Structure function analysis of O-antigen modifying enzymes in the bacterial pathogen *Salmonella*

Caroline Rose Pearson

PhD

University of York

Biology

September 2019

Abstract

Carbohydrates are essential for bacterial survival. As well as being taken up and used as a carbon source, they are synthesised by bacteria to protect them from environmental stresses and assist in colonisation of new environments. Bacteria possess a range of mechanisms to modify the diverse carbohydrate structures that they synthesise, a key example being O-acylation. Acylation of bacterial carbohydrates can have applications in processes such as antigenic variation, osmoregulation, virulence and cell division, and can also have transferrable applications in clinical and industrial processes.

The aim of this research was to investigate the mechanism of carbohydrate O-acylation by Acyltransferase_3 (AT3) domain-containing proteins OafA and OafB of the bacterial pathogen *Salmonella*. These proteins contain a membrane bound AT3 domain fused to an extra cytoplasmic SGNH domain. They O-acetylate different residues in the variable repeating carbohydrate O-antigen of lipopolysaccharide.

The research presented in this thesis demonstrates that both the AT3 and SGNH domains are required for O-antigen acetylation in these proteins. It also highlights functional residues that support a conserved mechanism of transmembrane acyl group transport by AT3 domain-containing proteins, and suggests residues within the AT3 domain specifically adapted for function with a fused SGHN domain. Furthermore, structural and functional characterisation of the SGNH domain of OafA and OafB supports the hypothesis that this domain is responsible for the final step of acetyl group transport to the carbohydrate acceptor in this system, and also identifies how acceptor substrate specificity is achieved by this domain.

These findings have allowed a refined mechanistic model for AT3-SGNH fused proteins to be defined, enhancing our understanding of an important family of acyltransferase proteins which are found across all domains of life. This work therefore provides a framework for understanding and potentially manipulating these proteins, to enable carbohydrates of clinical and industrial significance to be engineered.

List of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, thank you to the BBSRC for funding my research project.

I would like to thank my supervisors Marjan van der Woude and Gavin Thomas for their invaluable help, advice and support during my PhD. Also, a huge thank you to all the members of the van der Woude and Thomas labs for their endless encouragement, baked goods contributions and camaraderie that kept me going through challenging times.

Thank you to my thesis advisory panel Daniela Barilla and Michael Plevin for their advice and the interesting discussions during TAP meetings that inspired new approaches in my research. To Reyme Herman, for always being happy to advise on experiments, and to Sarah Tindall, for solving the crystal structure which hugely helped with the progression of both of our research projects.

Thank you to Rebecca Hall and Stephen Thorpe for tearing apart my presentations and building them into something less “borderline incomprehensible”.

Thank you to the members of the Mottram lab, past and present, for adopting me into their group and making me feel at home in H block. With particular thanks to; Andreas Damianou, Chris Bower-Lepts, Elaine Brown, Manuel Saldivia, Nathaniel Jones, Nicola Baker, Rachel Neish, Rebecca Burge, Ridda Jabbar, and Vincent Geoghegan.

Thank you to all the members of the PhD community at the University of York who have provided an invaluable support network to see me through my PhD. With particular thanks to Aritha Dornau, Emma Stewart, Helen Davies, Jack Munns (and his tiny little worms), James Robson, Lotte van Beek, Nathaniel Holman, Sarah Rixham and Sophie Rugg.

Finally, I would like to thank my friends and family that definitely do not understand what I do but support me regardless. Especially to my parents, my sister and to my partner Ben, your support and encouragement has inspired me to keep going when I couldn’t see the light at the end of the tunnel.

Declaration

I 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.

This work has contributed to production of a joint first author manuscript with Sarah Tindall (University of York) which is submitted for publication (Pearson *et al.*, 2020). Work contributed by Sarah Tindall has been acknowledged and is unpublished at the time of thesis submission.

Creation of the *Salmonella* strain encoding chromosomal C-terminally His tagged OafA was carried out by Steinar Mannsverk under my supervision. Work carried out by the Technology Facility (Department of Biology, University of York) is indicated and NMR spectrometry was carried out by Pedro Aguiar in the centre for magnetic resonance at the University of York.

Chapter 1 Introduction

1.1 Bacterial cell surfaces

Bacteria inhabit a diverse range of habitats across aquatic, terrestrial and even aerial environments and can survive independently or within host organisms. In fact, there are few places on earth that bacteria have not been found to thrive. Within these environments, bacteria play a vast array of roles. Within host organisms, some are critical in providing essential nutrients to their host in exchange for their maintenance within the organism (Webster, 2014), whereas, others cause disease and even death of the host (Wilson *et al.*, 2002). In all cases, bacteria must adapt to survive in the diverse environments that they inhabit and their first contact with these environments is often through their cell surfaces.

This introduction outlines the carbohydrate structures at bacterial cell surfaces and the significance of their roles in bacterial physiology and survival. Modification of these carbohydrates can have diverse impacts on bacterial interactions. A key example being the modification of surface antigens in *Salmonella*. Many bacteria share a conserved approach to acylation of the carbohydrates that they synthesise and the focus of this thesis is to enhance the knowledge on the mechanism of action of the proteins involved in these specific carbohydrate O-acylation modifications.

1.1.1 Gram-positive and Gram-negative bacterial cell envelopes

The composition of the bacterial cell envelope is fundamental in defining bacterial cell shape (Cabeen and Jacobs-Wagner, 2005), resistance to environmental stresses (Yadav *et al.*, 2018), and interaction with the environment (Nikaido, 1999). Bacteria are divided into two broad classes, defined by the composition of their cell envelope **Figure 1.1**.

Gram-positive bacteria possess a single phospholipid bilayer surrounded by a thick layer of peptidoglycan - an integral polysaccharide component of the cell envelope that will be discussed in further detail in section 1.1.4. This thick peptidoglycan layer, often referred to as the cell wall, is interspersed with proteins and secondary cell wall glycopolymers called teichoic acids (Silhavy *et al.*, 2010). Proteins covalently attached to the peptidoglycan (CAP - **Figure 1.1**) display vast functional diversity in bacterial

interactions with host tissues, immune system components and other bacteria, and they can also play roles in nutrient acquisition and protein processing (Navarre and Schneewind, 1999). Teichoic acids are anionic carbohydrate containing polymers which can be anchored in the cell membrane by glycolipids (LTA - **Figure 1.1**) or covalently attached to peptidoglycan (WTA - **Figure 1.1**) (Brown *et al.*, 2013; Percy and Gründling, 2014). Teichoic acids have been demonstrated to contribute to regulation of autolysin activity and are therefore involved in growth and division of bacterial cells (Höltje and Tomasz, 1975).

The Gram-negative bacterial cell envelope, in contrast, is composed of three principal layers; two lipid bilayers which are separated by the periplasm (**Figure 1.1**). The periplasm is a multipurpose compartment which contains a thinner layer of peptidoglycan when compared to Gram-positive bacteria, as well as lipid anchored and many different soluble proteins (Miller and Salama, 2018). The inner and outer membranes of Gram-negative bacteria are also distinct from one another, the inner membrane is composed of a bilayer of phospholipids of varying composition (Sohlenkamp and Geiger, 2016), whereas the outer membrane is an asymmetric lipid bilayer, with the inner leaflet composed of phospholipids and the outer leaflet composed of lipopolysaccharide (Le Brun *et al.*, 2013; Abellón-Ruiz *et al.*, 2017). Lipopolysaccharide (LPS) is another important polysaccharide component of the cell envelope and will be discussed in further detail in section 1.1.6.

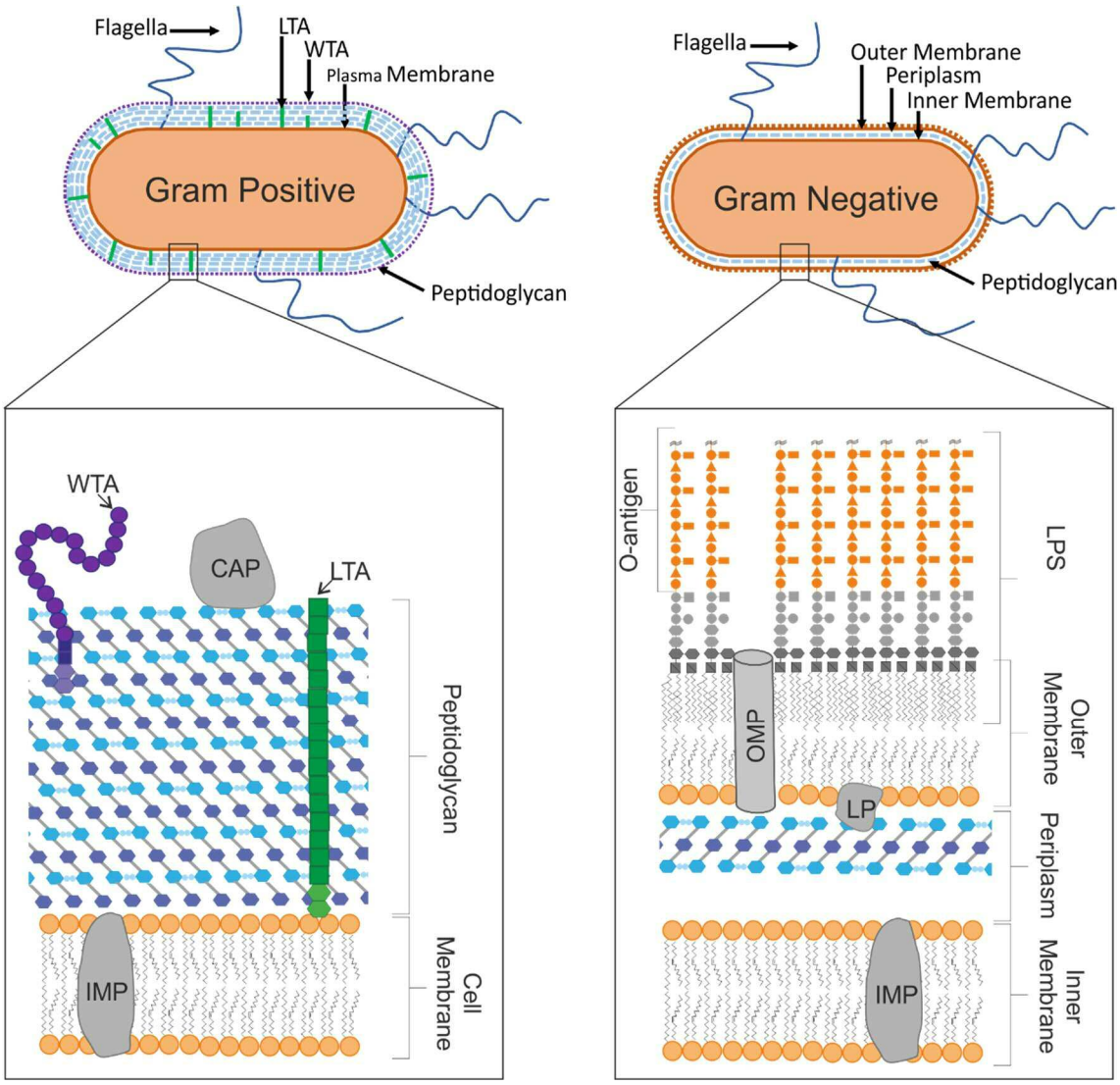


Figure 1.1| Schematic representation of Gram-positive and Gram-negative bacterial cell envelopes. CAP = covalently attached protein, IMP = integral membrane protein, LPS= Lipopolysaccharide, LP= Lipoporotein, LTA = Lipoteichoic acid, OMP = outer membrane protein, WTA = wall teichoic acid. Adapted from (Silhavy *et al.*, 2010).

1.1.2 The role of polysaccharides in the bacterial cell envelope

As highlighted in the previous section, bacterial cell envelopes are associated with a variety of polysaccharides which are integral to the survival and morphology of prokaryotic organisms. Discussed below are some of the key polysaccharides of the bacterial cell envelope and their role in cellular integrity.

1.1.3 Capsular polysaccharide

Immediately exterior to the components of the cell envelope, discussed in 1.1.1, Gram-positive and Gram-negative bacteria can also produce a polysaccharide capsule, defined

as a gel-like layer of polysaccharide that can be up to 10 µm thick (Davis, 1996). These capsular polysaccharides are often negatively charged and display vast structural diversity in their monosaccharide composition, branching and linkages (Wen and Zhang, 2014). They can be linked to the bacterial cell surface with a lipid moiety or covalently anchored to the peptidoglycan of Gram-positive bacteria (Sørensen *et al.*, 1990; Wen and Zhang, 2014).

Bacterial capsules can be major virulence determinants as they are antiphagocytic, due to mutual repulsion by the net negative charge, and can prevent complement mediated killing, through binding regulators of the complement pathway or masking underlying cellular components that could activate the complement cascade (Moxon and Kroll, 1990). Capsular polysaccharides can also be important for survival in, and colonisation of, new environments. These structures can promote adherence to innate surfaces and other cells for formation of biofilms (Roberts, 1996), they are important in resistance to desiccation and disinfectants (Tipton *et al.*, 2018), and they have also been implicated in escape of mucus-mediated clearance from the host (Nelson *et al.*, 2007).

Due to the many and varied roles that capsular polysaccharides play in survival of pathogenic bacteria, and the fact that these structures are the most exterior component of the bacterial envelope, they have been targeted in vaccine research. Indeed, capsular polysaccharides have been successfully targeted in vaccines against *Streptococcus pneumoniae* (Berical *et al.*, 2016), *Neisseria meningitidis* (Peltola *et al.*, 1977), *Salmonella enterica* serovar Typhi (Thiem *et al.*, 2011) and *Haemophilus influenzae* (Yogev *et al.*, 1990).

1.1.4 Peptidoglycan

Peptidoglycan, also referred to as murein, is a polymer of monosaccharide chains linked by amino acids that form a mesh like structure around the cell that is an important feature in maintaining cell shape and structural integrity, and can play roles in bacteriophage interactions (Höltje, 1998; Gaidelyte *et al.*, 2006). As mentioned in section 1.1.1, peptidoglycan is much more abundant in the Gram-positive bacterial cell envelope, accounting for 30-70% of the total cell envelope and measuring 15-30 nm (Schleifer and Kandler, 1972; Vollmer *et al.*, 2008). Whereas, there is comparatively little peptidoglycan in Gram-negative bacteria, which accounts for less than 10% of the total

cell envelope and is equivalent to just one to three layers of peptidoglycan chains. Peptidoglycan chains, or glycan strands, are composed of vast interlocking chains of peptidoglycan monomers, composed of a disaccharide repeat of N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) with a peptide stem. The glycan strand is elongated by linkage of disaccharide repeats and crosslinking of the strands occurs by transpeptidase reaction, which links the peptide stems between MurNac molecules (**Figure 1.2**)(Höltje, 1998).

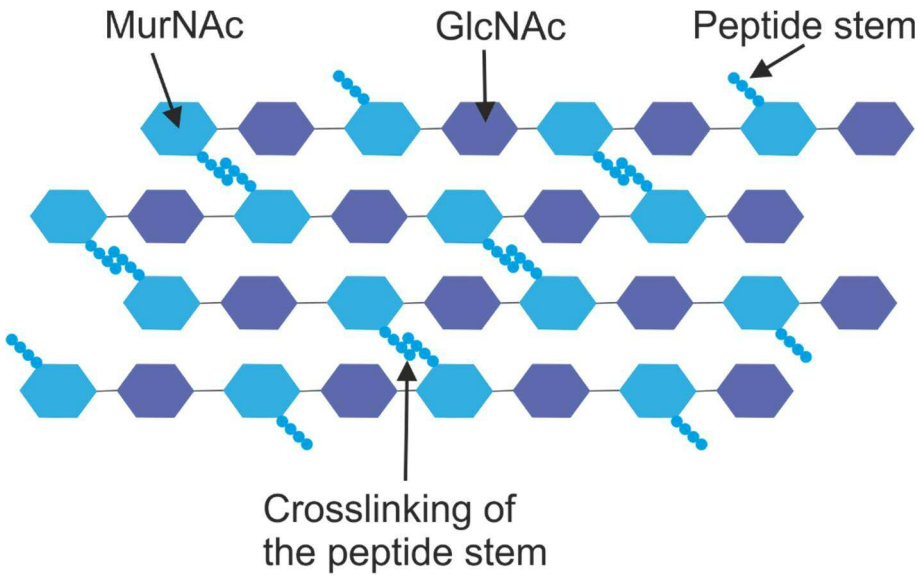


Figure 1.2| Composition of peptidoglycan in the cell envelope. N-acetylmuramic acid (MurNAc) residues are represented in blue and N-acetylglucosamine (GlcNAc) residues are represented in purple.

The carbohydrate composition of glycan strands is regarded as uniform across many species studied and it is the peptide stem which is the most variable component of peptidoglycan between species (Vollmer *et al.*, 2008). However, it has long since been established that glycan strand carbohydrates can be modified by addition of O-acetyl substituents (Schleifer and Kandler, 1972). These carbohydrate alterations can have impact on recognition of bacteria by the host as well as resistance of peptidoglycan to hydrolytic enzymes (Vollmer, 2008).

1.1.5 Osmoregulated periplasmic glucans

The periplasmic space of Gram-negative bacteria is also resident to structurally varied oligosaccharides composed of D-glucose. These glucans are suggested to play a role in providing resilience of bacterial cells against osmotic stress (Bohin, 2000), and can be

important for virulence of Gram-negative bacterial pathogens (Page *et al.*, 2001; Bhagwat *et al.*, 2009). Gram-negative bacteria inhabit diverse environments with varying osmotic pressures, for example, they can survive in pure water, human blood and sea water (Wood, 2015). Synthesis of osmoregulated periplasmic glucans (OPG – previously termed membrane derived oligosaccharides) were found to be regulated by osmotic pressure in a range of organisms (Kennedy, 1982; Miller *et al.*, 1986; Breedveld and Miller, 1994). Although, only composed of D-glucose, OPG's display structural diversity through the linkage of β-glycosidic bonds, and they can also be decorated with acetyl, succinyl and methylmalonyl substituents, resulting in the anionic properties of these polysaccharides (Bohin, 2000). It has been proposed that the anionic properties of these periplasmic glucans helps to maintain the high osmotic pressure of the periplasmic space (Kennedy, 1982). However, not all OPG's are anionically charged and alternative roles in closing porin channels (Delcour *et al.*, 1992), detergent resistance (Rajagopal *et al.*, 2003), and regulating environmental sensing proteins (Bontemps-Gallo *et al.*, 2013) have been suggested.

1.1.6 Lipopolysaccharide

Lipopolysaccharide (LPS) is the major lipid bound carbohydrate component of the outer leaflet of the outer membrane of Gram-negative bacteria (**Figure 1.1**). LPS is considered an essential component of the cell envelope of many Gram-negative bacteria as it plays roles in the correct assembly and function of outer membrane proteins and prevents the passage of toxic hydrophobic molecules into the cell (Zhang *et al.*, 2013; Klein and Raina, 2019). LPS is composed of a membrane anchored lipid A, a core oligosaccharide portion, and a distal O-antigen (**Figure 1.3**)(Osborn *et al.*, 1964; Miller *et al.*, 2005). Its lipid A portion is a key activator of innate immunity (Raetz and Whitfield, 2002; Miller *et al.*, 2005). The core oligosaccharides connect lipid A to the O-antigen and are generally well conserved in composition between related bacterial species, particularly in the inner core region (Raetz and Whitfield, 2002). The 3-deoxy-d-*manno*-2-octulosonic acid (KDO) sugar is a mostly invariable constituent of the LPS core and is only replaced with D-glycero-D-talo-oct-2-ulosonic acid (Ko) in some organisms (Holst, 2011). It can therefore be used to detect LPS presence in colourimetric and biophysical assays (Unger, 1981; Rybka and Gamian, 2006).

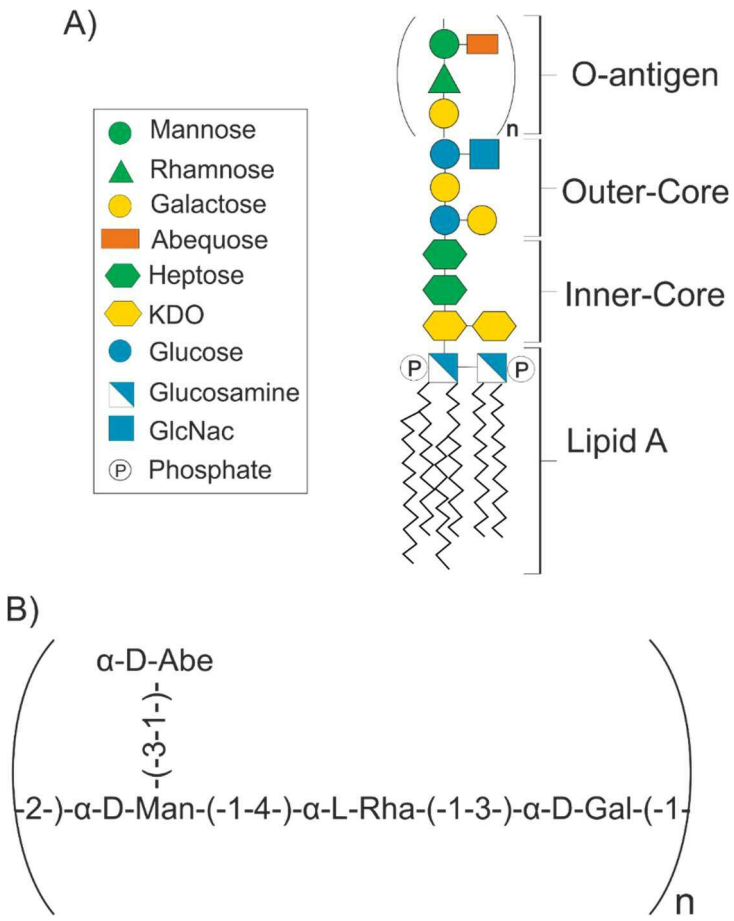


Figure 1.3| Structure of lipopolysaccharide. A) Schematic representation of *Salmonella enterica* ser. Typhimurium LPS. Glycans are represented according to symbol nomenclature for graphical representation of glycans (Hart *et al.*, 2015). B) Linkage of monosaccharides in the O-antigen repeat unit of *Salmonella enterica* ser. Typhimurium LPS.

Some organisms which reside in mucosal surfaces, such as *N. gonorrhoeae* and *H. influenzae* possess LPS that does not contain the distal repeating O-antigen and instead has a single non-repeating oligosaccharide chain (Jacques, 1996; Preston *et al.*, 1996). This alternative glycolipid is termed lipooligosaccharide (LOS). Although it does not possess the highly variable O-antigen structure, LOS is still a major virulence factor in these bacteria (Kimura *et al.*, 1987; Brandtzaeg *et al.*, 2004), and its oligosaccharide composition can vary depending on expression of biosynthesis glycosyltransferase genes (Zhu *et al.*, 2001; Patrone and Stein, 2007) .

The O-antigen, when present, is the most distal and variable component of LPS and it plays a pivotal role in bacterial interactions, survival, and pathogenicity. The O-antigen is composed of repeating oligosaccharide units composed of 3-6 monosaccharides, which can be linear or branched (Lerouge, 2001). The number of O-antigen repeat units on each LPS molecule can vary, giving rise to the characteristic ladder pattern obtained

after visualisation of LPS that has been run on SDS-PAGE gel (**Figure 1.4**)(Rezania *et al.*, 2011).

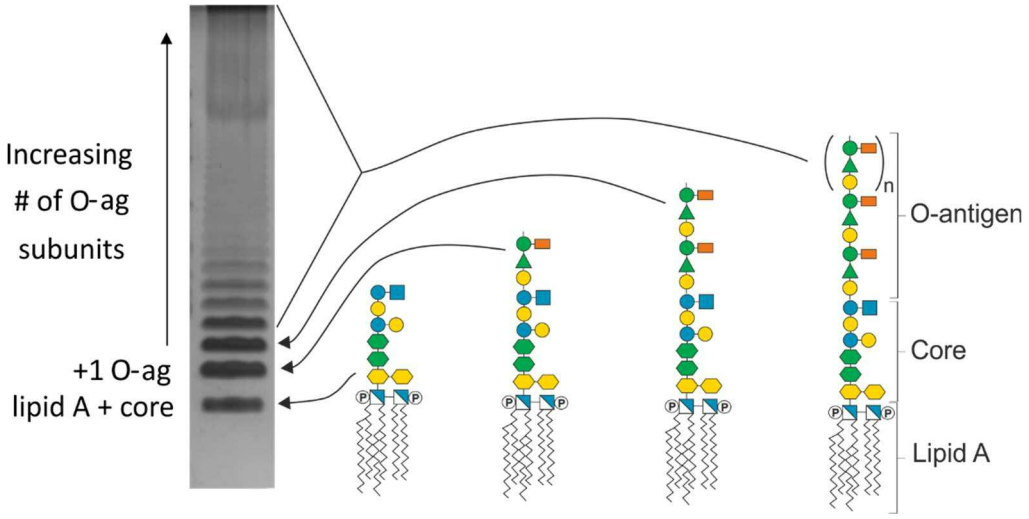


Figure 1.4| LPS banding pattern on SDS-PAGE gel. *Salmonella enterica* ser. Typhimurium LPS run on tricine SDS-PAGE and visualised by silver staining shows the typical laddering pattern of LPS after size separation. Each rung on the ladder represents LPS with an additional O-antigen repeat.

O-antigens can be up to 100 units long in some cases (Stevenson *et al.*, 1995). The length of the LPS O-antigen is important for many interactions with host immune defences. Very long (>90 units) LPS is implicated in evading complement mediated killing through steric hindrance of complement (Murray *et al.*, 2006). Conversely, shorter O-antigen lengths are required to allow invasion of host cells (Hölzer *et al.*, 2009) and interaction of cell surface proteins for actin based motility of intracellular pathogens (Morona *et al.*, 2003). Therefore, the architecture of LPS on the outer membrane of Gram-negative bacteria must be intricately balanced, to facilitate survival inside a host.

As the most exposed molecules to host defences, O-antigens can be highly immunogenic and they can also be the target of bacteriophage interactions to initiate infection (Broeker and Barbirz, 2017). For this reason, bacteria possess mechanisms to alter their O-antigen structure beyond the carbohydrate composition of the repeating unit. A diverse range of additional modifications can be performed which include, altering the linkage of sugars (Wang *et al.*, 2002) and adding extra moieties such as glucose (Davies *et al.*, 2013) or acetyl groups (Slauch *et al.*, 1996). O-antigen diversity is of particular significance in *Salmonella* biology as it is a determinant in the classification of the >2600 serovars that have currently been characterised (Grimont and Weill, 2008; Issenhuth-Jeanjean *et al.*, 2014) and has been linked with variation in the production of protective

antibodies and altered bacterial virulence. *Salmonella* O-antigen diversity will be discussed in further detail in the following section.

1.2 Salmonella and the LPS O-antigen

1.2.1 Overview of *Salmonella*

Salmonella are Gram-negative bacteria of the Enterobacteriaceae family which also includes other pathogenic examples such as *Escherichia coli* and *Shigella flexneri*. The *Salmonella* genus contains two species, *Salmonella enterica* and *Salmonella bongori* (Grimont and Weill, 2008). *S. enterica* is most commonly associated with infection of warm-blooded animals, whereas *S. bongori* is mostly associated with the diseases that occur within cold-blooded animals, with limited incidence of human infection (Fookes *et al.*, 2011). *S. enterica* is a major cause of foodborne illness in humans and is the most commonly identified pathogens from food in the USA after *Campylobacter* (Tack *et al.*, 2019).

S. enterica is further divided into six subspecies which have been subdivided into over 2600 serovars by the White-Kauffmann-Le Minor scheme (further discussed in section 1.2.2), based on antigenic properties of their flagellar H-antigen and lipopolysaccharide (LPS) O-antigen (**Figure 1.5**) (Grimont and Weill, 2008; Issenhuth-Jeanjean *et al.*, 2014). *S. enterica* subspecies *enterica* is the most clinically relevant subspecies and is the dominant subspecies in causing human *Salmonella* infections (salmonellosis) (Fierer and Guiney, 2001). In the US, there are over 10,000 cases annually of these infections (Lynch *et al.*, 2009; Boore *et al.*, 2015). Thus, this subspecies will be the focus of further discussion.

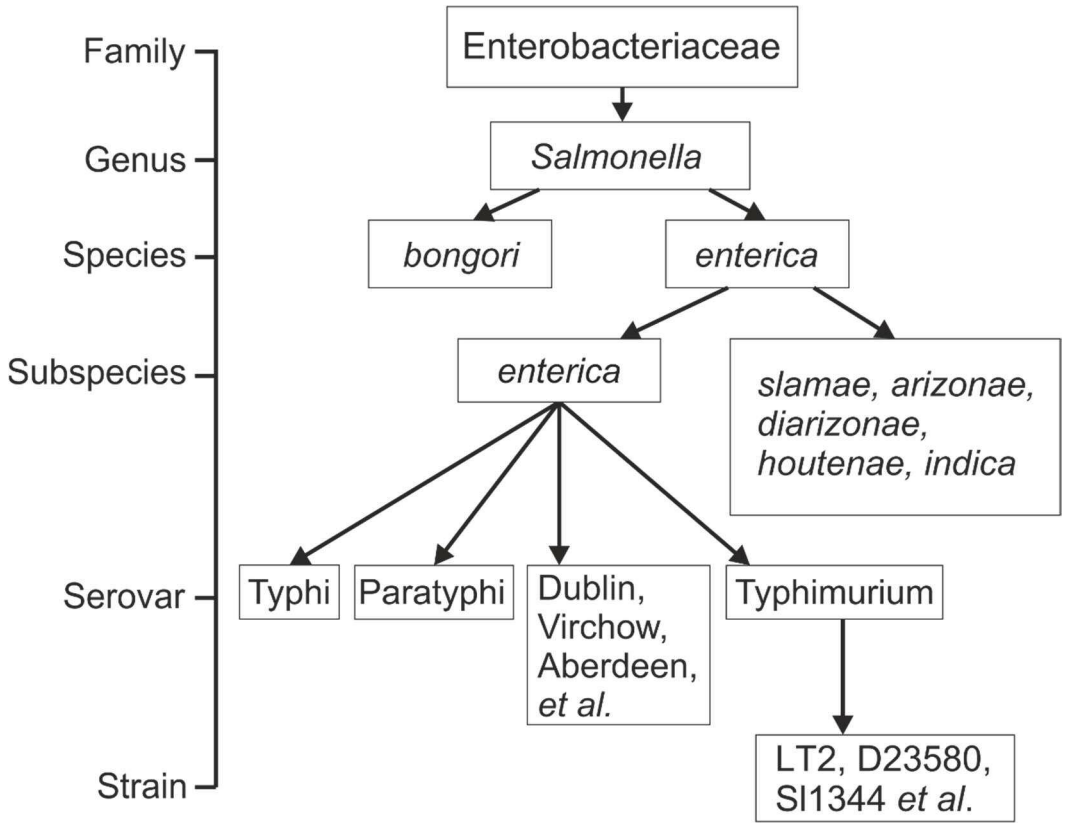


Figure 1.5| Divisions of the *Salmonella* genus.

S. enterica subspecies *enterica* infections display a wide variety of clinical manifestations, from asymptomatic disease to gastroenteritis and systemic infections, and the severity of these infections can be dependent on the serotype of *Salmonella* and the health status of the human host (Jones *et al.*, 2008; Eng *et al.*, 2015). Invasion of eukaryotic cells and intracellular survival and replication are essential requirements in the infection cycle of *S. enterica*. Following ingestion and survival of the acidic stomach environment and host defences in the small intestine (Álvarez-Ordóñez *et al.*, 2011, 2012), cells breach the gut epithelium using a variety of mechanisms depending on the cells that they encounter (Garai *et al.*, 2012). *Salmonella* can be phagocytosed by specialist phagocytic cells or they can actively induce their uptake using a Type III secretion system, a 'molecular syringe' which can inject effector proteins into epithelial cells to induce cytoskeletal rearrangements to promote their uptake into a *Salmonella* containing vacuole (SCV) (Haraga *et al.*, 2008). Once inside the host, *Salmonella* cells maintain themselves in the SCV through use of a second Type III secretion system to inject effectors into the host cell which maintain the conditions of the SCV (Beuzón *et al.*, 2000).

Self-limiting gastroenteritis is the prominent clinical symptom of non-typhoidal *Salmonella,* which is caused by serovars such as Typhimurium, Enteritidis and Newport (Tack *et al.*, 2019). These serovars tend to infect a broad host range and maintain reservoirs in livestock and other animals, with asymptomatic carriage in some cases (Cheong *et al.*, 2007; Díaz-Sánchez *et al.*, 2013; Kagambèga *et al.*, 2013). Non-typhoidal *Salmonella* is a considerable public health burden in both developing and developed countries, with serovar Typhimurium estimated to cause more than 150,000 human deaths from gastroenteritis each year (Majowicz *et al.*, 2010; Hiyoshi *et al.*, 2018). In some cases, non-typhoidal serovars can invade the bloodstream and cause bacteraemia which results in high fever and can cause septic shock (Eng *et al.*, 2015). In these cases the infecting strains are termed invasive non-typhoidal *Salmonella* (iNTS). A sub-lineage of the Typhimurium serovar is the dominant cause of invasive non-typhoidal *Salmonella* infections in Africa (Reddy *et al.*, 2010). These infections are a significant cause of morbidity and mortality, with the very young, elderly and immunocompromised individuals at particular risk (Uche *et al.*, 2017). There is also increasing evidence that iNTS infections are becoming more prevalent in developed countries with 17.3 cases per 100,000 population recorded in Queensland Australia, caused prominently by serovars Virchow, Typhimurium and Aberdeen (Parisi *et al.*, 2019). The significance of non-typhoidal *Salmonella* in human disease and livestock infection make it a substantial health and economic problem.

In contrast to non-typhoidal serovars, *Salmonella enterica* serovar Typhi and Paratyphi can result in enteric (Typhoid or Paratyphoid) fever, where, following initial invasion of intestinal epithelial cells, they disseminate to the lymph nodes, liver and spleen (de Jong *et al.*, 2012). The clinical presentation of typhoid and paratyphoid fever is distinct from non-typhoidal *Salmonella* serovars, with Typhoidal strains moderating intestinal inflammation upon infection, to prevent the acute gastroenteritis which is associated with non-typhoidal *Salmonella* strains and allow dissemination to other sites in the body (Winter *et al.*, 2010; Hiyoshi *et al.*, 2018). Typhoidal *Salmonella* strains are highly host restricted and spread from human to human by the faecal-oral route. As, these pathogens can disseminate to immune privileged niches within the body such as the gallbladder (Hiyoshi *et al.*, 2018), they can persist within apparently healthy hosts for extended periods of time and be shed from the body over a longer duration, promoting their maintenance within a population (Kingsley and Baumler, 2000). *Salmonella* ser.

Typhi is a major cause of enteric fever, in 2017 there were over 14 million cases of Typhoidal *Salmonella* globally, resulting in over 135,000 deaths, with 76.3% of these cases caused by serovar Typhi (Stanaway *et al.*, 2019).

1.2.2 Classification of *Salmonella* serovars

As mentioned in section 1.2.1, *Salmonella* serovars are classified using the White-Kauffmann-Le Minor scheme. This scheme was first published in 1934 and listed 44 *Salmonella* serovars (Salmonella Subcommittee of the Nomenclature Committee of the International Society for Microbiology, 1934; Grimont and Weill, 2008). In its most recent published update this number had risen to over 2600 classified serovars, with the majority of these (1586) arising from subspecies enterica (Issenhuth-Jeanjean *et al.*, 2014). This scheme is important for surveillance of *Salmonella* and provides a consistent method to determine the prevalence of different serovars across the globe. It characterises *Salmonella* isolates based on the antigenic properties of their surface antigens, and in recent years phage typing has been employed to further differentiate frequently isolated serovars such as Typhimurium and Enteritidis (Rabsch, 2007; Grimont and Weill, 2008). Phage typing involves subjecting *salmonella* isolates to infection with specific bacteriophage to record which phage they are susceptible to lysis with, however, its usefulness as a surveillance technique has been questioned (Baggesen *et al.*, 2010).

Salmonella surface antigens used for classification are the H-antigen, Vi-antigen and O-antigen and designation of the H, O and Vi antigens of *Salmonella* isolates is determined by agglutination reactions with specific antibodies. The H-antigen is determined by the structure of the flagellin protein which makes up flagellum filaments required for bacterial motility. By expression of different flagellin genes different *Salmonella* isolates can produce over 100 different H-antigens (McQuiston *et al.*, 2008). The Vi antigen denotes strains which express a polysaccharide capsule and much like many other encapsulated bacteria (section 1.1.3), possession of the genes encoding the Vi capsular antigen are thought to be connected with virulence (Seth-Smith, 2008). The O-antigen serotype is determined by the structure of the LPS O-antigen polysaccharide (see section 1.1.6) (**Figure 1.3**). There are currently 67 different O-antigen structures recognised by serotyping antibodies (Grimont and Weill, 2008), however incidence in which expression

of O-antigen modification genes does not alter the O-antigen serotype have been recorded (Davies *et al.*, 2013). This highlights the diversity of O-antigen structural modifications that are overlooked by current standard serotyping procedures. Therefore, understanding the modifications of these O-antigen structures could have significance in understanding *Salmonella* virulence, and epidemiology.

1.2.3 *Salmonella* LPS O-antigen biosynthesis

The general components of LPS are discussed in section 1.1.6 and the enzymes involved in synthesis of LPS and delivery to the cell surface have been well characterised (Simpson *et al.*, 2015; Maldonado *et al.*, 2016; Okuda *et al.*, 2016; Owens *et al.*, 2019). Briefly, the lipid A and core components of LPS are synthesised on the cytoplasmic side of the inner membrane through sequential activity of conserved enzymes which first produce the lipid A molecule then consecutively add each core sugar through activity of defined glucosyltransferase enzymes. The lipid A + core is then flipped to the periplasmic side of the inner membrane by the MsbA ABC transporter. The O-antigen is synthesised separately and the pathway for O-antigen synthesis will be outlined in more detail in the following paragraphs. Once synthesised, the polymerised O-antigen and is added to the Lipid A + core by the O-antigen ligase protein WaaL. This forms the completed LPS molecule on the periplasmic side of the inner membrane. The fully synthesised LPS molecule is then transported to the outer membrane through action of 7 proteins which form the lipopolysaccharide transport (LPT) complex (LptABCDEFG) (**Figure 1.6**).

There have been significant developments in the understanding of the mechanism of LPS transport in recent years, initiated by the development of *in vivo* photo-crosslinking stratergies to trap LPS on Lpt proteins and distinguish their interactions (Okuda *et al.*, 2012). Subsequent resolution of the crystal structure of multiple LPT complex proteins has allowed a detailed model of the mechanism of LPS transport to the outer membrane to be outlined (Sherman *et al.*, 2014; Luo *et al.*, 2017; Li *et al.*, 2019; Owens *et al.*, 2019). Briefly, LPS present in the inner membrane inserts into the LPT complex through the entry point formed by the transmembrane proteins LptG and LptF and the single transmembrane helix of LptC. ATP hydrolysis by the two LptB proteins, interacting at the cytoplasmic face of the inner membrane complex, then provide the energy to move LPS molecules along the LPT complex onto the chain of LptA proteins that span the

periplasm. LptD and LptE constitute the outer membrane LPS translocon and once LPS has been passed from the LptA chain to this complex they move LPS to the outer leaflet of the outer membrane (Sperandeo *et al.*, 2017; Owens *et al.*, 2019)(**Figure 1.6**).

There are three known pathways for O-antigen biosynthesis in Gram-negative bacteria; Wzy dependent, ABC transporter dependent, and synthase dependent, and these are reviewed in depth by Kalynych *et al,.* (2014). In all cases, the composition of the O-antigen repeat unit is determined by the organisation of the O-antigen biosynthesis gene cluster. O-antigen biosynthesis genes first encode the proteins involved in synthesis of nucleotide sugar precursors of the O-antigen, however, if the monosaccharide components of the O-antigen are 'housekeeping' sugars, already encoded by bacterial genomes, these are not usually duplicated (Reeves, 1994). Following the nucleotide sugar biosynthesis genes, glycosyltransferase proteins encoded in the O-antigen biosynthesis gene cluster then determine the sequential transfer of these various precursor sugars to the lipid carrier (Samuel and Reeves, 2003).

Almost all characterised *Salmonella* O-antigens are synthesised by the Wzy dependent pathway, which involves synthesis of individual repeat O-units on the cytoplasmic side of the inner membrane followed by transport of these units to the periplasmic side by the Wzx flippase (Liu *et al.*, 2013). The Wzy glycosyltransferase then polymerises these O-units on the periplasmic side of the inner membrane to form the completed O-antigen polymer (Islam and Lam, 2014) (**Figure 1.6 A**).

The ABC transporter dependent pathway is only described for *Salmonella* strains which produce the O:67 O-antigen (Liu *et al.*, 2013) but is more commonly identified in *E. coli* O-antigen biosynthesis (Greenfield and Whitfield, 2012). Rather than polymerisation of the O-antigen on the periplasmic side of the inner membrane, the complete O-antigen chain is synthesised on the cytoplasmic side and transported across the inner membrane by a specialised ABC transporter (**Figure 1.6 B**).

In the synthase dependent pathway, O-antigen polysaccharide synthesis and transport are paired by the synthase enzyme. This pathway is only characterised in one *Salmonella* serovar (Keenleyside and Whitfield, 1996). It is not yet well understood in this role and is usually associated with synthesis of exopolysaccharides other than LPS (Whitney and Howell, 2013).

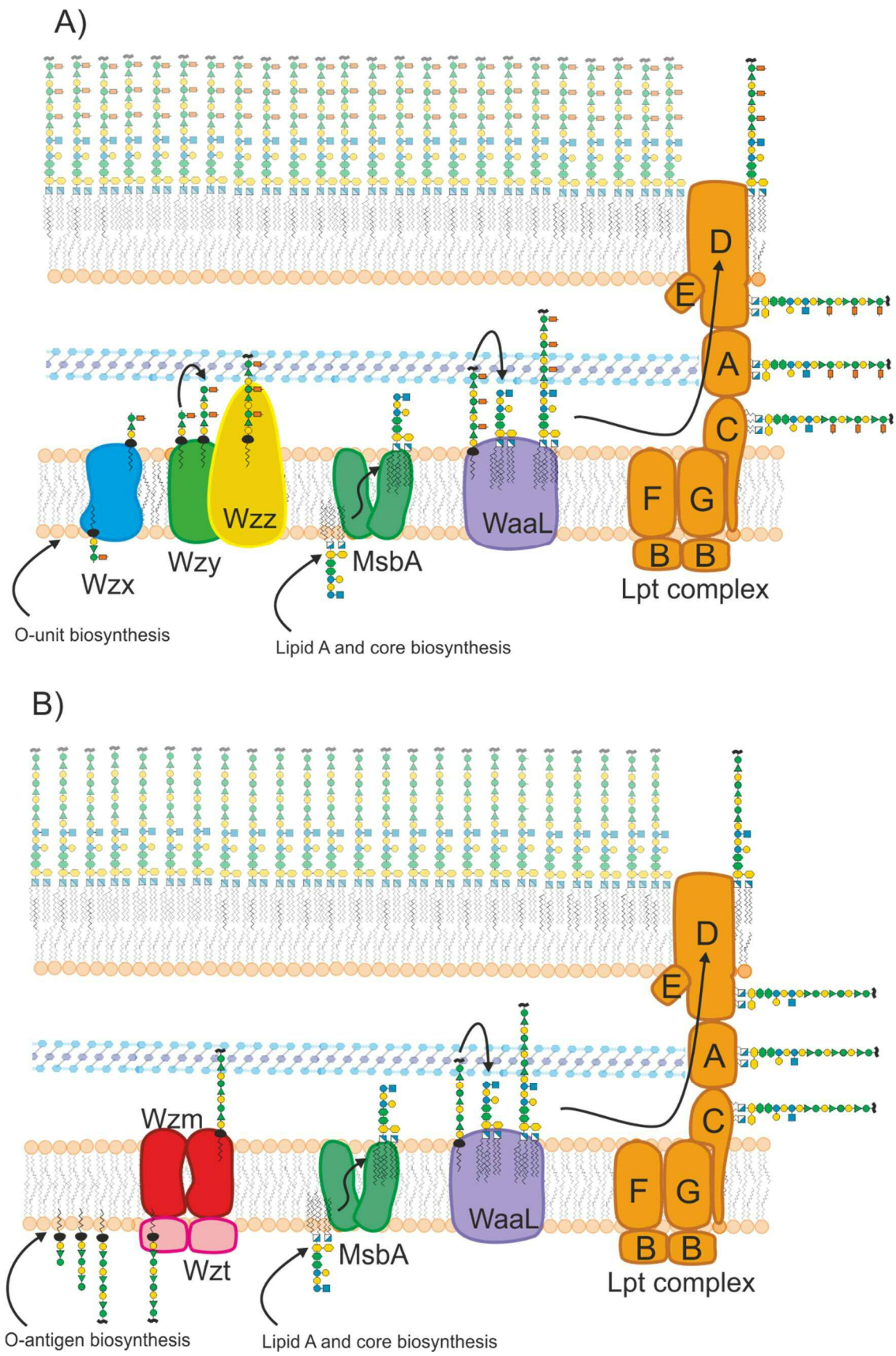


Figure 1.6| Pathways of LPS biosynthesis. A) Wzy dependent O-antigen biosynthesis. O-antigen units are synthesised on the cytoplasmic side of the inner membrane linked to a UndPP carrier. These O-units are flipped to the periplasmic side through the action of Wzx flippase then polymerised to form the full length O-antigen through the action of Wzy polymerase, with Wzz being the O-antigen chain length determinant.

Meanwhile MsbA flips LipidA + Core (synthesised on the cytoplasmic side of the inner membrane) to the periplasmic side then the O-antigen is added to the form the full LPS molecule by WaaL. The Lpt complex of proteins then transports fully formed LPS to the outer leaflet of the outer membrane. B) ABC transporter dependent O-antigen biosynthesis. Rather than O-units being polymerised to form the full length O-antigen on the periplasmic side of the inner membrane, the full O-antigen is synthesised on the cytoplasmic side linked to a UndPP carrier. The fully synthesised O-antigen is then flipped to the Periplasmic side of the inner membrane by the ABC transporter composed of Wzm and Wzt proteins. Ligation of the O-antigen to lipidA and core and its subsequent delivery to the outer leaflet of the outerm membrane then follows the same pathway as (A).

1.2.4 O-antigen modifications of *Salmonella*

Variation of O-antigen structure is not simply determined by O-unit synthesis on the cytoplasmic side of the inner membrane. Following transport of the O-antigen to the periplasmic side of the inner membrane, further modifications can be performed.

In Wzy dependent O-antigen biosynthesis, variation in the O-antigen structure can be inferred by altered linkage between O-units. The O:27 serotype of *Salmonella* contains the same repeating O-unit as *Salmonella enterica* subspecies *enterica* serovar Typhimurium (STM) (**Figure 1.3 B**) but each O-unit is connected by 1-6 rather than 1-2 glycosidic bonds (Wang *et al.*, 2002). This variation was initially thought to be determined by a serotype converting bacteriophage however was later shown by Wang *et al.,* (2002) to be determined by variants of the Wzy O-antigen ligase protein encoded outside of the *rfb* gene cluster (Grimont and Weill, 2008).

Another prevalent modification caused by genes outside the standard *rfb* gene cluster is glycosylation. Glycosyltransferase (*gtr*) operons encode proteins which decorate the basic O-antigen repeat structure with monosaccharides such as glucose (Davies *et al.*, 2013). The *gtr* operon encodes three proteins, GtrA, GtrB and the serotype specific glycosyltransferase (Guan *et al.*, 1999). GtrB is responsible for transfer of UDP-glucose to the UndP lipid carrier on the cytoplasmic side and GtrA is the flippase that flips the UndP-glucose to the periplasmic side of the inner membrane for the serotype specific glucosyltransferase (GtrC) to perform the specific glucosyltransferase modification of the O-antigen (Allison and Verma, 2000).

Glycosylation of *Salmonella* O-antigens has been linked with bacteriophage interactions, and serotype switching due to the genes being controlled by phase variation (Broadbent *et al.*, 2010). It can also have implications in *Salmonella* persistence. For example, glucosylation of the O:12 antigen of STM has been suggested to enhance long-term intestinal colonisation in the murine intestine (Bogomolnaya *et al.*, 2008).

In addition to glycosylation, O-antigen monosaccharides can be substituted with acetyl groups (Liu *et al.*, 2013). These acetylation modifications can have impact on bacteriophage interactions and have been linked to immunogenicity and virulence of pathogenic strains. For example, rhamnose acetylation by lysogenic *Salmonella* bacteriophage A3 and A4 were discovered to prevent bacteriophage adsorption by Wollin *et al*., (1987), rhamnose and abequose acetylated O-antigen glycoconjugate vaccines are suggested to be more immunogenic than unacetylated forms (Baliban *et al.*, 2017), and Hitri *et al.,* (2019) suggested that the O-antigen of *Salmonella* serovar Typhi was more rigid when acetylated, to prevent recognition of buried epitopes by the innate immune system.

The mechanism of O-antigen acetylation is not yet well characterised, therefore understanding this mechanism could enhance understanding of the O-antigen biosynthesis pathway and have applications in producing *Salmonella* strains with specifically acetylated O-antigens. O-antigen acetyltransferases of *Salmonella* will be discussed in section 1.2.5.

1.2.5 O-antigen acetyltransferases OafA and OafB

In STM there are currently two defined O-antigen acetyltransferases OafA and OafB, which are the focus of this research. OafA was discovered by Slauch *et al.,* (1996) to be responsible for acetylating the 2-hydroxyl group on the abequose moiety of the STM O-antigen. The definition of the name of OafA is not currently published in the literature and a suggested definition relevant to current understanding is given later in this section.

The O-antigen modification catalysed by OafA has been shown to affect the immunological properties of *Salmonella* LPS, therefore affecting the interaction of the bacterium with its host organism (Slauch *et al.*, 1995). Its clinical importance has been demonstrated by Lanzilao *et al.,* (2015) who found that O-antigen abequose acetylation

was required for production of protective antibodies. Abequose acetylation by OafA is recognised by serotyping antibodies and causes conversion to the O:5 serotype of the Kauffmann White Lee Minor scheme (Slauch *et al.*, 1995).

