sequence was aligned against the original sequence for *gp21* to check for homology. The multiple aligned sequence tool Clustal omega was used to for this purpose. <https://www.ebi.ac.uk/Tools/msa/clustalo/>.

3.3.7.9 Protein expression and purification for Gp21 protein

The Gp21 protein was expressed and purified as stated above (3.2.4.1), (3.2.6.1) and (3.2.6.2) with slight modifications. This including the use of 0.5 mM IPTG to induce the expression of the protein, and also using thrombin protease (GE Healthcare, UK) to cleave the GST tag from Gp21 protein (the cleavage process was performed according to the manufacturer's protocol). Briefly, one unit of thrombin protease was used to cleave 100 µg of the GST fusion Gp21 in 1 X PBS and incubated at 22°C for 16 hours.

3.3.8 Protein expression and purification of the tail protein for CDMH1

This assay was carried out in order to produce and purify the predicted tail proteins Gp29 and Gp30 from phage CDMH1. This in order to determine which one of these tail proteins is responsible for binding to *C. difficile*. To amplify the genes of interest (*gp29*, *gp30*), the same approach that was followed for the CDHS1 phage tail proteins (section 3.2.1.3) was applied here. The resulting constructs were confirmed & transformed in to *E. coli* BL21 (DE3) for expression, followed by affinity chromatography purification using GST tags.

3.3.8.1 Protein expression of Gp29 and Gp30

Gp29 and Gp30 were tagged with a GST tag at the N- terminus using the pLEICS02 vector as shown in Figure 3-2 above (PROTEX, University of Leicester). The obtained construct was confirmed and transformed in to *E. coli* BL21 (DE3) as described above in section 3.3.2, thereafter the expression was conducted as mentioned in 3.2.4.1.

3.3.8.2 GST tagged Gp29 purification

The Gp29 protein was purified using a batch method purification, this purification is also an affinity chromatography based method. The procedure was conducted as followed; cells were harvested from a 1 L culture at 5000 x g for 20 minutes, then the pellet was resuspended in 50 ml of lysis buffer containing 1X PBS buffer and 5 mM DTT and 1 Complete EDTA-free Protease inhibitor tablet (Roche). Thereafter, the cells were

Cloning, expression and purification of the tail proteins for CDHS1 & CDMH1 phage

sonicated using 7 pulses for 20 seconds each with a one-minute rest between each pulse.

After which centrifugation was conducted for 30 minutes at 20,000 x g at 4 °C.

The glutathione-Sepharose 4B for batch separation was prepared by mixing 4 ml of glutathione-Sepharose 4B resin in a 50 ml centrifuge tube, washed with 10 ml distilled water and equilibrated using 15 ml of 1X PBS. Then the supernatant was added to the equilibrated resin and rocked gently at 4 °C for 2 hours. After incubation, centrifugation was performed to remove unbound protein and then resin was washed 4 times using 20 ml of 1X PBS. For each wash, 5 ml of the washing buffer (1X PBS) was added to the resin, rocked for 5 minutes and then centrifuged for 4000 x g at 4 °C, this process was repeated 4 times. Then TEV protease enzyme (PROTEX) was added to cleave the GST tag, the digestion was conducted overnight. Finally, the elution buffer (50 mM Tris-HCL pH 8.0, 10 mM reduced glutathione, 5 mM DTT) was added, and the elution was done 4 times, each time 2 ml of the elution buffer was added. The purified protein was viewed using SDS page as described above.

3.3.8.3 GST tagged Gp30 purification

Gp 30 protein was purified using the same protocol that are described above in sections 3.2.6.1 and 3.2.6.2.

3.4 Results

3.4.1 Amplification of the putative genes that encode tail proteins for CDHS1 and CDMH1 phages

3.4.1.1 Amplification of the putative genes *gp18*, *gp19*, and *gp22* of CDHS1

PCR was used to amplify the *gp18*, *gp19*, and *gp22* that encode the putative tail proteins for phage CDHS1. The results from PCR reactions are presented in Figure 3-4, it was found that the respective primers were all successful in amplifying the phage DNA at the expected sizes. These three PCR products (*gp18*, *gp19*, and *gp22*) extracted from the gel and sent for cloning to the Protein Expression Laboratory (Protex) in Biochemistry Department, University of Leicester

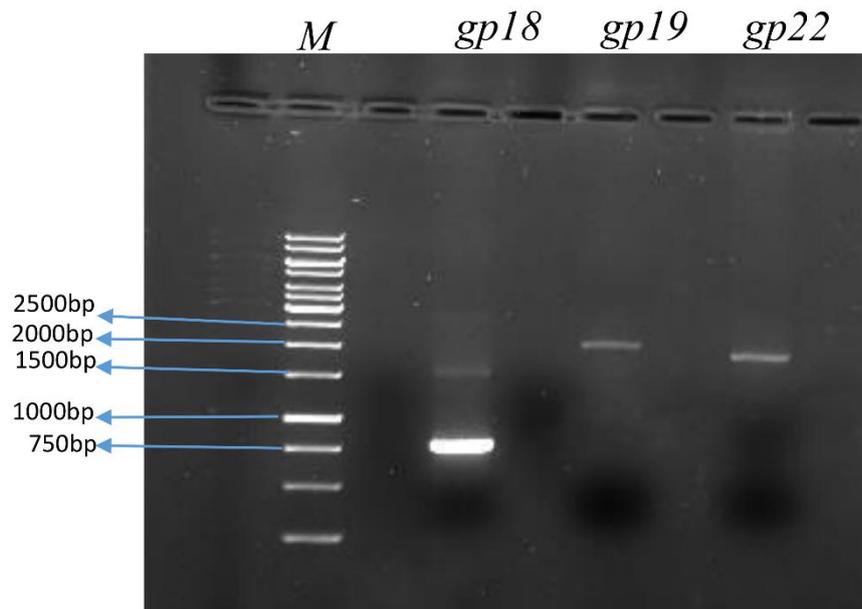


Figure 3-4:- PCR products for the three putative genes encoding tail fiber proteins for phage CDHS1 separated on a 1% agarose gel

The results presented in this Figure show that the designed primers were all successful in amplifying the genes of interest. Lane1 is a 1 kb DNA ladder with size indicators. Lane 2 is the PCR product for *gp18* that is 750 bp in size, lane 3 is *gp19* with size of 2010 bp and lane 4 indicates *gp22* with size of 1851 bp

3.4.1.2 Cloning

The amplified genes were sent for cloning to the Protein Expression Laboratory (Protex) and sequenced by The Protein Nucleic Acid Chemistry Laboratory (PNACL). Sequences were verified by alignment against the original sequence using the ClusterW2 tool. The three genes *gp18*, *gp19* and *gp22* were matched to the original sequences of these genes. In the case of *gp21*, the cloning attempt failed after several attempts by PROTEX so the cloning procedure was carried out internally within our lab.

3.4.1.3 Cloning *gp21* genes

The gene *gp21* was amplified from the CDHS1 genome. The result illustrated in Figure 3-5 shows that the designed primers were successful in amplifying the gene of interest *gp21*, thereafter the PCR product was cleaned and was ready to be cloned.

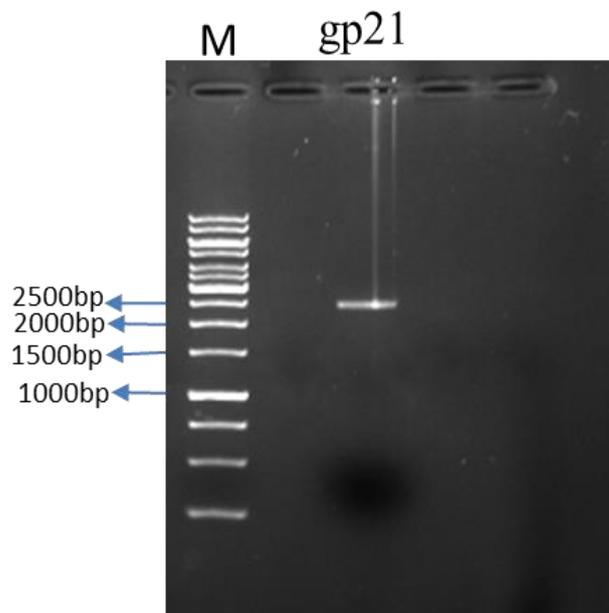


Figure 3-5: PCR products of CDHS1 *gp21* gene separated on a 1% agarose gel

This Figure shows that the successful amplification of the *gp21* of interest. Lane1 is 1 a kb DNA ladder with size indicators. Lane 3 is the PCR product for *gp21* that is 2472 bp in size.

3.4.1.4 Amplification of the two putative genes that predicted to be encoding tail proteins for CDMH1

PCR was also used to amplify genes *gp29* and *gp30* which encode for putative tail proteins for phage CDMH1. The results presented in Figure 3-6 demonstrate that products of expected sizes were generated via PCR. Finally, these two PCR products were cleaned up and sent to the PROTEX service for cloning.

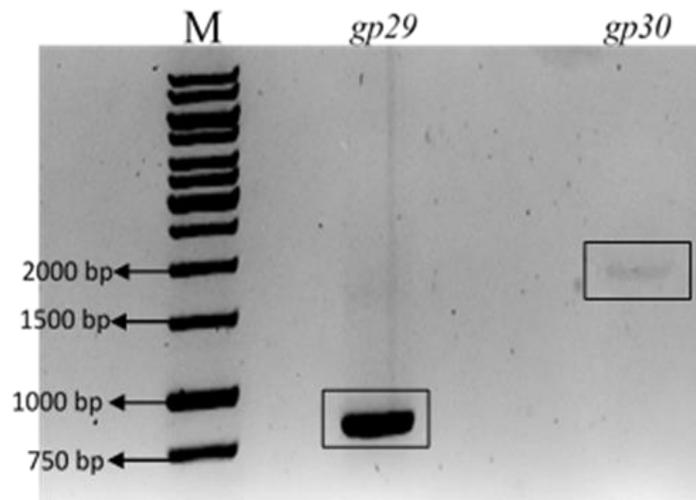


Figure 3-6: PCR products for the two putative genes encoding tail fiber proteins for phage CDMH1

The results presented in this figure shows that the designed primers were successful in amplifying the genes of interest. Lane1 is a 1 kb DNA ladder with size indicators. Lane 2 is the PCR product for *gp29* that is 805bp in size, lane 3 is *gp30* with size of 1695 bp

3.4.2 Protein expression and purification of tail proteins of CDHS1 phage

3.4.2.1 Expression and purification of His tagged Gp18

To express the Gp18 protein, *E. coli* BL21 (DE3) was grown at 37 °C and expression was induced using 0.1 mM of IPTG. Gp18 was purified using affinity chromatography on a nickel column as the first step. Figure 3-7a shows the SDS-PAGE gel of the eluted nickel column fractions. The Gp18 protein migrated to the expected band size of 28 kDa. Thereafter, the Gp18 protein was further purified using gel filtration chromatography. Figure 3-7b represents the elution profile of Gp18 protein obtained from the gel filtration purification step. The peak corresponding to the Gp18 protein eluted at 67ml which was expected from the size of the 20 kDa protein, based on the elution position of molecular weight standards. Fractions were collected across the elution profile of the peak corresponding to the Gp18 protein for analysis by SDS- PAGE as shown in Figure 3-7c which represents pure Gp18 at the expected size (28 kDa). Gp18 protein fractions obtained from the gel filtration process were pooled together and concentrated to 13.9 mg/ml.

Cloning, expression and purification of the tail proteins for CDHS1 & CDMH1 phage

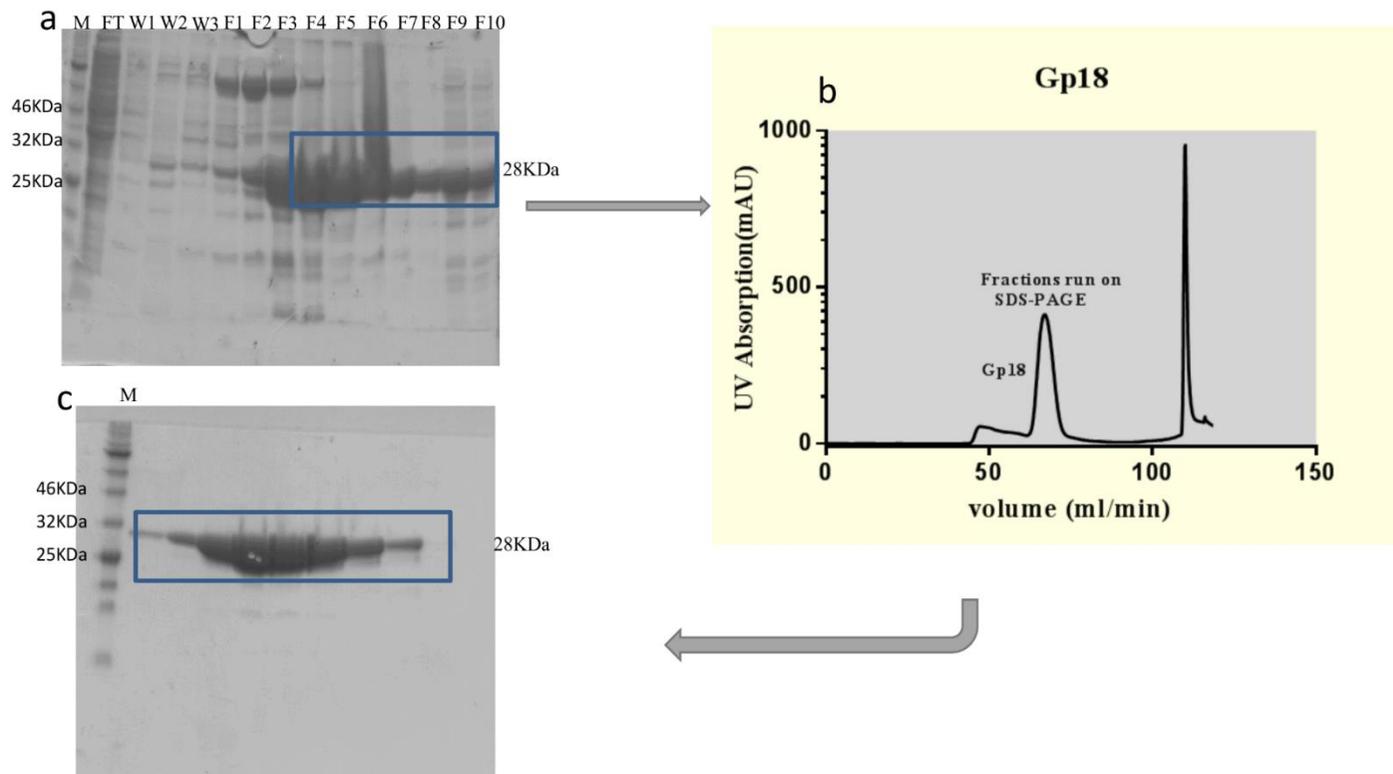


Figure 3-7 : Purification of the His tagged Gp18 protein

(a) SDS- PAGE Gel shows the step gradient affinity chromatography purification for Gp18 protein, the molecular weight of this protein is 28 kDa, and the desired bands are highlighted in blue, the protein was eluted using step gradient imidazole 80mM (F1, F2A and F3), 100mM (F4, F5 and F6), and 250mM (F7, F8, F9 and F10). Then the desired fractions were pooled together, spun down and purified further using Gel Filtration. (b) Represents the elution profile of the Gel Filtration purification fractions of the Gp18 protein on a Superdex 75 column. (c) Shows the SDS-PAGE Gel of the pure version of the Gp18 protein after gel filtration at the expected size of the proteins 28kDa.

3.4.2.2 Expression and purification of GST tagged proteins from CDHS1

3.4.2.2.1 GST tagged Gp19

Gp19 was expressed in the same way as the Gp18 protein, the result in Figure 3-8 shows that the Gp19 protein was not present at the expected band size, possibly due to degradation during the purification steps. The bands that are highlighted with blue represent the GST tag at the right size (26 kDa) and was identified as the tag by sequencing using the PNACL service at the University of Leicester.

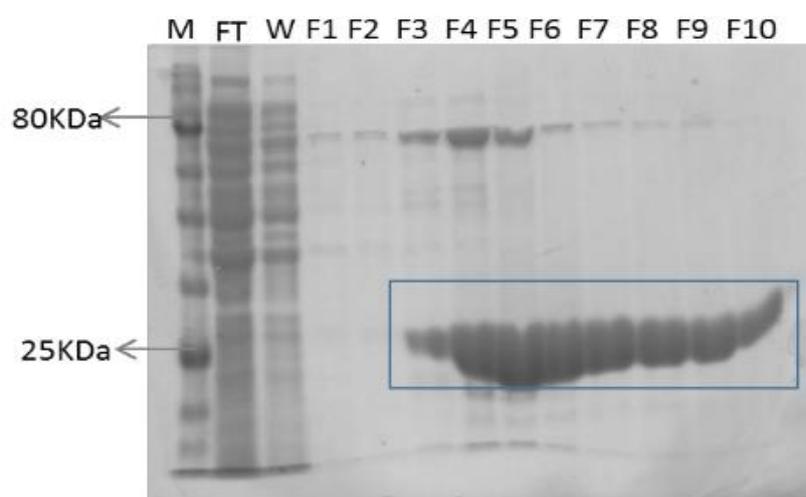


Figure 3-8 : SDS-PAGE for the purification of GST tagged Gp19

The expected size of the recombinant Gp19 is 106kDa, however after the purification step only the GST tag could be identified, represented by the band present at the 26 kDa band region.

3.4.2.2.2 GST tagged Gp22

The purification of the Gp22 protein was carried out using a glutathione affinity chromatography column as a first step, and the eluted fractions were ran on 12% SDS PAGE gels. The expected bands of the recombinant Gp22 protein were observed at 94KDa as shown in Figure 3-9a. Then the Gp22 protein was cleaved from the GST tag via the TEV protease activity. Then, for further purification, gel filtration chromatography was conducted, the elution profile is presented in Figure 3-9b. Two sharp peaks were eluted from the column; the first peak eluted at 66.7ml which corresponded to the expected Gp22 protein size (68 kDa) based on the known elution volumes of molecular weight standards. The second peak was eluted at 82.7ml which matched the molecular weight of the GST tag (26 kDa). Then the desired fractions (corresponding to Gp22 form peak one) from the gel filtration were separated using a 12% SDS PAGE gel Figure 3-9c. Then the fractions of interest were concentrated using Amicon Ultra-4 10K Centrifugal Filter Devices (Millipore, UK). The concentration of this protein was measured and was found to be 11mg/ml.

Cloning, expression and purification of the tail proteins for CDHS1 & CDMH1 phage

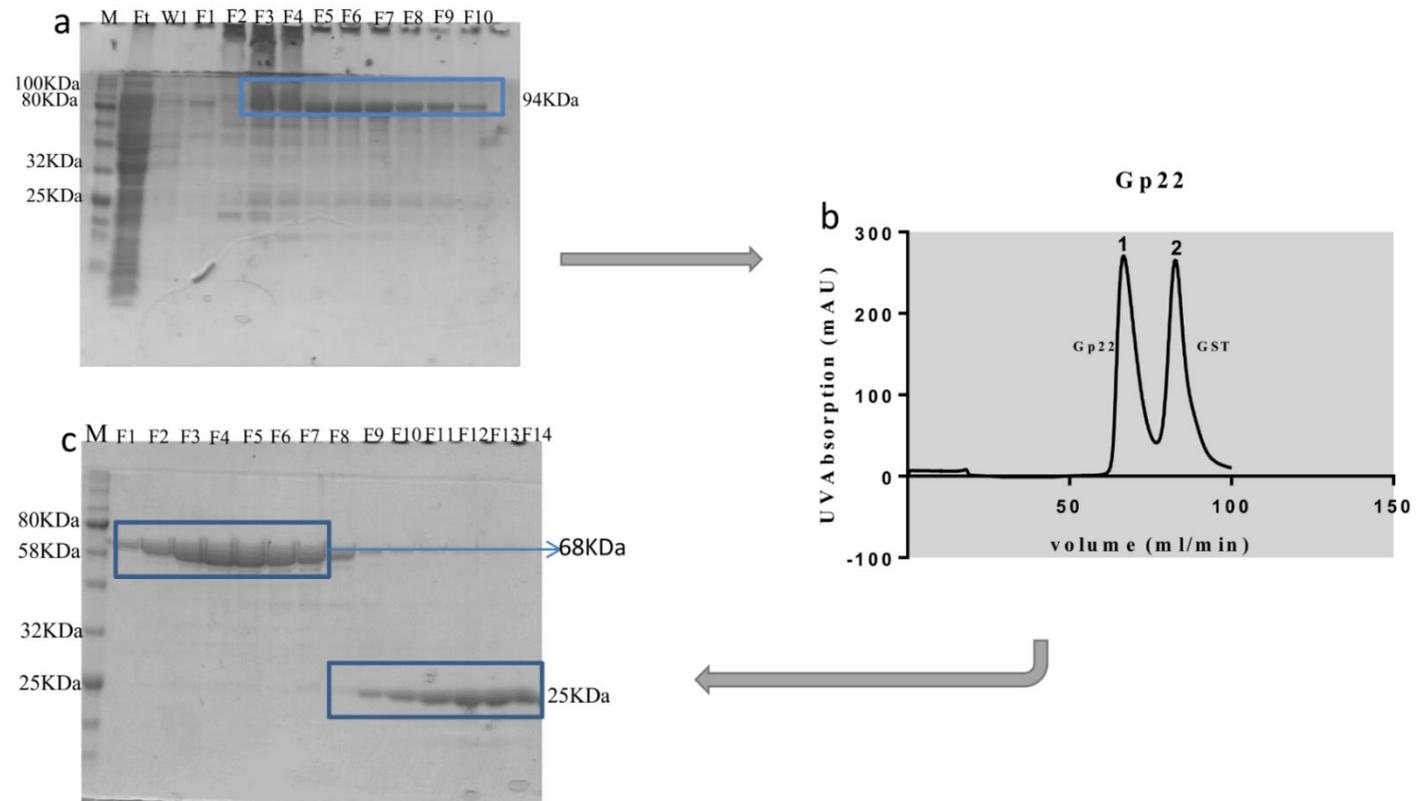


Figure 3-9: Purification of GST tagged Gp22

(a) Represents the 12% SDS-PAGE Gel of the affinity chromatography purification of recombinant Gp22, the size of the recombinant protein is 94 kDa and the corresponding bands are highlighted in blue. M represents protein marker, FT is the flow through, W1 is washing step and F1 to F10 are the elution fractions. (b) Representation of the gel filtration column elution profile of the Gp22 protein. (c) Shows the fractions of the pure Gp22 protein after gel filtration purification separated on a 12 % SDS-PAGE gel. The bands from F1 until F8 represents the size of the protein Gp22 (68kDa), whereas the bands in F9 to F14 relate to the size of the GST tag (26kDa).

3.4.2.2.3 GST tagged GP21

The expressed Gp21 was purified using affinity chromatography purification first, and the result in Figure 3-7a show the SDS-PAGE Gel analysis of the Gp21 protein fractions obtained from the glutathione affinity chromatography purification. In the second step of purification gel filtration was used, and the elution profile of the Gp21 protein is shown in Figure 3-7b. The Figure shows that there were three peaks eluted from the column, the first one was at 52 ml, which represents the recombinant Gp21 protein, most likely due to uncompleted digestion with thrombin protease. However, the second eluted peak was at 61.9 ml which corresponds to the size of Gp21 protein (83 kDa), based on the elution of known standards. Whereas the third peak was the GST tag. Figure 3-7c represent the SDS-PAGE Gel of Gp21 fractions obtained from the gel filtration purification (peak 2).

Cloning, expression and purification of the tail proteins for CDHS1 & CDMH1 phage

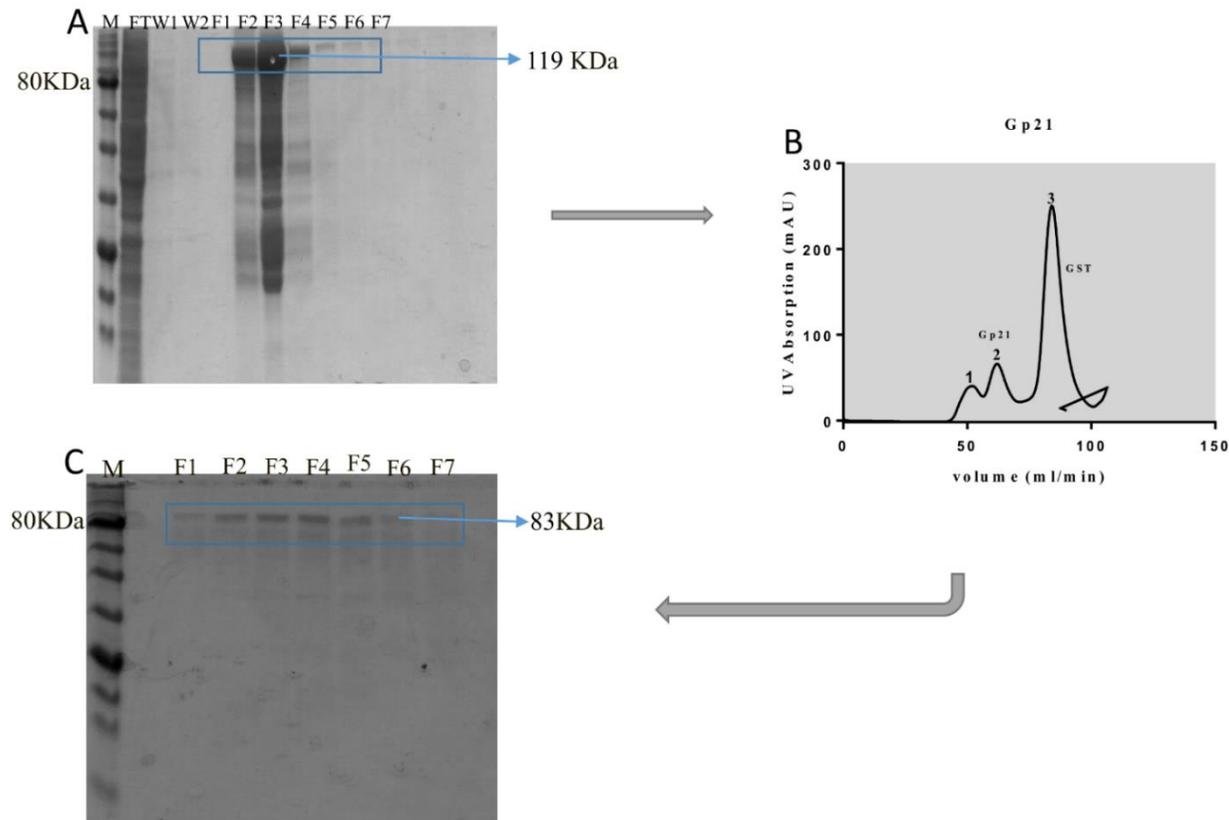


Figure 3-10: Purification of the GST tagged Gp21 protein

(A) Shows the SDS-PAGE of the affinity chromatography purification of recombinant Gp21, the expected size of the recombinant protein is 119 kDa and the corresponding bands are highlighted with blue. M represents protein marker, FT is the flow through, W is the washing steps and F1 to F7 are the elution fractions. (B) Represents the elution profile of the gel filtration purification of Gp21 protein. (C) Represents the fractions resulted from gel filtration purification for recombinant protein Gp21. The bands from F1 until F7 relate to the size of Gp21 (83kDa) separated on a 12% SDS- PAGE.

