Characterisation of Lipopolysaccharide O-antigen Acetyltransferases in the Bacterium *Salmonella*

Reyme Herman

MSc by Research

University of York

Biology

March 2015

Abstract

Bacteria are known to possess various modifications on the cell envelope for a variety of reasons, including pathogenicity and survivability. Components on the Gram-negative bacteria cell envelope such as the lipopolysaccharide (LPS) and peptidoglycan, among others, are commonly modified in a multitude of ways. In *Salmonella*, the LPS itself can possess different modifications on the three components of the molecule; the O-antigen, core polysaccharide and lipid A. This study focuses on two LPS O-antigen acetyltransferases, namely OafA and GtrC family II proteins, which target different sugar residues within the O-antigen.

Findings from this project suggest that OafA and GtrC family II proteins are highly similar and possess two domains in the form of an N-terminal inner membrane bound acyltransferase 3 domain and a C-terminal SGNH hydrolase domain. SGNH hydrolase domains possess a characteristic α/β/α fold. In addition, a catalytic triad of serine, aspartate and histidine is also present with the latter being separated by a maximum of two amino acids (DXXH). Using the GtrC family II protein from the invasive non-typhoidal *Salmonella* (iNTS) strain *S.* Typhimurium D23580, $GtrC^{BTP1}$, as a model for GtrC family II proteins, the amino acids within the catalytic triad were individually replaced by alanine and the functions were analyzed by phage susceptibility. Each of the single replacement mutations resulted in the abrogation of the protein function, suggesting that each of the catalytic residues tested are important for protein function.

To further describe these acetyltransferases, biochemical and biophysical experiments would be performed and these methods require pure samples of proteins. This study developed a method for the successful expression and purification of these membrane bound acetyltransferases with an overall yield of approximately 1.2 mg of the purified proteins per litre of bacterial culture. Expression of these proteins were performed in an *E. coli* based expression culture using an L-arabinose inducible vector which also introduced a deca-His tag on the C-terminal end of the expressed proteins. The methods described in this thesis for the expression and purification of OafA and $GtrC^{BTP1}$ can be brought forward to aid in the preparation of these proteins for future biochemical and biophysical experiments.

List of Contents

Chapter 2: Materials and Methods

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to thank my supervisor Marjan van der Woude for all the help, patience and guidance she has provided me with throughout my degree. I would like to show my appreciation to Erica Kintz and other members of the van der Woude lab, past and present, for their constant support and making my stay in the lab an enjoyable one.

Also, I would like to thank Gavin Thomas and Jen Potts, along with their respective lab members, from the University of York for arranging and supporting me during my stay in their labs for the protein expression and purification experiments. I would like to give a special thanks to Andrew Bretnall and Judith Hawkhead for all the help and support they have provided me for the protein expression and purification experiments.

I would also like to thank my thesis advisory panel members, Gavin Thomas and Nathalie Signoret, for their invaluable opinions and help throughout the progress of my project.

Finally, I would like to thank my family and friends for the continuous support they have given me throughout my degree. Their care and motivation made me who I am today.

Authors Declaration

The work presented in this thesis is my own and I am the sole author. This work has not been published for an award at any University. All sources are acknowledged as References.

Chapter 1

Introduction

1.1 Salmonella

Gastrointestinal infections are continuously problematic across developing and also, developed countries (Ternhag et al., 2008). These infections are known to be caused by members of the enterobacteriaceae family of Gram negative bacteria which, as seen in the name of the family, are enteric in nature (Kauffmann, 1966). They are able to invade target host cells and, in turn, able to cause infections to the host, depending on their characteristics. The pathogenic groups of bacteria within this family include the genera *Klebsiella*, *Escherichia* and *Salmonella*.

The genus *Salmonella* is closely related to the genus *Escherichia* and causes salmonellosis (Coburn et al., 2007). This genus is further divided to two species, namely *Salmonella bongori* (*S. bongori*) and *Salmonella enterica* (*S. enterica*). The latter is the better studied of the two species due to the association of this species to human and warm blooded animal infections, making it clinically significant. While many of the strains of *Salmonella* are incapable of being pathogenic, there are select strains which are major causes of food borne illnesses such as typhoid fever and gastroenteritis. On the other hand, *S. bongori* is associated with the diseases that occur within cold-blooded animals such as lizards (Fookes et al., 2011). The classification of the bacteria family *Salmonella* is depicted on Figure 1.1.1.

1.1.1 Salmonella enterica

Both species of *Salmonella* mentioned in Section 1.1 are able to invade host cells due to the presence of the Salmonella Pathogenicity Island 1 (SPI1) (Galan and Curtiss, 1989, Mills et al., 1995). However, the presence of SPI2 encoding the type 3 secretion system 2 (T3SS-2) in *S. enterica* allows the bacteria within this family to cause enterocolitis and infect systemically. This species of *Salmonella* uses proteins secreted using the T3SS-1 and T3SS-2 proteins to cause the invasion into the host cells and proliferation within these cells (Bispham et al., 2001, Coombes et al., 2005, Coburn et al., 2005). *S. enterica* is further divided into 6 subspecies in *S. enterica* subsp. *arizonae*, *diarizonae, houtenae, indica, salamae* and *enterica.*

1.1.2 Salmonella enterica subspecies *enterica*

The *S. enterica* subsp. *enterica* is the only one of the six subspecies of *S. enterica* which is responsible for the occurrence of salmonellosis in warm blooded animals. This subspecies can also be referred to as Subspecies I and is further divided into serovars based on their antigenic properties. Over 2600 serovars of *S. enterica* subsp. *enterica* have been identified (Gal-Mor et al., 2014).

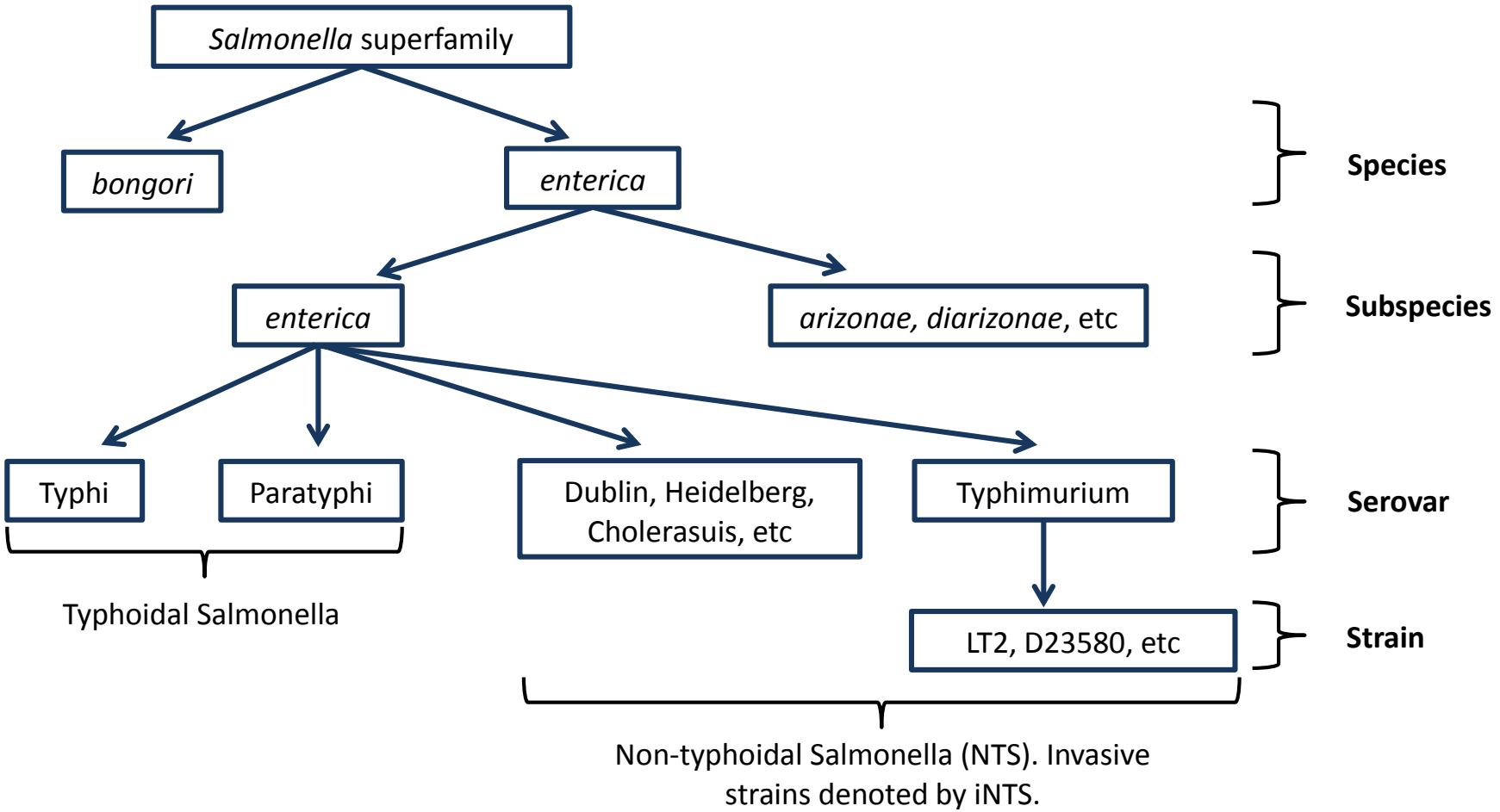


Figure 1.1.1. Classification of *Salmonella* bacterium.

1.1.3 Salmonella enterica subspecies *enterica* serovars

Bacteria within the enterica subspecies of *Salmonella enterica* are further classified into serovars, depending on the antigenic structure of the bacteria. The factors that affect the serotyping are molecules found on the surface of the bacteria and hence, the antigenic properties. The Kauffman-White scheme (Microbiology, 1934) utilises the combination of the O-antigen of the lipopolysaccharide (LPS) and the H-antigen of the flagella to classify bacteria within *S. enterica* subsp. *enterica* into serovars (Cdc.gov, 2015).

Some of the clinically relevant serovars include *S. enterica* subsp. *enterica* serovar Typhi, Paratyphi and Typhimurium. The names of these serovars are shortened to *S.* Typhi, *S.* Paratyphi and *S.* Typhimurium respectively. The different disease outcomes would be determined by the host and the infecting strain of *S.* enterica (subspecies I) (Fierer and Guiney, 2001).

The host adapted *S.* Typhi and Paratyphi are typhoidal salmonella which cause typhoid fever in humans (Huang and DuPont, 2005). Typhoid fever is characterised by bacteraemia which occurs when the bacteria which invade the gastrointestinal tract then escape into the body and blood stream. Other *S. enterica* serovars are classified as non-typhoidal *Salmonella* (NTS) and includes both pathogenic and non-pathogenic strains. *S.* Typhimurium, a form of pathogenic NTS, is a common cause of gastroenteritis in humans. Also, as seen in the name, *S.* Typhimurium is able to cause typhoid fever in mice. Despite being non-typhoidal, a few strains of NTS are also invasive and cause bacteraemia. These strains of NTS are collectively referred to as invasive non-typhoidal salmonella (iNTS) (Gordon, 2011).

1.2 Cell Components of Gram Negative Bacteria including *Salmonella*

Bacteria have an overall structure of one or two membranes surrounding the cytoplasm which contains the genome and various cytoplasmic proteins and enzymes (Figure 1.2.1). Bacteria are classified as either Gram-positive or Gram-negative based on the thickness of the peptidoglycan and the absence or presence of an outer membrane (Hucker and Conn, 1923). This classification is based on the Gram staining of the bacteria using crystal violet staining and counter staining. Gram-positive bacteria such as *Staphylococcus* and *Bacillus* have thick peptidoglycan layers while gram-negative bacteria such as *Escherichia* and *Salmonella* have a thin peptidoglycan layer surrounded by an outer membrane.

1.2.1 Gram-negative bacteria cell envelope

The cytoplasm of the Gram-negative bacteria is surrounded by a cell envelope which consists of the inner membrane and the cell wall, which is made up of the peptidoglycan layer and the outer membrane (Wilson et al., 2011). The inner membrane and the outer membrane are made up of

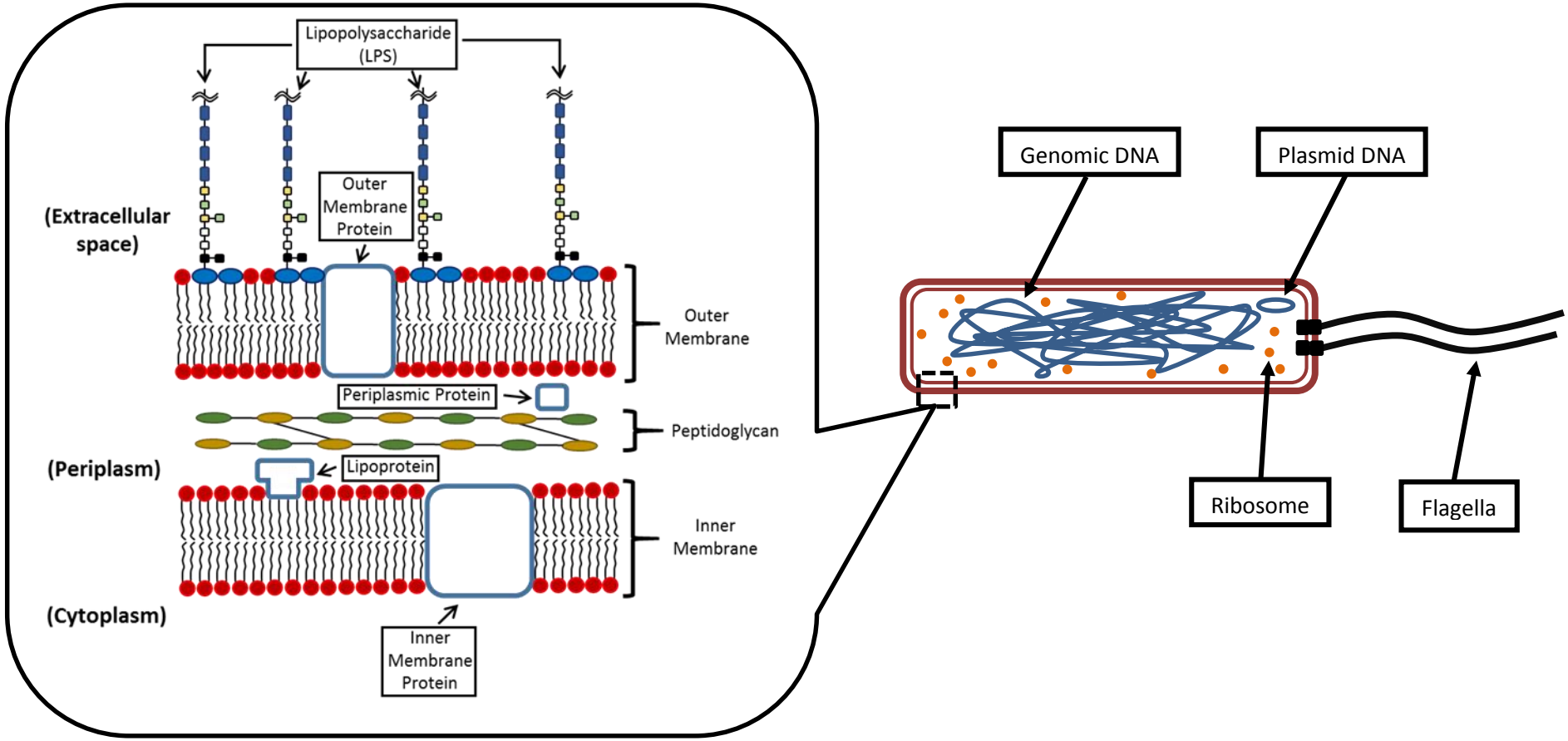


Figure 1.2.1. Structure of the Gram-Negative Bacteria. The composition of the inner membrane and the cell wall is depicted inset.

lipid bilayers and can possess various bound proteins and molecules which perform a range of functions. Along with the peptidoglycan, these proteins and other molecules were discussed in Section 1.2.1.1 to 1.2.1.3.

1.2.1.1 Membrane proteins, lipoproteins and periplasmic proteins

A vast number of proteins exist within the bacteria's cell wall and inner membrane. These proteins can be transmembrane proteins, lipoproteins and periplasmic proteins (Wilson et al., 2011). Transmembrane proteins span the entire width of the membrane and could also include a hydrophilic periplasmic or cytoplasmic domain. Lipoproteins contain a lipid domain which anchors the protein, while allowing the hydrophilic domain to perform the protein function. Lastly, periplasmic proteins are soluble proteins found in the periplasm.

These proteins are functionally important to ensure bacterial survivability and in some cases, pathogenicity. Examples of these proteins are the inner membrane bound cell wall modifying enzymes and the outer membrane bound β-barrel shaped porins which allows for the transport of molecules across the lipid bilayer.

1.2.1.2 Peptidoglycan

The peptidoglycan of bacteria is a structurally important feature of the cell wall in order to maintain cell structure integrity (Höltje, 1998). As mentioned in Section 1.2, Gram-positive bacteria have a thick peptidoglycan layer while Gram-negative bacteria have a thinner peptidoglycan layer, surrounded by a lipid bilayer outer membrane (Figure 1.2.1). The peptidoglycan is formed from layers of polysaccharides made up of alternate *N*-acetylmuramic acid (MurNAc) and *N*-acetylglucosamine (GlcNAc) molecules (Figure 1.2.2). Each MurNAc molecule possesses a short 4 amino acid tail which, with the help of transpeptidase, are cross linked with the amino acid tail of another MurNAc.

The peptidoglycan is a common target for bactericidal β-lactam antibiotics such as penicillin and cephalosporins which act by inhibiting the formation of the peptidoglycan and hence, affecting the cell wall integrity (Morin and Gorman, 2014).

1.2.1.3 Lipopolysaccharide – S*. Typhimurium*

The lipopolysaccharide (LPS) is a polysaccharide molecule found in abundance on the outer membrane of Gram-negative bacteria. These molecules have multiple functions for bacterial survival and pathogenicity. They are made up of lipid A, the core polysaccharide and the O-antigen (Erridge et al., 2002). Using the *S.* Typhimurium LPS as an example, the structure of the LPS is depicted on Figure 1.2.3.

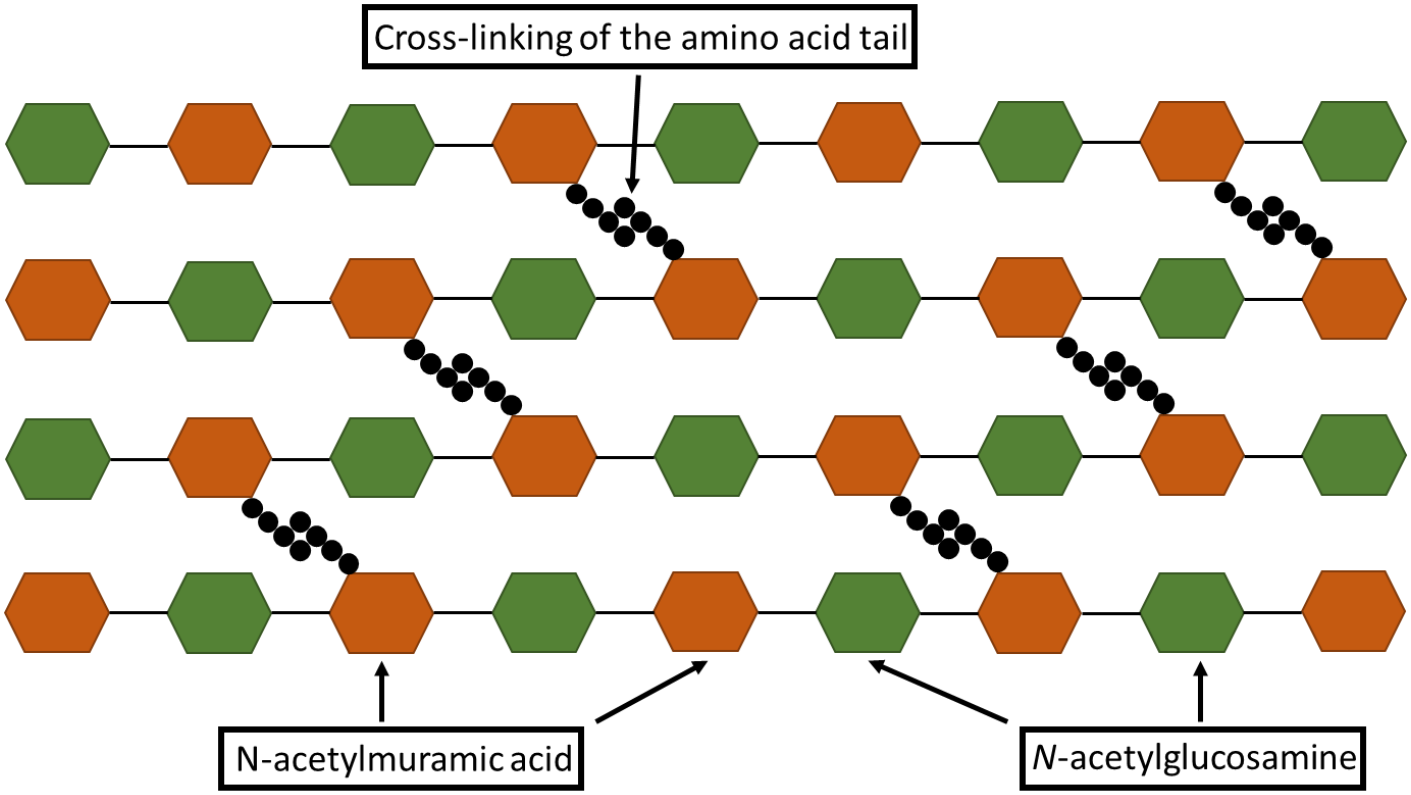


Figure 1.2.2. Structure of the Peptidoglycan. N-acetylglucosamine (NAG) molecules were represented by green hexagons. N-acetylmuramic acid (NAM) molecules were represented by orange hexagons and the 4 amino acid long tails were represented in a chain of 4 black circles. Cross linking of these amino acid chains between the layers were identified. Structural information were attained from Boneca (2005).

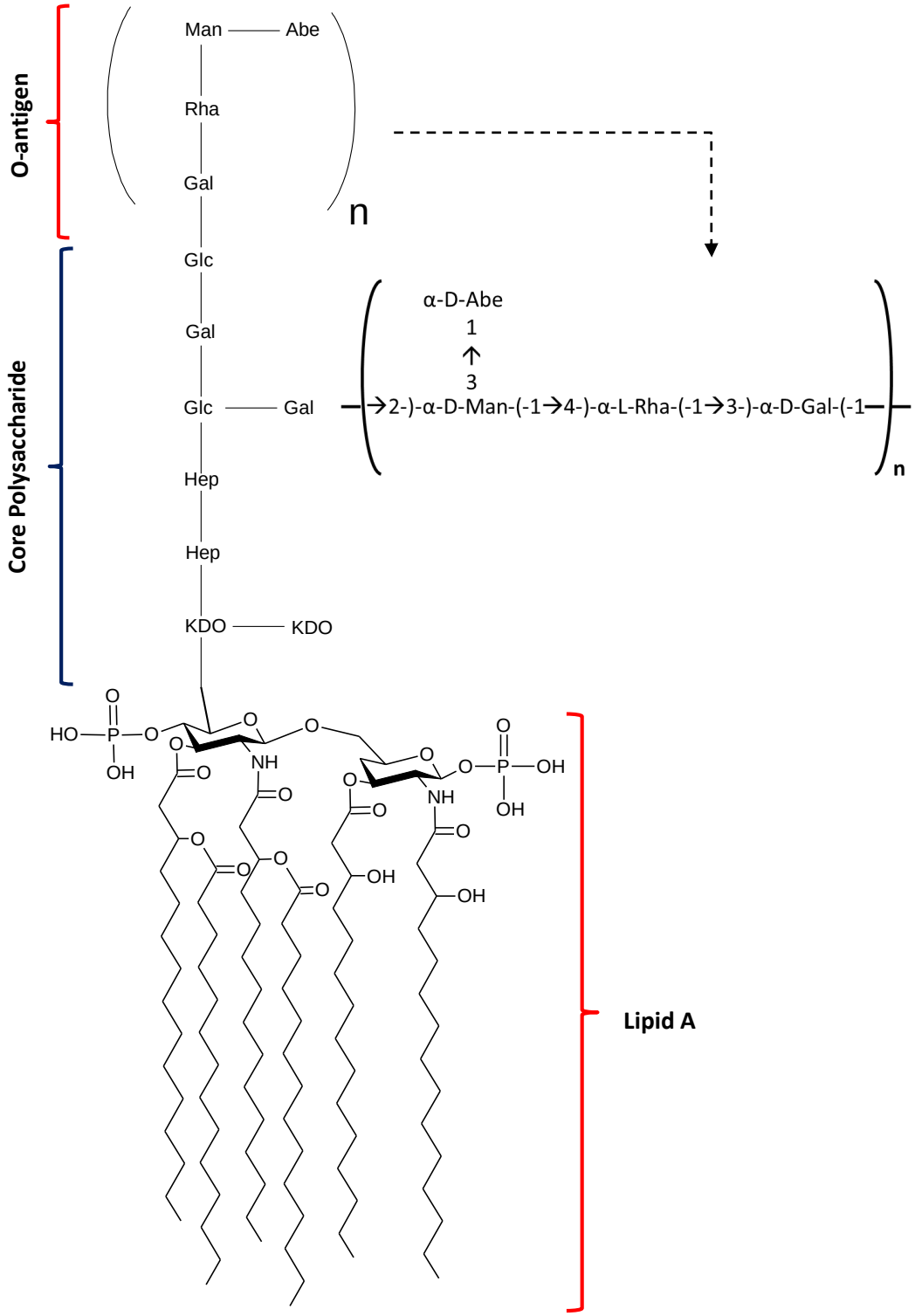


Figure 1.2.3. Structure of the Unmodified *S.* Typhimurium (1,4,12:i:1,2) Lipopolysaccharide (LPS). The different components of the LPS (i.e. Lipid A, Core Polysaccharide and O-antigen) were identified by their respective names in bold with a coloured brace. The linkages of the different sugar residues within the O-antigen were also presented as identified by the black dotted arrow. 3-deoxy-D-mannooctulosonic acid (KDO), galactose (Gal), heptose (Hep), glucose (Glc) rhamnose (Rha), mannose (Man), abequose (Abe). Structural information were attained from Miller et al. (2005). Diagram designed using ChemSketch v14.01.

Lipid A anchors the LPS molecule in the lipid bilayer of the outer membrane and provides the toxicity of the LPS. Using the lipid A tail, the LPS is able to cause the triggering of the immune system by activating the human pathogen-associated molecular patterns (PAMPs) receptor, Toll-like receptor 4 (TLR4), when in abundance in the blood stream (Miller et al., 2005). This endotoxemia can cause septic shock (Raetz and Whitfield, 2002).

The O-antigen subunits differ between different serovars of Salmonella and other bacteria. In the case of *S.* Typhimurium, the basal O-antigen subunit shown on Figure 1.2.3 is made up of an α-D-mannose joined via a (1-4) linkage to an α-L-rhamnose which is also joined with an α-D-galactose via a (1-3) linkage (Hellerqvist et al., 1968). The α-D-mannose is also bound to an α-D-abequose the with a (3-1) linkage. This forms one O-antigen subunit which is linked to another O-antigen subunit via a (1-2) linkage from the α-D-galactose from one O-antigen subunit to the α-D-mannose of another O-antigen subunit. The unmodified form of the LPS O-antigen forms the O:4 serotype in the presence of the abequose residue (Grimont and Weill, 2007). On the other hand, the O:12 serotype is the result of the absence of modifications and the abequose residue on the LPS O-antigen.

The makeup of the components of the LPS can vary between different bacteria. These include the use of a combination of different saccharide molecules to create the LPS backbone and the presence of different modifications on the LPS structure. LPS structure modifications will be discussed in Section 1.3. Some pathogenic bacteria, namely *Neisseria* and *Haemophilus*, possess the lipooligosaccharide (LOS) rather than the LPS. This molecule is defined as a low molecular weight LPS, possessing only the lipid A and core oligosaccharides, which mimics host proteins. (Moran et al., 1996).

1.3 Modifications Within Components in the Cell Wall

To ensure survivability, the various components of the cell wall can be modified. The peptidoglycan, flagella and the LPS can be modified in multiple ways. For this thesis, the modifications of the LPS will be focused on and discussed further. Using the *S.* Typhimurium LPS as a model, the LPS structure with the respective modifications was depicted on Figure 1.3.1 which can be compared to the unmodified LPS structure shown on Figure 1.2.3.

1.3.1 LPS modifications

1.3.1.1 Lipid A and core polysaccharide modifications

Multiple modifications of the Lipid A and the core polysaccharide in *S.* Typhimurium are regulated by the two component regulatory systems in PhoP-PhoQ and PmrA-PmrB (Miller et al., 2005). PagL and PagP are two modification enzymes regulated by PhoP-PhoQ which deacylates lipid A and transfers a palmitate to lipid A respectively

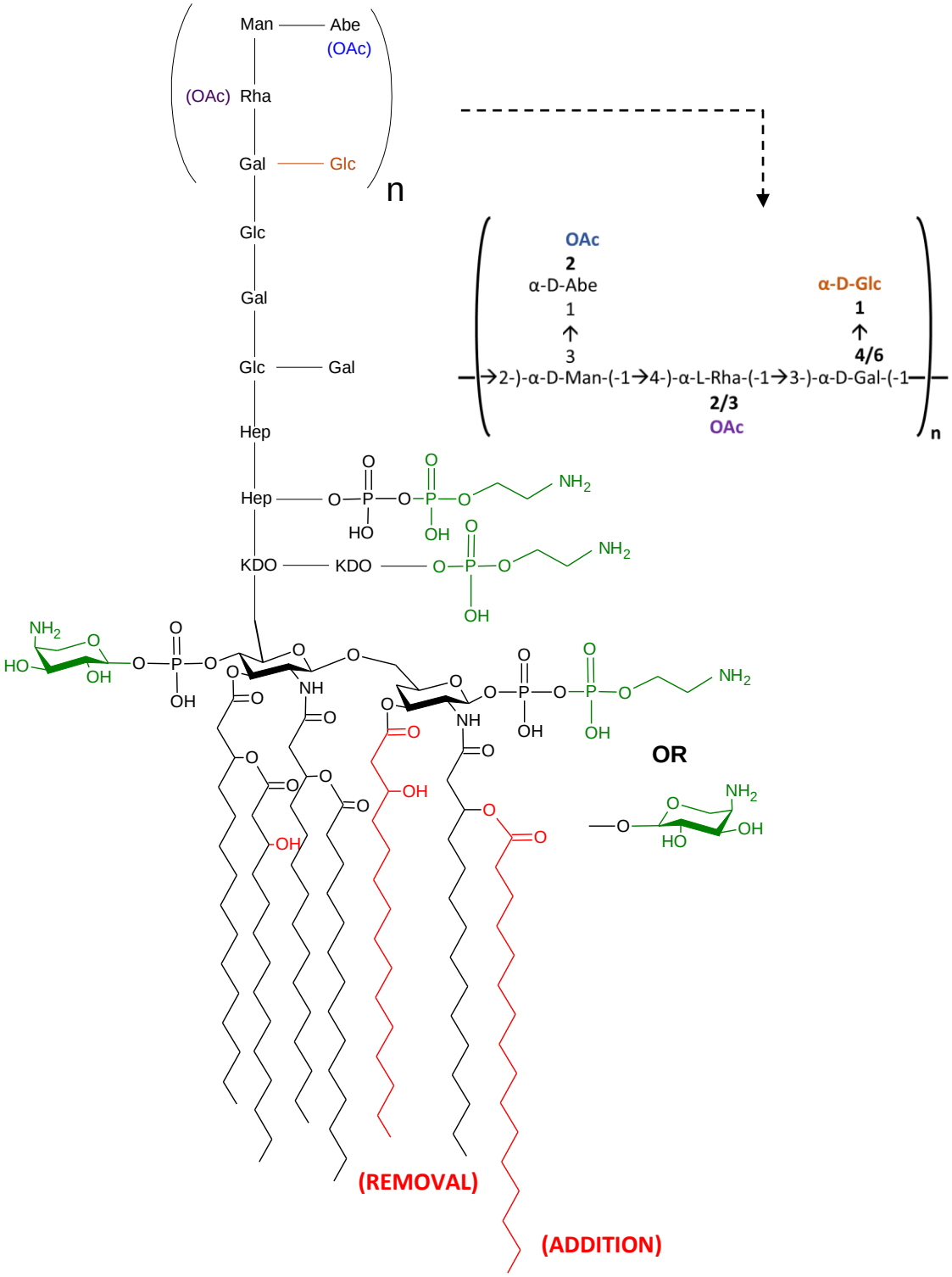


Figure 1.3.1. Structure of the *S.* Typhimurium LPS with modifications. PhoP-PhoQ mediated modifications were coloured in red. PmrA-PmrB mediated modifications were coloured in green. O-antigen modifications were colour coded in either blue, purple or orange and in bold. A diagram showing the O-antigen modifications was identified with a dotted arrow. 3-deoxy-D-mannooctulosonic acid (KDO), galactose (Gal), heptose (Hep), glucose (Glc), rhamnose (Rha), mannose (Man), abequose (Abe). Structural information were attained from Hellerqvist et al. (1968), Miller et al. (2005) and Kintz et al. (2015). Diagram designed using ChemSketch v14.01.

(Guo et al., 1998, Trent et al., 2001). LpxO is also a PhoP-PhoQ regulated enzyme which transfers a hydroxyl group to a lipid A tail (Gibbons et al., 2000). These modifications alter lipid A recognition by human TLR4. PmrA-PmrB regulates a set of enzymes that confer increased resistance to cationic antimicrobial peptides such as polymyxin. These enzymes can modify the LPS by the addition of 4-aminoarabinose (Ara4N) or phosphoethanolamine to different sites on lipid A and core polysaccharide.

1.3.1.2 O-antigen modifications

As discussed in Section 1.1.3, the O-antigen of the LPS and the H-antigen of the flagella are used to determine the serotype of the Salmonella in the Kauffman-White classification. Multiple modifications on the *S.* Typhimurium LPS O-antigen subunits are known and produce different serotypes. Glucosylation of the galactose residue is a form of modification which could be present and are thought to be caused by the activity of the *gtr* operon (Davies et al., 2013). Acetylation of the O-antigen subunit of the *S.* Typhimurium LPS is of particular interest in this project with this modification known to exist on abequose (Abe) (Hellerqvist et al., 1968) and rhamnose (Rha) (Wollin et al., 1987; Micoli et al., 2014). These O-antigen modifications will be further discussed in Section 1.4.

1.4 O-antigen Modifications in *S.* Typhimurium

1.4.1 Glucosylation - Gtr operon

Glycosylation of the O-antigen subunit is caused by proteins from the *gtr operon*. This operon typically encodes for three Gtr proteins, namely GtrA, GtrB and GtrC (Davies et al., 2013). These three proteins are inner membrane bound and act as a mechanism for the addition of the sugar to the O-antigen. Using studies from the *Shigella flexneri* gtr operon, the GtrA and GtrB proteins are responsible for the transfer of the glucose from the cytoplasm, across the inner membrane and into the periplasm where the GtrC then adds this sugar onto the O-antigen (Allison and Verma, 2000).

Davies et al. (2013) performed phylogenetic analyses of the *gtr operons* found in *Salmonella* (Figure 1.4.1). 10 GtrC, 7 GtrB and 5 GtrA clusters were identified in the different strains of interest and each strain can possess more than one form of the *gtr* operon. Comparing the amino acid identities between each of the clusters of the Gtr proteins separately, it was determined that the GtrC clusters are the most significantly different and the identities between the GtrA at 84% and GtrB at 90% are significantly higher. The significant differences between the GtrC clusters suggest that the differences between the serotypes in *S. enterica* (subspecies I) might be a result of the various GtrC protein activities. These clusters of GtrC will, henceforth, be individually referred to as a GtrC family and each family assigned a roman numeral. Within the GtrC families for *S. enterica* (subspecies I), strains possessing family I to IV are not confined

within one serogroup which is a group of serotypes with common antigens. This could mean that the different families mentioned are responsible for different O-antigen modifications.

1.4.1.1 Family I and III and Family IV

Consistent with the glycosyltransferase activity, both GtrC Family I and III were determined to facilitate the addition of a glucose to the galactose moiety of the O-antigen backbone with a (1-6) linkage to give the O:1 serotype (Young et al., 1964, Vander Byl and Kropinski, 2000, Pedulla et al., 2003) and (1-4) linkage to give the O:12_2 serotype respectively (Bogomolnaya et al., 2008). This modification is shown on the modified *S.* Typhimurium O-antigen subunit shown on Figure 1.3.1.

