An Insight into the Function of Family IV GtrABC: Putative Glycosyltransferase Proteins Present in Some Serovars of Non-Typhoidal Salmonella

Grace Taylor-Joyce

MSc by Research

University of York

Biology

January 2016

ABSTRACT

The O-antigen is a component of the LPS, a surface molecule that covers the surface of bacteria including Salmonella. The O-antigen has an important role in infection as it is recognized by the host immune system. Salmonella is a Gram-negative foodborne bacterial pathogen causing inflammation in the gut and diarrhea or an invasive typhoidal disease. The O-antigen of Salmonella has been known to be modified by *gtr* genes, a group of three genes: *gtrA, gtrB* and *gtrC* found within an operon, which have been found to mediate the addition of a glucose molecule on to the repeating unit of the O-antigen. However, the family IV Gtr proteins' function has not been identified but a study has shown that they may be needed for persistence within a host. This project studied the expression of these genes, using a beta-galactosidase assay and RT-qPCR, and looked at their ability to mediate modifications to the *Salmonella* O-antigen using LPS extraction and analysis techniques. The results of this study show expression from transcriptional start sites within the operon suggesting that GtrC(IV) can be expressed independently of the remainder of the operon. The results also show that GtrAB(IV) can replace the function of other GtrAB proteins of known function indicating that they are able to bind and transfer a glucose to a GtrC protein. However, this study was not able to show that GtrC(IV) was targeting the O-antigen, it also found that there is no common sugar amongst the family IV core O-antigens. This suggests that either further unidentified modifications are needed to be made to the O-antigen or the O-antigen is not the target of GtrC(IV).

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Marjan van der Woude, for her support, knowledge and guidance throughout my research degree. I would also like to thank past and present members of the van der Woude lab for all their help and support, they have made my stay in the lab enjoyable and memorable.

I would also like to thank the members of my thesis advisory panel, Pegine Walrad and Paul Pryor, for their opinions and advise through my research.

I would finally like to thank family and friends for their support throughout my studies.

AUTHOR'S DECLARATION

I, Grace Taylor-Joyce, declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

1. INTRODUCTION

1.1 SALMONELLA

Salmonella is a Gram-negative foodborne bacterial pathogen. The genus *Salmonella*, part of the Enterobacteriaceae family is made up of two species: *Salmonella bongori* and *Salmonella enterica*. *S. enterica*, the most well studied *Salmonella* species, contains six subspecies (Coburn et al., 2007). Within Subspecies are serovars, for example *Typhimurium* is a serovar of the subspecies *S. enterica subsp. enterica* and is referred to as *S.* Typhimurium. The enterica subspecies is the most well studied as it contains 99% of all serovars responsible for disease in warm blooded animals (Chan et al., 2003). The serovars in the *enterica* subspecies are grouped into serotypes determined by the structure of their lipopolysaccharide (LPS) and flagella. The serotype is determined by agglutination with specific sera to these two structures (Grimont and Weill, 2007).

The specific symptoms of salmonellosis depend on host, their susceptibility and the infecting serovar. Typhoidal serovars cause an invasive, systemic disease, which is accompanied by an absence of an inflammatory mucosal response in the intestine (Dougan and Baker, 2014). In humans, typhoidal disease is caused by the *S. enterica* serovars *Typi* and *Paratyphi A*. Non-typhoidal serovars are less invasive, causing inflammation in the gut and diarrhea and generally causing a less life threatening disease (Kolodziejek and Miller, 2015). However, in recent years some strains of non-typhoidal *Salmonella* (NTS) have been found to cause a more invasive disease similar to typhoidal serovars in immunocompromised patients in Sub Saharan Africa (Crump and Heyderman, 2015). In humans NTS is caused by multiple serovars of *S. enterica* including *S.* Typhimurium and *S. Enteritidis* (Bangtrakulnonth et al., 2004) However, in mice *S.* Typhimurium causes a typhoidal disease. The main focus if this project was NTS serovars.

Inflammation in the gut, caused by the invasion of epithelial cells by NTS, can be utilized to the advantage of the *Salmonella* pathogen. During inflammation and episodes of diarrhea the gut can become devoid of nutrients (Santos et al., 2009), affecting both the pathogen and commensal gut bacteria. To overcome this problem *S.* Typhimurium can use available

ethanolamine as a nutrient during anaerobic growth. This requires tetrathionate as an electron acceptor, which is a molecule produced by the inflamed gut (Thiennimitr et al., 2011). This metabolic pathway is not available to commensal gut bacteria allowing *Salmonella* a growth advantage, which aids colonization of the gut, persistence and transmission (Rivera-Chavez and Baeumler, 2015).

Salmonella is an intracellular bacterium that can invade host cells such as epithelial cells and macrophages to avoid clearance by the immune system. To facilitate its invasion *Salmonella* utilizes a type three-secretion system (T3SS) needle to transport proteins across a membrane. *S.* Typhimurium has two T3SS contained within *Salmonella* pathogenicity island 1 (SPI-1) or 2 on the chromosome (Kolodziejek and Miller, 2015). The SPI-1 T3SS is used for invading cells and injects effector molecules into a host cell to stimulate its uptake by the cell. Effector molecules released through the T3SS are also important in activating the inflammatory response in the gut needed for colonization of NTS (LaRock et al., 2015).

Once inside the host cell *Salmonella* is contained within a *Salmonella* containing vacuole (SCV), which matures into a unique niche for *Salmonella* to reside within the cell. Effector proteins injected by the SPI-1 T3SS can remain in the invaded cell and aid the formation of the SCV (Hernandez et al., 2004). The SPI-2 T3SS needle is assembled whilst in the SCV and is used by the bacteria to deliver effector proteins into the cytosol of the host cell to promote intracellular survival of the pathogen. As well as the bacteria in the SCV there is a select population of *Salmonella* that are free in the host cell cytosol. These pathogens quickly replicate to fill the host cell with bacteria, which are equipped and ready to invade new host cells (Knodler et al., 2010). Host cells, full with cytosolic *Salmonella* are shed into the gut lumen, where they undergo inflammatory cell death. This process releases the invasion primed Salmonella into the gut lumen, which facilitates further invasion of host cells from the gut and shedding of the bacteria from the gut so that the pathogen can complete its life cycle (Knodler et al., 2010).

1.1.1 Components of the Gram-negative bacterial cell envelope

As bacteria are single celled organisms they need to be able to withstand stress and changes in their environment. The Gram-negative bacterial cell wall has two phospholipid bilayers: an inner membrane and an outer membrane. Between the two membranes is the periplasmic region

containing many proteins and the peptidoglycan cell wall (Silhavy et al., 2010). The peptidoglycan layer is made up of repeating units of two sugars: N-acetyl glucosamine (NAG) and N-acetyl muramic acid (NAM). Multiple chains of NAG and NAM are linked by peptide cross bridges (Vollmer et al., 2008). The flagellum is an important surface component for motility and chemotaxis and can be recognized by the host immune system (Eom et al., 2013). Attached to the outer membrane are lipopolysaccharide (LPS) molecules, which are important for immune recognition of the pathogen and are responsible for eliciting an immune response to infection (Galanos et al., 1984). The LPS will be discussed in more detail below.

NTS, along with many other bacteria, can form biofilms to aid their persistence within the environment and their host. Bacterial biofilms can be made up of various components, the components that are necessary for formation depends on the colonization site. For example, colanic acid and cellulose have been found to be needed for biofilm formation on chicken intestinal tissue and human epithelial type 2 cells but only cellulose is required for formation on a plastic surface (Ledeboer and Jones, 2005). The O-antigen capsule, among other types of capsules, can comprise part of the biofilm. Recently, the O-antigen capsule of *S.* Typhimurium has been shown to confer resistance to killing by human serum (Marshall and Gunn, 2015). The O-antigen capsule is very similar to the O-antigen of the LPS (discussed below) but is a much larger molecule with a higher molecular weight allowing it to form a capsule across the bacterial surface (Snyder et al., 2006).

1.2 THE O-ANTIGEN OF LIPOPOLYSACCHARIDE (LPS)

The outer membrane of Gram-negative bacteria is mostly covered by LPS, a lipid molecule that is anchored in to the membrane whilst binding sugar repeats that extend out from the outer cell surface. A bacterium's LPS plays an important role in activating an inflammatory response. The innate immune system will recognize LPS molecules using pattern recognition receptors that will recognize and bind LPS, activating an immune response (Dixon and Darveau, 2005; Gyorfy et al., 2013). For NTS activation of an immune response is an important step during infection (Thiennimitr et al., 2011).

The LPS is made up of three constitutive parts: the lipid A, the core oligosaccharide and the O-antigen. Lipid A is the domain that anchors LPS into the outer membrane and contains the

hydrophobic, lipid component of the LPS. The structure of this lipid is highly conserved between Gram-negative bacteria and is almost constant within enterobacteria. As well as acting as an anchor, lipid A is the toxic part of the LPS, when released from the membrane the LPS can diffuse through the host environment via its hydrophilic O-antigen whilst the lipid A domain will activate a strong immune response. The second domain is the core oligosaccharide, which joins the O-antigen to the lipid A anchor. It is composed of a short chain of sugars, the outer region connects to the O-antigen and the inner region binds the lipid A anchor (Lerouge and Vanderleyden, 2002).

The third component is the O-antigen repeating unit. This hydrophilic molecule is composed of a repeating unit of sugars forming a chain with as many as 40 repeats. The repeating unit has between 3-5 oligosaccharide subunits which are specific to each serovar of *Salmonella* (Lerouge and Vanderleyden, 2002). Variations in O-antigen sugar content can produce multiple antigenic types and give rise to specific immunological responses. The basic *S.* Typhimurium O-antigen repeating structure is galactose-rhamnose-mannose with an abequose side chain attached to mannose and in serotype O5, for example, the abequose sugar is acetylated (Hauser et al., 2011) (figure 1).

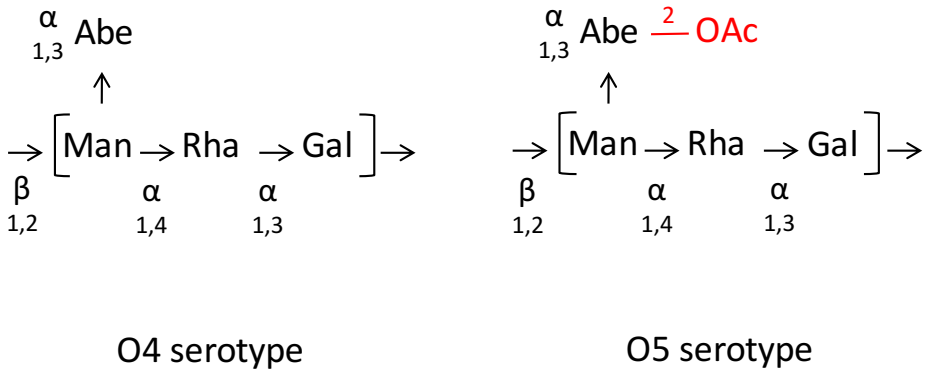


Figure 1: O-antigen structure of *S.* Typhimurium O4 and O5 serotypes

Serotype O5 has a modification of an acetyl group attached to the abequose side chain shown in red. Sugar moieties have been abbreviated: mannose (Man), rhamnose (Rha), galactose (Gal), abequose (Abe) and OAc indicates the acetyl group. The anomeric configuration of the sugar is given to the left and the carbon numbers involved in the bond are given underneath. Carbon 2 of abequose is bound by the acetyl group in serotype O5. Structural information taken from Micoli et al., 2014.

1.2.1 The O-antigen role in infection

The O-antigen, being the variable component of the LPS, can determine specific virulence factors. For example, some O-antigen structures and sugar components can promote adhesion of the bacterial cell to the epithelial cell wall. In *Helicobacter pylori* the Le(X) structure of the O-antigen enables the bacteria to adhere to the human gastric epithelium which may promote colonization of the mucosal surface (Edwards et al., 2000). It has also been shown that specific sugars can slow the pathogens uptake by macrophages (Liang-Takasaki et al., 1982). As well as variation in composition, the length and number of repeating sugars in the O-antigen can provide defense against the lytic action of compliment by keeping it away from the bacterial cell surface and preventing it forming pores in the membrane (Grossman et al., 1987).

1.2.2 Genes Involved in determining the structure and composition of the O-antigen.

The genes responsible for determining the sugars that make up the O-antigen are the *rfb* genes. This cluster of genes encodes biosynthesis enzymes, for the biosynthesis of the sugar precursors, and specific sugar transferases that will transfer the sugars onto a lipid carrier, undecaprenyl pyrophosphate (UndPP), ready to be flipped across the membrane (Jiang et al., 1991; Kalynych et al., 2014). The *wzx/wzy* group of genes is involved in the next step of LPS formation. The Wzx protein flips the lipid carrier linked chain of sugars across the membrane into the periplasm and is transferred to the Wzy protein. This protein then adds the subunit of sugars to a growing chain that will make up the O-antigen. The length of the chain is regulated by the Wzz protein (Batchelor et al., 1991).The O-antigen chain of sugars is then ligated onto the lipid A and core oligosaccharide to from a complete LPS molecule (Islam and Lam, 2014). Variations in these genes can produce multiple O-antigen structures and affect the virulence of a pathogen.

There are also genes outside of the *rfb* cluster that can modify the O-antigen. Such as the *oafA* gene in *S. Typhimurium* or the *oacD* gene in *Shigella flexneri* (Sun et al., 2014) that can add an acetyl group onto a specific sugar. For example in *S.* Typhimurium the OafA protein will add an acetyl group onto the abequose moiety of the O-antigen which will define this bacterium as an O5 serotype (Slauch et al., 1996).

Modifications of the O-antigen can affect how the pathogen is recognized by the immune system and may be important for O-antigen associated virulence factors. *Shigella flexneri*, causing shigellosis associated with diarrhea and fever, has been shown to have multiple modifications to its O-antigen repeating unit. For example, the rhamnose II of the core O-antigen unit can be modified by the addition of a glucose at a 1-3 position. This results in the conversion of *S. flexneri* from serotype Y to serotype 5a. This modification is mediated by the *gtr* operon which encodes the serotype specific protein GtrV (Allison and Verma, 2000).

Bacteriophage are known to carry genes that can positively affect the virulence of their bacterial hosts. For example, it has recently been shown that, as well as encoding the shiga toxin, shiga-toxigenic bacteriophage can effect the expression of the bacterial host's acid response genes improving its survival in the stomach (Veses-Garcia et al., 2015). But bacteriophage can also carry genes that prevent the co-infection of the bacterial host. The O-antigen is often used as a receptor for bacteriophage binding to the bacterial cell surface, therefore to avoid co-infection it is thought that some bacteriophage can mediate the modification of the O-antigen. For example, the P22 *Salmonella* bacteriophage carries a *gtr* operon that mediates the addition of a glucose molecule onto the galactose moiety of the S. Typhimurium O-antigen. O-antigen modification by these genes are thought to prevent co-infection with other P22 bacteriophage as the P22 tail-spike protein cleaves the O-antigen to gain better access to the bacterial surface. This cleavage activity is prevented by the glycosylation of the galactose moiety of the O-antigen (Kintz et al., 2015).

1.3 THE *GTR* OPERON AND THE ENCODED PROTEINS

Many *Salmonella* serovars contain glucosyltransferase (*gtr*) operons. The *gtr* operons of known function encode proteins that can change the serotype of the strain by specifically modifying the O-antigen through addition of a glucose molecule (Davies et al., 2013). This project will focus on the *gtr* genes present in multiple serovars of *S. enterica subsp. enterica*.

1.3.1 Model of Gtr proteins function

gtrB, the second gene in the *gtr* operon, is proposed to be a bactoprenol glucosyltransferase whose function is to transfer a glucose molecule from UDP-glucose onto a undecaprenyl-pyrophosphate(UnDP) lipid carrier(Allison and Verma, 2000). The GtrA protein is proposed to be a flippase that flips UnDP-glucose across the inner membrane of the cell wall so that it can be accessed by GtrC in the periplasm (Allison and Verma, 2000; Nair et al., 2011). In the periplasm GtrC will transfer the glucose from the UnDP lipid carrier onto a specific point on the O-antigen. This is thought to occur during the synthesis of the O-antigen before it is attached onto the lipid A core (Allison and Verma, 2000). A diagram showing the model of the Gtr proteins activity can be found in figure 2. All known function Gtr proteins in both *Salmonella* and *S. flexneri* have been found to mediate the addition of a glucose molecule (Allison and Verma, 2000). However, it is possible that *gtr* families of unidentified function could be binding and transferring a sugar other than glucose.

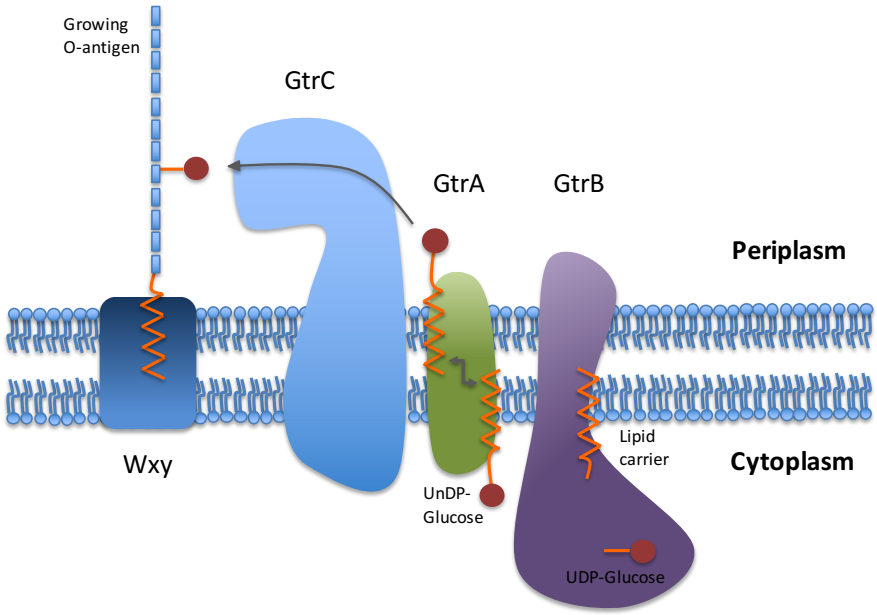


Figure 2: A model of *gtrABC* protein activity.

The GtrB protein is a bactoprenol glucosyltransferase, which transfers a glucose molecule from UND-glucose onto a undecaprenyl-pyrophosphate(UnDP) lipid carrier. UnDP-Glucose is then bound by GtrA which flips it across the membrane into the periplasm so that it can be accessed by GtrC. GtrC is a glucosyl transferase that transfers the glucose onto the growing O-antigen chain, at a specific point, whilst it is being formed in the periplasm. (Adapted from Lerouge and Vanderleyden, 2002)

1.3.2 The *gtr* proteins are split into 10 families

The P22 phage encoded Gtr proteins were used to identify other *gtr*-like genes across 57 *Salmonella* strains including both *S. enterica* and *S. bongori* and four genomes of bacteriophage known to modify the O-antigen of *Salmonella* (Davies et al., 2013). This was done using a BLASTn analysis which found 59 other *gtr*-like genes, mainly within *S. enterica subs. enterica*. The 59 operons cluster in to 10 families based on the sequence of *gtrC,* as this is the proposed serotype specific gene. The *gtrA* and *gtrB* genes do not follow the same pattern of clustering as their *gtrC* counterparts. This was predicted to be due to recombination events between operons of different families as *gtrA* and *gtrB* sequences are well conserved across the families so function of the operon would not be disrupted by recombination (Davies et al., 2013). To date the functions and specific modifications mediated by the family I, II, III and VIII proteins have been elucidated. Also, Davies et al., has demonstrated that the family VI and V proteins mediate the addition of a molecule to the *S. Infantis* O-antigen repeating unit but details of the modification are not known. As the GtrC protein of each family is targeting a specific moiety of the basal O-antigen they are confined to serovars that produce an O-antigen containing their target. For example, all family I and III containing strains have a galactose moiety as part of their core O-antigen structure, the sugar target of both these families.

More than one *gtr* family can be found in the same strain, for example in the *S.* Typhimurium serovar the family IV proteins usually share the genome with the family III genes. In S. Infantis family IV, family V and VI gtr genes have been found in the same genome (Davies et al., 2013).

1.3.3 Glycosylation by the Family I, III and VIII *gtr* operons

Family I, III and VIII all mediate very similar modifications to the core *Salmonella* O-antigen. Family I includes the *gtr* operon encoded by the P22 bacteriophage, the most well studied *gtr* genes. These genes (previously known as the a1 gene) mediate the addition of glucose onto the galactose moiety of the basal unit via a 1-6 linkage (Mäkelä, 1973) (figure 3). Their function was first identified by Young and his group who showed that a1 mutants are unable to produce antigen 1. This is now referred to as an O:1 serotype, which is produced after conversion by GtrABC proteins encoded in the P22 genome (Byl and Kropinski, 2000; Fukazawa and Hartman, 1964).

The role of the family III Gtr proteins was shown in the S. Typhimurium serovar. The STM0557/0558/0559 genes, which are *gtr* genes CBA respectively, were also shown to transfer a glucose molecule onto a galactose moiety but via a 1-4 linkage as apposed to a 1-6 linkage (figure 3). It was also shown that these genes were needed for persistence within the murine intestine (Bogomolnaya et al., 2008). Although this result should be considered with the knowledge that this operon phase varies (Broadbent et al., 2010), which was not controlled for during their experiments. The phase variation of the surface proteins may have ensured persistence and not the specific modification.

The family VIII *gtr* operon is encoded by the *Salmonella* bacteriophage epsilon 34. Proteins encoded in the epsilon 34 genome modify the O-antigen of *S. Anatum* by addition of a glucose at a 1-4 position to the beta-galactose moiety of the core O-antigen unit (Reeves et al., 2013) (Villafane et al., 2008). This is almost identical to the modification mediated by family III but is specifically targeting a beta-galactose moiety. Before conversion by epsilon 34, *S. Anatum* must first be converted by *Salmonella* bacteriophage epsilon 15. Epsilon 15 blocks the activity of the host cell O-polysaccharide alpha polymerizing enzyme to replace it with its own O-polysaccharide beta polymerizing enzyme, which produces an O-antigen core structure with a beta, rather than alpha, galactose moiety (Kropinski et al., 2007), which can be targeted by GtrC of epsilon 34.

1.3.4 Acetylation by the Family II *gtr* operon

Unlike the other *gtr* operons the family II Gtr protein is thought to transfer an acetyl group onto the O-antigen. It is proposed that GtrC(II), produced from the BTP1 prophage, mediates the addition of the acetyl group found at the C-3 and C-2 positions of the rhamnose moiety of the S. Typhiumurium str. D23580 O-antigen (Kintz et al., 2015; Micoli et al., 2014) (figure 3). Interestingly the family II GtrC does not require GtrAB to function and consequently, the *gtr* operons that have been studied have had either deleted or mutated *gtrAB* gene sequences (Kintz et al., 2015; Davies et al., 2013).

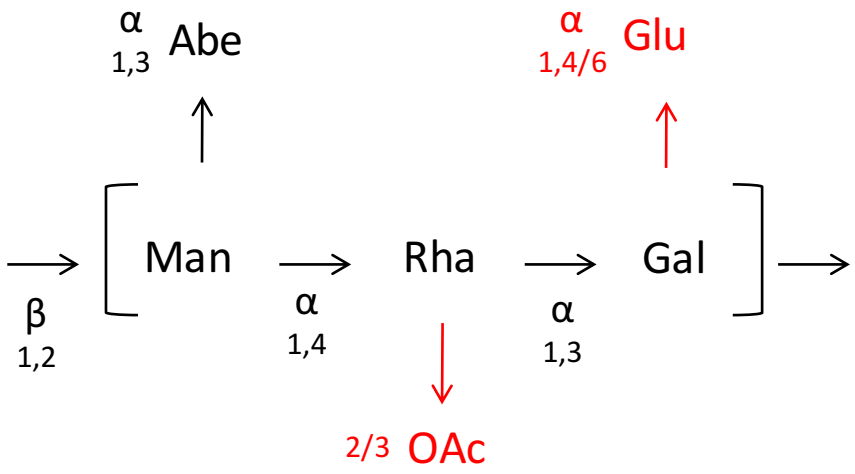


Figure 3: Modifications to the O-antigen structure of *S.* Typhimurium mediated by Gtr proteins of family I, II, and III.

Family I mediates the addition of a glucose at the 1, 6 position and family III at the 1, 4 position on the galactose moiety. Family II mediates the addition of an acetyl group onto the rhamnose moiety at either carbon 2 or 3. Modifications are indicated in red. Sugar moieties have been abbreviated: mannose (Man), rhamnose (Rha), galactose (Gal), abequose (Abe) and OAc indicates the acetyl group. The anomeric configuration of the sugar is given to the left and the carbon numbers involved in the bond are given underneath. Structural information taken from Micoli et al., 2014.

1.3.5 Expression of the *gtr* operon and phase variation

Several *gtr* gene clusters, including the P22 *gtrABC* genes, have been shown to regulate their expression through phase variation via a Dam and OxyR dependent system. Two pairs of GATC sites, that can be methylated by Dam, were identified upstream of the *gtrA* transcriptional start site, which overlaps three OxyR binding sites. Differential methylation of each GATC pair determines whether the promoter is in an off or on state. When the downstream GATC pair in methylated OxyR is unable to bind at the downstream sites leaving it free for RNA polymerase to bind. However, when the upstream GATC pair is methylated the OxyR can't bind and instead binds to the downstream GATC sites blocking RNA polymerase binding to the promoter preventing transcription (Broadbent et al., 2010). It is thought that the on or off state is heritable due to the methylated GATC sites but the method of switching is still unknown. The heterogeneous population, created by phase variation, may be an important virulence strategy of the bacteria, for example in evasion of the adaptive immune response (Woude and Bäumler, 2004).

1.4 THE FAMILY IV *GTR* OPERON

1.4.1 The family IV Gtr proteins importance in infection

This project studied the family IV *gtr* genes. Very little is known about these genes and their proteins' functions, but it has been hypothesized that they work to modify the O-antigen repeating unit based on their sequence similarity with Gtr families with known function. As mentioned above, the O-antigen can have an important role during infection. Further studies into its structure and modifications have the potential to inform therapeutic methods of tackling *Salmonella* infection.

More specifically the family IV proteins are important to study as they have recently been shown to be required for persistence within four different host animals (Chaudhuri et al., 2013). In this study each gene was separately mutated and used to infect chickens, pigs, cattle and mice. *gtrA(IV)* and *gtrB(IV)* mutants showed a significantly reduced persistence within all animals apart from the mouse. Whereas, the *gtrC(IV)* mutant showed significantly reduced persistence within all host animals including the mouse. Therefore, the family IV genes may be needed for virulence within a host.

1.4.2 Expression of the family IV *gtr* genes under infection relevant conditions

A recent study has used RNA-seq to look at the transcriptional profile of *S.* Typhimurium under various infectious relevant growth conditions and generated heat maps for every gene (Kröger et al., 2013). The heat maps generated for the family IV *gtr* genes don't show any points of very high expression. And the level of mRNA does vary from each gene, which could be due to mRNA degradation after transcription. However, there are a few conditions that show higher levels of transcription than others: under nitric oxide shock *gtrA(IV)* shows its highest level of transcription with *gtrC(IV)* also showing high levels of transcription. pH3 shock also induces some of the highest levels of transcription particularly in *gtrC(IV)*. (Kröger et al., 2013).

When *Salmonella* is infecting a host it will experience conditions of NO shock from the immune response mounted against it. Intracellular and extracellular *Salmonella* can trigger the expression of inducible nitric oxide synthase (iNOS) causing production of NO from immune cells like macrophages and dendritic cells (MacMicking et al., 1997; Kalupahana et al., 2005). This response has also been seen in human intestinal epithelial cells after bacterial invasion (Salzman et al., 1998). pH3 conditions are relevant to infection because this is the pH of the gastric system of humans (Goktepe et al., 2005) and other common hosts of *S.* Typhimurium like pigs (Crévieu-Gabriel et al., 1999; Wilkins et al., 2010) is around pH3.

Experiments done alongside the RNA-seq transcriptome study (Kröger et al., 2013) looked globally for transcriptional start sites within the *S.* Typhimurium genome. Terminator exonuclease was used to enrich a sample of RNA for primary transcripts by selectively degrading 5'-monophosphorylated transcripts before RNAseq. This experiment found two possible transcriptional start sites (TSS) within the *gtrB*(IV) gene upstream of *gtrC*(IV) (personal correspondence with Disa Hammerlof – exact positions of the TSS's were not available). This shows that *gtrC(IV)* could be being transcribed independently of *gtrAB(IV)*.

1.4.3 The function of the family IV Gtr proteins

Very little is known about the function of the family IV genes, but some hypotheses can be made based on the results of the studies mentioned above. As well as demonstrating the genes importance in infection, the Chaudhuri et al., (Chaudhuri et al., 2013) results suggest that GtrC(IV)'s function doesn't always rely on the expression of GtrAB(IV) as the *gtrC*(IV) mutant was the only family IV *gtr* mutant attenuated in the mouse model. This difference may be a result of the conditions the pathogen is under; it will have encountered a different environment in the mouse as it was infected intravenously whereas the other three animals were infected orally. Also, *S.* Typhimurium causes a typhoid like infection in mice, unlike the other animals studied. It may be that at certain points during infection only GtrC is needed for virulence. The data showing possible TSS within *gtrB*(IV) also supports the theory that GtrC(IV)'s function may not always rely on GtrAB(IV). If *gtrC*(IV) is being transcribed independently it may be because it is need at different times during the infection process.

Unlike most of the other families, family IV's *gtrA* and *gtrB* genes cluster into the same groups as the family IV *gtrC'*s. This is probably because there has been no recombination of the family IV operon with other family operons, this may be because GtrC(IV) is unable to function with GtrAB(IV) from other families and can only function with GtrAB(IV). This could also mean that the role of GtrC differs from other family GtrC's. These conclusions are in contrast to the conclusion that GtrC may work independently of GtrAB mentioned previously. It could be that GtrC can function by itself at certain points during infection and at others it relies on GtrAB.

Another dissimilarity of the family IV *gtr* genes is that the operon's expression does not phase vary in the same way as some other *gtr* operons. No GATC sites, necessary for phase variation through a OxyR and Dam dependent mechanism, were found up stream of *gtrA*(IV). Nor was phase variation seen when the operons regulatory region was in front of the *lacZ* gene. Therefore, they have a different mechanism of controlling gene expression compared to some other *gtr* families which have been shown to phase vary. This could mean that their role in virulence is different as well.

Davies et al (Davies et al., 2013) demonstrated that the family I, III, V and VI Gtr proteins functioned to modify the O-antigen by extracting the LPS from strains (*S.* Typhimurium was used when expressing families I and III and *S. Infantis* for families V and VI) that were

constitutively expressing only one of these *gtr* families, all other O-antigen modifying genes had been removed. The LPS was analysed using gel electrophoresis, which produces a banding pattern showing varying lengths of O-antigen repeating units. When comparing this banding pattern to one belonging to a basal strain, which contains no known O-antigen modifying genes, there is a shift seen between the two banding patterns showing that a molecule has been added to the O-antigen repeating unit. However, when the family IV genes were analysed in this way, in an *S.* Typhimurium strain, no shift was seen between the banding pattern of a strain constitutively expressing the family IV genes and the banding pattern of a basal strain expressing no known O-antigen modifying genes.

This could be due to multiple reasons. The GtrC substrate or target is not available, for example GtrC(IV) may target a modification of the *S*. Typhimurium O-antigen, possibly one provided by another Gtr family. The GtrC is not adding anything to the LPS but just changing the bonds between the sugars. The mRNA transcript of the constitutively expressed family IV operon is not being translated into protein under normal laboratory conditions. There is also the possibility that GtrC is not acting on the LPS and its target is something else.

