Study of O-acetyltransferase B (*oacB*) and three novel *orfs* encoded by Sf101 bacteriophage of *Shigella flexneri*

Munazza I. Rajput

June 2021

A thesis submitted for the degree of Doctor of Philosophy at
The Australian National University

Australian
National
University

Declaration

I declare that this thesis is my original work and the results presented within this thesis are, except otherwise acknowledged, the outcome of the research conducted by myself while enrolled as a PhD candidate at the Australian National University.

Munazza I. Rajput

June 2021

Dedicated in the loving memories of my father

S.M Sharif Farooqui

1934 -2009

Acknowledgments

I am thankful to my Creator, who helped me sustained in all the difficult times. Secondly, my Ph.D. supervisor Assoc professor Naresh Verma, who allowed me to work in his laboratory and follow my dream. His untiring support and patience throughout my project helped me complete my Ph.D. During all the hard times he was always there to give me advice. Without his support, I would not have been able to achieve my goal.

I would like to thank my supervisory panel members, Dr. Carolyn Behm and Dr. Ulrike Mathesius, for their valuable inputs for my project. Their guidance and encouragement supported me during all the bad times. I am very grateful to Professor Ryland Young (USA) and Dr. Jeremy Bar (Monash University) for their help via emails on my bacteriophage-related aim of the Ph.D. project.

I would like to thank Dr. Denisse and Dr. Joe Brock from RSB for the timely help while I was troubleshooting my experiments. My big thanks to my mentors Dr. Kia Chan, Dr. Simon William, and Dr. Christina Spry, who also helped me troubleshoot research experiments and guided me for careers post-PhD. I am grateful to Dr. Benjamin Schwessinger and Dr. Ashley Jones for their help in sequence and assembly of the lysogenic strain for my bioinformatics analysis.

Sincere thanks must go to my lab members, Samiya, Pamodha, Dr. Pawan, and Tanuka, who helped me till the end of my Ph.D. I am also thankful to all my friends at the Research School of Biology (RSB); Dr. Nazia, Shagufta, Dr. Renate, Holly, Kiran Javed, Haseena Khan, Aditya Yadav, and Sanduni Hapuarachchi who were always there to listen to my problems and supported me morally. I would like to thank Dr. Dawar (CSIRO) who helped me during starting years of my Ph. D by helping me analyse my vectors and all his support.

My thanks to Dr. Farid Rahimi, for all technical support during my experiments and proofreading of my manuscript and thesis. Thanks very much for all your moral support. I

would like to thank Jenni Hayward and Dr. Stephenson Fairweather for proofreading my thesis chapters. I would also thank Uschi and Graham, who were always there to cheer me up in times of need. Uschi's help along with Sanduni at the time of my 17th-week miscarriage is unforgettable.

My thanks to the Statistical unit -RSB; Dr. Teresa Neeman, Robert Cope, and Hwan-Jin Yoon for their kind help in analyzing my data. Dr. Candida Spence and Jason Murdoch for helping me putting my thesis together. I like to thank the whole RSB IT team for their help throughout my Ph.D. especially during COVID-19 lockdown time.

My Ph.D. funding was provided by the Islamic development bank (IDB) and RSB. I can never forget how my beautiful university (The Australian National University-ANU) helped me continue my studies even after my primary scholarships exhausted and provided me with the HDR-Fee remittance scholarship till the completion of my Ph.D. I would like to thank my parent university 'University of Sindh, Jamshoro, Pakistan, for granting me study leave to pursue a Ph.D. at ANU. I am extremely thankful for travel support from the VC-Chancellor travel grant to travel to the USA and present my work at American Society for Microbiology conference.

On personal note, I would like to thank my family, who was big support system during my Ph.D. All this would not have been possible without your unwavering love and support. I love to say thanks to my late father, who gave me dream to pursue PhD., and my mother, back home whose prayers never let me quit even when it was seemingly impossible to stand. Thank you, "Abu and Ami", for believing in me. My husband, Imran whose support, love, and commitment to my aim helped me achieve this target. In the end, I will thank my two little daughters, Shiza and Eshal, who strived hard throughout this journey, and they never let me feel, how badly they missed me during extended hours of my work in the lab. My thanks to my parents-in-law, my siblings and their kids whose constant support and prayers helped me struggle. Love you all so much.

Abstract

Shigella flexneri is the leading cause of bacillary dysentery in developing countries and is associated with significant morbidity and mortality. Bacteriophages are known to play important role in the pathogenesis of *S. flexneri* by encoding O-antigen-modifying genes involved in serotype conversion. Bacteriophage Sf101 encoded *O*-acetyltransferase B (*oacB*) is a serotype converting gene that adds an acetyl group at either 3/4 position of Rhamnose III of the O-antigen. In some serotype 1c strains *oacB* is carried by Sf101 bacteriophage integrated within *sbcB* gene on the host chromosome. In contrast, in several serotypes 1c strains, and other strains of other serotypes carrying 3/4-O-acetylation, the location of *oacB* gene is upstream of *adrA* gene within *proA*-*adrA* region, where all *gtr*-carrying phages integrated.

The first part of the study deals with the investigation of the topological features of OacB and the identification of important residues in OacB. For this purpose, multiple alignments of OacB were performed with other acetyltransferases from related bacterial species, and several conserved domains and motifs were identified. Site-directed mutagenesis carried out on the selected residues of the OacB revealed seven amino acids that were critical for the function of OacB.

In the second part of the study, the distribution of the *oacB* gene in the serotype 1c strains was investigated with reference to the Sf101 lysogenic strain (SFL1683). The complete genome sequence of SFL1683 was generated using MinION flow technology to develop a reference genome, and regions carrying *oacB* in 1c strains were investigated. To gain insight into the origin of *oacB* gene in serotype 1c strains, whole-genome sequences of strains collected from various geographical regions were inspected for the presence of Sf101 phage or its remnants. The results revealed that in only two lysogenic strains Sf101 phage integrated within *sbcB* gene, and in other strains within *proA*- *adrA* region. All the

analysed strains carried conserved Sf101 attachment sites within the *sbcB* gene, and *sbcB* gene was flanked by similar insertion sequences (IS). The genetic arrangement upstream *adrA* gene in SFL1683 was different from other 1c strains due to the absence of IS elements flanking *oacB* gene. It was also identified that the SfI phage attachment site shared six base pairs (bp) homology with the Sf101 attachment site. In the Y394 strain, the SfI phage attachment site was identified at three different locations within the *proA-adrA* region which might have helped the Sf101 phage to integrate into this region. The abundance of IS elements in this region was indicative of insertion or deletion events and resulted in the deletion of Sf101 phage leaving behind *oacB* gene.

In a previous study, the four novel *orfs (oacB/orf16*, *orf17*, *orf41,* and *orf56*) were found to have no phage-related functions and limited homologies with *Shigella* and *E. coli* proteins. The physiological effect of novel *orfs* encoded by Sf101 on the virulence of *S. flexneri* was also explored in the last part of the study. The lysogenic strain of Sf101 lacked a large virulence plasmid (VP) due to which virulence studies were performed in another serotype 1c strain harbouring VP. Three different virulence assays were employed involving *C. elegans* as an *in vivo* model and HeLa cells as an *in vitro* model. The results showed the presence of these *orfs* did not affect the virulence phenotype of the host. This study has provided a detailed characterization of *oacB* of serotype converting Sf101 phage of *S. flexneri* and opened avenues for the upcoming research to understand the serotype conversion in *S. flexneri*.

Publications

Published

Pawan Parajuli, Munazza I. Rajput and Naresh K. Verma: Plasmids of *Shigella flexneri* serotype 1c strain Y394 provide advantages to bacteria in the host. ***BMC Microbiology***, 2019, 19(1): 86

Under review

Munazza I. Rajput and Naresh K. Verma: Identification of critical residues of O-antigen modifying *O*-acetyltransferase B (OacB) of *Shigella flexneri.* BMC Molecular and Cell Biology 92021)**.**

In preparation

1) Distribution and acquisition of *O*-acetyltransferase (*oacB*) gene in *S. flexneri* 1c serotype.
2) Review: Transferases of *S. flexneri*

List of Abbreviations

aa	amino acid
ABC	ATP-binding cassette
Amp	ampicillin
ANU	Australian National University
attp	Attachment site of the phage
AP	Alkaline phosphatase
bp	base pairs
BCA	Bicinchoninic Acid Assay
BRF	Biomolecular Resource facility
BSA	bovine serum albumin
CDD	Conserved domain database
Cm	chloramphenicol
CFU	Colony-forming unit
DLA	Double layered agar

EDTA	ethylenediamine tetracaetic acid
DMEM	Dulbecco'd modified Eagle's medium
DNA	Deoxyribonucleic acid
EF	effector proteins
ETEC	Enterotoxigenic *E. coli*
GlcNAc	N-acetylglucosamine
Gtr	glucosyltransferase
HRP	Horse Radish Peroxidase
Hela	carcinoma of patient Henrietta Lacks
IgA	immunoglobulin A
IgG	immunoglobulin G
IL	interleukin
int	integrase
Ipa	Invasion plasmid antigen
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kb	kilobase
kDa	kilodalton
Kan	kanamycin
LB	Luria Bertani
LBA	Luria Bertani agar

LDC	lysine decarboxylation
LPS	lipopolysaccharides
M	Molar
mM	Millimolar
M cell	micro fold cell
mg	milligram
MQH_2O	Milli Q water
µl	microliter
ml	millilitre
NGM	Nematode growth medium
OD_{600} nm	optical density at 600 nm
Oac	*O*-acetyltransferase
OacB	*O*-acetyltransferase B
Orf	open reading frame
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEtN	Phosphoethanolamine
PFU	plaque forming unit
Phage	bacteriophage
PHASTER	phage search tool- enhanced release

PMN	polymorphonuclear
PVDF	polyvinylidene fluoride
RAST	rapid annotations using subsystem technology
Rha	rhamnose
rpm	revolution per minute
RT-PCR	reverse transcriptase polymerase chain reaction
SHI	*Shigella* pathogenicity island
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
STEC	Shiga toxin-producing *E. coli*
T3SS	type three secretion system
TM	transmembrane
T_m	Melting temperature
UDP	uridine Diphosphate
UV	ultraviolet
VP	virulence plasmid
WT	wild type
Xis	excisionase

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 General introduction

The members of the genus *Shigella* are Gram-negative, non-motile, nonsporulating, and facultative anaerobic bacilli belonging to the family *Enterobacteriaceae* [1]. There are four species of the genus *Shigella*: *S. flexneri, S. boydii, S. sonnei and S. dysenteriae*. All *Shigella* species are clinically important because they are causative organisms of shigellosis, which is the most infectious form of bacillary dysentery in humans and primates.

All four *Shigella* species are further subdivided into serotypes based on biochemical and serological differences. According to recent classification, S. *dysentriae* (serogroup A) has 15 serotypes; *S. flexneri* (serogroup B) has 19; *S. boydii* (serogroup C) has 20, and *S. sonnei* (serogroup D) has 1 serotype [2]. *Shigella* species account for 190 million shigellosis cases annually with 70,000 deaths [3]. Among the four species, *S. dysenteriae* is important for large regional epidemics (WHO, 2005) whereas, *Shigella flexneri* is the most prevalent causative agent of gastroenteritis in low-income countries. The most prevalent serotype of *S. flexneri* is 2a followed by 1b,3a, and 4a [4]. Most of the outbreaks occur in developing countries and affect children under the age of five years. The emergence of new serotypes hampers the development of vaccines [5]. Furthermore, the increasing antibiotic resistance emerging among *Shigella* species limits treatment options. *Shigella* spread in most developing countries is associated with unhygienic conditions, overcrowded areas with people and unavailability of clean drinking water [6]. The, the World Health Organization (WHO) has urged scientists to develop a vaccine for the prevention of shigellosis.

1.1.2 Shigellosis and symptoms

Shigellosis is a clinical syndrome associated with the acute onset of colorectal inflammation that leads to watery or mucoid diarrhea. Other manifestations of the disease include high fever, generalized toxicity, anorexia, vomiting, nausea, abdominal pain, and diarrhea accompanied with blood. The presence of erythrocytes, neutrophils, and mucus reflects the invasive character of *Shigella*. While the symptoms of shigellosis typically appear after 1-2

days of contact with the bacteria (Centre for Disease Control-CDC), the incubation period can range from 1-4 days [7-9]. In healthy individuals, the infection is self-limiting and resolves in 5-10 days. However, infection in immunocompromised and unhealthy children can lead to death due to a lack of nutrients /electrolytes and dehydration. Other manifestations of shigellosis include intestinal perforations, septicemia, and toxic megacolon [10].

1.1.3 Mode of transmission

Sir William Osler defined dysentery as among the four major epidemic diseases on the earth. The natural hosts for *Shigella* are humans. Shigellosis is a highly contagious disease and an infectious dose of as low as 10–100 viable *Shigella* cells can cause infection [9, 11]. The infection spreads in overcrowded areas with poor sanitary conditions or because of food or water being contaminated with the pathogen in under-developed countries. Transmission predominantly occurs through person-to-person contact by the faecal-oral route. Mechanical transmission of shigellosis also takes place by common houseflies in settings where removal of human excreta is not done properly [7]. Further, the transmission of *Shigella* also has recently been reported among men who have sex with men (MSM) [12].

1.1.4 Host immune response against Shigella spp

The pathogenesis of *Shigella* infection relies on the ability to circumvent the host's innate immune response. The immune response (acquired natural immunity) to *Shigella* infection is serotype-specific, and humoral immunity is considered an important aspect of protective immunity [13, 14]. The antibody response is largely directed against the O-antigen of lipopolysaccharide (LPS) whereas; in the case of natural infection, mucosal secretory immunoglobulin (Ig) A antibodies and serum IgG antibodies are also directed against the virulence invasion plasmid antigens (*ipa*). In the human *S. flexneri* 2a challenge model, antigen specific (LPS and IpaB) memory B cells are important in protecting the host from *Shigella* [15, 16]. Inflammation at the site of infection results in the overproduction of cytokines by macrophages (IL-1, TNF-α, IL-6, IFN-γ, TNF-β, IL-4, IL-10, TGF-β, and IL-8), which in turn recruit polymorphonuclear leukocytes (PMNs) from circulation at the site of

infection (gut). Entry of PMNs enables infiltration of more bacteria and results in mucosal inflammation to confine the infection to the submucosa [14]. To counteract this immune suppression, *Shigella spp*. releases a series of effectors like MkaD, OspG, OspI and others, at the site of infection inside the inside the intestinal cells (entrocytes) that help bacteria to escape from the host immune system and assist in intracellular proliferation. Moreover, type III secretory system (T3SS) effector IpgD has been reported to impair T-cell migration, which obstructs T-cell contact with the antigen-presenting cell and hence diminishes adaptive immune response [2, 17]. In another study by Sellge et al., (2010), it was observed that in a murine model $CD4^+$ Th17 cells confer limited protection against reinfection [18].

1.2 Virulence genes of *S. flexneri*

S. flexneri contains a highly conserved large (220 kb) virulence plasmid (VP) which encodes genes that contribute to invasion of the epithelial cells [19, 20]. Specifically, VP encodes genes that are involved in invasion, intracellular spread, and gene regulation (Figure 1.1). This single-copy plasmid is maintained in the bacterial population through the post-segregationally killing (PSK) mechanisms involving toxin: antitoxin (TA) systems. One of the characteristics of the plasmid is a 32 kb pathogenicity island (PAI) also called the *ipa-mxi-spa* locus. This locus is highly conserved and flanked by insertion sequences (IS) IS100 and IS600. The PAI region consists of 39 genes, which are arranged in two oppositely transcribed clusters [21]. The genes encoded by VP can be split into four groups based on their functions. The first group of PAI encoded proteins are constituents of the T3SS apparatus, which functions to inject bacterial effector proteins into the host cytoplasm. The expression of the T3SS is temperature-dependent and the best expressed temperatures is that of the human gastrointestinal tract. Injected T3SS effector proteins facilitate invasion, survival inside the epithelial cells and induction of pyroptosis (lytic programmed cell death due to inflammation) in the macrophages [22]. These T3SS effectors include invasion plasmid antigens IpaA to IpaD; IpaB, IpaC and IpaD are crucial virulence factors that help alter host cell functions to facilitate bacterial entry.

The second group of PAI encoded proteins are called membrane expression of ipa (mxi) and surface presentation of ipa (spa) antigens, which encode apparatuses for assembly and function of the T3SS. Group 3 consists of transcriptional activators, VirB and MxiE which assist T3SS-associated genes. Group 4 genes code for chaperones (ipgA, ipgC, ipgE and spa15) for the stability of T3SS substrates inside the cytoplasm of bacteria.

In addition to T3SS encoded by PAI, the virulence genes that modulate the host immune response post-invasion, include the outer *Shigella* proteins Osp (11 *osp* putative genes), sepA, and IpaH (five alleles) gene families. Whereas *virA*, *virG*, *sopA* and *phoN2* are involved in actin-guided movement of *Shigella*. The enterotoxin ShET2 is associated with the classic symptom of watery diarrhoea during shigellosis (Table 1.1)

1.2.1 Shigella pathogenicity islands (SHI)

Pathogenic bacteria like *Shigella* acquire pathogenicity islands (SHI) through horizontal gene transfer. In *S. flexneri,* five different pathogenicity islands have been reported. The first pathogenicity island, SHI-O, is comprised of bacteriophage-encoded genes (*oac, oacB* and *gtrs*) which determine the serotype of *Shigella* [23], and each serotype elicits a unique immune response. The second pathogenicity island, SHI-1, is 46.5 kb in size and encodes the genes *sigA, pica, set1A,* and *set1B,* located down-stream of the *pheV tRNA* gene. SHI-1 is responsible for intestinal fluid accumulation in a rabbit ileal loop model associated with mucus permeabilization, and development of watery diarrhoea [24]. The third pathogenicity island, SHI-2, is located within the *arg-mtl* region at the *selC* locus and encodes several genes including *iucA* to *iucD* (aerobactin operon) and *iutA*, which are involved in the iron acquisition using siderophore system and downregulation of inflammation by suppressing T-cell signaling. The fourth pathogenicity island, SHI-3, is present in *S. boydii* strain 0-1392. It encodes genes for the synthesis and transport of the hydroxamate siderophore aerobactin and is located within the 21-kb island between *lysU* and *pheU tRNA* gene [25]. The fifth pathogenicity island (*Shigella* resistance locus (SRL) located downstream of *serX tRNA* gene, encodes genes for antibiotic resistance and ferric dicitrate systems for iron acquisition.

SRL in *S. flexneri* 2a strain was found to be is 66.2 kb sequences comprises of 59 ORFs [26] (Table 1.2).

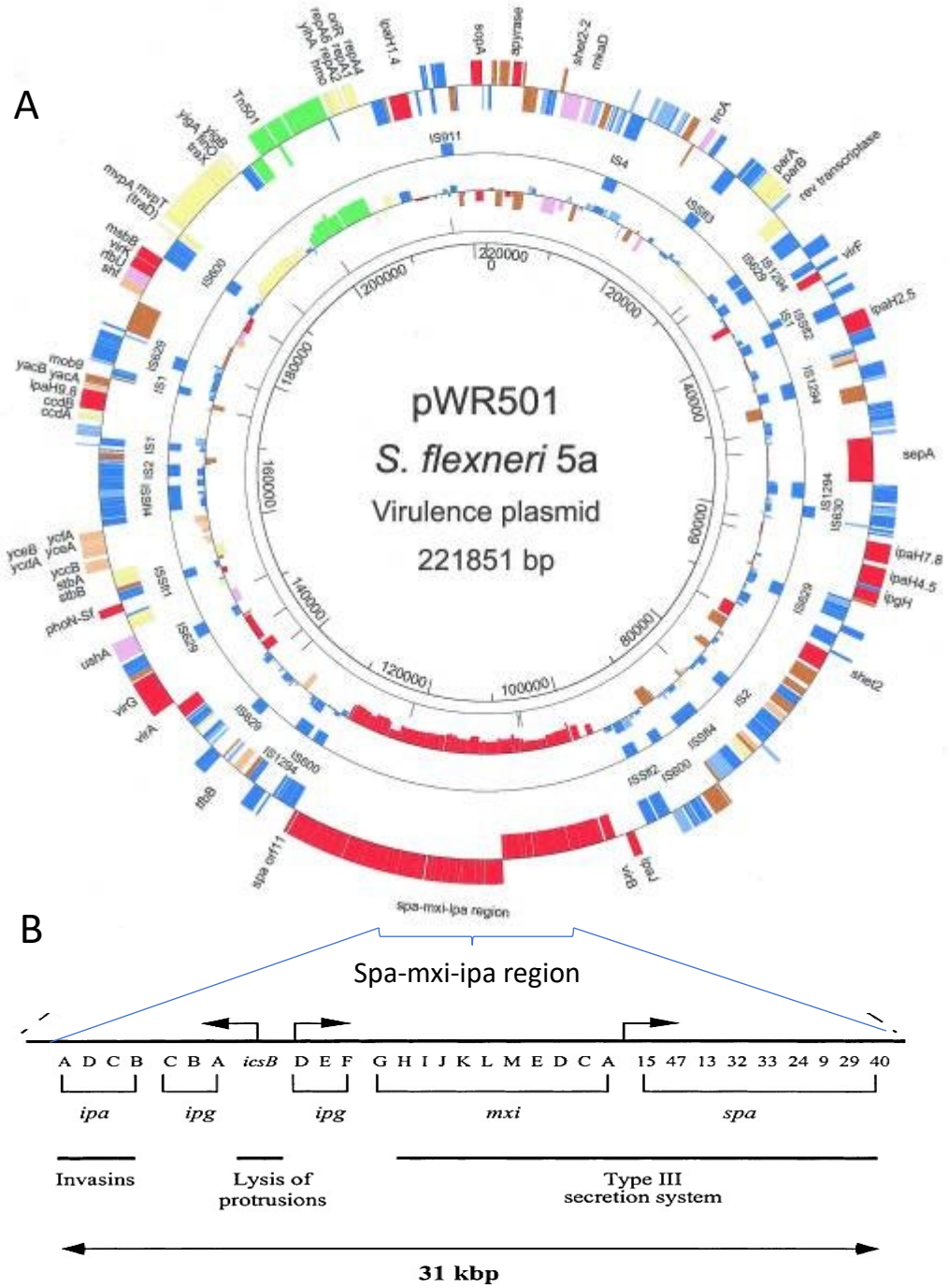


Figure 1.1: Circular map of the large (221 kb) virulence plasmid (VP) of *S. flexneri*.
A) The outermost ring contains virulence-associated proteins which are coloured in red. **B)** The 31 kb entry region of 221 kb virulence plasmid is shown in an enlarged view. Modified from [27] and [28].

Table 1.1 Virulence factors encoded on the *S. flexneri* virulence plasmid (220 kb). Adapted from [21]

Effector	Biochemical activity	Host cell target (s)	Virulence function and /or phenotype	References
IpaA	Vinculin activation	Vinculin, β_1-integrins, Rho signaling	Efficient invasion, actin cytoskeleton rearrangements, disassembly of cell-matrix adherence	[29], [30], [31], [32]
IpaB	Membrane fusion	Cholesterol, CD44, caspase 1	Control of type III secretion, translocon formation, phagosome escape, macrophage apoptosis	[33], [34, 35]
IpaC	Actin polymerization	Actin, β-catenin	Translocon formation, filopodium formation, phagosome escape, disruption of EC tight junctions	[33], [36]
IpaD	Effector protein		Control of type III secretion, membrane insertion of the translocon	[37]
IcsA (VirG)	Polymerization of Actin	N-WASP, vinculin	Recruitment of actin-nucleating complex required for actin-based motility and intercellular spread	[38]
IcsB			Camouflage of IcsA to prevent autophagic recognition	[39]
IcsP	Serine protease		Cleavage of IcsA, modulation of actin-based motility	[40, 41]
IpgB1	RhoG mimicry	ELMO protein	Induction of Rac1-dependent membrane ruffling	[42]
IpgB2	RhoA mimicry	RhoA ligands	Induction of actin stress fiber-dependent membrane ruffling	[43]
IpgD	Phosphoinositide 4-phosphatase	Phosphatidylinositol 4,5-bisphosphate	Facilitation of entry, promotion of host cell survival	[44]
OspB			T3SS substrate, unknown function	[45]
OspC1		Nucleus and cytoplasm	Induction of PMN migration	[46]
OspD1			T3SS substrate, unknown function in host cells, antiactivator of MxiE	[47]
OspE2		Focal contacts	Maintenance of EC morphology, efficient intercellular spread	[48]
OspF	Phosphothreonine lyase	MAPKs Erk and p38	Inhibition of histone phosphorylation and NF-κB-dependent gene expression, reduction of PMN recruitment	[49],[46]
OspG	Protein kinase, ubiquitination inhibitor	Ubiquitin-conjugating enzymes	Downregulation of NF-κB activation, reduction of inflammation	[50]

Effector	Biochemical activity	Host cell target (s)	Virulence function and /or phenotype	References
PhoN2[a]	Apyrase		Unipolar localization of IcsA	[45]
SepA[a]	Serine protease		Promotion of intestinal tissue invasion and destruction	[51]
VirA	Cysteine protease	α-Tubulin	Facilitation of entry and intracellular motility by the degradation of microtubules	[46]

1.2.2 Pathogenesis of Shigella

Shigella enters the human body via the faecal-oral route. After surviving the acidic milieu of the stomach, *Shigella* gains access to the small intestine where it needs to resist bile salts degradation. Interestingly, bile salts have recently been found to enhance the virulence of *Shigella* by induction of biofilm formation [52]. After reaching the colon, *Shigella* strives to bind to the thick mucus layer, which is at the greatest thickness in the colon as compared to other parts of the digestive tract [53]. The ability of *Shigella* to inhabit and colonize the intestinal epithelium is a crucial determinant of disease establishment. Pathogenesis is a multistep process (Figure 1.2) that starts with the invasion of the intestinal epithelium through a microfold cell (M-cell) from the basolateral side [54]. M cells are specialized epithelial cells within the follicle-associated epithelium (FAE), that can endocytose the luminal antigens and microorganisms to transfer them to the sub-epithelial spaces [55]. *Shigella* exploits M-cells to cross the physical protective barrier of the host intestinal epithelium via transcytosis and is released into the intraepithelial pocket of M-cells. *Shigella* then reaches the submucosa of the gastrointestinal tract and is engulfed by macrophages [56]. *S. flexneri* induces capase-1 dependent macrophage apoptosis by secreting virulence plasmid encoded invasion plasmid antigens (Ipa) IpaB, IpaC and IpaD and escapes into the cytoplasm [21, 53].

Macrophage apoptosis induces a severe immune response, which is mediated by the release of pro-inflammatory cytokines, IL-1β and IL-8. IL-1β secretion results in an influx of polymorphonuclear (PMN) cells at the site of infection, which destroys the integrity of the epithelial lining. While allowing PMN cells to access and phagocytose *Shigella,* this breach also paves the way for more bacteria to reach the sub-epithelial space without accessing M-cells. Indeed, the harmful intestinal inflammation that is the characteristic of shigellosis is associated with IL-1β, whereas IL-8 is responsible for the generation of a beneficial antibacterial response by activation of natural killer cells and production of interferon-gamma.

Following these responses, bacterium is now exposed to the basolateral side of an intestinal epithelial cell (IEC). The cytoplasm of the IEC serves as a replication niche and, after utilizing

the available nutrients of one IEC, *Shigella* invades the neighbouring cells [2]. Movement within the host cells is supported by the intracellular spread protein (IcsA). IcsA mediates the actin-based motility of *Shigella* by hijacking two host proteins: namely, the Wiskott-Aldrich syndrome protein (N-WASP) and the actin-related protein 2/3 (APR2/3) complex. This complex serves as the nucleation site and helps bacterial movement through the host cell cytoplasm [57]. Intracellular spread takes place through the same mechanism and initiates with the endocytosis of *Shigella* at tricellular tight junctions, and a new cycle of release-replication and spread takes place [2].

Moreover, *Shigella* secretes at least three types of effectors that promote the survival of IEC, enabling *Shigella* to multiply within the cell. Within IECs, IpaB controls mitotic arrest deficiency 2 like protein 2 (MAD2L2 or MAD2B) to arrest cell maturation, while IpgD stops cell death and OspE aids in detachment. [58].

Diarrhoea symptoms are associated with the production of two *Shigella* enterotoxin1 (ShET) encoded in the *Shigella* pathogenicity island 1 (SHI-1), while ShET2 is encoded by the virulence plasmid VP [2].

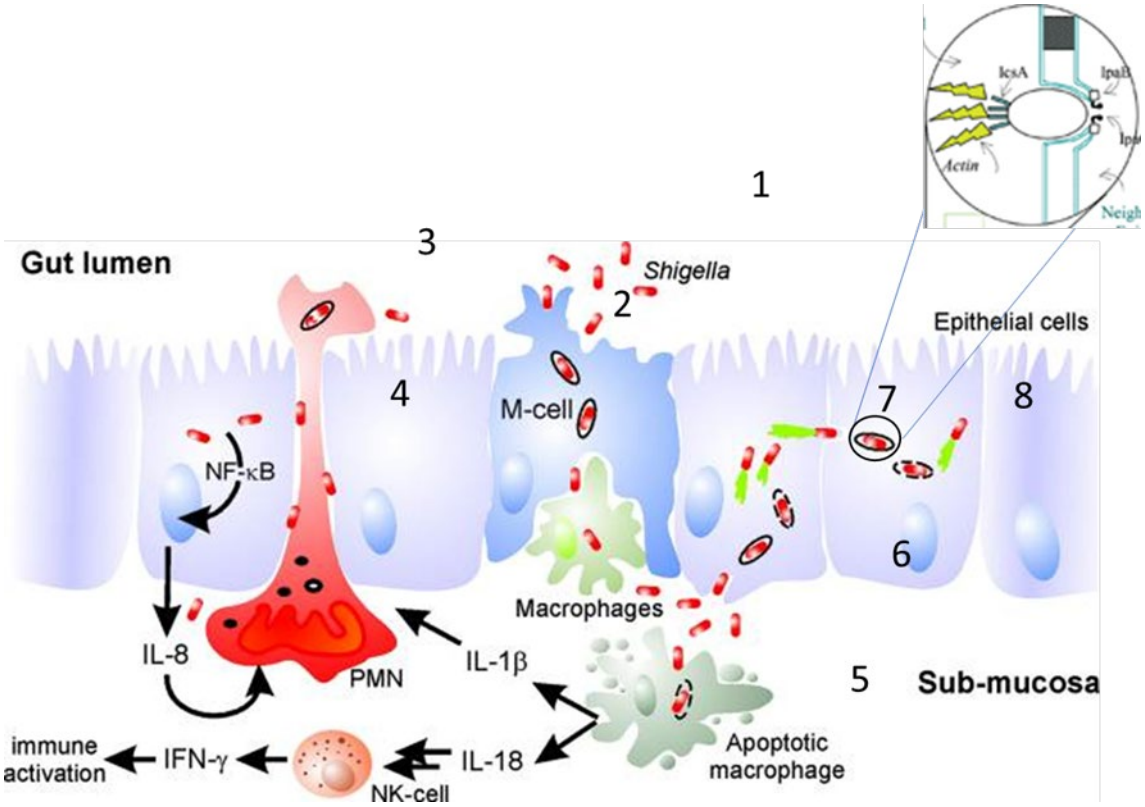


Figure 1.2: Cellular pathogenesis of S. *flexneri*.

1. Transcytosis through M cells
2. Infiltration of polymorphonuclear leukocytes (PMNs) to create gaps for *Shigella* entry
3. Manipulation of tight junction proteins to allow movement of bacteria into the submucosa
4. Inflammatory response due to macrophages apoptosis
5. Entry into the cell via basolateral membrane
6. Lysis of vacuole by IpaB and IpaC and release of *Shigella* into epithelial cell's cytoplasm and actin polymerization
7. Intercellular spread

Figure Adapted from [12, 17]

Table 1.2 Genes and protein functions involved in virulence on the *Shigella* chromosome. Modified from [59]

PAI	Gene(s)	Protein function	References
SHI-1	*sigA* *pic* *set1A, set1B*	Putative enterotoxin Intestinal colonization ShET1 enterotoxin	[24, 60] [61]
SHI-2	*iucA-D* *iutA* *shiA-G*	Siderophore, complexes with iron Bacterial receptor for iron-siderophore complex Novel ORFS, ShiA involved in a reduction in host inflammatory response	[62] [63]
SHI-3	*iccA-D* *iutA*	Siderophore, complexes with iron Bacterial receptor for iron-siderophore complex	[25]
Shi-O	*gtrA, gtrB, gtr*	Serotype conversion and O-antigen modification	[23]
SRL	*tetDCAR, cat, dhfrI and ant1*	Resistance to tetracycline, Chloramphenicol, trimethoprim, and streptomycin	[26]

1.2.3 Evolution of Shigella flexneri

Kiyoshi Shiga isolated and characterized *Shigella (S. dysenteria)* for the first time in 1897 from the stool sample of a patient during the dysentery epidemic in Japan and soon after, other species were discovered and added under the umbrella of genus *Shigella* [2]. Before 1987, S*higella* species were placed in a similar taxonomic position as the non-pathogenic *E. coli* but, due to biochemical differences between them, *Shigella* is placed in a separate major group. The early genomic era had divided *Shigella* into three major clades C1-C3, which arose as the lineages of *E. coli* 35000-270000 years ago. The C1 cluster contains the majority of S. *boydii*, *S. dysenterie,* and serotype 6 strains; C2 includes *S. boydii* 7 and one *S. dysenterie* strain, and C3 includes all *S. flexneri* except serotype 6 and one strain from *S. boydii*. *S. sonnei* is grouped with the *E. coli* clade [64]. It was also revealed that the sequence divergence between *Shigella* and *E. coli* K-12 is about 1.5% which is marginal when compared with 15% of *Salmonella enterica* (a close relative of *E. coli*) [21]. However, free-living *E. coli* lacks virulence factors that make *Shigella* a human restricted pathogen [2].

Bacteria have the ability to transform from a non-pathogenic to a pathogenic state by gaining new genetic material which allows them to gain new phenotypic traits. Mobile genetic elements (MGEs), phages, and plasmids play an important role in such genetic modification; for instance, the acquisition of a large 120-140 MDa virulence plasmid (VP) and chromosomal pathogenicity islands (PAI) are the hallmarks in the evolution of various *Shigella spp*. because the VP encodes genes that are important for invasion and intracellular survival. While the transition from a free-living to an intracellular lifestyle offers many benefits, *Shigella* has important catabolic genes which were present in ancestral *E. coli* strains; for example, *Shigella* shed genes of the *lac o*peron which are required for lactose metabolism; *ompT* and lysine decarboxylase have similarly been either deleted or inactivated due to their interference with *Shigella* spread and hinderance in enterotoxin activity due to *cadA* (encodes lysine decarboxylation) [65, 66]. Likewise, *argT* was also deleted because of its inhibitory effect on the invasive capacity of *Shigella* [2]. Over time, *Shigella* also lost surface structures like fimbriae and flagella, potentially to avoid host immune system activators and thereby facilitating epithelial cells invasion [67, 68].Tominaga et al. (1994) identified expression of flagellar genes from *Shigella* in *E. coli* K-12. Figure 1.3 depicts the evolutionary events in the development of *Shigella* from commensal *E. coli*.

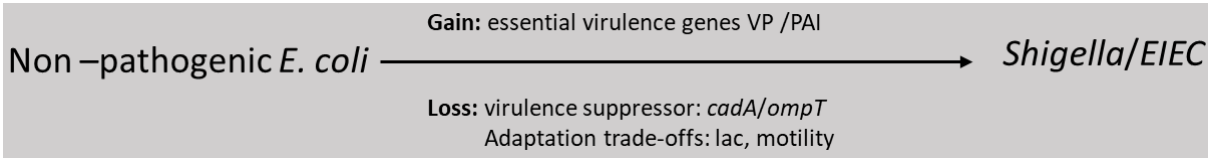


Figure 1.3: Evolutionary events during the development of *Shigella* /EIEC.
Prominent events include the gain of virulence plasmid and virulence genes (pathogenicity islands). Other important events involved loss of genes; *cadA* and other genes, which are not needed by the bacteria to survive in the niche like motility and utilization of the lactose. Modified from [69].

The study of the pathogenic forms of *E. coli* such as Enteroinvasive *E. coli* (EIEC) can help understand the evolution of *Shigella.* EIEC causes dysentery like *Shigella*. Indeed, both organisms are genetically, biochemically, and pathogenically related to each other and it is difficult to distinguish *Shigella* from EIEC. Both pathogens are considered as pathovars of *E. coli* and EIEC are precursors of full-blown *Shigella* strains [70]**.** However, some useful biochemical tests differentiate *Shigella* from EIEC, for example, most *Shigella spp* are lactose, mucate, and acetate negative whereas EIEC may be positive for one or all the properties. Moreover, xylose fermentation and production of gas are positive for EIEC and not in *Shigella* [70]. It is reported that *Shigella* has emrged from *E. coli* ancestors in several independent events and EIEC forms a distinguishing pathovar [69].

1.3 Treatment, prevention, control, and vaccine development

In recorded incidents of shigellosis, oral rehydration is the treatment of choice to replace the fluids and electrolytes lost because of diarrhea. Patients in shock or coma and those with severe vomiting are treated with intravenous replacement of fluid and electrolytes [71]. Young children, adults, and immune-deficient individuals with severe infections require antibiotics treatment, with the drugs of choice being either fluoroquinolones (ciprofloxacin) first-line or β-lactams (ceftriaxone) and macrolides (azithromycin) second line [72, 73]. Antibiotic therapy of shigellosis reduces the duration of fever, diarrhea, and fecal excretion of the pathogen [71]. However, the alarming problem of antibiotic-resistant and evolving serotypes of *Shigella spp* emphasize the need to develop an effective vaccine. A vaccine against a single serotype of *Shigella* will not be sufficient, rather, a multivalent vaccine (having more than one serotype) is needed to combat shigellosis caused by different serotypes of *Shigella spp.*[74]. Designing a successful vaccine, that is efficacious against all serotypes with conformation epitopes of their O polysaccharide antigens continues to present a great challenge. Several vaccine development strategies have been utilized since 1940 but no licensed vaccine is yet available for field release.

1.3.1 Progress in vaccine development

The previous infection by a wild type *Shigella* can confer resistance to the subsequent infections, this provides a compelling explanation that vaccination is a possible strategy for the prevention of shigellosis [7]. Shigellosis affects children under 5 years of age in developing countries and if children are subjected to recurrent infection they become susceptible to abnormal physical and cognitive development [75]. The unavailability of the appropriate animal model(s) and evolving *Shigella* serotypes have contributed to the lack of a licensed vaccines for bacillary dysentery.

1.3.1.1 Cellular candidates

Mutation in the genes involved in virulence or metabolic processes represents a strategy to design attenuated vaccines [76]. Live attenuated bacteria are highly immunogenic and offer broad antigenic exposure to the human immune system. The two examples in this regard are *virG* and *guaBA*-based mutants. *virG* is a toxin gene and required for cellular dissemination of bacteria. The Walter Reed Army Institute of Research (WRAIR) has designed a vaccine against a strain with an essential mutation in *virG (icsA)* gene. The vaccine has been shown to evoke a strong immune response along with host protection. The Centre for Vaccine Development (CVD) at the University of Maryland, Baltimore has created a series of *Shigella* strains (*S. flexneri* 2a 2457T) containing mutations in the gua*BA* gene which encodes enzymes for bacterial metabolism and also created deletion mutants of enterotoxin genes *sen* and *set* genes [5]. Clinical trials have shown this vaccine induces a strong immune response and has a greater safety profile.

1.3.1.2 Glycoconjugate candidates

Laboratory of Developmental and Molecular Immunology, USA has proposed the idea of a vaccine directed against *Shigella* O polysaccharides complexed to a protein carrier. Clinical trials have discovered that *S. flexneri* serotype 2a LPS conjugated to recombinant Pseudomonas exoprotein A (rEPA) protein elicits a serotype-specific immune response in humans [14]. Another bioconjugate vaccine used synthetic oligosaccharides conjugated to a carrier protein,developed by the Pasteur Institute has shown promising results in the preclinical trials [77]. A recombinant glycoconjugate vaccine developed by Limatech Biologics is in phase 2b clinical evaluation [78].

1.3.1.3 Novel antigen candidate

Outer membrane vesicles (OMV) are also being used as vaccine candidates using Generalized Modules of Membrane Antigens (GMMA), which is a high-yielding vesicle technology. These vaccines contain 40% LPS and 60% of outer membrane protein antigens including OmpA, OmpC/OmpF and, IpaB-D. In preliminary trials, the vaccine candidates have shown 65-100% protection in mouse studies [76].

1.3.1.4 Subunit candidates

To achieve homologous and heterologous protection, the purified T3SS proteins, *IpaB* and *IpaD* were delivered with an adjuvant to a *Shigella* infected mouse model. These vaccines protected animal models and will progress to clinical trials. Another subunit vaccine candidate is a 34 kDa outer membrane protein (OmpA) of *Shigella* which has shown protection in an animal model [79].

1.4 Serotype conversion in *Shigella flexneri*

The characteristic component of the cell wall of Gram-negative bacteria is LPS, which is exposed on the cell surface and contributes to the integrity of the outer membrane. LPS is

also one of the main virulence determinants as it protects from host defenses like complement-mediated killing [80]. LPS is composed of three regions: 1) the membrane anchor lipid A; 2) the core oligosaccharide; 3) a polymer of glycosyl (repeat) units known as the O-antigen (Figure 1.4).

1.4.1 Lipid A

Lipid A is the hydrophobic domain of LPS embedded in the outer membrane (Figure 1.4). This glucosamine-based phospholipid moiety is highly conserved in enteric bacteria [81]. Lipid A is responsible for the endotoxic activity of the bacteria and causes fever, diarrhoea, or septic shock [82].

1.4.1.1 Core

The core oligosaccharide is a negatively charged region of LPS that joins lipid A to the O-antigen (Figure 1.4). The core is divided into two parts, the inner core is conserved and closer to lipid A, while the outer core is present between the inner core and O-antigen is less conserved [82].

1.4.1.2 O-antigen

The O-antigen is the most distal part of LPS and carries a tri-rhamnose (Rha)-N-acetyl glucosamine tetra-saccharide subunit [83] (Figure 1.4). Each chain of O-antigen varies in length consisting of up to 40 repeat units [84]. O-antigen protrudes to the external surface of the cell and is a highly immunogenic and prime target of host defenses. O-antigen triggers the complement activation [85]. Compared to lipid A and the core A, the O-antigen is a highly diverse molecule. The diversity in O-antigen structure is based on the composition, sequences, and linkages of sugar residues, which provide the source for serotyping [86]. Moreover, variation is also contributed by the substitution of monomers with either sugar or non-sugar residues [87]. Due to variations in O-antigen, *S. flexneri* is divided into 19 serotypes and except for serotype 6, all the serotypes have the same O-antigen structure.

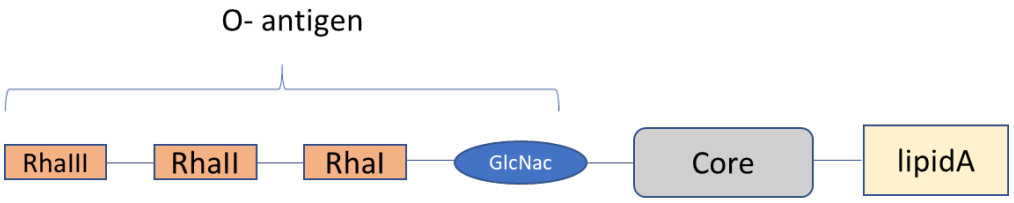


Figure 1.4: Structure of LPS of Gram-negative bacteria.
LPS consists of lipid A, core polysaccharide and, the outer most O-antigen. GlcNac= N-acetylglucosamine; Rha= rhamnoses

1.5 Synthesis of O-antigen

O-antigen synthesis takes place by a series of complex biochemical mechanisms and the genes associated with synthesis are located at the *rfb* locus. There are three model pathways for O-antigen synthesis: the Wzx/Wzy dependent pathway, the ATP binding cassette (ABC) transporter dependent pathway, and the synthase pathway. The first, two pathways will be discussed in detail in the following paragraphs.

In both pathways, the first sugar is added to a lipid carrier undecaprenyl phosphate (UndP) by the enzyme polyisoprene-phosphate N-acetylaminosugar-1-phosphate transferase (PNPT) [88]. However, Wzx/Wzy and ABC transport-dependent pathways differ in the cellular location where O-antigen chain polymerization occurs.

In the Wzx/Wzy pathway, synthesis begins at the cytoplasmic face of the inner membrane. Here, UndP bound sugars are translocated to the periplasmic side of the inner membrane by the flippase activity of Wzx [88]. Later pre-formed O-antigen units are joined by the Wzy polymerase to generate a long O-antigen chain. Here the polymer grows in sets with the addition of new repeat units at the inner core oligosaccharide of lipid A [89]. The chain length regulator, Wzz then directs the length of the growing chain [86] and the assembly process ends when the elongated O-antigen chain ligates to lipid A core polysaccharide complex by O-antigen ligase WaaL. The LPS-O antigen complex is transported to the outer membrane by

a group of seven conserved lipopolysaccharide transport proteins (Lpt) LptB2FGCADE which together make a protein bridge to transfer LPS from the inner to the outer membrane (Figure 1.5). The ATPase LptB from the ATP binding cassette (ABC) transporter family (LptB2FG) in the inner membrane hydrolyses ATP to provide energy to drive LPS through the periplasmic bridge. Meanwhile, the single transmembrane helix protein, LptC in the inner membrane makes a complex with LptB2FG, receives LPS from LptF-LptG and passes it to LptA. LptA helps connect the inner membrane complex to the outer membrane translocon, formed by LptD-LptE. At this stage, LPS is accepted by LptD-LptE and transferred to the outer leaflet of the outer membrane [89] (Figure 1.5).

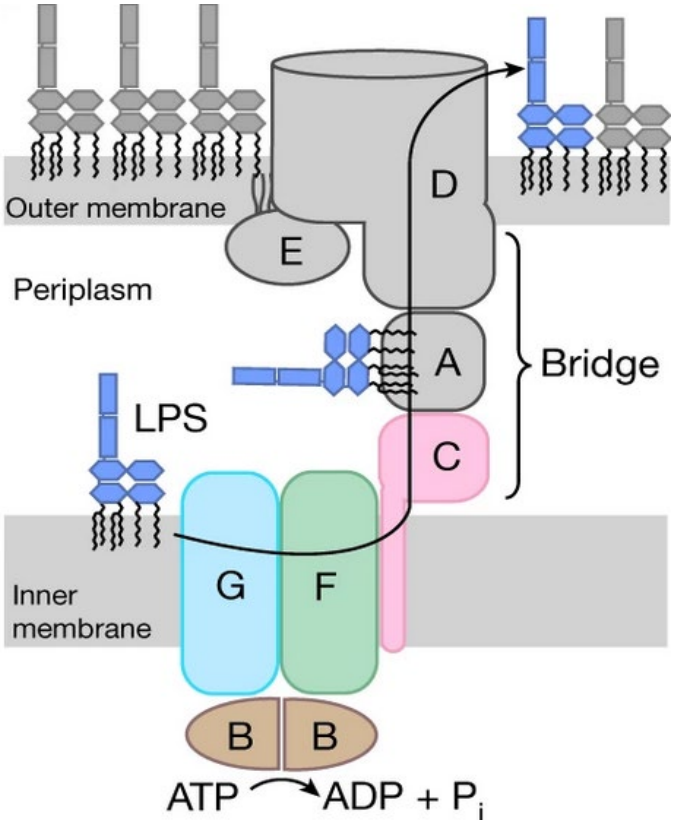


Figure 1.5: A schematic showing the LptB2FGCADE components of the LPS transport machinery.
The LPS moves from the inner membrane to the outer leaflet of the outer membrane with the help of seven lipopolysaccharide transport proteins (Lpt). In the process inner membrane component proteins (containing ATP-binding cassette) LptB2 (provides energy), LptF and, LptG form a steady sub-complex with an anchoring LptC, containing a single transmembrane helix. Following receipt of LPS from LptF and LptG, LptC transfers LPS to LptA which is positioned centrally and provides a connection between the inner membrane and LptD and LptE (outer membrane proteins). Image is taken from [90].

The ABC transporter-dependent pathway involves a similar initiation as Wzx/Wzy pathway; however, the mode of O-antigen polymerization differs between the two pathways. In the ABC transporter pathway, polymerization takes place in the cytoplasm by serotype-specific glycosyltransferase enzymes [91] without the use of Wzy polymerase. Furthermore, entire O-antigen polymers are translocated across the inner membrane by the ATP binding cassette (ABC) transporter [89].

1.6 Shigella flexneri O- antigen and structural diversity

The basic backbone structure of O-antigen in serotype Y consists of N-acetylglucosamine - rhamnose-rhamnose-rhamnose tetrasaccharide, which is common for all serotypes except serotype 6, which has N-acetylgalactose amine in place of N-acetylglucosamine and three rhamnoses [86]. When O-antigen basic subunits undergo modifications by the addition of glucosyl/acetyl groups (phage-encoded) or phosphoethanolamine (plasmid-encoded) residues new serotypes evolve. Due to differences in O-antigen structure, 19 serotypes of *S. flexneri* have been reported (Figure 1.6). Like in other Gram-negative bacteria, the highly immunogenic nature of the O-antigen of *S. flexneri* induces a strong protective immunity [23]. Moreover, every serotype elicits a unique immune response and subsequent infection by one serotype protects against the same serotype but not against other serotypes [92]. Three types of O-antigen altering processes change the antigenic signatures of *S. flexneri*: namely, O-glucosylation, O-acetylation, and the addition of phosphoethanolamine.

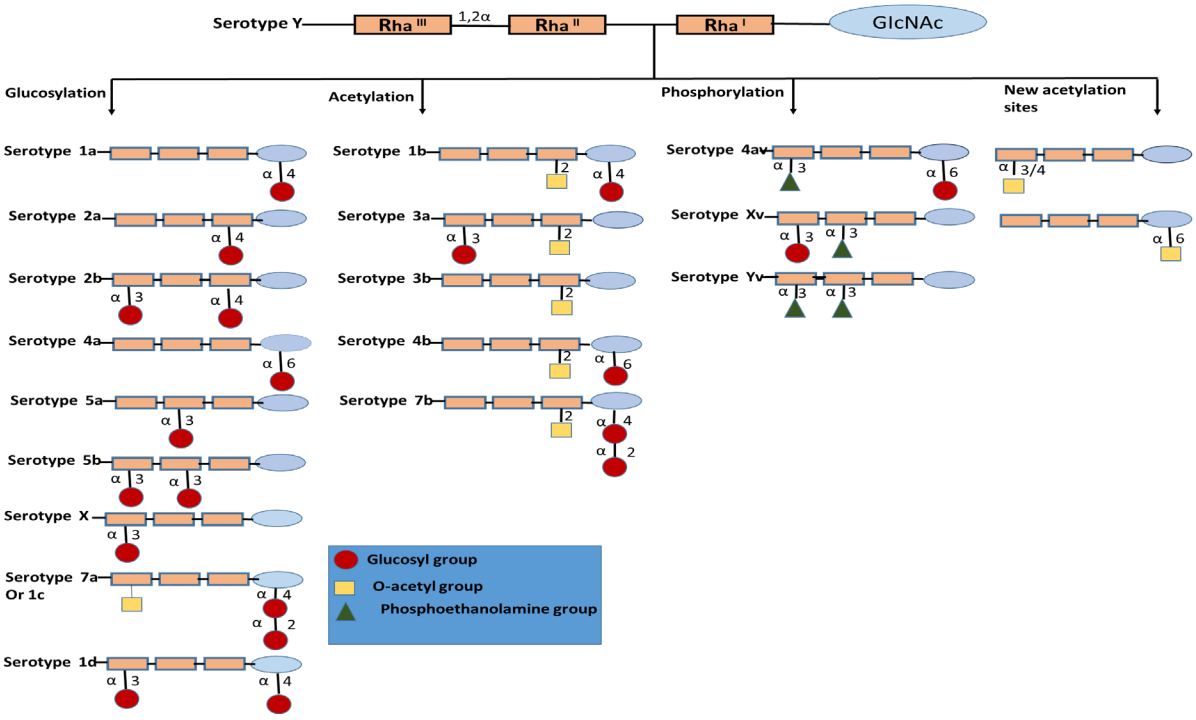


Figure 1.6: Serotypes of *S. flexneri* and their O-antigen structures.
Serotype Y (basic O-antigen structure) comprises repeating units of the rhamnose-rhamnose-rhamnose-N-acetylglucosamine tetra saccharide. The addition of glucosyl, *O*-acetyl, or phosphoethanolamine groups to different sugars within the tetrasaccharide repeat units through linkages create different serotypes. GlcNac: N-acetylglucosamine, Rha: rhamnose.

1.6.1 O-antigen glucosylation

Glucosylation is mediated by a cluster of three genes called the *gtr* cluster encoded by temperate bacteriophages (Figure 1.7). Currently, five phages (SfI, SfII, SfIV, SfV, and SfX) are known to encode *gtr* clusters in *S. flexneri* serotypes 1a, 2a, 4a, 5a, and X respectively [93-97]. The first two genes of the operon (*gtrA* and *gtrB*) are highly conserved and interchangeable between serotypes, whereas, the third gene, *gtr (type),* is serotype-specific [23, 92]. The product of *gtrA* is a 120 amino acid transmembrane (TM) domain-containing protein GtrA, which is comprised of four TM helices with both N- and C- termini in the cytoplasm. GtrA is a hypothetical protein, and its proposed function is flipping the UndP-glucose precursor from the cytoplasm to the periplasm [93, 98].

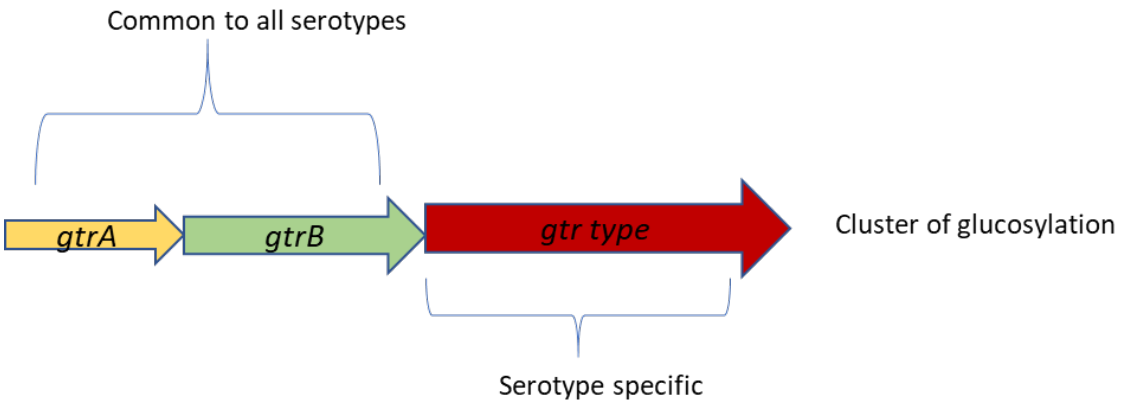


Figure 1.7: Organization of the O-antigen glucosylation gene cluster in *S. flexneri*.
Conserved *gtrA* and *gtrB* encode glucosyltransferases. By contrast, *gtr(type)* encodes for a glucosyltransferase that is serotype specific. The direction of arrows indicates the orientation of the genes. [gtr: glucosyltransferase].

The GtrB protein encoded by *gtrB* is a two membrane helices protein with varying numbers of amino acids from 305-309 [93]. GtrB has been shown experimentally to add UDP-glucose to the bactoprenol carrier in the cytosol to yield the UndP-glucose precursor [93, 94]. The third gene of the operon, *gtr (type),* encodes a serotype-specific glucosyltransferase (Gtr) membrane protein. While the size of the Gtr proteins depends on serotype (ranging from 416-526 amino acids), they generally have 9-11 putative TM regions and are up to 60% identical at the protein level [99]. Gtr proteins are involved in glucosylating specific rhamnose of the O-antigen by a special linkage [98]. Most of the serotypes of *S. flexneri* have evolved by the addition of glucosyl groups to one or multiple sugars. The serotype specific Gtr proteins differ in their primary structure and this diversity enables them to recognize different substrates to glucosylate. According to Whitfield et al., (1995), the secondary structure of Gtr (type) has similarity to flipases, O-antigen polymerases, and O-antigen ligases. Figure 1.8 summarises the proposed mechanism of O-antigen glucosylation.

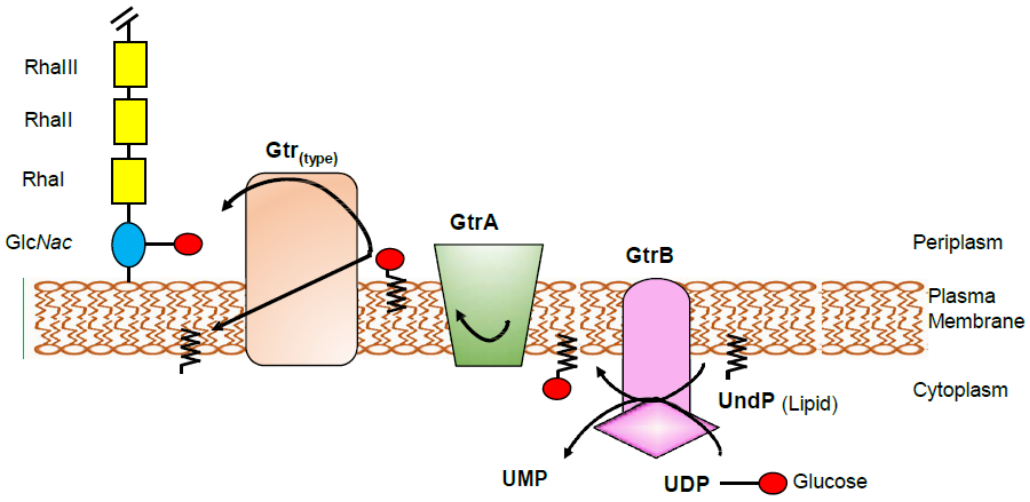


Figure 1.8: The hypothetical model of O-antigen glucosylation.
GtrB transfers a glucosyl group from UDP-glucose to the membrane-linked-lipid UndP. The GtrA protein itself or in conjunction with the Gtr(type), flips the lipid-associated glucose into the periplasm. The Gtr(type) protein then specifically transfers the glucosyl residue to a particular site on the O-antigen chain and returns the bactoprenol to the cytoplasmic side of the membrane. GlcNAc, N-acetylglucosamine; Rha, Rhamnose; UDP, uridine diphosphate; UndP, undecaprenyl phosphate. Image taken from [86, 100].

1.6.2 O-antigen acetylation

O-acetylation genes coding for O-acetyltransferase (Oac) enzymes are carried by the temperate bacteriophages. Gemski et al., (1975) reported for the first time that Sf6 phage could be isolated from the 3a strain of *S. flexneri* [101].

Later on, a 1002 bp size gene coding for the 333 amino acid protein Oac was identified by Verma et al., in 1991 [102]. Sf6 Phage genome analysis revealed that *oac* genes map next to the phage integrase gene and are transcribed as a monocistronic unit [103]. The Sf6 phage integrated at *argW tRNA* gene next to a conserved *yfdC* gene [104]. Harbouring the Sf6 encoded *oac* gene converts serotypes X, Y, 1a, and 4a to 3a, 3b, 1b, and 4b, respectively, and brings about 2-O-acetylation at Rhamnose I molecule of the O-antigen. Oac is a 333 aa TM protein with 10 alpha-helical membrane-spanning regions. Another transferase, *O*-acetyltransferase B *(oacB*) encoded by the Sf101 phage, adds an acetyl residue at either position 3 or 4 of Rhamnose III (3/4-O-acetylation) in serotypes 1a, 1b, 2a, 5a, 1c, Y, and 6 at position 6 of N-acetyl glucosamine (6-O-acetylation) in serotypes 2a, 3a, Y and Yv of the O-antigen subunits [105]. The Sf101 phage has recently been isolated from two 1c strains of *S.*

flexneri and in both the strains, the *oacB* is found next to the *sbcB* gene. However, the location of the *oacB* gene in some 1c strains and strains of serotype 1a, 1b, 2a, 5a, and Y is upstream of the *adrA* gene in the *proA-adrA* region [105, 106].

So far, five acetyltransferases, Oac [103], OacB [105], OacC [107], OacD [108], and Oac1b [109] have been reported in *S. flexneri*. Homologs of OacB from different acetyltransferases of *S. flexneri* are grouped under protein superfamily COG1835 (predicted acetyltransferases) having acyl_tranf_3 (AT3) conserved domain. Not much is known about how *O*-acetyltransferases acetylate O-antigen in *S. flexneri.*

Recently, topological features of O-acetyltransferase enzyme OafA from *Salmonella enterica* have been identified and O-acetylation model by Pearson et al., has been proposed [110]. OafA comprised of two interconnected domains: AT3 (transmembrane) and SGNH (periplasmic), which mutually perform O-acetylation of O-antigen residues (Figure 1.9). AT3 domain translocates acetyl group across the membrane and SGNH domain attaches it to the monosaccharide of O-antigen. The 3D structure of SGNH domain of OafA has been solved whereas crystal structure of AT3 domain needs to be determined yet. OafA shares similarities with O-acetyltransferases of *S. flexneri* [110, 111].