P22 is a phage which also possesses a *gtr operon* with GtrC family I. Prior to infection, the P22 phage is able to cleave the LPS O-antigen of the target cell using the endorhamnosidase activity of the tailspike protein in order to improve the phage's access to the surface of the bacteria. The formation of the O-antigen of the O:1 serotype, caused by GtrC family I, inhibits the endorhamnosidase activity (Figure 1.4.1). On the other hand, the O:12_2 serotype, caused by GtrC family III, is not able to block the endorhamnosidase activity (Wollin et al., 1981, Svenson et al., 1979). This suggests that *S. enterica* strains possessing the O:1 serotype could have acquired *gtr operon* from the P22 phage in order to avoid multiple infections within the same host cell, defined as superinfections.

1.4.1.2 Family IV

To date, the function of GtrC family IV has yet been characterised. An LPS band profiling was performed by Davies et al. (2013) using TSDS-PAGE separation of the samples of LPS from a wild type strain of *S.* infantis, another with an unmodified O-antigen and another complemented with a plasmid expressing GtrC family IV from a constitutive promoter to determine if any glucosylation occurs. Glucosylation would increase the mass of the LPS bands and would appear at a higher molecular weight on the gel. However, this outcome was not observed for the complemented strain of *S.* infantis. Hence, the function of the GtrC family IV would likely to not be glucosylation.

1.4.1.3 Family II – $GtrC^{BTP1}$ and the BTP1 phage

Davies et al. (2013) identified that family II *gtr operons* possess a truncated or absent *gtrB* gene with a truncated *gtrA* gene. This suggests that GtrA and GtrB might not be necessary for the function of GtrC unlike that of family I and III where GtrA and GtrB are required for the transport of the glucose for the addition onto the galactose moiety. An LPS band profiling detecting the band shifts as mentioned in Section 1.4.1.2 was also used to assess the outcome of the expression of GtrC family II and like GtrC family IV, no shifts were observed and hence, no

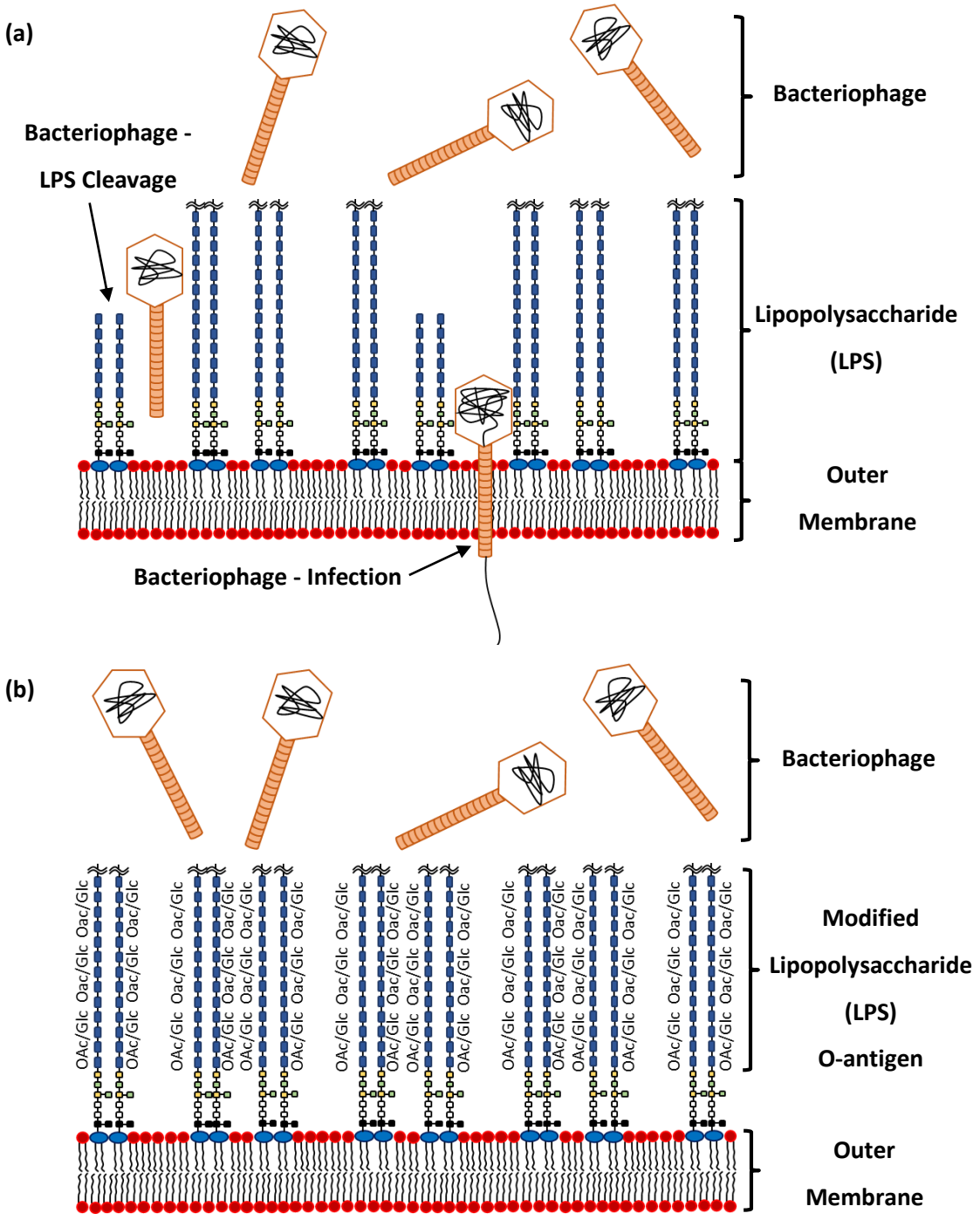


Figure 1.4.2. Infection of Bacteria by Bacteriophages. Bacteria with an unmodified O-antigen **(a)** are susceptible to specific bacteriophage infections. Specific LPS O-antigen modifications in acetylation (OAc) and glucosylation (Glc) **(b)** can block the ability for phages to perform LPS cleavage to allow for better access to the cell surface.

addition of glucose was thought to be present. This suggests that GtrC family II would possess another function.

S. Typhimurium D23580 is a newly identified invasive non-typhoidal *Salmonella* (iNTS) (Kingsley et al., 2009) which possesses a glucosylation modification on the galactose residue and the presence of acetyl groups on the abequose and rhamnose residues of the LPS O-antigen (Figure 1.3.1). Through genomic analysis, the family II *gtr operon* on the BTP1 prophage was identified (Davies et al., 2013, Kintz et al., 2015) . Like the P22 phage and strains of *S. enterica* expressing the GtrC family I protein, it was hypothesised that GtrC family II functions to block the BTP1 tailspike protein mediated O-antigen cleavage by modifying the O-antigen serotype (Figure 1.4.1). From here henceforth, GtrC family II protein of *S.* Typhimurium D23580 will be referred to as $GtrC^{BTP1}$. Kintz et al. (2015) performed a BLAST analysis of $GtrC^{BTP1}$ and identified one *Shigella flexneri* O-antigen acetyltransferase called Oac and a *S.* Typhimurium LPS O-acetyltransferase called OafA. The latter is responsible for the acetylation of the abequose residue of the LPS O-antigen of *S.* Typhimurium to give the O:5 antigen (Slauch et al., 1995, Slauch et al., 1996). This protein will be discussed further in Section 1.4.2.

Through bioinformatics, protein analysis and comparisons with the *S. flexneri* Oac and *S.* Typhimurium OafA, Kintz et al. (2015) suggested that $GtrC^{BTP1}$ is an acetyltransferase. $GtrC^{BTP1}$ was predicted to possess 11 transmembrane domains and a C-terminal periplasmic tail which is 273 amino acids long with similar characteristics with O-antigen acetyltransferases discussed above. Strains of *S.* Typhimurium deficient of $GtrC^{BTP1}$ was susceptible to BTP1 phage associated lysis as determined using a phage assay. $GtrC^{BTP1}$ was able to function on its own in the absence of GtrA and GtrB. Using the biochemical analysis of the LPS O-antigen of *S.* Typhimurium D23580 by Micoli et al. (2014), data from Kintz et al. (2015) suggests that $GtrC^{BTP1}$ is responsible for the acetylation of the rhamnose residue.

1.4.2 OafA

The acetylation on the abequose residue of the *S.* Typhimurium LPS O-antigen is thought to be caused by a putative acetyltransferase protein called OafA producing the O:5 serotype (Slauch et al., 1995; Slauch et al., 1996). The presence of acetylation on the abequose residue was immunologically significant to the bacteria and host but is not important to the virulence of the bacteria (Michetti et al., 1992, Michetti et al., 1994). This protein was predicted to possess 9 transmembrane helices with a C-terminal tail. Despite being extensively biologically characterised, the molecular basis of OafA has yet to be characterised.

1.5 Aim of the Project

Comparing OafA to GtrC^{BTP1}, very similar characteristics were observed in their structure and function. Both proteins were thought to be membrane bound with a C-terminal periplasmic tail with the predicted ability to add an acetyl group to a sugar residue of the LPS O-antigen of various *S. enterica* strains. The main difference between these proteins is the sugar substrate specificity with abequose for OafA and rhamnose for GtrC^{BTP1}.

The aim of this project is to characterise OafA and GtrC^{BTP1} using various bioinformatical, microbiological and biochemistry methods. It was hypothesised that OafA and GtrC^{BTP1} are closely related both structurally and functionally. However, it was predicted that differences would lie in their respective substrate binding sites. To understand OafA and GtrC^{BTP1}, functional and in-depth bioinformatics analysis of these proteins were performed followed by appropriate *in vivo* and *in vitro* assays to support the bioinformatical findings.

Chapter 2

Materials and Methods

2.1 Bacterial Strains, Plasmids and Oligonucleotides

Each item mentioned on Tables 2.1.1 to 2.1.4 was identified by a specific code and the description was included.

Table 2.1.1 and 2.1.2 respectively contains the list of *E. coli* and *Salmonella* used in this project sorted according to genus and whether they are harbouring a plasmid.

Table 2.1.3 contains the list of plasmid vectors used.

Table 2.1.4 contains the list of oligonucleotides used. They are sorted according to which experiments they are used in or if they are sequencing primers.

2.2 Products

2.2.1 Chemicals and kits

The kits used for DNA preparation (i.e. the QIAprep Spin Miniprep kit, the QIAquick PCR Purification Kit, QIAEX II Gel Extraction Kit and the QIAquick Gel Extraction Kit) were purchased from Qiagen.

Other chemicals and reagents were acquired from Fisher Scientific, Sigma Aldrich, Life Technologies, Melford, VWR international, Merck-Millipore and Bio-Rad, among others.

2.2.2 Molecular biology reagents

Two commercially available chemically competent strains of *E. coli* were used in this project. These were the *E. cloni*® 10G Chemically Competent Cells (Lucigen) and the XL1-Blue Competent Cells (Agilent).

Enzymes and other biological agents were attained from New England Biolabs (NEB), Qiagen, Sigma-Aldrich, Promega, Roche and Merck-Millipore.

2.3 In silico

Bioinformatics analyses were mainly performed using the Geneious Basic 5.6.7 bioinformatics suite.

2.3.1 Data mining

Nucleotide and amino acid sequences were attained from NCBI using the NCBI search function on Geneious Basic 5.6. Protein domain searches were performed using InterPro (Available at: www.ebi.ac.uk/interpro/) using the amino acid sequences attained from NCBI. Protein crystal structures were attained from the RCSB Protein Data Bank (Available at: www.rcsb.org).

Table 2.1.1. List of E. coli strains used in this study.

Name	*Strain Code*	*Genotype*	*Plasmid Harboured*	*Source*
Wild-type *E. coli* strains used				
***E. coli* DH5α**	DL433	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–	-	Life Technologies
***E. coli* MC1061**	MV1651	F^- Δ(ara-leu)7697 $[araD139]_{B/r}$ Δ(codB-lacI)3 galK16 galE15 λ^- $e14^-$ mcrA0 relA1 rpsL150(strR) spoT1 mcrB1 hsdR2(r^-m^+)	-	(Casadaban and Cohen, 1980)
***E. coli* XL-1 Blue (commercially prepared, chemically competent cells)**	-	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–	-	Agilent
E. coli strains harbouring plasmids				
***E. coli* DH5α pOafA**	MV1634	See above	pMV433	This study
***E. coli* DH5α pOafA (C-tail)**	MV1636	See above	pMV444	This study
***E. coli* XL-1 Blue pGtrC^{BTP1}**	MV1635	See above	pMV434	This study
***E. coli* DH5α pGtrC^{BTP1} (C-tail)**	MV1637	See above	pMV445	This study
***E. coli* MC1061 pOafA**	MV1644	See above	pMV433	This Study
***E. coli* MC1061 pGtrC^{BTP1}**	MV1645	See above	pMV434	This Study

Table 2.1.2. List of Salmonella strains used in this study.

Name	*Strain Code*	*Genotype*	*Plasmid Harboured*	*Source*
Wild-type *Salmonella* strains used				
STM LT2	sMV346	*Salmonella* Typhimurium LT2 (ATCC strain number 19585, Lot number 215096)	-	ATCC
STM D23580	sMV189	WT (Kingsley et al., 2009)	-	Sanger Institute
Salmonella strains harbouring plasmids				
STM LT2 pGtrC^{BTP1}	sMV888	WT	pMV434	This study
STM LT2 pGtrC^{BTP1} (S430A)	sMV890	WT	pMV437	This study
STM LT2 pGtrC^{BTP1} (D618A)	SMV889	WT	pMV438	This study
STM LT2 pGtrC^{BTP1} (H621A)	sMV892	WT	pMV439	This study

Table 2.1.3. List of plasmids use in this study.

Name	*Description*	*Antibiotic Resistance Cassette*	*Plasmid Code*	*Primers used to Generate Inserts*	*Source*
Expression vectors used					
pBADcLIC2005	Expression vector, encodes for a C-terminal deca-His tag	Amp^r	pMV432	-	(Geertsma and Poolman, 2007)
Expression vectors with various inserts					
pOafA	pMV432, with an *oafA* gene insert	Amp^r	pMV433	oMV1074 + oMV1075	This study
pOafA (C-tail)	pMV432, with an *oafA* insert containing the sequence encoding the C-terminal tail only	Amp^r	pMV444	oMV1076 + oMV1075	This study
pGtrC^{BTP1}	pMV432, with an *gtrC^{BTP1}* gene insert	Amp^r	pMV434	oMV1077 + oMV1079	This study
pGtrC^{BTP1} (C-tail)	pMV432, with an *gtrC^{BTP1}* insert containing the sequence encoding the C-terminal tail only	Amp^r	pMV445	oMV1078 + oMV1079	This study
pGtrC^{BTP1} (S430A)	pMV432, with an *gtrC^{BTP1}* insert possessing the nucleotide changes resulting in an S430A mutation	Amp^r	pMV437	oMV1158 + oMV1159	This study
pGtrC^{BTP1} (D618A)	pMV432, with an *gtrC^{BTP1}* insert possessing the nucleotide changes resulting in an D618A mutation	Amp^r	pMV438	oMV1156 + oMV1157	This study
pGtrC^{BTP1} (H621A)	pMV432, with an *gtrC^{BTP1}* insert possessing the nucleotide changes resulting in an H621A mutation	Amp^r	pMV439	oMV1154 + oMV1155	This study

Table 2.1.4. List of oligonucleotides/primers used in this study.

Primer Code	*Sequence (5' to 3')*	*Direction*	*Purpose*
Protein expression - LIC			
oMV1074	ATG GGT GGT GGA TTT GCT ATG ATC TAC AAG AAA TTC AGA C	F	Isolate the *oafA* gene, introduce LIC cassette
oMV1076	ATG GGT GGT GGA TTT GCT ATG ACA AAA GGT GTT AGT TTT AG	F	Isolate the *oafA* gene encoding the C-terminal tail only, introduce LIC cassette
oMV1075	TTG GAA GTA TAA ATT TTC TTT TGA AAT CTG CTT TTT CAC	R	Isolate the *oafA* gene (encoding both the full length protein and C-terminal tail only), introduce LIC cassette
oMV1077	ATG GGT GGT GGA TTT GCT ATG GAA CAC TTA AAA TAC AGA C	F	Isolate the *gtrC*^{BTP1} gene, introduce LIC cassette
oMV1078	ATG GGT GGT GGA TTT GCT ATG GGA ATT AAA GAA AGA AGC GTA AAC	F	Isolate the *gtrC*^{BTP1} gene encoding the C-terminal tail only, introduce LIC cassette
oMV1079	TTG GAA GTA TAA ATT TTC TCT TAT TAT CAA ATG CCC TAT C	R	Isolate the *gtrC*^{BTP1} gene (encoding both the full length protein and C-terminal tail only), introduce LIC cassette
Mutagenesis			
oMV1154	AGA TTG GGG CGC TTT GAC AAA GCC TG	F	Introduce the H621A mutation to the GtrC^{BTP1} protein expressed from the pGtr^{BTP1} vector. Mutagenesis carried out by inverse PCR and blunt end ligation
oMV1155	ACA GCT GTA ACA AAA TCT G	R	"
oMV1156	ACA GCT GTA GCT TGG GGC CAT	F	Introduce the D618A mutation to the GtrC^{BTP1} protein expressed from the pGtr^{BTP1} vector. Mutagenesis carried out by inverse PCR and blunt end ligation
oMV1157	AAC AAA ATC TGG GCC ATC AC	R	"

Primer Code	*Sequence (5' to 3')*	*Direction*	*Purpose*
Mutagenesis (cont'd)			
oMV1158	AAT CGG TGA TGC ATA TGC AGC	F	Introduce the S430A mutation to the GtrC^{BTP1} protein expressed from the pGtr^{BTP1} vector. Mutagenesis carried out by inverse PCR and blunt end ligation
oMV1159	ATA AAA ATA TTA TTG CGG CTA TTT TTA ATG	R	"
Sequencing			
oMV1160	CAA AAG TGT CTA TAA TCA CGG CAG	F	Verification of gene inserts in the pBADcLIC2005 vector
oMV1161	ATC CGC CAA AAC AGC CAA GCT G	R	"
oMV966	GGC CAA GCT ATT CAG CAC TC	F	Verification of the *oafA* gene
oMV967	GCA ATA CGG CCT GTC ATC TT	R	"
oMV616	AAT CGC CAA TAC TAA TGC CAG C	R	Verification of the *gtrC^{BTP1}* gene
oMV617	TGT TGC ACC TTT GGT CTT TAG G	F	"

All primers used in this study were designed and then, purchased in the freeze dried form from Sigma-Aldrich. Stocks of primers were in 100 µM concentration.

2.3.2 Sequence alignments

Nucleotide and amino acid sequence alignments were performed using the Geneious Alignment tool of the Geneious Basic 5.6.7 bioinformatics suite.

For nucleotide alignments, the following parameters were set as follows. Cost matrix at 65% similarity (5.0/-4.0), gap open penalty at 12, gap extension penalty at 3 using the global alignment with free end gaps alignment type.

For the amino acid alignments, the following parameters were set as follows. Cost matrix at the Identity setting, gap open penalty at 5, gap extension penalty at 0, using the global alignment with free end gaps alignment type.

2.3.3 Structural analyses and predictions

In this study, the structure of various proteins were predicted for their characterisation. These predictions were used to determine potential protein domains and active sites by comparing with known structures of various proteins derived from the literature and accessed from RCSB Protein Data Bank (Available at: www.rcsb.org).

2.3.3.1 Transmembrane helix predictions

Transmembrane helix predictions were performed using TMHMM Server v2.0 (Available at: www.cbs.dtu.dk/services/TMHMM/).

2.3.3.2 3D structural predictions

3D structural predictions were performed using SWISS-MODEL (Available at: swissmodel.expasy.org/). Amino acid sequences from the proteins to be predicted that were attained from data mining (Section 2.3.1) were threaded through the SWISS-MODEL server and modelled using a known protein structure as a template.

2.3.3.3 Structural visualisation

Protein structures, predicted or known, were attained in the PDB format (.pdb) and visualised on the Swiss-PdbViewer v4.1.0.

2.4 Common Materials and Methods

2.4.1 Bacterial Culturing – Liquid and Solid Media

Bacteria cultures were prepared using Lennox broth (Fisher Scientific). Liquid media were prepared by dissolving the Lennox broth (LB) powder in ddH_2O at a concentration of 20 g/L. If LB agar plates were to be prepared, granulated agar (Fisher Scientific) were added to the liquid culture mix at 17.5 g/L.

Both mixes were autoclaved prior to use. For solid media, the molten LB agar mix were cooled to 55°C and poured onto petri dishes.

If antibiotics were required for selection, the appropriate antibiotics (e.g. ampicillin (Sigma-Aldrich)) were added to a volume of liquid media prior to bacteria culturing or to the molten LB agar prior to the preparation of the LB agar plates. In this project, only ampicillin was used and was added to the liquid media at a final concentration of 100 µg/mL.

Bacterial culture incubation was performed in specific temperature incubators with liquid cultures being grown using an orbital shaker.

2.4.2 Genomic DNA isolation

Bacterial strains of interest were grown to saturation using liquid media (Section 2.4.1). Next, 1.5 mL of the culture was centrifuges for 2 minutes at 13,000 rpm in the Heraeus™ Pico™ microcentrifuge. The supernatant was then removed. The remaining pellet was resuspended in 400 µL of salt homogenising buffer (final concentrations 0.4 M NaCl, 10 mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0). 40 µL of 20% SDS (Melford) and 4 µL Proteinase K (Roche) at 200 µg/mL were added to the resuspended bacterial pellet. The mix was then incubated at 55°C overnight. The next day, 300 µL of 5M NaCl was added and the resulting mixture was vortexed for 30 seconds. After which, the mixture was centrifuged at 10,000 rpm using the microcentrifuge for 30 minutes. The supernatant was separated from the pellet which contains the bacterial cell wall debris. An equal amount of isopropanol (Sigma-Aldrich) was mixed well with the supernatant before incubation at -20°C for 1 hour. After incubation, the mixture was centrifuged for 20 minutes at 10,000 rpm at 4°C in the Eppendorf Centrifuge 5415 R microcentrifuge. The supernatant was thoroughly pipetted out, leaving a pellet. The pellet containing the gDNA was washed in 70% ethanol (AnalaR NORMAPUR) and then sufficiently dried before resuspension in autoclaved ultrapure water (Milli-Q by Millipore Corporation). The resuspended gDNA was further incubated at 65°C for 1 hour to complete resuspension. When required, storage of the gDNA was performed at 4°C.

2.4.3 Plasmid DNA isolation

Bacterial strains possessing the plasmid of interest were grown to saturation in liquid culture (Section 2.4.1) overnight at an appropriate temperature and with appropriate antibiotics. Plasmid isolation was performed using the QIAprep Spin Miniprep kit (Qiagen) according to the protocol provided in the QIAprep Miniprep Handbook (Available at: www.qiagen.com).

2.4.4 Polymerase Chain Reaction (PCR) and purification

PCR were performed on either isolated gDNA (Section 2.4.2) or plasmid DNA (Section 2.4.3). Primers were designed and then, ordered from Sigma-Aldrich. Primers were delivered in the freeze dried form and solubilised with ultrapure H_2O (MilliQ by Millipore).

High fidelity PCR was performed using the Q5® Hot Start High-Fidelity DNA Polymerase (NEB) or the KOD Hot Start DNA Polymerase (Merck-Millipore). Otherwise, analytical PCR was performed using the GoTaq® or GoTaq® G2 Flexi DNA polymerase (Promega). For details, see product specific protocols which are available online.

PCR purification was performed using the QIAquick PCR Purification Kit (Qiagen). For full protocol, refer to the QIAquick Spin Handbook (Available at: www.qiagen.com).

2.4.5 DNA concentration – vacuum concentration

DNA concentration was performed using the miVac DNA Concentrator (Genevac). Prior to DNA concentration, the machine was pre-warmed to 50°C. Concentration was performed at 50°C for an appropriate amount of time.

2.4.6 Restriction enzyme digest

In this project, all restriction enzyme mediated DNA digestion were performed using a single enzyme in each experiment. All restriction enzymes were procured from NEB and used according to the suppliers recommended protocol (Available at: www.neb.com).

2.4.7 DNA analyses

2.4.7.1 Agarose gel electrophoresis

For DNA analyses, 0.8% or 1% agarose gels were prepared by dissolving 0.8% or 1% (w/v) Agarose (Sigma-Aldrich) into TAE buffer (Life Technologies). Ethidium Bromide was added to the mixture at a final concentration of 0.2 µg/mL . Once the gel has been solidified in a cast, the gel was placed in a gel electrophoresis tank filled with TAE buffer. DNA samples and the DNA reference ladder marker, GeneRuler DNA Ladder Mix (Thermo Scientific), were then loaded and gel electrophoresis was performed at 100 V. The resulting gel was visualised in a G:BOX gel doc (Syngene).

2.4.7.2 DNA concentration measurement

Measurement of DNA concentration was performed using the NanoDrop 1000 (Thermo Scientific). The absorbance at 260 nm were measured to determine the concentration of the DNA in the test sample. The 260/230 and 260/280 ratios were also analysed to determine the purity of the test sample.

2.4.7.3 Sequencing

Sequencing was carried out by Eurofins Genomics using the SmartSeq kit using a primer specific to the region of DNA that is of interest. Sequencing primers were mentioned in Table 2.1.4 under the section of the table referred to as “Sequencing”.

2.4.8 Gel Excision

To isolate DNA of interest from unwanted DNA, gel excision was performed using 0.8% agarose gels prepared as mentioned in Section 2.4.7.1. However, ethidium bromide was replaced by the Nancy-520 dye (Sigma Aldrich) at a final concentration of 0.4 µg/mL. The DNA samples and the DNA reference ladder marker, GeneRuler DNA Ladder Mix (Thermo Scientific), were loaded and gel electrophoresis was performed at 25 V. The resulting gel was visualised on the Dark Reader Transilluminator (Clare Chemical) and the appropriate gel pieces which contains the DNA of interest were excised using a scalpel. Once isolated, the gel pieces were processed and the DNA contained within were isolated using either the QIAquick Gel Extraction Kit or QIAEX II Gel Extraction Kit (Qiagen). Refer to the QIAquick Spin Handbook and QIAEX II Handbook respectively for detailed protocols (Available at: www.qiagen.com).

2.4.9 Bacterial Transformation

2.4.9.1 Electroporation

Bacteria strains to be used for transformation, were grown in liquid media (Section 2.4.1) overnight (approximately for 19 hours) to saturation at an appropriate temperature. A new liquid media was set up the next day and inoculated with the saturated bacteria culture at a 1/50 to 1/100 dilution. The newly prepared liquid culture was grown to an OD600 of 0.5 at an appropriate incubation temperature in an orbital shaker. The OD600 value was measured using a Jenway 6505 spectrophotometer. Once an OD600 value of 0.5 was reached for the inoculated culture, between 1 to 5 mL of the culture was centrifuged for 7 minutes at 3,000 rpm at 4°C in the Eppendorf Centrifuge 5415 R microcentrifuge. The supernatant was pipetted out and discarded, leaving behind a bacterial pellet. The pellet was resuspended in 1 mL 10% (v/v) Glycerol and centrifuged for 7 minutes at 3,000 rpm at 4°C in the Eppendorf Centrifuge 5415 R microcentrifuge before the removal of the supernatant. This step was repeated twice more. After the final resuspension and centrifugation step, the bacterial pellet was resuspended in 100 µL of 10% Glycerol.

50 µL of the resuspended bacterial pellet was used for each transformation reaction along with either 1 µL to 5 µL of plasmid DNA (Section 2.4.3) or ligation independent cloning (LIC) mix (Section 2.5.1). The bacterial suspension and DNA were mixed and loaded into a Gene Pulser/Micropulser Electroporation Cuvette with a 0.2 cm gap width (BioRad). The cuvette was

Table 2.4.9. List of media mentioned in section 2.4.9

Media Name	*Components at their final concentrations (Autoclaved/Filter Sterilised)*
SOC Medium	2% Tryptone, 0.5% (w/v) Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ (Autoclaved) + 20 mM D-Glucose (Filter Sterilised)

Table 2.4.12.1 List of buffers mentioned in section 2.4.12.1

Buffer Name (volume)	*Components and their respective amounts/volumes (Autoclaved/Filter Sterilised)*
12% Separating Gel (approx. 10 mL)	3 mL of 40% Acrylamide/Bis-acrylamide 29:1 solution, 2.5 mL of 1.5 M Tris-HCl at pH 8.8, 100 µL of 10% (w/v) SDS, 4.35 mL of ddH_2O + 5 µL of TEMED, 50 µL of 10% (w/v) ammonium persulphate
4% Stacking Gel (approx. 5mL)	0.5 mL of 40% Acrylamide/Bis-acrylamide 29:1 solution, 1.26 mL of 0.5 M Tris-HCl at pH 6.8, 50 µL of 10% (w/v) SDS, 3.16 mL of ddH_2O + 5 µL of TEMED, 25 µL of 10% (w/v) ammonium persulphate
2X Sample Loading Buffer (20 mL)	1.2 g of SDS, 1.2 mL of 2-mercaptoethanol, 0.03 g of dithiothreitol, 9.2 mL of glycerol, 1.2 mL of 1M Tris-HCl at pH 8.0, small number of flakes of bromophenol blue, ddH_2O to 20 mL (filter sterilise)
5X Running Buffer (1 L)	15.1 g of Tris, 72 g of glycine, 950 mL of ddH_2O (Autoclaved) + 50 mL of 10% (w/v) SDS (Filter Sterilise)

Table 2.4.12.2. List of buffers mentioned in section 2.4.12.2

Buffer Name (volume)	*Components and their respective amounts/volumes (Autoclaved/Filter Sterilised)*
Coomassie Brilliant Blue Staining Solution (1 L)	450 mL of methanol, 2.5g of Brilliant blue R, 100 mL acetic acid, 450 mL of ddH_2O
Coomassie Destaining Solution (1 L)	100 mL of ethanol, 100 mL of acetic acid, 800 mL of ddH_2O

Table 2.4.12.3. List of buffers mentioned in this section

Buffer Name	*Components at their final concentrations (Autoclaved/Filter Sterilised)*
Towbin/Transfer Buffer	25 mM Tris, 192 mM glycine, 20% (v/v) methanol

inserted into the BioRad MicroPulser™ and electroporation was then performed according to the suggested settings in the MicroPulser™ Electroporation Apparatus Operating Instructions and Applications Guide (Available at: www.bio-rad.com). In this project, the program EC2 of the MicroPulser™ was used. After electroporation, the mix was supplemented with 1 mL of SOC media (Table 2.4.9) and allowed to recover by incubation at an appropriate temperature for 1 hour in an orbital shaker. After recovery, 100 µL to 150 µL were plated on an LB agar plate containing the appropriate selection antibiotics (Section 2.4.1) which would then be allowed to incubate overnight at a suitable temperature.

2.4.9.2 Chemically competent cell transformation

Chemically competent cell transformation for commercially attained bacterial strains (Section 2.2.2.1) were performed according to the respective manufacturer's recommended protocol. Otherwise, the following protocol was performed.

A 50 µL aliquot of chemically competent bacteria cells was thawed on ice for 30 minutes. Once thawed, 1 µL to 5 µL of plasmid DNA (Section 2.4.3) or ligation independent cloning (LIC) mix (Section 2.5.1) were added to the aliquot of bacteria cells and mixed. The mixture was left on ice for 30 minutes, after which, heat-shock at 42°C was performed for 1.5 minutes. 500 µL of SOC media (Table 2.4.9) was then added to the heat-shocked mixture, which was then allowed to recover by incubation at an appropriate temperature for 1 hour in an orbital shaker. After recovery, 200 µL were plated on an LB agar plate containing the appropriate selection antibiotics (Section 2.4.1) which would then be allowed to incubate overnight at a suitable temperature.

2.4.10 Induction of protein expression

2.4.10.1 Small scale

Small scale induction of protein expression was performed for the protein expression steps mentioned in results Section 4.1, 4.2 and 5.2.

The bacterial strain of interest was grown overnight (approximately 19 hours) with the appropriate antibiotics. The next day, 10 mL liquid media (Section 2.4.1) was inoculated with the bacterial strain of interest at a dilution of between 1/50 and 1/100 the total volume and grown to an OD600 of between 0.5 to 0.8 in an orbital shaker at an appropriate temperature. An aliquot of the bacterial culture was then removed to represent the culture of uninduced bacteria, a control. An appropriate concentration of inducer (e.g. L-arabinose (Sigma Adrich)) was added to the remaining culture. The culture was then incubated to saturation at a pre-determined temperature for a specified amount time.

2.4.10.2 Large scale

See Section 2.8.1 for further details.

2.4.11 Bacteria cell lysis

2.4.11.1 BugBuster™ protein extraction reagent (Merck-Millipore)

Several cultures of induced and uninduced bacteria strains from Section 2.4.10.1 were lysed using the BugBuster™ Protein Extraction Reagent (Merck-Millipore). 10X BugBuster™ Protein Extraction Reagent was diluted to 1X concentration using phosphate buffered saline (PBS) (Life Technologies). A protocol was adapted from Merck's BugBuster™ Protein Extraction Reagent User Protocol as is mentioned below. The protein samples representing the different fractions were prepared for SDS-PAGE analysis which will be discussed in Section 2.4.12.

1 mL of the bacteria culture was spun down at 10,000 rpm for 10 minutes using the Heraeus™ Pico™ microcentrifuge. The supernatant was removed and the pellet was resuspended in 1X BugBuster™ Protein Extraction Reagent. The mixture was places on a rotating mixer for 20 minutes. After which, a 20 µL aliquot of the mixture was separated to represent the 'total protein' fraction. The remaining 180 µL of the mix was centrifuged at 13,000 rpm for 5 minutes using the Heraeus™ Pico™ microcentrifuge. For the 'soluble protein' fraction, 20 µL of the supernatant was isolated. For the 'insoluble protein' fraction, the supernatant was initially removed, the pellet was resuspended in 180 µL PBS and 20 µL of the final resuspended pellet was isolated.

2.4.11.2 Sonication

See Section 2.8.3 for further details.

2.4.12 Protein sample analysis

2.4.12.1 SDS-PAGE

The SDS-PAGE gels were prepared first. A sufficient volume of the 12% separating gel (Table 2.4.12.1) was added first into a set up with the Bio-Rad Mini-PROTEAN® 3 Glass Plates. A layer of isopropanol was added to the top of the gel. The gel was allowed to polymerise and once it had, the isopropanol was removed and the top of the gel was rinsed with ddH_2O. After which, an appropriate amount of the 4% stacking gel (Table 2.4.12.1) was added before placement of a gel comb. The gel was allowed to complete polymerisation before use. If needed, the gels were stored at 4°C within the plates, wrapped with a moist paper towel.

Prior to sample loading, the sample loading buffer (Table 2.4.12.1) was added to the protein samples to get a final concentration of 1X. The mixture was boiled for 7 minutes before loading into the gel. However, samples containing membrane proteins were not boiled but instead,

heated to a temperature of between 55°C to 65°C for a longer period of time (approximately 20 minutes) before loading into the gel.

For SDS-PAGE, the Bio-Rad Mini-PROTEAN® 3 vertical or the Mini-PROTEAN® Tetra systems were used. The abovementioned casted gels were set up accordingly into the system. The 5X running buffer (Table 2.4.12.1) was diluted to 1X using ddH_2O before being added in an appropriate amount into the system. For western blot and coomassie, 7 µL and 20 µL of the protein sample mix was loaded into the gels respectively. At least one lane on each gel was loaded with 7 µL of the protein marker, Precision Plus Protein™ Kaleidoscope™ Standards or Precision Plus Protein™ All Blue Standards (Bio-Rad). Gel electrophoresis was allowed to run for approximately 40 to 50 minutes at 200 V or when the dye front has reached the bottom of the gel.

2.4.12.2 Coomassie staining

The gel designated for coomassie staining was removed from the glass plates and sufficiently rinsed with ddH_2O. A small volume of coomassie brilliant blue staining solution (Table 2.4.12.2) was added to cover the top of the gel. Staining was performed for 1 hour on a rocker at room temperature. After which, the staining solution was removed and the stained gel was sufficiently rinsed with ddH_2O. An adequate amount of the coomassie destaining solution (Table 2.4.12.2) was then added to the stained gel. Destaining was allowed to occur overnight (approximately 19 hours). The gel was then visualised using the G:BOX gel doc (Syngene).

2.4.12.3 Western transfer and antibody staining

Western transfer was performed in either of the following two systems, the iBlot® Dry Blotting System (Life Technologies) or the Trans-Blot® SD Semi-Dry Transfer Cell (Bio-Rad).

Western transfer using the iBlot® Dry Blotting System was performed with the iBlot transfer stack containing a nitrocellulose membrane. The recommended protocol was adhered to (Available at: www.lifetechnologies.com). Western transfer using the Trans-Blot® SD Semi-Dry Transfer Cell was performed with a Immobilon-P PVDF membrane (Merck-Millipore) and the transfer buffer (Table 2.4.12.3) following the manufacturer's recommended protocol (Available at: www.bio-rad.com).