3.4.3 Protein expression and purification of the tail proteins of CDMH1 phage

3.4.3.1 GST tagged Gp29 from CDMH1

The overexpressed Gp29 was purified using batch method purification, then the eluted fractions were ran on a 12% SDS PAGE gel and the bands corresponding to the size of the Gp29 protein was highlighted as shown in Figure 3-11. Then fractions of interest were concentrated using Amicon Ultra-4 10K Centrifugal Filter Devices (Millipore, UK), the concentration of this protein was found to be 3 mg/ml.

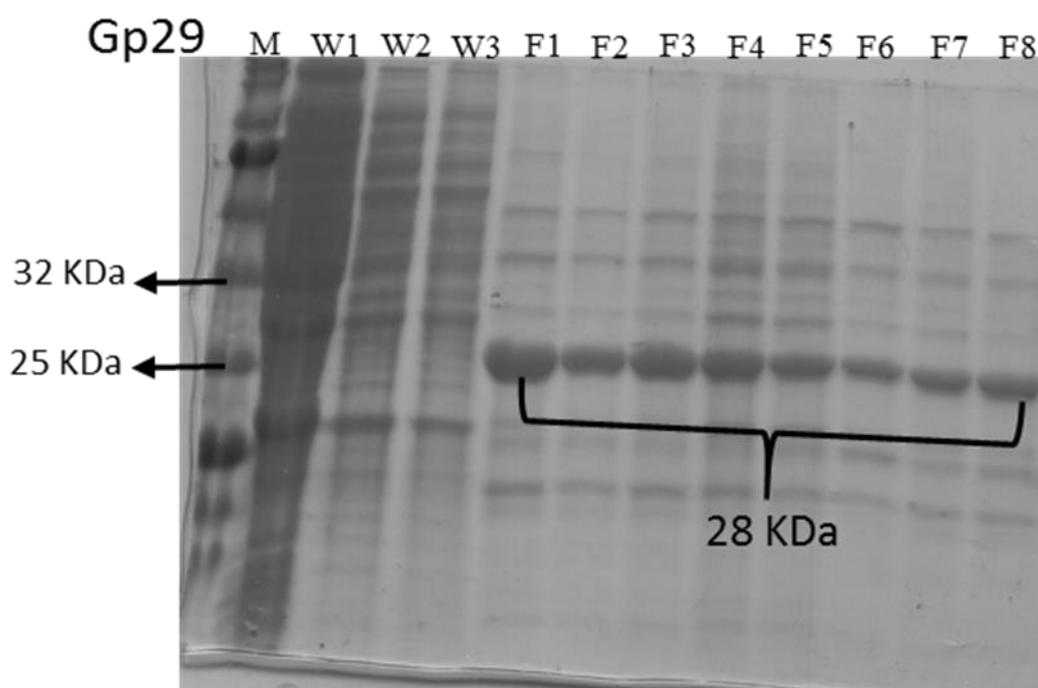


Figure 3-11: SDS PAGE for CDMH1 GP 29 purification

Shown above is the results of the batch method purification of Gp29, the size of the protein is 28 kDa and the corresponding bands were highlighted. M represents protein marker, W1 to W3 are samples from washing steps and F1 to F8 are the elution fractions.

3.4.3.2 GST tagged Gp30 CDMH1

The purification process was carried out using a glutathione affinity chromatography column and the eluted fractions were ran on a 12% SDS PAGE Gel as shown in Figure 9a. For further purification, a gel filtration separation was conducted. Figure 9b shows the elution profile of Gp30 after gel filtration, the profile pattern of the Gp30 protein presented as three peaks. The first peak was an aggregate, the second peak represented the Gp30 protein based on the expected size of the Gp30 protein (60 kDa), whereas the third peak was for the GST tag protein. Then the column fractions that corresponded to the proteins of interest were ran on a 12% of SDS PAGE Gel as shown in Figure 9c. Surprisingly the GST tagged Gp30 protein was found in the fractions that were collected from peak 2 which was supposed to represent the protein Gp30 only. The fractions of interest were concentrated using Amicon Ultra-4 10K Centrifugal Filter Devices (Millipore, UK).

Cloning, expression and purification of the tail proteins for CDHS1 & CDMH1 phage

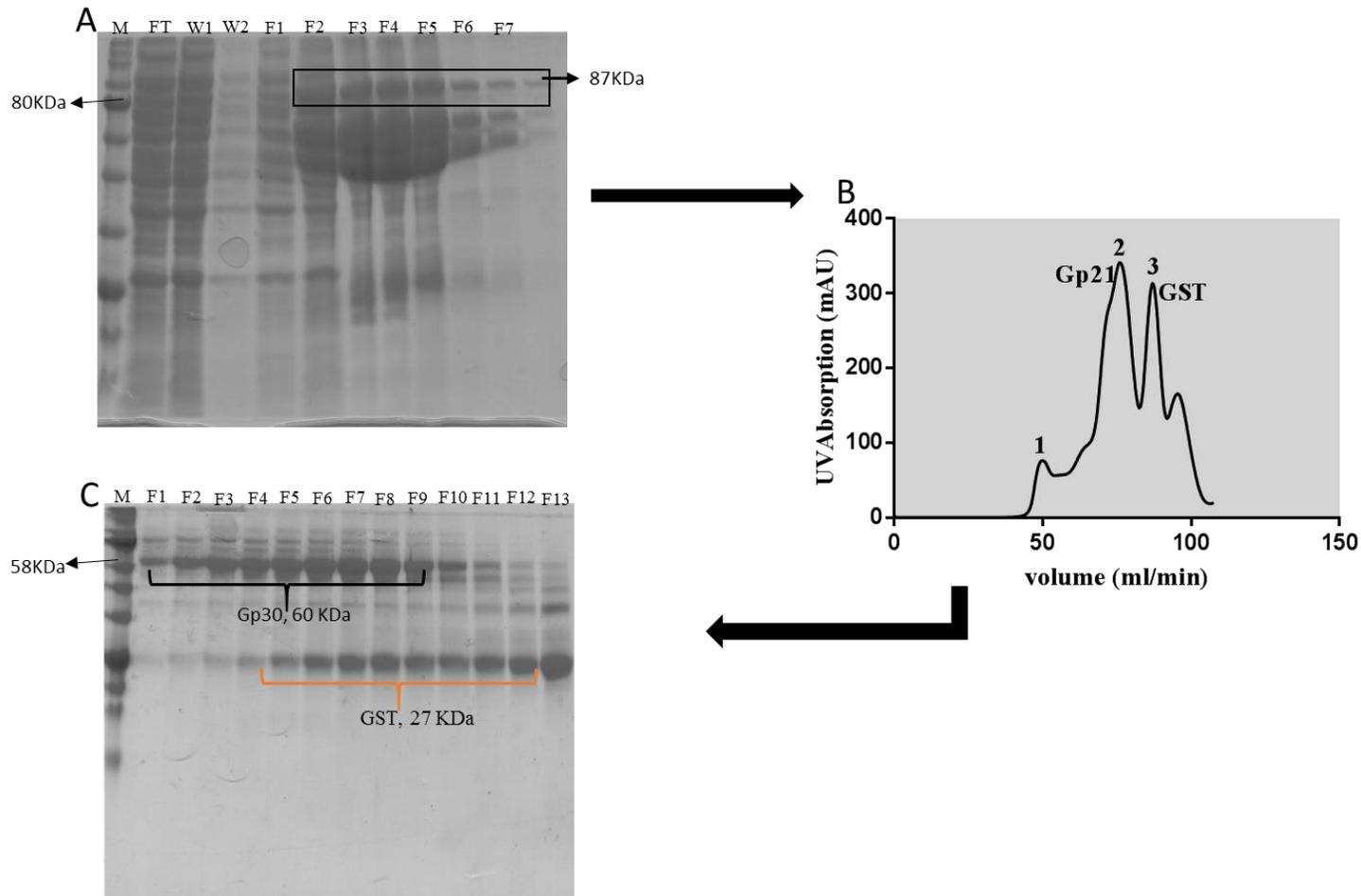


Figure 3-12: purification of GST tagged Gp30 from CDMH1

(A) Shows the SDS-PAGE of the affinity chromatography purification for Gp30 protein, the molecular weight of recombinant protein is 87 KDa, and the desired bands are highlighted in black. (B) Represents the elution profile of Gp30 protein fractions resulted from gel filtration purification (C) Illustrate is the SDS PAGE of the Gel filtration purification fraction of Gp30.

3.5 Discussion

To date, the interaction between *C. difficile* and its phages has not been characterised yet. This work is the first attempt to identify the receptors binding proteins for phages that infect *C. difficile*. In this chapter the proteins that are involved in the phage baseplate structure and in phage-bacterial binding were overexpressed and purified for two phages CDHS1 and CDMH1, both of which infect different *C. difficile* strains.

3.5.1 Protein expression and purification for the tail proteins of CDHS1

The genes that encode the proteins that form the baseplate for siphoviruses of Gram-positive bacteria are located downstream to the gene encoding the tape major protein (*tmp*) and upstream of the genes encoding the holin and endolysin proteins (Bielmann *et al.*, 2015). Therefore, for the siphovirus CDHS1 of *C. difficile*, there are four genes *gp18*, *gp19*, *gp21* and *gp22* that are located within this region. They were amplified and cloned. Cloning for the genes of interest *gp18*, *gp19* and *gp22* was successfully carried out by PROTEX service with no difficulties. However, the gene *gp21* was cloned in our lab due to the inability of PROTEX to clone it.

Several challenges were encountered during the expression and purification process of those proteins. As these two processes were very time consuming for most of the proteins, IPTG was used to induce the proteins overexpression. Therefore, in order to find the optimum concentration of the IPTG, different concentration of IPTG were used to induce the expression, also different temperatures were used to obtain the optimum conditions for the production of the four proteins. Despite the challenges that have been faced during the expression process, the four proteins of interest were successfully expressed.

The first protein to be expressed and purified was the Gp18 protein. As in chapter 2 section (2.3.1.1), the *in silico* analysis of Gp18 protein showed that the protein have

Cloning, expression and purification of the tail proteins for CDHS1 & CDMH1 phage

similar structural homology to the distal tail protein (Dit) from phage TP901-1 that infects *L. lactis*. The Dit protein is considered to be conserved in the baseplate of siphoviruses that infect Gram-positive bacteria, and it forms the central hub of the baseplates of those Siphoviruses (Veesler *et al.*, 2012).

Gp18 was expressed easily, however the purification was quite challenging due to the presence of non-specific bands during the affinity chromatography purification. These non-specific bands resulted from the nonspecific binding which occurred between the histidine of the *E. coli* and the Nickel column, this interaction occurs with weak binding. Therefore, such non-specificity was overcome using linear gradient of the elution buffer, thereafter, gel filtration purification was carried out to obtain highly pure Gp18 protein and to also remove the contaminants.

The second protein to be expressed was Gp22 which was fused with a GST tag. This protein was expressed and purified more easily than Gp18 protein. However, challenges were encountered when Gp21 proteins was further processed. The *in silico* analysis of Gp21 protein revealed that this protein had structural homology to the upper baseplate protein of TP901-1. The cloning service (PROTEX, University of Leicester) found it difficult to clone the gene encoding Gp21 protein after several attempts. Therefore, cloning process was successfully done in our lab using the PGEX-4T plasmid, and the gene was fused with a GST tag. The resultant construct was sequenced to confirm the presence of the insert.

On the other hand, the purification process of Gp21 protein was challenging too, due to the instability of this protein. As Gp21 protein tended to degrade during the digestion of the Gp21 recombinant protein using the thrombin protease enzyme to cleave the protein from the GST tag. As this process of cleaving Gp21 from the GST tag was performed at

Cloning, expression and purification of the tail proteins for CDHS1 & CDMH1 phage

room temperature for 16h, Gp21 was not stable and subsequently degraded during this process. An attempt to conduct the digestion process at 4 °C for a longer time duration was performed, however the digestion was not completely successful. Therefore a large volume (6 liters) of culture was used and by this way, 6 mg/ml of the Gp21 protein was obtained.

Gp19 protein was cloned successfully and the insert was confirmed using sequencing, however, this protein was difficult to purify. As shown in the results 3.3.2.2.1, this protein was lost, most likely degraded during the purification process as the outcome of the affinity chromatography showed that the only bands present were those that corresponded to the size of GST tag (26 kDa). Whereas the expected band size of the recombinant protein was supposed to be 106 kDa. The definitive cause of Gp19 loss is unknown. However, protein degradation is a frequent problem, and it occurs during any stage of protein expression and purification. One of the key causes for protein degradation is the activity of proteases that are produced by *E. coli*. Therefore, several precautions were taken to prevent the loss of this protein by protease activity, such as using a protease inhibitor cocktail, conducting the purification in the cold room and processing the protein as fast as possible after breaking the cells (Ryan *et al.*, 2013). Despite all the above, the loss of this protein still occurred.

The gene encoding this Gp19 protein is located downstream to the gene encoding Gp18 protein. In addition to this, using the HHpred software, Gp19 reveal homology similarity to an endopeptidase. The location of the gene encoding Gp19 protein as well as the results from the *in silico* analysis of this protein could suggest that Gp19 may belong to the virion-associated peptidoglycan hydrolases (VAPGHs) analogous. This is a group of enzymes that phages use to degrade the bacterial peptidoglycan so that the phage may inject DNA during phage infection (Drulis-Kawa *et al.*, 2015).

3.5.2 Protein expression and purification for tail proteins Gp29 and Gp30 of CDMH1

The work that has been done to identify and characterize the receptors binding protein (RBPs) of myoviruses that infect Gram-positive bacteria is still rare compared to the amount of work that have been done for siphoviruses of Gram-positive bacteria. Therefore, CDMH1 phage, a myovirus that infects *C. difficile* was targeted in this study to identify the RBPs for this phage. The approaches that have been used to identify the RBPs for CDMH1 were similar to the one used in identification of RBPs for CDHS1 phage. Two genes that are predicted to encode tail proteins are *gp29* and *gp30* were targeted. The cloning of these two genes was performed successfully using PROTEX and the resulting constructs were sequenced to confirm the desired insert.

The two-resulting proteins were overexpressed successfully, whereas the purification was quite challenging, especially in the case of Gp30. Both proteins (Gp29 and Gp30) were fused with a GST tag. In the case Gp29, the molecular weight of this protein is 28 kDa in size, which means that this protein and GST (26kDa) are close to each other in term of size, and that would lead to difficulty in the separation of Gp29 protein from the GST tag, and removing the GST tag using Gel filtration method. Therefore, the best option to purify and then separate Gp29 protein from the GST tag was to use a Batch purification method.

Whereas, for the Gp30 protein, the purification was time consuming as difficulty was experienced after the successful cleavage of the GST tag from the Gp30 protein. When the sample was subjected to gel filtration purification to enhance the purity of this protein, the outcome was that there was difficulty in removing the GST tag completely from the Gp30 protein fraction as shown in Figure 3-9c. As shown, both Gp30 and the GST tag kept eluting together at the same peak, which was at the expected peak of Gp30 based on its size. The reason behind the co-purification of GST with Gp30 is unknown. The Batch

Cloning, expression and purification of the tail proteins for CDHS1 & CDMH1 phage

purification method would be the best option to overcome this problem. Due to the possibility that the GST tag can be cleaved while the recombinant protein would still be bound to the glutathione beads. Then the Gp30 protein will be found in the flow through. However, due to shortages in time, it was difficult to perform this method for the Gp30 protein.

3.6 Conclusion

In conclusion, this chapter described the gene amplification, production and the purification of the proteins that were predicted to be tail proteins involved in phage attachment to *C. difficile*. Two phages were used, CDHS1 and CDMH1, and the main finding of this chapter were:

- 1- Out of the four tail proteins that have been targeted for the study for CDHS1 phage, three of them Gp18, Gp21 and Gp22 were successfully produced and purified.
- 2- For Gp19, there was difficulties faced in the production of this protein.
- 3- On the other hand, for phage, the two proteins Gp29 and Gp30 that were targeted were successfully produced and purified.

Chapter 4 Identifying the Receptor binding
proteins for *C. difficile* phages CDHS1 and
CDMH1

4.1 Introduction

In chapter 3, it was shown that three of the four CDHS1 phage proteins predicted to be tail proteins (Gp18, Gp21, and Gp22) that may be involved in phage binding with *C. difficile* were produced and purified successfully. Furthermore, for phage CDMH1, the two possible candidate proteins (Gp29 and Gp30) that may act as RBPs were also produced and purified successfully. The purified proteins were further investigated in order to identify which one of these are the RBPs for these two phages. This chapter describes the steps to identify the phage RBPs for CDHS1 and CDMH1.

As stated in the introduction in chapter 1, the first step of phage infection is the attachment of phages to the bacterial host cells. Specifically, this attachment occurs between the proteins located at the end of the phage tail known as receptor binding proteins (RBPs) and the cell wall associated ligands that are known as phage receptors. The recognition that occurs between RBPs and receptors on bacterial host is specific, and has a high binding affinity and is essential for the phage infection (Bielmann *et al.*, 2015). Several phages of Gram-positive bacteria have had their RBPs identified. However, to date no studies have identified the RBPs of *C. difficile* phages or their corresponding receptors on the bacterial surface that phages bind or adsorb to, in order to establish infection. This is the first study to report the identification of the RBPs for *C. difficile* phages.

Several approaches have been used to identify phage RBPs, such as Bioinformatic analysis, followed by antibody based studies. In general, polyclonal antibodies are raised against overexpressed putative tail fiber proteins and then used to neutralise phage infection (Li *et al.*, 2016a). Molecular approaches are also employed, for example, a chimeric phage was produced to identify the RBPs of the *L. lactis* phages TP901-1 and Tuc2009. In this approach, the gene encoding TP901-1 lower baseplate protein (bppL) was replaced with the gene (*orf53*) of phage Tuc2009 (Vegge *et al.*, 2006). The results

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

showed that the chimeric TP901-1 phage was able to infect Tuc2009 host strain efficiently, indicating that the TP901-1 lower baseplate protein (bppL) and (orf53) of phage Tuc2009 are both responsible for the phage attachment to the host (Vegge *et al.*, 2006).

Another approach that has been used to identify the RBPs for phages P35 and A118 of *L. monocytogenes*, is the use of purified predicted tail fiber proteins that were tagged with Green fluorescent protein (GFP) and incubated with *L. monocytogenes* cells. The results obtained from this assay, showed that the proteins (A118 GFP-gp19, GFP-gp20, and P35 GFP-gp16) coated the specific strain of *L. monocytogenes* fluoresced green under the confocal microscope. Indicating that the proteins A118 GFP-Gp19, GFP-Gp20, and P35 GFP-Gp16 may play a key role in phages A118 and P35 binding with *L. monocytogenes* (Bielmann *et al.*, 2015).

In this study several approaches were followed to identify the RBPs for CDHS1 and CDMH1. The first approach was to use the GFP tagged putative tail proteins, the same approach used for *L. monocytogenes* phages (Bielmann *et al.*, 2015). However, this was not successful due to difficulties in the construction of the plasmid for some of the genes encoding tail proteins that were targeted in this study. In addition, *C. difficile* naturally generates green fluorescence when excited with a blue/ultraviolet (UV) light. This would make any visualisation of GFP along with *C. difficile* problematic. Due to these challenges other approaches were followed.

The second approach applied utilised the binding affinity between biotin and streptavidin (Li *et al.*, 2016b). In brief, an enzyme-linked immunosorbent assay (ELISA) based principle was applied, where the purified predicted tail proteins were biotinylated and incubated with *C. difficile*. Then the streptavidin conjugated with alkaline phosphatase was added to the mixture, thereafter a substrate was added. The ELISA reader was then

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

used to determine the binding, which would be indicated by a colour generated after adding the substrate. However, this assay was not successful, due to the lack of a positive control, as well as due to a strong background colour that was generated after adding the substrate in all the proteins tested. This made it difficult to distinguish whether the colour change was due to the binding with *C. difficile* or a false positive result.

The final and most successful approach used in this project was to purify tail proteins and to raise polyclonal antibodies against them. Subsequently, these generated antibodies were incubated with phages to determine which protein could neutralise the phage infection. This approach has been used to identify RBPs of many phages such Phage $\phi 11$ of *S. aureus* and *L. monocytogenes* phages A118 and P35 as (Li *et al.*, 2016b, Biemann *et al.*, 2015). Further use of these antibodies was in immunogold labelling to visualise the location of these proteins on the phage particle using Transmission electron microscopy (TEM).

4.2 The aims of this Chapter

The aim of this chapter was to determine which of the purified tail proteins of CDHS1 and CDMH1 act as RBPs and are involved in phage binding to *C. difficile*.

To obtain the aim of this study: -

- 1- The purified Gp18, Gp21 and Gp22 proteins of CDHS1, in addition to Gp29 and Gp30 from CDMH1 were sent to Eurogentec (Brussels, Belgium) to produce polyclonal antibody against these proteins.
- 2- To determine which of these proteins act as RBPs, the produced antibodies would be tested to see if they could neutralize phage infection.
- 3- The antibodies were then used in the immunogold labelling of the phage particles to identify the specific location of these proteins on the phage particle using Transmission Electron Microscopy (TEM).

4.3 Methods

4.3.1 Strains and Culturing

Three *C. difficile* strains were used in this study: CD105LC1 ribotype 027 which is the propagating host for phage CDHS1, CDR20291 also ribotype 027 as a second host for CDHS1. In addition to this, strain CD105HE1 ribotype 076 was used to propagate phage CDMH1. These strains were harvested by growing them in anaerobic conditions (10% H₂, 5% CO₂, and 85% N₂) on Brain-Heart Infusion (BHI) agar plates, supplemented with 7% defibrinated horse blood (DHB) for 24 hours at 37 °C. Liquid cultures were prepared by taking a single colony from the blood agar plate, and then inoculated in either 7 ml bijou tube containing 5 ml of Fastidious Anaerobic Broth (FAB) or to a 50 ml centrifuge tube containing pre-reduced BHI broth, based on the purpose of use. The liquid cultures were left to grow overnight in anaerobic conditions.

4.3.2 Phage propagations

To obtain a sufficient number of phages to be used in the experiments within this project, two methods described below were used to amplify the phages.

4.3.2.1 Plaque Assay

Plaque assays were performed using 120 mm square Petri dishes with 1% BHI Agar. These plates were overlaid with a mixture of 300 µl of FA broth *C. difficile* culture and 150 µl of the phage CDHS1 or CDMH1 phage respectively, in addition to this 8 ml of BHI 0.4% agar supplemented with 0.4 M MgCl₂ and 0.1 M CaCl₂ (to aid phage attachment) was also poured onto the agar plate. When the plates were set they were placed in an anaerobic chamber at 37 °C for 24 hours.

To collect the phages, the plates were taken out from the anaerobic incubator and the overlay of BHI 0.4% agar was scraped into a 50 ml centrifuge tube and 5 ml of BHI broth was added. This was then placed on a mechanical rocker for 20 minutes, after which it was stored at 4 °C overnight to allow the phage to diffuse from the agar to the broth. Then

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

the tube was centrifuged two times at 3390 x g for 10 minutes and, the supernatant filtered, using a 0.22 µm filter, into another clean centrifuge tube. Then the obtained filtrated supernatant containing the newly harvested phages were spot tested to determine the titration of the phage (the number of plaque forming units (PFU) within the filtrate), this method is described below.

4.3.2.2 Phage propagation using liquid culture

Phage stock were also obtained using phage propagation in liquid culture. Briefly, 500 µl of overnight culture in FAB was added to 50 ml pre reduced BHI and the bacteria were allowed to grow until it reached an OD₅₅₀ of 0.2. Thereafter 500 µl of the desired phage was added to the culture and incubated for 24 hours. Then, the phage was harvested by centrifuging at 3,400 x g for 10 min, the supernatant was filtered using a 0.22 µm filter, finally the phage concentration was determined using the spot test assay. This method was used as it gives high concentration of phage used.

4.3.3 Spot test

These tests were conducted to determine the phage concentration (PFU/ml). For each sample of the phage to be tested, serial ten-fold dilutions were prepared. A bacterial lawn was prepared by adding 550 µl of a *C. difficile* overnight culture (FAB) to 8 ml of the BHI 0.4% agar mixed with salt. The tube was inverted a few times to mix well, then poured onto a plate and allowed to set. Thereafter, 10 µl of each dilution was spotted onto the lawn. The spots were then left to dry and the plates were incubated anaerobically at 37 °C overnight. After incubation the plates were removed from the anaerobic chamber and the PFU/ml was calculated from the plaques observed.

4.3.4 Phage purification

After the phages were propagated. Phages were purified using anion-exchange chromatography with Convective Interactive Medias monoliths (CIMs). The column

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

used in this assay was a 1 ml quaternary amine (QA), which is a monolith column developed by BIA Separations (Slovenia).

QA was attached to the AKTA purifier machine and then was equilibrated with the 30 ml of binding buffer (50 mM Tris-HCL pH 7.5 and 8 mM MgSO₄·7 H₂O). 5 ml of the phage sample was loaded to the machine. Thereafter a linear gradient was ran to elute the phage using high salt buffer (50 mM Tris-HCL pH 7.5, 8 mM MgSO₄·7 H₂O and 2 M NaCl). Then the fractions corresponding to the desire peak were titrated using the spot test.