1.5 AIMS OF THE PROJECT

The aim of the project was to gain an insight into the function of the family IV Gtr proteins, a range of techniques were used to do this. Modifications made to the O-antigen by other Gtr families was studied to test whether they provide a target for GtrC(IV) activity. GtrAB(IV) was studied to test whether their function is unique to GtrC(IV) or whether they could replace the function of GtrAB proteins from other families. The expression of the gtr(IV) genes was studied under infection relevant conditions to understand what role the proteins might have in infection. The independent expression of GtrC(IV) was also investigated as this suggests that GtrC(IV) may work independently of GtrAB(IV). A bioinformatics analysis was carried out to predict the function of the Gtr(IV) proteins and compare their amino acid (AA) sequence to Gtr proteins of known function.

Collectively, these studies may give an insight into the function of the family IV Gtr proteins. The family IV proteins may be important for infection in certain hosts (Chaudhuri et al., 2013)

and as they are thought to be involved in surface modification, an understanding of their function and role in infection could inform development of therapeutic methods against salmonellosis.

2. MATERIAL AND METHODS

2.1 DNA ANALYSIS

2.1.1 Agarose gel electrophoresis

1% agarose gel was prepared by melting 1% (W/V) agarose (Sigma-Aldrich) into TAE buffer (Life Technologies). The gel was left to cool slightly before pouring into a cast then adding ethidium bromide to a final concentration of 0.2 μg/ml. Once set the gel was placed into a gel tank filled with TAE buffer. DNA samples were loaded along with a GeneRuler DNA Ladder Mix (Thermo Scientific) and ran at 100V. Visualisation was done in G:BOX gel doc (Syngene).

2.1.2 DNA quantification

A NanoDrop 1000 (Thermo Scientific) was used to measure DNA concentration by measuring the light absorbance at an optical density (OD) 260. Readings at OD280 and OD230 were also taken to measure the contamination of the sample. A 1μl sample was loaded for measurement.

2.1.3 Sequencing

A SmartSeq kit from Eurofins Genomics Company was used to sequence lengths of up to 1100bp of purified PCR products or extracted plasmid. Samples were prepared according to the manufacturers instructions.

2.2 BACTERIAL CULTURE AND GROWTH MEDIA

2.2.1 Solid and liquid Lennox broth (LB) media

LB powder (Fisher Scientific) was dissolved in ddH2O to a final concentration of 20g/L. when making solid media; powdered agar (Fisher Scientific) was added to a final concentration of 17.5g/L. The solution was then autoclaved. Antibiotics were added to the required concentration (table 1) once the media had cooled to 55°C. Whilst solid media was still melted it was poured into plates and dried in a lamina flow hood.

Table 1: Antibiotic concentrations

Antibiotic	Final cocentration
Ampicillin	100µg/ml
Chloramphenicol	34µg/ml

2.2.2 Solid and liquid M9 minimal media

M9 salts (Sigma) were dissolved in ddH2O and autoclaved. Once cooled to 55°C nutrients, listed in table 3, were added to the required concentration. If solid media was being prepared powdered agar was dissolved in ddH2O to a final concentration of 17.5g/L, autoclaved then cooled to 55°C. This was then added to M9 minimal media in a 1:1 ratio, mixed and poured into plates before drying in a lamina flow hood.

2.2.3 InSPI2 (Inducing Salmonella Pathogenicity Island II) media

See table 2 for components and concentrations. Once all components were dissolved the solution was adjusted to pH 5.8 and filter sterilized. The recipe for InSPI2 media was taken from Kröger et al., 2013.

2.2.4 Growth conditions

Strains were cultured overnight in a 10 ml volume tube by inoculating 5 ml of LB with one colony (colonies were grown on an LB agar plate containing appropriate antibiotics) of bacteria in a 37°C incubator with shaking at 25 rpm unless otherwise specified. Appropriate antibiotics were added to the required concentration.

2.3 POLYMERASE CHAIN REACTION (PCR) AND PURIFICATION

High fidelity PCR was performed using either Q5® Hot Start High-Fidelity DNA Polymerase (NEB) or KOD Hot Start DNA Polymerase (Merck-Millipore).

For diagnostic PCR GoTaq® G2 Flexi DNA polymerase (Promega) was used. Product specific protocols were followed for all. Colony, plasmid or purified PRC product were used as template. Primers were purchased in freeze-dried form from Sigma and resuspended in ultra pure H2O (milliQ by Millipore) to a concentration of 100μM. PCR products were purified using QIAquick PCR Purification Kit (QIAGEN). The protocol provided with the kit was followed.

2.4 PLASMID EXTRACTION AND CONCENTRATION

Strains containing the plasmid were cultured overnight in a 10 ml volume tube by inoculating 5 ml of LB with one colony (colonies were grown on an LB agar plate containing appropriate antibiotics) of bacteria in a temperature specific incubator with shaking at 25rpm. Appropriate antibiotics were added to the required concentration. Plasmid was then extracted from the saturated culture using the QIAprep Spin Miniprep kit (QIAGEN). The protocol provided with the kit was followed.

The concentration of the extracted plasmid was determined using the method described in section 2.1.2. If a higher concentration of plasmid was required a miVac DNA Concentrator

(Genevac) was pre-warmed to 50°C before spinning the plasmid for an appropriate amount of time to evaporate water from the sample.

2.5 RESTRICTION ENZYME DIGESTS

All restriction enzymes were from NEB and used according to the manufacturer's instructions. Restriction digests were performed on either extracted plasmids or purified PCR products.

2.6 LIGATION

When ligating insert to vector a 3:1 molar ratio was used. T4 DNA ligase from NEB was used for all ligations and the protocol provided was followed.

2.7 BACTERIAL TRANSFORMATION BY ELECTROPORATION

Strains were grown to saturation in liquid media overnight (O/N) with appropriate antibiotics and at the required temperature with shaking. The next morning new liquid media was inoculated with a 1 in 100 dilution of O/N culture and grown to an OD600 of between 0.4-0.6. OD600 was measured using a Jenway 6505 spectrophotometer.

Once the OD600 0.4-0.6 was reached between 5-50ml (depending on the efficiency of transformation needed) was centrifuged for 15 min at 5000xg at 4°C. Supernatant was removed then an equivalent volume of autoclaved ultra pure H2O was used to re-suspend the pellet before centrifuging again for 8 min under the same conditions. This step was repeated a further 2-4 times. The final pellet was re-suspended in 100µl. 50µl of this solution was pipetted into a Gene Pulser/Micropulser Electroporation Cuvette with a 0.2 cm gap width (BioRad) along with 1-4µl of plasmid/ligation solution. A BioRad MicroPulserTM was used to apply a current for electroporation and was operated according to the instructions provided with the machine, the EC2 setting was used for all transformations. 1ml of SOC media (table 4) was added immediately after electroporation.

The culture was then incubated for 1 hour at 37°C with shaking unless specified otherwise. 100µl of culture was then spread on appropriately prepared plates. The remaining culture was centrifuged at 5000xg for 5 min. The supernatant was removed and the pellet was re-suspended in 100µl of LB before spreading. Plates were then incubated O/N at the required temperature. These methods were adapted from a protocol developed by Erica Kintz.

2.8 GENOME INTEGRATION AND THE CRIM VECTOR

lacZ gene constructs were integrated into the chromosome using the CRIM vector as described by Haldimann and Wanner, 2001. Briefly, after cloning of a desired region into the CRIM vector, in front of the *lacZ* gene between the Pst1 and Acc651 restriction sites, the plasmid was transformed by electroporation into a pir+ strain. This strain was then grown to saturation in LB liquid media containing chloramphenicol (34µg/ml) before extraction of the plasmid. This was then electroporated into a *S*. Typhimurium LT2 strain containing a pINT-temperature sensitive (ts) plasmid, which can express the phage integrase protein for integration into the chromosomal attB phage insertion site. Before electroporation this strain was grown at 30°C with ampicillin (100µg/ml) to maintain the pINT-ts plasmid. Once the CRIM vector had been transformed into this strain the culture was immediately grown at 37°C when integration of the CRIM vector took place and then at 42°C at which the pINT-ts plasmid was lost. Any un-integrated CRIM vectors were also lost as they were Pir dependent.

2.9 LPS EXTRACTION

2.9.1 Extraction and preparation of LPS

If the strain was induced before LPS extraction an O/N culture grown in LB was diluted 1 in 100 and grown to OD600 0.6 then induced with 1mM of IPTG then grown O/N before LPS was extracted. 1ml of O/N culture was centrifuged for 5 min at 8000xg and the supernatant discarded. The pellet was re-suspended in 100µl of LPS buffer A (table 5), vortexed and boiled for 5 min. 400µl of LPS buffer B (table 5) was added and vortexed before adding 2µl of RNase (Roche) and DNase (Sigma) and incubated at 37°C for 5 hours. 5µl of proteinase K (10 mg/ml) was then added, votexed and incubated at 50°C O/N.

25µl of each sample was mixed to a 1:1 ratio with loading buffer (table 12), vortexed, boiled for 6 min then cooled to room temperature before loading into a TSDS-PAGE gel. These steps were modified from the methods used by Davies et al., 2013.

2.9.2 Preparation and running of TSDS-PAGE gel.

The separating gel (table 6) was pipetted between the glass casting plates leaving around a 5cm space above. A layer of water-saturated butanol was pipetted on top and the gel was left to polymerize. Once polymerized the water saturated butanol was washed off and the stacking gel (table 7) pipetted into the cast and the well comb was inserted. Once the stacking gel had polymerised the comb was removed and the gel, between the casting plates, was attached to the tank. Anode (0.2M tris, pH8.9) and cathode buffers (table 8) were poured into the bottom and top chambers respectively. Once loaded, the gel was ran at 50V until the die front had passed through the stacking gel when it was turned up to 100V and ran overnight then stopped once the die front had reached the end of the gel. This protocol was modified from the methods used by Lesse et al., 1990.

2.9.3 Silver staining of TSDS-PAGE gel.

All steps were performed at room temperature with slow agitation. The gel was first fixed in fixative solution (table 9) for 3-5 days. It was then placed in oxidizer (table 10) for 10 minutes then washed three times with milli-Q water for 15 minutes. The gel was stained with silver stain solution (6mM silver nitrate) for 30 minutes in a foil-covered container. The gel was rinsed with milli-Q water before transferring into the developer solution (table 11) and developing for 10-45 min or until sufficiently developed. The silver reaction was then stopped with 1% acetic acid. This protocol was modified from the methods used by Kittelberger and Hilbink, 1993.

2.10 NITRIC OXIDE SHOCK (NO) SHOCK

Growth conditions followed those used by Kröger et al., 2013. O/N cultures grown in M9 minimal media (see section 2.2.2) were inoculated 1 in 100 in 50 ml of InSPI2 media (see section 2.2.3 and table 2). The culture was grown at 37°C with shaking until an OD600 of around 0.3 was reached, during the beta-galactosidase assay a sample was taken at this point. The culture was then split into pre-warmed glassware and one half was given 250μM of spermine NONOate. The cultures were grown for a further 20 min before sampling. During the beta-galactosidase assay a sample was also taken 60 min after spermine NONOate addition.

2.11 PH3 SHOCK

Growth conditions followed those used by Kröger et al., 2013. O/N cultures were inoculated 1 in 100 in 50 ml of LB media. The culture was grown at 37°C with shaking until on OD600 of around 0.3 was reached. The culture was then split into two pre warmed falcon tubes and centrifuged for 7 min at 5000xg. The supernatant was removed and one tube of pelleted culture was re-suspended in LB at pH3, acidified by addition of HCl, and the other with LB at pH7. This was then poured into pre-warmed glassware and grown at 37°C with shaking for 10 min/5, 10 and 15 min before sampling.

2.12 BETA-GALACTOSIDASE ASSAY

A protocol developed by Miller, 1972 was followed: samples of culture were taken and their OD600 was recorded. They were chilled on ice for 10-20 min before dilution in Z buffer to a final volume of 1 ml at the relevant ratio. 50μl and 100μl of 0.1% SDS and chloroform respectively were added and the mixture was vortexed for 10 sec. These samples were then pre-incubated at 28°C for around 5 min. 200μl of ortho-nitrophenyl-β-galactoside (ONPG) solution (4mg/ml ONPG dissolved in a phosphate buffer solution at pH7 made by combining 0.1M of NaH_2PO_4 and Na_2HPO_4 at a 1 to 1.56 ration) was added and vortexed for 10 sec before returning to incubation at 28°C. The time of ONPG addition was recorded. Once the solution had turned a pale yellow colour the reaction was stopped with 0.5ml of 1M Na_2CO_3 and the time was

recorded. The solution was centrifuged for 15 min at 8000xg. The supernatant was removed and its OD420 was recorded.

A formula was used to calculate the expression of beta-galactosidase in Miller units: Beta-galactosidase units = (1000 x OD420) / (T x V x OD600), where 'T' is time of the reaction in minutes and 'V' is volume of the culture used in ml.

2.13 RNA EXTRACTION

The RNeasy mini Kit (QIAGEN) and Bacteriaprotect reagent (QIAGEN) were used for RNA extraction. The protocol provided with the Bacteria protect reagent was followed. 2.5ml of culture was sampled for RNA extraction.

RNA sample was then checked for genomic DNA contamination. A 2µl aliquot of extracted RNA (or 2µl of genomic DNA used as a positive control) was mixed with 125µl of a sterile buffer (60mM Tris-HCl, 1mM EDTA, at pH6.8). 2 µl of Dnase-free Rnase (Roche) was added. The solution was incubated O/N at 37°C. A diagnostic PCR was performed using primers to be used in the RT-qPCR reaction and GoTaq® G2 Flexi DNA polymerase (Promega).

2.14 CDNA SYNTHESIS

SuperScript™ II Reverse Transcriptase (Invitrogen) was used and the accompanying protocol was followed.

2.15 RT-QPCR

An Applied Biosystems Fast qPCR machine was used. Each reaction contained 20µl total volume containing: 10µl fast SYBR green master mix (Applied Biosystems), 7µl nuclease free

water, 0.5µl of each primer (250nM), 2µl cDNA or nuclease free water as the negative control. The thermal cycling conditions used were: initial denaturation for 20 sec at 95°C followed by 40 cycles of denaturation for 3 sec at 95°C then a combined annealing and extension step at 60°C for 30 sec. After cycling was complete a melt curve was performed: initial denaturation was performed at 95°C for 5 sec. The temperature was then held at 60°C for 1 min before gradually increasing at 0.3°C increments from 60°C to 95°C then held at 95°C for 15 sec. The level of fluorescence is recorded across the gradual heating period to detect the temperature at which the double stranded product denatures. The change in fluoresces over time is plotted against the temperature to produce a melt curve.

2.16 BIOINFORMATICS ANALYSIS

Functional domain prediction of GtrABC family IV proteins was done using the InterPro: protein sequence analysis & classification tool (Mitchell et al., 2014).

All AA and nucleotide sequence aliments were generated using Cobalt Constraint-based Multiple Protein Alignment Tool (Papadopoulos and Agarwala, 2007a) and Align Sequences Nucleotide BLAST tool (Altschul et al., 1990) respectively. Parameters of alignments are given in figure legends. All transmembrane domain prediction was done using TMHMM Server v. 2.0 available at http://www.cbs.dtu.dk/services/TMHMM/ unless specified otherwise.

3. BIOINFORMATICS ANALYSIS OF THE FAMILY IV *GTR* GENES AND PROTEINS

Very little is known about the family IV operon and the three encoded proteins. The operon has previously been found in the genomes of multiple strains from different eight serovars (Davies et al., 2013). The family IV proteins are predicted to modify the O-antigen due to sequence similarity with other known O-antigen modifying genes but this function was not able to be demonstrated in previous studies (Davies et al., 2013). The family IV gtr operon has not undergone any recombination events with other family's operons (mentioned in section 1.4.3) unlike gtr operons of other gtr families (Davies et al., 2013). This could suggest that the function of these proteins is slightly different, and therefore recombination of the operon with gtr genes from other families would result in a non-functional operon. This section compares the family IV proteins with proteins from other families. It also identifed newly sequenced strains containing the gtr(IV) genes and analyse their O-antigen structures.

3.1 THE SPREAD OF *GTR* GENES ACROSS THE SALMONELLA SEROVARS

In 2013 Davies et al identified 57 *gtr* operons in the genomes of *S.* enterica, *S.* bongori and *Salmonella* bacteriophage and assigned them to ten different families based on their *gtrC* sequence. Since then more Salmonella strains have been sequenced. BLASTn analysis was done using a *gtrC* nucleotide sequence from families one to seven. This identified many other *gtr* operons in newly sequenced *S.* enterica strains. A list of the O-antigen groups of each serovar was collected from the: 'Antigenic formulae of the *Salmonella* serovars' compiled by the WHO Collaborating Centre for Reference and Research on Salmonella (Grimont and Weill, 2007). The O-antigen group is determined through agglutination with specific antisera and refers to the core O-antigen structure (European Food Safety Authority, 2010) (Grimont and Weill, 2007). This showed the range of O-antigen structures of each family's serovars.

Table 2: List of serovars containing each family of *gtr* genes

Family I			Family II			Family III			Family IV		
Serovar	No. of strains	O-antigen group	Serovar	No. of strains	O-antigen group	Serovar	No. of strains	O-antigen group	Serovar	No. of strains	O-antigen group
Heidelberg	5	O:4	Gallinarum	3	O:9	Dublin	2	O:9	Tennessee	1	O:7
Dublin	2	O:9	Pullorum	3	O:9	Pullorum	3	O:9	Thompson	4	O:7
Typhimurium	7	O:4	Dublin	2	O:9	Gallinarum	3	O:9	Paratyphi C	1	O:7
Abony	1	O:4	Paratyphi A	5	O:2	Typhimurium	21	O:4	Infantis	1	O:7
Pullorum	1	O:9	Typhi	12	O:9	Agona	3	O:4	Choleraesuis	3	O:7
Paratyphi A	5	O:2	Typhimurium	3	O:4	Heidelberg	5	O:4	Typhimurium	22	O:4
YU39*	1		Enteritidis	≥100	O:9	Paratyphi B	1	O:4	Agona	3	O:4
						Panama	1	O:9	Heidelberg	5	O:4
						Javiana	1	O:9	YU39*	1	-
						Enteritidis	≥100	O:9	Bovismorbificans	1	O:8
						Typhi	6	O:9	Hadar	1	O:8
						Paratyphi A	5	O:2	Newport	12	O:8
						Abony	1	O:4	Anatum	4	O:3,10
						YU39*	1	-	Senftenberg	1	O:1,3,19
									Ouakam	1	O:9,46
									Cubana	1	O:13
Total strains	22		Total strains	≥128		Total strains	≥153		Total strains	62	

Family V			Family VI			Family VII		
Serovar	No. of strains	O-antigen group	Serovar	No. of strains	O-antigen group	Serovar	No. of strains	O-antigen group
Choleraesuis	3	O:7	Choleraesuis	3	O:7	Newport	12	O:8
Paratyphi C	1	O:7	Paratyphi C	1	O:7	Bovismorbificans	1	O:8
Typhimurium	1	O:4	Typhimurium	1	O:4	Hadar	1	O:8
Thompson	4	O:7	Thompson	4	O:7			
Infantis	1	O:7	Infantis	1	O:7			
Bareilly	1	O:7	Bareilly	1	O:7			
Tennessee	1	O:7	Tennessee	1	O:7			
Montevideo	3	O:7	Montevideo	3	O:7			
Total strains	15		Total strains	15		Total strains	14	

Green cells indicate newly identified serovars.
*YU39 indicates the strain not the serovar and is refereeing to Salmonella enterica subsp. enterica strain YU39

Table 2 lists the serovars that each family were found in and the number of strains in each serovar that contain the genes and also the O-antigen group of each serovar. The cells highlighted in green show newly identified serovars that contain *gtr* genes. But in previously identified serovars there were also newly sequenced strains that were found to contain *gtr* genes. The Typhimurium serovar in particular had many newly sequenced strains that contained *gtr* operons. One strain of Typhimurium was found to have *gtr* genes of family V and VI *gtr* families not previously found in Typhimurium. Another notable observation about the family V and VI *gtr* families is that they were always found together in a strain. This could indicate that they are carried by the same bacteriophage, which is supported by the observation that they both map to the same, phage associated SPI-16, genomic region in *S.* Typhimurium.

The BLASTn analysis showed that, to date, the family IV *gtr* genes are present in the largest number of serovars. Family IV was also one of the most widely spread families in terms of the number of strains it has been found in. But Family III and family II are found within the largest number of strains. Although, this is mainly because family II and III are found in the Enteritidis serovar, which make up the bulk of the strains in these families.

As well as being wide spread across serovars, Table 2 showed that the family IV *gtr* genes came from 7 different O-antigen groups. This was a large amount compared to the other *gtr* families that had between one and three different O-antigen groups. Davies et al., observed that the *gtr* families were contained within limited O-antigen groups that share the same core O-antigen structure. This is most likely because the GtrC protein targets a specific moiety of the O-antigen structure and can only function in certain O-antigen groups. Although, the family IV genes seem to be maintained within strains of a range of O-antigen groups.

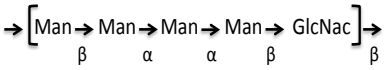
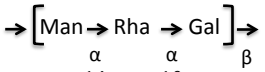
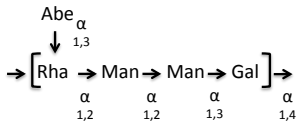
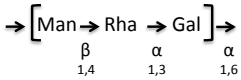
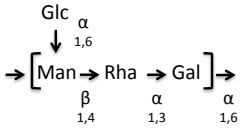
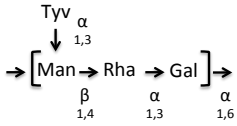
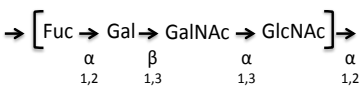
3.2 THE O-ANTIGEN STRUCTURE OF THE SEROVARS CONTAINING THE FAMILY IV *GTR* GENES

The GtrC is thought to add a sugar molecule onto a specific residue of the O-antigen repeating unit. To predict which residue is modified by the family IV proteins the basic structure of the repeating units of all serovars containing the family IV *gtr* genes were compiled into a table and compared to identify conserved sugars. The basic O-antigen structure of each serovar was identified from their O-antigen group (Liu et al., 2014a). As the family IV gtr genes come from a range of different O-antigen groups, shown above, this could indicate that the O-antigen structures are also diverse.

Information on other *gtr* families in that serovar was included in the table as they can provide modifications to the basic O-antigen structure. For example, family I and III are known to add a glucose molecule onto a galactose at different positions (Young et al., 1964) (Bogomolnaya et al., 2008). The family V and VI are known to modify the O-antigen in some way (Davies et al., 2013) but the specific modifications are not known.

When comparing the O-antigen structures of all the serovars containing the family IV genes all but the Cubana serovar contained a mannose moiety and all but the O:7 group contained a galactose moiety (table 3). Importantly, there was no common sugar between the O-antigen structures. As the GtrC protein is predicted to transfer a sugar molecule to a specific position this means that either some or all of the core O-antigens need a further modification to provide a common target for GtrC(IV). On the other hand, the diversity in O-antigen structures could indicate that the target of the GtrC(IV) is not the O-antigen, which would mean that the *gtr*(IV) genes are not restricted to a particular O-antigen group. In table 3 the list of the other *gtr* genes contained within the strains showed that the family IV genes shared a genome with all six other families at some point. But in some strains they were found to be the only *gtr* gene within that genome.

Table 3: Family IV strains' *gtr* composition and their O-antigen group and structure

Family IV				
Serovar	**No. of strains**	**Other *gtrs* in strain**	**O-antigen group***	**O-antigen structure**
Tennessee	1	V/VI	O:7 (C1)	
Thompson	4			
Paratyphi C	1			
Infantis	1			
Choleraesuis	3	V/V/VI		
Typhimurium	18	III	O:4 (B)	
Agona	3			
Heidelberg	5			
Typhimurium	1 (Str. L-3553)	II/III		
	1 (Var. 5-Str. CFSAN001921)			
	1 (Str. D23580)			
	1 (Str. FORC_015)	V/VI		
YU39**	1	I/III	-	-
Bovismorbificans	1	VII	O:8 (C2-C3)	
Hadar	1			
Newport	12			
Anatum	4	-	O:3,10 (E1)	
Senftenberg	1	-	O:1,3,19 (E4)	
Ouakam	1	-	O:9,46 (D2)	
Cubana	1	-	O:13 (G)	

*The new O-group classification scheme is given first and the old scheme is in brackets.

3.3 ANALYSIS OF THE FAMILY IV GTRA PROTEIN

GtrA has been shown to be a hydrophobic protein with four transmembrane domains (Korres et al., 2005). It is proposed to be the UnDP-glucose translocase, which flips the UnDP-glucose across the membrane so that it can be accessed by GtrC (Nair et al., 2011). To get a better understanding of the function of the family IV operon a bioinformatics analysis of the GtrA(IV) proteins was done. This analysis aimed to determine whether the function of GtrA(IV) is as a UnDP-glucose translocase and how its AA sequence relates to GtrA proteins from operons with a known function. For example the family I and III operon's proteins are known to modify the LPS by adding a glucose molecule onto the O-antigen (Bogomolnaya et al., 2008) (Young et al., 1964).

All family IV GtrA AA sequences were aligned using the Cobalt Constraint-based Multiple Protein Alignment tool. The alignment (figure 4) shows the family IV GtrA proteins have an almost identical AA sequence with only two points having a variable AA. For this reason, only one sequence, from *S.* Typhimurium str. LT2, was used in further bioinformatics analysis.

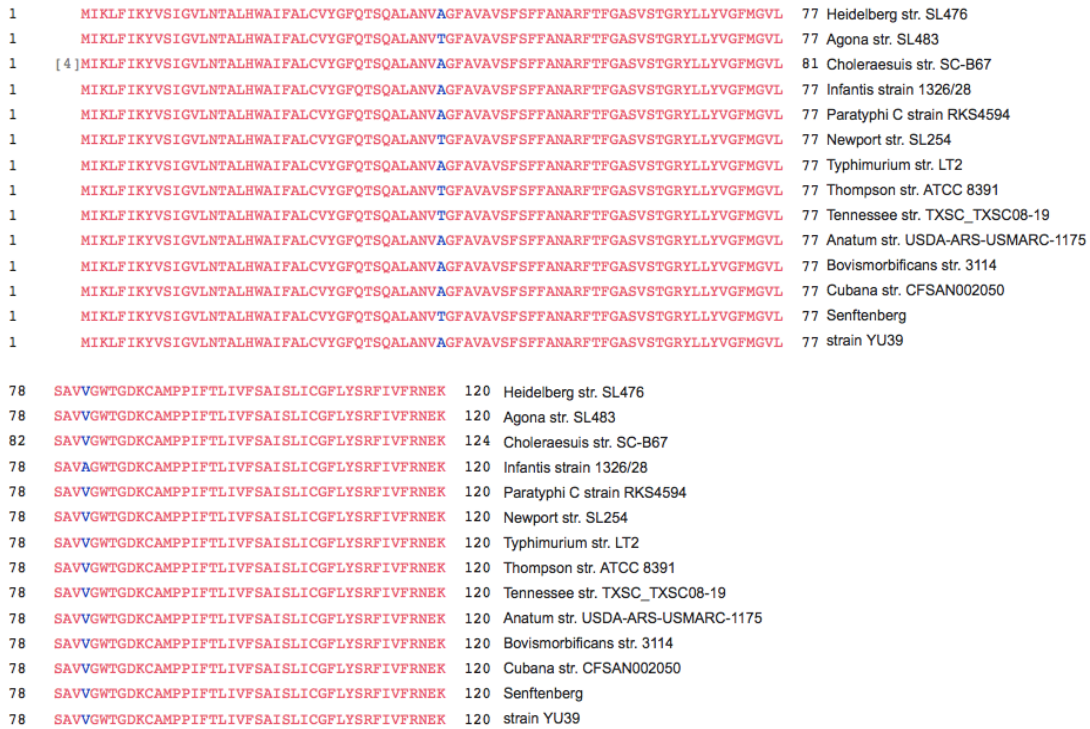


Figure 4: Alignment of GtrA(IV) AA sequences.

After each sequence is the name of the serovar and strain the sequence belongs to. The numbers either side of the sequences is the first and last AA number in that block. Numbers in grey within the sequence represents AA that are not shown and indicates how many. Red residues indicate that the AA is the same across all sequences and blue means that the AA changes across the sequences.

3.3.1 Functional analysis of family IV GtrA

A BLASTp analysis was used to identify proteins with similar AA sequence to the family IV GtrA protein so that a possible function could be identified. The search brought up many glucosyl translocases and bactoprenol-linked glucosyl translocases with high sequence identity to GtrA. Functional regions of the protein were identified using InterPro (Mitchell et al., 2014). This program identified a domain, which spans the majority of the protein, belonging to a GtrA-like family of proteins that trans locates undecaprenyl phosphate linked glucose (UndP-Glc) across the cytoplasmic membrane. These results match with the proposed function of GtrA as a bactoprenol-linked glucosyl translocase.

3.3.2 Transmembrane domain analysis of family IV GtrA

As GtrA proteins are predicted to be membrane proteins so that they can bind a sugar molecule and flip it across the membrane to be accessed by GtrC the GtrA(IV) protein was analysed to see if it has any transmembrane domains (see section 2.16 for methods). Like other GtrA proteins transmembrane domains (Korres et al., 2005), GtrA(IV) had four transmembrane domains (figure 5) with three cytoplasmic and two periplasmic loops. The transmembrane domains of the family I and III GtrA proteins also showed the same structure (figures not shown).

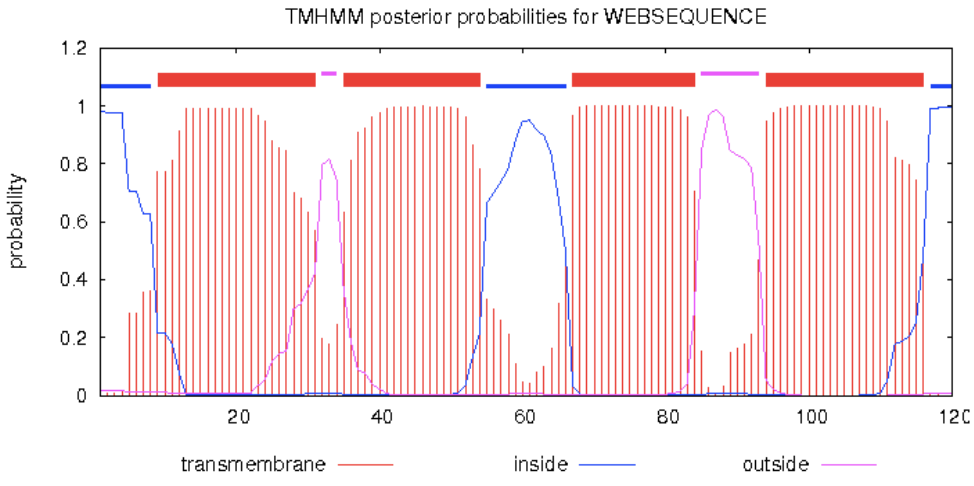
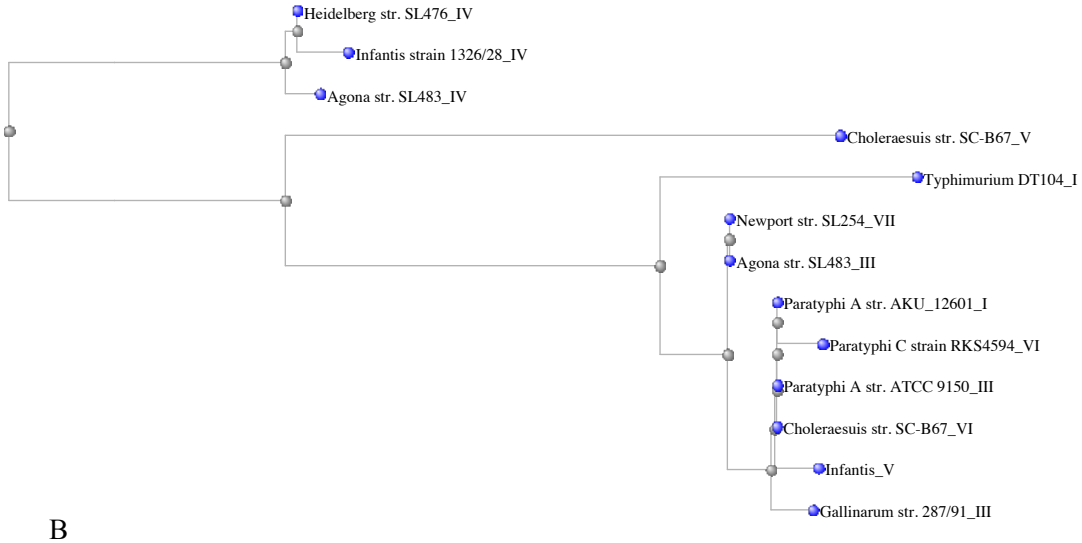


Figure 5: Transmembrane domain prediction of GtrA(IV).