After the completion of the western transfer from either of the two systems, the nitrocellulose or PVDF membranes were prepared for western blotting using the protocol mentioned in the QIAexpress Detection and Assay Handbook (Available at: www.qiagen.com). The primary antibody used was the Tetra·His Antibody (Qiagen) and the secondary antibody was the goat anti-mouse IgG with a HRP conjugate (Sigma-Aldrich).

The western blotting of the membranes were then detected via chemiluminescent detection. This was performed with the Luminata Classico Western HRP substrate (Merck-Millipore) according to the protocol suggested by the manufacturer (Available at:

www.mercekmillipore.com). The results were developed in the dark room and onto an Amersham Hyperfilm ECL (GE Healthcare Life Sciences) film.

2.5 Ligation Independent Cloning (LIC)

LIC is a method of cloning which involves the insertion of a gene into an expression plasmid without the use of a DNA ligase. Instead, this method utilises long exposed tails (LIC tails) of complementary DNA from both the insert and the plasmid which binds with more strength than the usual complementary ends after restriction enzyme digest. After annealing, the resulting nick would be fixed by the host strain after transformation.

In this project, genes of interest were cloned into the pBADcLIC2005 vector possessing regions of DNA flanking the insert site which are able to be converted into LIC tails. The sequences to be converted to the LIC tail were (5' ATG GGT GGT GGA TTT GCT 3') and (5' GAA AAT TTA TAC TTC CAA 3').

In order to engineer the gene of interest with the LIC tails, forward and reverse primers were designed to possess a region complementary to the gene and with the remaining being the LIC tail sequence. High fidelity PCR (Section 2.4.4) was performed using the isolated gDNA (Section 2.4.2) of the strain possessing the gene of interest as a template and the designed primers. The PCR products were then purified using the QIAquick PCR Purification (Qiagen).

The plasmid vector, pBADcLIC2005, was isolated from the host strain using the protocol mentioned in Section 2.4.3. To expose the two regions of DNA which would be converted to the LIC tail, the plasmid DNA was digested by the *SwaI* restriction enzyme (NEB) using the suggested time-saver protocol (Available at: www.neb.com).

At this point, both the insert and the plasmid were linear with the two regions of DNA, for the LIC tail, flanking the gene or the rest of the plasmid respectively. To generate the LIC tails from these regions of DNA, the 3' to 5' exonuclease activity of the T4 DNA polymerase (NEB) was used. 200 ng of PCR products or plasmid DNA were made up to 10 µL by direct dilution with ddH_2O or by concentrating the DNA sample first (Section 2.4.5) if need be. The 10 µL of DNA sample were then mixed with 3 µL of NEBuffer 2 (NEB), 25 mM of dCTP for the sample of vector DNA or dGTP for the PCR products (Fermentas) and 0.5 µL of T4 DNA Polymerase (NEB). The mixture was then incubated for 30 minutes in a water bath at 20°C. To stop the enzymatic reactions, the mixture was then treated at 75°C for 20 minutes.

The T4 DNA polymerase treated samples of plasmid DNA and PCR products were then mixed in a molar ratio of 1:2 for annealing. The mixture was then incubated at 20°C in a water bath for 10 minutes before the addition of 25 mM of EDTA. A further 10 minute incubation at 20°C was performed.

The annealing mixture was then used for transformation into either a chemically competent *E. coli* or by electroporation using the methods mentioned in Section 2.4.9. For the pBADcLIC2005 vector, antibiotic selection was performed using ampicillin (Sigma-Aldrich) at a concentration of 100 μg/mL.

2.6 Site Directed Mutagenesis by inverse PCR and blunt end ligation

To introduce specific mutations to proteins of interest, mutagenesis by inverse PCR and blunt end ligation was performed. This utilises expression plasmids containing the gene for the protein of interest and mutagenic primers.

Firstly, divergent primers (Table 2.1.4) were designed to amplify the plasmid while also introducing a specific mutation by utilising a mutagenic primer containing the corresponding base changes. The mutation was designed to be in the middle of the mutagenic primer in order to maximise the binding of the primer to the template plasmid DNA. The expression plasmid containing the gene was purified (Section 2.4.3) from the corresponding bacterial strains. Using the purified expression plasmid, high fidelity PCR (Section 2.4.4) with the designed divergent mutagenic primers was performed. The resulting products were then purified by PCR purification (Section 2.4.4).

Blunt end ligation was set up with 1 μL of purified PCR products, 2 μL of T4 DNA ligase buffer (NEB), 1 μL T4 DNA ligase (NEB), 1 μL of T4 Polynucleotide Kinase (NEB) and 15 μL of ddH_2O. The mixture was allowed to incubate at 16°C overnight (approximately 19 hours).

The ligated products were then transformed into a strain of chemically competent *E. coli* (Section 2.4.9.2) before being transformed via electroporation (Section 2.4.9.1) into a *S.* Typhimurium strain for downstream *in vivo* analyses. The mutant genes were confirmed by sequencing (Section 2.4.7.3).

2.7 BTP1 phage plaque assay

To analyse the susceptibility of a strain of *Salmonella* to the BTP1 phage, a phage plaque assay was performed.

2.7.1 Phage stock preparation

Phage stocks of BTP1 was attained by first culturing the *S.* Typhimurium D23580 strain until saturation (Section 2.4.1). Next, the culture was centrifuged at 10,000 rpm for 10 minutes in the Heraeus™ Pico™ microcentrifuge to pellet the bacteria. The phage containing supernatant was isolated from the pellet and a few drops of chloroform was added in order to kill any remaining

bacteria. When necessary, the phage stocks were diluted with PBS to the desired fold concentration.

2.7.2 Top agar preparation

The phage plaque assay was performed using an LB agar plate that is layered with a LB agar called the 'top agar' (final concentrations 20% Lennox Broth Powder, 0.7% Agar).

To prepare the plate, an LB agar plate was initially prepared following the protocol mentioned in Section 2.4.1 with the necessary concentration of supplements (i.e. inducer and antibiotics) when required. For each LB agar plate, 3.5 mL of top agar was prepared, supplemented with inducer and antibiotics at the proper concentrations when necessary and mixed with 100 µL of pre-prepared saturated culture of the test bacteria strain. The final mixture was then layered on the pre-prepared LB agar plate and was allowed to solidify on the existing LB agar.

2.7.3 BTP1 phage treatment

When the top agar layer has solidified as mentioned in Section 2.10.2, 5 µL of BTP1 phage stocks and dilutions were added drop wise onto the top agar layer at designated positions. The drop was allowed to be absorbed into the top agar layer before incubation at an appropriate temperature overnight (approximately 19 hours). The resulting plates were then visualised.

2.8 Large scale protein expression, bacteria cell lysis and vesicle/membrane fraction isolation

2.8.1 Large scale protein expression

Large scale induction of protein expression was performed as mentioned in results Section 4.3. The bacterial strain of interest was grown to saturation overnight (approximately 19 hours) with the appropriate antibiotics. The next day, pre-warmed 625 mL liquid media in a 2 L baffled flask was inoculated with the bacterial strain of interest at a dilution of between 1/50 and 1/100 the total volume and grown to an OD600 of between 0.5 to 0.8 in an orbital shaker at an appropriate temperature. An aliquot of the bacterial culture was then removed to represent the culture of uninduced bacteria, a control. An appropriate concentration of inducer (e.g. L-arabinose) was added to the remaining culture. The culture was then incubated to saturation at a pre-determined temperature for a specified amount time.

2.8.2 Bacterial harvest

The bacteria cultures were allowed to reach saturation after induction. After which, the cells were harvested by centrifugation at 5000 rpm for 15 minutes at 4°C using the Heraeus™ Megafuge™ 40 centrifuge (Thermo Scientific). The supernatant was discarded and the pellet was resuspended in 35 mL resuspension buffer (Table 2.8.1). From here henceforth, the cells were kept on ice.

2.8.3 Sonication

For sonication, 1 mM of AEBSF (Sigma-Aldrich), 1mM $MgCl_2$ and approximately 100 µg/mL of DNase I (Sigma-Aldrich) was added to the 35 mL of resuspended bacteria. The mixture was attached to a set up with the Misonix Sonicator 3000 Homogenizer equipped with the ½ inch horn. Sonication was performed at 4 W with a 3 second pulse time and a 7 second rest time for 3 minutes in total.

2.8.4 Vesicle isolation

After sonication, the tube containing the sonicated bacteria suspension was kept on ice for 5 minutes before the addition of 5 mM EDTA at pH 7.5. A low speed centrifugation of the lysate was performed at 13,000 rpm for 20 minutes at 4°C using the JA-25.50 rotor in the Beckman Avanti J-26 High Speed Centrifuge. The resulting supernatant was retained. Next, the supernatant was centrifuged at 40,000 rpm for 1 hour at 4°C in the Beckman L7-65 Ultracentrifuge and the Type 70 Ti rotor. The supernatant was discarded and the pellet was resuspended in 10 mL resuspension buffer (Table 2.8.1). The suspension contains the vesicles/insoluble fraction and was distributed into 0.5 mL aliquots and snap-frozen with liquid nitrogen and stored in -80°C.

2.9 Detergent based membrane protein solubilisation and column based immobilised metal affinity chromatography (IMAC)

2.9.1 Detergent based solubilisation of membrane proteins

Vesicles isolated as mentioned in Section 2.5.2.4 were mixed appropriately with the solubilisation buffer (Table 2.9.1), mixed by inversion and incubated on ice for 30 minutes. The mixture was then centrifuged at 53,000 rpm for 20 minutes at 4°C using the Beckman L7-65 Ultracentrifuge and the Type 70 Ti rotor.

2.9.2 Column based immobilised metal affinity chromatography (IMAC)

For IMAC, the HisTrap™ HP (GE Healthcare Life Sciences) column was attached to the ÄKTA Prime (GE Healthcare Life Sciences) to regulate the flow rates of buffers and samples loaded.

Prior to IMAC, the HisTrap™ column was charged with fresh nickel (Ni^{2+}) ions and the column was equilibrated with the column wash buffer (Table 2.9.1).

The supernatant from the centrifugation step mentioned in Section 2.9.1 was then allowed to pass through the column. The flow through was collected and saved for western blot and coomassie staining analyses (Section 2.4.12). The pellet from the centrifugation step was resuspended in PBS in an equal volume to the supernatant and a sample was analysed by western blot and coomassie staining analyses (Section 2.4.12).

The column was then washed with 20 column volumes of column wash buffer (Table 2.9.1). The wash through was also collected and saved for western blot and coomassie staining analyses (Section 2.4.12).

The elution of the proteins were performed by passing through 1 mL of the elution buffer (Table 2.9.1) three times. The elution fractions were collected and were analysed by western blotting and coomassie analyses (Section 2.4.12) along with a sample of the vesicles isolated as mentioned in Section 2.8.4, the resuspended pellet from the centrifugation step which represents the insoluble fraction, the column flow through and wash through.

2.10 Protein purification by size exclusion chromatography (SEC)

To further purify the samples of proteins of interest isolated by IMAC as mentioned in Section 2.5.3.2, size exclusion chromatography (SEC) by gel filtration was performed. The HiLoad 26/600 Superdex 200 prep grade Gel Filtration column (GE Healthcare Life Sciences) was attached to an ÄKTA Purifier (GE Healthcare Life Sciences) to regulate the flow of buffer and samples. The gel filtration column was equilibrated with the SEC buffer (Table 2.10.1) before the addition of the samples of proteins. After loading, the SEC buffer was allowed to flow through the column again until 1 column volume of buffer had flowed through. Fractions of the flow through were continuously collected until the flow of the SEC buffer was halted. The A_{280} values detected by the sensor attached to the ÄKTA Purifier was used to determine which fractions might contain the protein of interest and these fractions were analysed by coomassie analysis (Section 2.4.12).

Table 2.8.1. List of buffers mentioned in section 2.8

Buffer Name	*Components at their final concentrations (Autoclaved/Filter Sterilised)*
Potassium Phosphate (KPi) Buffer	Add 1 M KH_2PO_4 (Acidic) drop wise to 1 M K_2HPO_4 (Basic) until pH reaches 7.8
Resuspension Buffer	50 mM KPi Buffer pH 7.8, 20% (v/v) glycerol

Table 2.9.1. List of buffers mentioned in section 2.9

Buffer Name	*Components at their final concentrations (Autoclaved/Filter Sterilised)*
Potassium Phosphate (KPi) Buffer	Add 1 M KH_2PO_4 (Acidic) drop wise to 1 M K_2HPO_4 (Basic) until pH reaches 7.8
Solubilisation Buffer	50 mM KPi Buffer pH 7.8, 200 mM NaCl, 20% (v/v) glycerol, 10 mM Imidazole, 3 mM dithiothreitol, 0.5% (w/v) n-Dodecyl β-D-maltoside + Vesicles (Section 2.5.1.4)
Column Wash Buffer	50 mM KPi Buffer pH 7.8, 200 mM NaCl, 20% (v/v) glycerol, 40 mM Imidazole, 3 mM dithiothreitol, 0.04% (w/v) n-Dodecyl β-D-maltoside
Elution Buffer	50 mM KPi Buffer pH 7.8, 200 mM NaCl, 20% (v/v) glycerol, 500 mM Imidazole, 3 mM dithiothreitol, 0.04% (w/v) n-Dodecyl β-D-maltoside

Table 2.10.1 List of buffers mentioned in section 2.10

Buffer Name	*Components at their final concentrations (Autoclaved/Filter Sterilised)*
SEC Buffer	50 mM KPi Buffer pH 7.8, 200 mM NaCl, 5 mM dithiothreitol, 0.05% (w/v) n-Dodecyl β-D-maltoside

Chapter 3

In silico Analyses of *Salmonella* O-antigen Acetyltransferases, OafA and GtrC family II

Chapter 3 Introduction

Literature based studies identified two *Salmonella* lipopolysaccharide (LPS) O-antigen membrane bound acetyltransferases in GtrC family II proteins (Davies et al., 2013) and OafA (Slauch et al., 1996).

The GtrC family II proteins are thought to acetylate the rhamnose residue of the LPS O-antigen. Previous studies by Kintz et al. (2015) on the GtrC family II protein of *S.* Typhimurium D23580, $GtrC^{BTP1}$, revealed the importance of this protein in the acetylation of the rhamnose residue and hence, protection against BTP1 bacteriophage associated killing.

OafA, on the other hand, is only found in *S.* Typhimurium and is thought to acetylate the abequose residue of the LPS O-antigen, conferring the *S.* Typhimurium O:5 serotype. Extensive microbiological and immunological studies have been performed on OafA with its function being linked to the adherence and invasion (Michetti et al., 1994), the structure of the LPS and the triggering of the immune response (Slauch et al., 1995).

The primary function of GtrC family II proteins and OafA differ only in their substrate specificity, with the former being specific for the rhamnose and the latter for the abequose residue of the LPS O-antigen. To determine what sets these two proteins apart, *in silico* studies were first performed to determine any similarities and differences between these proteins while also identifying any potential associations to known protein families which would provide a better understanding of their respective characteristics.

3.1 Sequence similarities and differences between GtrC family II proteins and OafA

To identify the similarities within the sequences of OafA and the GtrC family II proteins, the nucleotide and amino acid sequences were compared using an alignment tool. The amino acid sequences for each protein were aligned pairwise and percentage identity values were tabulated onto Table 3.1.1 for the amino acid sequences.

The percentage amino acid identity values between each of all of the GtrC family II proteins tabulated on Table 3.1.1 and depicted on Figure 3.1.1 showed that $GtrC^{BTP1}$ is particularly divergent from the rest of the GtrC Family II proteins. This is consistent with the findings of (Davies et al., 2013) which identified an earlier speciation event from a common ancestor than that which led to the formation of the rest of the *gtrC* family II genes. The *gtr* operon that $gtrC^{BTP1}$ is located on is found on the BTP1 prophage of *S*. Typhimurium D23580 while the rest of the *gtrC* family II genes are not thought to lie in any identified prophage. For the ease of future analysis, the amino acid sequences of all GtrC family II proteins except for $GtrC^{BTP1}$ were aligned using Geneious Basic 5.6.7 and the resulting consensus sequence, referred to as 'GtrC family II consensus', will be used to represent all of these proteins.

The *oafA* gene encodes for OafA, a protein that is functionally similar but with a different substrate specificity and hence, the amino acid sequence was expected to be significantly different than the GtrC family II proteins. From Table 3.1.1, the percentage similarity values between the amino acid sequences of OafA and the GtrC family II proteins are only between 28.1% and 29.2%. Figure 3.1.1 shows that the amino acid sequence for OafA is least conserved among all of the amino acid sequences analysed. The acetyltransferase function of OafA and GtrC family II proteins could be carried out by any of the few conserved amino acid regions of the proteins while the cause for the substrate specificity might lie elsewhere in the amino acid sequences.

Figure 3.1.2 highlights the various conserved regions between OafA and the GtrC family II proteins. The sequences that contribute to the functional similarities of OafA and GtrC family II might lie within the identified regions of conserved amino acids. To determine if these conserved amino acids are within any known domains, various domain prediction analyses were performed and discussed in the following sections.

	SEN P125109	SG 287/91	SeD CT02021853	SSPA 12601	SPA 9150	STY CT18	STY TY2	STM D23580 (GtrC^{BTP1})
SG 287/91	99.8%	-	-	-	-	-	-	-
SeD CT02021853	99.5%	99.4%	-	-	-	-	-	-
SSPA 12601	98.9%	98.8%	98.4%	-	-	-	-	-
SPA 9150	98.9%	98.8%	98.4%	100%	-	-	-	-
STY CT18	98.6%	98.4%	98.1%	99.4%	99.4%	-	-	-
STY TY2	98.6%	98.4%	98.1%	99.4%	99.4%	100%	-	-
STM D23580 (GtrC^{BTP1})	77.4%	77.2%	76.9%	77.5%	77.5%	77.2%	77.2%	-
STM (OafA)	29.2%	29.1%	28.1%	29.2%	29.2%	28.9%	28.9%	28.3%

Table 3.1.1. The percentage identity between amino acid sequences of GtrC family II and OafA proteins when aligned pairwise with the corresponding proteins. *S. enterica* strains with GtrC family II were identified black with specific names within brackets, while the serovar with OafA was identified by the red name followed by (OafA*)*. SG = *S.* Gallinarum, SeD = *S.* Enteritidis, SSPA/SPA = *S.* Paratyphi A, STY = *S.* Typhi, STM = *S.* Typhimurium. For alignment parameters, see Section 2.3.2. Alignments were performed using Geneious Basic 5.6.7 (Drummond et al., 2012).

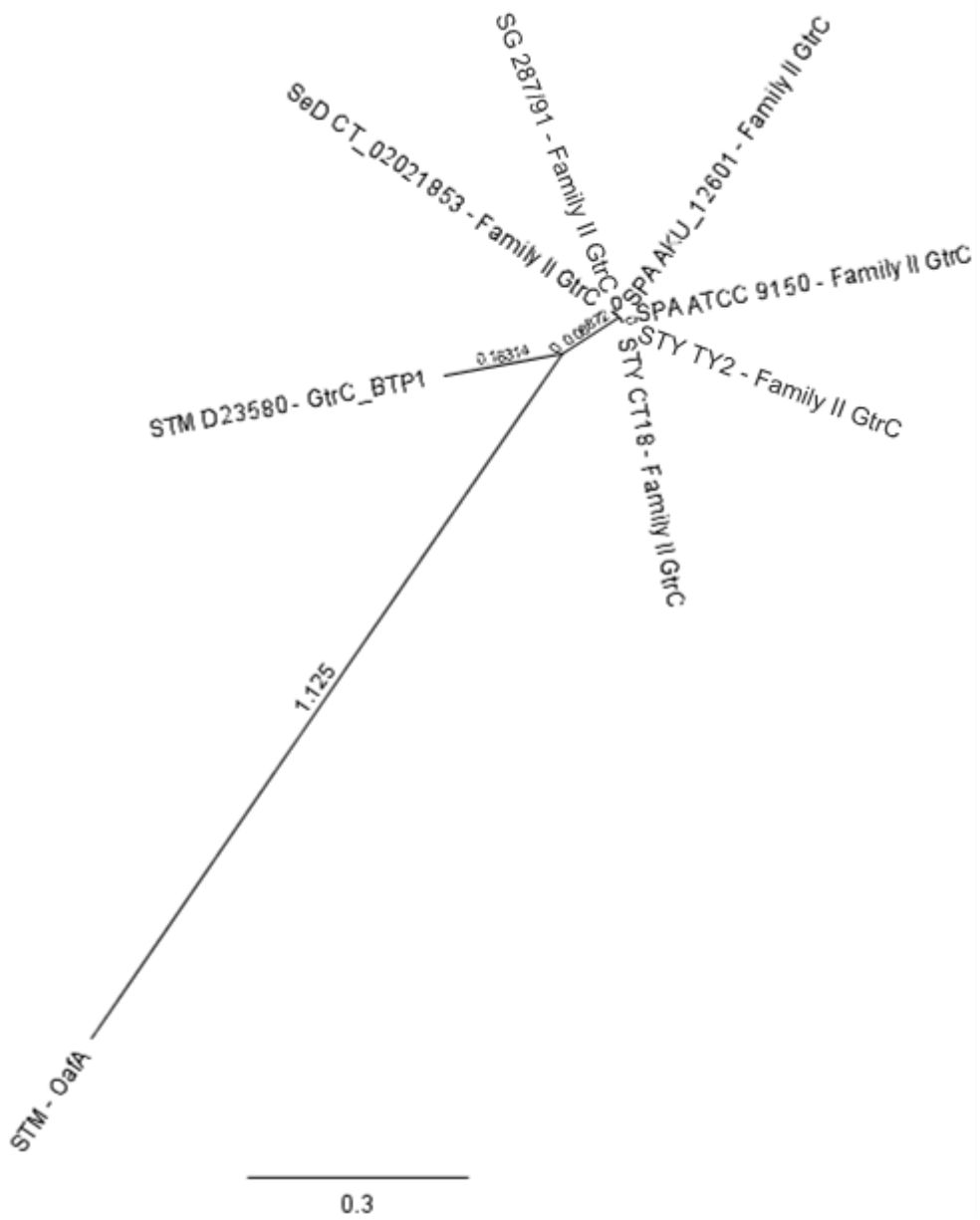


Figure 3.1.1. Phylogenetic analysis of the amino acid sequences of GtrC family II proteins, including GtrC^{BTP1}, and OafA from *Salmonella.* Depicted, is an unrooted tree with each branch being identified by the *Salmonella* strain followed by the name of the protein. Where applicable, each branch possesses a substitutions per site value and is drawn to a scale. OafA (denoted by STM - OafA) is the most divergent among the analysed proteins while GtrC^{BTP1} (denoted by STM D23580 - GtrC_BTP1) is the most divergent among the GtrC family II proteins. SG = *S.* Gallinarum, SeD = *S.* Enteritidis, SSPA/SPA = *S.* Paratyphi A, STY = *S.* Typhi, STM = *S.* Typhimurium. For alignment parameters, see Section 2.3.2. Alignments were performed and the tree was built using Geneious Basic 5.6.7 (Drummond et al., 2012).

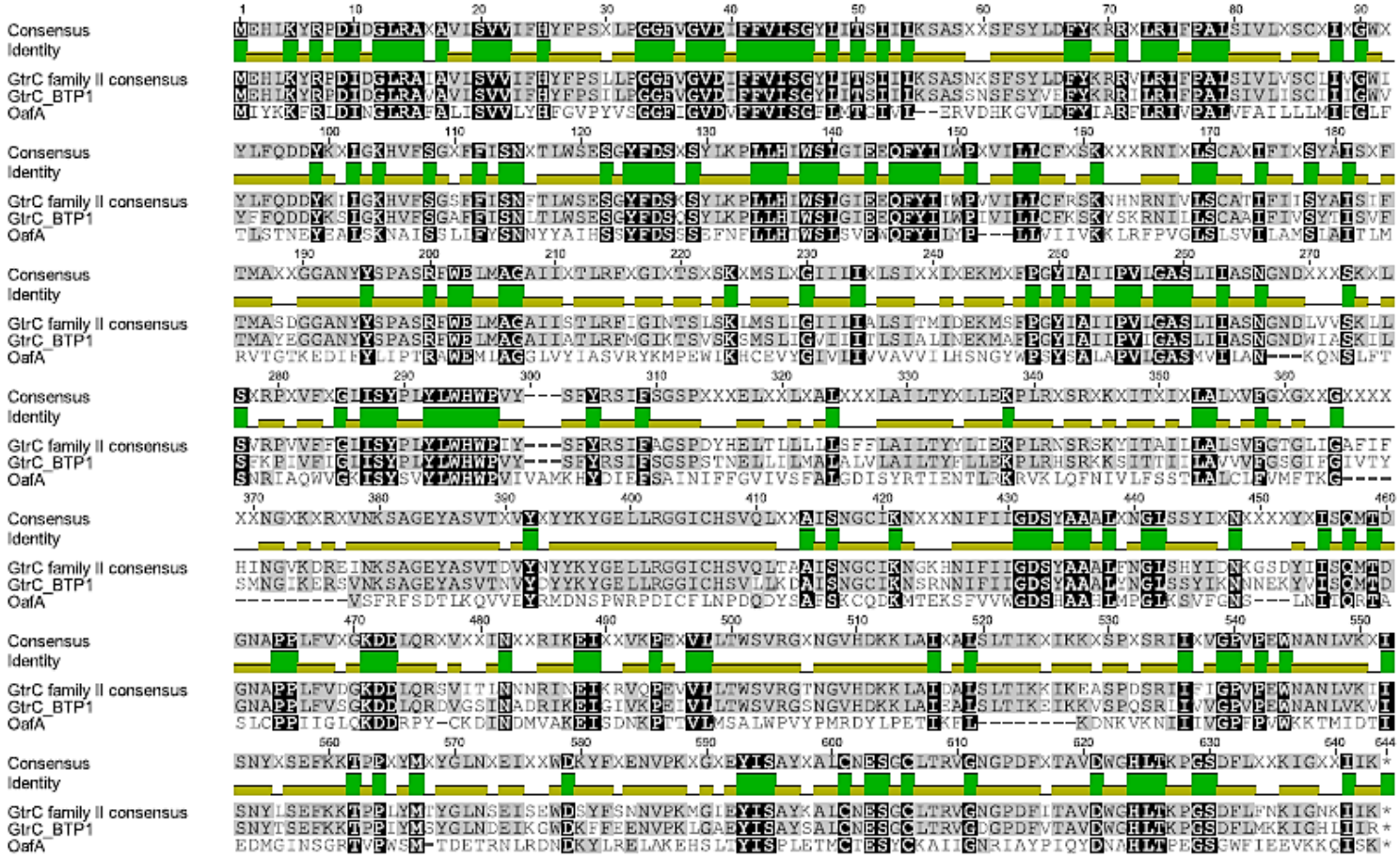


Figure 3.1.2. Amino acid sequence alignments of OafA, GtrC^{BTP1} (GtrC_BTP1) and GtrC family II consensus. Conserved amino acid residues identified within all three sequences are highlighted in black and identified by the green identity bars, the residues aligned in only 2 out of the 3 sequences are highlighted in grey and identified by the brown identity bars and residues with no identity are not highlighted and do not have an identity bar. For alignment parameters, see Section 2.3.2. Alignments were performed using Geneious Basic 5.6.7 (Drummond et al., 2012).

3.2 Functional domain predictions of OafA, GtrC^{BTP1} and other GtrC family II proteins

OafA and the GtrC family II proteins are thought to be functionally similar with their main functions being acetyltransferases involved in the addition of an acetyl group onto specific sugar residues on the LPS O-antigen. However, the differences lie in the sugar substrate specificity. Conserved amino acids identified in Section 3.1 between OafA and GtrC family II proteins might be conserved between proteins possessing certain functional domains. To determine if these proteins of interest might possess any known functional domains, the amino acid sequences were searched against the InterPro database and their identified functional domains were aligned below the respective amino acid regions. The results are depicted in Figures 3.2.1, 3.2.2 and 3.2.3 for GtrC family II consensus, GtrC^{BTP1} and OafA respectively.

OafA, GtrC^{BTP1} and other GtrC family II proteins (i.e. GtrC family II consensus) have the same functional domains and the same domain architectures with the acyltransferase (acetyltransferase) 3 domain (Pfam: PF01757) spanning most of the N-terminal half and SGNH hydrolase type esterase spanning the C-terminal half of the protein.

3.2.1 Acyltransferase (acetyltransferase) 3 domain

The acyltransferase 3 family of enzymes (Pfam: PF01757) consist of proteins that are involved in the transfer of acyl groups other than amino-acyl (i.e. from tRNAs). Proteins possessing this functional domain family are mostly uncharacterised and usually possess approximately 300 amino acids. Hence, there is a need to expand our understanding of proteins within this domain family, structurally and functionally. Some of the proteins within this family are further classified into the membrane-bound *O*-acetyltransferase (MBOAT) group of enzymes. Acetyl is a form of acyl group, specifically possessing a methyl group bound to a carboxyl group. The N-terminal domain of OafA and all GtrC family II proteins will be referred to as the acyltransferase 3 domain while their function will continue to be referred to as acetyltransferases.

3.2.2 SGNH hydrolase type esterase domain

The SGNH hydrolase type esterase (SUPERFAMILY: SSF52266), or SGNH esterase for short, is a divergent family of esterases with underlying functions similar to the classical GDSL family of lipases and esterases. However, rather than having the usual α/β fold depicted on Figure 3.2.4(a) where alpha helices stack above about eight beta strands (Ollis et al., 1992), it possesses an α/β/α fold where about 8 alpha helices occupy both sides of 5 beta strands as shown on Figure 3.2.4(b) using the rhamnogalacturonan acetylesterase (RGAE) of *Aspergillus aculeatus* (Mølgaard et al., 2000) for example. Both GDSL and SGNH esterases have the classical serine

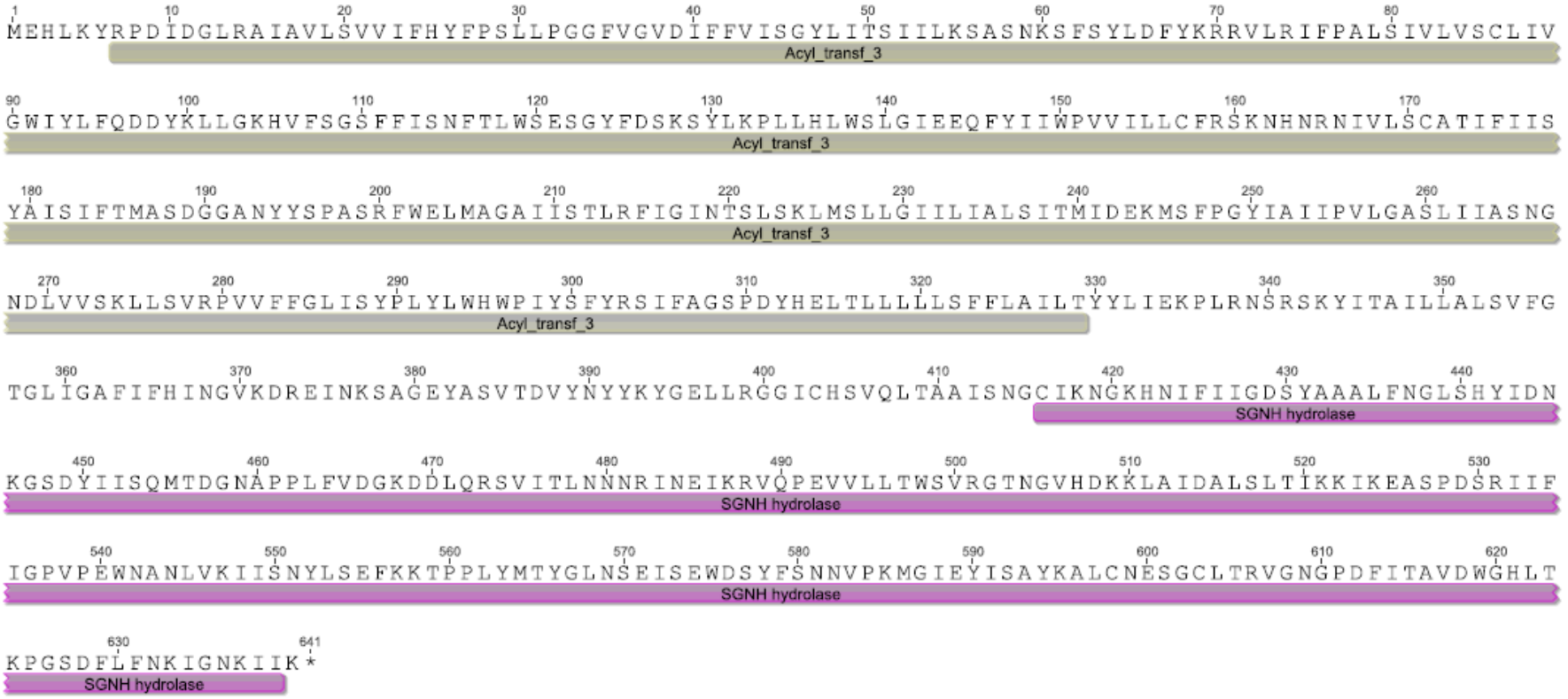


Figure 3.2.1. Predicted functional domains of GtrC family II consensus. Predicted functional domains of within the protein were annotated as a coloured bar spanning the regions thought to represent the respective domains. Acyl_transf_3 refers to the group of proteins containing the acyltransferase 3 domain annotated on the PFAM database (PF01757). SGNH hydrolase refers to the group of proteins possessing the SGNH hydrolase type esterase domain annotated on the SUPERFAMILY database (SSF52266). Predictions were done using InterProScan (Jones et al., 2014, Hunter et al., 2011) on Geneious Basic 5.6.7 (Drummond et al., 2012).

Figure 3.2.2. Predicted functional domains of GtrC^{BTP1}. Predicted functional domains of within the protein were annotated as a coloured bar spanning the regions thought to represent the respective domains. Acyl_transf_3 refers to the group of proteins containing the acyltransferase 3 domain annotated on the PFAM database (PF01757). SGNH hydrolase refers to the group of proteins possessing the SGNH hydrolase type esterase domain annotated on the SUPERFAMILY database (SSF52266). Predictions were done using InterProScan (Jones et al., 2014, Hunter et al., 2011) on Geneious Basic 5.6.7 (Drummond et al., 2012).

Figure 3.2.3. Predicted functional domains of OafA. Predicted functional domains of within the protein were annotated as a coloured bar spanning the regions thought to represent the respective domains. Acyl_transf_3 refers to the group of proteins containing the acyltransferase 3 domain annotated on the PFAM database (PF01757). SGNH hydrolase refers to the group of proteins possessing the SGNH hydrolase type esterase domain annotated on the SUPERFAMILY database (SSF52266). Predictions were done using InterProScan (Jones et al., 2014, Hunter et al., 2011) on Geneious Basic 5.6.7 (Drummond et al., 2012).