In addition to abequose acetylation, multiple *Salmonella* serovars contain a rhamnose moiety on the O-antigen which can be acetylated (Liu *et al.*, 2013). The protein responsible for rhamnose acetylation in STM was first discovered through analysis of a family II g*trC* gene encoded on the BTP1 prophage associated with STM D23580, a major cause of invasive multi drug resistant iNTS in sub-Saharan Africa (Kingsley *et al.*, 2009). Rather than being responsible for the O-antigen glycosylation associated with other Gtr operons, this protein was found to be responsible for acetylation of O-antigen rhamnose at the 2- and 3- hydroxyl groups (Kintz *et al.*, 2015, 2017). This modification is shown to protect *Salmonella* from lysis by the BTP1 bacteriophage and is thought to act through steric hindrance of the BTP1 tailspike protein that has rhamnosidase activity and would usually cleave the O-antigen, allowing access to the cell surface (Kintz *et al.*, 2015). *Salmonella enterica* subspecies *enterica* serovar Paratyphi A (SPA) also possesses acetylation at rhamnose 2- and 3- hydroxyl groups (Ravenscroft *et al.*, 2015) and a homologue of OafB has been identified through sequence analysis which is shown to be responsible for this modification in *Salmonella ser.* Typhi (Kintz *et al.*, 2017). This rhamnose acetylation was shown to be controlled by epigenetic phase variation, allowing the phenotype to be switched ‘on’ and ‘off’ within a bacterial population, and it was found to be an immunodominant modification (Kintz *et al.*, 2017). Thus, O-antigen rhamnose acetylation is involved in modulation of bacteriophage interactions and its phase variable expression could promote persistence in the host.

It is clear that the rhamnose acetylating family II GtrC protein of STM is an acetyltransferase with no functional relationship to the GtrABC glycosylating proteins (Kintz *et al.*, 2015). The gene encoding this acyltransferase protein could have been incorporated into the Gtr operon as a result of a recombination event or lysogenization of phage DNA. Therefore, we propose to rename this and orthologous rhamnose acetyltransferases as OafB. The name reflects the protein architecture (O-antigen acetyltransferase fused B), similar to that we suggest for OafA (O-antigen acetyltransferase fused A).

Both *Salmonella* O-antigen abequose and rhamnose acetyltransferases contain an N-terminal multiple membrane spanning domain characterised as an Acyltransferase-3 (AT3) domain (IPR002656, PF01757), fused to a significant extra-cytoplasmic region which bears homology to an SGNH domain (IPR013830, PF14606 or PF13472) (Finn *et al.*, 2014; Mitchell *et al.*, 2015)(**Figure 1.7**).

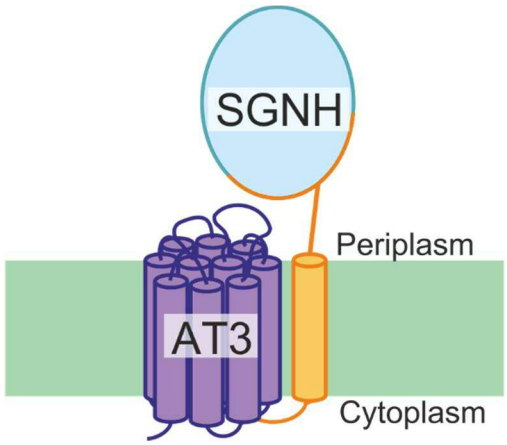


Figure 1.7| Domain architecture of *Salmonella* O-antigen acetyltransferases. The membrane bound AT3 domain (purple) is connected to the periplasmic SGNH domain (blue) via a membrane spanning linking region (orange).

AT3 domains are a common characteristic of many bacterial polysaccharide O-acetyltransferases which play diverse roles in bacterial interactions and survival, and will be discussed further in section 1.3. SGNH domains have a slightly confusing history and were first identified as a subgroup of the GDSL hydrolase family, with their transferase ability being characterised at a later date. The GDSL family, first described by Upton and Buckley (1995), is characterised by five catalytic blocks containing Ser-His-Asp/Glu catalytic triad and an essential GDSL motif in catalytic block I. The SGNH-hydrolase family was subsequently identified as a subgroup of this family. Firstly, the acetyl esterase proteins RGAE and SsEst were classified as members of a new family of hydrolases due to conservation of catalytic residues within three of the five catalytic blocks (I, III, V) identified previously for GDSL proteins (Dalrymple *et al.*, 1997). Subsequently, the crystal structure of RGAE was solved (Mølgaard *et al.*, 2000), this enabled identification of related protein structures that had been submitted to the protein data bank with a high structural homology α/β/α hydrolase fold, but relatively low sequence identity. Alignments with this wider list of acyl esterase proteins revealed conserved residues corresponding to block II of the GDSL catalytic blocks (Mølgaard *et al.*, 2000). The SGNH-hydrolase family name was then coined due to the finding that one single residue in

each of the four catalytic blocks identified was conserved and catalytically important (S-Block I, G-Block-II, N-Block III, H-Block V) (Mølgaard *et al.*, 2000).

1.3 Acyltransferase_3 family of bacterial proteins: important in many biological processes but mechanistically undefined

OafA and OafB belong to a large family of bacterial proteins which contain membrane bound acyltransferase_3 (AT3) domains. This family of proteins are responsible for O-acylation of diverse oligosaccharide substrates. O-acylation is the result of formation of a covalent bond between an oxygen atom in a substrate, and an acyl group, whereas, N-acylation occurs when the acyl group is covalently bound to a nitrogen atom. O-acylation of sugar moieties is a common occurrence across multiple bacterial species involving acetyl-, propionyl-, butyryl-, succinyl-, and many other acyl substituents. These modifications have applications in many biological processes such as antigenic variation, (Verma *et al.*, 1991; Slauch *et al.*, 1996; Kintz *et al.*, 2015), osmoregulation (Lacroix *et al.*, 1999), virulence (Bera *et al.*, 2005), cell division (Laaberki *et al.*, 2011) and modification of secreted or extracellular oligosaccharides involved in many bacterial interactions; from competition (Hara and Hutchinson, 1992) to initiation of symbiosis (Davis *et al.*, 1988).

AT3 domains are around 300-350 amino acids long and their molecular function is classified by the InterPro database as having 'transferase activity, transferring acyl groups other than amino-acyl groups' (Mitchell *et al.*, 2015). It appears that the name acyltransferase_3 emerged as an arbitrary identifier, originally defined by the InterPro database as it was the third acyltransferase domain to be identified as a distinct functional unit. The first paper to refer to this functional domain through reference to the InterPro database was by Zhu *et al.,* (2007), however proteins containing these functional domains have been studied for many years previous, with little cross reference between the papers studying them.

These proteins were first proposed to be a family of integral membrane proteins responsible for trans-acylation modifications by Slauch *et al.,* (1996), who identified a number of proteins involved in acylation of sugar moieties that showed significant homology to OafA. These were termed integral membrane trans-acylase family proteins, later to be defined by the InterPro database as acyltransferase_3 family proteins.

Although initial recognition of what became the AT3 proteins was acknowledged back in 1996, the mechanism of action of the proteins responsible for these specific and biologically important acylation modifications has still not been well defined and there is no comprehensive review compiling the current knowledge of the acyltransferase_3 family proteins. In the following section a detailed summary of the diverse carbohydrate modifications performed by prokaryotic AT3 family proteins is outlined.

1.3.1 Key examples of prokaryotic AT3 domain containing carbohydrate O-acetyltransferases

1.3.1.1 LPS O-antigen

1.3.1.1.1 WbaK (orf17.4)

WbaK is responsible for acetylation of carbon six of the galactose residue of the *S. enterica* serogroup E1 O-antigen (Hong *et al.*, 2013). In contrast to *oafA* and *oafB*, which reside at loci distinct from the O-antigen biosynthesis gene cluster in the *Salmonella* genome, the *wbaK* gene is closely associated with O-antigen synthesis genes and is predicted to have been inserted into the end of this gene cluster by homologous recombination (Hong et al., 2013). The WbaK acetyltransferase may play roles in bacteriophage interaction as it is the target of the epsilon15 bacteriophage which converts group E1 *Salmonella enterica* to group E2 by inhibition of the O-antigen acetylation modification (Hong *et al.*, 2013).

1.3.1.1.2 Oac proteins (OacA,B,C,D)

In addition to its significance in *Salmonella,* O-antigen acetylation is an integral feature of serotype conversion in the Gram-negative bacterial pathogen *S. flexneri.* Oac, one of the most well studied members of the AT3 family, has been shown to catalyse 2-O-acetylation of RhaI of the tetra-saccharide repeat (RhaIII-RhaII-RhaI-GlcNAc) of the *Shigella* O-antigen shared by all serotypes of *S. flexneri* except from serotype 6. It was independently identified by both Verma *et al.*, (1991) and Clark *et al.,* (1991) through investigating the ability of bacteriophage *SF6* to antigenically convert *S. flexneri* serotypes through acetylation of the LPS O-antigen. O-antigen acetylation caused by the *SF6* phage causes immunity to superinfection by further *SF6* phage as the modification sterically hinders its endorhamnosidase activity.

A second *oac* gene was identified in *S. flexneri* strain 1b by Sun *et al.,* (2012), which was named *oac1b* and thought to be carried by a different phage that is only present in 1b strains. Two further *S. flexneri* O-antigen acetylation modifications were identified by Perepelov *et al.,* (2012), including 3/4-O-acetylation of RhaIII and 6-O-acetylation of GlcNAc. This finding resulted in the identification of further Oac proteins in *S. flexneri* responsible for these modifications, prompting the original Oac protein to be re-named to OacA. OacB was found to be responsible for the 3/4-O-acetylation of RhaII in serotypes 1a, 1b, 2a, 5a, and Y but not serotype 6 (Wang *et al.*, 2014) whereas OacC is responsible for the 3/4-O-acetylation of RhaII specifically in serotype 6 (Knirel *et al.*, 2014). OacD was found to be present on the serotype converting bacteriophage *SfII* genome and identified as the *oac* homologue responsible for 6-O-acetylation of GlcNAc by Sun *et al.,* (2014).

Homologues of Oac have been identified in *Burkholderia,* responsible for 4-O-acetylation of 6-deoxy-α-l-talopyranose of the LPS O-antigen (Brett *et al.*, 2011), and a second gene, denoted *wbiA*, was found to be responsible for acetylation of the same O-antigen sugar at the C2 position (Brett *et al.*, 2003).

Although proven to be responsible for a range of O-antigen acetylation modifications through mutation and complementation, the mechanism of action of these O-antigen acetyltransferase proteins is yet to be discovered. With multiple proteins that cause specific O-antigen acetylation in *S. flexneri* and *Burkholderia* spp.*,* it is clear that they are likely to play important biological roles, for example, in bacteriophage interaction.

1.3.1.1.3 Lag1

Lag1 was identified by Zou *et al.,* (1999) as the gene responsible for altering the reactivity of *Legionella pneumophila* LPS to serogroup 1 LPS monoclonal antibodies. NMR analysis revealed that *lag1* mutants failed to acetylate position 8 of legionaminic acid moiety in the LPS O-antigen, which has a similar structure to sialic acid (Knirel *et al.*, 1994; Zou *et al.*, 1999). The acetylation modification of *Legionella* LPS by Lag1 has been associated with virulence of serogroup 1 LPS strains (Kozak *et al.*, 2009), highlighting the clinical importance of this modification. Zähringer *et al.,* (1995) determined that 8-O-acetylation of leigionaminic acid causes increased hydrophobicity of *Legionella* cell surface and suggest that this may enable stable aerosol production, which is a major source of transmission of Leigonnaires' disease. Some functional analysis of Lag1 has

been conducted by Luck *et al.,* (2001); they showed that a serine to leucine amino acid change in the highly conserved motif V-X-X-F-F-X-(I/V/L)-S-G-(F/W/Y), which is shared among many AT3 proteins from both Gram-positive and negative bacteria, was responsible for loss of O-acetyltransferase activity of Lag1. The mechanistic role of this single amino acid is not yet characterised and there is no *in vitro* assay for the acetylation modification.

1.3.1.2 LOS core oligosaccharides

The HI_0392 protein was included in the analysis which first identified the AT3 domain protein family (Slauch *et al.*, 1996) but at the time had unknown function. It was later suggested that the *HI0392* gene in *H. influenzae* strain Rd was originally joined with its adjacent gene *HI0391* to create an OafA homologue but two single base insertions into the genome have split the larger reading frame (Fox *et al.*, 2005). In other *H. influenza* strains such as strain Egan, this is a single ORF which functions as an O-acetyltransferase to acetylate HeptoseIII of the inner core. This modification promotes serum resistance to complement mediated killing (Fox *et al.*, 2005).

Lot3 in *N. meningitidis* is responsible for C-3 acetylation of GlcNac in the inner core of meningococcal lipooligosaccharide (LOS) (Kahler *et al.*, 2006). LOS of *N. meningitidis* is an important virulence factor. It has been implicated in interaction with host epithelial cells and induces proinflammatory responses during human infection (Kahler and Stephens, 1998; Van Deuren *et al.*, 2000). Jennings, Lugowski and Ashton (1984) found that O-acetylated LOS are more immunoreactive, producing higher antibody titres than their non-acetylated counterparts.

1.3.1.3 Enterobacterial common antigen

The enterobacterial common antigen (ECA) was discovered in the 1960s to be a highly conserved polysaccharide found across all enterobacteria (Kunin, 1963). ECA can be presented in the outer membrane by linkage to phosphoglyceride (Kuhn *et al.*, 1983), as a cyclic form in the periplasm (Kajimura *et al.*, 2005), and predominantly in bacterial strains incapable of producing O-antigens, it can be bound to lipid A (Kuhn *et al.*, 1988). *Escherichia coli* K-12 ECA is composed of repeating units of GlcNAc - (1-4)-ManNAcA-(1-4)- Fuc4NAc with an acetyl substituent on C6 of the GlcNAc residue (Lugowski *et al.*, 1983). WecH was found to be responsible for this O-acetylation of the GlcNAc residue (Kajimura *et al.*, 2006). ECA has been suggested to play a role in *Salmonella* virulence,

connected to resistance against bile salts (Ramos-Morales *et al.*, 2003), however, a clear role for the acetylation modification of ECA has yet to be presented.

1.3.1.4 Capsular polysaccharide

The *wcjE* gene lies within the capsular polysaccharide synthesis (cps) locus of *Streptococcus pneumoniae* and is highly conserved across many serotypes. Calix and Nahm (2010) identified WcjE as the acyltransferase responsible for the O-acetylation of the 1-phosphoglycerol residue in *S. pneumoniae* serotype 11Aα capsular polysaccharide. Calix *et al., (2012)* subsequently found that WcjE in serotype 9V is responsible of 6-O-acetylation of capsular polysaccharide β-N-acetylmannosamine, thus, it is possible that various *wcjE* genes across different serotypes could perform different acetyl modifications to the capsular polysaccharide, modulating interactions with its environment. Indeed, WciG was identified to specifically acetylate carbon 2 of 6-β-galactofuranose in the capsular polysaccharide of a serotype 35C *S. pneumoniae* isolate (Geno, Saad, *et al.*, 2017), and the WcjE homologue of this isolate specifically acetylated carbon 5 and 6 of the 3-β-galactofuranose sugar (Geno, Bush, *et al.*, 2017).

The modifications resulting from expression of WciG and WcjE produce varied biological properties within pneumococcal serotypes, resulting from altered biochemical properties and barrier function of the capsular polysaccharide (Spencer *et al.*, 2017).

Additionally, loss of capsular polysaccharide acetylation is associated with invasive pneumococcal disease, where *S. pneumoniea,* which usually colonise the nasopharynx, invade the bloodstream. This change is associated with the ability of non-acetylated capsular polysaccharide to bind ficolin, a pattern-recognition molecule in human serum which initiates phagocytosis (Brady *et al.*, 2014). This demonstrates the intricate link between acetylation of capsular polysaccharide with the clinical behaviours of pneumococcal serotypes.

1.3.1.5 Peptidoglycan

Peptidoglycan acetylation is an important factor in controlling autolysins and lytic enzymes across many bacterial species; reviewed by Moynihan & Clarke (2011). Bera *et al*., (2005) were able to identify *oatA* as the gene responsible for peptidoglycan acetylation in *Staphylococcus aureus.* They showed that OatA was able to O-acetylate carbon 6 of MurNAc in peptidoglycan, which is responsible for lysozyme resistance in pathogenic staphylococci. Peptidoglycan acetylating homologues of OatA have also

been found in other Gram-positive bacteria following this study (Crisóstomo *et al.*, 2006; Veiga *et al.*, 2007; Aubry *et al.*, 2011; Bernard *et al.*, 2011; Laaberki *et al.*, 2011). Through these analyses, further roles of peptidoglycan acetylation in regulating extracellular protein anchoring and cell division were demonstrated (Laaberki *et al.*, 2011).

In parallel with OafA and OafB, OatA contains an extra cytoplasmic C-terminal SGNH domain (Aubry *et al.*, 2011) and the crystal structure of the SGNH domain of *S. pneumoniae* has been solved (Sychantha *et al.*, 2017)(**Figure 1.8**). This provides a valuable comparison for analysis of the fused SGNH domain of OafA and OafB, however, this domain was investigated without the context of the AT3 domain. Thus, the juxtaposition of these two functional domains has not been further studied.

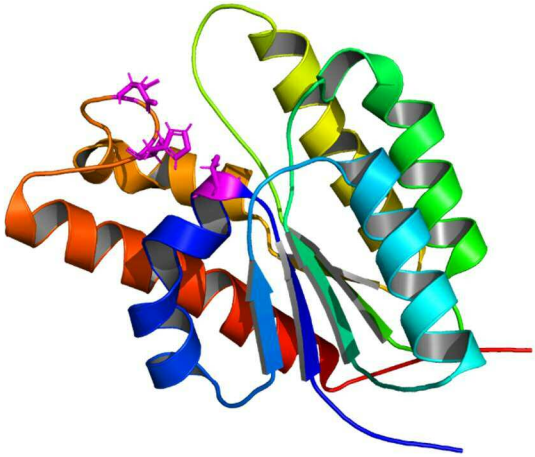


Figure 1.8| Crystal structure of the SGNH domain of *S. pneumoniae* solved by Sychantha *et al.* (2017).

1.3.1.6 Osmoregulated periplasmic glucans

The sole constituent sugar of OPG's is D-glucose but they can display a variety of structures and multiple substituents including acetyl and succinyl groups (Bohin, 2000) (section 1.1.5). The first protein to be found responsible for succinylation of OPG's was MdoC. It was identified by Lacroix *et al.,* (1999), through a screen for *E. coli* K-12 mutants lacking OPG succinylation. A similar approach was used by Cogez *et al.,* (2002) to identify the corresponding OPG acetyltransferase in *Rhodobacter sphaeroides* named OpgC. Homologues of OpgC have been identified in other Gram-negative bacteria following this, however the mechanism of action is as yet uncharacterised (Roset *et al.*, 2006; Bontemps-Gallo *et al.*, 2016).

1.3.1.7 Poly-N-acetylglucosamine

Biofilms are a community of surface associated bacterial cells that are embedded in a matrix of extracellular components (Donlan, 2002). This matrix is mostly produced by the bacteria and is composed of a range of biopolymers including polysaccharides, protein and DNA which define the biofilm architecture (Flemming and Wingender, 2010). The ability of bacteria to form biofilms is closely linked to their ability to cause chronic persistent infections through decreased antimicrobial susceptibility and resistance to phagocytosis or mechanical clearance (Bjarnsholt, 2013).

Biofilm forming staphylococci are frequently associated with infections of inserted medical devices (von Eiff *et al.*, 2002) and an essential component of their biofilm matrix is the partially N-deacetylated polymer of N-acetylglucosamine (poly-N-acetylglucosamine - PNAG) (Rohde *et al.*, 2010). GlcNAc moieties of PNAG can also be O-succinylated and this modification is suggested to be catalysed through the action of the IcaC protein (Sadovskaya *et al.*, 2005; Atkin *et al.*, 2014). IcaC was originally thought to be responsible for transport of extracellular PNAG through the *s*taphylococcal cell membrane during biofilm formation (Heilmann *et al.*, 1996). However, due to its homology to other AT3 domain O-acetyltransferases IcaC has since been suggested to O-succinylate the GlcNAc residues of PNAG (Atkin *et al.*, 2014). Acetylation of PNAG alters the net charge of the polymer and the degree of PNAG N-deacetylation and O-succinylation varies with growth condition. Therefore, it is hypothesised that regulation of the physiochemical properties of PNAG could assist in adherence and biofilm formation on a range of different surfaces and therefore enhance the *in vitro* survival of staphylococcal species (Sadovskaya *et al.*, 2005). A role for IcaC in formation of mature biofilm has also been identified (Heilmann *et al.*, 1996). Therefore, exploring the mechanism of this acyltransferase protein could reveal its significance as a therapeutic target.

1.3.1.8 Pilin

By process of elimination Warren *et al.*, (2004) determined that PglI was responsible for acetylation of the tetrasaccharide structure which is O-linked to Ser63 of Pilin in *N. meningitidis* strain C311#3. Pilin is a surface exposed adhesin protein which is a major virulence factor (Mattick, 2002). It can be post-translationally modified by addition of a covalently-linked glycan structure which, in the case of C311 strains of *N. meningitidis,*

is a tetrasaccharide repeat composed of Gal(β1-4),Gal(α1–3),[2,4-diacetamido-2,4,6-tri-deoxyhexose] (DATDH) (Stimson *et al.*, 1995). Only the DATDH of this tetrasaccharide repeat is acetylated and Warren *et al.*, (2004) concluded that PglI is the acetyltransferase responsible for DATDH acetylation of the pilin tetrasaccharide. This was also supported by the experiments of Aas *et al.,* and Anonsen *et al.,* (2007; 2017) for *N. gonorrhoeae*.

1.3.1.9 Xanthan

GumF and GumG were both identified by Katzen *et al.,* (1998) as the proteins responsible for O-acetylation of the exopolysaccharide xanthan produced by *Xanthomonas campestris* pv. Campestris. Xanthan has a wide commercial application as a thickening agent and is composed of a 1-4 linked repeating D-Glucose backbone with side chains composed of D-Mannose(β1-4), D-Glucoronic acid(β1-2), and D-Mannose is attached to alternate glucose residues of the backbone by α1-3 linkage (Jansson *et al.*, 1975). GumF was found to be responsible for acetylation of the mannose residue closest to the glucose backbone of xanthan whereas GumG was found to be responsible for acetylation of the terminal mannose of the side chain (Katzen *et al.*, 1998). Acetylation of xanthan has been shown to affect its viscosity and this has many industrial applications (Hassler and Doherty, 1990; Lopes *et al.*, 1992).

1.3.1.10 Root nodulation factors

1.3.1.10.1 NodX

Nod factors (NFs) are rhizobial lipo-chitinoligosaccharide signals that trigger root nodule development in legumes and their structural modification can affect their biological activity (Dénarié *et al.*, 1996). The *nodX* gene was first identified as it provided *Rhizobium leguminosarum* biovar viciae strain TOM with the ability to nodulate Afghanistan peas (Davis *et al.*, 1988). This work was followed up by Firmin *et al.,* (1993a) who suggested that NodX was responsible for O-acetylation of C6 of the reducing GlcNAc residue of the *R. leguminosarum* viciae nodulation factor Nod*RIv*-V. This modification is speculated to have some role in Nod factor perception through interaction with the *sym2* gene product, whose activity depends on the structure of Nod factors secreted by the infecting bacterium (Geurts *et al.*, 1997; Geurts and Bisseling, 2013). However, multiple other mechanisms have been suggested and further characterisation is required.

1.3.1.10.2 NolL

The NolL protein obtained from *Rhizobium loti* (*Mesorhizobium loti*) functions as an acetyl transferase, transferring an acetyl group specifically to the GlcNAc5 species of fucose lipo-chitinoligosaccharides (Pacios Bras *et al.*, 2000). Acetyltransferase activity of NolL has been associated with increased Nod factor and lipo-chitooligosaccharide production involved in root nodulation.

1.3.1.10.3 ExoH and ExoZ

Succinoglycan is an acidic calcoflour-binding exopolysaccharide (EPS) which is important for effective induction of root nodule formation by *Sinorhizobium meliloti* (or *Rhisobium meliloti*) on alfalfa (Leigh *et al.*, 1985). It is composed of repeating octasaccharide subunits which have acetyl, succinyl, and pyruvyl substituents (Reinhold *et al.*, 1994). Succinylation and acetylation of succinoglycan have been shown to be performed by ExoH and ExoZ respectively. ExoH was identified to be responsible for C-6 succinylation of glucose 7 of succinoglycan by Leigh *et al.,* (1987) in a genetic screen for *R. meliloti* with altered succinoglycan formation. Lack of this succinylation modification resulted in ineffective root nodulation. ExoZ was identified by Buendia *et al*., (1991) to affect succinoglycan production and it was later discovered that this protein was responsible for the acetylation of C6 of glucose of *S. meliloti* succinoglycan (Reuber and Walker, 1993).

R. meliloti produces both a high molecular weight (HMW) and low molecular weight (LMW) form of succinoglycan, the latter having been suggested to mediate root nodule invasion (Battisti *et al.*, 1992). ExoK and ExsH glycanases have been proposed to contribute to the cleavage of HMW succinoglycan to its LMW form (York and Walker, 1998a). The succinyl and acetyl modifications performed by ExoZ and ExoH respectively have been shown to influence the susceptibility of succinoglycan to cleavage by ExoK and ExsH, with succinylation enabling this cleavage and acetylation inhibiting it (York and Walker, 1998b). Hence, cooperative action of AT3 domain containing carbohydrate acyltransferases plays a significant role in root nodulation, an important factor in enhancing crop yields.

1.3.1.11 Macrolide antibiotics

Macrolide antibiotics are compounds which contain a macrocyclic lactone ring (Ōmura, 2002). These molecules are structurally related but can contain different sugar residues

and acyl substituents to create slightly different antibiotics with differing antibiotic ability. CarE was identified when Epp *et al.,* (1989) investigated the possibility of creating hybrid macrolide antibiotics through introduction of genes from different species of *Streptomyces*. Carbomycin and Spiramycin are structurally related macrolide antibiotics synthesised by *Streptomyces thermotolerans* and *Streptomyces ambofaciens* respectively (Vazquez, 1967; Epp *et al.*, 1989). Carbomycin contains an isovaleryl group attached to position four of the mycarose sugar which is absent on spiramycin. Epp *et al.,* (1989) demonstrated that CarE was responsible for this acyl modification. There is also evidence of some substrate promiscuity in CarE; Arisawa *et al.,* (1993) demonstrated that CarE expressed in *Streptomyces lividans* was able to produce 4-O-isovaleryltylosin as well as 4-O-acetyl, 4-O-propionyl, and 4-O-butyryltylosin. Tylosin is another structurally related macrolide antibiotic.

In similar experiments MdmB was identified from *Streptomyces mycarofaciens* as a 3-O-acyltransferase that catalyses the addition of acetyl and propionyl groups to position 3 of the lactone ring in 16-member macrolide antibiotics like midecamycin and spiramycin (Hara and Hutchinson, 1992). Subsequently, further macrolide antibiotic acyltransferases AcyA, Mpt and MidE have been discovered in *Streptomyces* species and these proteins have been extensively used to manipulate the structures of common macrolide antibiotics (Arisawa *et al.*, 1994; Cong and Piepersberg, 2007).

1.3.1.12 Antitumour drugs

The *cmmA* gene was first identified through analysis of the biosynthetic gene cluster of Chromomycin A_3, an aureolic acid type antitumour drug, from *Streptomyces griseus* (Menéndez *et al.*, 2004a). CmmA was then functionally characterised by Menéndez *et al.*, (2004b) to be responsible for the O-acetylation of C-4 of L-chromose B and D-oliose of Chromomycin A_3. The acetylation modifications on Chromomycin A_3 proved to be functionally important as the parental compound lacking acetyl groups had reduced antibiotic and antitumour activity. Menéndez *et al.*, (2004b) were also able to show that this modification takes place after the sugar residues have been added to the Chromomycin A3 precursor. García *et al.,* (2011) showed that CmmA exhibits both acyl donor and acceptor substrate flexibility as it is able to produce differently acetylated derivatives of exogenously added mithramycin, another structurally related antitumour compound and is also able to use acetyl-, isobutyryl- and propionyl-CoA to perform

these modifications.

Ansamitocins are antitumour drugs synthesised by *Actinosynnema pretiosum*. Acyl side chains are important for activity of the compound as Maytansinol, which does not have these side chains, is inactive. Asm19 was proposed to be the 3-O-acyltransferase responsible for catalysing the attachment of the acyl side chain of the ansamitocins (Yu *et al.*, 2002). This was supported by Moss *et al.,* (2002) who found that Asm19 had high substrate acceptor specificity paired with donor substrate flexibility as it was able to use acetyl-, propionyl-, isobutyryl-, butyryl-, or isovaleryl-CoA as a substrate but could only add these groups to C3 of N-desmethyl-4,5-desepoxymaytansinol (an ansamitocin precursor).

1.4 Characterising the mechanism of AT3 domain containing O-acyltransferases

It is clear from section 1.3 that AT3 domain containing O-acyltransferases play varied and important roles in bacterial physiology, adaptations and interactions. Although increasing numbers of these proteins have had their acyltransferase activity experimentally confirmed, there is limited knowledge about the mechanism of action of any of these proteins. Although acyl-CoA derivatives have been used for *in vitro* characterisation of many of these enzymes, the *in situ* acyl donor has not been experimentally confirmed. In LPS polysaccharide acetylation, the stage of LPS biosynthesis at which these acyl modifications are performed is unknown and although some conserved functional residues have been highlighted, their role in the mechanism of O-acylation of bacterial carbohydrates is still unknown. An intriguing further question raised by the proteins within this family is the significance of the extracytoplasmic SGNH domain. As is the case for OafA and OafB, some of the AT3 family acyltransferase proteins discussed in section 1.3.1 contain a fused SGNH domain. In some instances, importance of this extracytoplasmic domain has been demonstrated through catalytic inactivation or truncation of the protein (Thanweer and Verma, 2012; Kintz *et al.*, 2015). Additionally, in the proteins that do not contain fused SGNH domains, no alternative fused domains are suggested. Therefore, determining the role of this domain in SGNH-fused proteins is of keen interest.

1.5 Aims and objectives

The aim of this thesis is to improve the mechanistic understanding of AT3 domain containing O-acyltransferases through the study of OafA and OafB from STM and will address the following questions:

- Can identification of essential residues in the membrane-bound AT3 domain and periplasmic SGNH domain give clues to their role in acetyl transfer?
- Can we obtain insight into the juxtaposition of the AT3 and SGNH domains by elucidating the structure of the periplasmic domain of OafA or OafB?
- What is the function of the SGNH domain and can it function independently of the AT3 domain?
- Can answering these questions inform a refined mechanistic model for the acetylation of the STM O-antigen by OafA and OafB?

Chapter 2 Materials and Methods

2.1 Suppliers

All chemicals, reagents and media were purchased from Sigma-Aldrich, Merck, Fisher Scientific, Amersham, Thermo Fisher, New England Biolabs (NEB), GE Heathcare, Qiagen, Bio-Rad, Scientific Laboratory Supplies (SLS), Santa Cruz Biotechnology, Invitrogen, and Oxford Biosystems.

2.2 Bacterial strains, plasmids and oligonucleotides

The bacterial strains used in this research are listed in **Table 2.1**. *Salmonella enterica* subspecies *enterica* serovar Typhimurium (STM) LT2 Basal O-ag strains expressing OafA point mutants are listed in **Table 2.10** and OafB point mutant expression strain (Path 892) is described in section 2.5.9. *Escherichia coli* BL21(DE3) strains for production of OafA and OafB protein are listed in **Table 2.16**. All plasmids are listed in (**Table 2.2**). Oligonucleotides used in this research are listed in their relevant sections but in all cases they were sourced from Sigma-Aldrich in dry format and re-suspended in sterile 10 mM Tris, 0.1 mM EDTA pH 8.0 to 100 µM concentration then diluted to 10 µM with ddH_2O for use in PCR. Oligonucleotides were stored at -4 °C in their stock or working concentrations.

Table 2.1| Bacterial strains used

Strain	Genome	Source	Use
Escherichia coli strains			
XL1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacIq ZΔM15* Tn*10* (Tet^r)]. (Genes listed signify mutant alleles. Genes on the F´ episome, however, are wild-type unless indicated otherwise).	Agilent	Subcloning of all plasmids engineered for this project.

BL21(DE3)	F–, *ompT, hsd*SB (r_B–, m_B–), *gal, dcm,* (DE3)	Novagen	Expression strain for C-terminal OafA and OafB constructs.
Salmonella enterica subspecies enterica serovar Typhimurium strains			
Strain	**Genome**	**Source**	**Use**
LT2 (Path 346)	STM LT2 (ATCC strain number 19585, Lot number 215096)	ATCC	Cloning the *oafA* gene for expression plasmid construction.
LT2 Basal O-ag (Path 293)	STM LT2 Δ*gtr* all, Δ*oafA*	(Davies et al., 2013)	Test strain for plasmid expressed O-antigen modification genes.
D23580 (Path 189)	STM D23580 Malawi isolate	Sanger institute	Cloning the *oafB* gene for expression plasmid construction.
LT2 oafA-His (Path 1102)	STM LT2 with 10XHis tag added to C-terminal end of *oafA* by lambda red recombination (2.5.14)	This project	Confirmation of OafA expression from wild type STM LT2
Path 86	STM LT2 transformed with pKD46 (Table **2.2**)	Mark R Davies (University of York)	Lambda red recombination of STM LT2

Table 2.2| List of plasmids used. Amp = Ampicillin, Kan = Kanamycin.

Plasmid name	Description	Antibiotic resistance	Source / Reference
pBADcLIC2005 (PMV432)	pBAD vector with LIC cloning site. Adds a C-terminal 10xHis tag to target protein.	Amp	University of York technology facility (Geertsma and Poolman, 2007)
PMV433	pBADcLIC2005 vector encoding OafA from STM LT2 (# WP_000639473)	Amp	Created by Reyme Herman using ligation independent cloning (2.5.8)
PMV434	pBADcLIC2005 vector encoding OafB from STM D23580 (# SIU02679) created by Reyme Herman using ligation independent cloning	Amp	Created by Reyme Herman using ligation independent cloning (2.5.8)
pETFPP_30 (PMV463)	Adds PelB leader for periplasmic expression and 3C cleavable C-terminal 10xHis tag.	Amp	University of York technology facility
pDHL_1029	Contains FRT site flanked Kanamycin resistance cassette for lambda red recombination	Kan	(Ke *et al.*, 2016)
pKD46	Temperature sensitive lambda red recombinase expression plasmid	Amp	(Datsenko and Wanner, 2000)
pCP20	Temperature sensitive plasmid that gives thermal induction of FLP recombinase synthesis	Amp	(Datsenko and Wanner, 2000)

2.3 Growth media, culture conditions and buffer recipes

2.3.1 Growth Media

Table 2.3| Growth media recipes.

Growth Media	Recipe
LB Media	20 g/L LB broth powder - Lennox (Fisher)
Freeze Media	20 g/L LB broth powder – Lennox (Fisher), 20% v/v Glycerol
SOC Media	2% w/v Tryptone (Fisher), 0.5% w/v Yeast extract (Sigma), 10mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM Glucose
Agar plates	When solid media was required, 17 g/L agar was added to the media solutions before autoclaving.

2.3.2 Bacterial culture conditions and storage

Bacteria were cultured at 37 °C, and with 200 rpm shaking for liquid media, unless otherwise stated. Glycerol stocks of bacterial strains were stored at -80 ° C. Bacterial cultures on solid LB media were removed with a swab and re-suspended in 1 mL freeze medium (**Table 2.3**). Alternatively, 400 µL 50% glycerol was mixed with 800 µL stationary phase liquid LB media culture before storage at -80 °C.

Antibiotic stocks were made up at 1000x working concentration in deionised water, or in the case of Kanamycin in ethanol, and filter sterilised through a 0.2 µM syringe filter. Aliquots of antibiotic were stored at -20 °C. The working concentrations of antibiotics used are Ampicillin (Amp) = 100 µg /ml, Kanamycin (Kan) = 50 µg /ml.

2.3.3 Buffer recipes

Table 2.4| Buffer components.

Buffer Name	Buffer components
25X Protease inhibitor cocktail	1x cOmplete™ protease inhibitor cocktail tablet (Roche) dissolved in 2 ml diH_2O
4x SDS sample loading buffer	12 g Glycerol, 3 ml diH_2O, 10 ml 10% SDS, 1 ml 1 M Tris-HCL pH 7.2, 0.06 g Bromophenol Blue, 3% v/v ß-mercaptoethanol
Coomassie brilliant blue staining solution	45% (v/v) methanol, 0.25% (w/v) Brilliant blue R (Sigma), 1% (v/v) acetic acid
KPi Buffer	200 mM NaCl, 50 mM Potassium Phosphate buffer pH 7.8
LPS Blocking Buffer	5% (w/v) Milk in PBS-T
LPS sample buffer	60 mM Tris-HCL, 1 mM EDTA, pH 6.8
Ni^A-Elution buffer	20 mM Tris , 300 mM NaCl, 40 mM Imidazole pH7.5
Ni^A-Equilibration buffer	20 mM Tris , 300 mM NaCl, 12 mM Imidazole pH 7.5
Ni^A-Wash buffer	20 mM Tris , 300 mM NaCl, 40 mM Imidazole pH 7.5
NPI-10 lysis buffer	50 mM NaH_2PO_4, 100 mM NaCl, 10 mM Imidazole, pH 8.0
NPI-20 wash buffer	50 mM NaH_2PO_4, 100 mM NaCl, 20 mM Imidazole, pH 8.0
NPI-500 elution buffer	50 mM NaH_2PO_4, 100 mM NaCl, 500 mM Imidazole, pH 8.0
PBS	5 g Gibco® PBS Tablet (Fisher Scientific CAT#18912014) dissolved in 500 mL of distilled water. (pH 7.45)
PBS-T	1% (v/v) Tween® 20 (Sigma) in PBS
Protein Blocking Buffer A	3% (w/v) BSA, 20 mM Tris-HCL pH 8, 150 mM NaCl
Protein Blocking Buffer B	5% (w/v) Milk, 20 mM Tris-HCL pH 8, 150 mM NaCl
Salt homogenising buffer	0.4 mM NaCl, 2 mM EDTA, 10 mM Tris-HCL pH 8.0
Sucrose solution	25% w/v sucrose, 5 mM EDTA, 20 mM Tris-HCL pH 8.0
TBS buffer	20 mM Tris, 300 mM NaCl pH 7.5
TBS^{IM}	20 mM Tris-HCL pH 8, 150 mM NaCl
TBS-T	20 mM Tris-HCL pH 8, 150 mM NaCl, 1% (v/v) Tween® 20
TE	10 mM Tris, 0.1 mM EDTA pH 7.0

Transfer Buffer	0.3% (w/v) Tris, 1.44 % (w/v) Glycine, 20% Methanol

2.4 In silico analysis

2.4.1 Alignments

TCoffee (Notredame *et al.*, 2000) with default settings was used for all multiple sequence alignments in Chapter 3 and for alignment of $OafA_{STM}$, $OafB_{STM}$ and $OafB_{SPA}$ protein sequences for direct comparison (Chapter 5).

A survey of the literature identified 30 experimentally-characterised bacterial carbohydrate acetyltransferases, these sequences were aligned along with OafB from *Salmonella* ser. Paratyphi A, using TCoffee. Protein details and their accession numbers are listed in Appendix I. All TCoffee alignments were conducted using the Jalview workbench (Waterhouse *et al.*, 2009).

Structure based sequence alignment was conducted by Sarah Tindall using PROMALS-3D with default settings. The two closest structural homologues to the crystal structure of the periplasmic domain of $OafB_{SPA}$ (6SE1 = OafB_SPA) were identified using the DALI server (5B5S, 2VPT). These structures were aligned along with a selection of typical SGNH domains for which structural information is available (1IVN = TAP1_E.col, 4K40 = Ape1_N.men, 1DEX = RGAE_A.acu) and the crystal structure of the SGNH domain of OatA (5UFY = OatA_S.pne). Five further representative sequences of OafA, OafB, and OatA were included in the structure-based sequence alignment (WP_00400612 = OafB_SPA, SIU02679 = OafB_STM, WP_000639473 = OafA_STM, AAX87447 = OafA_HI, Q2FV54 = OatA_Saur).

2.4.2 Phylogenetic Trees

To produce a phylogenetic tree of the AT3 domains of O-acetyltransferase proteins, sequences from the TCoffee alignment of 31 AT3 domain containing acetyltransferases (Appendix I) were taken and the C-terminal sequence representing the SGNH domain in AT3-SGNH fused proteins was removed up to the beginning of TMH11 (335_OafA$_{STM}$). These truncated sequences were re-aligned with TCoffee.

The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model (Jones *et al.*, 1992) with assistance from Rebecca Hall (University of York). The bootstrap consensus tree inferred from 100 replicates is taken to represent the evolutionary history of the taxa analysed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches (Felsenstein, 1985). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. This analysis involved 31 amino acid sequences. There were a total of 742 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar *et al.*, 2018).

2.4.3 Measurement of atom distance in Coenzyme A molecule

PDB files for crystal structures of proteins complexed with CoA ligands (1H16, 2WL4, 5G17, 5TVJ) as well as the ideal coordinates for a Coenzyme A from PDB chemical component COA (RCSB, 1999) were analysed using the PyMOL molecular graphics system (version 2.3.0, Schrödinger, LLC). The measurement wizard was used to select P3B and S1P atoms in the structure to record the distance between them.

2.5 General cloning techniques

2.5.1 Polymerase chain reaction

Q5, Phusion, and GoTaq G2 flexi DNA polymerase enzymes were used for PCR. The reaction and thermocycler conditions are listed in **Table 2.5**, **Table 2.6** and **Table 2.7**.

Table 2.5| Q5 High Fidelity DNA polymerase (NEB) PCR reaction mixture and thermocycling conditions.

Component	Volume (50 µl reaction)
5x Q5 Reaction Buffer	10 µl
10 mM dNTPs (Thermo)	1 µl
10 µM Forward Primer	2.5 µl
10 µM Reverse Primer	2.5 µl
Template DNA: Genomic DNA = 1 ng/µl Plasmid DNA = 25	1-2.5 µl
Q5 Hot startHF DNA polymerase (2 U/µl)	0.5 µl
Nuclease free water	Reaction volume to 50 µl
Thermocycler: 98 °C for 30 sec, 35x (98 °C for 10 sec, A °C for 30 sec, 72 °C for X sec), 72 °C for 5 min 'A' indicates annealing temperature and was set depending on the annealing temperature of the primers in use. 'X' indicates extension time and was calculated based on the size of the fragment to be amplified at 20-30 sec per Kb	

Table 2.6| Phusion High Fidelity DNA polymerase (Thermo Scientific) PCR reaction mixture and thermocycling conditions.

Component	Volume (50 µl reaction)
5x Phusion HF buffer	10 µl
10 mM dNTPs (Thermo)	1 µl
10 µM Forward Primer	2.5 µl
10 µM Reverse Primer	2.5 µl
Template DNA: Genomic DNA = 1 ng/µl Plasmid DNA = 25 ng/µl	1-2.5 µl
Phusion HF DNA polymerase 2 U/µl	0.5 µl
Nuclease free water	Reaction volume to 50 µl

Thermocycler:

98 °C for 30 sec, 35x (98 °C for 10 sec, A °C for 30 sec, 72 °C for X sec), 72 °C for 5 min

'A' indicates annealing temperature and was set depending on the annealing temperature of the primers in use. 'X' indicates extension time and was calculated based on the size of the fragment to be amplified at 20-40 sec per Kb

Table 2.7| GoTaq® G2Flexi DNA polymerase (Promega) PCR reaction mixture and thermocycling conditions.

Component	Volume (50 µl reaction)
5x Green GoTaq Flexi Buffer	10 µl
25 mM $MgCl_2$	3 µl
10 mM dNTPs (Thermo)	1 µl
10 µM Forward Primer	1 µl
10 µM Reverse Primer	1 µl
Template DNA: Genomic DNA = 1 ng/µl Plasmid DNA = 25 ng/µl	1-2.5 µl
GoTaq Flexi G2 DNA polymerase 5 U/µl	0.5 µl
Nuclease free water	Reaction volume to 50 µl
Thermocycler: 95 °C for 3 min, 35x (95 °C for 30 sec, A °C for 30 sec, 72 °C for X sec), 72 °C for 5 min 'A' indicates annealing temperature and was set depending on the annealing temperature of the primers in use. 'X' indicates extension time and was calculated based on the size of the fragment to be amplified at 1 min per Kb	

2.5.2 Agarose gel electrophoresis

DNA samples were mixed with 6x DNA loading dye (Thermo Fisher) to a final concentration of 1x loading dye. For general sample analysis, prepared samples were loaded into the wells of 1% agarose gel, made with 1X TAE buffer from 10X stock (Thermo Fisher: 40 mM Tris-acetate, 1 mM EDTA at pH 8.3). 0.5 µg/mL Ethidium Bromide or 1 µg/mL Nancy-520 DNA gel stain (Sigma-Aldrich) was incorporated into agarose gels for DNA visualisation by UV or Blue light, respectively. For gel extraction, 0.8% agarose gels were used with Nancy-520 DNA gel stain. Gels were run in 1x TAE buffer on the RunOne™ horizontal electrophoresis system at 100 V as standard or 25 V for gel extraction.

2.5.3 Preparation of genomic DNA

1.5 ml stationary phase bacterial culture was centrifuged at 17,000 xg for 2 min. Supernatant was removed and the cell pellet re-suspended in 400 µl of salt homogenising buffer (**Table 2.4**) to which 2% (w/v) SDS and 200 µg/ml proteinase K (Promega) was added. Cells were incubated at 65 °C overnight. 300 µl 5M NaCl was added and the solution vortexed at maximum speed (3200 RPM) for 30 sec using Scientific Industries SI™ Vortex-Genie™ 2 then centrifuged at 10,000 xg for 30 min. Supernatant was transferred to a fresh tube and an equal volume of isopropanol added then incubated at -20 °C for 1 hr. The mixture was centrifuged at 10,000 xg for 20 min at 4 °C. The pellet was washed with 70% ethanol then re-suspended in 100 µl sterile TE (**Table 2.4**) after residual ethanol had evaporated. Genomic DNA isolates were stored at 4 °C until further use.

2.5.4 Preparation of plasmid DNA

Plasmid DNA was isolated from stationary phase bacterial cultures using the NEB Monarch miniprep kit (NEB) following the manufacturer's instructions. Plasmid DNA was eluted with ddH_2O for the final elution step.

2.5.5 Clean up of PCR products for cloning

Reaction components and buffers were removed from PCR products using Qiagen PCR clean-up kit following the manufacturer's instructions. Elution from the filter was carried out with ddH_2O.

2.5.6 Gel extraction

For isolation of a specific DNA fragment for cloning, the DNA sample or PCR products were subject to agarose gel electrophoresis (see section 2.5.2). The required DNA band was visualised using the BioRad ChemiDoc MP imaging system and cut out of the gel. The excised band was subject to gel extraction using the Qiagen gel extraction kit following manufacturer's instructions. Elution from the filter was carried out with ddH_2O.

2.5.7 DNA Quantitation

Concentration of DNA samples was quantified using a NanoDrop™ spectrophotometer (Thermo Scientific). This measures the absorbance of the sample at 260 nm and the purity of DNA through calculating the 260/280 and 260/230 absorbance ratios. 280 nm absorbance can indicate presence of protein, phenol or other contaminants that absorb light at 280 nm and a ratio of ~1.8 is usually considered pure. 230 nm absorbance can indicate organic compounds, salts or other contaminants that absorb light at 230 nm. A ratio of ~2 is usually considered pure.

2.5.8 Ligation independent cloning

Ligation independent cloning (LIC) to produce the OafA and OafB expression plasmids (PMV433 and PMV434, **Table 2.2**) was carried out by Reyme Herman. Methods are adapted from the Master's thesis of Reyme Herman. Briefly, PCR using Q5 High Fidelity DNA polymerase (NEB) (**Table 2.5**) was carried out using genomic DNA (2.5.3) from respective STM strains encoding OafA and OafB (Path 346 and Path 189, **Table 2.1**). LIC primers were designed to amplify the *oafA* and *oafB* genes and add long tails of complementary sequence to the LIC cassette of the pBADcLIC vector at the *Swa*I restriction site. These long tails of complementary sequence allow insert and plasmid DNA to anneal following generation of complementary overhangs to produce the complete expression vector without need for ligation.

The pBADcLIC2005 plasmid was digested with *Swa*I restriction enzyme (NEB) following manufacturer's instructions. To expose the complementary LIC tails of the plasmid and insert DNA the 3'-5' exonuclease activity of T4 DNA polymerase (NEB) was used. 10 µl [200 ng] insert or plasmid DNA were mixed with 3 µL of NEBuffer 2 (NEB), 25 mM of dCTP for the plasmid DNA or dGTP for the PCR products (Fermentas), and 0.5 µL of T4 DNA Polymerase (NEB). The mixture was incubated for 30 min at 20 °C then the reaction stopped by incubation at 75 °C for 20 min.

T4 DNA polymerase treated samples were mixed at a molar ratio of 1:2 plasmid:insert and allowed to anneal at 20 °C for 10 min before 25 mM EDTA was added for a further 10 min incubation at 20 °C. The annealing mixture was then transformed into the relevant bacterial strain by electroporation or heat shock (2.5.11).

Table 2.8| Primers designed for LIC of OafA and OafB into the pBADcLIC2005 vector by Reyme Herman. Details of Path 346 and Path 189 can be found in **Table 2.1**.