Produced using TMHMM Server v. 2.0. The red lines show the transmembrane regions while pink and blue indicate regions outside and inside the cell respectively. The y axis shows the probability that the prediction is correct and the x axis gives the AA number.

A



B

```
1    ---MLKLFAKYTSIGVLNTLIHWGVFAFCVYGMHTHQALANFSGFVIAVSFSFYANARFTFNATTTTLRYMMYVGFMGTL  77  Typhimurium DT104_I
1    ---MLKLFAKYTSIGVLNTLIHWGVFAFCVYGMHTHQALANFSGFVIAVSFSFYANARFTFNASTTTLRYMMYVGFMGTL  77  Paratyphi A str. AKU_12601_I
1    mftMLKLFAKYTSIGILNTLIHWVVFAVCLYGLHTNQALANFAGFVIAVSFSFFANARFTFNASTTTMRYMLYIGFMGTL  80  Choleraesuis str. SC-B67_V
1    ---MLKLFAKYTSIGVLNTLIHWGVFAFCVYGMHTHQALANFSGFVIAVSFSFYANARFTFNASTTTLRYMMYVGFMGTL  77  Infantis_V
1    ---MLKLFAKYTSIGVLNTLIHWGVFAFCVYGMHTHQALANFSGFVIAVSFSFYANARFTFNASTTTLRYMMYVGFMGTL  77  Paratyphi A str. ATCC 9150_III
1    ---MLKLFAKYTSIGVLNTLIHWGVFAFCVYGMHTHQALANFSGFVIAVSFSFYANARFTFNASTTTLRYMMYVGFMGTL  77  Agona str. SL483_III
1    ---MLKLFAKYTSIGVLNTLIHWGVFAFCVYGMHTYQALANFSGFVIAVSFSFYANARFTFNASTTTLRYMMYVGFMGTL  77  Gallinarum str. 287/91_III
1    ---MLKLFAKYTSIGVLNTLIHWGVFAFCVYGMHTHQALANFSGFVIAVSFSFYANARFTFNASTTTLRYMMYVGFMGTL  77  Choleraesuis str. SC-B67_VI
1    ---MLKLFAKYTSIGVLNTLIHWGVFAFCVYGMHTHQALANFSGFVIAVSFSFYANARFTFNASTTTLRYMMYVGFMGTL  77  Paratyphi C strain RKS4594_VI
1    ---MLKLFAKYTSIGVLNTLIHWGVFAFCVYGMHTHQALANFSGFVIAVSFSFYANARFTFNASTTTLRYMMYVGFMGTL  77  Newport str. SL254_VII
1    ---MIKLFIKYVSIGVLNTALHWAIFALCVYGFQTSQALANVAGFAVAVSFSFFANARFTFGASVSTGRYLLYVGFMGVL  77  Heidelberg str. SL476_IV
1    ---MIKLFIKYVSIGVLNTALHWAIFALCVYGFQTSQALANVTGFAVAVSFSFFANARFTFGASVSTGRYLLYVGFMGVL  77  Agona str. SL483_IV
1    ---MIKLFIKYVSIGVLNTALHWAIFALCVYGFQTSQALANVAGFAVAVSFSFFANARFTFGASVSTGRYLLYVGFMGVL  77  Infantis strain 1326/28_IV

78   SAVVGWMADQCSLPPLITLITFSAISLVCGFIYSRFIVFRDIR  120  Typhimurium DT104_I
78   SAVVGWMADKCSLPPLLTLVTFSAISLICGFIYSKFIVFRDAK  120  Paratyphi A str. AKU_12601_I
81   SATVGWVADKSAFPPIITLITFSLISLICGFIYSKFIVFRDT-  122  Choleraesuis str. SC-B67_V
78   SAVVGWMADKCSLPPLLTLVTFSAISLTCGFIYSKFIVFRDAK  120  Infantis_V
78   SAVVGWMADKCSLPPLLTLVTFSAISLICGFIYSKFIVFRDAK  120  Paratyphi A str. ATCC 9150_III
78   SAVVGWMADKCSLPPLLTLITFSAISLICGFIYSKFIVFRDAK  120  Agona str. SL483_III
78   SAVVGWMADKCSLPPLLTLVTFSAISLICGFIYSKFIVFRDAK  120  Gallinarum str. 287/91_III
78   SAVVGWMADKCSLPPLLTLVTFSAISLICGFIYSKFIVFRDAK  120  Choleraesuis str. SC-B67_VI
78   SAVVGWMADKCSLPPLLTLVTFSAISLICGFIYSKFIIFRDAK  120  Paratyphi C strain RKS4594_VI
78   SAVVGWMADKCSLPPLLTLITFSAISLICGFIYSKFIVFRDAK  120  Newport str. SL254_VII
78   SAVVGWTGDKCAMPPIFTLIVFSAISLICGFLYSRFIVFRNEK  120  Heidelberg str. SL476_IV
78   SAVVGWTGDKCAMPPIFTLIVFSAISLICGFLYSRFIVFRNEK  120  Agona str. SL483_IV
78   SAVAGWTGDKCAMPPIFTLIVFSAISLICGFLYSRFIVFRNEK  120  Infantis strain 1326/28_IV
```

Figure 6: A comparison of GtrA proteins AA sequence from different families.

A. Is a phylogenetic tree created from the alignment of GtrA proteins (figure 3B) using the neighbor joining method. The blue nodes indicate each sequence used in the alignment and the name of the serovar and strain and the family the protein is from is given next to each node. B. Is an alignment of GtrA AA sequences from 6 different *gtr* families. After each sequence is the name of the serovar and strain the sequence belongs to and which *gtr* family it came from. The numbers either side of the sequences is the first and last AA number in that block. Red residues indicate that the AA is the same across all sequences and blue means that the AA changes across the sequences. The blue lines underneath the sequences indicate the transmembrane regions of the family IV GtrA protein.

3.3.3 Comparison of family IV GtrA with other family GtrA proteins

To compare GtrA(IV) to GtrA proteins of the other *gtr* families AA sequences from family I to VII were aligned, excluding family II. Family II functions as an acetyl transferase and the GtrC protein does require GtrAB for activity (Kintz et al., 2015). Before this, every GtrA AA sequence was aligned individually within their family. This analysis showed that within families the GtrA AA sequence is very conserved (figures not shown). To represent all variations in a family's GtrA sequences two to three sequences were taken from each family and used in the alignment in figure 6B. A phylogenetic tree was also generated from the alignment using the neighbor joining method (figure 6A).

Figure 6B showed that the family IV AA sequences had the most changes from the consensus sequences. The coloured regions indicated where the sequences had some conservation where red indicated a region conserved across all of the proteins and blue residues showed where some proteins have a change in AA. When studying the blue regions it is clear that the majority of AA differences came from the family IV sequence. Many of the AA changes, outside of predicted transmembrane regions, were changes in a property, mainly between hydrophobic and neutral. The phylogenetic tree produced from the alignment, where the family IV proteins were clearly the most distant family, supported the differences seen in family IV sequences. The tree also shows that the GtrA proteins of the other families had mixed together and didn't cluster into one group like the family IV proteins did, which is consistent with what Davies et al., saw when analysing the phylogenetic relationship between the *gtr* families. This observation could indicate that the family IV proteins have a function that is distinct from the other GtrA proteins for example they might be binding a different sugar to glucose.

3.4 ANALYSIS OF THE FAMILY IV GTRB PROTEIN

GtrB proteins have been previously shown to have two transmembrane domains (Korres et al., 2005) and its proposed function is as a glucosyl transferase, which transfers glucose from UDP-glucose to UnDP-glucose so that it can be flipped across the membrane by GtrA (Nair et al., 2011). As before, the purpose of this bioinformatics analysis is to establish whether GtrB(IV) functions as a glucosyl transferase and how its AA sequence relates to GtrB proteins from other *gtr* families.

`

The family IV GtrB AA sequences were aligned (figure 7) using the Cobalt Constraint-based Multiple Protein Alignment Tool (Papadopoulos and Agarwala, 2007b). The family IV GtrB proteins had the same AA sequence apart from four points where the residue varied between two AA. Because the sequences were so similar only one sequence was used in further bioinformatics analysis of GtrB. The AA sequence used was from *S*. Typhimurium LT2.

Figure 7: An alignment of GtrB(IV) AA sequences.

After each sequence is the name of the serovar and strain the sequence belongs to. The numbers either side of the sequences is the first and last AA number in that block. Red residues indicate that the AA is the same across all sequences and blue means that the AA changes across the sequences. The blue line indicates the glucosyl transferase 2-like domain identified by Interpro (Mitchell et al., 2014).

3.4.1 Transmembrane domain analysis of family IV GtrB

Based on their proposed function GtrB proteins are predicted to be membrane proteins. GtrB(IV) was analysed to see if it had any transmembrane domains (see section 2.16.). GtrB(IV) had two transmembrane domains (figure 8). The loop between them was in the periplasm with the c-terminal end and long N-terminal end in the cytoplasm. This transmembrane structure is the same as other GtrB transmembrane structures observed previousely (Korres et al., 2005). The transmembrane domains of the family I and III GtrB poteins, which are known to glycosylate the O-antigen (Bogomolnaya et al., 2008) (Young et al., 1964), also show that same structure as seen in figure 8 (figures not shown).

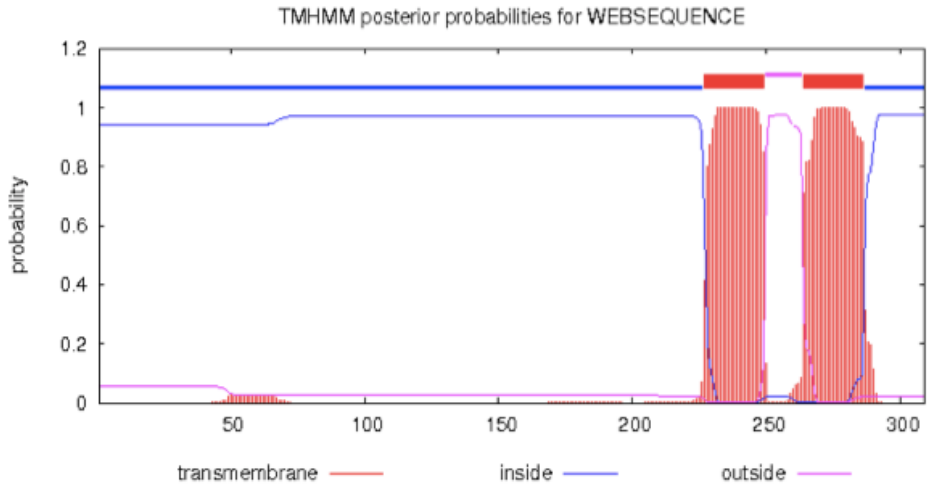


Figure 8: Transmembrane domain prediction of GtrB(IV).

Produced using TMHMM Server v. 2.0. The red lines show the transmembrane regions while pink and blue indicate regions outside and inside the cell respectively. The y axis shows the probability that the prediction is correct and the x axis gives the AA number.

3.4.2 Functional analysis of family IV GtrB

A BLASTp analysis was carried out to identify AA sequences similar to the family IV GtrB protein. BLASTp analysis matched GtrB with bactoprenol glucosyl transferase proteins with a high sequence identity. Interpro (Mitchell et al., 2014) identified a glycosyl transferase 2-like domain (figure 4) which is a domain found in a diverse family of glycosyl transferases that transfers UDP-glucose to a range of substrates.

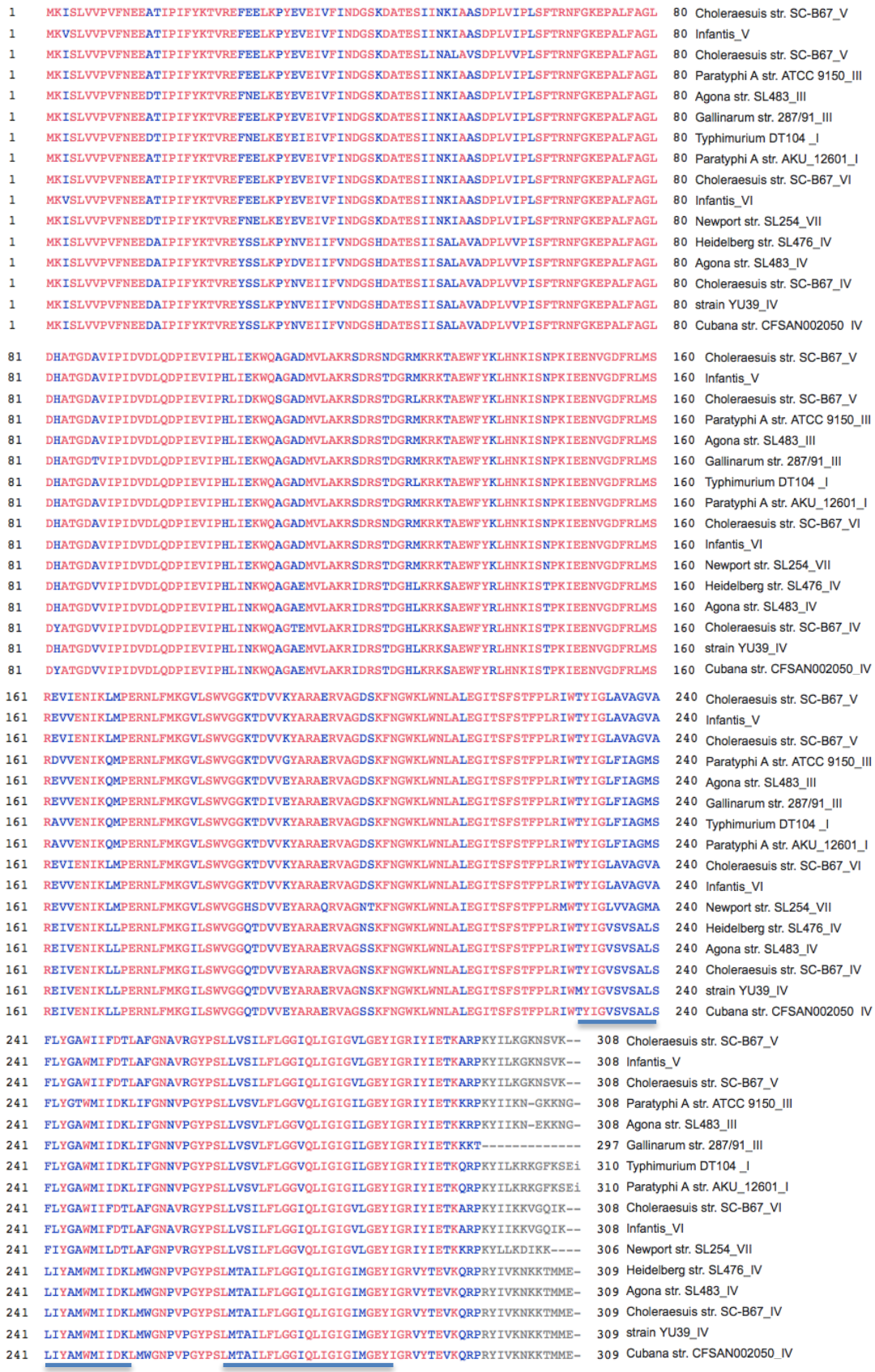


Figure 9: Alignment of GtrB AA sequences from 6 different Gtr families.

After each sequence is the name of serovar and strain the sequence belongs to and the *gtr* family it came from. Numbers either side of the sequences is the first and last AA number. Red residues indicate the AA is the same across all sequences and blue means the AA changes across the sequences. The blue lines underneath the sequences indicate the transmembrane regions of the family IV GtrB protein.

3.4.3 Comparison of family IV GtrB with other family GtrB proteins

An alignment of all the GtrB proteins was done to compare GtrB(IV) with GtrB proteins from other families. GtrB proteins used in the alignment were picked using the same method as with GtrA. A phylogenetic tree was also generated from the alignment using the neighbor joining method.

Figure 10 showed that the family IV AA sequences had the most changes from the consensus sequences as the majority of AA differences, shown in blue, came from the family IV sequence. Many of the AA changes, outside of predicted transmembrane domains, were changes in a property, mainly between hydrophobic and neutral or negative to neutral. This difference in sequences is supported by the phylogenetic tree produced from the alignment (figure 9) where the family IV proteins a clearly the most distant family. Similarly to GtrA(IV) this observation could indicate that the family IV GtrB proteins have a function that is distinct from the other GtrB proteins.

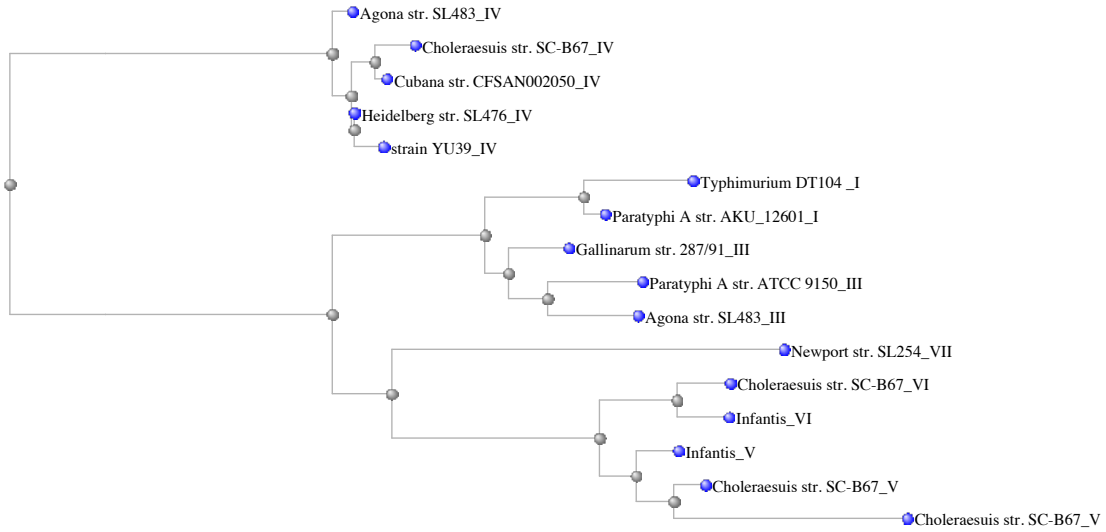


Figure 10: Phylogenetic tree of GtrB proteins.

Created from the alignment of GtrB proteins (figure 10) using the neighbor joining method. The blue node indicates each sequence used in the alignment and the name of the serovar, strain and the family the protein is from, is given next to each node.

3.5 ANALYSIS OF THE FAMILY IV GTRC PROTEIN

The GtrC protein is the protein that ensures specificity of the O-antigen modification and has been shown to add a particular sugar onto a certain moiety of the O-antigen using a specific linkage by GtrAB (Davies et al., 2013) (Young et al., 1964). For example the Gtr III proteins mediate the addition of a glucose molecule onto a galactose moiety at a C-4 position, this activity is specific to the GtrC(III) protein (Bogomolnaya et al., 2008). Therefore the proposed function of the GtrC proteins is as a glycosyl transferase that takes the sugar from GtrA that has been flipped across the membrane and transfers it onto the O-antigen in the periplasm. All GtrC proteins have 9-11 transmembrane regions and a long C-terminal tail in the cells periplasm, which is thought to have catalytic activity (Allison and Verma, 2000).

The family IV GtrC AA sequences were aligned (figure not shown) using the Cobalt Constraint-based Multiple Protein Alignment Tool. Like GtrAB, the AA sequences of all the family IV GtrC proteins had almost identical AA sequence. There was only three points in the consensus sequence where the residue varied between two AA. Because the sequences are so similar only one sequence was used in further bioinformatics analysis of GtrC. The AA sequence used was from *S.* Typhimurium LT2.

3.5.1 Transmembrane domain analysis of family IV GtrC

A transmembrane domain prediction (see section 2.16 for methods) was done using the GtrC *S.* Typhimurium LT2 sequence to look for transmembrane domains and a long C-terminal tail (figure 11). TMHMM transmembrane domain finder identified 9 transmembrane domains with a short N-terminal tail in the cytoplasm and a long 151 AA C-terminal tail in the periplasm. There were 4 periplasmic loops and 4 cytoplasmic loops.

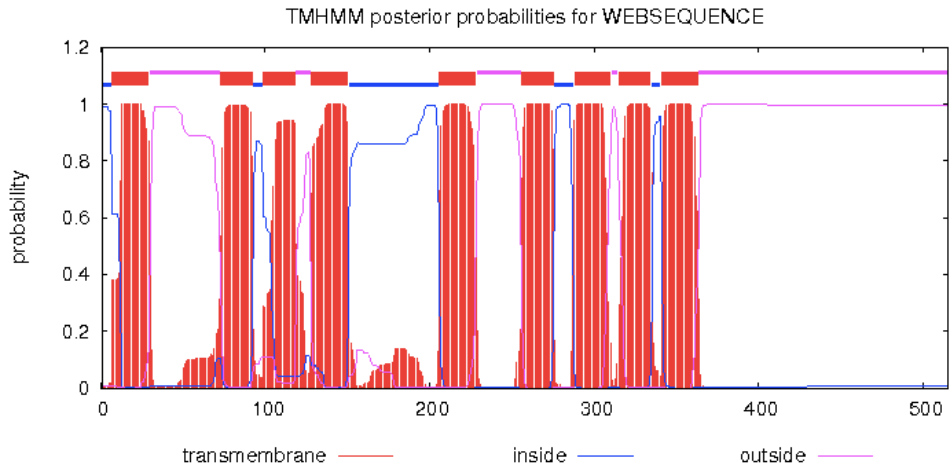


Figure 11: Transmembrane domain prediction of GtrC(IV).

Produced using TMHMM Server v. 2.0. The red lines show the transmembrane regions while pink and blue indicate regions outside and inside the cell respectively. The y axis shows the probability that the prediction is correct and the x axis gives the AA number.

3.5.2 Functional analysis of family IV GtrC

Interpro (see section 2.16. for methods) results for family IV GtrC found no functional domains. Although this result is not unexpected as the family III GtrC, which is known to modify the O-antigen by transferring a glucose (Bogomolnaya et al., 2008), also has no predicted functional domains when analysed by Interpro. The BLASTp results for GtrC were not conclusive, no glucosyl transferase proteins were identified and the search findings mostly consisted of membrane proteins of unknown function from *S. enterica*. However there was one putative GtrC protein identified from Salmonella phage epsilon34 and they had a low sequence identity of 30%. This protein has been found directly downstream of proposed *gtrAB* genes. Although its function hasn't been proven the epsilon34 bacteriophage is known to modify Salmonella's surface polysaccharide by adding glucose onto a galactose residue (Wright, 1971) (Villafane et al., 2008). When this protein is analysed using Interpro no functional domains are found.

It may be the case that the epsilon 34 phage causes a modification to host bacterial O-antigen, which is the same as the modification mediated by the family IV proteins as they have some sequence identity in their GtrC. However the sequence identity is only low and might be significant only if the sequence similarity is just within the family IV proteins. BLAST analysis within the Salmonella phage epsilon34 genome was done with GtrC proteins from the other families. This analysis showed that the family III GtrC protein had a 39% sequence identity, higher than the family IV GtrC protein. Therefore, the sequence identity could be due to a similarity in function as a sugar transferase rather than reflecting a specific function.

3.5.3 Comparison of family IV GtrC with other family GtrC proteins

An alignment of all the GtrC proteins was done to compare GtrC(IV) with GtrC proteins from other families. Before this, every GtrC AA sequence was aligned individually within their family. This analysis showed that within the families the GtrC AA sequence is very conserved (figures not shown). To represent all variations in a family's GtrC sequences two to three sequences were taken from each family and used in the alignment in figure S2. A phylogenetic tree was also generated from the alignment using the neighbor joining method (figure 12).

The alignment (figure S2) and the phylogenetic tree produced (figure 12) showed that the family IV GtrC protein is more closely related to the family I and III GtrC proteins than the other Gtr families and there is a higher sequence similarity between GtrC's from family III and IV than between family III and family I.

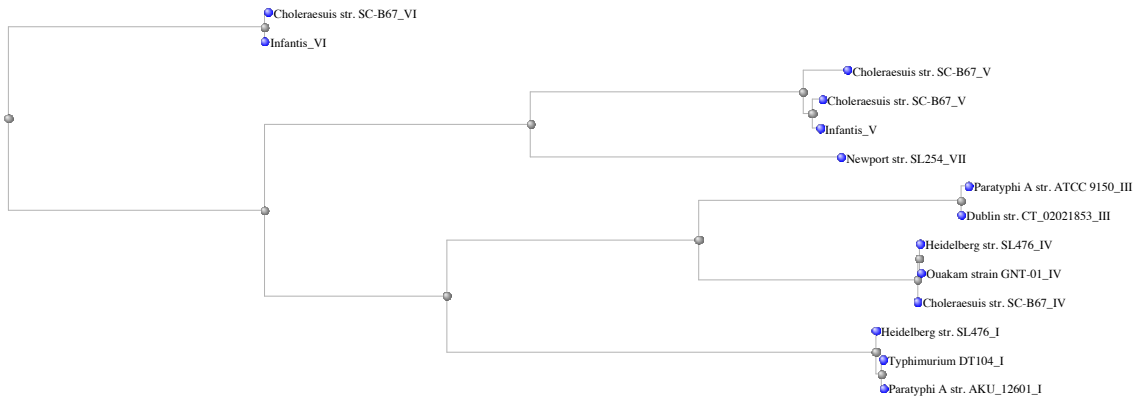


Figure 12: Phylogenetic tree of GtrC proteins.

Created from the alignment of GtrC proteins from 6 families in figure A3.1 using the neighbor joining method. The blue node indicates each sequence used in the alignment and the name of the serovar and strain and the family the protein is from is given next to each node.

A second alignment was done using only the family I, III and IV protein sequences so that their similarities could be more clearly seen (figure 13). The transmembrane regions of each family were plotted underneath the alignment: red, blue and orange for family I, III and IV respectively. The glucosyl transferase II domain within GtrC(I), predicted by InterPro (Mitchell et al., 2014), is indicated above the alignment by a green line. The glucosyl transferase II domain is found within the GtrC equivalent of Shigella phage SfII, which mediates the addition of a glucose molecule onto the third rhamnose of the *Shigella flexner* O-antigen (Mavris et al., 1997). Alignments of the C-terminal tails of the proteins were also done and orange boxes highlighted clusters of conserved AA identified by these alignments.

In figure 13 the coloured regions indicate that the sequences have some conservation where red indicates AA conserved across all of the proteins and blue residues shows that some proteins have a change in AAs. It is clear that the N-terminal part of the proteins, including the glucosyl transferase II domain within family I, have higher conservation than the C-terminal tail of the protein. This might indicate that the C-terminal tail is the specific part of the protein and relates to the position of the modification on the O-antigen repeating unit.

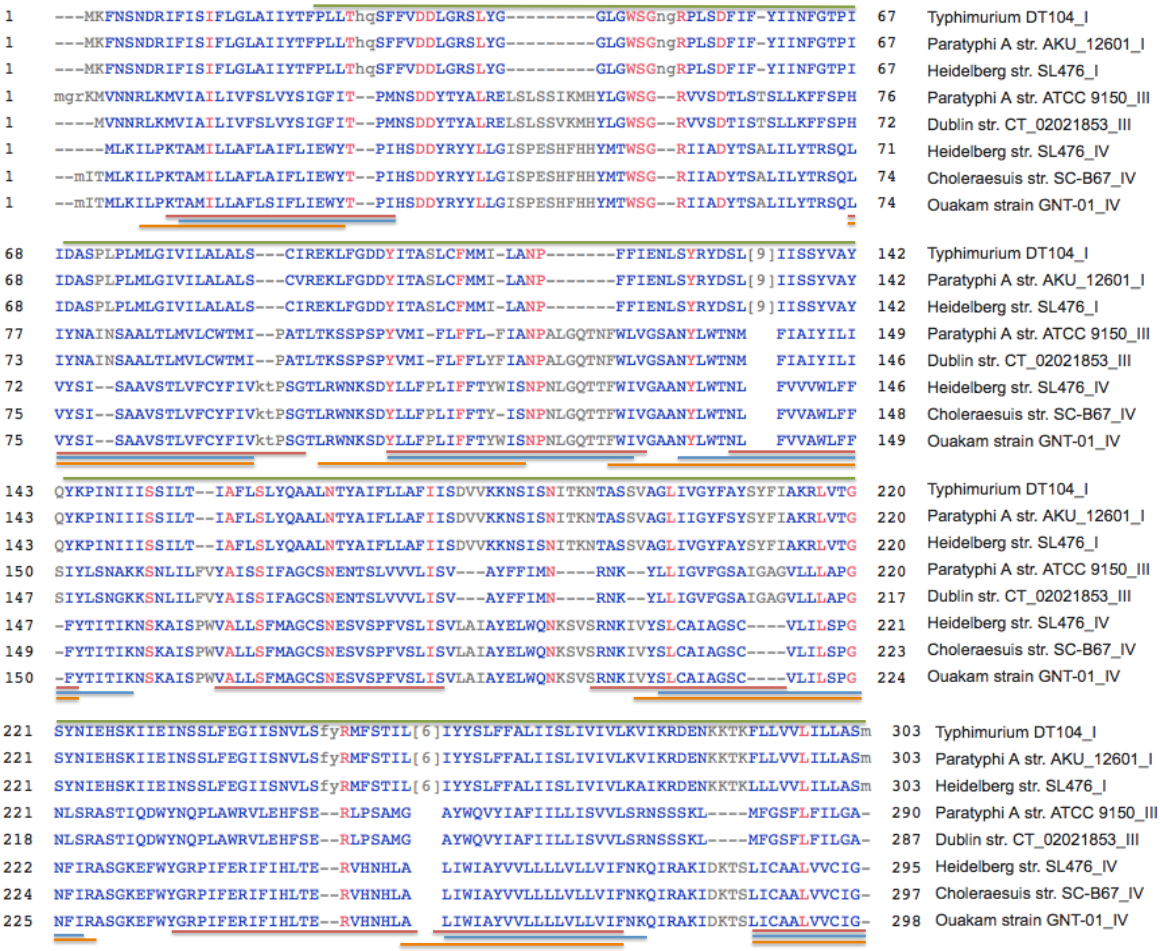


Figure 13. Continued on the next page

Figure 13: Alignment of GtrC AA sequences from Gtr families I, III and IV.

After each sequence is the name of the serovar and strain the sequence belongs to and which *gtr* family it came from. The numbers either side of the sequences are the first and last AA number in that block. Numbers in grey within brackets represents AA that are not shown and indicates how many. Lower case AA in grey show low-complexity sequences that have been filtered out when building the alignment. The coloured regions indicate that the sequences have some conservation where red indicates AA conserved across all of the proteins and blue residues shows that some proteins have a change in AAs. The transmembrane regions of each family were plotted underneath the alignment: red, blue and orange for family I, III and IV respectively. The glucosyl transferase II domain within GtrC(I), predicted by InterPro (Mitchell et al., 2014), is indicated above the alignment by a green line. Clusters of conserved AA in the C-terminal tail between the family III and IV proteins are highlighted by an orange box.