4.3.5 Phage neutralization using antibodies

This assay was used to determine which one of the tail proteins Gp18, Gp21 and Gp22 of phage CDHS1 and Gp29 and Gp30 of phage CDMH1 were responsible for binding to *C. difficile*. Once the aforementioned proteins were expressed, purified, and the concentration measured using the Nano drop, 50 µl of each protein suspended in 20 mM of NaCl and 20 mM Tris-HCL pH 7.5 was sent to the Eurogentec Company (Brussels, Belgium) for the generation of polyclonal antibodies. A custom Speedy 28-Day polyclonal service was used. After one month the anti-serum antibody were received in three forms; pre-immune, the second bleed and the final bleed for each protein. The anti-serum antibodies were used to determine which of these antisera was able to neutralize the phage infection.

Each antibody serum was diluted using a saline magnesium (SM) buffer (10 mM NaCl, 8 mM, MgSO₄·7H₂O, and 50 mM Tris-HCL pH 7.5, prepared in 1 Litre), the dilutions were 1:10, 1:100, 1:1000 and 1:10000. Then phages CDHS1 and CDMH1 were added to each of the dilutions mentioned above. The mixture was incubated for 20 minutes at 37° C, the mixture was serially diluted using SM buffer, and spot tested as described above in (4.3.3).

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

4.3.6 Immune gold labelling

After the phage was purified, Immunogold labelling was carried out to localise the tail protein on the phage. In an Eppendorf tube, each antibody serum was diluted 1:100 using SM buffer and then incubated with 10^9 PFU/ml CDHS1 phage for 20 minutes at room temperature. Then the mixture was adsorbed to glow discharged, carbon-coated grids. Then the grids washed once with SM buffer for 10 minutes. Then a 1:30 diluted goat anti-rabbit IgG coupled with 12 nm gold colloids (Dianova, Hamburg) was added on the grids were left for 20 minutes. Then the grids were negatively stained with 1% (w/v) uranyl acetate before examination under the TEM.

4.4 Results

4.4.1 Investigation of the role of Gp18, Gp21 and Gp22 in CDHS1 phage adsorption

To determine which one of the three proteins (Gp18, Gp21 and Gp22) is the RBP for phage CDHS1 and is involved in the phage binding with *C. difficile*, polyclonal antibodies were raised against these proteins, and then the generated antibody was used to neutralise the phage infection.

4.4.1.1 Gp18 protein

To investigate whether Gp18 protein has a role in CDHS1 phage binding to *C. difficile* strains CD105LC1 and CDR20291, anti-Gp18 serum was pre-incubated with CDHS1 phage, thereafter the mixture was serially diluted and used in the spot assay to evaluate whether the anti-Gp18 serum can neutralise the phage infection or not. Figure 4-1A shows the result of spot test of phage CDHS1 when pre-incubated with anti-Gp18 serum. No significant difference was found with either the negative control (the pre-immune sera, the black column) or the positive control (anti-Gp18 serum, the grey column). This suggests that Gp18 is not responsible for CDHS1 binding to CD105LC1. These results were further confirmed using another strain CDR2029, as shown in Figure 4-1B. Again the result shows no significant differences between the negative control (the pre-immune sera represented by the blue column) and the positive control (purple column), thus confirming that the anti- Gp18 serum doesn't inhibit CDHS1 phage infection.

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

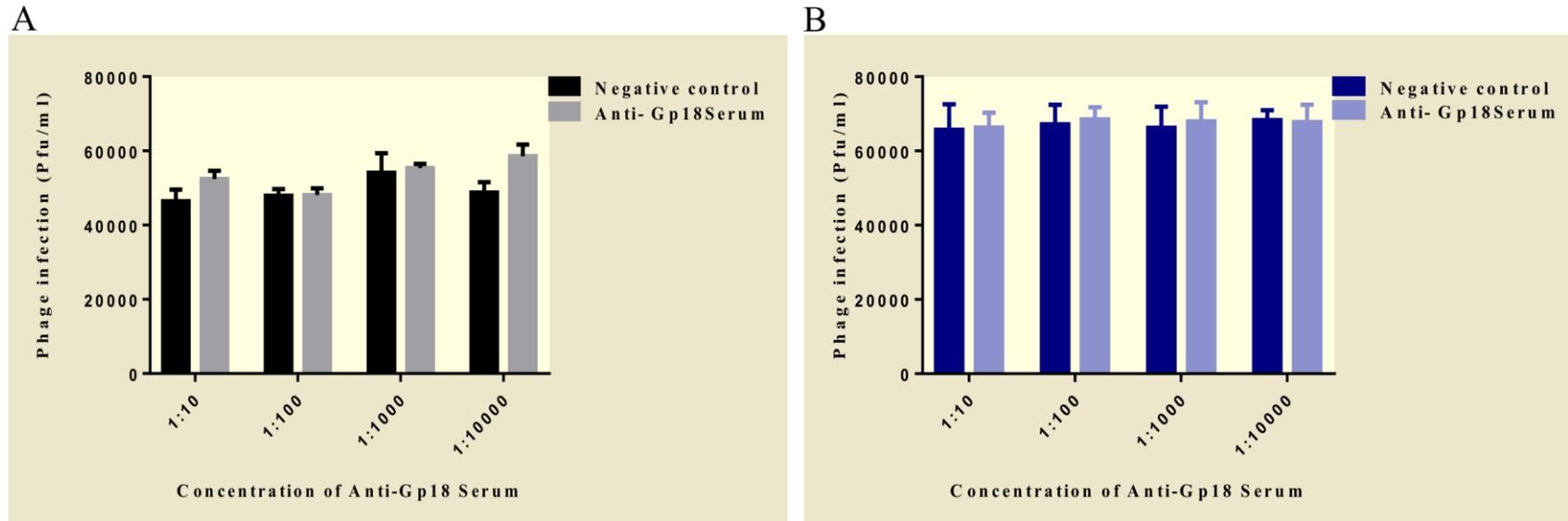


Figure 4-1 Investigation of the role of Gp18 protein in phage CDHS1 adsorption

(A) Shows the titres of phages following a spot assay analysis of phage CDHS1 on CD105LC1 strain incubation with concentrations of anti-Gp18 serum. The black column is the negative control, which was the pre-immune sera from rabbits where no anti-bodies were raised, the grey column is the Anti-Gp18 serum. (B) Shows the titres of phage following a spot assay analysis of phage CDHS1 on CDR2029 strain after incubation with different concentrations of anti-Gp18 serum, the blue column is the negative control and the purple column is the anti-Gp18. Error bars represent three biological repeats with three technical repeats. The data represents means \pm standard deviations (SD, $n = 3$) from the replicates. Statistical differences were calculated by two-way ANOVA, no significant differences between the positive control and the negative control were found.

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

4.4.1.2 Gp21 protein

To elucidate if the Gp21 protein has a role in CDHS1 phage binding with *C. difficile*, anti-Gp21 serum was tested as stated above with anti-Gp18 serum. Figure 4-2 demonstrates the result of the spot assay of CDHS1 after pre-incubation with anti-Gp21 on CD105LC1 and CDR2029 strains respectively. No significant differences between the negative control the pre-immune sera, (black column) and the positive control anti-Gp21 serum (grey column) was observed when CD105LC1 strain was used. The same result was obtained when CDR2029 was also used, no significant differences between the negative control and the positive control were found. This strongly suggests that the anti-Gp21 serum has no role in the inhibition of CDHS1 phage infection.

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

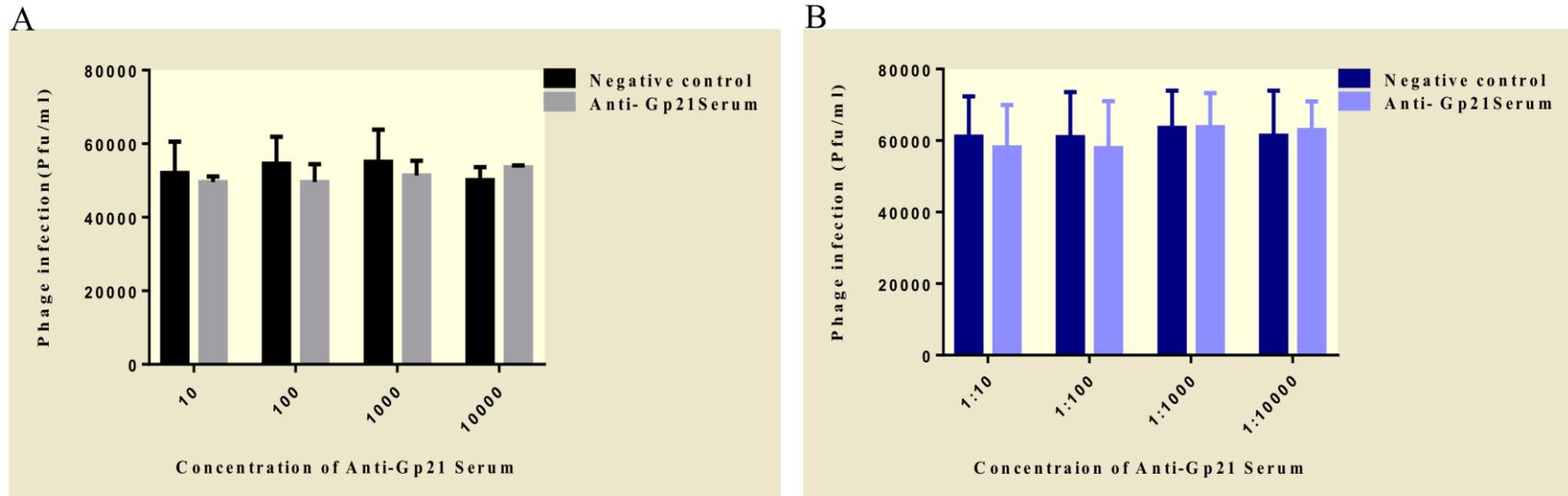


Figure 4-2 :- Investigation of the role of Gp21 protein in phage CDHS1 adsorption

The result demonstrated the spot assay of CDHS1 (after the pre-incubation with different concentrations of anti-Gp21) on CD105LC1 (A) and CDR2029 (B) strains. The black column in figure (A) represents the negative control whereas the grey column is the anti-Gp21 serum. (B) The blue column represents the negative control and the purple column is the anti-Gp21. The assay was conducted with three biological repeats and three technical repeats. Error bar represent means \pm standard deviations (SD, n = 3). Statistical differences were calculated via two-way ANOVA, no significant differences between the positive control and the negative control in both (A (and (B). The negative control was the pre-immune sera from rabbits where no anti-bodies were raised.

4.4.1.3 Gp22 protein

The Gp22 was the last protein to be tested to determine to see whether Gp22 plays a role in phage CDHS1 binding or not. To test this, anti-Gp22 serum was used to block the phage infection using the same approach that previously being used with Gp18 and Gp21. Figure 4-3 (A) shows the blocking assay carried out using anti-Gp22, the results show that the phage CDHS1 infection to CD105LC1 strain was completely inhibited with serial dilutions of 1:10, 1:100 and 1:1000 of anti-Gp22 serum. The phage started to infect at when a serial dilution of 1:10000 of anti-Gp22 was used, this indicating that Gp22 is the protein that may play an essential role for phage CDHS1 binding with the host strain. In Figure 4-3 (B), the blocking assay of phage CDHS1 infection to CDR20291 strain is shown, the result indicates that anti-Gp22 was able to block infection of phage CDHS1 to CDR20291 strain in the way same as in Figure 4-3 (A). This provides further evidence for the potential involvement of the Gp22 protein in phage CDHS1 attachment with *C. difficile* strains (CD105LC1& CDR20291). Therefore, the Gp22 protein is most likely the RBP for phage CDHS1. The pre-immune sera that used as negative control in this assay shows no inhibitory activity of the phage infection when compared to the anti-Gp22 serum.

Identifying the Receptor binding proteins for C. difficile phages CDHS1 and CDMH1

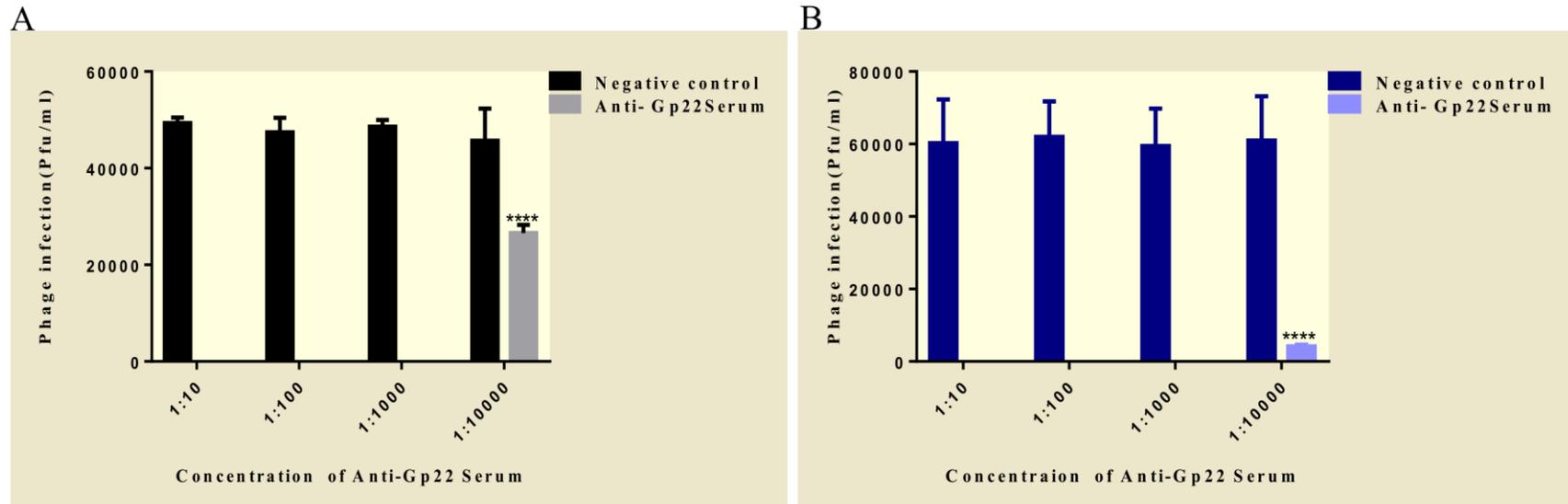


Figure 4-3 :- Neutralization of CDHS1 infection with rabbit anti-Gp22

The results demonstrate the inhibition assay of phage CDHS1 infection using different concentrations of anti-Gp22 serum. (A) Shows the inhibition of phage CDHS1 infection to CD105LC1 host, the black column represents the negative control whereas the grey column is the anti-Gp22 serum. (B) Shows the inhibition assay of CDHS1 infection to CDR2029, the blue column represents the negative control and the purple column represents the anti-Gp22. The assay was conducted with three biological repeats with three technical repeats. Error bar represent means \pm standard deviations (SD, n = 3). Statistical differences calculated by two-way ANOVA, significant differences between the positive control and the negative control was observed, P value= <0.0001. The negative control was the pre-immune sera from rabbits where no anti-bodies were raised.

4.4.2 Immunogold labelling of the baseplate proteins or tail proteins of CDHS1 using TEM

This assay was performed to identify the precise location of the tail proteins of CDHS1. To achieve this, the phage was pre-incubated with anti-serum antibody specific for the individual proteins (Gp18, Gp21 and Gp22). The mixture was then incubated with secondary anti-rabbit goat antibodies coupled with 12 nm gold colloids. The gold-labelled antibodies were displayed as black spots when viewed under TEM, as an indicator towards the location of the protein on the phage particle. In addition, as the negative control, CDHS1 was incubated with the pre-immune sera from rabbit where no anti-bodies were raised. Then the preparation was incubated with gold-conjugated anti-rabbit antibodies.

4.4.2.1 The location of Gp18 on CDHS1 phage

GP18 is one of the proteins involved in the phage CDHS1 baseplate structure, and to determine the specific location of Gp18 on CDHS1, anti-Gp18 serum was incubated with CDHS1 and then with secondary anti-rabbit goat antibodies coupled with 12 nm gold colloids. The result in Figure 4-4 (A, B and C) represents the location of Gp18 on CDHS1 indicated by the black spots. The position of gold-labelled antibodies (black spots) is found to be at the base of the CDHS1 phage, which is the location of the Gp18 protein. Whereas in Figure 4-4 (D, E and F), shown is the CDHS1 incubated with the pre-immune sera and then with the gold conjugated goat antibodies as a negative control. The result shows random scattering of black spots on the grids when viewed under the microscope. This indicates that the labelling of CDHS1 Gp18 protein occurs only in the presence of the anti-Gp18 serum (the primary antibody)

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

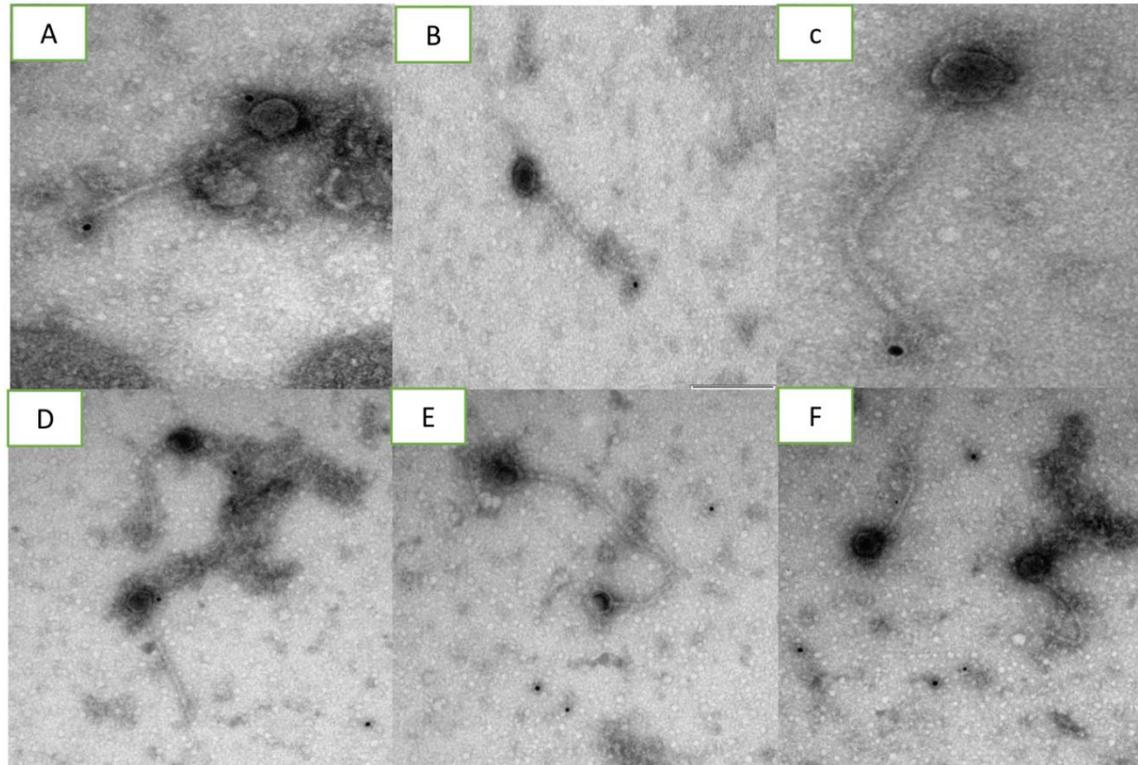


Figure 4-4 :- Immunogold labelling of tail protein Gp18

Transmission electron microscopy (TEM) image of negatively stained phage CDHS1 after immunogold labelling with Anti-Gp18 serum. A, B & C are the positive control which is Anti-Gp18 serum, D, E & F are the negative control (the pre-immune sera). CDHS1 was incubated with polyclonal anti-GP18 rabbit antibodies raised against the Gp18 protein and labelled with anti-rabbit secondary goat antibodies coupled with 12 nm gold.

Identifying the Receptor binding proteins for C. difficile phages CDHS1 and CDMH1

4.4.2.2 The location of Gp 21 on CDHS1 phage

To determine the specific location of Gp21 on CDHS1, the same approach that was carried out with anti-Gp18 was applied with anti-Gp21. In Figure 4-5 (A, B and C) the result shows that gold-conjugated antibodies (black spots) were found bound to the baseplate of CDHS1 phage, which is where the protein Gp21 is located. Whereas Figure 4-5 (D, E and F) shows the result of the negative control, where CDHS1 was incubated with the pre-immune sera and then with the gold conjugated anti rabbit goat antibodies. The result shows random black spots on the grids when viewed under the microscope. Indicating the localization of CDHS1 Gp21 protein occurs only in the presence of the polyclonal anti-GP21 serum.

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

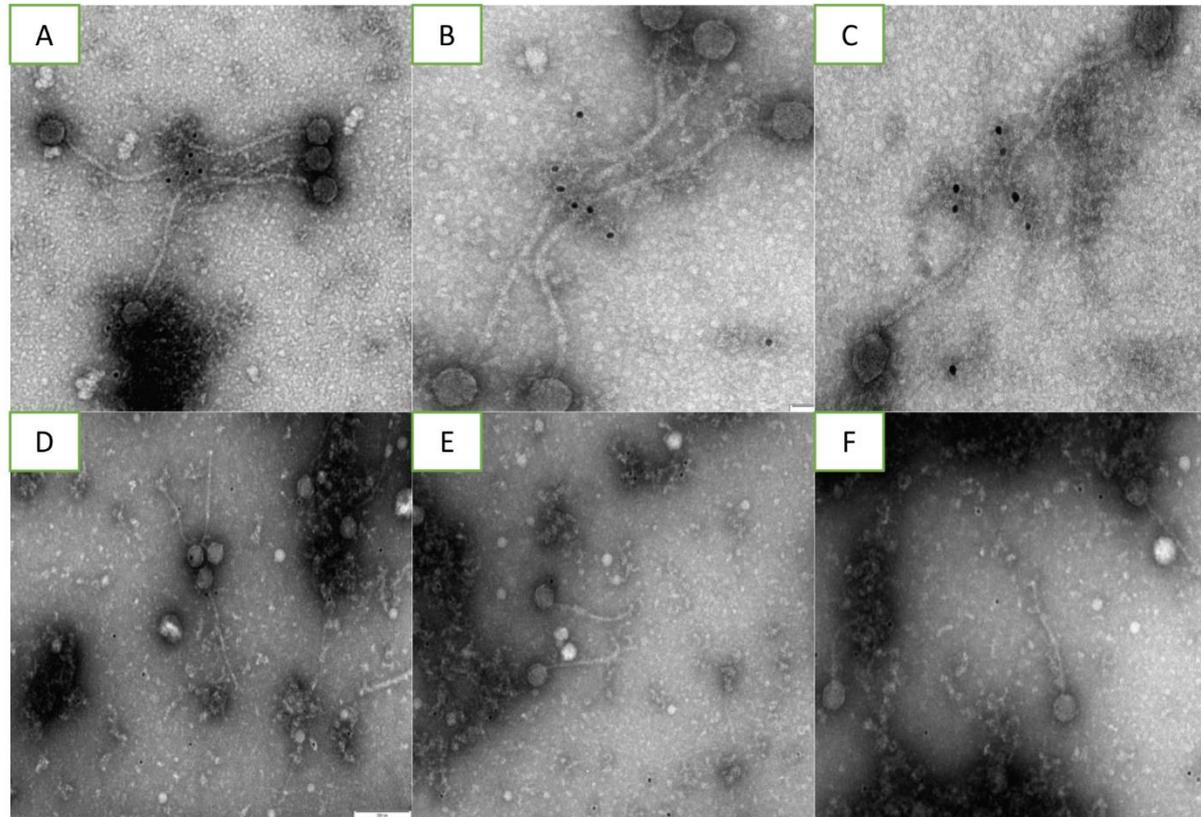


Figure 4-5 Immunogold labelling of tail protein Gp21

Transmission electron microscopy (TEM) image of negatively stained phage CDHS1 after immunogold labelling with Anti- Gp21 serum. A, B&C are the positive control which is Anti- Gp21 serum, D, E& F are the negative control. CDHS1 was incubated with polyclonal anti-GP21 rabbit antibodies raised against the Gp21 protein and labelled with anti-rabbit secondary goat antibodies coupled with 12 nm gold.

4.4.2.3 The localization of Gp22 on CDHS1 Phage

For the Gp22 protein localization on CDHS1, the result in Figure 4-6 (A, B and C) shows that gold-conjugated antibodies (black spots) bound to the structures associated with the baseplate (lower baseplate) of CDHS1 phage representing the specific location of Gp22 protein. Whereas in Figure 4-6 (D, E and F), where the negative control shows CDHS1 incubated with the pre-immune sera and then with the gold conjugated anti rabbit goat antibodies. The results show that there is no labelling on CDHS1 phage particle.