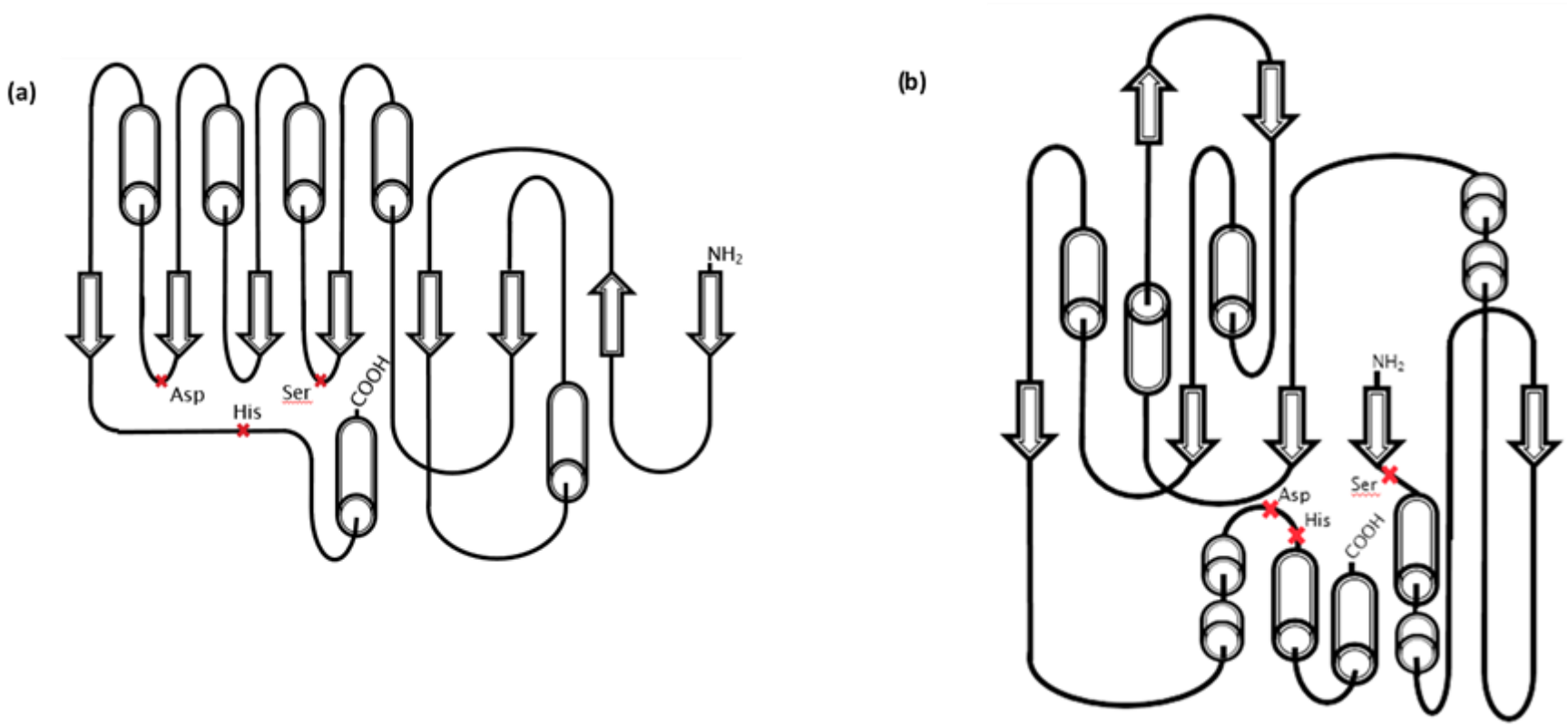


Figure 3.2.4. Topological diagram of (a) general classical GDSL esterase following an α/β fold, and (b) SGNH hydrolase type esterase following an α/β/α fold which is based on rhamnogalacturonan acetylesterase (RGAE) from *Aspergillus aculeatus* (Mølgaard et al., 2000). The α-helices are depicted as cylinders, β-strands are depicted as arrows and loops between α-helices and β-strands are represented by single solid black lines. Typical serine esterase catalytic residues of GDSL esterase and the known serine esterase catalytic residues of RGAE were identified by the red crosses with their respective names.

esterase Ser-Asp-His catalytic triad but they are located in different positions on their secondary structures. Notably, the catalytic aspartate and histidine residues in RGAE are separated by two amino acid residues rather than the more significant distance between the two catalytic residues in the GDSL esterase as shown on Figure 3.2.4. Mølgaard et al. (2000) also identified two amino acid residues in RGAE, Gly42 and Asn74, which form the oxyanion hole which functions as a stabiliser for the active site during binding of the substrate.

3.2.3 Domain architecture of OafA and GtrC family II proteins

The functional domain detection of OafA and $GtrC^{BTP1}$ using the InterProScan identified two potential functional domains, acyltransferase 3 (Pfam: PF01757) and SGNH hydrolase type esterase (SUPERFAMILY: SSF52266), on both proteins with the former being on the N-terminal end and the latter on the C-terminal end of the proteins. From here henceforth, the N-terminal Acyltransferase 3 and C-terminal SGNH hydrolase type esterase domain architecture will be referred to as the acyltr3-SGNH domain architecture. Proteins of similar domain architectures were determined by searching for either PF01757 or SSF52266 on the InterPro database. As of February 2015, the search returned with 13627 proteins predicted or known to have the Acyltr3-SGNH domain architecture.

One notable protein that was identified by this search is the *S. aureus* peptidoglycan *O*-acetyltransferase, OatA. *S. aureus* is known to be a pathogen that is fully unaffected by lysozyme, due to the presence of this protein. The modification caused by OatA gives *S. aureus* the ability to persist and colonise in the host (Bera et al., 2005).

3.2.4 Multiple component acetyltransferase systems using the acyltransferase 3 and SGNH hydrolase type esterase domains on separate proteins

Rather than having both the acyltransferase 3 and SGNH hydrolase type esterase domains on the same protein and hence, having the Acyltr3-SGNH domain architecture, some O-acetylation mechanisms utilise two separate proteins with each domain being on a separate subunit.

3.2.4.1 PatA and PatB of Neisseria gonorrhoeae (N. gonorrhoeae)

Like OatA in the gram-positive *S. aureus*, PatA and PatB are involved in the O-acetylation of the C6 hydroxyl group of muramic acid residues within the peptidoglycan of the Gram-negative *N. gonorrhoeae* and *N. meningitidis*. PatA is thought to be the MBOAT that transfers an acetyl to the periplasm where PatB then transfers the acetyl onto the peptidoglycan (Moynihan and Clarke, 2010). PatB was predicted to be an SGNH hydrolase type esterase enzyme, possessing the α/β/α fold and the Ser-Asp-His catalytic triad (Moynihan and Clarke, 2010) which are arranged in a similar way as that of RGAE as shown on Figure 3.4.2.

3.2.4.2 AlgF, AlgI, AlgJ and AlgX of Pseudomonas aeruginosa (P. aeruginosa)

In *P. aeruginosa*, alginate, a biofilm exopolysaccharide, is made up of a linear chain of D-mannuronic acid which can be acetylated at the C2 and C3 hydroxyl groups. The alginate is modified in the periplasm through *O*-acetylation by four proteins; AlgF, AlgI, AlgJ and AlgX. Like PatB and PatA in *N. gonorrhoeae* and *meningitidis*, AlgJ and AlgX are both predicted to be SGNH hydrolase type esterases with both predicted to have the α/β/α fold and the SGNH esterase type catalytic triad and AlgI is predicted to be an MBOAT enzyme respectively (Riley et al., 2013), (Baker et al., 2014). Replacement of each of the three catalytic residues with alanine in AlgX resulted in the complete absence of enzyme activity while replacement mutations performed in AlgJ only reduced enzymatic activities by 80%.

3.3 Transmembrane Structure Prediction of OafA and GtrC^{BTP1}

The functional domain prediction classified the N-terminal region of OafA and GtrC^{BTP1} as acyltransferase 3. This domain family consist of proteins that are within the inner membrane including those within the MBOAT family of enzymes and to determine if the same is true with OafA and the GtrC family II proteins, the sequences were uploaded onto the TMHMM v2.0 server which predicts the presence of transmembrane helices and the location of other regions which are not involved in a transmembrane helix using the hydrophobic and hydrophilic characteristics of the amino acids while also giving the value of likelihood that the region is actually as predicted by the software.

3.3.1 Transmembrane helices predicted in the N-terminal region of OafA and GtrC^{BTP1}

Figures 3.3.1 and 3.3.2 represents the transmembrane helix predictions for OafA and GtrC^{BTP1} respectively. The N-terminal region of the proteins, which are putatively classified in the acyltransferase 3 family of functional domains, are predicted to possess around 9 to 11 transmembrane helices. This is consistent with the hypothesis that the N-terminal region of both proteins are highly related to or within the MBOAT family of enzymes.

3.3.2 Periplasmic region predicted in the C-terminal region of OafA and GtrC^{BTP1}

Kintz et al. (2015) utilised a PhoA fusion on amino acid number 370, which is the start of the predicted C-terminal periplasmic tail, and at amino acid number 640, which is at the C-terminal end of the protein. PhoA is an alkaline phosphatase which can only fold when in the periplasm. Its activity can be assayed using XPhos (5-Bromo-4-chloro-3-indolyl-phosphate) supplemented

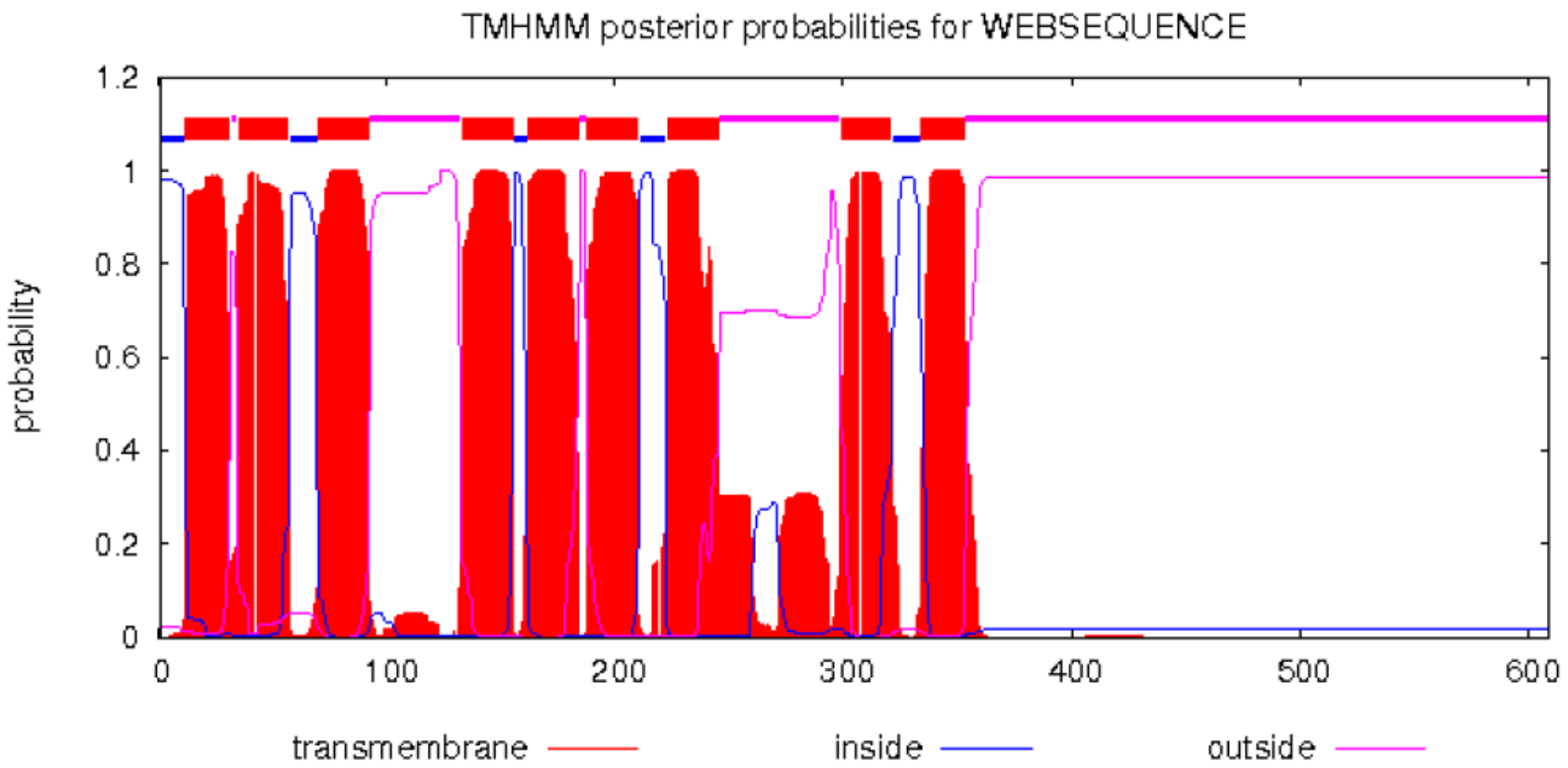


Figure 3.3.1. Transmembrane prediction of OafA. Amino acid sequence of OafA was put into the TMHMM v2.0 server (Sonnhammer et al., 1998, Krogh et al., 2001). Red blocks refer to regions of amino acids that are within a transmembrane helix, the blue line represents the region of amino acids that are within the inside of the membrane which refers to the cytoplasm of the bacteria and the purple line represents the region of amino acids that are outside the membrane which refers to the periplasm of the bacteria.

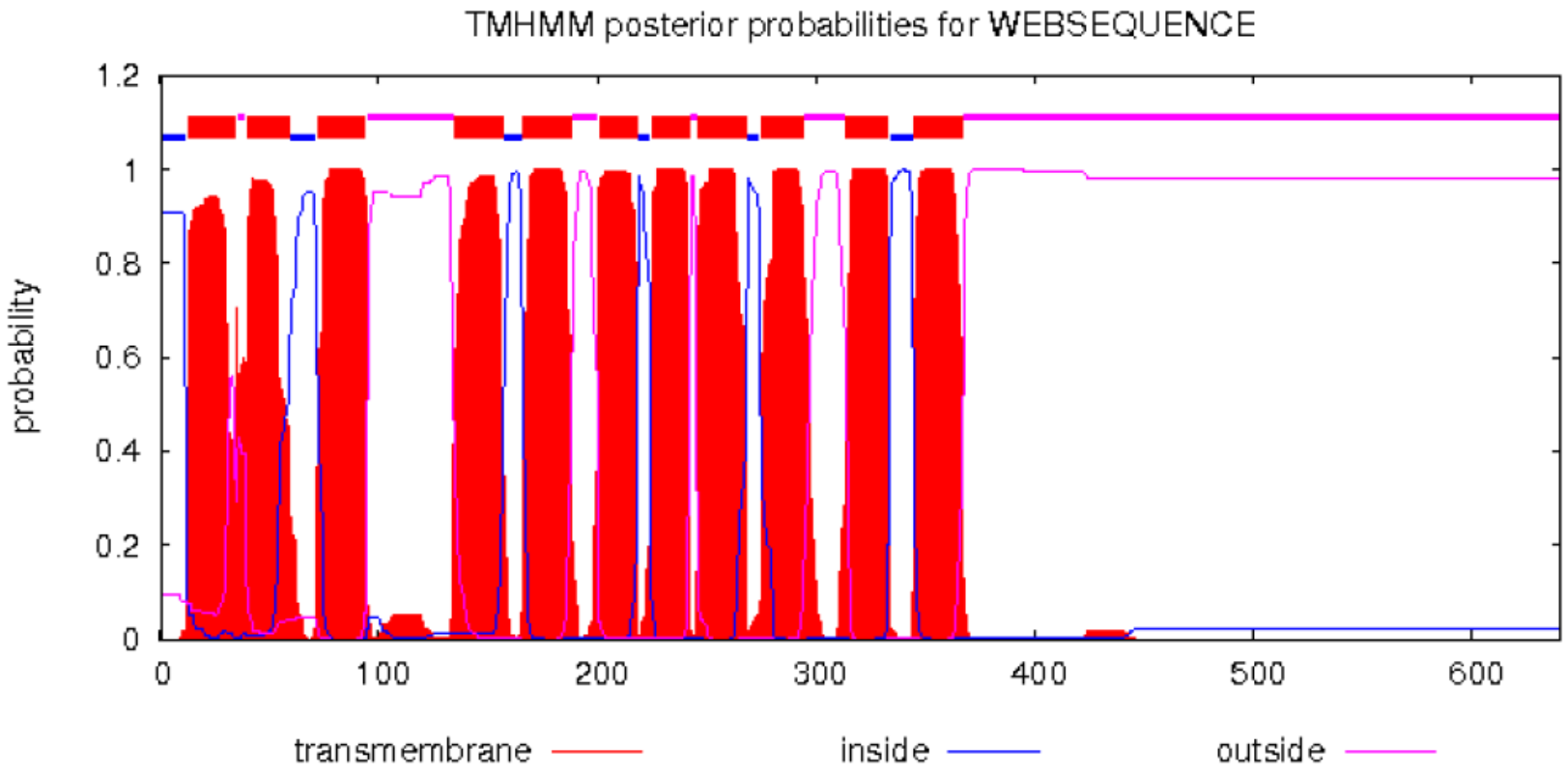


Figure 3.3.2. Transmembrane prediction of GtrC^{BTP1}. Amino acid sequence of GtrC^{BTP1} was put into the TMHMM v2.0 server (Sonnhammer et al., 1998, Krogh et al., 2001). 11 transmembrane helices, represented by the red blocks, were identified. The blue line represents the region of amino acids that are within the inside of the membrane which refers to the cytoplasm of the bacteria and the purple line represents the region of amino acids that are outside the membrane which refers to the periplasm of the bacteria.

LB agar plates. If the region of the protein PhoA has been fused into is in the periplasm, the colony grown on the XPhos supplemented LB agar plates will turn blue due to the successful cleavage of XPhos by PhoA and hence, deemed as PhoA positive. Both fusions were PhoA positive which suggests that they are within the periplasmic space.

The predicted SGNH-hydrolase type esterase domains are hence, to be within the periplasmic space. This could mean that this SGNH hydrolase domain might have a downstream function in the acetylation process of the LPS O-antigen sugar residues, after the predicted MBOAT like N-terminal acyltransferase 3 region. This characteristic is parallel to that of the single or multiple protein *O*-acetyltransferase systems of OatA in *S. aureus*, PatA/PatB in *N. gonorrhoeae* and AlgI/AlgJ and AlgX in *P. aeruginosa* mentioned in Section 3.2.3 and 3.2.4. To determine if the C-terminal regions might adopt a typical SGNH hydrolase type esterase α/β/α fold, a structural prediction of the proteins needed to be performed.

3.4 C-terminal tail structure predictions of OafA and GtrC^{BTP1}

The amino acid sequences of OafA and GtrC^{BTP1} were uploaded onto the SWISS-MODEL server (http://swissmodel.expasy.org/) (Arnold et al., 2006, Guex et al., 2009, Kiefer et al., 2009, Schwede et al., 2003) where models for both proteins are built using known protein structures from the Protein Data Bank (PDB) as the templates. The results were then visualised using the Swiss-PdbViewer 4.1.0 (Guex and Peitsch, 1997) and depicted in Figure 3.4.1. As an example of a known and characterised protein of the SGNH hydrolase type esterase family, the known crystal structure of rhamnogalacturonan acetylesterase (RGAE) of *Aspergillus aculeatus* (PDB: 1DEO) was used as a comparative structure and is illustrated on Figure 3.4.2. The predicted structure of the periplasmic C-terminal region of OafA, shown on Figure 3.4.1(a), was modelled against a GDSL-like lipase from *Parabacteroides distasonis* (PDB: 3P94) while that of GtrC^{BTP1}, shown on Figure 3.4.1(b) was predicted using a carbohydrate esterase from *Clostridium thermocellum* (PDB: 2VPT) as a template.

The predicted periplasmic C-terminal tail region of both OafA and GtrC^{BTP1} possesses the typical SGNH hydrolase α/β/α fold with the predicted catalytic residues being on the top of the α/β/α fold as shown on Figure 3.4.1. The known structure and catalytic sites of RGAE, depicted on Figure 3.4.2, is similar to that of the predicted structures of OafA and GtrC^{BTP1} and the catalytic sites are found in similar positions.

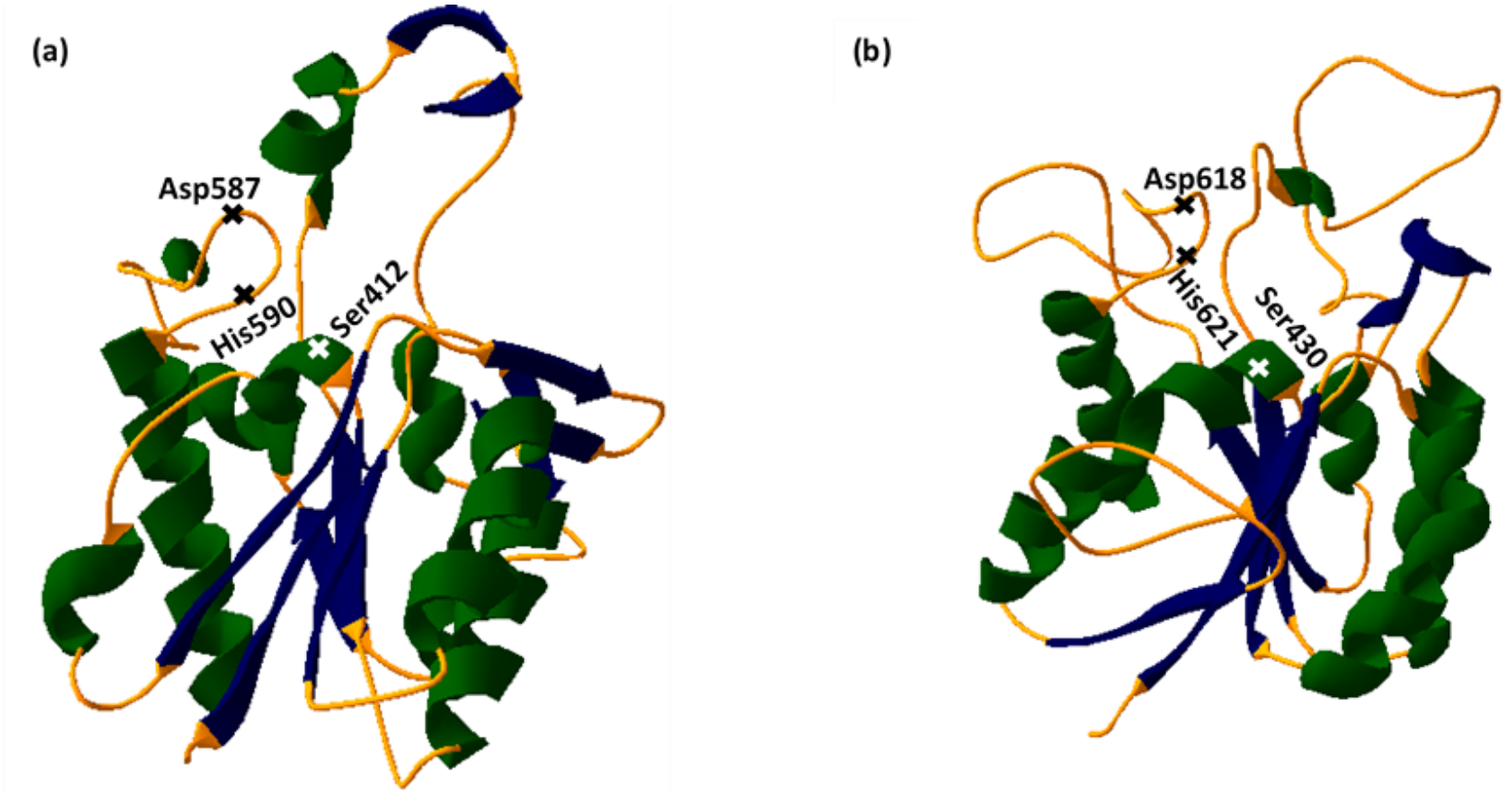


Figure 3.4.1. Predicted crystal structures of the C-terminal periplasmic tails of (a) OafA and (b) GtrC^{BTP1}. OafA was predicted using a GDSL-like lipase from *Parabacteroides distasonis* (PDB: 3P94) and GtrC^{BTP1} using a carbohydrate esterase from *Clostridium thermocellum* (PDB: 2VPT). α-helices are coloured in dark green, β-strands are coloured blue and coils are coloured orange. Amino acids of predicted catalytic triads are identified by crosses; black crosses are for amino acids that are found on coils and the white crosses are for amino acids that are found on the α-helices. Amino acid names and positions are also included with their respective crosses.

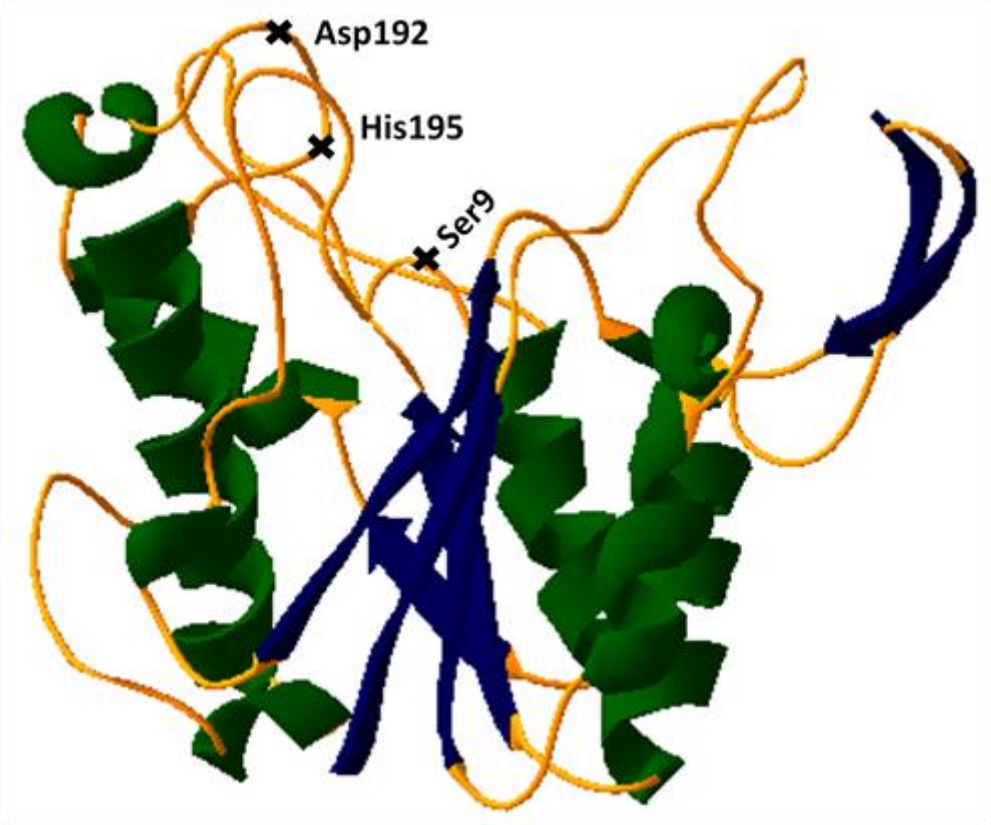


Figure 3.4.2. Crystal structure of rhamnogalacturonan acetylesterase (RGAE) of *Aspergillus aculeatus* (PDB: 1DEO). α-helices are coloured in dark green, β-strands are coloured blue and coils are coloured orange. Amino acids of the known catalytic triad are identified by black crosses. Amino acid names and positions are also included with their respective crosses.

Mølgaard et al. (2000) co-crystallised RGAE with the sulphate ion and deduced the position and identity of the residues making up the catalytic triad of the protein by identifying the presence of hydrogen bonding with the sulphate ion. PatB, the peptidoglycan *O*-acetyltransferase of *N. gonorrhoea*, possesses the same catalytic triad which when substituted individually with alanine, resulted in either in the complete inactivation of the enzyme or a significant decrease in protein activity (Moynihan and Clarke, 2014). Since the catalytic triad is also found on OafA and GtrC^{BTP1}, mutating these potential catalytic residues should abrogate or suppress the function of both proteins.

RGAE is an acetylesterase which functions as a deacetylase of the polysaccharide rhamnogalacturonan, as opposed to being an acetyltransferase which OafA and GtrC^{BTP1} are thought to be. The difference in function between the LPS acetyltransferases and RGAE suggest that the catalytic triad is an intermediate necessary for both acetyltransferase and acetylesterase functions with the site for the specific functions being elsewhere in the protein or the proteins act as both a transferase and esterase for the regulation of the amount of acetylation on the various. PatB was also found to have a higher specificity for the non-*O*-acetylated peptidoglycan molecules (Moynihan and Clarke, 2010). If the C-terminal domains of OafA and GtrC^{BTP1} are similar to PatB, there should be a binding site for the non-*O*-acetylated LPS molecule within the domain itself. Due to the highly specific nature of the O-antigen molecule of the LPS, no known O-antigen binding motifs were identified *in silico*. To determine where the binding site is, OafA and GtrC^{BTP1} can be co-crystallised with pure LPS O-antigen molecules.

3.5 Chapter 3 Conclusion

With the findings discussed in this chapter, OafA and GtrC^{BTP1} are predicted to be two domain proteins with the N-terminal domain being of the acyltransferase 3 domain and the periplasmic C-terminal tail being the SGNH hydrolase type esterase domain. Further analyses of the C-terminal periplasmic tail identified common SGNH hydrolase characteristics in the proteins of interest, such as the position of the catalytic triad of Ser-Asp-HIs and their predicted 3D structures.

To further support these hypotheses, extensive experimental protein analyses should be performed including, but not restricted to, structural studies, *in vitro* and *in vivo* functional analyses.

Chapter 4

Expression and Purification of OafA and GtrC^{BTP1} from *S.* Typhimurium

Chapter 4 Introduction

Bioinformatics analysis and further support from findings by Kintz et al. (Kintz et al., 2015) for GtrC^{BTP1} determined that OafA and GtrC^{BTP1} are likely to be highly similar with both proteins predicted to possess a membrane bound N-terminal domain and a SGNH hydrolase type esterase C-terminal periplasmic domain. Preliminary *in silico* comparisons with known SGNH hydrolase type esterase proteins suggests that these proteins possess the characteristics of this family of proteins in their respective periplasmic C-terminal domain.

To determine if the SGNH hydrolase C-terminal domain of both proteins can fold independently of their respective N-terminal inner-membrane bound domains, the C-terminal SGNH hydrolase type esterase region of the proteins were attempted to be expressed, purified and characterised alongside the full length proteins.

An excellent method for the characterisation of proteins is the determination of their various structures. This should provide confirmation on the different forms the protein can adopt and the sites of the ligand interactions. Structural determination of OafA and GtrC^{BTP1} is therefore one of the main objectives of this project. Structural determination methods like NMR and X-ray crystallography require pure samples of proteins. Alongside the biophysical structural determination experiments, biochemical experiments could also be performed to characterise these proteins *in vitro* to determine which substrates are able to interact with these proteins and whether or not the proteins are monomeric, among others. To obtain this, these proteins were expressed in an expression strain and purified using various methods which are discussed in the following sections.

4.1 Expression of OafA and GtrC^{BTP1} from the pBADcLIC Vector

Expression of both OafA and GtrC^{BTP1} were performed by inserting the isolated genes into the pBADcLIC vector possessing the arabinose-inducible *araBAD* promoter, P_{BAD}, upstream of the inserted *oafA* and *gtrC* ^{BTP} genes. Hence, expression of these proteins can be controlled using different concentrations of the inducer, L-arabinose.

The sequence encoding for the full length and the C-terminal putative SGNH hydrolase domains OafA and GtrC^{BTP1} were isolated via PCR from the genome of *S.* Typhimurium LT2 and *S.* Typhimurium D23580 respectively. All four sequences were inserted into the vector separately via ligation independent cloning (LIC) and then transformed into a standard *E. coli* cloning strain. The inserts were confirmed by sequencing (Section 2.4.7.3).

From here henceforth, the pBADcLIC vector containing the full length *oafA* gene will be called pOafA and the vector encoding the full length GtrC^{BTP1} protein will be called pGtrC^{BTP1}. The vector containing the sequence encoding the C-terminal tail region of OafA will be called pOafA(C-tail) and GtrC ^{BTP1} will be called pGtrC^{BTP1}(C-tail). When expressed from the vector, the proteins will possess a C-terminal deca-His tag which will allow the expressed proteins to be detected by anti-poly His antibodies and isolated from the expression strain by immobilised metal affinity chromatography (IMAC).

Trial protein expression were performed using *E. coli* DH5α for pOafA, pOafA(C-tail) and pGtrC^{BTP1}(C-tail) while pGtrC^{BTP1} was expressed from *E. coli* XL-1 Blue. A final concentration of 1mM L-arabinose was used with induction taking place for four hours after the cells reached an OD_{600} of 0.5 at 30°C. With the deca-His tag introduced, the full length OafA is 72.4 kDa, full length GtrC^{BTP1} is 74.1 kDa, the C-terminal domain of OafA is 32.9 kDa and the C-terminal domain of GtrC^{BTP1} is 32.3 kDa.

4.1.1 Trial expression of the full length proteins of OafA and GtrC^{BTP1}

Owing to the predicted presence of the transmembrane helices, the full length proteins OafA and GtrC^{BTP1} are expected to be in the insoluble fractions of the cell lysates. The western blot analysis from Figure 4.1.1(a) and 4.1.2(a) suggests that, with the trial expression conditions mentioned above, the full length OafA and GtrC^{BTP1} might have been successfully expressed as the anti-His antibody binding was detected via secondary antibody binding and chemiluminescent visualisation. In addition, the signal potentially representing these proteins were detected in the insoluble fraction of the cell lysates. A less intense signal was observed in the total protein fraction of only OafA and not GtrC^{BTP1}. This could be because more OafA was produced than GtrC^{BTP1}. The protein samples, however, were not in the same position after protein separation as the corresponding protein molecular weight marker bands of approximately 72 kDa and 74 kDa for OafA and GtrC^{BTP1} respectively.

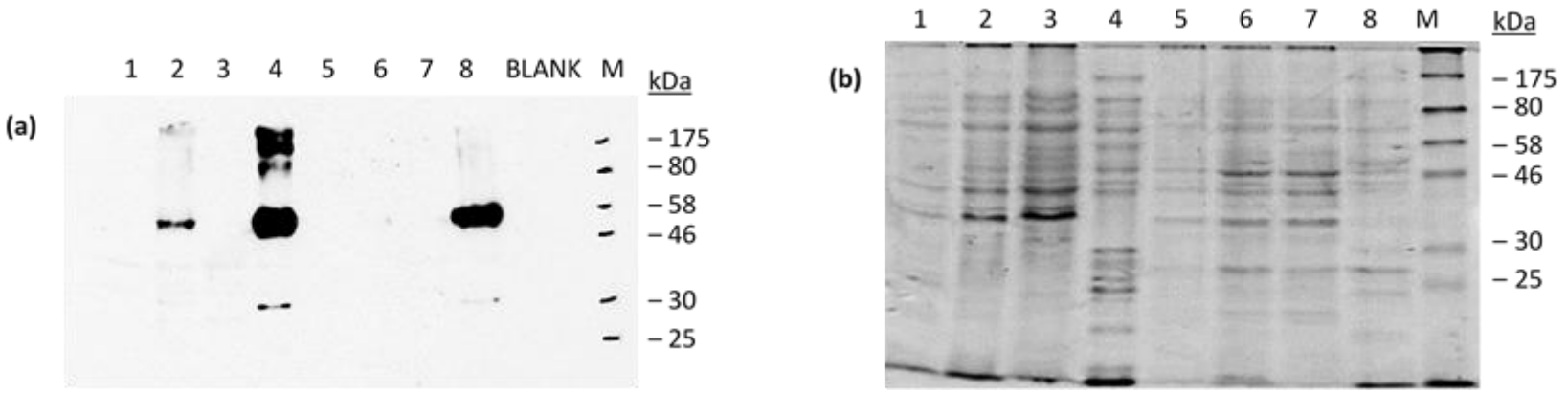


Figure 4.1.1. Expression Trials of the Full Length OafA and GtrC^{BTP1}. (a) Western analysis with the anti- tetra His antibody, and **(b)** coomassie stained 12% SDS-PAGE gels. Lane 1 = Total proteins from lysis of the uninduced DH5α with pOafA, Lanes 2 = Total proteins from the lysis of the induced cultures of DH5α with pOafA, Lane 3 = soluble fraction from the lysis of the induced cultures of DH5α with pOafA, Lane 4 = insoluble fractions isolated from the lysis of the induced cultures of DH5α with pOafA, Lane 5 = Total proteins from lysis of the non-induced XL-1 Blue with pGtrC^{BTP1}, Lanes 6 = Total proteins from the lysis of the induced cultures of XL-1 Blue with pGtrC^{BTP1}, Lane 7 = soluble fraction from the lysis of the induced cultures of XL-1 Blue with pGtrC^{BTP1} and Lane 8 = insoluble fractions isolated from the lysis of the induced cultures of XL-1 Blue with pGtrC^{BTP1}.