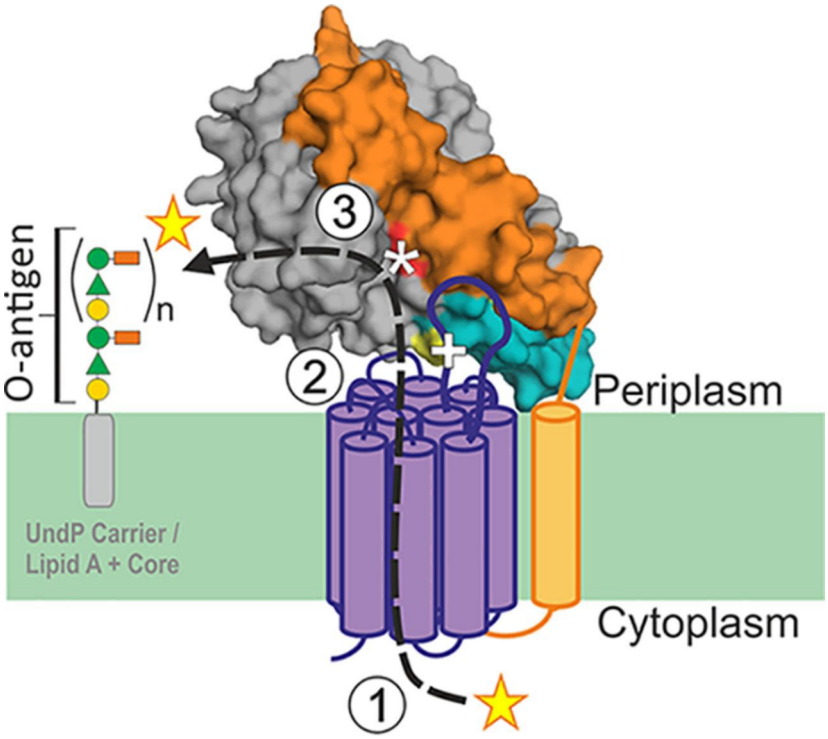


Figure 1.9: O-acetylation model of O-antigen acetyltransferases in *Salmonella*.
Acetyltransferase 3 domain (AT3) present as a transmembrane domain contains ten transmembrane domains (purple) and one extra linking region (teal). The extra membrane domain SGNH is shown in grey. (1) AT3 domain interacts with the cytoplasmic acetyl group donor and after processing of acetyl group transfers it to the periplasmic side of the inner membrane. (2) Conserved residues Asparginine and Serine transfer acetate to SGNH domain. (3) Acetate attachment to specific O-antigen residue is catalysed by SGNH domain. An asterik denotes active side of the SGNH domain and interaction is denoted by a plus sign Adapted from [110].

Investigation of topological features and catalytic residues of OacB will help understand the mechanism of action of this enzyme and O-antigen modification in *S. flexneri*.

1.6.3 O-antigen phosphoethanolamine

In 2009, Perepelov *et al*., identified the phosphoethanolamine (PEtN) modification of the O-antigen of *S. flexneri* [112]. The polymorphic *opt* gene encoding for PEtN transferase is carried by 6.85 kb plasmids and phosphorylates Rha II and Rha III in subtypes 4av, Xv, Yv, and yv1 and confers E1037 antigenic determinant. There are two alleles of *opt*, *optII* and *optIII* borne by double-stranded circular plasmids, pSfxv_2 and pSFyv_2 respectively. Opt is a 506 amino acid protein with four TM helices and belongs to the sulfatase superfamily. The sulfatase domain at the C- terminus of Opt putatively catalyses the transfer of PEtN to sugar residues of the O-antigen [113]. It has been reported that phosphorylation interferes with other O-antigen modifications [114]. PEtN modifications have also been reported in *N. meningitidis*, *Hemophilus inflenza*, *S. typhimurium*, *E. coli*, *Pasteurella multocida and Campylobacter jejuni* [112].

1.7 New serotypes of *S. flexneri*

Since the 1990s, numerous new serotypes such as 1c and Xv of *S. flexneri* have been reported [115]. Serotype 1c was first identified in 1989 in Bangladesh and afterward was found to be prevalent in Egypt, Indonesia, Vietnam, Pakistan, China, and the UK [116]. Serotype 1c is an emerging serotype in developing countries and bears a unique O-antigen structure with two glucosyl groups attached to N-acetyl-glucosamine residue [1, 117]. The *gtr(type)* gene mediates glucosylation of the O-antigen; however, the complete bacteriophage encoding *gtrIc* has not been identified so far in the *S. flexneri* genome. A variant of 1c serotype presents a slightly different structure of O-antigen after the addition of acetyl group at position 3 (major) or 4 (minor) of Rha III.

Serotype Xv (SFxv) first appeared in the Henan province of China in 2001, and then in 2003 became the most dominant serotype by replacing serotype 2a. Later in 2007, it reached other parts of China (Shanxi, Gansu, and Shanghai), and became the dominant serotype in these areas of China [118]. Afterward, a new untypeable serotype 4s was identified in Beijing, China

[118]. This serotype initially was identified as the clone of Xv strain because of similarities in some biochemical reactions. However, the differences were found in their agglutination reactions: specifically, SFxv was able to agglutinate with anti-7,8 group antisera, whereas 4s failed to react. Another serotype 4av, which is a variant of serotype 4a, has a PEtN group linked at the position I of Rhamnose residues [119]. Later analysis of the SFxv strain, LPS structure by NMR spectroscopy reveals the presence of PEtN group at position 3 of one of the rhamnose residues [120].

While the PEtN modification is encoded by the *opt* gene, present on a 6.8 kb plasmid, (pSFxv_2), the genes allowing glucosyl (*gtr*) and O-acetyl (*oacB*) modifications are encoded by bacteriophages. Temperate bacteriophages have thus been found to play a vital role in O -antigen modifications in *S. flexneri* [121].

1.8 Bacteriophages

Bacteriophages which infect and replicate only in bacterial cells, are the most abundant organisms on the earth and are present in every ecological niche. Phages were discovered by William Twort in 1915, and in 1917, Felix d'Herelle identified a microbe that caused their lysis in liquid media and caused discrete patches (plaques) on the solid media seeded with bacteria [122, 123].

A bacteriophage is made up of proteins that encapsulate nucleic acid in the form of either DNA or RNA. The virus genome is very variable and has different forms and sizes: for example, phage MS2 is, an RNA virus that has a genome of a few kilobases, whereas some DNA phages like T4 have large genomes with ≥100 genes. Moreover, the size of the genome determines the dimensions of the capsid, with larger genomes requiring larger capsids [124]. Based on their replication, phages are divided into lytic phages and temperate phages.

1.8.1 Lytic phages

The life cycle of virulent phages is based only on the lysis of the host cell. The cycle begins with the invasion of a bacterial cell. The phage then replicates by exploiting the host cell machinery to produce new phage particles. Programmed cell lysis occurs when the host cell becomes overburdened, releasing the phage particles and allowing them to re-infect neighbouring bacteria, and hence the cycle repeats [125].

1.8.2 Temperate phages

Lysogenic or temperate phages can integrate their genome into the bacterial genome as an inert prophage and replicate passively with their host as a unit. The infected bacteria or the lysogen reproduce and live normally without any harm, and the prophage is transmitted to each daughter cell at every cell division. Moreover, prophages frequently encode ‘morons’, which are transcriptionally independent units of DNA that can express while the phage is in the prophage state. Morons can increase the virulence of the host directly in the form of toxins or indirectly by enhancing the ecological fitness of hosts during the infection process [126]. Lysogens can remain in this state for generations but can re-enter the lytic cycle through a process of induction; inducing agents like UV radiation and the mutagen mitomycin trigger the cell DNA damage response (SOS response), leading to the formation and assembly of phage proteins which in turn results in the destruction of host cells with the release of phage particles [104, 127]. Lambda phages contain ~63 open reading frames and are an excellent example of temperate phages. Lambda phages can transport several important genes between bacteria by the process of transduction. This can occur either through generalised or specialized transduction. In generalised transduction, the phage mistakenly incorporates bacterial DNA instead of phage nucleic acid. On subsequent infection of another bacterial cell, the DNA is injected into the cytosol and combines with the host chromosome. In the case of specialised transduction, imprecise excision of the prophage from the bacterial genome results in the stealing of adjacent bacterial gene(s) by the phage and ultimately transfer to another lysogenic host [126].

1.8.3 Role of phages in bacterial pathogenicity

The lysogenic phage can drive bacterial diversity during its stay in the bacterial genome by carrying extra genes in their genome with each integration event [126]. Changes in the phenotype of the host due to phage-encoded genes, like the change from non-pathogenic to a pathogenic strain or an increase in virulence, are called lysogenic conversion [128]. Studies have shown that up to 20% of the genome of pathogenic bacteria are made up of integrated prophages and phage-encoded genes [129]. Protein products of phage-encoded genes can help bacteria to invade host tissues, avoid immune defenses, and damage host cells. Therefore, the acquisition of phage genes confers a selective advantage to the bacterial host for its survival and clonal expansion, which indirectly contributes to the dissemination of phage genes [130]. The lysogenic conversion was first observed in the 1950s in *Corynebacterium diphtheriae* when an avirulent strain became virulent after phage transduction of a potent toxin [126]. Several phage-encoded toxins and virulence factors have been reported in Gram-negative, as well as Gram-positive bacteria, (Table1.4), many of which play important roles in human diseases. Common diseases caused by bacteria carrying prophage-encoded toxins include botulism, diphtheria, cholera, and those associated with Shiga toxigenic *E. coli* e.g. *E. coli* 0157, STEC [131].

The most studied and well-characterized among these are the cholera toxin (CT) and Shiga toxin. *Vibrio cholera* is the etiological agent of cholera and the toxin responsible for the disease is encoded by a single-stranded DNA filamentous CTX phage. CT is an A-B type two-component exotoxin associated with the watery diarrhoea typical of cholera [132, 133].

Shiga toxins (Stx1 and Stx2) are important virulent determinants of STEC and are encoded by a group of Stx phages. Stx1 and 2 are among the most potent biological poisons and are responsible for the accumulation of fluid in rabbit ileal loops and renal damage in mice. Human infection is associated with diarrhoea, hemorrhagic colitis, and hemolytic uremic syndrome caused by *S. dysenteria* 1 *or E. coli 0157:H7* strains [134, 135].

Bacteriophages encode other virulence factors that can enhance bacterial virulence during the infection process. There is a growing list of such virulence factors (Table 1.4), which include T3SS effector proteins involved in the invasion process such as SopE2 (SopE phage), SspHI (Gifsy-3 phage), and gogB (Gifsy-1 phage). Stx phage encodes *lom* gene which helps *E. coli* bind to epithelial cells [136-139]. Some phage-encoded virulence factors contribute to the survival of bacteria in their hosts such as SodC1 (Gifsy-2 phage) which mediates bacterial defense against oxidative burst [130], *SseI* and *gipA* in *Salmonella enterica* help bacteria to survive in macrophages and Payer patches, respectively [126]. Stx phages encode *bor* gene, which helps bacteria to evade the immune system [140].

Many bacteriophages encode O-antigen modification genes which help a bacterial host to escape host defenses. Each gene decorates O-antigen with a unique antigenic epitope. Bacteriophages that have been isolated and characterized in *S. flexneri* include SfI, II, IV, V X, Sf6, and Sf101 [94, 96, 104, 105, 121, 141-143].

Table 1.3 Examples of prophage-associated virulence factors. Modified from [126]

Bacteria	Phage	Phage-encoded virulence genes	Reference
C. diptheriae	Beta	Diphtheria toxin (*tox*), cytotoxin	[144]
E. coli	Stx	Shiga toxin (*stx_1, stx_2*), cytotoxins	[145]
		stk—affects signal transduction	[146]
		TTSS effectors *cif, espI/nleA, espI, espK, espEU/tccP, nleI*	[147]
	λ	*lom*—binding to epithelial cells	[148]
		bor—outer membrane protein that aids bacterial immune evasion.	[149]
	CP-933C	Cryptic phage regulates TTSS	[150]
S. enterica	φSopE	TTSS effector (*sopE*) promotes invasion of epithelial cells	[136]
	Gifsy-1/2/3	*gipA, gogB*- survival and growth, sodC*1*- survival in macrophages and *sspHI*-TTSS effector	[137], [138],[139]
P. aeruginosa	D3	Altered outer membrane properties reduce phagocytosis	[151]
S. mitis	SM1	*pblA and pblB*—platelet binding	[152]
C. jejuni	CJIE1	Increased adherence and invasion	[153]
V. cholerae	CTX	*ctx*—cytotoxin	[154]
S. flexneri	SfI, SfII, SfV, SfX, Sf6, Sf101	*gtrI, gtrII, gtrV, gtrX, oac, oacB*-O-antigen modification	[94],[93],[96, 105, 121], [105]

1.8.4 Bacteriophages of Shigella flexneri

S. flexneri hosts several phages and, to date, seven phages, (SfI, SfII, SfIV, SfV, SfX, Sf101, Sf6), have been identified as involved in serotype conversion of serotype Y strain to serotype 1a, 2a, 4a, 5a, X, 1c,3b, and 4a respectively whereas, the Mu phage is a new addition in *S.*

flexneri phages [94, 96, 104, 105, 121, 141-143]. Morphologically, SfI, SfII, SfIV, SfV, and SfMu belong to the Myoviridae family of viruses, whereas SfX, Sf101, and Sf6 are members of Podoviridae family. Members of Myoviridae family have an icosahedral head and long contractile tail, while Pododviridae members bear an icosahedral head with a short tail. Except for the Mu phage, which is a mu-like phage, all other S. *flexneri* phages contain modular genetic organization of lambdoid phages. Almost all the *S. flexneri* have been sequenced and characterised. Phages SfI, SfII, SfIV, SfV, SfX, Sf101, Sf6 encode for *gtrI, II, IV, V, X, oacB,* and *oac* genes, respectively, and their genome size ranges from 37k to 39k base pairs.

1.8.4.1 Integration of serotype converting phages

Phages integrate their DNA into the host in two ways i) through transposition (Mu phage) or ii) site-specific recombination-SSR (lambda phages). SSR is responsible for both insertion and excision of prophages into and from the bacterial chromosome. In the case of integration, host integration factor (IHF) and phage-specific integrases mediate insertion of the phage into the bacterial chromosome at a specific attachment site *(attB)* and each lambdoid phage has a unique integration site [155]. In the case of the lambda phage, for the cross-over to happen, a segment of 21 bp of *attB* and a core sequence from the 240 bp segment of phage *attP* sequence is required [156]. Recombination of the core sequences (*attB* and *attP*) produces two new half-sites *attR* and *attL* on the right and left sides of the integrated sequence, respectively. If the lytic cycle recommences, the reverse reaction starts and reconstructs the original sites [157].

In all cases, genes of O-antigen modification are found next to the phage *int attP* region, and upon integration of phage into the bacterial chromosome, the int and O-integration genes are moved to the opposite ends of the phage DNA [23]. Almost all serotype converting phages integrate into the *thrW tRNA* gene of the host except Sf6 and Sf101. Sf6 integrates into *argW tRNA* gene, whereas Sf101 integrates at a newer location i.e. within *sbcB* gene [105]. Jakhetia *et al.,* 2014*,* reported that the integrase gene of the Sf101 phage is 99% identical to integrases

of HK544, mEp234, and phage 2851. Further, *attP* of both Sf101 and Phage 2851 are highly similar and both phages integrate into the *sbcB* locus [158]. Based on the integration site of the Sf101 phage, the location of the O-antigen converting gene *oacB* in the Sf101 lysogen (serotype 1c SFL1683 and SFL1684) is in the *sbcB* locus. By contrast, the location of *oacB* gene in other serotypes (such as,1a,1b, 2a, 5a, Y and other 1c strains of *S. flexneri),* at the *adrA* gene [106], consistent with other serotype converting genes (*gtrs/oac*) [93, 104, 121, 142, 159].

1.9 Structural and functional studies of O-antigen modifying genes

Currently available vaccines are serotype-specific, which limits their application on evolving serotypes. The development of a multivalent vaccine is desperately needed to combat shigellosis. There is a need to understand the mechanism by which the serotype converting enzymes (Gtrs/Oacs) of *S. flexneri* modify O-antigen to design an effective vaccine. Identification of structural elements (domains/loops) of proteins would allow a better understanding of the interaction of enzymes with each other. Previously, much work has been done to understand the topology of the glycosyltransferases (Gtrs) of *S. flexneri* [99, 160-162]. Sequence homology exists among Gtrs and it is shown that structurally related enzymes often catalyse similar reactions [163]. Studies have also shown that all Gtrs recognize identical acceptor or donor substrates. There are six $Gtr_{[type]}$ proteins; GtrI, GtrIc, GtrII, GtrIV, GtrV and, GtrX. Currently, the structures of five Gtrs have been elucidated using the *phoA-lacZ* dual reporter system. A Dual reported system was developed by Alexeyev and Winkler, 1999, and consists of a fusion of *E. coli* enzymes alkaline phosphatase-AP (*phoA*) and an alpha fragment of β-galactosidase-BG (*lacZα*) to study the topology of membrane proteins. AP is always active during localisation in the periplasm and BG in the cytoplasm [164]. Lehane *et al*. (2005) showed that GtrII has 486 amino acids, nine TM, a re-entrant loop, and three large periplasmic regions. There are four critical residues regions (Glu^{40}, Phe^{414}, Cys^{435}, and Lys^{478}) in the periplasmic region, three of which are conserved among other Gtrs [161]. Furthermore, GtrII

has structural homology with GtrI, which is a 506 amino acid protein with nine TM helices and cytoplasmic N-terminus and periplasmic C-terminus [161].

The topology of another serotype converting protein, GtrIV was solved experimentally by creating several fusions and using a dual-reporter system [160]. The study concluded that GtrIV consists of eight TM helices, two long periplasmic loops, two small cytoplasmic N- and C- termini, and a re-entrant between TM III and IV. It was also found that GtrIV has structural homology to GtrIc [160]. GtrIc is the largest of all Gtrs so far identified, having 526 amino acids arranged in 10 TM helices, cytoplasmic N- and C- termini, two large periplasmic loops, and a double TM dipping loop [99]. Korres and Verma (2006) using the dual-reporter technique found that GtrV has 417 amino acids consisting of cytoplasmic N-terminus, nine TM helices, a re-entrant loop, and C-terminus in the periplasm. There is also marked sequence similarity between GtrV and GtrX [165]. Moreover, it was reported that there are three motifs in GtrV; two in loop no. 2 and one in loop No.1, bear acidic residues critical to GtrV function [165]. However, the topology of GtrV was later verified using the substituted cysteine accessibility method (SCAM) and the results agreed with the previous topology proposed by Korres and Verma, 2006 [162]. SCAM allows the structural study of any protein in its native state whereas the gene fusion approach with dual reporter genes *phoA-laczα* has the disadvantage of disturbing protein native structure [166].

GtrX has 416 amino acids with nine TM. Chimera between GtrV and GtrX were also constructed to investigate the conserved function between two proteins. In this experiment, loop no. 2 of GtrX was changed into GtV and it was observed that this chimera retained its function and the N-terminal periplasmic region of the Gtrs holds a conserved role [165].

Studies conducted on Oac using dual reporter genes *phoA-laczα* showed that it is an integral membrane protein consisting of ten TM helices, a cytoplasmic N-terminus, and a periplasmic C-terminus [167]. By using site - directed mutagenesis three arginine residues, (R73, R75 and R76) in the cytoplasmic loop 3 were found to be critical to Oac function [111]. In this study,

two loops (3 and 11) of Oac were deleted to investigate the role of these loops in Oac function. Deletion of these loops interfered with Oac assembly in the membrane suggesting that these regions are structurally important to Oac. Further deletion experiments revealed that the loop is catalytically important for Oac [168]. Several conserved motifs (serine-glycine and phenylalanine -proline amino acids) in the TM segments were also found critical to Oac function [168].

Gathering all information generated by these studies on serotype converting proteins can help identify regions of interest in these proteins. This could help design a chimeric protein with the capability to convert two or more serotypes at a time and in turn, could be a good vaccine candidate.

1.9.1 Structural investigation

Membrane proteins are important and diverse constituents of the proteome and play roles in regulating the permeability of membranes. Integral membrane proteins (IMPs) are membrane proteins that account for 25-30% coding capacity of the genes in typical organisms [169]. IMPs are always attached to biological membrane and need a detergent or organic solvent to be solubilized. The most common type of IMPs is transmembrane proteins (TM). The hydrophobic nature of these proteins makes them difficult to study structurally, and only 1% of high-resolution structures are known in all proteins [169]. The lack of a 3D structure, which provides a snapshot of molecular interactions limits the understanding of functional analysis of any membrane protein. The structure facilitates the assignment of function, in the case of membrane proteins, it's difficult to solve 3D structure experimentally. Therefore, other alternative approaches are necessary, like determining two-dimensional structure bioinformatically and then testing the model using experimental approaches, for example employing reporter genes (*phoA -lacZ*, GFP, etc) and site-specific label binding (SCAM and oxidative labelling) [170]. Among all methods, SCAM is most feasible and involves working with functional variants of a protein.

1.9.1.1 Substituted cysteine accessibility method (SCAM)

SCAM is considered as one of the ideal methods to study the internal structures of integral membrane protein like protein channels, transporters, and binding site crevices. In comparison to other methods for a topological investigation like a dual reporter, SCAM is considered a preferred technique as it minimally alters the structure of the target protein. SCAM uses the ability of cysteine (Cys) to make disulphide bonds with the sulfhydryl reagents (SH) being used in SCAM [166]. The presence of only one cysteine in the target protein is essential for SCAM, hence a functional mutant of the target protein lacking cysteine residues needs to be generated by mutating all native cysteines to either alanine or serine [170]. Functional variants can then be produced using Cys-less mutant as a template by introducing a single cysteine at the position of interest. Following confirmation of cys-substituted mutants, *in vivo* labelling with thiol-reactive reagent (Oregon-green 488 maleimide carboxylic acid-OGM) is performed and the introduced cysteine can be located. When disulphide bonds are formed between cysteine and OGM, fluorescence can be detected under UV light. OGM is cytoplasmic membrane impermeable and enables detection of cysteine in the periplasm. Methanethiosulfonate ethyl trimethylammonium (MTSET) a membrane impermeable thiol-specific blocking reagent is used to detect cytoplasmic cysteines. MTSET binds with any cysteine present in the periplasmic region, the membrane is then permeabilized to add OGM so that it can reach the cytoplasm and bind with any cysteine residue available and fluoresce. Along with this, a control test is also run in parallel in which OGM is added after permeabilising the bacterial cells (without the use of MTSET). This labels cysteine both in cytoplasm and periplasm and assures successful cysteine detection and if the introduced cysteine is present within the membrane there will be no fluorescence detected in any stage of the SCAM [166, 171].

1.10 Aims of this study

Despite years research, *Shigella flexneri* remains a pathogen of concern, for which no licensed vaccine is currently available to control shigellosis. The main obstacle is the changing serotype behaviour of *S. flexneri***,** mainly serotype-conversion is attributed to bacteriophage encoded genes in *S. flexneri* genome, included glucosyltransferases and acetyltransferases. Not much is known about serotype conversion mediated by *O*-acetyltransferases belonging to *S. flexneri***.** Recently, *oacB* gene was isolated from Sf101 bacteriophage in serotype 1c strain of S. *flexneri*, which adds an acetyl residue to the O-antigen at either position 3 or 4 of Rhamnose III. The main aims of this study were to understand the mechanism by which *S. flexneri* OacB facilitates O-antigen acetylation, and which residues of OacB were critical to the function. Moreover, this study also aimed to understand the integration of Sf101 phage in 1c strains of *S. flexneri,* to understand the distribution of *oacB* gene in serotype 1c of *S. flexneri*. Finally, in this study, the role of Sf101 encoded novel genes (*oacB/16*, *orf17*, *orf41* and, *orf56*) in the virulence of *S. flexneri* was also determined.

AIM 1: Identification of critical residues of O-antigen-modifying *O*-acetyltransferase B (OacB) of *Shigella flexneri*

Aim 1 comprises of two result chapters. Result chapters 3 is based on the optimization of overexpression of OacB in *E. coli* (TOP10 cells) in two expression vectors and development of two-dimensional OacB model. Chapter 4 is based on the investigation of critical residues in OacB using SDM. The residues were selected based on their conserved location within the conserved domain of OacB and the contribution of the individual amino acids in OacB function was investigated. (Chapter 4 results/discussion are presented as a manuscript).

AIM 2: Acquisition and distribution of *O*-acetyltransferase B (*oacB*) gene in *S. flexneri* genome

Chapter 5 presents the bioinformatics analysis of MiSeq (Illumina) sequences of 1c strains and other serotypes collected from different geographical locations, to investigate the acquisition and distribution of the *oacB* gene in all the strains. The *oacB* gene is found at two (*adrA/sbcB* loci) different locations in 1c serotype strains, therefore in this chapter, the genomic organizations of *sbcB* and *adrA* loci in serotype 1c, 1a, 1b and Y strains were investigated to understand the integration of Sf101 phage in 1c strains of *S. flexneri*.

AIM 3: Study on the role of Sf101 phage-encoded novel genes in *S. flexneri* virulence

The role of many bacteriophages encoded virulence factors have been elucidated in previous studies. In a previous study, four Sf101 encoded genes (*orf16/oacB*, *orf17*, *orf41* and *orf56*) were identified, with no known phage-related function. In Chapter 6 of this thesis the role of unique Sf101 genes in *S. flexneri* virulence was determined, using *in vivo C. elegans*, (bacterial accumulation/ liquid killing), and *in vitro* Hela cells (invasion) assays.

Chapter 2: Material and Methods

2.1. Bacterial culture

Bacterial cultures were routinely grown in Luria Bertani (LB) broth in a shaking incubator at 37^{o}C at either 180 or 200 rpm (Appendix A). LB agar plates with antibiotics were used for the growth of bacterila cultures at 37 ^{o}C or 30 ^{o}C. Cultures for *S. flexneri* were grown at a temperature of 30 ^{o}C to maintain the virulence plasmid in LB agar plates or LB broth (in a shaker incubator), containing appropriate antibiotics (Appendix A).

For long-term storage of the bacterial strains, a single colony was used to make glycerol stocks in 50% LB and 50% glycerol and stored at - 80 ^{o}C temperature. For the propagation of the bacteriophages, bacterial cultures were grown in NZCYM broth (Appendix A).

Supplementary antibiotics used were 100 µg/ml ampicillin, 250 µg/ml erythromycin, 50 µg/ml kanamycin, 25 µg/ml and 200 µg/ml gentamicin.

2.2 Bacterial strains and Plasmid vectors

Bacterial strains such as the *E. coli* strains, wild-type *S. flexneri* strains, and recombinant *S. flexneri* strains used in the present study are shown in Tables 2.1, 2.2, 2.3, respectively. While Plasmid vectors used include the pBC SK+, pKD46, pBAD/*Myc*-HisA, pFLAG-CTC. For the protein expression, pFLAG-CTC and pBAD-*Myc*-HisA were used (Figure 2.1).

Table 2.1 Wild type *S. flexneri* strains used in this study

Strain	Serotype	Country of Isolation
SFL1287	1a	Japan
SFL1288	1a	Japan
SFL1299	Y	Japan
SFL1300	1b	Japan
SFL1309	1b	Japan
SFL1315	1b	Japan
SFL1353/SFL124	Y	Srilanka
SFL1416	1a	UK
SFL1417	1b	Japan
SFL1492	1a	Bangladesh
SFL1493	1a	Bangladesh
SFL1494	1a	Bangladesh
SFL1496	1b	Bangladesh
SFL1497	1b	Bangladesh
SFL1500	1c	Bangladesh
SFL1501	1c	Bangladesh
SFL1502	1c	Bangladesh
SFL1503	1c	Bangladesh
SFL1504	1c	Bangladesh
SFL1538	Y	Bangladesh
SFL1541	Yv	Bangladesh
SFL1561	1c	Vietnam
SFL1562	1c	Vietnam
SFL1564	1c	Vietnam
SFL1565	1c	Vietnam
SFL1566	1c	Vietnam
SFL1567	1c	Vietnam
SFL1572	1c	Vietnam
SFL1573	1c	Vietnam
SFL1578	1c	Vietnam
SFL1579	1c	Vietnam
SFL1580	1c	Vietnam
SFL1581	1c	Vietnam
SFL1582	1c	Vietnam
SFL1584	1c	Vietnam
SFL1585	1c	Vietnam
SFL1585	1c	Vietnam
SFL1613 (Y394)	1c	Bangladesh
SFL1683	1c	Egypt
SFL1684	1c	Egypt

Strain	Serotype	Country of Isolation
SFL1685	1c	Egypt
SFL1685	1c	Egypt
SFL1687	1c	Egypt
SFL1688	1c	Egypt
SFL1689	1c	Egypt
SFL1690	1c	Egypt
SFL1691	1c	Egypt
SFL1692	1c	Egypt
SFL2182	1c	Vietnam
SFL2191	1c	Vietnam
SFL2192	1c	Vietnam
SFL2242	1c	Vietnam
SFL2262	1c	Vietnam
SFL2263	1c	Vietnam
SFL2445	1c	United Kingdom
SFL2446	1c	United Kingdom
SFL2447	1c	United Kingdom
SFL2448	1c	United Kingdom
SFL2449	1c	United Kingdom
SFL2450	1c	United Kingdom
SFL2451	1c	United Kingdom
SFL2452	1c	United Kingdom
SFL2453	1c	United Kingdom
SFL2454	1c	United Kingdom
SFL2455	1c	United Kingdom
SFL2456	1c	United Kingdom
SFL2457	1c	United Kingdom
SFL2458	1c	United Kingdom
SFL2459	1c	United Kingdom
SFL2460	1c	United Kingdom
SFL2461	1c	United Kingdom
SFL2462	1c	United Kingdom
SFL2463	1c	United Kingdom
SFL2464	1c	United Kingdom
SFL2465	1c	United Kingdom
SFL2466	1c	United Kingdom
SFL2467	1c	United Kingdom
SFL2468	1c	United Kingdom
SFL2469	1c	United Kingdom
SFL2470	1c	United Kingdom
SFL2471	1c	United Kingdom
SFL2472	1c	United Kingdom

Strain	Serotype	Country of Isolation
SFL2473	1c	United Kingdom
SFL2474	1c	United Kingdom
SFL2475	1c	United Kingdom
SFL2477	1a	United Kingdom
SFL2485	1b	United Kingdom
SFL2494	1a	United Kingdom

Table 2.2 *E. coli* strains created and used in this study

Strain	Description	Source
JM109	*rec*A1 supE44 endA1 hsdR17 gyrA96 relA1 thi (lac-proAB) [F' traD36 proAB lacI^{q}lacZ M15	[172]
B2298	OP50 *E.coli* (CGC)	[173]
B2370 (TOP10)	Invitrogen cloning and expression host. F- mcrA	Invitrogen
B2372	pBAD/*Myc*-HisA	
B2375	pBAD/*Myc*-His/lacZ transformed into TOP10	This study
B1188	XLI-blue	This study
B2574	JM109 carrying pNV2110 (*oacB* gene in pFLAG-CTC vector)	This study
B2575	JM109 carrying pNV2111 (*oacB* gene in pBAD/*Myc*-HisA vector)	This study
B2596	JM109 carrying pNV2132	This study
B2598	XLI carrying pNV2133	This study
B2600	XLI carrying pNV2135	This study
B2602	XLI carrying pNV2137	This study
B2603	XLI carrying pNV2138	This study
B2604	XLI carrying pNV2139	This study
B2605	JM109 carrying pNV2140	This study
B2606	JM109 carrying pNV2141	This study
B2607	TOP10 carrying pNV2132	This study
B2608	TOP10 carrying pNV2137	This study
B2609	TOP10 carrying pNV2137	This study
B2610	TOP10 carrying pNV2137	This study
B2611	BL21 carrying pNV2140	This study
B2614	JM109 carrying pNV2142	This study
B2615	TOP10 carrying pNV2142	This study
B2619	XLI carrying pNV2146	This study
B2620	XLI carrying pNV2147	This study
B2621	XLI carrying pNV2148	This study
B2622	XLI carrying pNV2149	This study

Strain	Description	Source
B2623	XLI carrying pNV2150	This study
B2624	XLI carrying pNV2151	This study
B2625	XLI carrying pNV2152	This study
B2626	XLI carrying pNV2153	This study
B2627	XLI carrying pNV2154	This study
B2628	XLI carrying pNV2155	This study
B2629	XLI carrying pNV2156	This study
B2630	XLI carrying pNV2157	This study
B2631	XLI carrying pNV2180	This study
B2632	XLI carrying pNV2178	This study
B2633	XLI carrying pNV2179	This study
B2651	JM109 carrying pNV2167	This study
B2652	JM109 carrying pNV2168	This study
B2653	JM109 carrying pNV2169	This study
B2654	TOP10 carrying pNV2167	This study
B2655	TOP10 carrying pNV2168	This study
B2656	TOP10 carrying pNV2169	This study
B2658	XLI-blue carrying pNV2171	This study
B2659	XLI-blue carrying pNV2172	This study
B2660	XLI-blue carrying pNV2173	This study
B2661	XLI-blue carrying pNV2174	This study
B2662	DH5α carrying pNV2175	This study
B2663	DH5α carrying pNV2176	This study
B2664	DH5α carrying pNV2177	This study
B2666	DH5α carrying pNV2185	This study
B2674	TOP10 carrying pNV2193	This study
B2675	TOP10 carrying pNV2194	This study
B2676	TOP10 carrying pNV2195	This study
B2677	TOP10 carrying pNV2171	This study
B2678	TOP10 carrying pNV2196	This study
B2679	TOP10 carrying pNV2173	This study
B2680	TOP10 carrying pNV2197	This study
B2681	TOP10 carrying pNV2198	This study
B2682	TOP10 carrying pNV2199	This study
B2683	TOP10 carrying pNV2147	This study
B2684	TOP10 carrying pNV2178	This study
B2685	TOP10 carrying pNV2146	This study
B2686	TOP10 carrying pNV2152	This study
B2687	TOP10 carrying pNV2148	This study
B2688	TOP10 carrying pNV2149	This study
B2689	TOP10 carrying pNV2171	This study

Table 2.3 Recombinant *S. flexneri* strains used in this study

Strain	Description	Source
SFL2572	SFL1691 carrying pNV2132	This study
SFL2573	SFL1691 carrying pNV2133	This study
SFL2575	SFL1691 carrying pNV2135	This study
SFL2576	SFL1691 carrying pNV2137	This study
SFL2577	SFL1691 carrying pNV2138	This study
SFL2578	SFL1691 carrying pNV2139	This study
SFL2579	SFL1691 carrying pNV2140	This study
SFL2580	SFL1691 carrying pNV2141	This study
SFL2581	SFL1691 carrying pNV2142	This study
SFL2582	SFL1613 having VP tagged with kanamycin resistance gene	This study
SFL2583	SFL1691 carrying pNV2146	This study
SFL2584	SFL1691 carrying pNV2148	This study
SFL2585	SFL1691 carrying pNV2147	This study
SFL2586	SFL1691 carrying pNV2149	This study
SFL2587	SFL1691 carrying pNV2150	This study
SFL2588	SFL1691 carrying pNV2151	This study
SFL2589	SFL1691 carrying pNV2152	This study
SFL2590	SFL1691 carrying pNV2154	This study
SFL2591	SFL1691 carrying pNV2153	This study
SFL2592	SFL1691 carrying pNV2155	This study
SFL2593	SFL1691 carrying pNV2156	This study
SFL2594	SFL1691 carrying pNV2157	This study
SFL2595	SFL1691 carrying pNV2178	This study
SFL2596	SFL1691 carrying pNV2179	This study
SFL2597	SFL1691 carrying pNV2180	This study
SFL2604	SFL2456 carrying pNV2132	This study
SFL2613	SFL2456 carrying pNV2167	This study
SFL2614	SFL2456 carrying pNV2168	This study
SFL2615	SFL2456 carrying pNV2169	This study
SFL2616	SFL1691 carrying pNV2172	This study
SFL2617	SFL1691 carrying pNV2171	This study
SFL2618	SFL1691 carrying pNV2173	This study
SFL2619	SFL1691 carrying pNV2174	This study
SFL2620	SFL1691 carrying pNV2175	This study
SFL2621	SFL1691 carrying pNV2176	This study

Strain	Description	Source
SFL2622	SFL1691 carrying pNV2177	This study
SFL2626	SFL2456 carrying pNV2185	This study
SFL2627	SFL1691 carrying pNV2194	This study
SFL2628	SFL1691 carrying pNV2195	This study
SFL2629	SFL1691 carrying pNV2196	This study
SFL2630	SFL1691 carrying pNV2197	This study
SFL2631	SFL1691 carrying pNV2193	This study
SFL2632	SFL1691 carrying pNV2198	This study
SFL2633	SFL1691 carrying pNV2199	This study

Table 2.4 Plasmids used in this study

Plasmid	Characteristics	Source
pBAD/*Myc*-HisA	pBAD/*Myc*- His plasmids are derived from PBR322 expression vectors and designed for regulated, dose-dependent recombinant protein expression and purification in *E. coli.* PBAD promoter – araBAD helps to achieve optimum levels of soluble, recombinant from *E. coli.* The regulatory protein, *Ara*C, is provided on the pBAD/His and pBAD*/Myc*-His vectors allowing regulation of PBAD. Amp^R	Invitrogen
pBAD/*Myc*-His/lacZ	pBAD/*Myc*-His*/lacZ* is a 7242 bp control vector containing the gene for B-galactosidase fused to the C-terminal peptide. It was constructed by digesting the vector pTrcHis2*/lacZ* with *Nco* I and *Nsi* I to remove the *lac*I gene and the *trc* promoter and replacing with a *Nco* I-*Nsi* I fragment containing the *ara*C gene and the *ara* BAD promoter. The B-galactosidase portion of the fusion may be released by digestion with *Sfu* I (*Bst*B I). Amp^R	Invitrogen
pFLAG-CTC	5348bp expression vector for cytoplasmic expression of properly inserted open reading frame as a C-terminal FLAG fusion peptide. FLAG epitope is a hydrophilic 8 amino acid tag (DYKDDDDK). *Tac* promoter *lacO* and *lacI* repressors. Amp	Sigma-Aldrich
pNV2111	Wild type *oacB cloned into* pBAD/*Myc*-HisA vector using *Xho*I and *EcoR*I sites	This study
pNV2110	WT *oacB* cloned in pFLAG-CTC vector using *Xho*I and *Bgl*II.	This study
pNV2132	Erythromycin resistance gene introduced at *Sph*I site in pNV2111	This study
pNV2133	pNV containing OacB with cysteine129 mutated to Ala	This study
pNV2135	pNV containing OacB with cysteine284 mutated to Ala	This study
pNV2137	pNV containing OacB with cysteine326 mutated to Ala	This study

Plasmid	Characteristics	Source
pNV2138	pNV containing OacB with cysteine357 mutated to Ala	This study
pNV2139	pNV containing OacB with cysteine295 mutated to Ala	This study
pNV2140	pNV2111cloned with *Em* gene at *Nco*I site	This study
pNV2146	pNV2132 with OacB Arg47 mutated to Ala	This study
pNV2147	pNV2132 with OacB Arg116 mutated to Ala	This study
pNV2148	pNV2132 with OacB W71 mutated to Ala	This study
pNV2149	pNV2132 with OacB Arg119 mutated to Ala	This study
pNV2150	pNV2132 with OacB Lysine156 mutated to Ala	This study
pNV2151	pNV2132 with OacB tyrosine 96 mutated to Ala	This study
pNV2152	pNV2132 with OacB phenylalanine 98 mutated to Ala	This study
pNV2153	pNV2132 with OacB Aspartic acid 44 mutated to Ala	This study
pNV2154	pNV2132 with OacB His320 mutated to Ala	This study
pNV2155	pNV2132 with OacB Glutamic acid188 mutated to Ala	This study
pNV2156	pNV2132 with OacB Proline122 mutated to Ala	This study
pNV2157	pNV2132 with OacB FY191-192 mutated to Ala	This study
pNV2178	pNV2132 with OacB H58 mutated to Ala	This study
pNV2179	pNV2132 with OacB Valine 87 mutated to Ala	This study
pNV2180	pNV2132 with OacB WT183-184 mutated to Ala	This study
pNV2167	Orf17 cloned into pBAD/*Myc*-HisA at *Nco*I and *EcoR*I	This study
pNV2168	Orf41 cloned into pBAD/*Myc*-HisA at NCOI and EcoRI	This study
pNV2169	Orf56 cloned into pBAD/*Myc*-HisA at *Nco*I and *EcoR*I	This study
pNV2171	pNV2132 with OacB Ser146 mutated to Ala	This study
pNV2172	pNV2132 with OacB D173 mutated to Ala	This study
pNV2173	pNV2132 with OacB Gly164 mutated to Ala	This study
pNV2174	pNV2132 with OacB Gly140 mutated to Ala	This study
pNV2175	pNV2132 with OacB Ser139 mutated to Ala	This study
pNV2176	pNV2132 with OacB Ser153 mutated to Ala	This study
pNV2177	pNV2132 with OacB Ser174 mutated to Ala	This study
SFL1613 having pkD46 VP tagged with kanamycin resistance gene	pNV-Y394	This study
pNV2185	pBAD/*Myc*-HisA vector cloned with *Em* gene at *Sph*I site.	This study
pNV2193	Leucine34 mutated to Cysteine in pNV2139	This study
pNV2194	Alanine 73 mutated to Cysteine in pNV2139	This study
pNV2195	Valine114 mutated to Cysteine in pNV2139	This study
pNV2196	Ala162 mutated to Cysteine in pNV2139	This study
pNV2197	Isoleucine202 mutated to Cysteine in pNV2139	This study
pNV2198	Isoleucine251 mutated to Cysteine in pNV2139	This study
pNV2199	Asparginine335 mutated to Cysteine in pNV2139	This study

Table 2.5 Primers used in this study

Primer name	Primer sequence	Primer comments
oacB pBADF	CCGCTCGAGGATGCA TAT GAT	*Xho*I at 5' and *EcoR*I at 3'
oacB pBAD R	CCTGAATTCCGTTGATTGTTGTT	
oacB pFLAG F	CCGCTCGAGATGCATATGATTGAA	*Xho*I at 5' and *Bgl*II at 3'
oacB pFLAG R	CCGAGATCTTTGATTGTTGTTT	
oacBnewF	CATCCGTGATATTGATGT	RT-PCR
oacBnewR	ATTTGATGAATGGCGTCT	RT-PCR
Orf17newF	TGGTACACAATTTGGTAC	RT-PCR
Orf17newR	TACTGCTTTGTAGAATGG	RT-PCR
Orf41newf	TTTGGTTACAGTACGGAA	RT-PCR
Orf41newR	TTAGTTGCAATAGTACCC	RT-PCR
Orf56newF	CGATATGCACGGGCAAAA	RT-PCR
Orf56newR	TAATCACGACCTTTCTGA	RT-PCR
Cloning *orf17* in pBAD/*Myc*-HisA Fwd	CGCG CCATGG CGATTTTAATTTGGTACACAATTTG G	Forward cloning primer *orf17*
orf17 in pBAD/ *Myc*-HisA Rev	CGCG GAATTC CGCTGCTTTGTAGAATGGACAGA	Reverse cloning primer *orf17*
Cloning *orf41* in pBAD/*Myc*-HisA Fwd	CGCGCCATGG CGATCCGGTTTGGTTACAGTACG	Forward cloning primer *orf41*
Cloning *orf41* in pBAD/*Myc*-HisA Rev	CGCG GAATTC CG GTTGCAATAGTACCCATAGATAA	Reverse cloning primer *orf41*
Cloning *orf56* in pBAD/*Myc*-HisA Fwd	CGCGCCATGGCGACCGATATGCAC GGGCAAAA	Forward cloning primer *orf56*
Cloning *orf56* in pBAD/*Myc*-HisA Rev	CGCG GAATTC CGATCACGACCTTTCTGAAAGCA	Reverse cloning primer *orf56*
Apy1	CATAATCAAGAGACAAAACGATA	Forward *apy* -VP
Apy2	CCAGCCTTTCCAGTAATCCC	Reverse *apy*-VP
VirG1	CGGGTACTCAAGAACTTCAAT	Forward *virG* -VP
VirG2	TTCCGCCAAAATGAGAGTTCC	Reverse *virG* -VP

Primer name	Primer sequence	Primer comments
Em^R-*Sph*I-Fwd	AATGCATGCTAAGACGGTTCGTGTTCGT	Cloning of Em^R gene in pNV2111
Em^R-*Sph*I-Rev	AATGCATGCCATAGAATTATTTCCTCCCG	Cloning of Em^R gene in pNV2111
Final_VP_forward Kan region	ATTTGATGCACGTAGTAAATACCAGTGAACAAATAGCTTCGGATACTCCTCTTGAGCGATTGTGTAGGCT	Amplification of kanamycin gene for VP tagging
Final_VP_Rev Kan region	GATGCGACGGGCAGACTTTGAAAATGTTCGACCAGCAACGTGATGGCAATTGAATATCCTCCTTAGTTCC	Amplification of kanamycin gene for VP tagging
Site-directed mutagenesis primers		
Creation Cys-less OacB		
C129A-Fwd	GTTC ATAGTTAGCCTAGCTCTCATTTTTATC	To mutate residue C129 to alanine
C129A-Rev	GATAAAAATGAGAGCTAGGCTAACTATGAAC	To mutate residue C129 to alanine
C284A-Fwd	CACTGATTCTTGCTGGAATAACATTTATC	To mutate residue C284 to alanine
C284A-Rev	GATAAATGTTATTCCAGCAAGAATCAGTG	To mutate residue C284 to alanine
C295A-Fwd	CATCAGGTGCTATTTGTATGGAATATTAAG	To mutate residue C295 to alanine
C295A-Rev	CTTAATATTCCATACAAATAGCACCTGATG	To mutate residue C295 to alanine
C326A-Fwd	GATATTCCTTTATGCCCTAATGACGTG	To mutate residue C326 to alanine
C326A-Rev	CACGTCATTAGGGCATAAAGGAATATC	To mutate residue C326 to alanine
C357A-Fwd	CGTTCACATCAGCCTAACTTTTAAATTAATAG	To mutate residue C357 to alanine
C357A-Rev	CTATTAATTTAAAAGTTAGGCTGATGTGAACG	To mutate residue C357 to alanine
Mutation of conserved /selected residues of OacB		
D44F	GAATAATCAAATAG**C**TGGAATGCGGGGGTTCTTAG	To mutate residue D44 to alanine
D44R	CTAAGAACCCCCGCATTCCA**G**CTATTTGATTATTC	To mutate residue D44 to alanine
R47F	CAAATAGATGGAATGG**CG**GGGTTCTTAGCAATTTTC	To mutate residue R47 to alanine

Primer name	Primer sequence	Primer comments
R47R	GAAAATTGCTAAGAACCC**CG**CCATT CCATCTATTTG	To mutate residue R47 to alanine
H58F	CTTATTCAT**GC**CGCAGCAATTTGG	To mutate residue H58 to alanine
H58R	CCAAATTGCTGCG**GC**ATGAATAAG	To mutate residue H58 to alanine
W71F	CTTGTCATCTGGAGTAGCGGAAGC ACCTTCATCAAATC	To mutate residue W71 to alanine
W71R	GATTTGATGAAGGTGCTTCCGCTAC TCCAGATGACAAG	To mutate residue W71 to alanine
V87F	GCCAAGTTGGTG**CT**TCATTCTTTTT TATG	To mutate residue V87 to alanine
V87R	CATAAAAAAGAATGA**AG**CACCAACT TGGC	To mutate residue V87 to alanine
Y96F	CTTTTTTATGATTACTGGT**GC**TCTGT TCTTTTCAAAG	To mutate residue Y96 to alanine
Y96R	CTTTGAAAAGAACAGA**GC**ACCAGTA ATCATAAAAAAG	To mutate residue Y96 to alanine
F98F	GATTACTGGTTATCTG**GC**CTTTTCA AAGATTATCTC	To mutate residue F98 to alanine
F98R	GAGATAATCTTTGAAAAG**GC**CAGAT AACCAGTAATC	To mutate residue F98 to alanine
R116F	GACAAGGCTTTATGTATCA**GC**ATTA CTACGATTAACCC	To mutate residue R116 to alanine
R116R	GGGTTAATCGTAGTAAT**GC**TGATAC ATAAAGCCTTGTC	To mutate residue R116 to alanine
R119F	GTATCAAGATTACTA**GC**ATTAACCC CAATGTTC	To mutate residue R119 to alanine
R119R	GAACATTGGGGTTAAT**GC**TAGTAAT CTTGATAC	To mutate residue R119 to alanine
P122F	GATTACTACGATTAACC**GC**AATGTT CATAGTTAGCCT	To mutate residue P112 to alanine
P122R	AGGCTAACTATGAACATT**GC**GGTTA ATCGTAGTAATC	To mutate residue P122 to alanine
S139F	CATTGTAGGTTTTAAGGCT**GG**ATGG AGAATGCAGG	To mutate residue S139 to alanine
S139R	CCTGCATTCTCCATCCAG**CC**TTAAA ACCTACAATG	To mutate residue S139 to alanine
G140F	CATTGTAGGTTTTAA**G**TCTGCATGG AGAATGC	To mutate residue G140 to alanine
G140R	GCATTCTCCAT**G**CAGACTTAAAACC TACAATG	To mutate residue G140 to alanine
S146F	GAATGCAGGTA**G**CCACAGAAGAGC TTTTTGTG	To mutate residue S146 to alanine
S146R	CACAAAAAGCTCTTCTGTGG**C**TACC TGCATTC	To mutate residue S146 to alanine

Primer name	Primer sequence	Primer comments
S153F	CAGAAGAGCTTTTTGTG**G**CAATAAT GAAGTGG	To mutate residue S153 to alanine
S153R	CCACTTCATTATTG**C**CACAAAAAGC TCTTCTG	To mutate residue S153 to alanine
G164F	GCCATTCACTGCACTAG**C**TATGCC GAACATTAATGACG	To mutate residue G164 to alanine
G164R	CGTCATTAATGTTCGGCATAGCTA**G** TGCAGTGAATGGC	To mutate residue G164 to alanine
D173F	CGAACATTAATGACGTAAAAG**C**TTC ATTTACTATCAATGC	To mutate residue D173 to alanine
D173R	GCATTGATAGTAAATGAA**G**CTTTTA CGTCATTAATGTTCG	To mutate residue D173 to alanine
S174F	CATTAATGACGTAAAAGAT**G**CATTT ACTATCAATGCCGCTGTAAC	To mutate residue S174 to alanine
S174R	GTTACAGCGGCATTGATAGTAAATG **C**ATCTTTTACGTCATTAATG	To mutate residue S174 to alanine
WT183-184 F	CTGTAACA**GC**G**G**CACTTGTATATGA ATG	To mutate residues Wt183-184 to alanine/ alanine
WT183-184 R	CATTCATATACAAGT**GC**C**G**CTGT TACAG	To mutate residues Wt183-184 to alanine/ alanine
E188F	GTAACATGGACACTTGTATATG**C**AT GGTTCTTTTATTTTTC	To mutate residue E188 to alanine
E188R	GAAAAATAAAAGAACCATG**C**ATATA CAAGTGTCCATGTTAC	To mutate residue E188 to alanine
FY191-192F	GTATATGAATGGTTCGCTGCTTTTT CTCTTCCGGTAATTTC	To mutate residues FY191-192 to alanine/ alanine
FY191-192R	GAAATTACCGGAAGAGAAAAAGCA GCGAACCATTCATATAC	To mutate residues FY191-192 to alanine/ alanine
H320F	GCGTTTACCTTCTGGCTGGGATATT CCTTTATTGC	To mutate residue H320 to alanine
H320R	GCAATAAAGGAATATCCCAGCCAG AAGGTAAACGC	To mutate residue H320 to alanine
K372F	CTAATAATTACACTGAGCATACATTT ATCATACTGG	To mutate residue K372 to alanine
K372R	CCAGTATGATAAATGTATGCTCAGT GTAATTATTAG	To mutate residue K372 to alanine
Creation of single cysteine variants		
L34C Fwd	CCCCAATAAAT**TG**TGTTGAGCACGG CAGGAATAATC	To mutate residue L34 to cysteine

Primer name	Primer sequence	Primer comments
L34C Rev	GATTATTCCTGCCGTGCTCAAC**AC**A ATTTATTGGGG	To mutate residue L34 to cysteine
A73C Fwd	CATCTGGAGTATGGGAA**TGC**CCTT CATCAAATCTGTTAGC	To mutate residue A73 to cysteine
A73C Rev	GCTAACAGATTTGATGAAGG**GCA**TT CCCATACTCCAGATG	To mutate residue A73 to cysteine
V114CFwd	GGACAAGGCTTTATTGTTCAAGATT ACTACGATTAACCCC	To mutate residue V114 to cysteine
V114CFwd	GGGGTTAATCGTAGTAATCTTGAAC AATAAAGCCTTGTCC	To mutate residue V114 to cysteine
A162C Fwd	GAAGTGGTTGCCATTCACTTGTCTA GGTATGCCGAAC	To mutate residue A162 to cysteine
A162C Rev	GTTCGGCATACCTAGACAAGTGAAT GGCAACCACTTC	To mutate residue A162 to cysteine
I202C Fwd	CTCTTCCGGTAATTTCCGCGCTCTG CAAAAGAAAAGTCAG	To mutate residue I202 to cysteine
I202C Rev	CTGACTTTTCTTTTGCAGAGCGCGG AAATTACCGGAAGAG	To mutate residue I202 to cysteine
I251C Fwd	CAAAGATAGTCAATGGATGTGCCAA AGCAAAAG	To mutate residue I251 to cysteine
I251C Rev	CTTTTGCTTTGGCACATCCATTGAC TATCTTTG	To mutate residue I251 to cysteine
T275C Fwd	GACGTATTTCAAAACATGTTACGCA CCGCTACCACTG	To mutate residue T275 to cysteine
T275CRev	CAGTGGTAGCGGTGCGTAACATGT TTTGAAATACGTC	To mutate residue T275 to cysteine
L303C Fwd	CAGGTTGTGATTTGTATGGAATATT AAGATGC AATATAACCAG	To mutate residue L303 to cysteine
L303C Rev	CTGGTTATATTGCATCTTAATATTCC ATACAAATCACAACCTG	To mutate residue L303 to cysteine
A335C Fwd	GACGTGGATTATTCCTAATTGTTAC ACTGAAAATAC	To mutate residue A335 to cysteine
A335C Rev	GTATTTTCAGTGTAACAATTAGGAA TAATCCACGTC	To mutate residue 335 to cysteine
L20C Fwd	CATCCGTGATATTGATGTCGTGTTT AGCTGTGGG	To mutate residue L20 to cysteine
L20C Rev	CCCACAGCTAAACACGACATCAATA TCACGGATG	To mutate residue L20 to cysteine
V220C Fwd	GATTAGCGCAATATCGCTATTTTGT TTCATTTTATTTTTC	To mutate residue V220 to cysteine
V220C Rev	GAAAAATAAAATGAAACAAAATAGC GATATTGCGCTAATC	To mutate residue V220 to cysteine

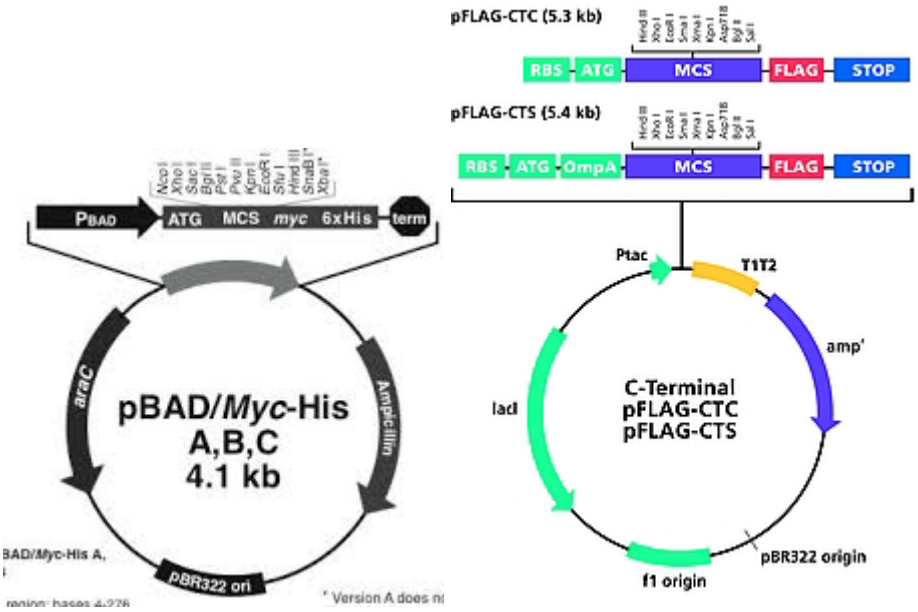


Figure 2.1: Maps of expression vectors pBAD/*Myc*-His A and pFLAG-CTC

2.3 DNA methods

2.3.1(a) Genomic DNA isolation was performed using REVOLUGEN for fire monkey

Following manufacturer's instructions genomic DNA was extracted using REVOLUGEN Fire monkey genomic DNA isolation kit. Approximately 2 ml of overnight culture was pelleted at 16,000 x g for 30 sec in an eppenddrouf tube. The supernatant was discarded, and the pellet was washed with 1x PBS and pelleted again same as above. The pellet was incubated for 37 °C for 10 minutes with 3 mg/ml lysozyme solution. After incubation 10µl of a 20 mg/ml, RNase A solution was added and mixed by vortexing and incubated for 5 minutes at room temperature.

2.3.1(b) Plasmid miniprep by alkaline lysis

The alkaline lysis method was performed for the isolation of plasmid DNA from the bacteria [174]. Approximately 1.5 ml of overnight bacterial culture was pelleted in a microcentrifuge at

6,000 x g for 30 sec. After discarding supernatant pellet was re-suspended in 100 µl of Solution I **(**Appendix A) and vortexed. This was followed by the addition of 200 µl of freshly prepared Solution II (Appendix A) to lyse the cells. 150 µl of cold Solution III (Appendix A) was added into the tubes and mixed by inverting the tubes gently. Precipitation of cell debris and chromosomal DNA was achieved by placing the tubes on ice for 5 minutes. The precipitate was centrifuged at 16,000 x g for 5 minutes and the supernatant was shifted to the new eppendorf tubes. Following this 0.5 µl of 1 mg/ml RNAase was added to the tubes and incubated for 10 minutes at room temperature to remove the RNA. DNA was precipitated out by adding two volumes of ice-cold 100% ethanol. The precipitate was pelleted at 16,000 x g for 5 minutes and washed with 300 µl of 70% ethanol and spun for 2 minutes. DNA pellet was vacuum dried using the Savant SC100 “Speed Vac”, and centrifuged. The pellet was resuspended in 20-25 µl of Milli-Q water and stored at -20 ^{o}C.

2.3.1. (c). Plasmid DNA isolation using Promega miniprep kit

Wizard plus SV Promega miniprep kit was used to isolate plasmid DNA for cloning and sequencing purposes. Briefly, approximately 1-10 ml of overnight bacterial culture was pelleted at full speed of 16,000 x g for 5 minutes in a microcentrifuge. Pellet was resuspended with 250 µl of cell resuspension solution followed by the addition of 250 µl of cell lysis solution. The tube was inverted 4 times to mix the contents and 10ul of alkaline protease solution was added and the tube was again inverted 4 times. The tube was incubated for 5 minutes at room temperature. Following this 350 µl of neutralization solution was added and the tube was inverted 4 times to mix and then centrifuged for 10 minutes at full speed. After spin, the cleared lysate was transferred into a spin column placed in a collection tube, and the column was centrifuged at full speed for 1 minute. The flow-through was discarded and the column was inserted back into the collection tube. For washing 750 µl of wash, the solution was added to the column and spun. Flow-through was discarded and the column was washed again with 250 µl of ethanol and centrifuge for 2 minutes at full speed at room temperature. The column was inserted into a new 1.5 ml microcentrifuge tube and 35-40 µl of nuclease-free water was

added and centrifuged for 1minute at room temperature to elute plasmid DNA and stored at -20 °C.

2.3.1(d). DNA for Colony PCR

LB plates containing appropriate antibiotics were used for the dilution streak of the required bacterial strains. A single colony was picked and re-suspended in 25 µl of 0.5 M Sodium hydroxide (NaOH) solution and were incubated for 30 minutes at room temperature. This was followed by the addition of 25 µl of 1 M Tris-HCl (pH 8.0) and 450 ul of sterilised Milli-Q water. 5 µl of this template was used for PCR reaction,

2.3.2 Determination of DNA concentration

NC-1000 spectrophotometer Nanodrop (Bioscience) was used to determine the concentration of DNA at 260 nm absorbance (A_{260}). The standard used for calculating the concentration of DNA in ng/µl as absorbance A_{260} of 1 represents 50 µg/ml. The purity of the DNA sample was indicated by determining the ratio of A260 nm / A280 nm. A ratio of approximately 1.8 was accepted for pure DNA without protein contamination.

2.3.3 Agarose gel electrophoresis

0.7% agarose gels were prepared in 0.5 x TBE buffer (Appendix A) containing ethidium bromide at a final concentration of 10 µg/ml. DNA samples were mixed with 1/10 volume of blue loading dye (Appendix A) to load onto the gel. All gels were run with 500 ng of *EcoRI* digested SPP-1 phage DNA (Figure 2.2) as a molecular size marker. Agarose gels were run in TBE buffer (0.5x) at 80-100 V (Volts), till the dye front reached 3/4th of gel. Gel-Doc was used to visualize the DNA from a sample under Ultraviolet light. Vision-capt (version 14.1 a) was used for the recording of photographs. Prints were taken on a thermal printer (Sony). SPP-1 marker was used for comparison and estimation of the size of DNA bands in the samples.

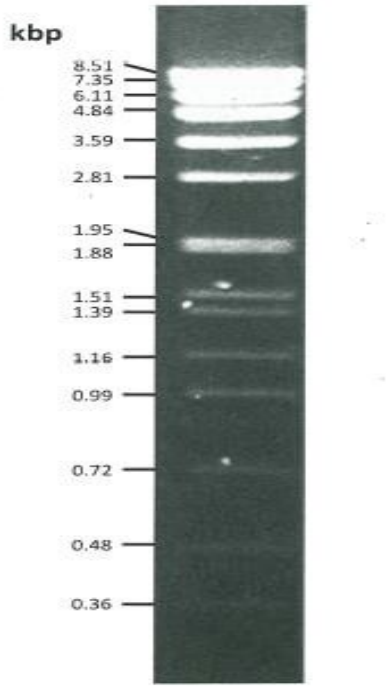


Figure 2.2: DNA marker ladder showing sizes of DNA fragments on left. DNA marker ladder was obtained by *EcoRI* digestion of SPP-1 bacteriophage DNA.