Transmembrane regions cover a lot of the N-terminal covered by the glucosyl transferase II domain. The only region not covered by a transmembrane domain and containing clusters of conserved AA is between the first two transmembrane domains. This region has eight conserved AA. This could indicate that this region is involved in a glucosyl transferase activity that is common between all proteins, for example binding the glucose. The C-terminal domain, which has less sequence conservation than the N-terminal, still has four residues conserved across all proteins, which are spread across the C-terminal rather than clustered. But between the family III and IV C-terminal tail there are two large clusters of conserved AA

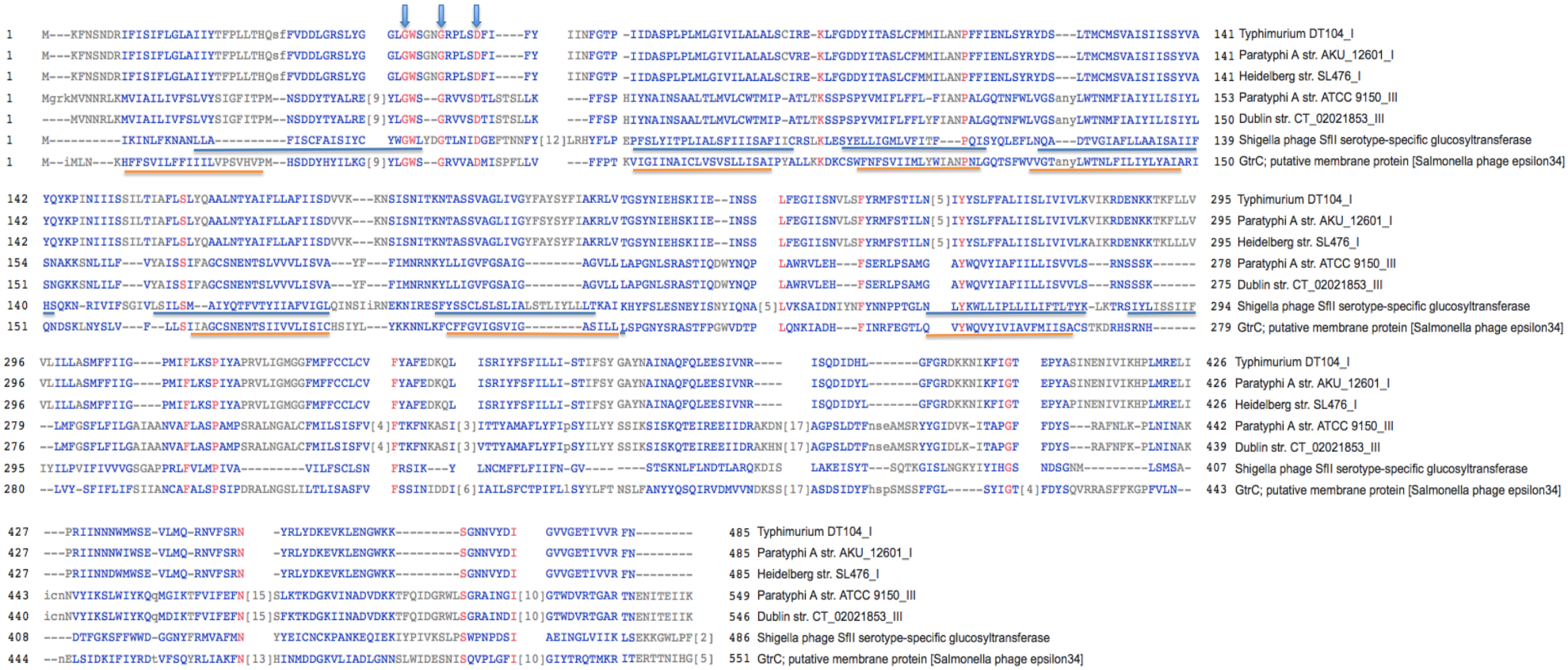


Figure 14: Alignment of glucosyl transferase proteins of known function.

After each sequence the origin of the sequence is specified, the first five sequences are all from Salmonella genomes and the name of the serovar and strain the sequence belongs to and which *gtr* family it came from is indicated. The numbers either side of the sequences are the first and last AA number in that block. Numbers in grey within brackets represents AA that are not shown and indicates how many. Lower case AA in grey show low-complexity sequences that have been filtered out when building the alignment. The coloured regions indicate that the sequences have some conservation where red indicates AA conserved across all of the proteins and blue residues shows that some proteins have a change in AAs. The transmembrane regions in the Shigella phage SfII glucosyltransferase are indicated in blue and the epsilon 34 GtrC protein's transmembrane domains are indicated in orange. Transmembrane domains were predicted using TMpred (http://www.ch.embnet.org/software/TMPRED_form.html). Conserved AA that were identified in the alignment of GtrC proteins from family I, III and IV (figure 10) are are indicated with blue arrows.

3.6 ANALYSIS OF KNOWN GLUCOSYL TRANSFERASE PROTEINS

To identify AAs involved in binding the glucose molecule in a glucosyl transferase protein an alignment of known glucosyl transferase proteins was generated. The proteins used in the alignment were the family I and III proteins used in the previous alignments, the family VIII GtrC proteins from the *Salmonella* phage epsilon 34 and the GtrC equvelent, serotype-specific glucosyl transferase protein from the SfII *Shigella* phage.

The alignment had one cluster of conserved AA at the N-terminal end of the proteins which might be involved in binding a glucose molecule. The AA cluster, indicated by arrows in figure 11, were also present in GtrC(IV). Part of the cluster of conserved AA (figure 14) were covered by a predicted transmembrane domain in the Shigella phage SfII GtrC equivalent. But Glycine and aspartic acid, part of the cluster of AAs, were predicted to be in the periplasm in all GtrC proteins studied. This alignment could indicate that these two AA might be needed to bind a gluc

3.7 CHAPTER CONCLUSION

3.7.1 Summary and conclusion of the spread of the family IV genes and the O-antigen structure of their serovars

Analysis of the family IV serovar's O-antigen structures showed that there is no common sugar between the serovars that contain them. This was reflected in the number of O-antigen groups that the family IV genes were found within. Unlike the other *gtr* families, who were found within a maximum of three O-antigen groups, family IV was found within eight different groups. This could suggest that the O-antigen is not the target of GtrC(IV). Or this could be because the family IV genes provide a selective advantage to the bacteria and has been well maintained within populations.

If GtrC(IV) is modifying the O-antigen at a specific moiety then further modifications to of all or some of the O-antigens basic structure would be needed. For example, if the Cubana serovar's O-antigen had a manose modification added to its repeating unit structure this would mean that there was a common O-antigen sugar between the serovars. Other *gtr* families could be providing a common modification to the basic O-antigen structure. Although, as some family IV serovars only contain the family IV *gtr* genes, this would not be the case for all serovars. As all but the O:7 group of serovars contain a galactose moiety it may be the case that the family V or VI gtr genes, contained within the serovars of group O:7, are mediating the addition of a galactose residue onto the O-antigen.

Alternatively, the difference in O-antigen structures could reflect that the O-antigen is not the target of the family IV proteins and they are targeting something else in the periplasm. Or it could be that GtrC(IV) has a broader range of activity and does not have a specific target. Both of these hypotheses could explain why the family IV genes are in a largest number of serovars and O-antigen groups as their activity would not be restricted by O-antigen structure.

Another contributing factor to their wide spread across serovars may be because they are required for persistence within a host (Chaudhuri et al., 2013) and for this reason have been well maintained within a strains genome.

3.7.2 Summary and conclusion of the function of the family IV proteins

The bioinformatics analysis of the family IV proteins has shown that they are all transmembrane proteins (figures 2., 5. and 8.) and seem to function broadly as predicted based on the model of the Gtr proteins activity. Based on the BLASTp InterPro results the GtrA protein is predicted to function as a glucosyl translocase that translocates UndP-Glc across the cytoplasmic membrane. GtrB is predicted to function as a glycosyltransferase that transfers UDP-glucose to a range of substrates. Based on GtrC's sequence similarity with the family III and I GtrC proteins it is predicted to function as a glucosyl transferase as both families are known to add a glucose onto the O-antigen repeating unit.

The alignments of six of the families GtrA and GtrB proteins (figures 3B. and 7.) showed that there was a lot of conservation between their AA sequences. This is expected, as these are the non-specific proteins that are expected to have broadly the same function. However, the family IV GtrA and GtrB proteins were both shown to have the most AA differences in its sequence compared to the other family GtrAB proteins whose sequences were more conserved with each other. This could clearly be seen in the phylogenetic trees generated from the protein's alignments (figures 3A. and 6.). This could be because the functions of the family IV proteins are slightly different from the other family IV GtrAB proteins. For example, they might bind a different sugar that isn't glucose. This hypothesis is supported by the observation that the family IV operon has not recombined with other family operons, unlike the other *gtr* families (Davies et al., 2013). If GtrAB(IV)'s function was different then recombination would make a non functional operon.

In contrast to the GtrAB(IV) proteins the GtrC(IV) protein had a higher sequence similarity with GtrC(I) and GtrC(III) than the other families. The GtrC(IV) and Gtr(III) protein were very closely related and had a higher sequence similarity than GtrC(III) and GtrC(I). This is interesting as both families I and III are known to mediate the addition of glucose onto different positions on a galactose moiety.

Therefore, the GtrC(IV) may have a very similar function to GtrC(III) and also bind and transfer a glucose sugar onto a galactose moiety. The galactose sugar is part of the core structure of all but the O:7 group of serovars occupied by the family IV genes. All of these O:7 group serovars contain the family V and VI *gtr* genes. It is not known what modifications the family V and VI proteins make to the O-antigen and could possibly be a galactose sugar, providing a common sugar amongst the O-antigens of the family IV serovars.

The alignment in figure 13 of family I, III and IV GtrC proteins showed clusters of conserved AA between all three proteins located between the first two transmembrane domains. Between all three proteins the C-terminal tail had less sequences conservation. This could be because the C-terminal tail is involved in specifically adding the sugar onto the O-antigen repeating unit whilst the clusters of conserved AA at the N-terminal could be involved in binding the sugar. For family I and III this in known to be the same sugar, a glucose, therefore there is expected to be a conserved glucose binding region. The conserved AA between the first two transmembrane domains in the alignment of GtrC proteins from families I, III and VI (figure 13), could be the AA necessary for binding a glucose molecule. As the family IV protein also shows conserved AA within this region it may be the case that GtrC(IV) also binds a glucose.

To strengthen this theory, glycosyl transferase proteins that are known to bind a glucose molecule, including the family I and III proteins, were aligned (figure 14). This alignment had just one clustering of conserved AA, which is likely to be the AA involved in binding the glucose molecule as this is the functional commonality between these proteins. The alignment identified two conserved AA predicted to be in the periplasm in all GtrC proteins studied in section 3.5. and could be involved in binding the glucose molecule. These AA are also conserved in the periplasm of GtrC(IV), which could indicate that it is also binding a glucose molecule.

The C-terminal tails of the family III and IV proteins has two clusters of conserved AA that aren't present in the family I proteins. This could indicate that the family III and IV proteins have a very similar function. The only difference between family I and III GtrC proteins is that GtrC(I) mediates a 1,6 linkage and GtrC(III) mediates a 1,4 linkage of glucose to the

galactose moiety. Therefor, it may be that family IV makes the same linkage as family III and mediates the addition of a glucose to a galactose sugar at a 1,4 position.

4. EXPRESSION OF THE FAMILY IV *GTR* OPERON UNDER INFECTION RELEVANT CONDITIONS

The family IV *gtr* genes have been shown to be required for persistence within a host (Chaudhuri et al., 2013) but we know nothing about what advantage they provide during an infection. To start to answer this question it would be useful to know at what point the operon is expressed when the bacteria is infecting a host. This chapter investigates whether any infection relevant conditions cause an up regulation in the family IV *gtr* genes expression. This information could give an indication of when the family IV Gtr proteins are needed during the infection process (discussed in section 1.4.2.).

The expression of the family IV *gtr* genes from a constitutive Ptac promoter in a basal *S.* Typhimurium strain does not cause a detectable modification of the LPS (Davies et al., 2013). One of the reasons for this may be because the family IV genes constitutively expressed mRNA is not being translated into protein due to post transcriptional controls under normal laboratory growth conditions. If the protein's function is important for growth and persistence during an infection then they may not be needed when grown under normal laboratory conditions. For example, the activity and production of the proteins may be energetically costly for the cell so their translation is restricted to times when they are need for persistence within a host. If the strain constitutively expressing the family IV genes was grown under conditions that triggered up regulation of the operon then this is likely to be a condition under which the cell would translate the transcript into protein. These conditions could then be used when growing up strains that are constitutively expressing the gtr(IV) genes and analysing their LPS banding pattern to look for modifications of the O-antigen.

A recent study (Kröger et al., 2013) has done a global transcriptome analysis of *S.* Typhimurium under various infection relevant conditions using an RNAseq method. The results for the family IV *gtr* operon showed that the point at which *gtrA* had the highest

recorded number of transcripts was under NO shock and *gtrC* is also up regulated under NO shock. Another condition where up regulation of *gtrC* is seen is pH3 shock where *gtrC*'s highest level of up regulation is seen as well as one of the highest transcript numbers.

4.1 EXPRESSION OF FAMILY IV *GTR* GENES UNDER PH3 SHOCK

One of the aims of these experiments was to gain an insight into when the family IV genes are needed during infection by testing whether certain infection relevant conditions cause up regulation of expression. This experiment tested whether pH3 shock would cause up regulation of the family IV genes. pH3 conditions are relevant to infection because this is the pH of the gastric system of humans (Goktepe et al., 2005) and other common hosts of *S.* Typhimurium like pigs (Crévieu-Gabriel et al., 1999; Wilkins et al., 2010) is around pH3. The Kroger et al., (Kröger et al., 2013)data showed that *gtrC* and *gtrA* had some of the highest transcript levels under pH3 shock compared to the other conditions measured. Also, the data produced by Chaurdhuri et al., (Chaudhuri et al., 2013), which showed that the family IV genes were important for persistence within a host, found that *gtrAB* were only necessary for persistence in the hosts that were orally infected, cows, chickens and pigs. The mouse, the host where *gtrAB* were not required for persistence, was infected intravenously.

To measure the levels of gene expression under pH3 shock total RNA was extracted from a culture of wild type *S.* Typhimurium (strain 380) that had been shocked with pH3 and from a control culture that had been kept at pH7 (see sections 2.11. and 2.13. for pH3 shock and RNA extraction methods). Reverse transcription (section 2.14.) was then performed on the RNA samples and the resulting cDNA was used from RT-qPCR analysis (section 2.15.). Comparative RT-qPCR was used and two genes that had a constant expression across conditions of pH7 and pH3 were used as reference genes. The two reference genes were *yibT* and *modF*. These genes were picked from *S*. Typhimurium's global transcriptome data which showed transcript levels under various conditions (Kröger et al., 2013). Both *yibT* and *modF* were chosen because their expression levels were constant between conditions of pH7 and pH3.

Relative RT-qPCR uses the levels of the reference gene's transcript to normalize the levels of the transcript of interest between the treated and control samples. This gives a value of expression level as relative quantification (RQ). This is because we assume the reference gene's expression will not change between the treated and control, as mentioned above the reference genes expression was constant between pH7 and pH3 conditions. Consequently, we can only compare between samples where the only difference between them is the conditions controlled for by the reference genes. When other conditions vary, such as time, only fold change can be compared between samples. The relative quantification of the control sample is always put to 1 as this sample represents a control level of gene expression which is expected to stay constant.

4.1.1 Expression of family IV *gtr*C ten minutes after pH3 shock

Figure 15 shows *gtrC* expression measured by RT-qPCR of samples that were acidified to pH3 and grown for 10 min compared to samples that were kept at pH7. The two biological replicates did not show the same trend in expression. Replicate 1 pH3-shocked sample had a slightly higher but insignificant change in *gtrC* expression level (a fold increase of 1.24) compared to the sample kept a pH7. In contrast, replicate 2 showed a significant decrease (3.06 fold change) in *gtrC* expression level under pH3 shock compared to the level of expression under pH7 conditions.

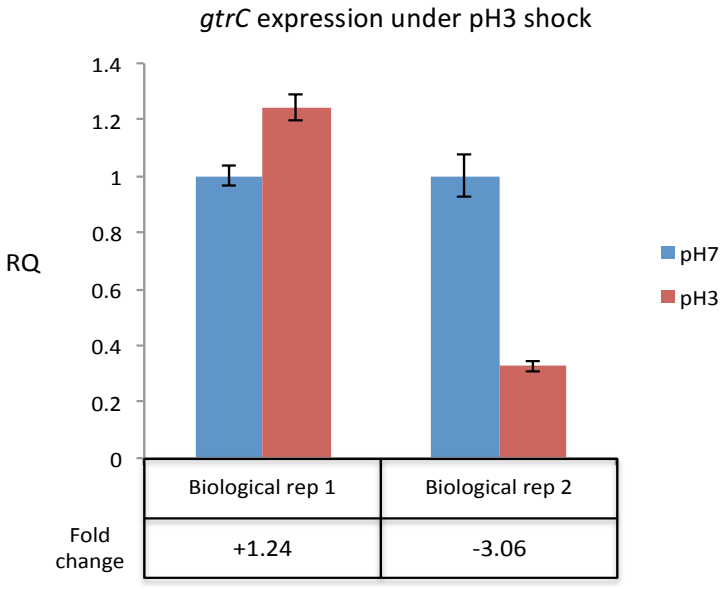


Figure 15: *gtrC* expression under pH3 shock.

Relative expression levels measured by RT-qPCR of *gtrC* under pH3 shock and the control condition of pH7 in biological replicates. Red bars relate to pH3 condition and blue bars to pH7 condition. The y-axis shows relative quantification of *gtrC* relative to the reference gene levels. The x-axis indicates which biological replicate the samples are from. Values in the boxes below the x-axis give the fold change in relative *gtrC* levels between pH3 and pH7 treated samples in the biological replicate. A + in front of the fold change value indicates the levels of *gtrC* under pH3 shock have increased compared to the levels at pH7 and a – means they have decreased.

4.1.2 Expression of family IV *gtr*A and *gtr*C at multiple time points under pH3 shock

The variation between the biological replicates in figure 15 could have been due to small differences in the time each sample's RNA was protected (see section 2.13. for methods) or fixed in the sample. Some bacterial mRNA transcripts can have a half life as short as 40 sec (Richards et al., 2008). The following experiment was designed so that RNA was extracted from a time course of samples to see how the levels changed over a short period of time. RNA was extracted at 5, 10 and 15 minute intervals.

The *gtr(IV)* operon was expected to be transcribed together from a promoter upstream of *gtrA*. The Kroger et al., (Kröger et al., 2013)data showed that there were differences in the *gtrABC* transcript levels of the family IV operon under specific conditions. To compare the levels of expression of the family IV operon's genes, both *gtrC* and *gtrA* expression levels were measured.

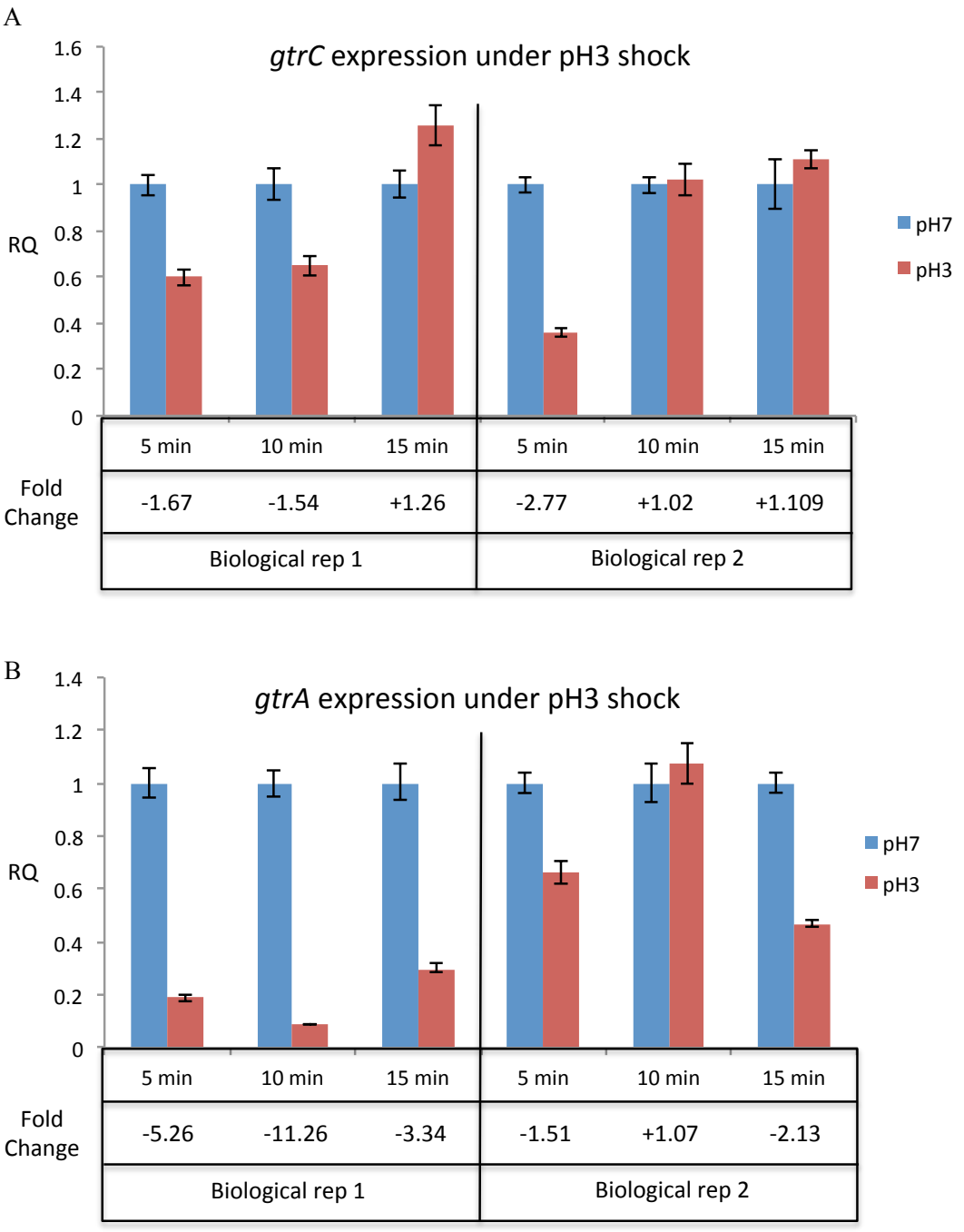


Figure 16: The relative expression levels of *gtrC* and *gtrC* under pH3 shock.

Expression was measured by RT-qPCR, under pH3 shock and the control condition of pH7 in two biological replicates at 5, 10 and 15 minutes after shock. The red bars relate to the pH3 condition and the blue bars relate to the pH7 condition. The y-axis shows the relative quantification of *gtrC*, which is relative to the reference gene levels. The x-axis gives the number of minutes after pH3 shock the samples were taken. The values in the boxes below the x-axis give the fold change in relative *gtrC* levels between the pH3 treated and pH7 treated samples in the biological replicate. A + in front of the fold change value indicates the levels of *gtrC* under pH3 shock have increased compared to the levels at pH7 and a – means they have decreased. The bottom boxes indicate which biological replicate the samples were taken from. A. *gtrC* expression under pH3 shock: shows the level of expression of the *gtrC* transcript. B. *gtrA* expression under pH3 shock: shows the level of expression of the *gtrA* transcript.

In figure 16A there was no significant fold increase in expression at pH3 compared to pH7. Both biological replicate 1 and 2 showed that *gtrC* fold change in expression between pH3 and pH7 conditions increased over the 15-minute period. In both replicates the 5-minute sample shows the biggest fold decrease in expression at pH3. In replicate 1 the fold change at 10 min stayed constant with the 5 min sample. The sample taken at 10 minutes in replicate 2 showed no fold change in *gtrC* expression between conditions of pH3 and pH7. At 15 minutes both replicate samples had no fold change in expression showing that *gtrC* expression was similar under both conditions. (Figure 16A)

The results in figure 16A show that, in both samples, the relative level of *gtrC* drops after pH3 shock but returns to the initial relative level after 15 min and that *gtrC* fold change in expression between pH3 and pH7 conditions can change over a period of 5 minutes. This supports the conclusion that the differences seen between the biological replicates in figure 15 may have been due to difference is the time taken to protect the RNA.

The level of gtrA (figure 16B) in biological replicate 1 consistently had a significant fold decrease in expression under pH3 conditions compared to pH7, at 10 min there was an 11-fold decrease in *gtrA* expression. In Biological replicate 2 the 5 minute and 10 minute sample did not have a significant change in *gtrA* expression between conditions of pH3 and pH7. At 15 minutes *gtrA* expression has decreased 2.13 fold in the pH3 shocked sample compared to the pH7 sample. There is a difference in the pattern of *gtrA* expression between the two biological replicates where biological replicate 1 showed a much bigger fold decrease of *gtrA* expression under pH3 than biological replicate 2, particularly in the 10 min sample.

A comparison of figure 16A and 16B shows that in biological replicate 1 there was a large difference in the expression patterns of *gtrA* and *gtrC*. *gtrA* had a greater difference in expression levels between conditions of pH3 and pH7 compered to the change in *gtrC* expression levels. Comparing *gtrA* and *gtrC* in biological replicate 2 their trends in expression were mostly similar. But At 15-minutes *gtrA*'s expression under pH3 conditions had decreased compared to the expression under pH7, whereas *gtrC* expression didn't changed between the two conditions.

4.2 EXPRESSION OF FAMILY IV *GTR* GENES UNDER NITRIC OXIDE (NO) SHOCK

These experiments looked at whether NO shock could cause an up-regulation of the family IV *gtr* operon. It was seen from the Kroger et al., (Kröger et al., 2013) data that cultures grown under nitric oxide shock had a 3.76 fold increase of *gtrA* compared to a control growth condition. As discussed in section 1.4.2., when *Salmonella* is infecting a host it will experience conditions of NO shock from the immune response mounted against it.

4.2.1 Expression from the family IV *gtr*A upstream region under NO shock

The effect of NO shock on expression from the up stream region of *gtrA* was studied by cloning the upstream region in front of the *lacZ* gene, a transcriptional reporter, and measuring the levels of LacZ produced using a beta-galactosidase assay (see section 2.12. for methods). The Conditional-Replication, Integration, Excision, and Retrieval (CRIM) vector (Haldimann and Wanner, 2001) was used to integrate the family IV regulatory region-*lacZ* construct into the genome of *S.* Typhimurium (see section 2.8 for methods). Strain 917 was created with a 355bp region upstream of the *lacZ* gene intograted into the chromosome. The 355bp region included the upstream region of *gtrA* and overlapped the *gtrA* start site.

Strain 917 was grown in Salmonella pathogenicity island II inducing media (InSPI2) media and shocked with NO at OD600 0.3 by the addition of spermine NONOate (see section 2.10. for methods). Before the addition of NO, the culture was split and one half was taken as a control, which received no NO shock. Samples were taken just before NO addition. Then again at 20 min after, which is when samples were taken for RNAseq analysis by Kroger et al., (Kröger et al., 2013) , and finally 60 min after NO addition to look for any long-term effects of the operons expression. Three biological replicates were used in this experiment and four technical replicates were taken from each biological replicate.

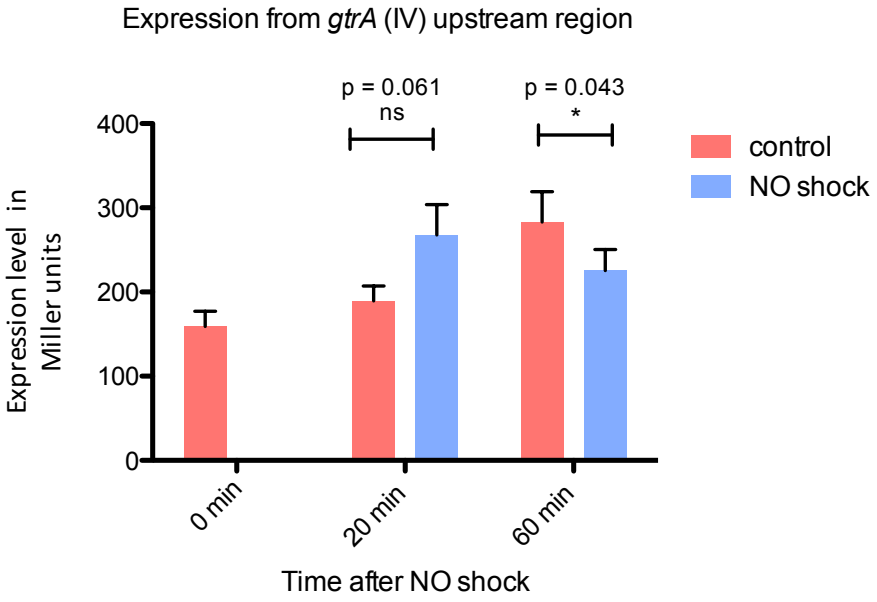


Figure 17: Expression of *gtrA* under NO shock measured through the *lacZ* reporter gene.

The level of beta-galactosidase measured in Miller units produced from strain 917 is shown on the y-axis. Strain 917 has the *lacZ* gene down stream of a 355pb *gtrA(*IV) regulatory region integrated into its genome. On the x-axis is the number of minutes after NO shock the sample was taken. 0 minutes indicates it was taken just before NO addition. The blue bars indicate the sample was shocked with NO and the red bars indicate a control sample that received no NO shock. A t-test was performed on the data from the 20-minute sample and the 60-minute sample with a significance of $p<0.05$. The P value is above the corresponding bars.

At 20 minutes after NO shock the level of beta-galactosidase expression, measured in Miller units, in the treated sample has increase compared to the control sample but the difference between their means is not significant (figure 17). The fold change in expression that was seen in the 20-minute sample was lower than the results reported by Kroger et al., (Kröger et al., 2013) who found a 3.76 fold increase in gtrA expression. At 60 minutes after NO shock beta-galactosidase expression in the treated sample has decreased compared to the control sample with a significance of $p=0.043$ (figure 17).

4.2.2 Expression of family IV *gtr*A under NO shock

To determine whether the increase seen in figure 17 at 20 min after NO shock is significant more biological replicates would need to be done. But the increase was lower than expected based on the Kroger et al., (Kröger et al., 2013) results, which showed a 3.76 fold increase in *gtrA* expression under NO shock. This could have been due to the differences in the beta-galactosidase assay and the RNAseq methods used by Kroger et al., (Kröger et al., 2013). RNAseq measures the genes transcript level whereas a beta- galactosidase assay is measuring protein which will have a lag in production compared to its corresponding mRNA. Therefore, a time point slightly after 20 minutes might have shown an increase in beta-galactosidase. Another reason for not seeing the increase in expression that was expected could have been that the region cloned in front of the *lacZ* gene did not contain the whole promoter region. To overcome this problem RT-qPCR was used. RT-qPCR directly measures the mRNA level of a gene transcribed from the chromosome so will show the actual level of transcript available to be translated into protein.

A culture of wild type *S.* Typhimurium (strain 380) was grown in InSPI2 media and shocked with NO at OD600 0.3 by the addition of spermine NONOate (see section 2.10. for methods). Before the addition of NO, the culture was split and one half was taken as a control, which received no NO shock. Samples from both the control and the treated sample were taken at 20 minutes after shock. Two biological replicates of this experiment were done. RNA was extracted from these samples and reverse transcription was performed before the resulting cDNA was used in an RT-qPCR experiment. The level of *gtrA*'s transcript was measured in this experiment. Out of the three family IV genes *gtrA* had the highest absolute transcript level in *S.* Typhimurium's global transcriptome data (Kröger et al., 2013).

Comparative RT-qPCR was used and two reference genes were chosen using the Kroger et al., (Kröger et al., 2013) data to ensure there expression levels were constant between conditions of NO shock and a control condition. The two reference genes chosen were kdgR and STM1575.