Primer Name	Sequence 5'-3'	Use
OMV1074	ATG GGT GGT GGA TTT GCT ATG ATC TAC AAG AAA TTC AGA C	Forward primer for LIC cloning of OafA from Path 346 genomic DNA
OMV1075	TTG GAA GTA TAA ATT TTC TTT TGA AAT CTG CTT TTT CAC	Reverse primer for LIC cloning of OafA from Path 346 genomic DNA
OMV1077	ATG GGT GGT GGA TTT GCT ATG GAA CAC TTA AAA TAC AGA C	Forward primer for LIC cloning of OafB from Path 189 genomic DNA
OMV1079	TTG GAA GTA TAA ATT TTC TCT TAT TAT CAA ATG CCC TAT C	Reverse primer for LIC cloning of OafB from Path 189 genomic DNA

2.5.9 Mutagenesis by inverse PCR and blunt end ligation

Primers were designed to amplify plasmid pMV433 (**Table 2.2**) whilst also introducing a specific mutation to the OafA sequence. The mutation was designed to be in the middle of the forward mutagenic primer to maximise template plasmid binding. See **Table 2.9** for details of mutagenic primer pairs. Mutagenesis PCR was carried out with either Q5 High Fidelity DNA polymerase (NEB) (**Table 2.5**) or Phusion High Fidelity DNA polymerase (Thermo Scientific) (**Table 2.6**). PCR products were digested with *Dpn*I (New England Biolabs) following manufacturer's instructions, to digest template DNA, then purified by gel extraction (2.5.6). Linear products were phosphorylated and ligated with polynucleotide kinase and T4 DNA ligase (New England Biolabs), respectively. The resulting OafA point mutant expression strains and their plasmid numbers are listed in **Table 2.10**.

Table 2.9| Primers for mutagenesis of *oafA* in the pBADcLIC2005 plasmid (**Table 2.2**) by inverse PCR and blunt end ligation. Mutated codon sequences are underlined.

OafA Mutant	F Primer	Sequence 5'-3'	R Primer	Sequence 5'-3'
R14A	oMV1403	AAT GGG CTA GCG GCT TTT G	oMV1404	TAT ATC GAG TCT GAA TTT CTT GTA GAT CAT AGC
H25A	oMV1407	GTG CTG TAT GCG TTC GGT GT	oMV1408	AAC ACT TAT CAA TGC AAA AGC TCT TAG CC
S32A	oMV1399	CT TAT GTG GCG GGT GG	oMV1400	GTACACCGAAGTGATACAGC A
G33A	oMV1409	CT TAT GTG TCA GCG GGC TTT	oMV1410	G TAC ACC GAA GTG ATA CAG CAC
G34A	oMV1382	GTG TCA GGT GCG TTT ATA GGT	oMV1383	ATA AGG TAC ACC GAA GTG ATA CAG
F35A	oMV1411	G TCA GGT GGC GCG ATA GG TG	oMV1412	ACA TAA GGT ACA CCG AAG TGA TAC AGC AC
G37A	oMV1388	GGC TTT ATA GCG GTA GAT GTT TTC	oMV1389	ACC TGA CAC ATA AGG TAC ACC
V38A	oMV1390	C TTT ATA GGT GCG GAT GTT TTC TTT	oMV1391	CCACCTGACACATAAGGTAC A
D39A	oMV1392	C TTT ATA GGT GTA GCG GTT TTC TTT GTA	oMV1391	CCACCTGACACATAAGGTAC A
V40A	oMV1393	GGT GTA GAT GCG TTC TTT GTA ATT TCT	oMV1394	TAT AAA GCC ACC TGA CAC ATA AGG TA
S45A	omv1505	C TTT GTA ATT GCG GGT TTT CTT ATG AC	oMV1506	AAAACATCTACACCTATAAA GCCACCT
G46A	oMV1395	GTA ATT TCT GCG TTT CTT ATG ACT G	oMV1396	AAA GAA AAC ATC TAC ACC TAT AAA GCC
R69A	oMV1553	GAT TTT TAT ATT GCA GCG TTC CTA AG	oMV1554	AAG TAC TCC TTT GTG GTC TAC
R72A	oMV1555	T ATT GCA AGA TTC CTA GCG ATT GTA	oMV1556	TAA AAA TCA AGT ACT CCT TTG TGG TC
S112A	oMV1507	G TCT TTA CTT TTT TAT GCG AAT AAT TA	oMV1508	GATATTGCGTTCTTGCTAAGT G
N113A	oMV1533	A CTT TTT TAT TCA GCG AAT TAT TAC G	oMV1534	AAGACGATATTGCGTTCTTG C
Y122A	oMV1509	CAC TCT AGT GCG TTC GAC TCA	oMV1510	AATTGCGTAATAATTATTTGA ATAAAAAAGTA

	F Primer	Sequence 5'-3'	R Primer	Sequence 5'-3'
G202A	oMV1415	GAA ATG CTG GCT GCG GGC	oMV1416	CCA TGC CCT GGT AGG GAT AAG ATA GA
E325A	oMV1475	C AGA ACA ATT GCG AAC ACG CT	oMV1476	TATGAAATATCCCCAAGCGC AAAAG
C383S	oMV1332	AGG CCT GAT ATT AGC TTC CTC AAT CCA	oMV1333	CCA GGG AGA GTT GTC CAT ACG ATA
C397S	oMV1351	GCA TTC TCA AAA AGC CAG GAT AAA TAG	oMV1352	TGA ATA ATC TTG ATC TGG ATT GAG GAA
C439S	oMV1336	GCA AGC TTG AGC CAA CCA AT	oMV1337	AGT TCT CTG CGT AAT GTT AAG TGA ATT TCC
C453S	oMV1334	GAC AGG CCG TAT AGC AAA GAC ATC AAT	oMV1378	ATC TTT TTG AAG CCC AAT GAT TGG TGG
C567S	oMV1353	CTT GAG ACT ATG AGC ACA GAA AGT TAT	oMV1366	TGG TGA AAT ATA AGT AAG GGA ATG CTC
C572S	oMV1355	GT ACA GAA AGT TAT AGC AAA GCA ATA ATA	oMV1356	CAT AGT CTC AAG TGG TGA AAT ATA AG
S437A	oMV1491	CAG AGA ACT GCA GCG TTG TGC	oMV1492	CGTAATGTTAAGTGAATTTCC AAATACCGATTTC
E569A	oMV1473	CT ATG TGT ACA GCG AGT TAT TGC	oMV1474	TCT CAA GTG GTG AAA TAT AAG TAA GGG
S412A	oMV1152	ATG GGG TGA CGC GCA TGC CGC	oMV1153	ACA ACA AAA GAC TTT TCA GTC ATT TTA TCC TGA CAT TTT G
D587A	oMV1150	ATA CAA TAT GCC AAT GCG CAC C	oMV1151	AGG GTA AGC AAT TCT ATT CC
H590A	oMV1148	T GAC AAT GCG GCC CTA ACA CCA GAA G	oMV1149	TAT TGT ATA GGG TAA GCA ATT C

Table 2.10| STM strains for OafA point mutant functional analysis. Details of LT2 basal O-antigen strain can be found in **Table 2.1**.

Strain #	Plasmid #	Strain details
Path 993	432	LT2 Basal O-ag + pBADcLIC2005
Path 932	433	LT2 Basal O-ag + pBADcLIC2005_WT-OafA
Path 1023	509	LT2 Basal O-ag + pBADcLIC2005_R14A-OafA
Path 1022	508	LT2 Basal O-ag + pBADcLIC2005_H25A-OafA
Path 1009	501	LT2 Basal O-ag + pBADcLIC2005_S32A-OafA
Path 1024	510	LT2 Basal O-ag + pBADcLIC2005_G33A-OafA

Strain #	Plasmid #	Strain details
Path 1014	506	LT2 Basal O-ag + pBADcLIC2005_G34A-OafA
Path 1025	511	LT2 Basal O-ag + pBADcLIC2005_F35A-OafA
Path 1026	512	LT2 Basal O-ag + pBADcLIC2005_I36A-OafA
Path 1010	502	LT2 Basal O-ag + pBADcLIC2005_G37A-OafA
Path 1011	503	LT2 Basal O-ag + pBADcLIC2005_V38A-OafA
Path 1012	504	LT2 Basal O-ag + pBADcLIC2005_D39A-OafA
Path 1015	507	LT2 Basal O-ag + pBADcLIC2005_V40A-OafA
Path 1035	521	LT2 Basal O-ag + pBADcLIC2005_S45A-OafA
Path 1013	505	LT2 Basal O-ag + pBADcLIC2005_G46A-OafA
Path 1085	540	LT2 Basal O-ag + pBADcLIC2005_R69A-OafA
Path 1086	541	LT2 Basal O-ag + pBADcLIC2005_R72A-OafA
Path 1036	522	LT2 Basal O-ag + pBADcLIC2005_S112A-OafA
Path 1078	539	LT2 Basal O-ag + pBADcLIC2005_N113A-OafA
Path 1037	523	LT2 Basal O-ag + pBADcLIC2005_Y122A-OafA
Path 1027	513	LT2 Basal O-ag + pBADcLIC2005_G202A-OafA
Path 1028	514	LT2 Basal O-ag + pBADcLIC2005_E325A-OafA
Path 1033	519	LT2 Basal O-ag + pBADcLIC2005_R379A-OafA
Path 1001	497	LT2 Basal O-ag + pBADcLIC2005_C383S,C397S-OafA
Path 1007	499	LT2 Basal O-ag + pBADcLIC2005_C439S,C453S-OafA
Path 1002	498	LT2 Basal O-ag + pBADcLIC2005_C567S,C572S-OafA
Path 1032	518	LT2 Basal O-ag + pBADcLIC2005_S437A-OafA
Path 1030	516	LT2 Basal O-ag + pBADcLIC2005_E569A-OafA
Path 933	452	LT2 Basal O-ag + pBADcLIC2005_S412A-OafA
Path 934	453	LT2 Basal O-ag + pBADcLIC2005_D587A-OafA
Path 1055	527	LT2 Basal O-ag + pBADcLIC2005_H590A-OafA

OafB point mutant H621A was produced using the same protocol by Reyme Herman using PMV434 (**Table 2.2**) as the template DNA with mutagenic primers OMV1154 (5' AGA TTG GGG CGC TTT GAC AAA GCC TG 3') and OMV1155 (5' ACA GCT GTA ACA AAA TCT G 3'). The resulting plasmid was recorded as PMV439 and the corresponding STM LT2 Basal O-ag + pBADcLIC2005_H621A-OafB is Path 892.

2.5.10 Restriction enzyme cloning

Restriction enzyme cloning was used to create OafA and OafB expression constructs in the pETFPP_30 vector (**Table 2.2**). Insert DNA was amplified from PMV433 for OafA and PMV434 by PCR (Section 2.5.1) using Q5 High Fidelity DNA polymerase (**Table 2.5**). Insert primers were designed to introduce appropriate restriction enzyme sites for insertion into the multiple cloning site of the pETFPP_30 vector.

Table 2.11| Primer sequences for restriction enzyme cloning of OafA and OafB constructs into the pETFPP_30 expression plasmid.

Primer Name	**Sequence (5'-3')**	**Function**	**Restriction site**
OMV1271	ACATTTGGCCATGAAAGGTGTTAGTTTTAGATTTTCAG	Forward OafA_Lys355	*Nco*I
OMV1272	ATATTTGGCCATGATGGACAACTCTCCCTGG	Forward OafA_Met373	*Nco*I
OMV1273	AATATTGGCCATGGAAAAGTCTTTTGTTGTATGGG	Forward OafA_Glu403	*Nco*I
OMV1266	GCATCTCGAGTTTTGAAATCTGCTTTTTCACTTC	Reverse All OafA	*Xho*I
OMV1259	TAGCCCATGGCAATGAATGGAATTAAAGAAAGAAGC	Forward OafB_Met367	*Mcs*I
OMV1260	ATGCCCATGGATGGTGAGCTATTGCGC	Forward OafB_Gly395	*Mcs*I
OMV1261	TTATCCATGGGCCGCAATAATATTTTTATAATCGG	Forward OafB_Arg421	*Mcs*I
OMV1262	TAATCTCGAGTCTTATTATCAAATGCCCTATCTTCT	Reverse All OafB	*Xho*I

PCR reaction products were cleaned up (2.5.5) then restriction digests were performed according to manufacturer's instructions (New England Biolabs) on insert and plasmid DNA. Each 50 µl restriction digest reaction contained up to 1 µg of DNA and 20U of restriction enzyme in the appropriate 1X restriction enzyme buffer. Restriction digest products were cleaned up (2.5.5) to remove cleaved DNA fragments.

Insert and plasmid DNA was ligated at 1:3 molar ratio with T4 DNA ligase (New England Biolabs) following the manufacturer's instructions. The ligation mixture was incubated overnight at 16 °C then 5 µl of ligation mix was transformed into a sub cloning *E. coli* strain (2.5.11.1). Colony PCR (2.5.12) was used to confirm correct insertion of the ligated plasmid before the plasmids were sent for sequencing.

2.5.11 Bacterial transformation

2.5.11.1 Heat shock

All plasmids were first subcloned into *E. coli* XL1-Blue competent cells (Agilent) (**Table 2.1**) following manufacturers instruction for heat shock transformation.

For other competent cells the following general heat shock protocol was followed.

50 µl aliquot of competent cells was thawed on ice for 30 min. 1 µl plasmid DNA was added and the cells were incubated on ice for a further 30 min. Cells were heat shocked at 42 °C for 1.5 min then 900 ml LB or SOC medium (**Table 2.3**) was added. Cells were incubated at 37 °C for 1 hr before 200 µl was spread on solid media with appropriate antibiotic selection for overnight incubation at 37 °C.

2.5.11.2 Electroporation

Stationary phase bacterial culture was diluted to 1:100 into 5 ml fresh LB and grown to OD_{600} 0.4-0.6. Cells were harvested at 3000 xg for 7 min at 4 °C and the remaining steps were completed at 4 °C and all centrifugations carried out at 3000 xg for 7 min. The cell supernatant was discarded, cells were washed by re-suspension in 1 ml 10% (v/v) glycerol, pelleting by centrifugation and removal of the supernatant. This was repeated twice more and after final centrifugation step the supernatant was discarded and cells re-suspended in 100 µl 10% (v/v) glycerol. 50 µl cell suspension were mixed with 5 µl of DNA sample and loaded into a chilled 0.2 cm gap MicroPulser electroporation cuvette (Bio-Rad) and pulsed with setting EC-2 of the MicroPulser electroporator (Bio-Rad). 1 ml SOC media (**Table 2.3**) was immediately added to cells and they were allowed to recover at 37 °C for 1 hour. Successfully recombined cells were selected for by plating on solid LB agar with appropriate antibiotic selection.

2.5.12 Colony PCR

Colony PCR was used as initial confirmation that the correct plasmid DNA had been transformed into bacterial cells before being sent for sequencing. 50 µl master mix of GoTaq Flexi G2 DNA polymerase reaction mix (**Table 2.7**), without template DNA, was set up for the number of colonies to be screened +2 (for positive and negative control). 50µl PCR master mix was aliquoted into 0.2 ml PCR tubes. Single colonies were picked from solid media with appropriate antibiotic selection and spotted onto fresh plates

before the remainder was re-suspended into the PCR master mix. The original template plasmid was used as a negative control and no DNA was added to the mastermix for the negative control. Following amplification with the thermocycler conditions in **Table 2.7** with 52 °C annealing temperature, PCR products were analysed by agarose gel electrophoresis (2.5.2) to confirm amplification products of the expected sizes.

2.5.13 DNA Sequencing

The OafA and OafB insert sequence of all expression plasmids were confirmed by sequencing by Eurofins Genomics using Mix2Seq sequencing tubes. Sequencing primers used to cover various regions of OafA and OafB sequence in different expression plasmids are listed in **Table 2.12**.

Table 2.12| Sequencing primers.

Primer Name	Primer sequence 3'-5'	Function
OMV1379	GCTTTATCCATTCTGGCATTTTGTA	Reverse sequencing primer starting at nucleotide C658 of OafA.
OMV966	GGCCAAGCTATTCAGCACTC	Forward sequencing primer starting at nucleotide G725 of OafA
OMV1160	CAAAAGTGTCTATAATCACGGCAG	Forward sequencing primer which binds upstream of the insertion site of the pBADcLIC2005 plasmid
OMV1161	CAGCTTGGCTGTTTTGGCGGAT	Reverse sequencing primer which binds downstream of the insertion site of the pBADcLIC2005 plasmid
OMV1269	TAATACGACTCACTATAGGG	Forward sequencing primer for inserts in the pETFPP_30 plasmid

OMV1270	TATGCTAGTTATTGCTCAGCGGT	Reverse sequencing primer for inserts in the pETFPP_30 plasmid

2.5.14 Lambda red recombination

STM LT2 with 10X-His tag added to the C-terminus of the chromosomal *oafA* gene (Path 1102, **Table 2.1**) was created using lambda red recombination. This process was carried out by Steinar Mannsverk under close supervision by the thesis author following the protocol for lambda red recombination described by Datsenko and Wanner (2000).

Primers were designed according to **Figure 2.1**, to amplify the kanamycin resistance cassette flanked by FRT sites from pDHL1092 plasmid, using priming site 1 and 4 (Datsenko and Wanner, 2000), and to add a 10xHis tag to the recombination site complementary to the C-terminal end of the *oafA* gene (**Table 2.13**). Phusion High Fidelity DNA polymerase PCR was used to amplify this fragment following the protocol in **Table 2.6**, using an annealing temperature of 68 °C.

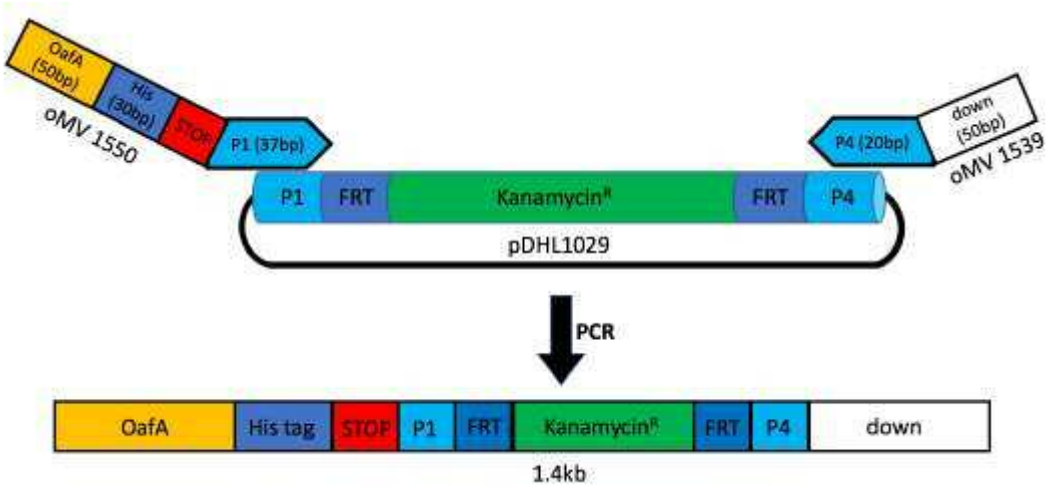


Figure 2.1| Lambda red recombination primer design for amplification of FRT flanked kanamycin resistance cassette from pDHL1029 to create the C-terminal OafA 10xHis tag chromosomal insert for lambda red recombination of STM LT2. Figure created by Steinar Mannsverk and reproduced with his permission.

Path 86 (STM LT2 transformed with the pKD46 plasmid, **Table 2.1**) was subject to electroporation with the 10xHis-FRT-KanR-FRT chromosomal insert fragment using the following protocol.

Stationary phase liquid culture of Path 86, grown at 30 °C, was diluted 1:100 into 50 ml fresh LB + 100 µg/ml Amp and grown to OD_{600} 0.4-0.6. 0.2% (w/v) L-arabinose was added 30-60 min before harvesting to induce expression of the lambda red recombineering proteins from the pKD46 plasmid. Cells were harvested at 3000 xg for 15 min at 4 °C and the remaining steps were completed at 4 °C and all centrifugations carried out at 3000 xg for 7 min. The cell supernatant was discarded and cells re-suspended in 20 ml 10% (v/v) glycerol then pelleted by centrifugation. Supernatant was discarded and cells re-suspended in 2 ml 10% (v/v) glycerol and pelleted by centrifugation and this process was repeated twice more. Following the final centrifugation step, the supernatant was discarded and cells were re-suspended in 100 µl 10% (v/v) glycerol. 50 µl cell suspension were mixed with 780 ng of the 10XHis insert PCR product and loaded into a chilled 0.2 cm gap MicroPulser electroporation cuvette (Bio-Rad) and pulsed with setting EC-2 of the MicroPulser electroporator (Bio-Rad). 1 ml SOC media (**Table 2.3**) was immediately added to cells and they were allowed to recover at 30 °C for 3 hr. Successfully recombined cells were selected for by plating on solid LB agar (**Table 2.3**) with 50 µg/ml Kan selection at 37 °C.

STM LT2 cells containing the successfully recombined 10xHis-FRT-KanR-FRT cassette were next transformed with the pCP20 plasmid (**Table 2.2**). The pCP20 plasmid encodes yeast FLP recombinase genes to excise the kanamycin resistance cassette flanked by FRT sites (Datsenko and Wanner, 2000) and was transformed using the above electroporation protocol with the following modifications. Stationary phase cultures were grown at 37 °C then diluted 1:100 in 25 ml fresh LB + 50 µg/ml Kan . This culture was grown at 37 °C to OD_{600} 0.4-0.6 without L-arabinose. 25 ml culture was centrifuged and all 10% (v/v) glycerol washing steps were carried out with 2 ml 10% (v/v) glycerol. Following the final resuspension of cells into 100 µl 10% (v/v) glycerol, 50 µl of cells were mixed with 140 ng pCP20 plasmid DNA before micropulsing. The cells were left to recover for 1.5 hr at 30 °C. Successfully transformed cells with their kanamycin resistance cassette excised by the LFP recombinase encoded on pCP20 were selected by first plating on solid LB agar with 100 µg/ml selection for the pCP20 plasmid at 30°C, then confirming loss of kanamycin resistance by lack of growth on LB 50 µg/ml Kan plates. The pCP20 plasmid was subsequently cured by growth on non-selective solid LB media at 40 °C. Excision of the KanR cassette was confirmed by colony PCR (section 2.5.12) using OMV1543 and OMV1548 (**Table 2.13**).

Table 2.13| Lambda red recombination primers.

Primer Name	Sequence (5'-3')	Function
OMV1550	CAGAAGGCTCAGGGTGGTTTATTGA GGAAGTGAAAAAGCAGATTTCAAAA CATCATCACCATCATCACCATCACCAT CATTAAGTGTAGGCTGGAGCTGCTTC GAAGTTCCTATACTTTC	Forward OafA C-terminus His tag + kanR priming site 1
OMV1539	GCATTATTGTTGTAGTTTTATAAAAT AAAAAGAGGGGCAAGCCCCTCTGT ATTCCGGGGATCCGTCGACC	Reverse OafA C-terminus + kanR priming site 4
OMV1543	AGTCCTCATTGTTGTTGCTGT	Forward primer for amplification of insert sequence from STM LT2 genome for sequencing
OMV 1548	GCGGTTGATACATCGGTTGC	Reverse primer for amplification of insert sequence from STM LT2 genome for sequencing
OMV1545	CCATGGAGCATGACGGATGA	Forward sequencing primers for insert confirmation
OMV1547	TTCAGTGACAACGTCGAGCA	
OMV1549	GGACCGCTATCAGGACATAG	Reverse sequencing primer for insert confirmation
OMV1546	CAGCAAGCGAACCGGAATTG	

2.6 *In situ* functional analysis of OafA

In situ functional analysis of OafA was carried out using strains outlined in **Table 2.10** or strains Path 346, Path 293 and Path 1102 listed in **Table 2.1**. Unless otherwise stated during assay optimisation in Chapter 4 , strains for *in situ* functional analysis were cultured at pH 7.0 in 100 mM sodium phosphate-buffered LB at 37 °C in a baffled conical flask with shaking at 200 rpm. Overnight cultures were diluted 100-fold and grown for 16 hr. Samples were normalised to (OD_{600}) of 3.0 per ml for LPS and protein sample preparation.

2.6.1 Slide agglutination

1 ml of stationary phase culture was pelleted by centrifugation at 3000 xg for 10 min and re-suspended in PBS (**Table 2.4**) to a milky suspension (200-600 µl). On microscope slides with two etched 10 mm rings (Thermo Scientific – Gold Seal) one drop of test antibody was added to one ring and one drop of PBS added to the other as a negative control. One drop of bacterial suspension was added to each of the test circles and the slide rocked for 30-60 sec and inspected for agglutination of cells. If cells auto agglutinated in the negative PBS control the test was discarded and repeated.

2.6.2 Functional analysis of OafA by LPS immunoblot

2.6.2.1 Crude LPS sample preparation

The method was adapted from (Davies *et al.*, 2013). Unless otherwise stated during assay development in Chapter 4, the following protocol was used.

1 ml of OD-normalised (OD_{600} 3.0) overnight culture was pelleted for 5 min at 16,000xg. Cell pellets were re-suspended in 100 µl LPS sample buffer (**Table 2.4**) containing 2% (w/v) SDS then boiled at 100 °C for 5 min. Samples were vortexed following each further step of preparation. 400 µl of LPS sample buffer was then used to dilute the solution before 2 µl RNAse (Roche) and DNAse (Sigma) was added and the solution incubated at 37 ˚C for 16 hr. Samples were then treated with 100 µg proteinase K (Promega) for 16 hr at 50 ˚C. All samples were stored at 4 °C until analysis and vortexed prior to use.

2.6.2.2 T-SDS PAGE of LPS samples

Tricine Sodium Dodecyl Sulphate Poly Acrylamide Gel Electrophoresis (T-SDS PAGE) gels were cast in the Mini-PROTEAN™ casting stand with 1 mm gel plates (BioRad). Separating and stacking gel were made up according to **Table 2.14**. Separating gel was poured and allowed to polymerise in the casting stand under a layer of 100% ethanol. The ethanol was removed and gel washed with diH_2O before the stacking gel was poured and allowed to polymerise with the appropriate 1 mm comb fitted in the top. Gels were stored at 4 °C until use. LPS samples were run in the Mini-PROTEAN™ electrophoresis cell (BioRad) in 1x SDS sample loading buffer (**Table 2.4**), with anode and cathode buffer (**Table 2.14**) in the corresponding chamber. Samples were run at 50V for 15 min followed by 100V for ~1.5 hr and at least one lane was loaded with 5 µl of the protein marker blue pre-stained protein standard broad range (NEB). LPS bands were either visualised by

silver staining (2.6.2.3.2) or Immunoblotting (2.6.2.3.1) techniques. To ensure sharp bands in silver stain gels (2.6.2.3.2), acrylamide/bis solution should be stored tightly closed at 4 °C and used within 2 months of opening.

Table 2.14| T-SDS-PAGE Gel components.

Buffer Name	**Components**
10% Separating Gel	10% (v/v) Acrylamide/Bis solution [37.5:1] (BioRad), 1 M Tris-HCL pH 8.45, 1% (w/v) SDS, 10% (v/v) Glycerol, 0.6% (w/v) APS, 0.06% (v/v) TEMED
4% Stacking Gel	4% (v/v) Acrylamide/Bis solution [37.5:1] (BioRad), 750 mM Tris-HCL pH 8.45, 0.75% (w/v) SDS, 0.8% APS, 0.08% TEMED
Anode Buffer	0.2 M Tris-HCL pH 8.9
Cathode Buffer	0.1 M Tris-HCL, 0.1 M Tricine, 0.1 % (w/v) SDS

2.6.2.3 Visualisation of LPS in T-SDS-PAGE gels

7.5 µl of crude LPS extracts (2.6.2.1) were run on 1.0 mm T-SDS PAGE (2.6.2.2) for analysis.

2.6.2.3.1 LPS Immunoblotting

Optimised protocol:

The TSDS-PAGE-separated LPS samples were transferred onto Immobilon-P PVDF membrane (Merck-Millipore) using the Trans-Blot® Semi-Dry Transfer Cell (BioRad). PVDF membrane was activated in methanol for 15 sec, washed 3x with diH_2O then equilibrated in transfer buffer (**Table 2.4**). The SDS-PAGE gel was equilibrated in transfer buffer and filter paper soaked in transfer buffer was stacked with the equilibrated membrane and gel according to manufacturer's instructions and the system was run at 20 V for 15 min. PVDF membrane was blocked in LPS Blocking Buffer (**Table 2.4**) for 1 hr at room temperature. The membrane was then incubated in primary O:5 serotyping antibody (1:10000) (Statens Serum Institute; 40272) and *Salmonella* core antigen (1:200) (Insight Biotechnology; 5D12A) in LPS Blocking Buffer (**Table 2.4**) for 1 hr at room temperature. Following primary antibody incubation, the membrane was washed 3 times in PBS-T for 10 min (**Table 2.4**) then incubated in secondary Goat Anti-Rabbit IgG StarBright Blue700 (1:5000) (Bio-Rad) and Goat anti-mouse IgG (H+L) DyLight 800

(1:5000) with LPS Blocking Buffer (**Table 2.4**) for 1 hr at room temperature. Following 4 10 minute washes in PBS-T, fluorescent LPS bands were visualised and imaged on the ChemiDoc MP imaging system (BioRad).

Pre-optimisation protocol:

Prior to optimisation of the protocol for LPS immunoblotting only single antibody immunoblotting was used with primary serotyping antibodies of O:5, O:4, O:12 and OMA (Statens Serum Institute) diluted 1:10000 in LPS blocking buffer. Goat anti-mouse IgG-HRP (Sigma-Aldrich) diluted 1:10000 with LPS blocking buffer (**Table 2.4**) was used as the secondary antibody and following secondary antibody incubation for 1 hr at room temperature and 4x 10 min membrane wash steps in PBS-T, LPS bands were visualised using Luminata Classico Western HRP substrate (Merck-Millipore) and imaged on the ChemiDoc MP imaging system (BioRad) or using X-ray film (GE Healthcare, Amersham™ Hyperfilm™ ECL).

Dot Blot:

For dot blot immunoblotting described in Chapter 4, LPS samples were spotted onto nitrocellulose membrane (Amersham). The nitrocellulose membrane was incubated in TBS-T for 10 min then mounted on top of filter paper for the LPS to be spotted onto the membrane. Following 1hr incubation at room temperature to allow the nitrocellulose membrane to fully dry, the protocol for LPS immunoblotting was followed (2.6.2.3.1) from the initial 1hr LPS Blocking Buffer incubation step.

For dot blot immunoblotting described in Chapter 5, 5 µl of LPS reaction samples from the *in vitro* acetyltransferase activity assay (2.8.4) were loaded onto nitrocellulose membrane using a BioRad Bio-Dot® microfiltration apparatus. Briefly, nitrocellulose membrane (Amersham) was incubated in TBS (**Table 2.4**) for 10 min then mounted into the BioRad Bio-Dot® microfiltration apparatus according to manufacturer's instructions. 5 µl LPS samples were mixed with 145 µl TBS and loaded into relevant wells of the apparatus then allowed to bind to the membrane under gravity for 1 hr. 100 µl TBS was then added to wells and allowed to flow through under gravity before the vacuum was applied and wells were washed twice with 300 µl TBS. The membrane was removed from the apparatus following manufacturer's instructions then the protocol for LPS detection

with O:5 serotyping antibodies and *Salmonella* core antigen was then followed as per LPS immunoblotting (2.6.2.3.1).

2.6.2.3.2 LPS Silver stain

Silver staining methods are adapted from (Kittelberger and Hilbink, 1993). All solutions were made up with ddH_2O and are listed in (**Table 2.15**). All incubation steps were carried out at room temperature with gentle agitation.

LPS samples in TSDS-PAGE gel (2.6.2.2) were fixed overnight in fixative solution. Fixative solution was discarded and gels were incubated in oxidiser solution for 10 min. Oxidiser solution was discarded and the gels were subject to 3x 15 minute washes in ddH_2O. Following the final wash step, H_2O was discarded and gels were incubated in silver solution for 30 min wrapped in foil to prevent light exposure. Silver solution was discarded and gels quickly rinsed with ddH_2O before developer solution was added. Gels were incubated in developer solution until LPS bands were adequately stained then the developer reaction was stopped by washing in 1% (v/v) Acetic acid.

Table 2.15| Silver staining solutions. All solutions are made fresh immediately prior to use.

Solution	Components
Fixative	30% (v/v) Ethanol, 10% (v/v) Acetic acid
Oxidiser	0.7% (w/v) Periodic acid, 30% (v/v) Ethanol, 10% (v/v) Acetic acid.
Silver	0.1% (w/v) Silver nitrate
Developer	3% (w/v) Sodium carbonate, 0.02% (v/v) Formaldehyde

2.6.2.4 Confirmation of OafA protein expression

Insoluble protein fractions were isolated from test strains using Bug Buster™ solution (Novagen). 1 ml of OD-normalised (OD_{600} 3.0) overnight culture was pelleted for 5 min at 16,000xg. The supernatant was discarded and the cell pellet re-suspended in 50 µl Bug Buster™ solution (Novagen). 1 µl DNase I (Thermo Fisher) was added to the suspension and this was incubated on a rocker at room temperature for 20 min. Lysed cells were centrifuged at 16,000 x g for 15 min. Supernatant (soluble fraction) was removed and the pellet re-suspended in 75 µl 4x SDS sample loading buffer. The re-

suspended pellet was incubated at 60 °C for 10 min then centrifuged at 16,000 x g for 10 min. 10 µl of supernatant from the insoluble fraction was analysed by SDS-PAGE and Western blotting (2.7.2, 2.7.3).

2.7 Recombinant protein expression and purification

2.7.1 Protein production in BL21 (DE3) cells

OafA and OafB protein constructs were cloned into the pETFPP_30 plasmid (**Table 2.2**) by restriction enzyme cloning (2.5.10). This plasmid adds an N-terminal PelB leader sequence to target protein expression to the periplasm and a C-terminal 6xHis tag to allow antibody detection and affinity purification of the expressed protein. Chemically competent *E.coli* BL21(DE3) (**Table 2.1**) were then transformed with the constructed expression plasmids by heat shock transformation (2.5.11.1). Protein expression strains are listed in **Table 2.16**. The OafB point mutant H621A produced by Reyme Herman in plasmid PMV 439 was used as the template for restriction enzyme cloning of OafB_Arg421^{H621A}.

Table 2.16| OafA and OafB protein expression strains. *E. coli* expression strain details can be found in **Table 2.1** and expression plasmid details in **Table 2.2**.

Strain	Expression Plasmid Info
BL21(DE3) pMV463	pETFPP_30
BL21(DE3) pMV467	pETFPP_30 OafB_Gly395
BL21(DE3) pMV471	pETFPP_30 OafB_Met367
BL21(DE3) pMV468	pETFPP_30 OafB_Arg421
BL21(DE3) pMV475	pETFPP_30 OafB_Arg421^{H621A}
BL21(DE3) pMV472	pETFPP_30 OafA_Lys355
BL21(DE3) pMV473	pETFPP_30 OafA_Met373
BL21(DE3) pMV474	pETFPP_30 OafA_Glu 403

2.7.1.1 Small scale protein expression trials for OafA and OafB

Single colonies were inoculated into 5 ml LB media (**Table 2.3**) with appropriate antibiotic selection and incubated at 37 °C with 250 rpm shaking overnight. Overnight cultures were inoculated into fresh media to OD_{650} 0.02 and cultured at the selected temperature with 250 rpm shaking. Protein expression was induced with the selected

IPTG concentration and cultures were returned to the conditions for the selected incubation time before samples were harvested.

2.7.1.2 Large scale production of OafA Lys355

Single colonies of BL21(DE3) pMV472 (**Table 2.16**) were inoculated into 20 ml LB media (**Table 2.3**) with 100 µg/ml Amp selection and incubated at 37 °C with 250 rpm shaking overnight. Overnight cultures were inoculated into 1L fresh media to OD_{650} 0.02 and cultured at 30 °C with 250 rpm shaking. Protein expression was induced with 0.1 mM IPTG once cells reached OD_{650} ~0.4 and culturing was continued for 16 hr (overnight) before cells were harvested and subject to periplasmic protein extraction (2.7.7).

2.7.2 Sodium Dodecyl Sulphate Poly Acrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE gels were cast in the Mini-PROTEAN™ casting stand with 1 mm gel plates (BioRad). Separating and stacking gel were made up according to **Table 2.17**. Separating gel was poured and allowed to polymerise in the casting stand under a layer of 100% ethanol. The ethanol was removed and gel washed with diH_2O before the stacking gel was poured and allowed to polymerise with the appropriate 1 mM comb fitted in the top. Gels were stored at 4 °C until use. Protein samples were run in the Mini-PROTEAN™ electrophoresis cell (BioRad) in 1x SDS sample loading buffer (**Table 2.4**) at 200V for 50 min at least one lane was loaded with 5 µl of the protein marker blue prestained protein standard broad range (NEB). Protein bands were either visualised by Coomassie staining or Immunoblotting techniques.

Table 2.17| SDS-PAGE Gel components and running buffer.

Buffer Name	Components
12% Separating Gel	12% (v/v) Acrylamide/Bis solution [37.5:1] (SLS), 375 mM Tris-HCL pH 8.8, 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.04 % (v/v) TEMED
2% Stacking Gel	2% (v/v) Acrylamide/Bis solution [37.5:1] (SLS), 0.1 mM Tris-HCL pH 6.8, 0.83% (w/v) SDS, 0.83% APS, 0.0083% TEMED
SDS Running Buffer	0.3 % (w/v) Tris, 1.4 % (w/v) Glycine, 0.1% (w/v) SDS

2.7.3 Visualisation of protein after Gel electrophoresis

Protein bands or target protein was either visualised by Coomassie staining or by Immunoblotting.

2.7.3.1 Coomassie staining

The SDS-PAGE gel was rinsed with diH_2O before being incubated in Coomassie brilliant blue staining solution (**Table 2.4**) for 1 hr. The gel was then washed with de-staining solution (10% (v/v) Ethanol, 10% (v/v) Acetic acid) until the background staining was removed.

2.7.3.2 Protein immunoblotting

For immunoblotting SDS-PAGE separated proteins were transferred to immobilon-P PVDF membrane (Merk-Millipore) using the Trans-Blot® Semi-Dry Transfer Cell (BioRad). PVDF membrane was activated in methanol for 15 sec, washed 3x with diH_2O then equilibrated in transfer buffer (**Table 2.4**). The SDS-PAGE gel was equilibrated in transfer buffer and filter paper soaked in transfer buffer was stacked with the equilibrated membrane and gel according to manufacturer's instructions and the system was run at 20 V for 15 min. PVDF membrane was blocked in protein blocking buffer A (**Table 2.4**) for 1 hr at room temperature. The membrane was then incubated in primary Tetra·His Antibody (Qiagen) diluted 1:1000 in protein blocking buffer A (**Table 2.4**) for 1 hr at room temperature. Following primary antibody incubation, the membrane was washed 3 times in TBS-T for 10 min then once in TBS^{IM} for 10 min (**Table 2.4**) then incubated in secondary goat anti-mouse IgG-HRP (Sigma-Aldrich) diluted 1:10000 with protein blocking buffer B (**Table 2.4**) for 1 hr at room temperature. Following 4x 10 minute washes in TBS-T, protein bands were visualised using Luminata Classico Western HRP substrate (Merck-Millipore) and imaged on the ChemiDoc MP imaging system (BioRad)

2.7.4 Total protein extraction

For total protein extraction 1 ml bacterial cell culture was pelleted by centrifugation 16,000 xg for 5 min. The supernatant was discarded and pellet stored at -20 °C until protein extraction. Cell pellets were re-suspended in 25 µl per OD unit of 4x SDS sample loading buffer (**Table 2.4**). Solutions were boiled at 95 ˚C for 10 minutes and centrifuged at 16,000 xg for 5 minutes. 5 µL of supernatant from cell culture samples were loaded for SDS-PAGE analysis.

2.7.5 Small scale periplasmic protein extraction

For periplasmic protein extraction from 1 ml of bacterial cells a scaled down protocol for periplasmic protein extraction was used. All centrifuge steps were carried out at 8500 xg for 20 min at 4 °C. 1 ml cell culture was centrifuged and supernatant discarded. The cell pellet was re-suspended in 30 µl sucrose solution (**Table 2.4**), incubated on ice for 15 min then centrifuged. The supernatant (sucrose fraction) was stored at -20 °C or discarded. The cell pellet was re-suspended in 30 µl ice-cold 5 µM $MgCl_2$, 1 µl 15 mg/ml Lysozyme (Sigma Aldrich) and 1.4 µl 25X Protease Inhibitor Cocktail (**Table 2.4**) and incubated on ice for 30 min then centrifuged a final time. The supernatant was analysed as the periplasmic fraction. The remaining cell pellet should contain the cell debris and cytoplasmic component and was discarded.

2.7.6 Small scale spin column nickel affinity purification

The Ni-NTA spin column purification kit (Qiagen) was used for small scale nickel affinity purification of expressed proteins. For BL21(DE3) cells expressing OafA and OafB constructs 30 ml of overnight cell culture was centrifuged at 4000 xg for 15 min at 4 °C. 10x Bug buster solution (Novagen) was diluted to 1x using NPI-10 lysis buffer (**Table 2.4**). Cells were lysed in 1.7 ml 1x BugBuster solution for 20 min at room temperature on a rocker. The protocol for protein purification under native conditions from *E.coli* lysates in the Ni-NTA spin kit handbook 01/2008 (Qiagen) was then followed from step 3. Briefly, lysed cells were centrifuged 12,000 xg for 20 min at 4 °C. Ni-NTA column was equilibrated in 600 µl NPI-10 lysis buffer (**Table 2.4**) and centrifuged for 2 min at 890 xg, then the clarified cell lysates were loaded onto the column 600 µl at a time and centrifuged 270 xg for 5 min. The spin column was washed twice with 600 µl NPI-20 wash buffer (**Table 2.4**) and centrifuged for 2 min at 890 xg. Bound protein was eluted from the column with 2 elution steps of 300 µl NPI-500 elution buffer (**Table 2.4**) centrifuged for 2 min at 890 xg. Column flow through from each step was stored at -20 °C until analysis.

2.7.7 Periplasmic protein extraction using EDTA and lysozyme

Cells from 1L culture were harvested by centrifugation at 5000 xg 15 min at 4 ˚C. The cell pellet was re-suspended in 20ml ice cold SET buffer (Table 1), 13 mg lysozyme were

added and the solution mixed by inverting then incubated for 2 hr at 30 ˚C. The solution was then centrifuged at 21000 xg for 15 min at 4 ˚C and the supernatant taken as the periplasmic fraction. A small sample of the remaining pellet was taken for SDS-PAGE analysis.

2.7.8 Protein dialysis

To exchange buffers for purified proteins Spectra/Por 12-14,000 MWCO dialysis tubing (Fisher Scientific) was used. Protein sample was sealed into Spectra/Por tubing and immersed in at least 1000x volume of new buffer solution with gentle stirring at room temperature for at least 1 hr then into fresh buffer at 4 °C overnight.

2.7.9 Large scale nickel affinity purification of periplasmic protein fractions

See **Table 2.4** for all buffers used in this protocol.

The isolated periplasmic fraction was dialysed overnight (2.7.8) into TBS buffer. The appropriate amount of Ni^A-Equilibration buffer was added to the periplasmic fraction to make a final concentration of 12 mM imidazole then the solution was clarified by centrifugation 21000 xg for 10 min 4 ˚C and filtered using a 0.45 µm filter. Purification was performed on the AKTA purification system using a HisTrap FF 5 ml column (GE healthcare). The column was equilibrated in Ni^A-Equilibration buffer then the protein sample was loaded on to the column and washed to remove contaminating proteins with Ni^A-Wash buffer. Protein was eluted from the column using Ni^A-Elution buffer buffer and collected in 3ml fractions which were analysed by SDS-PAGE to ensure protein purity. Following purification, protein samples were dialysed (2.7.8) into 50 mM sodium phosphate buffer 100 mM NaCl pH7.8 unless otherwise stated.

2.7.10 Protein quantification

Protein concentration was determined using A_{280} absorbance of the solution on the Jasco V-560 spectrophotometer. A_{280} absorbance was converted into protein concentration using the Beer-Lambert law ($A = \varepsilon cl$):

*A280 absorbance = (extinction coefficient of the protein) * (concentration of protein in solution) * (path length of the cuvette)*

2.8 *In vitro* characterisation of purified proteins

2.8.1 Biophysical analysis

For mass spectrometry analysis OafA_Lys355 was dialysed into 25 mM ammonium acetate (2.7.8) then concentrated to 69 µM using vivaspin 500 10,000 MWCO spin columns (GE Healthcare). For FT-ICR-MS (see below) OafA_Lys355 was dialysed (2.7.8) into 35 mM ammonium acetate.

All mass spectrometry was carried out by members of the Technology Facility at the University of York. Electrospray ionisation time of flight (ESI-TOF) analysis was carried out on the ABI Qstar tandem mass spectrometer.

Peptide mass fingerprinting was carried out using Matrix Assisted Laser Desorption Ionization Tandem Time-of-Flight (MALDI-TOF/TOF) mass spectrometry using a Bruker ultraflex-III. OafA_Lys355 was digested with trypsin, an MS spectrum was acquired between 800-5000 m/z and the ten strongest peaks, above a signal to noise threshold, were selected for MS/MS fragmentation. The tandem mass spectral data produced are searched against the NCBI database using the Mascot search program to generate peptide matches with associated expect values. Protein identifications are inferred from peptide matches.

Protein molecular mass measurement was carried out using fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) and MALDI In source decay mass spectrometry (MALDI-ISD-MS) was used for N-terminal sequencing of OafA_Lys355.

NMR spectrometry was carried out by Pedro Aguiar in the centre for magnetic resonance at the University of York.

2.8.2 Analysis of the crystal structure of $OafB_{SPA}$_377

Analysis towards solving the crystal structure of $OafB_{SPA}$_377 was carried out by Sarah Tindall (University of York) and the methods used are adapted from (Pearson *et al.*, 2020)(manuscript submitted for publication).

To crystallise $OafB_{SPA}$_377, a hanging-drop vapour diffusion method was used with 20 mg/mL $OafB_{SPA}$_377 in a drop ratio of 1:1 protein:reservoir solution. After incubation for 24 hr at 20 ^{o}C crystals grown in 100 mM BisTris pH 5.5, 0.25 M lithium sulfate, 25% PEG

3350 were cryoprotected by addition of glycerol to a final concentration of 20% and vitrified in liquid nitrogen.

X-ray diffraction data for crystals of $OafB_{SPA}$_377 were collected on beamline I04-1 (Diamond Light Source, UK) at a wavelength of 0.9282 Å using a Pilatus 6M-F detector. Data were integrated with XDS (Kabsch and IUCr, 2010), and scaled and merged with AIMLESS (Evans and Murshudov, 2013) via the Xia2 pipeline (Winter, 2010). Fragon molecular replacement (Jenkins, 2018) used Phaser (McCoy *et al.*, 2007) to place an ideal poly-alanine helix of 14 amino acids in length followed by density modification with ACORN (Jia-Xing *et al.*, 2005). ARP-wARP (Perrakis *et al.*, 1999) was used for automated chain tracing, and the model was refined using REFMAC5 (Garib N. Murshudov; Alexei A. Vagin; Eleanor J. Dodson, 1997; Murshudov *et al.*, 1999, 2011; Winn *et al.*, 2003; Vagin *et al.*, 2004; Nicholls *et al.*, 2012). Manual manipulation of the model between refinement cycles was performed using Coot (Emsley *et al.*, 2004, 2010). The final model was evaluated using MolProbity (Chen *et al.*, 2010) and PDB validate, secondary structure shown in Fig. 5A was annotated using STRIDE (Heinig and Frishman, 2004). The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB ID code 6SE1).

2.8.3 *In vitro* esterase activity assay

The catalytic activity of OafA and OafB C-terminal constructs was confirmed by acetyl esterase activity using p-nitrophenyl acetate (pNP-Ac) as a substrate. 100 µl of enzyme solution (10 µM $OafA_{STM}$, 40 µM $OafB_{STM}$ or 0.04 U/ml Acetylxylanesterase as the positive control) or appropriate control buffers were added to relevant wells of a 96 well plate and incubated at 37 ˚C for 10 min prior to addition of pNP-Ac. 100 mM pNP-Ac stock dissolved in ethanol was diluted to 1mM using the corresponding buffer to each protein sample. 100 µl 1mM pNP-Ac was added to matching sample and control wells and immediately placed into a plate reader incubated at 37 ˚C. Absorbance at 405 nm was measured at T=0, and then at 5 min intervals. The acetylxylanesterase positive control was an esterase from *Orpinomyces sp.* (Megazyme).

2.8.4 *In vitro* acetyltransferase activity assay

Crude LPS extracted from OafA-negative STM LT2 strain (Path993) was heated at 100°C for 20 min to inactivate the proteinase K (see above). Heat-treated LPS was mixed 1:1 with KPi buffer (**Table 2.4**). 10 µM $OafA_{STM}$ and 20 µM $OafB_{SPA}$ C-terminal constructs were incubated at 4 °C in LPS-KPi mixture with 4 mM pNP-Ac dissolved in ethanol (4% (v/v) final concentration in reaction). Samples of the reaction mix were taken after specified time points and inactivated by boiling for 10 min.

Chapter 3 *In silico* analysis of carbohydrate O-acetyltransferases

3.1 Introduction

The intention of this study is to characterise the mechanism of action of O-antigen acetyltransferases OafA and OafB with the aim of applying this knowledge both to O-antigen modification and to the array of related carbohydrate O-acyltransferase proteins that catalyse a diverse range of modifications across the domains of life, as discussed in the introduction (section 1.3). This chapter describes the *in silico* analysis of experimentally characterised bacterial O-acetyltransferases which allowed hypotheses to be formed about the mechanism of action and functional residues of O-antigen acetyltransferases. This work has contributed to production of a joint first author manuscript with Sarah Tindall (University of York) which is submitted for publication (Pearson *et al.*, 2020) therefore some of the text and figures from this manuscript have been reproduced in the text. Work contributed by Sarah Tindall has been acknowledged and is unpublished at the time of thesis submission.

3.2 Defining a working model for *Salmonella* O-antigen acetylation

The STM O-antigen acetyltransferases OafA (Slauch *et al.*, 1996) and OafB (Davies *et al.*, 2013) (formerly F2GtrC) are both predicted by InterPro to contain an N-terminal AT3 domain (IPR002656, PF01757) fused to an SGNH domain (IPR013830, PF14606 or PF13472) (Finn *et al.*, 2014; Mitchell *et al.*, 2015) (**Figure 3.1 A**). The AT3 domain has 10 trans membrane helices (TMH) and an additional 11th helix that is presumably required to localise the fused SGNH domain in the periplasm (Krogh *et al.*, 2001); this prediction is supported by experimental topology analysis of OafB (Kintz *et al.*, 2015).