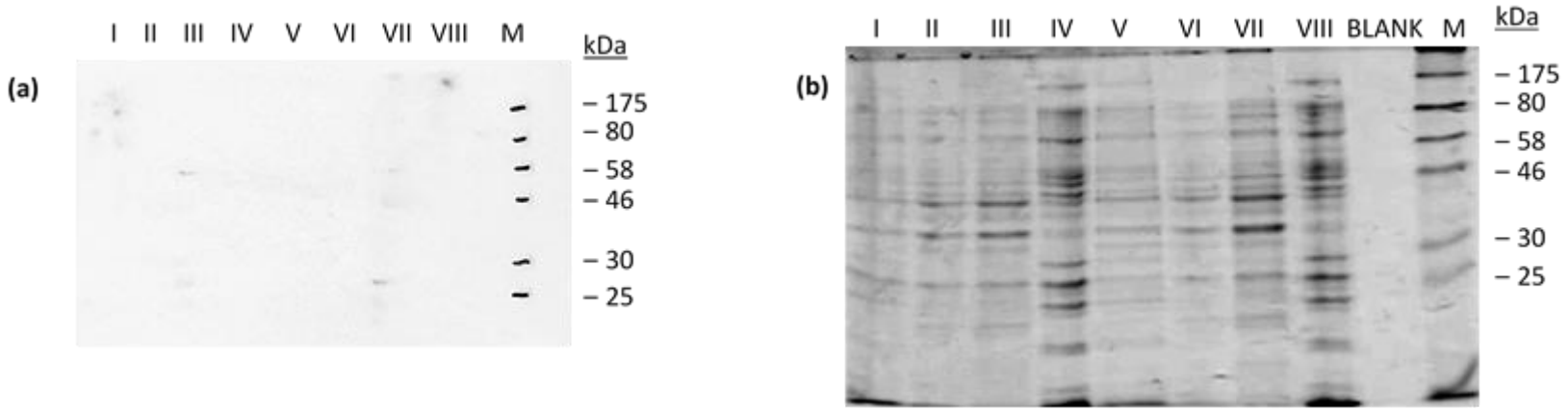


Figure 4.1.2. Expression Trials of the C-terminal Tail of OafA and GtrC^{BTP1}. (a) Western analysis with the anti- tetra His antibody, and **(b)** coomassie stained 12% SDS-PAGE gels. Lane I = Total proteins from lysis of the uninduced DH5α with pOafA(C-tail), Lanes II = Total proteins from the lysis of the induced cultures of DH5α with pOafA(C-tail) respectively, Lane III = soluble fraction from the lysis of the induced cultures of DH5α with pOafA(C-tail) respectively, Lane IV = insoluble fractions isolated from the lysis of the induced cultures of DH5α with pOafA(C-tail), Lane V = Total proteins from lysis of the non-induced DH5α with pGtrC^{BTP1}(C-tail), Lanes VI = Total proteins from the lysis of the induced cultures of DH5α with pGtrC^{BTP1}(C-tail), Lane VII = soluble fraction from the lysis of the induced cultures of DH5α with pGtrC^{BTP1}(C-tail) and Lane VIII = insoluble fractions isolated from the lysis of the induced cultures of DH5α with pGtrC^{BTP1}(C-tail).

This could be attributed to the interactions of these proteins with sodium dodecyl sulphate (SDS), an anionic detergent, in the PAGE gel used for separation of the proteins. Depending on the amount of SDS, the nature and number of hairpins formed by the transmembrane helices and the state of folding of the proteins (i.e. whether the proteins are fully denatured or not) (Rath et al., 2009). With the current SDS-PAGE protocol adopted, both OafA and GtrC^{BTP1} run at just under the level where 58 kDa band of the protein molecular weight marker lies. The primary antibody used is a poly-histidine specific anti-tetra His antibody which suggests that any signals observed on the western analyses will be a protein possessing at least 4 consecutive histidines, which should include the expressed proteins of interest. Both the total proteins of the uninduced and soluble fraction of the induced *E. coli* clones cell lysates expressing either OafA or GtrC^{BTP1} did not produce any signals after staining with the primary antibody. This provides further support that any signals observed on the western analysis would most likely be the proteins of interest. The smear above the ~58 kDa signal with respect to the protein marker seen in the western analysis of the insoluble fraction sample of the cell lysate of the strain expressing OafA could be caused by the aggregation of the expressed OafA protein. Prior to the SDS-PAGE protein separation, the insoluble fractions were not boiled due to the likelihood that the proteins will aggregate in the presence of heat and SDS (Andrew Bretnall, personal communication). Instead, they were heated at 65°C for approximately 10 minutes. Hence, future preparation of insoluble fractions prior to SDS-PAGE gel separation were performed at 55°C for approximately 20 minutes in an attempt to ensure that the proteins aggregate less.

A separate signal was observed at approximately 30 kDa in the insoluble fraction sample of the cell lysate of both expression strains with a stronger signal observed for the strain expressing OafA. Two reasons were hypothesised to have caused the presence of this extra signal. Firstly, GtrC^{BTP1} was confirmed to have a transmembrane region and a periplasmic tail region (Kintz et al., 2015) with the same predicted for OafA. When expressed, these proteins could have had some form of cleavage that resulted in the separation of these two regions, producing two ~30 kDa fragments. If this were to have happened, the deca-His tag would be in the periplasmic soluble ~30 kDa fragment. Since the signal at 30 kDa was only found in the insoluble fractions of the cell lysates, this would be an unlikely cause unless the periplasmic soluble fragment were to aggregate and form an inclusion body. A second and more likely reason why this 30 kDa signal is present could be due to a contaminating native transmembrane protein found in the *E. coli* expression strains. One such contaminating protein would be a zinc transporter protein, ZitB, which possess a molecular weight of 34.6 kDa. The presence of a native tetra histidine c-terminal end of ZitB could be a problem in western analysis using the anti-tetra His primary antibody if the expressed proteins were to be in the 30 kDa region and it could also be a major problem during protein purification.

To confirm if the proteins are indeed OafA and GtrC^{BTP1}, protein identification by peptide mass fingerprinting would have to be done after a large scale expression of these proteins, as a larger amount of proteins would be required for the successful detection of the peptide masses after protein digestion. Prior to protein identification, all future western blot analysis of the full length OafA and GtrC^{BTP1} proteins will based on the assumption that the proteins are running at just under 58 kDa.

4.1.2 Trial expression of the C-terminal tails of OafA and GtrC^{BTP1}

Due to the absence of the putative transmembrane helices, these proteins are expected to be in the soluble fraction of the various cell lysates. The putative periplasmic C-terminal SGNH hydrolase domain of OafA and GtrC^{BTP1} were not successfully expressed with only a very faint signal observed at the 30 kDa protein marker level for the insoluble fraction of the cell lysate of the strain expressing the C-terminal tail of GtrC^{BTP1}. However, the signal could also be a contaminating protein like ZitB. Due to the need for an expression protocol that would work for their expression and the potential confusion with contaminating proteins, the C-terminal tail proteins will not be purified and the following experiments will only be on the full length proteins of OafA and GtrC^{BTP1}.

4.2 Expression optimisation of OafA and GtrC^{BTP1}

4.2.1 Parameters for optimisation

Protein structure determination methods such as X-ray crystallography require pure proteins in the milligram (mg) scale. Photosynthetic proteins were some of the first crystallised membrane proteins due to their abundance in nature, unlike most other membrane proteins (Grisshammer and Tateu, 1995, Ostermeier and Michel, 1997). Under the same expression conditions, more soluble proteins are expressed multiple folds higher than insoluble proteins from an expression vector in an expression host. Naturally, a larger culture volume of the *E. coli* strains expressing the proteins would be prepared prior to protein isolation as compared to that of the host strains expressing soluble proteins. However, there is also a need to optimise the expression of these proteins in order to get as much proteins as possible from a certain volume of culture. Multiple parameters could be varied in order to find the optimal expression strategy.

4.2.1.1 E. coli expression strains

Firstly, a different expression strain could perform better in expressing the proteins of interest. The inducer used is L-arabinose, a sugar that is able to be metabolised by *E. coli* DH5α and XL-1 Blue strains. *E. coli* MC1061 is an *E. coli* K-12 based strain that notably possesses an inactive arabinose operon with the Δ(ara-leu)7697 deletion mutation, which means that the *araB*, *araA* and *araC* were removed, and *araD* possesses a mutation at amino acid 139 (Casadaban and

Cohen, 1980). With these mutations, the *E. coli* MC1061 strain is unable to metabolise L-arabinose, thus allowing for a constant amount of L-arabinose present for induction.

4.2.1.2 Induction temperature

The next parameter of protein expression that could be modified is the temperature at which the induction of the protein expression happens. Cells exposed to mild hypothermia have slower cell processes, which allows the coupling of transcription of the protein of interest with the translation and then, translation with post-translational processes (Chou, 2007). If these processes are not coupled, cell viability may be compromised. If transcription outruns translation, accumulation of mRNA would occur which may cause the destruction of ribosomes (Dong et al., 1995). In addition, the lower temperature would suppress the activity of most bacterial proteases, allowing for the reduced degradation of the potentially vulnerable protein of interest and hence, increasing the amount of protein for purification (Hunke and Betton, 2003, Pinsach et al., 2008). The expression optimisation of OafA and GtrC^{BTP1} included two sets of cultures grown to 37°C and underwent separate temperature shifts to 20°C and 30°C for induction.

4.2.1.3 Concentration of inducer

As mentioned, the coupling of transcription and translation is vital for the viability of the cell and hence, optimal protein production. Another factor that can affect the coupling of these processes is the concentration of the inducer, L-arabinose, added into the culture. The pBADcLIC vector used to harbour *oafA* and *gtrC^{BTP1}* possesses the *araBAD* promoter system from the arabinose operon which is titratable (Guzman et al., 1995). A low concentration would result in the insufficient production of the protein of interest while an excessive amount of the inducer might uncouple transcription and translation. Hence, an optimal concentration of inducer would be beneficial in the optimisation of expression of OafA and GtrC^{BTP1}. Three concentrations of L-arabinose at 0.001%, 0.01% and 1% (w/v) were tested in the following expression optimisation experiments.

4.2.1.4 Duration of incubation

A decreased induction temperature and a small concentration of inducer would result in a slower rate of protein production. Hence, a longer duration of induction might be necessary for the optimal expression of these protein. In the following optimisation experiments, one set of cultures was induced for 4 hours and another set of cultures was induced overnight (approximately 19 hours).

4.2.2 Protein expression optimisation

OafA and GtrC^{BTP1} were expressed in either the initial cloning strains mentioned in Section 4.1, *E. coli* DH5α and XL-1 Blue, or *E. coli* MC1061 using the inducer concentrations of either 0.001%, 0.01% or 1% (w/v) L-arabinose at either 30°C or 37°C for 4 hours or overnight (approximately 19 hours). Every permutation of the above mentioned conditions were tested and the protein expressions were assessed via western analysis shown on Figures 4.2.1 to 4.2.4 along with their respective coomassie stained gels.

4.2.2.1 Expression optimisation of OafA

The optimisation of the expression of OafA was performed and the protein expression levels were analysed via western analysis shown on Figures 4.2.1 and 4.2.2 with the former representing all samples from *E. coli* DH5α based protein expressions and the latter representing all samples from *E. coli* MC1061 based protein expressions. The western analyses shown on Figure 4.2.1 suggests that 0.0001% (w/v) L-arabinose is insufficient for the expression of OafA while 1% and 0.01% (w/v) L-arabinose resulted in the production of comparatively similar amounts of proteins from *E. coli* DH5α induced at the same temperature. The signal from the western analyses of proteins from the *E. coli* DH5α induced at 30°C seemed to be fractionally more intense than that of the strains induced at 20°C. This difference could be attributed to more proteins being loaded into the SDS-PAGE gel for the insoluble fraction protein separation of the former. The coomassie stained gels of the insoluble fraction protein separation of the *E. coli* DH5α induced at 30°C shown on Figure 4.2.1(b) had more intense staining for the protein samples from the cultures induced with 0.01% and 1% (w/v) L-arabinose as compared to the corresponding protein samples on Figure 4.2.1(d) for the *E. coli* DH5α induced at 20°C. From Figure 4.2.1, the cultures induced for 4 hours with either 1% or 0.01% (w/v) L-arabinose at both 20°C and 30°C seem to have produced a higher signal on the western analyses, suggesting that these cultures are capable of producing the largest amount of proteins within the *E. coli* DH5α based expression cultures.

Figure 4.2.2(c) represents the western analyses performed for the comparison of the relative expression levels of OafA from the cultures of *E. coli* MC1061. Unlike the expression of OafA from *E. coli* DH5α, the inability to metabolise arabinose allows protein expression from *E. coli* MC1061 to occur even when induced with 0.0001% (w/v) L-arabinose. Expression of OafA across all cultures of *E. coli* MC1061 induced at the same temperatures seem to be relatively similar as the signal intensities do not very significantly. For *E. coli* MC1061 induced at 20°C, the signal intensity for the protein samples of the cultures induced overnight is less than that of the 4 hour induced samples. Like that of the *E. coli* DH5α cultures, this could be attributed to the fact that less

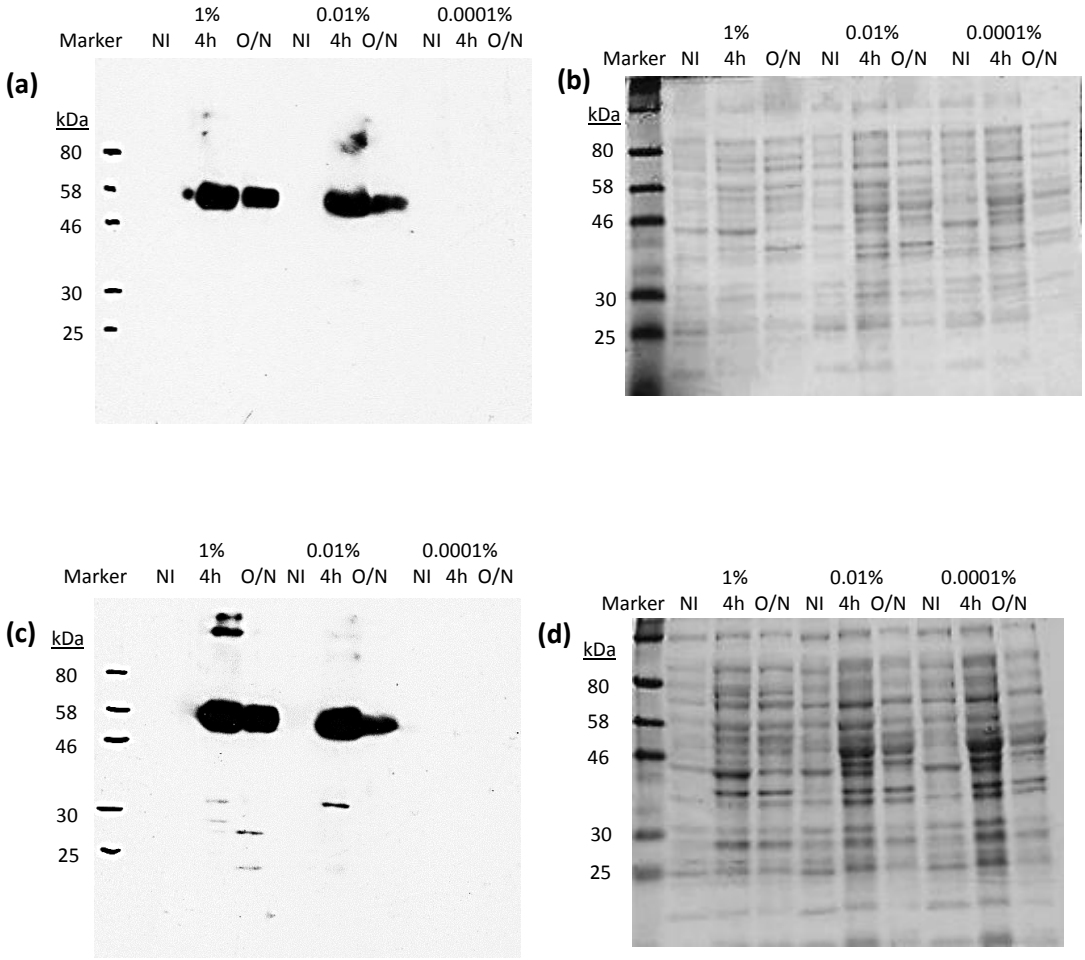


Figure 4.2.1. Western analyses of the expression of OafA from *E. coli* DH5α. Proteins expressed at **(a)** 20°C and **(c)** 30°C with their respective coomassie stained gels **(b)** and **(d)**. 1%, 0.01% and 0.0001% represents the concentration in w/v of the inducer, L-arabinose. NI = no inducer added, 4h = cultures were induced for 4 hours, ON = cultures were induced overnight. Samples were the insoluble fractions of the various cell lysates. SDS-PAGE was performed with a 12% SDS-PAGE gel, under denaturing and reducing conditions.

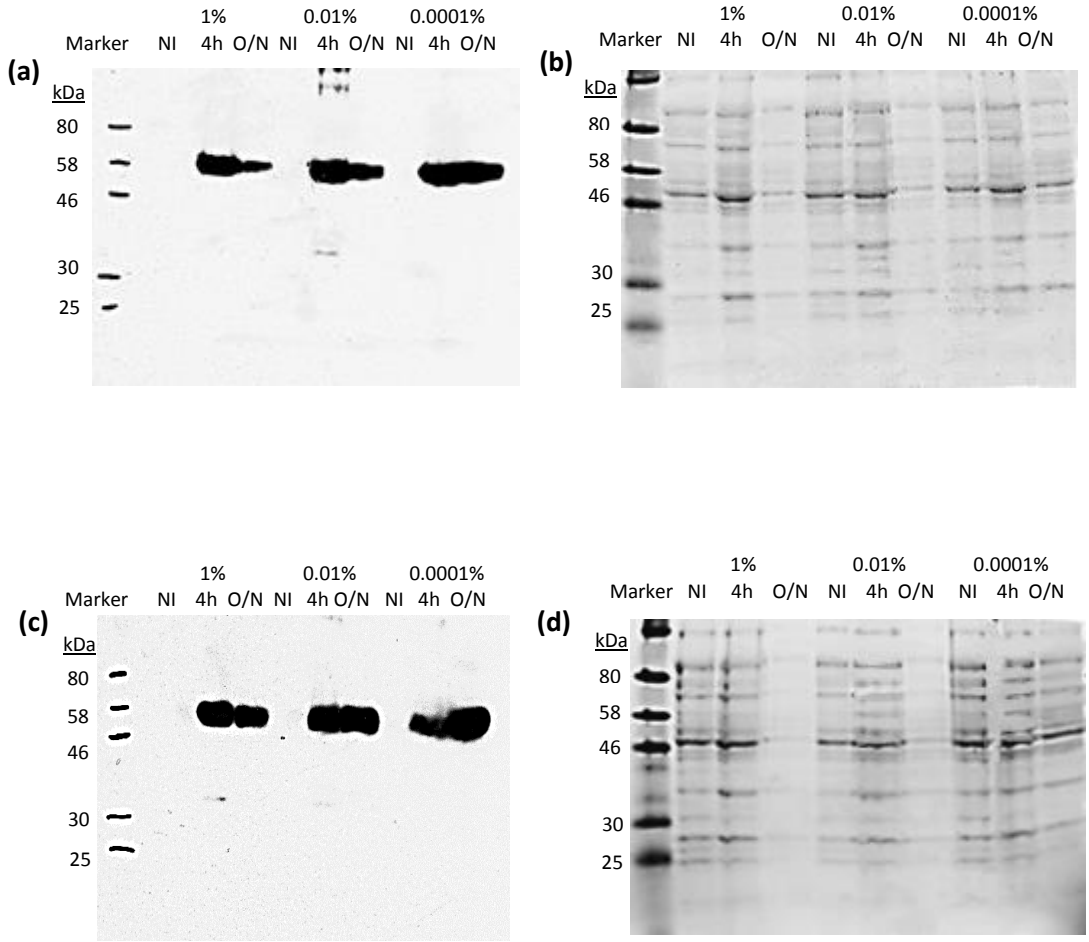


Figure 4.2.2. Western analyses of the expression of OafA from *E. coli* MC1061. Proteins expressed at **(a)** 20°C and **(c)** 30°C with their respective coomassie stained gels **(b)** and **(d)**. 1%, 0.01% and 0.0001% represents the concentration in w/v of the inducer, L-arabinose. NI = no inducer added, 4h = cultures were induced for 4 hours, ON = cultures were induced overnight. Samples were the insoluble fractions of the various cell lysates. SDS-PAGE was performed with a 12% SDS-PAGE gel, under denaturing and reducing conditions.

proteins from the overnight induced cultures were added into the SDS-PAGE gel for protein separation as shown in the photo of the coomassie stained gel on Figure 4.2.2(b). The same lack of intensity of the coomassie stained protein sample from the overnight induced culture as compared to the 4 hour induced culture of *E. coli* MC1061 induced at 30°C is observed on Figure 4.2.2(d). However, the western analyses shown on Figure 4.2.2(c) suggests that there was no reduction in signal intensities between the overnight and 4 hour induced cultures despite the stark differences in coomassie staining. Hence, the overnight cultures of *E. coli* MC1061 induced with either the 0.0001%, 0.01% or 1% (w/v) L-arabinose at 30°C seemed to have produced the largest amount of OafA among the *E. coli* MC1061 based expression cultures.

Comparing the western analyses and the respective coomassie stained gels containing proteins samples from the *E. coli* DH5α cultures induced for 4 hours with either 1% or 0.01% (w/v) L-arabinose at both 20°C and 30°C and the overnight cultures of *E. coli* MC1061 induced with either the 0.0001%, 0.01% or 1% (w/v) L-arabinose at 30°C, the latter was decided to be the optimum conditions with the various conditions tested. Using the results of the optimisation experiments, future expression of OafA was decided to be performed using the *E. coli* MC1061 expression host, with 0.01% (w/v) L-arabinose at 30°C overnight.

4.2.2.2 Expression optimisation of GtrC^{BTP1}

Expression optimisation of GtrC^{BTP1} was performed using the conditions mentioned in Section 4.2.1 and the levels of protein expression were analysed using western analyses which are depicted with the respective coomassie stained gels on Figure 4.2.3.

The results of the expression optimisation of GtrC^{BTP1} are similar to that of the expression optimisation of OafA. For protein expression cultures of *E. coli* XL-1 Blue strains, no or negligible expression of GtrC^{BTP1} was observed on the western analyses of cultures induced with 0.0001% (w/v) L-arabinose. Like *E. coli* DH5α, *E. coli* XL-1 Blue strains are able to metabolise arabinose and therefore, the 0.0001% (w/v) L-arabinose added to the cultures for induction was taken up by the host and metabolised before it was able to induce significant expression of GtrC^{BTP1}.

E. coli XL-1 Blue induced with 0.01% (w/v) L-arabinose overnight at both induction temperatures were observed to have produced much less GtrC^{BTP1} than the 4 hour induced culture. The coomassie stained gels for both induction temperatures shown on Figure 4.2.3(b) and (d) depicted a stronger stained band for the lanes representing the proteins from these samples and also that of the cultures of *E. coli* XL-1 Blue induced with 0.0001% (w/v) L-arabinose as compared to that of the cultures induced with 1% (w/v) L-arabinose. Due to its intensity, these bands might represent heat-shock proteins (Andrew Bretnall, personal communication). The peculiar presence of these bands on the select lanes on the coomassie stained gels, suggests that the proteins making up

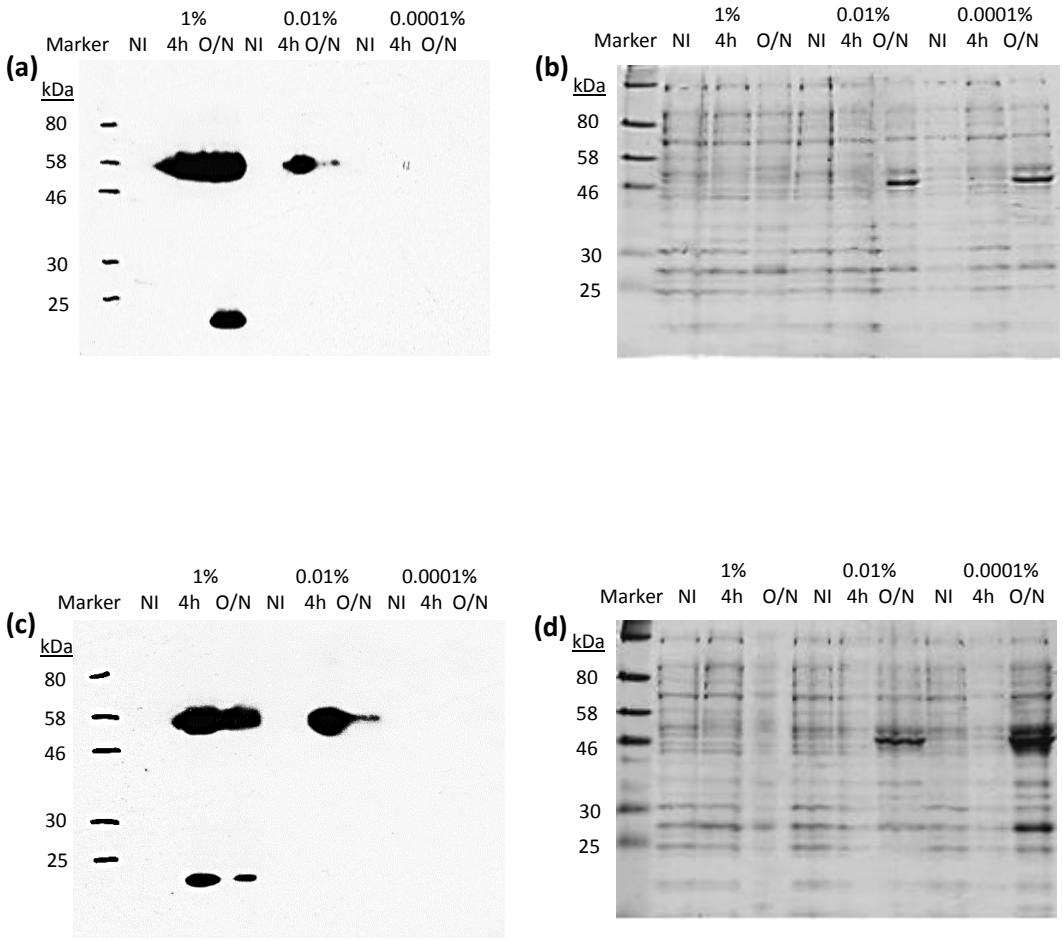


Figure 4.2.3. Western analyses of the expression of GtrC^{BTP1} from *E. coli* XL1-Blue. Proteins expressed at **(a)** 20°C and **(c)** 30°C with their respective coomassie stained gels **(b)** and **(d)**. 1%, 0.01% and 0.0001% represents the concentration in w/v of the inducer, L-arabinose. NI = no inducer added, 4h = cultures were induced for 4 hours, ON = cultures were induced overnight. Samples were the insoluble fractions of the various cell lysates. Samples were the insoluble fractions of the various cell lysates. SDS-PAGE was performed with a 12% SDS-PAGE gel, under denaturing and reducing conditions.

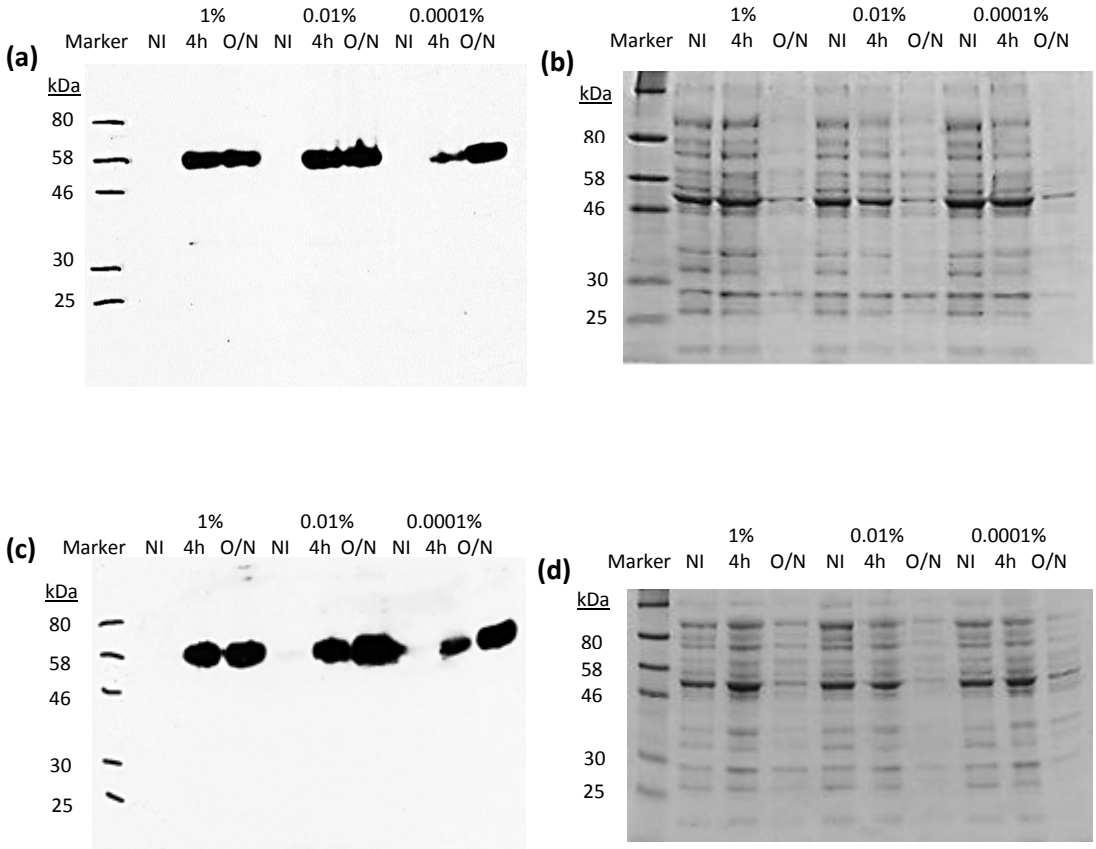


Figure 4.2.4. Western analyses of the expression of GtrC^{BTP1} from *E. coli* MC1061. Proteins expressed at **(a)** 20°C and **(c)** 30°C with their respective coomassie stained gels **(b)** and **(d)**. 1%, 0.01% and 0.0001% represents the concentration in w/v of the inducer, L-arabinose. NI = no inducer added, 4h = cultures were induced for 4 hours, ON = cultures were induced overnight. Samples were the insoluble fractions of the various cell lysates. Samples were the insoluble fractions of the various cell lysates. SDS-PAGE was performed with a 12% SDS-PAGE gel, under denaturing and reducing conditions.

the intensity of these bands are only present overnight and in the absence of a concentration of arabinose around or higher than 1% (w/v). The cultures of *E. coli* XL-1 Blue induced with 1% (w/v) L-arabinose at both tested temperatures were observed to have produced the highest yield of $GtrC^{BTP1}$ as shown by the western analyses of the protein expression of these cultures on Figure 4.2.3 (a) and (c). An extra single signal is observed on each of the lanes containing the following; *E. coli* XL-1 Blue induced with 1% (w/v) L-arabinose at 20°C overnight, *E. coli* XL-1 Blue induced with 1% (w/v) L-arabinose at 30°C for 4 hours and overnight. These single signals at less than 25 kDa with respect to the marker suggests that some form of degradation might have happened to the full length $GtrC^{BTP1}$ proteins. Nonetheless, among the cultures of *E. coli* XL-1 Blue used for the expression of $GtrC^{BTP1}$, the cultures induced with 1% (w/v) L-arabinose at either 20°C or 30°C for either 4 hours or overnight seemed to have produced the strongest signals on the western analyses which most likely translates to the highest output of $GtrC^{BTP1}$ with the various parameters tested.

Expression of $GtrC^{BTP1}$ performed using the host *E. coli* MC1061 was successful across all concentrations of L-arabinose as shown on Figure 4.2.4. The signal intensities on the western analyses that are thought to represent $GtrC^{BTP1}$ were consistently intense throughout the conditions tested with the exception of the cultures induced with 0.0001% (w/v) L-arabinose for 4 hours and overnight. The cultures induced for 4 hours had a weaker signal as compared to that of the overnight induced cultures. This is most likely due to the fact that less amount time is allowed for induction to occur and hence, less proteins were produced.

The protein samples loaded onto the SDS-PAGE gel for protein separation of the samples from the overnight induced cultures were less than that of the 4 hour induced cultures and the no induction negative control as seen in the lighter coomassie stained bands on the lanes representing these samples shown on Figure 4.2.4(b) and (d). However, the intensities of the signals on the western analyses were similar for the overnight induced cultures rather than the 4 hour cultures which suggests that the overnight induced cultures were expressing more $GtrC^{BTP1}$ than the 4 hour induced cultures. The cultures induced at 30°C also had more intense signals than that of the 20°C with the culture induced with 0.01% (w/v) L-arabinose overnight being the most intense. Hence, the *E. coli* MC1061 strain expressing $GtrC^{BTP1}$ induced with 0.01% (w/v) L-arabinose overnight would be chosen for future expression of $GtrC^{BTP1}$.

4.3 Large Scale Production and Purification of OafA and $GtrC^{BTP1}$

Using the chosen expression conditions from Section 4.2.2, OafA and $GtrC^{BTP1}$ were expressed from a much larger culture volume of 2.5 L each in an attempt to purify the milligrams of proteins required for future structural determination studies.

4.3.1 Solubilisation of the inner membrane bound OafA and GtrC^{BTP1}

For soluble proteins, isolation by immobilised metal affinity chromatography (IMAC) with sufficient washing would remove all other unwanted proteins, giving a very pure sample of correctly folded proteins that is necessary for the accuracy of future biochemical and biophysical studies. However, this is not sufficient for membrane proteins such as OafA and GtrC^{BTP1} due to the fact that they are bound onto the membrane. Hence, the liberation of these membrane proteins from the membrane is necessary before the purification by IMAC to ensure the purity of these proteins.

Detergent based solubilisation is a method of liberating these membrane proteins from the lipids that make up the membrane and is shown on Figure 4.3.1. Every detergent molecule possesses a hydrophilic head and a hydrophobic tail, allowing it to mimic the lipid molecules that make up membranes. At a high enough detergent concentration, these detergent molecules group together to form detergent micelles. The minimum concentration of detergent required for these micelles to form is the critical micellar concentration (cmc). Micelles are particularly important for the purification of membrane proteins as detergents in the micellar form are required for the solubilisation of these proteins while keeping the folding of these proteins. Hence, a concentration of detergents above the cmc is required for the successful liberation of membrane proteins (Seddon et al., 2004). Different detergent molecules have different cmc values. Environmental factors such as pH levels, temperature and the presence of other molecules could also affect the cmc value of a specific detergent (le Maire et al., 2000).

Detergents can be classified as ionic, non-ionic, zwitterionic and bile acid salts. Depending on the nature of the membrane proteins, different detergents are more effective in liberating different membrane proteins. Hence, a detergent solubility test should be performed using different detergents to determine which are the most appropriate for solubilisation of these proteins.

OafA and GtrC^{BTP1} were initially solubilised with 0.5% (w/v) n-dodecyl β-D-maltoside (DDM), a mild, non-ionic detergent with a cmc of 0.18mM (0.0009% w/v) in water (Privé, 2007) and a micellar molecular weight of approximately 60 kDa (VanAken et al., 1985).

Figure 4.3.2 depicts the western analyses (Figure 4.3.2(a) and (c)) and corresponding coomassie stained gels (Figure 4.3.2(b) and (d)) of the DDM based solubilisation of OafA and GtrC^{BTP1} between lanes 1 to 5 and the purification of these proteins by IMAC. The latter will be discussed in Section 4.3.2.

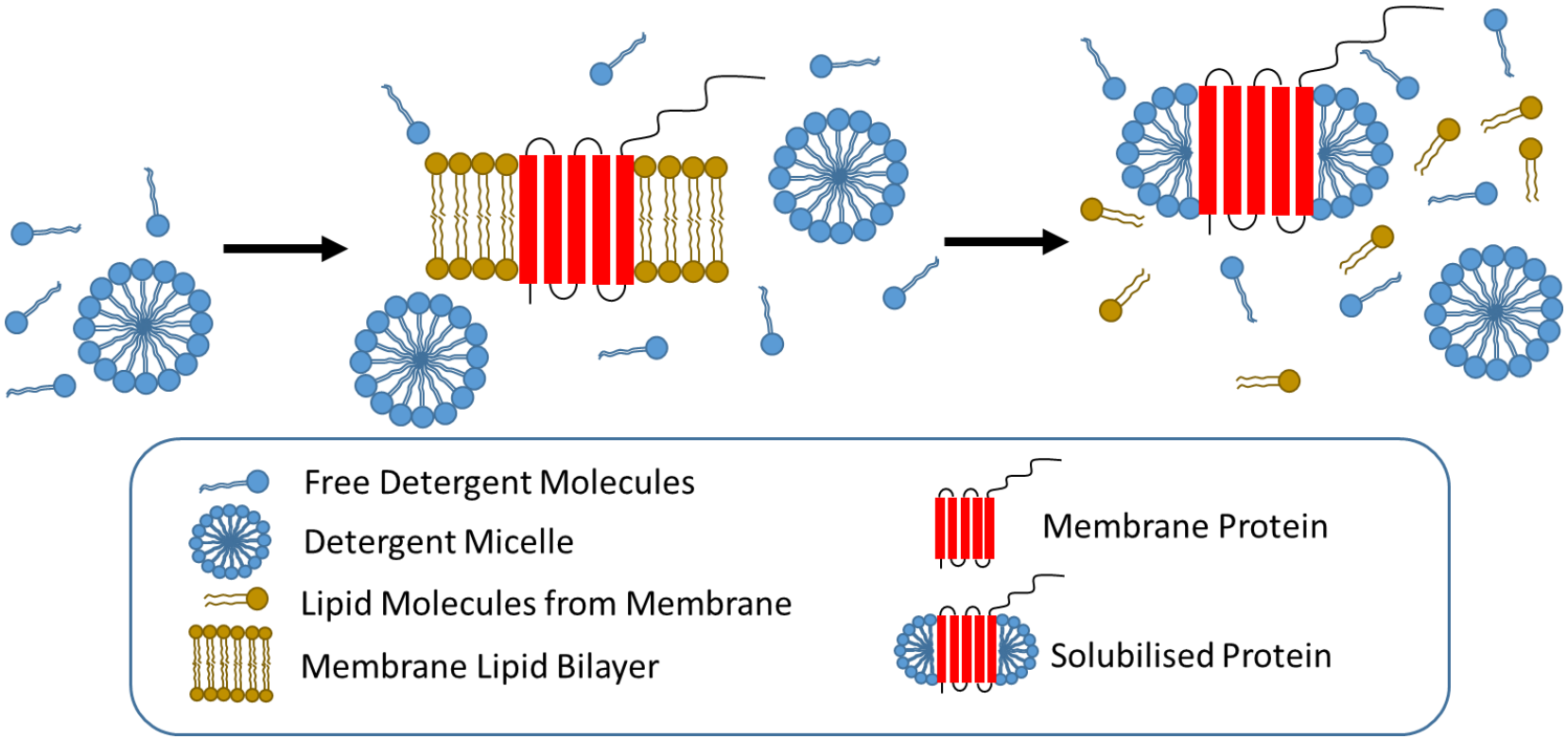


Figure 4.3.1. Detergent based Solubilisation of Membrane Proteins. Refer to text for explanation.