2.3.4 DNA purification

DNA bands of interest were cut from agarose gels under UV light with the help of a scalpel and transferred in a pre-weighed eppendorf tube. For purification of DNA from gel fragments and for enzyme reaction, Wizard SV gel and PCR kit (Promega) was used. Empty Eppendorf tubes, as well as gel slices, were weighed to determine the weight of the gel slice by subtracting the weight of the empty tube from the total weight. Per 10 mg weight of the gel slice, 10 µl of membrane binding solution was added. The tube was vortexed and kept in the incubator at 50-65 ^{o}C to dissolve the gel. In the case of an enzymatic reaction, an equal volume of membrane binding solution was added directly to the reaction. Treated DNA samples were transferred to the SV mini-column assembly and allowed to stand for 1 minute at room temperature. The column was centrifuged at 16,000 x g for 1 minute and flow-through was discarded. 700 µl of membrane binding solution was added to the column and was spun for a further one minute. Flowthrough was discarded again, and the column was washed with 500

µl of membrane wash solution and centrifuged for 5 minutes at 16,000 x g. Flowthrough was discarded again. To ensure removal of any residual membrane-wash solution mini-column was spun at 16,000 x g for a further 1 minute. The spin column was placed in a clean 1.5 ml eppendorf tube for elution of DNA and 40-50 µl of Milli-Q water was added to the column and left for 1 minute in standing position. Finally, the columns were spun at 16,000 x g for 1 minute to collect the DNA in eppendorf tube. The purified DNA was stored at -20 ^{o}C temperature.

2.4 Polymerase Chain Reaction (PCR)

2.4.1 Primer design

All primers used were manufactured by Sigma Aldrich (USA). Primers were 16-23 base pairs (bp) long with 50-64^{o}C melting temperature (T_m). In the case of cloning restriction enzymes sites were added to the primer sequences complementary to the template, and extra bases were also added where needed. The absence of secondary structure was confirmed on OligoEvaluator^{TM} Sigma-Aldrich. **Table 2.5** shows the list of primers used in the present research study and the restriction sites introduced are underlined.

2.4.2 Amplification of genes for cloning

Routine PCR was carried out using Taq polymerase while Stratagene (PfU Ultra II polymerase) was used for cloning and sequencing purposes and Phusion polymerase was used in some site-directed mutagenesis PCRs. The typical PCR reaction included 10x *polymerase* buffer, DNA polymerase µl (1 unit of Taq polymerase or 0.5 units of *Pfu* ultra II), Primer 1 (5 µM), Primer 2 (5 µM), 2.5 mM dNTPs, Template DNA (5 ul of colony DNA or approximately 10 ng of plasmid DNA), and Milli Q water up to 20 µl.

The following formula was used to calculate the annealing temperature (T_m) from the primer nucleotide composition. The following equation was used for Tm determination.

$T_m = 2 \times [A+T] + 4 \times [G+C]$.

Annealing temperature of PCR reaction was taken 5°C lower than the lowest Tm calculated for the used primer pairs. Other parameters used for reactions were as below:

Cycle	Temperature	Time	Repeat
Initial denaturation	95°C	5 min	1
Denaturation	95°C	30 sec	35-40
Annealing	T_m minus 5°C	30 sec	
Extension	72°C	1 min / kb for *PfU*	
Final extension	72°C	Extension + 2min	1
Hold	4°C	∞	∞

Amplified PCR products were analysed on 0.7% agarose gels purified using Wizard SV gel and PCR clean-up kit (Promega) and stored at -20°C.

2.4.4 Sequencing reaction

Big Dye Terminator sequencing method was used for the sequencing reaction according to the Biomedical Resources Facility (BRF), John Curtin School of Medical Research (JCSMR), ANU.

20 µl reaction mix was prepared by adding 1 µl of Big Dye Terminator (ready reaction premixed), 3.5 µl of 5x Big dye reaction buffer, 0.125 µM of primer and Template DNA. As per the requirement of the JCSMR protocol template concentration for plasmid and linear DNA used was 150-200 ng and 50 ng respectively. The parameters used to carry out the thermal cycle sequencing were 25 cycles of 96°C 10 sec, 50°C 5 sec for 4 minutes, 60°C for 4 minutes, The temperature was held at 4°C and the reaction was made up to 20 µl Milli Q water

20 µl of reaction was transferred to 1ml eppendorf tube and 2 µl of 3 M sodium acetate and 60 µl of absolute (100%) ethanol were added. The tube was briefly vortexed and left for incubation at room temperature for 15-25 min. This was followed by centrifugation for 20-40 minutes at 16, 000 x g. The pellet was carefully washed with 250 µl of freshly prepared 70% ethanol and spun for 5 minutes at 16,000 x g. The supernatant was aspirated, and the pellet

was dried in a vacuum for 15 minutes onto the SC100 Speed Vac centrifuge (Savant). Samples were stored at -20^{o}C. or were sent for automated sequencing at the BRF (JCSMR, ANU).

2.5 Site-directed mutagenesis.

Protocol from Quick change site-directed mutagenesis kit (Figure 2.3) was used to perform mutagenesis. In brief, the entire plasmid was amplified using primers containing the desired mutation(s). Details of the primers used are given in Table.2.5. DpnI digestion was performed to remove any parental DNA followed by transformation into XLI blue electrocompetent cells for the repair of nick DNA in the mutated plasmid.

As per the protocol the mutagenesis primers were designed with 25-45 bases having the desired mutation in the middle and a C or G at the 3' end. The T_m was kept >78 °C and GC content approximately 45-50%. The primers were designed to anneal to the same sequence but on opposite strands.

PCR parameters used are as: cycle 1 (I x) 95°C/30 seconds (Initial denaturation), Cycle 2 (18x), 95°C/30 seconds (denaturing DNA), (T_m) °C/30 seconds (primer annealing), and 68°C/2 minutes/kb (strand extension).

T_m was calculated using the following formula:

$T_m = 81.5+0.41(\%GC)-675/N -\%$ mismatch

where N is the length of the primer in bases and the values for %GC and % mismatch are whole numbers.

The amplification of the product was confirmed by running 5 µl onto an agarose gel. Only after getting amplified product on a gel, the procedure was continued. The DpnI treatment was carried out of the PCR product following transformation into XLI blue cells. 5-10 colonies were inoculated for plasmid isolation and sequencing.

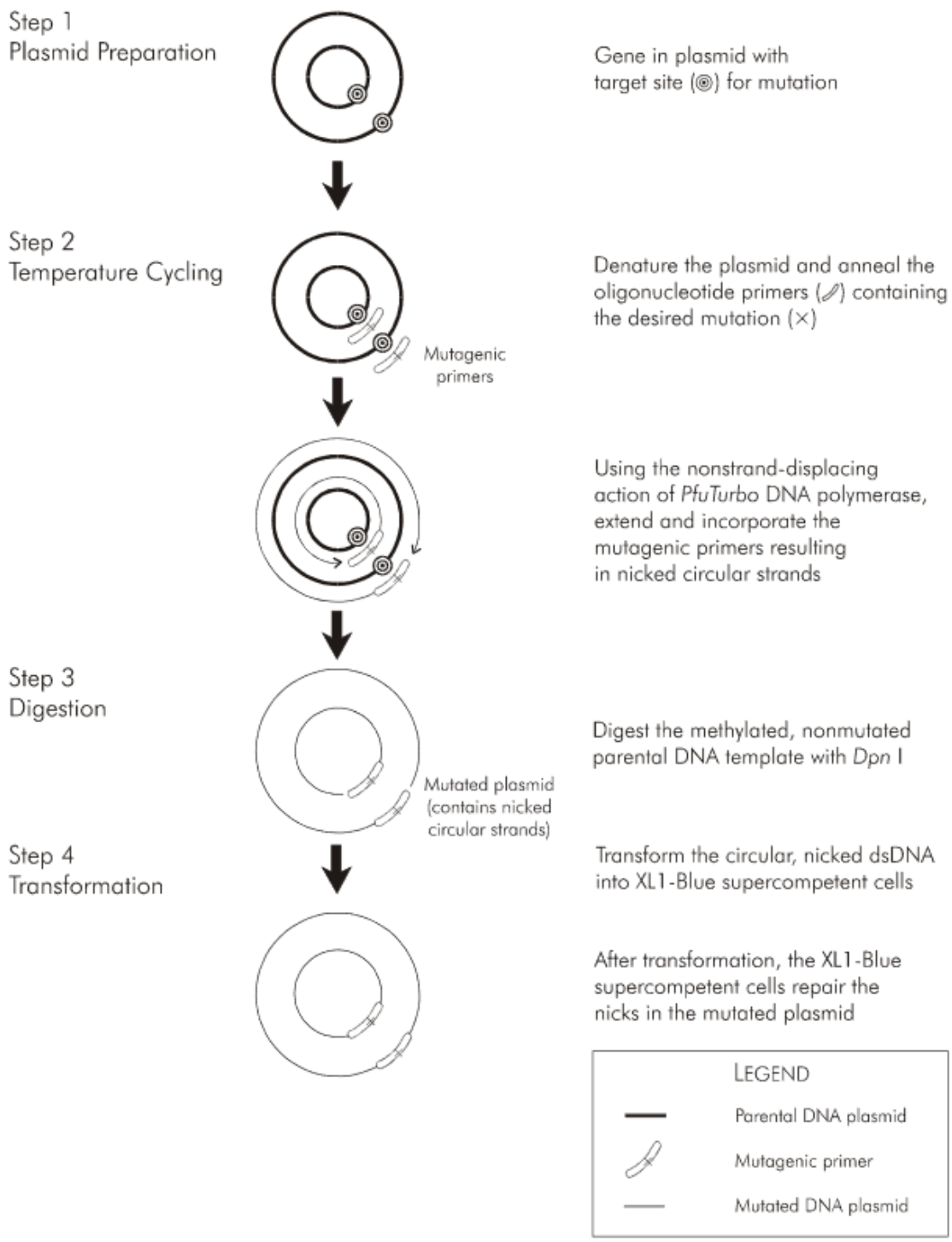


Figure 2.3: Summarized view of steps involved in site-directed mutagenesis. picture taken from Stratagene manual.

2.6 Cloning

2.6.1 Restriction Endonuclease (RE) digestion

Restriction enzymes were used to carry out single/double digestions of the plasmid DNA screening purposes. Digestion of plasmid DNA and PCR products were carried to produce cohesive ends for cloning. Enzymes and buffers were purchased from New England Biolabs and digestion reactions were performed as per the manufacturer's instruction. The composition of RE digestion mix for digestion of DNA for a single reaction carried DNA, 1 x RE buffer, 5-20 units of RE and sterile Milli-Q water to make up the total volume.

For double digestion, a similar reaction mix was prepared with two REs per reaction at the same time. Single or double digests were performed for screening purposes using 20 µl total volume containing 200-300 ng of DNA sample. The mixture was incubated for two hours at an optimal temperature. Digestion reaction was blocked by adding blue loading dye for running on the 0.7% agarose gel immediately. Total 50 µl volume using 1 µg DNA was incubated at optimal temperature for 3 hours, was prepared for the single or double digestion for the cloning purposes. DNA was purified directly or after running onto agarose gel following incubation (as described in Section 2.4.2). Heat inactivation was also performed in some cases to inactivate the enzymes.

2.6.2 DNA Dephosphorylation

Calf intestinal alkaline phosphatase (CIAP) by Fermentas was used to dephosphorylate DNA ends after digestion. This was done to stop digested plasmid DNA from recircularization during ligation reaction. Compatible Fermentas buffers were used with CIAP. After adding CIAP, DNA was incubated at 37 °C for 30 min and then heat-inactivated at 65 °C for 15 min.

2.6.3 Ligation Reaction

Both plasmid vector and insert DNA were digested with RE and ligated in a 10 µl reaction mixture to construct the recombinant plasmids. Other components of ligation reaction mix

include 1x of T4 DNA ligase buffer (Promega), and 0.5 unit of T4 ligase (Promega) in sterile Milli-Q water.

A molar ratio of 3:1 to 6:1 of vector DNA and insert DNA was used to achieve the best possible recombination outcome. The reaction mix contained 150 ng of vector DNA while the amount of insert DNA was calculated by formula as under.

$$\text{Insert amount (ng)} = \frac{3 \times \text{vector amount (150 ng)} \times \text{insert size (bp)}}{\text{Vector size (bp)}}$$

Concentrations of the insert and plasmid DNA were estimated by spectrophotometer.

The reaction was incubated overnight at 15°C. After incubation, the ligation reactions were used for bacterial transformations or were stored at -20°C for later use.

2.7 Transformation of DNA into competent cells

DNA was transformed into bacterial cells either by electroporation or heat shock methods.

2.7.1 Electrocompetent cells preparation

The cultures of *E. coli* and *S. flexneri* were grown overnight with shaking at 200 rpm at 37°C and 30°C respectively. The pre-warmed LB with or without appropriate antibiotic was used to dilute (1/100) overnight bacterial cultures and incubated at 30/37°C in a shaker incubator at 200 rpm, to obtain log phase growth OD_{600} (0.5-0.6). The growth was stopped by placing cultures on ice and incubated for 30 minutes at 4°C. All the onward steps were performed at 4°C temperature. Then the cultures were shifted into pre-chilled 50 ml falcons in equal volumes. These were centrifuged in a benchtop centrifuge at 3750 rpm for 15 min at 4°C. This was followed by re-suspension of pellet in 30 ml of sterile cold Milli-Q water and spun as above. The pellet was washed twice with 20 and 15 ml sterile cold Milli-Q water respectively. The cell pellet was re-suspended in 25 ml sterile cold glycerol (10%) and centrifuged as above.

Obtained pellet was re-suspended in sterile cold 10% glycerol (10 ml) and centrifuged again for 15 minutes.

In the last, the pellet was re-suspended in 100-200 µl sterile cold glycerol (10%) and 40 µl aliquots of cells were transferred to pre-chilled eppendorf tubes (1.5 ml) and were immediately stored at -80^{o}C.

2.7.2 Electroporation

Electrocompetent cell aliquots (40 µl) were thawed on ice before adding 0.5-1 µl of plasmid DNA or ligation mix and were moved to pre-chilled sterile electroporation (BioRad) cuvette. Electroporation was performed on BioRad Gene Pulser by passing a pulse of 200 ohms, 25 microfarads, and 2.5 kilovolts through the cells. Thereafter 1 ml of LB was added immediately to the cuvette and the mix was moved to the eppendorf tube (1.5 ml). The cells were incubated at 200 rpm for 1 hour at 37^{o}C and post-incubation spun to pellet at 16,000 x g for 30 sec. The supernatant was discarded, and the pellets were resuspended by vortexing and plated on LB agar plates containing appropriate antibiotics. LB agar plates were incubated overnight at 37^{o}C.

2.7.2.1 Calculation of efficiency of competent cells

To determine the efficiency of competent cells a known concentration of pBCSK (+) was transformed into freshly prepared competent cells and by determining the number of colonies obtained. 1/10,1/100,1/1000 and 1/10000 dilutions were performed to obtain a suitable number of colonies. The following formula was used to calculate efficiency:

Number of transformed cells (N) = Number of colonies x Dilution factor

N/µg DNA (efficiency) = N/amount of DNA transformed (µg)

2.8 Screening of cloned plasmids and disruption mutant strains

2.8.1 Antibiotic selection

The clones were screened for antibiotic resistance of the introduced plasmid and then and restriction enzyme digestion of the miniprep DNA. Gene disruption mutants containing integrated kanamycin-resistant gene were selected on the 50 µg/ml kanamycin LB plates (Appendix A). 25 µg/ml of chloramphenicol (Appendix A) was added to LB (LB agar) to cells containing plasmid pBC SK+ or derivative. Erythromycin-resistant gene containing recombinant plasmids were selected on 250 µg/ml of erythromycin (Appendix A).

2.9 Gene disruption by lambda red recombination approach

2.9.1 Construction of the knockout template

A PCR-based approach was used for constructing a knockout template (Figure 2.4). In a PCR-based approach, the *kan* (kanamycin gene) was amplified from pKD4. Primers containing 70 bp homology to the target gene were used for this purpose. The resulting PCR product was agarose gel purified. DpnI digestion was performed to remove the plasmid template, purified, and stored at -20^{o}C.

2.9.2 Transformation of the knockout template in S. flexneri red recombinase induced strains

The lambda red genes (*gam*, *beta* and *exo*) encoding plasmid pKD46 was transformed into *S. flexneri* strain for lambda red mediated homologous recombination. Strains of *S. flexneri* harbouring pKD46 were used to prepare electrocompetent cells. 2 ml of the overnight culture was used to inoculate 100 ml of SOB medium (Appendix A) containing 100 µg/ml ampicillin. After 1.5 hours of incubation at 30^{o}C in a shaker incubator (180 rpm), the induction of expression of lambda red genes was performed by adding 100 mM arabinose. Following an additional 2 hours of incubation at same temperature, the electrocompetent cells were prepared as described in section 2.7.1. The knockout template was transformed into *S. flexneri* as described in section 2.9.1. Cells were recovered following transformation in 1 ml of

SOB medium with arabinose (Sigma-Aldrich) added to a final concentration of 100 mM. After 3 hours of incubation at 30°C in a shaker incubator (180 rpm), the cells were spun down at 16,000 x g for 1 minute. After removing the supernatant, the pellet was resuspended plated on a LB agar plate containing 50 µg/ml kanamycin and incubated overnight at 30°C. The obtained colonies were screened for successful disruption using colony PCR (Figure 2.4).

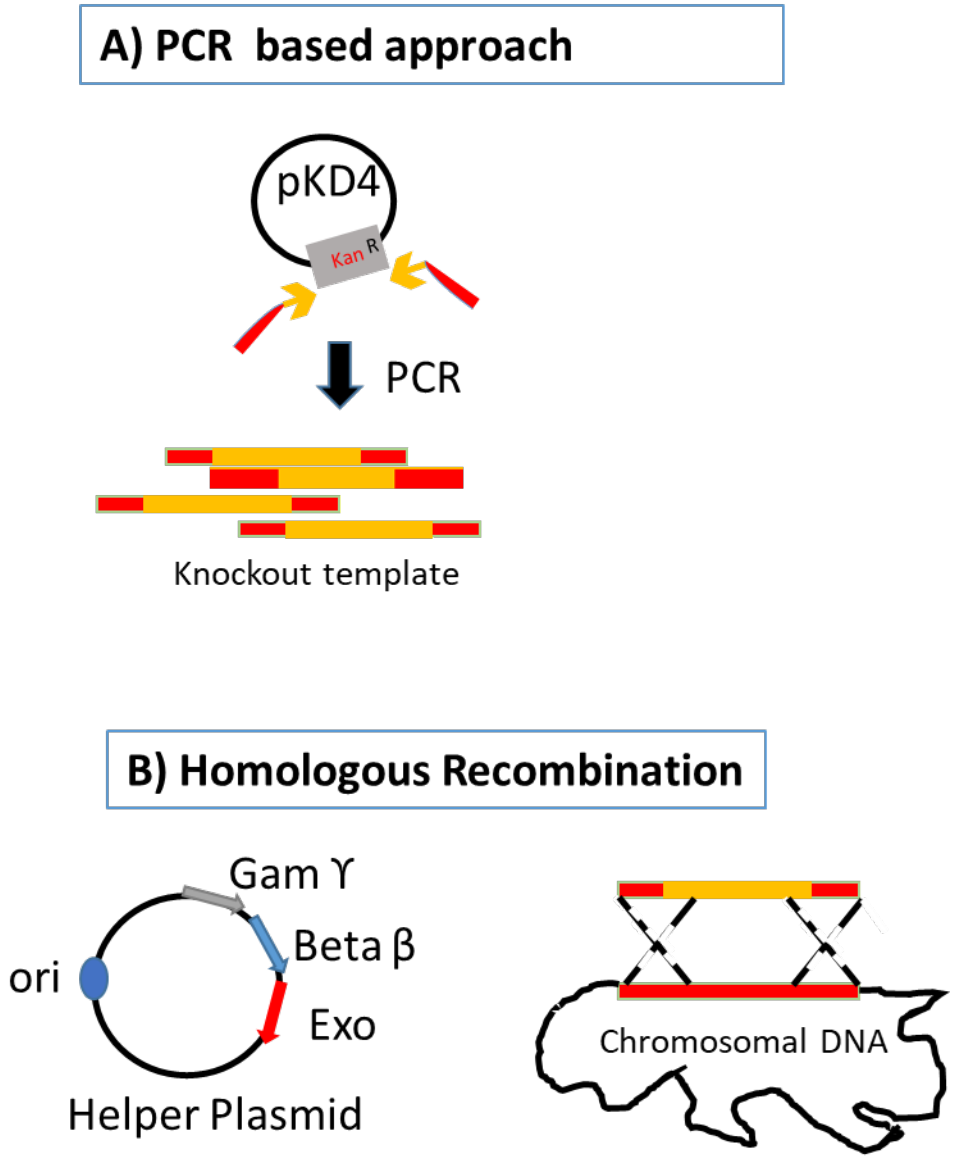


Figure 2.4: Schematic presentation of lambda red mediated recombineering approach for the gene disruption.
(A) Using 70 bp long primers containing homology to the target gene, the antibiotic cassette was amplified. (B) Expression of lambda red recombinase genes in the introduced *S. flexneri* strain

2.10 Absorption of antisera

Antiserum was raised in the sheep immunized with the serotype 1c strain, SFL1683 (positive for *oacB*). Crude antiserum collected was purified to render it specific to 3/4 O-acetylated Rha III epitope. All the nonspecific antibodies in the antiserum were absorbed by mixing antiserum with the heat killed SFL1691 cells (*oacB* negative 1c strain). Slide agglutination using prepared antiserum was performed and agglutination observed with strain SFL1683 only and not with SFL1691 within one minute was considered specificity of antiserum to 3/4 *O*-acetylation.

2.10.1 Slide agglutination assay

Dilution-streaked culture of required *Shigella* strain was grown overnight on LB agar plates with or without appropriate antibiotics at 30°C. Next day a single colony from the plate was picked with the help of a sterile toothpick or wire loop and mixed with a drop of antiserum (antibody) on a glass slide. The slide was rocked gently back and forth and observed for agglutination. 1x PBS was used in parallel for the negative agglutination test instead of antibody.

2.11 Conjugation

Exponential phase cultures of both donor and recipient strains were mixed in equal volumes (0.5 ml each) in an Eppendroff tube and the mating mixture was incubated at 37°C in a water bath for 40- 60 minutes. After incubation, the tube was spun at full speed for 2 minutes and the pellet was resuspended in 1ml PBS and spun again, the pellet resuspended in 1 ml of PBS. Serial dilutions were prepared and plated on LB plates supplemented with either kanamycin (50 µg/ml) or Cm (25 µg/ml) antibiotics to differentiate the conjugative transfer of virulence plasmid.

2.12 Protein expression

2.12.1 Whole-cell lysate

A single colony of required bacterial strain was grown in LB containing ampicillin (100 μg/ml) and incubated overnight at 37°C at 180 rpm. The next day 10 ml of warm LB_{Amp100} was inoculated with 150ul (1/66 dilution) of overnight culture in a 50 ml falcon tube and incubated at 37°C for ~3 hours in a shaker incubator until O.D. reached 0.4-0.6. The cultures were induced at OD_{600} 0.4/ 0.5/ 0.6 with either IPTG or L-arabinose. The concentration of IPTG used was 0.25 mM, 0.5 mM and, 1mM whereas, L-arabinose was used in 0.1%, 0.002%, 0.02% and 0.2% (w/v) concentrations. Following induction, cultures were grown for 4 hours at 37 °C at 180 rpm for four hours. After which 1 ml culture was taken in a sterilised eppendroff and spun down at full speed for 1 minute. The supernatant was discarded, and the pellet was resuspended in a 2x loading buffer. The prepared sample was used immediately to run the SDS-PAGE gel.

2.12.2 Membrane preparation

A single colony of required bacterial strain was grown in LB containing ampicillin (100 μg/ml) and incubated overnight at 37°C at 180 rpm. The next day 10 ml of warm LB_{Amp100} was inoculated with 150 μl (1/66 dilution) of overnight culture in a 50 ml falcon tube and incubated at 37°C for ~3 hours in a shaker incubator until O.D. reached 0.4-0.6. The cultures were induced at OD_{600} 0.4, 0.5, and 0.6 with either IPTG or L-arabinose. The concentration of IPTG used was 0.25mM, 0.5 mM, and 1mM whereas, 0.1%, 0.002%, 0.02% and 0.2% (w/v) concentrations of the L-arabinose was used. Following induction, cultures were grown for 4 hours at 37°C at 180 rpm for four hours and then harvested by centrifugation at 10,000 x g for 15 min at 4°C. The pellet was washed twice in a 50 ml falcon tube with 30mM Tris-HCl (pH 8.0) and spun 3000 x g for 10min at 4 °C and supernatant was discarded. Pellets were stored at -80°C.

2.12.3 Protein concentration estimation

Protein concentration was determined using Pierce bicinchoninic acid (BCA) Protein Assay kit (Thermo scientific). Details can be found at http://www.piercenet.com/files/l296as8.pdf.

2.12.4 (a) OacB purification-His affinity

The membrane protein pellets were resuspended in 400 µl of Buffer A containing 300 mM NaCl in 50 mM Tris-Cl and 0.5 % (w/v) n-dodecyl-beta-D-maltoside (DDM) and 5 mM imidazole and homogenized in a glass tissue homogenizer. Proteins were left overnight to homogenized at 4 °C. The next day solubilized protein was mixed with 100 µl Nickle -Nitro acetic acid- NTA UNOsphere base matrix (Bio-Rad Nuvia IMAC Resin). To remove other interfering proteins, a wash buffer containing imidazole 20-50 mM in Buffer A was used. Elution of His-tagged protein (OacB expressed in pBAD/*Myc*-HisA vector) was performed using a high concentration of imidazole (500 mM). Purified protein was used immediately or stored at -80°C.

2.12.4 (b) OacB purification- immunoprecipitation

OacB from membrane protein sample, was isolated using Anti-FLAG^{R} M2 Affinity Gel (Sigma-Aldrich). All the purification steps were performed within 2-8 °C and pre-cooled lysis/wash buffers and equipment were used. In a fresh 1.5 ml eppendrouf tube, 40 µl of well suspended Anti-FLAG^{R} M2 Affinity Gel was transferred and spun at 5000 x g for 30 seconds. With a narrow-end pipette tip, the supernatant was carefully removed without disturbing the resin. The resin in the pellet was washed 2 x with 500 µl of TBS (Appendix A). The tube was spun at 5000 x g for seconds to collect resin. The resin was then washed with 0.1M Glycine-HCl (0.1M Glycine, pH 3.5) to clear up any residual unbound Anti-FLAG antibody from the resin suspension. The tube was spun again same as above, and resin was collected. The resin was washed 3 x with TBS and collected as above. 200 µl of prepared membrane protein sample suspended in Buffer A (50mM Tris-HCl, pH 8.0, with 100mM NaCl) was added to the resin and incubated overnight on a rocker to enhance binding efficiency. Next day resin was collected through centrifugation as above and the supernatant was removed using narrow-

end pipette tip. The washing of resin was performed thrice with TBS and collected as above through centrifugation. Later, 20-40 µl of 2 x sample buffer was added to the mixture and heated at 100°C. The sample was spun (as above) and the supernatant was collected and was used either immediately for SDS-PAGE or stored at -80°C.

2.13 SDS-PAGE

2.13.1 SDS-PAGE gel preparation and electrophoresis

SDS-PAGE electrophoresis was performed by either making 4% stacking and 12% resolving SDS gels (Appendix A) or by using Mini-Protean pre-cast gels. Before loading membrane fraction samples, the gels were set in 1 X SDS-PAGE buffer (Appendix A). The pre-cast gel was removed from the storage pouch and prepared for assembly in the Bio-Rad Electrophoresis apparatus. The green tape was removed from the bottom of the cassette and the gel was kept in the electrophoresis tank. The comb was removed by pulling it gently outward and the wells were rinsed with 1 X SDS running buffer (Appendix A). Chambers from inside and outside were filled with the running buffer. Prepared protein samples with the loading buffer and β-mercaptoethanol were loaded and the gel was run at 80-100 V until the dye front reached the reference line imprinted on the bottom of the cassette. A Pierce blue pre-stained Protein Molecular weight marker (Thermo scientific) was also loaded along with the samples in each gel (Figure 2.5).

2.13.2 Coomassie Staining

After SDS-PAGE gel electrophoresis of membrane protein samples (Section 2.11.2), 200 ml Coomassie brilliant blue stain (Appendix A) was then used to stain the gel and left for overnight. The next day the gel was destained using 100 ml of destaining solution for 15 minutes (Appendix A).

2.13.3 Silver staining

Following SDS-PAGE gel electrophoresis, the SDS gel was placed in 500 ml of fixing solution (Appendix A) 2 x for 30 minutes on a rocking platform. The gel was then sensitized for an hour or overnight with the sensitizing solution (Appendix A) on a rocking platform. The gel was rinsed 3 x 30 minutes with MQ water (1L/wash) on a rotating platform. The gel was stained with freshly prepared silver staining solution for 2 hours on a rocker. After staining the gel was rinsed with 1L MQ water for 10 seconds. Then gel was developed with the development solution until the desired band intensity was achieved. To stop the development, the gel was blocked with the blocking solution (Appendix A) for 10 minutes on a rotating platform. The gel was stored in MQ water.

2.14 Western blot

Membrane protein samples were transferred to polyvinylidene fluoride (PVDF) membrane in Bio-Rad Trans-Blot Transfer Cell apparatus. Directly after SDS-PAGE, the gel was kept in transfer buffer (Appendix A) for 15 minutes and PVDF was charged by placing it in methanol. Two supporting pads along with two gel-sized Whatman papers were soaked in the transfer buffer for 15 minutes. The assembly of the transfer sandwich was performed according to the manufacturer's instructions (Bio-Rad) and kept in the Bio-Rad Mini Transfer cell apparatus and the tank was filled with transfer buffer up to the level needed for one or two gels. The transfer was performed at 4°C at 40V for 4 hours. Once the transfer was complete the membrane was soaked in 5% (w/v) skim milk (Diploma) prepared in phosphate buffer saline (PBS; Appendix A) for blocking for 1 hour or overnight. Following blocking membrane was

washed 3x for 15 minutes per wash, in PBS containing 0.05% (v/v) Tween-20 (Sigma - Aldrich). After washing membrane was incubated with either anti-His antibodies or anti-FLAG antibodies overnight and 1 hour, respectively. Both the antibodies were conjugated to horseradish peroxidase (R&D). The antibody dilutions were prepared in PBS with 0.1% (v/v) Tween-20 and 15 (w/v) skim milk.

The membrane was washed afterward three times (15 minutes each time) with PBS containing 0.05% Tween-20. To detect bound antibodies, Super Signal West Pico Chemiluminescent substrate (Pierce) and enhancer and peroxidase buffers mixed in 1:1 concentration were poured onto the membrane. The membrane was incubated for five minutes and bound proteins with antibodies were detected using chemiluminescent camera dock (Vilber-Lournat) and Vilber-Lournat software version 15.11.

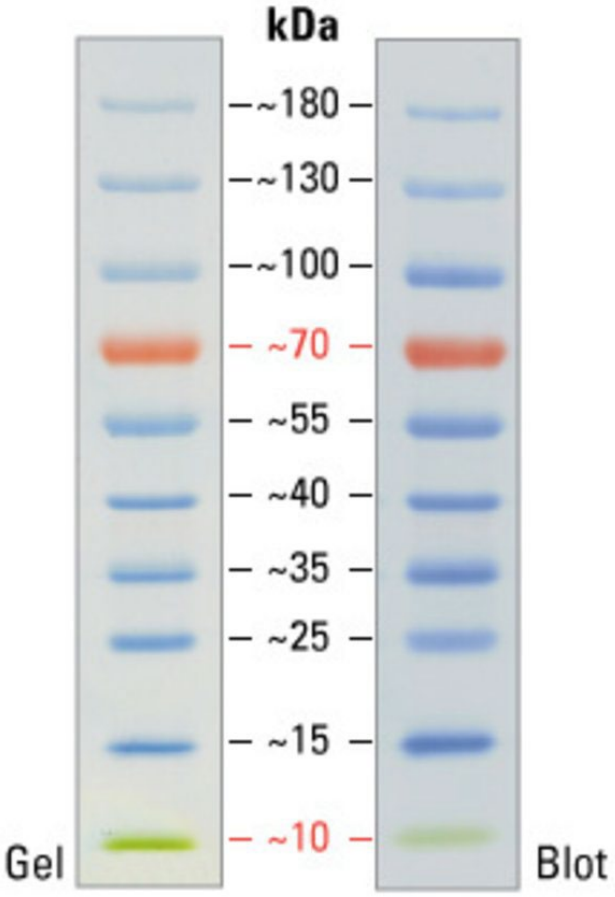


Figure 2.5: The Prestained Page-Ruler ladder (Thermofisher) used in SDS-PAGE gels.
The approximate sizes are shown in the middle of gel and blot images.

2.15 RNA methods

2.15.1 RNA Isolation (Hybrid method)

Bacterial RNA degrades very quickly and difficult to handle hence total RNA of the strain was extracted using trizol and RNAeasy mini kit (hybrid method). Briefly, the culture of the required *S. flexneri* strain was grown overnight at 30^{o}C, and the next day, overnight culture was mixed with LB agar in the 1:50 dilution to grow at 37^{o}C to an early log phage (OD_{600} = 0.6-0.65). RNAase Mini Kit (Qiagen) was used for the RNA isolation. The required volume of bacterial culture for 1 x 10^{9} CFU was calculated. Bacterial cells were centrifuged to harvest at 3,200 x g for 10 minutes at a temperature of 4^{o}C. The supernatant was discarded, and the pellet was re-suspended in 200 µl TE buffer containing lysozyme by vortexing. Following this RLT buffer

(700 µl) was added to the tubes and incubated at room temperature for 10-15 minutes and vortexed vigorously. Ethanol (500 µl) was added to the tubes and the solution was gently mixed by pipetting. Tubes were shifted to an assembly of RNAase Mini spin column. The column was centrifuged for 15 seconds at 8,000 x g and the flowthrough was discarded. Colum was washed first with RW1 buffer (700 µl) followed by RPE buffer (500 µl) and was centrifuged for 15 seconds at 8,000 x g for each time. After the final wash and centrifugation at the above standards, the column was shifted to a fresh Eppendorf tube. Following which 30- 50 µl of RNAase-free H_2O was added to the center of the column. The column was allowed to stand for 1 minute at room temperature. Then it was centrifuged at 8,000 x g for 1 minute to elute the RNA.

2.15.2 RNA quantification

NanoDrop ND- 1000 spectrophotometer (Bioscience) was used for the quantification of RNA samples. Absorbance was measured at 280 nm (A_{280}). RNA quantity was measured in ng/µl. The Standard of quantification was A_{280} of 1 represented quantity of 40 µg/ml. The purity of RNA was measured by the ratio of A_{260}/A_{280} nm. A ratio of ~2.0 represented pure RNA.

2.15.3 DNAase treatment of RNA

Removal of contaminating genomic DNA was achieved by treatment of RNA samples with DNAase. At 1x final concentration,10 U of Turbo DNAase (Ambion) along with Turbo DNAase buffer were added to the RNA samples and incubated at 37°C for 0.5-2 hours. DNAase inactivation reagent (1x volume) was then added to the tubes to inactivate the enzyme. Tubes were incubated for 5 minutes at room temperature. Inactivation reagent was pelleted by centrifugation for 1.5 minutes at 8,000 x g. The supernatant, containing RNA, was shifted into a new eppendorf tube. RNA samples were stored at -80°C.

2.15.4 Reverse Transcriptase PCR (RT-PCR)

Superscript II Reverse Transcriptase (Invitrogen) was used to perform the RT- PCR. The procedure was performed according to the manufacturer's instructions. Total RNA (DNAase treated) and random hexamer primers were used for the synthesis of 1st strand cDNA. 1 µl of

cDNA was used to set subsequent PCR using specific primers and *PfU* polymerase (Section 2.4). Two negative controls were used for all reactions, a negative RT-PCR control without reverse transcriptase enzyme and a PCR control for PCR reagent contamination.

2.16 Bacteriophage techniques

2.16.1 Bacteriophage induction

Induction of Sf101 bacteriophage was performed from wild type *S. flexneri* strain, SFL1683. Ultraviolet irradiation protocol was used for induction as described by Adams *et al*., [175]. Overnight bacterial cultures were diluted with 20 ml fresh LB agar in a 1:20 ratio. This could grow log phage at 37^{o}C with shaking at 200 rpm. 5 ml of culture was pelleted by centrifugation at 3,200 x g for 15 minutes. The supernatant was discarded, and the pellet was re-suspended in 2.5 ml of 10 mM magnesium sulfate ($MgSO_4$). Samples were shifted into a petri dish and kept under UV light for 2 minutes (UV light- Gelman Clemco, 10 cm distance, 254 nm). Samples were shifted to Falcon tubes, containing 15 ml LB, and wrapped with aluminum foil to protect from direct light, and incubated at 37^{o}C temperature. Tubes were spun for 15 minutes at 3,200 x g after incubation in the dark at 37^{o}C overnight. Supernatants containing any released phage were filtered to pass through 0.2 µm-pore-diameter filters (minisart) for removing any bacterial contamination.

2.16.2 Determination of phage titre

Phage titer was expressed as plaque-forming units (PFU). It was determined by the double-layer agar technique as described by Sambrook *et al.,* [174]. In short, phage stock serial dilutions of 10^{-1} to 10^{-8} were prepared in SM buffer (Appendix A). Soft agar was warmed in advance at 42^{o}C temperature (Appendix A). 100 µl phage sample was mixed with 100 µl of overnight bacterial culture containing susceptible strains. The phage/bacteria mix was allowed to incubate for 20 minutes at 37^{o}C temperature. Pre-warmed soft agar (3 ml) was then mixed into the phage/bacteria mixture and transferred to the LB agar plates. Agar plates were then swirled gently. Plates were air-dried at room temperature followed by overnight incubation at

37°C. Plaques in each plate were counted on the following day. The number of plaque-forming units per ml (PFU/ml) was determined to measure the phage titer in the original stock.

2.16.3 Bacteriophages propagation and purification

Stocks of bacteriophage were prepared by the propagation of phage on serotype Y strain SFL124 (SFL1353). Plaque assays were performed on the modified double-layer agar (DLA) as described by Santos *et al.,* [176]. In this technique, phage lysate (0.1 ml) was mixed with stationary phase culture (0.1 ml) of the host bacteria and allowed incubation for 20 minutes at 37^{o}C. Sterile tooth prick was used to pick a single plaque and was inoculated on NZCYM media (5 ml) (Appendix A). Now 5 µl of SFL1353 overnight culture was added to the tube. This was followed by incubation at 37^{o}C with shaking at 200 rpm till lysate was noted. The spinning of lysate was performed at 3,200 x g for 15 minutes to pellet the bacterial cells. The supernatant, containing phage, was passed through the filter for sterilization. The supernatant was then added to NZCYM media (100 ml) containing SFL1353 (100 µl) overnight culture. This was incubated at 37 ^{o}C temperature with shaking at 200 rpm till lysate was obtained.

Centrifuge bottle was used for collection of 100 ml culture and was centrifuged at 3,200 x g for 15 minutes. The supernatant was used to propagate phage in 500 ml and 1000 ml NZCYM media. 1000 ml of phage lysate was purified and concentrated using PEG/NaCl to get purified bacteriophage stocks.

2.16.4 Bacteriophage Sf101 lysogens

3 ml of soft agar was enriched with 100 ul of SFL1353 overnight culture. It was poured on LB agar plate and allowed to set. This plate was spotted with 100 µl of bacteriophage stock (10^{11} PFU). Plates were dried and incubated at temperature 37^{o}C for 16 hours. Obtained colonies of the clear zone were observed and were screened by *oacB* PCR and slide agglutination.

2.17 *C. elegans* methods

2.17.1 Nematode strain and maintenance

Wild type N2 *C. elegans* strains were used from the Behm laboratory, The Australian National University. Modified nematode growth medium (NGM) was used for the culture and growth of nematodes (Appendix A). Nematodes were cultured on plates containing *E. coli* strain OP50 at temperature 22°C. Leica MZ7.5 dissecting microscope was used for the observation of the worms.

2.17.2 Synchronisation of nematode

E. coli OP50 NGM plates chunked with a mixed population of worms and incubated at 22^{o}C for 2-3 days. Growth was observed by lots of eggs and gravid adults. Each plate was added with 10 ml of sterile S-Basal (Appendix A). Worms were collected from the plate by using a glass spreader and shifted to a 15 ml falcon bottle. The falcon containing worms was spun at 500 x g for 3 minutes at room temperature. Supernatant S-basal was aspirated without disturbing the worm pellets and only 2ml S-basal was left in the tube. 500 ul of the alkaline hypochlorite solution was then added to the tube (Appendix A). The tube was vortexed vigorously at room temperature for 90 seconds.10 ml of S-basal was added into the tube was then spun at 500 x g for 3 minutes to collect the eggs. The supernatant was aspirated without disturbing the pellet. To remove all the bleach eggs were washed three more times with 10 ml S- basal. At the final wash, 5 ml of S- basal was transferred to the pellet, and the tube was kept on a rocker overnight at room temperature to allow eggs to hatch. Next day, L1s were observed either by the naked eye or under a microscope and concentrated by centrifugation at 500 x g for 3 minutes and were quantified. A quantity of 200 worms was seeded in the *E. coli* OP50 NGM plates. Growth was allowed at 22^{o}C for reaching L4 stage.

2.17.3 C. elegans accumulation assay

A synchronized worm population of L4 stage of *C. elegans* was removed from the *E. coli* OP50 NGM plates by washing plates with S- basal. The worms were treated with 200 µg/ml of

gentamicin for 3 hours to remove any *E. coli* OP50 cells present on the worm surfaces. Later the worms were washed with sterile S- basal for removing traces of antibiotics. Bacterial strains being used for the accumulation assay were grown on NGM plates overnight at temperature 37^{o}C, this was to stimulate the expression of virulence plasmid-encoded genes. Plates were cooled at room temperature for ~2 hours before inoculation with 50-100 gentamicin-treated worms (L4 stage). Post inoculation plates were incubated for 24 hours at 22^{o}C. Next day, 20 worms were selected from each plate and were washed with S- basal thoroughly. S- basal contained 1 mM sodium azide to anesthetize the worms. Worms were left for 3 hours at room temperature in 200 µg/ml of gentamicin before being washed with sterile S- basal thoroughly. 200 µl of S-basal + 0.1% Triton-X was added to the tubes. Worms were lysed mechanically using silica glass beads. The lysate was diluted in 1x PBS for serial dilutions. LB agar plates were plated with appropriate dilutions. Plates were incubated overnight at temperature 37^{o}C. This was followed by counting colonies and the number of bacteria per nematode.

2.17.4 Liquid killing assay

Overnight cultures of *S. flexneri* were grown at 30°C and diluted at 1:50 in pre-warmed L.B. media. Log phase growth ($OD_{600)}$ was achieved at 37°C.*C. elegans*, L4 synchronized population of worms was collected from *E. coli* OP50 plates. L4s were treated with 200µg/ml of gentamycin for ~3 hours at room temperature to remove any surface bound OP50 *E. coli*. Worms were washed three times with S-basal to remove any residual antibiotic. A 24-well plate containing 100µl of the log phase bacterial culture was poured with 20-50 washed L4s. the volume of each well was maintained to 500µl with S-basal and the plate was incubated at 22°C for up to 48 hours. After every 12 hours, the interval number of live worms was counted, and percentage survival was calculated. The worms were considered dead when no pharyngeal pumping and movement on the tapping of the plate was observed. A minimum of three replicates for each tested strain was set up per trial.

2.18 Invasion assay

2.18.1 Sub-culturing of the cells

The Hela cells were offered by Dr. David Tscharke from John Curtin School of Medical Research, ANU. To grow cells 25 cm^2 tissue culture flasks (Thermo Fisher) were used. To grow cells to 70% confluency the cells were incubated in a CO_2 incubator with 5% CO_2 at 37 °C in Dulbecco's Modified Eagle's Medium (DMEM, Thermo Fisher) supplemented with 10% (v/v) Fetal Bovine Serum (FBS, Thermo Fisher), 1 x non-essential Amino Acids (Thermo Fisher) and 2 mM Glutamine (Thermo Fisher). Before subculturing Hela cells, 1 x PBS, the tissue culture media, and 1 x Trypsin were pre-warmed at 37 °C.

To subculture cells, the growth media was aspirated from the T25 flask and 5-10 ml of pre-warmed PBS was added gently to the side. PBS was aspirated back and 1 ml of pre-warmed 1 ml of Trypsin was added to cover the cells. The monolayer was rocked, and the flask was incubated in a CO_2 incubator for 3-4 minutes. Flask was observed to check if the cells were detached, and 9 ml of warm media was added to inhibit trypsin action.

The concentration of cells in the suspension was estimated using a hemocytometer and 1.5 x 10^5 cells were seeded in a new 25 cm^2 tissue culture flasks. The incubation was performed in 5% CO_2 at 37°C. Cells were diluted as per the requirement and 10 ml media was added to the cells. In a fresh tissue culture plate 9 ml of the media was added and 1 ml of above cells. 2 ml of this was dispensed in each well and incubated overnight.

2.18.2 Inoculum preparation for invasion assay

Bacterial cultures to be tested were dilution streaked on LBA plates and incubated overnight at 30°C in an incubator. The next day a single colony was picked to inoculate 5-10 ml of LB and incubated at 30°C in a shaker incubator. Next day the cultures were diluted 1:100 in LB with 0.1% deoxycholate (Sigma-Aldrich) and mid-log phase (OD_{600} = 0.6-0.8) cultures were grown at 37 °C. The CFU/ml of the bacterial cultures were calculated by using the growth curve of the strain(s). In the growth curve 1 OD_{600} = 8 x 10^8 CFU/ml. Afterward, cultures were

spun down at full speed to harvest a suitable number of bacterial cells and resuspended in 1 ml of 1 x PBS to wash the cells twice. Finally, in the pellet 250 µl of 1 x, PBS was added and the number of bacterial cultures corresponding to 2 x 10^9 CFU/ml was calculated. From this culture, 0.1 ml was used to inoculate a 6-well tissue culture plate.

2.18.3 Invasion assay

A confluent monolayer of Hela cells from the T25 flasks was divided 1:12 and 2 ml of the cells were poured into all wells of 6-well tissue culture plate in duplicates. To obtain 60% confluency the plates were incubated in 5% CO_2 at 37°C. 100 µl of prepared bacterial cultures in Section 2.17.2 were then inoculated into each well and the plate was spun for 10 minutes at 1000 rpm in a benchtop centrifuge with a plate carrier rotor. The plate was then incubated at 37°C/5% CO_2 for 30-40 min. After that media was aspirated and wells were washed 3 times with 2 ml 1X PBS and media containing 1x gentamycin and incubated at 37°C/5% CO_2 for 60-90 min. Media was then aspirated and wells were rinsed 2 x 1ml with PBS and stained with 500 µl Giemsa stain (Sigma-Aldrich) for 30 sec. The stain was removed, and cells were rinsed 1-2 times with distilled water, and the plate was left inverted on the bench at room temperature. Cells were examined at 100X (oil immersion) and 300 cells/well were counted. A cell was scored as invaded if it had 3 or more bacteria and percent invasion and the average for the three wells per strain. The statistical analysis was performed on R studio in consultation with the Statistical consulting unit, ANU.

2.19 Bioinformatics analysis

2.19.1 Prediction of membrane protein topology

Topology prediction program TOPCON (http://topcons.net/) was used to examine the amino acid sequence of OacB for the presence of hydrophobic regions. TOPCON uses five subset methods OCTOPUS, Philius, PolyPhobius, SCAMPI, and SPOCTOPUS to predict the presence of transmembrane and non-transmembrane regions in a protein (Chapter 3 and 4).

Clustal Omega and BioEdit Sequence Alignment Editor were used for the nucleotide/protein level alignment. PHASTER was used for identification of bacteriophage sequences. NCBI nucleotide/protein blast and IgV were also used.

2.19.2 Genome assembly

The raw fast files of *S. flexneri* strain SFL1683 (Sf101 lysogen) from Oxford Nanopore technologies MinION were base-called and demultiplexed using Guppy (Oxford Nanopore). Reads of high length and quality were filtered for assembly. MiSeq short reads and MinION long reads were used to carry out hybrid assembly using Unicycler Version 3 [177] and further improved by using Pilon (Total 12,110 improvements) [178]. Assembly of whole genome sequence (WGS) was kindly performed by Dr. Ashley Jones, Research school of Biology (RSB).

2.19.3 Sequence analysis and annotation

Bacterial genomes or nodes were annotated using Rapid Annotation using Subsystem Technology (RAST) version 2.0 [179]. BLASTp searches were then used for manual curation. The annotated sequences were visualized on Snap Gene Viewer (Version 3.3.1) and image files were created. Artemis Comparison tool was also used to generate image files. To search for bacteriophage-related genes, PHASTER (PHAge Search Tool Enhanced Release) database [180]. IS elements were identified using ISsaga [181].

Chapter 3: Optimization of OacB expression

3.1 Introduction

Membrane topology of polytopic membrane proteins is defined as the identification of hydrophobic transmembrane (TM) segments and their orientation in the membrane [182]. The hydrophobic segments are connected across the membrane by hydrophilic loops. This unique structure allows membrane proteins to perform vital cellular functions such as by transmitting signals across the biological membranes or by moving ions or small/macro molecules across the biological membranes. Topological studies of plasma membrane proteins may help elucidate their mechanisms of action by identifying hydrophobic TM regions and hydrophilic (periplasmic or cytoplasmic) loops of the protein which interact with the substrate molecules. Moreover, solved secondary structures also help in comparative studies with other characterized proteins to gain a better understanding of their function. Methods available to determine the topology of membrane proteins like the Pho/Lac reporter system and substituted cysteine accessibility method (SCAM) are of great help in this regard.

Serotype conversion is an important process whereby bacteria such as *S. flexneri* change their cell surface antigens. One mediator of serotype conversion is the *O*-acetylation of sugars that are incorporated into the LPS O-antigen which requires the activity of enzymes called *O*-acetyltransferases. Five *O*-acetyltransferase genes (*oac, oacB, oac1b, oacC,* and *oacD*) have been isolated from different serotypes of *S. flexneri*. The only *O*-acetyltransferase whose secondary structure has been defined is Oac of serotype 6 [168], which is responsible for 2 O-acetylation of rhamnose I. OacB, a homologue of Oac, has recently been isolated from 1c strain of *S. flexneri* (SFL1683). OacB catalyses the *O*-acetylation of rhamnose III of the O-antigen unit and in doing so alters the antigenic signature of the parent *S. flexneri* strain such that it undergoes serotype conversion [105]. Importantly, the topological features of OacB have not yet been elucidated.

To address this knowledge gap, this study initially aimed to predict and model the topological features of OacB with the help of computer-based topology programs. After this initial *in silico*

analysis, our long-term plan was to use SCAM to experimentally investigate the secondary structure of OacB. SCAM was chosen over other methods as it allows analysis of protein structure in its native state. SCAM involves the replacement of amino acids of interest with cysteine residues. The location of these cysteine residues is detected with SCAM-specific reagents. As has been mentioned before, SCAM requires the overexpression of OacB in *E. coli* to produce a detectable amount of protein. Hence, the *oacB* gene was cloned into two expression vectors (pBAD/*Myc*-HisA and pFLAG-CTC) to create plasmids pNV2132 and 2140 respectively. After confirmation of functionality of OacB in both systems using slide agglutination test, an optimisation process was undertaken which ultimately achieved the overexpression of OacB. Furthermore, a cysteine-less OacB mutant and single cysteine variants of OacB were created using site-directed mutagenesis. While the lengthy optimisation process prevented me from undertaking SCAM experiments, the successful expression of these OacB proteins will enable future studies to do so to validate the *in-silico* predictions made in this project.

The main aims of this chapter are as follows:

1. Generation of computer predicted model of OacB
2. Cloning of *oacB* gene into two expression vectors
3. Optimization of OacB overexpression in two vectors.

3.2 Results

3.2.1 Generation of a consensus computer predicted topology model of OacB

The topology of membrane proteins refers to the pattern of polypeptide chains that intertwine back and forth through the membrane [182]. The amino acid sequence of any protein can theoretically be used to determine the topology of the protein. Topology models are the visual descriptions of the approximate locations of TM helices, their orientation in the membrane, and help identify structural elements. Many computer-based prediction programs utilize the ‘positive –inside –rule ‘, according to this rule, most amino acids facing the cytoplasmic

compartment are positively charged, while, at the extra-cytoplasmic leaflet, the positively charged residues are comparatively less [182]. To predict the topology model of OacB, TOPCONS webserver was used. The amino acid sequence of OacB was used as an input sequence and the output was representing the hydrophobic and hydrophilic regions of the protein in the form of a graph [183]. TOPCON uses five sub-methods OCTOPUS, Philius, Polyphobius, SCAMI, and SPOCTOPUS to produce a consensus model. The working principle of all is based on the Hidden-Markov model (Table 3.1). According to the consensus model, OacB has 11 TM domains and 10 loops, and N-terminus located in the periplasm whereas, C-terminus in the cytoplasm (Figure 3.1B).

Table 3.1 Summary of the topology prediction of OacB, generated by the computer prediction programs.

Program	Number of transmembrane helices predicted	Location of N-terminus
OCTOPUS	**11**	**Periplasmic**
Philius	**11**	**Periplasmic**
Polyphobius	**11**	**Periplasmic**
SCAMI	**11**	**Periplasmic**
SPOCTOPUS	**11**	**Periplasmic**

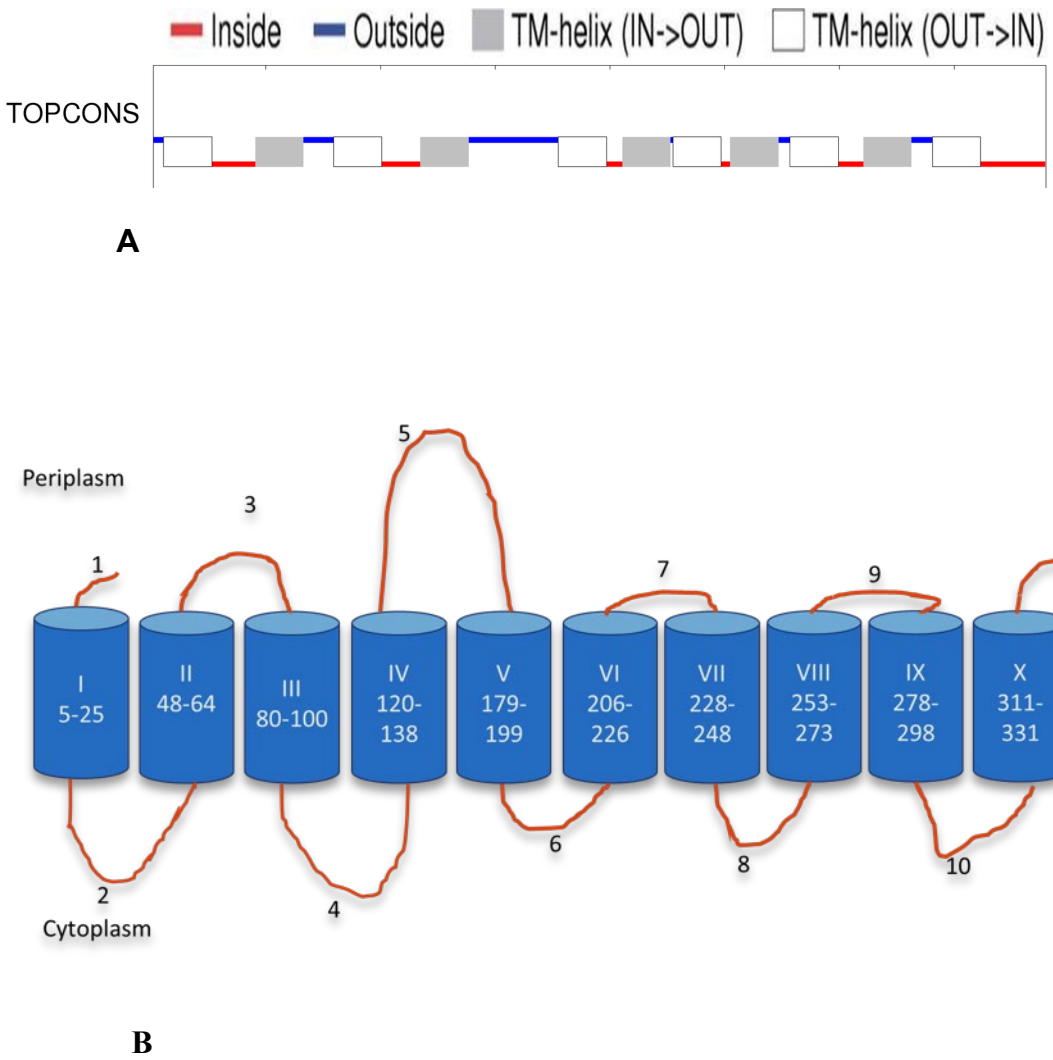


Figure 3.1: The consensus topology model of OacB is generated by the prediction programs (A) Consensus topology of OacB. The red lines indicate 'inside', blue lines indicate 'outside'; grey boxes indicate 'TM-helix inside-out' and empty boxes indicate 'TM-helix outside-in'. **(B)** OacB is predicted to have 11 TM domains, 10 loops with periplasmic C- and cytoplasmic N- termini. Roman numerals represent the TM domains whereas Arabic numerals represent the loops in periplasm and cytoplasm.

3.2.2 Cloning of oacB into pBAD/Myc-His A and FLAG-CTC vector

3.2.2.1 Amplification of oacB and isolation of pBAD/Myc- His A and FLAG-CTC

Colony PCR was carried out to amplify insert *oacB* from SFL 2535 using specific primers. After PCR, the PCR product was double digested (XhoI and EcoRI) and column purified. The vector pBAD/*Myc*-HisA was isolated from B2372. The vector was digested with XhoI and EcoRI restriction enzymes. Similarly, for cloning into pFLAG-CTC vector, the amplified product was double digested with XhoI and BglII and column purified. The vector isolated from strain B2381 was restriction digested with XhoI and BglII and prepared for ligation

reaction.

3.2.2.2 Ligation and transformation

The ligation reaction was set up for both the vectors to create pNV2111 (pBAD/*Myc*-His A containing *oacB* gene) and pNV2110 (pFLAG-CTC containing *oacB* gene). The potential clones were confirmed for the presence of insert (*oacB*) by restriction enzyme digest and Sanger sequencing (Figure 3.2A, B and 3.3A, B). The plasmids containing *oacB* gene were named B2574 (pNV2110) and B2575 (pNV2111).

To confirm the functionality of OacB in plasmids pNV2110 and pNV2111 both plasmids were transformed into *S. flexneri* serotype 1c strain SFL1691, lacking *O*-acetylation on rhamnose III (Rha III) of O-antigen due to the absence of *oacB* gene.

3.2.2.3 Cloning of Erythromycin gene in pNV2110 and pNV2111

To carry out agglutination assays the recombinant plasmids were transformed into a 1c strain lacking *oacB* gene. SFL1691 was selected for this purpose as it was PCR negative for *oacB* gene. However, the selected strain was found to be ampicillin-resistant when grown on LB plates containing ampicillin. Therefore, the erythromycin resistance (Em^R) gene was cloned into plasmids pNV2110 and pNV2111 to facilitate the selection of clones with the erythromycin gene. The erythromycin gene was amplified from pNV2074 (B2536) using specific primers. In the case of cloning Em^R into pNV2110, the PCR product was digested with SphI and for pNV2111 with NcoI. The digested products were column purified. Vector plasmids pNV2110 and pNV2111 were also digested with SphI and NcoI restriction enzymes respectively and dephosphorylated using calf intestinal alkaline phosphatase (CIAP). Both the vectors were loaded on the agarose gel and their bands were excised and purified. To create pNV2132 in (pBAD) and pNV2140 (pFLAG-CTC) ligation reactions were set up in either 1:6 or 1:3 vector to insert molar ratio (Section 2.6.3). Sanger sequencing and restriction digest confirmed the presence of insert (Figure 3.4A, B and 3.5A, B), and the plasmids containing erythromycin gene were named as B2596 and B2605 and used to make glycerol stocks. The plasmids (pNV2132 and pNV2140) were also transformed into *Shigella flexneri*

SFL1691 for functionality test and to produce glycerol stocks with unique strain names SFL2572 and SFL2579, respectively.

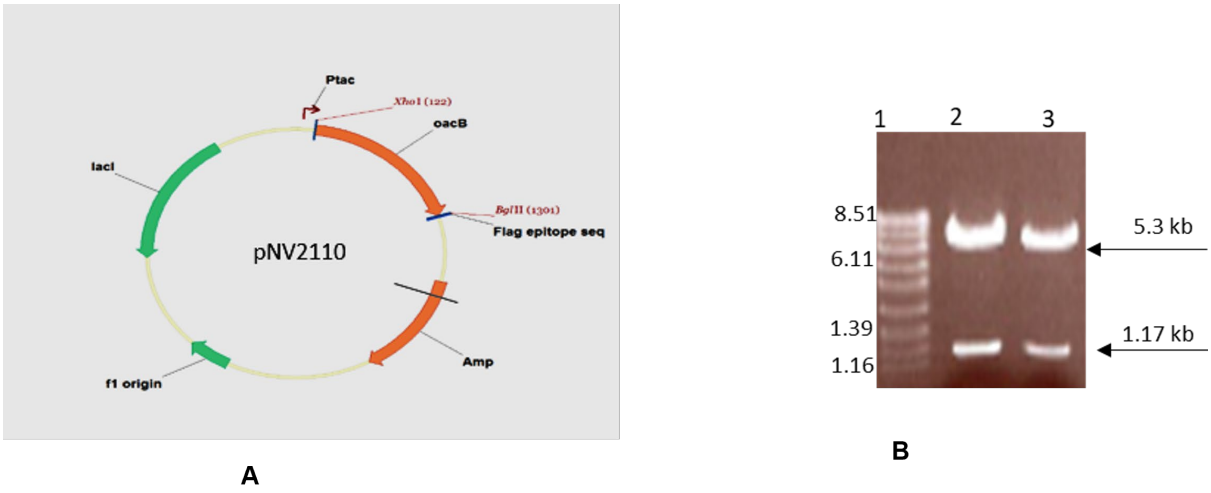


Figure 3.2: Cloning of *oacB* gene into pFLAG-CTC vector.
(A) Vector map of pNV2110 showing *oacB* gene cloned into pFLAG-CTC vector using *XhoI* and *BglII* restriction sites. **(B)** Screening of *oacB* gene in pNV2110 with XhoI and BglII.Alkaline lysis of the potential clones was performed to isolate plasmid DNA and double digested DNA was run on 0.7% agarose gel. Lane 1= Spp1 ladder ; lane2-3 = double digested pNV2110 . Expected sizes of *oacB* (1.17 kb) and pFLAG-CTC vector (5.3 kb) are indicated on the gel.