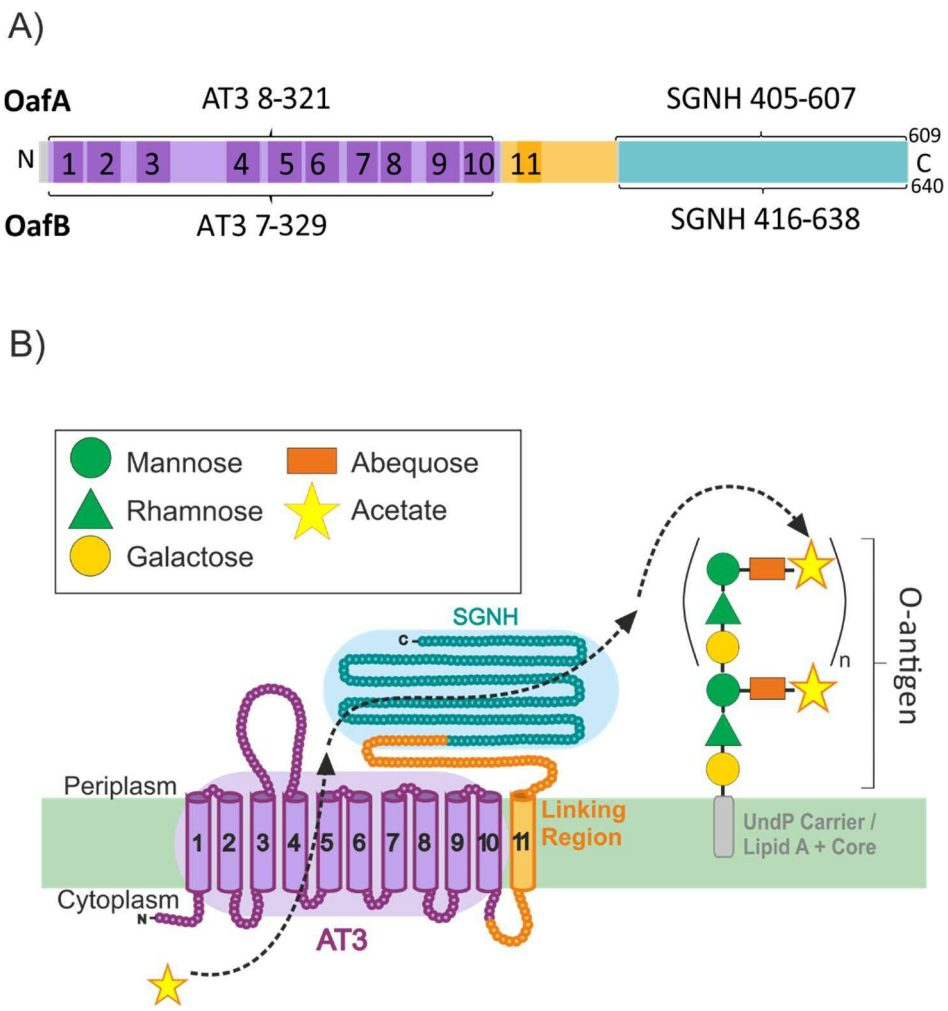


Figure 3.1| Schematic representation of *Salmonella ser.* Typhimurium OafA and OafB functional domains defined by InterPro and transmembrane helices predicted by TMHMM. AT3 indicates the Acyl_transf_3 InterPro domain (IPR002656) and this domain is coloured purple. SGNH indicates the SGNH hydrolase type esterase domain (IPR013830) in the SGNH superfamily (SSF52266) and is coloured teal. The linking region between the two domains is coloured orange. (A) Linear representation with domain boundaries for OafA indicated above and for OafB indicated below and transmembrane helices shaded and numbered. (B) Topological representation of OafA in the proposed mechanism of action of O-antigen acetyltransferases during maturation of the LPS in the periplasm. InterPro results were obtained – 07/10/15 and TMHMM results obtained – 13/01/17.

Phyre[2] protein fold recognition server (Kelley *et al.*, 2015) was able to model a predicted structure for the C-terminal periplasmic domain of OafA and OafB (389-609 for OafA and 410-636 for OafB). The predicted structures were modelled against the crystal structure of a putative hydrolase from *Bacteroides thetaiotaomicron* (PDB: bt3161) with confidence of 99.1% and 99.4% and sequence identity of 13% and 14% for OafA and OafB modelled regions, respectively (**Figure 3.2**). This analysis suggested that these proteins are likely to contain a C-terminal periplasmic domain with the characteristic α-

β-α fold of an SGNH domain with the catalytic triad residues Ser, Asp, His in close proximity (Mølgaard *et al.*, 2000) (**Figure 3.2**).

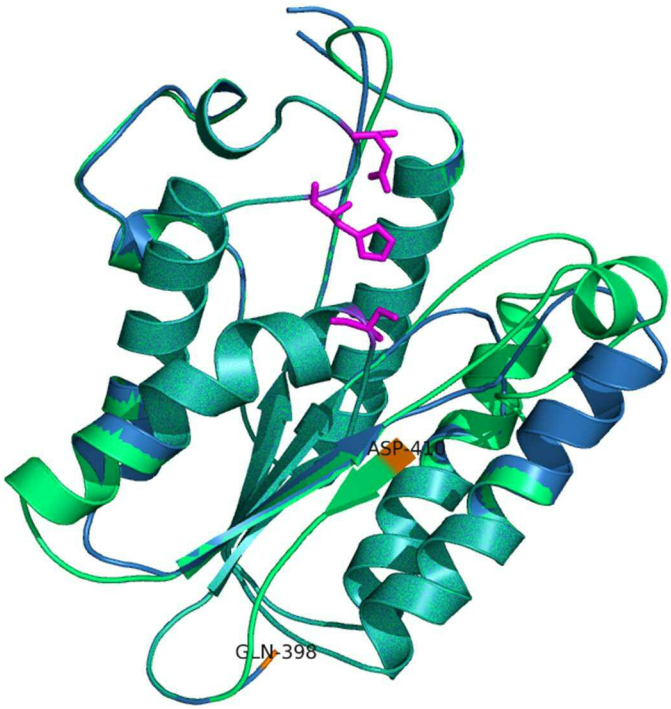


Figure 3.2| Predicted crystal structure for the C-terminal domain of *Salmonella* ser. Typhimurium OafA (blue) and OafB (green). Both proteins were modelled on template structure (PDB: bt3161). Catalytic triad residues are shown as magenta sticks and overlap exactly in both protein structure predictions. N-terminal residues are coloured orange and labelled.

Previous work showed that the SGNH domain is essential for acetyltransferase activity in OafA and OafB (Hauser *et al.*, 2011; Kintz *et al.*, 2015) and that functional residues on the cytoplasmic side of the AT3 domain can result in loss of O-antigen acetylation when mutated to alanine (Kintz *et al.*, 2015), suggesting that both defined functional domains of OafA and OafB are required for O-antigen acetylation. The SGNH domain is periplasmic and the O-antigen acetylation modification is proposed to occur on the periplasmic side of the inner membrane, yet the inner membrane spanning AT3 domain is still required for the acetylation modification (Kintz *et al.*, 2015). Therefore, the source of acetate is likely to be cytoplasmic and is transported to the periplasmic side by the AT3 domain before the SGNH domain subsequently transfers the acetyl group to the specific O-antigen acceptor carbohydrate (**Figure 3.1 B**).

3.3 Analysis of conserved residues across characterised O-acetyltransferases

In order to investigate the hypothesised mechanism of O-antigen acetylation (**Figure 3.1 B**), a review of the literature was conducted to identify bacterial AT3 domain-containing proteins whose carbohydrate acyltransferase activity had been demonstrated experimentally. This search was constrained to bacterial proteins initially in the hope of identifying proteins with close enough evolutionary relationships to maintain a conserved mechanism of action. These experimentally characterised proteins were compiled with the aim of identifying conserved features between them, which may give clues to a conserved mechanism of action or identify specific regions within these proteins which are adapted to their O-acyltransferase function.

41 bacterial AT3 domain-containing proteins, with experimentally confirmed carbohydrate O-acyltransferase activity were identified in the literature (Appendix I). These proteins span a diverse selection of Gram-positive and negative bacteria and have wide ranging biological roles from initiation of symbioses with plants to modification of antibacterial compounds (Summarised in **Figure 3.3**). 30 of the 42 proteins identified were characterised to have a native role of O-acetyltransferase activity (Appendix I). As, it was unknown whether the mechanism of transfer is conserved between proteins transporting diverse acyl substituents, sequence analysis was first constrained to these 30 proteins. Of the 30 proteins, 19 contain just the AT3 domain, while 11, including STM OafA and OafB, have the AT3-SGNH fused domain architecture (**Figure 3.3**).

For the 19 AT3-only O-acetyltransferase proteins there was no indication in the literature that these proteins functioned with a cognate periplasmic SGNH partner. Significantly, analysis of the genomic context of the gene encoding the AT3-only acetyltransferase OacA, implicated in acetylation of the rhamnose sugar of the *Shigella flexneri* LPS O-antigen (Verma *et al.*, 1991), did not reveal a candidate partner protein. This protein is encoded on the *Shigella* bacteriophage *SF6* genome. No significant hits were found when this bacteriophage genome (NCBI:txid10761) was searched against the protein sequences of the OafB SGNH domain, or other characterised periplasmic SGNH acetyltransferase proteins PatB (Moynihan and Clarke, 2010), AlgJ and AlgX (Baker *et al.*, 2014). This provides a key example of a protein which performs a similar

modification the AT3-SGNH fused OafB which, contrary to OafB, does not appear to require an SGNH domain to function. Thus, comparison between the conserved and functional residues between AT3-only and AT3-SGNH fused O-acetyltransferases could shed light on the key differences between the proteins which require fused SGNH domains to function and those which do not.

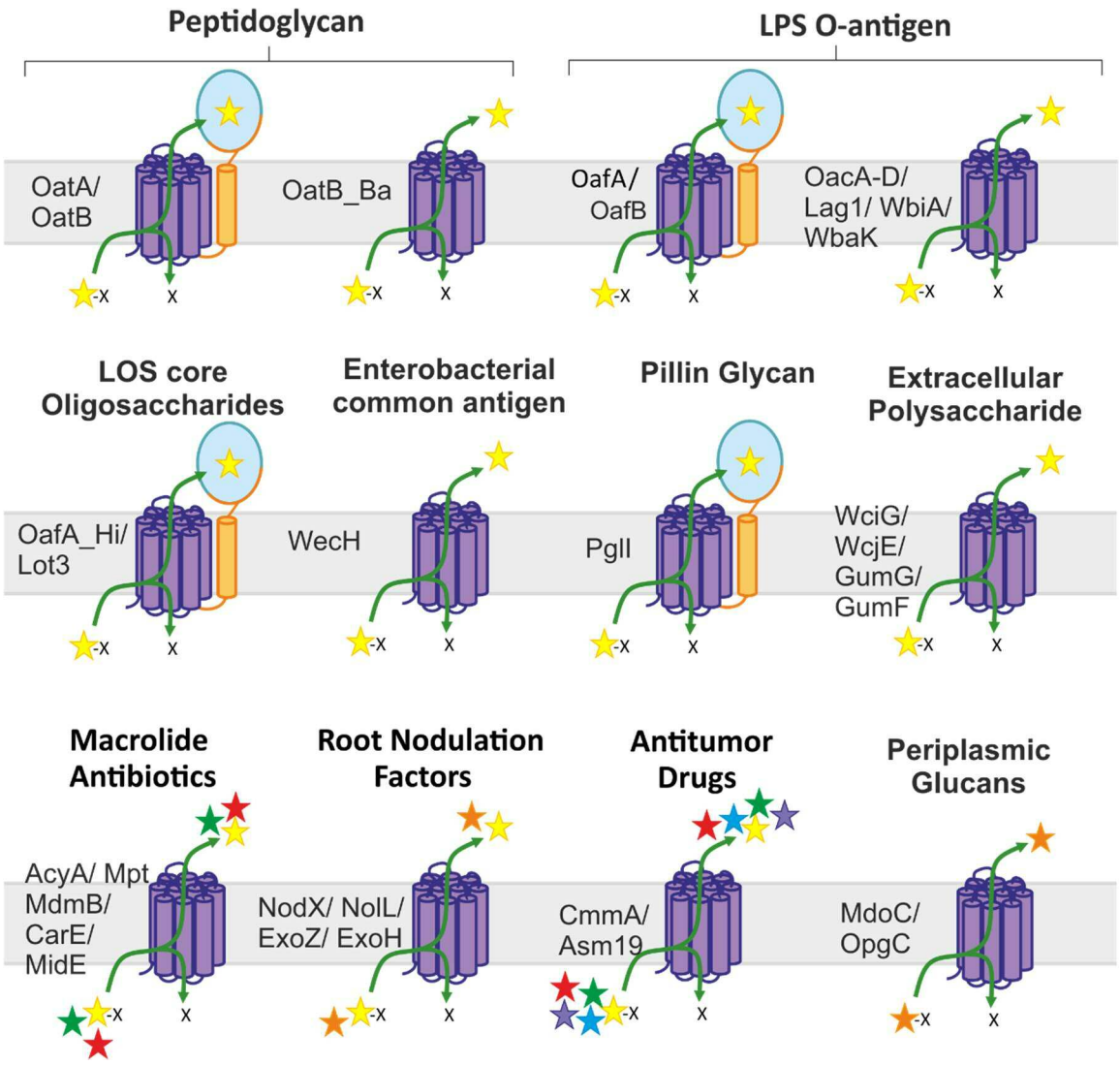


Figure 3.3| Diversity of acceptor substrates and protein architecture within experimentally characterised bacterial AT3 domain-containing acyltransferases.AT3 domain is coloured purple, the SGNH domain is coloured blue and the linking region between the two domains is coloured orange. Acyl groups are indicated by coloured stars. Yellow= Acetyl, Green = Propionyl, Blue = Isobutryl, Red = Isovaleryl, Orange = Succinyl, Purple = Butryl,.

3.3.1 Analysis of conserved residues in the AT3 domain

In order to investigate the conserved features between bacterial AT3 domain-containing O-acetyltransferases and to shed light on the specific adaptations of AT3-SGNH fused proteins for function with a cognate SGNH partner, alignments of all 30 identified acetyltransferase proteins, as well as an *S. enterica* serovar Paratyphi A (SPA) OafB homologue ($OafB_{SPA}$), were carried out using TCoffee (Section 2.4.1). The SPA OafB homologue was included in the analysis so that it could be applied to a collaborator's structural analysis. The full alignment output figure can be found in Appendix II.

Preliminary analysis of the evolutionary relationship of the AT3 domains of SGNH fused and AT3-only proteins, following removal of the SGNH domain from fused proteins and re-alignment using TCoffee (Section 2.4.2), revealed that the AT3-only and AT3-SGNH fused AT3 domains cluster separately, rather than according their association with Gram-positive or negative bacteria (**Figure 3.4**). This, suggests specific adaptation of AT3 proteins towards interaction with a fused SGNH domain.

As only 12 of the 31 sequences in the alignment contain a fused SGNH domain, analysis was initially centred on the membrane domains shared across all protein sequences in the alignment. Analysis of the residues conserved across the whole set of aligned proteins revealed that just 4 amino acids are invariant (100% identical) within the AT3 domain (Appendix II, **Figure 3.5 B**). These residues include a histidine residue in TMH1 (H25_OafA), a phenylalanine and glycine residue in TMH2 (F41, G46_OafA) and a second glycine residue in TMH6 (G202_OafA) (**Figure 3.5 A** coloured orange).

Residues which were similar in more than 90% of the sequences (>90% similar) were also highlighted in this alignment (coloured blue) (**Figure 3.5**). Similar residue groups were defined according to the characteristics of their side chains (Amino acid groups = FYW- Aromatic, ILVM- Small Hydrophobic, RK - Positive, DE- Negative, GA- Small, ST- Polar, NQ- Amide) and >90% similar small hydrophobic residues in TMH regions (A,I,L,M,P,V) were not highlighted for analysis due to their enrichment in these regions.

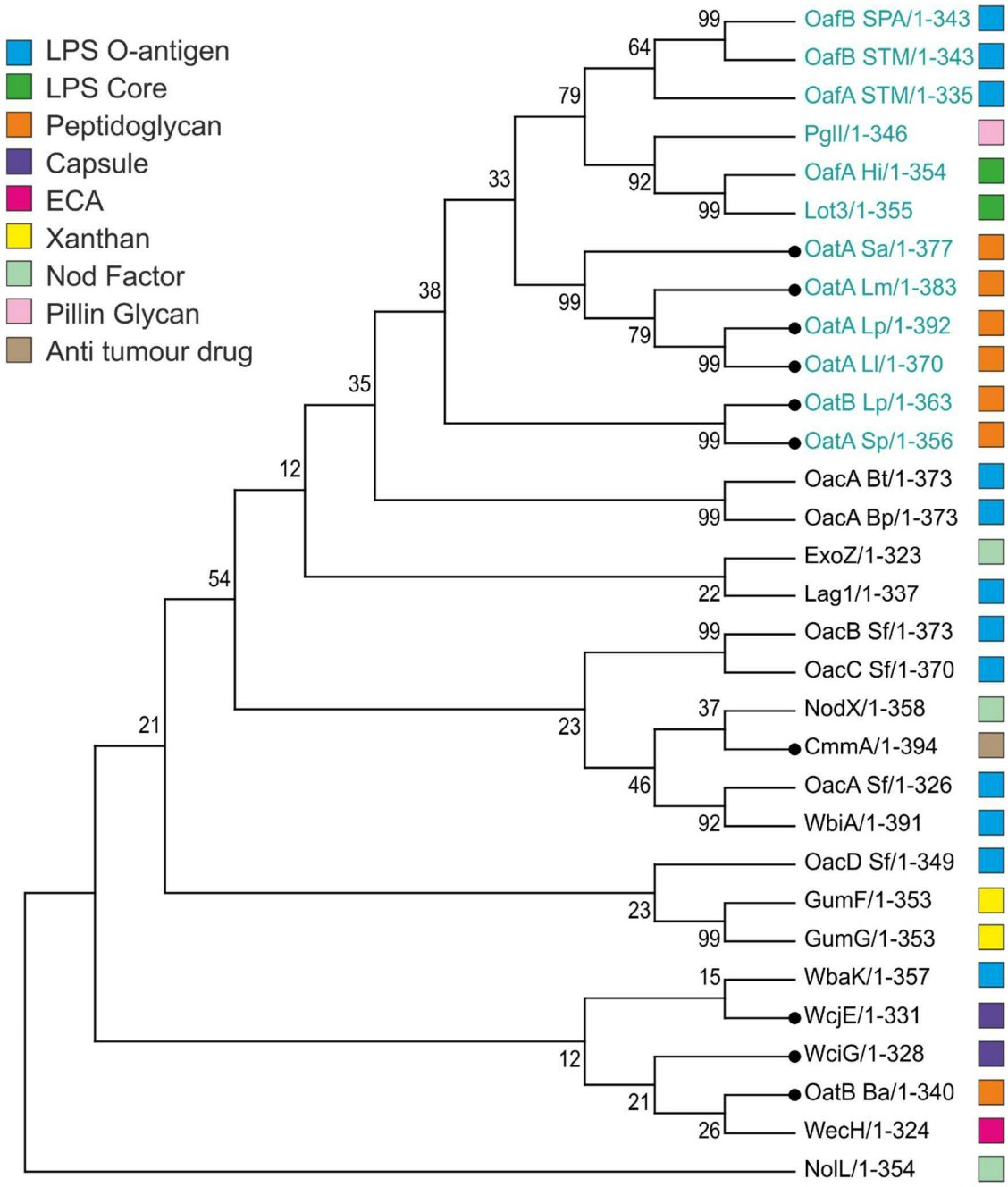


Figure 3.4| Evolutionary analysis of the AT3 domains of experimentally confirmed bacterial O-acetyltransferases. See section 2.4.2 for details of sequence selection and tree building parameters. Evolutionary analysis was conducted in MEGAX (Kumar *et al.,* 2018) using the Maximum Likelihood method and JJT matrix-based model. Protein names coloured teal represent AT3-SGNH fused proteins. Branches ending in solid circles represent proteins from Gram-positive organisms. Coloured boxes to the right of protein names indicate their acceptor substrate according to the Key in top left corner.

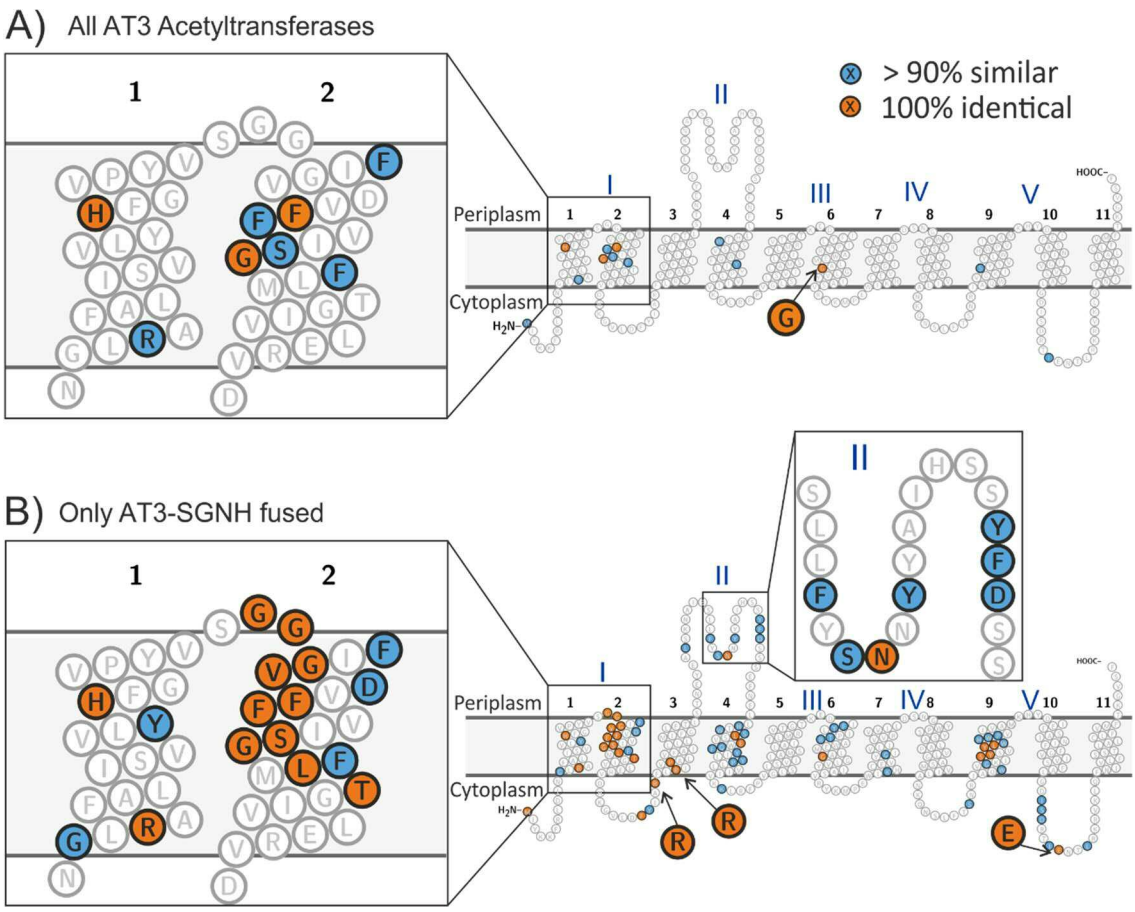


Figure 3.5| Conservation in transmembrane domains of experimentally characterised bacterial AT3 carbohydrate acetyltransferases. 100% identical residues are coloured orange, similar residues conserved in > 90% of sequences are coloured blue, conserved small hydrophobic residues A/I/L/M/P/V in transmembrane helices were not coloured unless 100% identical. (A) Conserved residues across all 30 currently known experimentally characterised proteins and $OafB_{SPA}$. (B) Conservation in only AT3-SGNH fused proteins in the alignment. See Appendix I for details of aligned sequences and Appendix II for the full alignment. Figure adapted with permission from (Pearson *et al.*, 2020).

In TMH1, an arginine residue was highlighted as >90 % similar across all proteins and was only replaced with lysine, suggesting a strong selection for a positively charged residue in this position. This residue is located 10 residues away from the invariant histidine residue shared across all protein sequences in the alignment (**Figure 3.5 A**). The HMM logo of all recorded AT3 domains in the Pfam database (PF01757) also highlights this R/K-X_{10}-H motif as highly conserved across the whole protein domain family (**Figure 3.6**). This motif had not previously been investigated by mutagenic analysis and its strong conservation across all AT3 domain-containing proteins could suggest that these are critical residues in a conserved mechanism for acyl group donor interaction or processing.

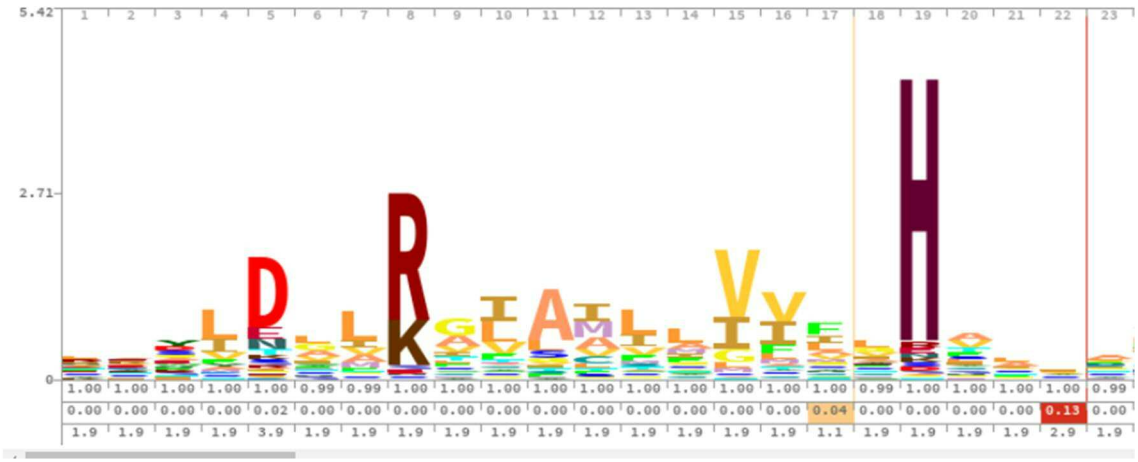


Figure 3.6| HMM logo of all AT3 domain-containing proteins (PF01757) in the Pfam database. Column height represents the invariance of that residue. Image taken – 01/08/2019.

The remaining residues identified as 100% identical across all proteins in the alignment were also selected as candidates for mutagenic analysis. F41_OafA and G46_OafA belong to the previously identified V-X-X-F-F-X-(I/V/L)-S-G-(F/W/Y), shared among many AT3 proteins from both Gram-positive and negative bacteria (Luck *et al.*, 2001), and the FFXISG motif identified in AT3-only O-antigen acetyltransferases with homology to *S. flexneri* Oac (Thanweer and Verma, 2012). The transmembrane glycine residue in TMH6 G202_OafA is also of interest as transmembrane glycine's are known to cause kinks in transmembrane helices and can be involved in helix to helix contacts (Javadpour *et al.*, 1999; Dong *et al.*, 2012). Therefore, the conserved glycine residue may play an important structural role in AT3 domain-containing acetyltransferases, either in stabilising the structure of the membrane spanning region or inducing transmembrane kinks to form substrate interaction or transport.

Residues specifically conserved in the AT3 domains of just the AT3-SGNH fused proteins were next analysed (**Figure 3.5 B**). The highest degree of conservation was seen within the first 4 transmembrane helices and in TMH 9. Conserved residues in TMH9 were mostly hydrophobic residues Y,V,L,W,H and invariant S280_OafA had previously been highlighted as essential in the AT3-only O-antigen acetyltransferase Oac (S742_Oac) (Thanweer and Verma, 2012).

The most striking shared feature of AT3-SGNH fused proteins is the highly conserved GG-F/Y-XGV-D/P/V motif located at the periplasmic side of TMH2 (G33-D39_OafA) (**Figure 3.5**). This replaces a longer and more divergent loop region between TMH1-2 in

the non-fused AT3-only proteins (Appendix II). Further conserved residues are seen in the periplasmic loop between TMH2-3, including S112_OafA, N113_OafA and Y122_OafA. The periplasmic location of these conserved residues which are specific to AT3-SGNH fused proteins could suggest specific adaptation to interaction with the SGNH domain.

An RXXR motif between TMH2-3 had previously been highlighted as essential for activity in the AT3-only *Shigella flexneri* Oac (R73, R75_Oac) (Thanweer *et al.*, 2008), and STM OafB (R71, R73_OafB) (Kintz *et al.*, 2015). This motif was not absolutely conserved across all proteins in the alignment but was invariant across the AT3-SGNH fused proteins.

AT3-only acetyltransferases do not contain an 11th TMH required to deliver the SGNH domain to the periplasm but a glutamate residue after the C-terminal end of TMH10 (E325_OafA) was invariant across AT3-SGNH protein sequences. Negatively charged glutamate on the cytoplasmic side of the transmembrane helix goes against the trend of positively charged residues being enriched around the cytoplasmic ends of trans membrane helices (von Heijne, 1986; Baker *et al.*, 2017). Therefore, the invariance of this residue may indicate some mechanistic involvement in AT3-SGNH fused proteins.

3.3.2 Analysis of conserved residues in the SGNH domain

SGNH domains are usually characterised by the presence of four blocks of sequence, containing conserved residues: block I – GDS, block II – G, block III – GxND and block V – DxxH (where 'x' is any non-proline residue) (Akoh *et al.*, 2004), with the catalytic triad, consisting of serine (block I), aspartic acid and histidine (block V) (Section 1.2.5). The alignment of characterised O-acetyltransferase proteins highlighted 6 invariant residues across the SGNH domains of AT3-SGNH fused proteins (Appendix III). Other than conservation of motifs found in catalytic blocks I and V (G419-S421_OafA and D587, H590_OafA), a glycine residue (G595_OafA) was also invariant across all sequences. Glycine residues have a single hydrogen atom as their side chain (Harvey *et al.*, 1995) allowing them to introduce conformational flexibility or tight turns into polypeptide chains. Indeed, Yan and Sun (1997) suggest that glycine residues may provide flexibility for enzyme active sites to allow substrate binding. This glycine residue is substituted for a serine in the SGNH domain of peptidoglycan acetyltransferase OatA (See figure 6 of

(Sychantha *et al.*, 2017)), which was the first structure of a bacterial AT3 fused SGNH domain-containing O-acetyltransferase to be published. Serine is another small amino acid which is common in tight turns within proteins. This residue was not identified to play a catalytic role in the SGNH domain of OatA through crystal structure analysis and sits at the beginning of an α helix at the end of a turn within the protein (**Figure 3.7**) (Sychantha *et al.*, 2017). Thus, it is likely that this conserved glycine residue plays a structural rather than catalytic role in the SGNH domain of AT3-SGNH fused O-acetyltransferase proteins.

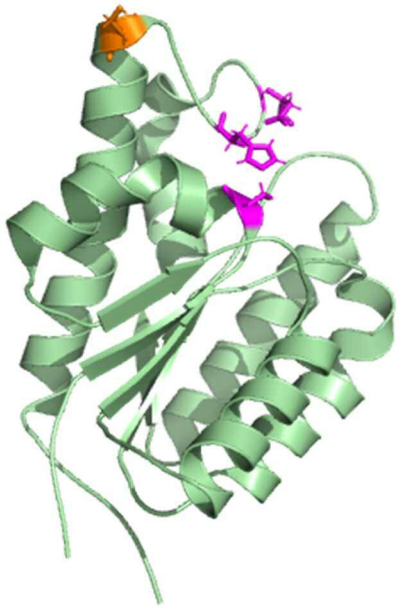


Figure 3.7| Crystal structure of OatA *Streptococcus pneumoniae* (PDB 5UFY). Catalytic triad residues Serine_438, Histidine_571, Aspartic acid_ 568 are shown as magenta sticks and Serine_577 is shown as Orange sticks.

During analysis of the predicted structure of OafA and OafB (**Figure 3.2**), pairs of cysteine residues were observed to be positioned within close proximity (**Figure 3.8**), highlighting them as potential disulphide bonding partners. As the C-terminal half of OafA, which includes the linking region and SGNH domain (**Figure 3.1**), has been experimentally confirmed to reside in the periplasm (Kintz *et al.*, 2015), these pairs of cysteine residues are highly likely to form disulphide bonds in this oxidising environment.

One disulphide bonding pair was conserved in position in both proteins (C567, 572_OafA and C598,603_OafB) (**Figure 3.8**), suggesting a potential conserved role in protein structure or function. Therefore, in order to investigate whether these cysteine pairs are

conserved across a range of carbohydrate acetyltransferase proteins, a broader set of AT3-SGNH fused proteins with homology to OafA were aligned (Section 2.4.1).

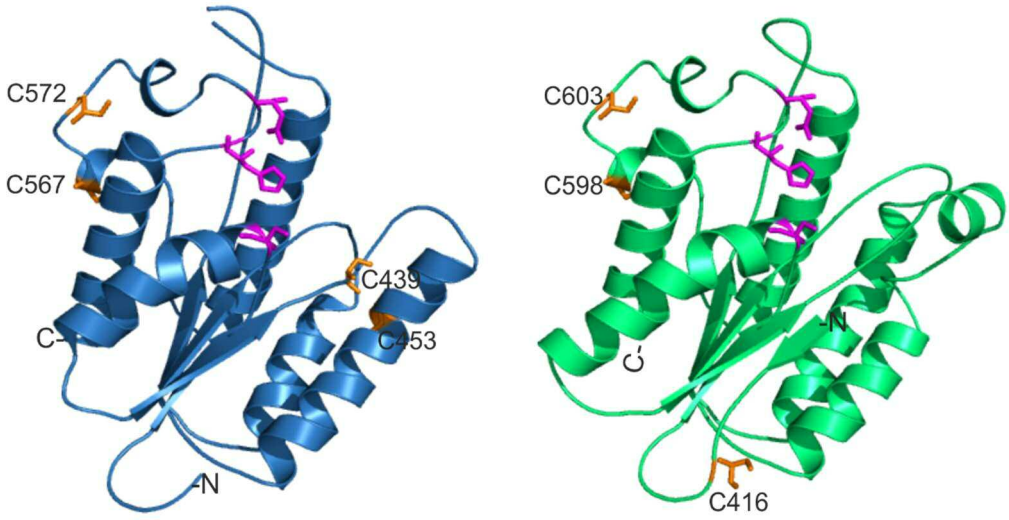


Figure 3.8| Position of cysteine residues in the predicted crystal structures of OafA (Blue) and OafB (Green). See **Figure 3.2** legend for details of structure prediction. Catalytic triad residues are shown as magenta coloured sticks and cysteine residues as orange coloured sticks.

The STM LT2 OafA protein sequence was used to blast search the uniref 50 database (Suzek *et al.*, 2015). The top 100 hits were taken and any sequences matching 100% or that only aligned to a fragment of the protein were removed. As a further filter, sequence hits coming from the same species were also removed then all 11 experimentally confirmed AT3-SGNH fused proteins were added to this list of proteins for alignment (Appendix IV).

Alignments of this broader set of AT3-SGNH proteins with homology to OafA highlighted conserved pairs of cysteine residues in the C-terminal periplasmic domain which aligned with those present in OafA (Appendix IV). One pair is in the predicted periplasmic portion of the linking region between the AT3 and SGNH domains (C383S, C397S_OafA), a second pair is found towards the middle of the SGNH domain (C439S, C453S_OafA), and a third pair towards the end of the SGNH domain (C567S, C572S_OafA) (Appendix IV). The central pair of cysteine residues, which are also suggested to be in close proximity by structure prediction (**Figure 3.8**), were not conserved in OafB (Appendix IV). In agreement with OafA and OafB, cysteine residues corresponding to the defined

putative pairs were only present or absent as pairs within the sequence of other aligned proteins.

The 'second' and 'third' pair of cysteine residues in the C-terminal half of OafA are found in close proximity in the predicted structure (**Figure 3.8**) supporting their predicted disulphide bond pairing. The structure prediction did not model the periplasmic linking region of OafA and OafB as this region is not present in the structure of the protein used to build this prediction (PDB: bt3161) (**Figure 3.2**). Neither of the linking region cysteine residues were modelled for OafA. However, as the only remaining un-modelled cysteine pair within the periplasmic domain, they are highly likely to interact. Additionally, one cysteine residue of the most N-terminal pair in OafB (C416_OafB) was modelled by the structure prediction and is not in close proximity to the other cysteine residues (**Figure 3.8**). This suggests that it would most likely be available to interact with the only remaining un-modelled cysteine. These conserved cysteine pairs could be important for catalytic function or structural integrity of the SGNH hydrolase domain as proteins in oxidising environments often rely on disulphide bonding to maintain their structural stability and facilitate correct folding (Bardwell, 1994).

3.4 Chapter summary

The working model for O-antigen acetylation by OafA and OafB (**Figure 3.1 B**) is in agreement with the predicted mechanism of periplasmic peptidoglycan acetylation by the AT3-SGNH fused acetyltransferase OatA (Sychantha *et al.*, 2017). Additionally, in the two component PatA/PatB peptidoglycan acetyltransferase system it is proposed that, PatB, a soluble SGNH protein, is responsible for transfer of the acetyl group onto the peptidoglycan substrate following delivery of the acetyl group to the periplasmic side by PatA, a membrane spanning O-acetyltransferase (MBOAT) protein (Moynihan and Clarke, 2014b). The membrane bound PatA MBOAT protein in this system is interchangeable with WecH, an AT3-only acetyltransferase (Kajimura *et al.*, 2006; Moynihan and Clarke, 2010). This provides an example of direct transfer of acetate between a membrane bound AT3 domain and soluble SGNH domain protein, and further supports the mechanistic model of the AT3 domain delivering the acetyl group to the SGNH domain for transfer onto the acceptor substrate in AT3-SGNH fused proteins.

This hypothesised mechanism raises the question of how AT3-only acetyltransferase proteins are adapted to acetylate extra cytoplasmic carbohydrates without the need of a fused SGNH domain. The finding that the SGNH domain is essential in O-antigen acetylation (Kintz *et al.*, 2015) would suggest that these AT3-only proteins function with an, as yet unidentified, periplasmic partner protein. Although no candidate SGNH partner proteins have been identified for these single domain acetyltransferases (Section 3.3.1), this search is limited by the fact that SGNH domains tend to have high structural homology but low sequence identity (Leščić Ašler *et al.*, 2010). Therefore, it may be difficult to identify SGNH domain proteins within bacterial genomes at a protein sequence level. Additionally, as it has been shown that an AT3-only domain-containing protein can donate acetyl groups to a periplasmic SGNH protein which usually interacts with a membrane bound acetyltransferase from a distinct family of transferases (Moynihan and Clarke, 2010), it is not unreasonable to suggest that these AT3-only proteins may interact with an alternative periplasmic acetyltransferase that belongs to a protein family other than SGNH.

Multiple sequence alignments identified conserved residues within the AT3 domain which were therefore hypothesised to be fundamental to the mechanism of all AT3 domain-containing acyltransferases. Significantly, the R/K-X_{10}-H motif in TMH1 was highly conserved across all AT3 domain-containing proteins, suggesting that this motif may play a conserved role in acyl group transfer. Residues with potential roles in defining the structure and interactions of transmembrane helices were also predicted and residues previously shown to be important for AT3 acetyltransferase function were confirmed as conserved in OafA and OafB.

This analysis also identified residues which were specifically invariant or conserved in the AT3 domain of AT3-SGNH fused proteins, highlighting adaptation of these AT3 domains to function with their fused periplasmic partner. Indeed, the GG-F/Y-XGV-D/P/V motif located towards the periplasmic side of the inner membrane could suggest an SGNH domain interaction site or contain catalytic residues involved in the transfer of the acetyl group between the two domains. A role which was also suggested for further conserved residues identified in the longer periplasmic loop between TMH2-3.

With respect to the SGNH domain, *in silico* analysis was able to confirm the presence of expected catalytic triad residues within the SGNH domain of OafA and OafB, however

the other catalytic blocks typical to SGNH domain-containing proteins were not highlighted (Akoh *et al.*, 2004). Pairs of cysteine residues predicted to form disulphide bonds in the oxidising environment of the periplasm were highlighted as a conserved feature within the aligned SGNH domains. These putative pairs of disulphide bonding cysteine residues have not previously been highlighted as conserved features of AT3-SGNH fused proteins and were predicted to be required for structural stability. Indeed, in the extracellular lipase from *Streptomyces rimosus,* which belongs to the SGNH hydrolase superfamily, three pairs of disulphide bonds were proposed to stabilise the SGNH fold (Leščić Ašler *et al.*, 2017).

Residues in the SGNH domain responsible for interaction with the acceptor substrate or the AT3 domain were not highlighted by the *in silico* analysis conducted. This is likely due to the fact that SGNH domains tend to have high structural homology but low sequence identity (Leščić Ašler *et al.*, 2010) and also because the aligned sequences belong to proteins which are responsible for acetylating a variety of carbohydrate acceptors (**Figure 3.3**). The residues integral to each specific modification are by definition unlikely to be identified as conserved between proteins that modify diverse substrates, therefore structural characterisation of the SGNH domain would be the best approach to identify candidate residues required for the mechanism of LPS O-antigen acetylation.

In order to investigate the functional importance of conserved residues which were identified by *in silico* analysis, a functional assay for these proteins was required so that point mutants of OafA and OafB could be functionally assessed. Although a functional assay for OafB had been described previously (Kintz *et al.*, 2015) and the serotype conversion resulting from abequose acetylation by OafA is detectable by serotyping antibodies (Slauch *et al.*, 1996), these assays had not been optimised for quantification of protein function. Therefore, the potential advantages and pitfalls of optimisation of each functional assay were investigated so that the best target for initial site directed mutagenesis could be chosen. This, and subsequent site directed mutagenesis and functional analysis of OafA and OafB is outlined in Chapter 4.

Chapter 4 Optimisation of an *in situ* functional assay and analysis of OafA functional residues

4.1 Introduction

OafA and OafB were targeted for mechanistic characterisation of AT3 domain-containing O-acetyltransferases as they are two different O-antigen acetyltransferases, from the same species and with the same functional domains, but which differ in the monosaccharides they modify. Additionally, for both systems their *in situ* function can be assayed due to the biological phenotypes that these O-antigen modifications infer (discussed further below). These properties give the potential for conclusions that are made for one protein, to be backed up by assessment in the other and they also provide a system for investigating the way that these proteins achieve substrate specificity.

OafA acetylates the abequose residues of the O-antigen at the 2 hydroxyl group and this modification defines the O:5 serotype of *Salmonella* (Slauch *et al.*, 1996). This means that specific O:5 serotyping antibodies can recognise the O-antigen when abequose monosaccharides have been acetylated by functional OafA. This antibody recognition is then lost when OafA is absent or non-functional. Conversely, acetylation of rhamnose by OafB is not detected by serotyping antibodies. However, this modification does confer resistance to lysis caused by the BTP1 bacteriophage, whose genome it is encoded on (Kintz *et al.*, 2015). This allows resistance to BTP1 phage lysis to be the readout of function of the OafB protein.

Assessing the enzyme activity of carbohydrate acetyltransferases presents a range of challenges as they are membrane bound proteins with a periplasmic domain. Expression, purification and reconstitution of integral membrane proteins into appropriate model membranes, for enzymatic characterisation, can be challenging and the composition of these model membranes can affect the activity of the protein (Shen *et al.*, 2013). Also, the fact that these proteins modify specific carbohydrate molecules which are not readily available as purified substrates, restricts the potential to assess the activity of these proteins against their physiological substrates. Therefore, negating the need for isolation of these proteins from their biological context and assessing their function in the native membrane would be highly valuable.

In situ functional assays were previously developed for both OafA and OafB and there were pros and cons for implementation of both assays for further characterisation of O-antigen acetyltransferases. A previously generated mutant of STM strain LT2 has all known glucosyltransferase operons (Gtr family III and Gtr family IV) and the o*afA* gene deleted (Path293, **Table 2.1**) (Davies *et al.*, 2013). This strain expresses a basal STM O-antigen with a standard, unmodified O-antigen repeating unit (**Figure 1.3**). The intention was to use this basal O-antigen strain to express point mutant variants of OafA or OafB *in trans* and assess their function, to identify functional residues within these proteins.

This chapter will outline the functional assays available for both proteins at the start of this project, as well as the logic of selection and optimisation of the OafA functional assay. This is followed by use of the optimised assay for characterisation of OafA functional residues from candidates identified through *in silico* analysis in Chapter 3.

4.2 Assessment of current approaches to functional analysis of *Salmonella* O-antigen acetyltransferases

4.2.1 Phage plaque assays for functional assessment of OafB

The functional assay for OafB involves administering aliquots of BTP1 bacteriophage suspension to a lawn of the STM test strain and observing clearance of the bacterial cells from the area that the phage was administered. These zones of clearance, or plaques, are a result of bacterial cell lysis by the BTP1 bacteriophage (Kintz *et al.*, 2015). A plaque is formed when a single bacteriophage infects a bacterium and starts its lytic cycle, resulting in subsequent infection and lysis of the surrounding cells, until a zone of clearance can be seen by eye (Abedon and Yin, 2009). If the STM test strain expresses functional OafB, the O-antigen will be decorated with enough acetylated rhamnose residues to prevent infection and subsequent lysis by the BTP1 bacteriophage, therefore no zones of clearance will be observed. However, if a non-functional variant of OafB is expressed or OafB expression is absent in the test strain, zones of clearance will be seen in the area that bacteriophage is administered (**Figure 4.1**).

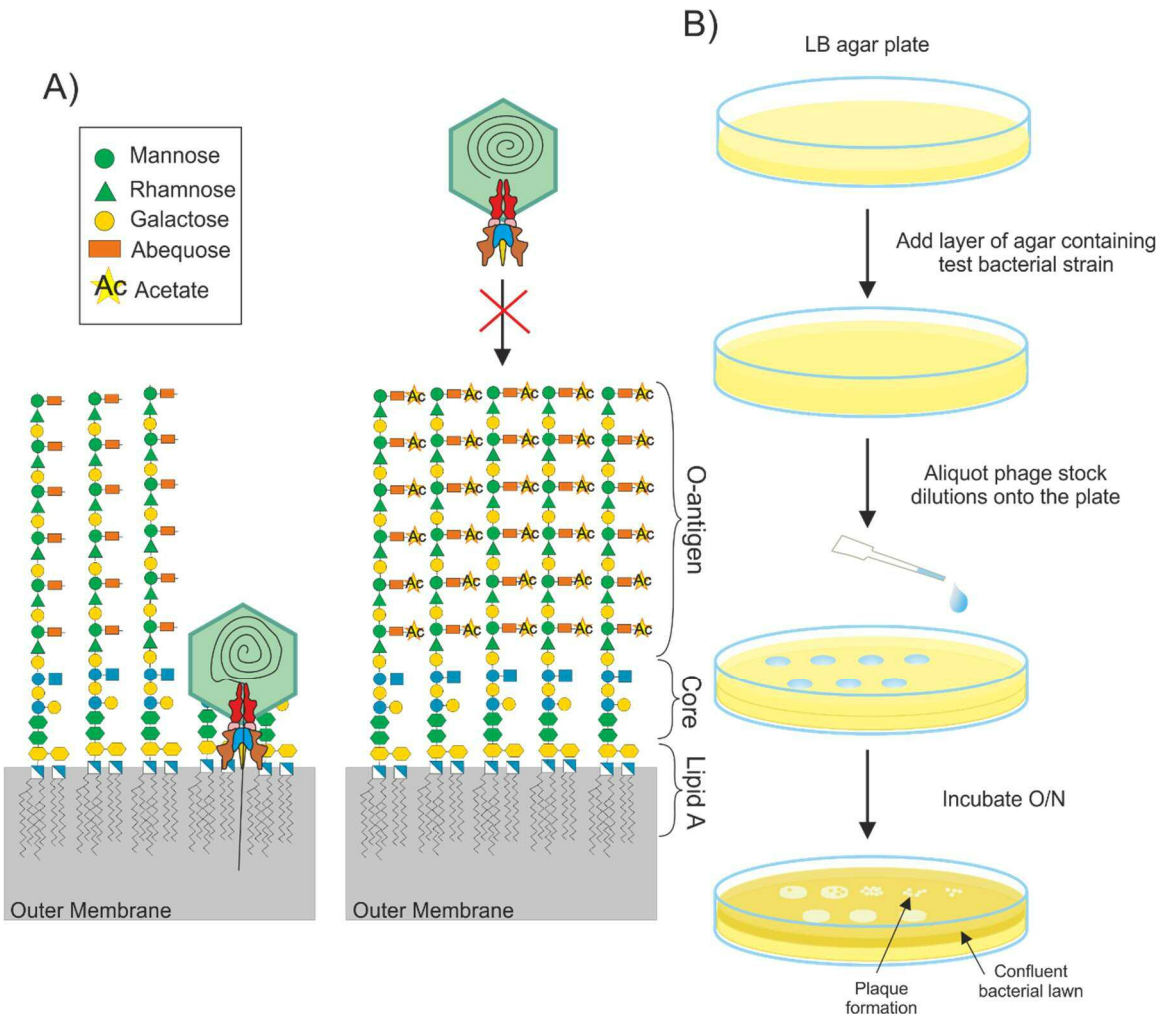


Figure 4.1|Phage plaque assay for functional analysis of OafB. A) Acetylation of O-antigen rhamnose residues by OafB prevents infection with BTP1 bacteriophage therefore preventing bacterial cell lysis. Diagram of BTP1 Phage is adapted from Tang *et al*., (2011) and Owen *et al.,* (2017). The phage particles are not to scale and are represented this way for demonstrative purposes. Glycan symbols for LPS molecules are in accordance with SNFG (Hart *et al.*, 2015). B) The phage plaque assay utilises the property illustrated in (A) by testing the ability of BTP1 bacteriophage suspensions to cause lysis of the test *Salmonella* Typhimurium strain expressing OafB or point mutated versions thereof.