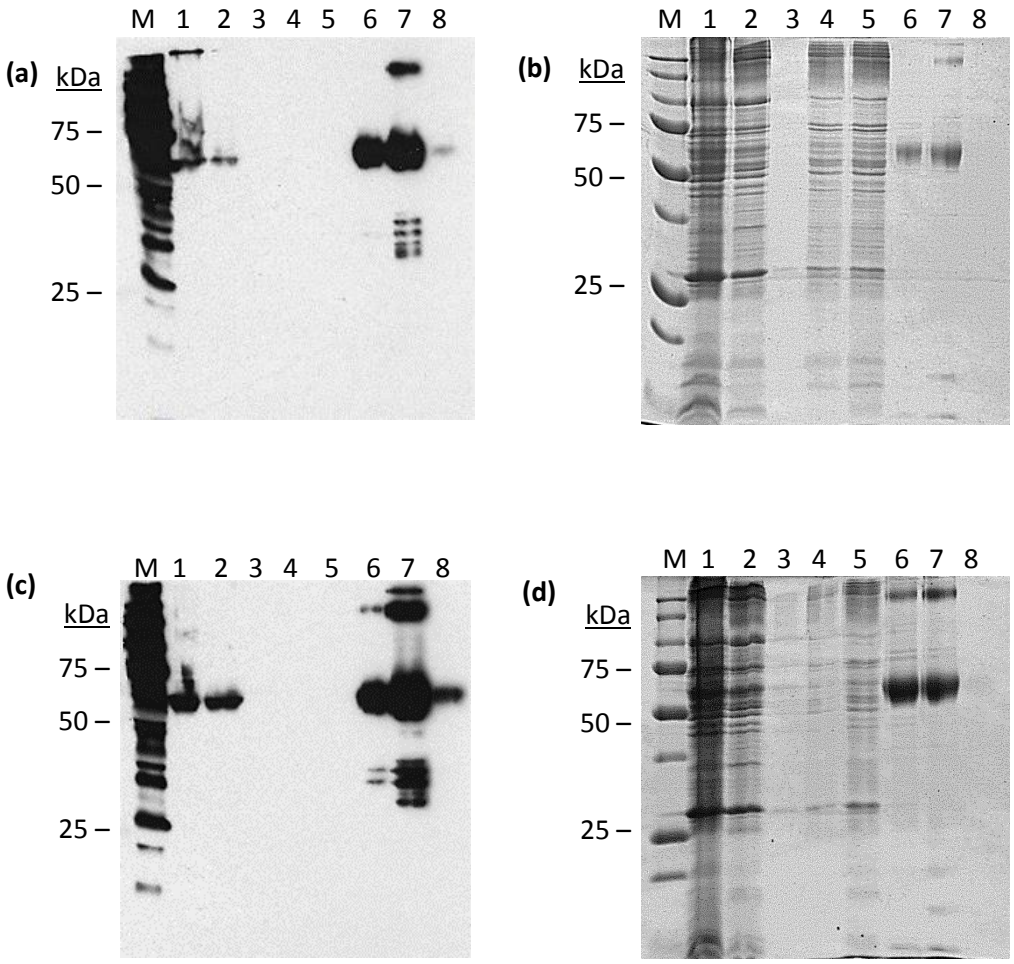


Figure 4.3.2. Western Analyses of the DDM Solubilisation and HisTrap™ Purification of **(a)** OafA and **(c)** GtrC^{BTP1} and their respective coomassie stained gels, **(b)** and **(d)**. M = Protein marker, Lane 1 = Vesicles isolated from cultures of expression host, Lane 2 = Soluble fraction after detergent solubilisation, Lane 3 = Insoluble fraction after detergent solubilisation, Lane 4 = HisTrap™ column flow through, Lane 5 = HisTrap™ column wash through, Lane 6 = Elution fraction 1, Lane 7 = Elution fraction 2, Lane 8 = Elution fraction 3. SDS-PAGE was performed with a 12% SDS-PAGE gel, under denaturing and reducing conditions.

To determine the level of solubilisation of OafA and GtrC^{BTP1} by DDM, the sample of proteins without DDM, the soluble fraction and insoluble fraction after DDM solubilisation were analysed as shown on lanes 1, 2 and 3 respectively. If the membrane proteins were successfully liberated from the lipid bilayer of the membrane, they would appear in the soluble fractions or lane 2 on Figure 4.3.2. This was indeed observed on the western analyses of the solubilisation of OafA and GtrC^{BTP1} shown on Figures 4.3.2(a) and (c) with a signal most likely representing these proteins of interest appearing in the soluble fraction on lane 2 with a complete absence of these proteins in the insoluble fraction shown on lane 3. The coomassie stained gels shown on Figure 4.3.2(b) and (d) also revealed that DDM was successful in the liberation of almost all of the membrane proteins with very few lightly visible bands observed in the insoluble fractions. Due to the highly successful solubilisation of both OafA and GtrC^{BTP1} by 0.5% (w/v) DDM, future solubilisation of both proteins would be carried out with 0.5% (w/v) DDM.

4.3.2 Purification of OafA and GtrC^{BTP1} by immobilised metal affinity chromatography (IMAC) – HisTrap™ nickel affinity column.

After detergent solubilisation, the proteins of interest would be purified by IMAC. This involves the interaction of the histidine side chain of the affixed His tag to metal ions with the valency of 2 (2+) bound to immobile beads in a column or magnetic beads. Every metal ion is able to bind to two histidine side chains. Therefore, the longer the His tag, the stronger the protein will bind to the beads. However, the His tag should not be excessively long due to the possibility that it would affect the folding and function of the protein. OafA and GtrC^{BTP1} were expressed with a c-terminal deca-His tag which should sufficiently bind onto the metal affinity beads for purification by IMAC. The principle behind nickel affinity purification is depicted on Figure 4.3.3.

IMAC purification was performed using the GE Healthcare Life Sciences HisTrap™ HP column which contains sepharose beads which were loaded with Nickel ions (Ni^{2+}) for His tag protein purification. Protein from the soluble fraction of the DDM solubilisation performed as mentioned in Section 4.3.1 were loaded on to the column. Sufficient washes were performed to remove as much proteins as possible and the bound proteins were eluted off with an elution buffer containing a high concentration of imidazole (Materials and Methods Section 2.9.2). Three consecutive 1 mL fractions were collected and western analyses of these fractions were performed as shown on lanes 6, 7 and 8 of Figure 4.3.2 with the first collected 1 mL fraction being represented by lane 6, second by lane 7 and the final fraction by lane 8. In addition, the flow through from the loading of the soluble fraction from the DDM solubilisation and the wash were also analysed and are depicted on lanes 4 and 5 respectively on Figure 4.3.2.

The western analyses of the flow and wash of the HisTrap™ purification of OafA and GtrC^{BTP1} shown on lanes 4 and 5 of Figure 4.3.2(a) and (c) suggests that the proteins of interest are not

present and hence, the interactions of the His tag of the expressed proteins is sufficient to bind these proteins to the HisTrap™ column while losing a negligible amount. Future protein purification experiments would be carried out using the HisTrap™ purification system charged with Ni^{2+} ions.

Lanes 6 and 7 of both the western analyses and coomassie stained gels of Figure 4.3.2 for the purification of both proteins showed an intense signal and significant coomassie staining for what are predicted to be OafA and $GtrC^{BTP1}$. Additional western signals and coomassie stained bands were present on all lane 7 and lane 6 for the purification of $GtrC^{BTP1}$ at approximately 250 kDa and a group of bands between 25 and 37 kDa. The group of bands between 25 kDa and 37 kDa would most likely represent amino acid fragments from the degradation of these proteins and the higher molecular weight bands may represent a form of protein aggregation.

The final elution fraction shown on lane 8 produced a weak signal on the western analyses for both proteins while producing a very faint coomassie stained band for $GtrC^{BTP1}$ and no staining for OafA. These information suggest that all proteins were successfully eluted within the 3 mL of elution buffer used.

The results from the purification of OafA and $GtrC^{BTP1}$ using the HisTrap™ column and the elution buffer suggests that this is a method to sufficiently purify and concentrate these proteins. However, the presence of protein aggregates and fragments would need to be overcome as future biochemical and biophysical experiments would require pure proteins in their native form.

4.3.3 Purification of OafA and $GtrC^{BTP1}$ by size exclusion chromatography (SEC)

Protein purification by IMAC is sufficient to separate the proteins of interest from almost all of the native proteins of the host strain. However, as observed in Section 4.3.2, it is insufficient to separate the aggregates and amino acid fragments from degraded proteins of interest.

In addition, native high histidine content proteins like ZitB as mentioned in Section 4.1.1 might also have be co-purified along with the proteins of interest. Hence, another purification step is required to isolate the solubilised proteins of interest from the aggregates, fragments and native high histidine content proteins.

Size exclusion chromatography (SEC) is a method where proteins are separated according to the various molecular weights. Gel filtration is one form of SEC that is commonly used in the further purification of proteins. This method involves the filtration of the proteins using beads that possess pores which are able to allow passage for proteins up to a certain molecular weight.

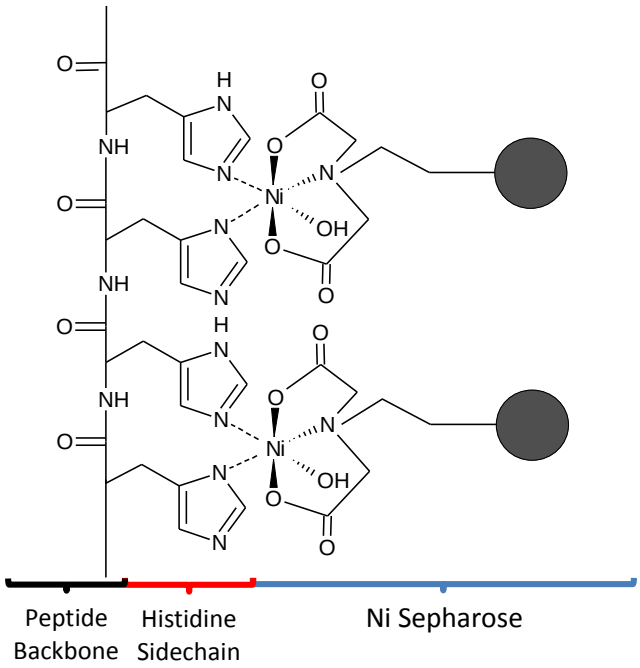


Figure 4.3.3. Binding of the His Tag to Ni^{2+} during Immobilised Metal Affinity Chromatography (IMAC). Refer to text for explanation. Diagram was designed using Microsoft PowerPoint 2013.

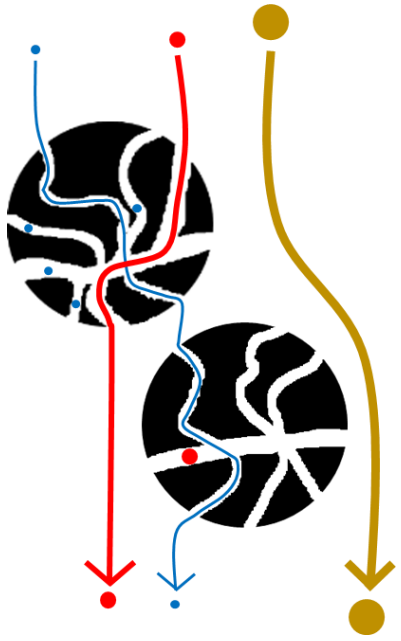


Figure 4.3.4. Principle of Size Exclusion Chromatography (SEC) by Gel Filtration. Refer to text for explanation. Diagram was designed using Microsoft PowerPoint 2013.

Proteins which are too big to fit through these pores travel straight down and are isolated in the initial few fractions. The smaller the protein, the more likely they will enter these pores and hence, travel a further distance down the gel filtration column. As a result, proteins with the biggest molecular weights and also protein aggregates will appear first, then possibly the proteins of interest and finally, the fragments from the degradation.

Presence of proteins in these fractions is measured by absorbance units at a wavelength of 280 nanometers (nm) (A_{280}). The detector that is used to measure these absorbance units detects the presence of tryptophan, tyrosine and to a smaller extent, phenylalanine groups on the proteins of interest. Therefore, the more proteins in the fraction will produce a larger A_{280} value. The A_{280} value measured can also be used to calculate the concentration of proteins within the fraction if their various extinction coefficients are known. Extinction coefficients can be calculated using the following equation (Gill and Von Hippel, 1989),

$$\begin{aligned} &Extinction\ coefficient\ (M^{-1}cm^{-1}) \\ &\quad = (number\ of\ tryptophan\ residues \times 5960) \\ &\quad + (number\ of\ tyrosine\ residues \times 1280) \end{aligned}$$

IMAC should concentrate these proteins of interest in the eluates as seen in Figures 4.3.2(b) and (d) if sufficient amounts of proteins were produced. Hence, the fractions containing the proteins of interest should produce the highest absorbance values if the proteins are not phenyl group amino acid deficient. OafA and GtrC^{BTP1} have extinction coefficients of 124680 $M^{-1}cm^{-1}$ and 121130 $M^{-1}cm^{-1}$ respectively. The purification steps mentioned in the following sections were performed using the GE Life Sciences S200 26/60 Prep Grade SEC column. Separate 5L expression cultures of *E. coli* MC1061 expressing OafA from pOafA and GtrC^{BTP1} from pGtrC^{BTP1} were prepared. The proteins were solubilised and purified by IMAC using the method mentioned in Sections 4.3.1 and 4.3.2. All eluates from IMAC which contained the proteins were loaded onto the SEC column and purified. Hence, the final yield of proteins calculated would be per 5L cultures of the expression strains prepared.

4.3.3.1 Purification of OafA by SEC – Gel Filtration

The SEC A_{280} trace from the purification of OafA by gel filtration is depicted on Figure 4.3.5(a) with the photos of the coomassie stained gels for the analysis of the selected fractions shown on Figure 4.3.5(b) and (c).

A clear strong peak at approximately 200 mAu was observed between samples from Fractions 30 to 39 as identified by the red arrow shown on Figure 4.3.5(a). These fractions were analysed by coomassie staining and the resulting photo was shown on Figure 4.3.5(c). This should represent the monomeric form of the solubilised OafA which will be used for future biochemical and biophysical analyses which would be performed for the characterisation of this protein.

Coomassie stained gels revealed that Fraction 33 possessed the highest concentration of OafA, followed by Fraction 34, 32, 35 and 36.

In addition to the strong peak at approximately 200mAu, there was a presence of a gradual slope as identified by the blue arrow on Figure 4.3.5(a) from between Fractions 21 to 29 before the steep increase starting from Fraction 30 that made up the 200mAu peak. This identified slope might represent protein aggregates of OafA since there is a presence of these proteins in the eluates from the IMAC purification as shown in Figure 4.3.2. Slight coomassie staining was observed in the lanes representing Fractions 28 and 29 shown on Figure 4.3.5(b) as identified by the green arrows. These coomassie stained bands appear at approximately 100 kDa but at much lower intensities as compared to the bands observed on lanes representing Fractions 32, 33 and 34 on Figure 4.3.5(c). These bands do not reappear on the lanes representing proteins within Fractions 30 and 39, hence suggesting that the protein aggregates should have been fully separated from the monomeric form of OafA.

Fractions 32 to 36 were pooled together in order to attain the highest yield of OafA possible, giving a 25 mL stock of proteins. The concentration of this 25 mL stock of OafA was determined using a Nanodrop 1000 which gave an A_{280} value of 0.091 Au. Using the known extinction coefficient and molecular weight of OafA, the concentration of OafA and yield was determined. The calculations for the concentration and yield of OafA were performed as follows,

$$Extinction\ coefficient\ of\ OafA = 124680 M^{-1} cm^{-1}$$

$$Molecular\ weight\ of\ OafA = 72391.5 Da$$

$$Concentration\ of\ OafA = 0.091 \div \frac{124680}{72391.5} = 0.0528 mg/mL$$

$$\therefore Total\ amount\ of\ OafA\ expressed\ per\ 5L\ culture = 0.0528 \times 25 = 1.32 mg$$

$$\therefore Yield\ of\ OafA\ per\ 1L\ culture = 1.32 \div 5 = 0.264 mg/L$$

The overall OafA expression yield of 0.264 mg per litre of culture is considerably lower than desired as a few milligrams of pure proteins would be ideal for protein crystallisation trials and other biochemical and biophysical experiments.

The reason for the low yield of OafA is unclear. The coomassie staining analyses of the elution fractions containing proteins isolated by IMAC as shown on Figure 4.3.2 revealed that there were more purified $GtrC^{BTP1}$ as compared to OafA as the coomassie stained bands on Figure 4.3.2(d) for $GtrC^{BTP1}$ was considerably more intense than that of OafA. Some form of degradation could have occurred prior to IMAC or less OafA was expressed as compared to $GtrC^{BTP1}$ to begin with. To remedy this situation, a larger expression culture volume for *E. coli* MC1061 expressing OafA from the pOafA vector is necessary to attain the appropriate yields of OafA. Concentration using a spin column with the PES membrane resulted in precipitation.

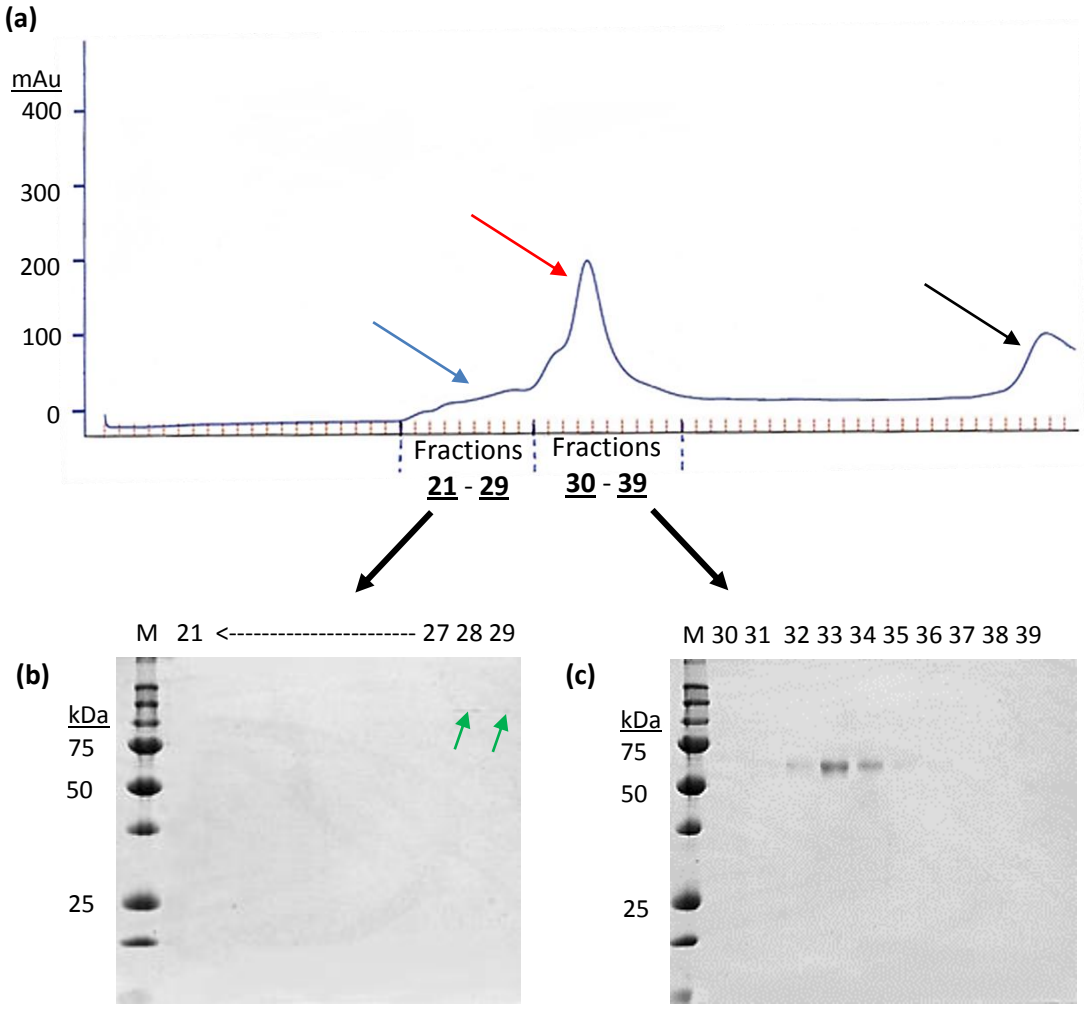


Figure 4.3.5. Size Exclusion Chromatography (SEC) Trace of the Purification of OafA. Absorbance values of each of the fractions collected were measured in milli-absorbance units (mAu) as shown in **(a)**. Fractions collected were then analysed by coomassie staining as shown in **(b)** for fractions 21 to 29 and **(c)** for fractions 30 to 39. Numbers on the top of the coomassie stained gel photo represents the fraction number. M = Protein marker. The red arrow was used to identify the peak that is thought to represent the protein of interest. The blue arrow was used to identify any regions of possible protein aggregation. The black arrow was used to identify absorbance peaks that might have been produced as a result of the presence of high concentrations of salt in fractions.

4.3.3.2 Purification of GtrC^{BTP1} by SEC – Gel Filtration

As per OafA, the SEC A_{280} trace from the purification of GtrC^{BTP1} by gel filtration is depicted on Figure 4.3.6(a) with the photos of the coomassie stained gels for the analysis of the selected fractions shown on Figure 4.3.6(b) and (c).

The A_{280} trace shown on Figure 4.3.6(a) revealed a very strong peak climaxing at slightly above 700mAu as shown by the red arrow with the peak covering Fractions 30 to 38. These fractions were then analysed by protein separation by gel electrophoresis followed by coomassie staining as depicted on Figure 4.3.6(c) which showed that the highest concentrations of proteins thought to be GtrC^{BTP1} was in Fraction 32, followed by 33, 31, 34, 35 and 36. Like OafA, a gradual slope was observed and identified by the blue arrow on the SEC A_{280} trace from Fraction 21 until the Fraction 30 where a sharp rise in the A_{280} value begins, forming the sharp peak. This was thought to represent the aggregation of the expressed GtrC^{BTP1} proteins. Analyses of Fractions 21 to 29 by SDS-PAGE protein separation and coomassie staining was performed and the photo of the resulting gel was depicted on Figure 4.3.6(b). Like OafA, very faint bands were observed at approximately 100 kDa with respect to the protein marker on lanes representing Fractions 28 and 29 as pointed out by the green arrows on Figure 4.3.6(b). These bands were also absent on the lanes representing Fractions 31 to 36 which would mean that the potential protein aggregates of what is thought to be GtrC^{BTP1} was successfully separated from the monomeric form of GtrC^{BTP1}.

Fractions 31 to 36 were pooled together in order to attain the highest yield of GtrC^{BTP1}. The A_{280} value of the resulting 30 mL stock was measured using the Nanodrop 1000 and was determined to be 0.325Au. Using the known values of the extinction coefficient and molecular weights of GtrC^{BTP1}, the concentration of proteins and yield were determined.

The calculations were as follows,

$$Extinction\ coefficient\ of\ GtrC^{BTP1} = 121130 M^{-1} cm^{-1}$$

$$Molecular\ weight\ of\ GtrC^{BTP1} = 74132.4 Da$$

$$Concentration\ of\ GtrC^{BTP1} = 0.325 \div \frac{121130}{74132.4} = 0.1989 mg/mL$$

$$\therefore Total\ amount\ of\ GtrC^{BTP1}\ expressed\ per\ 5L\ culture = 0.1989 \times 30 = 5.967 mg$$

$$\therefore Yield\ of\ GtrC^{BTP1}\ per\ 1L\ culture = 5.967 \div 5 = 1.1934 mg/L$$

As compared to OafA, the yield per litre of expression cultures of *E. coli* MC1061 expressing GtrC^{BTP1} from pGtrC^{BTP1} was approximately 4.5 folds higher. Even though the yield of GtrC^{BTP1} is higher than that of OafA, it is still a relatively lower than needed for extensive biochemical and biophysical characterisation. A larger expression culture volume of *E. coli* MC1061 would still be preferred to attain a higher yield prior to biochemical and biophysical experiments. Like for OafA, concentration using a spin column with the PES membrane resulted in precipitation.

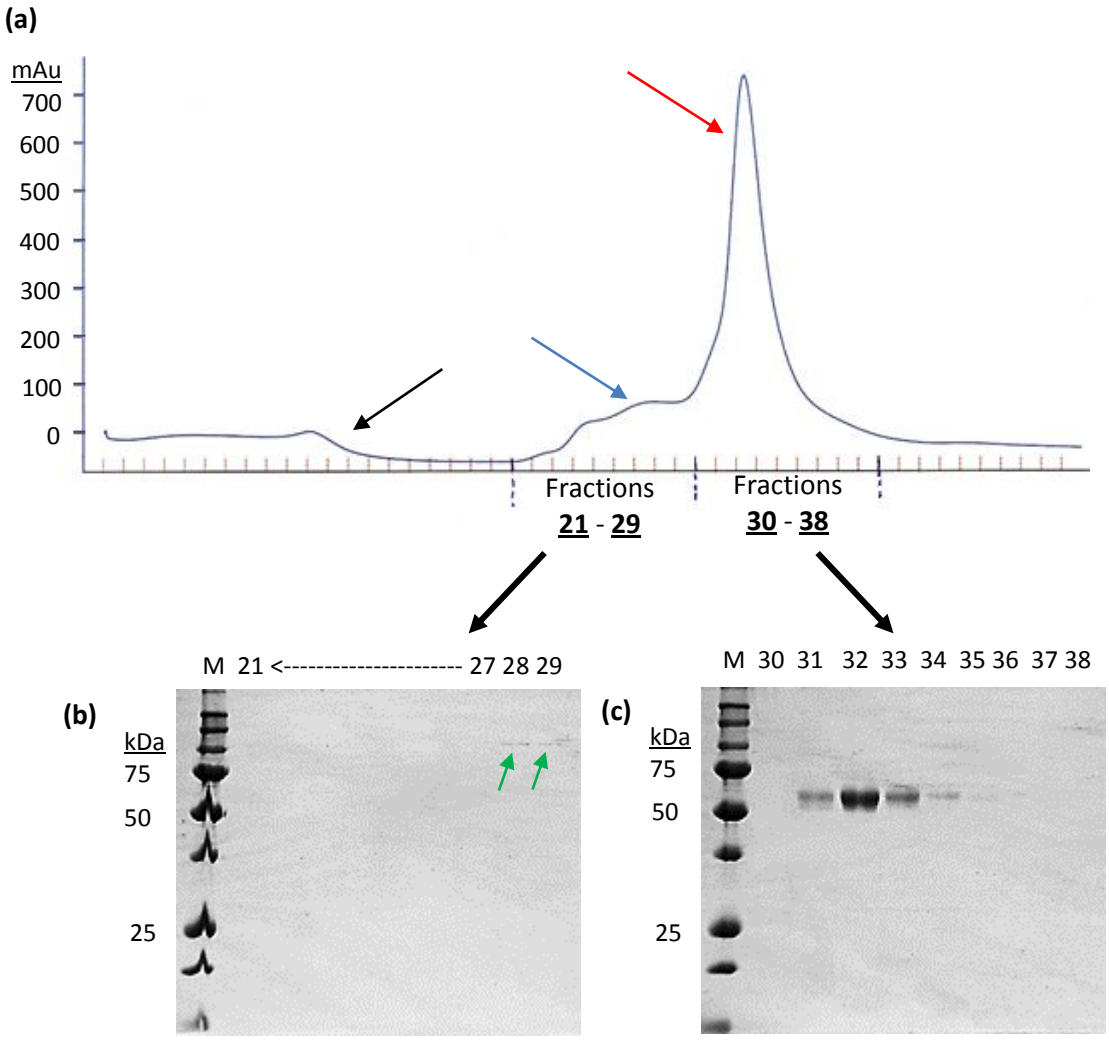


Figure 4.3.6. Size Exclusion Chromatography (SEC) Trace of the Purification of GtrC^{BTP1}. Absorbance values of each of the fractions collected were measured in milli-absorbance units (mAu) as shown in **(a)**. Fractions collected were then analysed by coomassie staining as shown in **(b)** for fractions 21 to 29 and **(c)** for fractions 30 to 38. Numbers on the top of the coomassie stained gel photo represents the fraction number. M = Protein marker. The red arrow was used to identify the peak that is thought to represent the protein of interest. The blue arrow was used to identify any regions of possible protein aggregation. The black arrow was used to identify absorbance peaks that might have been produced as a result of the presence of high concentrations of salt in fractions.

4.4 Chapter 4 Conclusion

From the results discussed in the sections within Chapter 4, it was concluded that the method used for the expression and purification of both OafA and GtrC^{BTP1} was sufficient to attain pure samples of both proteins. However, the nature of these proteins affects the overall yield of proteins attained from the expression cultures negatively with an achieved maximum protein yield of 1.2 mg/L. Larger culture volumes of *E. coli* MC1061 expressing OafA from the pOafA vector and GtrC^{BTP1} from the pGtrC^{BTP1} vector should be prepared for future expression and purification of these proteins.

Further expression optimisation could also be performed using more appropriate host strains such as *E. coli* C41 and C43 which were determined to be more successful at over-expressing membrane proteins due to the ability to produce membrane proteins as inclusion bodies at higher levels than common *E. coli* expression strains like *E. coli* BL21(DE3) without the toxicity (Miroux and Walker, 1996). These strains are able to consume L-arabinose due to the absence of the mutations necessary for the absence of arabinose metabolism. Hence, the optimisation experiments as mentioned in Section 4.2.2 would need to be repeated for all concentrations of L-arabinose, the duration and temperatures at which induction of protein expression is performed.

Due to the success of solubilisation of OafA and GtrC^{BTP1} using DDM based detergent solubilisation as seen in Section 4.3.1, no solubilisation optimisation is required to be performed. Purification of both proteins of interest by IMAC using the HisTrap™ column followed by SEC were also successful in attaining a highly pure sample of proteins which would be useful in any downstream biochemical and biophysical protein characterisation analyses.

Chapter 5

Serine Esterase Catalytic Triad In GtrC^{BTP1} from *S.* Typhimurium

Chapter 5 Introduction

The SGNH hydrolase family of domains are divergent from the classical GDSL family of serine lipases and esterases. One difference between these proteins is the position of the catalytic residues (i.e. serine, aspartate and histidine) which serine esterases are known to possess. While GDSL lipases have a large number of amino acids separating each of the catalytic resides, SGNH hydrolases notably have only a maximum of 2 amino acids separating the aspartate and histidine residues.

The role of these catalytic residues from the current literature were previously mentioned in Section 3.2. Here, the importance of these catalytic residues will be further discussed and suggesting how it might affect the function of OafA and $GtrC^{BTP1}$.

5.1 Analyses of the Catalytic Residues of GtrC^{BTP1} and OafA

5.1.1 Hypotheses for the replacement of the putative catalytic amino acids of OafA and GtrC^{BTP1}

As discussed in Chapter 3, OafA and GtrC^{BTP1} are thought to possess two domains with the N-terminal half being a member of the MBOAT family of O-acetyltransferases and the periplasmic C-terminal half being an SGNH hydrolase protein.

SGNH hydrolase type esterases possesses the serine esterase catalytic triad of serine, aspartate and histidine. The catalytic aspartate and histidine residues of SGNH hydrolase enzymes are found within three amino acids from one another (i.e. DXXH). Through bioinformatical analyses, OafA and GtrC^{BTP1} were hypothesised to possess the SGNH hydrolase type catalytic triad in their respective C-terminal periplasmic regions. The predicted structures of the respective C-terminal tails of both OafA and GtrC^{BTP1} were depicted along with the positions of the putative catalytic residues on Figure 3.4.1(a) and (b) respectively.

To assess the importance of the potentially catalytic amino acid residues of OafA and GtrC^{BTP1}, the catalytic residues in question were replaced with alanine and the enzymatic functions were assessed via *in vivo* studies.

Due to the presence of these characteristic SGNH hydrolase type catalytic amino acids, it is thought that these replacement mutations would abrogate the functions of these proteins.

5.1.2 Catalytic residues of GtrC^{BTP1}

5.1.2.1 Replacement mutations of the putative catalytic residues of GtrC^{BTP1}

The putative catalytic residues of GtrC^{BTP1} were individually replaced with alanine to produce mutants with the S430A, D618A and H621A mutations separately. This was performed by site directed mutagenesis of pGtrC^{BTP1} [pMV434] using the protocol described in Section 2.6. Hence, like the wild-type expressed GtrC^{BTP1} protein expressed and purified as mentioned in Chapter 4, the mutant GtrC^{BTP1} products would also possess a deca-his tag which would be useful in determining if the proteins were being expressed and in the future, could be used to express and purify these mutant proteins for biochemical or biophysical analyses. Resulting vectors of pBADcLIC expressing GtrC^{BTP1} possessing the abovementioned mutations will henceforth be referred to as pGtrC^{BTP1}(S430A) [pMV437], pGtrC^{BTP1}(D618A) [pMV438] and pGtrC^{BTP1}(H621A) [pMV439]. The sequences were confirmed by sequencing (Section 2.4.7.3).