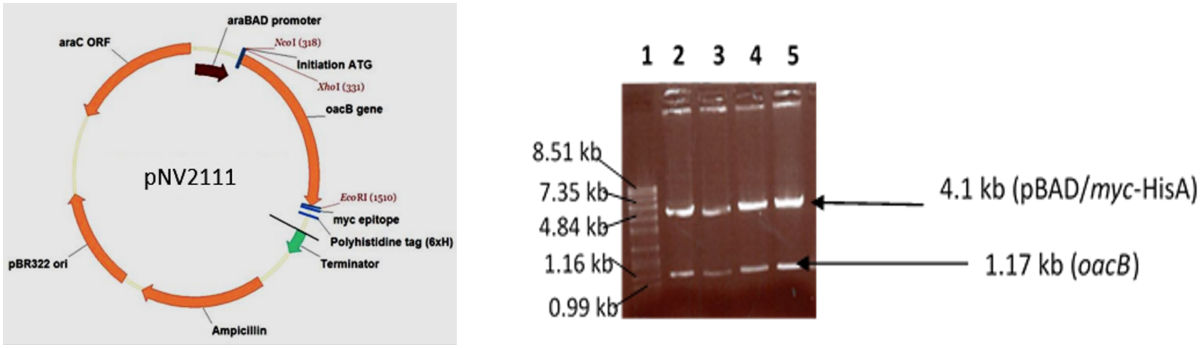


Figure 3.3: Cloninig of *oacB* gene in pNV2111.
(A) Vector map of pNV2111 showing *oacB* gene cloned into pBAD/*Myc*-HisA using *XhoI* and *EcoRI* restriction sites. **(B)** Screening of *oacB* gene in pNV2111 with XhoI and EcoRI.Alkaline lysis of the potential clones was performed to islolate plasmid DNA and double digested DNA was run on 0.7% agarose gel to isolate the inserted *oacB* . Lane 1= SppI ladder ; lane 2-5 = double digested . Expected sizes of *oacB* (1.17 kb) and pBAD/*Myc*-HisA (4.1 kb) are indicated on the gel.

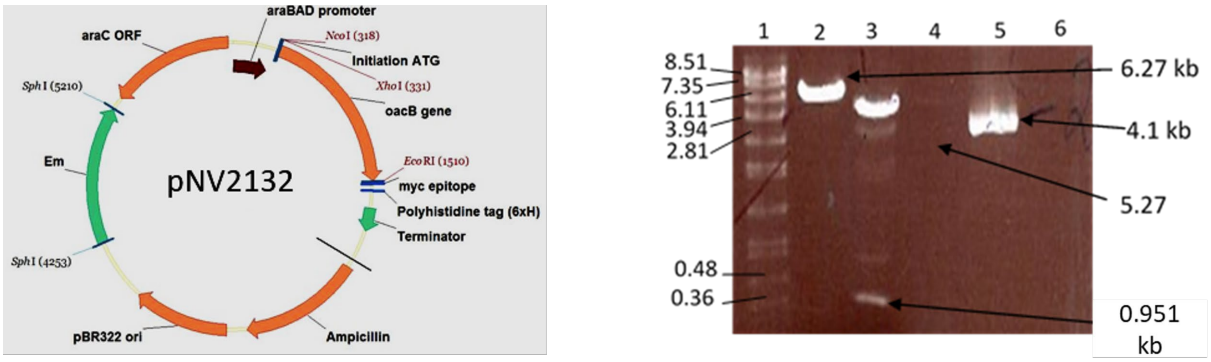


Figure 3.4: Cloning of erythromycin gene in pNV2111.
(A) pNV2132 map showing cloned erythromycin resistance gene in pNV2111 using SphI restriction site. **(B)** The screening of transformants containing erythromycin resistance gene cloned into pNV2111**.** Following the transformation of ligated plasmid 2132 and growth on selective media containing erythromycin (250 µg/ml), the plasmid DNA was isolated by alkaline lysis and restriction digested with SphI. The expected size erythromycin gene band of ~0.951 kilo base (kb) is indicated on 0.7% agarose gel. Lane 1 = SppI ladder; lane 2= Undigested pNV2111; lane 3= SphI digested pNV2132; lanes 4 and 6 = empty; lane = 5 uncut pBAD/*Myc*-HisA vector. 0.7% agarose gel was used to run the samples.

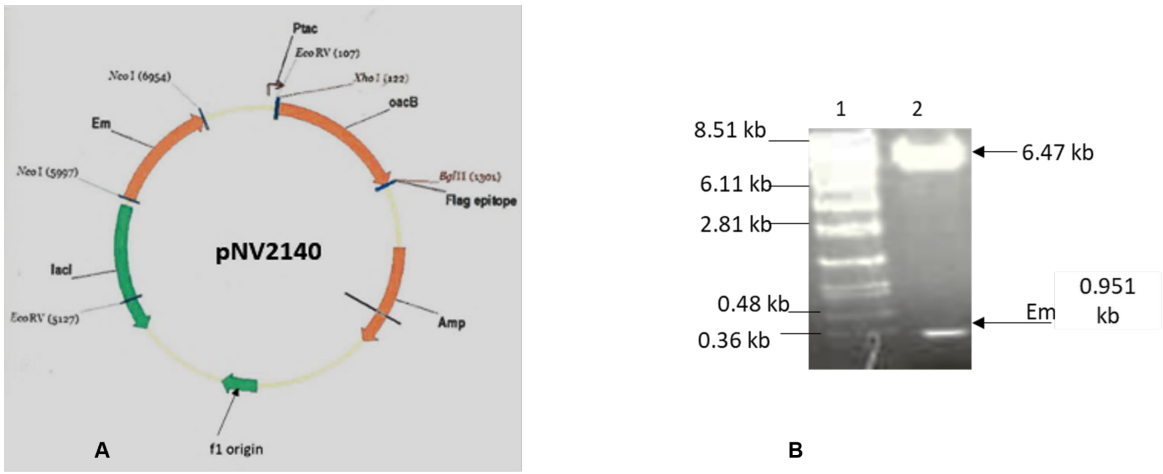
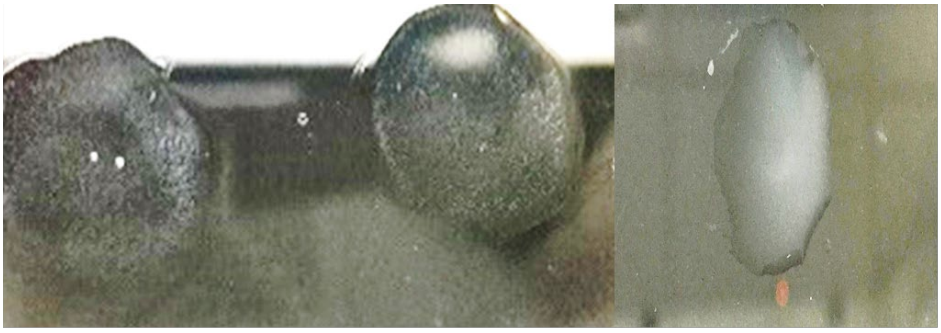


Figure 3.5: Cloning of erythromycin gene into pNV2110.
(A) Vector map of pNV2140 showing cloned erythromycin gene at NcoI site. (B) The screening of the transformants containing erythromycin resistance gene cloned into pNV2110. Following the transformation of ligated plasmid and growth on selective media containing antibiotics, the plasmid DNA was isolated by alkaline lysis and digested with SphI. The expected size of the erythromycin gene of ~0.951 kilo base (kb) is indicated on 0.7% agarose gel. Lane 1 = SppI ladder; lane 2 = pNV2140 digested with *NcoI*.

3.3 Functionality testing of OacB

The transformed strains SFL2572 containing pNV2132 and SFL2579 containing pNV2140 were used to carry out slide agglutination tests using antiserum specific to 3/4 *O*-acetylation modification of O-antigen for OacB functionality. If the OacB is functional, the *O*-acetyl residues (transferred via a functional OacB) modifying the O-antigen subunits interact with the antibodies and cause the bacterial cells to clump and produce a noticeable clumping reaction as observed in the positive control. Clumping recorded within 60 seconds of the addition of antiserum to the cells was considered as positive agglutination. No clumping after 60 seconds was considered a negative test. Results revealed the presence of a functional form of OacB in both the vectors (Figure 3.6).



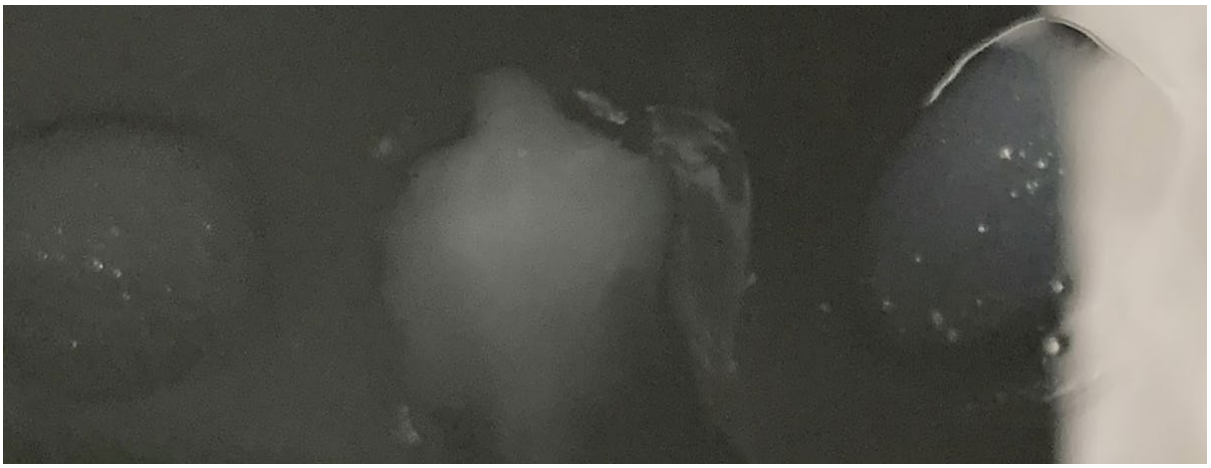
a) Test strain b) Positive control c) Negative control

Figure 3.6: Functionality of OacB. A representative image is shown for the functionality of OacB in pNV2312 and pNV2140.
To test the functionality of OacB, the plasmids were transformed into 1c strain SFL1691 (which lacks *oacB)* to produce strains SFL2572 and SFL2579 to perform agglutination assay using ¾ O-acetyl specific antiserum. A single isolated colony of the test strains was mixed with the antiserum and the glass slide was rotated back and forth for 60 seconds. Clumping/agglutination visible within 60 seconds was considered a positive test and no agglutination was indicated as a negative test. Results **(a)** Test strain clear agglutination indicates the functional form of OacB. **(b)** Positive control (SFL1683) with visible clumping/agglutination within 60 seconds. **c)** Negative control (SFL1691) no agglutination even after 2 minutes of adding antiserum. [Negative control was tested on a separate slide but at the same time and its imaged is attached together].

3.4 Creation of Cysteine-less mutant using site-directed mutagenesis.

For SCAM to be used there should only be one cysteine present in the OacB at any given time, as it mainly makes use of the ability of cysteine to form disulfide bonds with the sulfhydryl reagents being used in the SCAM. Amino acid sequence analysis of OacB revealed five native cysteines. The location of all the cysteines was determined using the predicted topology model and found that all cysteines were present in TM domains of OacB. Using Site-directed mutagenesis (SDM), all five native cysteine residues (C129, C284 C326, C357, and C295) present in the OacB were mutated to alanine to generate a functional cysteine-less form of this protein. Alanine was selected to be replaced with cysteine because of its small structure and non-polarity. Native cysteine residues were sequentially removed using plasmid pNV2132 as a template for the SDM PCR reaction. DpnI digestion of the amplified product was carried out following each mutation. All the transformants were screened and sequenced using pBAD/*Myc*-HisA sequencing primers using Sanger

sequencing to confirm the mutation. For the confirmation of OacB functionality, slide agglutination was performed as described in (Section 2.10.1). Functionality tests were done to ensure the native conformation of OacB even after mutation of all the cysteines in pNV2139 (Figure 3.7). Results revealed that strain (SFL2578) harbouring cysteine-less OacB mutant (pNV2139) demonstrated a strong agglutination reaction indicating fully functional OacB (Figure 3.7)



a) **b)** **c)**

Figure 3.7: Functionality of Cysteine less OacB in pNV2139.
To test the functionality of OacB in pNV2139, the plasmid was transformed into 1c strain SFL1691 (which lacks *oacB)* to produce strain SFL2572 to perform agglutination assay using ¾ O-acetyl specific antiserum. A single isolated colony of the test strains was mixed with the antiserum and the glass slide was rotated back and forth for 60 seconds. Clumping/agglutination visible within 60 seconds was considered as a positive test and no agglutination was indicated as a negative test. Results **(a)** Negative control (SFL1691) no agglutination even after 2 minutes of adding antiserum. **(b)** Test strain clear agglutination indicates the functional form of OacB. **(c)** Positive control (SFL1683) with visible clumping/agglutination within 60 seconds.

3.5 Creation of Single-cysteine variant of OacB using site-directed mutagenesis.

Site-directed mutagenesis was performed to introduce cysteine residues at pre-determined locations in OacB using pNV2139. The residues were selected based on their proximity to the critical residues in OacB . Hence, isoleucine 34 (I34) in cytoplasmic loop 2; alanine 73 (A73) in periplasmic loop 3; valine 114 (V114) in cytoplasmic loop 4; alanine 162 (A162) in periplasmic loop 5; isoleucine 202 (I202) in loop cytoplasmic loop 6; isoleucine 251 (I251) in cytoplasmic loop 8 and asparginine 335 (Asp335) in cytoplasmic loop 10 were selected to be mutated to cysteine. Mutations were confirmed via Sanger sequencing and plasmids created were later transformed into TOP10 *E. coli* cells for SCAM analysis, and into *S. flexneri* 1c strain SFL1691 cells for functionality assessment. The functionality of single cysteine OacB variants was determined using the slide agglutination test (same as above). All the single-cys variants demonstrated strong agglutination reactions except A73 and ISO34 which demonstrated a reduced level of clumping when compared with the positive control (SFL2578).

Table 3.2 Showing Agglutination reaction of single cysteine variants of OacB and their location in OacB.

Residue location	location	Agglutination reaction
isoleucine 34	cytoplasmic loop 2	+
alanine 73	periplasmic loop 3	+
valine 114	cytoplasmic loop 4	+++
alanine 162	periplasmic loop 5	+++
isoleucine 202	loop cytoplasmic loop 6	+++
isoleucine 251	loop cytoplasmic loop 8	+++
asparagine 335	loop cytoplasmic loop 10	+++

low agglutination reaction (+); Medium (++); strong agglutination (+++) reaction and no agglutination (-).

3.6 Optimization of OacB expression

As has been previously noted, SCAM analysis often requires heterologous overexpression to obtain enough protein for detection. Optimizing OacB overexpression and purification to maximize the amount of protein produced was an important prerequisite before SCAM analysis could be considered.

Overexpression of recombinant membrane proteins in correct folded form is one of the challenging problems in molecular biology and stems from a variety of factors. Theoretically, four simple steps are needed to achieve the expression of recombinant proteins: 1) cloning the gene of interest into an expression vector; 2) transformation of the plasmid into a suitable host; 3) induction of expression; and 4) purification of the protein. However, practically many factors can go wrong and, even with selection of the right combination of plasmid and expression host, it is often difficult to obtain high and soluble amounts of transmembrane proteins [184, 185].

In this section, the series of experiments performed to optimize the overexpression of OacB, and the cysteine-less or single cysteine OacB mutants are presented. Several parameters were tested to determine the optimal conditions for OacB overexpression, including, the concentration of inducer(s), the temperature of incubation, the optical density of the cultures at the time of induction, and the duration of incubation post-induction. Moreover, after achieving OacB expression, methods to prepare proteins sample for analysis on SDS-PAGE gel were also modified.

3.6.1 Optimization of expression of OacB in pFLAG-CTC vector

Plasmids pNV2110 (pFLAG-CTC vector carrying *oacB* gene) and pNV2140 (pNV2110 carrying Em^R gene) created in Section 3.2.2.3 were transformed into TOP10 *E. coli* cells to generate strains B2609 and B2608, respectively. B2539 [(pNV2075) Opt in pFLAG-CTC vector was used as a positive control for Western blot in most of the experiments. O-antigen phosphoethanolamine transferase (Opt) is a membrane protein and encoded by plasmid-

encoded gene *opt* of *S. flexneri*. Opt has four TM domains and it brings about modification of O-antigen by adding phosphoethanolamine to either Rha II or III or both of O-antigen and is responsible for producing variant serotypes Xv, 4av, and Yv from *S. flexneri* serotypes X, 4a, and Y, respectively.

3.6.1.2 Optimization of expression of OacB in the whole cell lysate of B2608 (TOP10 E. coli strain transformed with pNV2140).

B2608 (pNV2140 containing *oacB* and Em^R genes in pFLAG-CTC vector) cultures were grown in LB_{Amp} to an optical density (OD_{600}) of 0.3 and were induced with 0.5 mM and 1 mM IPTG. Whole-cell lysates were prepared after 4 hours of induction as described in Section 2.11.1. Samples prepared were mixed with 2x loading buffer (Appendix A) and heated at 100°C for 10 minutes and subjected to SDS-PAGE gel. B2539 having pNV2075 (*opt* gene in pFLAG-CTC vector) was used as a positive control. Western blot results suggested that OacB expression level was undetectable in either of the IPTG concentrations tested (Figure 3.8). A faint band of ~53 kDa was detected in the lane of positive control (B2539), and no signals were detected for the negative control (B2581) as expected. These results suggested that in whole-cell lysate samples, OacB is not expressed at detectable levels and that altering the IPTG concentration does not change this.

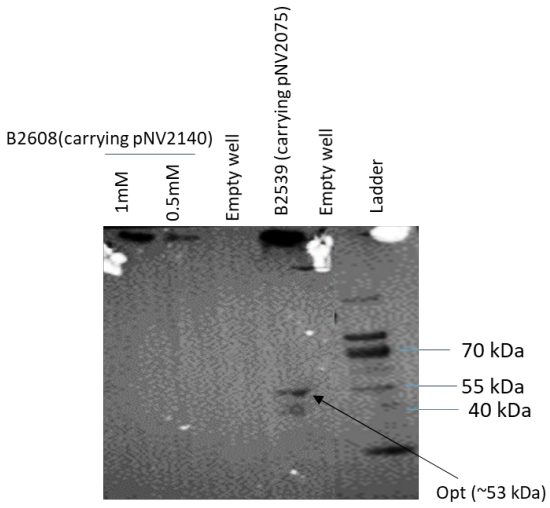


Figure 3.8: Optimization of OacB expression in B2608 (pFLAG-CTC vector containing OacB). Western Blot using anti-FLAG monoclonal antibodies conjugated to horseradish peroxidase. B2608 Cells were grown in $LB_{(Amp100)}$ at 37 ºC and induced at an optical density (OD_{600}) of 0.3 with either 0.5 mM or 1mM IPTG. Cultures were grown at 37 °C and the $OD_{(600)}$ was recorded after 4 hours of incubation and samples were collected to prepare whole-cell lysate samples in 2x loading buffer. Samples were subjected to SDS-PAGE gel. B2539 (containing pNV2075) was used as a positive control. Phosphoethanolamine (Opt) running size is 53 kDa, Empty vector pFLAG-CTC (strain B2581) was used as a negative control. No band was observed at the expected running size of OacB i.e., 44 kDa in the lanes of B2608 at the IPTG concentrations used. Positive control B2539 was detected running in the right position. The pre-stained molecular weight marker is shown on the right side of the gel**.** Anti-FLAG monoclonal antibodies conjugated to horseradish peroxidase in 1:6000 dilution was used.

3.6.1.3 Optimization of OacB in small scale membrane protein samples of B2608 (pFLAG-CTC vector carrying oacB and Em^R gene) and FLAG-affinity purification of OacB

Since OacB is an integral membrane protein, we thought that preparing membrane fractions of the cells rather than whole-cell lysates may help achieve higher concentration of OacB. Moreover, we reasoned that additional purification might assist detection of OacB. To test this, membrane fractions were prepared on small scale and immunoprecipitation of OacB using Anti-$FLAG^R$ M2 Affinity Gel (Sigma-Aldrich) was performed.

To do this, B2608 culture in 10 ml LB_{Amp} at 37 °C was induced at an OD_{600} of 0.4 with 0.5 mM of IPTG. After four hours, membrane protein samples were prepared from the induced culture (described in Section 2.11.2). An aliquot of 100 µl membrane fractions was mixed with washed

anti-FLAG^{R} M2 affinity gel. The gel contained purified IgG_1 which specifically recognises and binds to the FLAG fusion peptide attached to OacB. After binding to the gel and washes, elution of OacB was performed by the addition of a 2x loading buffer. OacB was released by heating the sample to 100°C for 10 minutes and loaded onto two SDS-PAGE gels for further analysis. Unpurified membrane protein samples were also prepared same way in 2x loading buffer and loaded onto the gel along with the purified samples. Following electrophoresis, gel was subjected to western blot analysis using 1:8000 anti-FLAG monoclonal antibodies.

No appropriately sized band of OacB (~44 kDa) was seen in the western blot for the purified samples (Figure 3.9). However, there was a high molecular weight diffused band in the lane of the unpurified sample of OacB, which could possibly suggest some expression of OacB. Similarly, a high molecular weight band was also present in the lane of the unpurified Opt sample, and the right size Opt band (~55 kDa) was also detected in the lane of the purified Opt (positive control) sample (Figure 3.9). No bands were observed in both the lanes of negative control (B2381) as expected (Figure 3.9). The presence of the high molecular weight diffused bands in the lanes of OacB and Opt could indicate some expression of both proteins. However, it might be the result of nonspecific signals produced due to the dilution of anti-FLAG antibodies. It is also possible that the sample prepared for SDS-PAGE had some problems. I therefore decided to investigate these possibilities in the next experiment.

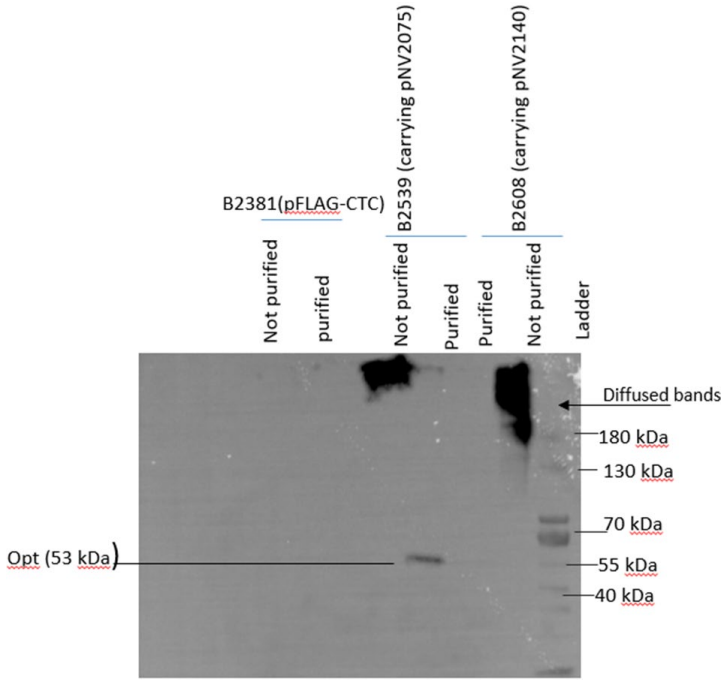


Figure 3.9: Optimization of OacB in small scale membrane protein samples of B2608 (pFLAG-CTC vector carrying oacB and Em^R gene) and FLAG-affinity purification of OacB.
Western blot using anti-FLAG monoclonal antibodies conjugated to horseradish peroxidase (1/8000) was carried out on the purified and unpurified protein samples obtained from B2608 and B2539 (positive control) when induced with 0.5 mM IPTG. Protein samples were purified using Anti-FLAGR M2 Affinity Gel (Sigma-Aldrich). The positive control running size is approximately ~53 kDa as shown in the figure. No right size band was observed for OacB (OacB expected running size is 44 kDa). Unpurified samples were also run on the gel to see the difference in the running pattern. The molecular weight marker is shown at the far right/left side in both images of the blot.

3.6.1.4 Effect of β mercaptoethanol on sample preparation

In the previous experiment, the presence of the high molecular weight diffused bands in the lanes of OacB and Opt could indicate some expression of both proteins; however, it was also possible that the sample prepared for SDS-PAGE had some problems. According to the M2 affinity gel protocol, it is suggested that the use of reducing agents may cause dissociation of heavy and light chains of the immobilized M2 antibodies. Hence in this experiment, the effect of the presence and absence of BME samples in a 2x sample buffer was observed. To do this, cultures of B2608 were prepared as described in section 3.6.1.3 and membrane proteins were prepared and purified using affinity resin.

In 2x loading buffer a 10 µl aliquot of membrane protein fraction was heated at 100°C for 10 minutes in the presence of BME and another 10 µl aliquot was heated for 10 minutes at 100°C in the absence of BME. Similarly, positive (B2539) and negative controls (B2381) were prepared. The samples were cooled and loaded onto the pre-cast gel. Western blot using anti-FLAG antibodies in 1:10 000 dilution showed aggregates of OacB in the gel and not much difference was observed in samples with or without BME. Fewer signals were observed in western blot in the lanes of pNV2140, and pNV2075 samples. Positive control bands of 53 kDa were observed for Opt in both the treatments (Figure 3.10). No bands were detected in the lanes of the negative control i.e., B2381.

It was concluded that the presence or absence of BME did not improve sample quality. The presence of the positive control Opt indicated the concentration of antibodies used for western blot was appropriate. The absence of appropriately sized OacB bands might be due to an undetectably small amount of OacB obtained from membrane preparations in these samples. Hence, we decided to scale up the experiment (200 mL cultures) to obtain more OacB protein.

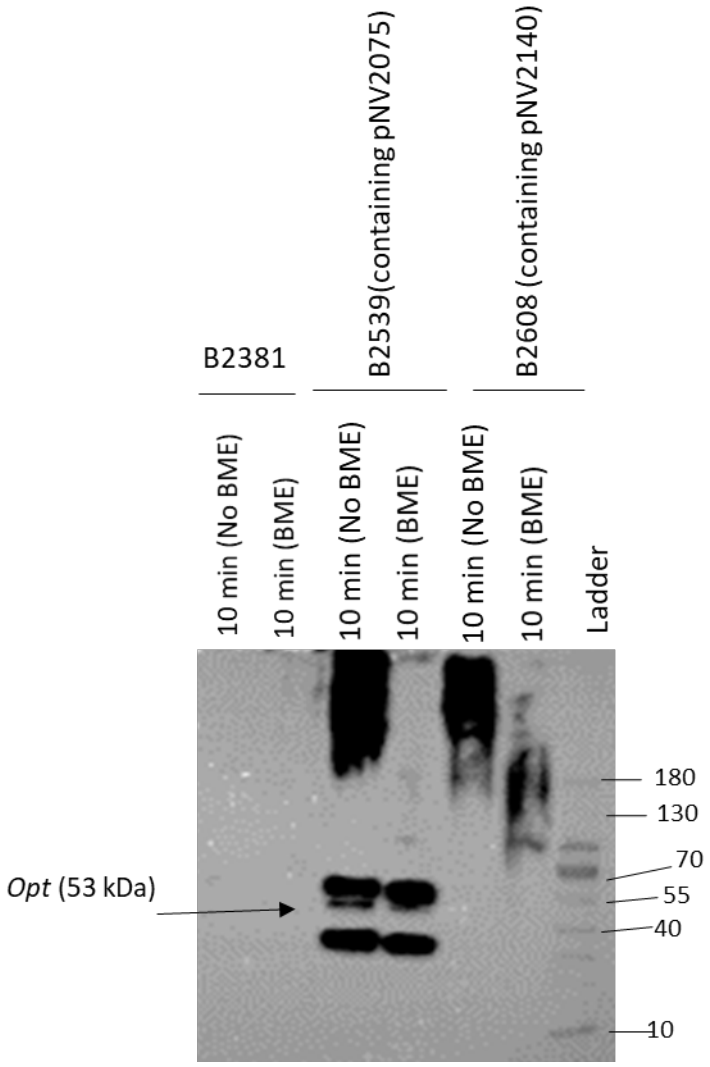


Figure 3.10: Effect of BME on purified OacB and Opt.
Western Blot using anti-FLAG monoclonal antibodies conjugated to horseradish peroxidase in 1/10,000 dilution. B2608 (pNV2140) and B2539 (pNV2075) cells were grown in $LB_{(Amp100)}$ at 37 ºC and induced at an optical density (OD_{600}) of 0.4 with 0.5 mM IPTG. Cultures were grown at 37 °C and the $OD_{(600)}$ was recorded after 4 hours of incubation. Membrane preparation was performed followed by purification using FLAG tag. Samples were prepared for each strain by heating for 10 minutes at 100°C in either absence or presence of BME. Western blot was performed after SDS-PAGE. Phosphoethanolamine (Opt) 53 kDa, expressed in pFLAG-CTC system (pNV2075) was used as a positive control (strain B2539). Uninduced empty vector pFLAG-CTC (strain B2581) was used as a negative control. No band was observed at the expected running size of OacB i.e., 45 kDa.

3.6.1.5 Optimization of OacB expression at large scale in pNV2140 and pNV2110

In previous experiments for whole-cell lysate and small-scale membrane preparations, the cultures were grown on small scale (10 ml) (Section 2.13). However, the OacB could not be detected at the expected size in any of the experiments and it was reasoned that might be due to the low concentration of OacB in these samples. Hence, we scaled up expression to

200 ml LB_{Amp} and grew cells at 37 °C to log phase (OD_{600} of 0.6) before induction with varying concentrations of IPTG (0.25, 0.5, and 1 mM) growth for a further 4 hours. In this experiment, two plasmids pNV2110 (*oacB* in pFLAG vector) and pNV2140 (*oacB* and Em^R in pFLAG) were used to optimize OacB expression in *E.coli* strains B2609 and B2608 respectively. The purpose of using pNV2110 is to see if the erythromycin gene in pNV2140 was putting an extra metabolic burden on the cells by requiring transcription and translation of an additional gene. The membrane protein sample of B2381 (pFLAG-CTC vector only) was also prepared in the same way as other samples. B2608 and B2609 membrane protein samples were purified using M2 affinity gel (Section 2.11.4b) and were prepared in 2x loading buffer as above. Following SDS-PAGE, western blot analysis was performed using anti-FLAG monoclonal antibodies conjugated to horseradish peroxidase (1:10,000).

Based on western blot analysis, there were diffused high molecular weight bands in all lanes except the B2381 empty vector control (Figure 3.11A), suggesting the presence of FLAG-tagged protein, perhaps in aggregates. However, the band of expected (44 kDa) size for OacB was not present in the lanes of pNV2110 and pNV2140 (Figure 3.11A).

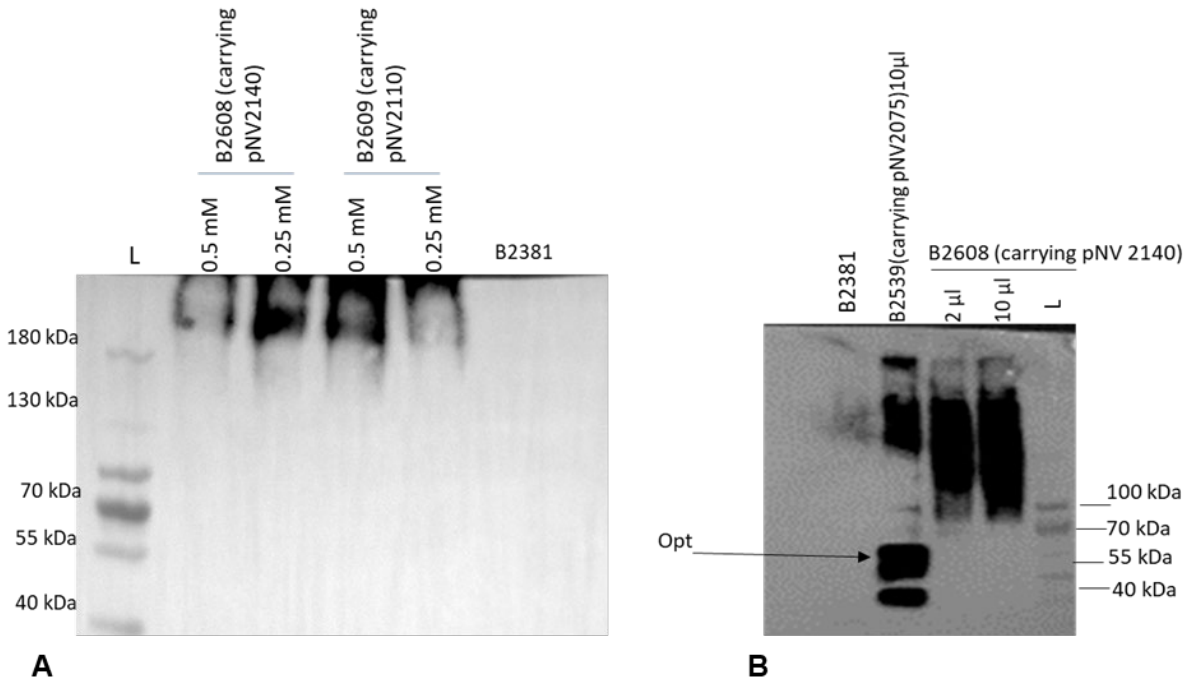


Figure 3.11: Expression of OacB in pFLAG-CTC vector (pNV2608 and pNV2140) in large-scale membrane preparation.
Proteins were collected and heated for 10 minutes at 100 °C before loading on gel**. A-B**, Western blot using anti-FLAG antibodies. Cells were grown in LB_{Amp} at 37 °C and induced at OD_{600} 0.6 with varying IPTG concentrations. Positive control pNV2075 (Opt ~53 kDa) was also expressed in pFLAG-CTC vector. **(A)** Lanes of pNV2140 and pNV2110 showing the presence of diffused bands. (**B)** Samples of pNV2140 loaded in two volumes 2 µl and10 µl in two wells. Sample of B2539 and B2381 (unpurified) were loaded 10µl volume each. Lanes of pNV2075 showing right size Opt (~53 kDa) band marker is given in the k Da. Few signals in the lane of negative control may due to overflow from the adjacent lane.

Another pre-cast gel was run (Figure 3.11B) using new aliquots of the same samples to see if loading less volume of samples might help resolve diffuse bands. This time samples of B2608 and positive and negative controls were loaded on the gel. Before loading, all samples were prepared as above in loading buffer; 2 µl and 10 µl of B2608 whereas, 10 µl of each positive and negative control were loaded. The western blot was performed as above, and blot analysis showed high molecular weight diffused signals for B2608 (pNV2140) and B2609 (pNV2110). In this experiment Opt was detected at the right size and no signals were recorded for B2381, there were some signals in the lane of the negative control indicating overflow from the adjacent lane. It was found that even at 1:10,000 dilution of antibodies the diffused bands

were present for OacB (Figure 3.11B). *Lin et al.,* (2001) had reported a correlation between protein aggregation the IPTG induction [186].

As the expression of OacB could not be achieved even after repeated attempts in pNV2110 and pNV2140 hence, it was decided to try to express OacB in pNV2132 (pBAD/*Myc*-HisA vector harboring *oacB* and erythromycin genes).

3.6.2 Optimization of OacB in pBAD/Myc His-A system

The plasmid pNV2132 was transformed into *E.coli* Top10 cells to create strain B2607. The arabinose-inducible *araBAD* promoter, P_{BAD}, upstream of the inserted *oacB* genes allows expression of the OacB by different concentrations of inducer i.e., L-arabinose. The histidine tag in the vector allows the purification of the protein. The optimum induction conditions for pNV2132 (B2607) were tested to get optimal OacB expression.

3.6.2.1 Optimization of OacB expression in pNV2132 using two concentrations of L-arabinose (0.002 and 0.2%) when induced at either OD_{600} 0.3 or 0.6

In this experiment, the cultures of B2607 carrying pNV2132 were grown in 10 ml LB_{Amp} to an OD_{600} of either 0.3 or 0.6 and induced with either 0.2% or 0.002% of L-arabinose. After induction, the cultures were grown for up to 4 hours at 37 °C. Samples were collected at different time points (0-4 hours) and whole cell lysates were prepared (Section 2.11.1). Before loading on SDS-PAGE gel the samples were mixed with 2x loading dye and boiled for 10 minutes at 100°C (as per the protocol used in the lab.).

Western blot using anti-His antibodies (1:5000) was performed, and results revealed the absence of OacB bands in the lanes of pNV2132 when induced with either 0.2% or 0.002% L-arabinose at 0.3 OD_{600} (Figure 3.12 A-B). However, high molecular weight diffused bands were observed in the lanes of pNV2132 (B2607) when induced with 0.2% or 0.002% L-arabinose at 0.6 OD_{600}, which suggested expression of OacB (Figure 3.12 A-C). Positive control protein (beta galactosidase- LacZ) was detected in the positive control lane at ~120 kDa. No signals were detected in the lane of the negative control (empty vector). The result

obtained in this experiment for OacB expression were not conclusive. It was assumed that the presence of aggregates in western blots (Figure 3.12C) might be due to the formation of inclusion bodies. Consequently, we decided to investigate whether lowering the temperature of the culture during induction.

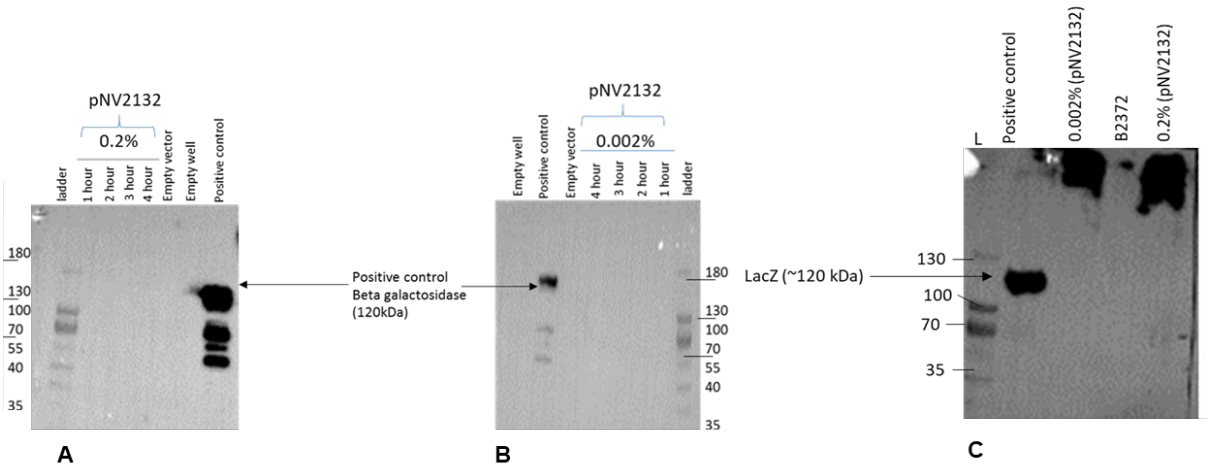


Figure 3.12: A-C. Optimization of expression of OacB in pNV2132 under various induction conditions:
A-B) Western blot using anti-His antibodies was conducted. Cells were grown in LB_{Amp} at 37° C and induced at an OD_{600} 0.3 with 0.002 and 0.2% L-arabinose.**(C)** Induction at an OD of 0.6 with either 0.002% or 0.2% L-arabinose.Samples collected at different time points post induction were subjected to SDA-PAGE gel. Positive control β - galactosidase (*LacZ*) cloned in pBAD/*Myc*-HisA (B2375) induced with 0.2% L-arabinose was detected in the positive control lane at the expected running size of ~120 kDa.OacB could not be detected at the expected running size of ~44 kDa. No signals were detected in the lane of empty vector (B2372) used as a negative control.

3.6.2.2 Optimization of OacB in pNV2132 at 18 °C and 37 °C (Time course experiment)

Given that we had so far observed only aggregates in the OacB overexpression conditions tested, we suspected that the temperature of incubation might be causing aggregation formation. Consequently, we decided to investigate whether lowering the temperature of the culture during induction would decrease aggregation of OacB.

To test this, B2607 cultures were grown as above and induced at OD_{600} 0.6 with 0.2% L-arabinose. Samples were collected every hour after induction until four hours and whole-cell lysates were prepared as above. The boiled samples with 2x loading buffer and BME were loaded onto gradient gel and western blot was performed with anti-His antibodies (1:5000). Results indicated the presence of the diffused/smeary bands in the lanes of pNV2132 indicating aggregation of the protein. The intensity of diffused bands was high in samples grown at 37 °C as compared to 18 °C (Figure 3.13), and the band of the expected size of OacB could not be detected in this experiment. The positive control was detected at the expected size (~120 kDa). Negative control was not used in this experiment. It was concluded that even at a lower temperature the high molecular bands were present and decreasing the temperature after induction does not reduce aggregate formation.

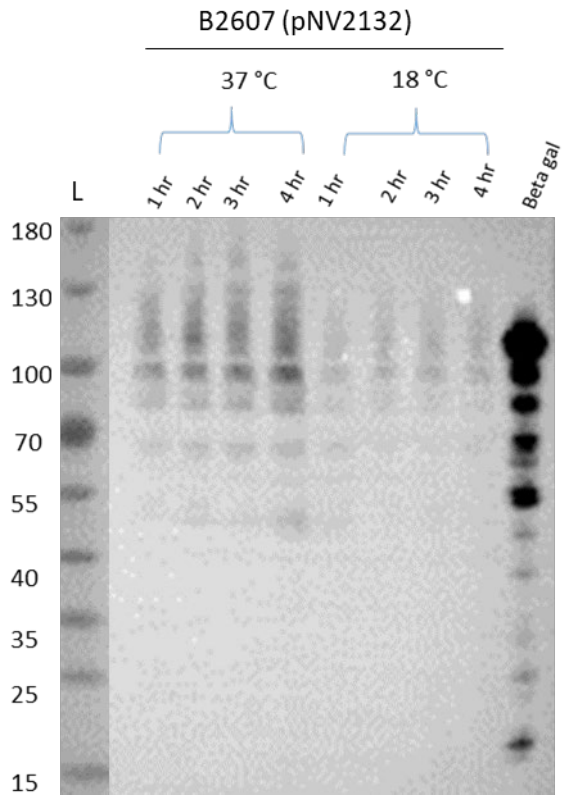


Figure 3.13: Expression of OacB in pBAD/*Myc*-HisA (pNV2132) at two incubation temperatures. Western blot using anti-His antibodies. Cells were grown in LB_{Amp100} and induced at an OD_{600} of 0.6 with 0.2% L-arabinose. Post induction samples were collected at different time points and prepared in a 2x loading buffer to load onto SDS-PAGE gel. Positive control β - galactosidase (*LacZ*) cloned in pBAD/*Myc*-HisA (B2375) induced with 0.2% L-arabinose was detected in the positive control lane at the expected running size of ~120 kDa (OacB expected running size is ~44 kDa).

3.6.2.3 Expression of OacB in pNV2132 in membrane preparations- small scale

As whole-cell lysate samples contain total proteins, we reasoned that the overabundance of other proteins in the samples might make OacB detection difficult. Hence, in this experiment I decided to prepare membrane protein samples.

To do this, Top10 *E. coli* containing pNV2132 was cultured in 10 ml LB_{Amp} and grown to OD_{600} of either 0.4 or 0.6. The cultures were then induced with 0.2% of L-arabinose at 37 °C and grown for 4 hours after induction. Membrane protein samples from these cultures were prepared and quantified using a BCA assay (Section 2.11.3). Two aliquots (25 µg of protein each) were mixed with 2x loading buffer (1:1). In tube one, DTT (100 mM) was added; in tube two BME (2.5%) was added. This was done to see if different reducing agents could help stop aggregate formation. Reducing agents break inter- and intra-disulfide bonds which are not

disrupted by SDS present in the loading buffer [187]. Before loading onto SDS-PAGE gel all three tubes were boiled for 10 minutes at 100 °C. Following electrophoresis, western blot was performed using anti-His antibodies (1:8000). Western blot results indicated the presence of aggregates in the gel in all four sample types (Figure 3.14). The PVDF membrane was Coomassie stained. No aggregates were detected in the stained membrane. The signals obtained in the lanes of B2608 samples indicated the presence of OacB; however, the bands were diffused and were not of the expected size. It was concluded that reducing agents BME and DTT could be used interchangeably but had no effect on aggregate formation. A subsequent experiment was performed on a large scale by growing cells in 200 ml $LB_{Amp.}$

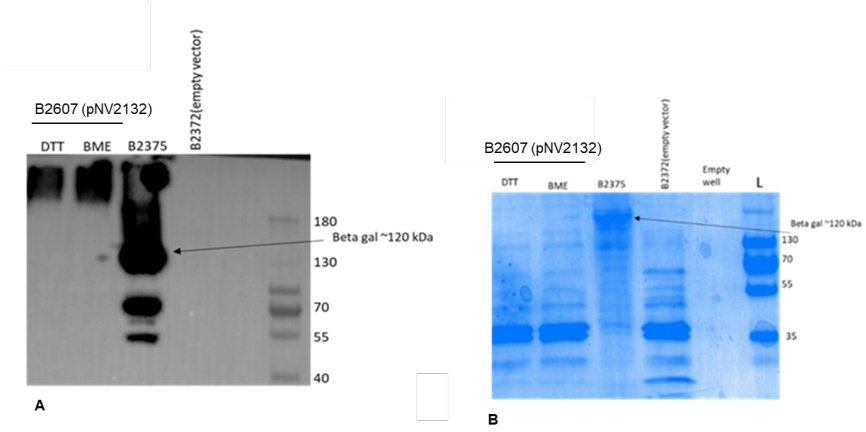
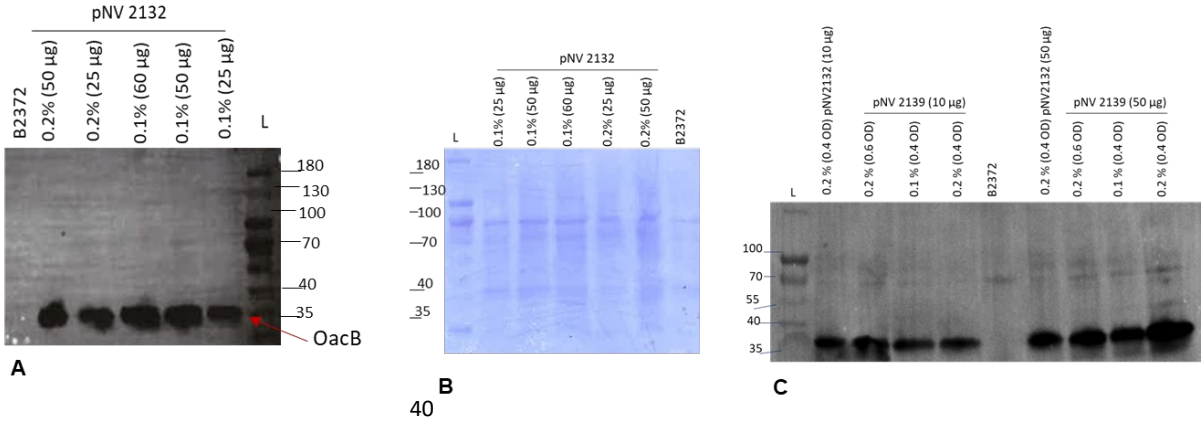


Figure 3.14: Effect of reducing agents on OacB membrane fraction.
A. Western blot using anti-His antibodies (1/8000) was performed. B2607 cultures were grown in LB_{Amp100} and induced at an OD_{600} of 0.6 with 0.2% L-arabinose. After 4 hours of incubation, membrane protein samples were prepared. Two tubes were prepared each containing 25 µg protein samples and 2x loading buffer. In two of the tubes, either DTT or BME was added. All samples were boiled before loading onto SDS-PAGE gel. Positive control β - galactosidase (*LacZ*) cloned in pBAD/*Myc*-HisA (B2375) induced with 0.2% L-arabinose was detected in the positive control lane at the expected running size of ~120 kDa (OacB expected running size is ~44 kDa). B2372 (empty vector) was used as a negative control. **B.** Coomassie-stained PVDF membrane to show equal loading of samples.

3.6.2.4 Expression of OacB in pNV2132 and pNV2139 (large scale)

B2607 cells were grown in 200ml LB_{Amp} and induced at OD_{600} of 0.4 with either 0.2% or 0.1 % of L-arabinose (as above). Membrane protein fractions were prepared, and BCA assay (Section 2.11.3) was performed and volumes corresponding to 25, 50, and 60 µg of protein were mixed with 2x loading dye having 100 mM BME. This time the samples were not boiled to observe the effect on not boiling on membrane proteins and loaded onto the precast gel. The western blot was performed using new anti-His antibodies in 1/10000 dilution (His Tag Horseradish Peroxidase-conjugated Antibody R&D systems). The blot was later stained with the Coomassie stain. Membrane fraction of B2372 was also loaded on the gel (10 µl), no band was present in the lane of the negative control. Signals were observed in the lanes of pNV2132 (Figure 3.15 A) and OacB was found running at the size of ~36 kDa and no aggregates were present in any of the lanes. However, no clear band of OacB is observed in the Coomassie-stained PVDF membrane (Figure 3.15 B). Western blot results confirmed; the best conditions for OacB expression were induction at a density of 0.4 with either 0.1% or 0.2% L-arabinose and incubation at 37 °C. It was observed that even loading at higher concentrations of proteins gave isolated bands as compared to the less concentration of protein in the samples. However, not boiling of samples had improved the sample quality and OacB was detected at the approximate running size. After the successful expression of OacB in pNV2132; an experiment was carried out to express cysteine less mutant of OacB in pNV2139 created in section 3.4. The samples were prepared in a 2x loading buffer with BME without heating. The expression of pNV2139 was achieved using the same optimal conditions as for pNV2132. Two gels were run one for western blot and the other for Coomassie staining. The western blot showed the expression of cys-less OacB in pNV2139 (Figure 3.15 C). The mass of OacB observed on western blots was ~36 kDa (Figure 3.15 A and C). The predicted molecular weight of OacB is 45 kDa, which agrees well with the running behaviour of membrane proteins on an SDS-PAGE gel. To explain, membrane proteins bind more detergent (SDS) and so the migration on SDS-PAGE does not strictly correlate with the molecular weight and is termed

"gel shifting" [188]. As an example, this abnormal behaviour on SDS-PAGE due to SDS binding has been observed for membrane proteins from *R. rubrum* chromatophores [189].



40

Figure 3.15: A-C. Expression of OacB in B2607 and B2665 (TOP10 *E. coli* cells transformed with pNV2132 and pNV2139 respectively).
B2607 and B2665 cells were induced at OD_{600} either of 0.4 or 0.6 with either 0.1% or 0.2 % L-arabinose, cultures were grown at 37° C for 04 hours, membrane fractions were prepared, and western blot was performed using 1/10000 anti-His antibodies. A) Different concentrations (25, 50, and 60 µg) of OacB from pNV2132 were loaded onto the pre-cast gel. B) Coomassie-stained PVDF membrane, C) Membrane samples of pNV2132 (50 µg) and pNV 2139 (10 µg) loaded onto the gel. B2372 was used as a negative control. The pre-stained PAGE ruler is marked as L. B2372 was used as a negative control. OacB has an expected running size of ~44 kDa.

The unexpected finding that unboiled samples contained monomers of OacB was intriguing, so we decided to further test the role of heat in OacB aggregate formation. To do this, OacB was expressed using the optimised conditions described above, membrane proteins were prepared again and quantified using BCA assay. Before loading onto SDS-PAGE gels, a volume containing 10, 25, and 50 µg proteins was mixed with 2x loading buffer containin BME and boiled for 10 minutes. Two separate tubes containing 10 µg OacB were also boiled for shorter times 3 and 5 minutes. Other protein samples containing the same concentration of protein were prepared in separate tubes and not boiled. Western blot revealed aggregation of protein in all boiled samples for 3, 5, and 10 minutes (Figure 3.16). This suggested that heating, even at lower protein concentrations, was the main reason for protein aggregation in the previous experiments in this study.

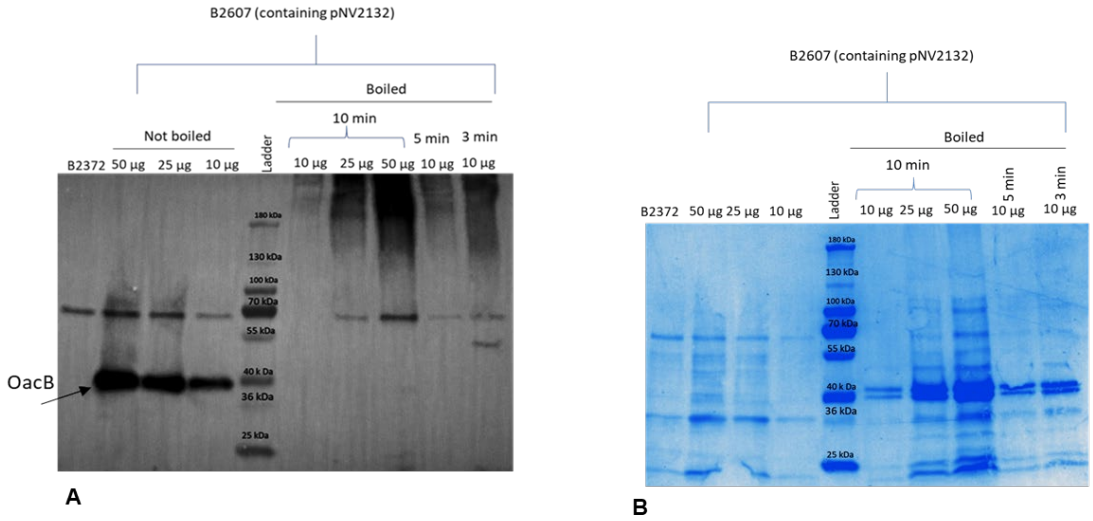


Figure 3.16A-B: Effect of heating on OacB aggregation.
A) Using optimized conditions for OacB overexpression, membrane protein samples were prepared and quantified using BCA assay. Samples containing 10, 25, and 50 µg were prepared in 2x loading buffer in duplicates or triplicates. A set of samples containing 10 µg (three separate tubes) were boiled for 3, 5, and 10 minutes. Another set of samples containing 25 and 50 µg protein were boiled for 10 minutes. A third set of samples containing 10, 25, and 50 µg did not boil/heat at all. All the samples were loaded onto SDS-PAGE gel. Following electrophoresis western blot was performed using 1/10,000 anti-HisA antibodies. **B)** Coomassie-stained PVDF membrane.

3.6.2.5 His-affinity purification of OacB

Once the expression of OacB was achieved successfully, His-affinity purification was performed in pNV2132 utilizing C-terminal 6x His-tag (Section 2.11.4). The purified protein was then subjected to SDS-PAGE for further analysis. Purified protein samples along with negative control (B2372) and unpurified pNV2132 were loaded onto three SDS-PAGE gels. After electrophoresis one gel was stained with silver stain, the second gel with Coomassie stain, and the third was subjected to western blot (Figure 3.17 A-C). In all the gels wash solutions (1 and 2) and eluted samples were loaded. In-gel for Coomassie stain flow through could not be loaded due to lack of sample and it was loaded in two other gels. Silver staining revealed a clear band of OacB (Figure 3.17A) in the lane of W2 (50 mM imidazole). This suggested most of the protein had eluted out with 50 mM imidazole concentration and less protein could be seen in elution with 250- and 500-mM imidazole. In Figure 3.17C western blot analysis showed no signal in the lane of flow-through sample whereas strong signals detected in the lanes of the lysate (membrane fractions mixed with buffer A) and wash2 (W2). Wash 1 (W1) and elution 1 &2 E1 and E2 also indicated the presence of OacB at the right size. In the Coomassie-stained gel, it was hard to detect OacB however the clearing of overabundant proteins could be seen after purification in the lanes of eluted samples. (Figure 3.17B).

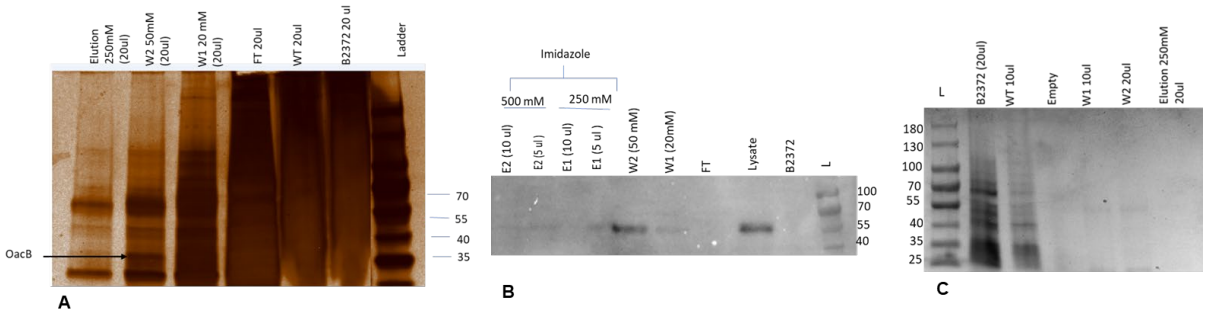


Figure 3.17A-C: His-affinity purified protein from pNV2132 expressed in *E. coli.*
***A*-B,** Purified protein samples were subjected to SDS-PAGE gels and stained with either silver stain or Coomassie stain. **A)** A reduced number of other protein bands are seen in the gels with a clear band of OacB running between 35 and 40 kDa size in the silver-stained gel. **B)** Western blot of purified OacB running at the expected size. **C)** In Coomassie-stained gel OacB was undetectable. (L=ladder; WT=wild type, W1= wash 1; W2= wash 2; FT= flow-through; E1 = elution 1; E2= elution2).

3.7 Conclusion

The OcaB expression experiments were performed to optimize the induction parameters. OacB expression was achieved in pNV2132 using 0.1 or 0.2 % of L-arabinose when added in cultures at an optical density of either 0.4 or 0.6 and incubated at 37 °C. During sample preparation for SDS-PAGE, only non-heated OacB samples worked for western blot with anti-His antibodies and better resolution/ strong band signals were observed in western blotting. Before this, samples were heated at 100 °C in the presence of loading buffer to denature the high order structures to ensure that the negative charge of amino acid is not neutralized and which let protein to migrate in an electric field [190]. OacB purification was also performed successfully using the IMAC technique. However, due to significant delay in achieving OacB expression SCAM experiments could not be performed. In a future study on OacB topology, recombinant plasmids pNV2132 and pNV2139 can be used and the parameters optimized for OacB can be employed to carry out OacB expression.

The topology model generated in this chapter was used to locate the position of conserved residues in OacB to perform site-directed mutagenesis in the next chapter. In chapter 4 of this thesis, the role of individual selected conserved residues in OacB was elucidated to unveil the mechanism of action of OacB.

Chapter 4: Identification of critical residues of O-antigen-modifying *O*-acetyltransferase B (OacB) of *Shigella flexneri*

Please note that the work presented in this Chapter is under review in the following article:

Munazza I. Rajput and Naresh K. Verma: Identification of critical residues of O-antigen modifying *O*-acetyltransferase B (OacB) of *Shigella flexneri.*

Manuscript submitted to BMC Molecular and Cell Biology 92021.

4.1 Introduction

Shigella flexneri, the predominant causative organism of bacillary dysentery in developing countries is accountable for 190 million shigellosis cases with 70,000 deaths annually [3]. Nineteen serotypes of *S. flexneri* have been recognized and each serotype differs from the other based on variations in lipopolysaccharide (LPS) structure. LPS plays an important role in bacterial virulence and the O-antigen is its most distal and variable domain followed by a core oligosaccharide and a lipid A. O-antigen is an immunodominant part of LPS and plays an important role during the infection process by evoking a serotype-specific immune response in the human host and any modification to the O-antigen helps bacteria escape the human immune system.

All serotypes of *S. flexneri* except serotype 6 bear a common O-antigen backbone of repeating tetrasaccharide units consisting of N-acetylglucosamine and three rhamnoses residues (serotype Y). This parent structure is modified by the addition of glucosyl, acetyl, or phosphoethanolamine residue(s) to one or more sugars to give rise to new serotypes [168, 191]. The O-antigen-modification by addition of acetyl or glucosyl residues is carried out by temperate phages. These phages integrate into the chromosomes of the host bacterium to form prophages that stably express serotype conversion genes [192]. Whereas phosphoethanolamine transferase (Opt) mediated O-antigen modification is plasmid-borne [193]. Gemski *et al.,* reported for the first time that Sf6 phage could be isolated from the 3a strain of *S. flexneri* [101]. Later, a 1002 bp gene coding for a 333 amino acid protein, Oac, was identified by Verma *et al.,* in 1991 [102]. Sf6- encoded *oac* gene converts serotypes X, Y, 1a, and 4a to 3a, 3b, 1b, and 4b, respectively, and brings about *O*-acetylation at Rhamnose III residue of the O-antigen. Oac is a TM protein with 10 alpha-helical membrane-spanning regions with N- and C- termini situated in the cytoplasm [168]. Like Oac a recently identified O-antigen modifying enzyme, *O*-acetyltransferase B (OacB), adds an acetyl residue at either position 3 or 4 of Rhamnose III (3/4-O-acetylation) in serotypes 1a, 1b, 2a, 5a, 7a, Y,

6, and position 6 of N- acetylglucosamine (6-O-acetylation) in serotypes 2a, 3a, Y and Yv of the O-antigen subunits [106]. The *oacB* gene is carried by a temperate bacteriophage Sf101 which is recently been isolated from a 1c strain of *S. flexneri* [105]. Serotype 1c (also known as 7a) was first reported from Bangladesh in 1980 and the O-antigen of this strain presents a unique architecture due to the presence of two glucosyl groups attached to N-acetylglucosamine and an acetyl group attached to Rha III. The addition of double glucosyl residues confers type 1C antigenic determinants [194]. While further modification of Rha III moiety on the same O-antigen confers a novel antigenic epitope called O-factor 9 known as “variant” factor in serotype 1c or 7a [195]. The antigenic diversity helps the bacterium escape the host immune system and hence is an important virulence factor. The mechanism behind the O-acetylation of O-antigen by OacB is yet unknown. Thus, the investigation of key features of OacB is needed to understand the mechanism of serotype conversion by OacB.