The simple methodology of this assay provides the advantage that many test strains can be screened in parallel, however the drawback is that it is not a direct quantitation of rhamnose acetylation on the O-antigen, as it uses phage resistance as a proxy. The level of rhamnose acetylation required for phage resistance as well as the location of acetylated repeats within the O-antigen polymer is not well characterised. NMR characterisation of O-antigen rhamnose acetylation levels in wild type *Salmonella* possessing the *oafB* gene are measured to be ~80% (Micoli *et al.*, 2013). It is not known at what percentage of O-antigen rhamnose acetylation the bacteria become susceptible to lysis. Therefore, although variations in the susceptibility of test strains to BTP1

infection may be seen, the percentage decrease in activity of OafB which results in loss of BTP1 phage resistance is unknown. For this reason, the phage plaque assay may not be the most appropriate means for quantitation of function of OafB point mutants. Thus, the potential for quantitative assessment of abequose acetylation by OafA using O:5 serotyping antibodies was explored.

4.2.2 O:5 serotyping antibodies for functional analysis of OafA

The functional assay for OafA uses O:5 serotyping antibodies to detect acetylated abequose residues on the LPS O-antigen. These serotyping antibodies are widely used for surveillance of *Salmonella* outbreaks and are used in a hierarchical process to narrow down the specific antigenic properties of *Salmonella* isolates. Reactivity of serotyping antibodies with specific strains is determined by an agglutination assay, where *Salmonella* suspensions will aggregate upon serotyping antibody binding due to antibody cross-linking. As O-antigen acetylation by OafA is the sole cause of serotype switching to the O:5 serotype (Slauch *et al.*, 1996), slide agglutination with O:5 serotyping antibodies can be used as a readout for OafA function. This was confirmed by loss of O:5 serotyping antibody agglutination in the basal STM O-antigen strain, in comparison to wild type STM LT2, or the basal STM strain complemented with plasmid expressed OafA (**Table 4.1**). Although this assay has the potential to confirm OafA activity, it cannot be used as a quantitative assay for the level of OafA acetyltransferase activity due to the binary readout of this results.

Table 4.1| Slide agglutination of *Salmonella* ser. Typhimurium test strains with serotyping antibodies OMA, O:4 and O:5. Strain details for each sample can be found in **Table 2.1**.

Sample	Strain	Slide agglutination result		
		OMA	O:4	O:5
WT	Path 346	+	+	+
Basal	Path 932	+	+	-
Basal + *oafA*	Path 993	+	+	+

O:5 serotyping antibodies can instead be used in an LPS immunoblot, where crude LPS extracts from STM test strains are run on TSDS-PAGE, transferred to PVDF membrane, and probed for O:5 antibody binding. To compare O-antigen acetylation levels between

different test strains, duplicate TSDS-PAGE gels are run, one of these is transferred to PVDF membrane for immunoblotting and the other is treated with silver stain reagents to visualise the total LPS sample loaded. To corroborate O-antigen acetylation with OafA protein expression, OafA protein is expressed from the multi copy, arabinose inducible, pBADcLIC expression plasmid which adds a C-terminal 10xHis tag to allow detection of OafA expression by Western blot (Appendix V).

Preliminary testing confirmed that plasmid encoded *oafA* could rescue O:5 antibody binding to LPS in basal STM (**Figure 4.2**). This complementation occurred without arabinose addition to the cells, corresponding with a faint positive signal for full length OafA in the insoluble protein extracts (**Figure 4.2 C,D**). The apparent 'leaky' expression of OafA from the pBADcLIC vector is likely due to the cells being cultured in rich media which is not explicitly defined in its compositions. Therefore, there is potential for the growth media to contain traces of arabinose. This is supported by the observation that arabinose inducible vectors in bacterial cells grown in rich media do not give complete repression of protein expression in the absence of added arabinose (Guzman *et al.*, 1995).

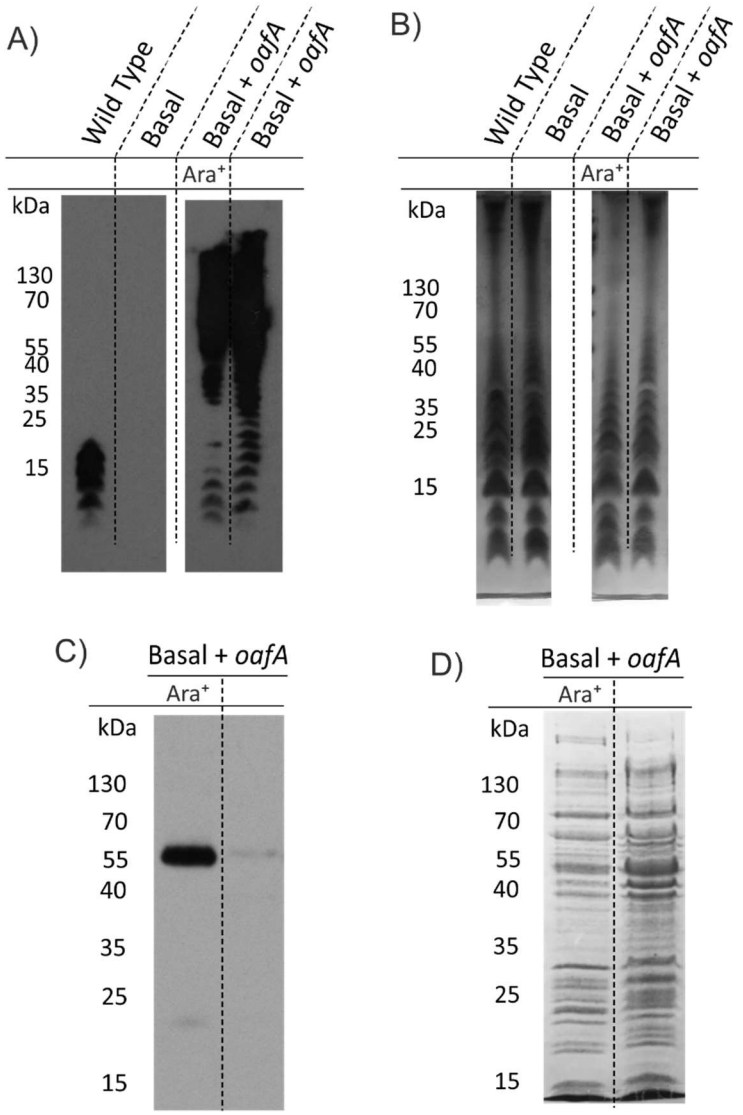


Figure 4.2|Preliminary analysis of the OafA functional assay. Basal = Basal STM + pBADcLIC (Path 993), Basal + *oafA* = Basal STM LT2 + pBADcLIC_*oafA* (Path 932), Wild type = Wild type STM LT2 (Path 346). Cultures of *Salmonella* test strains were grown for 4 hrs with or without 0.1% w/v arabinose induction at OD_{600} 0.5. 3 ml culture was subject to crude LPS extraction and 1 ml subject to insoluble protein isolation. Crude LPS extracts were run on TSDS-PAGE with sample loading normalised to harvest OD_{600}. A) LPS samples from one gel were transferred to PVDF membrane for immunoblotting with O:5 serotyping antibodies. Ara^+ indicates arabinose induction. B) A duplicate LPS gel was subject to silver staining to visualise total LPS loaded. C) OafA expression from the pBADcLIC plasmid was confirmed by anti 10xHis western blot with the insoluble protein faction and (D) protein loading was confirmed by Coomassie stained gel.

Wild type STM LT2 (Path346) displayed a different profile of O:5 antibody binding in the LPS immunoblot, with apparent lack of O-antigen acetylation in the longer O-antigen repeats, even though the full profile of LPS lengths could be seen by silver staining of this sample (**Figure 4.2 A,B**). This variation could be caused by the different genetic background of this strain as it still has functional O-antigen modification *gtr* operons in its genome.

Although LPS immunoblotting paired with protein Western blotting is a more direct method of visualising OafA activity and confirming OafA expression, the lab protocol in use for LPS immunoblotting was not optimised for quantification of OafA activity. Significantly, there is no internal control on the LPS immunoblot to allow normalisation across sample signals for their comparison (**Figure 4.2 A**). This means that sample loading and transfer efficiency cannot be considered in signal quantification. Silver staining does allow confirmation that LPS sample was loaded into the gel and that its O-antigen profile is comparable to other samples (**Figure 4.2 B**). However, using the silver staining to normalise LPS loading for the O:5 immunoblot is not appropriate as it is not possible to determine whether LPS samples have all been transferred with the same efficiency to the PVDF membrane for immunoblotting. Additionally, during preliminary analysis it was difficult to reproduce consistent results with LPS silver staining and some experimental repeats showed very feint and unclear LPS bands (**Figure 4.3**). This may have been caused by sensitivity of the silver stain to impurities in the crude LPS sample. These observations suggest that incorporating an internal loading control within the LPS immunoblot would be highly beneficial to enable quantitative comparison of O-antigen acetylation between *Salmonella* test strains.

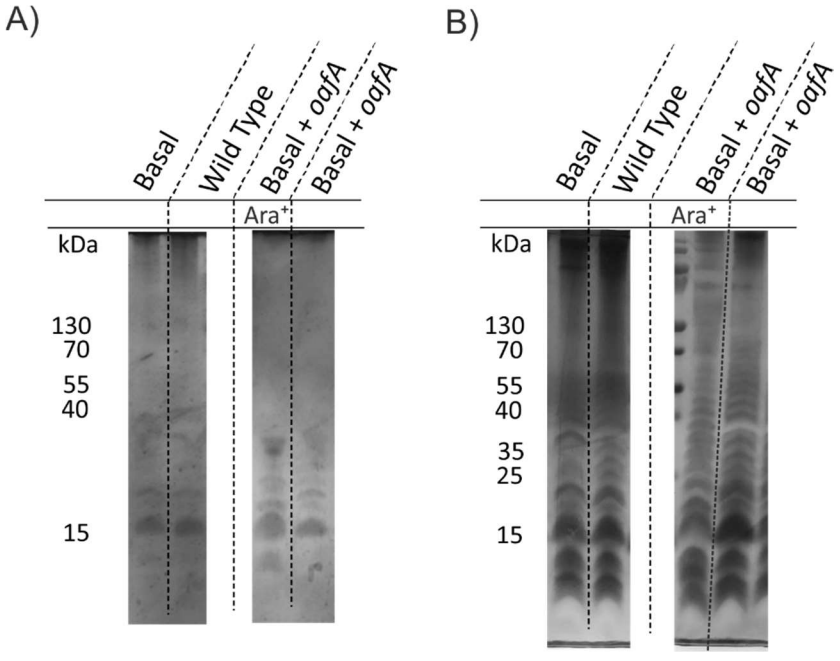


Figure 4.3| Inconsistency in LPS silver staining results. Sample details are the same as **Figure 4.2**. A-B show two repeats of the same procedure for LPS visualisation in TSDS-PAGE gels with identical samples and sample volumes run on separate gels on separate days. Ara^+ indicates cultures grown with arabinose induction.

4.3 Optimisation of LPS immunoblot for functional analysis of OafA

4.3.1 Identification of an internal LPS loading control for O:5 antibody immunoblotting

STM LT2 O-antigens should be reactive with O:4 and O:12 serotyping antibodies irrespective of their acetylation state determined by *oafA* (Hellerqvist *et al.*, 1969; Grimont and Weill, 2008). To this regard, O:4 serotyping antibodies were trialled as a control for LPS loading and transfer onto PVDF membrane in LPS immunoblots. The negative control in these experiments was Basal STM expressing a catalytically inactive point mutant of OafA from the pBADcLIC2005 vector (Basal + $oafA_{H590A}$), discussed further in section 4.4. Surprisingly, O:4 serotyping antibody binding appeared to correlate with acetylation of the O-antigen. O:4 serotyping antibody binding was almost undetectable for the basal, unmodified STM LPS, with just a faint signal in the sample without arabinose induction (**Figure 4.4**). In contrast, slide agglutination assays confirmed that all test strains from this experiment did agglutinate with O:4 antiserum and conversely, O:5 antiserum slide agglutination correlated with the results of the LPS immunoblot (**Figure 4.4 A, Table 4.1**). Despite possessing common O-antigen factors, serogroup O:4 *Salmonella* strains have been demonstrated to display varying affinity for O:4 antibodies when assessed by agglutination with serial dilutions of antiserum (Aribam *et al.*, 2015). This could explain the apparent lack of O:4 antibody binding in the LPS immunoblot.

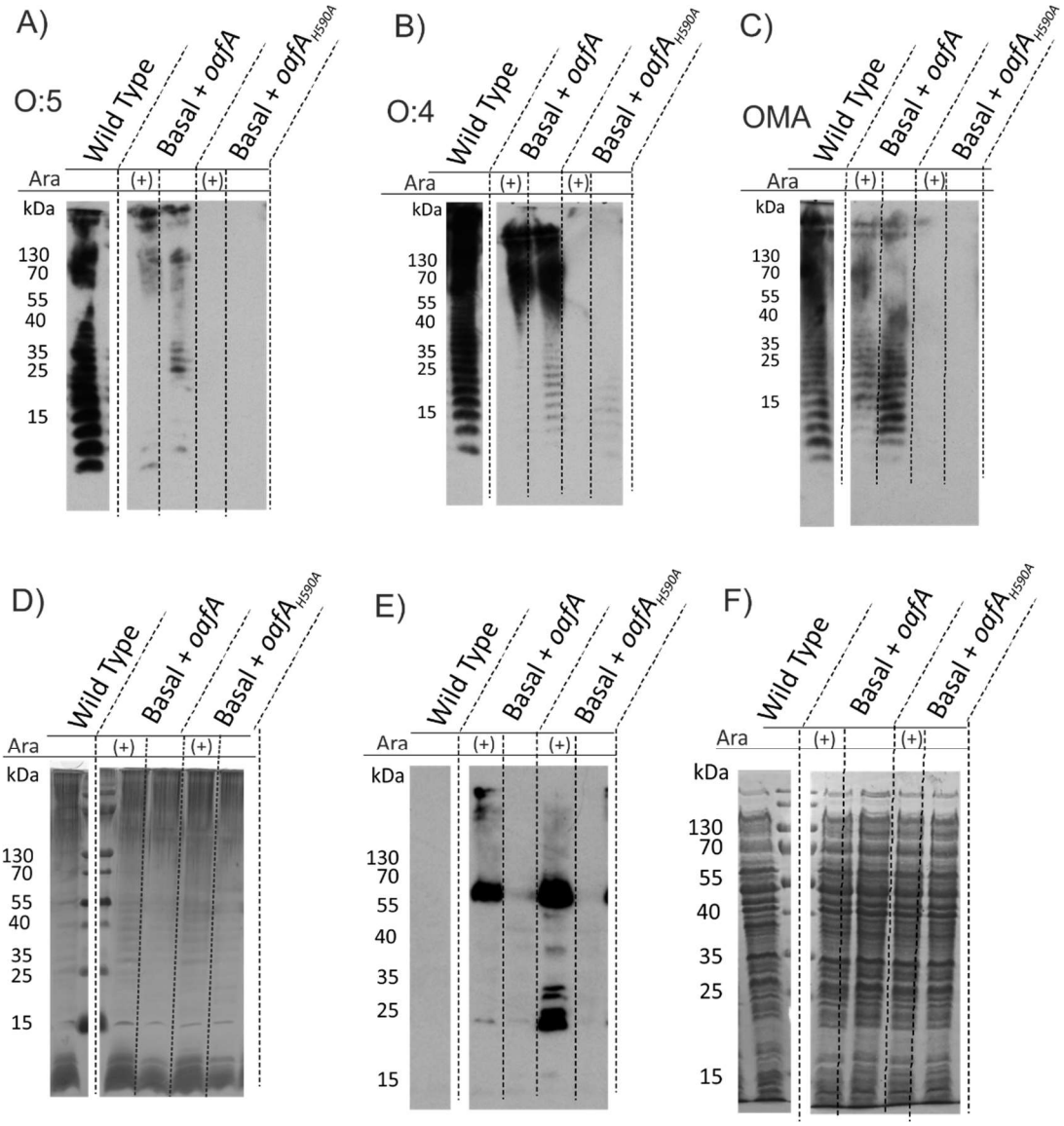


Figure 4.4| O:4 and OMA serotyping antibodies do not show uniform binding to acetylated and un-acetylated LPS. Wild type = Wild type STM LT2 (Path 346), Basal + *oafA* = Basal STM LT2 + pBADcLIC_*oafA* (Path 932), Basal + $oafA_{H590A}$ = Basal STM LT2 + pBADcLIC_$oafA^{H590A}$ (Path 1055) (negative control of catalytically inactive OafA protein). Cultures of *Salmonella* test strains were grown overnight with or without 0.1% w/v arabinose induction at OD_{600} 0.5. All samples grew to similar OD_{600} ~3.8, therefore 1 ml culture was subject to crude LPS extraction and 1 ml was subject to insoluble protein isolation. 15 µl crude LPS extracts were run on TSDS-PAGE. Ara (+) indicates arabinose induction. A-C) LPS immunoblot with (A) O:5, (B) O:4, and (C) OMA serotyping antibodies. D) Silver stain LPS gel to show uniform LPS loading. E) OafA expression from the pBADcLIC plasmid was confirmed by anti 10xHis western blot with the insoluble protein faction. F) Protein loading was confirmed by Coomassie stained gel.

As STM strains fall under the wider designation of O group B, they are expected to interact with OMA antiserum, which is used for preliminary identification of groups A, B, D, E, and L. Slide agglutination results confirmed this (**Table 4.1**). As OMA antiserum is higher up the serotyping hierarchy and recognises a range of different *Salmonella* O-

antigen structures (SSI Diagnostica, 2013) it was also trialled in the LPS immunoblot, as an alternative to O:4, for its ability to determine the presence of LPS in all lanes regardless of O-antigen acetylation state. Surprisingly, as with O:4 serotyping antibodies, OMA antibodies gave the same binding profile on the LPS immunoblot as O:5, with loss of signal for strains expressing catalytically inactive OafA which results in basal, un-acetylated, Typhimurium O-antigens (**Figure 4.4 C**).

Antibody binding in this assay may have been affected by the experimental conditions. Firstly, LPS immunoblotting could alter the affinity of serotyping antibodies as LPS is presented in a different context than it is for the slide agglutination assays which these antibodies are developed for. Also, acetylation may improve the efficiency of LPS transfer or binding to the PVDF membrane. PVDF membrane has a positive surface charge making it hydrophobic, as acetylation adds a negative charge to LPS it could result in improved transfer or binding of the LPS to the membrane.

For these reasons, an alternative technique of an LPS dot blot was trialled, where LPS samples were spotted and dried onto nitrocellulose membrane then probed for O:5 and OMA serotyping antibody binding (Section 2.6.2.3.1). These dot blot results correlate with slide agglutination assays, however, OMA antibody binding was reduced for O:5 negative samples (**Table 4.1**, **Figure 4.5**).

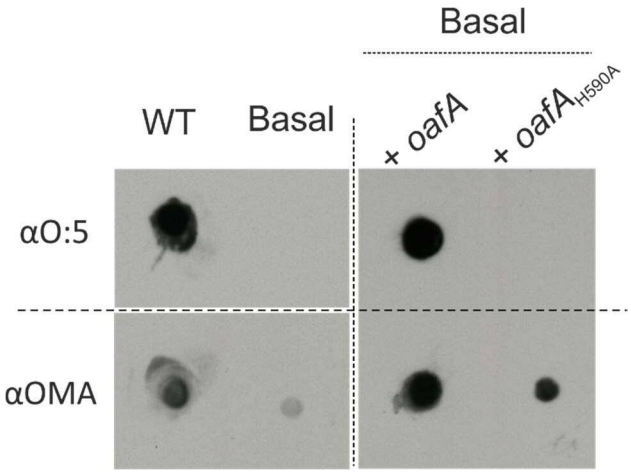


Figure 4.5| LPS dot blot using O:5 and OMA serotyping antibodies. 3 µl LPS samples from **Figure 4.2** (WT, Basal) and **Figure 4.4** (+*oafA,* +*oafA*^{H590A}) were spotted onto nitrocellulose membrane. After air drying, the membranes were blocked with 5% Milk PBS-T and probed with OMA or O:5 serotyping antibodies (Section 2.6.2.3.1).

This positive DOT blot result confirms that OMA serotyping antibodies will bind to all test sample LPS, however, it is not suitable as an internal LPS loading control due to

inconsistencies in binding affinity against acetylated and un-acetylated LPS (**Figure 4.4**). Therefore, alternative LPS antibodies were sought out that give more uniform LPS binding regardless of O-antigen acetylation.

A review of the literature uncovered evidence of a *Salmonella* core antigen (sc-52219, 5D12A), used successfully by Rondini *et al.,* (2013), which recognises the core oligosaccharides of *Salmonella* serogroups A, B, C1, C2, D, E1 and E2. As this antibody is raised in mice rather than rabbits it provides the potential for double immunostaining within the same PVDF membrane because *Salmonella* serotyping antibodies are raised in rabbits. To allow differentiation between O:5 and anti-*Salmonella* core antibody binding, fluorescent anti-Mouse and anti-Rabbit antibodies with divergent emission spectra were trialled in place of the HRP-conjugated antibodies used in previous experiments. Basal STM LT2 possessing either the empty pBADcLIC vector (Path 993) as a negative control or pBADcLIC_*OafA* vector (Path 932) as a positive control were used as the test strains for positive and negative O-antigen acetylation.

Double antibody staining with O:5 and anti-core antibodies was successful and demonstrated that the signal for anti-core antibody binding remains constant, irrespective of the abequose acetylation profile of the LPS samples (**Figure 4.6**). An equivalent silver stain LPS gel was used to confirm uniform LPS loading across all lanes (**Figure 4.6 D**), further supporting the use of the anti-S*almonella* Core antibody as an internal LPS loading control for *in situ* functional assessment of OafA. Now that a suitable internal LPS loading control had been found, quantification of LPS acetylation levels with this two-colour fluorescent antibody immunoblot was investigated.

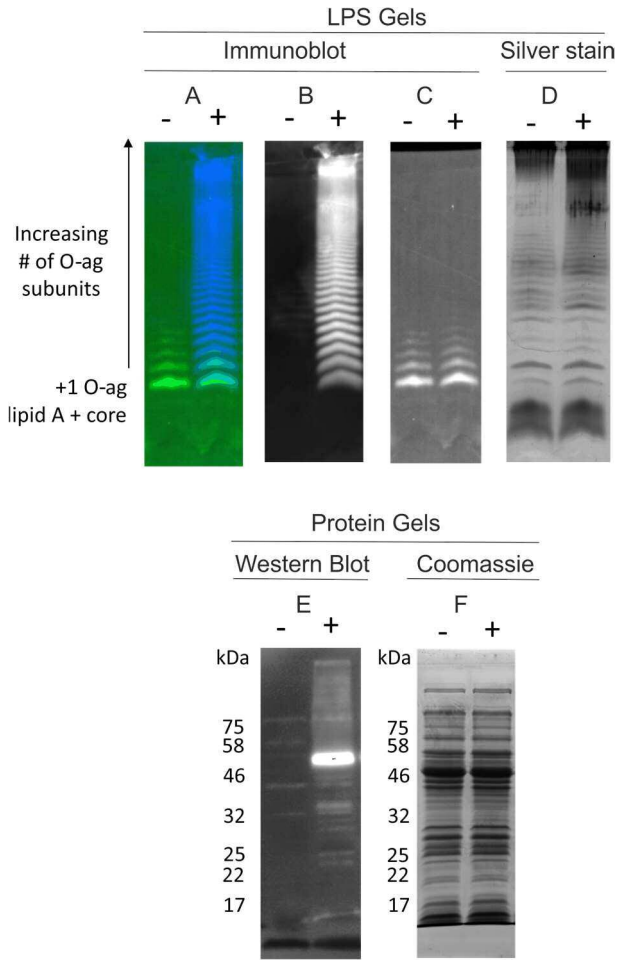


Figure 4.6| Anti-core serotyping antibodies as a loading control for LPS immunoblots. A) Two colour LPS immunoblot with O:5 (blue) and anti-Core (green) antibodies against LPS from (-) Basal STM LT2 + empty pBADcLIC vector (Path 993), (+) Basal STM LT2 + pBADcLIC_*oafA* (Path 932). B) O:5 only signal. C) Anti-Core only signal. D) Silver stain of the same samples as (A). E-F) 10xHis western blot and Coomassie stained gel of insoluble protein fraction respectively from the same strains as (A). Experimental conditions and LPS sample preparation are described in section 2.6.2.1.

4.3.2 OafA expression from the pBADcLIC vector is higher than chromosomal expression and complements OafA function without arabinose addition

Addition of 0.1% w/v arabinose at OD_{600} 0.5 for induction of OafA protein expression from the pBADcLIC_*oafA* vector resulted in a significant reduction in cell growth of Path 932 (T=4 hr, Induced OD_{600} = 1.35, Un-induced OD_{600} = 2.95). Cells containing the pBADcLIC_*oafA* plasmid gave a positive signal for O-antigen acetylation, and for His-tagged protein expression, regardless of whether OafA expression was induced by arabinose addition (**Figure 4.2**, **Figure 4.4**). Therefore, the level of OafA expression from the pBADcLIC_*oafA* plasmid in the absence of arabinose was compared to chromosomal

levels of OafA expression from wild type STM LT2 cells, to determine whether these were appropriate conditions for functional comparisons of OafA point mutants.

A 10xHis tag was added to the chromosomal copy of OafA using lambda red recombination (Section 2.5.14). Comparison of OafA expression levels from the chromosome and expression plasmid indicate that, in the absence of arabinose addition, the pBADcLIC_*oafA* plasmid gives a higher level of OafA protein expression than that which is expressed from the chromosome in wild type cells (**Figure 4.7 B**). In previous analysis the un-induced level of OafA expression was barely visible on the western blot in comparison to OafA signal from induced cells (**Figure 4.2**). However, this comparatively weak signal of OafA expression from the pBADcLIC plasmid, in the absence of arabinose induction, is significantly stronger than that from wild type cells and is sufficient to complement WT OafA function (**Figure 4.7**). Therefore, for subsequent *in situ* OafA functional assays using plasmid complemented *oafA,* arabinose induction was no longer used because as long as a positive signal for OafA protein could be seen by western blotting, there was sufficient protein to complement O-antigen acetylation.

Although it is possible to indicate OafA protein expression levels from experimental strains, the levels of OafA protein were not considered in our analysis. This is because plasmid expression levels of OafA are in high excess to the level of chromosomally expressed OafA from wild type cells and the lower level of expression is already sufficient to give full O-antigen acetylation (**Figure 4.7 A**). Therefore, it is not possible to attribute small alterations in protein levels to alterations of O-antigen acetylation within these assay conditions as the protein will still be in excess.

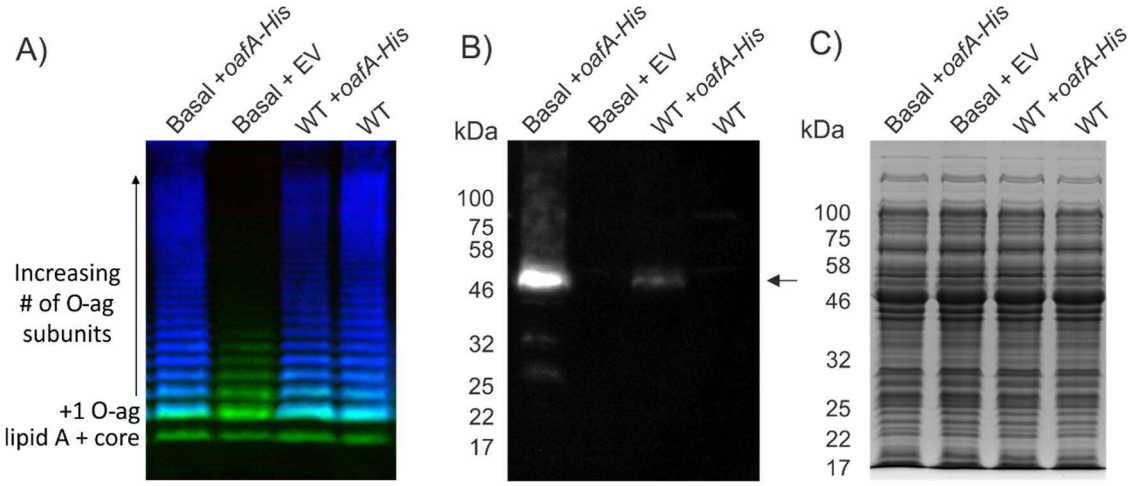


Figure 4.7| Chromosomal vs plasmid expression of OafA.All cultures were grown in the absence of arabinose. A) LPS western blot with crude LPS extracts from STM LT2 test strains: LT2 basal O-antigen strain expressing OafA from pBADcLIC plasmid (Basal + *oafA-His*) or containing an empty pBADcLIC plasmid (Basal + EV), LT2 WT O-antigen strain with a C-terminal 10xHis tag added to the chromosomal copy of OafA (WT + *oafA-His*) or the same strain with unmodified OafA (WT). O:5 antibody binding (Blue) shows abequose acetylation and loading control in green binds the *Salmonella* LPS core. B) Corresponding anti-His western blot and (C) Coomassie stained SDS-PAGE of insoluble protein fraction for detection of His tagged OafA. Arrow indicates full length OafA protein.

The experimental system of *oafA* complementation, producing comparatively higher levels of OafA protein to wild type cells, could prevent detection of point mutations that have impacts on different enzyme kinetic parameters of OafA. Reduction in K_m (enzyme-substrate affinity) of OafA could be abrogated by the fact that there is an abundance of OafA protein present in the cell, however it should be possible to determine point mutations that effect the k_{Cat} (enzyme turnover rate) as every OafA molecule will have the same- reduced level of turnover.

4.3.3 Quantification of LPS acetylation by two colour fluorescent antibody immunoblot

To reliably quantify activity of OafA point mutants by quantitative assessment of O-antigen acetylation, the protocol in use required optimisation to ensure consistency of experimental procedures and comparability between samples. Firstly, alkalisation of media, due to metabolism of amino acids as bacterial cells reach saturation, has been shown to cause the loss of acetyl groups from the O-antigen, and buffering the media can prevent this from happening (Ilg *et al.*, 2013). Therefore, all growth media was buffered to pH 7.0 for analysis of OafA function, to prevent any variation of results caused by random hydrolysis of O-antigen acetyl groups by alkalisation of the media. Secondly, overnight culture was normalised to OD_{600} 3.0 before LPS and protein preparation. This change was implemented in place of adjusting protein and LPS sample load volumes in accordance with harvest OD_{600} at the point of analysis. The rational being that this would provide a more uniform approach to sample preparation as the same number of cells will be treated with the same buffers for every sample. Treating different numbers of cells with the same volumes of LPS and protein preparation buffers could impact on efficiency of LPS or protein extraction, therefore making it less reliable to normalise sample loading after this step is completed.

Image lab™ analysis software was used to determine the optimum parameters for quantitation of two-colour fluorescent antibody immunoblot results, following optimisation of growth and sampling conditions. Image lab™ software can be used to define the lanes on gels and plot signal intensity changes along the length of each lane, so that the total signal intensity of the whole lane or a single band within that lane can be quantified. Background signal is subtracted from the plots for each lane using the rolling disk method (**Figure 4.8**). The default settings of Image lab™ set the disk size to 70 mm however 10 mm is usually considered optimal for removal of most background (Ghosh *et al.*, 2014). For quantitation, lane width was set to 8 mm and disk size for intensity calculations to 10 mm.

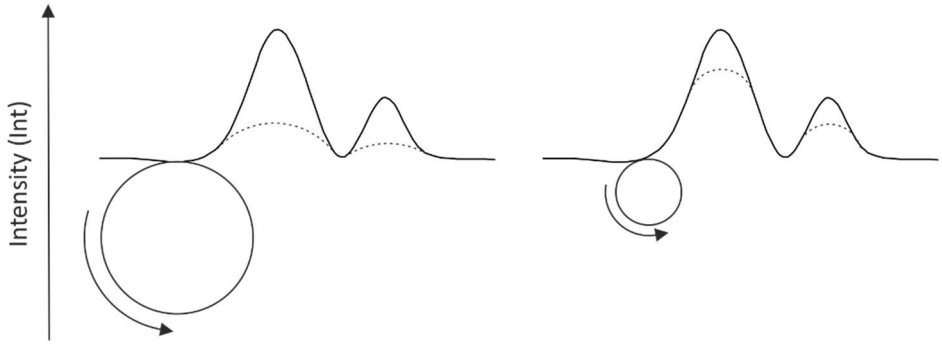


Figure 4.8| Rolling disk method of background subtraction used by Image lab™. The smaller the disk size, the more background will be subtracted from the signal. Adapted from (Gassmann *et al.*, 2009).

Serial dilutions of crude LPS extracts from Path 932 (*oafA*$^{+}$) were analysed by two colour fluorescent antibody western blot to create a calibration curve for assessment of the quantitative capabilities of the assay using Image Lab™ analysis (**Figure 4.9**). Comparison of signal intensity for O:5 and anti-core binding as a percentage of the undiluted LPS sample suggest that fluorescent antibody signal quantification gives an overestimation of the O:5 and anti-core binding, as the percentage LPS concentration calculated from Image Lab™ analysis was higher than the actual percentage concentration of LPS loaded. However, there is a positive correlation between amount of LPS loaded and antibody signal recorded (**Figure 4.9**). Whole lane quantification was used to measure the intensity of O:5 antibody binding per sample, as the distribution of acetylation between LPS with different numbers of O-antigen repeats is not known. For normalisation of the O:5 antibody signal between samples, using the αCore signal, single band quantification of the band representing LPS with a single O-antigen repeat was chosen. Using a single band for LPS quantitation makes the assumption that the distribution of O-antigen

lengths between samples is uniform. For comparison of strains expressing OafA point mutants, this assumption was considered fair due to the fact that all strains were isogenic, apart from the OafA point mutation, and would be exposed to the same conditions for growth and sample preparation. Additionally, each LPS molecule has just one core oligosaccharide region and the fact that the αCore antibodies used are monoclonal means the stoichiometry of binding to each LPS molecule is likely to be 1:1.

The aim was to use this assay to quantify the effect of OafA point mutations on O-antigen acetylation levels. O:5 antibody signal quantification gives an overestimation of the abequose acetylation signal, this assay can therefore be used to identify OafA point mutants with significant reduction in protein function but may not identify point mutants that cause a relatively small reduction in protein function.

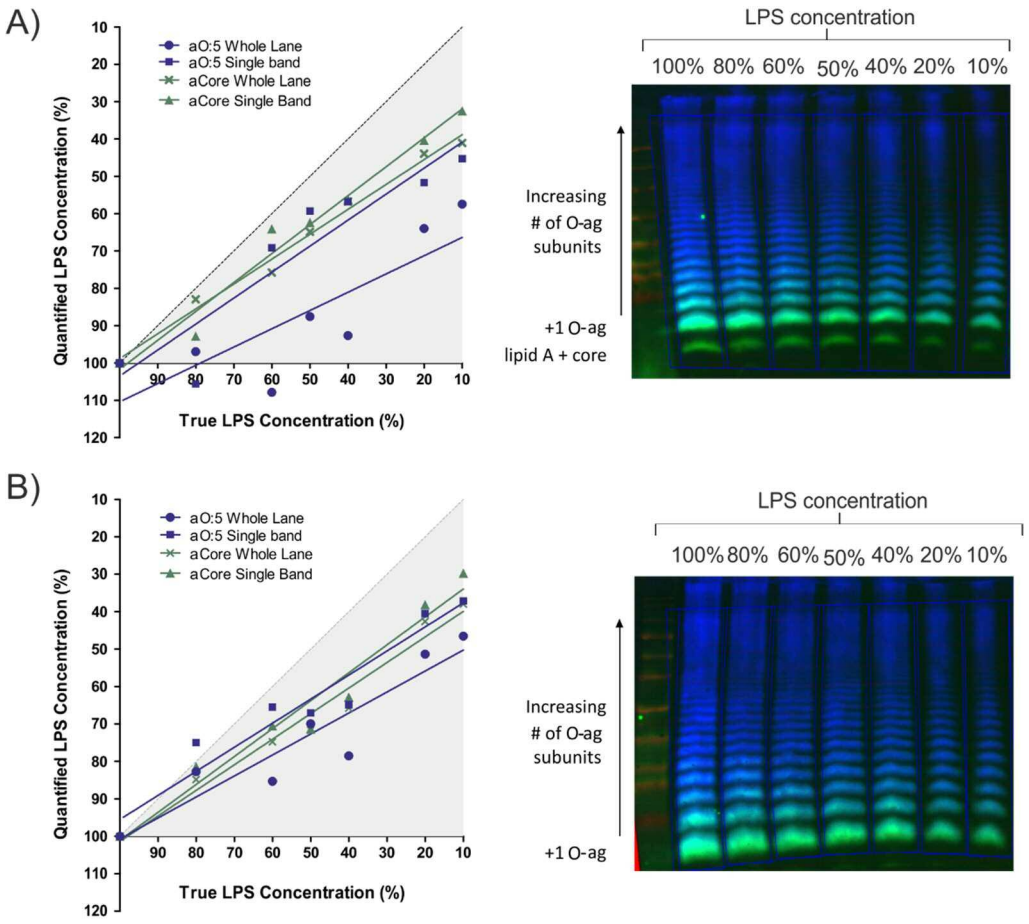


Figure 4.9| Validation of the LPS immunoblot. Quantification of LPS concentration in crude LPS preparations from Basal STM +pWT OafA. (A) and (B) are the same samples run and analysed in parallel. LPS concentration was calculated after quantification of signal intensity using ImageLab software (BioRad) from: O:5 antibody signal (blue) and *Salmonella* αCore antibody signal (green), from the whole lane (green crosses or blue circles) and a single band (+1 O-ag) (green triangles or blue squares). Results are plotted as a percentage of the signal intensity from the 100% concentration lane. The

grey dashed line represents expected values and the shaded region of the graph indicates results giving an over estimate of the LPS concentration. Right panel shows the quantified image with lane boundaries as blue boxes. The stacking gel section of the blot (top section) was not included in the lane boundaries for quantification. LipidA + Core band was not visible in repeat (B) therefore not included in the lane boundary. Single band quantification was repeated for LPS with >1 O-ag repeat and showed a comparable trend (data not shown).

Immunoblotting and silver stain of LPS were used to estimate the concentration of LPS in crude preparations (**Figure 4.10**). Purified and lyophilised LPS from STM LT2 was purchased to produce a known standard concentration series of LPS for comparison of signal intensity from immunoblotting and silver stain gels to the unknown crude LPS sample. Estimates of LPS yield as a percentage of cell dry weight by Darveau and Hancock (1983) (4.7%), and the cell dry weight per OD_{600} unit of *E. coli* from Glazyrina *et al.,* (2010) (0.39 g/L), were applied to give a rough estimate of the expected LPS concentration in crude preparations of 0.11 mg/ml. This result was reached using the following calculations:

Estimated LPS per 1 OD unit = 4.7% of dry cell weight

Estimated dry weight of Salmonella per 1 OD unit = 0.39 g/L

Mass of LPS harvested per 1 OD unit of Salmonella per L = $0.39 \times 0.047 = 0.01833$ *g*

Mass of LPS harvested per sample (1 ml OD_{600} 3.0) = $\frac{0.01833 \times 3}{1000}$ = 5.499×10^{-5} *g*

$\therefore \times 1000 = 0.05499$ *mg*

Concentration of LPS per ~500 µL sample = $0.05499 \times 2 = 0.10998$ *mg/ml*

A two-fold serial dilution series of pure LPS starting at 2 mg/ml was compared by LPS immunoblot and silver stain techniques to crude LPS extracts from basal O-antigen strain Path 932. Four T-SDS PAGE gels were set up in parallel for quantification, two of these were subject to silver stain and the other two were transferred to PVDF membrane for immunoblotting. Silver staining is a highly sensitive assay which can be used to visualise a range of molecules in polyacrylamide gel.

Calibration curves of the single band intensities for commercial LPS dilutions were plotted and the concentration of crude Path932 LPS was interpolated from this calibration curve using GraphPad Prism version 5.00 software, GraphPad Software, La Jolla California USA, www.graphpad.com. Bands corresponding to LPS with a single O-

antigen repeat were quantified for silver stain results as per previous assays. Bands corresponding to LPS with 3 and 2 O-antigen repeats were quantified for repeat 1 and 2 of immunoblots respectively, due to imperfections in the lower region of the gel. Silver stain results gave an estimated concentration of LPS per crude sample of 3.18 ng/µl and 4.18 ng/µl (**Figure 4.10 A,B**), whereas immunoblotting gave an estimated concentration of LPS per crude sample of 0.78 ng/µl and 1.99 ng/µl (**Figure 4.10 C,D**). Although silver staining is a direct quantitation of LPS that has been run in the gel, it is a very sensitive technique and impurities in the crude sample may skew the results compared to purified LPS, as silver ions can interact with any unsaturated compound with double bonds. Therefore, the concentration of LPS per crude LPS prep is estimated at 1-5 ng/µl. These results also demonstrate the high sensitivity anti-core antibodies in LPS immunoblotting as clear LPS bands could be seen when just 0.625 ng of LPS was loaded (**Figure 4.10**).

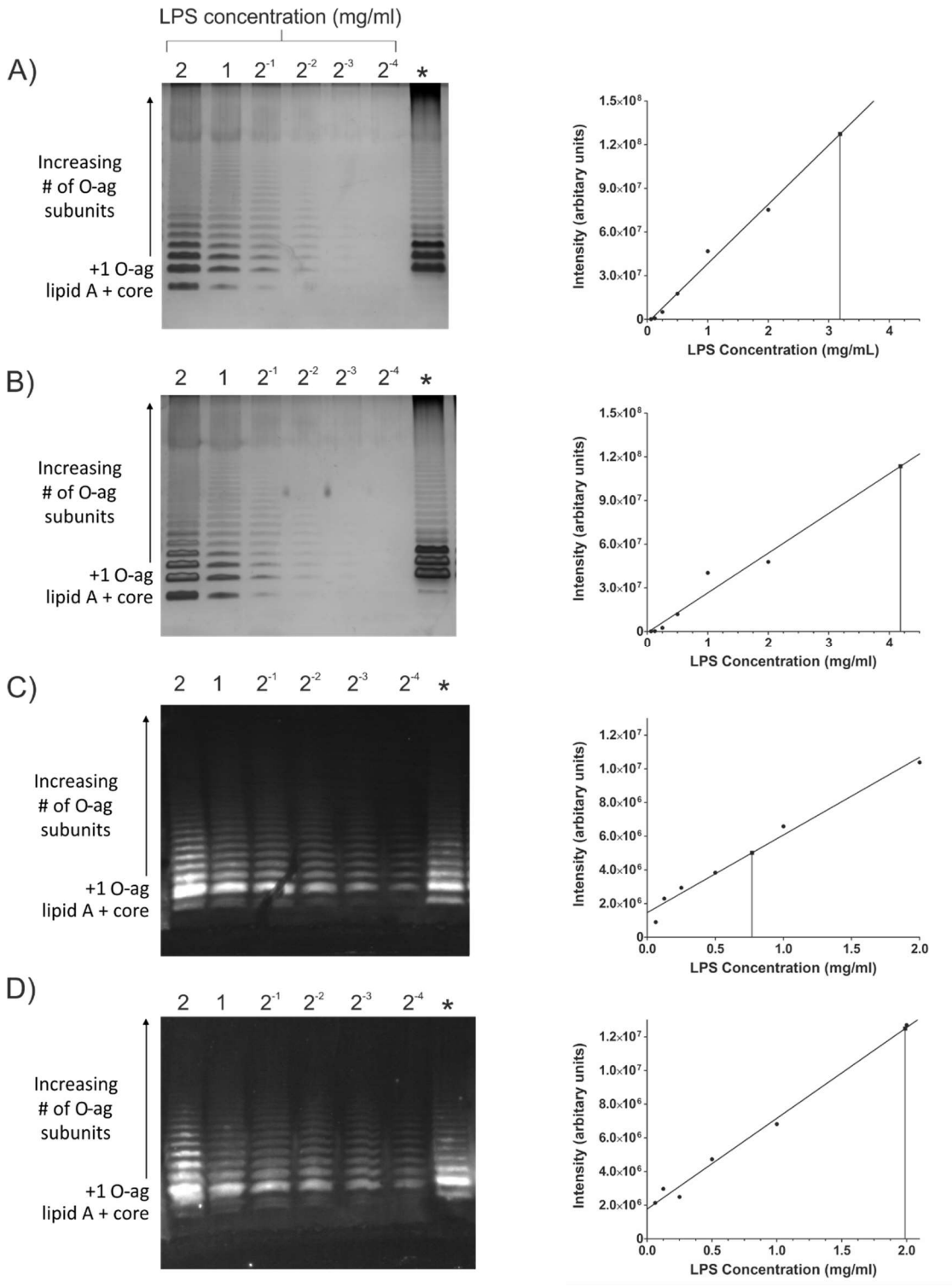


Figure 4.10| Quantification of LPS concentration in crude LPS preparations. Calibration curve of commercial STM LPS (2 mg/ml – 2^{-4} mg/ml) was set up and 10 µl of each concentration was run alongside 10 µl crude LPS extracts from Path 932 (lane denoted by a star). Right column A-B) = Parallel silver stain gels C-D) = Parallel LPS immunoblot with α *Salmonella* LPS Core antibodies. Corresponding right column shows lane intensity, quantified using ImageLab™ software (BioRad), plotted against concentration of the commercial LPS. Extrapolated crude LPS concentrations from each gel are indicated by a square.

4.4 Validation of the LPS immunoblotting protocol for assessment of OafA function

The fact that SGNH domains are soluble and can be functionally characterised using *in vitro* enzyme kinetic assays makes them more amenable to characterisation than membrane bound AT3 domains which are poorly characterised (Mølgaard *et al.*, 2000; Riley *et al.*, 2013; Baker *et al.*, 2014; Moynihan and Clarke, 2014a; Rauwerdink and Kazlauskas, 2015). OafA and OafB both display an AT3-SGNH fused domain architecture and the design of the current *in situ* OafA functional assay allows analysis of the importance of specific residues across the whole protein. Therefore, more well characterised catalytic residues within the SGNH domain were targeted by site directed mutagenesis for validation of the *in situ* OafA function assay. SGNH domains have a highly conserved catalytic triad of serine, aspartic acid and histidine (described further in section 1.2.5). Serine is the nucleophile which forms an Acyl-Ser intermediate, and aspartic acid aligns and polarises the histidine which activates the serine nucleophile (Berg *et al.*, 2002; Sychantha, Brott, *et al.*, 2018). Ser and His are strictly conserved in all functional SGNH domains and almost all Ser-Asp-His catalytic triads, however aspartic acid is not always required (Berg *et al.*, 2002; Rauwerdink and Kazlauskas, 2015). For this reason site directed mutagenesis of serine and histidine tends to abolish the catalytic activity of the SGNH domain, whereas some function can be retained in aspartic acid to alanine mutants (Lee *et al.*, 2006; Riley *et al.*, 2013; Moynihan and Clarke, 2014a).

To determine whether the difference between total functional knockout and partial reduction in function could be distinguished by this assay, O-antigen acetylation levels achieved by the catalytic triad mutants of OafA were quantified using the optimised two colour LPS immunoblot (**Figure 4.11**)(site directed mutagenesis of these residues was carried out by Reyme Herman). Full length OafA protein could be seen in the protein Western blot for all test strains (**Figure 4.11**). The predicted molecular weight of OafA with C-terminal His tag from the pBADcLIC2005 vector is 71.9 kDa (Gasteiger, Hoogland, Gattiker, S'everine Duvaud, *et al.*, 2005). However, aberrant migration is a common occurrence for transmembrane proteins in SDS-PAGE (Rath *et al.*, 2009) and full length OafA appears to run at ~50 kDa. A significant degradation band appeared around 28 kDa for OafA_S412A (**Figure 4.11 B**). This is likely a C-terminal fragment of OafA, as the His-tag is located at the C-terminal end, and could correspond to cleavage of the SGNH

domain from OafA. The molecular weight of the SGNH domain is calculated to be 25.6 kDa and the whole periplasmic portion to be 31.6 kDa using the ExPASy server (Gasteiger, Hoogland, Gattiker, S'everine Duvaud, *et al.*, 2005). Although this degradation band was seen, sufficient full length OafA was still present in the cells as a band for full length OafA could be seen on the Western blot (**Figure 4.11**).

OafA_S412A and OafA_H590A mutants resulted in less than 1% of wild type abequose acetylation, with 0.36% ± 0.3 and 0.87% ± 0.6 αO:5 antibody signal intensity compared to wild type respectively, whereas OafA_D587A retained ~10% of wild type activity (10.13% ± 1.70) (**Figure 4.11**). This is consistent with observations of the catalytic activity of catalytic triad mutants of other SGNH acetyltransferases (Moynihan and Clarke, 2014a; Sychantha *et al.*, 2017).

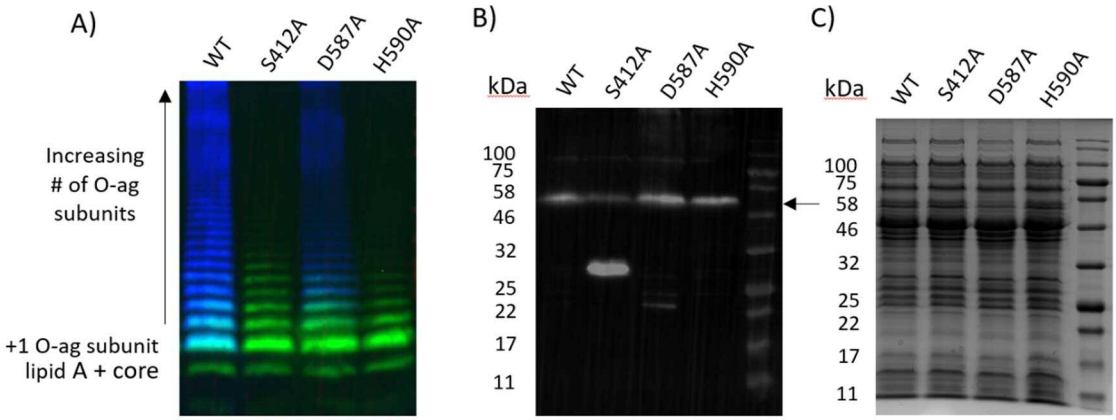


Figure 4.11| Validation of *in situ* function assay for OafA using SGNH domain catalytic triad mutants. A) LPS western blot with crude LPS extracts from *Salmonella* ser. Typhimurium basal O-antigen strain expressing OafA wild type (WT) and point mutant variants (S421A, D587A, H590A). O:5 antibody binding (blue) shows abequose acetylation and *Salmonella* LPS core antibody binding (green) acts as a loading control. B) Anti-His western blot for expression of His tagged OafA from insoluble protein fraction of same samples as (A). Arrow indicates full length OafA protein. C) Corresponding Coomassie stained protein gel with same samples as (B) to show uniform loading. A,B,C = representative of 3 technical replicates of LPS samples on three separate gels from 2 biological repeats.