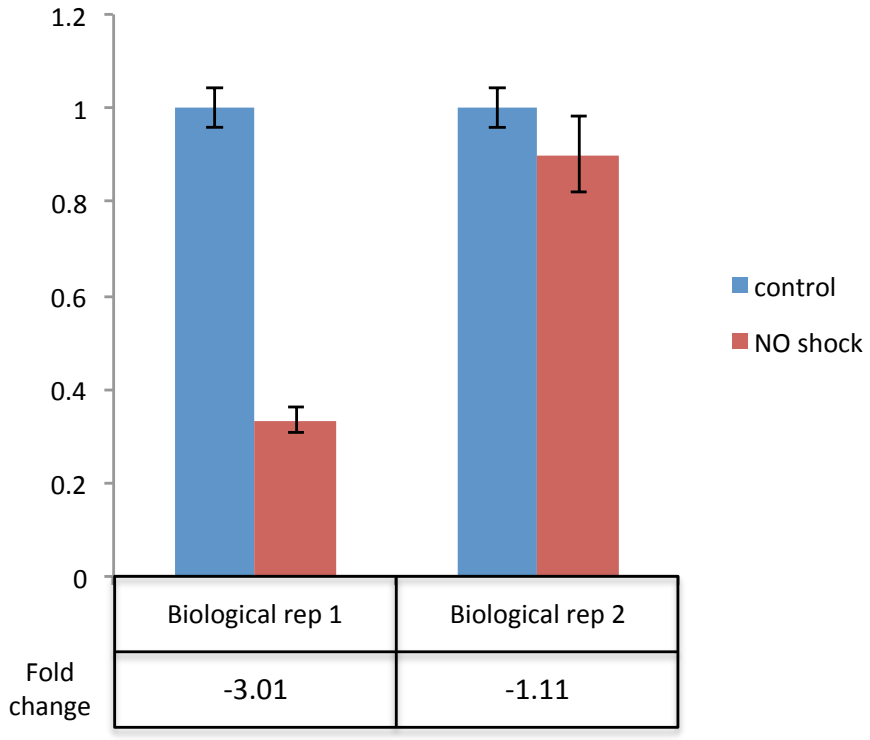


Figure 18: Expression of *gtrA* under NO shock measured by RT-qPCR

Relative expression levels, measured by RT-qPCR, of *gtrA* under NO shock and the control condition of no NO shock in two biological replicates. The red bars relate to the NO shock condition and the blue bars relate to the control condition. The y-axis shows the relative quantification of *gtrA,* which is relative to the reference gene levels. The x-axis indicates which biological replicate the samples are from. The values in the boxes below the x-axis give the fold change in relative *gtrA* levels between the NO shock and the control samples in the biological replicate. A + in front of the fold change value indicates the levels of *gtrA* under NO shock have increased compared to the levels at the control condition and – means they have decreased.

In figure 18, biological replicate 1 and 2 do not follow the same pattern of *gtrA* expression. In biological replicate 1 *gtrA* expression under NO shock is 3 fold less than *gtrA* expression in the control condition, whereas there is almost no change in *gtrA* expression levels in biological replicate 2. The pattern of expression of gtrA (figure 18) and expression from the gtrA upstream region (figure 17) are different. This could be due to the differences in the techniques used. The region cloned in front of the *lacZ* gene could be missing important regulatory regions. Also, the beta-galactosidase assay will not have accounted for any post-transcriptional regulation as the assay only measures the level of transcription from the *gtrA* upstream region and not the *gtrA* transcript unlike RT-qPCR.

4.3 IS FAMILY IV *GTRC* BEING EXPRESSED INDEPENDENTLY OF THE OPERON PROMOTER?

The global transcriptome study of *S.* Typhimurium (Kröger et al., 2013) also produced data showing the transcriptional start sites (TSS) of many *S.* Typhimurium genes. This was done using terminator 5'-phosphate dependent endonuclease, which was used to enrich a population of transcripts for primary transcripts. These transcripts were sequenced and this was used to identify the TSS. Two possible transcriptional start sites were identified within *gtrB*(IV) when the strain was grown in InSPI2 media (personal correspondence with Disa Hammerlof – exact positions of the TSS's were not available). This indicates that *gtrC* may be expressed independently from the main family IV operon promoter.

To confirm that there was expression from a promoter within *gtrB* a region of the gene thought to contain the promoter was cloned into the CRIM vector in front of the *lacZ* gene. A 581bp region within *gtrB* and overlapping the *gtrC* start site was amplified using high fidelity colony PCR. This region was then cloned into the CRIM vector in front of the *lacZ* gene (see section 2.5. and 2.6. for methods). The CRIM vector was then integrated into the *S.* Typhimurium (strain 79) genome at the attB phage insertion site making strain 944 (see section 2.8. for methods).

A beta-galactosidase assay (section 2.12.) was performed using strain 944 to test the expression from the site within *gtrB* and a control strain, strain 81, that was the same as strain 944 but had no region cloned in front of the *lacZ* gene. These two strains were cultured, in biological duplicates, in InSPI2 media to an OD600 of 0.3. Four technical replicate samples were taken from each culture to perform a beta-galactosidase assay.

The graph in figure 19 shows that the amount of beta-galactosidase (measured in miller units) produced by strain 944 was nearly 7 folds more compared to the control strain. A 7-fold increase in expression indicates that the region inserted in front of the *lacZ* gene is promoting transcription. Two possible promoter region recognized by sigma 70 have been predicted within the 581bp region and are shown in figure 20.

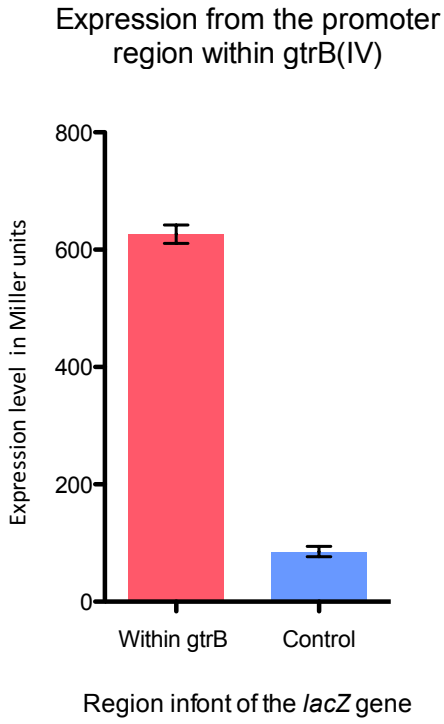


Figure 19: Expression from TSS within *gtrB*(IV).

The level of beta-galactosidase measured in miller units produced from strain 944 and strain 81. Strain 944 has the *lacZ* gene down stream of a 581pb region, taken from within *gtrB*(IV), integrated into its genome. Strain 81 is the same as strain 944 but has no region cloned in front of the *lacZ* gene. On the x-axis 'within *gtrB*' refers to strain 944 and 'control' refers to strain 81.

AAGCGCTGAGTGGTTCTACAGGCTGCATAACAAAATCAGTACGCCAAAGATTGAAGAGAATGTCGGTGATTTT
CGATTGATGTCGCGCGAGATTGTAGAAAATATCAAGCTATTACCAGAACGTAACCTTTTCATGAAAGGTATACT
TTCATGGGTTGGAGGTCAAACAGATGTGGTCGAATATGCCCGTGCTGAACGTGTCGCAGGTAACTCAAAATTTA
ATGGCTGGAAACTCTGGAACCTGGCGCTGGAGGGGATTACAAGTTTTTCTACTTTCCCTTTGCGTATCTGGATG
TATATAGGAGTGAGCGTTTCTGCCCTCTCCCTGATATATGCCATGTGGATGATCATTGATAAATTGATGTGGGG
AAACCCTGTTCCTGGTTATCCTTCGCTTATGACCGCGATTCTCTTCTTAGGCGGCATCCAGCTTATCGGCATAGG
CATCATGGGTGAATATATCGGACGCGTTTACACGGAGGTGAAGCAAAGACCCCGCTATATCGTGAAAAACAAA
AAAACAATGATGGAATAATGAACACTATGCTCAAGATATTACCGAAAACGGCGATGATACTACT

Figure 20: The 581bp region cloned in front of the *lacZ* gene.

An 581bp region, taken from within gtrB and overlapping gtrC start codon, that was cloned in front of the *lacZ* gene to make strain 944. The underlined bases show that primer regions used to amplify the fragment. The red lettering indicates the gtrB stop codon and the blue indicates the gtrC start codon. The highlighted bases show two possible sigma70 -10 and -35 boxes, one set in yellow and the other in green. The sigma70 promoter regions were predicted using BPROM - Prediction of bacterial promoters (V. Solovyev, 2011).

4.4 CHAPTER CONCLUSION

None of the results in this chapter have shown that there is an up regulation of the family IV *gtr* genes under pH3 or NO shock. After pH3 shock expression levels of *gtrC* and *gtrA* drop. After 15 minutes the expression of *gtrC* returned to the original level showing that the effect of the pH3 shock on *gtrC* expression is not sustained. Under pH3 and NO shock there was an indication that the promoter upstream of gtrA may have variable expression under certain conditions. As well as the main operon promoter, sites within *gtrB* were shown to promote transcription, which suggests that *gtrC* could be transcribed independently. This is supported by the results in section 4.1.2. which showed that *gtrA* and *gtrC* had different patters of expression.

The results of this chapter that looked at the family IV genes up regulation under pH3 and NO shock have not been able to replicate the findings of the study carried out by Kroger et al., (Kröger et al., 2013). This could have been because of difference between the methods of RNAseq and RT-qPCR. The two methods quantify expression by measuring different things, RT-qPCR only measures part of the transcript whereas RNA-seq measures the full transcript. Also, they both use different units of measurement to describe levels of expression, which are obtained by different normalization methods that could have different biases attached. Relative RT-qPCR uses the reference genes to normalize the data and gives the units in relative quantification (RQ). RNAseq analysis generates results in units of transcripts per million (TPM). Measurements in TPM are obtained by normalizing the data against sequencing depth and average read length of the sequencing run and gene of interest length and (Wagner et al., 2012). Therefore, relative RT-qPCR may have biases based on the expression of the reference gene whereas RNAseq will have biases based on the transcript pool of the sequencing run. This means that small fold changes in expression might not be reproducible between the two techniques.

There was a difference in the expression pattern of *gtrA,* under pH3 shock, between the two biological replicates in figure 16B. *gtrA* had consistently low expression at pH3 compared to pH7 in biological replicate 1, whereas in biological replicate 2 the fold change in expression between pH3 and pH7 was small. The difference between the biological replicates could be

due to the promoter having variable levels of expression under either pH3 conditions or pH7 conditions.

Similarly, *gtrA* expression under NO shock (figure 18) was not consistent between the two biological replicates. Biological replicate 1 showed a 3-fold decrease in *gtrA* expression under NO shock whereas there was no change in expression in biological replicate 2. This could be due to the time taken for the RNA to be fixed in each sample. As seen in figure 16A the levels of RNA can change over a short period of time. Or these results could also be showing that there is a variable level of expression from the *gtrA* upstream promoter.

As the *gtrABC* genes are together in an operon we would expect them to be expressed from the same promoter and therefor have the same expression pattern. However, the results in biological replicate 1 of figure 16A and 2B did not follow this assumption. The expression of *gtrA* has a higher decrease in expression under pH3 shock that *gtrC* compared to the control condition. The differences in these two genes patterns of expression could be due to the possible promoter region within *gtrB*. If this promoter responds differently to either pH3 or pH7 conditions then the expression pattern of *gtrC* will be different to *gtrA'*s expression.

Section 4.3 looked at whether there was a promoter region within *gtrB* that could be promoting expression of *gtrC* separately from the main operon promoter. Figure 19 showed the amount of beta-galactosidase produced from a strain with the proposed regulatory region in front of the *lacZ* gene. The presence of the proposed regulatory region in front of the *lacZ* gene increased beta-galactosidase production by up to 7 fold compared to the control strain, which had no regulatory region in front of the *lacZ* gene. *gtrC* expression from a promoter independent of the main operon promoter helps to explain the results in figure 16A and 2B where there is a large difference in *gtrA* and *gtrC'*s expression patterns in biological replicate 1. The promoter within *gtrC* may respond differently to the main promoter under certain growth conditions. This would mean the fold changes in *gtrA* and *gtrC* expression between two growth conditions could be different.

5. ARE MODIFICATIONS BY OTHER FAMILIES OF GTR PROTEINS NEEDED FOR GTR(IV) ACTIVITY?

The constitutive expression of the family IV *gtr* genes from a Ptac promoter in a basal *S.* Typhimurium strain does not cause a detectable modification of the LPS (Davies et al., 2013). One of the reasons for this could be that further modifications of the basal O-antigen are needed to provide GtrC with a target for addition of a sugar. In a wild type strain of *S.* Typhimurium, and other Salmonella strains containing the *gtr*(IV) operon, there are other families of *gtr* operons present (Davies et al., 2013). The modifications made by these proteins could be needed to provide GtrC(IV) with a target. This chapter looks at the effect of family III O-antigen modification on *gtrC*(IV) activity in *S.* Typhimurium and at the effect of either family V or family VI O-antigen modifications on *gtrC*(IV) activity in *S.* Infantis.

To study whether modifications by other gtr operons are needed for GtrC(IV) activity the following methods were used: LPS was extracted from strains and ran on a TSDS-PAGE gel to show the laddered pattern of varying lengths of O-antigen repeating units (see section 2.9. for materials and methods). In a strain expressing no other O-antigen modifying genes the *gtr*(III), *gtr*(V) or *gtr*(VI) operon was expressed from the Ptac promoter in the chromosomal DNA. The family III genes were expressed in *S.* Typhimurium and the family V and VI genes in *S.* Infantis, which is where the genes are naturally found. The *gtr*(IV) operon was expressed from the lac promoter carried on the Plac22 vector and had been transformed into the three strains expressing a *gtr* operon from the Ptac promoter. These three transformed strains combined the expression of the family IV operon with another *gtr* family.

The Plac22 vector expression system(Warren et al., 2000) is inducible as it carries the *lacIq* repressor gene, which binds the lac promoter and stops expression. IPTG is added to a culture to induce expression and works by binding the LacIq repressor protein, changing its conformation and preventing it from binding the lac promoter.

A strain expressing only *gtr*(III), *gtr*(V) or *gtr*(VI) from the Ptac promoter will be referred to as Gtr III, Gtr V or Gtr VI respectively or when referring to them as a collective, Ptac-Gtr will be used. The addition of +IV will be used to indicate the strain is also expressing *gtr*(IV) from the Plac22 plasmid or +Plac22 means the strain has been transformed with an empty Plac22 vector which contains no *gtr*(IV) operon.

As well as a comparison between Ptac-Gtr strains and Ptac-Gtr+IV strains a basal strain, one expressing no known O-antigen modifying genes, and a Ptac-Gtr+Plac22 strain were used as controls. Comparison of Ptac-Gtr with the basal strain shows sufficient modification of the O-antigen by the *gtr* operon expressed from the Ptac promotor. The Ptac-Gtr+Plac22 strain controls for any effect the presence of the vector could have on the strains O-antigen that was not due to its expression of the *gtr*(IV) operon.

5.1 DO MODIFICATIONS OF THE O-ANTIGEN BY THE FAMILY III, V OR VI PROTEINS PROVIDE A TARGET FOR FAMILY IV GTRC ACTIVITY?

5.1.1 Can Gtr family III modifications of the *S.* Typhimurium O-antigen provide family IV GtrC with a target for activity?

The *S.* Typhimurium LPS banding patterns (figure 21) showed a shift between the bands of the basal strain (strain 101), containing no know O-antigen modification genes, and the Gtr III strain (strain 122). This showed that the expression of *gtr*(III) from the Ptac promoter in strain 122 was modifying the LPS in *S.* Typhimurium. Between the LPS banding pattern of the Gtr III strain and the Gtr III+IV strain (strain 921) there was no detectable shift seen. The un-induced Gtr III+IV, where no IPTG was added to the culture, had the same LPS banding pattern as the induced Gtr III+IV strain but also had additional shadow bands that were level with bands of the basal strain. (Figure 21)

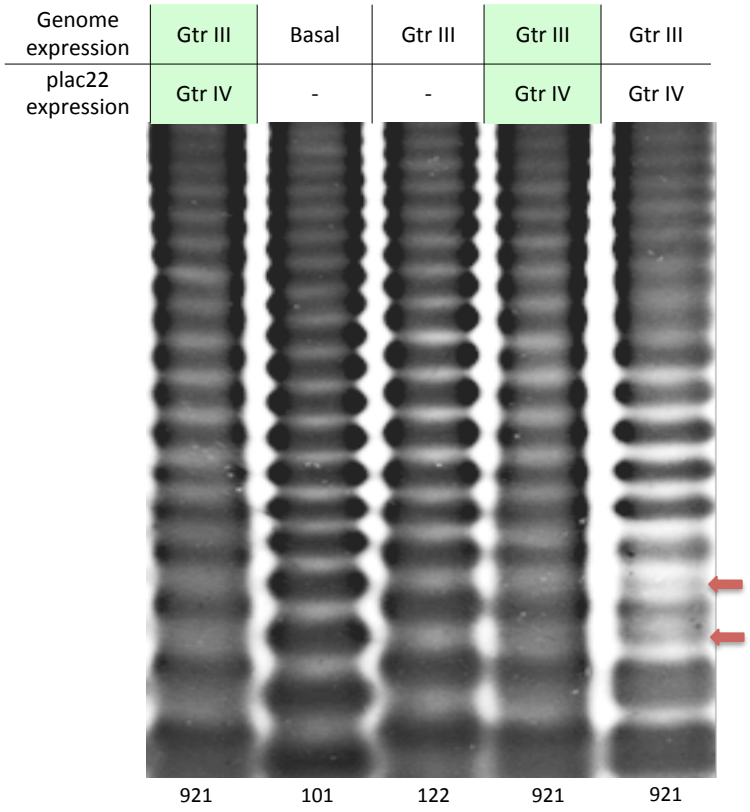


Figure 21: LPS analysis of a strain expressing the *gtr* families III and IV.

Extracted LPS ran on a TSDS-PAGE gel. The table above the lanes indicates which *gtr* genes are being expressed in that strain. The top row refers to the genes that are being expressed from the Ptac promoter in the genome, basal means that no *gtr* genes are expressed from that strain. The bottom row refers to the *gtr* genes that are being expressed from the lac promoter in the Plac22 plasmid, empty vector means that there are no genes expressed from the lac promoter but the Plac22 vector is still present. A dash indicates that there was no vector in that strain. Cells that are shaded green signify that that strain was induced with IPTG. The numbers below the lanes give the strain numbers, strain details can be found in table A2.3. The red arrows are pointing to shadow bands.

5.1.2 Can Gtr family V modifications of the *S.* Infantis O-antigen provide family IV GtrC with a target for activity?

There was a shift between the LPS banding patterns of the basal strain (strain 288) and the strain only expressing the *gtr*(V) genes (strain 328) (figure 22). Therefore, family V Gtr proteins were modifying the O-antigen repeating unit in *S.* Infantis.

When the *gtr*(IV) genes are expressed in a strain with gtr(V) under the Ptac promotor (strain 928) no detectable shift was seen. Additionally, the banding pattern of strain Gtr V+Plac22 (strain 927) had the same banding pattern as the other strains containing *gtr*(V) under the Ptac promoter. But as well as the main banding pattern there were two additional shadow bands, indicated with red arrows, that appeared to be running level with the basal bands. (Figure 22)

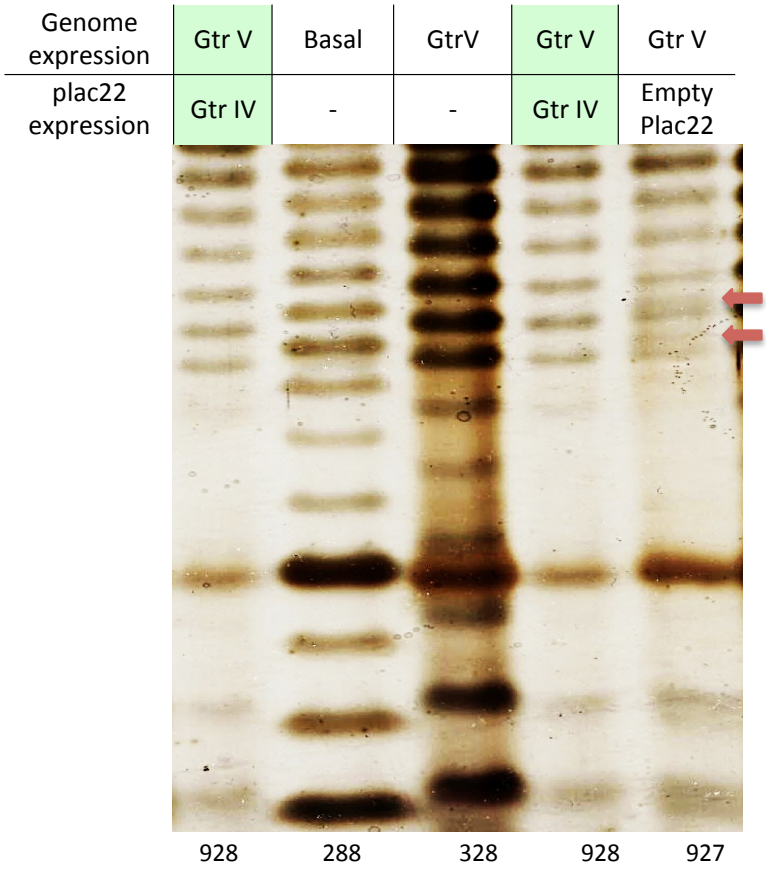


Figure 22: LPS analysis of a strain expressing the *gtr* families Vand IV.

Extracted LPS ran on a TSDS-PAGE gel. The table above the lanes indicates which *gtr* genes are being expressed. The top row refers to genes that are being expressed from the Ptac promoter in the genome, basal means that no *gtr* genes are expressed from that strain. The bottom row refers to the *gtr* genes that are being expressed from the lac promoter in the Plac22 plasmid, empty vector means there are no genes expressed from the lac promoter but the Plac22 vector is still present. A dash indicates that there was no vector in that strain. Cells that are shaded green signify that the strain was induced with IPTG. The numbers below the lanes give the strain numbers, strain details can be found in table 16. The red arrows are pointing to shadow bands.

5.1.3 Can Gtr family VI modifications of the *S.* Infantis O-antigen provide family IV GtrC with a target for activity?

There was a shift between the LPS banding patterns of the basal strain (strain 288) and the strain only expressing the *gtr*(VI) genes (strain321) (figure 23). Therefore, family VI Gtr proteins were modifying the O-antigen repeating unit in *S.* Infantis.

In figure 23 There was a clear shift between the banding pattern of strain Gtr VI (strain 321) and strain Gtr VI+IV (strain 930) showing that the additional presences of the Plac22 plasmid expressing *gtr*(VI) was changing the O-antigen of a strain 930 expressing *gtr*(VI) from the Ptac promoter. This banding pattern appeared to be running level with the basal strain bands. However, the same effect was seen when an un-induced empty Plac22 plasmid was present. The IPTG induced strain Gtr VI+IV had shadow bands, indicated by the red arrow, which was running at the same level as strain 321. (Figure 23)

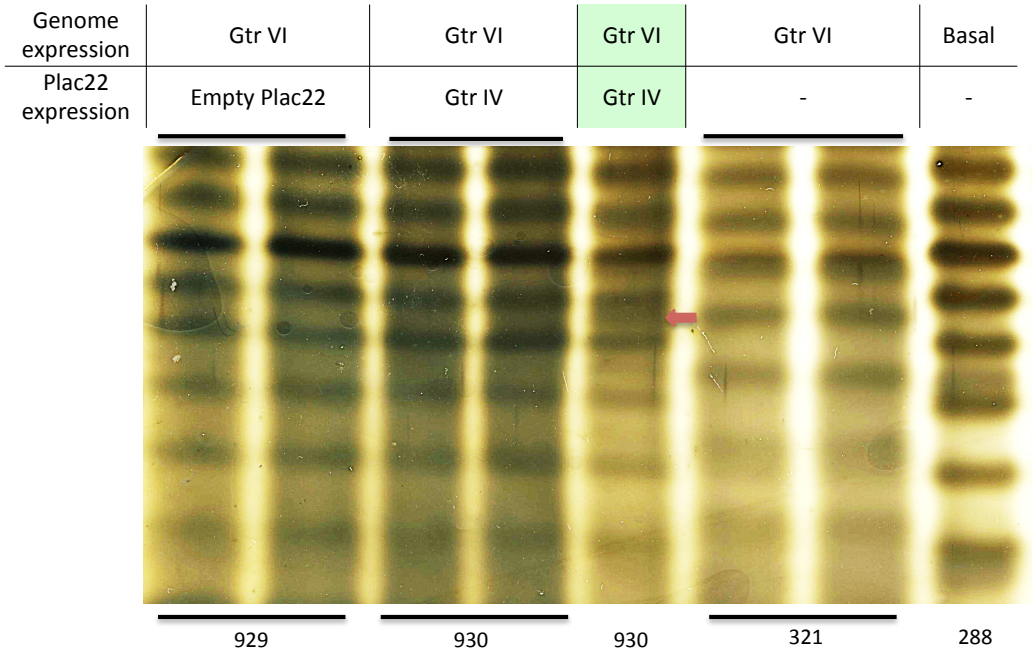


Figure 23: LPS analysis of a strain expressing the *gtr* families VI and IV.

Extracted LPS ran on a TSDS-PAGE gel. The table above the lanes indicates which *gtr* genes are being expressed in that strain. The top row refers to the genes that are being expressed from the Ptac promoter in the genome, basal means that no *gtr* genes are expressed from that strain. The bottom row refers to the *gtr* genes that are being expressed from the lac promoter in the Plac22 plasmid, empty vector means that there are no genes expressed from the lac promoter but the Plac22 vector is still present. A dash indicates that there was no vector in that strain. Cells that are shaded green signify that that strain was induced with IPTG. The numbers below the lanes give the strain numbers, strain details can be found in table 16. The red arrows are pointing to shadow bands.

5.1.4 Summary of results and interim conclusion

Taken together the data presented above shows that the presence of an empty Plac22 plasmid could cause a modification of the O-antigen of a strain expressing a *gtr* operon from a Ptac promoter in the genome. In a Gtr VI+Plac22 strain this could be seen though the whole of the LPS banding pattern which was running level with basal. In the un-induced Gtr V+Plac22 and un-induced Gtr III+IV strains the modifications caused by a Plac22 vector could be seen through the shadow bands that are also running at the basal level. The shadow bands show a proportion of extracted O-antigen that has differential modification to the rest of the extracted O-antigen. This was likely to be due to the LacIq repressor protein produced from the Plac22 plasmid.

The LacIq repressor protein represses the expression from the lac promoter on the Plac22 vector providing an inducible form of expression. However the chromosomal Ptac promoter, which *gtr*(III), (V) and (VI) were being expressed from, can also be bound and repressed by the LacIq repressor protein as the Ptac promoter is a combination of the trp and, importantly, the lac promoter (Amann et al., 1983) (figure 24). This relationship had not been taken into account when designing this experiment and strains containing an empty Plac22 vector were not induced with IPTG, which binds and inactivates the LacIq protein.

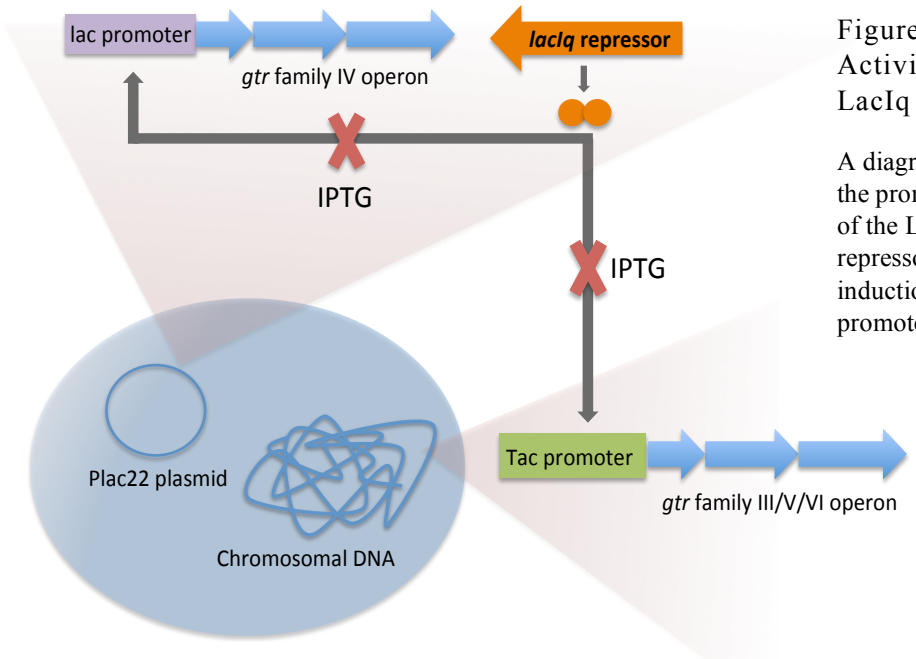


Figure 24: Activity of the LacIq repressor.

A diagram showing the promoter targets of the LacIq repressor and induction of the promoters by IPTG.

Therefore, the phenotype of a strain with a *gtr* operon under the Ptac promoter containing an un-induced vector is showing a basal phenotype which could clearly be seen in strain Gtr VI+Plac22 and un-induced Gtr VI+IV (figure 23). Induced Gtr VI+IV has shadow bands running level with strain Gtr VI which could be O-antigen that has been modified by the Gtr(VI) proteins. However, the main banding pattern is running at the basal level suggested that induction may not be sufficient to stop all repression from the Ptac promotor (figure 23). In strains with *gtr*(III) (figure 21) and *gtr*(V) (figure 22) under the Ptac promoter the presence of an un-induced vector and repression of the Ptac promoter seemed to have less of an effect on the phenotype; only shadow bands can be seen returning back to the basal level in both strains. Hence, LacIq repression was not preventing all expression from the Ptac promoter.

5.2 CAN INDUCTION STOP A PLAC22 VECTOR CAUSING MODIFICATION TO A PTAC-*GTR* STRAIN'S O-ANTIGEN?

5.2.1 Modifications of the O-antigen observed in the presence of Plac22 was due to its effect on the Ptac promoter.

To demonstrate that the effect of the Plac22 plasmid on the LPS banding pattern was due to repression of the Ptac promoter the banding patterns of a basal *S.* Typhimurium strain with and without the Plac22 plasmid was compared. An experiment that swapped the empty Plac22 vector for a pBAD vector in the Gtr VI strain was also done to control for the effect of ampicillin on the LPS banding pattern of a *S.* Infantis strain with the family VI modification, as this antibiotic targets cell wall synthesis.

In figure 25A there was no shift seen between the *S.* Typhimurium basal strain and the basal strains containing induced or un-induced empty Plac22 vector, this showed that without chromosomal *gtr* expression from the Ptac promoter Plac22 had no effect on the banding pattern. No shift was seen between the Gtr VI strain with and without pBAD (figure 25B), which showed that ampicillin is not having an effect on the LPS banding pattern of this strain.