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

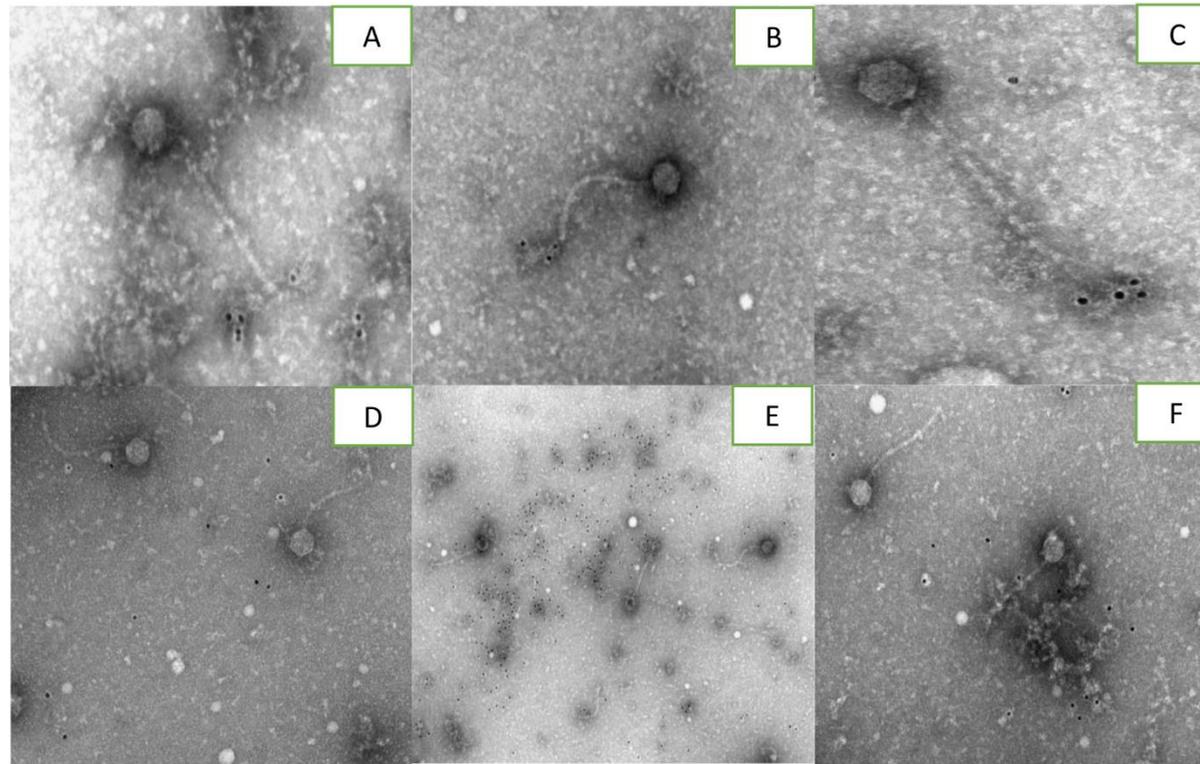


Figure 4-6 Immunogold labelling of tail protein Gp22

Transmission electron microscopy (TEM) image of negatively stained phage CDHS1 after immunogold labelling with Anti- Gp22 serum. A, B & C show the positive control which is anti-Gp22 serum. Figures D, E & F show the negative control. CDHS1 was incubated with polyclonal anti-Gp22 rabbit antibodies raised against the Gp22 protein and labelled with anti-rabbit secondary goat antibodies coupled with 12 nm gold colloids.

Identifying the Receptor binding proteins for C. difficile phages CDHS1 and CDMH1

4.4.3 Investigation the role of Gp29 and Gp30 in CDMH1 phage adsorption

The same approach was applied to the myovirus as was taken with the siphovirus, and thus the same approach to identifying the RBPs for CDMH1 was carried out.

4.4.3.1 Gp29 protein

In order to determine whether Gp29 protein can inhibit CDMH1 phage infection or not, different dilutions of anti-Gp29 serum was incubated with CDMH1 phage. Thereafter, the mixture was spotted on to a lawn of CD105HE1 strain. Figure 4-7 represents the inhibition assay of CDMH1 phage infection using Anti-Gp29 serum, the black column represents the negative control which is the pre-immune sera and the grey column is the anti-Gp29 serum. The result shows that anti-Gp29 was able to block the infection of the phage, indicating that Gp29 protein is the protein that most likely acts as the RBP for the phage CDMH1, thus having a key role in phage-bacterial attachment.

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

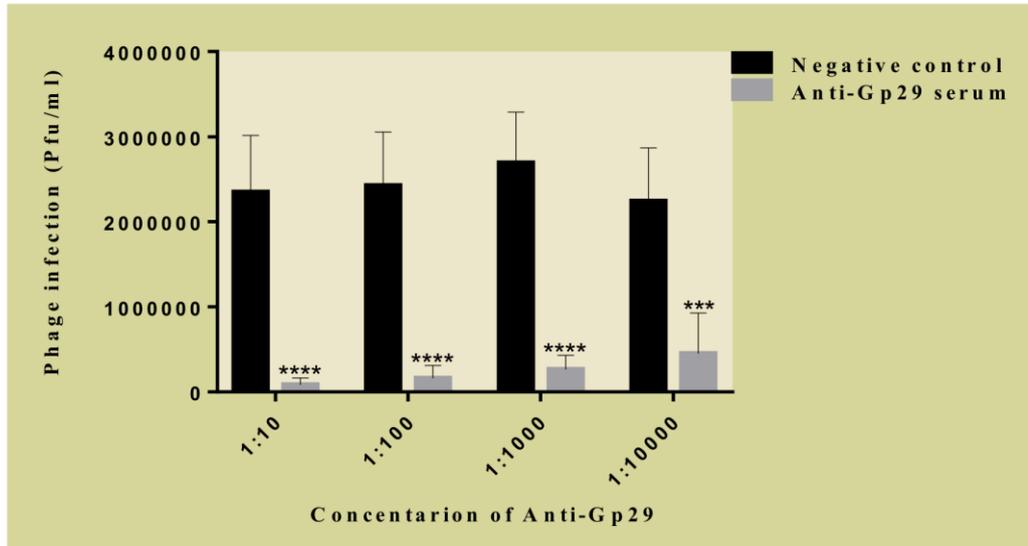


Figure 4-7 Neutralization of CDMH1 infection with rabbit anti-Gp29

The data in this graph shows the inhibition assay of phage CDMH1 infection using different concentrations of anti-Gp29, the black column represents the negative control (pre-immune sera) whereas the grey column is the anti-Gp29 serum. The assay was conducted with three biological repeats and with three technical repeats. Error bar represent means \pm standard deviations (SD, n = 3). Statistical differences calculated by two-way ANOVA, significant differences between the positive control and the negative control was observed, P value= <0.0001.

4.4.3.2 Gp30 protein

The second candidate protein that was targeted in this assay to identify its role in phage binding was the Gp30 protein. The result demonstrated in Figure 4-8 shows that there is no significant difference between the negative controls (pre-immune sera from rabbits where no anti-bodies were raised (Black column)) compared to the positive control (grey Column), indicating that Gp30 has no role in phage attachment with the *C. difficile* strain used.

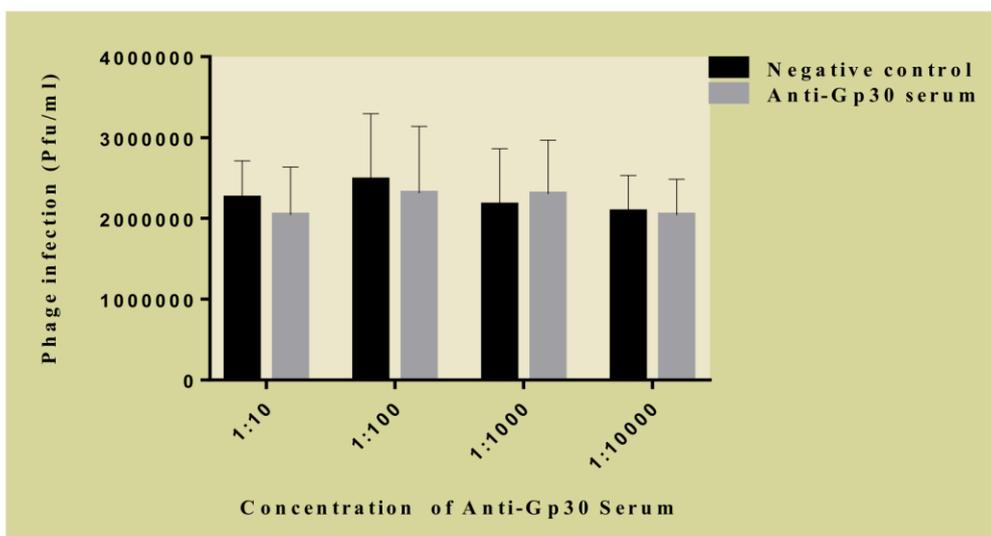


Figure 4-8 Neutralization of CDMH1 infection with rabbit anti-Gp30

The result demonstrates the inhibition assay of phage CDMH1 infection using different concentrations of anti-Gp30 serum, the black column represents the negative control and the grey column is the anti-Gp30 serum. The assay was conducted with three biological and technical repeats. Error bar represent means \pm standard deviations (SD, n = 3). Statistical differences calculated by two-way ANOVA, no significant differences between the positive control and the negative control. The negative control was the pre-immune sera from rabbits where no anti-bodies were raised.

4.4.4 Neutralization assay of phage CDHM3 & CDHM6 infection

CDHM3 and CDHM6 are phages that are members of the myovirus family, both infect *C. difficile* strain CD105HE1 in similar manner to phage CDMH1. So, it was worthy to test if Gp29 would be able to block the phage infection of these two phages using the polyclonal antibodies raised against the CDMH1 tail proteins.

4.4.4.1 Neutralization assay of phage CDHM3 infection using anti-Gp29 and anti-Gp30 serum

To establish if the anti-Gp29 and anti-Gp30 serum is able to block CDHM3 infection to strain CD105HE1, similar method that were applied with CDMH1 phage were also applied here. The results presented in Figure 4-9A shows the blocking assay using anti-Gp29 serum, the results illustrate that the anti-Gp29 serum was able to block the phage CDHM3 infection significantly, until a 1:1000 serum dilution was used. Whereas when anti-Gp30 was used, the result in Figure 4-9B shows that is no significant difference between the negative and positive controls.

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

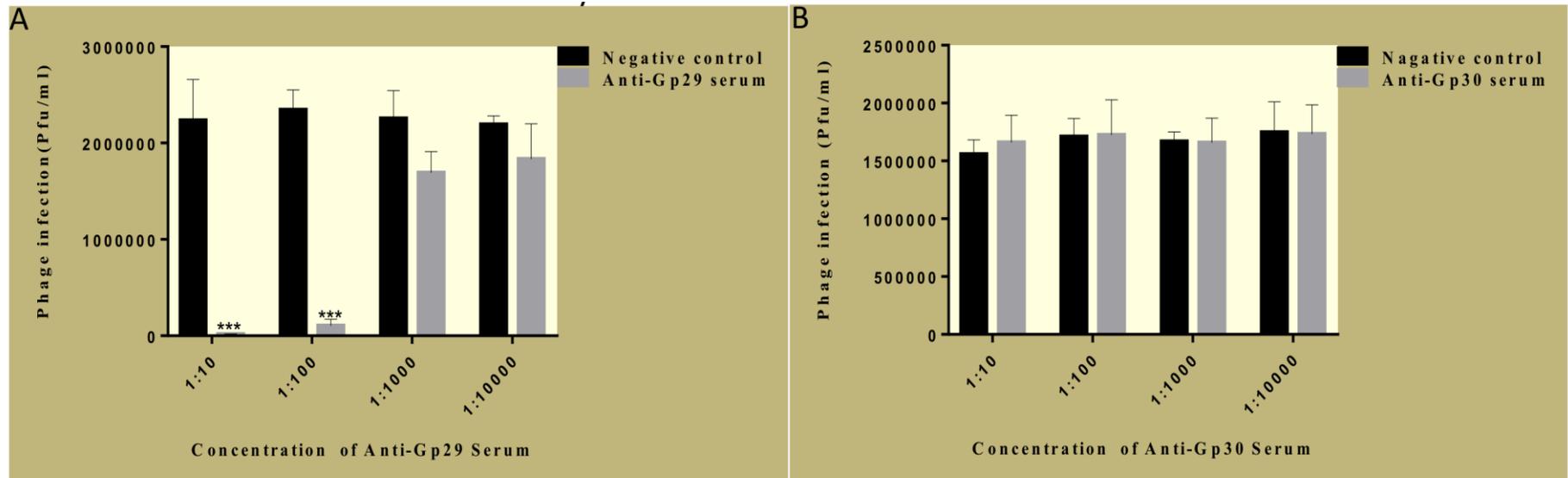


Figure 4-9 Neutralization of CDHM3 infection with rabbit anti-Gp29 & anti-Gp30

The result in graph (A) shows the inhibition assay of phage CDHM3 infection using different concentrations of anti-Gp29, the black column represents the negative control whereas the grey column shows the anti-Gp29 serum. (B) Shows the inhibition assay of phage CDHM3 infection using different concentrations of anti-Gp30 serum, the black column represents the negative control and the grey column represents the anti-Gp30 serum. The assay was conducted with three biological and three technical repeats. Error bar represent means \pm standard deviations (SD, $n = 3$). Statistical differences calculated by two-way ANOVA, significant differences between the positive control and the negative control, P value= <0.0001 in graph (A) whereas in graph (B) no significant differences between the positive control and the negative control. The negative control was the pre-immune sera from rabbits where no anti-bodies were raised.

4.4.4.2 Neutralization Assay of phage CDHM6 infection using Anti-Gp29 and Anti- Gp30 serum

To determine if the anti-Gp29 and anti-Gp30 serum can neutralise the CDHM6 infection to strain CD105HE1, a similar approach to the one used in the neutralisation assay of CDMH1 and CDHM3 infection was used. In Figure 4-10 (A) the result illustrates that the anti-Gp29 serum was able to block the phage CDHM6 infection too. However, when anti-Gp30 was used to block the phage CDHM6 infection, in Figure 4-10 (B) it shown that there were no significant difference between the negative and positive control. Indicating the anti-Gp30 was not able to block the phage infection.

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

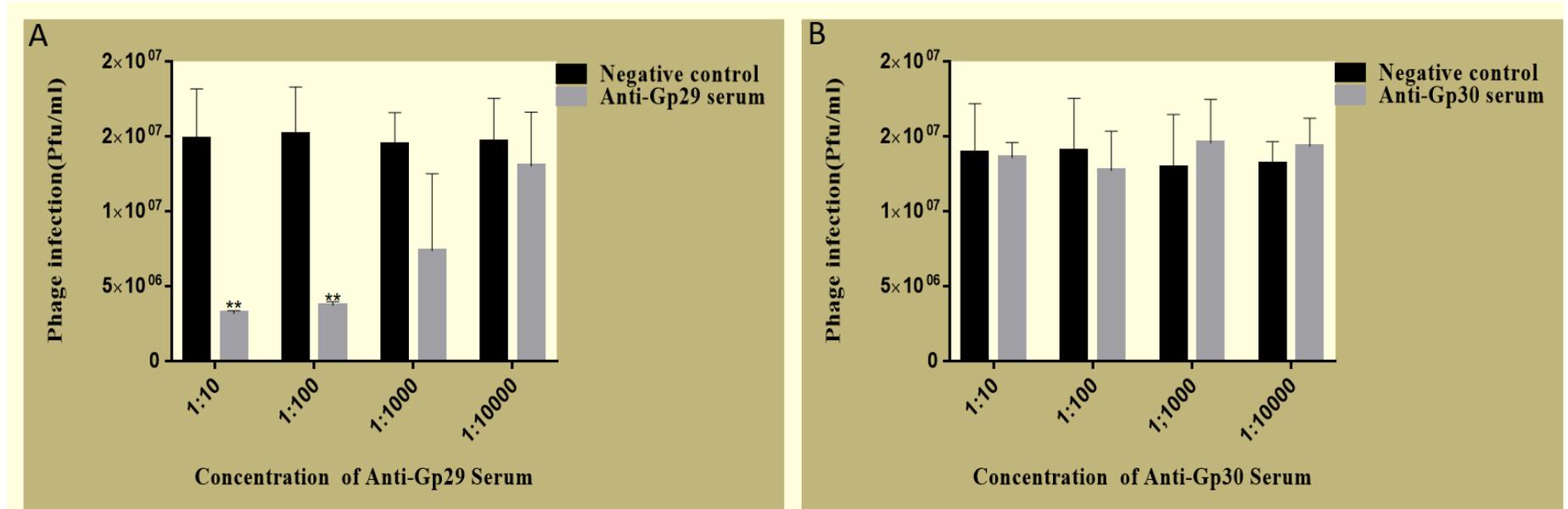


Figure 4-10 Neutralization of CDHM6 infection with rabbit anti-Gp29 & anti-Gp30

The result in graph (A) demonstrates the inhibition assay of phage CDHM6 infection using different concentrations of anti-Gp29, the black column represents the negative control whereas the grey column represents the anti-Gp29 serum. (B) Inhibition assay of phage CDHM6 infection using different concentration of anti-Gp30 serum, the black column represents the negative control and the grey column is the Anti-Gp30 serum. The assay was conducted with three and technical repeats. Error bar represent means \pm standard deviations (SD, n = 3). Statistical differences calculated by two-way ANOVA, significant differences between the positive control and the negative control, P value= <0.0001 in graph (A) whereas in graph (B) no significant differences between the positive control and the negative control. The negative control was the pre-immune sera from rabbits where no anti-bodies were raised.

4.5 Discussion

Phage-bacterial attachment is a key process to allow the phage to successfully infect the bacterial host. In this process, the phage uses the receptors binding proteins (RBPs) present at the end of the phage tail to attach to the host. As previously stated in the introduction (Chapter 1), these RBPs come in the form of tail fiber proteins or are present within the phage baseplate (Mahony *et al.*, 2016a). The mechanism of action by which the phages infect their hosts is poorly understood, especially with phages that infect Gram-positive bacteria (Bielmann *et al.*, 2015). Though, significant effort has been exerted in trying to identify and characterise the RBPs for phages that infect Gram-positive bacteria. However, the focus previously has mainly been on phages that infect *L. lactis*. Therefore, recently there has been an increased interest in studying and characterising the RBPs of phages that infect Gram-positive bacteria such as *Listeria* and *Staphylococcus* (Bielmann *et al.*, 2015, Li *et al.*, 2016b).

In this chapter, the receptor binding proteins of two phages was identified; the Siphovirus CDHS1 that infects *C. difficile* strains CD105LC and CDR2029 (Ribotype 027), and the myovirus CDMH1 that infect *C. difficile* CD105HE1 strain. This work is significant as it is the first study to have identified the RBPs for phages that infect *C. difficile*. Studying and characterising the RBPs of phages that infect *C. difficile* will enable us to improve our understanding by which the phages infect this organism.

4.5.1 Identification of RBPs of CDHS1 that infect *C. difficile*

The approach used to identify the RBPs for this phage was to over-express four predicted phage tail proteins Gp18, Gp19, Gp21 and Gp22 (identified via Bioinformatic analysis as having homology with other tail proteins from siphoviruses that infect other Gram-positive bacteria). After significant optimisation of all experimental parameters, three proteins (GP18, Gp21 and Gp22) were expressed, purified, and polyclonal antibodies were generated against them. The antibodies were then used to neutralize phage infection. The main significant finding of this chapter was that anti-Gp22 serum was the serum that blocked CDHS1 phage infection indicating that the corresponding protein Gp22 is the protein that possibly is responsible for phage CDHS1 binding with *C. difficile*, which means that Gp22 protein is the potential RBPs for CDHS1. Whereas, the other anti-serum antibodies tested in the neutralization assays for phage CDHS1 infection showed no blocking in the phage infection, indicating that Gp18, Gp21 proteins have no role in phage CDHS1 binding with its host.

Many Siphoviruses of Gram-positive bacteria have had their RBPs identified and characterised but this is the first time that a RBPs has been identified for a *C. difficile* phage.

In the literature, many approaches have been utilised to determine the RBPs of different Siphovirus as discussed in the introduction of this chapter. For example, the method used in the present study was to produce antibodies against the purified tail fiber protein, this method has been previously used to identify the RBPs of phage ϕ 11 that infects *S. aureus* (Li *et al.*, 2016b). An alternative method was used to identify the RBPs of other phages, this via the generation of a chimeric phages that possess an alternative RBP encoding genes. Subsequently, this can result in changing the host range of the wild type phages. Such method was used in identifying the RBPs of 936 group phages sk1 and bIL170 and

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

the P335 phages TP901-1 and Tuc2009 that infect *L. Lactis* (Dupont *et al.*, 2004, Vegge *et al.*, 2006).

From the studies and characterizations of the RBPs from different other phages that infect other Gram-positive bacteria (Bebeacqua *et al.*, 2010, Biemann *et al.*, 2015, Li *et al.*, 2016b). It has been highlighted before that genes encoding tail proteins or proteins involved in phage baseplate formation are located downstream to the gene encoding the tape major protein Tmp, and upstream to the genes encoding holin and endolysin proteins (Bielmann *et al.*, 2015). Amongst these genes, it has been found the gene encoding the RBPs is positioned directly up stream to the genes encoding holin and endolysin. Interestingly, the location of the gene encoding Gp22 protein is located directly up stream of the genes encoding holin and endolysin in CDHS1 phage genome. Indicating there could be conservation of the genomic architecture of these genes within different siphoviruses that infect different Gram-positive bacteria (Li *et al.*, 2016b).

4.5.2 Immunogold labelling of the baseplate proteins or tail proteins of CDHS1 using transmission electron microscope

To deduce the location of the tail protein or the protein involved in baseplate structure of CDHS1 phage, immunogold labelling of CDHS1 tail protein was under taken. The labelling of CDHS1 using anti-Gp18 anti serum demonstrated that the corresponding protein GP18 (which presents homology to the distal tail protein (Dit) in several siphoviruses such as TP901-1 of *L. Lactis* and A118 of *L. monocytogenes*) is located in the base of phage CDHS1, as shown in Figure 4-4. However, for protein Gp19 of CDHS1, the hypothetical tail protein for this phage was not tested in this project due to the difficulty of purifying this protein.

The labelling using anti-Gp21 anti serum showed that the Gp21 protein is located at the baseplate of this phage, as shown in Figure 4-5. The immunogold labelling using the anti-

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

Gp22 antiserum showed that Gp22, the strongest RBPs candidate of phage CDHS1 identified in this study, is located at the end of the baseplate, or it is representing the structure associated with the baseplate as shown in Figure 4-6. In addition to this, performing the immunogold labelling using a negative control where pre-immune sera was used, the results showed that the gold conjugated antibody was not able to label the phage. This is an indication that the primary antibodies must be present for the gold conjugated antibody to mark the phage CDHS1. Also, it is further confirmation that the corresponding protein is in that position.

Immuno-electron microscopy has been used to characterise the baseplate in several phages, for example for phages sk1, TP901-1 and Tuc2009 that infect *L. Lactis*, in addition to A118 and P35 which are phages specific for *L. monocytogenes* (Dupont *et al.*, 2004, Vegge *et al.*, 2006, Biemann *et al.*, 2015). The results of these studies have identified that the location of the genes encoding the tape measure protein (Tmp), Dit, Tail associated lysin (Tal) and the upper baseplate is conserved between the majorities of siphoviruses that infect other Gram-positive bacteria. That may suggest, there may be a cross link between these proteins (Bielmann *et al.*, 2015). It is also been demonstrated that, a TP901-1 phage mutant that lacked Tmp, Dit and Tal proteins lead to the formation of tail-less phage particles indicating the essential role of these proteins in phage tail formation and assembly (Vegge *et al.*, 2005).

Taking into consideration the conserved genomic architecture of the genes encoding Tmp, Dit and Tal proteins as well as the similarity of the amino acid sequences of these proteins in different Siphovirus that infect Gram-positive bacteria, it can be predicted that mechanism of the tail assembly is similar between these phages (Veesler *et al.*, 2010). Therefore, the location of these protein on the phage particle might be similar too. For instance, the Dit protein (corresponding to Gp18 in the case of CDHS1) is the protein

Identifying the Receptor binding proteins for *C. difficile* phages CDHS1 and CDMH1

responsible for the central hub formation in the baseplate of several Siphoviruses that have been characterised. Using Immunoelectron microscopy for A118 of *L. monocytogenes*, the precise D11 protein location was found below the interconnection of the tail tube and tail tip or the baseplate of the phage (Bielmann *et al.*, 2015).

The immunogold labelling for protein Gp21 (from CDHS1) which presents with homology to the upper base plate protein ORF48 (BppU) of TP901-1 that infects *L. Lactis*, was shown to be located at the baseplate of CDHS1 phage. The ORF48 (BppU) of TP901-1 from *L. Lactis* as well as the A118 Gp20 (homology to BppU) of *L. monocytogenes* are located at the upper part of the baseplate of the phage. As the structure of the baseplate from those two phages is composed of two parts; the upper part and the lower part that contains the RBPs (Mahony *et al.*, 2012, Bielmann *et al.*, 2015). Whereas, the RBPs of CDHS1 (Gp22) was located at the end of the baseplate, specifically at the structures associated with baseplate of CDHS1.

4.5.3 Identification of RBPs of CDMH1 that infect *C. difficile*

The approach that was used to identify the RBPs for *C. difficile* phage CDMH1 was similar to the one used to identify the RBPs of phage CDHS1. The main finding was that the anti-Gp29 serum was able to block the phage infection of CDMH1 phage to its propagation host CD105HE1, therefore the corresponding protein Gp29 could be the RBP for this phage. On the other hand, no neutralization occurred whilst using anti-Gp30 serum indicating that the Gp30 protein is not involved in phage attachment.

Within the literature, the research conducted for the characterization and research into the RBPs of myoviruses that infect Gram-positive bacteria is significantly less than the work that has been conducted for siphoviruses RBPs. Phage A511 that infects *L. monocytogenes* one of the myovirus whose RBPs has been identified (Habann *et al.*, 2014). The approach that was used for the identification of the RBPs for phage A511 was

Identifying the Receptor binding proteins for C. difficile phages CDHS1 and CDMH1

similar to the one applied in this project. For phage A511, five genes that are located between the *tmp* gene and Helicase were targeted for the study. These genes were amplified, cloned, and the resultant construct was transformed to an expression vector. The overexpressed proteins were purified and used for the production polyclonal antibodies, thereafter, these antibodies were then used to neutralise the phage infection (Habann *et al.*, 2014).

The main finding was that two proteins, Gp98 and Gp108 of phage A511 were found to be involved in phage A511 attachment. HHpred analysis revealed that the C terminal of Gp98 has a conserved peptidoglycan hydrolase domain which assists in host cell wall degradation, consequently, the phage will be able to inject its DNA material (Habann *et al.*, 2014). Whereas in the case of CDMH1, a blast protein search showed that the putative protein Gp24 CDMH1 has a conserved cell wall hydrolases domain at the C terminal of the protein, this has a role in peptidoglycan degradation. This knowledge can be used to hypothesize that Gp24 could be involved in the attachment of CDMH1 to its host. However, this hypothesis needs to be confirmed by applying the same approach that was used with Gp29 & Gp30 of CDMH1.

However, for phage A511, protein Gp108 alongside Gp106 protein were the most likely candidates as being RBPs, since RBPs are the least conserved component of the phage baseplate. Moreover, RBPs binding site are usually located at the C terminal of the protein. These characteristics are found in Gp108. Which has been confirmed to be the RBP for A511 phage (Habann *et al.*, 2014). RBPs characteristics that are mentioned above were the main reason to choose Gp29 and Gp30 proteins to be the most likely the RBPs for phage CDMH1. The neutralization assay revealed that Gp29 is a strong possible candidate as the RBPs of phage CDMH1.