These observations suggested that the optimised assay was suitable to identify residues within OafA which are critical for function or those that significantly impact the activity of the OafA protein. Therefore, an *in silico* inspired mutagenesis approach was used to identify further residues fundamental to the mechanism of OafA which may be broadly applicable to a wide range of AT3 domain containing O-acyltransferases.

4.5 Identification of functional residues in OafA using the optimised *in situ* acetyltransferase activity assay

Residues within OafA were selected for site directed mutagenesis based on the *in silico* analysis described in Chapter 3. **Table 4.2** describes the mutations made and the reason for their targeting.

Table 4.2| Site-directed mutagenesis of OafA. Residues for mutation were selected based on *in silico* analysis of experimentally characterised AT3 acetyltransferases (described in Chapter 3) and the reason for their selection are summarised below.

Mutant	Reason for mutation	Position within OafA
R14A H25A	R/K-X_{10}-H motif conserved across AT3 domain containing proteins	TMH1
S32A G33A G34A F35A I36A G37A V38A D39A V40A	XGG-F/Y-XGV-D/P/V-X motif conserved in AT3-SGNH fused acyltransferases	Periplasmic loop between TMH 1-2
S45A G46A	Conserved in SG in TMH2	TMH2
R69A R72A	RXXR motif identified as critical for function in other AT3 proteins	TMH3
S112A N113A Y122A	Conserved in periplasmic loop specific to AT3-SGNH fused proteins	TMH3-4 Periplasmic loop
G202A	Conserved trans membrane glycine, may be important for transmembrane domain structure	TMH6
E325A (Linker)	Conserved in AT3-SGNH fused proteins, negatively charged residue on the cytoplasmic side	TMH10-11 Cytoplasmic loop
C383S,C397S (Linker) C439S,C453S C567S,C572S	Conserved putative disulphide bonding pairs	SGNH
S412A D587A H590A	SGNH domain catalytic triad residues	

The results of site directed mutagenesis and *in situ* function analysis of OafA are summarised in **Table 4.3** and **Figure 4.12**. Representative LPS immunoblot and protein

Western blot results are presented in **Figure 4.13** with the OafA catalytic triad mutant results (S421A, D587A and H590A) presented in **Figure 4.11**. Critical residues were identified as those where mutation resulted in less than 1% O:5 antibody binding signal in LPS immunoblots compared to LPS from the strain expressing wild type OafA protein. Although some variation in O-antigen acetylation compared to wild type at the upper limits was observed (70-80%), the high level of variation meant that the cut off for significant inhibition of function was defined as point mutations resulting in less than 50% O:5 antibody signal intensity compared to wild type.

The protein degradation band seen for the OafA_S412A mutant by Western blot (**Figure 4.11**), suspected to be the cleaved SGNH domain, was present at varying levels in almost all OafA variant strains tested (**Figure 4.13**). Significantly, the OafA_G202A variant displayed a comparatively low signal for full length OafA expression and a significant degradation band at ~28 kDa (**Figure 4.13**). This mutant did not show significant reduction in OafA protein function, supporting the conclusion that as long as full length OafA is detectable by Western blot, there is sufficient protein present in cells to complement O-antigen acetylation.

Other than the catalytic triad residues of the SGNH domain characterised in section 4.4 (**Figure 4.11**), this analysis identified 6 residues critical for OafA function and 3 mutations that caused a significant reduction, but not complete knockout, of O-antigen abequose acetylation. The critical residues include: arginine and histidine that mark the R/K-X_{10}-H motif of TMH 1 (R14, H25), aspartic acid (D39) of the GG-F/Y-XGV-D/P/V motif which is specifically conserved in AT3-SGNH fused proteins, the TMH3 arginine residues (R69, R72), which form the RXXR motif that was previously identified as essential in OafB and the AT3-only O-antigen acetyltransferase of *S. flexneri* (Thanweer *et al.*, 2008), and the serine residue (S112) that was found to be specifically conserved in the TMH3-4 periplasmic loop in AT3-SGNH fused proteins (Section 3.3.1). The residues important but not critical for OafA function include: phenylalanine (F35) of the GG-F/Y-XGV-D/P/V motif, which is specifically conserved in AT3-SGNH fused proteins, glutamate in the cytoplasmic loop between TMH 10-11 that was specifically conserved in AT3-SGNH fused proteins and the most C-terminal putative disulphide bonding pair of cysteines (C567,572) in the SGNH domain (Section 3.3.1). The hypotheses formed about the role

of these OafA functional residues in the mechanism of O-antigen acetylation will be discussed in section 4.7.3.

Table 4.3| Summary of *in situ* mutagenesis analysis of STM OafA. Dark Grey= Point mutants with <1% wild type O:5 signal intensity, Light Grey = Point mutants with <50% wild type O:5 signal intensity, * = No OafA protein expression detected. Values represent the average of 2 biological repeats with 3 technical replicates.

Mutant	O:5 antibody binding signal compared to wild type LPS %(± SEM)	Position
R14A	0.07 ± 0.04	TMH1
H25A	0.33 ± 0.18	
S32A	105.25 ± 30.89	Periplasmic loop & TMH2
G33A	119.17 ± 18.72	
G34A	1.36 ± 0.88*	
F35A	19.24 ± 2.70	
I36A	101.47 ± 22.72	
G37A	118.13 ± 22.11	
V38A	86.38 ± 12.73	
D39A	0.31 ± 0.07	
V40A	121.28 ± 23.82	
S45A	98.18 ± 24.30	TMH2
G46A	99.59 ± 22.01	
R69A	0.10 ± 0.04	TMH3
R72A	0.07 ± 0.02	
S112A	0.24 ± 0.09	TMH3-4 Periplasmic loop
N113A	93.79 ± 14.92	
Y122A	85.76 ± 7.58	
G202A	74.14 ± 10.70	TMH6
E325A (Linker)	4.84 ± 1.13	TMH10-11 Cytoplasmic loop
C383S,C397S (Linker)	107.40 ± 26.80	SGNH domain
C439S,C453S	185.06 ± 54.63	
C567S,C572S	49.98 ± 4.33	
S412A	0.36 ± 0.26	
D587A	10.13 ± 1.70	
H590A	0.87 ± 0.62	

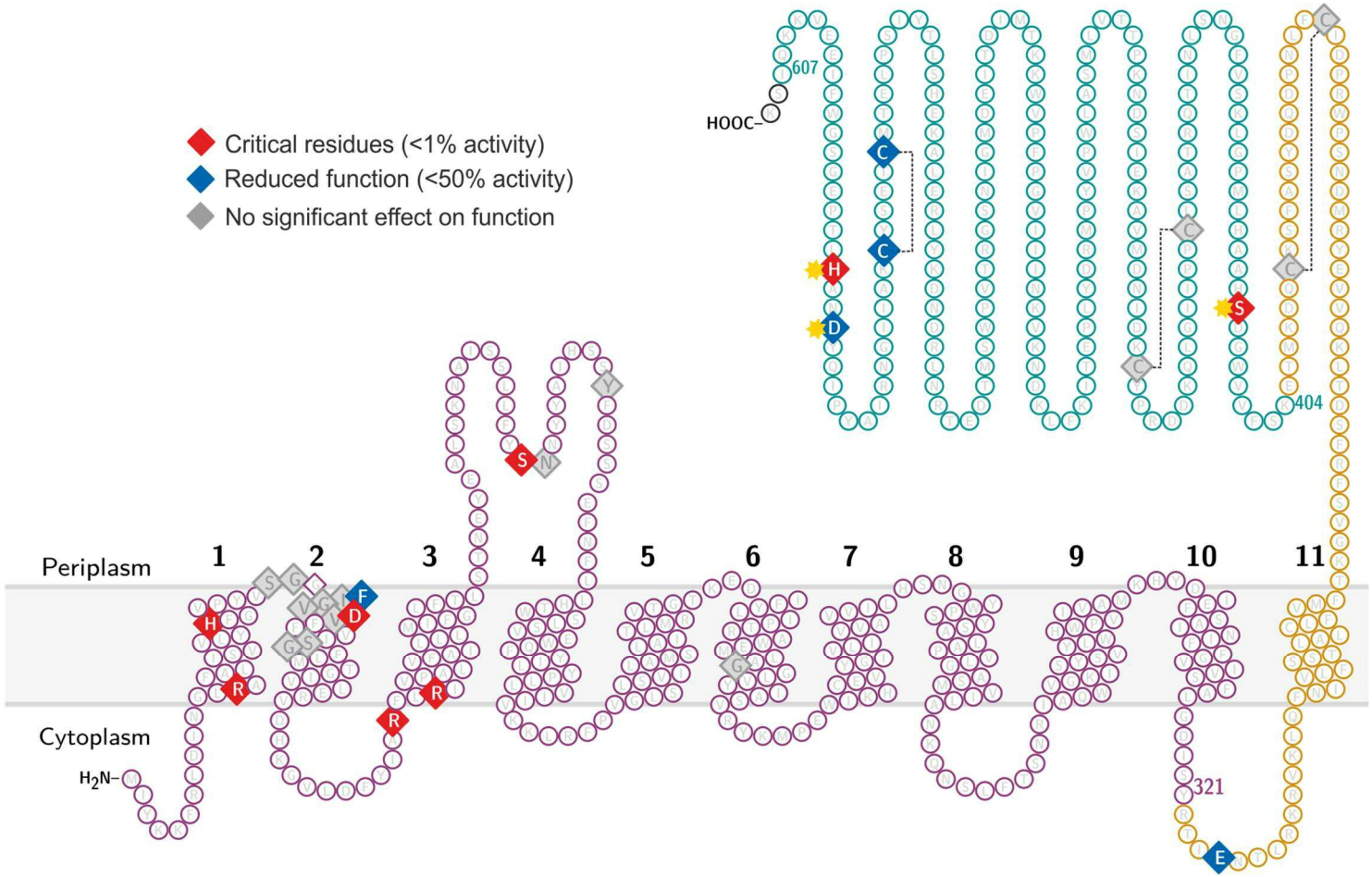


Figure 4.12| Summary of *in situ* mutagenesis analysis of STM OafA. Diamonds indicate mutated residues, cysteines were mutated to serine and all other residues were mutated to alanine. Dashed lines highlight putative disulphide bonding pairs and stars highlight the catalytic triad of the SGNH. Results relate to % O-antigen acetylation compared to wild type, mutants that caused loss of protein expression are diamond shaped but not coloured (G34A).

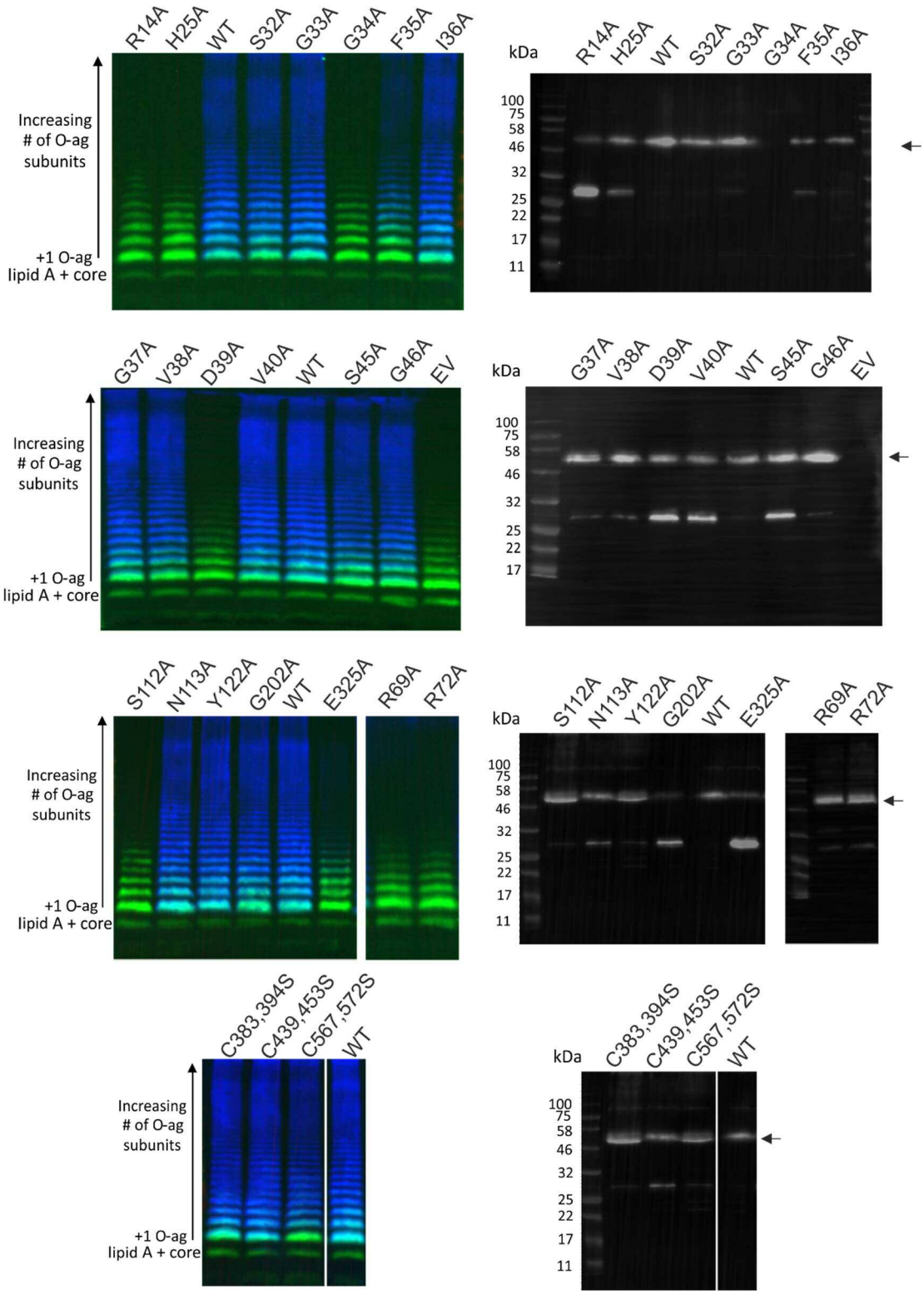


Figure 4.13| Functional analysis of OafA point mutants *in situ*. Left panel shows LPS western blot with crude LPS extracts from *Salmonella* ser. Typhimurium basal O-antigen strain expressing OafA point mutant variants. O:5 antibody binding (Blue) shows abequose acetylation and *Salmonella* LPS core antibody binding (Green) acts as a loading control. Right panel shows corresponding anti-His western blot for expression of His tagged OafA. Arrow indicates full length OafA protein.

4.5.1 Investigation of the role of the R/K-X_{10}-H motif

Both the arginine and histidine residues of the R/K-X_{10}-H motif of TMH1, identified as highly conserved across all AT3 containing proteins (Chapter 3), were critical for function in OafA (**Table 4.3**). Due to the integral role of this motif in O-antigen acetylation, its potential mechanistic role was investigated. Although not yet experimentally confirmed, Acyl-CoA derivatives have been suggested as the cytoplasmic acyl donor for AT3 domain-containing carbohydrate acyltransferase proteins (Bera *et al.*, 2005; Thanweer *et al.*, 2008). Therefore, the potential for these arginine and histidine residues to play a role in acyl-CoA interaction was investigated.

The average rise per residue in a transmembrane helix is 1.5 Å (Hildebrand *et al.*, 2004), this would position the arginine and histidine residues at a distance of ~18 Å. Helical wheel plot also positions these residues on the same side of the TMH (**Figure 4.14 A**). Arginine residues have previously been implicated in binding the 3' phosphate of acetyl-CoA and histidine residues in coordinating the thioester bond (Wu and Hersh, 1995; Jogl *et al.*, 2004). Therefore, the distance between the 3' phosphate and the thioester bond of CoA was measured in crystal structures of proteins in complex with CoA as a ligand (2.4.3). Across 4 different proteins in the PDB database (1H16, 5TVJ, 5G17, 2WL4) and the ideal coordinates for a CoA molecule from the PDB ligand summary page (RCSB, 1999), the distance between these two positions varied from 15.3 to 22.5 Å (**Figure 4.14 B-F**). Therefore, 18 Å is within the conformational flexibility of the distance between the 3' phosphate and the thioester bond of CoA. This highlights a potential acyl-CoA interaction site within the AT3 domain common to all AT3 proteins.

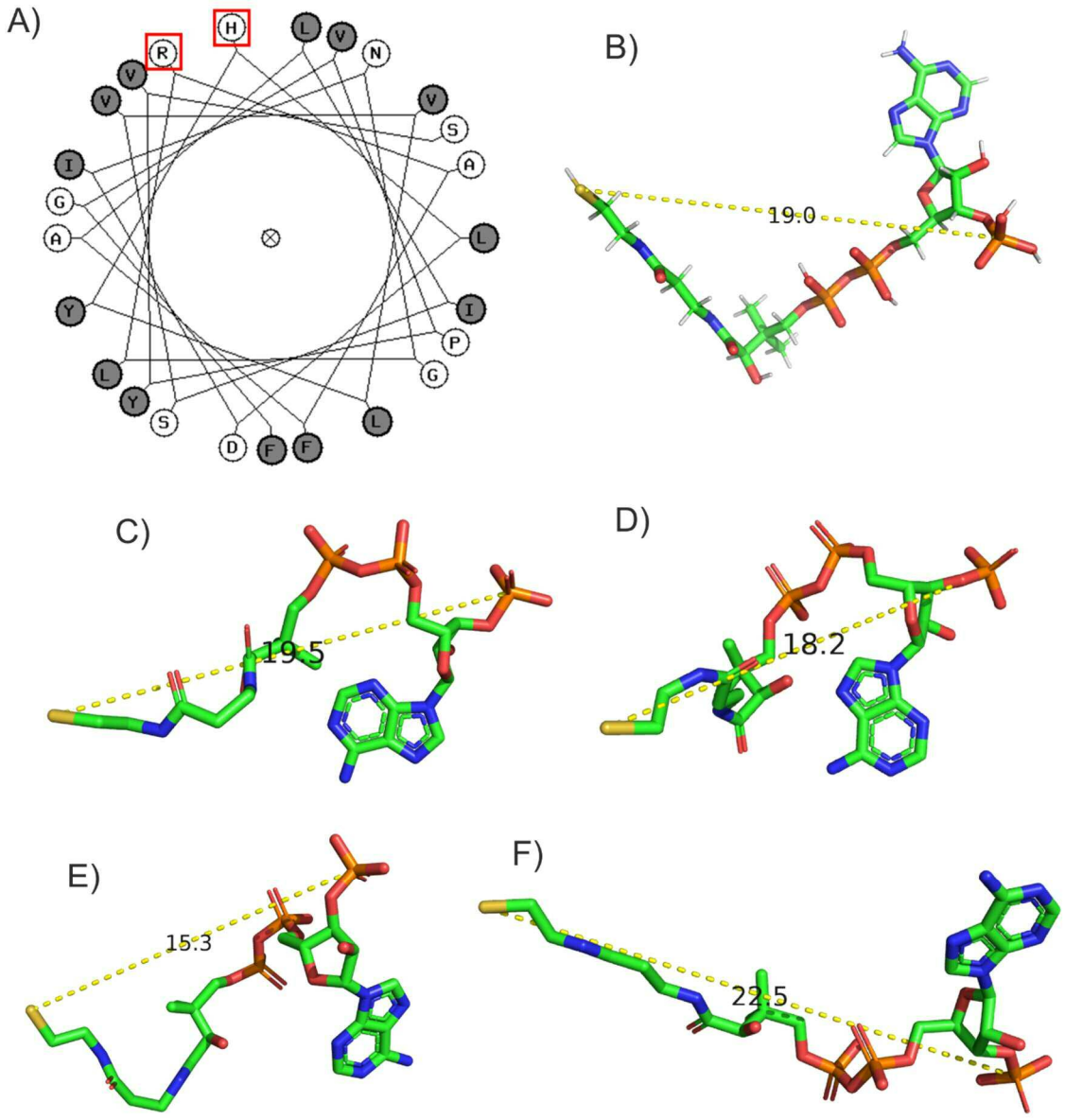


Figure 4.14| Does the R/K-X_{10}-H motif coordinate Coenzyme A. Helical wheel plot of OafA TMH1. Arginine and Histidine residues of the R/K-X_{10}-H motif are highlighted by red boxes. B-F) Distance, in Å, between 3' phosphate and thioester bond of CoA molecules using the PyMOL molecular graphics system measurement wizard (version 2.3.0, Schrödinger, LLC). B) Ideal coordinates for a CoA molecule, C-F) CoA in complex with crystallised acyltransferase proteins from PDB entries - (C) 1H16, (D) 5TVJ, (E) 5GI7, (F) 2W14.

4.6 Confirmation of the *in situ* substrate specificity of OafA and OafB

OafA and OafB have been characterised in the literature to perform acetylation of different monosaccharides of the STM LPS O-antigen. This suggests that these two proteins have adapted to be specific in the acetylation modification that they perform.

In order to confirm this *in situ* specificity, OafB, which is shown to acetylate rhamnose, was expressed in the basal STM O-antigen strain to determine whether any off-target acetylation of the abequose residue could be detected by LPS immunoblot with O:5 serotyping antibodies. Western blot for His-tagged OafB protein confirmed that full length OafB was expressed in the experimental conditions used (**Figure 4.15**). The expected molecular weight of His-tagged OafB is 73.6 kDa however, like OafA, this protein runs at the lower molecular weight of ~50kDa. The LPS immunoblot did not show detectable O:5 antibody binding to the basal O-antigen strain complemented with OafB (**Figure 4.15**), suggesting that, *in situ,* this protein is specific for rhamnose acetylation and does not show evidence of abequose acetylation. OafB expression from this vector has previously been confirmed to protect STM LT2 from lysis by the BTP1 bacteriophage by phage plaque assay therefore the OafB protein is functional in this experimental system (Reyme Herman, master's thesis, data not shown). These observations support the hypothesis that OafA and OafB are specifically adapted to acetylate different O-antigen residues on STM LPS.

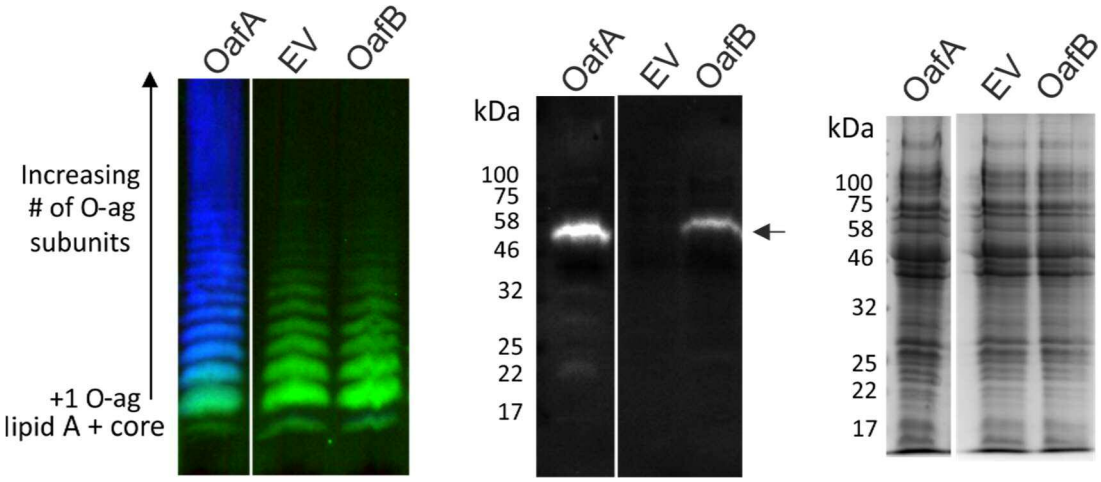


Figure 4.15| Confirmation of the in situ acceptor substrate specificity of OafA and OafB. A) LPS western blot with crude LPS extracts from *Salmonella* ser. Typhimurium basal O-antigen strain expressing OafA, OafB, or with an empty pBADcLIC vector (EV). O:5 antibody binding (blue) shows abequose acetylation and *Salmonella* LPS core antibody binding (green) acts as a loading control. B) Anti-His western blot for expression of His tagged OafA and OafB from insoluble protein fraction of same samples as (A). Arrow indicates full length OafA or OafB protein. C) Corresponding Coomassie stained protein gel with the same samples as (B) to show uniform loading. This experiment was carried out by Rebekah Eastwood with close supervision from the thesis author.

4.7 Chapter summary and future work

4.7.1 The two colour LPS immunoblot assay for OafA function is an optimised approach to functional analysis of O-antigen acetyltransferases

Previous investigations into the mechanism of O-acylation by AT3 domain-containing proteins in many other biological systems may have been hindered by lack of appropriate means of assessment of protein function. Although there are multiple published examples of bacterial AT3 domain-containing carbohydrate acyltransferases whose activity has been demonstrated experimentally (**Figure 3.3**), few studies move beyond knockout and complementation in characterisation of these proteins. Additionally, many of the techniques used to confirm loss or gain of carbohydrate acetylation involve purification of the carbohydrate acceptor and analysis by biophysical methods such as NMR and MS (Luck *et al.*, 2001; Anonsen *et al.*, 2017; Kintz *et al.*, 2017). These are costly and time-consuming approaches which require specialised expertise and do not lend themselves well to high throughput analysis of a range of point mutant variants of these proteins.

In the studies that further characterise essential residues of the AT3 domain containing acyltransferase proteins there is no quantification of function, rather a binary readout of active/inactive (Thanweer *et al.*, 2008; Thanweer and Verma, 2012). LPS immunoblotting as a functional readout for acetyltransferase activity has previously been used to identify critical residues in O-antigen acetyltransferase Oac from *Shigella flexneri* (Thanweer *et al.*, 2008; Thanweer and Verma, 2012). However, due to the same limitations as those identified here for assessment of OafA, such as lack of an internal immunoblot loading control, there is no quantification of the effect of point mutation. Optimisation of OafA functional analysis by immunoblotting techniques has provided a means of semi-quantitative assessment of protein function using relatively simple sample preparation and analysis procedures. This assay allows *in situ* analysis of the functional residues of OafA so that predictions can be made about their role in the mechanism of action of AT3 domain-containing acyltransferases.

This optimised approach to functional analysis of OafA is an improvement on previous analysis. Firstly, experimental conditions and sample preparation have been optimised for consistency and comparability between samples. Secondly, variables such as pH of the media have been considered to prevent variations in acetylation levels which are not caused by differences in OafA activity. Most importantly, a consistent approach to image quantitation and sample normalisation allows comparability of results between both samples on the same gel and between different experiments.

4.7.2 Limitations of the two colour LPS immunoblot assay for quantification of OafA function

Using the optimised dual antibody LPS immunoblot, it is possible to semi-quantitatively assess the acetylation level of the LPS O-antigen to give an indication of the acetyltransferase activity of OafA. This assay however, cannot give a direct readout of OafA catalytic activity due to a range of different factors. Firstly, although every effort has been made to standardise the assay procedure to allow comparative analysis between samples, there is still a degree of variation in results when the same LPS sample is run on different gels (Section 4.3.3, 4.4). These variations between technical repeats could arise at the point of sample loading into the gel, transfer of the LPS samples to PVDF membrane or at the point of image quantitation as lane boundaries are set manually.

Secondly, it is important to acknowledge that although we are confident that O:5 serotyping antibodies are specific to STM O-antigens with acetylated abequose residues, we do not know exactly how or where the antibody binds to the O-antigen. The O:5 serotyping antibodies used from SSI diagnostica are polyclonal which means that they will not all recognise exactly the same O-antigen epitope in the same way (Lipman *et al.*, 2005). A crystal structure of the Se155-4 IgG1 λ1 antibody interacting with a single STM O-antigen repeat (PDB: 1MFC) (Stanfield *et al.*, 1990; Haji-Ghassemi *et al.*, 2015) demonstrates that a single antibody Fab fragment is significantly wider than a single O-antigen repeat (**Figure 4.16**). Additionally, the O:5 serotyping antibodies will be composed of two Fab fragments linked to a single Fc region which imparts the antibody with its biological effector functions. This will further increase the region of the LPS O-antigen that is occluded by a single IgG antibody. The size of an IgG2 antibody is ~10-15

nm wide by ~10 nm long (PDB: 1IGT)(Reth, 2013). Thus, it is unlikely that the stoichiometry of antibody binding to each O-unit in the LPS O-antigen chain will be 1:1.

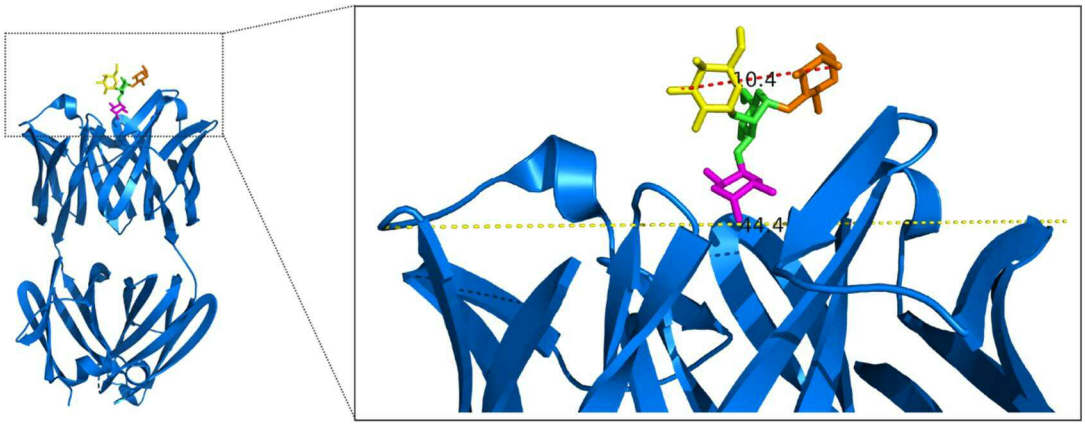


Figure 4.16| Size comparison of O-antigen and antibodies. Crystal structure of the Se155-4 antibody Fab fragment (blue) with *Salmonella* group B O-polysaccharide. Abequose = magenta, mannose = green, rhamnose = orange, galactose = yellow. Zoom in shows the width of a single Fab fragment of the IgG1 is ~44.4 Å (yellow dashed line), and an O-antigen repeat unit is ~10.4 Å (red dashed line). PDB file 1MFC edited using PyMOL software.

Previously, Surface Plasmon Resonance has been employed with bacterial polysaccharides to determine the stoichiometry of antibody binding (Brogioni and Berti, 2014), this requires the polysaccharide to be covalently attached to a surface or bound with a protein which is covalently attached to the surface. It may be feasible to use this approach to determine the stoichiometry of binding for crude LPS samples with polyclonal antibodies if an anti-Lipid A antibody was used to capture LPS on the surface. However, the fact that these antibodies are polyclonal may introduce too much variation in O-antigen binding for reliable quantification.

To determine the accuracy of quantitation of O-antigen acetylation by LPS immunoblot it would be beneficial to confirm these results with biophysical analysis of LPS from a selection of the STM OafA point mutant test strains. High-Resolution Magic-Angle Spinning NMR has been used previously to characterise the surface glycan structure of in-tact bacterial cells (Jachymek *et al.*, 1999) and has been used to analyse the O-antigen composition and acetylation state of STM O-antigens (Zandomeneghi *et al.*, 2012). This idea was pursued in collaboration with Nottingham University but unfortunately the O-antigen composition of different OafA variant strains has not yet been resolved.

The conditions identified for assessment of OafA function using this assay were selected to make it as quantitative as possible with the available resources. This assay was appropriate to identify residues within OafA which are critical for function or those that significantly impact the activity of the OafA protein.

4.7.3 Characterisation of functional residues of OafA informs their role in the mechanism of O-antigen acetylation

Analysis of conserved residues within OafA using the optimised *in situ* function assay allowed hypotheses to be formed about the mechanistic role of functional residues. These are described below.

The only mutation which resulted in negligible protein expression was OafA_G34A. A glycine residue in this position appears to be critical for correct protein production or insertion into the membrane as no OafA protein was detectable by anti-His western blot in the insoluble protein fraction, resulting in loss of O-antigen acetylation (**Figure 4.13**). This glycine residue is predicted to be located on the short periplasmic loop between TMH1 and 2 (**Figure 4.12**). Indeed, glycine is often found in loop regions due to its flexibility and the lack of steric hindrance caused by absence of a side chain (Javadpour *et al.*, 1999). Therefore, it is possible that mutation of G34 of OafA prevents correct insertion of the transmembrane helices, resulting in degradation of the incorrectly folded protein.

$OafA_{R14}$ and $OafA_{H25}$ within the R/K-X_{10}-H motif were essential for OafA activity (**Figure 4.12**). These residues were proposed to provide a conserved mechanism for acyl-donor interaction due to their strong conservation across all AT3 domain-containing proteins (Section 3.3.1, 4.5.1). Although cytoplasmic acetyl-CoA has not been confirmed as the donor for O-antigen acetylation, it occupies a central role in bacterial metabolism and is a prominent source of acetate in bacterial cells (Takamura Y, 1988; Krivoruchko *et al.*, 2015). The predicted positioning of these residues at opposite ends, but on the same surface, of the TMH and their predicted distance being within the conformational flexibility of the 3' phosphate and thioester bond of Coenzyme A (**Figure 4.14**) proposes the argument that acyl-CoA is the donor substrate in carbohydrate O-acylation. In support of this prediction, a role for a conserved intermembrane histidine residue, in coordinating the thioester bond of the acyl donor, has been suggested for membrane

bound O-acyltransferases containing an MBOAT (IPR004299) rather than AT3 domain (Ma *et al.*, 2018).

Additionally, a natural histidine to tyrosine point mutation in the equivalent location of the R/K-X_{10}-H motif was discovered to decrease the function of the *S. pneumonia* capsule acetylation protein WcjE in clinical isolates (Calix *et al.*, 2011). This supports the critical role of the R/K-X_{10}-H motif across diverse AT3 domain containing O-acyltransferases, in coordinating a cytoplasmic derived acetyl-CoA molecule within the membrane bound AT3 domain for transfer of the acetyl group to the SGNH domain, consistent with the working model (**Figure 3.1 B**).

AT3 domain-containing proteins are implicated in transferring a wide range of acyl groups such as succinate, isovalerate, and propionate (Arisawa *et al.*, 1993; Roset *et al.*, 2006; Cong and Piepersberg, 2007); these can all be carried by Coenzyme-A. The proposed mechanism of acetyl donor interaction would provide a potential conserved mechanism for transfer of any of these acyl substituents, further supporting the idea that the TMH1 arginine and histidine are fundamentally important for the mechanism of all AT3 domain-containing acyltransferases.

Only the phenylalanine and aspartic acid point mutations of the GG-F/Y-XGV-D/P/V motif, caused detectable reduction in OafA function *in situ*. OafA_F35A caused significant reduction and OafA_D39A caused loss of OafA function (**Table 4.3**, **Figure 4.12**). These residues are predicted to be found towards the periplasmic side of the trans membrane domain. The location of these functional residues within the transmembrane domain of OafA could suggest that they play a role in interaction with the periplasmic SGNH domain or carbohydrate O-antigen acceptor. Indeed, polar amino acids such as aspartic acid have long since been acknowledged to form hydrogen bonds with carbohydrate hydroxyl groups (Quiocho, 1986; Taroni *et al.*, 2000) and aromatic amino acids such as phenylalanine have been implicated in interacting with C-H bonds of carbohydrates (Hudson *et al.*, 2015). This proposed role, for AT3 domain residues to be involved in carbohydrate interaction, would suggest that the AT3 and SGNH domains are in close proximity during O-antigen modification.

OafA and OafB contain a linking region between the SGNH and AT3 domains with ~ 50 residues on the periplasmic side (**Figure 4.12**). In the absence of structural information about this domain or how it interacts with the AT3 or SGNH domain, it is not possible to

confirm whether the linking region is structured and holds the SGNH domain above the AT3 for direct transport and transfer of the acetyl group to the O-antigen, or whether the linking region is flexible and the SGNH first collects the acetyl group for the AT3 domain then moves to interact with the O-antigen to catalyse the modification. If the latter is true, these residues would be unlikely to interact with the carbohydrate acceptor and more likely to be important for interaction with the SGNH domain for the transfer of acetate.

In addition to the phenylalanine and aspartic acid of the GG-F/Y-XGV-D/P/V motif, OafA_S112 which is specifically conserved in the TMH3-4 periplasmic loop in SGNH fused proteins, was also critical for OafA function. Providing another potential periplasmic interacting partner for the SGNH domain interaction. Interestingly, *in silico* analysis did not identify significantly conserved motifs within the SGNH domain that indicated potential interaction sites for these conserved phenylalanine, aspartic acid and serine residues. Therefore, these interactions may be determined by the tertiary structure of the SGNH domain or by the carbohydrate substrate.

The SG (45-46) motif in TMH2 of OafA has been previously investigated for un-fused O-antigen acetyltransferases and suggested to be important for function (Luck *et al.*, 2001; Thanweer *et al.*, 2008; Thanweer and Verma, 2012). However, analysis of OafA_S45A and OafA_G46A did not identify these residues as critical for acetyltransferase activity (**Figure 4.12**). This highlights a clear difference between AT3-only and AT3-SGNH fused proteins and may suggest that they function with slightly different mechanistic residues.

The RXXR motif, previously shown to be critical for OafB function (Kintz *et al.*, 2015), was confirmed to be essential for OafA. These conserved arginine's could play a role in acetyl-CoA interaction, parallel to the suggested role for arginine in the R/K-X_{10}-H motif. However, these positive residues conserved on the cytoplasmic side between TMH 2-3 could also be important to maintain the membrane topology of the AT3 domain and act as an anchor point on the cytoplasmic side. A structural role for conserved cytoplasmic arginine's has been observed previously, where mutation of cytoplasmic arginine's resulted in aberrant topology of an integral membrane glucose transporter (Sato and Mueckler, 1999). A structural role of these conserved arginine's also fits with the fact that arginine residues interact with phospholipid head groups in the lipid bilayer (Li *et al.*, 2013; Robison *et al.*, 2016). It is also in keeping with the 'positive inside' rule where

positively charged amino acids are enriched on the cytoplasmic side (von Heijne, 1986; Baker *et al.*, 2017).

Of the putative disulphide bonding pairs identified within OafA by *in silico* analysis, only mutation of the most C-terminal disulphide bonding pair (C567, C572) caused a reduction in abequose acetylation levels, suggesting reduced acetyltransferase activity (**Figure 4.12**). Prevention of disulphide bond formation at this site could have reduced the structural stability of the catalytic domain or binding pocket of the SGNH domain, which may have resulted in reduced substrate interaction and turnover. As these cysteine pairs within SGNH domains of AT3-SGNH proteins appear or disappear as pairs and are not strictly conserved in all aligned proteins they are most likely to play a structural rather than catalytic role. It is possible that if the mutation of all cysteine pairs is combined, the structural integrity of the SGNH domain will be reduced. Therefore, a greater impact on O-antigen acetylation may be seen. This would support the hypothesis of their structural role.

4.7.4 Future Perspectives

The current method used to confirm protein expression in the *in situ* OafA function experiments involves Western blotting of insoluble protein fractions from each test strain to detect expression of His-tagged OafA. However, this does not prove conclusively that the protein is delivered to the inner membrane as it could be sequestered in inclusion bodies. For those mutants that retain some function but show reduced acetylation, a positive Western blot signal strongly suggests the protein has been properly inserted into the inner membrane, but for those with no function it would be beneficial to confirm their membrane localisation. Rather than total insoluble proteins being run, the membrane fraction could be isolated and run alone to give further support that the point mutation has caused loss of protein activity rather than prevented its correct expression and delivery to the inner membrane.

As this assay is able to assess the *in situ* function of OafA, as well as testing the role of specific functional residues, it could also be used to determine how different environments affect OafA function. For example, through the infection cycle, *Salmonella* cells are subject to a range of diverse environments, having to survive outside of an animal host during transmission and within different cellular environments once a host

is infected (Section 1.2.1). It could be possible to replicate some of these environmental stresses while culturing *Salmonella* in the lab and test how they affect the activity of OafA. One possible example is oxidative stress. The SGNH domain of OafA resides in the periplasm, an environment which can be subject to fluctuations in oxidative state (Van Der Heijden *et al.*, 2016). These fluctuations could determine activity of OafA so that it may be active during certain stages of the infection cycle and inactive during others. A feature seen in other periplasmic and regulatory proteins (Cremers and Jakob, 2013; Hillion and Antelmann, 2015).

This analysis was able to confirm the substrate specificity of OafA and OafB. It therefore provides an experimental system to investigate the domains or residues involved in substrate specificity. Following further characterisation of these proteins, to inform residues involved in acceptor substrate specificity, this experimental system could be used to test whether OafB can be engineered to catalyse abequose acetylation *in situ*.

This analysis was able to inform hypotheses on the role of specific residues within the membrane bound AT3 domain but was unable to identify residues within the SGNH domain which might be responsible for acceptor substrate interaction. In order to shed more light on the SGNH domain, structural characterisation may be more informative and could give insight into the role of conserved cysteine residues as well as whether the linking region is structured and thus, how the AT3 and SGNH domains are likely interact. Therefore purification, crystallisation and *in vitro* characterisation of the SGNH domain was investigated. This investigation is described in Chapter 5.

Chapter 5 *In vitro* analysis of the SGNH domain of O-antigen acetyltransferases

5.1 Introduction

Sequence-informed site-directed mutagenesis identified functional residues that play an integral role in the acetyltransferase activity of OafA (Chapter 4). Functional residues specifically conserved in AT3-SGNH fused proteins, that are located towards the periplasmic side of the inner-membrane, could suggest potential interaction sites for the acceptor substrate or the SGNH domain on the periplasmic side. We were yet to establish whether OafA and OafB act as predominantly static proteins with a structured linking region, where the SGNH domain would sit above the AT3 domain and both domains would interact with the carbohydrate acceptor simultaneously; or whether these proteins have a more dynamic two domain arrangement with a flexible linking region, where the SGNH domain first accepts the acetyl group from the AT3 domain then moves to interact with the O-antigen for the final acetyltransferase step.

Site directed mutagenesis also demonstrated that the SGNH domain is essential for O-antigen acetylation, due to the fact that OafA with a catalytically inactive SGNH domain is rendered non-functional (**Figure 4.13**). Although highly likely, we were yet to confirm the hypothesis that this domain is responsible for the final step of the transferase reaction. Furthermore, although the catalytic triad residues of the SGNH domain had been confirmed and putative disulphide bonding cysteine residues had been identified within the SGNH domain (Section 3.3.2, 4.5), the arrangement of the active site pocket and oxyanion hole residues was yet to be established. Therefore, the aim of this chapter was to express and purify the SGNH domains of OafA and OafB to perform *in vitro* biochemical and biophysical analysis to investigate the exact role of the SGNH domain in two domain AT3-SGNH fused O-antigen acetyltransferases.

5.2 Cloning and expression trials of OafA and OafB

The most direct approach to shed further light on the potential SGNH to AT3 domain interaction sites and the structural arrangement of the SGNH domain active site pocket would be to obtain crystal structures for full length OafA and OafB. Previous endeavours

to express and purify full length OafA and OafB proteins had been met with some success with a yield of 0.3 mg and 1.2 mg per litre of culture for OafA and OafB respectively (Reyme Herman, unpublished data). This did not yield sufficient protein for successful crystallisation trials and biophysical analysis such as ligand screening by saturation transfer difference (STD) NMR of OafB was unable to determine whether these proteins interacted with acetyl-CoA (Andrew Brentnall, University of York), due to complications caused by detergent molecule interactions.

Membrane proteins are notoriously difficult to crystallise or analyse biophysically due to the stretches of hydrophobic residues, their propensity to aggregate and the need to extract and stabilise them in a native conformation in a pseudo phospholipid bilayer (Carpenter *et al.*, 2008; Liang and Tamm, 2016; Birch *et al.*, 2018). Due to these foreseeable issues, the periplasmic portion of OafA and OafB, including the linking domain and SGNH domain, were targeted for *in vitro* functional analysis.

Initial bioinformatics analysis of OafA and OafB showed that both proteins contain a significant C-terminal periplasmic domain after an 11 transmembrane spanning membrane domain (**Figure 3.1**). Both proteins contain an N-terminal AT3 domain and C-terminal SGNH domain, connected by a single transmembrane spanning linking region (described further in Chapter 3).

51 and 49 residues of the linking region extend from the membrane before the start of the SGNH domain in OafA and OafB respectively. These extending residues were thought to provide a flexible region to allow interaction of the membrane bound and periplasmic domains and potential movement of the SGNH domain towards its carbohydrate acceptor for acetyl group transfer. Unstructured flexible regions within proteins can inhibit their ability to form crystals and therefore hinder their structural characterisation (Derewenda, 2004). However, if these linking region residues are structured and play a role in the mechanism of action, their removal could cause instability of the expressed protein and therefore difficulties in purification and crystallisation.

In order to define the boundaries for C-terminal OafA and OafB expression constructs the alignment of AT3-SGNH proteins with homology to OafA (Section 3.3.2) was used to identify conserved residues among the linking regions of AT3-SGNH fused proteins (Appendix IV). Analysis of this region of the alignment highlighted conserved glycine and arginine (OafA G356, R360) residues and two conserved cysteine residues (OafA C383,

C397) in the connecting sequence between the end of TMH11 and the beginning of the SGNH domain (**Figure 5.1**). Although *in situ* functional assays did not identify this putative disulphide bonding pair as essential for function in OafA (Section 4.5), they may be required to assist correct folding of the periplasmic domain in absence of the membrane domain.

In addition to the conserved linking region residues, it was also observed that TMH11 of the linking region is poorly conserved across AT3-SGNH fused proteins (Appendix IV). This TMH is proposed to act as a structural element to deliver the SGNH domain to the periplasmic side, rather than being involved in the catalytic activity of the transferase reaction (Chapter 3 introduction: 3.1). This observation supports the predicted role because other than containing hydrophobic residues with propensity to form an alpha helical TMH domain, strong selection of specific residues would not be required in a non-catalytic transmembrane helix with a solely structural role.

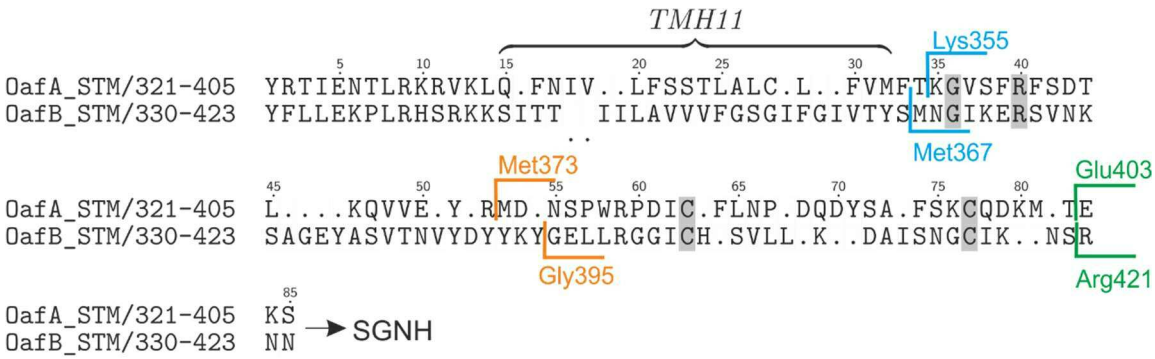


Figure 5.1| Analysis of the AT3-SGNH linking region sequence to inform OafA and OafB expression constructs. The linking region sequence was extracted from alignment (Appendix IV). Similar residues in >50% of sequences are boxed in grey and start sites for three OafA and OafB C-terminal expression constructs are indicated by blue, orange and green markers.

Three expression start sites were chosen for OafA and OafB, the first corresponding to a start just after TMH11 to include the whole of the periplasmic linking region, the second preserving the two conserved cysteine residues of the linking region, and the third beginning just upstream of the SGNH domain region defined by InterPro (**Figure 5.1**). The resulting proteins will be referred to by their start residue, for example OafA_Lys355.

The pETFPP_30 expression vector (Technology facility, University of York) was selected for initial expression trials of the C-terminal OafA and OafB constructs (Appendix V). Protein expression from this vector is inducible with IPTG. It adds an N-terminal PelB

leader sequence to target protein expression to the periplasm and also adds a 3C protease cleavable His tag to the C-terminus. It is important to target expression of the SGNH domain to an oxidising environment within the cell so that disulphide bonds can form (Berkmen, 2012). One of the putative disulphide bonding pairs in OafA causes reduced function when mutated to serine *in vivo* (C567-C572)(**Table 4.3**). Therefore, correct disulphide bond formation could be critical for folding and stability of the expressed proteins. The C-terminal cleavable His-tag was included to allow efficient purification by nickel affinity chromatography following successful expression of these proteins and also anti-His antibody detection for western blotting.

Table 5.1| *E. coli* BL21(DE3) C-terminal OafA and OafB expression strains. See **Table 2.1** for details of the BL21(DE3) expression strain.