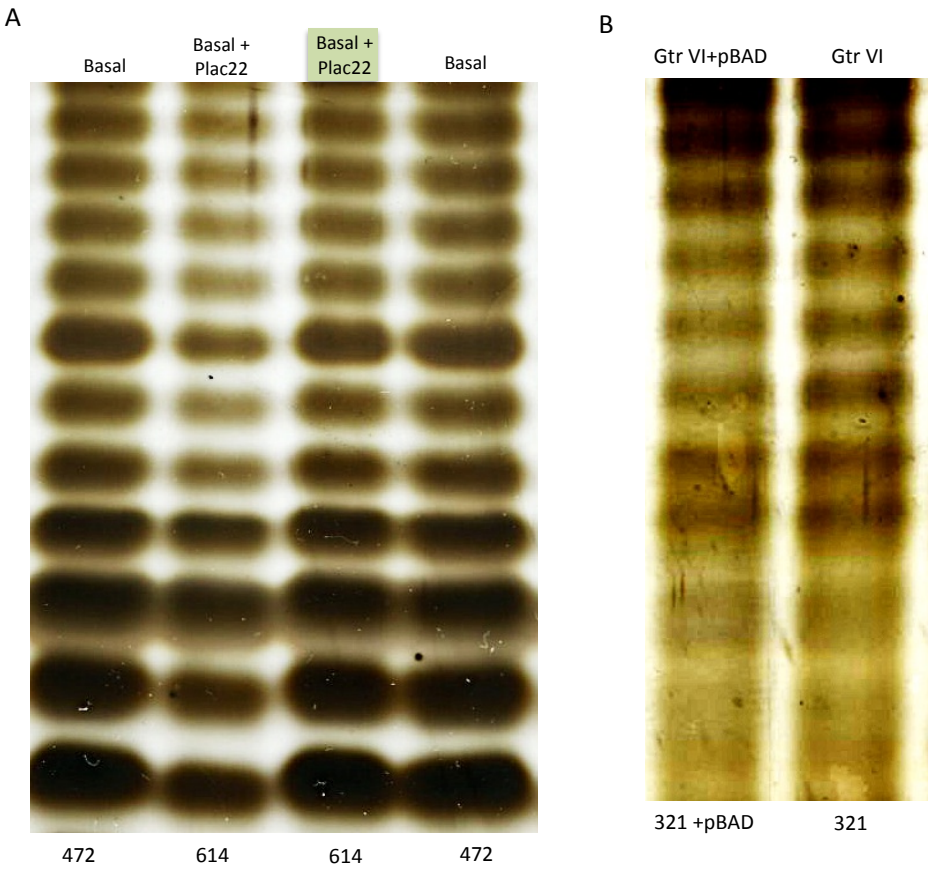


Figure 25: Plac22's effect on the O-antigen is likely due to repression of the ptac promoter.

Extracted LPS ran on a TSDS-PAGE gel. The labels below the lanes indicate which strain the LPS was extracted from, details of these strains can be found in table 14. A: All strains have a basal genotype with no known O-antigen modifying genes present. The labels above indicate whether the Plac22 vector is present. Green boxes behind the text highlight strains that have been induced with IPTG. B: Both strains have *gtr*(VI) under the Ptac promoter. The labels indicate whether the pBAD vector is present.

5.2.2 Is induction of a strain containing a Plac22 vector sufficient to stop repression from the Ptac promoter?

The previous experiments were repeated and every strain containing a Plac22 plasmid was induced with IPTG. The induced strain containing an empty Plac22 vector was compared to an un-induced form so that the effect of induction of the Ptac promoter could be seen.

In figure 26 There was a shift between the *S.* Infantis strains Gtr VI and un-induced Gtr VI+Plac22. In the induced form there was still a clear shift against the Gtr VI strain but there were shadows of bands running level with the Gtr VI strain. Similarly, the Gtr VI+IV strain was shifted compared to the Gtr VI strain but had shadow bands that were running at the same level. The main banding patterns of strains: Gtr VI+IV, Gtr VI+Plac22 induced and Gtr VI+Plac22 un-induced were running level with the basal strain (Figure 26). Therefore, the presence of a Plac22 vector in an induced or un-induced strain expressing *gtr*(VI) from the Ptac promoter caused the banding pattern to shift to the basal level. But, in the induced strains a small amount of the O-antigen did not shift against the Gtr IV strain creating shadow bands. The same experiment was done using strains with *gtr*(V) under the Ptac promoter. Unfortunately, there was an insufficient amount of LPS loaded to see the banding pattern (Figure not shown).

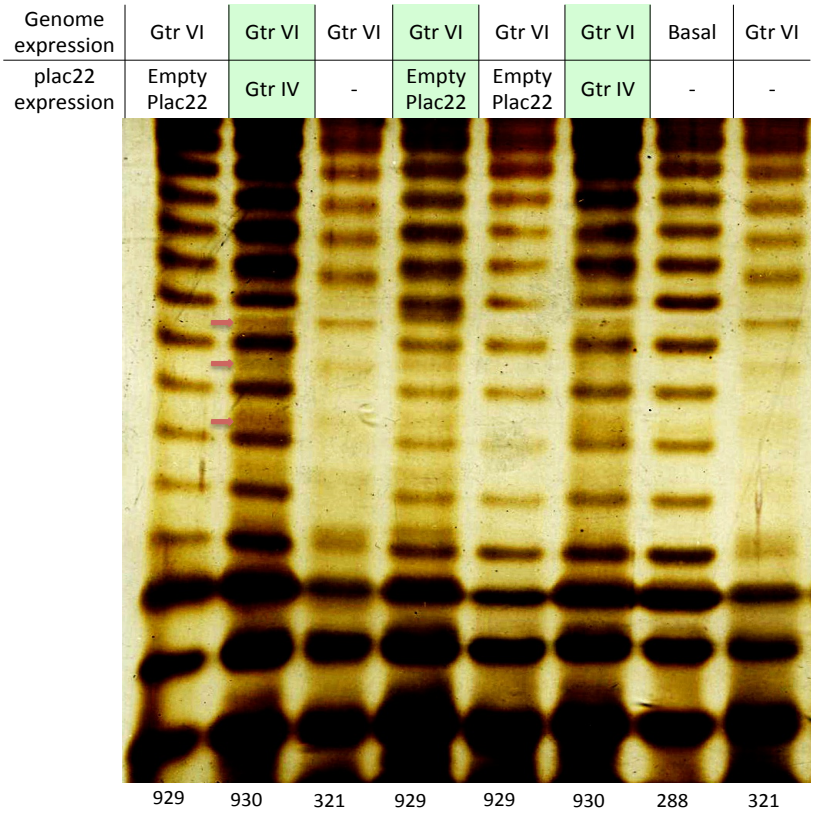


Figure 26: LPS analysis of strain expressing the *gtr* families VI and IV (all strains were induced).

Extracted LPS ran on a TSDS-PAGE gel. The table above the lanes indicates which *gtr* genes are being expressed in that strain. The top row refers to the genes that are being expressed from the Ptac promoter in the genome, basal means that no *gtr* genes are expressed from that strain. The bottom row refers to the *gtr* genes that are being expressed from the lac promoter in the Plac22 plasmid, empty Plac22 means that there are no genes expressed from the lac promoter but the Plac22 vector is still present. A dash indicates that there was no vector in that strain. Cells that are shaded green signify that that strain was induced with IPTG. The numbers below the lanes give the strain numbers, strain details can be found in table 16. The red arrows are pointing to shadow bands.

5.3 CHAPTER CONCLUSION

The addition of the Plac22 plasmid to a basal *S.* Typhimurium strain does not alter the LPS banding pattern (figure 25A). Thus the shifts due to its presence, seen previously (figure 23), must be because of its effect on the expression of the *gtr* operons from the Ptac promoter. Although, a further control using basal *S.* Infantis as well as *S.* Typhimurium would have been useful.

Induction with IPTG does not appear to be sufficient to stop all repression of the Ptac promoter. The induced Gtr VI strain containing empty Plac22 is shifted compared to the strain without Plac22 and is running level with the basal strain suggesting there is no modification of the O-antigen by the Gtr(VI) proteins (figure 26). Full modification of the O-antigen by the Ptac-Gtr proteins is necessary to draw reliable conclusions from the result, as it is an essential part of this experiment. To complete these experiments a vector that has no effect on the Ptac promoter should be used.

However, the un-induced Gtr III+IV strain in figure 21 and un-induced Gtr V+Plac22 strain in figure 22both show a complete banding pattern that is level with the corresponding Ptac-Gtr strain with only two shadow bands running level with the basal strain. This shows that even without induction there appears to be enough expression from the Ptac promoter for the Gtr(V) and Gtr(III) proteins to modify the majority of O-antigen that has been extracted. Both Gtr III+IV (figure 21) and Gtr V+IV (figure 22)were induced with IPTG so should be expressing enough Ptac-Gtr proteins to sufficiently modify the O-antigen. Therefore, although experiments using an alternative vector would be more reliable, conclusions can be drawn from the results in figure 21 and 2. In *S.* Typhimurium and *S.* Infantis Gtr(III) and Gtr(V) modifications respectively do not provide a target for detectable modification by the Gtr(IV) proteins.

Furthermore, if we make the assumption that the shadow bands of the induced Gtr VI+IV strain (figure 26) are showing a banding pattern that is being modified by the Gtr(VI) proteins then these results would show that the addition of the family IV proteins are not giving a detectable modification of the O-antigen.

It is interesting that the presence of the LacIq repressor had a stronger effect in the strain expressing *gtr*(VI) from the Ptac promoter. We would expect LacIq repression of the Ptac promoter to be the same in every strain. Therefor, the difference may have been due to differences between the ways each family of Gtr proteins works. For example, the Gtr(VI) proteins may need to be present in higher quantities to efficiently modify the O-antigen than Gtr(III) and (V), meaning that a drop in expression from the Ptac promoter can only be seen in the Gtr VI strain's phenotype.

6. IS THE FUNCTION OF THE FAMILY IV GTRAB PROTEINS DISTINCT FROM OTHER FAMILIES GTRAB PROTEINS?

A phylogenetic analysis, of the Gtr proteins in *S. enteriaca*, *S. bongori* and 4 phage genomes, done by Davies et al., 2013 used the GtrC protein to split the Gtr proteins in to ten families based on the clustering of the GtrC proteins. The analysis showed that family IV GtrA clustered into the same group and family IV GtrB clustered into the same group. Theses groups don't contain GtrB or GtrA proteins from another family. Unlike Family IV the other families (excluding family II which is an acetyltransferase (Kintz et al., 2015)) GtrA and GtrB proteins were split across different clusters showing that there has been recombination between operons. In the family IV operons this recombination hasn't happened. A reason for this may be that family IV GtrAB cannot replace the function of other GtrAB proteins. As well as this difference, alignments of GtrA and GtrB (figure X-don't have a figure number yet) highlighted that the family IV GtrA and GtrB proteins had the most AA changes from the consensus sequence.

These observations pointed to the hypothesis that family IV AB proteins were functionally different from other families' GtrAB proteins. This hypothesis was tested by replacing GtrAB(I) with GtrAB(IV) to see if the family IV GtrAB proteins could work with GtrC(I) and replace the function of GtrAB(I). If the family IV proteins had a function that is distinct from the other Gtr proteins then the replacement would lead to a non-functional *gtr* operon and a basal LPS phenotype. To do this a strain was used that had gtrAB(IV) expressed from the Ptac promoter in the genome and gtrC(I) expressed from the lac promoter on the Plac22 plasmid.

6.1 GENERATION OF STRAIN EXPRESSING THE FAMILY I AND FAMILY IV GENES

A strain was generated that expressed gtrAB(IV) in place of gtrAB(I) (Table 4). To do this strain 141 was used, which expressed gtrAB(IV) (gtrC(IV) had been removed from the operon) from the Ptac promoter but had no other known O-antigen modifying genes. Strain 141 was then transformed with plasmid 461, which is a plac22 vector that has gtrC(I) under the lac promoter. This generated strain 970, which expressed gtrAB(IV) from the Ptac promoter and gtrC(I) from the lac promoter on the Plac22 plasmid, but no other known O-antigen modifying genes were expressed. Stain 970 will be referred to as GtrAB(IV)C(I).

Four control strains were also generated (Table 4). Two of these strains were 967 and 968. Strain 967 expressed only gtrABC(I) and was made by transforming a basal strain (strain 101), with no known O-antigen modifying genes, with plasmid 460. Plasmid 460 was a plac22 vector with gtrABC(I) under the lac promoter. Strain 968 expressed only gtrC(I) and was made by transforming a basal strain with plasmid 461, which was a plac22 vector with gtrC(I) under the lac promoter. The two other control stains were 969 and 966. Strain 969's parent strain was strain 141, which has gtrAB(IV) under the Ptac promoter. Strain 966's parent strain was strain 101, which is a basal strain that expresses no known O-antigen modifying genes. Both strains were made by transformation with an empty plac22 vector, which had no genes under the lac promoter. These two strains were generated to control for the effect of the plac22 vector on the O-antigen.

Table 4: Table of strains used in chapter 6 experiments

Strain	Genes expressed from chromosomal Ptac promoter	Genes expression from Plac22 Vector
970/GtrAB(IV)C(I)	*gtrAB*(IV)	*gtrC*(I)
967/GtrABC(I)	-	*gtrABC*(I)
968/GtrC(I)	-	*gtrC*(I)
966	-	Empty vector
969	*gtrAB*(IV)	Empty vector

*All strains contain no other *gtr* genes except for those specified

The plac22 vector is an inducible expression system because the vector holds the lacIq gene; whose product will repress the lac promoter. Induction happens when IPTG is added to the culture and will bind the LacIq protein and stop it from binding the lac promoter. The LacIq repressor can also bind and repress the Ptac promoter, which was used in combination with the plac22 vector in this experiment. In previous experiments in chapter 5, induction with IPTG was not always sufficient to stop all repression by the LacIq protein. Therefore, before each strain was transformed with Plac22 the vector was modified by cutting out the *lacIq* gene to stop repression of both the lac promoter on the vector and the Ptac promoter in the genome.

Modification of the Plac22 vector was done by high fidelity PCR amplification (for methods see section 2.3.) of the whole Plac22 vector excluding the *lacIq* gene using primers oMV 1254 and oMV 1255 (table 14). The linear product was then digested (see section 2.5.) and ligated (see section 2.6.) together to create a *lacIq* free Plac22 vector that was used in all parts of this experiment (Figure 27). The vectors that were created were plasmids 461, 460 and 459.

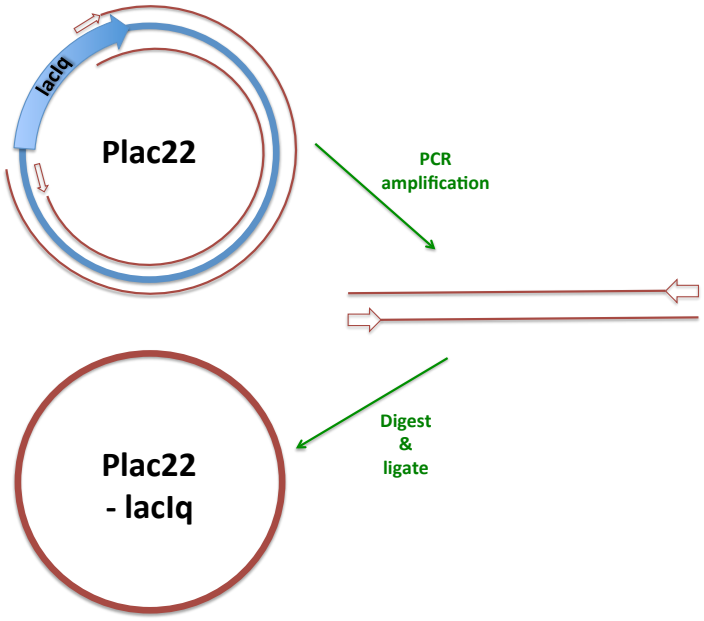


Figure 27: Removal of the *lacIq* gene from Plac22.

A diagram illustrating the method of removal of the *lacIq* gene from the Plac22 plasmid. The blue plasmid is the complete Plac22 plasmid with *lacIq* and the red plasmid is the *lacIq* free plasmid after modifications. The red arrows are showing the forward and reverse primers, which are oMV 1254 and oMV 1255 (table 14).

To show that the *lacIq* gene was successfully deleted from plasmids 461, 460 and 459 they were digested with restriction enzymes Bg1II and pstI generating two linear fragments of DNA. One of the fragments contained the site that the lacIq gene was deleted from. The plasmid digest was run on an agarose gel to separate out the two fragments and calculate their

lengths. All three plasmids would have had a DNA fragment length of 1141bp if the lacIq gene had been deleted or 2033pb if the lacIq gene was still present. The second larger fragment lengths of plasmids 459, 460 and 461 were approximately 3619bp, 6346bp and 4843bp respectively.

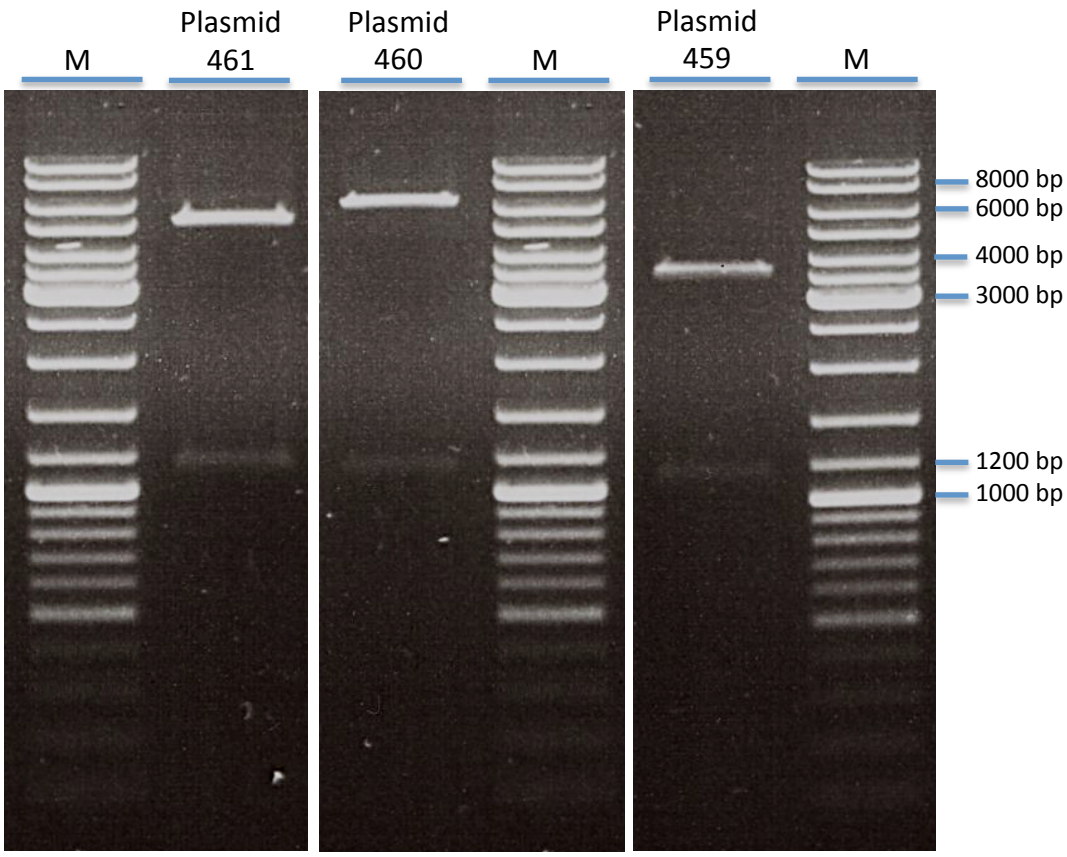


Figure 28: The *lacIq* gene was successfully deleted from plasmids 461, 460, and 459.

Agarose gel electrophoresis of plasmids: 461, 460, and 459 digested with restriction enzymes Bg1II and PstI. The length of the smaller fragment for all plasmids was 1141bp. This was the region that contained the lacIq gene. With the lacIq gene the fragment length would be 2033bp long. The second larger fragment lengths of plasmids 459, 460 and 461 were approximately 3619bp, 6346bp and 4843bp respectively. Above the lanes are labels giving the plasmid numbers and indicating the marker lanes (M). The marker used was GeneRuler DNA Ladder mix from Thermo

The gel in figure 28 shows that all three of the digested plasmids have one fragment length that is slightly below the 1200bp level of the marker indicating this is the 1141bp fragment and therefore the lacIq gene has been removed. The larger fragments for each plasmid are around the lengths 3619bp, 6346bp and 4843bp for plasmids 459, 460 and 461 respectively.

To demonstrate that the LacIq protein was no longer present in the strain and could not repress the promoters a comparison of the LPS banding patterns of induced and un-induced strain 967, expressing *gtrABC*(I) from the lac promoter was done. If the LacIq protein were present, then it would be repressing the lac promoter in a strain that had not been induced with IPTG, preventing *gtrABC*(I) expression. This would mean that the O-antigen would not be modified and the strains O-antigen banding pattern would be running at a different level to a strain that had been induced and was expressing the *gtrABC*(I) genes. However, if the LacIq protein is not present then the O-antigen banding pattern of an induced and un-induced strain 967 would be the same. A basal strain 966 expressing no *gtr* genes and containing an empty plac22 vector was used as a control to show that the family I Gtr proteins were modifying the O-antigen.

When comparing the induced and un-induced strains' banding patterns in figure 29 there was no shift between the induced and un-induced forms. If LacIq were still active a difference in the banding patterns of the induced and un-induced forms of strain 967 would have been seen. There was a shift between the O-antigen banding patterns of the 967 strain and the basal 966 strain which showed that in both the 967 samples the O-antigen had been modified by GtrABC(I).

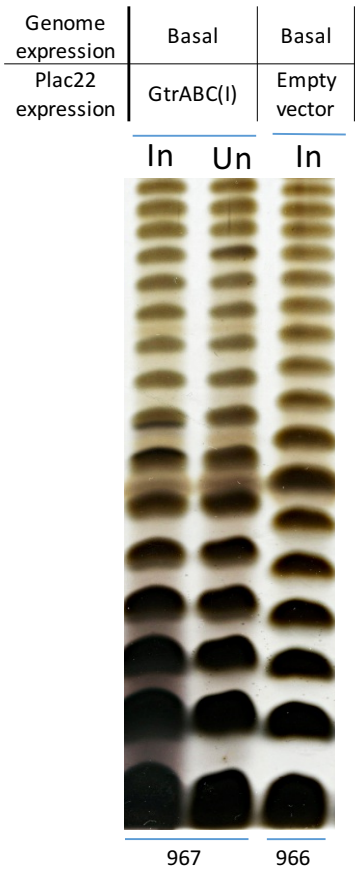


Figure 29: LPS analysis of induced and un-induced strain 967.

Extracted LPS ran on a TSDS-PAGEgel. The table above the lanes indicates which *gtr* genes are being expressed in the strain. The top row gives the chromosomal gtr expression: basal means that no *gtr* genes are expressed from the chromosome of that strain. The bottom row gives the plac22 vector gtr expression: *gtrABC*(I) is being expressed from the lac prompter on the plac22 vector. empty vector means that there are no genes expressed from the lac promoter but the Plac22 vector is still present. The labels above the lanes indicate if the strain was induced (In) or un-induced (Un).

6.2 CAN GTRAB(IV) WORK WITH GTRC(I) TO MODIFY THE O-ANTIGEN?

To study whether the function of GtrAB(IV) was distinct from other families' GtrAB proteins, the expression of *gtr*AB(I), from the family I operon, was replaced by expression of *gtr*AB(IV) making strain 970 as described above. The O-antigen of this strain was studied to see if it had modifications compared to a basal strain's O-antigen. If modification of the O-antigen was possible with this combination of genes then the function of GtrAB(IV) would be similar to that of GtrC(I).

To do this LPS was analysed from *S.* Typhimurium strains: 970 expressing gtrAB(IV) from the Ptac promoter in the genome and gtrC(I) from the lac promoter on the Plac22 vector, 967 expressing only gtrABC(I) from the lac promoter on the Plac22 vector, 968 expressing only gtrC(I) from the lac promoter on the Plac22 vector, 966, which is a basal strain and is transformed with an empty plac22 vector and 969 expressing gtrAB(IV) from the Ptac promoter in the genome and has been transformed with an empty Plac22 vector. Strains 970, 968 and 967 will be refered to as: GtrAB(IV)C(I), GtrC(I) and GtrABC(I) respectively. Details of all strains used can be found in table 16.

LPS was extracted from the strains and ran on a TSDS-PAGE gel (for methods see section 2.9.) that separated out varying lengths of O-antigen repeating units creating a laddered banding pattern. A shift between two strains banding pattern indicated that a sugar molecule had been added to the O-antigen repeating unit in one of the strains.

From comparing the LPS banding patterns of the strains in figure 30 there was a shift seen between strains GtrC(I) (strain 968) and GtrABC(I) (strain 967). Strain GtrC(I) was running level with the Basal+Plac22 strain (strain 966) whereas the GtrABC(I) banding pattern had shifted above the basal level banding pattern showing that the GtrC(I) strain had no modifications to its O-antigen whereas the strain expressing the full family I operon was able to modify the basal O-antigen. Importantly there was a shift between the strain GtrAB(IV)C(I) (strain 970) and GtrC(I). Also the GtrAB(IV)C(I) strain was running at the same level as GtrABC(I) so had modifications to its O-antigen repeating unit.

There was a shift between the banding patterns of strain 970 and 969, both expressing gtrAB(IV) from the genome, the only difference between these two strains was the expression of *gtrC*(I) from the lac promoter in strain 970 whereas strain 969 had an empty Plac22 plasmid. This demonstrates that it was the presence of GtrC(I) that caused a shift, not Plac22

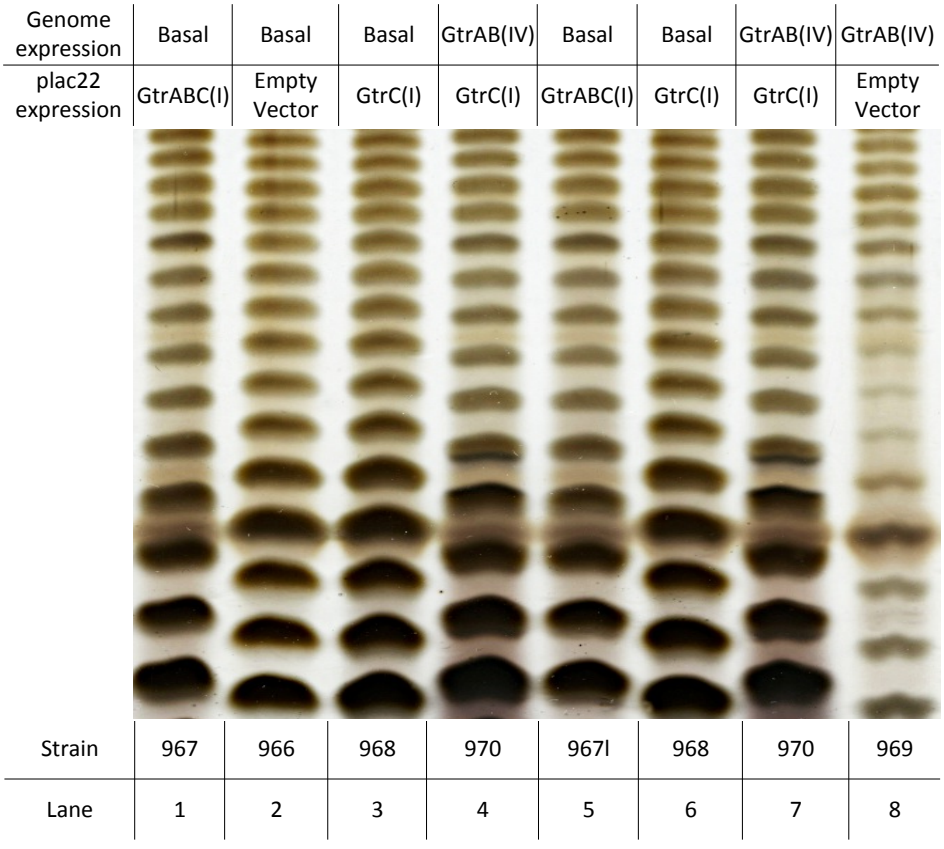


Figure 30: LPS analysis of a strain expressing *gtrAB*(IV) and *gtrC*(I).

Extracted LPS ran on a TSDS-PAGE gel. The tables above the lanes indicate which *gtr* genes are being expressed in that strain. The top row refers to the genes that are being expressed from the Ptac promoter in the genome, basal means that no *gtr* genes are expressed from the chromosome of that strain. The bottom row refers to the *gtr* genes that are being expressed from the lac promoter in the Plac22 plasmid, empty vector means that there are no genes expressed from the lac promoter. The table below the lanes indicate, on the top row, which strain number the LPS was extracted from, details of these strains can be found in table 16. The bottom row of the table gives the lane numbers. All strains were induced with IPTG.

6.3 CHAPTER CONCLUSION

The results in figure 29 have shown that the modification of the Plac22 plasmid successfully prevented the production of an active LacIq protein as there was no difference between an induced and un-induced strain 967 expressing *gtrABC*(I) from the lac promoter. Furthermore, In figure 30 the comparison between strain 970 and 969 demonstrated that *gtrC*(I) and *gtrAB*(IV) were being fully expressed as the modification of strain 970 must have been due to the additional expression of *gtrC*(I) and as *gtrC*(I) can't, by its self, modify the O-antigen then *gtrAB*(IV) must also have sufficient expression in this strain.

In figure 30 GtrC(I) (lane 3) and Basal+Plac22 (lane 4) strains were running equal with each other but shifted against the GtrABC(I) strain (lane 1) showing that the presence of the GtrC(I) protein by its self was not enough to modify the O-antigen and therefore had a basal O-antigen. This result was expected, as GtrC cannot modify the O-antigen without the assistance of GtrAB. But when the expression of *gtrAB*(IV) was introduced the O-antigen was modified and the strains LPS banding pattern ran level with that of a GtrABC(I) strain. This showed that the family IV GtrAB proteins could work with GtrC(I) and replace the function of GtrAB(I). This goes against the idea that the family IV GtrAB proteins have a function distinct from most other families and proves that their activities must be functionally the same. The function of a GtrC protein is specific to each family (Allison and Verma, 2000) and as family I proteins collectively mediate the addition of glucose onto the basal repeating unit the GtrAB(IV) proteins, in place of GtrAB(I), must have also bound a glucose and passed it to GtrC(I).

To improve the design of this experiment a strain that expressed gtrAB(I) from the Ptac promoter in the genome and gtrC(I) from the lac promoter on the Plac22 plasmid would have been a more reliable control. This control could replace strain 967 which expressed gtrABC(I) from the lac promoter on Plac22. This improvement would mean that when comparing it to strain 970 both gtrAB(I) and gtrAB(IV) would be expressed from the same place. This would control for any differences in expression from the two promoters. Also, it would control for any difference that came from gtrAB and gtrC being transcribed separately as apposed to in an operon. However, this control would have mainly been necessary if the results showed that strain 970 were not able to modify the O-antigen.

If the GtrAB(IV) can work with the GtrC(I) proteins then why has there not been any recombination of the family IV operon with other family proteins? It could be that the family IV *gtr* operon is evolutionarily younger than the other *gtr* families and there has not been as much time for the operon to recombine. Or the potential TSS within *gtrB* may have prevented any recombined operons from persisting within a population because *gtrC*(IV) needs to be down stream of *gtrB*(IV) to have the correct transcriptional control relevant to its role in infection. The differences observed in the family IV GtrAB AA sequence could reflect that the family IV operon is more evolutionarily distant from the other families.

7. DISCUSSION

7.1 EXPRESSION OF THE FAMILY IV *GTR* GENES

The results of chapter 4 did not identify and conditions that induce up regulation of the family IV gtr genes. However, the results provided other information about the expression of the *gtr* genes from the main operon promoter and a possible promoter within *gtrB*

7.1.1 Expression from the main operon promoter upstream of *gtrA*

The results in chapter 4 suggest that under certain conditions the family IV operon promoter shows variable levels of expression. It could be the case that this operon phase varies under certain conditions. Changes to switch frequency of phase variable genes due to changes in environmental stimuli has been observed in E. coli where temperature and growth media can modulate the on to off switch frequency of the type 1 fimbriae genes (Gally et al., 1993).