4.5.4 Determining the ability of anti-Gp29 to neutralize CDHM3 and CDHM6 phages that infect CD105HE1

There are several *C. difficile* phages that have been isolated in Professor Martha Clokie's laboratory, which along with CDMH1, use CD105HE1 as a manufacturing host. Therefore, it was worthy to examine if anti-Gp29 can prevent the infection of these phages to CD105HE1. The two other phages that were tested are CDHM3 and CDHM6, both of which infect CD105HE1 strain. Anti-Gp29 serum was incubated with CDHM3 and CDHM6 in the same way as with CDMH1, the results show that anti-Gp29 serum could block the infection by CDHM3 and CDHM6 indicating that the three phages may share the same RBPs.

It has been proven by (Nale *et al.*, 2016) that, the three phages CDMH1, CDHM3 and CDHM6, have the ability to infect several ribotypes of *C. difficile* with same efficiency such as RT018, RT 014/ 020 (Nale *et al.*, 2016). Taking in to the account that these three phages have similar host ranges, along with the fact that these phages are propagated in the same host manufacturer. Further study must be done to confirm the findings, such as using another host that these phages infect and to test the ability of these phages to infect in the presence of anti-Gp29 serum.

Comparable results were found in different systems, when the antibodies raised against the RBPs (ORF18) of phage P2 that belongs to *Lactococcus* phage group (936). These antibodies were used to neutralise the infection of the phages that have similar host ranges to phage P2 and belong to the same phage group (936). The results obtain from such cross reactivity show that the antibody was able to neutralize the infection of 936-like phages as efficient as when phage P2 was used. Which leads to the conclusion that those phages share the same RBPs (De Haard *et al.*, 2005).

4.6 Conclusion

In this chapter the aim was to identify the RBPs of two phages (CDHS1 & CDMH1) that infect *C. difficile* and used in this project. The main findings of this chapters were

- 1- The polyclonal antibodies that were raised against the tail proteins of both phages were successfully used to determine which antibody could neutralise phage infection
- 2- For phage CDHS1, the anti-Gp22 serum inhibited phage infection and thus it can be deduced that Gp22 is most likely to be the RBPs.
- 3- For CDMH1, anti-Gp 29 protein inhibited the phage infection so is the likely RBP for this phage.
- 4- Immunogold labelling assay was conducted to validate the neutralisation studies and to locate the tail proteins on the phage particles.
- 5- Anti-Gp29 was tested against another two phages CDMH3 and CDMH6 that infect the same host as CDMH1, and have similar host ranges. It was shown to inhibit CDMH3 and CDMH6 phages suggesting that these three phages may have the same RBPs, however, this finding require further confirmation.

Chapter 5 **Structural characterization of the
protein Gp22**

5.1 Introduction

The art of protein crystallisation in biology is one of the methods to determine the structure of proteins. The first attempt of protein crystallization was carried out 170 years ago by Friedrich Ludwig Handfield, when he accidentally crystallised haemoglobin from the blood of the earth worm (Holcomb et al., 2017). Since then, protein crystallisation has often been used as a way of characterising proteins (Giege, 2013). The principle of protein crystallization is based on the supersaturation state of a protein in solution. Using chemicals known as precipitants, the supersaturation of protein can be achieved by reducing its solubility (Holcomb et al., 2017).

Four stages can be considered in the crystallisation process:

1) Unsaturated stage - The protein and the precipitant are too low for nucleation or crystal growth. 2) Metastable stage - Where protein growth can occur, but is too dilute for nucleation. 3) Labile stage - Where both nucleation and crystal growth occurs. 4) Precipitation – This is when disordered aggregates form, as shown in Figure (5-1) (Giege, 2013, Holcomb et al., 2017)

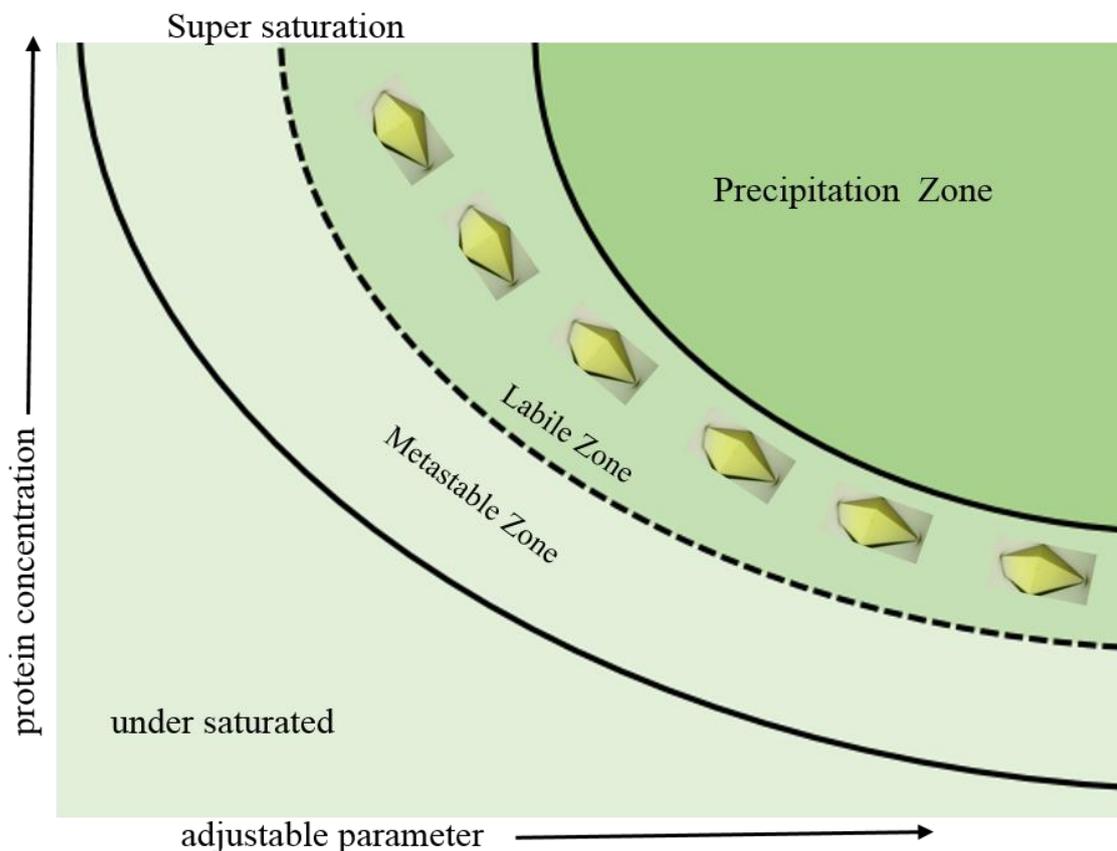


Figure 5-1 Schematic diagram of protein crystallization phase

The solubility diagram is divided in two regions, the under saturated region and the super saturated. The super saturated region is further divided into three Zones: the metastable, the labile, and finally precipitation the zone which occurs at a high super saturated stage.

There are many methods that are used to obtained protein crystallisation such as; vapour diffusion, batch crystallization, and liquid-liquid diffusion (Holcomb et al., 2017). The most common method is the vapour diffusion. This method can be achieved in two different way; hanging-drop vapour diffusion or sitting-drop vapour diffusion (McPherson et al., 2014b). Both methods rely on the equilibration of the protein and the precipitants against a reservoir solution (which is solution basically composed of the precipitants) (Holcomb et al., 2017).

The method used in this chapter is the sitting-drop vapour diffusion. In this method, a drop of protein and precipitant mixture are placed in a well above a reservoir solution.

Structural characterization of the protein Gp22

And the supersaturation in this method is achieved by evaporating the water from the drop of the mixture (protein and precipitant) in to the reservoir solution leading to the concentration of the proteins and precipitant, subsequently, leading to the nucleation and growth of the crystal (McPherson et al., 2014b).

X-ray crystallography was introduced in 1934 (Giege, 2013). It is the most common approach to determine the three-dimensional structure of a protein. This requires the production of high quality protein crystals (Dale et al., 2003). The determination of protein structures to atomic resolution enhances our understating of how proteins function, in addition to unravelling the mechanisms by which those proteins or macromolecules work (Dessau et al., 2011).

X- ray diffraction patterns are a fingerprint of the atomic arrangement inside a given sample. It occurs by the constructive interference of X-ray beams that are scattered at specific angles from each set of lattices within the protein structure, this further explained in Figure 5-2 (Bunaciu et al., 2015).

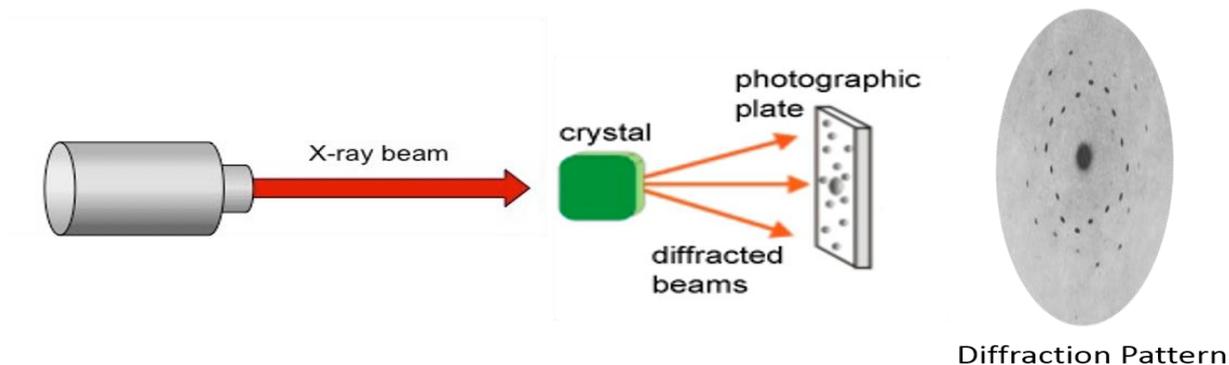


Figure 5-2 Schematic diagram represent the X- ray diffraction

As the protein crystal is hit with the X - ray beam, the beam is scattered. This scattering or diffraction pattern is a fingerprint of the arrangement of the atoms within the protein.

Structural characterization of the protein Gp22

X-ray scattering can be determined by Bragg's law: $n\lambda = d\sin\theta$

Where n is an integer, λ is the wavelength of the X-rays, d is the spacing between the planes in the atomic lattice, and θ is the angle between the incident ray and the scattering planes. Bragg's law is based on the relation between the wavelength of the X ray, the diffraction angle and the lattice spacing in a crystalline sample.

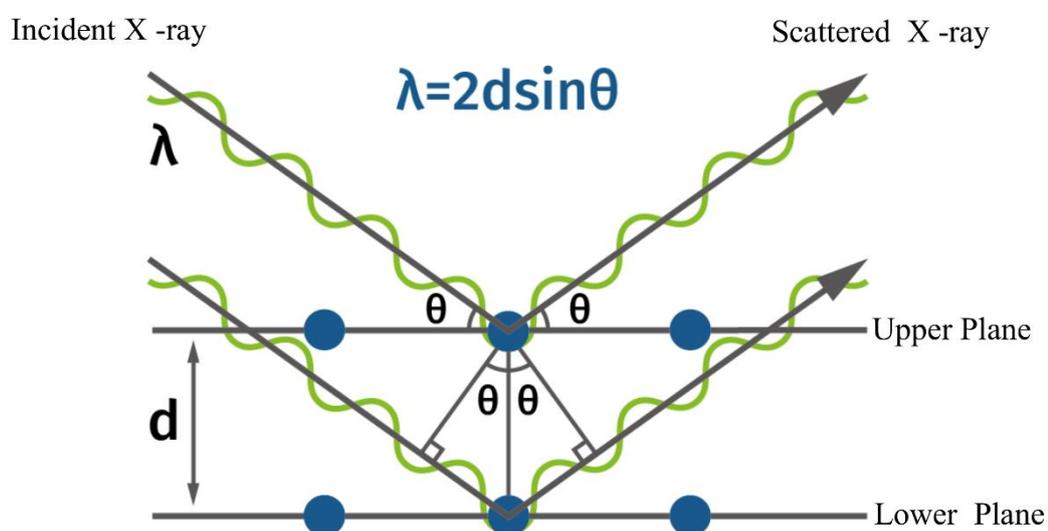


Figure 5-3: Schematic representation of Bragg's law

This law defines the definite relationship between the angles at which a beam of x rays must fall on the parallel planes of atoms in a crystal in order for there to be a strong reflection.

Structural characterization of the protein Gp22

Gp22 was shown to be the Receptor binding proteins (RBPs) for phage CDHS1. The aim in this chapter was to determine the three-dimensional structure of Gp22 to gain an insight of how this protein functions and its mechanism of action. Therefore, Gp22 was expressed and purified, and further crystallised using a commercial crystallization screen.

Bioinformatic analysis of Gp22 protein revealed no structural similarity with any other protein in the protein data bank (PDB). Consequently, the phases cannot be determined by molecular replacement (Driessen, 1996). Instead, selenomethionine was incorporated, enabling the phases to be solved by single-wavelength anomalous dispersion.

Selenomethionine labelling is a common practice used in phase's determination in protein crystallography, by using single or multiwavelength anomalous dispersion (Jones, 2007). To obtain a near to complete replacement of methionine to selenomethionine, different techniques can be applied based on the *E.coli* strain used, either using a *E.coli* methionine auxotrophic strain such DL41, or by inhibition the methionine biosynthesis pathway. In this case, any *E.coli* strain can be used to obtain the purpose (Doubl  , 2007). In this chapter, Gp22 protein was labelled by selenomethionine using the second method where the methionine biosynthesis pathway was inhibited.

5.2 The Aim

The aim in this chapter was to determine the three-dimensional structure of Gp 22 (the potential RBPs) from phage CDHS1.

To achieve this aim:

- 1- Gp22 proteins was cloned, expressed and purified.
- 2- Selenomethionine-Gp22, a Selenomethionine derivative of Gp22 was expressed and purified.
- 3- The purified Gp22 and selenomethionine-Gp22 were concentrated and used in the sitting-drop vapor diffusion method to crystallize the Gp22 protein. For this purpose, different commercial crystallization screens were used such as PACT, JCSG, and Proplex (Molecular Dimensions, UK).

5.3 Methods

5.3.1 Expression and purification of Gp22 protein

Expression and purification of Gp22 protein was performed as describe in chapter 3 sections 3.3.2.1 and 3.3.2.2.2.

5.3.2 Expression of Selenomethionine- Gp22 protein

Expression of Selenomethionine-Gp22 performed by the inhibition of the methionine pathway. Selenomethionine labelling is a powerful phasing technique used to help in the determination of the protein structure. The process to express Selenomethionine-Gp22 protein conducted as explained below.

Selenomethionine Medium Complete (M9) (Molecular Dimensions, UK) was prepared according to the manufacturer's guide. In brief, 21.6g of seleno Met medium base was dissolve in 1 L of distilled water and autoclaved. Then 5.1g of seleno Met nutrient mix was dissolve in 50 ml of distilled water, sterilised and then added to the base medium to make up the methionine minus medium. When the medium was ready, 4 ml of Selenomethionine solution was added per litre of medium. Then, the medium was supplemented with 0.15 mg/ml of Carbenicillin.

To express Selenomethionine-Gp22, cells were grown overnight in 5 ml of LB. The next day, the cells were then centrifuged for 5 minutes at 1300 g. Thereafter the pellet was resuspended in 1 ml of M9 medium and added to 1 litre of the same medium (M9). After this, the cells were left to grow in a 37 °C shaking incubator until they reached OD₆₀₀ 0.2, then amino acids were added; lysine, phenylalanine, and threonine at a final concentration of 100 mg/ml, and isoleucine, leucine and valine at a final concentration of 50 mg/ml.

Thereafter cells were grown until an optical density OD₆₀₀ of 0.4 at 37°C, then cells were induced using isopropyl β-D-1-thiogalactopyranoside (IPTG) with a concentration of 0.5 mM, and cells were incubated at 17°C overnight. After this growth period, the culture

Structural characterization of the protein Gp22

was centrifuged at 5000 x g for 20 min at 4°C, the supernatant was removed, and the cell pellet was stored at – 80 °C.

5.3.3 Purification of Selenomethionine- Gp22 protein

Proteins were purified as described previously in chapter 3, section 3.3.2.2.2.

5.3.4 Gp22 and Selenomethionine- Gp22 protein crystallization and optimization

To crystallise Gp22 and selenomethionine-Gp22, both proteins were concentrated to 11 mg/ml and 9 mg/ml respectively in a 20mM Tris-HCL pH 7.5 and 20 mM NaCL. Trial crystallisation plates were set up using a Mosquito robot by mixing 0.1 µL of protein with an equal volume of each buffer using the sitting-drop vapour diffusion technique. Different commercial crystallization screens were used, including PACT, JCSG, Proplex and Morpheus (Molecular Dimensions, UK). Screens were incubated at room temperature and also at 4°C. After two hours, protein crystals started to form in several different conditions. The screens that were chosen to be optimized were from JCSG and PACT.

5.3.5 The optimization of the conditions that produced crystals for Gp22 protein

Several attempts of optimisations were conducted to improve the quality of the crystals, by setting up larger drops (1 µl of the protein mixed with 1 µl of the reservoir buffer) or by changing the concentration of the precipitant, salts and by adding an additive screen. Table 5-1 below shows the conditions that were used to obtain crystals for the X-ray diffraction. The screens that gave the most promising crystals and used for further optimization are highlighted.

Structural characterization of the protein Gp22

Table 5-1:- List of the conditions that have been used to produce Gp22 protein Crystals

No	Conditions	Optimisations performed
1	<u>0.8M Succinic acid pH 7.0</u>	1- 1M, 0.9 M, 0.8 M, 0.7 M, 0.6 M, 0.5 M, 0.4 M, 0.3 M, 0.2 M Succinic acid pH 7.0. 2- 0.6 M Succinic acid pH 7.0 + additive screen
2	0.2 M Sodium chloride, 0.1 M HEPES pH 7.5 and 10% v/v 2-Propanol	0.2M Sodium chloride, 0.1 M HEPES pH 7.5 and 14%, 12%, 10%, 8%, 6%, 4% v/v 2-Propanol
3	1.0 M Ammonium phosphate dibasic, 0.1 M Sodium acetate pH 4.5	1M, 0.9 M, 0.8 M, 0.7 M, 0.6 M, 0.5 M, 0.4 M Ammonium phosphate dibasic, 0.1 M Sodium acetate pH 4.5
4	0.2 M Sodium malonate dibasic monohydrate and 20% w/v PEG 3350	0.2 M Sodium malonate dibasic monohydrate and 22%, 20%, 18%, 16%, 14%, 12% w/v PEG 3350
5	0.1 M Bicine pH 9.0 and 10% v/v MPD	0.1M Bicine pH 9.0 and 18%, 16%, 14%, 12%, 8%, 6% v/v MPD
6	<u>0.2 M Sodium citrate tribasic dehydrate and 20 % w/v PEG 3350</u>	1- 0.2 M Sodium citrate tribasic dehydrate and 22%, 20 %, 18%, 16%, 14% , 12%, 10% w/v PEG 3350 2- 0.5 M, 0.4 M, 0.3 M, 0.2 M, 0.1 M Sodium citrate tribasic dehydrate and 16% w/v PEG 3350 + additive screen. 3- <u>0.3M Sodium citrate tribasic dehydrate and 16% w/v PEG 3350 + additive screen.</u>
7	0.15 M DL-Malic acid and 20 w/vPEG 3350	0.15 M DL-Malic acid and 22%, 20 %, 18%, 16%, 14% , 12%, 10% w/v PEG 3350
8	0.1 M HEPES pH 7.0 and 30% v/v Jeff amine ED-2003.	0.1M HEPES pH 7.0 and 34%, 32%, 30%, 28%, 26%, 24% v/v Jeff amine ED-2003.

Structural characterization of the protein Gp22

9	0.2 M Potassium sodium tartrate tetrahydrate and 20 % w/v PEG 3350	0.2 M Potassium sodium tartrate tetrahydrate and 22%, 20 %, 18%, 16%, 14% , 12%, 10% w/v PEG 3350
---	--	---

5.3.6 Processes to improve the quality of the crystals after crystallization

5.3.6.1 Dehydration

This is one of the most common tools for a post-crystallization treatment to improve the quality of crystals. It has been reported to improve poor quality crystals for many different proteins (Heras *et al.*, 2003).

In the case of protein Gp22, crystals were grown in 0.3 M Sodium citrate tribasic dehydrate and 16% w/v PEG 3350, dehydration was conducted by replacing the reservoir solution with a solution containing with either 30 % w/v PEG 3000, 30 % w/v PEG 4000, 30 % w/v PEG 8000 or with 500 mM sodium citrate tribasic dehydrate as the dehydration agents. The plates were incubated overnight.

5.3.6.2 Micro seeding

To obtain a high-quality crystal that would be suitable for X- ray shooting, seeding was performed for Gp22 protein and Selenomethionine-Gp22. This technique is critical in aiding the improvement of the quality of the crystals. As seeding is used to introduce a previous nucleated crystal to a new drop of the protein and reservoir buffer from the conditions desired. This would enhance the efficiency of the nucleation and subsequently, the growth of the crystals.

To perform seeding. Crystals from two different conditions (0.6 M Succinic acid pH 7.0) and (0.3 M Sodium citrate tribasic dehydrate with 16% w/v PEG 3350) were smashed in to crystalline particles by overtaxing using seeding beads. Then the stock of the seeds was serially diluted 1:3, 1:9, 1:27, and 1:81, after which 0.3 μ l of each dilution was added to a drop of 1 μ l of protein and 0.7 μ l of the desired reservoir buffer. This in order to

Structural characterization of the protein Gp22

provide each drop with a different number of nuclei. Drops were prepared manually in a MRC crystallization plate.

5.4 Results

Several conditions from the four crystallisation screens (PACT, JCSG, Proplex and Morpheus) produced crystals in the initial screening. Thereafter, optimisation was conducted using each these conditions as a starting point. Two conditions yielded crystals upon optimisation. These conditions are described below.

5.4.1 The use of 0.8 M Succinic acid at a pH of 7.0 as a crystallization condition

Native Gp22 proteins was used to generate crystals initially. Figure 5-4 (A) shows crystals grown in 0.6 M succinic acid pH 7.0. This condition was used as the basis of an additive screen to try and generate better crystals. Figure 5-4 (B, C and D) shows crystals that were obtained using the additive screen (0.1 M ammonium sulphate, 0.1 M potassium sodium tartrate tetra hydrate and 0.01 M Spermine tetra hydrochloride) respectively . At this stage crystals were tested for diffraction using the in-house X-ray/detector set up. However, no diffraction was observed so other condition were explored.

Structural characterization of the protein Gp22

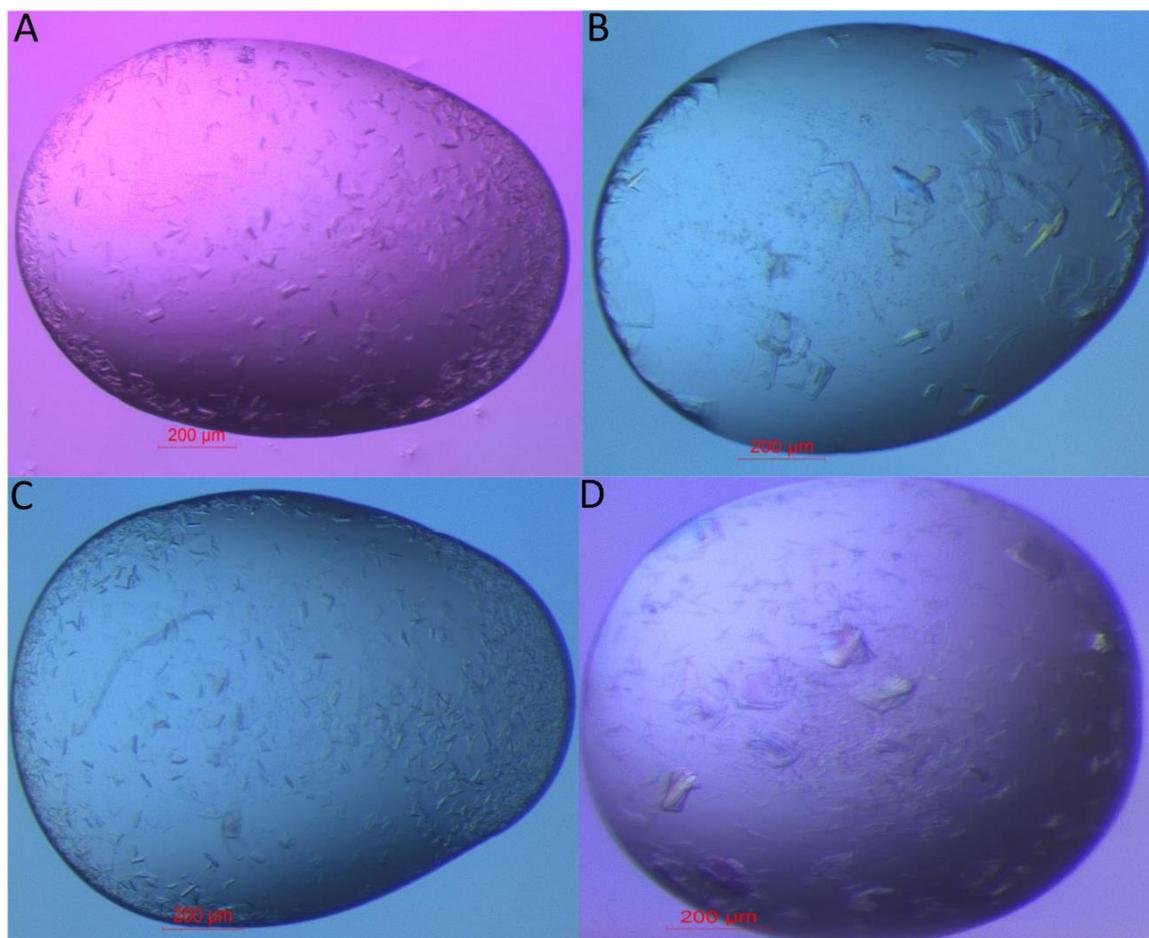


Figure 5-4: Gp22 crystal image

The Figure shows the optimisation of the condition of using 0.8M Succinic acid at pH 7.0. Image (A) represents the first step of optimisation using different concentration of the buffer (0.6 M Succinic acid pH 7.0). However the Images B, C and D are the representation of the second step of optimisation where additive screen was used. This additive screen being composed of 1.0 M ammonium sulphate, 1.0 M potassium sodium tartrate tetra hydrate and 0.1 M Spermine tetra hydrochloride respectively.