5.1.2.2 Functional analyses of GtrC^{BTP1}with catalytic residue replacement mutations

As mentioned in the introduction, GtrC^{BTP1} acetylates the rhamnose residues of the LPS O-antigen of *S.* Typhimurium D23580, hence protecting the bacteria from bacteriophage associated death by the phage BTP1. This modification enzyme originates from the BTP1 phage and is thought to be present in the bacteria to stop superinfection by blocking the BTP1 tail spike

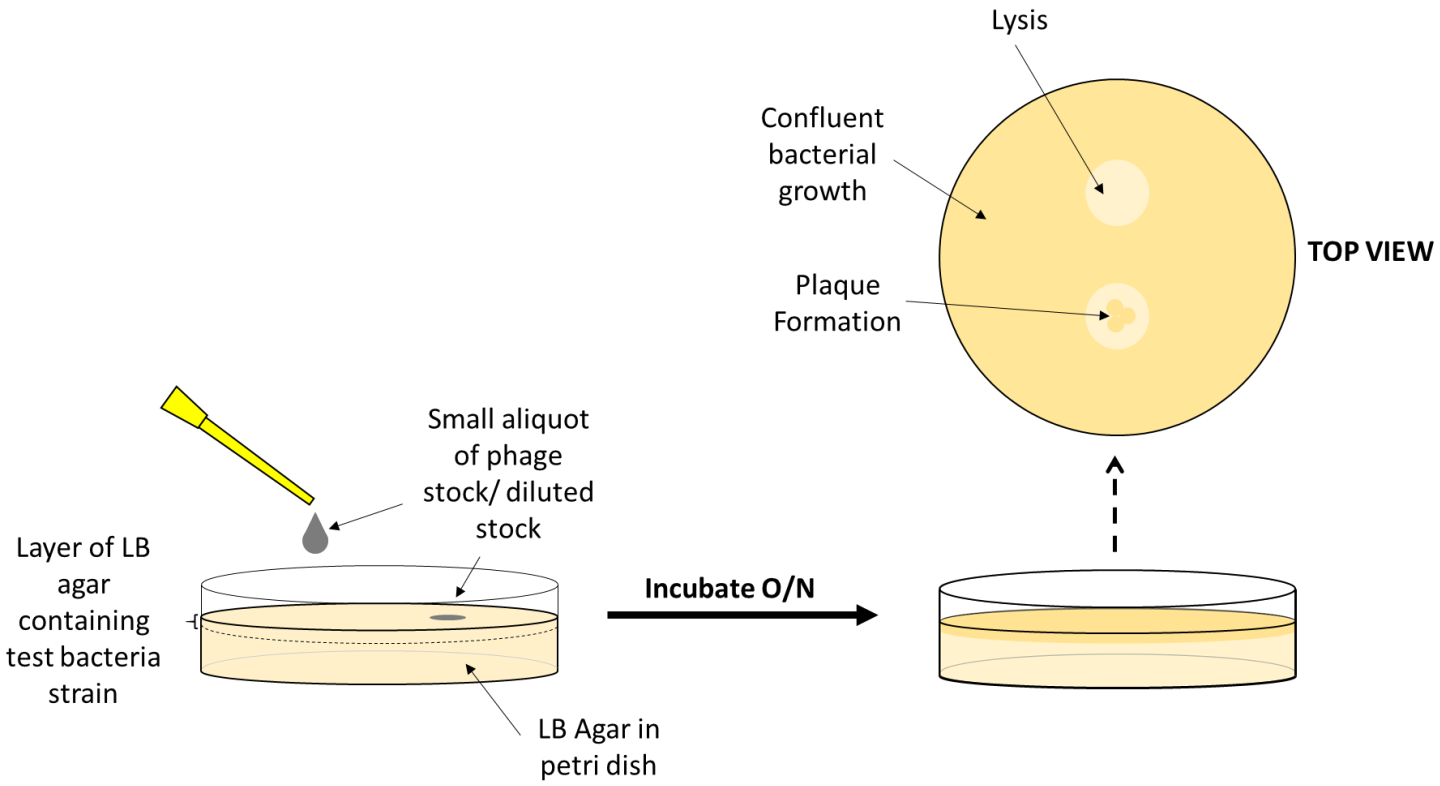


Figure 5.1.2.1 Phage Plaque Assay. See text for details.

protein endorhamnosidase activity (Kintz et al., 2015). In order to assess the functionality of the GtrC^{BTP1} proteins with the mutations mentioned in Section 5.1.2.1, an *in vivo* assay called the phage plaque assay was performed. This involves the addition of a small aliquot of bacteriophage onto LB agar containing the test bacteria layered on an LB agar plate as shown on Figure 5.1.2.1 and mentioned in Section 2.7.

The introduced bacteriophage would cause lysis of bacteria in a localised area of the layer of LB agar, identified by a clear area on the LB agar layer and the formation of plaques which are colonies of bacteria which would have gained resistance to the phage. Test bacterial strains which are resistant to phage infection would grow normally despite the presence of the bacteriophage and hence, no clear area on the LB agar layer would be observed.

The mutant GtrC^{BTP1} proteins with either the S430A, D618A or H621A mutations were expressed separately in *S.* Typhimurium LT2 which does not possess a native LPS O-antigen rhamnose modifying enzyme but possesses the necessary O-antigen substrate for GtrC^{BTP1} activity. Along with the mutant GtrC^{BTP1} proteins, the wild type GtrC^{BTP1} protein from pGtrC^{BTP1} was also expressed in *S.* Typhimurium LT2 to test for the function of GtrC^{BTP1} with a deca-his tagged C-terminal end that is expressed from the pBADcLIC vector. The clones were induced in liquid cultures overnight (approximately 19 hours) at the optimised expression conditions (0.01% w/v L-arabinose induction at 30°C) for the expression and purification of the wild type GtrC^{BTP1} proteins mentioned in Chapter 4.

To determine if the overnight cultures have produced the proteins, a western and coomassie analysis was performed and the results were depicted on Figure 5.1.2.2(a) and (b) respectively. The western analyses of the overnight cultures of *S.* Typhimurium LT2 expressing the wild type GtrC^{BTP1} and its various mutants revealed that the proteins of interest were indeed expressed in all clones, albeit in low amounts for that which was expressing the wild type and the mutants possessing the S430A and D618A replacements as shown by the red arrows in their respective lanes. The photo of the corresponding coomassie stained duplicate gel shown on Figure 5.1.2.2(b) showed that sufficient amount of proteins were loaded into the gel and hence, should produce more intense signals if the expression of these proteins in *S.* Typhimurium LT2 were comparable to the expression levels of the wild type GtrC^{BTP1} in *E. coli* as seen in Figure 4.2.3 and 4.2.4. Hence, expression of these proteins in *S.* Typhimurium LT2 was likely affected by a host strain characteristic. One likely reason for this could be the *S.* Typhimurium LT2 consumes arabinose at a faster rate than the *E. coli* expression strains utilised as mentioned in Section 4.2.2 due to the presence of the enzymes involved in the metabolism of arabinose. As the western and coomassie analyses for the expression of GtrC^{BTP1} and its mutant variants from the various clones were performed concurrently with the phage plaque assay, the low concentration of protein products in the overnight cultures were not known when the latter was performed.

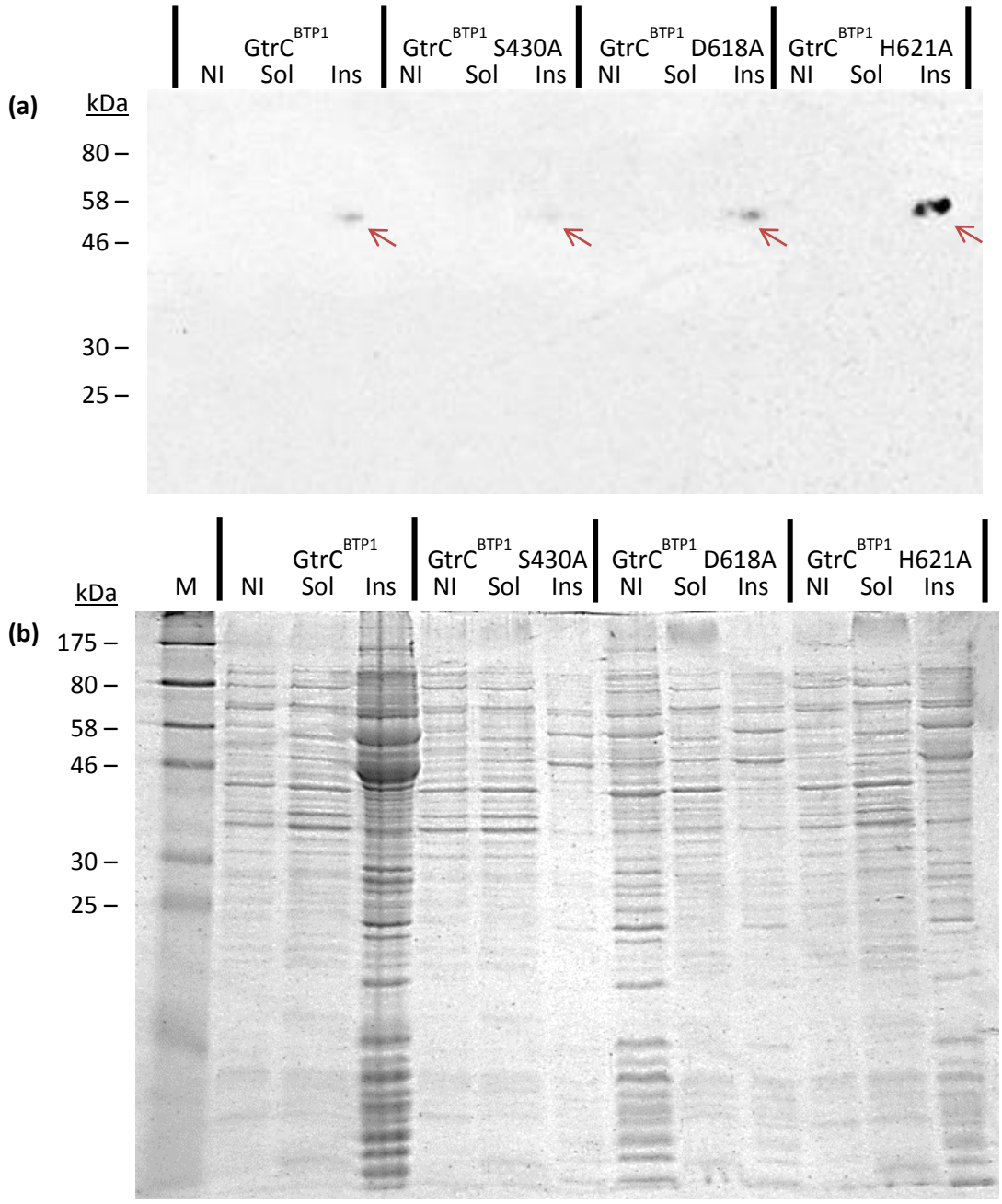
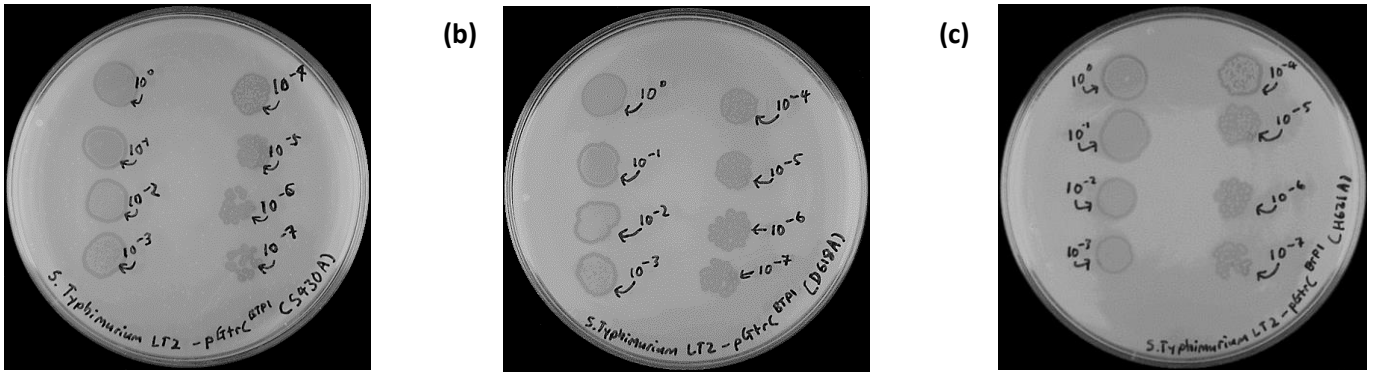


Figure 5.1.2.2 Western Analyses of *S.* Typhimurium LT2 Expressing GtrC^{BTP1} and its Mutant Variants. Performed to determine if the mentioned proteins are expressed in the overnight culture prior to addition into the LB agar for BTP1 phage plaque assay. **(a)** Western analyses using the anti-tetra his primary antibody and **(b)** coomassie staining of the soluble (Sol) and insoluble (Ins) fractions from the cell lysates of S. Typhimurium LT2 expressing GtrC^{BTP1} from pGtrC^{BTP1}, GtrC^{BTP1}S430A from pGtrC^{BTP1}(S430A), GtrC^{BTP1} D618A from pGtrC^{BTP1}(D618A) and GtrC^{BTP1} H621A from pGtrC^{BTP1}(H621A) along with the total protein sample from the cell lysates of their respective non – induced cultures (NI). M = protein standard marker. The red arrows in (a) identifies the signals which are thought to be the proteins of interest. Protein separation was performed with a 12% SDS-PAGE gel, under denaturing and reducing conditions.

Test Strains:



Control Strains:

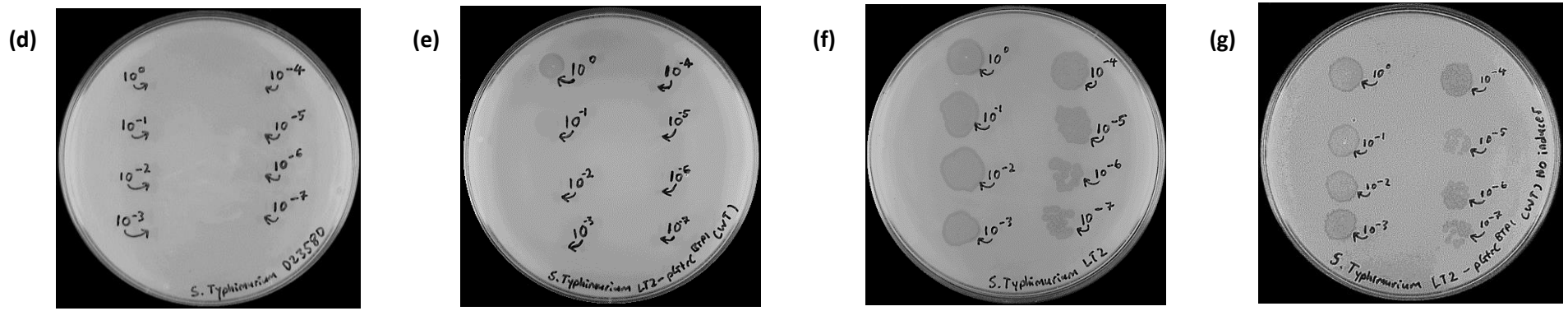


Figure 5.1.2.3. BTP1 Phage Plaque Assay for the Functional Analysis of Mutant GtrC^{BTP1} Proteins. *S.* Typhimurium LT2 strains expressing the GtrC^{BTP1} mutants possessing the **(a)** S430A, **(b)** D618A and **(c)** H621A replacement mutations. The negative controls with respect to BTP1 phage activity that were included were **(d)** *S.* Typhimurium D23580, possessing a native GtrC^{BTP1} and **(e)** *S.* Typhimurium LT2 pGtrC^{BTP1} expressing the wild-type GtrC^{BTP1}. The positive controls with respect to BTP1 phage activity include **(f)** *S.* Typhimurium LT2, parental strain for the test strains and **(g)** *S.* Typhimurium LT2 pGtrC^{BTP1} vector but no inducer was added for the expression of the wild-type GtrC^{BTP1} proteins. Dilutions of the phage stock were performed in 10-folds (i.e. 10^{0}, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6} and 10^{-7}) and the various spots where the aliquots were placed were identified by the black arrows. The LB agar in plates for (a), (b), (c) and (e) were supplemented with 0.01% L-arabinose.

In addition to protein expression level analyses, the comparison of the fold of the expressed proteins were performed using the results of the western analyses by looking at where their respective signals are with respect to the protein marker. No significant changes were observed between the wild type and the mutant proteins which could have led to a shift in the position of the signal with respect to the protein marker. However, this method would only detect significant changes in protein folding which would affect the interactions of the membrane protein with the detergent in the SDS-PAGE gel (Rath et al., 2009). Structurally minor but functionally significant changes might not be detected in the shift in signal position with respect to the protein marker. A more definitive method of determining structural changes such as circular dichroism would require pure protein samples which would prove to be cumbersome to produce due to the complex membrane protein purification procedures.

The induced overnight cultures were then added to molten LB agar with the inducer forming a mixture which was then layered on a prepared LB agar plate, also containing the inducer. Small aliquots of BTP1 phage in various dilutions were added in spots on the layer of LB agar containing the different induced overnight cultures. The resulting LB agar plates were incubated overnight and the results were depicted on Figure 5.1.2.3. The results from the BTP1 phage plaque assay is depicted on Figure 5.1.2.3. The BTP1 phage stocks were able to infect susceptible *S.* Typhimurium strains as shown by the positive control in the form of *S.* Typhimurium D23580 (Figure 5.1.2.3(d)) and the negative control in the form of *S.* Typhimurium LT2 (Figure 5.1.2.3(f)) with respect to BTP1 phage activity. The control in the form of *S.* Typhimurium LT2 pGtrC^{BTP1} was performed to determine if the GtrC^{BTP1} expressed from the pBADcLIC vector is sufficient in protecting the initially susceptible *S.* Typhimurium LT2 strain against BTP1 phage associated killing. As shown in Figure 5.1.2.3(e) for the induced and Figure 5.1.2.3(g) for the uninduced cultures of *S.* Typhimurium LT2 pGtrC^{BTP1}, the expressed GtrC^{BTP1} seemed to have been sufficient for the acetylation of rhamnose and hence, protection against BTP1 phage associated lysis at phage dilutions of 10^{-2} and lower of the initial stock despite the low expression levels seen in the western analyses (Figure 5.1.2.2(a)). This suggests that the expression of the mutant proteins in *S.* Typhimurium LT2 should be sufficient in the protection of the host strain from BTP1 phage associated lysis if the mutant GtrC^{BTP1} is able to function despite the mutations introduced.

Results from the phage assays performed for the *S.* Typhimurium LT2 strains expressing all three mutant proteins from their respective vectors mentioned above were depicted in Figure 5.1.2.3(a), (b) and (c) for GtrC^{BTP1} S430A, GtrC^{BTP1} D618A and GtrC^{BTP1} H621A respectively. All three mutant GtrC^{BTP1} proteins were not able to protect the host strain from BTP1 phage associated lysis, unlike the wild type GtrC^{BTP1} protein, as significant plaque formation were seen throughout all dilutions of the phage stocks, like that of the positive controls. The results of the BTP1 phage plaque assay qualitatively suggests that a functional from of GtrC^{BTP1} can be

expressed from the pBADcLIC vector and single replacement mutations for the catalytic triad are sufficient to stop the function of GtrC^{BTP1}.

5.1.3 Catalytic residues of OafA

Due to the lack of time, the investigation into the importance of the putative catalytic residues in Ser412, Asp587 and His590 of OafA was not be performed. However, if given time, the mutagenesis of the mentioned catalytic residues of OafA would closely follow that of GtrC^{BTP1} and also assessed using an *in vivo* based method. Plasmid vectors from which OafA and its mutants are expressed, would be complemented into an *S.* Typhimurium strain which was engineered to not express OafA. Unlike GtrC^{BTP1}, there are no phage assays for the assessment of OafA protein activity. However, another *in vivo* experiment is available for assessment of OafA activity. This involves the expression of the various proteins in the *S.* Typhimurium strain and isolate the LPS of the bacteria which would be tested for the presence of the modification by western analyses using the anti-O:5 primary antibody.

5.2 Catalytic Triads of Other SGNH Hydrolase type Esterase Proteins

To date, the catalytic triad of a number of SGNH hydrolase proteins have been analysed either biochemically using mutagenesis and protein activity assays or biophysically by crystallisation. In the following sections, functional studies of the catalytic residues of a selection of these proteins would be discussed in order to compare with the results of the mutation of the catalytic residues GtrC^{BTP1} and hypothesise possible outcomes of the future experiment involving the replacement mutations of the catalytic triad of OafA.

5.2.1 Escherichia coli - thioesterase I/protease I/phospholipase L_1 (TAP)

TAP is a multifunctional protein of *E. coli* with multiple lipase and esterase functions. Crystallisation of this protein revealed an SGNH hydrolase fold (PDB: 1IVN) with the relative positions of the catalytic residues of serine at amino acid position 10, histidine at position 157 and aspartate at position 154 as shown in Figure 5.2.1(a) (Lee et al., 2006).

To assess the importance of these catalytic residues, Lee et al. (2006) first performed amino acid replacement mutations on these residues by substituting these residues with alanine (i.e. S10A, H157A and D154A). These mutant proteins were expressed, purified and the protein activities were tested using an esterase-activity assay. The proteins with the S10A and H157A mutations resulted in a retainment of only 0.57% and 0.09% of activity respectively as compared to that of the wild type. The mutant protein with the D154A mutation retained 18% of the activity. Circular dichroism of these proteins suggest that no structural changes were present as a result of the amino acid replacements. Hence, the loss in activity would most likely be attributed to the amino acid substitutions.

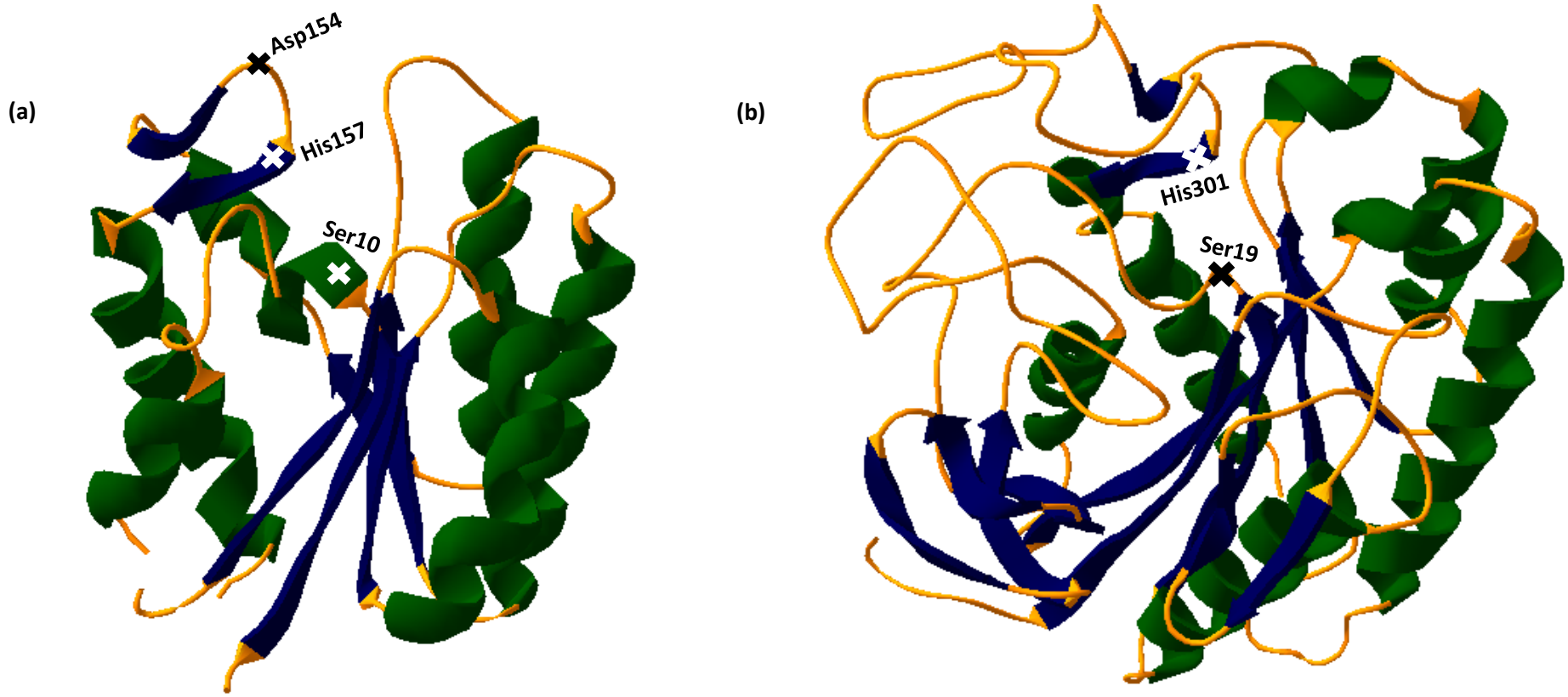


Figure 5.2.1. Crystal structure of (a) thioesterase I/protease I/phospholipase L_1 (TAP) of *E. coli* (PDB: 1IVN), and (b) NanS of *E. coli* O157:H7 (PDB: 3PT5). α-helices are coloured in dark green, β-strands are coloured blue and coils are coloured orange. Amino acids of catalytic triads are identified by crosses; black crosses are for amino acids that are found on coils and the white crosses are for amino acids that are found on the α-helices and β-strands. Amino acid names and positions are also included with their respective crosses.

5.2.2 E. coli O157:H7 - NanS (YjhS)

Sialic acid is a common molecule expressed on the surface of eukaryotic cells and possesses multiple functions. As a method of protection, *E. coli* O157:H7 and some other bacterial pathogens also express sialic acid in the capsular polysaccharide (CPS) or lipopolysaccharide (LPS) as a method to evade the immune system(Severi et al., 2007). Some of these bacteria are also able to catabolise the abundant sialic acid found in the host environment (Vimr and Lichtensteiger, 2002, Almagro-Moreno and Boyd, 2009) using enzymatic processes including the enzyme NanS, which deacetylates 9-*O*-sialic acid esterase(Steenbergen et al., 2009).

The structure of NanS was determined and analysed (PDB: 3PT5) and was found to possess an SGNH hydrolase type esterase fold as shown in Figure 5.2.1(b) (Rangarajan et al., 2011). However, unlike the *E. coli* TAP protein and most other SGNH hydrolase proteins, NanS only possesses two out of the three catalytic residues of the serine esterase in Ser19 and His301.

Rangarajan et al. (2011) generated mutant NanS proteins with S19A and H301N replacement mutations separately and tested their enzymatic activity using an esterase activity. Both mutant NanS proteins were not able to retain any esterase activity as compared to the wild type proteins. The absence in protein activity would most likely be attributed to the mutations introduced to the protein as the analyses of the tested proteins by dynamic light scattering, native gel analysis and circular dichroism revealed that these mutant proteins most likely possess the native fold of NanS and were in the monomeric form.

5.2.3 Pseudomonas aeruginosa - AlgX and AlgJ

Alginate is an exopolysaccharide that is excreted by *P. aeruginosa* to form biofilms in the lungs of cystic fibrosis patients for protection against environmental stresses and the host immune system (Simpson et al., 1988, Pedersen et al., 1990). Alginate can be modified by O-acetylation to produce a more stable biofilm matrix as compared to that of the non-O-acetylated alginate (Nivens et al., 2001). As mentioned in Section 3.2.4.2, there are two periplasmic proteins that are thought to cause the acetylation of alginate, namely AlgX and AlgJ. These proteins were analysed structurally and biochemically to understand their various functions better.

The structure of AlgX from *P. aeruginosa* PAO1 was determined to possess an SGNH hydrolase domain and a carbohydrate binding domain (PDB: 4KNC) as shown on Figure 5.2.2(a). The serine esterase catalytic triad of the SGNH hydrolase domain was identified at Ser269, His176 and Asp174 (Riley et al., 2013). To assess the importance of these catalytic residues, separate alanine replacement mutations were performed on these potentially catalytic amino acids (Riley et al., 2013). The resulting mutant AlgX possessed either the S269A, H176A or D174A mutations. The acetylesterase activity of these mutants were assessed in comparison with the wild type AlgX using a fluorometric assay.

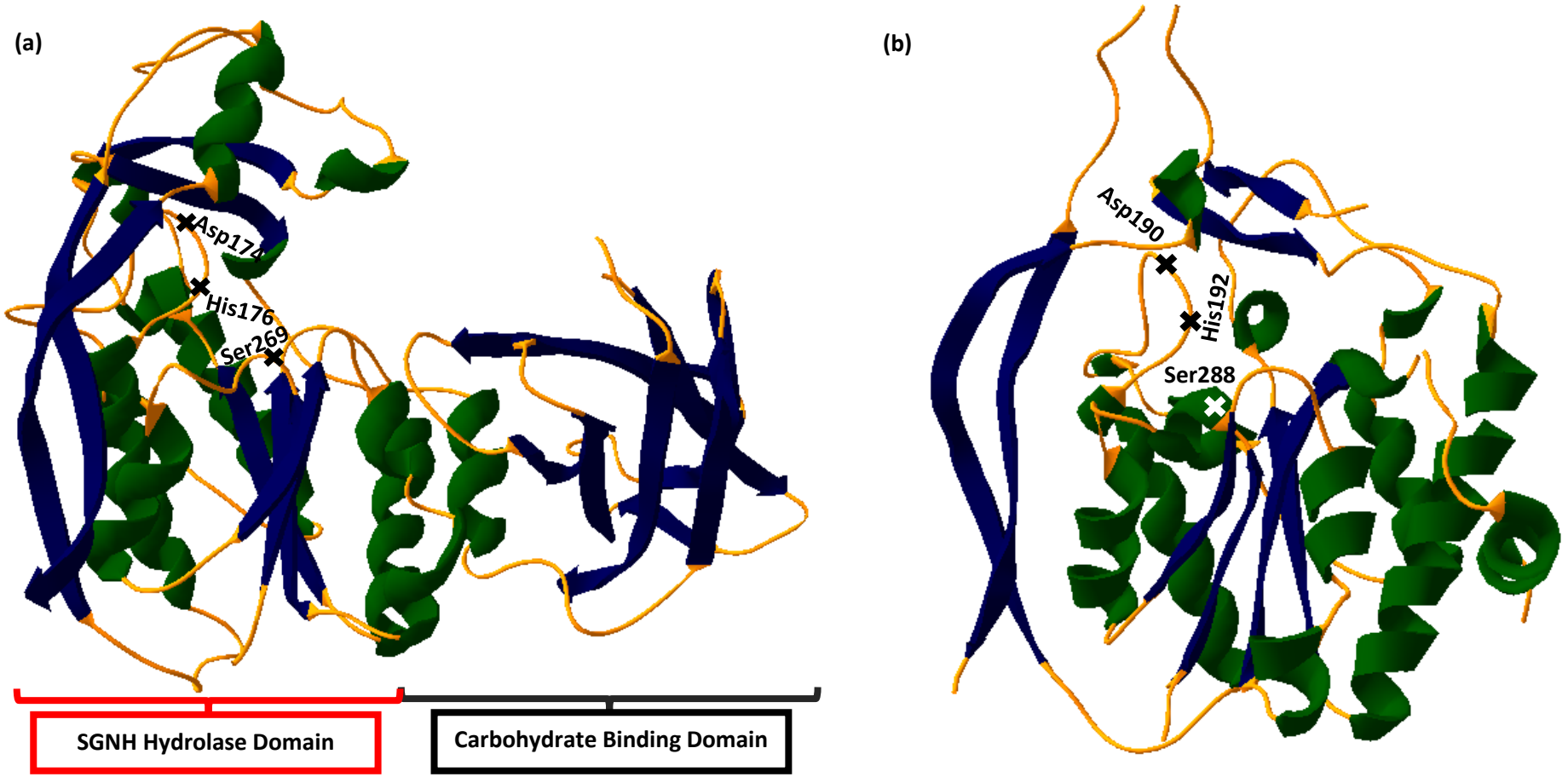


Figure 5.2.2. Crystal structure of (a) AlgX of *P. aeruginosa* PAO1 (PDB: 4KNC), and (b) AlgJ of *P. putida* (PDB: 4O8V). The two domains of AlgX were identified and named using either the red or black boxes. α-helices are coloured in dark green, β-strands are coloured blue and coils are coloured orange. Amino acids of catalytic triads are identified by crosses; black crosses are for amino acids that are found on coils and the white crosses are for amino acids that are found on the α-helices and β-strands. Amino acid names and positions are also included with their respective crosses.

The wild type AlgX possessed an activity value of 0.0068 ± 0.00057 µmol min^{-1} mg^{-1}. The mutant AlgX with the D174A mutation had an activity value of 0.00032 ± 0.00028 µmol min^{-1} mg^{-1} while that with the H176A mutation had an activity value of 0.00027 ± 0.00027 µmol min^{-1} mg^{-1} which were 4.71% and 0.397% of the wild type AlgX activity respectively. Acetylesterase activity of the mutant AlgX with the S269A replacement was absent. Circular dichroism analysis revealed that these mutant proteins retained the structure of AlgX despite the various introduced mutations. Hence, the decrease or absence of the acetylesterase activity measured would be due to the absence of these catalytic amino acids and not structural alterations.

The identified catalytic residues of *P. aeruginosa* AlgJ, a functionally similar protein to AlgX but without the carbohydrate binding domain, which are Ser288, H192 and D190 were also separately replaced by alanine and the mutants were analysed by acetylesterase activity using the pseudosubstrate, 3-carboxyumbelliferyl acetate (Baker et al., 2014). However, the esterase function for each of the mutant proteins retained approximately 20% of the function as compared to the wild type AlgJ. This was significantly more than the respective mutations in AlgX as mentioned above. Due to the recalcitrance of *P. aeruginosa* AlgJ to crystallisation for X-ray crystallography, the orthologous domain from *Pseudomonas putida* AlgJ was crystallised in its place (PDB: 4O8V) and was depicted on Figure 5.2.2(b) with the corresponding catalytic residues.

5.2.4 Neisseria gonorrhoeae - PatB

The peptidoglycan layer of bacteria in the cell wall and the various modifications involved are essential for their survival (Boneca, 2005). O-acetylation of the peptidoglycan is a common form of modification. This process is carried out by different groups of enzymes and in *N. gonorrhoeae*, PatA and PatB are responsible for the modification of the peptidoglycan.

As mentioned in Section 3.2.4.1, PatA is a member of the membrane bound O-acetyltransferase (MBOAT) family of enzymes while PatB is thought to be a member of the SGNH hydrolase type esterase family of enzymes.

The predicted structure of PatB was determined, and for this project was repeated, by using the known structure of the SGNH hydrolase type esterase *Sacchromyces cerevisiae* isoamyl acetate-hydrolysing esterase (PDB: 3MIL) as a template using SWISS-MODEL server (http://swissmodel.expasy.org/) (Arnold et al., 2006, Guex et al., 2009, Kiefer et al., 2009, Schwede et al., 2003). The results of the structure prediction of PatB was visualised on the Swiss-PdbViewer 4.1.0 (Guex and Peitsch, 1997) and depicted on Figure 5.2.3 with the relative positions of the putative catalytic residues of the protein.

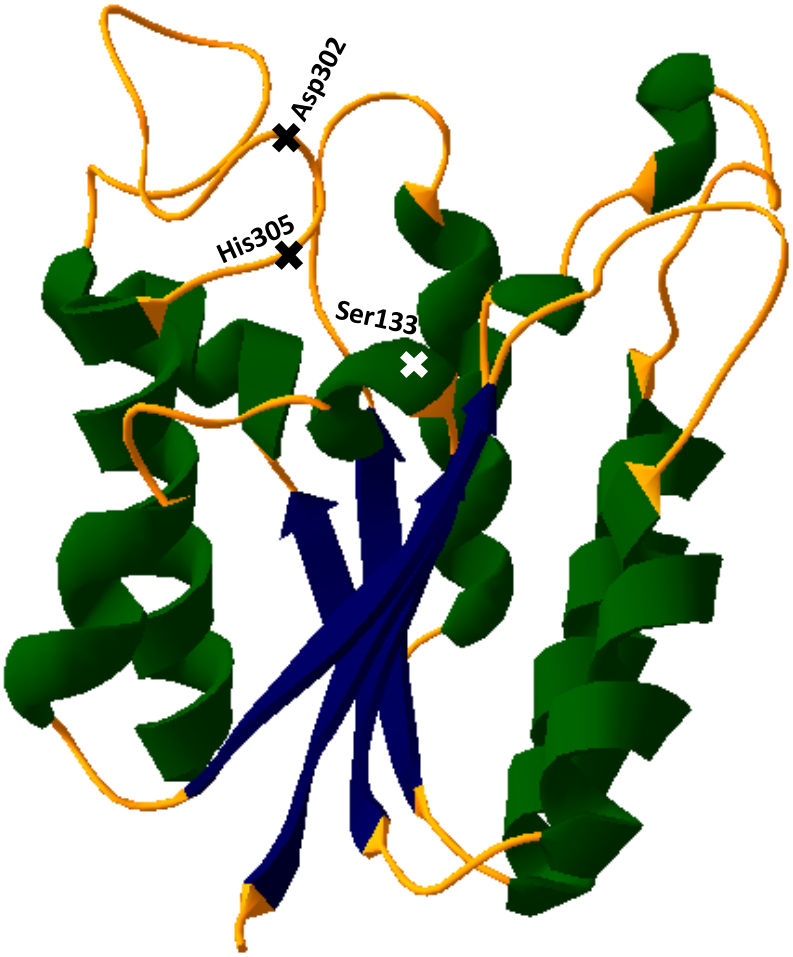


Figure 5.2.3. Predicted crystal structure PatB of *N. gonorrhoeae*. Prediction performed using isoamyl acetate-hydrolysing esterase of *Sacchromyces cerevisiae* (PDB: 3MIL) as a template. α-helices are coloured in dark green, β-strands are coloured blue and coils are coloured orange. Amino acids of predicted catalytic triads are identified by crosses; black crosses are for amino acids that are found on coils and the white crosses are for amino acids that are found on the α-helices. Amino acid names and positions are also included with their respective crosses.

The putative catalytic residues of Ser133, Asp302 and His305 were identified (Moynihan and Clarke, 2014) and their importance were analysed by separate alanine replacement mutations followed by steady state kinetic analysis of the determination of the activity of resulting mutant proteins. The activity of the mutant proteins with the S133A or H305A mutations were completely abolished while that with the D302A mutation retained 20% of the activity with respect to the activity profile of the wild type PatB protein. To ensure that no structural changes occurred as a result of the replacement mutations, circular dichroism analysis of the mutants were compared to that of the wild type and no changes were observed. Hence, the absence or decrease in protein activity was a result of the mutations rather than any structural changes.