However, phase variation was not observed when the family IV operon promoter region was put in front of the *lacZ* gene in this study and others(Broadbent et al., 2010). But it may be the case that a larger up stream region is needed for phase variation that was not included in the *lacZ* constructs. Or, phase variation of the *lacZ* construct was not seen because the switch frequency is too quick for a difference between two colonies expression of the *lacZ* gene to be appreciated through the colour of the colony. The family IV operon can not phase vary in the same way as many other *gtr* families can; through a OxyR and Dam dependent mechanism. The upstream region of the family IV operon has no GATC sites, which are methylation sites need for phase variable expression through the Dam and OxyR mechanism. Other DNA motifs associated with phase variation have not been identified either; no tandem repeats were found within the sequence upstream of the *gtrA* start codon using a tandem repeat finder (Benson, 1999). Tandem repeats are associated with slipped-strand mispairing; short repeating DNA

sequences cause mispairing of strands during DNA replication altering the number of repeats which can effect transcription if the repeats are in a promoter region or translation if the repeats are after the transcriptional start site (Torres-Cruz and van der Woude, 2003) (van der Woude, 2011).

But there are other methods of phase variation that have been observed in bacteria that could possibly be the case for the family IV genes. The family IV genes could be under the control of a transcriptional promoter or suppressor that has a phase variable expression. An example of this method of phase variation is the *filC* gene encoding the flagellin protein. Its expression is repressed by the FljA protein, which is produced from a phase variable promoter controlled by a DNA inversion mechanism. The FljA protein also represses translation of the protein by binding to the 5'-untranslated region and consequently promotes degradation of the transcript (Bonifield and Hughes, 2003) (Yamamoto and Kutsukake, 2006). This method of phase variable control would not be detectable using a beta-galactosidase assay. Other mechanisms of phase variation that don't involve the phase variation from the gene's promoter have been observed. The *cmpV* gene in E. coli is controlled by a DNA inversion mechanism which is down stream of the promoter region. It is predicted that in the off state the transcript, containing the switch region, forms a stable stem loop followed by a poly U tract, which causes transcription to be terminated (Emerson et al., 2009).

7.1.2 Expression from a possible promoter within *gtrB*

As well as the main operon promoter it was predicted that *gtrC* could be expressed independently from TSS within *gtrB* based on RNA-seq data of *S.* Typhimurium primary transcripts (personal correspondence with Disa Hammerlof). Expression from a region within *gtrB* was measured and showed a 7-fold increase in expression compared to a control (section 4.3.). Possible promoter regions were also found upstream of the *gtrC* start codon. These results indicate that the *gtrC*(IV) can be expressed by its self independently of the main promoter. This is supported by the results seen in section 4.1.2., which suggested that the two genes were being expressed independently of each other and responding differently to environmental ques.

Independent and differential expression of *gtrC* from the main operon promoter suggests that *gtrC* may have a function independent of *gtrAB* under certain conditions. This is supported by the findings of Chaudhuri et al., who reported that *gtrABC* were needed for persistence within

chickens, pigs and cattle but only *gtrC* was needed for persistence within a mouse (Chaudhuri et al., 2013). This implies that GtrC does not need GtrAB for persistence within a mouse. GtrC could have a secondary function that does not require GtrAB, for example it could be functioning with other proteins that transfer a different sugar or are localized to a different part of the membrane.

The RNA-seq data produced by Kröger et al., (Kröger et al., 2013) also showed that the family IV genes had different transcript levels under different conditions. Consistent with the results in section 4.1.2. the RNA-seq data showed that *gtrC* had a higher up regulation than *gtrA* under pH3 conditions compared to the control. Also, when comparing expression from the region within *gtrB* with the *gtrA* upstream region, there appears to be more expression from the promoter within *gtrB* than the main operon promoter (sections 4.2.1. and 4.3.). However it would be useful to repeat these assays in parallel so that all conditions can be kept constant. In contrast to this, the RNA-seq data shows that over all the transcript level, measured in transcripts per million, is highest in *gtrA* and lowest in *gtrB*, with *gtrC* in the middle. These differences could be due to a combination of post-transcriptional controls and differential expression of the genes through different promoters.

7.2 WHAT SUGAR ARE THE FAMILY IV GTR PROTEINS BINDING?

To date only the function of four out of ten of the *gtr* families has been elucidated. Both families I and III mediate the addition of a glucose onto the O-antigen repeating unit of *S.* Typhimurium. The phylogenetic tree generated from the alignments of all family GtrC proteins in section 3.5.3. showed that the family I, III and IV clustered into a group showing that family IV GtrC had a higher sequence similarity with the family I and III GtrC proteins than with the other Gtr families. The phylogenetic tree further showed that there was a higher sequence similarity between GtrC proteins from family III and IV than between family III and I. An alignment of just the family I, III and IV GtrC proteins showed that the N-terminal region was more conserved between all proteins with the long periplasmic C-terminal tail having less sequence similarity. As both family I and III proteins mediate the addition of a glucose then we would expect there to be a conserved region binding the glucose sugar. There is a region between the first two trans-membrane domains, in the periplasm, that is the only non trans-membrane domain with a cluster of conserved AA. As these AA are conserved in both the

family I and III proteins they may be involved in binding the glucose molecule. To strengthen this theory, glycosyl transferase proteins that are known to bind a glucose molecule, including the family I and III proteins, were aligned. This alignment had just one clustering of conserved AA, which is likely to be the AA involved in binding the glucose molecule as this is the functional commonality between these proteins. This clustering of conserved AA was at the same position as the one found in the alignment of the family I, III and IV GtrC proteins. Although, in family IV one of the conserved AA was not present but the other three: tryptophan, glycine and aspartic acid were conserved. The presence of three conserved AA that appear to be needed for glucose binding could indicate that GtrC(IV) is binding a glucose sugar. The conclusions of chapter 6 also support this.

The results in chapter 6 suggest that GtrAB(IV) bound a glucose molecule. They were able to replace the function of GtrAB(I) despite both GtrAB(IV) being the most distant in terms of AA sequence compared to the other families studied.

Although it is possible that GtrAB(IV) are not specific in the sugar they bind and the result in chapter 6 does not show that GtrC(IV) binds a glucose, this result combined with the bioinformatics analysis strongly suggests that the GtrABC(IV) collectively bind and transfer a glucose molecule.

7.3 IS THE O-ANTIGEN THE TARGET OF THE FAMILY IV GTR PROTEINS?

It was predicted that the family IV GtrABC proteins function to modify the O-antigen repeating unit of the LPS based on their similarity with other families of Gtr proteins know to modify the O-antigen by glycosylation. Also, the bioinformatics analysis of GtrC(IV) (section 3.5.3.) showed that its AA sequence was very similar to the GtrC proteins of family I and III. The sequence similarity between GtrC(IV) and GtrC(III) was higher than the similarity between GtrC(III) and GtrC(I), which appeared to be due to clusters of AA in the C-terminal tails of GtrC(III) and GtrC(IV) which were not present in GtrC(I). The C-terminal tail is predicted to determine the specific modification and bind the O-antigen at a specific position due to a lower sequence similarity in this area between the family I and III proteins. Both are known to add a glucose molecule on to a galactose moiety but at different positions; family I makes a 1-6

linkage whereas family III makes a 1-4 linkage. Therefore, the clusters of similar sequences in the C-terminal tails shared only by GtrC(III) and GtrC(IV) could indicate that the family IV proteins also mediate the addition of a sugar molecule through a 1-4 linkage to a galactose moiety.

However, when a cell constitutively expressing the family IV genes is analysed by running its LPS on a gel against basal LPS (extracted from a cell with no known O-antigen modifying genes) there is no difference between the banding patterns of each LPS extract. If the O-antigen were being modified we would expect to see a shift between the two banding patters. There could be many possible reasons for this result; the following sections will explore the arguments for and against each possibility.

7.3.1 The *gtr* transcript is not being translated into protein.

One of the reasons for not seeing a shift may be because the family IV gene's constitutively expressed mRNA was not being translated into protein due to post-transcriptional controls under normal laboratory growth conditions. If the proteins function is important for growth and persistence during infection, then they may not be needed when grown under normal laboratory conditions. For example, the activity and production of the proteins may be energetically costly for the cell so their translation is restricted to times when they are need for persistence within a host.

However, the results in section 6.2., which showed that GtrAB(IV) could work with GtrC(I) to modify the O-antigen, also showed that *gtrAB*(IV) must have been transcribed and translated into protein. In both this experiment and the Davies et al., experiments the same laboratory conditions were used and the genes were being expressed from the same constitutive promoter. Therefore, in the Davies et al., experiment *gtrAB*(IV) will have been transcribed into protein.

However, this still does not rule out the possibility that the *gtrC*(IV) transcript was not being translated into protein as the experiments in section 6.2 only studied GtrAB(IV) activity. The results in section 6.2. showed that GtrAB(IV) can work with GtrC(I). GtrAB(IV) might also be able to work with GtrC of family III, which are also present in the S. Typhimurium genome along with the family IV genes, as both family I and III modify the O-antigen by addition of a

glucose. So translation of GtrAB(IV) could still be favorable for the cell even if GtrC(IV) is not under normal laboratory growth conditions. The argument for the possibility that the *gtrC* and *gtrAB* transcripts have different post-transcriptional controls is supported by the observation that the promoter within *gtrB* seems to have different transcriptional controls to the main operon promoter discussed above.

7.3.2 Are other modifications to the O-antigen are needed to provide the GtrC(IV) with a target for activity?

One of the reasons a shift from the basal banding pattern was not seen could have been because GtrC(IV)'s target includes a modification of the core repeating unit of the O-antigen. The O-antigen core repeating units of each serovar containing a family IV operon do not have a common sugar moiety. If GtrC(IV) is specifically targeting the O-antigen, then further modifications are almost certainly needed on at least one of the serovar's O-antigens. All but the Cubana serovar have a mannose moiety in their core O-antigen and all but the O:7 group of serovars have a galactose moiety as part of their core O-antigen. The bioinformatical analysis done in section 3.5.3 identified the galactose moiety as a possible target of GtrC(IV) based on the proteins similarity with GtrC(III). While the Cubana serovar has no other *gtr* operons in its genome, the O:7 group of family IV serovars all contain the family V and VI operons as well as the family IV genes. The family V and VI Gtr proteins are known to modify the O-antigen (Davies et al., 2013) but the specific modification each family makes is unknown. Therefore, it is possible that the O-antigen of all family IV, O:7 group serovar's core is being modified by either the family V or VI Gtr proteins by addition of a galactose sugar. This would provide a common sugar amongst all know O-antigen cores of family IV serovars. However, this does not explain why there was no shift seen when *S.* Typhimurium expressed the family IV genes. The *S.* Typhimurium serovar used in the experiment carried out by Davies et al. has an O:4 group O-antigen that has a galactose sugar as part of the core O-antigen.

Chapter 5 looked at whether any of the other *gtr* genes that share a genome with the family IV *gtr* genes were providing a modification to the O-antigen need for GtrC(IV) activity. The results did not show that the family III proteins in *S.* Typhimurium or the family V proteins in *S. Infantis* provided GtrC(IV) with a target for activity. The results of the experiment that looked at the family VI modification indicated that it did not provide a target for GtrC(IV) activity but the result was not conclusive and needed to be repeated.

If neither family V or VI are adding a galactose moiety to the basal O-antigen, then there may be other proteins present within the O:7 serovars that mediate the addition of a galactose. Or, a galactose moiety is not needed and another sugar is the common target of GtrC(IV).

As well as the Gtr proteins there are other proteins that can modify the LPS. Bacteriophage often carry bacterial virulence factors, many of these target the LPS (Bondy-Denomy and Davidson, 2014). For example the epsilon 15 phage modifies the core O-antigen of group O:3,10 by changing the alpha-galactose to a beta-galactose (Losick and Robbins, 1967), which allows the GtrC protein of epsilon 34 to glucosylate the beta-galactose at carbon four(Kropinski et al., 2007; Weinbaum et al., 2013). This demonstrates that for some O-antigen modifying proteins the anomeric form of the sugar is a determinant of specificity. This could be the case for the family IV proteins. Furthermore, the galactose moiety of the *S.* Typhimurium O-antigen has been reported to have both an alpha and a beta anomeric configuration (Bogomolnaya et al., 2008; Huang et al., 2012; Liu et al., 2014b; Reeves, 1993). This suggests that this is a modification that is made to the *S.* Typhimurium O-antigen. If the GtrC(IV) is targeting the galactose at carbon 4, as was suggested earlier, and is specific to one anomeric configuration, then this could be the reason why modification of the *S.* Typhimurium O-antigen was not seen by Davies et al., as the galactose may have been in the wrong anomeric configuration. The majority of the galactose containing family IV O-antigen structures are reported to have a galactose in the alpha configuration (section 3.2.). This could mean that the family IV proteins target an alpha-galactose rather than a beta-galactose if galactose is their target.

The bioinformatics analysis suggested that GtrC(IV) might have a very similar function and even make the same linkage as GtrC(III). However, the results of Davies et al., suggest that this isn't the case because GtrABC(III) can mediate a detectable modification of the O-antigen while GtrABC(IV) can not. But if GtrC(III) was either targeting an opposite anomeric configuration to Gtr(IV) or was not specific for anomeric configuration then it would be possible that the two proteins could have an almost identical function but GtrC(III) would be able to modify the *S.* Typhimurium O-antigen whilst the target of Gtr(IV) is not present.

7.3.3 The modification mediated by the family IV Gtr proteins does not give a detectable shift.

An alternative reason for why a strain constitutively expressing the Gtr(IV) proteins did not have a detectable change in its O-antigen compared to a basal strain could have been because the modification mediated by the family IV proteins doesn't change the size of the O-antigen repeating units enough to cause a visible shift to the O-antigen banding pattern.

This could be because the sugar molecule added to the O-antigen is very small. For example, some of the smallest monosaccharaides are d- and l-glyceraldehydes and dihydroxyacetones, which consist of only three carbons as opposed to six carbons that make up a glucose molecule (Berg et al., 2002). The results of chapter 6 showed that GtrAB(IV) could bind a glucose and the bioinformatics analysis supported the idea that the family IV Gtr proteins bind and transfer a glucose. If this is the case then a shift should be seen similar to the addition of a glucose molecule by Gtr(I) and Gtr(III), where there is a clear shift in the O-antigen banding pattern compared to a basal O-antigen. However, the family IV proteins might be modifying a smaller number of repeating units. For example, GtrC(IV) might have a lower activity to the other Gtr proteins of known function. This would mean that the modification would only be increasing the size of the O-antigen by a small proportion and might not visibly alter the position of the band in the gel compared the basal O-antigen of the same length.

Another type of modification that would not cause a shift would be one that added a sugar to the O-antigen which is replacing the position of a sugar that is part of the basal O-antigen. For example, the protein produced by the *Salmonella* phage epsilon 15, replaces the function of a host cell O-polysaccharide alpha polymerizing enzyme with a beta polymerizing enzyme. This changes the galactose moiety from alpha to beta (Kropinski et al., 2007). As this phage protein is replacing the function of a host protein it does not require any GtrAB like proteins to function. But GtrAB(IV) have been maintained within multiple *Salmonella* serovar genomes as functional proteins, which has been demonstrated in chapter 6, suggesting that they have a functional role in a cell. However, GtrC(IV) could still have the same purpose as the epsilon 15 protein but function differently and still require GtrAB(IV) to transfer a sugar molecule.

7.3.4 The O-antigen is not the target of family IV GtrC(IV)

The reason there is no detectable modification made to the basal O-antigen when GtrABC(IV) is constitutively expressed could be because the O-antigen is not the target of GtrC(IV). This is supported by the observation that there is no common sugar between the family IV core O-antigen structures and there is a wider diversity of O-antigen groups occupied by the family IV genes compared to the other *gtr* families. This diversity could indicate that the family IV genes are not restricted by the O-antigen structure of a strain because the O-antigen is not the target of GtrC(IV). If this is the case we would expect the GtrC(IV) target to be at the surface of the cell as GtrABC(IV) are all membrane proteins. We would also expect the target to be a polysaccharide because GtrC(IV) has a similar sequence to proteins known to target O-antigen polysaccharide units of the LPS.

The bacterial biofilm is an important feature of bacterial persistence within the environment and a host and contains various polysaccharide structures: cellulose, colanic acid, and the O-antigen capsule (Gibson et al., 2006) (Steenackers et al., 2012). Cellulose and colanic acid have been found to be important components when forming a biofilm on chicken epithelial cells (Ledeboer and Jones, 2005), one of the hosts used in the Chaudhuri et al., study (Chaudhuri et al., 2013), which identified *gtrABC*(IV) as important genes during host persistence. In *Salmonella Enteritidis* strains the O-antigen capsule has been associated with persistence within the environment and resistance to desiccation (Gibson et al., 2006). In contrast, the O-antigen capsule of *S.* Typhimurium has been shown to confer resistance to killing by human serum (Marshall and Gunn, 2015).

Whilst cellulose consists of only glucose molecules, both colanic acid and the O-antigen capsule of *S.* Typhimurium contain galactose, (Danese et al., 2000) which, after bioinformatical analysis, may be the target of GtrC(IV). As well as containing galactose, the O-antigen capsule's trisaccharide backbone and the core LPS O-antigen share the same repeating unit in *S.* Typhimurium (Marshall and Gunn, 2015). If the O-antigen capsule is the target of GtrC(IV) then it makes sense that the sequence of GtrC(IV) and GtrC(III) are conserved if there targets are structurally very similar. Also, if GtrC(IV) is making the same modification as GtrC(III), which has been suggested above, it seems unlikely that they would both be maintained within the same serovar. But this problem can be explained if one protein is targeting the O-antigen and the other the O-antigen capsule.

If the O-antigen capsule was the target of GtrC(IV) we would expect them to have a similar expression pattern. The recent study by Marshall and Gunn (Marshall and Gunn, 2015) reported that the O-antigen capsule of *S.* Typhimurium appeared to have a heterogeneouse expression amoungst a population of cells. The results of chapter 4 sugested that *gtrA* could phase vary, but no evidence of phase variation can be seen in the promoter regoin. This could be because its expression is being controlled alongside the O-antigen capsule genes through a phase variable promoter or suppressor.

7.4 FINAL CONCLUSIONS AND FUTURE EXPERIMENTS

The bioinformatics analysis of GtrC(IV) and results in section 6.3. strongly suggest that the sugar transferred by the GtrABC(IV) is a glucose molecule and is likely to be transferring it onto a galactose moiety at a 1-4 position. The analysis of the serovars containing the family IV *gtr* genes showed that all but the O:7 group of serovars contained a galactose as part of their core O-antigen. Also, there was no common sugar between all the family IV O-antigen core structures. This means that if GtrC(IV) has a specific target then either there are further modifications needed to the core O-antigen or the O-antigen is not the target.

If the O-antigen of the LPS is not the target, then it is likely that GtrC(IV) is targeting the O-antigen capsule as it is structurally very similar and contains a galactose sugar. To test whether this is the case the O-antigen capsule of a cell constitutively expressing the family IV genes needs to be analysed. The O-antigen capsule can be separated from the LPS based on their charge. Once fatty acids are removed from the O-antigen capsule it has a low charge whereas the LPS O-antigen is small with a high charge. O-antigen capsule analysis is more difficult than LPS analysis as the capsules repeating units can be as many as 2,300 or more. Therefore, when analysed using gel electrophoresis the O-antigen capsule can not separate out into a banding pattern and will migrate into a thick band at the separating and stacking gel interface (Snyder et al., 2006). But the composition of the O-antigen capsule can be determined by NMR spectroscopy.

Studies of the expression of the family IV genes demonstrated that *gtrC*(IV) could be expressed independently of the main operon. This could mean that it can function independently of GtrAB(IV) as results in section 4.1.2. showed that the two promoters might have different

transcriptional controls. This could be a sign that expression of GtrC(IV) from the main operon promoter does not fit the needs of its function and has adapted to function slightly differently from the original operons function. For example, it may be functioning at a different location within the membrane and with other glucosyl transferase proteins. This could support the theory that GtrC(IV) is not targeting the O-antigen and has adapted to target another membrane polysaccharide with different expression needs like the O-antigen capsule for example, unlike the LPS the O-antigen capsule is not always present on the cell surface (Marshall and Gunn, 2015).

This theory suggests that GtrAB(IV) is no longer working with GtrC(IV) due the possible difference in their expression controls. As GtrAB(IV) appears to still be functional, as demonstrated in chapter 6, it may be the case that GtrAB(IV) is working with other GtrC proteins in the cell, which was also shown to be possible in chapter 6.

On the other hand, if the O-antigen is the target of GtrC(IV) and a further modification is needed then presumably both the *S.* Typhimurium and *S. Infantis* don't have the target present in their basal O-antigen. Strains from both serovars have shown no detectable modification of the basal O-antigen when *gtrABC*(IV) were constitutively expressed (Davies et al., 2013). The *S.* Typhimurium core O-antigen contains a galactose sugar but the sugar has been reported to have both beta and alpha anomeric forms. As anomeric configuration can be a determinant of specificity for some O-antigen modifying proteins, this could be the reason why no modification was seen in the *S.* Typhimurium serovar. *S. Infantis* is part of the O:7 group serovars so does not contain a galactose, a possible target of GtrC(IV). Although chapter 5 did not show that either the modifications mediated by the family V or VI proteins provided GtrC(IV) with a target there is still the possibility that other genes on the *S. Infantis* genome mediate the addition of a galactose, if this is the target of GtrC(IV). Or, the family V or VI proteins transfer a galactose onto the O-antigen but this is also in the wrong anomeric configuration and is able to be changed.

If the O-antigen is the target of GtrC(IV) it maybe the case that not all of the O-antigen core structures of the family IV serovars need a modification to the O-antigen to provide GtrC(IV) with a target. Hence, other experiments could look at the effect of *gtrABC*(IV) constitutive expression on the O-antigen of other serovars that naturally contain the family IV genes that aren't part of O-antigen groups that have already been studied. The O:8 group seems the most

promising group to study as there are three serovars containing the family IV genes that belong to the O:8 group, whereas the other un-studied groups only contain one family IV serovar.

7.5 CONCLUDING REMARKS

From the results of this project it seems more likely that the target of the family IV Gtr proteins may not be the O-antigen. There is no common sugar between the identified core O-antigen structures and GtrABC(IV) has not been shown to modify the O-antigen after modifications mediated by other Gtr proteins. Also supporting this conclusion is the possibility of GtrC(IV) being independently expressed and having different transcriptional controls from the main promoter suggesting that its function is distinct from other studied Gtr proteins and needs to be expressed at different times and may function alone. The O-antigen capsule is a good candidate for the possible target of GtrC(IV) because its structure is very similar to the O-antigen.

However, there is still the possibility that the basal O-antigen of *S.* Typhimurium and S. Infantis needs to be modified in some way to provide a target for GtrC(IV) activity. As the serovars containing the family IV genes have diverse core O-antigen structures future experiments could look at the effect of constitutive expression of the family IV genes in other serovars that naturally contain the family IV genes.

Structures on the surface of bacteria can play crucial roles in infection, as the family IV proteins are thought to modify the surface of the *Salmonella* pathogen knowledge of their function could give us a deeper understanding of salmonellosis. Furthermore, this body of work has given an insight into the function of the family IV proteins and provided a basis for future studies.

APPENDICES

APPENDIX 1 – TABLES OF BUFFERS, GELS AND MEDIA

Table A1.1: PCN media

Component	Final concentration
MES	0.08 M
Tricine	4 mM
NaCl	0.05 M
D-Glucose	0.022 M
NH_4Cl	0.015 M
$MgSO_4$	1 mM
$CaCl_2$	0.01 µM
Na_2MoO_4	0.01 µM
$Na2SeO_3$	0.3 µM
$CoCl_2$	0.1 µM
$CuSO_4$	0.8 µM
$MnCl_2$	0.001 µM
$ZnSO_4$	0.004 µM
H_3BO_3	0.1 mM
$FeCl_3$	0.376 mM
K_2SO_4	0.4 mM
K_2HPO4/KH_2PO4*	0.4 mM

* K_2HPO_4 and KH_2PO_4 were combined at a ratio of 1:10.76

Table A1.2: M9 minimal media

Component	Final Concentration
M9 Minimal Salts	
Fe Citrate	0.002 mM
$CaCl_2$	0.1 mM
$MgSO_4$	1 mM
Vitamin B1	0.03 mM
Glucose	10 mM
Xgal*	0.1 mM

*was dissolved in DMF. Only added when making solid media

Table A1.3: SOC media

Component	Final Concentration
Tryptone	2% (W/V)
Yeast Extract	0.5% (W/V)
NaCL	10 mM
KCL	2.5 mM
$MgCl_2$	10 mM
$MgSO_4$	10 mM
Glucose	20 mM

Filter Sterilized after dissolving

Table A1.4: LPS buffer A/B

Component	Final Concentration
Tris-HCl	60 mM
EDTA	1 mM
SDS*	2% (W/V)

pH 6.8

*Only in LPS Buffer A

Table A1.5: Separating gel

Component	Final Concentration
Gel buffer (Table X)	33.2% (V/V)
Acrylamide	10% (V/V)
Glycerol	10.4% (V/V)
APS*	0.06% (W/V)
TEMED*	0.6 µl/ml

* were added just before casting

Table A1.6: Stacking gel

Component	Final Concentration
Gel buffer (Table X)	24.8% (V/V)
Acrylamide	4% (V/V)
APS*	0.08% (W/V)
TEMED*	0.8 µl/ml

* were added just before casting

Table A1.7: Cathode buffer

Component	Final Concentration
Tris	0.1 M
Tricine	0.1 M
SDS	0.1% (W/V)

Table A1.8: Fixative solution

Component	Final Concentration
Ethanol	30% (V/V)
Acetic Acid	10% (V/V)

Table A1.9: Oxidizer solution

Component	Final Concentration
Periodic acid	0.033 M
Ethanol	30% (V/V)
Acetic Acid	10% (V/V)

Table A1.10: Developer solution

Component	Final Concentration
Sodium Carbonate	0.28 M
Formaldehyde	0.0074% (V/V)

Table A1.11: Loading buffer

Component	Final Concentration
SDS	6% (W/V)
2-mercaptoethanol	6% (V/V)
Dithiothreitol	10 mM
Glycerol	46% (V/V)
Tris (pH8)	60 mM
Bromophenol blue	0.1% (W/V)

Table A1.12: Gel buffer

Component	Final Concentration
Tris	3 M
SDS	0.3% (W/V)
pH to 8.45	

Table A1.13 Z buffer

Component	Final Concentration
Anhydrous Na_2HPO_4	0.06 M
Anhydrous NaH_2PO_4	0.04 M
KCl	0.01 M
$MgSO_4$	0.001 M
b-Mercaptoethanol	0.05 M

APPENDIX 2 – TABLES OF OLIGOS, PRIMERS AND STRAINS

Table A2.1: Oligos

Code	Sequence	Direction	Purpose
oMV1216	GATTGTTTCGCCGCTTCCTC	R	RT-qPCR primer. Aplify a region within modF, refrence gene for pH3 shock
oMV1217	CTAAATCGCGTCAGGCAACG	F	RT-qPCR primer. Aplify a region within yibT, refrence gene for pH3 shock
oMV1218	GCTCAAAACGGATGAACGGG	R	RT-qPCR primer. Aplify a region within yibT, refrence gene for pH3 shock
oMV1246	GATGTGGCATGAGGAAACGG	F	RT-qPCR primer. Aplify a region within STM1575, refrence gene for NO shock
oMV1247	GCACACCAGTGCGATAAGC	R	RT-qPCR primer. Aplify a region within STM1575, refrence gene for NO shock
oMV1248	GCTTCAGTGCTGGTAATCGTG	F	RT-qPCR primer. Aplify a region within kdgR, refrence gene for NO shock
oMV1249	GCCATCGGTAAAGTGCTTCTG	R	RT-qPCR primer. Aplify a region within kdgR, refrence gene for NO shock
oMV1254	TACTACGGTACCCGTCAACCACC ATCAAACAGG	F	Forward primer to amplify the whole of the Plac22 plasmid excuding the *lacIq* gene. Is positioned at the end of the *lacIq* coding region
oMV1255	TGATTTGGGTACCTGCCATACCG CGAAAGGTT	R	Reverse primer to amplify the whole of the Plac22 plasmid excuding the lacIq gene. Is positioned upstream of the *lacIq* start codon.