5.4.2 The use of 0.2 M Sodium citrate tribasic dehydrate, and 20 % w/v PEG 3350 as a crystallization condition

In this condition, several optimisation steps were conducted. The first step was to change the original concentration of the precipitant (20 % w/v PEG 3350) gradually to different concentrations. 16% w/v PEG 3350 was found to be best for producing Gp22 crystals. Furthermore, different salt (0.2 M Sodium citrate tribasic dehydrate) variations were also tested. The result show that 0.3 M Sodium citrate tribasic dehydrate was the best to produce crystals. As the crystals were slightly bigger than the other crystals that were produced from other salt concentration.

As the last step of optimization for this condition, the additive screen was combined with the previous optimisation. The best conditions chosen are as follows; firstly, 0.3 M Sodium citrate tribasic dehydrate, 16 % w/v PEG 3350 and 0.1 M potassium chloride. The crystal produce by this condition is shown in Figure 5-5 (A). Secondly, 0.3 M Sodium citrate tribasic dehydrate, 16 % w/v PEG 3350 and 0.1 M Lithium chloride was used, the obtained crystal for this condition is shown in Figure 5-5 (B). Finally, 0.3 M Sodium citrate tribasic dehydrate, 16 % w/v PEG 3350 and 0.1 M Caesium chloride was also used.

These three aforementioned conditions were found to be the best in aiding Gp22 and Selenomethionine- Gp22 crystal formation. When subjected to x- ray, a diffraction of 5.8 Å was observed. Therefore, to improve the quality of the crystals, dehydration and seeding were performed.

Structural characterization of the protein Gp22

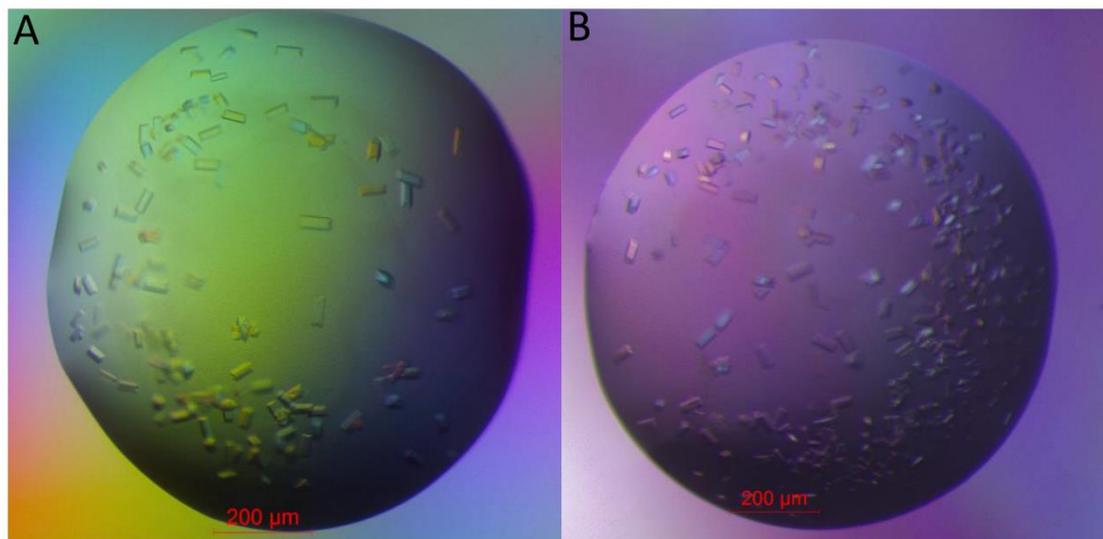


Figure 5-5 : Gp22 protein crystal image represent

The optimisation of the second condition targeted for this study (0.2 M Sodium citrate tribasic dehydrate and 20 % w/v PEG 3350). (A) and (B) images are representation of manual preparation of the conditions using different concentration of PEG, salt and an additive screen (0.3 M sodium citrate tribasic dehydrate, 16 % w/v PEG 3350 and potassium chloride along with 0.3 M sodium citrate tribasic dehydrate, 16 % w/v PEG 3350 and lithium chloride. The two preparations were found to be the most promising conditions.

5.4.3 Dehydration

The condition of using 0.3 M Sodium citrate tribasic dehydrate and 16 % w/v PEG 3350 with the two additives (0.1 M potassium chloride and 0.1 M lithium chloride) were used in the dehydration treatment to improve the quality of the Gp22 crystals, this in order to obtain a crystal that diffracts at higher level which subsequently would lead to resolve Gp22 protein structure. However, for the Selenomethionine-Gp22 protein, crystals were difficult to reproduce using this condition. Therefore, the focus here was on the native Gp22 protein. Four different dehydration agents were used. The first condition highlighted in Figure 5-6 (A) represent the crystals after the treatment with 30% of PEG 3350. The second condition is shown in Figure 5-6 (B) where 30% of PEG 4000 was used. The third condition in Figure 5-6 (C) shows the crystal when 30% of PEG 8000 used, and finally Figure 5-6 (D) shows the crystals formed after using 0.5 M Sodium citrate tribasic dehydrate. An X- Ray trial has not been performed yet for any of these conditions.

Structural characterization of the protein Gp22

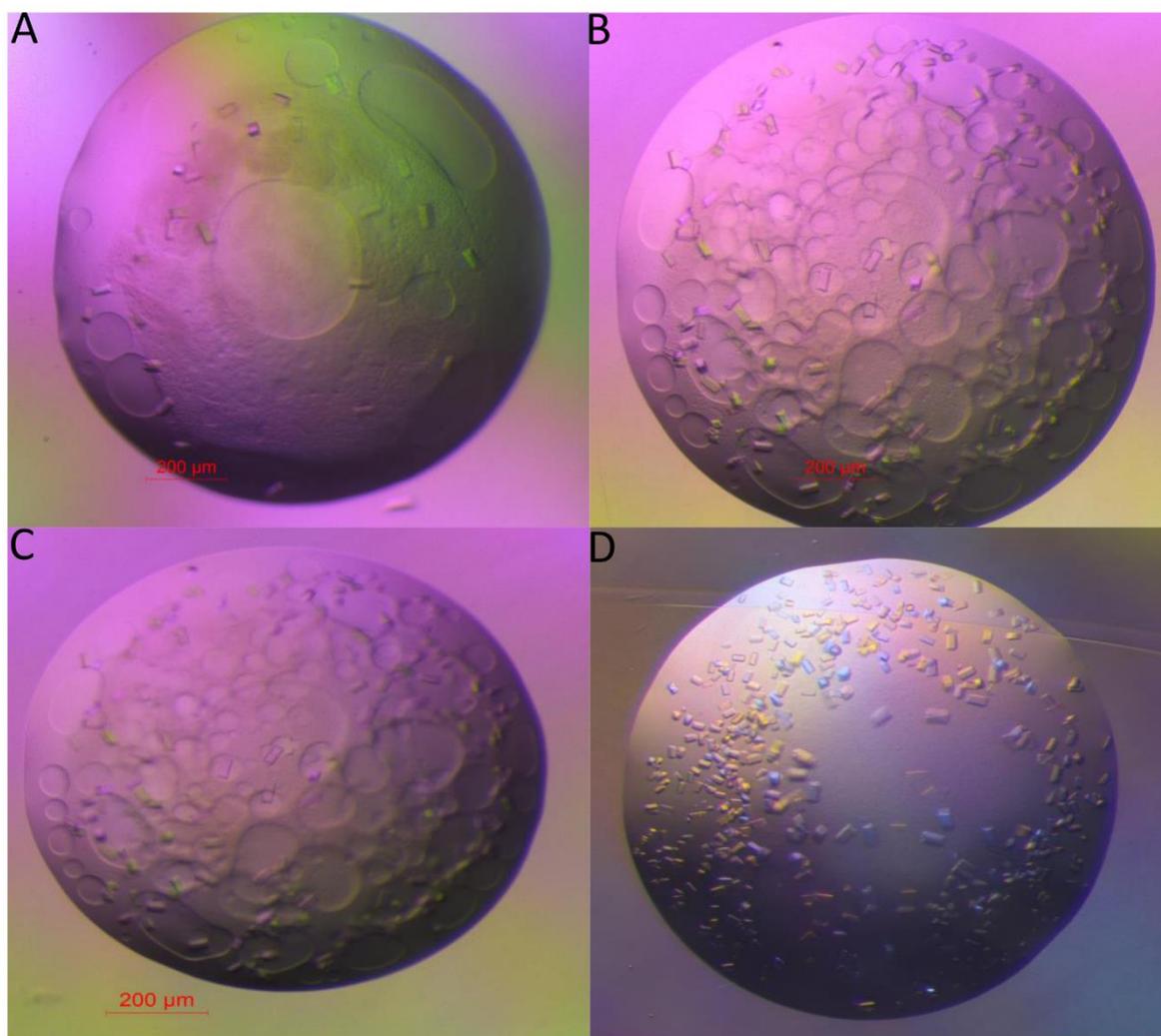


Figure 5-6 : Gp22 crystal images

Represent above is the data obtained from the dehydration process of using 0.3 M Sodium citrate tribasic dehydrate and 16 % w/v PEG 3350 as different conditions along with several dehydration agents; (30% of PEG 3350 (A), 30% of PEG 4000 (B), 30 % of PEG 8000 (C) and (D) 0.5 M Sodium citrate tribasic dehydrate.)

Structural characterization of the protein Gp22

5.4.4 Micro-seeding

To enhance the quality of the crystals micro-seeding was performed, and the best condition that was able to produce crystals was when 0.1 M HEPES (pH 7.0) and 26% v/v Jeff amine ED-2003 were used. The result shown in Figure 5-7 shows the image of the Selenomethionine-Gp22 protein crystals that were obtained after the seeding process using the above condition. X- Ray diffraction has not yet been performed for the crystals obtained from the above mentioned condition.

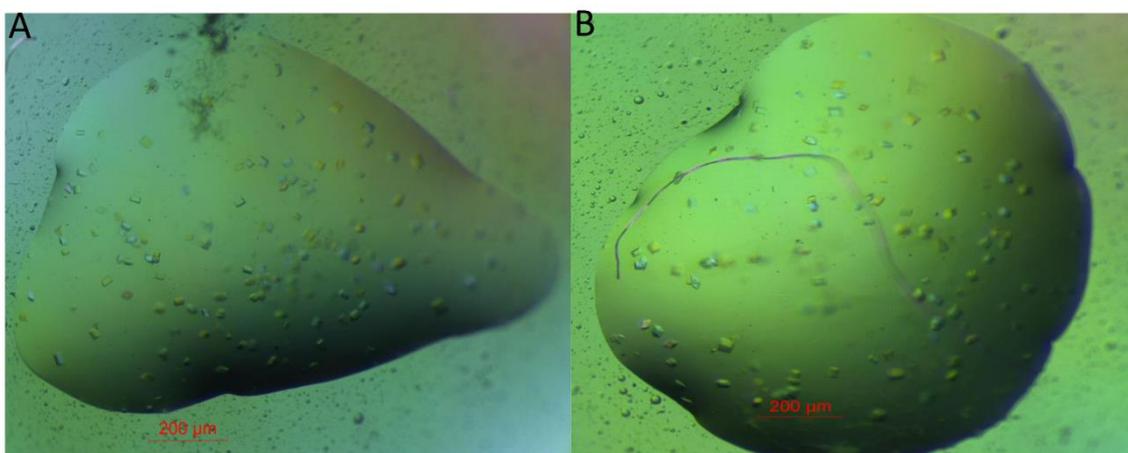


Figure 5-7: Gp 22 Crystals image

Micro-seeding was used to produce the crystals shown in (A) & (B), this was when a 0.1 M HEPES (pH 7.0) solution containing 26 % v/v Jeff amine ED-2003 was used.

Structural characterization of the protein Gp22

5.4.5 X- ray diffraction

Although some crystals diffracted, the diffraction was not of sufficient resolution to determine the structure of Gp22. The conditions that produced the best crystals were with the following conditions:

1 - 0.3 M sodium citrate tribasic dehydrate with 16 % w/v PEG 3350 and 0.1 M potassium chloride.

2 - 0.3 M sodium citrate tribasic dehydrate, 16 % w/v PEG 3350 and 0.1M lithium chloride

3- 0.3 M sodium citrate tribasic dehydrate, 16 % w/v PEG 3350 and 0.1 M caesium chloride.

Prior to diffraction, the crystals were transferred to a reservoir solution containing 30 % glycerol as a cryo-protectant before freezing in liquid nitrogen. The diffraction data collection was collected from Diamond light source. The best data set diffracted to a 5.8 Å resolution for the native Gp22 protein (shown in table 5-2). Furthermore, a diffraction pattern is shown in Figure 5-8.

Table 5-2:- X ray data collection of Gp 22 protein

Data Collection	Native Gp22
Space group	P212121
Unit cells dimension	a 180.1Å, b 220.7Å, c 231.8Å,, α 90°, β 90°, λ 90°
Resolution	5.8 Å

Structural characterization of the protein Gp22

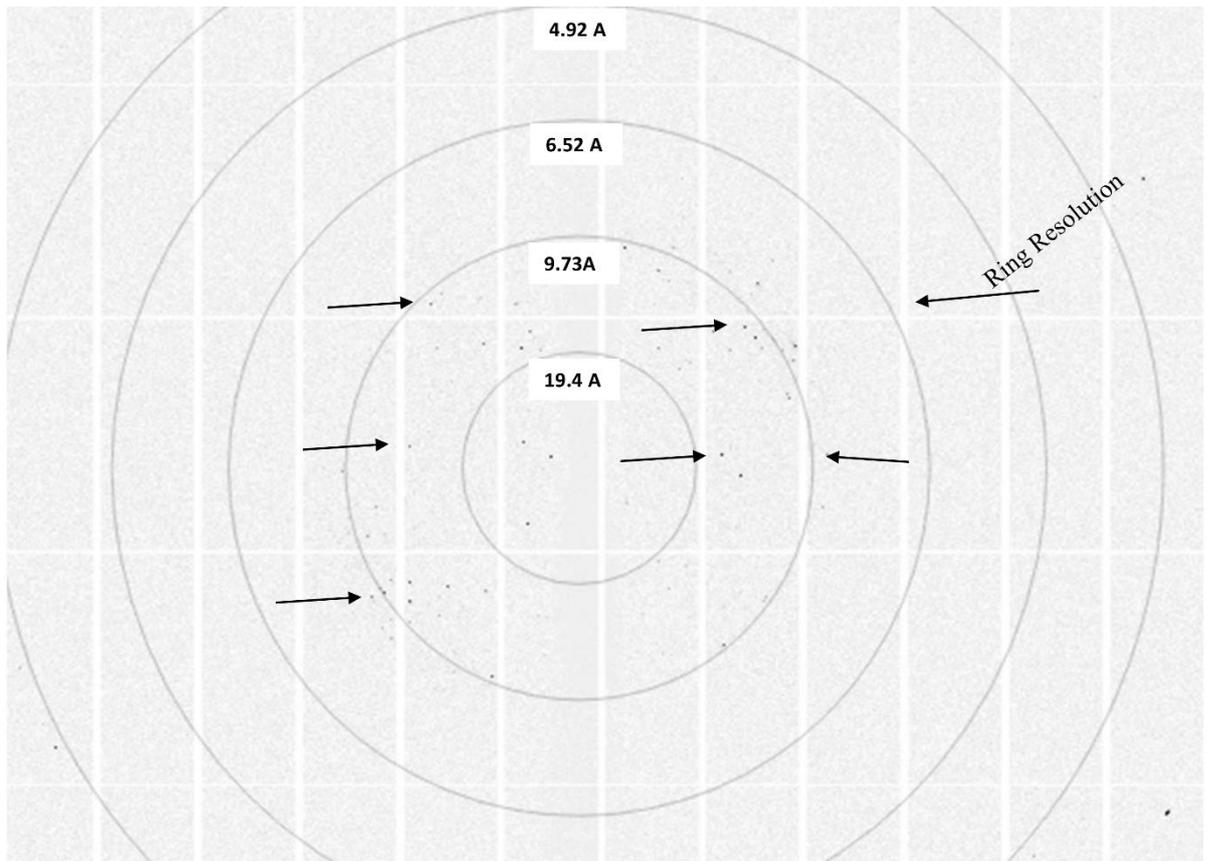


Figure 5-8: diffraction pattern for the native GP22 protein

The Gp22 protein diffracted up to 5.8° . The black spot represents the Gp22 protein scattering obtained, this is a reflection of the Gp22 atom arrangement.

5.5 Discussion

X-ray crystallography is a common and effective technique to determine the three-dimensional structure of proteins, however, in order to achieve the desired result. It is important to obtain a very high quality crystal that would be suitable to diffract at a level at which enables us to resolve the three-dimensional structure of the protein (Dale *et al.*, 2003). With this taken into consideration, in this chapter I aimed to identify the three-dimensional structure of the Gp22 protein, the RBP for phage CDHS1. This will empower us to understand the function of this protein, moreover the mechanism by which this protein attaches to the *C. difficile*. Due to the lack of structural homology for Gp22 protein, Selenomethionine labelling was performed to resolve the phasing problems and help in resolving the structure of the Gp22 protein.

Both native and selenomethionine-labelled Gp22 proteins crystallised. Crystals formed within 2-3 hours of setting up the crystallisation screens. Several conditions from different crystallization screens were able to produce Gp22 protein crystals, however the condition of using 0.3 M sodium citrate tribasic dehydrate, 16 % w/v PEG 3350 with three different additives (0.1 M potassium chloride, 0.1 M lithium chloride and 0.1 M Caesium Chloride) yielded the best crystals for both the native and selenomethionine-enriched Gp22 protein. The best crystal diffracted to 5.8 Å, other crystals diffracted weakly or not at all.

The production of high quality crystals that are able to diffract at a high resolution is one of the key requirements for protein structure determination (McPherson *et al.*, 2014). Also it is one of the stumbling blocks that is usually encountered when conducting protein crystallisation (Huang *et al.*, 2016). As a result, one of the major problems encountered when trying to determine the Gp22 protein structure was to obtain crystals that would diffract at a high resolution, which would allow to resolve the Gp22 protein structure.

Structural characterization of the protein Gp22

There are many factors that influence the quality of the protein crystals, however, the difficulty is the uncertainty of knowing which of these factors or parameters are key in influencing protein crystallisation (McPherson *et al.*, 2014). Therefore, several steps of optimisations have been applied in order to overcome this problem. Such as using different temperatures, changing the concentration of the protein, salt, and precipitants, varying the volume of the drop and adding an additive screen. This is known as Pre-crystallization optimization, and this aims to enhance the quality of the crystals (Lobley *et al.*, 2016).

The results that were obtained from such optimizations, revealed that the conditions mentioned above is the most promising, as both, the native Gp22 protein and Selenomethionine-Gp22 were both able to diffract up to 5.8 Å in these conditions. This level of resolution is not sufficient to obtain the full three-dimensional structure for Gp22. Ideally, the aim would be to produce crystals that diffract up to 3 Å at most. For this reason, further optimisation steps were applied. Such optimisation were performed post Gp22 protein crystallisation. Several techniques that can be performed on the obtained crystals, such as dehydration and annealing, these optimisation steps are known as post crystallisation treatments (Lobley *et al.*, 2016).

The dehydration step was the main post crystallisation optimisation procedure that was applied to improve the quality of the Gp22 crystal diffractions. This technique is considered as one of the key and most common post crystallisation treatments. It has been reported to improve the quality of protein crystals diffraction remarkably. As dehydration will assist in overcoming the poorly packed and unarranged crystals. By reducing the amount of water inside the crystal. As this is one of the major reasons as to why protein crystals diffract poorly (Heras *et al.*, 2005).

Structural characterization of the protein Gp22

Attempts were made to apply dehydration techniques to improve Gp22 protein crystal diffractions, this by using high concentration of precipitants and salts as mentioned in the methods section (5.3.6.1). However, X- ray has not yet been performed, therefore the effect of the dehydration on the Gp 22 crystals is not yet known. Many proteins in the Protein Data Bank (PDB) have had their three-dimensional structures resolved after using dehydration treatments (Heras *et al.*, 2005, Lobley *et al.*, 2016, Hellmich *et al.*, 2014).

Another optimisation step applied to enhance the diffraction of Gp22 protein crystals was the use of seeding. This technique has long been reported as a widely used method to enhance the quality of the protein crystals, thus improving the diffraction of those crystals (Till *et al.*, 2013, Zhu *et al.*, 2005) The principle behind using seeding comes from the paradox that sometimes occur in protein crystallisation, that the optimum condition for nucleation is not suitable for the crystal growth. This is because crystal nucleation occurs at during a highly supersaturated state. However, crystal growth occurs during low levels of the saturation (Bergfors, 2003). Therefore, seeding is used, as this technique works in a separate manner in both events, to fulfil the different requirements of nucleation and growth (Zhu *et al.*, 2005).

In the case of Gp22 protein, seeding was used separately and in conjunction with dehydration. Only one condition produced crystals after the seeding procedure, however, X- ray has not yet been performed in order to see if there is an improvement in the diffraction. Gp22 protein is the most potential candidate to be a RBPs for CDHS1 Phage. Therefore, identifying the three-dimensional structure of this protein will assist in understanding the function of this protein, moreover, provide an insight into the mechanism of action by which this protein attaches to *C. difficile* strain used.

Structural characterization of the protein Gp22

Several RBPs of different siphoviruses that infect other Gram-positive bacteria have had their structure determined. Such as the ORF 18 of phage P2 and ORF 49 of phage Tp901-1, both of which are phages infect *L. lactis*. In addition to this, the structure of Gp45 protein from phage ϕ 11 that infect *S. aureus* has also been determined. The common feature that has been concluded from such studies reveal that the RBPs of these phages are composed of three domains, the shoulder (N-terminal), the nick, and the head (C-terminal) which contains the binding site that the RBPs use to attach to the bacterial host (Dowah *et al.*, 2018, Koc *et al.*, 2016, Sciara *et al.*, 2008).

5.6 Conclusion

The conclusion that can be drawn from this chapter are that:

- 1- The protein crystallisation for Gp22 and Selenomethionine-Gp22 protein has been successful. Several crystallisation conditions have been producing crystals for these two proteins.
- 2- The quality of the desired crystals needs to be improved as the diffraction resolution obtained was up to 5.8 Å. Therefore, several optimisation steps were performed.
- 3- Dehydration and seeding techniques were conducted to improve the quality of the obtained crystals. However, X- ray diffraction has not yet been performed, there is no statement can be made on the effectiveness of these methods in improving the quality of the crystals.

Conclusion and future work

Chapter 6 Conclusion and future work

Conclusion and future work

6.1 Key finding of this study

Studying phage-bacterial interactions is essential, as this improves our understanding of phage behaviour, and enables downstream exploitation of phages. The key phage proteins responsible for phage-bacterial interactions are the RBPs. Phage RBPs have been extensively studied and characterised for phages that infect Gram-negative bacteria such as *E. coli* T4 phage (Mahony *et al.*, 2012). However recently there has been an increased interest in studying the RBPs for phage that infect Gram-positive bacteria such as phages that infect *L. lactis*, *S. aureus* and *L. monocytogenes* (Spinelli *et al.*, 2006, Biemann *et al.*, 2015, Li *et al.*, 2016b). Therefore, the main aim of this PhD project was to identify the RBPs for two phages that infect *C. difficile*; the first one being CDHS1, a siphovirus that infects clinically prevalent *C. difficile* strains (CD105LC1 & CDR20291) that belong to ribotype 027. The being second being phiCDMH1, which is a myovirus that infects CD105HE1, ribotype 076.

Several approaches have been used to identify the RBPs for phages, such as performing bioinformatics analysis, antibody based studies, molecular approaches and structural studies (Koc *et al.*, 2016, Vegge *et al.*, 2006). In this study, bioinformatics analysis, antibody based studies and structure based studies were attempted.

6.1.1 Identification of the putative function of the tail proteins

It has been proven that the genes located between the gene encoding the tape measure protein (Tmp), and the genes encoding the holin and endolysin are the genes that encode the tail proteins, or the proteins that are involved in the phage baseplate structure (Biemann *et al.*, 2015). Therefore, for phage CDHS1, four genes that are said to encode tail proteins (Gp18, Gp19, Gp21, and Gp22) were targeted for analysis in this project. However, eight genes that encoded tail proteins (Gp23, Gp24, Gp25, Gp26, Gp27, Gp28, Gp29 and Gp30) from CDMH1 phage were targeted during this research project. Using *in silico* analysis for these proteins, it was found that the Gp18 protein from phage

Conclusion and future work

CDHS1 has strong structural homology to the Dit protein from other siphoviruses, such as the TP901-1, Tuc2009 and P2 phages of *L. lactis*, phages A118 and P35 that infect *L. monocytogenes*, and $\phi 11$ phage that infects *S. aureus*. However for the Gp19 protein, strong homology to the Tal associated lysin protein (Tal) was found. Gp21 showed a strong level of homology to the N-terminal of the ORF48, the upper baseplate of phage TP901-1 which infects *L. lactis*. However, for protein Gp22 the protein with the highest level of interest, the *in silico* analysis revealed that Gp22 has no homology to any other protein found in the current database.

For proteins from phage CDMH1, *in silico* analysis showed that the majority of the proteins targeted have structural homology to the tail proteins from the T4 phage that infects *E. coli*. Which indicates that the baseplate of the CDMH1 phage may have a T4 like baseplate structure. Interestingly, it was found that Gp24 from phage CDMH1 has structural homology to the tail-associated lysosome of the T4 phage (Gp27). Thereby suggesting that Gp24 may bind to the cell wall of *C. difficile*. In addition, the Gp24 protein has a cell wall hydrolase domain at the C-terminal, which binds to peptidoglycan. However, for Gp29 and Gp30 proteins, no homology to any other phage proteins was found, which made them proteins of interest.