Strain	**Expression Plasmid Info**	**Expressed protein size (kDa)***
BL21(DE3) pMV463	pETFPP_30 plasmid acquired from TF	-
BL21(DE3) pMV467	pETFPP_30 OafB_Gly395	32.2
BL21(DE3) pMV471	pETFPP_30 OafB_Met367	29.2
BL21(DE3) pMV468	pETFPP_30 OafB_Arg421	26.5
BL21(DE3) pMV475	pETFPP_30 OafB_Arg421^{H621A}	26.5
BL21(DE3) pMV472	pETFPP_30 OafA_Lys355	31.4
BL21(DE3) pMV473	pETFPP_30 OafA_Met373	29.2
BL21(DE3) pMV474	pETFPP_30 OafA_Glu 403	25.7
BL21(DE3) pMV548	pETFPP_30 STM LT2 SapA	59.4
BL21(DE3) pMV585	pET20b_Oant_2186	-

*Expected protein size relates to theoretical molecular weight calculated from the expressed protein sequence.

Initial expression trials for OafB_Gly395 indicated that induction of protein expression with 1 mM IPTG caused a reduction in cell growth when cultured at 37 °C, indicating a potentially deleterious effect of expressed protein on the viability of BL21(DE3) cells (T= 4hr Induced OD_{650} = 0.51, Uninduced OD_{650} = 3.63) (**Figure 5.2 A**). The induced cultures did however, reach a similar stationary phase OD_{650} to un-induced cells (OD_{650} ~3) (**Figure 5.2 A**). This suggested that the initial lag in cell growth may have been due to metabolic burden of protein expression, either the protein was toxic to cells and

required degradation for cell growth to continue or high levels of protein expression increased the metabolic load on protein production machinery of the cell, reducing the efficiency of cell growth. Inhibited cell growth was also seen with BL21(DE3) pMV463 containing the empty pETFPP_30 expression plasmid following IPTG treatment. This may again have been caused by a metabolic burden on the cells. Over expression of the PelB signal sequence could have overloaded the Sec protein secretion machinery so that extra-cytoplasmic proteins could not be secreted as efficiently for cell growth and those un-secreted proteins cause deleterious effects if they accumulate in the cytoplasm (Rosenberg, 1998; Mergulhão *et al.*, 2004; Horga *et al.*, 2018).

No visible band of expressed protein was seen in total protein extracts on Coomassie stained SDS-PAGE, however protein expression was visible after 4 hr induction by western blot (**Figure 5.2 B**). Multiple immunoreactive bands could be seen on the anti-His-Tag western blot, this could indicate multimers or aggregates of expressed protein. ß-mercaptoethanol was not included in the sample buffer for this gel. ß-mercaptoethanol is responsible for reducing disulphide bonds in proteins and OafB_Gly395 contains two pairs of cysteine residues which are predicted to form disulphide bonds (Section 3.3.2). Lower molecular weight bands may have been seen due to incomplete denaturation of the protein resulting in a smaller hydrodynamic radius or due to cleavage at the N-terminal end of the protein. Random interaction between different cysteine residues could result in protein multimers, giving rise to the higher molecular weight bands.

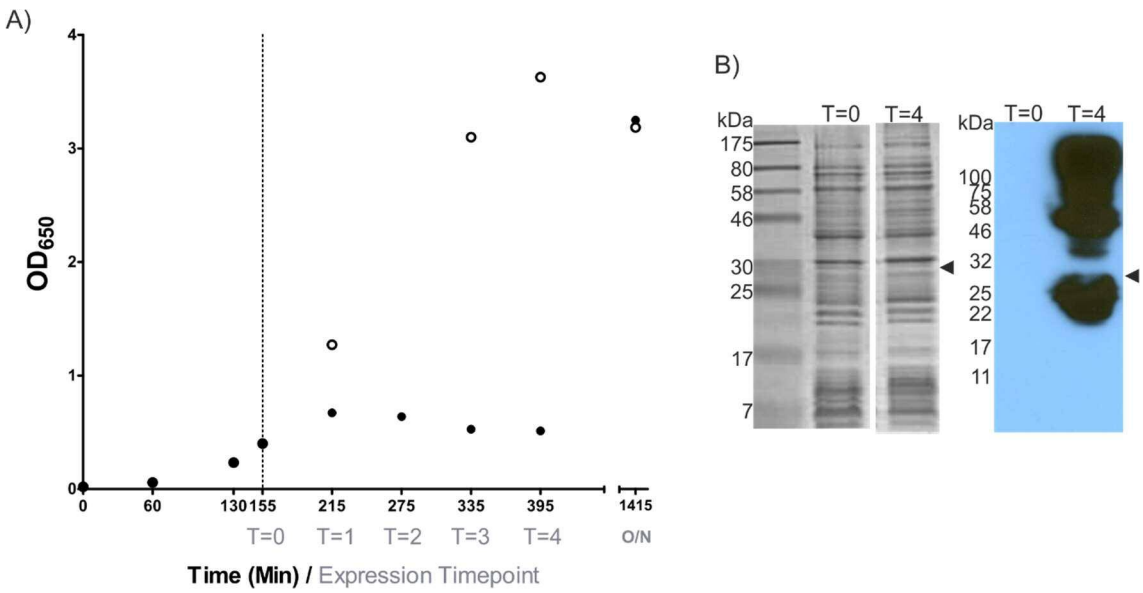


Figure 5.2| Effect of 1 mM IPTG induction on growth of BL21(DE3) pMV467 and expression of OafB_Gly395. A) Optical densities of (OD_{650}) two BL21(DE3) pMV467 cultures grown in LB to mid-log phase (OD_{650} 0.4-0.6) induced with 1 mM IPTG (closed circles) or left uninduced (open circles). B) Coomassie stained SDS-PAGE gel (left) and anti-His-Tag western blot (right) of total protein extracts from time point zero and 4 hours post 1 mM IPTG induction of samples as in (A). Expected molecular weight of expressed OafB_Gly395 is indicated by an arrow head (29.2 kDa). See **Table 5.1** for strain details. Overnight induction samples could not be analysed due to problems with sample preparation.

To reduce the deleterious effect of protein expression and optimise production of soluble OafB_Gly395 for subsequent purification, a further expression trial was conducted. Lower concentrations of IPTG (0.5 mM and 0.1 mM) were used for induction during early log phase as lower IPTG concentrations have shown to benefit protein expression conditions by reducing the negative effect on cell growth and productivity (Larentis *et al.*, 2014). Three different expression temperatures were also trialled 37 °C, 30°C and 25 °C. Reducing the temperature for protein expression is a well known technique for reduction of protein aggregation (Schein, 1989) and has proven to increased the soluble yields of aggregation prone proteins (Vasina and Baneyx, 1997). As confirmation of the expression protocol viability, a positive control of Bl21(DE3) pMV548 (**Table 5.1**) which expresses SapA, an STM LT2 periplasmic binding protein, already successfully expressed using the pETFPP_30 expression system (Bryony Ackroyd, University of York), was tested in parallel (**Figure 5.3**).

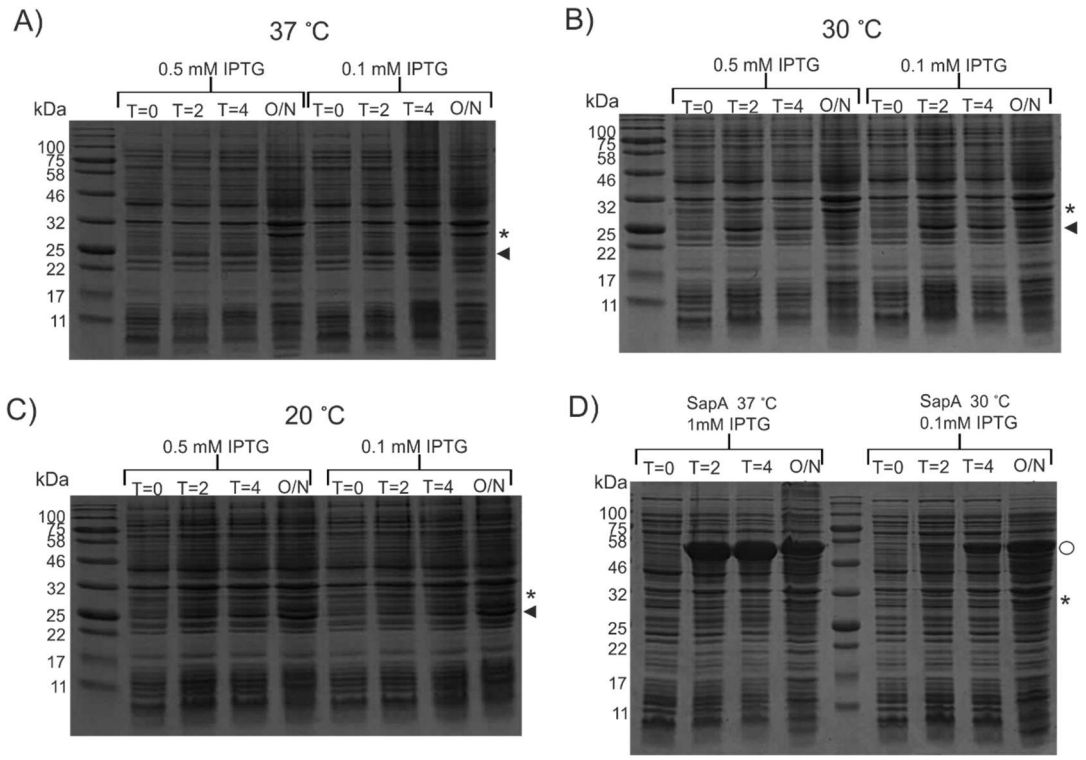


Figure 5.3| Effect of IPTG concentration and temperature on expression of OafB_Gly395 and SapA from E. coli BL21(DE3). A-C) Coomassie stained SDS-PAGE gel of total protein extracts from BL21(DE3) pMV467 (OafB_Gly395) after growth at 37 °C, 30 °C or 20 °C and induction with 0.5 mM or 0.1 mM IPTG in early log phase (OD_{650} 0.4-0.6). D) Coomassie stained SDS-PAGE gel of total protein extracts from BL21(DE3) pMV548 (SapA) after growth at 37 °C and induction with 1mM IPTG or 30 °C and induction with 0.1 mM IPTG in early log phase. See Table 5.1 for strain details. Arrow heads indicate potential OafB_Gly395 expression, asterisk indicates potential ß – lactamase expression and circle indicates SapA expression. ß-mercaptoethanol was not included in the sample buffer for these gels.

Positive over-expression of SapA confirmed that the protocol for induction, sample preparation and analysis would show target protein expression if it was occurring (**Figure 5.3**). A more intense ~30 kDa protein band in the overnight samples for all OafB_Gly395 expression conditions, which was also seen for SapA expressing BL21(DE3) cells, is likely to represent ß -lactamase which has a molecular weight of 31.5 kDa and is shown to migrate at a lower molecular weight when disulphide bonds are not reduced (Pollitt and Zalkin, 1983).

Potential OafB_Gly395 expression bands could be seen at ~25 kDa for 20 ˚C and 30 ˚C induction (**Figure 5.3**) which correlate with the bands seen by western blotting (**Figure 5.2**) and the deleterious effects of OafB protein expression were reduced by lower temperature and IPTG inducer concentrations. The T=4 hr time point sample for 30 ˚C

expression with 0.1 mM IPTG induction had a comparable OD_{650} to BL21(DE3) cells expressing SapA (OD_{650} 2.3 and 2.4 respectively).

As 30 ˚C expression with 0.1 mM IPTG induction showed potential for reducing the toxic effects of protein expression and resulting in successful expression of target protein in preliminary Coomassie stained SDS-PAGE gel analysis (**Figure 5.3**), it was trialled for all three of the OafB expression constructs. Rather than total protein, the periplasmic fraction from induced cells was isolated (Section 2.7.4) to enable clearer visualisation of induced periplasmic protein expression (**Figure 5.4**). The quantity of loaded protein was only partially correlated with the cell density of the analysed samples (twice as much sample was loaded for T=0 and the sample diluted by half for T=O/N) however, comparison of the relative intensity of constitutively expressed protein bands allows some indication of the level of target protein expression in each sample.

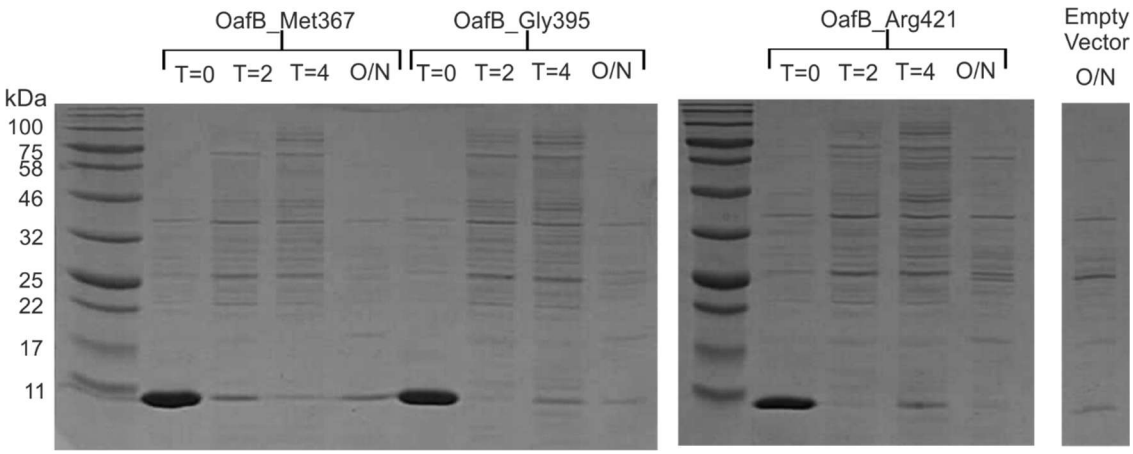


Figure 5.4| Periplasmic expression trial of OafB C-terminal constructs. Periplasmic extraction by ice cold osmotic shock of time point samples for BL21(DE3) expression trials of OafB_Met367, OafB_Gly395, OafB_Arg421, and Empty Vector negative control (pMV463). Samples were induced with 0.1mM IPTG and cultured at 30 °C. See Table 5.1 for strain details. Bands were visualised by Coomassie staining of SDS-PAGE gels.

Again, there was no indication of over-expression of any of the target proteins in these samples (**Figure 5.4**). A ~25 kDa band did appear in the overnight sample of OafB_Arg421 (MW=26.5 kDa) which was not seen in the uninduced sample, possibly indicating target protein, however the intensity of this band did not suggest that high yields could be achieved from purification (**Figure 5.4**). The ~10 kDa band in the T=0 sample for all expression constructs was presumed to be lysozyme which was added to the sample during periplasmic protein extraction. It appears more intense in this sample as twice as much lysozyme containing supernatant and was loaded into the well.

Proteins containing SGNH domains have been known to catalyse promiscuous activities and can catalyse both esterase and transferase reactions (Leščić Ašler *et al.*, 2010; Moynihan *et al.*, 2014). It was hypothesised that heterologous expression of C-terminal OafB could cause off target esterase or transferase activity within the cell, resulting in toxic effects and inhibition of protein over-expression. To address this, a catalytically inactive version of OafB_Arg421 with H621A mutation of the catalytic triad histidine residue (**Table 5.1**) was subject to expression trial in *E. coli* BL21 (DE3).

Growth of BL21(DE3) pMV475 cells following OD_{650} 0.4 induction of the catalytically inactive protein was comparable to the wild type OafB_Arg421 expressing strain BL21(DE3) pMV468 (**Figure 5.5**). A protein band at ~28 kDa appeared for OafB_Arg421^{H621A} which was strongest after T=4 hr induction and correlated with loss of the ~25 kDa band seen for the wild type protein. However, small scale nickel affinity purification (Section 2.7.6) from BL21(DE3) pMV475 after T=4 hr induction did not reveal clear signs of OafB_Arg421^{H26A} over-expression (**Figure 5.5 C**). A periplasmic binding protein from *Ochrobactrum anthropi* expressed under the same conditions was used as a positive control.

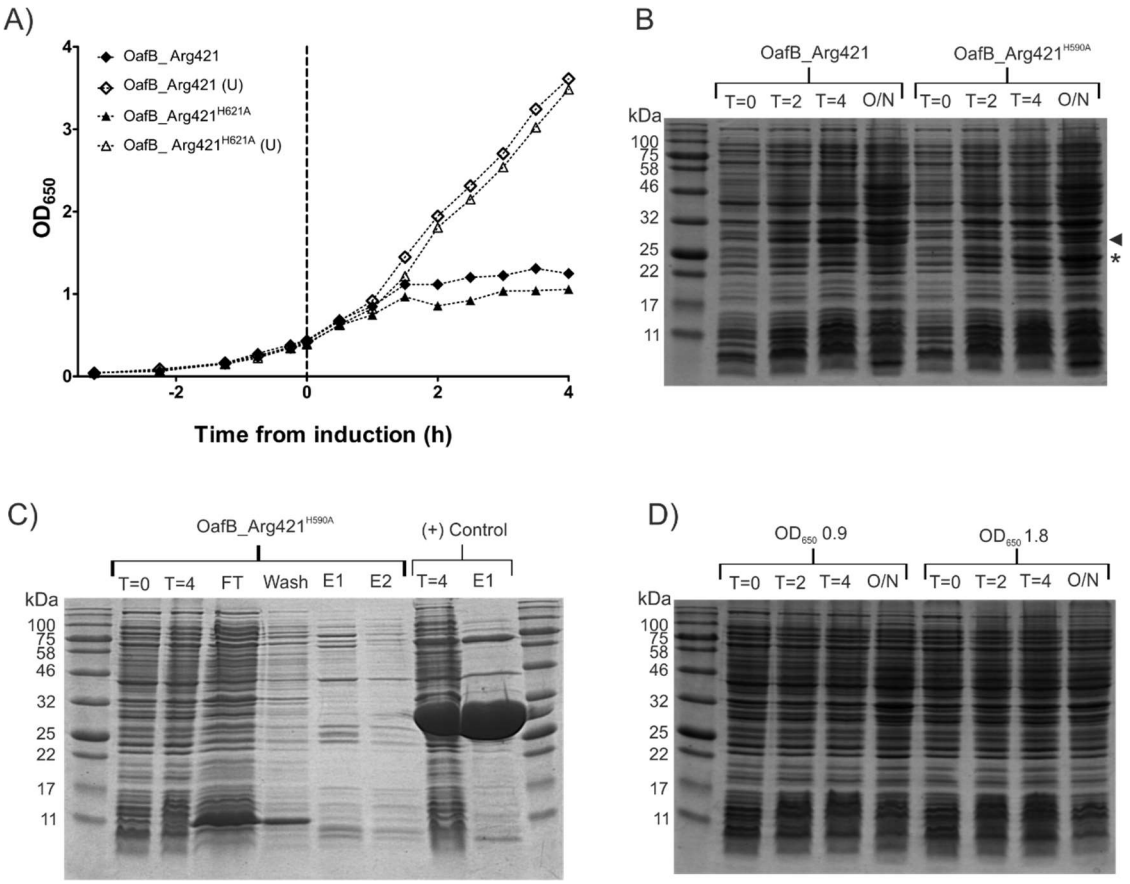


Figure 5.5| Effect of catalytic activity of OafB_Arg421 on cell growth and protein expression in BL21(DE3). A) Growth curve of BL21(DE3) expressing wild type OafB_Arg421 (diamonds) and catalytically inactive OafB_Arg421^{H590A} (triangles). Cells were grown at 30 °C and induced with 0.1 mM IPTG at OD_{650} 0.4. Uninduced samples are indicated by open shapes and induced samples by filled shapes. B-D) Coomassie stained SDS-PAGE gel of: B) total protein extracts from induced cultures as in (A) at T=0, T=2, T=4 hours and overnight arrow head and asterisk indicate potential target protein expression bands. C) NiNTA (qiagen) spin column purification of T=4 hr induction of OafB_Arg421^{H590A} as in (A) and positive control BL21(DE3) pMV585. FT= flow through, E= elution steps. D) Total protein extracts from BL21(DE3) expressing wild type OafB_Arg421 induced at OD_{650} 0.9 and OD_{650} 1.8. See Table 5.1 for strain details.

IPTG induction was conducted at three different optical densities of cells, OD_{650} 0.4, OD_{650} 0.9 and OD_{650} 1.8. The theory was that if cells did express protein then become stalled in their growth, the more cells present at the time of induction, the more protein produced in a single hit. Cells induced at OD_{650} 0.9 and 1.8 were abruptly halted in their growth following induction, reaching T= hr OD_{650} of 1.3 and 1.9 and did not produce sufficient yields of OafB_Arg421^{H25A} for detection by Coomassie stained SDS-PAGE analysis (**Figure 5.5 D**).

Before pursuing alternative expression systems to improve the yield of expressed C-terminal OafB proteins, an expression trial was conducted for the OafA C-terminal

expression constructs. Growth curves of IPTG treated cells suggested that the growth inhibitory effects of OafB expression were reduced for OafA constructs and SDS-PAGE analysis of total protein extracts suggested that over-expression of all three protein constructs had been successful, with highest yields of expressed protein for OafA_Lys355 in the overnight induced sample (**Figure 5.6**). Small scale nickel affinity purification suggested that OafA_Lys355 again gave the highest yield of soluble periplasmic protein from BL21(DE3) cells cultured at 30 °C with overnight 0.1 mM IPTG induction (**Figure 5.6 C**). The exact yield was not calculated due to the protein sample requiring further purification, however, the intensity of the band at ~27 kDa in the first elution fraction from this sample was visibly more intense than that of the other samples.

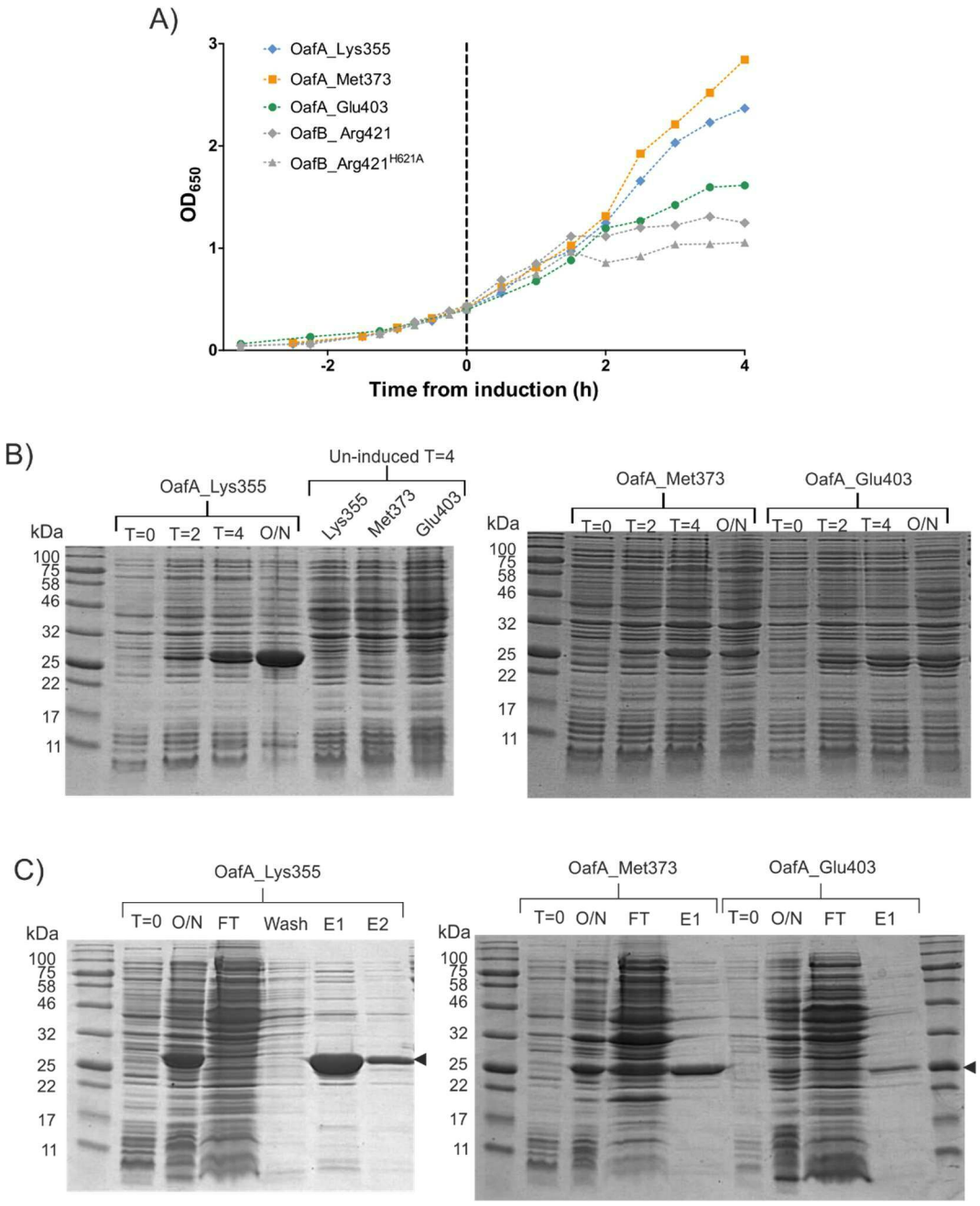


Figure 5.6| Expression trial of Bl21(DE3) cells expressing C-terminal OafA constructs. A) Growth curve of induced BL21(DE3) cultures expressing OafA C-terminal expression constructs in comparison to those expressing catalytic mutant (H590A) and wild type OafB_Arg421. B) Coomassie stained SDS-PAGE gel of total protein extracts from strains as in (A) and T= 4hr un-induced samples. C) Coomassie stained SDS-PAGE gel of small scale nickel affinity spin column purification. FT= Flow through, E = Elution fraction. Arrow heads indicate putative target protein expression bands

5.3 Purification and *in vitro* characterisation of OafA_Lys355

The expression of OafA_Lys355 was scaled up to 1 L, the periplasmic fraction isolated (Section 2.7.7) and purified by nickel affinity chromatography using Tris based buffers at pH 7.5 (Section 2.7.9) (**Figure 5.7 A**). The isoelectric point (pI) of OafA_Lys355 is

calculated by ExPASy server to be 6.68 (Gasteiger, Hoogland, Gattiker, Séverine Duvaud, *et al.*, 2005). As proteins tend to be least soluble near their pI (Shaw *et al.*, 2002), elution fraction E56 of purified OafA_Lys355 was dialysed into 20 mM sodium phosphate buffer pH 7.8 for further analysis. A_{280} absorbance of the dialysed elution fraction was recorded and yield of expressed protein calculated at 1.71 mg per litre of culture (Section 2.7.10).

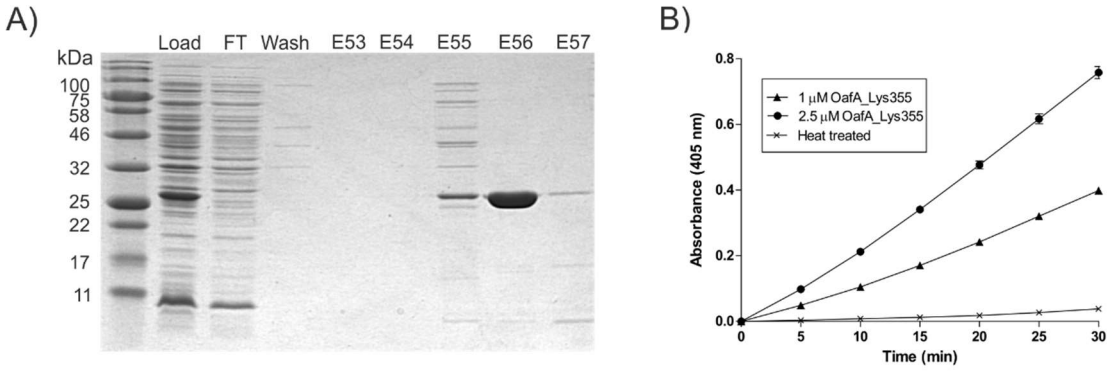


Figure 5.7| Purified OafA_Lys355 is catalytically active *in vitro*. A) Coomassie stained SDS-PAGE gel of nickel affinity purification of the periplasmic fraction from BL21(DE3) cells expressing OafA_Lys355. Load = total periplasmic fraction, FT = Column flow through after loading, Wash = proteins removed from the column during wash steps, E = elution fraction. B) *In vitro* Acetyl-esterase activity of OafA_Lys355 (fraction E56) assessed by hydrolysis of pNitorphenyl acetate (pNP-Ac) in 20 mM Sodium Phosphate Buffer pH 7.8. Heat treated control of 1 µM OafA_Lys355 was incubated at 90 °C for 5 min prior to testing. Error bars = SEM, N=3 technical repeats. Some error bars are occluded by the data points.

In vitro catalytic activity was confirmed for OafA_Lys355 via its ability to hydrolyse the ester substrate p-nitrophenyl acetate (pNP-Ac) (**Figure 5.7 B**), an assay commonly used to test SGNH domain function (Moynihan and Clarke, 2013; Baker *et al.*, 2014). The specific activity for OafA_Lys355 in these conditions was calculated at 4.9 nmol min^{-1} mg^{-1} and 4.2 nmol min^{-1} mg^{-1} for the 1 µM and 2.5 µM concentrations of OafA respectively, and the specific activity for the heat-treated control was 0.2 nmol min^{-1} mg^{-1}. This activity suggested that OafA_Lys355 was correctly folded and catalytically active in the conditions tested and heat denaturation of the sample abolishes this activity.

The small difference in specific activity between the two OafA concentrations tested could be caused by technical error introduced through sample preparation and also variation in the relative ratio of enzyme to substrate. Decreasing the concentration of enzyme in the reaction effectively increases the number of available substrate

molecules per enzyme, which may allow for a faster turnover of substrate per enzyme molecule. Enzyme kinetic parameters were not calculated here, however, as this would require an abundance of protein to repeat the assay over a range of substrate concentrations to calculate Michaelis-Menten parameters.

5.3.1 Biophysical and Biochemical analysis of OafA_Lys355

A second large scale expression and purification of OafA_Lys355 was conducted to provide enough protein for biochemical and biophysical analysis, unfortunately protein precipitation was observed following this purification, resulting in a reduction of protein yield to 0.74 mg per litre of culture (data not shown). Increasing salt concentrations for protein storage can help to shield surface charges on proteins to increase their stability or prevent aggregation (Zhang, 2012). Therefore, subsequent OafA_Lys355 purification fractions were dialysed into 50 mM sodium phosphate buffer pH 7.8, 100 mM NaCl . This resulted in a yield of 4.56 mg OafA_Lys355 per litre of culture (data not shown).

ESI-TOF MS (Section 2.8.1) was the first biophysical approach used to confirm that the purified protein was indeed OafA_Lys355. This gave a major peak with mass 31469.3 Da (Technology facility, University of York). The expected average molecular weight of Oafa_Lys355 is 31374.9 Da, giving a major peak that is 94.4 Da larger than the expected molecular weight of the expressed protein. This unexpected mass change indicated that OafA_Lys355 may have undergone post translational modification or be bound by a 94.4 Da adduct. Peptide mass fingerprinting was carried out by MALDI-TOF/TOF MS with trypsin digestion to confirm the identity that the 94.4 Da larger protein (Section 2.8.1). The result gave 45% sequence coverage and confirmed that the analysed protein was indeed OafA_Lys355 (**Figure 5.8**).

```
  1 MKGVSFRFSD TLKQVVEYRM DNSPWRPDIC FLNPDQDYSA FSKCQDKMTE
 51 KSFVVWGDSH AAHLMPGLKS VFGNSLNITQ RTASLCPPII GLQKDDRPYC
101 KDINDMVAKE ISDNKPTTVL MSALWPVYPM RDYLPETIKF LKDNKVKNII
151 IVGPFPVWKK TMIDTIEDMG INSGRTVPWS MTDETRNLRD NDKYLRELAK
201 EHSLTYISPL ETMCTESYCK AIIGNRIAYP IQYDNAHLTP EGSGWFIEEV
251 KKQISKLEVL FQGPSSGHHH HHH
```

Figure 5.8| Peptide mass fingerprinting of OafA_Lys355. Sequence coverage (highlighted in red) of OafA_Lys355 by peptide mass fingerprinting following trypsin digest carried out by Technology Facility, University of York.

Mass spectra were further analysed to identify any unassigned peaks which matched with the mass of predicted peptide fragments plus 94.4 Da but no substantial peak had been unassigned (data not shown). To give better resolution of the exact mass increase of OafA_Lys355, the protein sample was analysed by FT-ICR-MS which gives better resolving power than time of flight instruments. The spectra achieved from this analysis gave three main peaks, major peak= 31449.79, minor peak A = 31511.53, minor peak B = 31350.43 (**Figure 5.9 A**). Compared to the monoisotopic mass of OafA_Lys355 (31354.60 Da) minor peak B was only 4.17 Da lower than the expected mass. This could indicate the presence of two disulphide bonds (mass reduction of 2 Da per disulphide). Therefore, minor peak B was predicted to represent a small percentage of unmodified OafA_Lys355 in the sample. The major peak was 99.36 Da larger than minor peak B. The abundance of this species suggested that a consistent single covalent modification was present on the OafA_Lys355 protein. Minor peak A was 61.74 Da larger than the major peak, which is consistent with sodium and potassium ion adducts on the protein caused by the ESI ionisation method (Draper *et al.*, 2009).

N-terminal sequencing of OafA_Lys355 by MALDI-ISD-MS (Suckau and Resemann, 2003) (Section 2.8.1) revealed that the monoisotopic mass recorded for the N-terminal 'MKGVSFRF' fragment of OafA_Lys355 (1070.64 Da) was 99.14 Da larger than expected (**Figure 5.9 B**), however, the location of this modification could not be resolved using this technique. To target the N-terminal fragment of OafA_Lys355, peptide mass fingerprinting after AspN digestion was carried out (Section 2.8.1). This identified a 9 residue N-terminal fragment of OafA_Lys355 with a calculated monoisotopic mass of 1157.55 Da which was 99.31 Da larger than the expected monoisotopic mass of this fragment. Analysis of the ms/ms spectra (Proteomics group, University of York Technology facility) revealed that the 99.31 Da modification was most likely due to succinylation of the most N terminal methionine.

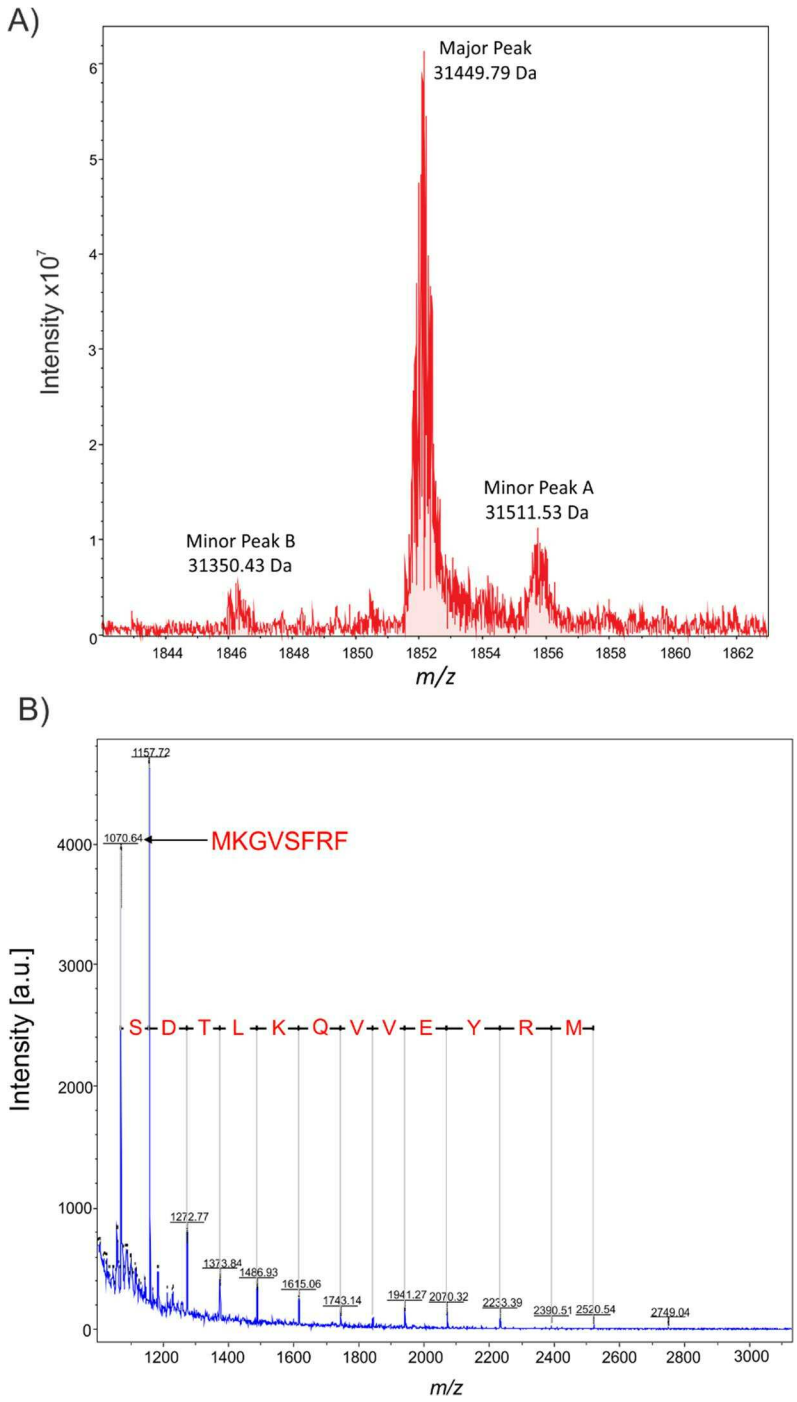


Figure 5.9| Mass spectrometry analysis of OafA_Lys355 by A) FT-ICT-MS and B) MALDI-ISD-MS. Monoisotopic mass expected for OafA_Lys355 is 31354.597 Da and for the MKSFRF fragment is 971.5 Da.

The N-terminal methionine of OafA_Lys355 was added to the native protein sequence as a result of restriction enzyme cloning and is not present in the native protein. There is evidence for bacterial proteins to be N-terminally succinylated (Chan *et al.*, 1993), however, this process is not well documented in the literature. Although an intriguing protein posttranslational modification, the N-terminal succinylation modification was

not relevant to the function of OafA and lays outside the scope of this project. Therefore, it was not pursued further as part of the *in vitro* functional analysis of C-terminal OafA.

1 dimensional NMR analysis of OafA_Lys355 confirmed that the protein was structured, due to visible amide proton and methyl group shifts above 8.5ppm and below 0.5ppm respectively (**Figure 5.10 A**). In addition, reducing the pH from pH 7.8 to pH 7.0 or adding DTT at 10:1 or 20:1 DTT:protein ratio indicated that changing the pH or reducing conditions of the sample did not significantly disrupt protein structure (**Figure 5.10 B**).

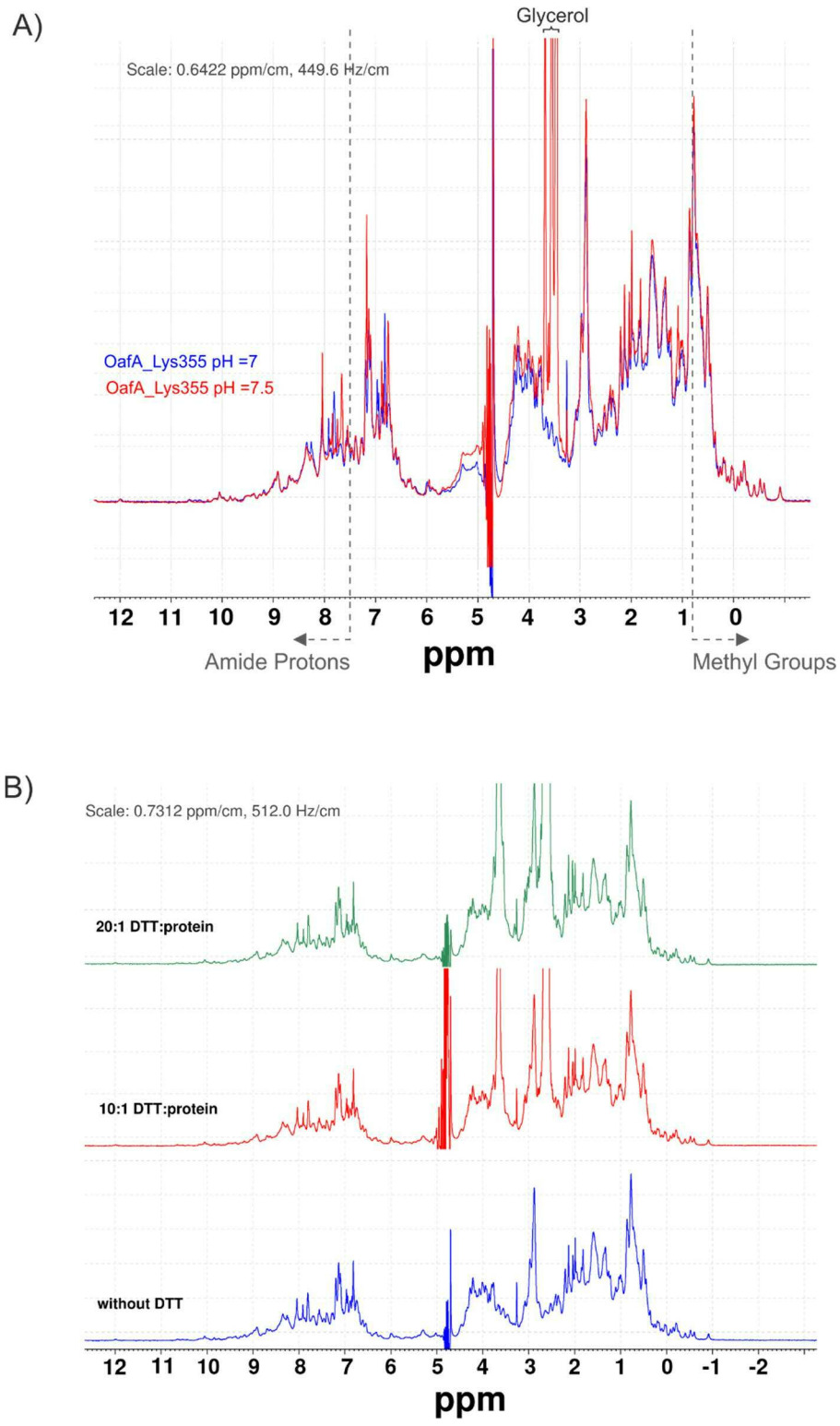


Figure 5.10| The effect of varying buffer conditions on OafA_Lys355, assessed by 1D NMR. A) 1D NMR spectra of OafA_Lys355 in 50 mM sodium phosphate, 100 mM NaCl at pH 7.0 (blue line) and pH 7.8 (red line). Extra peake between 3-4 ppm indicate contamination with glycerol due to spin column concentration of the protein sample. B) 1D NMR spectra of OafA_Lys355 in 50 mM sodium phosphate pH 7.0, 100 mM NaCl in the presence of 20:1 DTT:Protein (green line), 10:1 DTT:Protein (red line) and no DTT control (blue line). Inconsistencies in the peaks observed between 2.8 and 5 ppm are a result of DTT addition.

It should be noted that, although addition of reducing agent to the protein sample did not result in significant structural changes, it is not possible to determine whether the disulphide bonds that are predicted to be present in OafA_Lys355 (Section 3.3.2) were present before addition of the reducing agent or absent afterwards. FT-ICT-MS analysis did suggest the presence of two disulphide bonds within the purified protein (**Figure 5.9 A**) however further MS analysis would be required to determine the location of these bonds. For example, tandem mass spectrometry after fragmentation of the peptide backbone could reveal which pairs of cysteine residues form disulphide bonding pairs (Gorman *et al.*, 2002). In order to confidently determine the structural role of disulphide bonding in OafA_Lys355, the correct conditions for reduction of the disulphide bonds must be determined by mass spec analysis then the protein would need to be subject to these conditions during NMR analysis.

^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) NMR spectroscopy would give more detailed information about the structure of C-terminal OafA, such as substrate interactions and whether the linking region residues are structured or flexible, however, this approach requires 15N labelled protein. Expression trials of OafA_Lys355, OafA_Met373 and OafA_Glu403 in minimal media suggested that it would only be possible to express and purify sufficient 15N labelled OafA_Lys355 for HSQC NMR analysis, as minimal media expression trials for the two shorter OafA fragments with progressively fewer linking region residues were unsuccessful (data not shown). As HSQC NMR of just OafA_Lys355 would not provide conclusive evidence as to the organisation of the linker region residues without comparison to the shorter proteins, the amount of insight into the biophysical properties of OafA that could be gained from this analysis was deemed to be outweighed by the time and cost involved. For this reason, crystallisation of OafA_Lys355 was prioritised as the next approach.

Crystal trials using sitting drop vapour diffusion with PEG/Ion HT or JCSG+ HT 96 well plate buffers did not produce viable crystals for analysis. Before further crystallisation attempts were made for OafA_Lys355, the crystal structure of the C-terminal half of the *Salmonella enterica* serovar Paratyphi A OafB homologue, residues 377-640, which will henceforth be termed $OafB_{SPA}$_377, was solved by Sarah Tindall (University of York). Therefore, analysis of $OafB_{SPA}$_377 crystal structure was conducted to investigate the mechanistic role of the periplasmic region of O-antigen acetyltransferases.

5.4 Analysis of the crystal structure of the periplasmic domain of $OafB_{SPA}$

$OafB_{SPA}$ shares 78% sequence identity with $OafB_{STM}$ and 31% sequence identity with $OafA_{STM}$ (**Figure 5.11**). As SGNH domains usually have high structural homology but relatively low sequence identity, the structural data from $OafB_{SPA}$ was analysed to draw conclusions on the potential functional residues of all three proteins.

Expression, purification and X-ray crystallography of $OafB_{SPA}$_377 was carried out by Sarah Tindall (University of York). Subsequent bioinformatics-based analysis of this structure was also conducted by Sarah Tindall and results of this analysis were used to inform further experimental analysis of C-terminal OafA and OafB constructs by the thesis author. The remainder of section 5.4 is adapted from (Pearson *et al.*, 2020)(manuscript submitted for publication).

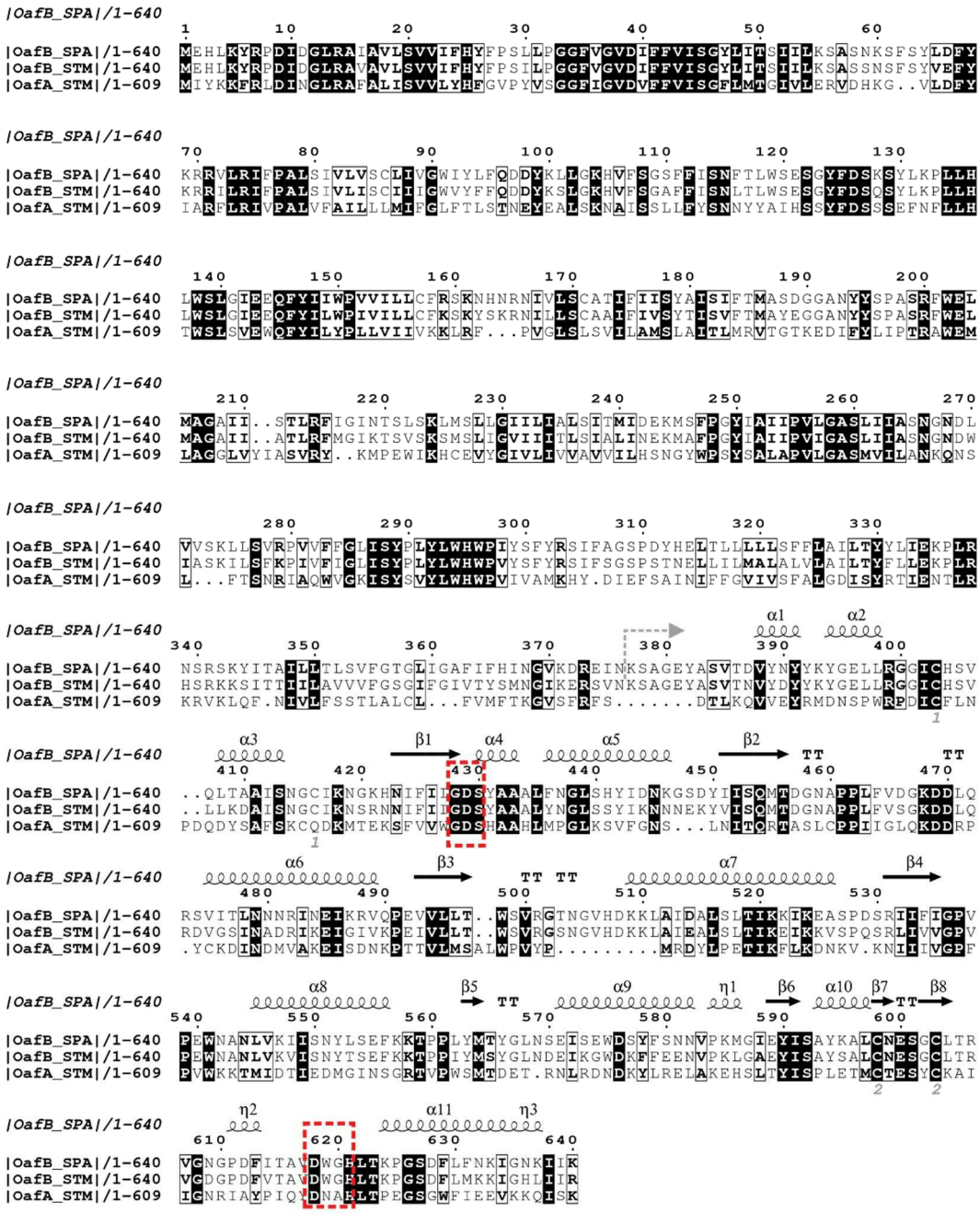


Figure 5.11| Alignment of *Salmonella* O-antigen acetyltransferases. SPA = *Salmonella enterica* ser. Paratyphi A, STM = *Salmonella enterica* ser. Typhimurium. Dashed grey arrow represents the crystallised region of $OafB_{SPA}$ and the structural features are outlined above the alignment. Dashed red boxes represent catalytic blocks I (GDS) and V (DXXH) of typical SGNH domains. Disulphide bonding pairs are indicated by grey numbers below the alignmnet.