Table A2.1: Oligos (continued)

Code	Sequence	Direction	Purpose
oMV1065	GGTAGCT**CTGCAG**TTCAGCGCA ACGTGATTAGT	F	Aplify the 355pb up stream and overlapping region of gtrA(IV) to clone into the CRIM vector infront of the lacZ gene. Contains the Pst1 restriction site
oMV1166	CATGAT**GGTACC**AAACTGACA GCGACAGAAA	R	Aplify the 355pb up stream and overlapping region of gtrA(IV) to clone into the CRIM vector infront of the lacZ gene. Contains the Acc651 restriction site
oMV1184	CATGAT**GGTACC**AGTAGTATCA	R	Aplify a 581bp region within gtrB(IV) containing possible TSS for gtrC(IV) to clone into the CRIM vector infront of the lacZ gene. Contains the Acc651 restriction site
oMV1186	GGTAGCT**CTGCAG**AAGCGCTG AGTGGTTCTACAG	F	Aplify a 581bp region within gtrB(IV) containing possible TSS for gtrC(IV) to clone into the CRIM vector infront of the lacZ gene. Contains the Pst1 restriction site
oMV1209	GTTTACCTTTGGAGCCAGCG	F	RT-qPCR primer. Aplify a region within gtrA(IV)
oMV1210	CATAGCACACTTGTCGCCTG	R	RT-qPCR primer. Aplify a region within gtrA(IV)
oMV1213	TGTTCGTTGTTGCATGGCTG	F	RT-qPCR primer. Aplify a region within gtrC(IV)
oMV1214	GAGACGAAAGGTGAGACGCT	R	RT-qPCR primer. Aplify a region within gtrC(IV)
oMV1215	CAGATCGTCGGTCCTAACGG	F	RT-qPCR primer. Aplify a region within modF, refrence gene for pH3 shock

Table A2.2: Plasmids

Name	Code	Descriotion	Anitbiotic Resistance	Primers used to modify	Source
CRIM vector	p243	Contains the *lacZ* gene, an attp site for site specific intogration, replication origin of R6K (needs pir+ strain)	Chloramphenicol	-	
Plac22	-	Vector containing the lac promoter and lacIq repressor gene	Ampacilin, Tetracyclin	-	
reg gtrA(IV)-lacZ	p440	CRIM vector with a 355bp region taken from upstream and overlapping gtrA(IV) cloned upstream of lacZ	Chloramphenicol	oMV1065, oMV1166	van der Woude Stocks (modified-this project)
within gtrB(IV)-lacZ	p442	CRIM vector with a 581bp region taken from within gtrB(IV) cloned upstream of lacZ	Chloramphenicol	oMV1184, oMV1186	van der Woude Stocks (modified-this project)
lac-grtABC(IV)	p447	LT2 gtrABC(IV) cloned into Plac22 downstream of the lac promotor.	Ampacilin, Tetracyclin		van der Woude Stocks
plac22-lacIq	p459	Plac22 plasmid with the lacIq repressor gene cut out using oMV 1254, oMV1255	Ampacilin, Tetracyclin	oMV 1254, oMV1255	van der Woude Stocks (modified-this project)
lac-gtrABC(I)	p460	p459 with gtrABC(I) under the lac promoter	Ampacilin, Tetracyclin	oMV 1254, oMV1255	van der Woude Stocks (modified-this project)
lac-gtrC(I)	p461	p459 with gtrC(I) under the lac promoter	Ampacilin, Tetracyclin	oMV 1254, oMV1255	van der Woude Stocks (modified-this project)

Table A2.3: Strains

Name	Code	Descriotion	Anitbiotic Resistance	Primers used to	Source
CRIM vector	p243	Contains the *lacZ* gene, an attp site for site specific intogration, replication origin of R6K (needs pir+ strain)	Chloramphenicol	-	
Plac22	-	Vector containing the lac promoter and lacIq repressor gene	Ampicillin, Tetracyclin	-	
reg gtrA(IV)-lacZ	p440	CRIM vector with a 355bp region taken from upstream and overlapping gtrA(IV) cloned upstream of lacZ	Chloramphenicol	oMV1065, oMV1166	van der Woude Stocks (modified-this project)
within gtrB(IV)-lacZ	p442	CRIM vector with a 581bp region taken from within gtrB(IV) cloned upstream of lacZ	Chloramphenicol	oMV1184, oMV1186	van der Woude Stocks (modified-this project)
lac-grtABC(IV)	p447	LT2 gtrABC(IV) cloned into Plac22 downstream of the lac promotor.	Ampicillin, Tetracyclin		van der Woude Stocks
plac22-lacIq	p459	Plac22 plasmid with the lacIq repressor gene cut out using oMV 1254, oMV1255	Ampicillin, Tetracyclin	oMV 1254, oMV1255	van der Woude Stocks (modified-this project)
lac-gtrABC(I)	p460	p459 with gtrABC(I) under the lac promoter	Ampicillin, Tetracyclin	oMV 1254, oMV1255	van der Woude Stocks (modified-this project)
lac-gtrC(I)	p461	p459 with gtrC(I) under the lac promoter	Ampicillin, Tetracyclin	oMV 1254, oMV1255	van der Woude Stocks (modified-this project)

Table A2.3: Strains (continued)

Name	Srain code	Genotype	Plasmid	source
SIN Ptac-gtrABC(VI)	Strain 321	Salmonella Infantis S1326/2, delta:gtrABC(IV)::kan resistant, delta::gtrABC(V), delta:reg-gtrABC(Vi)::Ptac	-	van der Woude Stocks
SIN Ptac-gtrABC(VI)/plac22	Strain 929	Salmonella Infantis S1326/2, delta:gtrABC(IV)::kan resistant, delta::gtrABC(V), delta:reg-gtrABC(Vi)::Ptac	Plac22	van der Woude Stocks
SIN Ptac-gtrABC(VI)/p447	Strain 930	Salmonella Infantis S1326/2, delta:gtrABC(IV)::kan resistant, delta::gtrABC(V), delta:reg-gtrABC(Vi)::Ptac	p447	van der Woude Stocks
STM Ptac-gtrABC(III)	Strain 122	Salmonella Typhimurium LT2, delta:oafA, delta:gtrABC(IV)::kan resistant, delta:reg-gtrABC(III)::Ptac	-	van der Woude Stocks
STM Ptac-gtrABC(III)/gtrABC(IV)	Strain 921	Salmonella Typhimurium LT2, delta:oafA, delta:gtrABC(IV)::kan resistant, delta:reg-gtrABC(III)::Ptac	p447	van der Woude Stocks
Basal STM /gtrABC(I)	Strain 967	Basal Salmonella Typhimurium LT2, delta:oafA, delta:gtrABC(III)::tet resistant, delta:gtrABC(IV)::Kan resistant	p460	van der Woude Stocks (transformation - this project)
Basal STM /plac22	Strain 966	Basal Salmonella Typhimurium LT2, delta:oafA, delta:gtrABC(III)::tet resistant, delta:gtrABC(IV)::Kan resistant	p459	van der Woude Stocks (transformation - this project)
Basal STM /gtrC(I)	Strain 968	Basal Salmonella Typhimurium LT2, delta:oafA, delta:gtrABC(III)::tet resistant, delta:gtrABC(IV)::Kan resistant	p461	van der Woude Stocks (transformation - this project)
STM Ptac-gtrAB(IV)/gtrC(I)	Strain 970	Basal Salmonella Typhimurium LT2, delta:oafA, delta:gtrABC(III)::kan resistant, delta:reg-gtrABC(IV)::Ptac delta:gtrC(IV)::tet resistant	p461	van der Woude Stocks (transformation - this project)
STM Ptac-gtrAB(IV)/Plac22	Strain 969	Basal Salmonella Typhimurium LT2, delta:oafA, delta:gtrABC(III)::kan resistant, delta:reg-gtrABC(IV)::Ptac delta:gtrC(IV)::tet resistant	p459	van der Woude Stocks (transformation - this project)
Basal STM/Ptac-gtrAB(IV)	Strain 141	Basal Salmonella Typhimurium LT2, delta:oafA, delta:gtrABC(III)::kan resistant, delta:reg-gtrABC(IV)::Ptac delta:gtrC(IV)::tet resistant	-	van der Woude Stocks
STM LT2	Strain 917	Salmonella Typhimurium LT2. p440 intograted into the lambda att site	-	This Project

APPENDIX 3 – BIOINFORMATICS FIGURES

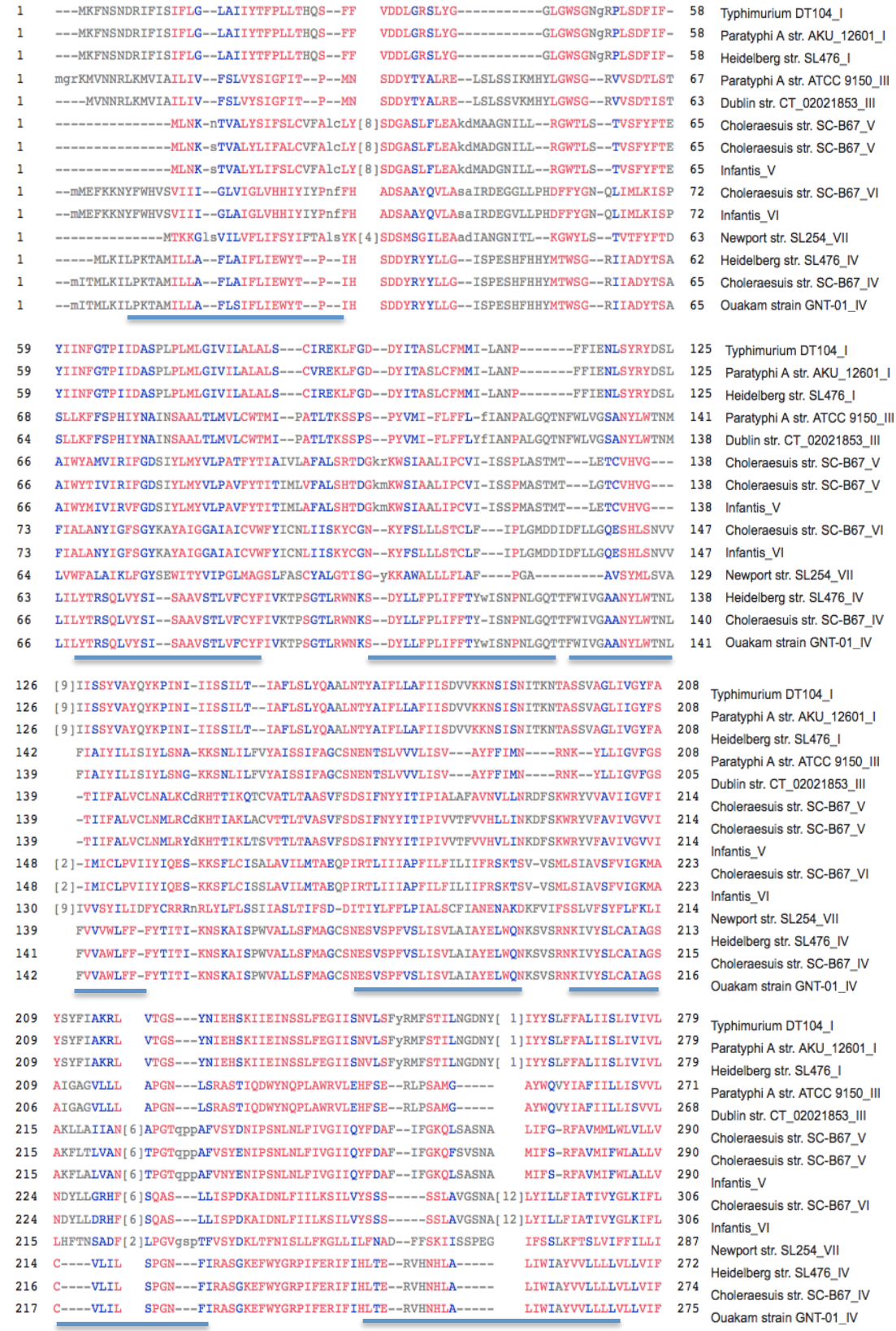


Figure Continued on next page.

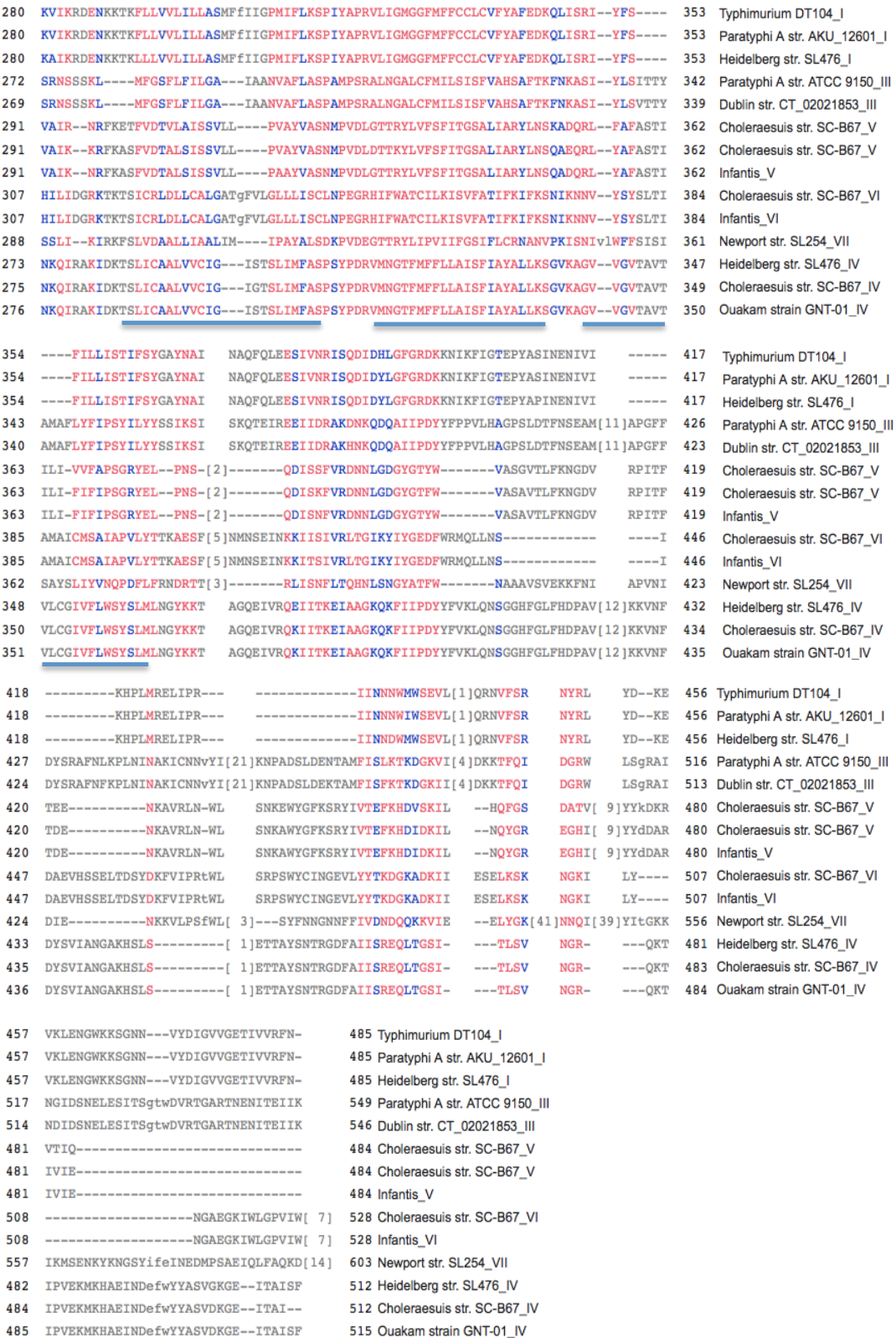


Figure A3.1: GtrC, from 6 different Gtr families, AA alignment.

After each sequence is the name of the serovar, strain and *gtr* family the sequence belongs to. Numbers either side of the sequences are the first and last AA number. Bracketed numbers in grey are AA that are not shown and indicates how many. Lower case AA in grey show low-complexity sequences that have been filtered out when building the alignment. Coloured regions indicate the sequences have some conservation; red indicates AA conserved across all of the proteins and blue shows that some proteins have a change in AAs. The blue lines underneath the sequences indicate the transmembrane regions of the family IV GtrB protein.

REFERENCES

Allison, G.E., and Verma, N.K. (2000). Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri. Trends Microbiol. *8*, 17–23.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Amann, E., Brosius, J., and Ptashne, M. (1983). Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in Escherichia coli. Gene *25*, 167–178.

Bangtrakulnonth, A., Pornreongwong, S., Pulsrikarn, C., Sawanpanyalert, P., Hendriksen, R.S., Wong, D.M.A.L.F., and Aarestrup, F.M. (2004). Salmonella Serovars from Humans and Other Sources in Thailand, 1993–2002. Emerg. Infect. Dis. *10*, 131–136.

Batchelor, R.A., Haraguchi, G.E., Hull, R.A., and Hull, S.I. (1991). Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of Escherichia coli. J. Bacteriol. *173*, 5699–5704.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.

Berg, J.M., Tymoczko, J.L., Stryer, L., Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). Biochemistry (W H Freeman).

Bogomolnaya, L.M., Santiviago, C.A., Yang, H.-J., Baumler, A.J., and Andrews-Polymenis, H.L. (2008). "Form variation" of the O12 antigen is critical for persistence of Salmonella Typhimurium in the murine intestine. Mol. Microbiol. *70*, 1105–1119.

Bondy-Denomy, J., and Davidson, A.R. (2014). When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. J. Microbiol. *52*, 235–242.

Bonifield, H.R., and Hughes, K.T. (2003). Flagellar phase variation in Salmonella enterica is mediated by a posttranscriptional control mechanism. J. Bacteriol. *185*, 3567–3574.

Broadbent, S.E., Davies, M.R., and van der Woude, M.W. (2010). Phase variation controls expression of Salmonella lipopolysaccharide modification genes by a DNA methylation-dependent mechanism. Mol. Microbiol. *77*, 337–353.

Byl, C.V., and Kropinski, A.M. (2000). Sequence of the Genome of SalmonellaBacteriophage P22. J. Bacteriol. *182*, 6472–6481.

Chan, K., Baker, S., Kim, C.C., Detweiler, C.S., Dougan, G., and Falkow, S. (2003). Genomic Comparison of Salmonella enterica Serovars and Salmonella bongori by Use of an S. enterica Serovar Typhimurium DNA Microarray. J. Bacteriol. *185*, 553–563.

Chaudhuri, R.R., Morgan, E., Peters, S.E., Pleasance, S.J., Hudson, D.L., Davies, H.M., Wang, J., van Diemen, P.M., Buckley, A.M., Bowen, A.J., et al. (2013). Comprehensive Assignment of Roles for Salmonella Typhimurium Genes in Intestinal Colonization of Food-Producing Animals. PLoS Genet *9*, e1003456.

Coburn, B., Grassl, G.A., and Finlay, B.B. (2007). Salmonella, the host and disease: a brief review. Immunol. Cell Biol. *85*, 112–118.

Crévieu-Gabriel, I., Gomez, J., Caffin, J.P., and Carré, B. (1999). Comparison of pig and chicken pepsins for protein hydrolysis. Reprod. Nutr. Dev. *39*, 443–454.

Crump, J.A., and Heyderman, R.S. (2015). Y A Perspective on Invasive Salmonella Disease in Africa. Clin. Infect. Dis. *61*, S235–S240.

Danese, P.N., Pratt, L.A., and Kolter, R. (2000). Exopolysaccharide Production Is Required for Development of Escherichia coli K-12 Biofilm Architecture. J. Bacteriol. *182*, 3593–3596.

Davies, M.R., Broadbent, S.E., Harris, S.R., Thomson, N.R., and van der Woude, M.W. (2013). Horizontally Acquired Glycosyltransferase Operons Drive Salmonellae Lipopolysaccharide Diversity. PLoS Genet. *9*.

Dixon, D.R., and Darveau, R.P. (2005). Lipopolysaccharide heterogeneity: innate host responses to bacterial modification of lipid a structure. J. Dent. Res. *84*, 584–595.

Dougan, G., and Baker, S. (2014). Salmonella enterica Serovar Typhi and the Pathogenesis of Typhoid Fever. Annu. Rev. Microbiol. *68*, 317–336.

Edwards, N.J., Monteiro, M.A., Faller, G., Walsh, E.J., Moran, A.P., Roberts, I.S., and High, N.J. (2000). Lewis X structures in the O antigen side-chain promote adhesion of Helicobacter pylori to the gastric epithelium. Mol. Microbiol. *35*, 1530–1539.

Emerson, J.E., Reynolds, C.B., Fagan, R.P., Shaw, H.A., Goulding, D., and Fairweather, N.F. (2009). A novel genetic switch controls phase variable expression of CwpV, a Clostridium difficile cell wall protein. Mol. Microbiol. *74*, 541–556.

Eom, J.S., Kim, J.S., Jang, J.I., Kim, B.-H., Yoo, S.Y., Choi, J.H., Bang, I.-S., Lee, I.S., and Park, Y.K. (2013). Enhancement of Host Immune Responses by Oral Vaccination to Salmonella enterica serovar Typhimurium Harboring Both FliC and FljB Flagella. Plos One *8*, e74850.

European Food Safety Authority (2010). Scientific Opinion on monitoring and assessment of the public health risk of “Salmonella Typhimurium-like” strains. EFSA J. *8*, Article 1826.

Fukazawa, Y., and Hartman, P.E. (1964). A P22 BACTERIOPHAGE MUTANT DEFECTIVE IN ANTIGEN CONVERSION. Virology *23*, 279–283.

Galanos, C., Lehmann, V., Lüderitz, O., Rietschel, E.T., Westphal, O., Brade, H., Brade, L., Freudenberg, M.A., Hansen-Hagge, T., Lüderitz, T., et al. (1984). Endotoxic properties of chemically synthesized lipid A part structures. Eur. J. Biochem. *140*, 221–227.

Gally, D.L., Bogan, J.A., Eisenstein, B.I., and Blomfield, I.C. (1993). Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in Escherichia coli K-12: effects of temperature and media. J. Bacteriol. *175*, 6186–6193.

Gibson, D.L., White, A.P., Snyder, S.D., Martin, S., Heiss, C., Azadi, P., Surette, M., and Kay, W.W. (2006). Salmonella Produces an O-Antigen Capsule Regulated by AgfD and Important for Environmental Persistence. J. Bacteriol. *188*, 7722–7730.

Goktepe, I., Juneja, V.K., and Ahmedna, M. (2005). Probiotics in Food Safety and Human Health (CRC Press).

Grimont, P.A.D., and Weill, F.-X. (2007). ANTIGENIC FORMULAE OF THE SALMONELLA SEROVARS.

Grossman, N., Schmetz, M.A., Foulds, J., Klima, E.N., Jimenez-Lucho, V.E., Leive, L.L., Joiner, K.A., and Jiminez, V. (1987). Lipopolysaccharide size and distribution determine serum resistance in Salmonella montevideo. J. Bacteriol. *169*, 856–863.

Gyorfy, Z., Duda, E., and Vizler, C. (2013). Interactions between LPS moieties and macrophage pattern recognition receptors. Vet. Immunol. Immunopathol. *152*, 28–36.

Haldimann, A., and Wanner, B.L. (2001). Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J. Bacteriol. *183*, 6384–6393.

Hauser, E., Junker, E., Helmuth, R., and Malorny, B. (2011). Different mutations in the oafA gene lead to loss of O5-antigen expression in Salmonella enterica serovar Typhimurium. J. Appl. Microbiol. *110*, 248–253.

Hernandez, L.D., Hueffer, K., Wenk, M.R., and Galán, J.E. (2004). Salmonella modulates vesicular traffic by altering phosphoinositide metabolism. Science *304*, 1805–1807.

Huang, J.X., Azad, M.A.K., Yuriev, E., Baker, M.A., Nation, R.L., Li, J., Cooper, M.A., Velkov, T., Huang, J.X., Azad, M.A.K., et al. (2012). Molecular Characterization of Lipopolysaccharide Binding to Human α-1-Acid Glycoprotein, Molecular Characterization of Lipopolysaccharide Binding to Human α-1-Acid Glycoprotein. J. Lipids J. Lipids *2012, 2012*, e475153.

Islam, S.T., and Lam, J.S. (2014). Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. Can. J. Microbiol. *60*, 697–716.

Jiang, X.M., Neal, B., Santiago, F., Lee, S.J., Romana, L.K., and Reeves, P.R. (1991). Structure and sequence of the rfb (O antigen) gene cluster of Salmonella serovar typhimurium (strain LT2). Mol. Microbiol. *5*, 695–713.

Kalupahana, R.S., Mastroeni, P., Maskell, D., and Blacklaws, B.A. (2005). Activation of murine dendritic cells and macrophages induced by Salmonella enterica serovar Typhimurium. Immunology *115*, 462–472.

Kalynych, S., Morona, R., and Cygler, M. (2014). Progress in understanding the assembly process of bacterial O-antigen. FEMS Microbiol. Rev. *38*, 1048–1065.

Kintz, E., Davies, M.R., Hammarlöf, D.L., Canals, R., Hinton, J.C.D., and van der Woude, M.W. (2015). A BTP1 prophage gene present in invasive non-typhoidal Salmonella determines composition and length of the O-antigen of the lipopolysaccharide. Mol. Microbiol. *96*, 263–275.

Kittelberger, R., and Hilbink, F. (1993). Sensitive silver-staining detection of bacterial lipopolysaccharides in polyacrylamide gels. J. Biochem. Biophys. Methods *26*, 81–86.

Knodler, L.A., Vallance, B.A., Celli, J., Winfree, S., Hansen, B., Montero, M., and Steele-Mortimer, O. (2010). Dissemination of invasive Salmonella via bacterial-induced extrusion of mucosal epithelia. Proc. Natl. Acad. Sci. U. S. A. *107*, 17733–17738.

Kolodziejek, A.M., and Miller, S.I. (2015). Salmonella modulation of the phagosome membrane, role of SseJ. Cell. Microbiol. *17*, 333–341.

Korres, H., Mavris, M., Morona, R., Manning, P.A., and Verma, N.K. (2005). Topological analysis of GtrA and GtrB proteins encoded by the serotype-converting cassette of Shigella flexneri. Biochem. Biophys. Res. Commun. *328*, 1252–1260.

Kröger, C., Colgan, A., Srikumar, S., Händler, K., Sivasankaran, S.K., Hammarlöf, D.L., Canals, R., Grissom, J.E., Conway, T., Hokamp, K., et al. (2013). An Infection-Relevant Transcriptomic Compendium for Salmonella enterica Serovar Typhimurium. Cell Host Microbe *14*, 683–695.

Kropinski, A.M., Kovalyova, I.V., Billington, S.J., Patrick, A.N., Butts, B.D., Guichard, J.A., Pitcher, T.J., Guthrie, C.C., Sydlaske, A.D., Barnhill, L.M., et al. (2007). The Genome of ε15, a

Serotype-Converting, Group E1 Salmonella enterica-Specific Bacteriophage. Virology *369*, 234–244.

LaRock, D.L., Chaudhary, A., and Miller, S.I. (2015). Salmonellae interactions with host processes. Nat. Rev. Microbiol. *13*, 191–205.

Ledeboer, N.A., and Jones, B.D. (2005). Exopolysaccharide sugars contribute to biofilm formation by Salmonella enterica serovar typhimurium on HEp-2 cells and chicken intestinal epithelium. J. Bacteriol. *187*, 3214–3226.

Lerouge, I., and Vanderleyden, J. (2002). O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. FEMS Microbiol. Rev. *26*, 17–47.

Lesse, A.J., Campagnari, A.A., Bittner, W.E., and Apicella, M.A. (1990). Increased resolution of lipopolysaccharides and lipooligosaccharides utilizing tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis. J. Immunol. Methods *126*, 109–117.

Liang-Takasaki, C.J., Mäkelä, P.H., and Leive, L. (1982). Phagocytosis of bacteria by macrophages: changing the carbohydrate of lipopolysaccharide alters interaction with complement and macrophages. J. Immunol. Baltim. Md 1950 *128*, 1229–1235.

Liu, B., Knirel, Y.A., Feng, L., Perepelov, A.V., Senchenkova, S.N., Reeves, P.R., and Wang, L. (2014a). Structural diversity in SalmonellaO antigens and its genetic basis. FEMS Microbiol. Rev. *38*, 56–89.

Liu, B., Knirel, Y.A., Feng, L., Perepelov, A.V., Senchenkova, S.N., Reeves, P.R., and Wang, L. (2014b). Structural diversity in Salmonella O antigens and its genetic basis. FEMS Microbiol. Rev. *38*, 56–89.

Losick, R., and Robbins, P.W. (1967). Mechanism of ϵ15 conversion studied with a bacterial mutant. J. Mol. Biol. *30*, 445–455.

MacMicking, J., Xie, Q.W., and Nathan, C. (1997). Nitric oxide and macrophage function. Annu. Rev. Immunol. *15*, 323–350.

Mäkelä, P.H. (1973). Glucosylation of Lipopolysaccharide in Salmonella: Mutants Negative for O Antigen Factor 1221. J. Bacteriol. *116*, 847–856.

Marshall, J.M., and Gunn, J.S. (2015). The O-Antigen Capsule of Salmonella enterica Serovar Typhimurium Facilitates Serum Resistance and Surface Expression of FliC. Infect. Immun. *83*, 3946–3959.

Mavris, M., Manning, P.A., and Morona, R. (1997). Mechanism of bacteriophage SfII-mediated serotype conversion in Shigella flexneri. Mol. Microbiol. *26*, 939–950.

Micoli, F., Ravenscroft, N., Cescutti, P., Stefanetti, G., Londero, S., Rondini, S., and Maclennan, C.A. (2014). Structural analysis of O-polysaccharide chains extracted from different Salmonella Typhimurium strains. Carbohydr. Res. *385*, 1–8.

Miller, J.H. (1972). Experiments in molecular genetics (Cold Spring Harbor Laboratory).

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., et al. (2014). The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. gku1243.

Nair, A., Korres, H., and Verma, N.K. (2011). Topological characterisation and identification of critical domains within glucosyltransferase IV (GtrIV) of Shigella flexneri. BMC Biochem. *12*, 67.

Papadopoulos, J.S., and Agarwala, R. (2007a). COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics *23*, 1073–1079.

Papadopoulos, J.S., and Agarwala, R. (2007b). COBALT: constraint-based alignment tool for multiple protein sequences. Bioinforma. Oxf. Engl. *23*, 1073–1079.

Reeves, P. (1993). Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale. Trends Genet. *9*, 17–22.

Reeves, P.R., Cunneen, M.M., Liu, B., and Wang, L. (2013). Genetics and Evolution of the Salmonella Galactose-Initiated Set of O Antigens. PLoS ONE *8*, e69306.

Richards, J., Sundermeier, T., Svetlanov, A., and Karzai, A.W. (2008). Quality Control of Bacterial mRNA Decoding and Decay. Biochim. Biophys. Acta *1779*, 574–582.

Rivera-Chavez, F., and Baeumler, A.J. (2015). The Pyromaniac Inside You: Salmonella Metabolism in the Host Gut. In Annual Review of Microbiology, Vol 69, S. Gottesman, ed. (Palo Alto: Annual Reviews), pp. 31–48.

Salzman, A.L., Eaves-Pyles, T., Linn, S.C., Denenberg, A.G., and Szabó, C. (1998). Bacterial induction of inducible nitric oxide synthase in cultured human intestinal epithelial cells. Gastroenterology *114*, 93–102.

Santos, R.L., Raffatellu, M., Bevins, C.L., Adams, L.G., Tükel, Ç., Tsolis, R.M., and Bäumler, A.J. (2009). Life in the inflamed intestine, Salmonella style. Trends Microbiol. *17*, 498–506.

Silhavy, T.J., Kahne, D., and Walker, S. (2010). The Bacterial Cell Envelope. Cold Spring Harb. Perspect. Biol. *2*.

Slauch, J.M., Lee, A.A., Mahan, M.J., and 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. J. Bacteriol. *178*, 5904–5909.

Snyder, D.S., Gibson, D., Heiss, C., Kay, W., and Azadi, P. (2006). Structure of a capsular polysaccharide isolated from Salmonella enteritidis. Carbohydr. Res. *341*, 2388–2397.

Steenackers, H., Hermans, K., Vanderleyden, J., and De Keersmaecker, S.C.J. (2012). Salmonella biofilms: An overview on occurrence, structure, regulation and eradication. Food Res. Int. *45*, 502–531.

Sun, Q., Knirel, Y.A., Wang, J., Luo, X., Senchenkova, S.N., Lan, R., Shashkov, A.S., and Xu, J. (2014). Serotype-converting bacteriophage SfII encodes an acyltransferase protein that mediates 6-O-acetylation of GlcNAc in Shigella flexneri O-antigens, conferring on the host a novel O-antigen epitope. J. Bacteriol. *196*, 3656–3666.

Thiennimitr, P., Winter, S.E., Winter, M.G., Xavier, M.N., Tolstikov, V., Huseby, D.L., Sterzenbach, T., Tsolis, R.M., Roth, J.R., and Bäumler, A.J. (2011). Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota. Proc. Natl. Acad. Sci. *108*, 17480–17485.

Torres-Cruz, J., and van der Woude, M.W. (2003). Slipped-Strand Mispairing Can Function as a Phase Variation Mechanism in Escherichia coli. J. Bacteriol. *185*, 6990–6994.

Veses-Garcia, M., Liu, X., Rigden, D.J., Kenny, J.G., McCarthy, A.J., and Allison, H.E. (2015). Transcriptomic Analysis of Shiga-Toxigenic Bacteriophage Carriage Reveals a Profound Regulatory Effect on Acid Resistance in Escherichia coli. Appl. Environ. Microbiol. *81*, 8118–8125.

Villafane, R., Zayas, M., Gilcrease, E.B., Kropinski, A.M., and Casjens, S.R. (2008). Genomic analysis of bacteriophage epsilon 34 of Salmonella enterica serovar Anatum (15+). BMC Microbiol. *8*, 227.

Vollmer, W., Blanot, D., and Pedro, M.A.D. (2008). Peptidoglycan structure and architecture. FEMS Microbiol. Rev. *32*, 149–167.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. Theor. Den Biowissenschaften *131*, 281–285.

Warren, J.W., Walker, J.R., Roth, J.R., and Altman, E. (2000). Construction and Characterization of a Highly Regulable Expression Vector, pLAC11, and Its Multipurpose Derivatives, pLAC22 and pLAC33. Plasmid *44*, 138–151.

Weinbaum, G., Kadis, S., and Ajl, S.J. (2013). Bacterial Endotoxins: A Comprehensive Treatise (Elsevier).

Wilkins, W., Rajić, A., Waldner, C., McFall, M., Chow, E., Muckle, A., and Rosengren, L. (2010). Distribution of Salmonella serovars in breeding, nursery, and grow-to-finish pigs, and risk factors for shedding in ten farrow-to-finish swine farms in Alberta and Saskatchewan. Can. J. Vet. Res. Rev. Can. Rech. Vét. *74*, 81–90.

Woude, M.W. van der, and Bäumler, A.J. (2004). Phase and Antigenic Variation in Bacteria. Clin. Microbiol. Rev. *17*, 581–611.

van der Woude, M.W. (2011). Phase variation: how to create and coordinate population diversity. Curr. Opin. Microbiol. *14*, 205–211.

Wright, A. (1971). Mechanism of conversion of the Salmonella O antigen by bacteriophageepsilon 34. J. Bacteriol. *105*, 927–936.

Yamamoto, S., and Kutsukake, K. (2006). FljA-Mediated Posttranscriptional Control of Phase 1 Flagellin Expression in Flagellar Phase Variation of Salmonella enterica Serovar Typhimurium. J. Bacteriol. *188*, 958–967.

Young, B.G., Fukazawa, Y., and Hartman, P.E. (1964). A P22 bacteriophage mutant defective in antigen conversion. Virology *23*, 279–283.