In this study, functional elements of OacB were determined using the site-directed mutagenesis (SDM) approach to explore the role of individual amino acids in OacB to understand their role in the acetylation of *S. flexneri O*-antigen. In addition, we identified that OacB belongs to the family of proteins that contain the acyltransferase-3 (AT3) domain (InterPro IPR002656 and Pfam Pf01757) which is commonly found in eukaryotes and prokaryotes involved in numerous acylation modifications [105, 110]. In prokaryotes, AT3 domain-containing proteins are involved in the acetylation of secondary cell wall polysaccharides or acetylation of peptidoglycan in the periplasm to induce lysozyme resistance [196-198]. In some organisms, AT3 domain-containing membrane proteins are linked with catalytic SGNH (Serine, Glycine, Asparginine, and Histidine) domain whereas in others only the AT3 domain is present. It is not clear how proteins containing only the AT3 domain function to modify substrates. OacB contains only the AT3 domain and in the present study using SDM we have identified critical residues in OacB to understand the mechanism of action of one of the important serotype-converting enzymes in *S. flexneri*.

4.2 Results

4.2.1 Physicochemical properties of OacB

To predict the physicochemical properties of the OacB, the Protparam tool in the Expasy server (http://web.expasy.org/protparam) was used. The amino acid sequence of OacB was used to calculate molecular weight and pI. Results predicted the molecular weight of OacB as 44,231.03 kDa and pI as 9.26. The instability index was computed as 29.56 indicating that protein is stable.

4.2.2 Topology analysis of OacB using prediction programs

TOPCONS webserver was used, which used five sub-methods (OCTOPUS, Philius, Polyphobius, SCAMI, and SPOCTOPUS) to produce a consensus model topology. A consensus model predicted OacB has 11 TM helices and 10 loops with the longest loop five at the periplasmic side. All five programs defined the orientations of the N-terminus and C-terminus as periplasmic and cytoplasmic, respectively (Chapter 3, Figure 1).

4.2.3 OacB similarities with other proteins

The bioinformatics tools NCBI-CDD search revealed that OacB and its homologs are clustered under protein superfamily COG1835 or acyl_trans_3 (AT3), having a functional conserved acetyltransferase domain Figure 4.1 [105]. The proteins having an acyl_tranf_3 domain are not uncommon in prokaryotes and eukaryotes and are involved in acylation modifications. Moreover, OacB protein BLAST returned with the OafA domain (corresponds to a group of proteins having two functional domains); AT3 domain; and SGNH hydrolase type esterase domain (Superfamily: SSF52266). Nevertheless, OacB lacks the SGNH domain while the acyltransferase 3 domain (Pfam: 01757) spans from 38-372 amino acids.

Proteins containing the SGNH domain along with AT3 or having either of the domains are involved in acetyltransferase activity [110, 199, 200]. In OacB, four conserved motifs are present: DGxRGxLAxxVxxHH, FFxxITG (Y or F) LFxxK, WxLxYEWxFY (x = any amino acid) at the N terminus, and fourth motif YSXYLxHG at the carboxyl terminus**.** Multiple alignments

of OacB with acetyltransferases of other 16 evolutionarily related bacteria revealed that OacB shares identity with acetyltransferase proteins from *E. coli, Pseudomonas spp, Shigella spp*, *Acinetobacter baumanii*, *Flavobacterium,* and several other bacterial AT3 proteins (Figure 4.2A).

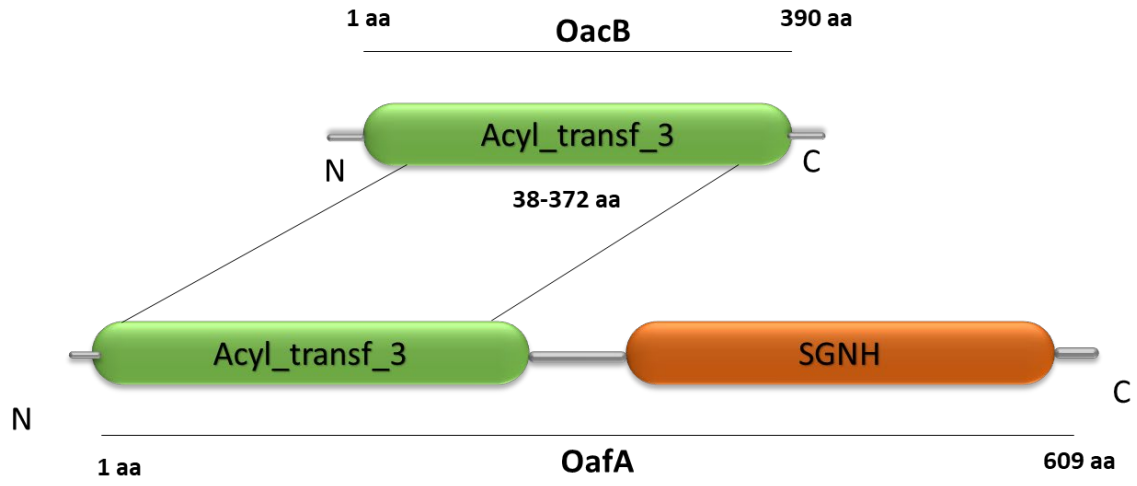


Figure 4.1: BLASTp hits for OacB.
Domain architecture of OacB (390 amino acids). OacB Acyl_trans_3 domain spans from 38-372 amino acids. Blast hits identified OacB like OafA (609 amino acid) containing acyltransferase and SGNH-hydrolase domains in OafA (609 amino acids). [amino acid: aa]

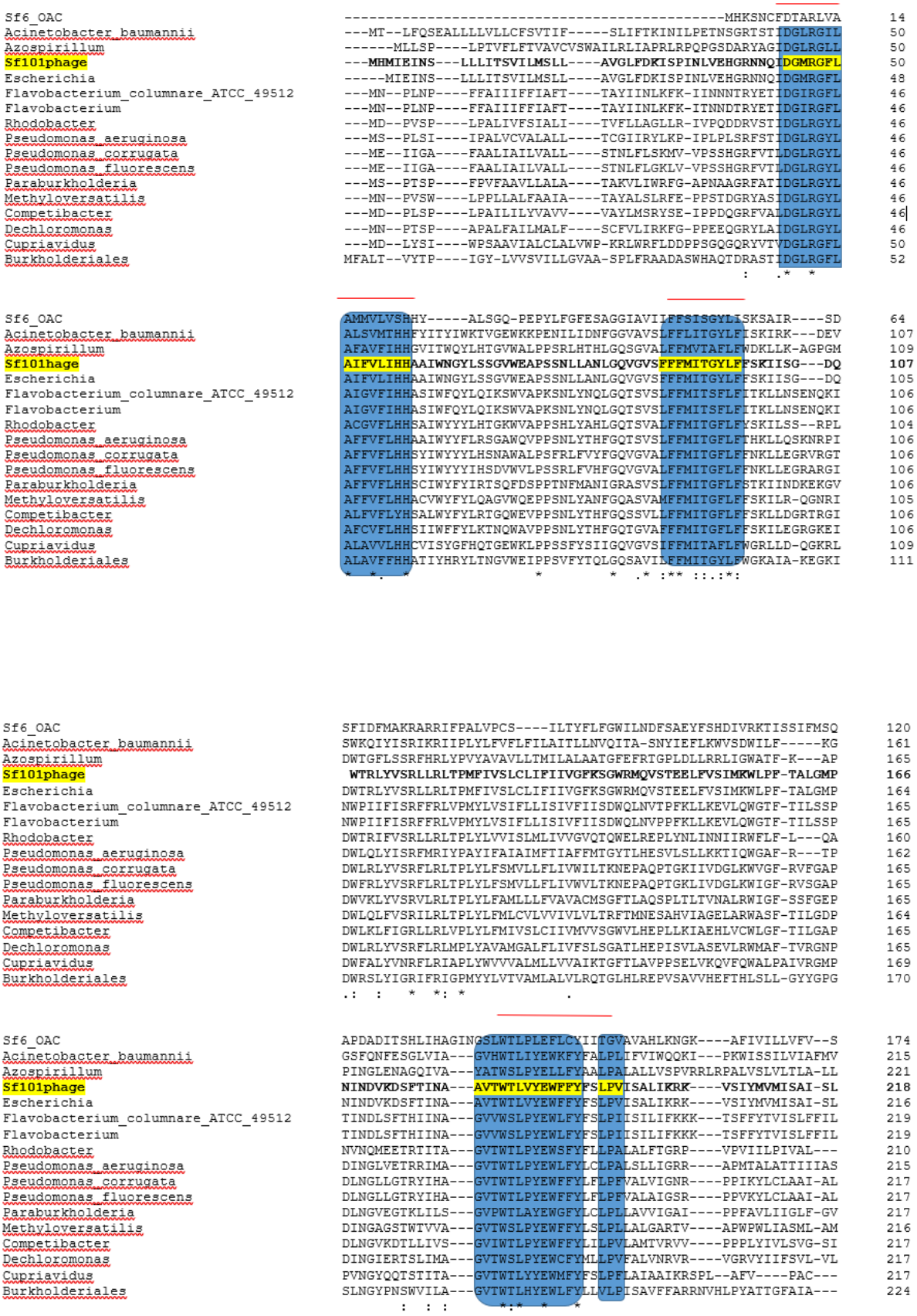

Sf6_OAC
Acinetobacter_baumannii
Azospirillum
Sf101phage
Escherichia
Flavobacterium_columnare_ATCC_49512
Flavobacterium
Rhodobacter
Pseudomonas_aeruginosa
Pseudomonas_corrugata
Pseudomonas_fluorescens
Paraburkholderia
Methyloversatilis
Competibacter
Dechloromonas
Cupriavidus
Burkholderiales

Figure 4.2A: Multiple alignment of Sf101 OacB with homologues acyltransferases.
Clustal Omega was used to align OacB protein from Sf101 phage with its homologues in other species. Conserved motifs are shown in blue boxes and red lines on top. Yellow highlighted amino acids in blue boxes belong to Sf101 OacB.

The pairwise alignment of OacB with its closest homolog *O*-acetyltransferase (Oac) has shown conservation of critical arginines R116 and R119 in cytoplasmic loop 4 which correspond to R73 and R76 in cytoplasmic loop 3 of Oac respectively [168] (Figure 4.2B).

Figure 4.2B: Pairwise alignment of Oac and OacB at the protein level using Clustal Omega. Three conserved motifs in both the proteins are shown in blue boxes. Arginine 116 and 119 (R116 and R119) are shown in red boxes with their equivalent residues in Oac arginine 73 and 76 (R73 and R76).

.

Moreover, protein sequences of O-acetyltransferases of *S. flexneri* (Table 4.1) were obtained from GenBank to perform multiple alignments. Three motifs were found to be conserved in OacB, Oac1b, Oac, and OacC. Moreover, two arginines found to be critical in Oac function (R73 and R76) are very well conserved in all acetyltransferases of *S. flexneri* belonging to different serotypes (Figure 4.2C). The conservation of the motifs and other amino acids in acetyltransferases of *S. flexneri* warrants their conserved role in *O*-acetylation. Additionally, OacC showed marked homology (97%) with OacB among all others whereas, OacD showed

the least identity (13%) when compared in pairwise alignment using CLUSTAL Omega. Both OacB and OacC acetylate Rha III moiety of O-antigen in various serotypes of *S. flexneri* and homology between them indicates their close relationship (Figure 4.2C). Oac and Oac1b showed 25% homology with OacB.

Table 4.1 O-acetyltransferases of *S. flexneri* from different serotypes

Gene	Serotypes	Acetylated residue	Phage involved	Reference
oac	1b, 3a, 3b, 4b and 7b	2-O-acetylation at Rhamnose (Rha) I	Sf6	[201]
oacB	1a, 1b, 2a, 5a, Y, 6 and 1c. 2a, 3a, Y and Yv	¾ O-acetylation of Rha III Position 6 of N acetyl glucosamine (GlcNAc)	Sf101	[105]
oacC	Serotype 6	¾ O-acetylation Rh III	Phage-like structure	[194]
oacD	Serotype 2	6-O-acetylation of N-acetylglucosamine (GlcNAc)	SfII	[108]
oac1b	1b	2-O acetylation at Rha III	Novel phage	[109]

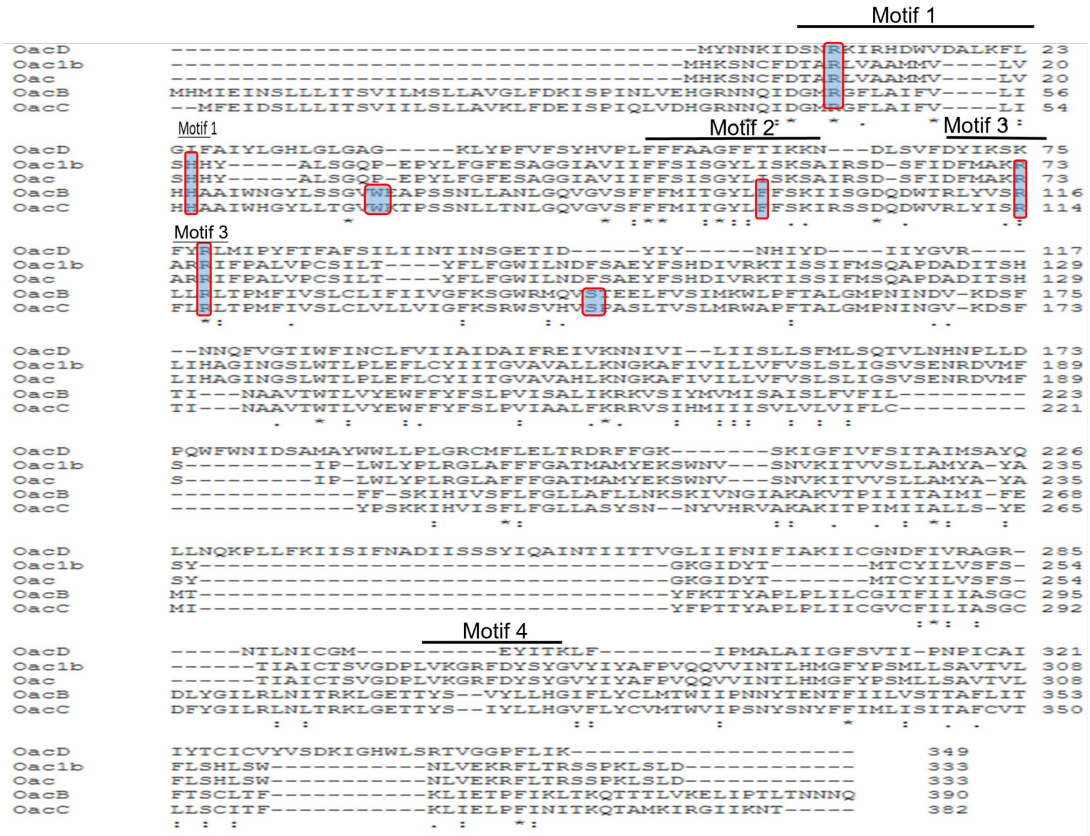


Figure 4.2C: Multiple alignments of acetyltransferases (Oac, OacB, OacC, OacD, and Oac1b) from Shigella flexneri strains.
The alignment was done using Clustal Omega. Asterisks and dots represent the amino acids that are identical or similar, respectively. Black lines above amino acids indicate four conserved motifs. Blue boxes with red outlines showing critical residues (R47, H58, W71, F98, R116, R119, and S146) identified in OacB.

4.2.4 Identification of critical residues in OacB

Multiple positions in membrane-bound OacB were targeted to mutate wild-type residues to alanine using SDM. The rationale for the selection of residues for mutagenesis largely was their conservation in acetyltransferases and chemical properties. However, several non-conserved residues were also selected to probe their role in OacB mechanism of action. Details of residue location and reason for selection are summarized in Table 4.2**.** In total 22 residues of different charges were mutated to neutral amino acid alanine. Alanine was chosen for replacing an amino acid in OacB as it maintains the native spatial structure of the protein. pNV2132 harboring wild-type *oacB* and erythromycin resistance genes in pBAD *Myc*-HisA

vector (Chapter 2 Table 2.5) was used as a template to create single/double amino acids point mutations. For mutagenesis Aspartic acid 44 (D44) and basic amino acid arginine 47 (R47) were selected in cytoplasmic loop 2, which are part of the conserved motif 2. Arginine 116 (R116) and 119 (R119) were selected in cytoplasmic loop 4. Aromatic residue tryptophan 71 (W71) from periplasmic loop 3, phenylalanine 98 (F98) polar residue, and tyrosine 96 (Y96) from TM III were selected to be mutated to alanine. Moreover, residues of different charges were also mutated to alanine; proline 122 (P122) in TM IV; lysine 156 (K156) in long loop five; aromatic tryptophan 183 and threonine 184 (WT183-184); glutamic acid 188 (E188), and aromatic pair of tyrosine and phenylalanine (YF191-191) in TM V and histidine 320 (H320) in TM X. Details of the reason of residue selection and location are summarized in Table 4.2.

Putative loop five in OacB is predicted to be the largest loop comprised of 40 amino acids. Most of the residues in loop five were not found to be conserved in other acetyltransferases of other bacteria (Figure 4.2A). However, the long loops have been found to play important roles in glucosyltransferases [160, 165]. Hence residues in loop five of OacB were targeted for mutagenesis to determine their role in OacB mechanism of action [165, 202]. Overall, seven mutations were carried out in loop five. These included polar amino acid serine 139 (S139), glycine 140 (G140), serine 146 (S146), serine 153 (S153), glycine (G164), an acidic aspartic acid 173 (D173), and serine 174 (S174) is replaced with alanine.

4.2.4.1 Functionality assay:

To confirm the function of OacB in mutants, slide agglutination was performed using 3/4 O-acetyl-specific antiserum. Mutated plasmids were electroporated into serotype 1c strain SFL1691 for functional assessment by conversion of serotype 1c into 1c “variant” strain due to the catalytic action of OacB. *S. flexneri* 1c serotype “variant” strain SFL1683 harboring *oacB* wild type gene was used as a positive control. Seven strains namely SFL2583 (mutation R47), SFL2595 (mutation H58), SFL2589 (mutation F98), SFL2584 (mutation W71), SFL2585 (mutation R116), SFl2586 (mutation R119) and SFL2617 (mutation S146) failed to agglutinate

with the antiserum. Whereas, low agglutination reactions were observed in strains SFL2618 (mutation G164), SFL2620 (mutation S139), SFL2621 (mutation S153), and SFL2590 (mutation H320) (Table 4.2). All the agglutination negative mutants were also found to be conserved in other acetyltransferases of *S. flexneri* (Figure 4.2C). Locations of critical residues are highlighted in the 2D model of OacB (Figure 4.3). However, other mutants D44, V87, Y96, P122, K156, D173, S174, WT183-184, E188, and FY191-192 reacted positively with the serotype-specific antiserum indicating functional enzyme.

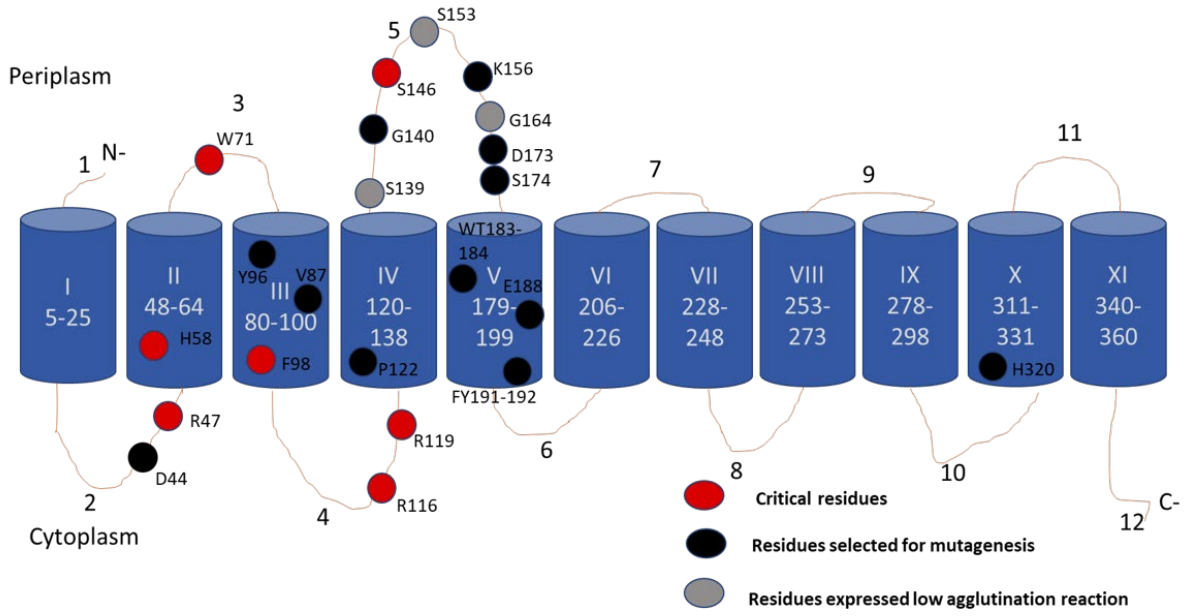
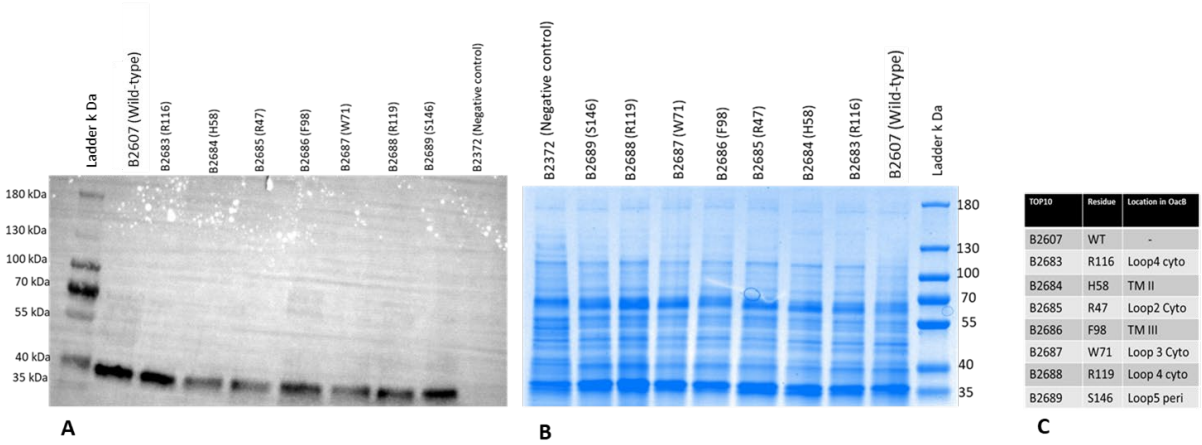


Figure 4.3: Topology model of OacB showing amino acids selected for site-directed mutagenesis in OacB.

Topology model showing amino acids selected for mutagenesis. Red circles indicate amino acids critical for OacB whereas, black circles indicate non-critical residues. Grey colored circle indicating residues showed low agglutination reaction in slide agglutination test using serotype-specific antisera.

4.2.5. Role of residues in the assembly of protein in the membrane

To confirm the assembly and synthesis of non-functional mutants of OacB in the plasma membrane, membrane proteins of all the mutants were prepared. An equal amount (~25 µg) protein of all OacB non-functional mutants and wild-type OacB proteins were subjected to SDS-PAGE gel. Western blot using anti-His antibodies was performed. All the non-functional mutants of OacB were found to be assembled in the membrane as the wild-type protein (Figure 4.4).



TOP10	Residue	Location in OacB
B2607	WT	-
B2683	R116	Loop4 cyto
B2684	H58	TM II
B2685	R47	Loop2 Cyto
B2686	F98	TM III
B2687	W71	Loop 3 Cyto
B2688	R119	Loop 4 cyto
B2689	S146	Loop5 peri

Figure 4.4A-C: SDS-PAGE and western blot of membrane protein of non-functional mutants of OacB.
A. Membrane proteins of mutants of OacB were prepared and subjected to western blot using anti-His antibodies. An equal amount of ~25 µg of each protein was loaded onto an SDS-PAGE gel. B2607 (wild type =WT) was used as positive control and B2372 was used as a negative control. Pre-stained PageRuler^{TM} was used as a molecular mass marker. Approximately 44 kDa OacB protein band is indicated. **B**, Equal amount of proteins loaded onto another SDS-PAGE gel and Coomassie-stained. **C**, Table summarizing location of critical residues in OacB.

Table 4.2 Purpose of the selection of different residues and results of slide agglutination assay using specific antisera to 3/4 O-acetylation against *S. flexneri* strains expressing different site-directed mutants of OacB

S. flexneri	Residue mutated to Alanine	Location in OacB	Reason for mutation	Agglutination level*	Functionality
SFL1691	N/A			-	
SFL1683	N/A			+++	Yes
SFL2591	D44	Cytoplasmic loop 2	Part of conserved Motif 1	+++	Yes
SFL2583	R47	Cytoplasmic loop 2	Part of conserved Motif 1	-	No
SFL2595	H58	TM II	Conserved in motif 1	-	No
SFL2584	W71	Periplasmic loop 3	Not conserved	-	No
SFL2596	V87	TM III	conserved	+++	Yes
SFL2588	Y96	TM III	Part of conserved motif 2	+++	Yes
SFL2589	F98	TM III	Part of conserved motif 2	-	No
SFL2585	R116	Cytoplasmic loop 4	Part of conserved motif 3	-	No
SFL2586	R119	Cytoplasmic loop 4	Part of conserved motif 3	-	No
SFL2593	P122	TM IV	Conserved	+++	Yes
SFL2620	S139	Periplasmic loop five	Not conserved Charge and part of large loop five	+	Yes
SFL2619	G140	Periplasmic loop five	Not conserved part of large loop five	+++	Yes
SFL2617	S146	Periplasmic loop 5	Part of conserved motif 3	-	No
SFL2621	S153	Periplasmic loop five	Not conserved part of large loop five	+	Yes

S. flexneri	Residue mutated to Alanine	Location in OacB	Reason for mutation	Agglutination level*	Functionality
SFL2587	K156	Periplasmic loop 5	Polar not conserved	+++	Yes
SFL2618	G164	Periplasmic loop five	Not conserved	++	Yes
SFL2616	D173	Periplasmic loop five	Part of conserved motif 1	+++	Yes
SFL2622	S174	Periplasmic loop five	Not conserved part of large loop five	+++	Yes
SFL2592	E188	TM V	Part of conserved motif 3	+++	No
SFL2594	FY191-192	TM V	Part of conserved motif 3	+++	Yes
SFL2597	WT183-184	TM V	Part of conserved motif 3	+++	Yes
SFL2590	H320	TM X	Not conserved catalytic residue	+++	Yes

Agglutination levels were detected as being either strong (+ + +), medium (+ +), weak (+), or no agglutination (−). The level of functionality of OacB corresponds to the level of agglutination. The absence of agglutination shows OacB is not functional or not present.

4.3 Discussion

In the present study, several critical residues in OacB were identified using site-directed mutagenesis (SDM) and agglutination assay. Alignment of proteins having homology with OacB helped identify several conserved residues present in TM helical regions and periplasmic and cytoplasmic loops. Selected residues were then subjected to SDM to confirm their role in OacB function. The functionality assays revealed the presence of seven critical residues in OacB. The assembly and synthesis of all the non-functional mutants of OacB in the plasma membrane were confirmed by western blotting on the membrane fractions of these mutants.

The non-essential residues identified in OacB were D44, Y96, V87, K156, P122, D173, S174 WT183-184, E188, and FY191-192. The critical amino acids detected through negative agglutination reaction include three arginine residues (R47, R116, and R119); phenylalanine (F98); tryptophan residue (W71); histidine (H58), and serine (S146). The importance of arginine residues in protein assembly and catalysis is well-documented [167, 198]. Hence, we targeted three conserved arginine residues for substitution with alanine R47, R116, and R119 to confirm their role in OacB. The results revealed that the replacement of any of the three arginine residues affected the function of the protein, as the mutants failed to react with serotype-specific antiserum. Residues R47, R116, and R119 are also found to be fairly conserved among proteins containing single (AT3) /double (AT3-SGNH) domain (s) acetyltransferases in other domains of life [110, 203]. In another study in which topological features of Oac were investigated, it was revealed that arginines can play diverse roles depending on their location in the protein and identified three critical arginine residues important for Oac function in the conserved cytoplasmic RxxR motif [167, 201]. R75 and R76 were found to play a structural role whereas R73 of Oac was attributed to the catalytic role. In the present study, two arginine residues of OacB, R116, and R119 in cytoplasmic loop 4, corresponding to Oac residues R73, and R76 respectively, were mutated to alanine. The functionality assays revealed a non-functional protein, and it is possible to correlate their essential role in OacB. Recently, corresponding /equivalent arginine residues R69 and R72 in TM helix I are also found to play an essential role in dual-domain protein OafA which acetylates abequose residue of the O-antigen in *Salmonella enterica serovar* P*aratyphi* [110]. Similarly, in another protein, OafB when corresponding arginine R71 and R74 in conserved RxxR motif were replaced with alanine the resultant protein was failed to complement the short O-antigen phenotype [203]. Moreover, the substitution of R47 in OacB also affected the functionality and hence suggests its importance in OacB. However, the role of the corresponding residue to R47 was not evaluated in Oac. R47, R116, and R119 are conserved in almost 30 other acetyltransferases belonging to other bacteria and plants acetyltransferases [110]. Among five

acetyltransferases of *S. flexneri*, both arginines (R116 and R119) are found to be conserved in all with only one difference that instead of arginine (R) another similar amino acid lysine (K) in OacD is present at the corresponding position of R119 of OacB.

Arginines have been reported as catalytic residue and served as a general base catalyst [204]. In a study, the role of the arginine residue (R359) is explained as the oxyanion hole forming residue in the PatB1 enzyme, which belongs to SGNH domain-containing protein and responsible for *O*-acetylation of secondary cell wall polysaccharides in *Bacillus cereus* [198]. When Arg359 was mutated to alanine to investigate its role in an experiment, the variant protein with mutated R359 (R359A) devoid of esterase or transferase activities confirming the importance of R359 as an active site residue. It appears that cytoplasmic R116 in OacB might be involved in a catalytic role, whereas R47 and R119 might have structural roles. However, further biochemical experiments are required to be performed to confirm the roles played by these amino acids in OacB.

In this study, W71 present in the periplasmic loop 3, was targeted for mutagenesis because of its conserved position in other acetyltransferases. The results obtained in this study showed the critical role of W71 in the function of OacB. Hence, it is yet to be determined that W71 plays either a catalytic or structural role in OacB. Tryptophan is an aromatic amino acid with large cyclic side chains, and capable of making protein to protein/ligand interactions with their neighbouring residues by its large hydrophobic polar side chain [205] [206]. Replacement of W71 residues in OacB resulted in the loss of protein function and maybe this residue also responsible for maintaining the structural integrity of the protein.

Another important residue identified in the current study was F98. The phenylalanine side chain is considered critical for full catalytic activity in a protein. A conserved phenylalanine residue F98 in TM III was replaced with alanine in OacB. Mutation of this residue affected OacB function which was discernible with the slide agglutination test. In a study to evaluate the role of phenylalanine residue, Shmara *et al.,* performed alanine scanning mutagenesis to

the aminoglycoside 6' -N-acetyltransferase type 1b that confers resistance to aminoglycoside antibiotics, and the replacement of F171 resulted in the loss of resistance to kanamycin and amikacin, in the resultant derivatives [207]; they postulated the participation of this residue in acceptor substrate specificity. Phenylalanine is hypothesized to play a central role in the alignment of the acetyl group for the transfer to the substrate [208]. Likewise, in glucosyltransferases, F414 in periplasmic loop 10 of the GtrII of *S. flexneri* was identified as critical to GtrII function and corresponding phenylalanine residues in GtrI, GtrII, GtrIV, GtrV, and GtrX of *S. flexneri* are postulated as an important residue. Similarly, in another transferase enzyme, Oac when FP78-79 and FPV 282-84 in the TM domains III and IX, respectively, were substituted with alanine, the function of the protein was affected, and it was hypothesized that phenylalanine is involved in making active site [201]. Furthermore, the study on arylamine acetyltransferases NAT1 and NAT2 supported that phenylalanine 125 residue involved in forming active site and played role in substrate selectivity in human arylamine acetyltransferases NAT1 and NAT2 [209]. A similar role of F98 can be attributed in OacB due to its location close to conserved arginine residues (R116 and R119).

The role of four serine residues S139, S146, S153, and S174 found in large periplasmic loop 5 in OacB was also investigated. Serine residues are found to be critical for activity in many proteins [110, 201, 210]. Conserved serine residues in Ser-Lys dyad or catalytic triad in SGNH hydrolases are involved in catalytic activities in enzymes. In a study on the O-acetylation of secondary cell wall polysaccharide (SCWP) by PatB1, serine 337 was identified as a catalytic nucleophile in the crystal form of PatB1. Similarly, in another study serine residue in OafA of *Salmonella enterica serovar Typhi* (O:5) was found to constitute an oxyanion hole in the C-terminal SGNH periplasmic domain [211]. In the present study, when four non-conserved serine residues S139, S146, S153, and S174 present in large periplasmic loop 5 were replaced with alanine, and functionality of each mutant was tested in slide agglutination assay, only S146 mutant (SFL2617) failed to agglutinate with the serotype-specific antiserum. Amino acid, S174 was replaced with alanine without affecting the function of OacB, whereas mutants

S139 and S153 showed low agglutination reactions. The substitution of the latter two may have resulted in disruption of the catalytic site involved with the substrate interaction in the periplasmic loop 5. The complete absence of function observed in the case of S146 mutation indicates the essential role played by this residue. Considering the O-acetylation occurs in periplasm we hypothesize that residue S146 in the large periplasmic loop 5 is responsible for adding an acetyl group to the Rha III moiety of O-antigen. The critical role of large periplasmic loops has previously been reported in glucosyltransferases of *S. flexneri* [160, 212].

Furthermore, in the present study when two glycines in large periplasmic loop 5 were mutated to alanine, medium agglutination reaction was observed for G164 as compared to the positive control, whereas G140 agglutinated strongly. Glycine is a non-polar amino acid having a single hydrogen atom as its side chain instead of carbon. This feature permits glycine to fit into tight turns of the protein that might limit other amino acids. The exact role of glycine in OacB is not yet clear and it is thought that the reduced agglutination reaction detected for mutant G164 might be due to the structural connection of this residue with active site residue (S146) and when mutated to alanine the interaction with active site residues was affected. Fanny *et al.,* investigated a network of five residues forming the active site in alkaline phosphatase enzyme and identified three dynamically independent but architecturally interconnected residues [213].

The definite role of Histidine 58 (H58) in OacB is not clear, however, the conserved status of this residue in thirty other acetyltransferases found in other organisms indicates its essential role [110]. In the present study, the loss of function of OacB after replacing H58 with an alanine confirmed its critical role. In OafA and OatA histidine is part of the catalytic triad of the SGNH domain and is involved in catalysis [110, 214]. Moreover, an equivalent residue H25 in OafA of *Salmonella enterica serovar Typhimurium* has recently been found to play a role in processing the bound acetate to arginine residue-R14 in TMH1 and transferring it to the periplasmic side of the inner membrane [110].

Conclusion:

This study assessed the role of conserved and non-conserved amino acids in OacB and has yielded valuable information on the functional roles played by individual amino acids. Functionally important residues were identified in TM domains, cytoplasmic and periplasmic loops in OacB. Substitution of R47, R116, R119, H58, W71, F98, and S146 changed the phenotype of the protein. Our results reiterate the importance of strictly conserved arginine residues in acetyltransferases. The knowledge gained about the critical residues and domain of this protein can help support future studies to understand the O-antigen modification mechanism by OacB and subsequently other acetyltransferases.

Chapter 5: Acquisition and distribution of *O*-acetyltransferase B (*oacB)* gene in *S. flexneri* genome

5.1 Introduction

Bacterial evolution is mainly brought by either natural selection through beneficial mutations or via mobile genetic elements (MGE) including bacteriophages, insertion sequences (IS), pathogenicity islands, integrons, and plasmids [215, 216]. MGE are powerful drivers of rearrangements and reshuffling of bacterial genomes, and *Shigella* genomes are reported to contain many insertion sequences [217, 218]. In *S. flexneri* strain 2457T, IS constitute up to 6.7% of the chromosome. Moreover, bacteriophages constitute 10-20% of bacterial genomes and are one of the major sources of exogenous bacterial genetic information. *E. coli* strain 0157:H7 harbors 18 prophage elements constituting 16% of its total genome [219].

In *S. flexneri* genome, seven phages SfI, SfII, SfIV, SfV, SfX, Sf101, and, Sf6 have been characterized which carry serotype-converting genes encoding glucosyltransferase (*gtr) I, II, IV, V, X,* and *O-acetyltransferase (oacB/oac),* respectively. These phages are responsible for evolving new serotypes of *S. flexneri*. The O-antigen of serotype 1c represents three different modifications, first two modifications represent two glucosyl residues attached to 04 of the GlcNAc residue via α-D-Glc*p*-(1→2)-αD-Glc*p* linkage. The addition of each glucosyl group is mediated by two different phages, Addition of the first glucosyl group is mediated by SfI phage which encodes *gtrI* operon. However, the phage responsible for mediating the addition of the second glucosyl group which encodes *gtrIC* operon has not yet been characterized [100, 220]. The third modification of O-antigen in serotype 1c strains is mediated by an *O*-acetyl transferase gene (*oacB*) carried by Sf101 phage giving rise to the 1c variant. *oacB* adds acetyl group at either position 3/4 of rhamnose III molecule of the O-antigen [105].

In all the phages of *S. flexneri*, the genes for O-antigen modification are found next to the phage integrase gene and attachment sites. Lambdoid phages integrate into bacterial chromosomes via site-specific recombination and the cross-over takes place between homologous sequences on bacterial chromosomes also known as attachment site (*attB)* and homologous sequence on phage genome, called as phage attachment site (*attP)*. The *attP*

location on the phage genome is typically next to the phage integrase gene. In the Sf101 phage, the serotype converting gene "*oacB*", is found next to the *attP* and phage integrase gene [105]. Except for two phages, Sf6 and Sf101, all other phages responsible for O-antigen glucosylation, integrate into the host chromosome at the *tRNA-thrW* site present between *proA* and *adrA* [23, 175, 220]. Sf6 phage (carrying *oac* gene) integrates into *argW tRNA* gene in the host chromosome and Sf101 phage, integrates within *sbcB* gene present next to *yeeD* gene [104, 105]. In two strains of serotype 1c (SFL1683 and SFL1684), *oacB* gene is located at the *sbcB* locus, whereas, in other 1c, 1a,1b, 2a, 5a, and Y serotype strains, *oacB* maps upstream of *adrA* gene and is flanked by transposases and integrase genes, representing a transposon like structure. It is thought that *oacB* gene moved from one location to other in the chromosomes of *S. flexneri* strains via a transposon activity and subsequently disseminated in other serotypes [105, 113]. Two alternative locations for *oacB* gene in *S*. *flexneri* chromosome invite questions about the origin of Sf101 phage and the evolution of the 1c serotype.

In our lab collection of sequenced bacterial genomes, eight 1c serotype strains contained *oacB* gene. Out of these, two strains SFL1683 and SFL1684 contained a complete copy of the Sf101 phage, whereas, in the other six strains *oacB* gene is located within the *proA-adrA* region.

To understand the integration of Sf101 phage, the complete genome of Sf101 lysogenic strain SFL1683 isolated from Egypt, was sequenced in this study using long reads MinIOn flow (Oxford Nanopore Technologies) sequencing platform (Section 2.19.2). The genome was annotated using RAST server and analysed on PHASTER to identify bacteriophage regions (2.19.3). The chromosomal regions of SFL1683 carrying *oacB* gene were analysed with a reference genome of serotype 1c strain, Y394 (GenBank: CP020753) in which, *oacB* was located upstream of the *adrA* gene. Additionally, Miseq sequences of other 1c (n=66),1a (n=9), and 1b (n=6) strain (Section 2.2), collected from different geographical locations were also analysed for the presence of *oacB* gene and other Sf101 phage remnants.

5.2 Results

5.2.1 Sf101 integration site sbcB locus

5.2.1.1 Analysis of sbcB locus in Sf101 lysogenic strain SFL1683

Sf101 genomic sequence was obtained from GenBank (KJ832078) to locate the prophage regions in the lysogenic strain SFL1683 (Section 2.19.2). Jakhetia *et al*., 2014, have reported the integration of Sf101 phage into the 5' end of the *sbcB* gene at *attB* in the bacterial chromosome [105]. The obtained Sf101 genome was mapped against the whole genome sequence (WGS) of Sf101 lysogenic strain, SFL1683 (this study). The analysis revealed that Sf101 integrated within the 5' end of the *sbcB* gene and resulted in the creation of two recombinant sites attachment right (*attL)* and attachment left (*attR).* The downstream region of *sbcB* gene (next to *attL*) comprised of conserved housekeeping genes including bacterial permease transport protein (*yeeD).*

Whereas the upstream region of *sbcB* gene contained whole Sf101 phage sequences starting from an integrase gene (encoded by *orf18)* until the last gene *orf17* of Sf101 phage which was flanked by *attR* (Figures 5.1 and 5.2). There were 227 sequences between *attL* and integrase/*orf18* which contained 46 bp of *sbcB* gene and 181 bp intergenic region (Figure 5.2).

Analysis of the right-side junction (~271 bp) spanning from an intergenic region between *orf17* and an insertion sequence (IS) in the lysogen revealed the presence of 25 bp sequence of *sbcB* gene, including *attR* with a base difference from *attL* (*attR* contained "T" instead of "C"). The BLASTn of this region with Sf101 genome identified 97% identity (262/271) at 18399-18130 bp (270 bp) in phage genome, corresponding to the intergenic region between *orf17* and *orf18* (Figure 5.2). In the lysogen, *oacB* gene was found upstream of the *sbcB* gene, and followed by *attR*. There was an abundance of mobile elements, spanning approximately 9259 bp region next to *attR* sequence. Hence, upstream region of *sbcB* gene was analysed carefully to identify individual IS elements using ISsaga [181]. IS66 was identified next to *orf17*, followed

by IS629 (Figure 5.1). There were three open reading frames (*orfA, B, C)* of IS66 present in this region spanning ~3 kb region of the lysogen chromosome. Once the Sf101 phage region was identified and analysed in the WGS of the lysogen; another 1c strain carrying Sf101 phage was analysed.

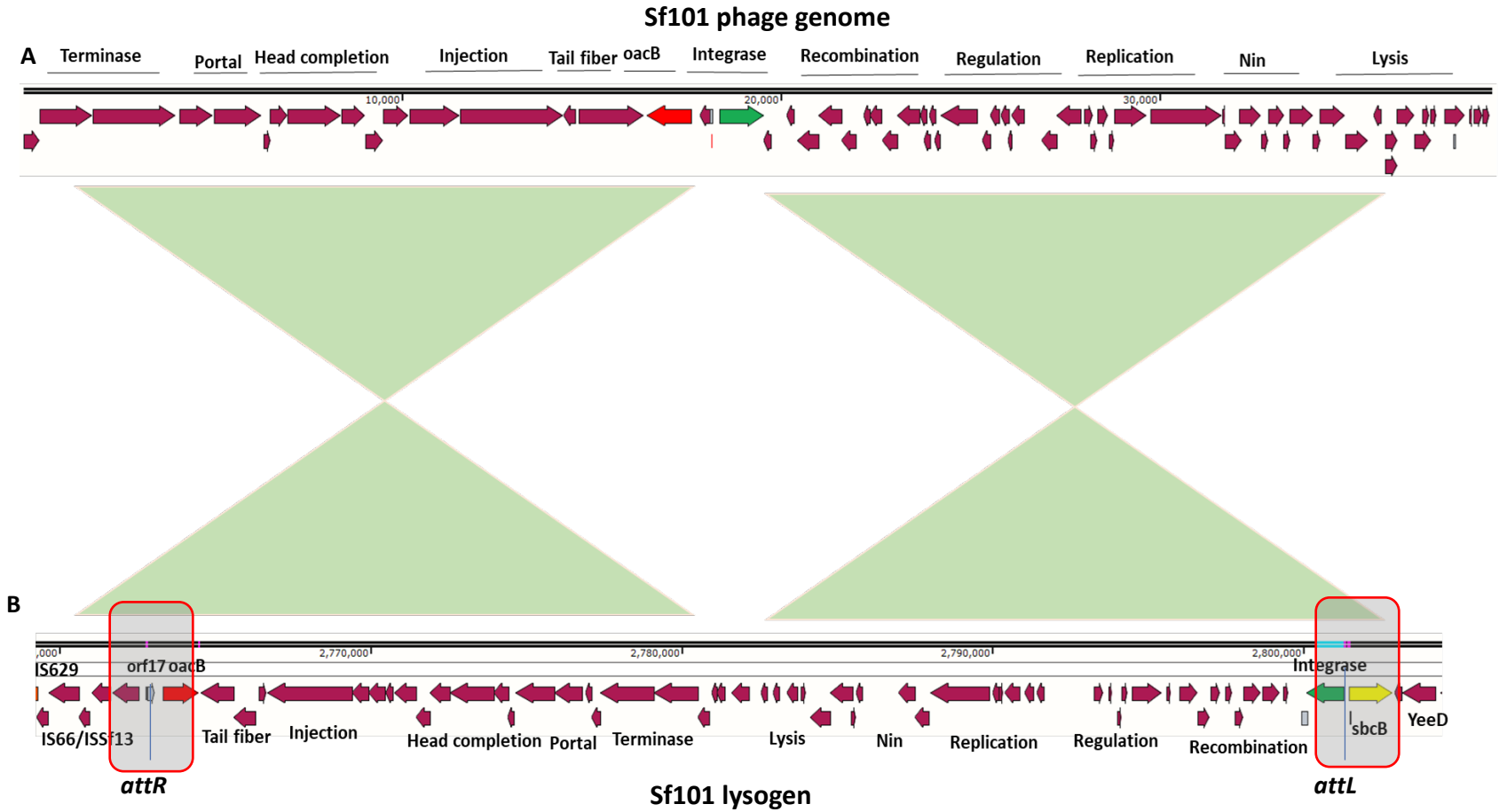


Figure 5.1: Comparison of Sf101 region in lysogenic strain SFL1683.
Sf101 genome and SFL1683 genomes are shown to scale **(A and B).** Open reading frames are directed with the arrows. All genes including IS elements/transposases are shown in plum colors except *oacB* (red); *integrase* (green) and *sbcB* (Yellow). **(A)** Complete genome sequence of Sf101 phage. The *oacB* gene is shown in red arrow and an integrase in green present next to each other **(B)** Complete Sf101 lysogen in SFL1683 strain of serotype 1c. The Sf101 phage is integrated into the *sbcB* gene (shown in yellow and enclosed in red box). downstream sequences of *sbcB* gene comprised of housekeeping gene (*yeed*) and at upstream there was an integrase (*orf18*) of Sf101 phage. The *attL* is shown within a red box in between *integrase* and *sbcB* genes. The Integrase gene (green) and *oacB* gene (red and enclosed in red box) are present on the opposite sides of the lysogen genome. Following *oacB* gene there is *orf17*, *attR* and a stretch of insertion sequences (IS), including IS66/ISSf13 and IS629 (only ~4.5 kb region shown in this image). [kb: kilo base].

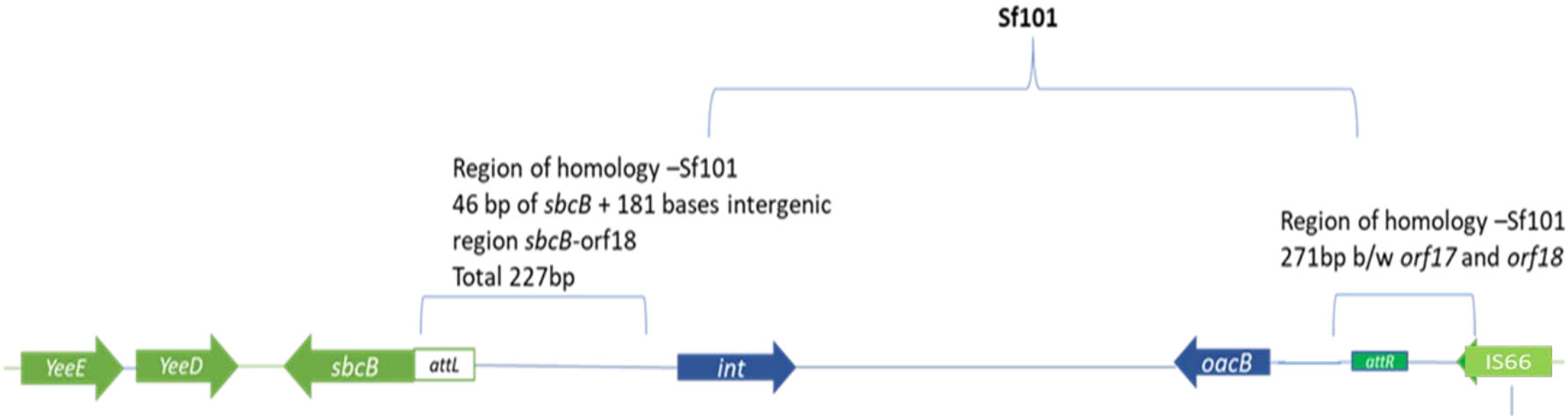


Figure 5.2: Prophage junction regions in lysogenic strain SFL1683.
Analysis of integration site of Sf101 phage in SFL1683 strain identified ***attL*** and ***attR*** sequences on left and right sides of phage integration. Both *attL/attR* separated by Sf101 sequences. There were 46 base pairs (bp) (13 bp of *attL present* within *sbcB* gene and 33 bp of *sbcB* gene) having sequence homology with Sf101 phage sequences. At *attL* there was 181 (bp) intergenic region present between *sbcB* gene and Sf101 integrase (encoded by *orf18*). At *attR* there were 271 bp present between *orf17* and IS66. Green lines represent phage regions in the host genome. Red outlines genes belong to the bacterial genome. Partial integrase gene is shown next to *attR*.

5.2.1.2 Analysis of sbcB locus in Sf101 lysogenic strain SFL1684

Another 1c lysogenic strain, SFL1684 was analysed for the integration of Sf101 phage within *sbcB* gene. Miseq sequence of SFL1684 was used to carry out BLASTn against the Sf101 sequence (KJ832078). A node 268 of 38295 bp, was identified carrying an almost complete copy of the Sf101 phage. The node was visualized using SnapGene viewer and found that all 66 *orfs* of Sf101 phage were present completely in node 268 except *orf18* (encoding integrase gene), whose first 260 bp were deleted /absent leaving behind last 919 bases (Figure 5.4A/B). Node 268 also contained 37 bp sequences which were 100% identical to the *sbcB* sequence (9-45 bp at 5' *sbcB*) including the Sf101 attachment site. It was concluded that the genetic arrangement at *sbcB* locus in SFL1684 was similar to another lysogen (SFL1683).

To determine the presence of complete or cryptic Sf101 phage, 1c, 1a, and1b strains from our lab strain collection along with the published complete genome sequences of *S. flexneri* and *Escherichia spp*. strains were also analysed.

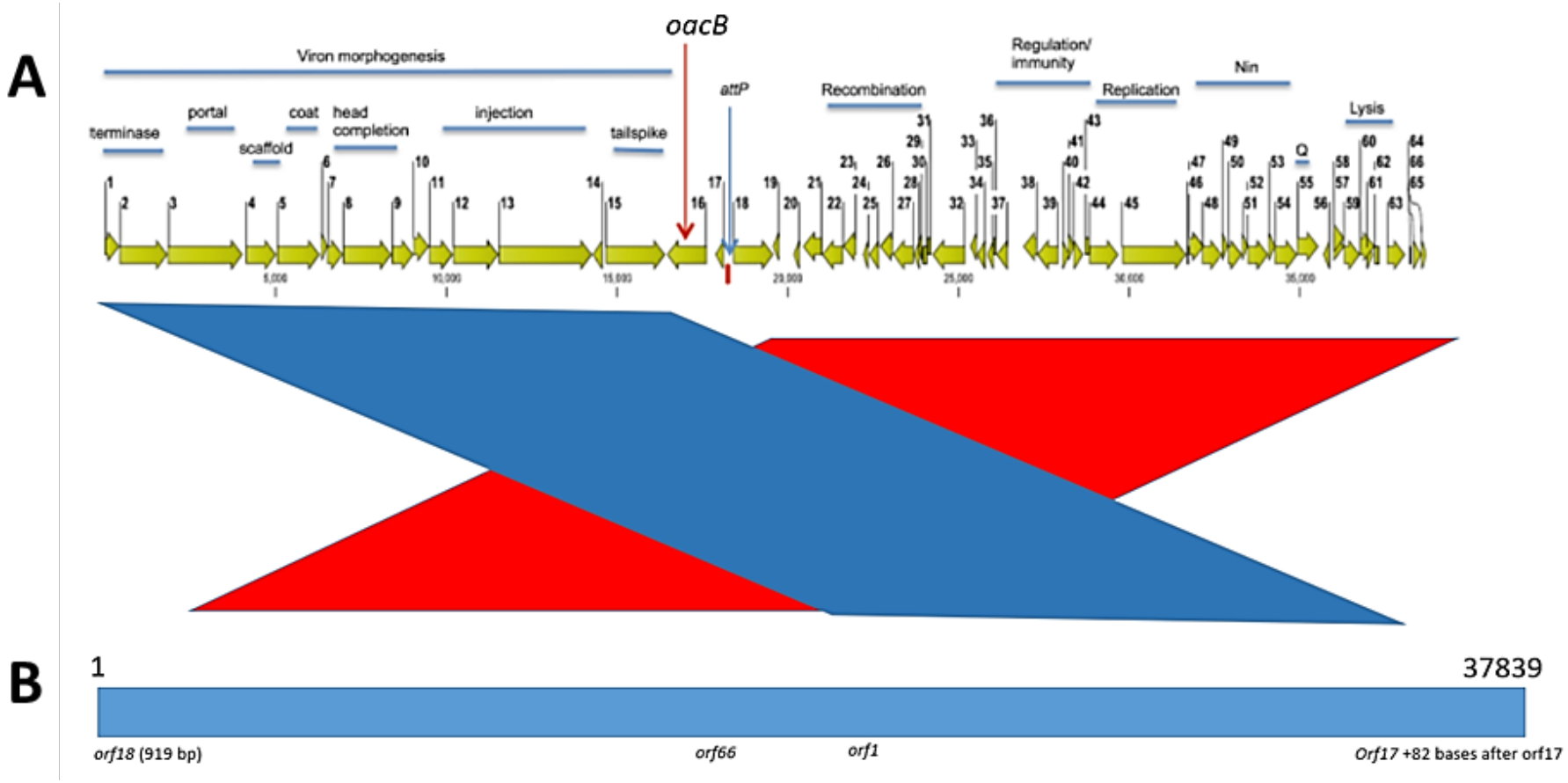


Figure 5.3: Sf101 lysogen (SFL1684) genome representing integrated Sf101 phage.
(A) Complete Sf101 genome containing *orf1-66*. The functional modules are indicated above the arrows [105] **(B)** Complete Sf101 lysogen sequence in *S. flexneri* 1c strain SFL1684.The Sf101 phage integrated within *sbcB* gene and *orf1-17* (blue) and *orf18-66* (Red). Red = collinear and blue = inverted synteny.

5.2.1.3 Analysis of sbcB locus in other serotypes of S. flexneri and E. coli

Whole-genome sequences of published *S. flexneri* / *E. coli* strains, our lab collection serotype 1c, and of other serotype strains (1a, 1b, Y) collected from different geographical locations were used to analyse genetic organization at *sbcB* locus. For this purpose, *sbcB* gene sequence was pulled out from the Sf101 lysogen (SFL1683) to carry out NCBI nucleotide (BLASTn) for the identification of the nodes carrying *sbcB* gene in all the selected strains (Table 5.1). Nodes identified were of varying sizes and nodes ranging in size ~4-5 kb were annotated using RAST server. Results revealed that like in the lysogen, downstream region of *sbcB* gene in all the strains including Y394 was very well conserved and comprised of housekeeping genes (y*eeD, yeeY*). Similarly, the upstream region comprised of insertion elements (IS) and and contained IS66 and IS629 (Figure 5.4). In particular, WGS of serotype 1c strain, Y394, was analysed in detail. In Y394 a ~4.8 kb upstream region to *sbcB* was found to contain insertion elements (IS66/IS629) from 2090160 to 2095006 bp followed by two hypothetical protein sequences. However, no *attR* sequence was identified in Y394 and any other sequences. Presence of *attR* indicates insertion of phage. Moreover, Sf101 phage attachment site "*attB"* was identified within the *sbcB* gene (as part of *sbcB* gene) in all the strains (Figure 5.4). A similar genetic arrangement from *sbcB* to IS elements was found in most of the genomes, however a few strains in which the sizes of the nodes were very small could not be analysed in much detail. Moreover, no Sf101 phage remnants including *attR* were identified in these strains. In next sections *proA-adrA* region in all the strains were analysed.

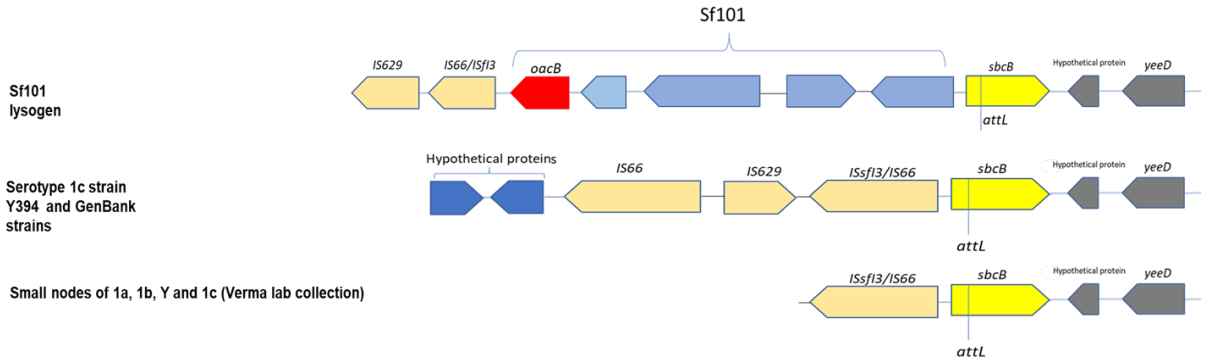


Figure 5.4: Genetic arrangement at *sbcB* locus in serotype 1c, 1a, 1b and Y strains of *S. flexneri* strains.

The whole-genome sequence of the lysogenic strain (SFL1683) carrying a complete copy of the Sf101 phage, was used to show (at the top in light blue) genetic arrangement at the *sbcB* locus. The direction of arrows indicates the orientation of the genes. Sf101 attachment *(attL*) is shown within *sbcB* gene. The *oacB* (in red color) is flanked by IS66, IS629, and IS66 on one side. The published sequence of another 1c strain Y394 is also included in the analysis. Nodes from 1c (n= 65), 1a (n= 9),1b (n= 6) and serotype Y from our lab. collection are also showing the presence of IS elements in the *sbcB* locus along with the Sf101 attachment site. Due to the small sizes of several nodes, the region beyond IS66 could not be analysed in serotype 1a, 1b, Y, and some 1c strains. Genome sequences of *S. flexneri* and *Escherichia spp*. were obtained from NCBI database which presented same genetic arrangement as in Y394 except few differences (which are not shown for simplicity). In all other strains except SFL1683, Sf101 phage is deleted /absent. The arrangement of IS elements in all strains is the same as lysogen. Color scheme: *sbcB* gene, yellow color; housekeeping genes y*eeD/yeeE*, grey color; beige color, IS elements and blue' hypothetical proteins.

Table 5.1 *S. flexneri* / *E. coli* strains obtained from GenBank for *sbcB* locus analysis

S. flexneri* / *E. coli	**Accession no.**	**Sf101 attachment intact**	***oacB***
S. flexneri FDAARGOS_716	CP050985.1	Yes	NO
S. flexneri	CP024984.1	Yes	Yes
S. flexneri	CP024983.1	Yes	Yes
S. flexneri 2a	CP045941.1	Yes	Yes
Serotype Y	CP042980.1	Yes	No
S. flexneri 7b	CP024473.1	Yes	No
S. flexneri 5a	CP037923.1	Yes	No
S. flexneri 2457T	AE014073.1	Yes	Yes
S. flexneri 301	JF813188.1	Yes	Yes
S. flexneri 1a	CP012735.1/ CP020342	Yes	Yes
S. flexneri 5908.2	CP045522.1	Yes	No
S. flexneri	CP033510	Yes	Yes
E. coli O68H12	CP061758	Yes	No
E. coli strain RHB31-C15	CP057259	Yes	No

5.2.2 Analysis of proA-adrA region in Y394 and SFL1683

Another reported site for *oacB* gene in *S. flexneri* genomes is within *proA-adrA* region. In a published serotype 1c strain Y394, *oacB* gene was located near SfI region within *proA-adrA* locus [221]. In Y394, SfI phage integrated into *tRNA-Thr* gene at *proA* locus via a site-specific recombination using a 46 bp SfI attachment site *attB*.