6.1.2 Production and purification of the tail proteins of phages CDHS1 and CDMH1

After *in silico* analysis of the tail proteins from the two phages used in this study was performed. The four proteins Gp18, Gp19, Gp21 and Gp22 from phage CDHS1 and two proteins Gp29 and Gp30 from phage CDMH1 were targeted to be expressed, produced and purified as part of this study. The main result of this chapter of study was that all the proteins targeted for the production were successfully overexpressed and purified aside from. Gp19. The Gp19 protein from CDHS1 was found to be challenging and was not produced. The pure version of the proteins that were successfully expressed (Gp18, Gp21,

Conclusion and future work

and Gp22 of phage CDHS1) and (Gp29 and Gp30 from phage CDMH1) were sent to Eurogentec Company (Brussels, Belgium) in order obtain to polyclonal antibodies against these proteins.

6.1.3 Determining the RBPs of phages CDHS1 and CDMH1

When the polyclonal antibody raised against the proteins targeted in this study were received. The research was aimed at identifying which of these proteins (Gp18, Gp21, and Gp22 of phage CDHS1 along with Gp29 and Gp30 from phage CDMH1) act as the RBPs for these phages. The approach applied was to incubate the polyclonal antibodies raised against these proteins with the phages. Then, the spot test was used to see if the phage was able to infect the specific strain after incubation with the antibody. The key finding of this chapter and the most interesting part of this whole project was that the anti-Gp22 antibodies were able to block the infection of phage CDHS1 to strains CD105LC1 and CDR20291. The anti Gp29 from CDMH1 also was able to inhibit the infection of CDMH1 to CD105HE1 strain. This indicates that, proteins Gp22 of phage CDHS1, and Gp29 from CDMH1 are the RBPs for both phages. To date, this is the first study that has identified the RBPs of phages that infect *C. difficile*.

Furthermore, anti-Gp29 was tested against another two other phages; CDHM3 and CDHM6 both of which infect strain CD105HE1 and have a similar host range to phage CDMH1. Interestingly, it was found that anti-Gp29 serum was able to inhibit CDHM3 and CDHM6 infection to CD105HE1 as well. Which indicates that the three phages CDHM3, CDHM6 and CDMH1 may share the same RBPs.

6.1.4 Determining the three-dimensional structure of protein Gp22

This PhD research provided evidence that the Gp22 protein is the RBPs for phage CDHS1. Therefore, it was important to continue the research to understand the mechanism of action and function of this protein. To achieve this purpose, protein crystallization was performed for the Gp22 protein. In this approach different commercial

Conclusion and future work

protein crystallization screens were used in the sitting-drop vapour diffusion technique. To date, the main finding was that Gp22 diffracted up to 5.8 Å, was the best level of diffraction so far. However, this resolution is still not sufficient to obtain the full three-dimensional structure for Gp22 protein. Therefore, work is still underway to obtain a better resolution for this protein. Due to limitations in time, further work could not be carried out, which would have been aimed at determining the protein structure of protein Gp29 from phage CDMH1.

6.2 Future work

Based on the results obtained from this PhD project. The following recommendations can be considered for future work:

6.2.1 Expression and purification of the tail associated lysin for the two phages used in this project

Based on the *in silico* analysis performed for protein Gp19 of CDHS1 and Gp24 from CDMH1. It was concluded that these two proteins may act as a tail associated lysin protein (Tal) for both phages. The main functions of Tal proteins in other phages, is to degrade the peptidoglycan layer of the host, in order to aid the injection of phage DNA in to the bacteria. The possibility to degrade the peptidoglycan layer of the host could potentially be exploited, these proteins could be used for therapeutic purposes against *C. difficile*. Therefore, it is worth trying to overexpress these proteins and test their anti-microbial activity against *C. difficile*.

In this project, there was an attempt to produce Gp19 of CDHS1; however, it was difficult to produce the Gp19 protein, this due to the possible degradation of this protein. To overcome this problem, it could be possible to try to produce a truncated version of this protein. Alternatively, another option is to produce this protein in combination with Gp18 to avoid any possible degradation.

Conclusion and future work

6.2.2 Resolving the structure of the RBPs

Despite the hard work and effort exerted in trying to resolve the structure of Gp22 from phage CDHS1, to date the Gp22 crystals diffracted at resolution of 5.8Å. Which is not sufficient to resolve the three dimensional structure for this protein. Therefore, work is still needed to improve the quality of the Gp22 crystals. Many techniques have been carried out to obtain high quality crystals that can diffract at a higher resolution, these techniques include dehydration and seeding. It would be possible to overexpress truncated versions of Gp22 (C-terminal end and N-terminal end), and then crystalize terminal end alone.

In addition to this, Cryo-electron microscopy for Gp22 could also be used to help resolve its structure. Moreover performing Cryo-electron microscopy for the baseplate for the two phages (CDMH1 and CDHS1) used in this project would help to understand the mechanism by which these two phages attach to their respective hosts. Due to limitations in time, it was difficult to crystalize protein Gp29 of CDMH1; however this could be considered for future work.

6.2.3 Exploiting the RBPs of the two phages as diagnostic tools

One of the major advantages of using phages as therapeutic tools is that the phages attach to bacterial host in a highly specific manner. The key factor for such specificity is the RBPs at the end of the phage tail proteins. Therefore, this property of having high specificity could be exploited within the diagnostic field.

6.2.4 Exploiting RBPs of the two phages to identify the binding ligand on the *C. difficile* cell wall

The first attachment between the bacterial host and phage occurs when the RBPs at the end of the phage tail binds to the receptor on the surface of the bacterium. However, bacterial hosts may block such attachment, preventing the phage or the RBPs of the phage to reach the ligand on the surface of the bacteria. Therefore, further understating the

Conclusion and future work

biology of the phage, it would be interesting to identify the receptors on the surface of the bacteria. This could possibly be achieved by exploiting the known RBPs by using different approaches that focus on protein-protein interactions, such as pull down assays and the thermal melt shift assay.

Appendices

Chapter 7 Appendices

7.1 Appendix 1: Purification of bacteriophages

1- CDHS1 Purification

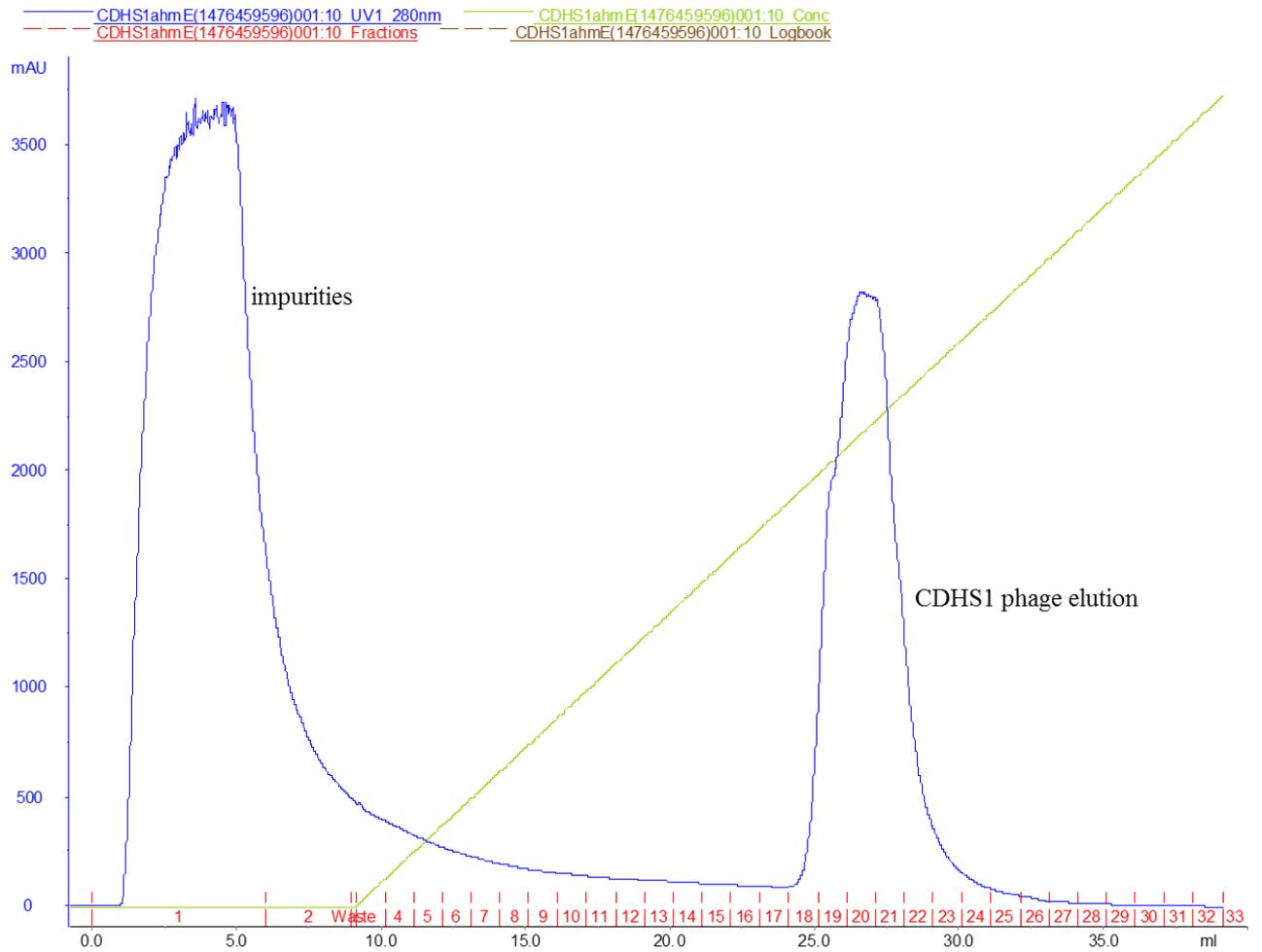


Figure 1. CDHS1 phage purification
Linear gradient output diagram of the purification of phage CDHS1.

2- CDMH1 purification

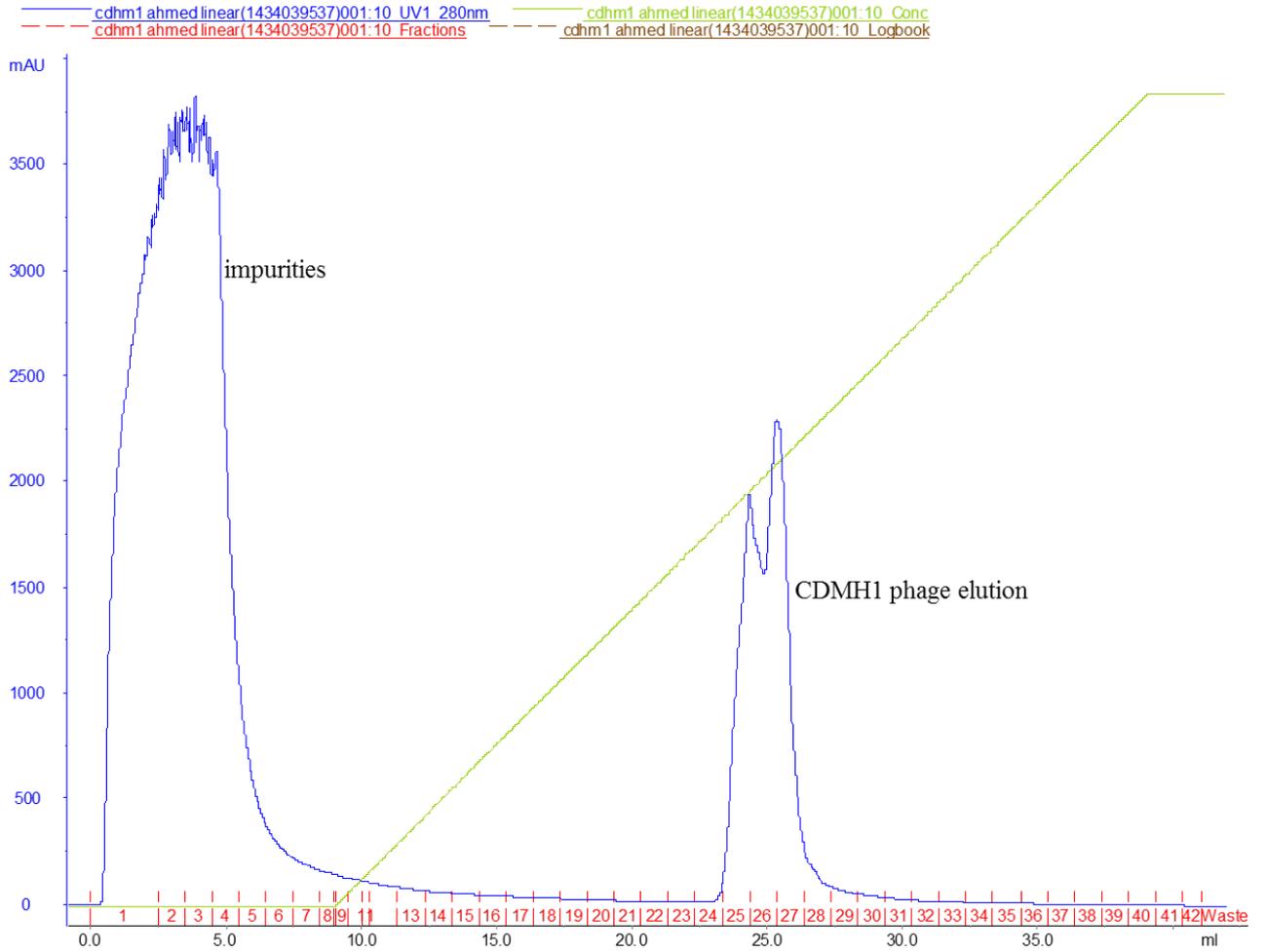


Figure 2. CDMH1 phage purification
Linear gradient output diagram of the purification of phage CDMH1.

Appendices

7.2 Appendix 2– Protein Sequences confirmation

The bold peptides are the peptides that matched with the original protein sequence

1- Gp18 protein sequences from CDHS1 phage

1 HHHHHHSSGV DLGTENLYFQ SMRGGHK**AMF FVYNGRDSRE** FGLKIYNIND
51 **LSAPQMEVER** VSVPGKDGDL LLK**KGFENFT LTIECDIDAR** QSNIEEVATE
101 IKK**WLQDIS YKKLFLNSD FYYLASCNNK** LDITRNFKNF **ASCLLTFCY**
151 **PFRYAEEEEII SLNVLNLKSA TITNFYRESK** PVLIEGAGD ISIKINTQSI
201 **VLRGVAENGI LSDLIIDSEQ MNVYRINKEN** NIIVNENK**L FSDFP**ILEEG
251 **ENQISWEGDI** KSIKINPRWN I

2- Gp21 sequence from CDHS1 phage

1 VINLRDRIYT VDINTKSYQV AKYKQYDNAI EFKINLLENN IEK**DLTGYTA**
51 **IANFQRPDGK** IVYQSC TIEN SIATTTIENN ITEVAGDVIV EFTFYKDDL
101 VTTFSLKINI EKSIDKNSIT EEPK**WDYISV TINQVKEVVE** **GIEEIKETEE**
151 **ARKEAEIKRV** EEFNSIKETF DSKVTEVTD KNSMISDVNT TKDTLTKEVT
201 DTKNDLTVV TTAKESMISE VTTVKEELEI AEGKRVVEFN SIKETFDSKV
251 TEVTDKNSM ISDVNTTKDT LTKEVTDTKN DLTNVVTTAK ESMISEVTTV
301 KEELEIAEGK **RVEEFNSIKE** TFDSKVTEVT DAKNSMISDV NTKTTLIDE
351 VNTVKAEVTT AKNTMISEVT TAKETMQTEV TDAINAIPK EELKGVGIEI
401 KGSLDNISAL PVNPTLSDAY FVKSATIENQ IDLYVWDNTN WVKVPDIKIK
451 GENGGLLEFN WDGTRLGIRI EGQENYTYTD LKGQKGDKGD SIEFNWNGTR
501 LGIKIEGQEN **YSYTELKGEQ** GYTPTIGENG NWWINNIDTQ KPARGASLRI
551 LGKLSIDNL PLDPTIGDCW IIGR**NIYIYQ** **TKWEDLGSLA** GVDGK**NLEFN**
601 **WDGTQLGVRQ** QYELDYKYID LKGD**NIEFAW** **DGTRLGVRIE** GQENYTYTDL
651 KGQK**GDSIEF** **SWDGT**ELGVR IEGQEDYSYT NLKGATGNKL EFNWNGSQLG
701 **IREEGQTEYI** **YTELRGEQGY** TPAIGENGNW FINGEDTGKA SKG**KVTW**NEL
751 **LEKPKELDYI** KKSTTFNSDG SITDILDSVS KTITKFNADG TIVDEKYIDN
801 VLVSKVKTF KGNQIEEIKE EVS

Appendices

3- Gp22 sequence from CDHS1 phage

```

1 MSWAETYKVN SDLQGEPLNF LSYLQDIKLN GLDSYVLFIG NARIWEELYL
51 NSLYLFSDRG IRETVYTAFS ETDIDNLFNK STKLGEQLNA FYRTDIFSLG
101 NADNVVKEMT IEHYSNLEEK FKAGYDRYVT REQEKSTIGA WFNSTFSLDN
151 TDLENLTIE EILANVEATN AILNNSNAIV ALTMCKSSMD AVVASSNAMD
201 LLGQYILRV TESPVIRAIL KNNVIRDAII NSDEAMTQIS SNENSVMEIF
251 NDLEATKVLV QNONSINKIL TNNVTVEKII PNLLEMKYNL QTSLNYINTI
301 KSNIASGKGQ IMAITYNEEI FPILKNAVKN YDGMETTRNI SQRDIEEKIK
351 ISDAILESSI AMATFANNSI IVNKVGDRVG IIESIFSKTV SLNAFMKSTT
401 AINILVNKTT AFTKIANNST AFNAMLTISE NNVTIANNNTT AMGIIANNAQ
451 AMSTVANNDT SISVFNNTT AMGIIANSST AMTKITLTGL ALNRMVKSNT
501 AKSILISKNS TLQTYKNNIQ NTIQGSTAYF RTITGFADAD DNPPQINST
551 YVGITYCYGY KNSYYGIVY HGYNTSIEAG RGNGYKDETK KFITLGGARY
601 DQSGDGYFTY AMYQAI

```

4- Sequence of selenomethionine labelling Gp 22 protein

```

1 MSWAETYKVN SDLQGEPLNF LSYLQDIKLN GLDSYVLFIG NARIWEELYL
51 NSLYLFSDRG IRETVYTAFS ETDIDNLFNK STKLGEQLNA FYRTDIFSLG
101 NADNVVKEMT IEHYSNLEEK FKAGYDRYVT REQEKSTIGA WFNSTFSLDN
151 TDLENLTIE EILANVEATN AILNNSNAIV ALTMCKSSMD AVVASSNAMD
201 LLGQYILRV TESPVIRAIL KNNVIRDAII NSDEAMTQIS SNENSVMEIF
251 NDLEATKVLV QNONSINKIL TNNVTVEKII PNLLEMKYNL QTSLNYINTI
301 KSNIASGKGQ IMAITYNEEI FPILKNAVKN YDGMETTRNI SQRDIEEKIK
351 ISDAILESSI AMATFANNSI IVNKVGDRVG IIESIFSKTV SLNAFMKSTT
401 AINILVNKTT AFTKIANNST AFNAMLTISE NNVTIANNNTT AMGIIANNAQ
451 AMSTVANNDT SISVFNNTT AMGIIANSST AMTKITLTGL ALNRMVKSNT
501 AKSILISKNS TLQTYKNNIQ NTIQGSTAYF RTITGFADAD DNPPQINST
551 YVGITYCYGY KNSYYGIVY HGYNTSIEAG RGNGYKDETK KFITLGGARY
601 DQSGDGYFTY AMYQAI

```

5- Positions where the Selenomethionine modifications occurred

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss Sequence
29 - 43	1651.8390	1650.8317	1650.8729	-25	0 K.LNGLDSYVLFIGNAR.I
63 - 80	2106.9540	2105.9467	2105.9793	-15	0 R.ETVYTAFSETDIDNLFNK.S
84 - 93	1210.6000	1209.5927	1209.6142	-18	0 K.LGEQLNAFYR.T
94 - 107	1492.7510	1491.7437	1491.7569	-9	0 R.TDIFSLGNADNVVK.E
108 - 120	1670.6590	1669.6517	1669.6738	-13	0 K.EMTIEHYSNLEEK.F Delta:S(-1)Se(1)
(M)					
209 - 217	1001.5600	1000.5527	1000.5553	-3	0 R.VTTESPVIR.A
258 - 268	1256.6780	1255.6707	1255.6884	-14	0 K.VLVQNONSINK.I
288 - 301	1684.8620	1683.8547	1683.8831	-17	0 K.YNLQTSLSNYINTIK.S
330 - 338	1134.3890	1133.3817	1133.3892	-7	0 K.NYDGMETTR.N Delta:S(-1)Se(1) (M)
379 - 388	1092.6300	1091.6227	1091.6227	0	0 R.VGIIIESIFSK.T
485 - 494	1071.6510	1070.6437	1070.6448	-1	0 K.ITLTGLALNR.M
517 - 531	1726.8010	1725.7937	1725.8434	-29	0 K.NNIQNTIQGSTAYFR.T
562 - 581	2220.9460	2219.9387	2220.0236	-38	0 K.GNSYYGIVYHGYNTSIEAGR.G
591 - 599	962.5660	961.5587	961.5709	-13	1 K.KFITLGGAR.Y
592 - 599	834.4750	833.4677	833.4759	-10	0 K.FITLGGAR.Y

Appendices

6- Sequence of Gp29 protein from CDMH1

1 MAIDK**SY**YTI ITDVGKAKIA NASVTGNKVG FVKIQLGDGG GSEYTP**TESQ**
51 TALKNVVWEG NIGNTTTTDET APNCIILES**L** IPSSVGGFMI REIGYLD**DDEN**
101 N**L**IAISKYKE CYKPSIEQGA VVDMKVKT**VL** IVSNVNNI**EL** KIDPTIIF**AT**
151 LKDIQDLETK IGTVNTKIDT TKTELK**S**NIE TAKTEID**E**KI GD**TT**QL**TT**TD
201 KTNIVGALNE VKTSVDSI**E**T TAEKTSY**N**NA TSKLTATT**VQ** GAID**EV**VAKI
251 ENFNEVN**IS**I QNDML**PI**

7- Sequence of Gp30 protein from CDMH1

1 MTTEW**N**FN**Y**I GTGKKVILKP GK**Y**KLE**C**WGA SGGGRFDE**WT** E**CA**KGGY**S**KG
51 ELTLK**K**ETIL Y**V**YAGES**G**YK KFSN**IS**DWAG FNGGGRG**P**NE GVDPK**FT**TC**G**
101 **G**GATDIR**L**IG **G**VWNDE**Q**GLL **S**R**I**I**V**AGGGG SIGTSS**F**SSI GLGGGFAG**G**M
151 GVGAGTTCTG GTQYEGG**V**TV NSNGNG**S**FGK GGIGN**V**CAGG GGWYGGAG**A**S
201 SSGVGGGG**S**G YVLT**K**DSY**K**P KGYIPT**S**EYW LENV**S**IAGD NTSNAHGY**A**K
251 ITLLQAL**P**FL NISSY**S**STA TFKADHT**D**PT LLTK**IE**Y**F**ID **D**VL**K**ETIT**T**D
301 LTLEK**T**IN**Y**T **L**ED**N**AL**H**TL**K** IVVTDSANAT VEK**V**SV**S**RG IAP**L**PSG**S**TT
351 DEVTNK**W**IEI KDAFK**T**G**K**TS IINTL**L**AK**N**I EASLN**N**TL**V**E LSEK**I**K**T**S**F**D
401 SSDAS**V**Q**D**LM NQLTQAN**T**I SQLNA**K**Y**K**VA GG**W**TT**P**V**Y**TN GTEK**A**L**V**Y**N**S
451 TREAIR**Y**D**W**I TISNL**G**F**I**PN VFYA**E**CD**Y**VN SYTK**S**KN**K**L**F** VFAC**Y**D**V**PT**S**
501 SSKDF**V**FT**C**D IQLGSS**D**SEY KVT**Y**C**G**LY**Q**H NKLD**I**H**M**T**D**N LIHL**P**AF**A**T**I**
551 NAVAS**Y**Y**W**R**A** IK**I**Y

Bibliography

Chapter 8 Bibliography

Bibliography

- Aguado, J. M., Anttila, V. J., Galperine, T., Goldenberg, S. D., Gwynn, S., Jenkins, D., Noren, T., Petrosillo, N., Seifert, H., Stallmach, A., Warren, T., Wenisch, C., Consortium, C. D. I. C., 2015. Highlighting clinical needs in *Clostridium difficile* infection: the views of European healthcare professionals at the front line. *J Hosp Infect*, **90**, 117-125.
- Aksyuk, A. A., Leiman, P. G., Shneider, M. M., Mesyanzhinov, V. V., Rossmann, M. G., 2009. The structure of gene product 6 of bacteriophage T4, the hinge-pin of the baseplate. *Structure*, **17**, 800-808.
- Alvarez, B. & Biosca, E. G., 2017. Bacteriophage-Based Bacterial Wilt Biocontrol for an Environmentally Sustainable Agriculture. *Front Plant Sci*, **8**, 1218.
- Aminov, R., Caplin, J., Chanishvili, N., Coffey, A., Cooper, I., De Vos, D., Doškař, J., Friman, V.-P., Kurtböke, İ., Pantucek, R., 2017. Application of bacteriophages. *Microbiology Australia*, **38**, 63-66.
- Arisaka, F., Yap, M. L., Kanamaru, S., Rossmann, M. G., 2016. Molecular assembly and structure of the bacteriophage T4 tail. *Biophys Rev*, **8**, 385-396.
- Bartual, S. G., Otero, J. M., Garcia-Doval, C., Llamas-Saiz, A. L., Kahn, R., Fox, G. C., van Raaij, M. J., 2010. Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proc Natl Acad Sci U S A*, **107**, 20287-20292.
- Bebeacua, C., Bron, P., Lai, L., Vegge, C. S., Brondsted, L., Spinelli, S., Campanacci, V., Veesler, D., van Heel, M., Cambillau, C., 2010. Structure and molecular assignment of lactococcal phage TP901-1 baseplate. *J Biol Chem*, **285**, 39079-39086.
- Bebeacua, C., Lai, L., Vegge, C. S., Brondsted, L., van Heel, M., Veesler, D., Cambillau, C., 2013. Visualizing a complete Siphoviridae member by single-particle electron microscopy: the structure of lactococcal phage TP901-1. *J Virol*, **87**, 1061-1068.
- Bergfors, T., 2003. Seeds to crystals. *J Struct Biol*, **142**, 66-76.
- Bertozzi Silva, J., Storms, Z., Sauvageau, D., 2016. Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett*, **363**.
- Bielmann, R., Habann, M., Eugster, M. R., Lurz, R., Calendar, R., Klumpp, J., Loessner, M. J., 2015. Receptor binding proteins of *Listeria monocytogenes* bacteriophages A118 and P35 recognize serovar-specific teichoic acids. *Virology*, **477**, 110-118.
- Borren, N. Z., Ghadermarzi, S., Hutfless, S., Ananthakrishnan, A. N., 2017. The emergence of *Clostridium difficile* infection in Asia: A systematic review and meta-analysis of incidence and impact. *PLoS One*, **12**, e0176797.