5.3 Conclusion – Serine Esterase Catalytic Triad Analysis of OafA and GtrC^{BTP1}

The analysis of the serine esterase catalytic triad residues of confirmed or putative SGNH hydrolases as discussed in Sections 5.2.1 to 5.2.4 shows that replacing these catalytic serine or histidine residues result in the significant abrogation of enzymatic function of the enzyme while the effects of replacing the aspartate on the enzymatic function were less impactful. This, along with the fact that NanS of *E. coli* O157:H7 only possesses the catalytic dyad of serine and histidine, suggests that the aspartate residue, while potentially important, is not necessary for enzymatic function.

As discussed above and in the introduction, GtrC^{BTP1} is known to acetylate the rhamnose residue of the LPS O-antigen of *S.* Typhimurium D23580 and hence, protecting the bacteria from BTP1 phage associated killing. The *in vivo* phage plaque assay demonstrated that the catalytic triad of GtrC^{BTP1} is essential for the function of the protein in the protection against BTP1 associated lysis. The BTP1 phage plaque assay proved to be a useful tool in determining if GtrC^{BTP1} and its mutant variants are able to carry out the primary function or not. If the catalytic triad analysis of OafA were to be performed, the results are hypothesised to be consistent to GtrC^{BTP1} as *in silico* analyses revealed high levels of similarities between these proteins.

Chapter 6

Discussion

6.1 *Salmonella enterica* (subspecies I) Lipopolysaccharide (LPS) O-antigen Modifications - OafA and GtrC^{BTP1}

Gtr family II of the newly identified invasive non-typhoidal *Salmonella* (iNTS), *S.* Typhimurium D23580, was shown to be an O-antigen acetyltransferase responsible for the addition of an acetyl group onto the rhamnose residue of the LPS O-antigen backbone (Erica Kintz, Personal Communication). In this strain, the family II *gtr* operon is found on the BTP1 prophage. Like the glucosylation by the family I *gtr operon*, the acetylation of the rhamnose residue of the *S.* Typhimurium D23580 LPS O-antigen provides protection for the bacteria against BTP1 bacteriophage infection by blocking the endorhamnosidase activity of the phage tail spike protein. GtrC^{BTP1} was found to share similar properties to a known *S.* Typhimurium LPS O-antigen acetyltransferase, OafA. They were both predicted to have an inner membrane bound domain and a periplasmic C-terminal domain which consists of close to half the total number of amino acids that make up the proteins. OafA acetylates the abequose residue on the LPS O-antigen of *S.* Typhimurium which gives the O:5 serotype.

Predicted structural similarities and subtle functional differences between OafA and GtrC family II, represented by GtrC^{BTP1}, calls for the need to understand these proteins further and attempt to identify the basis for their functional difference.

6.2 Characterisation of OafA and GtrC^{BTP1}

As discussed in Chapter 3, *in silico* analyses of these proteins confirmed that OafA, GtrC^{BTP1}, and in fact, other GtrC family II proteins have a similar predicted two domain overall structure of an N-terminal inner membrane bound domain and a periplasmic C-terminal domain.

To further analyse these proteins, their various functional domains were predicted using the InterProScan tool (Jones et al., 2014, Hunter et al., 2011). The same results returned for OafA and the GtrC family II proteins with functional families identified for each of the two predicted domains. The N-terminal inner membrane bound domains were predicted to be an acyltransferase 3 (Pfam: PF01757), which is a largely uncharacterised domain family, and the C-terminal periplasmic domains were predicted to be from the SGNH hydrolase type esterase family (SUPERFAMILY: SSF52266). SGNH hydrolase proteins are known to have a distinct α/β/α fold and a serine esterase catalytic triad of serine, aspartate and histidine with the latter two being characteristically at a maximum of three amino acids apart from one another as shown in Figures 3.4.2, 5.1.1 and 5.1.2. The same was predicted for OafA and GtrC^{BTP1} as shown in Figure 3.4.1. These findings suggest a myriad of characterisation experiments which could be performed to further understand the mechanism behind the function of these proteins. This includes ligand and substrate binding experiments, domain complementation experiments,

structural characterisation along with multiple other biochemical and biophysical experiments. Biochemical and biophysical experiments would require pure samples of these proteins.

SGNH hydrolase type esterases have a distinct serine esterase catalytic triad as mentioned in Chapter 3 and Section 6.2.1 and was predicted to be present in both $GtrC^{BTP1}$ and OafA. Prior studies on various other proteins involved in acetyltransferase or acetylesterase activities have suggested that this catalytic triad is necessary for protein function. To determine if these catalytic residues are necessary in $GtrC^{BTP1}$, mutant proteins were engineered and each of them possessed either the S430A, D618A or H621A mutations. The function of these proteins were tested using the phage plaque assay discussed in Section 5.1.2.2. As shown on Figure 5.2.3(d) to (g), the mutant $GtrC^{BTP1}$ failed to retain the function of the proteins despite the presence of a single specific replacement mutation as plaque formation and spots of bacterial lysis were observed across all phage dilutions. The results from this study on the catalytic residues of $GtrC^{BTP1}$ were similar to that on other studies performed on other SGNH hydrolase proteins mentioned in Sections 5.2.1 to 5.2.4 which suggests that these proteins are very likely to possess the SGNH hydrolase domain with the characteristic catalytic triad.

Various biochemical or biophysical *in vitro* methods requiring sample of purified proteins were utilised in the previous studies on other SGNH hydrolase proteins mentioned in Sections 5.2.1 to 5.2.4. Due to the fact that the SGNH hydrolase portion of both OafA and $GtrC^{BTP1}$ is attached to a membrane bound O-acetyltransferase domain, performing the more common *in vitro* methods might be more cumbersome. This is because the protein yield from the purification of membrane bound proteins is significantly lower than that of soluble proteins. This would require additional expression optimisation to increase the yield and detergent tests to find a more biochemical analysis friendly detergent. Instead, *in vivo* strategies like the BTP1 phage plaque assay and the LPS analysis by western blotting are available for use to analyse the function of the mutant OafA and $GtrC^{BTP1}$ proteins respectively. Unlike *in vitro* tests, this allows for a more biologically relevant functional analyses of these proteins within a native environment.

Previous studies also showed that about 20% of the function was retained for the mutant proteins possessing the catalytic aspartate replacement mutations. Hence, it was hypothesised that $GtrC^{BTP1}$ possessing the D618A mutations would also retain some functional activity especially if lesser amounts of phage is present for infection (i.e. spots with lower concentrations of phage). However, the extent of phage activity was similar throughout the *S.* Typhimurium LT2 strains expressing each of the three mutant proteins mentioned above. One reason for this could be that the phage dilutions were still too high at 10^{-8} of the original phage stock. Another reason for this phenomenon could be that the decreased rhamnose acetylation, if any, were only present on a few LPS molecules and not randomly spread out between all LPS molecules. This allows the BTP1 phage to perform its endorhamnosidase activity to cleave the unmodified LPS

in order to gain access to the cell surface. If the modification was spread out randomly, some form of protection against the BTP1 phage endorhamnosidase activity could still be present. Finally, unlike the other SGNH hydrolase proteins mentioned in previous studies, Asp618 could be essential for the function of $GtrC^{BTP1}$.

6.2.1 Putative Mechanism of Action

With the current understanding of OafA and $GtrC^{BTP1}$, including findings from this study, it is hypothesised that an acetyl group from a source, likely to be acetyl-CoA, would be transported from the cytoplasm to the periplasmic C-terminal tail domain via the acyltransferase activity of the N-terminal domain. The acetyl group would then be transported to the LPS O-antigen at an unknown stage of the LPS biosynthesis via the activity of the serine esterase catalytic triad of serine, aspartate and histidine using an unidentified sugar binding motif which contributes to the specificity of these proteins (Figure 6.3.1). To date, the addition of the acetyl groups and in fact, the glucose to the specific sugar residues occur at an unknown stage of the LPS biosynthesis pathway. In addition, it is not definitively known if the modifications occur on all LPS molecules.

6.3 Future Experiments

Further analysis would be required in order to determine the complete mechanism of action and characteristics of these acetyltransferases. In this section, a few unanswered questions were raised and experiments were suggested.

6.3.1 LPS O-antigen binding sites

Firstly, the O-antigen binding sites of the proteins are still unknown. One method to determine where these sites could be is by extensive search in the current literature for proteins which are known to bind to these sugars to determine the common sequential, chemical or structural characteristics that could contribute to the sugar binding. Once identified, mutations can be performed on the respective sites and the functions of the proteins can be tested using the *in vivo* tests mentioned in Section 5.1. Another method is to narrow down the region of amino acid residues which possesses the sugar binding residues by performing a protein domain complementation which involves engineering a protein with the N-terminal acyltransferase 3 domain of either OafA or $GtrC^{BTP1}$ and the C-terminal SGNH hydrolase domain of the other. The function of these hybrid proteins can be tested using the same *in vivo* tests mentioned in Section 5.1. If the functions are retained despite having a C terminal domain that is from the other protein, the sugar binding site should not be within the SGNH hydrolase C terminal domain.

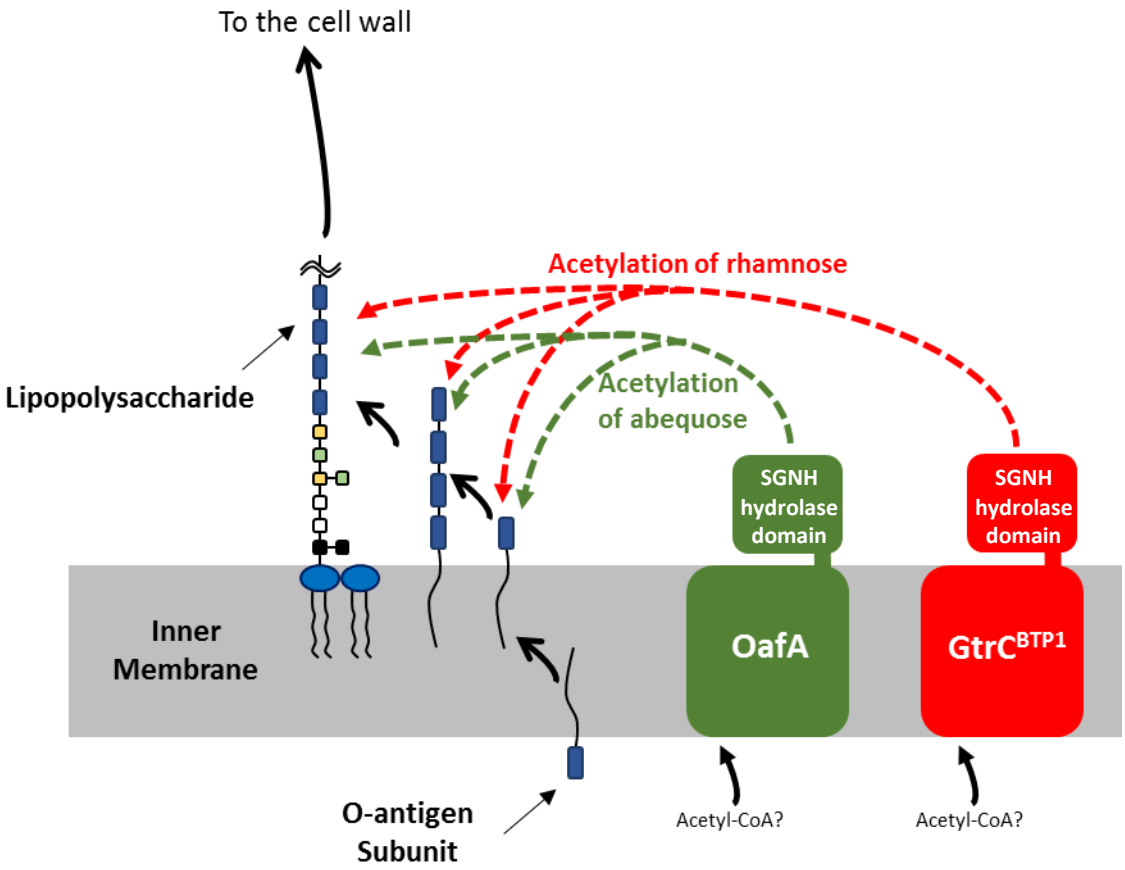


Figure 6.2.1. Proposed mechanism of OafA and GtrC^{BTP1}. A general lipopolysaccharide (LPS) biosynthesis pathway is shown using the black arrows with an open head. For both acetyltransferases, the acetyl group from an source (suggested as Acetyl-CoA on the diagram) is transported through the inner membrane via the membrane bound acyltransferase 3 domain and passed through to the periplasmic C-terminal SGNH hydrolase domain. Using the catalytic triad within these proteins, the acetyl group is then transported to the specific sugar residues at an unknown stage of the LPS biosynthetic pathway as shown by the red and green dashed arrows.

6.3.2 Serine esterase catalytic residues of OafA

As mentioned in Chapter 5 and Section 6.2.3, the catalytic triad of GtrC^{BTP1} were mutated which resulted in the loss of protein function. A similar approach to OafA should be taken in order to analyse the importance of the putative catalytic residues. Due to the similarities of OafA and GtrC^{BTP1}, it was hypothesised that mutating the catalytic residues in OafA using replacement mutations would also abolish protein function.

6.3.3 Biochemical and biophysical analyses of OafA and GtrC^{BTP1}

Biochemical and biophysical analyses of these proteins of interest would reveal different molecular and physical characteristics which would then clarify the underlying functional mechanisms of these proteins. Biochemical experiments such as size exclusion chromatography with multi angle light scattering (SEC-MALS) (Striegel et al., 2009) for the detection of the absolute molecular mass of these proteins (i.e. whether they are monomeric or otherwise) and saturation transfer difference nuclear magnetic resonance (STD-NMR) for the identification of potential ligands of these proteins can be performed (Viegas et al., 2011). The structure of these proteins can also be determined using X-ray crystallography or NMR.

However, these biochemical and biophysical experiments require pure samples of proteins in order to get accurate results. Hence, the expression of OafA and GtrC^{BTP1} should be further optimised based on the current expression conditions mentioned in Chapter 4 in order to get a sufficiently high yield of proteins.

6.3.4 Other GtrC family II proteins

Among GtrC family II proteins, GtrC^{BTP1} is the least similar at approximately 77% amino acid conservation with each of the rest of the members of the family (Table 3.1.2). Nevertheless, the characteristics of GtrC^{BTP1} should be parallel with that of the other members of the GtrC family II. To determine if this holds true, the various proteins within GtrC family II can be complemented into the *S.* Typhimurium LT2 strain and the resulting clones expressing the various proteins can be tested using the BTP1 phage assay as per the method mentioned in Section 5.2.1. The proteins which provide protection of the *S.* Typhimurium LT2 strain against BTP1 phage activity would suggest that these orthologous proteins are functionally the same and hence, should share multiple characteristics with GtrC^{BTP1}.

6.3.5 Oac, LPS acetyltransferase of Shigella flexneri

Oac is an LPS acetyltransferase form *S. flexneri* which, like GtrC^{BTP1}, acetylates a rhamnose in the LPS (Thanweer and Verma, 2012). However, Oac is only a single domain membrane bound protein of the acyltransferase 3 family and hence, does not possess an SGNH hydrolase domain. If this modification system utilises a separate periplasmic protein like what PatB is to PatA in

Neisseria, it has yet to be identified be it an SGNH hydrolase family protein or of a different family. This protein could also be functional on its own unlike the various proteins mentioned in this thesis including OafA and GtrC^{BTP1}.

The diversity in LPS acetyltransferase systems calls for the need to further understand the proteins involved and what sets each system or individual proteins apart.

6.4 Closing Remarks

Equipped with biologically relevant functional assays, the findings from this project has provided a deeper understanding of the structural and functional characteristics of OafA and GtrC^{BTP1} and yet, it has also suggested more experiments which could be perform to build upon the current knowledge of these proteins. To date, not much is known about the mechanism of LPS O-antigen acetyltransferases, be it in *Salmonella* or other genus of Gram-negative bacteria. Hence, the results from this project and future work could go on to characterising other LPS O-antigen acetyltransferases and perhaps, identify a new family of LPS modification enzymes.

List of Abbreviations

cmc	Critical micellar concentration
DDM	*n*-dodecyl- β-D-maltoside
EDTA	Ethylenediaminetetraacetic acid
IMAC	Immobilised metal affinity chromatography
kDa	Kilodalton
KPi	Potassium phosphate
LB	Lysogeny broth/Lennox broth
LIC	Ligation independent cloning
LPS	Lipopolysaccharide
PBS	Phosphate buffered saline
PVDF	Polyvinylidene fluoride
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SOC	Super optimal broth with catabolite repression
STM	*Salmonella enterica* subspecies *enterica* serovar Typhimurium
TAE	Tris-acetate EDTA
TEMED	N,N,N',N'-Tetramethylethylenediamine
v/v	Volume per volume
w/v	Weight per volume

List of References

ALLISON, G. E. & VERMA, N. K. 2000. Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri. *Trends Microbiol,* 8**,** 17-23.

ALMAGRO-MORENO, S. & BOYD, E. F. 2009. Insights into the evolution of sialic acid catabolism among bacteria. *BMC evolutionary biology,* 9**,** 118.

ARNOLD, K., BORDOLI, L., KOPP, J. & SCHWEDE, T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics,* 22**,** 195-201.

BAKER, P., RICER, T., MOYNIHAN, P. J., KITOVA, E. N., WALVOORT, M. T., LITTLE, D. J., WHITNEY, J. C., DAWSON, K., WEADGE, J. T. & ROBINSON, H. 2014. P. aeruginosa SGNH Hydrolase-Like Proteins AlgJ and AlgX Have Similar Topology but Separate and Distinct Roles in Alginate Acetylation. *PLoS pathogens,* 10**,** e1004334.

BERA, A., HERBERT, S., JAKOB, A., VOLLMER, W. & GÖTZ, F. 2005. Why are pathogenic staphylococci so lysozyme resistant? The peptidoglycan O-acetyltransferase OatA is the major determinant for lysozyme resistance of Staphylococcus aureus. *Molecular microbiology,* 55**,** 778-787.

BISPHAM, J., TRIPATHI, B., WATSON, P. & WALLIS, T. 2001. Salmonella Pathogenicity Island 2 Influences Both Systemic Salmonellosis andSalmonella-Induced Enteritis in Calves. *Infection and immunity,* 69**,** 367-377.

BOGOMOLNAYA, L. M., SANTIVIAGO, C. A., YANG, H.-J., BAUMLER, A. J. & ANDREWS-POLYMENIS, H. L. 2008. 'Form variation' of the O12 antigen is critical for persistence of Salmonella Typhimurium in the murine intestine. *Molecular Microbiology,* 70**,** 1105-1119.

BONECA, I. G. 2005. The role of peptidoglycan in pathogenesis. *Current opinion in microbiology,* 8**,** 46-53.

CASADABAN, M. J. & COHEN, S. N. 1980. Analysis of gene control signals by DNA fusion and cloning in< i> Escherichia coli</i>. *Journal of molecular biology,* 138**,** 179-207.

CDC.GOV (2015). *Serotypes and the Importance of Serotyping Salmonella | Salmonella Atlas | Reports and Publications | Salmonella | CDC*. [online] Available at: http://www.cdc.gov/salmonella/reportspubs/salmonella-atlas/serotyping-importance.html [Accessed 12 Feb. 2015]

CHOU, C. P. 2007. Engineering cell physiology to enhance recombinant protein production in Escherichia coli. *Applied microbiology and biotechnology,* 76**,** 521-532.

COBURN, B., GRASSL, G. A. & FINLAY, B. 2007. Salmonella, the host and disease: a brief review. *Immunology and cell biology,* 85**,** 112-118.

COBURN, B., LI, Y., OWEN, D., VALLANCE, B. A. & FINLAY, B. B. 2005. Salmonella enterica serovar Typhimurium pathogenicity island 2 is necessary for complete virulence in a mouse model of infectious enterocolitis. *Infection and immunity,* 73**,** 3219-3227.

COOMBES, B. K., COBURN, B. A., POTTER, A. A., GOMIS, S., MIRAKHUR, K., LI, Y. & FINLAY, B. B. 2005. Analysis of the contribution of Salmonella pathogenicity islands 1 and 2 to enteric disease progression using a novel bovine ileal loop model and a murine model of infectious enterocolitis. *Infection and immunity,* 73**,** 7161-7169.

DAVIES, M. R., BROADBENT, S. E., HARRIS, S. R., THOMSON, N. R. & VAN DER WOUDE, M. W. 2013. Horizontally acquired glycosyltransferase operons drive Salmonellae lipopolysaccharide diversity. *PLoS genetics,* 9.

DONG, H., NILSSON, L. & KURLAND, C. G. 1995. Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. *Journal of bacteriology,* 177**,** 1497-1504.

DRUMMOND, A., ASHTON, B., BUXTON, S., CHEUNG, M., COOPER, A., DURAN, C. & FIELD, M. 2012. Geneious v5. 6. Created by Biomatters. *See http://www. geneious. com*.

ERRIDGE, C., BENNETT-GUERRERO, E. & POXTON, I. R. 2002. Structure and function of lipopolysaccharides. *Microbes and infection,* 4**,** 837-851.

FIERER, J. & GUINEY, D. G. 2001. Diverse virulence traits underlying different clinical outcomes of Salmonella infection. *Journal of Clinical Investigation,* 107**,** 775.

FOOKES, M., SCHROEDER, G. N., LANGRIDGE, G. C., BLONDEL, C. J., MAMMINA, C., CONNOR, T. R., SETH-SMITH, H., VERNIKOS, G. S., ROBINSON, K. S. & SANDERS, M. 2011. Salmonella bongori provides insights into the evolution of the Salmonellae. *PLoS pathogens,* 7**,** e1002191.

GAL-MOR, O., BOYLE, E. C. & GRASSL, G. A. 2014. Same species, different diseases: how and why typhoidal and non-typhoidal Salmonella enterica serovars differ. *Frontiers in microbiology,* 5.

GALAN, J. E. & CURTISS, R. 1989. Cloning and molecular characterization of genes whose products allow Salmonella typhimurium to penetrate tissue culture cells. *Proceedings of the National Academy of Sciences,* 86**,** 6383-6387.

GEERTSMA, E. R. & POOLMAN, B. 2007. High-throughput cloning and expression in recalcitrant bacteria. *Nature methods,* 4**,** 705-707.

GIBBONS, H. S., LIN, S., COTTER, R. J. & RAETZ, C. R. 2000. Oxygen Requirement for the Biosynthesis of theS-2-Hydroxymyristate Moiety in Salmonella typhimurium Lipid A FUNCTION OF LpxO, A NEW Fe2+/α-KETOGLUTARATE-DEPENDENT DIOXYGENASE HOMOLOGUE. *Journal of Biological Chemistry,* 275**,** 32940-32949.

GILL, S. C. & VON HIPPEL, P. H. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Analytical biochemistry,* 182**,** 319-326.

GORDON, M. A. 2011. Invasive Non-typhoidal Salmonella Disease–epidemiology, pathogenesis and diagnosis. *Current opinion in infectious diseases,* 24**,** 484.

GRIMONT, P. A. & WEILL, F.-X. 2007. Antigenic formulae of the Salmonella serovars. *WHO Collaborating Centre for Reference and Research on Salmonella, Institut Pasteur, Paris, France*.

GRISSHAMMER, R. & TATEU, C. G. 1995. Overexpression of integral membrane proteins for structural studies. *Quarterly Reviews of Biophysics,* 28**,** 315-422.

GUEX, N. & PEITSCH, M. C. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *electrophoresis,* 18**,** 2714-2723.

GUEX, N., PEITSCH, M. C. & SCHWEDE, T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis,* 30**,** S162-S173.

GUO, L., LIM, K. B., PODUJE, C. M., DANIEL, M., GUNN, J. S., HACKETT, M. & MILLER, S. I. 1998. Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell,* 95**,** 189-198.

GUZMAN, L.-M., BELIN, D., CARSON, M. J. & BECKWITH, J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology,* 177**,** 4121-4130.

HELLERQVIST, C. G., LINDBERG, B., SVENSSON, S., HOLME, T. & LINDBERG, A. A. 1968. Structural studies on the O-specific side-chains of the cell-wall lipopolysaccharide from< i> Salmonella typhimurium</i> 395 ms. *Carbohydrate Research,* 8**,** 43-55.

HÖLTJE, J.-V. 1998. Growth of the stress-bearing and shape-maintaining murein sacculus of Escherichia coli. *Microbiology and Molecular Biology Reviews,* 62**,** 181-203.

HUANG, D. B. & DUPONT, H. L. 2005. Problem pathogens: extra-intestinal complications of Salmonella enterica serotype Typhi infection. *The Lancet infectious diseases,* 5**,** 341-348.

HUCKER, G. J. & CONN, H. J. 1923. Methods of Gram staining.

HUNKE, S. & BETTON, J. M. 2003. Temperature effect on inclusion body formation and stress response in the periplasm of Escherichia coli. *Molecular microbiology,* 50**,** 1579-1589.

HUNTER, S., JONES, P., MITCHELL, A., APWEILER, R., ATTWOOD, T. K., BATEMAN, A., BERNARD, T., BINNS, D., BORK, P. & BURGE, S. 2011. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic acids research***,** gkr948.

JONES, P., BINNS, D., CHANG, H.-Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A. & NUKA, G. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics,* 30**,** 1236-1240.

KALYNYCH, S., MORONA, R. & CYGLER, M. 2014. Progress in understanding the assembly process of bacterial O-antigen. *FEMS microbiology reviews,* 38**,** 1048-1065

KAUFFMANN, F. 1966. The Bacteriology of Enterobacteriaceae. Collected Studies of the Author and his Co-Workers. *The bacteriology of enterobacteriaceae. Collected studies of the author and his co-workers.*

KIEFER, F., ARNOLD, K., KÜNZLI, M., BORDOLI, L. & SCHWEDE, T. 2009. The SWISS-MODEL Repository and associated resources. *Nucleic acids research,* 37**,** D387-D392.

KINGSLEY, R. A., MSEFULA, C. L., THOMSON, N. R., KARIUKI, S., HOLT, K. E., GORDON, M. A., HARRIS, D., CLARKE, L., WHITEHEAD, S. & SANGAL, V. 2009. Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome research,* 19**,** 2279-2287.

KINTZ, E., DAVIES, M., HAMMARLÖF, D. L., CANALS, R., HINTON, J. C. & VAN DER WOUDE, M. 2015. A BTP1 prophage gene present in invasive non-typhoidal Salmonella determines composition and length of the O-antigen of the LPS. *Molecular microbiology*.

KROGH, A., LARSSON, B., VON HEIJNE, G. & SONNHAMMER, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology,* 305**,** 567-580.

LE MAIRE, M., CHAMPEIL, P. & MØLLER, J. V. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta (BBA) - Biomembranes,* 1508**,** 86-111.

LEE, L., LEE, Y., LEU, R. & SHAW, J. 2006. Functional role of catalytic triad and oxyanion hole-forming residues on enzyme activity of Escherichia coli thioesterase I/protease I/phospholipase L1. *Biochem. J,* 397**,** 69-76.

LIU, B., KNIREL, Y. A., FENG, L., PEREPELOV, A. V., SENCHENKOVA, S. Y. N., REEVES, P. R. & WANG, L. 2014. Structural diversity in Salmonella O antigens and its genetic basis. *FEMS Microbiology Reviews,* 38**,** 56-89.

MICHETTI, P., MAHAN, M., SLAUCH, J., MEKALANOS, J. & NEUTRA, M. 1992. Monoclonal secretory immunoglobulin A protects mice against oral challenge with the invasive pathogen Salmonella typhimurium. *Infection and Immunity,* 60**,** 1786-1792.

MICHETTI, P., PORTA, N., MAHAN, M. J., SLAUCH, J. M., MEKALANOS, J. J., BLUM, A., KRAEHENBUHL, J.-P. & NEUTRA, M. R. 1994. Monoclonal immunoglobulin A prevents adherence and invasion of polarized epithelial cell monolayers by Salmonella typhimurium. *Gastroenterology,* 107**,** 915-923.

MICROBIOLOGY, S. S. O. T. N. C. O. T. I. S. F. 1934. The genus Salmonella lignieres, 1900. *The Journal of hygiene,* 34**,** 333.

MILLER, S. I., ERNST, R. K. & BADER, M. W. 2005. LPS, TLR4 and infectious disease diversity. *Nature Reviews Microbiology,* 3**,** 36-46.

MILLS, D. M., BAJAJ, V. & LEE, C. A. 1995. A 40 kb chromosomal fragment encoding Salmonella typhimurium invasion genes is absent from the corresponding region of the Escherichia coli K-12 chromosome. *Molecular microbiology,* 15**,** 749-759.

MIROUX, B. & WALKER, J. E. 1996. Over-production of Proteins in< i> Escherichia coli</i>: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *Journal of molecular biology,* 260**,** 289-298.

MØLGAARD, A., KAUPPINEN, S. & LARSEN, S. 2000. Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases. *Structure,* 8**,** 373-383.

MORIN, R. B. & GORMAN, M. 2014. *The Biology of B-Lactam Antibiotics*, Elsevier.

MORAN, A. P., PRENDERGAST, M. M., & APPELMELK, B. J. 1996. Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease. *FEMS immunology and medical microbiology*, *16*(2), 105-115.

MOYNIHAN, P. J. & CLARKE, A. J. 2010. O-Acetylation of Peptidoglycan in Gram-negative Bacteria IDENTIFICATION AND CHARACTERIZATION OF PEPTIDOGLYCAN O-ACETYLTRANSFERASE IN NEISSERIA GONORRHOEAE. *Journal of Biological Chemistry,* 285**,** 13264-13273.

MOYNIHAN, P. J. & CLARKE, A. J. 2014. The Mechanism of Action of Peptidoglycan O-Acetyltransferase B Involves a Ser-His-Asp Catalytic Triad. *Biochemistry*.

NIVENS, D. E., OHMAN, D. E., WILLIAMS, J. & FRANKLIN, M. J. 2001. Role of alginate and its O acetylation in formation of Pseudomonas aeruginosa microcolonies and biofilms. *Journal of bacteriology,* 183**,** 1047-1057.

OLLIS, D. L., CHEAH, E., CYGLER, M., DIJKSTRA, B., FROLOW, F., FRANKEN, S. M., HAREL, M., REMINGTON, S. J., SILMAN, I. & SCHRAG, J. 1992. The α/β hydrolase fold. *Protein Engineering,* 5**,** 197-211.

OSTERMEIER, C. & MICHEL, H. 1997. Crystallization of membrane proteins. *Current opinion in structural biology,* 7**,** 697-701.

PEDERSEN, S. S., KHARAZMI, A., ESPERSEN, F. & HØIBY, N. 1990. Pseudomonas aeruginosa alginate in cystic fibrosis sputum and the inflammatory response. *Infection and immunity,* 58**,** 3363-3368.

PEDULLA, M. L., FORD, M. E., KARTHIKEYAN, T., HOUTZ, J. M., HENDRIX, R. W., HATFULL, G. F., POTEETE, A. R., GILCREASE, E. B., WINN-STAPLEY, D. A. & CASJENS, S. R. 2003. Corrected sequence of the bacteriophage P22 genome. *Journal of bacteriology,* 185**,** 1475-1477.

PINSACH, J., DE MAS, C., LÓPEZ-SANTÍN, J., STRIEDNER, G. & BAYER, K. 2008. Influence of process temperature on recombinant enzyme activity in< i> Escherichia coli</i> fed-batch cultures. *Enzyme and Microbial Technology,* 43**,** 507-512.

PRIVÉ, G. G. 2007. Detergents for the stabilization and crystallization of membrane proteins. *Methods,* 41**,** 388-397.

RAETZ, C. R. & WHITFIELD, C. 2002. Lipopolysaccharide endotoxins. *Annual review of biochemistry,* 71**,** 635.

RANGARAJAN, E. S., RUANE, K. M., PROTEAU, A., SCHRAG, J. D., VALLADARES, R., GONZALEZ, C. F., GILBERT, M., YAKUNIN, A. F. & CYGLER, M. 2011. Structural and enzymatic characterization of NanS (YjhS), a 9-O-Acetyl N-acetylneuraminic acid esterase from Escherichia coli O157: H7. *Protein Science,* 20**,** 1208-1219.

RATH, A., GLIBOWICKA, M., NADEAU, V. G., CHEN, G. & DEBER, C. M. 2009. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences,* 106**,** 1760-1765.

RILEY, L. M., WEADGE, J. T., BAKER, P., ROBINSON, H., CODÉE, J. D., TIPTON, P. A., OHMAN, D. E. & HOWELL, P. L. 2013. Structural and Functional Characterization of Pseudomonas aeruginosa AlgX ROLE OF AlgX IN ALGINATE ACETYLATION. *Journal of Biological Chemistry,* 288**,** 22299-22314.

SCHWEDE, T., KOPP, J., GUEX, N. & PEITSCH, M. C. 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic acids research,* 31**,** 3381-3385.

SEDDON, A. M., CURNOW, P. & BOOTH, P. J. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes,* 1666**,** 105-117.

SEVERI, E., HOOD, D. W. & THOMAS, G. H. 2007. Sialic acid utilization by bacterial pathogens. *Microbiology,* 153**,** 2817-2822.

SIMPSON, J. A., SMITH, S. E. & DEAN, R. T. 1988. Alginate inhibition of the uptake of Pseudomonas aeruginosa by macrophages. *Journal of general microbiology,* 134**,** 29-36.

SLAUCH, J. M., LEE, A. A., MAHAN, M. J. & MEKALANOS, J. J. 1996. 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,* 178**,** 5904-5909.

SLAUCH, J. M., MAHAN, M. J., MICHETTI, P., NEUTRA, M. R. & MEKALANOS, J. J. 1995. Acetylation (O-factor 5) affects the structural and immunological properties of Salmonella typhimurium lipopolysaccharide O antigen. *Infection and immunity,* 63**,** 437-441.

SONNHAMMER, E. L., VON HEIJNE, G. & KROGH, A. A hidden Markov model for predicting transmembrane helices in protein sequences. Ismb, 1998. 175-182.

STEENBERGEN, S. M., JIRIK, J. L. & VIMR, E. R. 2009. YjhS (NanS) is required for Escherichia coli to grow on 9-O-acetylated N-acetylneuraminic acid. *Journal of bacteriology,* 191**,** 7134-7139.

STRIEGEL, A., YAU, W. W., KIRKLAND, J. J. & BLY, D. D. 2009. *Modern size-exclusion liquid chromatography: practice of gel permeation and gel filtration chromatography*, John Wiley & Sons.

SVENSON, S., LÖNNGREN, J., CARLIN, N. & LINDBERG, A. 1979. Salmonella bacteriophage glycanases: endorhamnosidases of Salmonella typhimurium bacteriophages. *Journal of virology,* 32**,** 583-592.

TERNHAG, A., TÖRNER, A., SVENSSON, Å., EKDAHL, K. & GIESECKE, J. 2008. Short-and long-term effects of bacterial gastrointestinal infections. *Emerging infectious diseases,* 14**,** 143.

THANWEER, F. & VERMA, N. K. 2012. Identification of critical residues of the serotype modifying O-acetyltransferase of Shigella flexneri. *BMC biochemistry,* 13**,** 13.

TRENT, M. S., PABICH, W., RAETZ, C. R. & MILLER, S. I. 2001. A PhoP/PhoQ-induced Lipase (PagL) That Catalyzes 3-O-Deacylation of Lipid A Precursors in Membranes ofSalmonella typhimurium. *Journal of Biological Chemistry,* 276**,** 9083-9092.

VANAKEN, T., FOXALL-VANAKEN, S., CASTLEMAN, S. & FERGUSON-MILLER, S. 1985. Alkyl glycoside detergents: synthesis and applications to the study of membrane proteins. *Methods in enzymology,* 125**,** 27-35.

VANDER BYL, C. & KROPINSKI, A. M. 2000. Sequence of the Genome of SalmonellaBacteriophage P22. *Journal of bacteriology,* 182**,** 6472-6481.

VIEGAS, A., MANSO, J., NOBREGA, F. L. & CABRITA, E. J. 2011. Saturation-transfer difference (STD) NMR: a simple and fast method for ligand screening and characterization of protein binding. *Journal of Chemical Education,* 88**,** 990-994.

VIMR, E. & LICHTENSTEIGER, C. 2002. To sialylate, or not to sialylate: that is the question. *Trends in microbiology,* 10**,** 254-257.

WILSON, B. A., SALYERS, A. A., WHITT, D. D. & WINKLER, M. E. 2011. *Bacterial pathogenesis: a molecular approach*, American Society for Microbiology (ASM).

WOLLIN, R., ERIKSSON, U. & LINDBERG, A. A. 1981. Salmonella bacteriophage glycanases: endorhamnosidase activity of bacteriophages P27, 9NA, and KB1. *Journal of virology,* 38**,** 1025-1033.

YOUNG, B. G., FUKAZAWA, Y. & HARTMAN, P. E. 1964. A P22 bacteriophage mutant defective in antigen conversion. *Virology,* 23**,** 279-283.