In this study, when the SfI region (containing complete *gtrI* cluster) in Y394 was analysed it was found that the *oacB* gene was present downstream of *adrA* (conserved gene) within SfI region and was flanked by insertion sequences. On one side of the *oacB* gene there was an integrase gene and other side IS600 followed by IS629. Notably, within *proA-adrA* region, three coipies of SfI phage attachment site were present (Figure 5.5A). In the Y394 genome, when the IS600 sequence was analysed, it was found that 1145 bp consisted of OrfB, which was duplicated and present on either side of the *oacB* gene (at the distance of 802 bp on either side of *oacB*). This OrfB on the left side of *oacB* also had disrupted integrase gene and hence only 465 bp of integrase were left.

The region of *pro-adrA* in the lysogen was also investigated to analyse the genomic arrangement in this region in the absence of *oacB* gene. The results revealed that, like Y394, the intact SfI (*gtrI* cluster) was present near *proA* gene. A 4 kb region immediately downstream of SfI region comprised of IS elements which was similar to the IS elements present in Y394. In the lysogen some IS elements were not present, which were found to be flanking *oacB* in Y394 (due to the absence of *oacB* gene in this location in the lysogen). Moreover, the SfI attachments site, was identified at two locations, one within the *tRNA* gene next to *proA* and another downstream *gtrA* of the *gtrI* operon in the lysogen (Figure 5.5B). The presence of triplicates of SfI attachment site was interesting and hence this attachment site was compared with Sf101 attachment site, using CLUSTAL Omega and found that out of 13 bp of Sf101 attachment site, 6 bp overlap with SfI attachment site.

The analysis of *adrA* locus in SFL1683 revealed that the downstream region of *adrA* comprised of conserved housekeeping genes (pyrolline D carboxylate, Yai, etc) whereas, upstream to *adrA*, there was IS629, a member of the IS66 family. This IS element was IS629 (1210 bp) present next to *adrA* in Y394. Downstream to IS629 in lysogen conserved housekeeping genes coding for phosphate starvation inducible protein (PsiF), protein ira P were present. No other IS elements were found in this ~6 kb stretch of chromosome in the lysogen. The *adrA* and *proA* genes in the lysogen were 50 kb apart from each other. In Y394 there was an abundance of IS elements near *adrA* gene due to the presence of *oacB* gene and SfI region and both were flanked by common IS elements (Figure 5.5B).

A

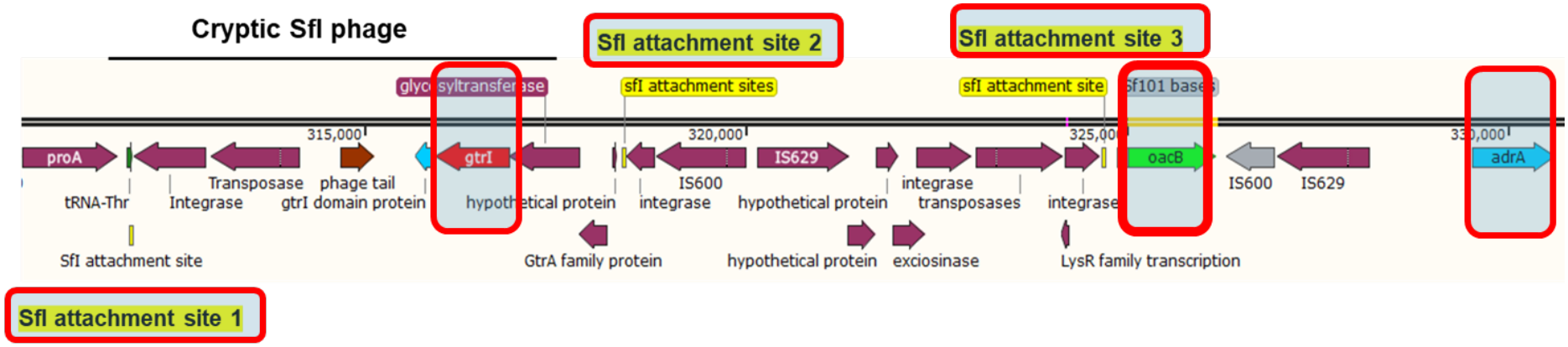
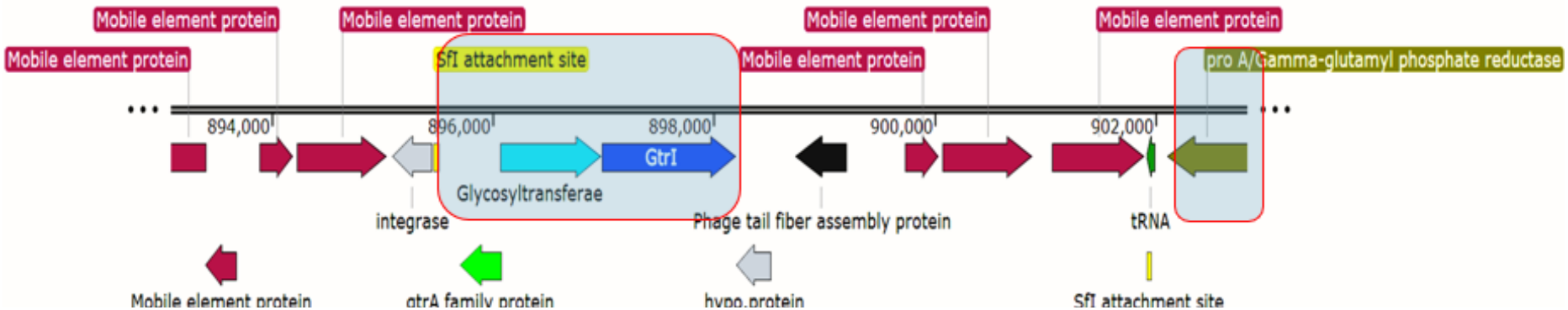


Figure 5.5A: Genomic region of serotype 1c strain Y394 carrying *oacB* gene.
The numbers below the scale bar show the positions in the genome in base pairs. The direction of the arrows indicates the orientation of open reading frames (ORFs). The labels represent major proteins encoded by the ORFs. The important proteins encoded by the ORFs are indicated in the labels. The important genes are colour coded; *GtrI,* red; *tRNA, dark green*; *oacB*, light green; SfI attachment sites, yellow; *adrA*, blue. Three genes *gtrI*, *oacB* and *adrA* are enclosed in red bordered boxes; all other orfs are shown in plum color including *proA* gene.



B

Figure 5.5B: *GtrI* /SfI region in Sf101 lysogen (SFL1683).
The arrows indicate the orientation of open reading frames (ORFs). The numbers below the scale bar show the positions in the chromosome. The important proteins encoded by the ORFs are indicated in the labels. The important genes are colour coded; *GtrI,* blue; glucosyltransferase, light blue; phage tail, black; *tRNA*, green; *proA*, dark green; mobile elements, plum; SfI attachment sites, yellow. The gtrI cluster and *proA* gene are enclosed in the red box

5.2.2.1. Analysis of region flanking oacB gene in other strains of 1c, 1a, 1b and Y serotypes and S. flexneri strains obtained from Genbank

In six 1c serotype strains (SFL1500-1503, SFL2448 and Y394), *oacB* gene was found to be located upstream of *adrA* gene between the *proA-adrA* region on the host chromosome. However, oacB and adrA genes were all present on two separate nodes of the short read sequences of these genomes. The nodes containing *oacB* gene was found to be ~ 3 kb in size and *adrA* carrying nodes were ~40-46 kb long. Hence, regions flanking *oacB* gene in all those starins were analysed in all these strains using Y394 as a reference genome on IGV.

BLASTn was performed to identify 1c strains positive for *oacB* gene and the nucleotide sequence of *oacB* gene was obtained from GenBank (Accession No. KJ832078). Five nodes of varying lengths (~3kb) from serotype 1c strains (SFL1500-1503 and SFL2448) containing *oacB* gene were annotated. It was observed that similar to Y394, in all the strains, there was an integrase gene flanking *oacB* gene on one side and IS elements on the other side. At the integrase side, there was an intergenic region of 348 bp out of which 132 bp were identical (98% homology) with Sf101 phage sequence (KJ832078:17766-17635 bp) corresponding to the intergenic region between *oacB* and *orf17* in the Sf101 phage genome.

On the other side of *oacB* IS600, there was a 173 bp sequence comprised of 53 bp of the intergenic region (between the *oacB* gene and IS600), and 120 bp of IS600, which had identities to intergenic region between *oacB* and *orf18* (integrase) in Sf101 phage genome (98%) and *orf15* (phage tail protein) 96% respectively. All 6 strains were found to have a similar genetic arrangement and *oacB* was flanked by integrase and IS elements (Figure 5.6).

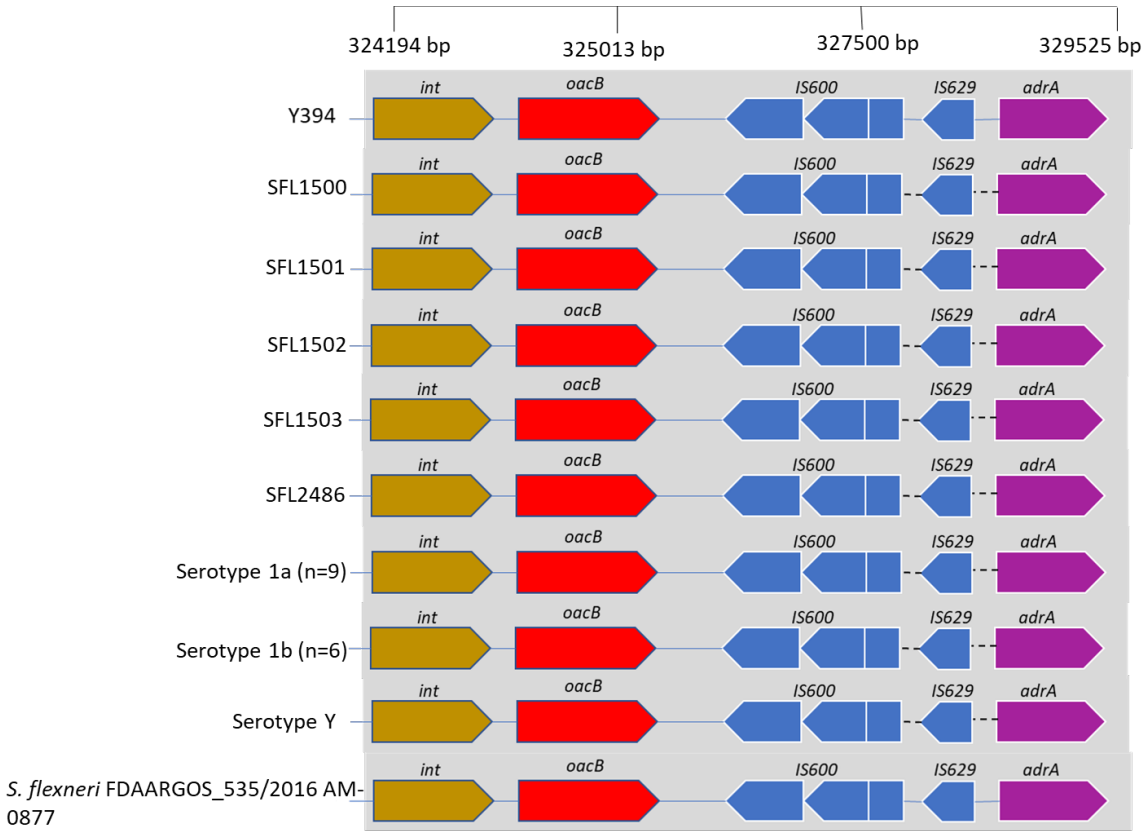


Figure 5.6: Genome arrangement of chromosomal regions carrying *oacB* gene in serotype 1a, 1b,1c, and other *S. flexneri* strains.
Genomic structure of regions flanking *O*-acetyltransferase gene *oacB*. Sequences of *oacB* positive strains from serotype 1c, 1a, and 1b were obtained from our lab collection of strains. The *oacB* positive genome sequences of *S. flexneri* FDAARGOs_535 (Accession: CP034060.1), and *S. flexneri* 2016 Am-0877 (Accession no. CP033510) were obtained from GenBank. The IS elements were identified using ISSaga and the genes are shown in different colors: *adrA*, magenta; IS629/IS600, blue; *oacB*, red; integrase (*int*), dark yellow. *oacB*-carrying transposon at *adrA* locus is highlighted in the grey box. The genomes are shown to scale. The numbers below the scale bar indicates positions in the genome in base pairs (bp).

In all the five strains, *adrA* gene was present on separate nodes and those were obtained by performing another BLASTn with *adrA* gene sequence (Table 5.2). In nodes from strains SFL1500-1503 and SFL2448, there was ~156 bp of IS elements upstream to *adrA*, whereas the downstream region comprised of conserved housekeeping genes coding for Pyrolline D carboxylate, DUF protein188, Shikimate kinase III and YaiA. These nodes from drafted genomes were compared with the Y394 sequence and it was presumed that these genomes also contain identical genetic arrangement as in Y394**.** As these genomes are short reads and

hence further confirmation is needed with long reads. Flanking sequences to *oacB* gene were also analysed in other *S. flexneri* genomes available in GenBank and found a similar genetic arrangement at *adrA* gene and *oacB* was flanked by a transposon-like structure (Figure 5.6). In all these strains no other Sf101 remnants were identified.

Table 5.2 Nodes of *adrA* and *oacB* in 1c serotype strains of *S. flexneri* (SFL)

Strains	origin	Node for *oacB*	Node for *adrA*
Y394	Bangladesh	complete genome	complete genome
SFL1500	Bangladesh	953 length 2440 bp	397, length ~45 kb
SFL1501	Bangladesh	1046 length 2985 bp	334, length ~45 kb
SFL1502	Bangladesh	537 length 2932 bp	206, length ~45 kb
SFL1503	Bangladesh	1966 length 2999 bp	1670 length 3562 bp
SFL2448	Egypt	964 length 3000 bp	96 length ~45 kb

5.3 Conclusion

The results presented here explained genomic arrangement of various serotypes of *S. flexneri* carrying *oacB* gene in their chromosomes. The analysis was performed by generating the whole-genome sequence of the Sf101 lysogen in this study, which contained Sf101 phage upstream of *sbcB* gene, and available MiSeq sequences of 1c, 1a, 1b, and Y strains were compared to analyse the chromosomal regions carrying *oacB* gene. The results demonstrated that in all the analysed strains genetic arrangement at *sbcB* locus was conserved including a 13 bp Sf101 attachment site within the intact *sbcB* gene. However, as the attachment site was part of the conserved *sbcB* gene, irrespective of the presence or absence of *oacB* gene, it did not determine the presence of the Sf101 phage and its remnants. It was found that the *sbcB* downstream region in all the strains, comprised of housekeeping genes and the upstream region comprised of IS elements similar to the lysogen but no Sf101 sequences or *oacB* gene. Another lysogenic strain was also analysed in this study and a node was identified carrying a complete copy of the Sf101 phage near *sbcB* locus. Only two lysogenic strains were found to contain a complete Sf101 phage downstream *sbcB* gene. No other strains were found to contain *oacB* gene or any Sf101 phage remnants upstream *sbcB* gene.

The *gtrI* and *adrA* regions in the lysogen (SFL1683), were analysed using Y394 as a reference genome. It was found that the lysogen contained complete SfI phage region downstream *proA* gene. There *gtrI* and *adrA* genes were ~50 kb apart from each other and in the lysogen. In Y394 and other strains, the same genes were only ~18 kb away from each other. Moreover, the *gtrI* region in Y394 contained *oacB* gene near *adrA* and was flanked by transposons. The *gtrI-adrA* regions were also analysed in other strains from serotype 1a, 1b, serotype Y from our lab. and published *S. flexneri* genomes from GenBank. The genomic regions within *proA-adrA* regions were similar and contained SfI region. Six serotype 1c strains and all 1a and 1b strains contained *oacB* gene within SfI phage region and in these strains, *oacB* gene was

possibly flanked by IS600 and IS629 on one side and an integrase at the other side (as inferred from Y394 genome). In the Y394 genome *proA-adrA* region was analysed in detail and found the presence of repetitive SfI attachment sites. The pairwise alignment of attachment sites of SfI and Sf101 phages identified six bp homology between them. Homology between two attachment sites indicated the integration of Sf101 phage in this region via SfI attachment site and explained the presence of *oacB* gene in some 1c strains.

.

Chapter 6: Study on the role of Sf101 phage-encoded novel genes in *S. flexneri* virulence.

Part of the work presented in this chapter appears in the following published article:

Pawan Parajuli, **Munazza I. Rajput** and Naresh K. Verma: Plasmids of *Shigella flexneri* serotype 1c strain Y394 provide advantages to bacteria in the host. ***BMC Microbiology***, 2019, 19(1): 86

6.1 Introduction

Prophages integrate into bacterial chromosomes and constitute important gene elements for horizontal gene transfer between bacteria and phages. Many phages encoded factors in bacterial genomes have evolutionary benefits for the host and contribute to bacterial fitness. Moreover, in some cases, prophage-encoded genes may increase the virulence of their bacterial host and result in a shift in the host phenotype, a process known as lysogenic conversion. Phage-encoded factors have been found to play an important role in host adhesion/colonization, intracellular survival and escape from host defenses [149, 151, 153]. The role played by genes encoded by temperate phages during the infection process has been recognized and the advancements in whole bacterial genome sequencing have revealed that a higher frequency of prophage integration has occurred in pathogenic bacteria than in non-pathogenic bacteria [126]. Among characterized phage-encoded factors, there is a growing list of uncharacterized unique genes encoded by serotype converting bacteriophages of *S. flexneri*.

Sf101 phage is a serotype converting phage of *S. flexneri* and encodes a serotype converting gene, *oacB*, which is found in serotypes 1a, 1b, 2a, 5a, 1c, and Y of *S. flexneri*. In a previous study comparative genomic analysis of Sf101 phage with phages of *S. flexneri* and *E. coli* identified four proteins of unknown functions, encoded by *orf16, orf17, orf41* and *orf56* [105]. These proteins were absent in any other phages and do not relate to any phage-related functions. It is known that *orf16/oacB* mediates 3/4 *O*-acetylation of Rha III of O-antigen and changes the antigenic epitope, resulting in the increase in the virulence of the host [23]. Hence, in this study, the role of *oacB (orf16)* and three *orfs* (*orf 17, orf41* and *orf56*) were evaluated with respect to the virulence of the host.

Firstly, a Sf101 lysogenic strain (SFL1683) was used to perform a reverse transcriptase-polymerase chain reaction (RT-PCR) to confirm the expression of these genes. The lysogenic strain SFL1683 was tested for the presence of virulence plasmid (VP) and found that VP was

not present. Since the use of this strain was not possible, attempts were made to restore VP by making Sf101 lysogen in the serotype Y strain (SFL1353) or conjugation. Finally, the role of these *orfs* were tested using another approach, in which all four genes were cloned individually into an expression vector and transformed into another 1c strain harboring VP to create four recombinant *S. flexneri* strains. These strains were subsequently used for virulence assays involving *C. elegans* and HeLa cells.

6.2 Results

6.2.1 Role of Sf101 unique genes in host virulence

6.2.1.1 Expression of oacB, orf17, orf41, and orf56 in the Sf101 lysogen

A previous study had identified four proteins in Sf101 phage that were absent in any other serotype-converting phages of *S. flexneri* [105]. These unique proteins of Sf101 phage were encoded by *orf16*/*oacB*, *orf17, orf41* and *orf56.*The role of *oacB* in O-antigen modification in *S. flexneri* was confirmed however, it's involvement in the virulence of *S. flexneri* is yet to be determined. The bioinformatics analysis of other three Orfs (Orf17, Orf41 and Orf56*)* showed absence of any conserved domains in these three Orfs. Additionally, Orf17 and Orf56 were found to have homology with hypothetical proteins of *E. coli*; and Orf41 had no protein homologue in *Shigella* and *E. coli*. Moreover, these *orfs* were also not identified as part of *S. flexneri* pathogenicity islands. Since no function is associated with these genes, their role was determined in the virulence of the host.

To determine whether phage genes are involved in the virulence, they must be expressed in the lysogen, so RT-PCR was performed to check the expression of all four genes in the Sf101 lysogen. RT-PCR results showed expression of all four genes (Figure 6.1). Expected size products were obtained in all four genes-specific primers indicating the expression of all four genes in the lysogen.

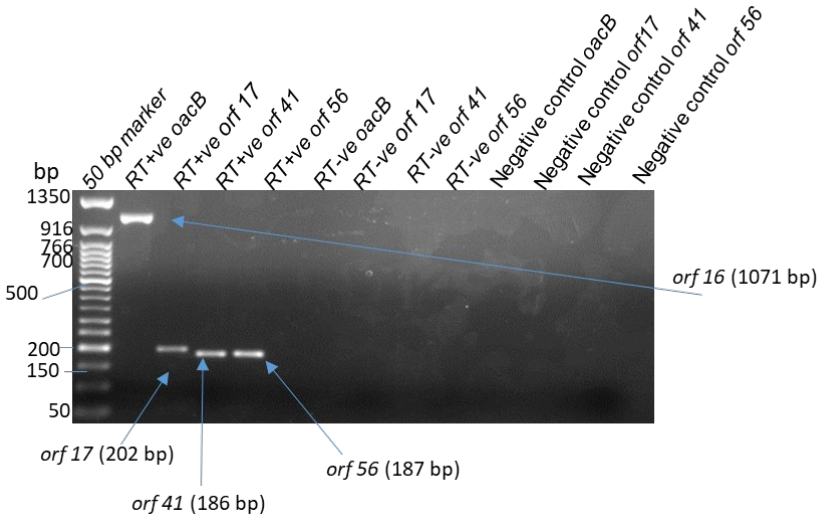


Figure 6.1: RT-PCR for the confirmation of expression of Sf101 orfs in the lysogen.
Total RNA of SFL1683 was used to produce randomly primed cDNA. Gene-specific primers were used to amplify the gene of interest from the cDNA. Lane 1 has 50 base pairs DNA marker, RT+: reaction with reverse transcriptase, RT-: reaction without reverse transcriptase (negative control for DNA contamination). The negative control is PCR control having all PCR reagents with sterile MQ water as a template. Expected sizes: *orf16/oacB* = 1071 bp, *orf17* =202 bp, *orf41*= 186 and *orf56*=187bp. (base pairs =bp).

6.2.2 Restoration of virulence plasmid

Several strategies were employed to restore /reintroduce VP into the SFL1683 strain. Experiments were performed to make Sf101 lysogen in serotype Y strain (Section 2.2) and even, after multiple attempts Sf101 lysogen in serotype Y could not be obtained. Hence, it was decided to restore the VP in the lysogen using conjugation.

6.2.2.1 Virulence plasmid tagging with kanamycin gene.

Conjugation of SFL1683 with another 1c strain carrying VP was considered as a possible way to reintroduce VP. However, the process of screening for colonies harboring VP would potentially consume an enormous amount of time. Additionally, a large number of potentially transconjugant colonies would need to be screened with 3/4 *O*-acetyl antisera to isolate SFL1683 strains with no guarantee of the presence of VP. Hence, to confirm the presence of VP in these colonies, colony PCR would be needed to perform using primers targeted to the

VP sequence (*apy/virG*). And afterward, the VP restored strain could be used to make knockouts for virulence assays. Nonetheless, plasmid could be lost in this strain too, due to the absence of any selection to maintain it. Therefore, it was decided to introduce a selection marker (antibiotic cassette) into the non-coding region of the VP using the lambda red recombinase method. The tagging of VP with kanamycin cassette was successfully performed and the strain containing tagged VP was used for mating experiments. To restore the VP in SFL1683 conjugation experiments were performed using another serotype 1c strain SFL 1613 (Y394) containing the VP, as a donor strain. Antibiotic susceptibility testing confirmed that SFL1683 and SFL1613 were sensitive to kanamycin antibiotics. Hence, before conjugating, the VP of SFL1613 was tagged with the kanamycin *(kan)* resistance gene to facilitate the selection of resultant transconjugants. The putative reverse transcriptase (RT) gene, within the large virulence plasmid, was selected for the disruption by inserting *the kan* gene using homologous recombination.

Disruption of a gene by the lambda red system requires a helper plasmid such as pKD46 or pKM208 so, SFL1613 was transformed with purified helper plasmid pKD46 (containing ampicillin resistance gene-Amp^R) to generate SFL2606. The helper plasmid encodes the proteins that are required for lambda red mediated homologous recombination. To disrupt the RT gene in VP, 70 bp long primers were used which had 50 bp sequence homology to the RT gene and 20 bp sequence homology to the kanamycin gene (Figure 6.2). The *kan* gene was amplified using pKD4 miniprep DNA with the 70 bp long primers and the amplicon (Figure 6.3) was then transformed into SFL2606 expressing lambda red genes and plated on kanamycin/ampicillin plates for selection of the mutants. The obtained colonies were first screened by colony PCR using forward primer (KO test Fwd) and reverse primer (KO test reverse). The primer sets were chosen so that the forward primer annealed at the end of the *kan* gene, and the reverse primer bound to the downstream region of the RT gene (Figure 6.4 and 6.5). The mutants were further confirmed by Sanger sequencing and one of the colonies was used to make the glycerol stock and named SFL2582. The primers were designed to

amplify two genes; apyrase gene (*apy*) and an outer membrane protein IcsA also known as *VirG*, present on the VP. As these genes are present on opposite ends of VP, their amplification would confirm the presence of VP. Hence, the presence of virulence plasmid in SFL2582 was confirmed by colony PCR using *apy* and *virG* specific primers PCR (Figure 6.6). Thus, virulence plasmid was tagged successfully and was used for subsequent conjugation experiments.

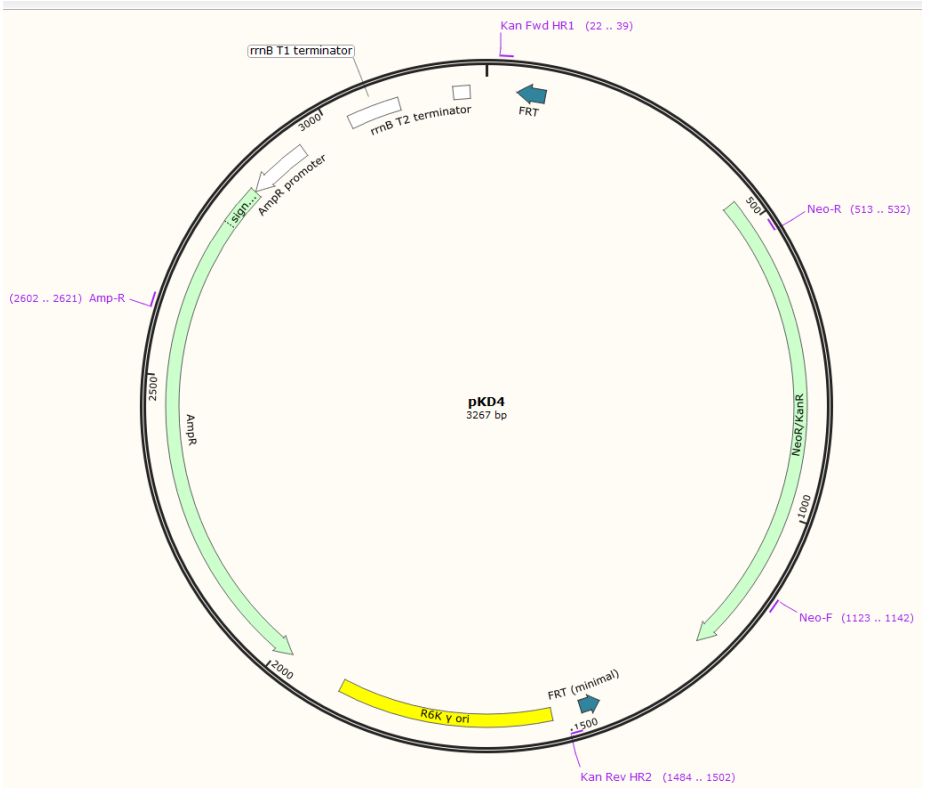


Figure 6.2: Map of pKD4 with Forward (kan Fwd HR1) and Reverse (kan Rev HR2).
HR1 and HR2 refer to the primer flanking the region homologous to the upstream of the reverse transcriptase gene.

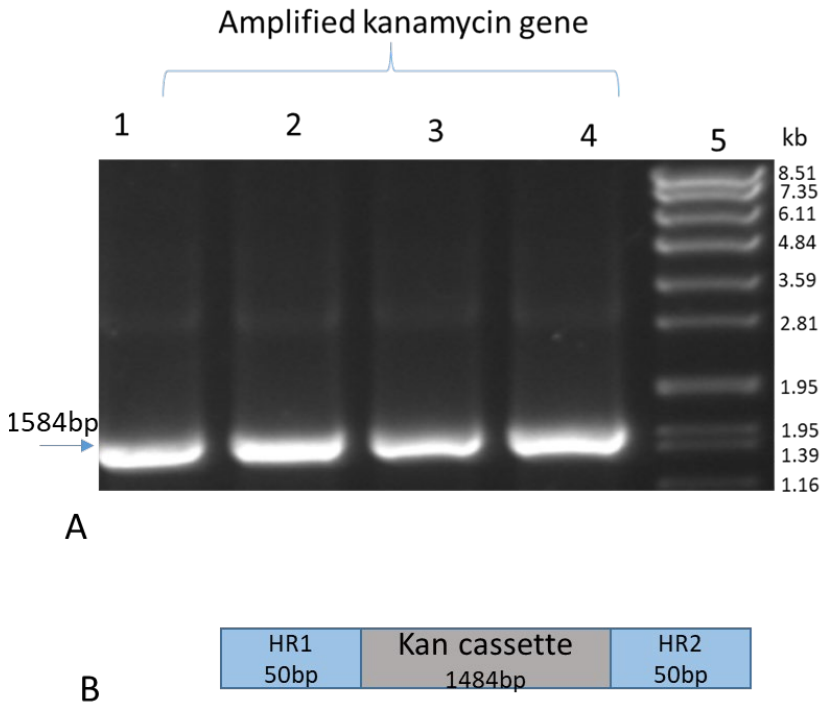


Figure 6.3(A-B): Screening for templates for RT gene disruption in VP.
Kanamycin gene was amplified using primers containing 50 bp homology to Reverse transcriptase (RT) gene. The amplified product was run onto 0.7% agarose gel. Wells 1-4 contain amplified templates. Expected product size 1584 bp (1484 bp of *kan* cassette + 100 bp homologous region to RT gene). Well 5= Spp1 ladder used as size marker and fragments sizes are in base pairs (bp). **B.** Schematic representation of template of RT disruption. The template comprises a 1484 bp sequence from pKD4 plasmid including *Kan* gene and 50 bp homologous region (HR1) upstream and 50 bp homologous region (HR2) downstream RT gene.

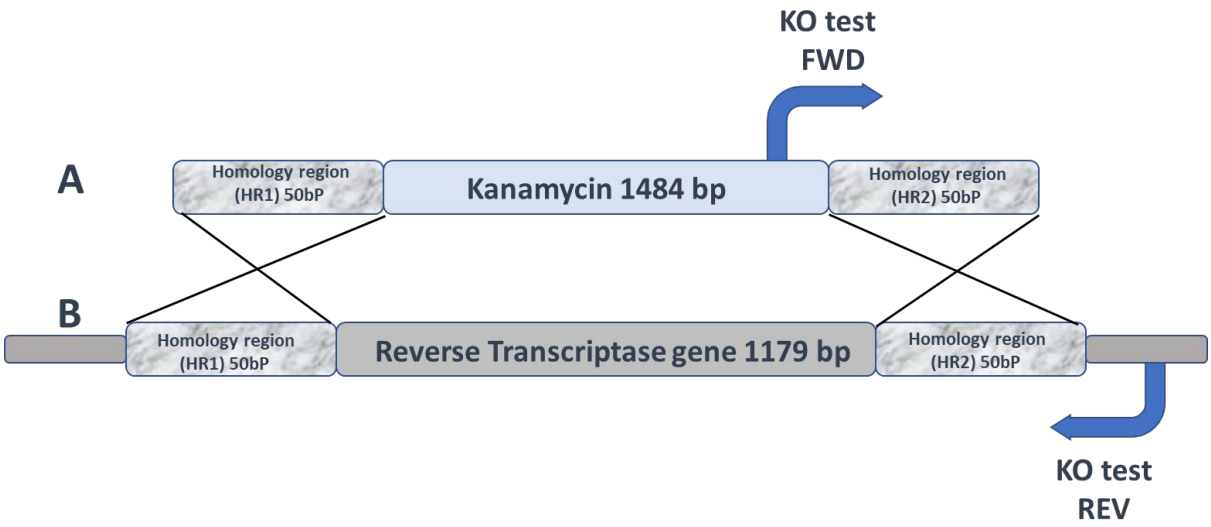


Figure 6.4(A-B): Reverse transcriptase (*RT)* gene disruption by the lambda red recombination method.
A. Representation of the KO template. **B**. Representation of wild-type chromosomal DNA. The RT gene was disrupted by the kanamycin gene. KO test FWD and KO test REV are priming sites for confirmation of disrupted strain.

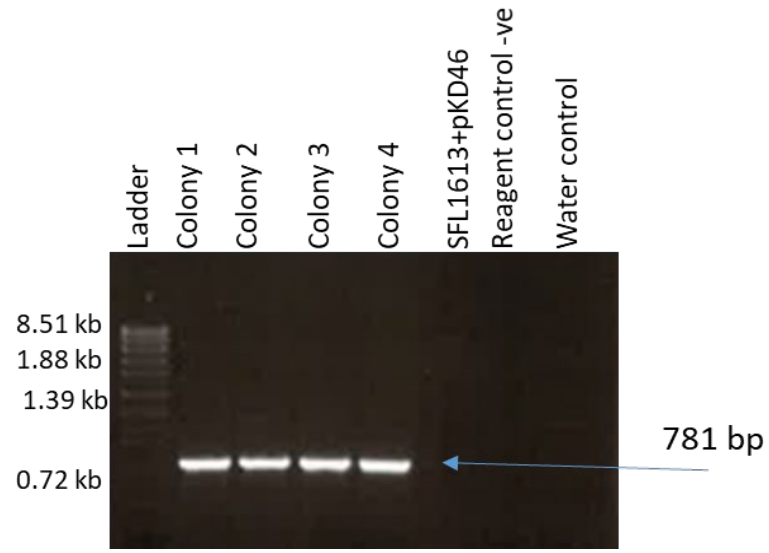


Figure 6.5: KO test Confirmation of Reverse transcriptase gene disruption by PCR screening.
Image of agarose gel (0.7%) showing expected product size ~781 base pairs (kan=274 bp, HR2=50 bp, and downstream RT=457bp) of transformant colonies (1-4). The first well from the left was loaded with SppI ladder. SFL1613 containing pKD46 was used as a negative control. PCR reagents control containing all PCR ingredients except template is shown in last well.

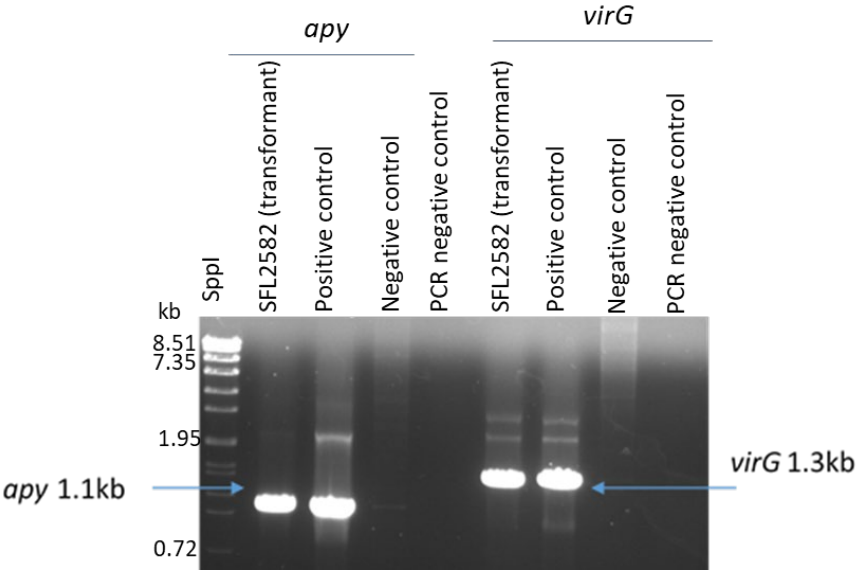


Figure 6.6: Colony PCR for presence of *Apy* and *virG* genes in the transformants (VP tagged with kanamycin gene).
One of the transformants was tested for the presence of *apy* and *virG* via colony PCR. Expected size for *apy* = 1.1 kb and *virG* = 1.3 kb. PCR negative control showed no band. Negative control (SFL1683) with no VP is negative for both *apy* and *virG* genes. Positive control SFL1353 (with VP) showed bands of expected sizes. The first well from the left shows the SppI DNA ladder and fragment sizes in base pairs (bp).

6.2.2.2 Transfer of virulence plasmid in lysogenic strain (SFL1683) of Sf101 via conjugation

Mating experiments were performed to reintroduce the virulence plasmid; the SFL2582 strain containing tagged VP was used as the donor strain and lysogenic strain as a recipient strain. However, after repeated efforts transfer of the large VP was not successful, and to investigate the role of unique genes of Sf101 in the host virulence, another approach was tried.

6.3 Alternate approach to study the role of the unique genes in virulence

The four *orfs* were individually cloned into the pBAD/*Myc*-HisA vector. First, 1c strains were analysed for the presence of VP and absence of the unique *orfs (orf16, orf17, orf41,* and *orf56)* to ensure that these genes were not present in the new 1c strain harbouring VP. MiSeq assembled sequences of serotype 1c strains were used to perform nucleotide blast against nucleotide sequences *orf16/oacB, orf17, orf41* and *orf56*. Subsequently, a 1c strain SFL2456 was selected for the transformation of recombinant plasmids which also contained the VP but without *oacB*/*orf16*. Colony PCR of SFL2456, using gene-specific primers for *apy/virG* and other *orfs* was also performed for the confirmation of VP and *orfs*, respectively. The antibiotic susceptibility test confirmed that SFL2456 was resistant to ampicillin and sensitive to erythromycin. Hence a plasmid (pNV2185) was constructed carrying the erythromycin gene for subsequent cloning of *orf16/oacB, orf17, orf41,* and *orf56*.

6.3.1 Construction of vector pNV2185 (pBAD/Myc-HisA vector with erythromycin resistance (Em^R gene at SphI site)

To create a vector carrying the erythromycin resistance gene, the B2596 strain carrying pNV2132 (containing *oacB* and *erythromycin* genes in pBAD-*Myc* HisA vector) was used to obtain an erythromycin gene. The erythromycin resistance gene was successfully cloned at *Sph*I site into pBAD/*Myc*-HisA vector. A single colony containing insert (pNV2185) was used to make glycerol stock and named B2666.

6.3.2 Cloning of orf17, orf41 and orf56

A series of sticky end ligations was carried out in the cloning experiments. The colony PCR was performed to amplify the *orfs*, using lysogenic strain as a template. The primers, containing *Nco*I and *EcoR*I restriction sites, were designed specifically for each *orf* (Section 2.2). For the vector plasmid, pNV2185 was digested with *Nco*I and *EcoR*I. The ligation reactions were set, and single colonies of each positive clone were used to make glycerol stocks and named B2651 (carrying *orf17* pNV2167), B2652 (carrying *orf41* pNV2168), and B2653 (carrying *orf56* pNV2169) Figure 6.7A-C. The plasmids were also transformed into the 1c strain, SFL2456 to produce recombinant *S. flexneri* strains and single colonies of each were used to make glycerol stocks and named SFL2613, SFL2614, and SFL2615 carrying pNV2167, pNV2168, and pNV2169, respectively. Plasmid pNV2132 carrying *oacB/orf16*, created in 3.2.2.3 was also transformed into SFL2456 strain to produce SFL2604.

To confirm the presence of VP in the transformed *Shigella* strains (SFL2604, SFL2613, SFL2614, and SFL2615), colony PCR using the *apy* and *virG* primers were performed and the results confirmed the presence of VP in each strain as expected (Figure 6.8).

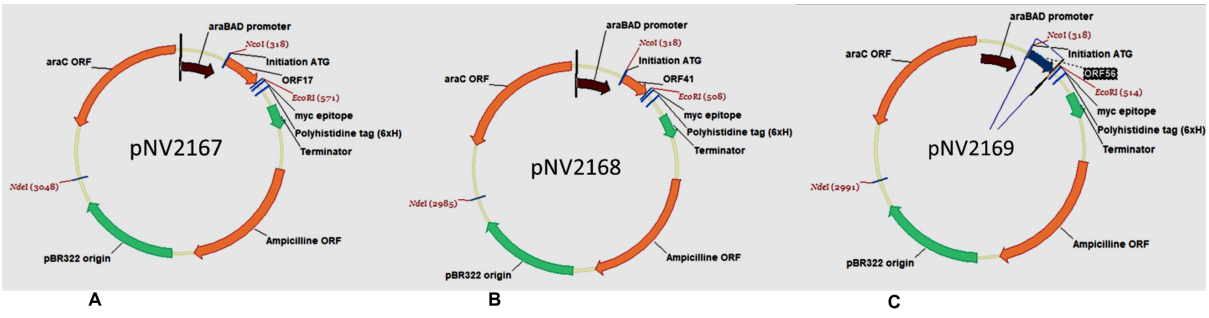


Figure 6.7A-C: Plasmids maps of pNV2167 (*orf17*), pNV2168 (*orf41*), and pNV2169 (*orf56*).
In-frame cloning of all three *orfs* was performed. All the *orfs* were cloned individually into pNV2185 (pBAD/*Myc*-HisA vector cloned with erythromycin resistance gene at *Sph*I site). All three *orfs* (*orf17, orf41,* and *orf56*) were cloned at *Nco*I and *EcoR*I sites to generate pNV2167 (A), pNV2168 (B), and pNV2169 (C) respectively.

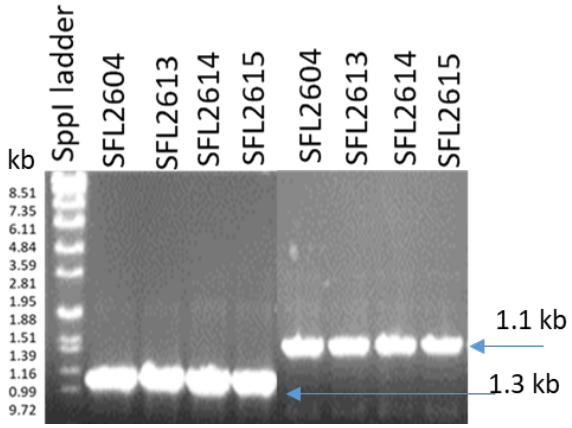


Figure 6.8: Agarose gel electrophoresis for the Confirmation of VP.
Colony PCR for a*py* and *virG* genes was performed for the presence of VP in the transformed strains SFL2604, SFL2613, SFL2614, and SFL2615. All strains showed positive results for the presence of *apy* and *virG* genes. Expected size for *apy* =1.1 kb and *virG* = 1.3 kb. Spp1 ladder was loaded as a molecular size marker in the first well from the left.

6.3.3 Overexpression of Orf 17, Orf41 and Orf56

The plasmids pNV2167 (*carrying orf17*), pNV2168 (*carrying orf41*), and pNV2169 (*carrying orf56*) were introduced into the expression strain TOP10 *E. coli* cells to generate B2654, B2655, and B2656, respectively, to study the roles these *orfs* in the virulence of host strain. This was done to evaluate the ability of the plasmids to express the cloned genes, evidence of expression of proteins the in these *E. coli* strains would predict the successful expression of the cloned genes in the transformed *S. flexneri* strain (SFL2456). In the case of *oacB*, overexpression of OacB in pNV2132, has been discussed in chapter 3 in detail and pNV2132 was transformed into SFL2456 to create *Shigella,* strain SFL2604. In this section, overexpression of Orf17, Orf41 and Orf56 is discussed.

6.3.3.1 Expression of Orf17, Orf41 and Orf56 in B2654, B2655, and B2656

B2654 (TOP10 *E. coli* strain transformed with pNV2167) cultures were grown in LB_{Amp} to an OD_{600} of 0.55 and induced with two concentrations of L-arabinose (0.1%, and 0.2%). The whole-cell lysate samples (Section 2.11.1) were then prepared at 2-, 3-, and 4-hours post induction. The overexpressed Orf17 can be seen in Figure 6.9A, at the expected running size of ~11.7 kDa. In the case of B2655, the maximum detection of overexpressed Orf41 was achieved in cultures when induced at OD_{600} 0.6 with 0.1% and 0.01% of L-arabinose in a sample prepared from whole cell lysate (Figure 6.9B). In the case of B2656 (Orf56), no band was seen in the whole cell lysate sample and membrane protein samples were prepared (Section 2.11.1 and 2.11.2) to perform western blot. Optimum overexpression of Orf56 in pNV2169 was achieved when induced at OD_{600} 0.6 with 0.1% of L-arabinose (Figure 6.9C). In all experiments B2372 (pBAD/*Myc*-HisA empty vector in TOP10 cell) was used as a negative control

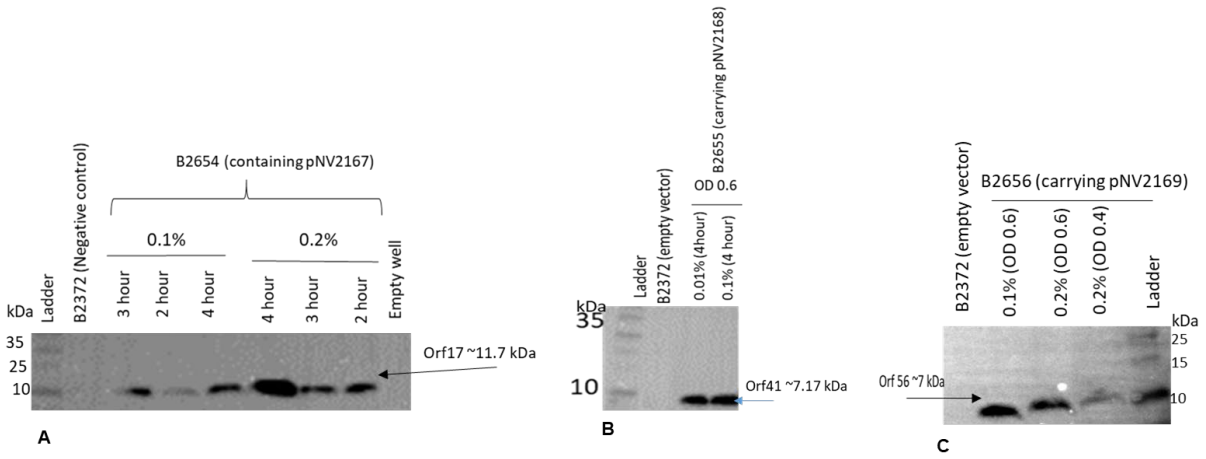


Figure 6.9A-C: Overexpression of Orf17 in B2654 (Top10 *E. coli* strain transformed with pNV2167); Orf41 in B2655 (TOP10 *E. coli* strain transformed with pNV2168) and B2656 (TOP10 *E. coli* strain transformed with pNV2169)

Western blot was carried out using anti-His antibodies in 1/10000 dilution. **A.** Cultures were induced at OD_{600} 0.55 with either 0.1% or 0.2% of L-arabinose. Whole-cell lysate samples were prepared at different time points (from1 to 4 hours). The signals were detected at the expected running size of 11.7 kDa for Orf17. **B.** Cultures of B2655 were induced with 0.01 /0.1% L-arabinose concentrations at an optical density of (OD_{600}) 0.6. After 4 hours of incubation (post-induction), whole-cell lysate samples were prepared, and protein samples were loaded onto an SDS-PAGE gel. The signals were detected at the expected running size of ~7.17 kDa for Orf41. **C.** Cultures of B2656 were induced with 0.1% and 0.2% L-arabinose concentrations at OD_{600} of either 0.4 or 0.6. After 4 hours of induction, membrane proteins were prepared and ~10 µg of protein was loaded onto an SDS-PAGE gel. The signals were detected at the expected running size of ~7 kDa for Orf56. All the samples were prepared in 2X loading buffer. In all the blots no signals were detected in the lane of the negative control of B2372 (TOP 10 *E. coli* strain transformed with pBAD/*Myc*-HisA vector).

6.3.3.3 RT-PCR of SFL2604, SFL2613, SFL2614 and SFL2615

The plasmids created in Section 6.3.2, could not be used for overexpression purpose in *S. flexneri* because the *Shigella* system does not support the induction of cloned genes. Hence, RT-PCR was performed, and RNA was extracted from SFL2604 (pNV2132), SFL2613 (pNV2167), SFL2614 (pNV2168), and SFL2615 (pNV2169) using the same protocol as described in section 6.2.1.1 All the transformed strains showed positive results for the amplification of all four *orfs* when amplified with *Taq* polymerase using gene-specific primers (Figure 6.10) indicating expression of all the *orfs* in the introduced plasmids (pNV2167, pNV2168 and pNV2169).

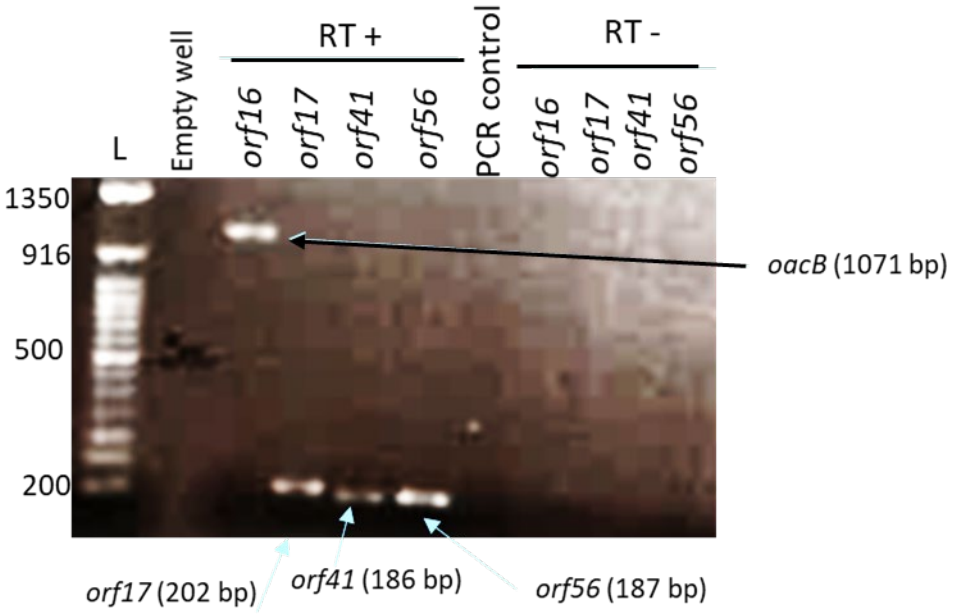


Figure 6.10: RT-PCR: Expression of Sf101 unique genes in the transformed *Shigella* strains. Total RNA of transformed strains of *S. flexneri* (SFL2604, SFL2613, SFL2614, and SFL2615) were used to produce randomly primed cDNA. Gene-specific primers were used to amplify the gene of interest from the cDNA. Expected size bands of *oacB* = 1071 bp, orf*17*=202 bp, *orf41* =186 bp and *orf56*= 187 bp are seen in the respective lanes. Lane 1 (left) loaded with 50 base pairs DNA marker, RT+: reaction with reverse transcriptase, RT-: reaction without reverse transcriptase (negative control for DNA contamination).

6.4 *in vivo* virulence assays using *Caenorhabditis elegans*

The natural, and only, hosts for *S. flexneri* include humans and primates, which makes it difficult to study the pathogenesis of *Shigella* in the laboratory. A limited number of *in vivo* animal models, such as murine pulmonary and guinea pig keratoconjunctivitis have been used to study the pathogenesis of *S. flexneri*. Nevertheless, these models lack clinical relevance because the site of infection is different from that of S. *flexneri* infection in humans. Recently a soil-dwelling nematode has been employed as a model to study the infection process in many enteric pathogens due to its similarities with human intestinal cells [222]. Also, the *C. elegans* immune system shares some characteristics with the human innate immune system which is suggestive of utilisation of similar response mechanism to bacterial infection by *C. elegans* as humans [223]. *C. elegans* were used as an animal model to study virulence of different *S. flexneri* strains the results showed that the virulent strains of *Shigella* accumulate in the gut of the nematode, whereas, avirulent strains were digested by the worms [224].. Similarly, in another study conducted in 2019 by Somasiri P. *et al*., *C. elegans* were used as an animal model to determine the differentially expressed genes in *S. flexneri* infected *C. elegans* strains [225]. Additionally, C. elegans transparent anatomical structure, quick regeneration time and economical suitability make it an ideal animal model to study the pathogenesis of several pathogens including *Shigella*, *Pseudomonas aeruginosa*, *Serratia marcescens* and others [226, 227]. Consequently, *C. elegans* was used to study the role of unknown genes in the virulence of *Shigella*.

Two assays- the bacterial accumulation assay and the liquid killing assay- were used in this study to help understand the role of the unique genes (*orf16*/*oacB, orf17, orf41* and *orf56*) in the virulence of *S. flexneri*. In both assays, OP50 *E. coli,* which is considered a standard food source for the worms in the laboratory was used as the negative control.

6.4.1 Bacterial Accumulation and liquid killing assays

Synchronized growth of wild-type (N2) nematodes to the L4 larval stage was achieved and the worms were then fed with SFL2456 (wild type), SFL2626 (SFL2456 + empty vector), SFL2604 (SFL2456+ plasmid expressing *orf16/oacB*), SFL2613 (SFL2456+ plasmid expressing *orf17),* SFL2614 (SFL2456+ plasmid expressing *orf41*), or SFL2615 (SFL2456 + plasmid expressing *orf*56) strains. The results showed that the accumulation *Shigella* strains (CFU/worm) in worms and the killing rate in the killing assay, was similar to the wild-type strain in all the strains. The CFU/nematode for strain SFL2604 is shown higher than the wild type in Figure 6.11A but statistical showed no significant difference and p value recorded was 0.3133. Moreover, no significant statistical differences were found among other strains across assays when compared with wild type (Figure 6.11 A and B).

These observations would suggest that *orf16/oacB, orf17, orf41* and *orf56* might not be required for the virulence of *S. flexneri* in *C. elegans*.

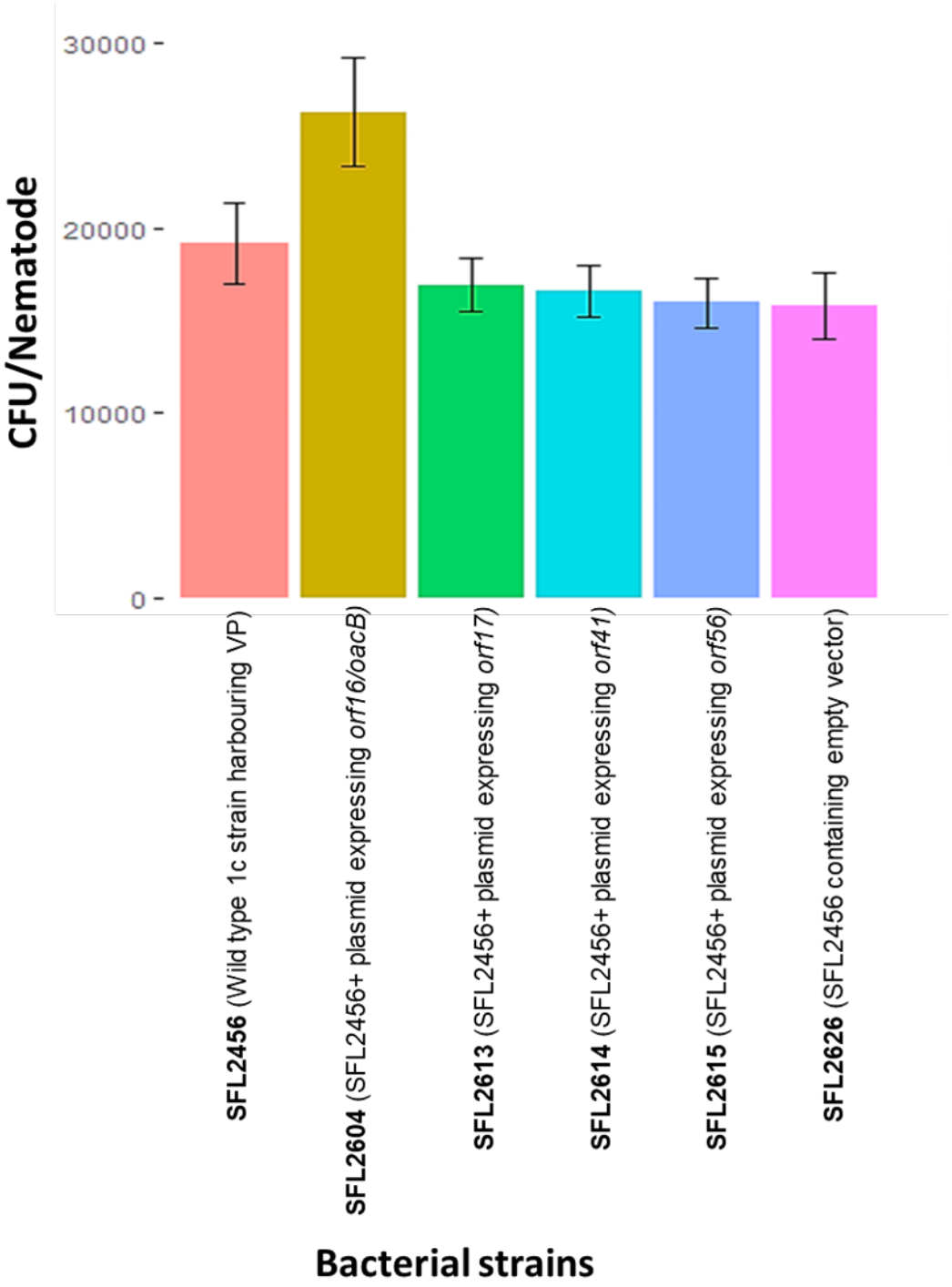


Figure 6.11 A: *C. elegans* bacterial accumulation assay.
Young adult hermaphrodite nematodes were fed with SFL2456 (wild type 1c strain), SFL2626 (SFL2456 + vector), SFL2604 (SFL2456+ plasmid expressing *orf16/oacB*), SFL2613 (SFL2456+ plasmid expressing *orf17),* SFL2614 (SFL2456+ plasmid expressing *orf41*), and SFL2615 (SFL2456+ plasmid expressing *orf56*) strains for 24 hours. 20 worms were picked from each plate and disrupted mechanically using silica carbide beads to release accumulated bacteria. From each lysate, appropriate dilution was plated on LB agar plates, and colonies were counted to calculate the number of *S. flexneri* cells internalized in each nematode. The analysis was based on six experimental repeats. No statistically significant differences were observed at p. value < 0.05 (t-test). Error bars represent Standard Error (SE) of the means.

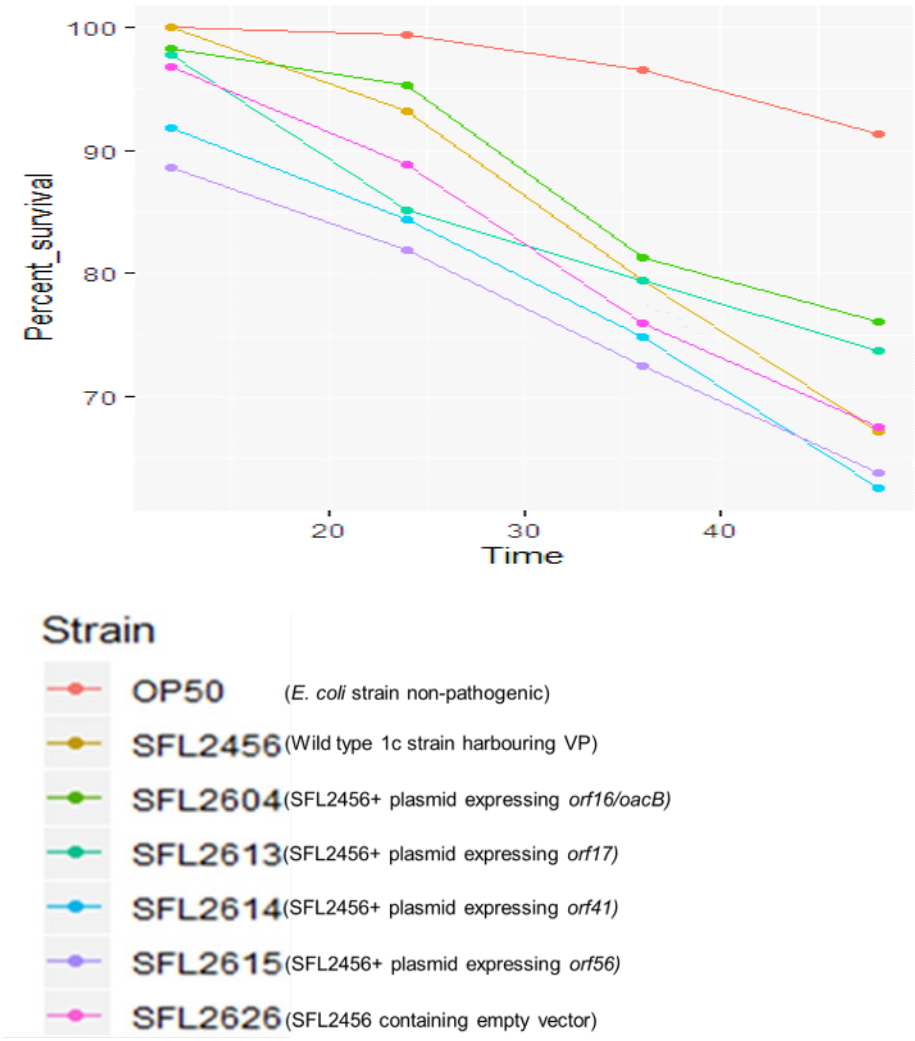


Figure 6.11 B: *C. elegans* liquid killing assay.
The graph represents the result of *C. elegans* liquid killing assays. Log phase cultures of SFL 2456 (wild type 1c strain), SFL2626 (vector only), SFL2604 (SFL2456+ plasmid expressing *orf 16/oacB*), SFL2613 (SFL2456+ plasmid expressing *orf 17),* SFL2614 (SFL2456+ plasmid expressing *orf 41*), and SFL2615 (SFL2456+ plasmid expressing *orf56*) grown at 37 °C for the expression of virulence plasmid was used to treat synchronized L4 nematodes. Post-infection, the worms were observed for 48 hours, and survival was scored every 12 hours. The results presented here are the mean of six independent repeats. The worms treated with any of the *Shigella* strains show a reduced survival rate compared with the OP50 *E. coli* strain (non-pathogenic). Log rank test was carried out to assess the differences between survival of *C. elegans* exposed to various bacterial mutants.