Bibliography

- Brandt, L. J., 2013. American Journal of Gastroenterology Lecture: Intestinal microbiota and the role of fecal microbiota transplant (FMT) in treatment of *C. difficile* infection. *Am J Gastroenterol*, **108**, 177-185.
- Burke, K. E. & Lamont, J. T., 2014. *Clostridium difficile* infection: a worldwide disease. *Gut Liver*, **8**, 1-6.
- Buttimer, C., McAuliffe, O., Ross, R. P., Hill, C., O'Mahony, J., Coffey, A., 2017. Bacteriophages and Bacterial Plant Diseases. *Front Microbiol*, **8**, 34.
- Chatterjee, S. & Rothenberg, E., 2012. Interaction of bacteriophage 1 with its *E. coli* receptor, Lamb. *Viruses*, **4**, 3162-3178.
- Christie, G. E. & Calendar, R., 2016. Bacteriophage P2. *Bacteriophage*, **6**, e1145782.
- Clokie, M., Millard, A. D., Letarov, A. V., Heaphy, S., 2011. Phages in nature. *Bacteriophage*, **1**, 31-45.
- Cornely, O. A., Crook, D. W., Esposito, R., Poirier, A., Somero, M. S., Weiss, K., Sears, P., Gorbach, S., 2012. Fidaxomicin versus vancomycin for infection with *Clostridium difficile* in Europe, Canada, and the USA: a double-blind, non-inferiority, randomised controlled trial. *The Lancet Infectious Diseases*, **12**, 281-289.
- Crowther, G. S. & Wilcox, M. H., 2015. Antibiotic therapy and *Clostridium difficile* infection - primum non nocere - first do no harm. *Infect Drug Resist*, **8**, 333-337.
- Czaplewski, L., Bax, R., Clokie, M., Dawson, M., Fairhead, H., Fischetti, V. A., Foster, S., Gilmore, B. F., Hancock, R. E. W., Harper, D., Henderson, I. R., Hilpert, K., Jones, B. V., Kadioglu, A., Knowles, D., Ólafsdóttir, S., Payne, D., Projan, S., Shaunak, S., Silverman, J., Thomas, C. M., Trust, T. J., Warn, P., Rex, J. H., 2016. Alternatives to antibiotics—a pipeline portfolio review. *The Lancet Infectious Diseases*, **16**, 239-251.
- Dale, G. E., Oefner, C., D'Arcy, A., 2003. The protein as a variable in protein crystallization. *J Struct Biol*, **142**, 88-97.
- Dang, V. T. & Sullivan, M. B., 2014. Emerging methods to study bacteriophage infection at the single-cell level. *Front Microbiol*, **5**, 724.
- De Haard, H. J., Bezemer, S., Ledebouer, A. M., Muller, W. H., Boender, P. J., Moineau, S., Coppelmans, M. C., Verkleij, A. J., Frenken, L. G., Verrips, C. T., 2005. Llama antibodies against a lactococcal protein located at the tip of the phage tail prevent phage infection. *J Bacteriol*, **187**, 4531-4541.

Bibliography

- Dowah, A. S. A. & Clokie, M. R. J., 2018. Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. *Biophys Rev*.
- Drulis-Kawa, Z., Majkowska-Skrobek, G., Maciejewska, B., 2015. Bacteriophages and Phage-Derived Proteins—Application Approaches. *Current medicinal chemistry*, **22**, 1757-1773.
- Dupont, K., Vogensen, F. K., Neve, H., Bresciani, J., Josephsen, J., 2004. Identification of the receptor-binding protein in 936-species lactococcal bacteriophages. *Appl Environ Microbiol*, **70**, 5818-5824.
- Fagan, R. P. & Fairweather, N. F., 2014. Biogenesis and functions of bacterial S-layers. *Nat Rev Microbiol*, **12**, 211-222.
- Flayhan, A., Vellieux, F. M., Lurz, R., Maury, O., Contreras-Martel, C., Girard, E., Boulanger, P., Breyton, C., 2014. Crystal structure of pb9, the distal tail protein of bacteriophage T5: a conserved structural motif among all siphophages. *J Virol*, **88**, 820-828.
- Garbe, J., Bunk, B., Rohde, M., Schobert, M., 2011. Sequencing and characterization of *Pseudomonas aeruginosa* phage JG004. *BMC microbiology*, **11**, 102.
- Garcia, P., Martinez, B., Obeso, J. M., Rodriguez, A., 2008. Bacteriophages and their application in food safety. *Lett Appl Microbiol*, **47**, 479-485.
- Gonzalez-Garcia, V. A., Pulido-Cid, M., Garcia-Doval, C., Bocanegra, R., van Raaij, M. J., Martin-Benito, J., Cuervo, A., Carrascosa, J. L., 2015. Conformational changes leading to T7 DNA delivery upon interaction with the bacterial receptor. *J Biol Chem*, **290**, 10038-10044.
- Habann, M., Leiman, P. G., Vandersteegen, K., Van den Bossche, A., Lavigne, R., Shneider, M. M., Biemann, R., Eugster, M. R., Loessner, M. J., Klumpp, J., 2014. *Listeria* phage A511, a model for the contractile tail machineries of SPO1-related bacteriophages. *Mol Microbiol*, **92**, 84-99.
- Haq, I. U., Chaudhry, W. N., Akhtar, M. N., Andleeb, S., Qadri, I., 2012. Bacteriophages and their implications on future biotechnology: a review. *Virol J*, **9**, 9.
- Hargreaves, K. R. & Clokie, M. R., 2014. phages: still difficult? *Front Microbiol*, **5**, 184.
- Hellmich, J., Bommer, M., Burkhardt, A., Ibrahim, M., Kern, J., Meents, A., Muh, F., Dobbek, H., Zouni, A., 2014. Native-like photosystem II superstructure at 2.44 Å

Bibliography

- resolution through detergent extraction from the protein crystal. *Structure*, **22**, 1607-1615.
- Heras, B., Edeling, M. A., Byriel, K. A., Jones, A., Raina, S., Martin, J. L., 2003. Dehydration Converts DsbG Crystal Diffraction from Low to High Resolution. *Structure*, **11**, 139-145.
- Heras, B. & Martin, J. L., 2005. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallographica Section D*, **61**, 1173-1180.
- Ho, T. D. & Slauch, J. M., 2001. OmpC is the receptor for Gifsy-1 and Gifsy-2 bacteriophages of *Salmonella*. *Journal of bacteriology*, **183**, 1495-1498.
- Hu, B., Margolin, W., Molineux, I. J., Liu, J., 2015. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proc Natl Acad Sci U S A*, **112**, E4919-4928.
- Huang, Q. & Szebenyi, D. M., 2016. Improving diffraction resolution using a new dehydration method. *Acta Crystallogr F Struct Biol Commun*, **72**, 152-159.
- Hunt, J. J. & Ballard, J. D., 2013. Variations in virulence and molecular biology among emerging strains of *Clostridium difficile*. *Microbiol Mol Biol Rev*, **77**, 567-581.
- Ivarsson, M. E., Leroux, J.-C., Castagner, B., 2015. Investigational new treatments for *Clostridium difficile* infection. *Drug Discovery Today*, **20**, 602-608.
- Javed, M. A., Poshtiban, S., Arutyunov, D., Evoy, S., Szymanski, C. M., 2013. Bacteriophage receptor binding protein based assays for the simultaneous detection of *Campylobacter jejuni* and *Campylobacter coli*. *PLoS One*, **8**, e69770.
- Jun, J. W., Giri, S. S., Kim, H. J., Yun, S. K., Chi, C., Chai, J. Y., Lee, B. C., Park, S. C., 2016. Bacteriophage application to control the contaminated water with *Shigella*. *Sci Rep*, **6**, 22636.
- Kenny, J. G., McGrath, S., Fitzgerald, G. F., van Sinderen, D., 2004. Bacteriophage Tuc2009 encodes a tail-associated cell wall-degrading activity. *J Bacteriol*, **186**, 3480-3491.
- Khanna, S., Baddour, L. M., Huskins, W. C., Kammer, P. P., Faubion, W. A., Zinsmeister, A. R., Harmsen, W. S., Pardi, D. S., 2013. The epidemiology of *Clostridium difficile* infection in children: a population-based study. *Clin Infect Dis*, **56**, 1401-1406.

Bibliography

- Killmann, H., Braun, M., Herrmann, C., Braun, V., 2001. FhuA barrel-cork hybrids are active transporters and receptors. *Journal of bacteriology*, **183**, 3476-3487.
- Klumpp, J., Dorscht, J., Lurz, R., Biemann, R., Wieland, M., Zimmer, M., Calendar, R., Loessner, M. J., 2008. The terminally redundant, nonpermuted genome of *Listeria* bacteriophage A511: a model for the SPO1-like myoviruses of gram-positive bacteria. *J Bacteriol*, **190**, 5753-5765.
- Knight, D. R., Elliott, B., Chang, B. J., Perkins, T. T., Riley, T. V., 2015. Diversity and Evolution in the Genome of *Clostridium difficile*. *Clin Microbiol Rev*, **28**, 721-741.
- Koc, C., Xia, G., Kuhner, P., Spinelli, S., Roussel, A., Cambillau, C., Stehle, T., 2016. Structure of the host-recognition device of *Staphylococcus aureus* phage varphi11. *Sci Rep*, **6**, 27581.
- Kostyuchenko, V. A., Leiman, P. G., Chipman, P. R., Kanamaru, S., van Raaij, M. J., Arisaka, F., Mesyanzhinov, V. V., Rossmann, M. G., 2003. Three-dimensional structure of bacteriophage T4 baseplate. *Nat Struct Biol*, **10**, 688-693.
- Kutateladze, M. & Adamia, R., 2008. Phage therapy experience at the Eliava Institute. *Med Mal Infect*, **38**, 426-430.
- Leiman, P. G., Arisaka, F., van Raaij, M. J., Kostyuchenko, V. A., Aksyuk, A. A., Kanamaru, S., Rossmann, M. G., 2010. Morphogenesis of the T4 tail and tail fibers. *Virology Journal*, **7**, 355.
- Li, X., Koç, C., Kühner, P., Stierhof, Y.-D., Krismer, B., Enright, M. C., Penadés, J. R., Wolz, C., Stehle, T., Cambillau, C., 2016a. An essential role for the baseplate protein Gp45 in phage adsorption to *Staphylococcus aureus*. *Scientific reports*, **6**, 26455.
- Li, X., Koc, C., Kuhner, P., Stierhof, Y. D., Krismer, B., Enright, M. C., Penades, J. R., Wolz, C., Stehle, T., Cambillau, C., Peschel, A., Xia, G., 2016b. An essential role for the baseplate protein Gp45 in phage adsorption to *Staphylococcus aureus*. *Sci Rep*, **6**, 26455.
- Lobley, C. M., Sandy, J., Sanchez-Weatherby, J., Mazzorana, M., Krojer, T., Nowak, R. P., Sorensen, T. L., 2016. A generic protocol for protein crystal dehydration using the HC1b humidity controller. *Acta Crystallogr D Struct Biol*, **72**, 629-640.
- Mahony, J., McDonnell, B., Casey, E., van Sinderen, D., 2016a. Phage-Host Interactions of Cheese-Making Lactic Acid Bacteria. *Annu Rev Food Sci Technol*, **7**, 267-285.

Bibliography

- Mahony, J., Stockdale, S. R., Collins, B., Spinelli, S., Douillard, F. P., Cambillau, C., van Sinderen, D., 2016b. *Lactococcus lactis* phage TP901-1 as a model for Siphoviridae virion assembly. *Bacteriophage*, **6**, e1123795.
- Mahony, J. & Van Sinderen, D., 2012. Structural aspects of the interaction of dairy phages with their host bacteria. *Viruses*, **4**, 1410-1424.
- Mahony, J. & van Sinderen, D., 2015. Gram-positive phage-host interactions. *Front Microbiol*, **6**, 61.
- Marti, R., Zurfluh, K., Hagens, S., Pianezzi, J., Klumpp, J., Loessner, M. J., 2013. Long tail fibres of the novel broad-host-range T-even bacteriophage S16 specifically recognize Salmonella OmpC. *Molecular microbiology*, **87**, 818-834.
- Martinez, F. J., Leffler, D. A., Kelly, C. P., 2012. *Clostridium difficile* outbreaks: prevention and treatment strategies. *Risk Manag Healthc Policy*, **5**, 55-64.
- Maxwell, K. L., Hassanabad, M. F., Chang, T., Paul, V. D., Pirani, N., Bona, D., Edwards, A. M., Davidson, A. R., 2014. Structural and Functional Studies of gpX of *Escherichia coli* Phage P2 Reveal a Widespread Role for LysM Domains in the Baseplates of Contractile-Tailed Phages. *Journal of bacteriology*, **196**, 2122.
- McPherson, A. & Cudney, B., 2014. Optimization of crystallization conditions for biological macromolecules. *Acta Crystallogr F Struct Biol Commun*, **70**, 1445-1467.
- MEADOW, P. M. & WELLS, P. L., 1978. Receptor sites for R-type pyocins and bacteriophage E79 in the Core Part of the lipopolysaccharide of *Pseudomonas aeruginosa*PAC1. *Microbiology*, **108**, 339-343.
- Merrigan, M. M., Venugopal, A., Roxas, J. L., Anwar, F., Mallozzi, M. J., Roxas, B. A., Gerding, D. N., Viswanathan, V. K., Vedantam, G., 2013. Surface-layer protein A (SlpA) is a major contributor to host-cell adherence of *Clostridium difficile*. *PLoS One*, **8**, e78404.
- Nale, J. Y., Spencer, J., Hargreaves, K. R., Buckley, A. M., Trzepinski, P., Douce, G. R., Clokie, M. R., 2016. Bacteriophage Combinations Significantly Reduce *Clostridium difficile* Growth In Vitro and Proliferation In Vivo. *Antimicrob Agents Chemother*, **60**, 968-981.
- O'Flynn, G., Ross, R. P., Fitzgerald, G. F., Coffey, A., 2004. Evaluation of a cocktail of three bacteriophages for biocontrol of *Escherichia coli* O157:H7. *Appl Environ Microbiol*, **70**, 3417-3424.

Bibliography

- Oleastro, M., Coelho, M., Gião, M., Coutinho, S., Mota, S., Santos, A., Rodrigues, J., Faria, D., 2014. Outbreak of *Clostridium difficile* PCR ribotype 027-the recent experience of a regional hospital. *BMC infectious diseases*, **14**, 209.
- Olia, A. S., Casjens, S., Cingolani, G., 2007. Structure of phage P22 cell envelope-penetrating needle. *Nat Struct Mol Biol*, **14**, 1221-1226.
- Paul, V. D., Rajagopalan, S. S., Sundarrajan, S., George, S. E., Asrani, J. Y., Pillai, R., Chikkamadaiah, R., Durgaiyah, M., Sriram, B., Padmanabhan, S., 2011. A novel bacteriophage Tail-Associated Muralytic Enzyme (TAME) from Phage K and its development into a potent antistaphylococcal protein. *BMC microbiology*, **11**, 226.
- Pirnay, J. P., De Vos, D., Verbeken, G., Merabishvili, M., Chanishvili, N., Vaneechoutte, M., Zizi, M., Laire, G., Lavigne, R., Huys, I., Van den Mooter, G., Buckling, A., Debarbieux, L., Pouillot, F., Azeredo, J., Kutter, E., Dublanchet, A., Gorski, A., Adamia, R., 2011. The phage therapy paradigm: pret-a-porter or sur-mesure? *Pharm Res*, **28**, 934-937.
- Pruitt, R. N. & Lacy, D. B., 2012. Toward a structural understanding of *Clostridium difficile* toxins A and B. *Front Cell Infect Microbiol*, **2**, 28.
- Quiberoni, A., Stiefel, J., Reinheimer, J., 2000. Characterization of phage receptors in *Streptococcus thermophilus* using purified cell walls obtained by a simple protocol. *Journal of applied microbiology*, **89**, 1059-1065.
- Rakhuba, D., Kolomiets, E., Dey, E. S., Novik, G., 2010. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol. J. Microbiol*, **59**, 145-155.
- Ricagno, S., Campanacci, V., Blangy, S., Spinelli, S., Tremblay, D., Moineau, S., Tegoni, M., Cambillau, C., 2006. Crystal structure of the receptor-binding protein head domain from *Lactococcus lactis* phage bIL170. *J Virol*, **80**, 9331-9335.
- Rodriguez-Palacios, A., Borgmann, S., Kline, T. R., LeJeune, J. T., 2013. *Clostridium difficile* in foods and animals: history and measures to reduce exposure. *Anim Health Res Rev*, **14**, 11-29.
- Rupnik, M., Wilcox, M. H., Gerding, D. N., 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol*, **7**, 526-536.
- Ryan, B. J. & Henehan, G. T., 2013. Overview of approaches to preventing and avoiding proteolysis during expression and purification of proteins. *Curr Protoc Protein Sci*, **Chapter 5**, Unit5 25.

Bibliography

- Sao-Jose, C., Baptista, C., Santos, M. A., 2004. *Bacillus subtilis* operon encoding a membrane receptor for bacteriophage SPP1. *J Bacteriol*, **186**, 8337-8346.
- Schmelcher, M. & Loessner, M. J., 2014. Application of bacteriophages for detection of foodborne pathogens. *Bacteriophage*, **4**, e28137.
- Schofield, D. A., Sharp, N. J., Westwater, C., 2012. Phage-based platforms for the clinical detection of human bacterial pathogens. *Bacteriophage*, **2**, 105-283.
- Sciara, G., Blangy, S., Siponen, M., Mc Grath, S., van Sinderen, D., Tegoni, M., Cambillau, C., Campanacci, V., 2008. A topological model of the baseplate of lactococcal phage Tuc2009. *J Biol Chem*, **283**, 2716-2723.
- Shan, J., Patel, K. V., Hickenbotham, P. T., Nale, J. Y., Hargreaves, K. R., Clokie, M. R., 2012. Prophage carriage and diversity within clinically relevant strains of *Clostridium difficile*. *Appl Environ Microbiol*, **78**, 6027-6034.
- Sillankorva, S. M., Oliveira, H., Azeredo, J., 2012. Bacteriophages and Their Role in Food Safety. *International Journal of Microbiology*, **2012**, 1-13.
- Simpson, D. J., Sacher, J. C., Szymanski, C. M., 2016. Development of an Assay for the Identification of Receptor Binding Proteins from Bacteriophages. *Viruses*, **8**.
- Sorensen, M. C., van Alphen, L. B., Harboe, A., Li, J., Christensen, B. B., Szymanski, C. M., Brondsted, L., 2011. Bacteriophage F336 recognizes the capsular phosphoramidate modification of *Campylobacter jejuni* NCTC11168. *J Bacteriol*, **193**, 6742-6749.
- Spinelli, S., Bebeacua, C., Orlov, I., Tremblay, D., Klaholz, B. P., Moineau, S., Cambillau, C., 2014a. Cryo-electron microscopy structure of lactococcal siphophage 1358 virion. *J Virol*, **88**, 8900-8910.
- Spinelli, S., Campanacci, V., Blangy, S., Moineau, S., Tegoni, M., Cambillau, C., 2006. Modular structure of the receptor binding proteins of *Lactococcus lactis* phages. The RBP structure of the temperate phage TP901-1. *J Biol Chem*, **281**, 14256-14262.
- Spinelli, S., Veessler, D., Bebeacua, C., Cambillau, C., 2014b. Structures and host-adhesion mechanisms of lactococcal siphophages. *Front Microbiol*, **5**, 3.
- Stockdale, S. R., Mahony, J., Courtin, P., Chapot-Chartier, M. P., van Pijkeren, J. P., Britton, R. A., Neve, H., Heller, K. J., Aideh, B., Vogensen, F. K., van Sinderen, D., 2013. The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate

Bibliography

- forms of their tail fiber into their virions for infection specialization. *J Biol Chem*, **288**, 5581-5590.
- Tasteyre, A., Barc, M. C., Collignon, A., Boureau, H., Karjalainen, T., 2001. Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect Immun*, **69**, 7937-7940.
- Taylor, N. M., Prokhorov, N. S., Guerrero-Ferreira, R. C., Shneider, M. M., Browning, C., Goldie, K. N., Stahlberg, H., Leiman, P. G., 2016. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature*, **533**, 346-352.
- Till, M., Robson, A., Byrne, M. J., Nair, A. V., Kolek, S. A., Shaw Stewart, P. D., Race, P. R., 2013. Improving the success rate of protein crystallization by random microseed matrix screening. *J Vis Exp*.
- Tremblay, D. M., Tegoni, M., Spinelli, S., Campanacci, V., Blangy, S., Huyghe, C., Desmyter, A., Labrie, S., Moineau, S., Cambillau, C., 2006. Receptor-binding protein of *Lactococcus lactis* phages: identification and characterization of the saccharide receptor-binding site. *J Bacteriol*, **188**, 2400-2410.
- Vaishnavi, C., 2015. Fidaxomicin-the new drug for *Clostridium difficile* infection. *The Indian journal of medical research*, **141**, 398.
- van Raaij, Mark J., Le, S., He, X., Tan, Y., Huang, G., Zhang, L., Lux, R., Shi, W., Hu, F., 2013. Mapping the Tail Fiber as the Receptor Binding Protein Responsible for Differential Host Specificity of *Pseudomonas aeruginosa* Bacteriophages PaP1 and JG004. *PLoS One*, **8**, e68562.
- Vedantam, G., Clark, A., Chu, M., McQuade, R., Mallozzi, M., Viswanathan, V. K., 2012. *Clostridium difficile* infection. *Gut Microbes*, **3**, 121-134.
- Veesler, D., Robin, G., Lichiere, J., Auzat, I., Tavares, P., Bron, P., Campanacci, V., Cambillau, C., 2010. Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): a baseplate hub paradigm in gram-positive infecting phages. *J Biol Chem*, **285**, 36666-36673.
- Veesler, D., Spinelli, S., Mahony, J., Lichiere, J., Blangy, S., Bricogne, G., Legrand, P., Ortiz-Lombardia, M., Campanacci, V., van Sinderen, D., Cambillau, C., 2012. Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. *Proc Natl Acad Sci U S A*, **109**, 8954-8958.
- Vegge, C. S., Brondsted, L., Neve, H., Mc Grath, S., van Sinderen, D., Vogensen, F. K., 2005. Structural characterization and assembly of the distal tail structure of the temperate lactococcal bacteriophage TP901-1. *J Bacteriol*, **187**, 4187-4197.

Bibliography

- Vegge, C. S., Vogensen, F. K., Mc Grath, S., Neve, H., van Sinderen, D., Brondsted, L., 2006. Identification of the lower baseplate protein as the antireceptor of the temperate lactococcal bacteriophages TP901-1 and Tuc2009. *J Bacteriol*, **188**, 55-63.
- Vinga, I., Baptista, C., Auzat, I., Petipas, I., Lurz, R., Tavares, P., Santos, M. A., Sao-Jose, C., 2012. Role of bacteriophage SPP1 tail spike protein gp21 on host cell receptor binding and trigger of phage DNA ejection. *Mol Microbiol*, **83**, 289-303.
- Walter, M., Fiedler, C., Grassl, R., Biebl, M., Rachel, R., Hermo-Parrado, X. L., Llamas-Saiz, A. L., Seckler, R., Miller, S., van Raaij, M. J., 2008. Structure of the receptor-binding protein of bacteriophage det7: a podoviral tail spike in a myovirus. *J Virol*, **82**, 2265-2273.
- Wang, J., Hofnung, M., Charbit, A., 2000. The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of *Escherichia coli* K-12. *Journal of bacteriology*, **182**, 508-512.
- Waseh, S., Hanifi-Moghaddam, P., Coleman, R., Masotti, M., Ryan, S., Foss, M., MacKenzie, R., Henry, M., Szymanski, C. M., Tanha, J., 2010. Orally administered P22 phage tailspike protein reduces salmonella colonization in chickens: prospects of a novel therapy against bacterial infections. *PLoS One*, **5**, e13904.
- Washizaki, A., Yonesaki, T., Otsuka, Y., 2016. Characterization of the interactions between *Escherichia coli* receptors, LPS and OmpC, and bacteriophage T4 long tail fibers. *Microbiologyopen*, **5**, 1003-1015.
- Wittebole, X., De Roock, S., Opal, S. M., 2014. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. *Virulence*, **5**, 226-235.
- Xia, G., Corrigan, R. M., Winstel, V., Goerke, C., Grundling, A., Peschel, A., 2011. Wall teichoic Acid-dependent adsorption of staphylococcal siphovirus and myovirus. *J Bacteriol*, **193**, 4006-4009.
- Yap, M. L., Klose, T., Arisaka, F., Speir, J. A., Veesler, D., Fokine, A., Rossmann, M. G., 2016. Role of bacteriophage T4 baseplate in regulating assembly and infection. *Proceedings of the National Academy of Sciences*, **113**, 2654-2659.
- Zhu, D.-Y., Zhu, Y.-Q., Xiang, Y., Wang, D.-C., 2005. Optimizing protein crystal growth through dynamic seeding. *Acta Crystallographica Section D: Biological Crystallography*, **61**, 772-775.

Bibliography

Zucca, M., Scutera, S., Savoia, D., 2013. Novel avenues for *Clostridium difficile* infection drug discovery. *Expert opinion on drug discovery*, **8**, 459-477.