6.5 *in vitro* invasion assays

Invasion of the colonic and rectal epithelium is the pre-requisite of *S. flexneri* infection. To determine the role of any gene(s) in the invading ability of *S. flexneri*, an invasion assay using HeLa/Baby hamster kidney (BHK) can be performed. This assay measures the number of invading bacteria for a 2-4-hour infection process. Post-infection the cells can be stained using

Giemsa stain for microscopy and counting. To determine the role of *oacB, orf17, orf41,* and *orf56* in the invasion of epithelial cells, HeLa cells were used in this study. HeLa cells were infected for 2 hours, with SFL 2456 (wild type 1c strain), SFL2626 (SFL2456 +vector only), SFL2604 (SFL2456+ plasmid expressing *orf 16/oacB*), SFL2613 (SFL2456+ plasmid expressing *orf 17),* SFL2614 (SFL2456+ plasmid expressing *orf41*), or SFL2615 (SFL2456+ plasmid expressing *orf56*). Approximately 300 HeLa cells were scored for each experimental repeat (total 2 repeats) (Figure 6.14 A-F).

Statistical analysis showed a significant decrease in the invading ability in the strains SFL2613 and SFL2615 compared with the wild-type strain, SFL2456 (Figure 6.15) and suggest that orf*17 and orf56* might play a role in decreasing the invasive potential of the host strain. However, due to time constraints only preliminary findings are presented here for invasion assay and further repeats are required to confirm the obtained results.

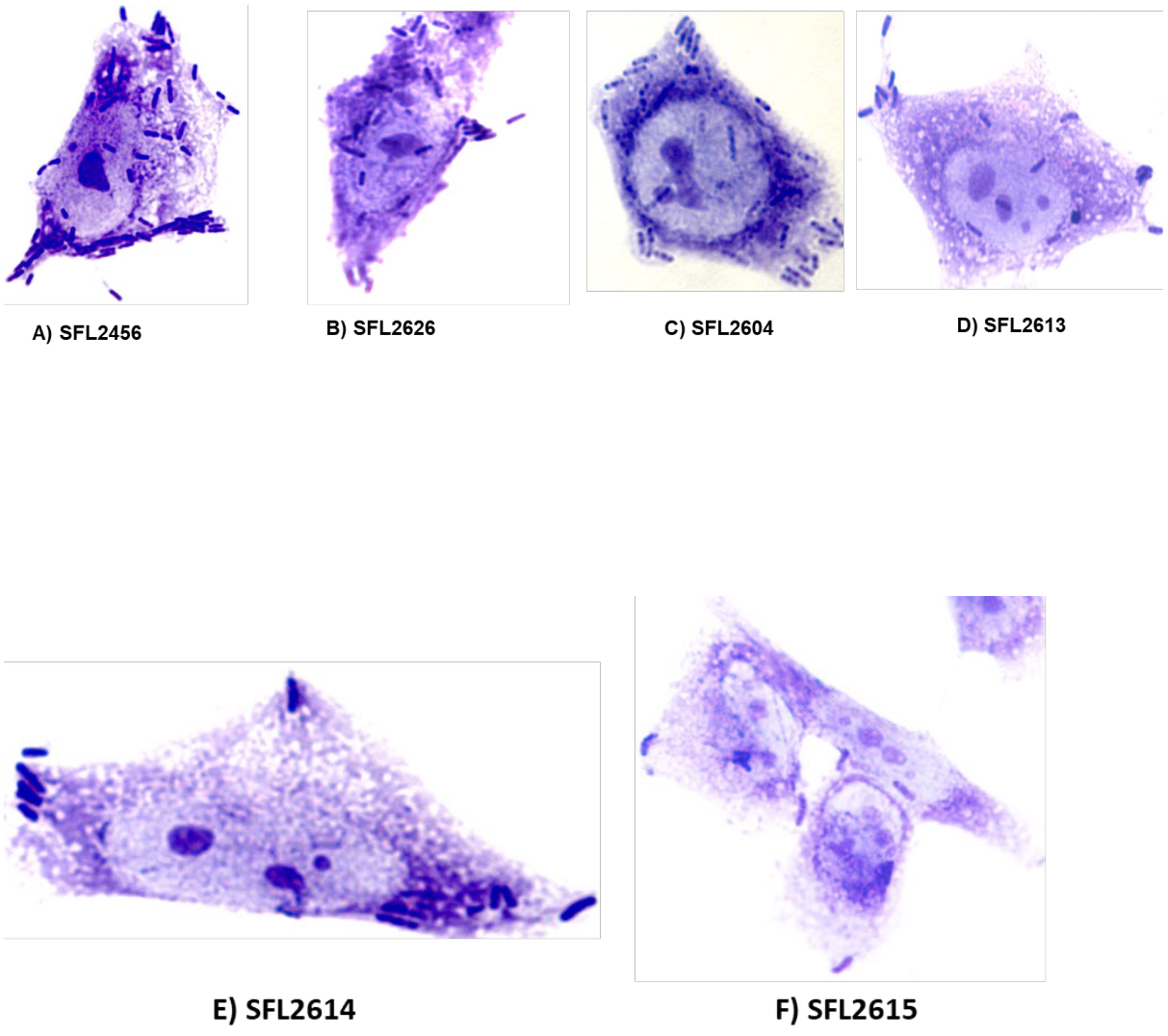


Figure 6.12 A-F: Microscopic images of HeLa cells infected with *S. flexneri* strains.
Each well of a tissue culture plate containing a monolayer of HeLa cells with 0.1 ml of 1x 10^8 CFU/ml log-phase cultures were infected for 2 hours at 37°C /5% CO_2. Post-infection monolayers were stained with Giemsa stain, and cells were examined under a Leica microscope and rod-shaped bacteria were observed. The bacterial strains were **A**. SFL2456 (WT), **B**. SFL2626 (SFL2456+ empty vector), **C**. SFL2604 (*oacB*), **D**. SFL2613 (*orf17*), **E**. SFL2613 (*orf41*), and **F**. SFL2615 (*orf56*). In all the images *Shigella* bacilli invading HeLa cells are visible.

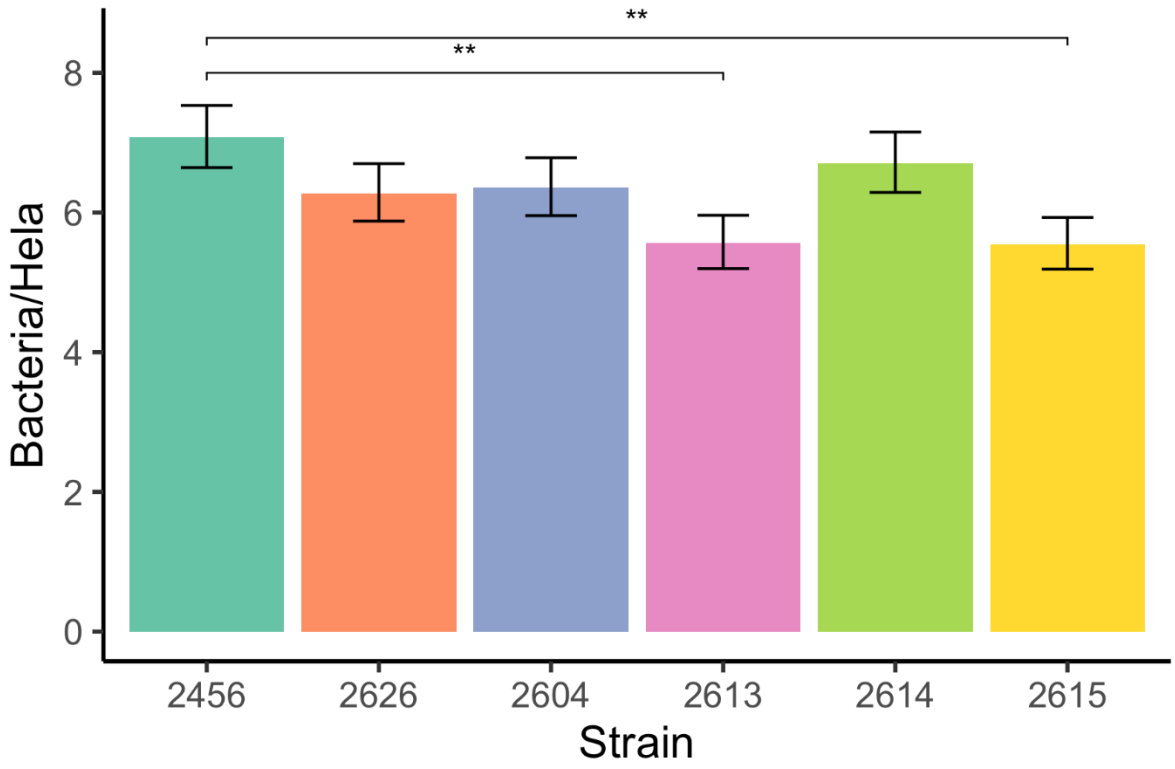


Figure 6.13: Invasion of HeLa cell monolayers by *S. flexneri*.
HeLa cells were infected for 2 hours at 37°C /5% CO_2 with SFL 2456 (wild type 1c strain), SFL2626 (SFL2456+vector only), SFL2604(SFL2456+ plasmid expressing *orf16/oacB*), SFL2613 SFL2456+ plasmid expressing *orf17),* SFL2614 (SFL2456+ plasmid expressing *orf41*) and SFL2615 (SFL2456+ plasmid expressing *orf56*). The number of internalized bacteria per HeLa cells is represented by Y-axis whereas X-axis represents bacterial strains. Two independent repeats were performed, and results were obtained by scoring 300 cells per repeat. Two strains SFL2613 (p. value = 0.0040) and SFL2615 (p. value = 0.0031) showed reduced ability to invade epithelial cells when compared with SFL2456/SFL2626. The statistical differences in the invasion were determined using a t-test. Asterisks denote a significant statistical difference ($p<0.05$ was considered statistically significant). Error bars represent standard errors.

6.6 Conclusion

In this study, the role of four novel genes of Sf101 phage in the virulence of *S. flexneri* was investigated. The expression of unique genes (*orf16/oacB, orf17, orf41,* and *orf56*) was confirmed in the lysogen (SFL1683) by RT-PCR. Due to the absence of VP in SFL1683, efforts were made to restore VP by making a Sf101 lysogen in serotype Y strain or conjugation between VP donor strain Y394 (carrying VP) and SFL1683. However, due to the contamination of high titre Sf101 phage lysate, lysogen could not be made; and non-conjugative nature of the virulence plasmid these strategies could not be employed. Hence, it was decided to investigate the role of novel genes of Sf101 phage in another 1c strain (SFL2456) containing VP and lacking Sf101 unique *orfs*. Three genes *orf17, orf41,* and orf*56* were individually cloned into pNV2185 (pBAD/*Myc*-HisA carrying the erythromycin gene) to create three plasmids. Before transforming plasmids into the 1c strain SFL2456 all the plasmids were transformed into TOP10 *E. coli* cells to perform protein expression experiments for confirmation of translation of cloned genes into functional proteins (presumably). After achieving successful overexpression of all the Orfs, the three plasmids including *oacB* were transformed into the selected S. *flexneri* 1c strain SFL2456 to create four *S. flexneri* recombinant strains. Following transformation, RT-PCR was performed on total RNA extracted from all four strains for confirmation of expression of *oacB, orf17, orf41* or *orf56,* in the newly created *Shigella* strains. The role of these genes in the host virulence was determined, by testing against the wild-type strain using invasion assay and *C. elegans* assays. Results of the liquid killing assay and bacterial accumulation assays showed no difference in virulence in the strains when compared with the wild-type strain. However, the results of the invasion assay revealed a possible reduction in invading ability in the presence of *orf17* and *orf56*, these preliminary findings need further confirmation.

Chapter 7: General discussion

7.1 Optimization of OacB overexpression

Chapter 3 presents results based on the creation of a two-dimensional model of OacB, cloning, and optimization of expression of OacB. Experiments were carried out to achieve the optimal expression of OacB in two expression vectors (pBAD/*Myc*-His A and pFLAG-CTC). In this chapter, various parameters were tested including the concentration of inducer(s), the temperature of incubation, the optical density of the cultures at the time of induction, and the duration of incubation post-induction. Moreover, after achieving OacB expression, methods to prepare proteins sample for analysis on SDS-PAGE gel were also modified.

The overexpression of integral membrane proteins is a daunting problem compared to soluble proteins, and this is one of the reasons that membrane proteins represent only 1.7% of atomic structures present in the Protein Data Bank [28]. Optimization of protein expression involves multiple parameters, such as the cell density of the cells for induction, the concentration of inducer, duration, and temperature of incubation [184, 228]. Moreover, after setting optimal conditions for expression the preparation of samples for SDS-PAGE and analysis on western blot/Coomassie staining is also a critical step to get the best results [190]. However, even after achieving the desired results for expression, many problems can arise during the western blot procedure and that could lead to an unexpected result, for example, the absence of any band on the blot, presence of unusual bands, weak signals, or high background [190]. During this study, OacB expression was found to be very challenging, and initially, the expression could not be achieved in either of the vectors. However, several conventional parameters were optimized and are discussed below.

7.1.1 Selection of expression vectors

For the expression of OacB two expression vectors were selected, each with different features. In the pBAD/*Myc*-HisA vector, the expression of the in-frame cloned gene is under the control of the *ara*BAD promoter (P_{BAD}), and transcription can be turned 'ON' by the addition of L-arabinose which activates the promoter. The vector has a C-terminal *Myc* epitope

(EQKLSEEDL) followed by a His-tag (-HHHHHH) [229]. Whereas, in the pFLAG-CTC system, the expression of the inserted gene is controlled by an inducible *tac* promoter whose activity is regulated by *lacO* and *lacI.* The addition of inducer IPTG (a molecular mimic of allolactose) allows optimum interaction between RNA polymerase and tac promoter resulting in the expression of the gene of interest. The vector carries eight amino acid (DYKDDDDK) C-terminal FLAG tag. In both vectors, the tagged protein can be detected in western blot analysis using a tag-specific antibody. Furthermore, the peptide tags in both vectors allow affinity purification using resins [184].

7.1.2 Optimization of optical density (OD_{600}) at the time of induction

The density of growing cells at the time of induction plays an important role in the optimization process of heterologous protein expression [230]. Optical densities between 0.3-0.8 were considered as best to induce protein expression. Induction at the exponential growth phase yields a high amount of overexpressed protein, as during this phase the entire cell is wired for growth and the population of growing cells is uniform. By contrast, induction at an early log phase or just after inoculation may affect bacterial growth due to the overexpression of recombinant protein and thus reduces the yield [184]. Similarly, if induction is carried out after passing the log phase there will be a reduction in protein yield due to depletion of nutrients and building up of wastes in the medium [231].

Therefore, in this study, expression was induced at multiple optical densities to test which condition yields the maximum expression of OacB. For the expression of OacB in pFLAG-CTC and pBAD/*Myc*-HisA vectors, cells were induced at various optical densities ranging from 0.3 to 0.6. The optimal expression of OacB was achieved in pNV2132 when induced at either OD_{600} 0.4 or 0.6. In the case of pNV2110 (pFLAG-CTC), the right size band of OacB could not be detected in any of the experiments when induced at OD_{600} 0.4 or 0.6. However, OacB aggregates were seen in some experiments (Section 3.6.1.4) indicating the expression of OacB.

7.1.3 Optimization of concentration of inducer

The concentration of the inducer is also one of the parameters which needed to be optimized [232]. To achieve the best results, the induction concentration recommended for IPTG is 50-100 μM [233]. However, at high concentrations IPTG can be detrimental to the host cells and lead to reduced protein production [234]. For example, a higher concentration of IPTG (1 mM) for inducing LigB in *E. coli* BL21 strain was found to negatively affect cell growth and hence decreased protein yield and they suggested 0.1 mM concentration should be used [235]. In a study arabinose-inducibility was considered better than IPTG for *lac*-derived promoter systems [236] On the other hand, insufficient inducer concentration may not titrate the repressor molecule completely and result in low protein yield. Hence, achieving the right balance between inducer concentration and protein yield can be a troublesome task.

Two inducers L-arabinose and IPTG were used for the induction of OacB expression in pBAD/*Myc*-HisA and pFLAG-CTC vectors, respectively. In pNV2132 and pNV2139 concentrations of L-arabinose 0.02%-0.2% were considered sufficient for protein induction and in the case pFLAG-CTC vector, various IPTG concentrations (0.5 mM-1 mM) were used. The expression of positive control (Opt) was achieved at 0.5-1 mM concentrations of IPTG. As there was a problem of aggregate formation during OacB expression, the optimal IPTG concentration could not be quantified. However, the presence of signals in blots of pNV2110 was indicative of OacB expression. The L-arabinose minimal concentration required for the induction was successfully optimised for pNV2132 and pNV2139 in these experiments.

7.1.4 Optimization of incubation temperature after induction

In all the OacB expression experiments 37°C temperature was confirmed as the finest incubation temperature for protein production. *E. coli* cells have a maximum growth rate at 37-39°C; however, in some experiments of pNV2132 (pBAD/*Myc*-HisA containing *oacB*) a lower incubation temperature (18°C) was also tested during the optimization process. It was done because higher temperatures generally favour aggregate formation due to the dependency of hydrophobic interactions on temperature [184], and we reasoned that lower temperatures

might reduce aggregation of OacB. Additionally, the formation of inclusion bodies results from the disturbance of the equilibrium between protein aggregation and solubilization. In a study, the soluble yield of cytoplasmic $_{Fab}$ fragments was enhanced 10-fold when induced and grown at 21 °C instead of 37 °C [237]. Hence optimization of incubation temperature plays a significant role in the protein expression process.

7.1.5 Heating membrane protein samples caused aggregate formation.

Protein samples are usually heated before loading onto SDS-PAGE gel to denature the protein and allow even coating with SDS detergent [238]. Heating protein samples allow to denature high order structures and ensures that the negative charge of amino acids is not neutralised which lets migration of protein in the electric field [190]. In this study, the control membrane protein Opt ran well on SDS-PAGE gels after heating at 100° C for 10 minutes. However, OacB was found to aggregate upon heating, making it impossible to observe OacB monomers upon western blotting. Similar results were observed in a study conducted on transmembrane iron transporter protein divalent metal transporter 1 (DMT1) and ferroportin (1 Fpn1) [239]. In that study, membrane protein samples were prepared in two ways and compared to determine the effect of heating and not heating on samples. Results revealed that when proteins were heated for 5 minutes at 95 °C protein aggregates were formed, and aggregated protein smears were observed upon western blotting (like OacB). Proteins were found to be stuck on top of the separation gel and the study concluded that heating was the main reason for aggregate formation [239]. Another study conducted by Yi-Nung Lee *et al*., (2005) while working on SARS-CoV membrane protein, also found that boiling treatment (100°C for 10 minutes) caused aggregation of the protein and observed smears upon western blotting. They found that spike and nucleocapsid proteins were detectable in western blots using the same denaturing conditions [240].

Hence, these studies and my own experiments suggest that boiling the sample caused aggregation of OacB and this explained the diffuse, high mass bands observed in most western blots. Moreover, this suggests that the expression of OacB was successfully achieved

in most of the conditions tested in this chapter, and that aggregate formation was only due to heating in the final step of sample preparation. We can therefore conclude that heating OacB is not appropriate to prepare samples for SDS-PAGE analysis.

Finally, the overexpression of cysteine-less OacB was also attempted in similar conditions to the wild-type OacB. Results revealed successful expression of cysteine-less OacB in pNV2139. To carry out SCAM, Single-cysteine variants of OacB were also created in this chapter and these mutants can be used in future studies to perform SCAM experiments for the verification of orientation of loops in OacB.

7.2 *O*-acetylation of O-antigen by OacB

In chapter four of this thesis, site-directed mutagenesis (SDM) of selected residues in OacB was performed to determine their role in *O*-acetylation of O-antigen of serotype 1c of *S. flexneri*. SDM has been accepted as a promising tool to probe the crucial sites in a protein. Several critical and non-essential residues were identified in the functionality assay using serotype-specific antiserum. For functionality assays, 3/4- *O*-acetylation specific antiserum prepared in this study (Section 2.10) was used.

7.2.1 Is there any other protein involved in the O-acetylation of O-antigen with OacB?

The findings and observations of this study and other acetyltransferases 3 (AT3) domain-containing proteins culminated in developing a hypothetical model of O-acetylation by OacB. OacB is a bi-functional protein that is responsible for both the transfer and translocation of the acetyl group to the acceptor substrate in *O*-antigen and does not require any other protein to facilitate this process. This mechanism agrees with other enzymes containing only the AT3 domain in other bacteria. For example, WecH, a single domain enzyme without SGNH C-terminal domain in *E. coli* encoded by *Yiah* gene acetylates *Enterobacterial* common antigen (ECA) polysaccharide, without the help of any other protein [241]. Similarly, trehalose *Corynebacterium diphtheriae* (TmaT) acetylates mycolyl moiety of trehalose corynomycolates

(TMCM) and facilitates periplasmic export of TMCM [242]. Furthermore, a plant pathogen, of *Xanthomonas campestris,* two inner membrane acetyltransferases GumG and GumF acetylate mannose moieties independently and like OacB lack C-terminal SGNH domain [243]. Other notable protein, containing AT3 domain only responsible for *O*-acetylation is the Wbak protein from *Salmonella enetrica* group E which acetylates Gal residue of the subgroup E1 O-antigen [244]. In another study, WciG was identified as *O*-acetyltransferase when two trivial mutations caused inactivation of WciG and resulted in the production of the different capsules in serotype 35D of *Streptococcus pneumoniae* [245]. This is yet to be discovered for all these acetyltransferases that what is the source of acetyl moiety and how AT3 only domain-containing proteins perform *O*-acetylation without the aid of any other protein. However, in *Neisseria gonorrhea* O-acetylation takes place with the help of two proteins PatA and PatB and genes encoding both proteins can be found on the chromosome next to each other [246]. PatA is an integral membrane protein that translocates acetyl moiety to the periplasmic side and PatB then transfers the acetyl moiety to the designated substrate on peptidoglycan [246].

Similarly, OafA enzyme of *S. enterica* modifies O-antigen by the coordination of two fused proteins i.e., AT3 and SGNH. When transmembrane AT3 domain receives acetyl group it translocates it to its partner periplasmic domain for acetylation of the designated residue of the O-antigen [110]. There are remarkable similarities between OacB and OafA in terms of their interacting critical residues with the acetyl group. In both enzymes arginine residues are found as first point of contact with the acetyl group. However, both enzymes lack crystal structures, and these findings need to be confirmed further. Furthermore, both enzymes share some important motifs in common for example RxxR and DGxRGxLAxxVxxHH motif [105, 110]. One of the most significant differences between two enzymes are their mechanism of actions; OafA translocates and transfers acetyl group by the coordination of two domains whereas OacB performs O-antigen acetylation alone.

To explore the possibility of the involvement of another protein during the O-acetylation process by OacB. The whole-genome sequence of lysogenic strain (SFL 1683) of Sf101

bacteriophage was analysed to determine the presence of any other coordinating protein surrounding OacB. However, no other interfering protein was found on the locus near *oacB* gene. Likewise, the origin of Oac and other acetyltransferases of *S. flexneri* is of bacteriophage origin and no other coordinating gene was present on the loci of these that could help in the O-acetylation process.

7.2.2 Hypothetical model of O-antigen O-acetylation by OacB

As mentioned above, the source of acetyl CoA in a reaction catalyzed by OacB and other acetyltransferases is not yet known [110, 198]. It is thought that it is presumably made available from the donor in the cytoplasmic pool. It has been reported that the *O*-acetylation of peptidoglycan and O-antigen by OatA (in *Staphylococcus aureus*) and OafB (in *Salmonella entrica*) takes place in periplasm as a maturation event. Hence, OacB mediated O-antigen modification is also most likely to take place in the periplasm [110, 214]. The process involves the translocation of the acetyl group, acquired from the cytoplasmic source to the periplasm, and finally transfer to a designated substrate i.e., Rha III in the periplasm.

Considering the structural and mutagenesis data presented in Chapter 4, we propose that the O-antigen *O*-acetylation process by OacB begins when R116 residue in the cytoplasmic loop 4 interacts with the available acetate (active center 1) and processes it with the help of His58 in TMII (active center 2). A critical residue F98 in TM III helps in the stabilization of the active center 1. From this point, the bound acetyl group is transferred across the membrane to the periplasmic loop 5 conserved residue S146 (active center 3). The serine residue might serve as a nucleophile to attack the carbonyl center of the substrate (bound acetyl donor). This active site mimics the oxyanion hole of other acetyltransferases like OafA and OatA. Following this S146 catalyzes the addition of the translocated acetyl group to the Rha III residue of the O-antigen within LPS to generate *O*-acetylated Rha III with the collapse of the oxyanion hole and release of the acetyl donor and CoA would recycle back to the cytoplasm (Figure 7.1). Further research into determining the exact location of critical amino acids with the help of the 3D crystal structure of OacB is required to support the proposed model.

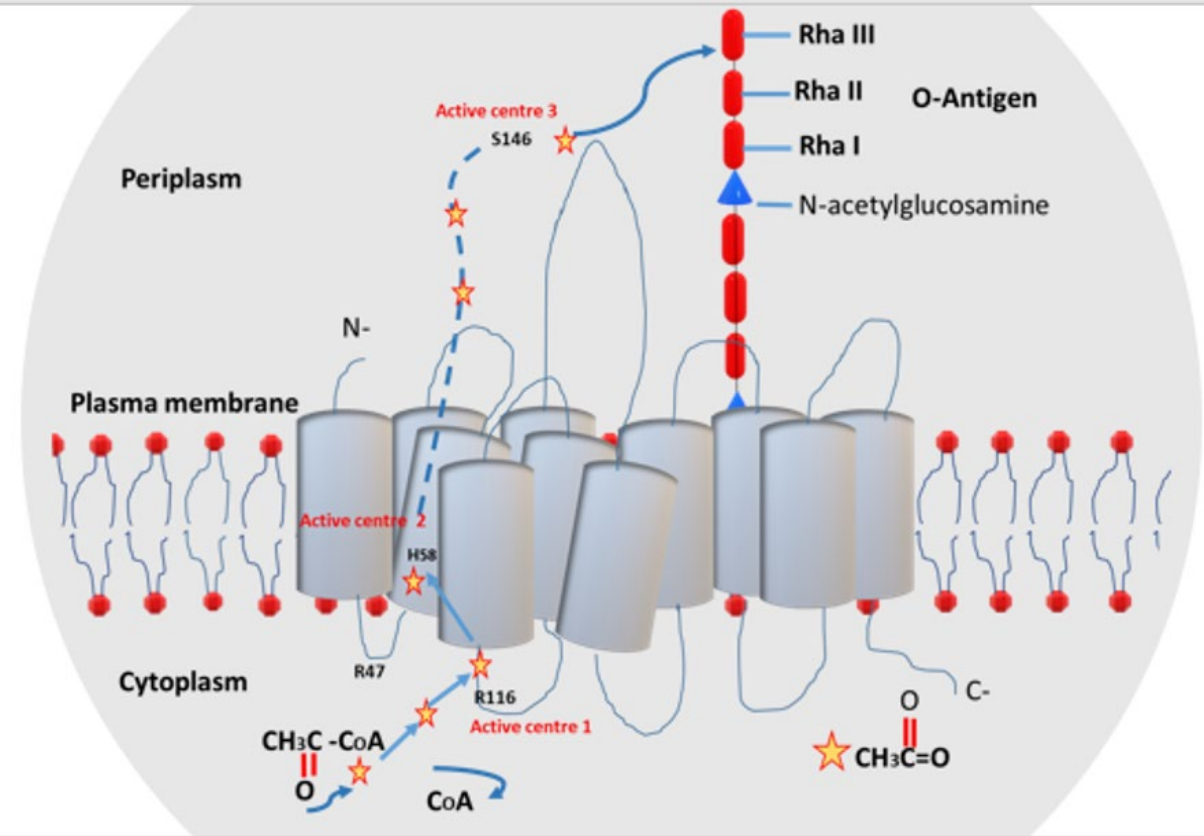


Figure 7.1: O-acetylation model by OacB:
Hypothetical model of mechanism of action of OacB. Acetyl CoA from cytoplasmic pool would serve as the source of acetate for the translocation by integral membrane protein OacB a) OacB catalytic residue arginine 116 (R116) in cytoplasmic loop four (active site 1) would interact with acetyl CoA and acquire the acetyl group which then translocate through the membrane after interacting with residue histidine 58 (H58) in TMII (active site 2). Conserved Serine 146 (S146) in periplasmic loop 5 (active site 3) mediates transfer of acetate to Rha III sugar of the O-antigen at the periplasmic site.

[Rha, Rhamnose; N-=amino/NH_2- terminus; C-= carboxyl/COOH- terminus]

7.3. oacB gene localized to two genomic locations in the strains of serotypes 1c, 1a, 1b, and Y

Serotype 1c of *S. flexneri* represents a unique O-antigen structure due to the presence of two glucosyl groups at GluNAc. The two glucosyl groups were introduced by two phages SfI and Sf1C and resulted in O-antigen modification [193, 220]. In the presence of two phages in serotype 1c strains, a third phage Sf101 was inserted, and this resulted in the O-antigen modification due to the *oacB*- mediated acetylation of Rha III of O-antigen. Infection of a single bacterium by multiple phages promotes the genetic reshuffling, evolution of the pathogenic

host and leads to the emergence of new serotypes due to the combination of homologous modules/sequences between diverse phages [247-249]. The chromosome of serotype 1c strains demonstrates the best example of lysogenization of a single host by multiple phages. In serotype 1c strain, Y394, it was found that the two cryptic phage regions (SfI and Sf1C) present 2 Mb apart from each other and the corresponding phages SfI and SfIC integrated at *tRNA-thr* and *tRNA-pro* genes, respectively [116]. The *oacB* gene was found to be present adjacent to SfI region and flanked by IS elements [106]. However, in another study the *oacB* gene was found upstream of the *sbcB* gene and the integration of the *oacB* encoding phage (Sf101) within *sbcB* gene was confirmed [105]. The presence of the *oacB* gene at two different locations in 1c strains has raised questions for the origin of *oacB* in these strains. The results presented in this study answered these questions by providing an overview of both the regions (*sbcB /adrA*) carrying *oacB* in *S. flexneri* chromosomes. Moreover, it was recently found by Parajuli *et al*., 2020, that all the serotype 1c strains were derived from two ancestral serotype 1a clones and contained cryptic SfI regions (*gtrI*) [221].

7.3.1 A 6 kb region near oacB gene integration site at sbcB locus in lysogen

In a typical event, when a temperate bacteriophage integrates into the host chromosome, the phage integrase and O-antigen modification genes once present next to each other on the phage chromosome, move to opposite ends of phage DNA yet being transcribed in the same direction [121, 250]. For the integration of a phage to happen, three factors play a role: phage-encoded recombinases (integrase/excionase), conserved phage *attR* and *attL,* and the host-encoded IHF (integration host factor) [251]. The integration of phage generates *attL* and *attR* sites, which are separated by a complete phage genome. When the Sf101 phage integrated into its host chromosome, the insertion resulted in the creation of *attL*/*attR,* and the presence of *attL* region was reported by Jakhetia *et al*., 2014 [105]. However, the *attR* region could not be analysed in that study due to the unavailability of WGS of the lysogenic strain. In the current study using WGS of Sf101 lysogen, sequences next to *attL* were revisited. Integration of bacteriophages into the coding region of the gene results in the disruption of the gene [251].

In *E. coli* C600 strain, integration of a bacteriophage at the 5' of *sbcB* gene resulted in the modification of 5' end of the *sbcB* gene [158]. The core attachment sequence of Sf101 phage overlapped with 13 bps of 5' end of *sbcB* gene and resulted in a base pair change.

In the lysogen, the region downstream of *sbcB* comprised of conserved housekeeping genes and the upstream region contained the whole Sf101 phage sequence. All *S. flexneri* strains analysed in this study were found to carry similar sequences of conserved genes downstream of *sbcB* genes, as present in the lysogen. The upstream sequences to *sbcB* genes in these strains contained numerous IS elements and no Sf101 sequences (including *attR)* were identified, in this region indicating the absence of complete or cryptic Sf101 phage. The phage remnant genes in bacterial genomes indicate the presence of the phage(s) which once had resided in the bacterial hosts [128] .

7.3.2 Transposon-like structure carrying oacB gene in 1c strains

In six serotype 1c and all 1a, 1b and Y strains, the *oacB* gene was flanked by an integrase gene and IS600/IS629 insertion sequences, giving a transposon-like structure. Similar flanking sequences to the *oacB* gene were reported by Wang *et al*., 2014, in the serotypes 1a, 1b, 5a, and Y strains [106]. It was thought that this transposon-like structure derived the mobilization of *oacB* gene in *S. flexneri* serotypes [105].

To confirm the mobilization of *oacB* gene from *adrA* to *sbcB* locus or vice versa, the flanking sequences of *oacB* gene in the Sf101 lysogen and other serotype 1c strains were analysed. It was carried out to identify any sequence(s) homologies that might have resulted in the localisation of *oacB* gene at either location (data not shown) via homologous recombination between similar sequences. However, no homology was found in sequences flanking *oacB* gene at either location, whereas three copies of SfI attachment site within *proA-adrA* region was identified in other serotype 1c strains carrying the *oacB* gene within the *proA-adrA* region, indicative of IS mediated insertion/deletion activity.

The findings from this study confirmed that in some serotype 1c strains, Sf101 phage integrated near *proA* gene within SfI region, using SfI attachment site as an anchoring point (SfI phage integrated into the *S. flexneri* chromosome at *tRNA-Thr* gene). Due to the occurrence of a large number of mobile elements in this vicinity, it is plausible to propose that Sf101 bacteriophage integrated within the *proA-adrA* region and was subsequently disrupted by insertion elements rendering a cryptic phage. Moreover, the SfI phage complete sequence was also not identified in any of the strains used in this study.

7.3.4 Sf101 phage integration in serotype 1c strains

Based on the findings of bioinformatics analysis in chapter 5, it was hypothesised that Sf101 phage integrated at two different locations in serotype 1c strains via two independent events. Nevertheless, the presence of *oacB* gene due to the insertion of the Sf101 phage within *sbcB* gene is an uncommon event for the integration of serotype converting phages of *S. flexneri*. All known characterised serotype converting phages of *S. flexneri* (SfI, SfII, SfIV, SfV, and SfX) are known to be integrated into *tRNA* gene within the *proA-adrA* region [23, 100, 193, 220] except for Sf6 which integrates into the *argW* tRNA gene next to conserved *yfdC* gene [104]. The *tRNA* genes used by many phages are target because of the features they offer, like the presence of multiple copies in a genome, small size, and their conserved sequences [252]. However, attachment at an alternate site is not unusual in lambda phages. A study identified several integration sites other than the *tRNA* gene, for Stx-phage $\Phi 24_B$ on the chromosome of *E. coli,* and *tRNA* was rarely a preferred site by the phage [253]. In a robust analysis of 471 *E. coli* and *Salmonella* prophages exposed 58 alternate integration loci for 369 *E. coli* phages and 102 for *S. enterica* phages, it was hypothesised that natural selection forms the basis of prophage integration patterns in relation to theie host genomes [254].

Furthermore, in *stx* bacteriophages of Shiga toxin-producing *E. coli* (STEC), five different attachment sites have been identified including *wrbA, yeh*, *sbcB, yecE, and Z2577* genes. It was also found that *wrbA* being the preferred site, however, if this site was occupied by another phage in a multi-lysogen, the phage would use an alternate attachment site [255]. In

another unknown phage of *S. flexneri*, encoding acetyltransferase gene *oac1b* was found to be integrated at a new location between *torT* and *ycmA* genes on the bacterial chromosome and resulted in the disruption of the *ycmA* gene [109, 256].

In the current study, *attL* and *attR* sites at the two ends of Sf101 phage in the Sf101 lysogen indicated host-phage junction sequences and confirmed the phage Sf101 integration within the *sbcB* gene, reiterating the findings of Jakhetia *et al*., 2014. The *attP* of Sf101 phage was found to complement the 5' end of *sbcB* gene (*attB*) in both the lysogens and resulted in the homologous recombination and subsequent integration of Sf101 phage into the host genome.

The homologies between two or more phages existing on the same host offer homologous recombination and result in the genome rearrangement of the host [257]. In some examples in *Streptococcus pyogenes* and *Xylella fastidiosa*, prophages with even limited sequence identity to bacterial hosts resulted in homologous recombination and subsequent genomic DNA rearrangement [258] . Recombination events involving regions of host or regions of other phages (active or cryptic) are considered to be a driving force of genomic diversity. This seems the same scenario where similarities between SfI and Sf101 attachment sites resulted in Sf101 integration in some 1c strains. Moreover, it was thought that SfI integration produced a duplicate copy of the SfI attachment site and then a third copy (present between *oacB* and *proA*) was created due to the IS-mediated insertion event (Figure 7.1).

Analysis clearly indicated that the sequences were rearranged in 1c strains by mobile elements and this rearrangement resulted in deletion of Sf101 *attL*/*attR* in 1c strains within *proA*-*adrA* region. It was found that other prophage (SfI/Sf101) sequences were disrupted by insertion elements or deleted during genome rearrangement events and resulted in the presence of cryptic phages. In a previous study, it was found that deletion of *stx* region in the prophage of *S. dysenteriae* resulted in defective phages [259]. Moreover, prophage genes are subjected to selection pressure (the beneficial genes are kept by the host bacteria, while others are deleted) resulting in the presence of cryptic phages [258].

The analysis also revealed that some of the 1c strains did not have *oacB* gene or Sf101 remnants, suggesting that Sf101 phage was never inserted into those strains. But the strains which received Sf101 phage at either site (*sbcB /adrA*), underwent severe gene disruption and deletion events due to the presence of multiple IS elements surrounding these regions (Figure 7.2). However, the existence of secondary attachment site for the *oacB* gene encoding phage offers potential implications for the mobility of *oacB* gene and ultimate serotype-diversity of *S. flexneri*.

Future research involving more serotype 1c strains from diverse geographical regions would help to understand further the origin of *oacB* gene in 1c strains.

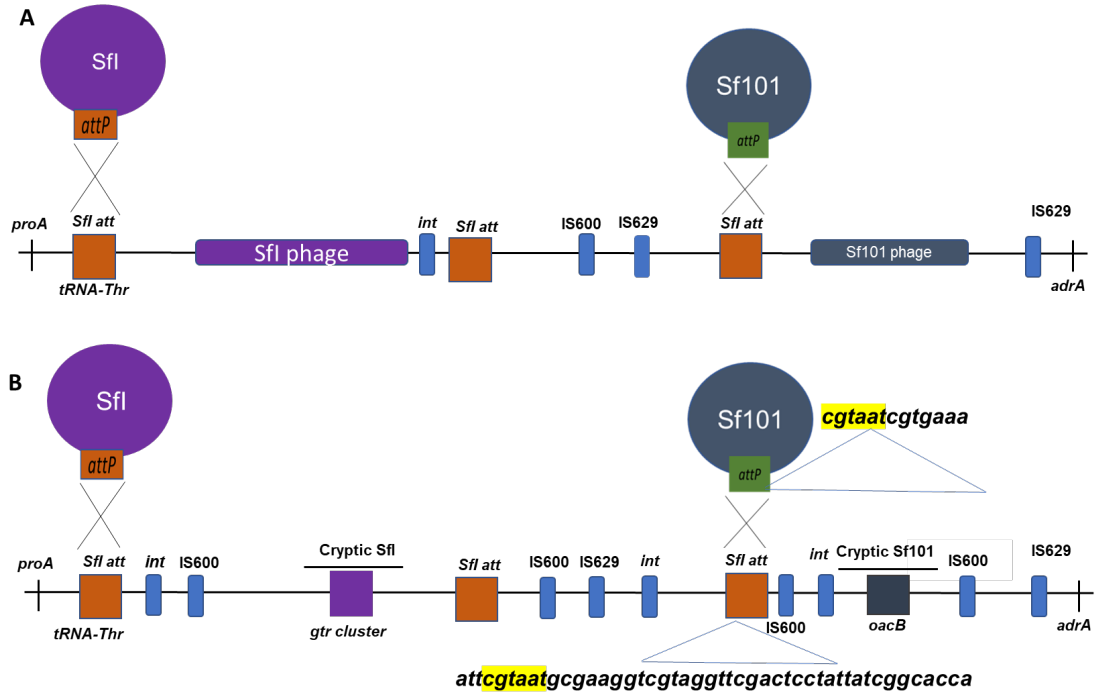


Figure 7.2: Acquisition of *oacB* gene in serotype 1c of *S. flexneri* strains.
A) The integration of SfI phage (purple) using phage *attP* (dark orange) into the *tRNA-thr* gene (dark orange). The presence of insertion sequences-IS (blue) resulted in the creation of three identical copies of the SfI attachment site (dark orange). Sequence homology between attachment sites of SfI and Sf101 provided an opportunity for Sf101 phage to integrate using Sf101 *attP* (green) at the triplicate copy of the SfI attachment site. The Sf101 genome is shown in dark grey color. **B)** Over time, due to the combination of multiple IS elements/integrases (*int*), the SfI and Sf101 sequences underwent deletion events leaving behind cryptic phages with representative genes *gtrI* gene cluster (purple) and *oacB* (dark grey). The similar sequences between SfI and Sf101 attachment sites are highlighted in yellow.

7.4 Novel *orfs* of Sf101 phage and virulence of *S. flexneri*

This study focused on investigating the role of four unique genes of Sf101 phage in the *S. flexneri* virulence. The expression of four novel genes (*orf16/oacB, orf17*, *orf41,* and *orf56*) was confirmed in the lysogen by RT-PCR. However, due to the absence of VP in the lysogenic strain, the use of this strain for virulence assays was not possible. Another, Sf101 lysogen-SFL1684 was also analysed for the presence of VP and found that this strain had also lost the VP. The presence of VP is essential for any *S. flexneri* strain to be virulent as VP encodes genes for type three-secretory system (T3SS) along with other virulence-associated genes [260]. After several attempts of mating experiments, VP could not transfer in the lysogen. Later, bioinformatics analysis of VP found that VP of Y394 lacked complete *tra* locus (*traD, traM, traY,* and *traI*) necessary for the plasmid exchange between two bacteria [261]. The lack of conjugative transfer machinery is not uncommon and reported in VP (pCO301) *S. flexneri* 2a strain 301 and pWR100 of *S. flexneri* strain M90T [262, 263]. As the VP could not be restored, the role of novel genes was studied in another serotype 1c strain harboring VP. For this purpose, the bacterial accumulation/ liquid killing assays using *C. elegans* and invasion assay involving Hela cells were used.

7.4.1 oacB modification and presence of orf17, orf41, and orf56 do not affect the virulence

C. elegans have been used as a promising animal model to study the contribution of unknown genes in the virulence of bacterial pathogen since 1963, when Sydney Brenner, first time used *C. elegans* as an animal model to study the behavior and neurobiological processes in eukaryotes [264]. Since then, *C. elegans* are being used as the preferred model to study the virulence of bacteria including *Shigella*, *Pseudomonas aeruginosa*, *Salmonella enterica*, *Staphylococcus aureus,* etc [224, 265-267]. Despite the use of *C. elegans* for virulence studies in different pathogens the complete pathogenesis and the response of the worms to the infection is not fully understood, hence limiting the use of this *in vivo* model. For example, in the case of human's enteric infection, by *S. flexneri* other environmental factors such as gastric

pH, gut microbiome, presence of food in the stomach, and the host physiology play a critical role in the establishment of the initial stages of the infection [268]. However, the morphological similarities between the intestinal cells of humans and *C. elegans* make *C. elegans* a model of choice to study some aspects of host-pathogen interaction *in vivo* [269]. In the present study two different *C. elegans* assays; bacterial accumulation and liquid killing assays were employed for preliminary investigation of the role of Sf101 novel *orfs* in the host virulence.

Burton *et al*., has shown that *C. elegans* digests avirulent bacteria whereas virulent bacteria accumulate in intestinal lumina and kill worms [269]. The bacterial accumulation assay was performed which determines the overall virulence of the pathogen. It has been shown that nematode killing takes place with the increasing bacterial load (CFU/worm) within the intestinal lumina [269]. In bacterial accumulation assay after 24 hours of infection, the intact bacteria were found to be present in the gut of the worms (Section 6.4.1) which indicated successful escape of bacteria from the pharyngeal grinding. For the nematodes, the first line of defense against invading bacteria is pharyngeal grinding. LPS of Gram-negative bacteria are known to play an important role during *C. elegans* infection and help escape host defenses [224, 270]. The modified O-antigen part of LPS has been shown to protect cells from the macerating effects of the worm's pharynx grinder [271]. In a study by Browning *et al*., 2013, when the effect of O-antigen on *C. elegans* was investigated it was found that killing of worms by *E. coli* strain was associated with the bacterial resistance to the mechanical shearing in the pharynx, and bacterial density in the gut resulted in rupture of the intestinal cavity [272]. The LPS of 1c strain presents a unique O-antigen structure which might help overcome the worms immune system and resulted in the accumulation. However, the accumulation assay was failed to differentiate the affect produced by either of the *orfs*.

A liquid killing assay was also performed to determine if the accumulation of bacterial pathogen in the gut resulted in the killing of worms [224]. It has the advantage over bacterial accumulation by providing continuous exposure of bacteria to the worms. Kesika *et al*., 2011 have shown that nematodes become more susceptible to S. *flexneri* in liquid medium [222].

This assay was performed for up to 48 hours and survival rates of the worms were compared. The assay was not performed beyond 48 hours due to the presence of L1/L2 larval forms, which were hindering the scoring of L4 worms. Nevertheless, in both the *C. elegans* assays no significant differences were observed among *S. flexneri* strains harbouring Sf101 novel *orfs* when compared to the controlled strain. This might have happened due to the shorter duration of exposure of worms to the *Shigella* in both the *C. elegans* assays. In a study, variations in virulence between *Shigella* and enteroinvasive *E. coli* (EIEC) were determined, using bacterial accumulation/killing assays. In which worms were exposed to bacteria for 72 hours and then differences in overall virulence were observed [273]. Similarly, Kesika *et al*., 2011 have demonstrated that when N2 worms were fed with lawns of *S. flexneri*, and complete killing of worms was achieved at153 hours [222]. In another study when the survival rate of worms was investigated it was found that when nematodes were exposed to *Salmonella enterica* for several hours and then transferred to OP50 plate the titers of *S. enetrica* remained high in the intestine and died over the course of 5-7 days [274].

It is also likely that these *orfs* (*orf16, orf17, orf41,* and *orf56*) might be needed in later phases of the infection process involving invasion of intestinal cells and that can be determined by increasing the infection time in worms and then a time-course experiment should be performed to evaluate CFU/worm and worms survival at each time point. This will also help understand the role of any of these *orfs* in adhesion and the establishment of persistent infection. It could also be possible that these *orfs* might be needed to invade the specialized processes during human infection, which are not part of the worms infection process. For example, during the human infection process, *S. flexneri* takes advantage of the transcytotic properties of specialized M cells to invade intestinal cells. Following which reaches the sub-epithelial spaces to invade the epithelial cells of the colon and rectum. However, *C. elegans* does not have specialized M cells and while infecting worms, *S. flexneri* uses the apical side to gain access to the intestinal cells [224].

7.4.2 Need more sensitive assays to evaluate role of Sf101 novel orfs in the host virulence

The results presented in chapter 6 also demonstrated the role of four Sf101 *orfs* in the invasive capability of *S. flexneri*. During the invasive process, *S. flexneri* invades epithelial cells of the large intestine, and invasion is achieved with the help of expression of genes located on virulence plasmid (VP), for type three secretory system (T3SS) to inject effector proteins into the host cells [52, 275]. In this study, Sf101 novel *orfs* did not affect the invasive potential when compared with the wild type of strain. The invasion assay measures the total number of infecting bacteria and due to the limited sensitivity of the assay, subtle phenotypic changes could not be determined. Hence, more sensitive virulence assays like Hela cell plaque assay, which determines attachment and internalisation of *Shigella* in the epithelial cells, followed by, intracellular multiplication in the cytosol and subsequent spread to the nearby cells [276]; assay involving radiolabelled bacteria as described by Guhathakurta *et al*., can be performed to investigate the role of individual *orfs* [275]; or a Sereny test, which determines the invasive ability of *Shigellae* by causing keratoconjunctivitis in the corneal epithelium of guinea pigs would help understand the role of these *orfs* in *S. flexneri* pathogenesis [277, 278]. Ideally, the role of novel Sf101 *orfs* should be investigated in the context to complete the Sf101 phage, and Sf101 lysogen should be made in serotype Y (with no O-antigen modifications), and then knock out strains should be created to evaluate their role against wild type strain.

7.4.3 Novel orfs of Sf101 phage may play role in phage biology

As the role of Sf101 novel *orfs* could not be determined using *C. elegans* survival and accumulation assays, and Hela cells invasion assays. It is also possible that these *orfs* are involved in phage-related functions which are yet to be discovered.

In one such study conducted by Seesandra *et al*., 2011, protein functions were investigated by identifying protein to protein interactions within phage proteins. In that study 68 lambda, ORFs out of 73 were cloned into Gateway vectors and protein interactions were determined in an array-based yeast two-hybrid (Y2H) screens. The study concluded by screening 97 interactions and identified functionally related proteins. Several unknown proteins were found to interact with the well-characterized proteins in the lambda genome. For example, CI (lambda repressor), CII (transcriptional activator), and CIII (regulatory protein) interacted with various components of virion related to the late [279]. Similarly, a partner with a known function of a hypothetical protein (*orf16*, *orf17*, *orf41,* and *orf56*) can be identified, and then a role could be assigned to an unknown protein in association with its partner. In the whole genome of the Sf101 phage, there are a total of 66 genes coding for proteins. Out of which only 43 were assigned putative functions whereas, the other 23 *orfs* belonging to various modules, showed similarities to uncharacterized proteins [105]. It would be interesting to create vectors carrying various genes from the same module(s) and observe protein to protein interactions as performed in the lambda virus.

7.5 Future directions

- Results obtained in Chapter 3 for the overexpression of OacB, can be utilised in future studies involving SCAM, to investigate the locations of loops in OacB.
- OacB should be crystallized, and structure should be analysed using X-ray. To produce a 3D structure, a pure form of protein is needed, and hence the optimized parameters for the production and purification of OacB in this study can be used at a large scale to produce OacB to perform X-ray crystallography. Alternatively, the over-expressed protein can be used to perform solid-state NMR (SS-NMR) to obtain high-resolution structural information. SS-NMR does not require crystals to determine structure [280].
- Once a 3D structure of OacB is solved, the mutants of OacB created in this study can be used to identify the catalytic site(s) in OacB. On the other hand, the mutants created in this study can be used in molecular docking to find the binding site in OacB. However, the prerequisite for this technique is the availability of the solved crystal structure of the protein. Once the 3D structure is obtained the molecular docking can then also be performed.
- Results obtained in chapter 5 helped explain the integration of Sf101 phage in 1c serotype strains and the origin of *oacB* in serotype 1c. To support the findings of this study, the serotype 1c strains from more geographical locations in the world be collected and sequenced for a better understanding of the origin of *oacB* gene.
- The results from *in vivo* assays performed in chapter 6 confirmed that *orf16/oacB, orf17, orf41* and, *orf56* had no role in the virulence of the host. However, for further confirmation of these results, Sf101 lysogen should be made in the serotype Y that does not have any other O-antigenic modification. The knockouts could then be created in the lysogen to be used in the virulence assays. Alternatively, the role of four novel genes of Sf101 phage should be determined by performing more sensitive

assays including plaque assay using Hela cells, which would help understand inter- and intra-cellular bacterial spread; rabbit ligated ileal loop, to find out the extent of alterations of mucosal surfaces due to the destruction of epithelial cells of the intestine and Sereny test to test the invasiveness of *Shigella* strains.

7.6 Concluding remarks

This study successfully demonstrated the overexpression of an integral membrane domain protein OacB which is responsible for O-antigen modification. The created 2D topology model of OacB also helped in the visual representation of the loops and domains in the protein. The key features of OacB were studied by probing the role of important residues and their possible interplay during the process of *O*-acetylation. The knowledge gained about the critical residues and domain of OacB can help support future studies to understand the O-antigen modification mechanism by OacB and subsequently other acetyltransferases.

The findings of this study contributed to understand the ability of Sf101 phage to integrate into two distinct chromosomal sites which could contribute to the spread of the *oacB* gene and increase in sero-diversity. The findings of this study have provided valuable information about the origin and evolution of serotype 1c strains.

The role of four novel genes in the host virulence using *invivo / in vitro* assays was also investigated. The results indicated that the role of these *orfs* in host virulence should be evaluated in more sensitive assays or in bacteriophage biology.

This study has provided a detailed characterization of *oacB* and three other novel *orfs* of serotype converting Sf101 phage of *S. flexneri*. The origin of *oacB* gene in serotype 1c strains investigated herein has opened avenues for the upcoming research to understand the serotype conversion in *S. flexneri*.

Appendix A

Bacterial Cell Culture

LB broth

1.0% (w/v) tryptone

0.5% (w/v) yeast extract

0.5% (w/v) NaCl

Milli Q water

LB broth agar

1.5% Bacto-agar to LB Broth

SOB medium

2.0% (w/v) tryptone

0.5% (w/v) yeast extract

0.5% (w/w) NaCl

10 mM $MgCl_2$

2.5 mM KCl

10 mM $MgSO_4$

Add tryptone, yeast and NaCl.

Make up the volume to 1L with Milli-Q water and autoclave, and then add filter-sterilised KCL, $MgCl_2$, $MgSO_4$.

SOC medium

20 mM glucose added to a final concentration to autoclave SOB medium.

Ampicillin: 100 mg/ml stock.

1000 mg/10 ml of Milli-Q water

Filter sterilise.

Chloramphenicol

25 mg/ml stock

250 mg of chloramphenicol powder added to 10 ml 100% Methanol.

Store at -20°C

Erythromycin

50 mg/ml stock

500 mg of ampicillin powder added to 10 ml Milli-Q water.

Vortex, filter, and store at -20°C

Kanamycin

50 mg/ml stock

500 mg of kanamycin powder added to 10 ml Milli-Q water.

Mix well, filter, and store at -20°C

Alkaline Lysis Solution for Plasmid DNA Isolation

Alkaline Lysis solution I

50 mM Glucose

25 mM Tris-HCL (pH 8.0)

10 mM EDTA (pH 8.0)

Autoclaved and stored at 4°C.

Alkaline Lysis Solution II

0.2 M NaOH

1% (w/v) SDS

Prepare fresh (Do not autoclave)

Alkaline Lysis Solution III

3 M CH_3OOK

11.5% (v/v) Glacial acetic acid

Autoclaved and stored at 4°C.

Agarose Gel Electrophoreses

0.5 x TBE Buffer

45 mM Tris-HCl

45 mM Boric Acid

1 mM Na_2EDTA

Blue Loading Dye

1 mg/mL Bromophenol blue

20% glycerol

SDS Gel Electrophoresis

12% Resolving Gel.

3.2 mL Milli-Q Water

2.1 mL Bis Acrylamide (acryl/bis 29:1)

1.9 mL Tris-HCl (1.5M pH8.8)

75 µL 10% SDS (w/v)

75 µL Ammonium Persulfate

3 µL Tetramethyl ethylenediamine

4% Stacking Gel

1.6 mL Milli-Q Water

250 µL Bis Acrylamide (acryl/bis 29:1)

625 µL Tris HCl (0.5M pH6.8)

25 µL 10% SDS (w/v)

12.5 µL Ammonium Persulfate

2.5 µL tetramethyl ethylenediamine

2X Sample loading buffer

10% (w/v) SDS

20% glycerol

0.1% (w/v) bromophenol blue

0.5 M Tris-HCl, pH 6.8

2.5% 2-β-mercaptoethanol (must be added fresh before use)

Western blot

Tris Base 50 mM

Glycine 384 mM

SDS 0.1%

Methanol 20%

Make fresh.

Phosphate buffer Saline (PBS)

NaCl

KCl

Na_2HPO_4 (10 mM)

KH_2PO_4 (2 mM)

Skim milk

Skim milk 5%

Made up to volume in 1 X PBS.

Tris-Buffered Saline (TBS)

50mM Tris-HCl, pH 7.4

150mM NaCl

Coomassie brilliant blue stain

Coomassie brilliant blue 0.5% (w/v)

Methanol 50%

Acetic acid 10%

Destaining solution

Methanol 40%

Acetic acid 10%

Silver staining (all made freshly)

Fixing solution

7.5% (v/v) glacial acetic acid

25% (v/v) propan-2-ol

Oxidizing solution

0.7% (w/v) periodic acid

7.5% (v/v) glacial acetic acid

Silver staining solution (all made freshly)

0.0187 M NaOH

1.3% (v/v) NH_4OH

0.67% (w/v) $AgNO_3$

Developer solution

0.0222% (v/v) formaldehyde

0.005% (w/v) citric acid

Stop solution.

1% (v/v) glacial acetic acid

Bacteriophage methods

NZCYM media

0.1% (w/v) Caseamino acids

0.066% (w/v) $MgSO_4.7H_2O$

0.167% (w/v) yeast extract

0.167% (w/v) NaCl

0.5% soft agar

1% (w/v) tryptone

0.5% (w/v) yeast extract

0.5% (w/v) NaCl

0.5% (w/v) agar

SM buffer

100 mM NaCl

25 mM Tris-HCL (pH 7.5)

8 mM $MgSO_4$

0.002% (w/v) gelatin

TE Buffer

10 mM Tris-HCL (pH 7.5)

1 mM EDTA (pH 8.0)

Tissue culture

1 x PBS

0.8% (w/v) NaCl

0.02% (w/v) KCl

0.144% (w/v) Na_2HPO_4

0.0024% (w/v) KH_2PO_4

Adjust the pH to 7.4 and sterilize by autoclaving.

***C. elegans* methods**

Modified nematode growth media (mNGM)

50 mM NaCl

0.35% (w/v) peptone

2% (w/v) agar

Sterilised by autoclaving and cooled to 55°C

1 mM $CaCl_2$

1 mM $MgSO_4$

5 μ/ml cholesterol

25 mM potassium phosphate

S-Basal

0.1 M NaCl

0.05 M potassium phosphate (pH 6.0)

1 ml cholesterol (5mg/ml in ethanol)

1 M potassium phosphate pH 6.0

0.132 M KH_2PO_4

0.868 M KH_2PO_4

Alkaline hypochlorite solution

200 µl 4M NaOH

300 µl bleach

Bibliography

1. Stagg RM, Cam PD, Verma NK: **Identification of newly recognized serotype 1c as the most prevalent Shigella flexneri serotype in northern rural Vietnam**. *Epidemiology and Infection* 2008, **136**(8):1134-1140.
2. The HC, Thanh DP, Holt KE, Thomson NR, Baker S: **The genomic signatures of Shigella evolution, adaptation and geographical spread**. *Nature Reviews Microbiology* 2016, **14**(4):235-250.
3. Pires SM, Fischer-Walker CL, Lanata CF, Devleesschauwer B, Hall AJ, Kirk MD, Duarte ASR, Black RE, Angulo FJ: **Aetiology-Specific Estimates of the Global and Regional Incidence and Mortality of Diarrhoeal Diseases Commonly Transmitted through Food**. *PloS one* 2015, **10**(12):e0142927-e0142927.
4. Kotloff KL, Winickoff JP, Ivanoff B, Clemens JD, Swerdlow DL, Sansonetti PJ, Adak GK, Levine MM: **Global burden of Shigella infections: implications for vaccine development and implementation of control strategies**. *Bull World Health Organ* 1999, **77**.
5. Mani S, Wierzba T, Walker RI: **Status of vaccine research and development for Shigella**. *Vaccine* 2016, **34**(26):2887-2894.
6. Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, Brewer TG, Engmann CM, Houpt ER, Kang G *et al*: **Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016**. *The Lancet Infectious Diseases* 2018, **18**(11):1229-1240.
7. Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi AK: **Shigellosis**. *The Lancet* 2018, **391**(10122):801-812.
8. DuPont HL, Hornick RB, Dawkins AT, Snyder MJ, Formal SB: **The response of man to virulent Shigella flexneri 2a**. *The Journal of infectious diseases* 1969, **119**(3):296-299.
9. Tribble DR: **Antibiotic Therapy for Acute Watery Diarrhea and Dysentery**. *Military medicine* 2017, **182**(S2):17-25.
10. Marteyn B, Gazi A, Sansonetti P: **Shigella: a model of virulence regulation in vivo**. *Gut microbes* 2012, **3**(2):104-120.
11. Ko C-F, Lin N-T, Chiou C-S, Wang L-Y, Liu M-C, Yang C-Y, Lee Y-S: **Infrequent cross-transmission of Shigella flexneri 2a strains among villages of a mountainous township in Taiwan with endemic shigellosis**. *BMC Infectious Diseases* 2013, **13**(1):1-8.
12. Ingle DJ, Easton M, Valcanis M, Seemann T, Kwong JC, Stephens N, Carter GP, Gonçalves da Silva A, Adamopoulos J, Baines SL *et al*: **Co-circulation of Multidrug-resistant Shigella Among Men Who Have Sex With Men in Australia**. *Clinical Infectious Diseases* 2019, **69**(9):1535-1544.
13. Ashkenazi S, Cohen D: **An update on vaccines against Shigella**. *Therapeutic advances in vaccines* 2013, **1**(3):113-123.
14. Jennison AV, Verma NK: **Shigella flexneri infection: pathogenesis and vaccine development**. *FEMS Microbiology Reviews* 2004, **28**(1):43-58.
15. Wahid R, Simon JK, Picking WL, Kotloff KL, Levine MM, Sztein MB: **Shigella antigen-specific B memory cells are associated with decreased disease severity in subjects challenged with wild-type Shigella flexneri 2a**. *Clinical Immunology* 2013, **148**(1):35-43.
16. Venkatesan MM, Buysse JM, Kopecko DJ: **Characterization of invasion plasmid antigen genes (ipaBCD) from Shigella flexneri**. 1988, **85**(23):9317-9321.
17. Lu T, Das S, Howlader DR, Zheng Q, Ratnakaram SSK, Whittier SK, Picking WD, Picking WL: **L-DBF Elicits Cross Protection Against Different Serotypes of Shigella spp**. 2021, **2**.
18. Salgado-Pabón W, Celli S, Arena ET, Nothelfer K, Roux P, Sellge G, Frigimelica E, Bousso P, Sansonetti PJ, Phalipon A: **Shigella impairs T lymphocyte dynamics in vivo**. *Proceedings of the National Academy of Sciences* 2013, **110**(12):4458-4463.

19. Pilla G, Tang CM: **Going around in circles: virulence plasmids in enteric pathogens**. *Nature Reviews Microbiology* 2018, **16**(8):484-495.
20. Sansonetti PJ, Kopecko DJ, Formal SB: **Involvement of a plasmid in the invasive ability of Shigella flexneri**. *Infection and Immunity* 1982, **35**(3):852-860.
21. Schroeder GN, Hilbi H: **Molecular pathogenesis of Shigella spp.: controlling host cell signaling, invasion, and death by type III secretion**. *Clinical microbiology reviews* 2008, **21**(1):134-156.
22. Pilla G, McVicker G, Tang CM: **Genetic plasticity of the Shigella virulence plasmid is mediated by intra- and inter-molecular events between insertion sequences**. *PLOS Genetics* 2017, **13**(9):e1007014.
23. Allison GE, Verma NK: **Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri**. *Trends in Microbiology* 2000, **8**(1):17-23.
24. Al-Hasani K, Navarro-Garcia F, Huerta J, Sakellaris H, Adler B: **The immunogenic SigA enterotoxin of Shigella flexneri 2a binds to HEp-2 cells and induces fodrin redistribution in intoxicated epithelial cells**. *PloS one* 2009, **4**(12):e8223-e8223.
25. Purdy GE, Payne SM: **The SHI-3 iron transport island of Shigella boydii 0–1392 carries the genes for aerobactin synthesis and transport**. *J Bacteriol* 2001, **183**.
26. Turner SA, Luck SN, Sakellaris H, Rajakumar K, Adler B: **Molecular Epidemiology of the SRL Pathogenicity Island**. *Antimicrobial Agents and Chemotherapy* 2003, **47**(2):727-734.
27. Dorman CJ, McKenna S, Beloin C: **Regulation of virulence gene expression in Shigella flexneri, a facultative intracellular pathogen**. *International Journal of Medical Microbiology* 2001, **291**(2):89-96.
28. Venkatesan MM, Goldberg MB, Rose DJ, Grotbeck EJ, Burland V, Blattner FR: **Complete DNA sequence and analysis of the large virulence plasmid of Shigella flexneri**. *Infection and immunity* 2001, **69**(5):3271-3285.
29. DeMali KA, Jue AL, Burridge K: **IpaA targets β1 integrins and rho to promote actin cytoskeleton rearrangements necessary for Shigella entry**. *Journal of Biological Chemistry* 2006, **281**(51):39534-39541.
30. Hamiaux C, van Eerde A, Parsot C, Broos J, Dijkstra BW: **Structural mimicry for vinculin activation by IpaA, a virulence factor of Shigella flexneri**. *EMBO reports* 2006, **7**(8):794-799.
31. Izard T, Tran Van Nhieu G, Bois PRJ: **Shigella applies molecular mimicry to subvert vinculin and invade host cells**. *The Journal of cell biology* 2006, **175**(3):465-475.
32. Tran Van Nhieu G, Ben-Ze'ev A, Sansonetti PJ: **Modulation of bacterial entry into epithelial cells by association between vinculin and the Shigella IpaA invasin**. *Embo j* 1997, **16**(10):2717-2729.
33. Blocker A, Gounon P, Larquet E, Niebuhr K, Cabiaux V, Parsot C, Sansonetti P: **The tripartite type III secreton of Shigella flexneri inserts IpaB and IpaC into host membranes**. *J Cell Biol* 1999, **147**(3):683-693.
34. Chen Y, Smith MR, Thirumalai K, Zychlinsky A: **A bacterial invasin induces macrophage apoptosis by binding directly to ICE**. *Embo j* 1996, **15**(15):3853-3860.
35. Hume PJ, McGhie EJ, Hayward RD, Koronakis V: **The purified Shigella IpaB and Salmonella SipB translocators share biochemical properties and membrane topology**. *Mol Microbiol* 2003, **49**(2):425-439.
36. Harrington AT, Hearn PD, Picking WL, Barker JR, Wessel A, Picking WD: **Structural characterization of the N terminus of IpaC from Shigella flexneri**. *Infect Immun* 2003, **71**(3):1255-1264.
37. Veenendaal AK, Hodgkinson JL, Schwarzer L, Stabat D, Zenk SF, Blocker AJ: **The type III secretion system needle tip complex mediates host cell sensing and translocon insertion**. *Mol Microbiol* 2007, **63**(6):1719-1730.
38. Rohde JR, Breitkreutz A, Chenal A, Sansonetti PJ, Parsot C: **Type III secretion effectors of the IpaH family are E3 ubiquitin ligases**. *Cell Host Microbe* 2007, **1**(1):77-83.

39. Ogawa M, Yoshimori T, Suzuki T, Sagara H, Mizushima N, Sasakawa C: **Escape of intracellular Shigella from autophagy**. *Science* 2005, **307**(5710):727-731.
40. Shere KD, Sallustio S, Manessis A, D'Aversa TG, Goldberg MB: **Disruption of IcsP, the major Shigella protease that cleaves IcsA, accelerates actin-based motility**. *Mol Microbiol* 1997, **25**(3):451-462.
41. Egile C, d'Hauteville H, Parsot C, Sansonetti PJ: **SopA, the outer membrane protease responsible for polar localization of IcsA in Shigella flexneri**. *Mol Microbiol* 1997, **23**(5):1063-1073.
42. Handa Y, Suzuki M, Ohya K, Iwai H, Ishijima N, Koleske AJ, Fukui Y, Sasakawa C: **Shigella IpgB1 promotes bacterial entry through the ELMO-Dock180 machinery**. *Nat Cell Biol* 2007, **9**(1):121-128.
43. Alto NM, Shao F, Lazar CS, Brost RL, Chua G, Mattoo S, McMahon SA, Ghosh P, Hughes TR, Boone C *et al*: **Identification of a bacterial type III effector family with G protein mimicry functions**. *Cell* 2006, **124**(1):133-145.
44. Pendaries C, Tronchere H, Arbibe L, Mounier J, Gozani O, Cantley L, Fry MJ, Gaits-Iacovoni F, Sansonetti PJ, Payrastre B: **PtdIns5P activates the host cell PI3-kinase/Akt pathway during Shigella flexneri infection**. *Embo j* 2006, **25**(5):1024-1034.
45. Santapaola D, Del Chierico F, Petrucca A, Uzzau S, Casalino M, Colonna B, Sessa R, Berlutti F, Nicoletti M: **Apyrase, the product of the virulence plasmid-encoded phoN2 (apy) gene of Shigella flexneri, is necessary for proper unipolar IcsA localization and for efficient intercellular spread**. *Journal of bacteriology* 2006, **188**(4):1620-1627.
46. Zurawski DV, Mitsuhata C, Mumy KL, McCormick BA, Maurelli AT: **OspF and OspC1 are Shigella flexneri type III secretion system effectors that are required for postinvasion aspects of virulence**. *Infect Immun* 2006, **74**(10):5964-5976.
47. Parsot C, Ageron E, Penno C, Mavris M, Jamoussi K, d'Hauteville H, Sansonetti P, Demers B: **A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in Shigella flexneri**. *Mol Microbiol* 2005, **56**(6):1627-1635.
48. Miura M, Terajima J, Izumiya H, Mitobe J, Komano T, Watanabe H: **OspE2 of Shigella sonnei is required for the maintenance of cell architecture of bacterium-infected cells**. *Infect Immun* 2006, **74**(5):2587-2595.
49. Arbibe L, Kim DW, Batsche E, Pedron T, Mateescu B, Muchardt C, Parsot C, Sansonetti PJ: **An injected bacterial effector targets chromatin access for transcription factor NF-kappaB to alter transcription of host genes involved in immune responses**. *Nat Immunol* 2007, **8**(1):47-56.
50. Kim DW, Lenzen G, Page AL, Legrain P, Sansonetti PJ, Parsot C: **The Shigella flexneri effector OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes**. *Proc Natl Acad Sci U S A* 2005, **102**(39):14046-14051.
51. Benjelloun-Touimi Z, Sansonetti PJ, Parsot C: **SepA, the major extracellular protein of Shigella flexneri: autonomous secretion and involvement in tissue invasion**. *Mol Microbiol* 1995, **17**(1):123-135.
52. Faherty CS, Redman JC, Rasko DA, Barry EM, Nataro JP: **Shigella flexneri effectors OspE1 and OspE2 mediate induced adherence to the colonic epithelium following bile salts exposure**. *Mol Microbiol* 2012, **85**(1):107-121.
53. Brunner K, Samassa F, Sansonetti PJ, Phalipon A: **Shigella-mediated immunosuppression in the human gut: subversion extends from innate to adaptive immune responses**. *Human vaccines & immunotherapeutics* 2019, **15**(6):1317-1325.
54. Mounier J, Vasselon T, Hellio R, Lesourd M, Sansonetti PJ: **Shigella flexneri enters human colonic Caco-2 epithelial cells through the basolateral pole**. *Infect Immun* 1992, **60**(1):237-248.

55. Wassef JS, Keren DF, Mailloux JL: **Role of M cells in initial antigen uptake and in ulcer formation in the rabbit intestinal loop model of shigellosis**. *Infection and Immunity* 1989, **57**(3):858-863.
56. Islam D, Veress B, Bardhan PK, Lindberg AA, Christensson B: **In situ characterization of inflammatory responses in the rectal mucosae of patients with shigellosis**. *Infect Immun* 1997, **65**(2):739-749.
57. Egile C, Loisel TP, Laurent V, Li R, Pantaloni D, Sansonetti PJ, Carlier M-F: **Activation of the Cdc42 Effector N-Wasp by the Shigella flexneri Icsa Protein Promotes Actin Nucleation by Arp2/3 Complex and Bacterial Actin-Based Motility**. *Journal of Cell Biology* 1999, **146**(6):1319-1332.
58. Kim M, Ogawa M, Fujita Y, Yoshikawa Y, Nagai T, Koyama T, Nagai S, Lange A, Fassler R, Sasakawa C: **Bacteria hijack integrin-linked kinase to stabilize focal adhesions and block cell detachment**. *Nature* 2009, **459**(7246):578-582.
59. Mattock E, Blocker AJ: **How Do the Virulence Factors of Shigella Work Together to Cause Disease?** *Frontiers in Cellular and Infection Microbiology* 2017, **7**(64).
60. Navarro-Garcia F, Gutierrez-Jimenez J, Garcia-Tovar C, Castro LA, Salazar-Gonzalez H, Cordova V: **Pic, an Autotransporter Protein Secreted by Different Pathogens in the Enterobacteriaceae Family, Is a Potent Mucus Secretagogue**. *Infection and Immunity* 2010, **78**(10):4101-4109.
61. Fasano A, Noriega F, Liao F, Wang W, Levine M: **Effect of shigella enterotoxin 1 (ShET1) on rabbit intestine in vitro and in vivo**. *Gut* 1997, **40**(4):505-511.
62. Moss JE, Cardozo TJ, Zychlinsky A, Groisman EA: **The selC-associated SHI-2 pathogeniCity island of Shigella flexneri**. *Mol Microbiol* 1999, **33**.
63. Ingersoll MA, Moss JE, Weinrauch Y, Fisher PE, Groisman EA, Zychlinsky A: **The ShiA protein encoded by the Shigella flexneri SHI-2 pathogenicity island attenuates inflammation**. *Cellular Microbiology* 2003, **5**(11):797-807.
64. Pupo GM, Lan R, Reeves PR: **Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics**. *Proc Natl Acad Sci USA* 2000, **97**.
65. Prosseda G, Di Martino ML, Campilongo R, Fioravanti R, Micheli G, Casalino M, Colonna B: **Shedding of genes that interfere with the pathogenic lifestyle: The Shigella model**. *Research in microbiology* 2012, **163**:399-406.
66. Nakata N, Tobe T, Fukuda I, Suzuki T, Komatsu K, Yoshikawa M, Sasakawa C: **The absence of a surface protease, OmpT, determines the intercellular spreading ability of Shigella: the relationship between the ompT and kcpA loci**. *Mol Microbiol* 1993, **9**.
67. Yang F, Yang J, Zhang X, Chen L, Jiang Y, Yan Y, Tang X, Wang J, Xiong Z, Dong J *et al*: **Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery**. *Nucleic Acids Research* 2005, **33**(19):6445-6458.
68. Bravo V, Puhar A, Sansonetti P, Parsot C, Toro CS: **Distinct Mutations Led to Inactivation of Type 1 Fimbriae Expression in Shigella spp**. *PLOS ONE* 2015, **10**(3):e0121785.
69. Lan R, Reeves PR: **Escherichia coli in disguise: molecular origins of Shigella**. *Microbes Infect* 2002, **4**(11):1125-1132.
70. Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR: **Molecular evolutionary relationships of enteroinvasive Escherichia coli and Shigella spp**. *Infect Immun* 2004, **72**.
71. Ashkenazi S: **Shigella infections in children: New insights**. *Seminars in Pediatric Infectious Diseases* 2004, **15**(4):246-252.
72. Williams PCM, Berkley JA: **Guidelines for the treatment of dysentery (shigellosis): a systematic review of the evidence**. *Paediatrics and international child health* 2018, **38**(sup1):S50-S65.

73. Bhattacharya D, Bhattacharya H, Sayi DS, Bharadwaj AP, Singhania M, Sugunan AP, Roy S: **Changing patterns and widening of antibiotic resistance in Shigella spp. over a decade (2000–2011), Andaman Islands, India**. *Epidemiology and Infection* 2015, **143**(3):470-477.
74. Kaminski RW, Oaks EV: **Inactivated and subunit vaccines to prevent shigellosis**. *Expert Review of Vaccines* 2009, **8**(12):1693-1704.
75. Van de Verg LL, Venkatesan MM: **Editorial Commentary: A Shigella Vaccine Against Prevalent Serotypes**. *Clinical Infectious Diseases* 2014, **59**(7):942-943.
76. Barry EM, Pasetti MF, Sztein MB, Fasano A, Kotloff KL, Levine MM: **Progress and pitfalls in Shigella vaccine research**. *Nat Rev Gastroenterol Hepatol* 2013, **10**(4):245-255.
77. Bélot F, Guerreiro C, Baleux F, Mulard LA: **Synthesis of two linear PADRE conjugates bearing a deca- or pentadecasaccharide B epitope as potential synthetic vaccines against Shigella flexneri serotype 2a infection**. *Chemistry - A European Journal* 2005, **11**(5):1625-1635.
78. Kämpf MM, Braun M, Sirena D, Ihssen J, Thöny-Meyer L, Ren Q: **In vivo production of a novel glycoconjugate vaccine against Shigella flexneri 2a in recombinant Escherichia coli: Identification of stimulating factors for in vivo glycosylation**. *Microbial Cell Factories* 2015, **14**(1).
79. Pore D, Chowdhury P, Mahata N, Pal A, Yamasaki S, Mahalanabis D, Chakrabarti MK: **Purification and characterization of an immunogenic outer membrane protein of Shigella flexneri 2a**. *Vaccine* 2009, **27**(42):5855-5864.
80. Frirdich E, Whitfield C: **Review: Lipopolysaccharide inner core oligosaccharide structure and outer membrane stability in human pathogens belonging to the Enterobacteriaceae**. *Journal of Endotoxin Research* 2005, **11**(3):133-144.
81. Raetz CRH, Whitfield C: **Lipopolysaccharide endotoxins**. *Annual review of biochemistry* 2002, **71**:635-700.
82. Knirel YA, Kondakova AN, Vinogradov E, Lindner B, Perepelov AV, Shashkov AS: **Lipopolysaccharide core structures and their correlation with genetic groupings of Shigella strains. A novel core variant in Shigella boydii type 16**. *Glycobiology* 2011, **21**(10):1362-1372.
83. West NP, Sansonetti P, Mounier J, Exley RM, Parsot C, Guadagnini S, Prévost M-C, Prochnicka-Chalufour A, Delepierre M, Tanguy M *et al*: **Optimization of Virulence Functions Through Glucosylation of Shigella LPS**. *Science* 2005, **307**(5713):1313-1317.
84. Lerouge I, Vanderleyden J: **O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions**. *FEMS Microbiol Rev* 2002, **26**(1):17-47.
85. Rosa Eugenia Reyes CRGl, Rafael Coria Jiménez, Maribel Ortiz Herrera and Alejandra Aquino Andrade (October 31st 2012). Mechanisms of O-Antigen Structural Variation of Bacterial Lipopolysaccharide (LPS), The Complex World of Polysaccharides, Desiree Nedra Karunaratne, IntechOpen, DOI: 10.5772/48147. Available from: https://www.intechopen.com/books/the-complex-world-of-polysaccharides/mechanisms-of-o-antigen-structural-variation-of-bacterial-lipopolysaccharide-lps-.
86. Brockhausen I, Hu B, Liu B, Lau K, Szarek WA, Wang L, Feng L: **Characterization of Two β-1,3-Glucosyltransferases from Escherichia coli Serotypes O56 and O152**. *Journal of Bacteriology* 2008, **190**(14):4922-4932.
87. Whitfield C: **Biosynthesis of lipopolysaccharide O antigens**. *Trends in Microbiology* 1995, **3**(5):178-185.
88. Carter JA, Blondel CJ, Zaldívar M, Álvarez SA, Marolda CL, Valvano MA, Contreras I: **O-antigen modal chain length in Shigella flexneri 2a is growth-regulated through RfaH-mediated transcriptional control of the wzy gene**. *Microbiology* 2007, **153**(10):3499-3507.
89. Kalynych S, Morona R, Cygler M: **Progress in understanding the assembly process of bacterial O-antigen**. *FEMS Microbiology Reviews* 2014, **38**(5):1048-1065.

90. Owens TW, Taylor RJ, Pahil KS, Bertani BR, Ruiz N, Kruse AC, Kahne D: **Structural basis of unidirectional export of lipopolysaccharide to the cell surface**. *Nature* 2019, **567**(7749):550-553.
91. Kos V, Cuthbertson L, Whitfield C: **The Klebsiella pneumoniae O2a Antigen Defines a Second Mechanism for O Antigen ATP-binding Cassette Transporters**. *Journal of Biological Chemistry* 2009, **284**(5):2947-2956.
92. Alf AL, Anders K, Weintraub A: **The Lipopolysaccharide of Shigella Bacteria as a Virulence Factor**. *Reviews of Infectious Diseases* 1991, **13**:S279-S284.
93. Mavris M, Manning PA, Morona R: **Mechanism of bacteriophage SfII-mediated serotype conversion in Shigella flexneri**. *Molecular Microbiology* 1997, **26**(5):939-950.
94. Guan S, Bastin DA, Verma NK: **Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX**. *Microbiology* 1999, **145**(5):1263-1273.
95. Stagg RM, Tang S-S, Carlin NIA, Talukder KA, Cam PD, Verma NK: **A Novel Glucosyltransferase Involved in O-Antigen Modification of Shigella flexneri Serotype 1c**. *Journal of Bacteriology* 2009, **191**(21):6612-6617.
96. Sun Q, Lan R, Wang Y, Wang J, Wang Y, Li P, Du P, Xu J: **Isolation and genomic characterization of SfI, a serotype-converting bacteriophage of Shigella flexneri**. *BMC Microbiology* 2013, **13**(1):39.
97. Jakhetia R, Talukder K, Verma N: **Isolation, characterization and comparative genomics of bacteriophage SfIV: a novel serotype converting phage from Shigella flexneri**. *BMC Genomics* 2013, **14**(1):677.
98. Korres H, Mavris M, Morona R, Manning PA, Verma NK: **Topological analysis of GtrA and GtrB proteins encoded by the serotype-converting cassette of Shigella flexneri**. *Biochemical and Biophysical Research Communications* 2005, **328**(4):1252-1260.
99. Ramiscal RR, Tang SS, Korres H, Verma NK: **Structural and functional divergence of the newly identified GtrIc from its Gtr family of conserved Shigella flexneri serotype-converting glucosyltransferases**. *Mol Membr Biol* 2010, **27**(2-3):114-122.
100. Guan S, Bastin DA, Verma NK: **Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX**. *Microbiology (Reading)* 1999, **145 (Pt 5)**:1263-1273.
101. Gemski P, Jr., Koeltzow DE, Formal SB: **Phage conversion of Shigella flexneri group antigens**. *Infect Immun* 1975, **11**(4):685-691.
102. Verma NK, Brandt JM, Verma DJ, Lindberg AA: **Molecular characterization of the O-acetyl transferase gene of converting bacteriophage SF6 that adds group antigen 6 to Shigella flexneri**. *Molecular Microbiology* 1991, **5**(1):71-75.
103. Clark CA, Beltrame J, Manning PA: **The oac gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of Shigella flexneri bacteriophage Sf6**. *Gene* 1991, **107**(1):43-52.
104. Casjens S, Winn-Stapley DA, Gilcrease EB, Morona R, Kühlewein C, Chua JEH, Manning PA, Inwood W, Clark AJ: **The Chromosome of Shigella flexneri Bacteriophage Sf6: Complete Nucleotide Sequence, Genetic Mosaicism, and DNA Packaging**. *Journal of Molecular Biology* 2004, **339**(2):379-394.
105. Jakhetia R, Marri A, Stahle J, Widmalm G, Verma N: **Serotype-conversion in Shigella flexneri: identification of a novel bacteriophage, Sf101, from a serotype 7a strain**. *BMC Genomics* 2014, **15**(1):742.
106. Wang J, Knirel YA, Lan R, Senchenkova SN, Luo X, Perepelov AV, Wang Y, Shashkov AS, Xu J, Sun Q: **Identification of an O-acyltransferase gene (oacB) that mediates 3- and 4-O-acetylation of rhamnose III in Shigella flexneri O antigens**. *J Bacteriol* 2014, **196**(8):1525-1531.
107. Knirel YA, Wang J, Luo X, Senchenkova SN, Lan R, Shpirt AM, Du P, Shashkov AS, Zhang N, Xu J *et al*: **Genetic and structural identification of an O-acyltransferase gene (oacC)**

responsible for the 3/4-O-acetylation on rhamnose III in Shigella flexneri serotype 6. *BMC Microbiol* 2014, **14**:266.
108. Sun Q, Knirel YA, Wang J, Luo X, Senchenkova SN, Lan R, Shashkov AS, Xu J: **Serotype-converting bacteriophage SfII encodes an acyltransferase protein that mediates 6-O-acetylation of GlcNAc in Shigella flexneri O-antigens, conferring on the host a novel O-antigen epitope**. *J Bacteriol* 2014, **196**(20):3656-3666.
109. Sun Q, Lan R, Wang Y, Wang J, Xia S, Wang Y, Zhang J, Yu D, Li Z, Jing H *et al*: **Identification of a divergent O-acetyltransferase gene oac 1b from Shigella flexneri serotype 1b strains**. *Emerg Microbes Infect* 2012, **1**(9):e21-e21.
110. Pearson CR, Tindall SN, Herman R, Jenkins HT, Bateman A, Thomas GH, Potts JR, Van der Woude MW: **Acetylation of Surface Carbohydrates in Bacterial Pathogens Requires Coordinated Action of a Two-Domain Membrane-Bound Acyltransferase**. *mBio* 2020, **11**(4):e01364-01320.
111. thanweer f: **identification of critical residues of Oac**.
112. Sun Q, Knirel YA, Lan R, Wang J, Senchenkova SyN, Jin D, Shashkov AS, Xia S, Perepelov AV, Chen Q *et al*: **A Novel Plasmid-Encoded Serotype Conversion Mechanism through Addition of Phosphoethanolamine to the O-Antigen of Shigella flexneri**. *PLOS ONE* 2012, **7**(9):e46095.
113. Knirel YA, Sun Q, Senchenkova SN, Perepelov AV, Shashkov AS, Xu J: **O-Antigen modifications providing antigenic diversity of Shigella flexneri and underlying genetic mechanisms**. *Biochemistry (Moscow)* 2015, **80**(7):901-914.
114. Sun Q, Knirel YA, Lan R, Wang J, Senchenkova SN, Shashkov AS, Wang Y, Wang Y, Luo X, Xu J: **Dissemination and serotype modification potential of pSFxv_2, an O-antigen PEtN modification plasmid in Shigella flexneri**. *Glycobiology* 2014, **24**(3):305-313.
115. Foster RA, Carlin NIA, Majcher M, Tabor H, Ng LK, Widmalm G: **Structural elucidation of the O-antigen of the Shigella flexneri provisional serotype 88-893: structural and serological similarities with S. flexneri provisional serotype Y394 (1c)**. *Carbohydrate Research* 2011, **346**(6):872-876.
116. Parajuli P, Adamski M, Verma NK: **Bacteriophages are the major drivers of Shigella flexneri serotype 1c genome plasticity: a complete genome analysis**. *BMC Genomics* 2017, **18**(1):722.
117. Ahmed K, Shakoori FR, Shakoori AR: **Aetiology of shigellosis in northern Pakistan**. *J Health Popul Nutr* 2003, **21**.
118. Qiu S, Wang Z, Chen C, Liu N, Jia L, Liu W, Wang L, Hao R, Zhang L, Wang Y *et al*: **Emergence of a novel Shigella flexneri serotype 4s strain that evolved from a serotype X variant in China**. *J Clin Microbiol* 2011, **49**(3):1148-1150.
119. Perepelov AV, L'vov VL, Liu B, Senchenkova SyN, Shekht ME, Shashkov AS, Feng L, Aparin PG, Wang L, Knirel YA: **A similarity in the O-acetylation pattern of the O-antigens of Shigella flexneri types 1a, 1b, and 2a**. *Carbohydrate Research* 2009, **344**(5):687-692.
120. Sun Q, Lan R, Wang Y, Wang J, Xia S, Wang Y, Zhang J, Yu D, Li Z, Jing H *et al*: **Identification of a divergent O-acetyltransferase gene oac1b from Shigella flexneri serotype 1b strains**. *Emerging Microbes & Infections* 2012, **1**(1):1-7.
121. Allison GE, Angeles D, Tran-Dinh N, Verma NK: **Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of Shigella flexneri**. *Journal of bacteriology* 2002, **184**(7):1974-1987.
122. Clokie MR, Millard AD, Letarov AV, Heaphy S: **Phages in nature**. *Bacteriophage* 2011, **1**(1):31-45.
123. Summers WC: **Bacteriophage research: early history**. *Bacteriophages: Biology and applications* 2005:5-27.
124. Black LW, Thomas JA: **Condensed Genome Structure**. In: *Viral Molecular Machines.* Edited by Rossmann MG, Rao VB. Boston, MA: Springer US; 2012: 469-487.

125. Penadés JR, Chen J, Quiles-Puchalt N, Carpena N, Novick RP: **Bacteriophage-mediated spread of bacterial virulence genes**. *Current Opinion in Microbiology* 2015, **23**:171-178.
126. Davies EV, Winstanley C, Fothergill JL, James CE: **The role of temperate bacteriophages in bacterial infection**. *FEMS Microbiology Letters* 2016, **363**(5).
127. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB: **Lysogeny in nature: mechanisms, impact and ecology of temperate phages**. *The ISME Journal* 2017, **11**(7):1511-1520.
128. Brüssow H, Canchaya C, Hardt W-D: **Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion**. *Microbiology and molecular biology reviews : MMBR* 2004, **68**(3):560-602.
129. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han C-G, Ohtsubo E, Nakayama K, Murata T *et al*: **Complete Genome Sequence of Enterohemorrhagic Eschelichia coli O157:H7 and Genomic Comparison with a Laboratory Strain K-12**. *DNA Research* 2001, **8**(1):11-22.
130. Boyd EF, Brüssow H: **Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved**. *Trends in Microbiology* 2002, **10**(11):521-529.
131. Abedon ST, Lejeune JT: **Why bacteriophage encode exotoxins and other virulence factors**. *Evolutionary bioinformatics online* 2007, **1**:97-110.
132. Waldor MK, Mekalanos JJ: **Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin**. *Science* 1996, **272**(5270):1910-1914.
133. Boyd EF, Davis BM, Hochhut B: **Bacteriophage–bacteriophage interactions in the evolution of pathogenic bacteria**. *Trends in Microbiology* 2001, **9**(3):137-144.
134. Melton-Celsa AR: **Shiga Toxin (Stx) Classification, Structure, and Function**. *Microbiology spectrum* 2014, **2**(4):10.1128/microbiolspec.EHEC-0024-2013-2013.
135. Bergan J, Dyve Lingelem AB, Simm R, Skotland T, Sandvig K: **Shiga toxins**. *Toxicon* 2012, **60**(6):1085-1107.
136. Mirold S, Rabsch W, Rohde M, Stender S, Tschäpe H, Rüssmann H, Igwe E, Hardt W-D: **Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic Salmonella typhimurium strain**. *Proceedings of the National Academy of Sciences* 1999, **96**(17):9845-9850.
137. Stanley TL, Ellermeier CD, Slauch JM: **Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects Salmonella enterica serovar typhimurium survival in Peyer's patches**. *Journal of bacteriology* 2000, **182**(16):4406-4413.
138. Figueroa-Bossi N, Uzzau S, Maloriol D, Bossi L: **Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella**. *Molecular Microbiology* 2001, **39**(2):260-272.
139. Ehrbar K, Hardt W-D: **Bacteriophage-encoded type III effectors in Salmonella enterica subspecies 1 serovar Typhimurium**. *Infection, Genetics and Evolution* 2005, **5**(1):1-9.
140. Barondess JJ, Beckwith J: **A bacterial virulence determinant encoded by lysogenic coliphage lambda**. *Nature* 1990, **346**(6287):871-874.
141. George DT, Stephenson DP, Tran E, Morona R, Verma NK: **Complete Genome Sequence of SfII, a Serotype-Converting Bacteriophage of the Highly Prevalent Shigella flexneri Serotype 2a**. *Genome announcements* 2013, **1**(5):e00626-00613.
142. Jakhetia R, Talukder KA, Verma NK: **Isolation, characterization and comparative genomics of bacteriophage SfIV: a novel serotype converting phage from Shigella flexneri**. *BMC Genomics* 2013, **14**(1):677.
143. Jakhetia R, Verma NK: **Identification and Molecular Characterisation of a Novel Mu-Like Bacteriophage, SfMu, of Shigella flexneri**. *PLOS ONE* 2015, **10**(4):e0124053.
144. Holmes RK, Barksdale L: **Genetic analysis of tox+ and tox- bacteriophages of Corynebacterium diphtheriae**. *Journal of virology* 1969, **3**(6):586-598.

145. Wagner PL, Neely MN, Zhang X, Acheson DW, Waldor MK, Friedman DI: **Role for a phage promoter in Shiga toxin 2 expression from a pathogenic Escherichia coli strain**. *Journal of bacteriology* 2001, **183**(6):2081-2085.
146. Plunkett G, 3rd, Rose DJ, Durfee TJ, Blattner FR: **Sequence of Shiga toxin 2 phage 933W from Escherichia coli O157:H7: Shiga toxin as a phage late-gene product**. *Journal of bacteriology* 1999, **181**(6):1767-1778.
147. Lavigne J-P, Blanc-Potard A-B: **Molecular evolution of Salmonella enterica serovar Typhimurium and pathogenic Escherichia coli: From pathogenesis to therapeutics**. *Infection, Genetics and Evolution* 2008, **8**(2):217-226.
148. Vaca Pacheco S, Garcıća González O, Paniagua Contreras GL: **The lom gene of bacteriophage λ is involved in Escherichia coli K12 adhesion to human buccal epithelial cells**. *FEMS Microbiology Letters* 1997, **156**(1):129-132.
149. Barondess JJ, Beckwith J: **bor gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein**. *Journal of bacteriology* 1995, **177**(5):1247-1253.
150. Flockhart AF, Tree JJ, Xu X, Karpiyevich M, McAteer SP, Rosenblum R, Shaw DJ, Low CJ, Best A, Gannon V *et al*: **Identification of a novel prophage regulator in Escherichia coli controlling the expression of type III secretion**. *Molecular microbiology* 2012, **83**(1):208-223.
151. Holloway BW, Cooper GN: **Lysogenic conversion in Pseudomonas aeruginosa**. *Journal of bacteriology* 1962, **84**(6):1321-1324.
152. Bensing BA, Siboo IR, Sullam PM: **Proteins PblA and PblB of Streptococcus mitis, which promote binding to human platelets, are encoded within a lysogenic bacteriophage**. *Infection and immunity* 2001, **69**(10):6186-6192.
153. Clark CG, Grant CCR, Pollari F, Marshall B, Moses J, Tracz DM, Gilmour MW: **Effects of the Campylobacter jejuni CJIE1 prophage homologs on adherence and invasion in culture, patient symptoms, and source of infection**. *BMC microbiology* 2012, **12**:269-269.
154. Faruque SM, Asadulghani, Alim AR, Albert MJ, Islam KM, Mekalanos JJ: **Induction of the lysogenic phage encoding cholera toxin in naturally occurring strains of toxigenic Vibrio cholerae O1 and O139**. *Infection and immunity* 1998, **66**(8):3752-3757.
155. Wang H, Yang CH, Lee G, Chang F, Wilson H, del Campillo-Campbell A, Campbell A: **Integration specificities of two lambdoid phages (21 and e14) that insert at the same attB site**. *Journal of bacteriology* 1997, **179**(18):5705-5711.
156. Campbell AM: **Chromosomal insertion sites for phages and plasmids**. *Journal of bacteriology* 1992, **174**(23):7495-7499.
157. Menouni R, Hutinet G, Petit M-A, Ansaldi M: **Bacterial genome remodeling through bacteriophage recombination**. *FEMS Microbiology Letters* 2015, **362**(1):1-10.
158. Strauch E, Hammerl JA, Konietzny A, Schneiker-Bekel S, Arnold W, Goesmann A, Pühler A, Beutin L: **Bacteriophage 2851 is a prototype phage for dissemination of the Shiga toxin variant gene 2c in Escherichia coli O157:H7**. *Infect Immun* 2008, **76**(12):5466-5477.
159. Knirel Y, Lan R, Senchenkova Sy, Wang J, Shashkov A, Wang Y, Perepelov A, Xiong Y, Xu J, Sun Q: **O-antigen structure of Shigella flexneri serotype Yv and effect of the lpt-O gene variation on phosphoethanolamine modification of S. flexneri O-antigens**. *Glycobiology* 2013, **23**.
160. Nair A, Korres H, Verma NK: **Topological characterisation and identification of critical domains within glucosyltransferase IV (GtrIV) of Shigella flexneri**. *BMC biochemistry* 2011, **12**:67-67.
161. Lehane AM, Korres H, Verma NK: **Bacteriophage-encoded glucosyltransferase GtrII of Shigella flexneri: membrane topology and identification of critical residues**. *The Biochemical journal* 2005, **389**(Pt 1):137-143.

162. Rusden AD, Stephenson DP, Verma NK: **Topological investigation of glucosyltransferase V in Shigella flexneri using the substituted cysteine accessibility method**. *Biochemistry* 2013, **52**(15):2655-2661.
163. Breton C, Imberty A: **Structure/function studies of glycosyltransferases**. *Current Opinion in Structural Biology* 1999, **9**(5):563-571.
164. Alexeyev MF, Winkler HH: **Membrane topology of the Rickettsia prowazekii ATP/ADP translocase revealed by novel dual pho-lac reporters**. *J Mol Biol* 1999, **285**(4):1503-1513.
165. Korres H, Verma NK: **Identification of essential loops and residues of glucosyltransferase V (GtrV) of Shigella flexneri**. *Mol Membr Biol* 2006, **23**(5):407-419.
166. Bogdanov M, Zhang W, Xie J, Dowhan W: **Transmembrane protein topology mapping by the substituted cysteine accessibility method (SCAMTM): Application to lipid-specific membrane protein topogenesis**. *Methods* 2005, **36**(2):148-171.
167. Thanweer F, Tahiliani V, Korres H, Verma NK: **Topology and identification of critical residues of the O-acetyltransferase of serotype-converting bacteriophage, SF6, of Shigella flexneri**. *Biochem Biophys Res Commun* 2008, **375**(4):581-585.
168. Thanweer F, Verma NK: **Identification of critical residues of the serotype modifying O-acetyltransferase of Shigella flexneri**. *BMC biochemistry* 2012, **13**:13-13.
169. Daley DO, Rapp M, Granseth E, Melen K, Drew D, von Heijne G: **Global topology analysis of the Escherichia coli inner membrane proteome**. *Science* 2005, **308**(5726):1321-1323.
170. Islam ST, Lam JS: **Topological mapping methods for α-helical bacterial membrane proteins--an update and a guide**. *MicrobiologyOpen* 2013, **2**(2):350-364.
171. Liapakis G: **Obtaining structural and functional information for GPCRs using the substituted-cysteine accessibility method (SCAM)**. *Curr Pharm Biotechnol* 2014, **15**(10):980-986.
172. Schmidt JJ: **DNA cloning: A practical approach Volumes 1 and 2: Edited by D M Glover. pp 190 and 245. IRL Press, Oxford. 1985. £13.50 each or £24 for the two ISBN 0-947945-18-7 and-19-5**. 1986, **14**(2):91-91.
173. Brenner S: **The genetics of Caenorhabditis elegans**. *Genetics* 1974, **77**(1):71-94.
174. Sambrook J, Fritsch EF, Maniatis T: **Molecular cloning: a laboratory manual**. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.
175. Adams MM, Allison GE, Verma NK: **Type IV O antigen modification genes in the genome of Shigella flexneri NCTC 8296**. *Microbiology* 2001, **147**(4):851-860.
176. Santos SB, Carvalho CM, Sillankorva S, Nicolau A, Ferreira EC, Azeredo J: **The use of antibiotics to improve phage detection and enumeration by the double-layer agar technique**. *BMC Microbiol* 2009, **9**:148.
177. Wick RR, Judd LM, Gorrie CL, Holt KE: **Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads**. *PLOS Computational Biology* 2017, **13**(6):e1005595.
178. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK *et al*: **Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement**. *PloS one* 2014, **9**(11):e112963.
179. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M *et al*: **The RAST Server: rapid annotations using subsystems technology**. *BMC Genomics* 2008, **9**:75.
180. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS: **PHASTER: a better, faster version of the PHAST phage search tool**. *Nucleic Acids Res* 2016, **44**(W1):W16-21.
181. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M: **ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes**. *Genome Biol* 2011, **12**(3):R30.
182. von Heijne G: **Membrane-protein topology**. *Nature Reviews Molecular Cell Biology* 2006, **7**(12):909-918.

183. Bernsel A, Viklund H, Hennerdal A, Elofsson A: **TOPCONS: consensus prediction of membrane protein topology**. *Nucleic acids research* 2009, **37**(Web Server issue):W465-W468.
184. Rosano GL, Ceccarelli EA: **Recombinant protein expression in Escherichia coli: advances and challenges**. *Frontiers in microbiology* 2014, **5**:172-172.
185. Marshall SS, Niesen MJM, Müller A, Tiemann K, Saladi SM, Galimidi RP, Zhang B, Clemons WM, Jr., Miller TF, 3rd: **A Link between Integral Membrane Protein Expression and Simulated Integration Efficiency**. *Cell reports* 2016, **16**(8):2169-2177.
186. Lin WJ, Huang SW, Chou CP: **DegP-coexpression minimizes inclusion-body formation upon overproduction of recombinant penicillin acylase in Escherichia coli**. *Biotechnol Bioeng* 2001, **73**(6):484-492.
187. Bass JJ, Wilkinson DJ, Rankin D, Phillips BE, Szewczyk NJ, Smith K, Atherton PJ: **An overview of technical considerations for Western blotting applications to physiological research**. *Scandinavian journal of medicine & science in sports* 2017, **27**(1):4-25.
188. Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM: **Detergent binding explains anomalous SDS-PAGE migration of membrane proteins**. *Proceedings of the National Academy of Sciences* 2009, **106**(6):1760-1765.
189. MIYAKE J, OCHIAI-YANAGI S, KASUMI T, TAKAGI T: **Isolation of a Membrane Protein from R. rubrum Chromatophores and Its Abnormal Behavior in SDS-Polyacrylamide Gel Electrophoresis Due to a High Binding Capacity for SDS1**. *The Journal of Biochemistry* 1978, **83**(6):1679-1686.
190. Mahmood T, Yang P-C: **Western blot: Technique, theory, and trouble shooting**. *North American Journal of Medical Sciences* 2012, **4**(9):429-434.
191. Knirel YA, Wang J, Luo X, Senchenkova SN, Lan R, Shpirt AM, Du P, Shashkov AS, Zhang N, Xu J *et al*: **Genetic and structural identification of an O-acyltransferase gene (oacC) responsible for the 3/4-O-acetylation on rhamnose III in Shigella flexneri serotype 6**. *BMC microbiology* 2014, **14**:266-266.
192. Perepelov AV, Shekht ME, Liu B, Shevelev SD, Ledov VA, Senchenkova SN, L'Vov V L, Shashkov AS, Feng L, Aparin PG *et al*: **Shigella flexneri O-antigens revisited: final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity**. *FEMS Immunol Med Microbiol* 2012, **66**(2):201-210.
193. Sun Q, Lan R, Wang Y, Wang J, Wang Y, Li P, Du P, Xu J: **Isolation and genomic characterization of SfI, a serotype-converting bacteriophage of Shigella flexneri**. *BMC Microbiol* 2013, **13**:39-39.
194. Sun Q, Knirel YA, Wang J, Luo X, Senchenkova SN, Lan R, Shashkov AS, Xu J: **Serotype-Converting Bacteriophage SfII Encodes an Acyltransferase Protein That Mediates 6-O-Acetylation of GlcNAc in Shigella flexneri O-Antigens, Conferring on the Host a Novel O-Antigen Epitope**. *Journal of Bacteriology* 2014, **196**(20):3656-3666.
195. Wang J, Lan R, Knirel YA, Luo X, Senchenkova SyN, Shashkov AS, Xu J, Sun Q: **Serological identification and prevalence of a novel O-antigen epitope linked to 3- and 4-O-acetylated rhamnose III of lipopolysaccharide in Shigella flexneri**. *Journal of clinical microbiology* 2014, **52**(6):2033-2038.
196. Moynihan PJ, Clarke AJ: **O-acetylation of peptidoglycan in gram-negative bacteria: identification and characterization of peptidoglycan O-acetyltransferase in Neisseria gonorrhoeae**. *The Journal of biological chemistry* 2010, **285**(17):13264-13273.
197. Bera A, Herbert S, Jakob A, Vollmer W, Götz F: **Why are pathogenic staphylococci so lysozyme resistant? The peptidoglycan O-acetyltransferase OatA is the major determinant for lysozyme resistance of Staphylococcus aureus**. *Mol Microbiol* 2005, **55**(3):778-787.
198. Sychantha D, Little DJ, Chapman RN, Boons GJ, Robinson H, Howell PL, Clarke AJ: **PatB1 is an O-acetyltransferase that decorates secondary cell wall polysaccharides**. *Nat Chem Biol* 2018, **14**(1):79-85.

199. Kintz E, Davies MR, Hammarlöf DL, Canals R, Hinton JCD, van der Woude MW: **A BTP1 prophage gene present in invasive non-typhoidal Salmonella determines composition and length of the O-antigen of the lipopolysaccharide**. *Molecular microbiology* 2015, **96**(2):263-275.
200. Jones CS, Sychantha D, Howell PL, Clarke AJ: **Structural basis for the O-acetyltransferase function of the extracytoplasmic domain of OatA from Staphylococcus aureus**. *Journal of Biological Chemistry* 2020, **295**(24):8204-8213.
201. Thanweer F, Verma NK: **Identification of critical residues of the serotype modifying O-acetyltransferase of Shigella flexneri**. *BMC Biochem* 2012, **13**:13.
202. Lairson LL, Henrissat B, Davies GJ, Withers SG: **Glycosyltransferases: Structures, Functions, and Mechanisms**. *Annual Review of Biochemistry* 2008, **77**(1):521-555.
203. Kintz E, Davies MR, Hammarlof DL, Canals R, Hinton JC, van der Woude MW: **A BTP1 prophage gene present in invasive non-typhoidal Salmonella determines composition and length of the O-antigen of the lipopolysaccharide**. *Mol Microbiol* 2015, **96**(2):263-275.
204. Watanabe M, Sofuni T, Nohmi T: **Involvement of Cys69 residue in the catalytic mechanism of N-hydroxyarylamine O-acetyltransferase of Salmonella typhimurium. Sequence similarity at the amino acid level suggests a common catalytic mechanism of acetyltransferase for S. typhimurium and higher organisms**. *Journal of Biological Chemistry* 1992, **267**(12):8429-8436.
205. Englert M, Nakamura A, Wang Y-S, Eiler D, Söll D, Guo L-T: **Probing the active site tryptophan of Staphylococcus aureus thioredoxin with an analog**. *Nucleic Acids Research* 2015, **43**(22):11061-11067.
206. Gao C, Lan D, Liu L, Zhang H, Yang B, Wang Y: **Site-directed mutagenesis studies of the aromatic residues at the active site of a lipase from Malassezia globosa**. *Biochimie* 2014, **102**:29-36.
207. Shmara A, Weinsetel N, Dery KJ, Chavideh R, Tolmasky ME: **Systematic Analysis of a Conserved Region of the Aminoglycoside 6 N-Acetyltransferase Type Ib**. *Antimicrobial Agents and Chemotherapy* 2001, **45**(12):3287-3292.
208. Dyda F, Klein DC, Hickman AB: **GCN5-Related N-Acetyltransferases: A Structural Overview**. *Annual Review of Biophysics and Biomolecular Structure* 2000, **29**(1):81-103.
209. Goodfellow GH, Dupret JM, Grant DM: **Identification of amino acids imparting acceptor substrate selectivity to human arylamine acetyltransferases NAT1 and NAT2**. *Biochem J* 2000, **348 Pt 1**(Pt 1):159-166.
210. Bardy SL, Ng SYM, Carnegie DS, Jarrell KF: **Site-directed mutagenesis analysis of amino acids critical for activity of the type I signal peptidase of the archaeon Methanococcus voltae**. *Journal of bacteriology* 2005, **187**(3):1188-1191.
211. Slauch JM, Lee AA, Mahan MJ, Mekalanos JJ: **Molecular characterization of the oafA locus responsible for acetylation of Salmonella typhimurium O-antigen: oafA is a member of a family of integral membrane trans-acylases**. *Journal of Bacteriology* 1996, **178**(20):5904-5909.
212. Moscoso JA, Korres H, George DT, Verma NK: **Identification of active site residues in the Shigella flexneri glucosyltransferase GtrV**. *Mol Membr Biol* 2010, **27**(2-3):104-113.
213. Sunden F, Peck A, Salzman J, Ressl S, Herschlag D: **Extensive site-directed mutagenesis reveals interconnected functional units in the alkaline phosphatase active site**. *eLife* 2015, **4**:e06181.
214. Sychantha D, Clarke AJ: **Peptidoglycan Modification by the Catalytic Domain of Streptococcus pneumoniae OatA Follows a Ping-Pong Bi-Bi Mechanism of Action**. *Biochemistry* 2018, **57**(16):2394-2401.
215. Muthuirulandi Sethuvel DP, Veeraraghavan B, Vasudevan K, Devanga Ragupathi NK, Murugan D, Walia K, Anandan S: **Complete genome analysis of clinical Shigella strains**

reveals plasmid pSS1653 with resistance determinants: a triumph of hybrid approach. *Gut Pathogens* 2019, **11**(1):55.
216. Ferenci T: **The spread of a beneficial mutation in experimental bacterial populations: the influence of the environment and genotype on the fixation of rpoS mutations**. *Heredity* 2008, **100**(5):446-452.
217. Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W, Fournier G, Mayhew GF, Plunkett G, Rose DJ, Darling A: **Complete genome sequence and comparative genomics of Shigella flexneri serotype 2a strain 2457T**. *Infect Immun* 2003, **71**.
218. Hall JPJ, Harrison E, Baltrus DA: **Introduction: the secret lives of microbial mobile genetic elements**. 2022, **377**(1842):20200460.
219. Ramisetty BCM, Sudhakari PA: **Bacterial 'Grounded' Prophages: Hotspots for Genetic Renovation and Innovation**. *Frontiers in genetics* 2019, **10**:65-65.
220. Stagg RM, Tang S-S, Carlin NIA, Talukder KA, Cam PD, Verma NK: **A Novel Glucosyltransferase Involved in O-Antigen Modification of Shigella flexneri Serotype 1c**. 2009, **191**(21):6612-6617.
221. Parajuli P, Minh BQ, Verma NK: **Newly Emerged Serotype 1c of Shigella flexneri: Multiple Origins and Changing Drug Resistance Landscape**. *Genes* 2020, **11**(9):1042.
222. Kesika P, Karutha Pandian S, Balamurugan K: **Analysis of Shigella flexneri-mediated infections in model organism Caenorhabditis elegans**. *Scand J Infect Dis* 2011, **43**(4):286-295.
223. Joshua GWP, Karlyshev AV, Smith MP, Isherwood KE, Titball RW, Wren BW: **A Caenorhabditis elegans model of Yersinia infection: biofilm formation on a biotic surface**. 2003, **149**(11):3221-3229.
224. George DT, Behm CA, Hall DH, Mathesius U, Rug M, Nguyen KCQ, Verma NK: **Shigella flexneri Infection in Caenorhabditis elegans: Cytopathological Examination and Identification of Host Responses**. *PloS one* 2014, **9**(9):e106085.
225. Somasiri P, Behm CA, Adamski M, Wen J, Verma NK: **Transcriptional response of Caenorhabditis elegans when exposed to Shigella flexneri**. *Genomics* 2020, **112**(1):774-781.
226. Ermolaeva MA, Schumacher B: **Insights from the worm: The C. elegans model for innate immunity**. *Seminars in Immunology* 2014, **26**(4):303-309.
227. Chen K, Franz CJ, Jiang H, Jiang Y, Wang DJBg: **An evolutionarily conserved transcriptional response to viral infection in Caenorhabditis nematodes**. 2017, **18**(1):1-10.
228. Kusuma SAF, Parwati I, Rostinawati T, Yusuf M, Fadhlillah M, Ahyudanari RR, Rukayadi Y, Subroto T: **Optimization of culture conditions for Mpt64 synthetic gene expression in Escherichia coli BL21 (DE3) using surface response methodology**. *Heliyon* 2019, **5**(11):e02741-e02741.
229. Guzman LM, Belin D, Carson MJ, Beckwith J: **Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter**. *J Bacteriol* 1995, **177**(14):4121-4130.
230. SHOJA AS, Varedi KS, Babaeipour V, Farnoud AM: **Recent advances in high cell density cultivation for production of recombinant protein**. 2008.
231. Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q: **Practical protocols for production of very high yields of recombinant proteins using Escherichia coli**. *Protein science : a publication of the Protein Society* 2009, **18**(5):936-948.
232. Donovan RS, Robinson CW, Glick BR: **Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter**. *J Ind Microbiol* 1996, **16**(3):145-154.
233. Baneyx F: **Recombinant protein expression in Escherichia coli**. *Current opinion in biotechnology* 1999, **10**(5):411-421.
234. Hannig G, Makrides SC: **Strategies for optimizing heterologous protein expression in Escherichia coli**. *Trends in Biotechnology* 1998, **16**(2):54-60.

235. Larentis AL, Nicolau JFMQ, Esteves GdS, Vareschini DT, de Almeida FVR, dos Reis MG, Galler R, Medeiros MA: **Evaluation of pre-induction temperature, cell growth at induction and IPTG concentration on the expression of a leptospiral protein in E. coli using shaking flasks and microbioreactor**. *BMC Research Notes* 2014, **7**(1):671.
236. Narayanan N, Hsieh M-Y, Xu Y, Chou CP: **Arabinose-Induction of lac-Derived Promoter Systems for Penicillin Acylase Production in Escherichia coli**. 2006, **22**(3):617-625.
237. Donovan RS, Robinson CW, Glick BR: **Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of thelac promoter**. *Journal of Industrial Microbiology* 1996, **16**(3):145-154.
238. Yuan X, Johnson MD, Zhang J, Lo AW, Schembri MA, Wijeyewickrema LC, Pike RN, Huysmans GHM, Henderson IR, Leyton DL: **Molecular basis for the folding of β-helical autotransporter passenger domains**. *Nature Communications* 2018, **9**(1):1395.
239. Tsuji Y: **Transmembrane protein western blotting: Impact of sample preparation on detection of SLC11A2 (DMT1) and SLC40A1 (ferroportin)**. *PLOS ONE* 2020, **15**(7):e0235563.
240. Lee Y-N, Chen L-K, Ma H-C, Yang H-H, Li H-P, Lo S-Y: **Thermal aggregation of SARS-CoV membrane protein**. *Journal of virological methods* 2005, **129**(2):152-161.
241. Kajimura J, Rahman A, Hsu J, Evans MR, Gardner KH, Rick PD: **O acetylation of the enterobacterial common antigen polysaccharide is catalyzed by the product of the yiaH gene of Escherichia coli K-12**. *J Bacteriol* 2006, **188**(21):7542-7550.
242. Yamaryo-Botte Y, Rainczuk AK, Lea-Smith DJ, Brammananth R, van der Peet PL, Meikle P, Ralton JE, Rupasinghe TW, Williams SJ, Coppel RL *et al*: **Acetylation of trehalose mycolates is required for efficient MmpL-mediated membrane transport in Corynebacterineae**. *ACS Chem Biol* 2015, **10**(3):734-746.
243. Vorhölter F-J, Schneiker S, Goesmann A, Krause L, Bekel T, Kaiser O, Linke B, Patschkowski T, Rückert C, Schmid J *et al*: **The genome of Xanthomonas campestris pv. campestris B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis**. *Journal of Biotechnology* 2008, **134**(1):33-45.
244. Hong Y, Duda KA, Cunneen MM, Holst O, Reeves PR: **The WbaK acetyltransferase of Salmonella enterica group E gives insights into O antigen evolution**. *Microbiology* 2013, **159**(Pt_11):2316-2322.
245. Geno KA, Saad JS, Nahm MH: **Discovery of Novel Pneumococcal Serotype 35D, a Natural WciG-Deficient Variant of Serotype 35B**. *Journal of Clinical Microbiology* 2017, **55**(5):1416-1425.
246. Moynihan PJ, Clarke AJ: **O-acetylation of peptidoglycan in gram-negative bacteria: identification and characterization of peptidoglycan O-acetyltransferase in Neisseria gonorrhoeae**. *The Journal of biological chemistry* 2010, **285**(17):13264-13273.
247. Allison HE: **Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens**. *Future Microbiol* 2007, **2**(2):165-174.
248. Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H: **Prophage genomics**. *Microbiol Mol Biol Rev* 2003, **67**(2):238-276.
249. Subramanian S, Parent KN, Doore SM: **Ecology, Structure, and Evolution of Shigella Phages**. *Annual review of virology* 2020, **7**(1):121-141.
250. Adhikari P, Allison G, Whittle B, Verma NK: **Serotype 1a O-antigen modification: molecular characterization of the genes involved and their novel organization in the Shigella flexneri chromosome**. *J Bacteriol* 1999, **181**(15):4711-4718.
251. Campbell A: **Prophage insertion sites**. *Research in Microbiology* 2003, **154**(4):277-282.
252. Groisman EA, Ochman H: **Pathogenicity Islands: Bacterial Evolution in Quantum Leaps**. *Cell* 1996, **87**(5):791-794.
253. Fogg PCM, Gossage SM, Smith DL, Saunders JR, McCarthy AJ, Allison HE: **Identification of multiple integration sites for Stx-phage Phi24B in the Escherichia coli genome, description**

of a novel integrase and evidence for a functional anti-repressor. *Microbiology (Reading, England)* 2007, **153**(Pt 12):4098-4110.
254. Bobay LM, Rocha EP, Touchon M: **The adaptation of temperate bacteriophages to their host genomes**. *Mol Biol Evol* 2013, **30**(4):737-751.
255. Serra-Moreno R, Jofre J, Muniesa M: **Insertion Site Occupancy by stx2 Bacteriophages Depends on the Locus Availability of the Host Strain Chromosome**. 2007, **189**(18):6645-6654.
256. Canchaya C, Fournous G, Brüssow H: **The impact of prophages on bacterial chromosomes**. *Mol Microbiol* 2004, **53**(1):9-18.
257. Brüssow H, Hendrix RW: **Phage Genomics: Small Is Beautiful**. *Cell* 2002, **108**(1):13-16.
258. Brüssow H, Canchaya C, Hardt WD: **Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion**. *Microbiol Mol Biol Rev* 2004, **68**(3):560-602, table of contents.
259. McDonough MA, Butterton JR: **Spontaneous tandem amplification and deletion of the shiga toxin operon in Shigella dysenteriae 1**. *Mol Microbiol* 1999, **34**(5):1058-1069.
260. Sasakawa C, Kamata K, Sakai T, Makino S, Yamada M, Okada N, Yoshikawa M: **Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in Shigella flexneri 2a**. *J Bacteriol* 1988, **170**(6):2480-2484.
261. Parajuli P, Rajput MI, Verma NK: **Plasmids of Shigella flexneri serotype 1c strain Y394 provide advantages to bacteria in the host**. *BMC Microbiology* 2019, **19**(1):86.
262. Zhang J, Liu H, Zhang X, Yang J, Yang F, Yang G, Shen Y, Hou Y, Jin Q: **Complete DNA sequence and gene analysis of the virulence plasmid pCP301 of Shigella flexneri 2a**. *Sci China C Life Sci* 2003, **46**(5):513-521.
263. Buchrieser C, Glaser P, Rusniok C, Nedjari H, D'Hauteville H, Kunst F, Sansonetti P, Parsot C: **The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri**. *Molecular Microbiology* 2000, **38**(4):760-771.
264. Meneely PM, Dahlberg CL, Rose JK: **Working with Worms: Caenorhabditis elegans as a Model Organism**. *Current Protocols Essential Laboratory Techniques* 2019, **19**(1):e35.
265. Tan MW, Mahajan-Miklos S, Ausubel FM: **Killing of Caenorhabditis elegans by Pseudomonas aeruginosa used to model mammalian bacterial pathogenesis**. *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(2):715-720.
266. Aballay A, Yorgey P, Ausubel FM: **Salmonella typhimurium proliferates and establishes a persistent infection in the intestine of Caenorhabditis elegans**. *Current Biology* 2000, **10**(23):1539-1542.
267. Sifri CD, Begun J, Ausubel FM, Calderwood SB: **Caenorhabditis elegans as a model host for Staphylococcus aureus pathogenesis**. *Infection and Immunity* 2003, **71**(4):2208-2217.
268. Sirivichayakul C, Thisyakorn U: **Severe shigellosis in childhood**. *The Southeast Asian journal of tropical medicine and public health* 1998, **29**(3):555-559.
269. Burton EA, Pendergast AM, Aballay A: **The Caenorhabditis elegans ABL-1 tyrosine kinase is required for Shigella flexneri pathogenesis**. *Appl Environ Microbiol* 2006, **72**(7):5043-5051.
270. Kamaladevi A, Balamurugan KJRa: **Lipopolysaccharide of Klebsiella pneumoniae attenuates immunity of Caenorhabditis elegans and evades by altering its supramolecular structure**. 2016, **6**(36):30070-30080.
271. Aballay A, Drenkard E, Hilbun LR, Ausubel FM: **Caenorhabditis elegans innate immune response triggered by Salmonella enterica requires intact LPS and is mediated by a MAPK signaling pathway**. *Current biology : CB* 2003, **13**(1):47-52.
272. Browning DF, Wells TJ, França FLS, Morris FC, Sevastsyanovich YR, Bryant JA, Johnson MD, Lund PA, Cunningham AF, Hobman JL *et al*: **Laboratory adapted Escherichia coli K-12 becomes a pathogen of Caenorhabditis elegans upon restoration of O antigen biosynthesis**. 2013, **87**(5):939-950.

273. Fung CC, Octavia S, Mooney A-M, Lan R: **Virulence variations in Shigella and enteroinvasive Escherichia coli using the Caenorhabditis elegans model**. *FEMS Microbiology Letters* 2015, **362**(3):1-5.
274. Sifri CD, Begun J, Ausubel FM: **The worm has turned--microbial virulence modeled in Caenorhabditis elegans**. *Trends Microbiol* 2005, **13**(3):119-127.
275. Guhathakurta B, Sasmal D, Ghosh AN, Kumar R, Saha P, Biswas D, Khetawat D, Datta A: **Adhesion and invasion of a mutant Shigella flexneri to an eukaryotic cell line in absence of the 220-kb virulence plasmid**. *FEMS Microbiol Lett* 1999, **181**(2):267-275.
276. Oaks EV, Wingfield ME, Formal SB: **Plaque formation by virulent Shigella flexneri**. *Infection and immunity* 1985, **48**(1):124-129.
277. Ranganathan S, Doucet M, Grassel CL, Delaine-Elias B, Zachos NC, Barry EM: **Evaluating Shigella flexneri Pathogenesis in the Human Enteroid Model**. *Infection and Immunity* 2019, **87**(4):e00740-00718.
278. Shim D-H, Suzuki T, Chang S-Y, Park S-M, Sansonetti PJ, Sasakawa C, Kweon M-N: **New Animal Model of Shigellosis in the Guinea Pig: Its Usefulness for Protective Efficacy Studies**. *The Journal of Immunology* 2007, **178**(4):2476-2482.
279. Rajagopala SV, Casjens S, Uetz P: **The protein interaction map of bacteriophage lambda**. *BMC Microbiol* 2011, **11**(1):213.
280. Baldus M: **Magnetic resonance in the solid state: applications to protein folding, amyloid fibrils and membrane proteins**. *European Biophysics Journal* 2007, **36**(1):37-48.