The core structure of $OafB_{SPA}$_377 resembles an SGNH domain, with an α/β/α hydrolase fold consisting of five central β-strands surrounded by six α-helices (**Figure 5.12 A**). This structure also confirms the presence of two disulphide bonds which were predicted previously by sequence alignments (Section 3.3.2) (**Figure 5.12**) and these were verified using mass spectrometry (Sarah Tindall). The closest structural homologues to

$OafB_{SPA}$_377, as identified by the DALI server, are carbohydrate esterases from *Talaromyces cellulolyticus* (5B5S) and *Clostridium thermocellum* (2VPT). Each have an RMSD of 2.5 Å over 207 and 201 backbone residues, respectively.

Due to the low sequence identity but high structural homology among SGNH domains, a structure-based sequence alignment was conducted to assist in structural comparisons with other crystallised SGNH domains. The crystal structures of seven SGNH domains were used in this alignment including, $OafB_{SPA}$_377, the two closest structural homologues (5B5S, 2VPT), the only other SGNH domain from a fused acyltransferase with a solved crystal structure, OatA (5UFY) (Sychantha *et al.*, 2017), and three crystal structures of soluble extra cytoplasmic SGNH domain containing proteins (1IVN, 4K40, 1DEX) (Mølgaard *et al.*, 2000; Lo *et al.*, 2003; Williams *et al.*, 2014). The sequences of the C-terminal domain of four OafA/OafB homologues which span the $OafB_{SPA}$_377 crystallised region were also included in this alignment (Appendix VI). The comparison of the $OafB_{SPA}$_377 crystal structure to these structural and sequence homologues revealed important observations.

Firstly, the periplasmic portion of the linking region of $OafB_{SPA}$_377 (residues 377-421 which is visible from residue 380 in the crystal structure) is structured and forms an extension of the SGNH domain that we now term the SGNH extension (Ext^{SGNH}) (**Figure 5.12**). Importantly, an equivalent region of sequence is seen in the structure-based sequence alignment for OatA *S. pneumoniae* (residues 395-429 of OatA-S.pne Appendix VI) but this was not included in the solved crystal structure of the SGNH domain of this protein (Sychantha *et al.*, 2017).

The Ext^{SGNH} interacts extensively with the SGNH domain covering 1500 $Å^2$ of the SGNH domain with 38 amino acids of the SGNH domain, interacting with 32 (of 41) residues in the structured Ext^{SGNH} (Sarah Tindall). The Ext^{SGNH} also appears to partly occlude the active site of the SGNH domain as the solvent accessible surface area (SASA) of the catalytic triad residues (40 Å) is significantly lower than those in OatA (which does not include the Ext^{SGNH}), 2VPT and 5B5S (132 Å, 110 Å and 126 Å, respectively). Removing the 22 most N-terminal residues from the structure of $OafB_{SPA}$_377 ($OafB_{SPA}$_399, **Table 5.2**) increases the SASA of the catalytic triad residues of OafB to 107.9 Å (Sarah Tindall).

There are only 12 periplasmic residues between the C-terminal end of the final TMH of the membrane bound domain ($OafB_{SPA}$_367) and the start of the crystallised region of

$OafB_{SPA}$_377 ($OafB_{SPA}$_380). This restricts the potential orientations that the SGNH domain could take with respect to the AT3 domain, and suggests that the AT3 and SGNH domain are located in close proximity due to the structured Ext^{SGNH}. These observations raised the question of whether the Ext^{SGNH} plays a structural role to orient the SGNH domain towards the AT3 domain for acetyl group transfer or whether it plays a specific role in catalytic activity or substrate specificity, due to its close association with the SGNH domain active site.

Furthermore, an additional helix (α8) (**Figure 5.11**) is present in $OafB_{SPA}$_377 which comprises 10% of the SGNH domain volume (**Figure 5.12**). The region representing helix α8 in the structure based sequence alignment is present only for AT3-SGNH fused proteins involved in acetylation of LPS O-antigens (**Figure 5.13**, Appendix VI). The extra helix (α8) is in close proximity to the N-terminal end of the structured Ext^{SGNH} (**Figure 5.12**). This would also place this region close to the membrane domain due to the likely proximity of the Ext^{SGNH} to the membrane domain as discussed above. Therefore, α8 could play an integral role in acetyl group transfer specific to O-antigen acetylation.

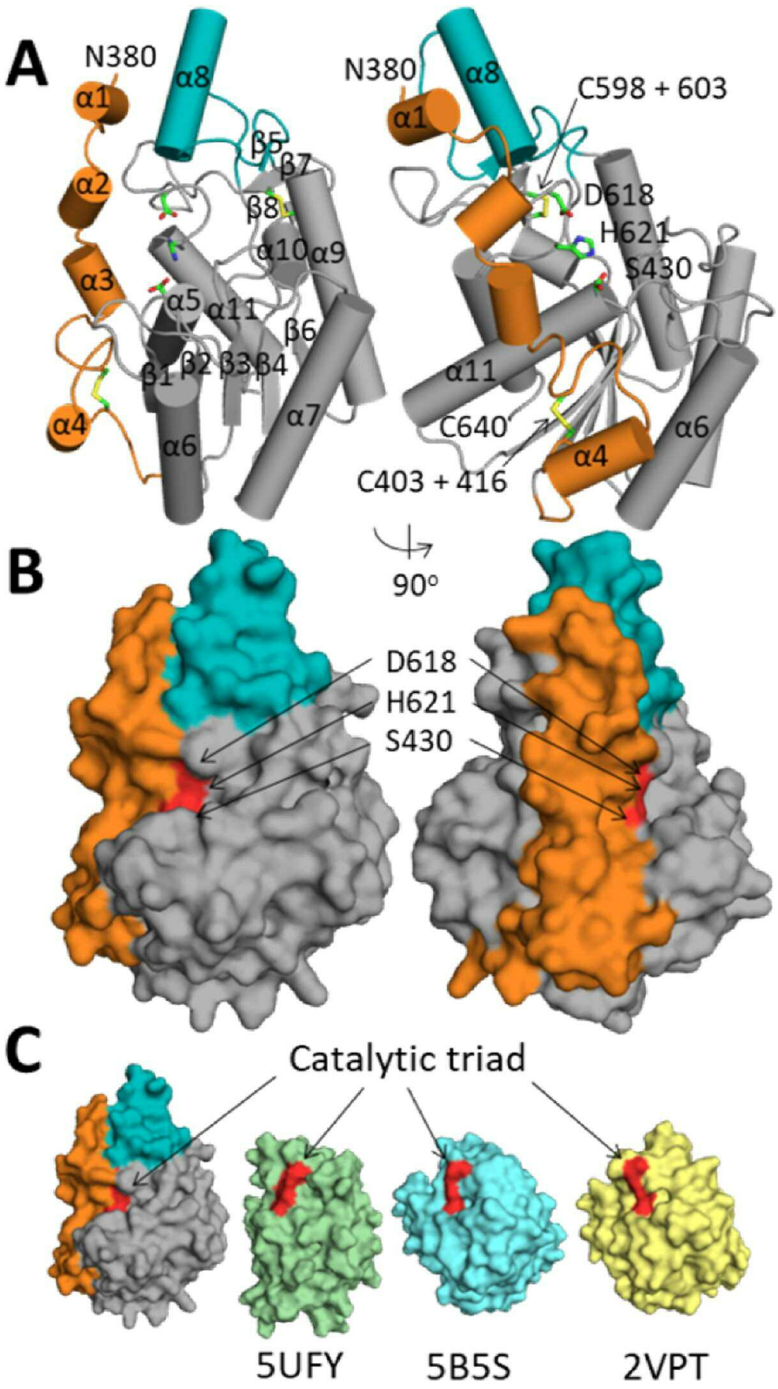


Figure 5.12| Analysis of the crystal structure of $OafB_{SPA}$_377. A) Cartoon representation of $OafB_{SPA}$_366 with α-helices and β-sheets numbered. The additional helix (α8) is coloured teal and the structured portion of the linking region (Ext^{SGNH}) is coloured orange. Catalytic triad residues and disulphide bonds are shown as sticks and labelled. B) Surface representation of $OafB_{SPA}$_366 with features coloured as above and the catalytic triad coloured red. C) Comparison of $OafB_{SPA}$_366 as in (B) to 5UFY (OatA), 5BBS and 2VPT. Figure produced by Sarah Tindall.

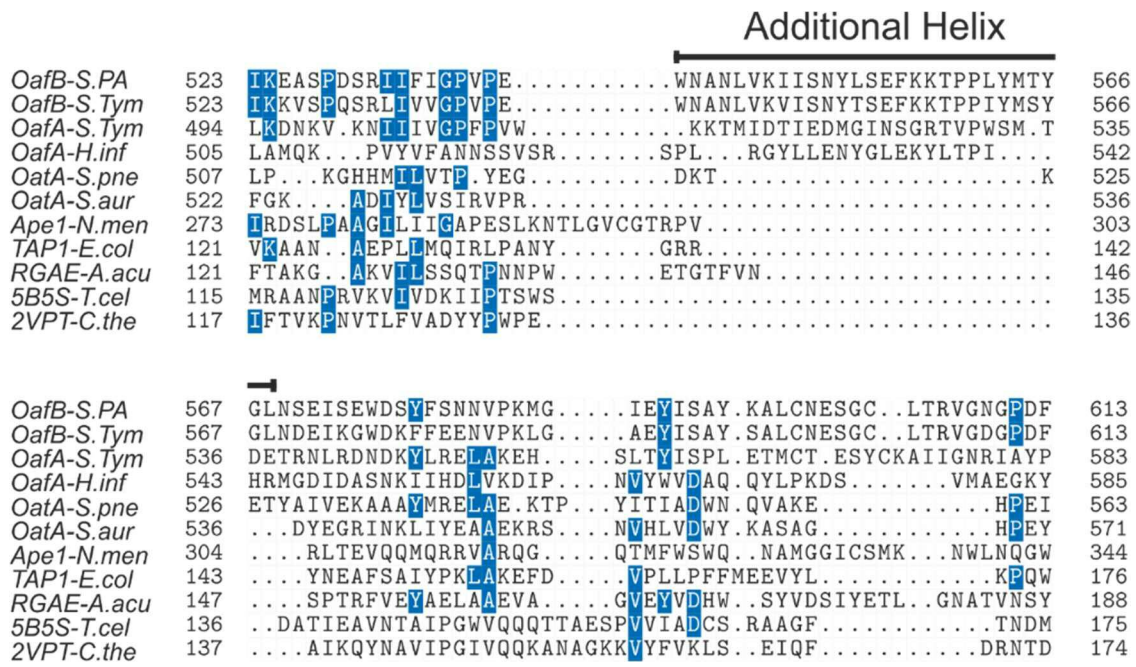


Figure 5.13| Structure based sequence alignment of OafB, OafA and structural homologues showing the additional helix ($OafB_{SPA}$_366 α8). Residues conserved in >50% of sequences are highlighted blue. See Appendix VI for full alignment and further sequence details.

A glutamate residue E600 $OafB_{SPA}$ is held in an outward facing position by the most C-terminal disulphide bond (C598-C603) (**Figure 5.14**). As glutamic acid is a polar amino acid with propensity to be involved in proton shuttling or stabilisation of reaction intermediates (Holliday *et al.*, 2009) it was proposed that this residue may be involved in the transferase reaction or may interact with positively charged residues in the periplasmic regions of the Acyltransferase 3 domain to facilitate acetyl group transfer. However, mutation of the equivalent residue in $OafA_{STM}$ (E569) did not result in significant reduction of *in situ* activity as the mutant protein retained 99.77% ±7.01 O-antigen acetylation level compared to wild type protein (**Figure 5.15**).

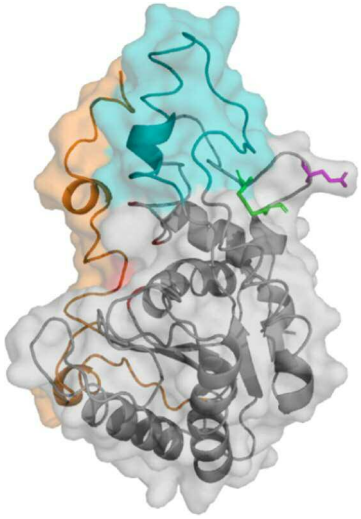


Figure 5.14| Crystal structure of $OafB_{SPA}$_366 shows that the C598-C603 disulphide bond (green sticks) holds E600 (magenta) in an outward facing conformation.

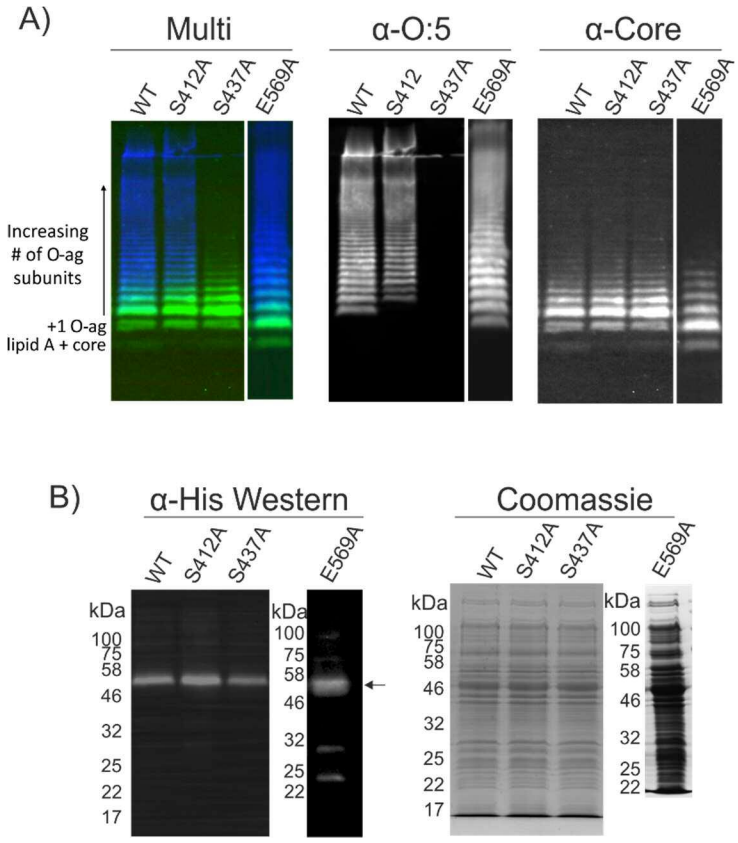


Figure 5.15| Functional analysis of S412A_$OafA_{STM}$ and E569A_$OafA_{STM}$ *in situ*. A) LPS immunoblot with crude LPS extracts from *Salmonella* ser. Typhimurium basal O-antigen strain expressing WT-OafA or point mutant variants (S412A, S437A, E569A). S437A_$OafA_{STM}$ sample is included here as a negative control of a known catalytically inactive OafA variant (Chapter 4). O:5 antibody binding (Blue) shows abequose acetylation and *Salmonella* LPS core antibody binding (Green) acts as a loading control. B) Corresponding anti-His western blot for expression of His tagged OafA and Coomassie stained SDS-PAGE gel to show uniform loading. Arrow indicates full length OafA protein.

SGNH domains are usually characterised by the presence of four blocks of sequence, containing conserved residues: block I – GDS, block II – G, block III – GxND and block V – DxxH (where 'x' is any non-proline residue) (Akoh *et al.*, 2004). In the *in silico* analysis described in section 3.3.2, only the catalytic triad residues of the SGNH domain (serine of block I, and aspartic acid and histidine in block V) had been highlighted by multiple sequence alignment (**Figure 5.11**), and the other catalytic blocks, expected to contain the oxyanion hole residues, were not well aligned (Appendix IV). *In situ* functional analysis revealed that the catalytic triad residues gave consistent results to analyses of typical catalytic triad activity in other SGNH proteins, with mutation of block I serine and Block V histidine resulting in loss of OafA function and mutation of block V aspartic acid reducing but not abolishing function (Lee *et al.*, 2006; Moynihan and Clarke, 2014b).

Structure-based sequence alignments with OafA and OafB gave a clearer view of the catalytic blocks of both proteins and again showed that while the catalytic triad residues are conserved in position in the structure, the oxyanion hole residues, glycine (block II) and asparagine (block III), are not (**Figure 5.16 A**). The conserved glycine of Block II is replaced by an asparagine in both OafB homologues and in the *H. influenzae* OafA homologue (N459_OafB$_{SPA}$) (See arrow in Block II alignment **Figure 5.16 A**). The structure of OafB$_{SPA}$_377 shows N459_OafB$_{SPA}$ to be within hydrogen bonding distance of a co-crystallised sulfate ion (**Figure 5.16 B**) suggesting that N459_OafB$_{SPA}$ could interact with bound substrate and participate in oxyanion hole formation. The alignment of OafA and OafB protein sequences (**Figure 5.11**) highlighted S437_OafA$_{STM}$ as a potential replacement for the block II glycine in the oxyanion hole. Indeed, the *in situ* OafA functional assay showed S437A_OafA$_{SPA}$ had <50% activity in comparison to wild type OafA (**Figure 5.15**).

Aspartic acid of the GxND motif (block III), typically involved in oxyanion hole formation (Mølgaard *et al.*, 2000), is not evident in OafA or OafB in the structure-based alignment (**Figure 5.16**). OafB$_{SPA}$ contains a GTNG motif (G502-G505_OafB$_{SPA}$) close to sequence block III, but the side chains of these residues are oriented away from the catalytic triad (**Figure 5.16**). These observations suggest that**,** although OafA and OafB display the typical catalytic triad of an SGNH domain, their oxyanion hole arrangement is atypical.

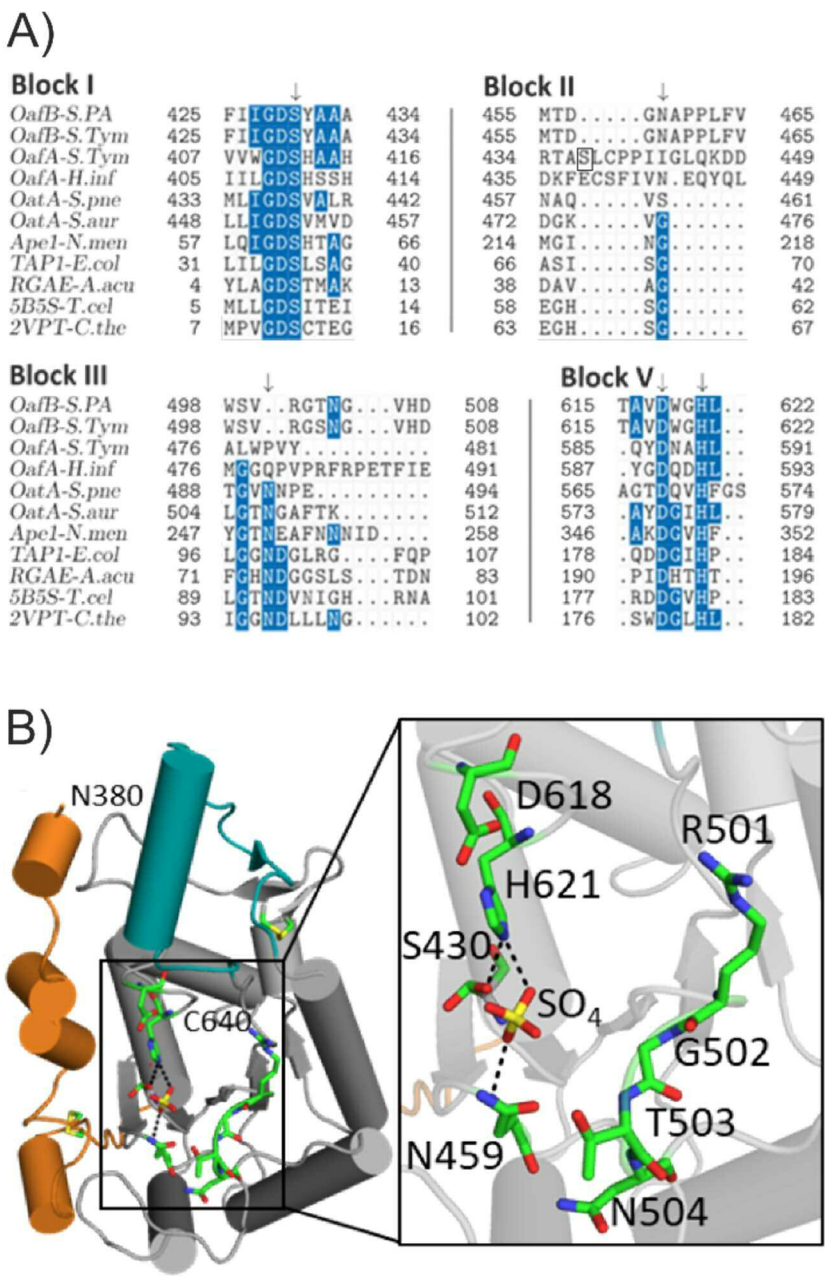


Figure 5.16| Analysis of the SGNH domain active site of OafB. A) Structure based sequence alignment as in Appendix VI showing conservation in the typical SGNH domain catalytic blocks. Residues conserved in >50% of sequences are highlighted in blue and the typical SGNH catalytic and oxyanion hole resides are indicated by arrows and putative oxyanion residue OafA_S437 is indicated by a box. B) Cartoon representation of $OafB_{SPA}$_377 with catalytic triad and potential oxyanion hole residues shown as sticks and hydrogen bonds to the co-crystallised sulfate ion shown as dashed black lines. Figure produced by Sarah Tindall and edited by thesis author.

5.5 *In vitro* acetyltransferase activity of the SGNH domain of O-antigen acetyltransferases

Our hypothesised mechanism of OafA and OafB suggests that the SGNH domain is responsible for the final step of the transferase reaction, in transferring the acetyl group

to the acceptor sugar (**Figure 3.1**). To support this hypothesis, an *in vitro* O-antigen acetyltransferase activity assay was developed to demonstrate whether the SGNH domain alone could perform acetyltransferase activity in solution.

O:5 serotyping antibodies had been demonstrated to specifically detect abequose acetylation of the STM O-antigen by immunoblotting (reviewed in Chapter 4) and OafA_Lys355 had been shown to be catalytically active against the esterase substrate pNP-Ac *in vitro* (**Figure 5.7**). Therefore, a reaction mixture of OafA_Lys355 (acetyltransferase), pNP-Ac (acetyl group donor) and crude LPS extracts from Basal O-antigen STM LT2 strain Path 993 (acetyl group acceptor) was set up and an LPS dot blot used to determine whether abequose acetylation could be detected (Section 2.8.4) (**Figure 5.17**). An excess of pNP-Ac was used in the reaction mixture in an effort to minimise the possibility that the reverse esterase activity would be catalysed by the SGNH domain, which would mask net acetyltransferase activity.

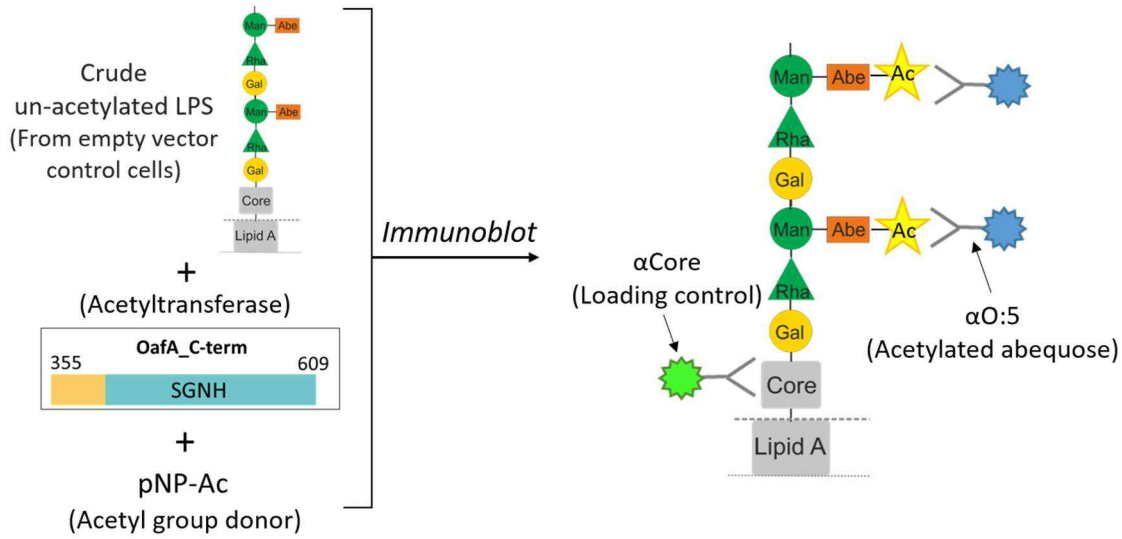


Figure 5.17| Experimental design of *in vitro* O-antigen acetyltransferase activity assay. See section 2.8.4 for details of experimental conditions.

Although the immunoblot quality is poor, the positive signal for O:5 serotyping antibody binding after incubation of OafA_Lys355 with un-acetylated LPS suggested that the SGNH domain of OafA alone is able to perform a transferase reaction and acetylate its physiological substrate *in vitro,* with the strongest signal being seen when LPS was diluted 50% in Tris buffer (**Figure 5.18**). This results supports the working model that the SGNH domain performs the last step in the transferase reaction; the transfer of the acetyl moiety to the acceptor carbohydrate (**Figure 3.1**).

Reaction component	Reaction mixture composition (µl)			
	A	B	C	D
871 µM OafA_Lys355	5	5	5	
100 µM $OafB_{SPA}$_377				5
Basal LPS (Path993)	30	15	15	15
50 mM KPi pH 7.8, 200 mM NaCl		15		
20 mM Tris pH 7.8, 100 mM NaCl			15	15
100 mM pNP-Ac	5	5	5	5

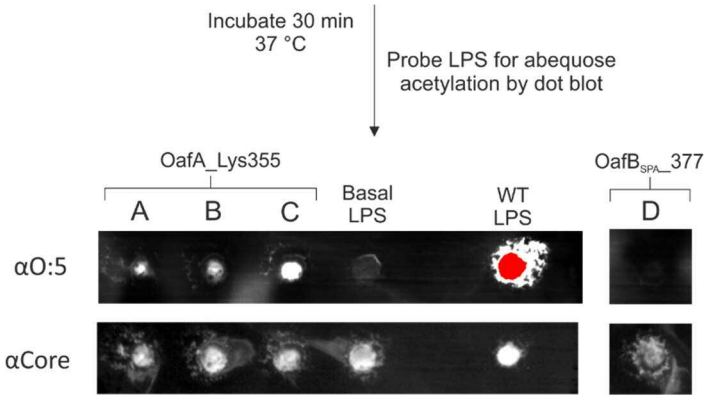
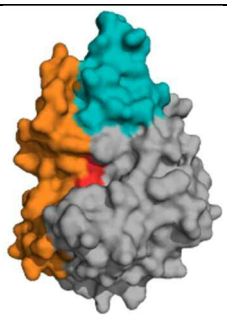
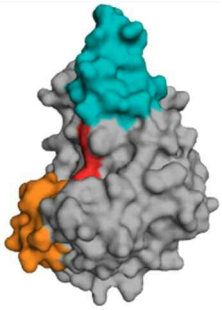
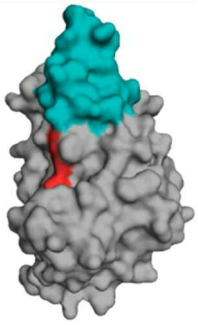


Figure 5.18| *In vitro* O-antigen acetyltransferase activity assay. Reaction mixtures indicated by the table were incubated at 37 °C for 30 minutes before 10 µl was spotted onto nitrocellulose membrane for immunoblotting with O:5 serotyping antibodies (abequose acetylation) and anti-*Salmonella* core antibodies as a loading control. As a controls basal LPS from Path 993 and wild type LPS from Path 932 was incubated with 12.5 mM pNP-Ac and 10 µl spotted onto PVDF. Red spots indicate overexposed areas.

$OafB_{SPA}$_377 did not show evidence of abequose acetyltransferase activity in the same assay conditions, however, abequose is not the native acetyl group acceptor of this protein, which acetylates rhamnose *in situ* and does not give rise to the O:5 serotype. As mentioned previously, SGNH domain containing proteins characterised in the literature are known to catalyse promiscuous acyltransferase or acylesterase activities (Section 1.3). Due to the structured Ext^{SGNH} resulting in an occluded active site of $OafB_{SPA}$, it was hypothesised that this region may play a role in determining acceptor substrate specificity of the SGNH domain. To test this hypothesis, further expression constructs of OafA and OafB were designed with fewer residues of the Ext^{SGNH} which should expose the active site (**Table 5.2**).

Table 5.2| OafA and OafB protein expression constructs and their predicted resulting structures. Constructs were cloned, expressed and purified by Sarah Tindall, University of York. Schematics of OafB structure were composed by Sarah Tindall and the summary table was composed by the thesis author.

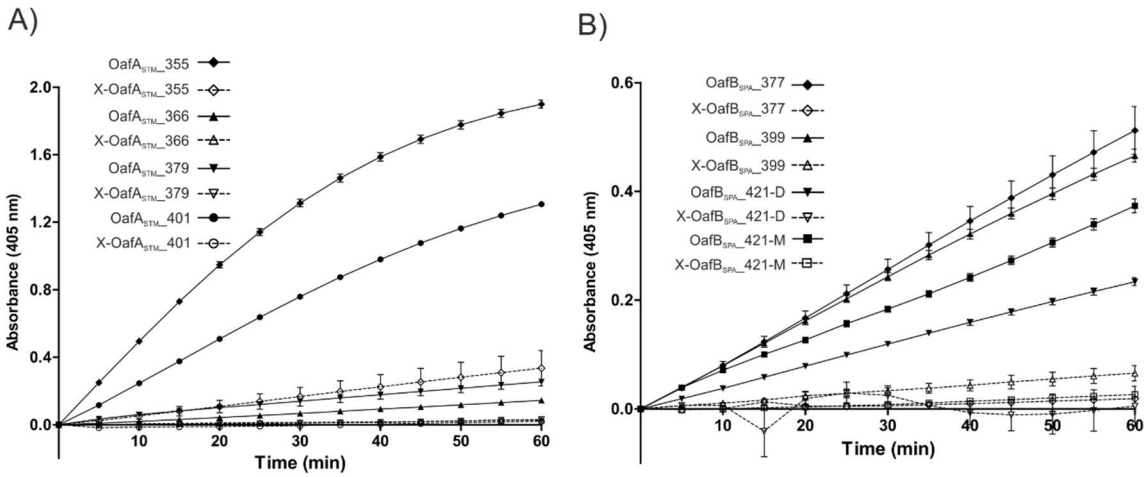
Name	Residues	Length	Description	Schematic representation (Inferred from $OafB_{SPA}$_377 structure)
$OafA_{STM}$_366	366-609	243	Full Ext^{SGNH}	
$OafB_{SPA}$_377	377-640	263		
$OafA_{STM}$_379	379-609	230	Half Ext^{SGNH} (exposed active site)	
$OafB_{SPA}$_399	399-640	241		
$OafA_{STM}$_401	401-609	208	No Ext^{SGNH} SGNH domain only	
$OafB_{SPA}$_421	421-640	219		

The longest OafA construct was reduced from residues 355-409 to residues 366-409 to better represent the equivalent crystallised region of $OafB_{SPA}$. The new constructs and their names are described in table (**Table 5.2**). All proteins were expressed and purified by Sarah Tindall. In fitting with this nomenclature, OafA_Lys355 will now be re-named $OafA_{STM}$_355. When the Ext^{SGNH} was completely removed from OafA and OafB, to express and purify just the SGNH domain, the resulting proteins ($OafA_{STM}$_401 and $OafB_{SPA}$_421) showed evidence of dimerization in solution. Size exclusion chromatography with multi-angle laser light scattering (SEC-MALLS) by Sarah Tindall

showed that $OafA_{STM}$_401 ran at 1.5 times its molecular weight (37 kDa, expected size = 24 kDa) indicating a transient dimer, whereas $OafB_{SPA}$_421 maintained two populations of dimerised (51 kDa, expected size = 50 kDa) and monomeric (26 kDa, expected size = 25 kDa) OafB which, when separated by size exclusion chromatography, were stable and did not equilibrate (Sarah Tindall, unpublished data).

To confirm the catalytic activity of these newly expressed proteins, ahead of further functional characterisation, their acetylesterase activity was assessed using the pNP-Ac hydrolysis assay. All proteins were able to demonstrate catalytic activity in the conditions tested (**Figure 5.19**), with the highest catalytic activity seen for constructs with the full Ext^{SGNH} for both OafA and OafB ($OafA_{STM}$_366, $OafB_{SPA}$_377). OafB showed consistently lower catalytic activity in these conditions in comparison to OafA, however, this approach is only able to detect acetylesterase activity of the SGNH domain, rather than the transferase activity that it is predicted to perform *in situ* (**Figure 3.1**). The specificities of these proteins for different acceptor sugars may also impact the availability of the active site for cleavage of acetyl-ester bonds of non-native substrates. Therefore, differences in catalytic activity of OafA and OafB in this assay may not reflect their ability to perform acetyltransferase reactions *in vitro,* but the results of this assay do suggested that the SGNH domain of OafA and OafB are catalytically active, irrespective of the presence or absence of the Ext^{SGNH}.

Heat treatment did not consistently abolish catalytic activity of the proteins in this assay (for example X-$OafA_{STM}$_355, **Figure 5.19 A**). However, significant reduction in catalytic activity of proteins after heat treatment was consistently seen. This suggests that OafA was not completely denatured by this heat treatment or that there was some propensity for the protein to refold following denaturation. Trichloroacetic acid treatment was considered an alternative negative control to precipitate out all the protein (Rajalingam *et al.*, 2009). However, as there is no indication that SGNH domains require cofactors to function, this would be effectively equivalent to a no protein control. As the proteins were purified by affinity chromatography followed by size exclusion chromatography, it is unlikely that contaminating proteins are present in these samples. Therefore, for subsequent analysis a no protein control reaction was considered appropriate and used as a definite negative result for the assays.



C)

Protein	Specific activity (nmol min^{-1} mg^{-1})	Standard deviation (±)
$OafA_{STM}$_355	4.07	0.55
$OafA_{STM}$_366	5.06	0.29
$OafA_{STM}$_379	0.58	0.15
$OafA_{STM}$_401	0.30	0.02
$OafB_{SPA}$_377	0.64	0.16
$OafB_{SPA}$_399	0.27	0.02
$OafB_{SPA}$_421-D	0.07	0.02
$OafB_{SPA}$_421-M	0.27	0.03

Figure 5.19| Catalytic activity of OafA and OafB constructs for investigation of the Ext^{SGNH}. A-B) Representative results of one biological replicate of pNP-Ac hydrolysis assay with OafA (5 µM) and OafB (20 µM) in 20 mM Tris-HCl pH 7.8, 100 mM NaCl. 'X' in figure legend denotes heat treatment at 100 °C for 10 min prior to assay and D/M after $OafB_{SPA}$_421 denoted to dimer or monomer population. Error bars = SEM, N=3 technical repeats. Some error bars are occluded by the data points. C) Specific activity of all constructs from at least 3 biological repeat assays. Note – $OafA_{STM}$_355 was assayed twice in Sodium Phosphate buffer and once in Tris buffer.

The *in vitro* O-antigen acetyltransferase activity assay (**Figure 5.17**, **Figure 5.18**) was repeated with all OafA and OafB constructs after dot blotting apparatus was sourced to improve the comparability of *in vitro* acetyltransferase activity results. This apparatus allowed concentration of sample loading within a confined area of the immunoblot membrane and reduced the opportunity for dilution of signal by diffusion of the sample across a larger surface area. Preliminary analysis suggested that 2 µl of LPS was the appropriate amount to be loaded in a single 'dot' as a balance between detection of LPS and reduction of debris which can block the membrane due to the LPS sample being a crude preparation (data not shown).

Strikingly, a positive signal for O:5 antibody binding was gained after incubation of un-acetylated STM LPS (derived from Path 993 **Table 2.1**) with all proteins tested, apart from $OafB_{SPA}$_377 (**Figure 5.20 A**). Thus, $OafA_{STM}$ SGNH domain is able to acetylate its native substrate in solution in both the presence and absence of the Ext^{SGNH} domain,

whereas OafB requires removal of the Ext^{SGNH} residues which occlude the SGNH active site in order to acetylate its non-native acceptor substrate. These results strongly indicated that the acceptor substrate specificity of the SGNH domain is constrained by the presence of the structured Ext^{SGNH}.

To obtain further insight into the effect of the Ext^{SGNH} on catalytic acetyltransferase activity of the SGNH domain, the reactions were scaled up and the products run on TSDS-PAGE for an LPS immunoblot. Rather than visualising a positive or negative spot for abequose acetyltransferase activity in this assay, this allowed the LPS bands to be separated and the profile of LPS O-antigen acetylation to be visualised. The profile of O-antigen acetylation is markedly altered by the presence of the Ext^{SGNH} residues covering the active site for both OafA and OafB (**Figure 5.20**). For OafA, the longest construct ($OafA_{STM}$_366) appears to have affinity for acetylation of LPS with 7 to 8 O-antigen repeats, however with the half and no Ext^{SGNH} proteins (**Table 5.2**) this profile changes to include prominent bands indicating acetylation of LPS with 5, 7, ~11 and ~14 O-antigen repeats for $OafA_{STM}$_379, and 5, 7 and ~14 O-antigen repeats for $OafA_{STM}$_401 (**Figure 5.20**). In the case of OafB, the inability of $OafB_{SPA}$_377 to acetylate the O-antigen abequose residues was confirmed, and a similar profile of LPS acetylation to the OafA proteins was revealed for the half and no Ext^{SGNH} OafB proteins (**Table 5.2**, **Figure 5.20**). Interestingly, the half and no Ext^{SGNH} OafA and OafB proteins also caused O:5 serotyping antibodies to bind to LPS which consists of just the lipidA and core oligosaccharides (See appearance of low molecular weight band in **Figure 5.20 C** below LPS with 1 O-antigen repeat). These observations support the hypothesis that Ext^{SGNH} determines substrate specificity, as removal of this domain results in promiscuity of LPS acetylation.

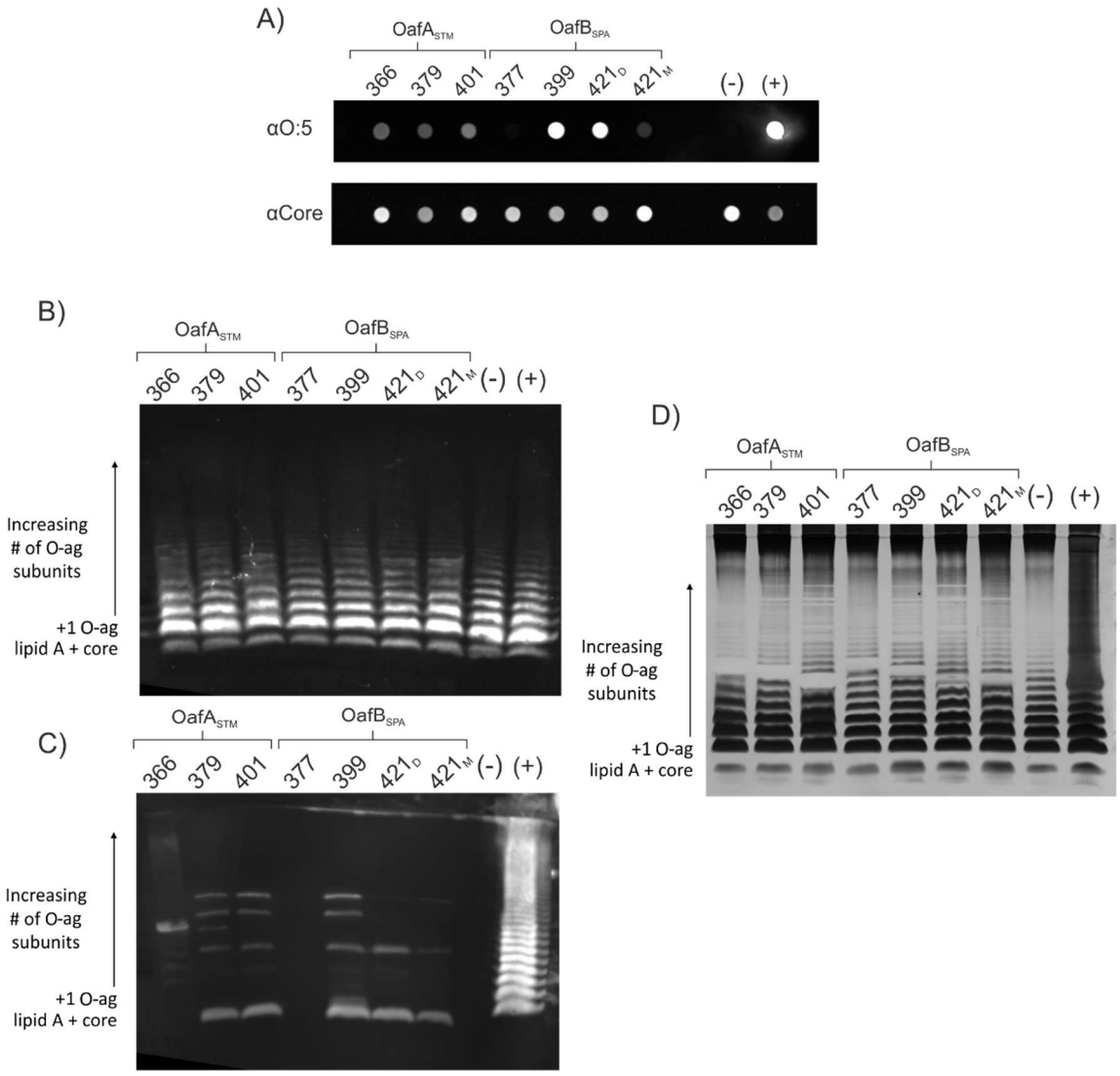


Figure 5.20| Effect of Ext^{SGNH} length on substrate specificity of OafA and OafB SGNH domain. A) Dot blot for acetylated abequose (αO:5) on basal LPS (Path 993) after incubation with purified the OafA and OafB SGNH domain constructs and pNPA as an acetyl group donor for the *in vitro* acetyltransferase activity assay. 10 µM OafA and OafB were used in these reactions. Number above each sample indicates the construct starting residue, see **Table 5.2** for further details. α-*Salmonella* Core antibody is used as a loading control. WT acetylated LPS (Path 932) is used as a positive control (+) and basal LPS from Path 993 is used as a negative control (-). Representative image of N=2 repeats. B-D) Representative image of LPS immunoblot N=2 repeats of scaled up *in vitro* acetyltransferase activity assay for constructs as in (A). B= LPS immunoblot α-Core signal as confirmation of LPS loading, C = LPS Immunoblot αO:5 signal, D= Silver stain TSDS-PAGE gel of LPS samples used in the assay. Gaps in the LPS banding pattern correlate with the molecular weight of the purified protein used in the assay.

5.6 Chapter Summary and future directions

This chapter demonstrates that the SGNH domains of OafA and OafB are able to acetylate the abequose sugar of the O-antigen of *Salmonella* ser. Typhimurium *in vitro* without the presence of the cognate fused AT3 domain. This supports the predicted role for the SGNH domain in the final step of acetyl group transfer to the acceptor substrate in AT3-SGNH fused acetyltransferases (**Figure 3.1**). In agreement with this, in the two component PatA/PatB peptidoglycan acetyltransferase system, PatB, a soluble SGNH protein, is responsible for transfer of the acetyl group onto the peptidoglycan substrate (Moynihan and Clarke, 2014b). As discussed in Chapter 3, this protein has also been suggested to interact and accept an acetyl group from an AT3-only acetyltransferase, further supporting the hypothesis that acetate is directly transferred from a membrane bound AT3 to a periplasmic SGNH for subsequent transfer to the acceptor substrate (Kajimura *et al.*, 2006; Moynihan and Clarke, 2010).

This chapter also investigates the importance of the periplasmic portion of the linking region (Ext^{SGNH}), which extends the SGNH structure and was not included in the first published structure of a bacterial AT3 fused SGNH domain involved in peptidoglycan acetylation (Sychantha *et al.*, 2017). Removal of the structured Ext^{SGNH} results in promiscuity of carbohydrate modification in *in vitro* O-acetyltransferase reactions (**Figure 5.20**). These findings suggest that, rather than acting as a flexible linking domain, the Ext^{SGNH} plays an intricate role in O-antigen acetylation, by modifying the accessibility and structure of the SGNH active site binding pocket to determine the carbohydrate acceptor that is acetylated.

The observation that these proteins have affinity for acetylation of LPS with specific numbers of O-antigen repeats is intriguing and raises questions about the way that the SGNH domain interacts with LPS. This pattern of LPS O-antigen acetylation is not seen when OafA acetylates LPS *in situ* (See positive control (+) in **Figure 5.20 B**), and the LPS molecules that are favoured are not the most abundant species in the sample. Therefore, the way that the O-antigen is presented to the SGNH domain *in vitro* may be the reason for this pattern. Due to its amphiphilic nature, LPS can form micelles or vesicles in solution (Sweadner *et al.*, 1977; Santos *et al.*, 2003). LPS with very short O-antigens is suggested to pack more tightly together in 3D structures, which could have affected the accessibility of these short O-antigens to the SGNH domain (Peterson *et al.*,

1986). Additionally, if mixtures of long and short O-antigens packed together, the longer O-antigen structures could have masked the shorter O-antigens from the SGNH domain. With regard to the longer O-antigen structures, these structures may have higher propensity to fold into more globular structures in solution, which again could inhibit interaction with the SGNH domain active site.

Appearance of low molecular weight band in **Figure 5.20 C** below LPS with 1 O-antigen repeat suggests modification of lipid A or core oligosaccharides by the SGNH domain of OafA and OafB in solution. It is possible for core sugars to be acetylated in other organisms, for example *Pseudomonas aeruginosa* (Knirel *et al.*, 2001). Therefore, we propose that this reactivity against LPS core for O:5 serotyping antibodies is caused by acetylation modification of the core oligosaccharides by the promiscuously acting SGNH domain, a modification not previously characterised for STM LPS.

In the absence of the Ext^{SGNH} it is possible that any hydroxyl group on carbohydrate molecules of the LPS O-antigen or core oligosaccharides might be acetylated in the *in vitro* acetyltransferase activity assay (**Figure 5.17**). Characterising the diversity of these promiscuous modifications by high resolution mass spectrometry might allow further understanding of how the structural changes around the catalytic triad can affect carbohydrate acetylation. Further characterisation of the acetylation modifications catalysed by SGNH domains with altered Ext^{SGNH} domains could also open up the possibility for rational design of Ext^{SGNH} residues which allow the SGNH domain to target a broader substrate range, or modification of the Ext^{SGNH} to restrict the active site for targeting of an alternative acceptor substrate. Thus, there are clear applications for this research in carbohydrate engineering.

Both monomer and dimer populations of $OafB_{SPA}$_421 were able to catalyse abequose acetylation *in vitro*. This raises the question of whether these proteins require dimerisation to function *in situ*. VanT is a protein with an AT3 fused cytoplasmic domain which requires dimerisation for function (Meziane-Cherif *et al.*, 2015), however this dimerisation has only been shown when the cytoplasmic domain is expressed without the AT3. Equally, dimerization of OafA and OafB SGNH domains is only observed when the Ext^{SGNH} is completely removed. Therefore, it is not yet clear whether this is a physiologically relevant interaction.

Dimerization may be an approach to minimize the access of water to the active site. Water acts as an acceptor for the acetyl group in SGNH hydrolase enzymes which act to hydrolyse acetyl groups and release them as acetate (Pfeffer *et al.*, 2013; Sychantha, Little, *et al.*, 2018). However, in the case of acetyltransferases, loss of the donor as acetate would reduce the efficiency of acetyl-transferase reactions. It is possible that rather than dimerisation, a water-limiting environment could be induced by interaction of the structured Ext^{SGNH} with the active site as well as interaction of the SGNH domain with the periplasmic surface of the AT3 domain. There are only 12 residues between the crystallised region of OafB and the AT3 domain, and thus, the AT3 and SGNH domains are likely to be located in close proximity.

The close proximity of AT3 and SGNH domains positions the periplasmic loops of the AT3 domain close to the SGNH additional helix (α8), suggesting a potential role for this functionally specific structural feature in the transferase reaction or for carbohydrate acceptor interaction. The close proximity of SGNH and AT3 domains is similar to the arrangement of domains seen in PglB, an oligosaccharide transferase from *Campylobacter lari* (Lizak *et al.*, 2011), with 13 TMH and a periplasmic domain. The periplasmic domain interacts via periplasmic loops in the transmembrane domain and both domains are hypothesised to interact with the peptide substrate (Lizak *et al*., 2011). It is possible that the same process occurs with OafA and OafB where the O-antigen substrate is modified at the interface of the AT3 and SGNH domains.

In addition to the α8 helix which is specifically conserved in O-antigen acetylating SGNH domains, further evidence for adaptation of fused SGNH domains towards specific carbohydrate acceptors is demonstrated by the non-canonical oxyanion hole structure of $OafB_{SPA}$_377. An atypical arrangement of oxyanion hole residues within the SGNH domain of O-acetyltransferases has become increasingly typical in recent years as further X-ray crystal structures emerge. There are examples of alternative oxyanion hole residues in the crystal structures of both AT3 fused and soluble SGNH domain O-acetyltransferases involved in acetylation of cell wall polysaccharides (Sychantha *et al*., 2017; Sychantha, Little, *et al*., 2018). Even OatA homologues which specifically acetylate N-acetyl muramic acid in *Staphylococcus aureus* and *Pseudomonas aeruginosa* have divergent oxyanion hole residues. These proteins were, however, found to have different specificities for the length of muroglycan molecules they are able to acetylate

(Sychantha *et al.*, 2017), highlighting further intricacies in the specificity of these previously assumed promiscuous catalytic domains.

Intricately tuning the specificity of carbohydrate acetylation by SGNH domains may be important to limit off-target acetylation and could be important for bacterial success. Inadvertent acetylation of complex carbohydrates could potentially have diverse and undesired biological effects due to the wide variety of cellular processes that can be affected by acetylation (Kahler *et al.*, 2006; Laaberki *et al.*, 2011; Bernard *et al.*, 2012; Kintz *et al.*, 2015; Knirel *et al.*, 2015; Baranwal *et al.*, 2017).

The crystal structure of $OafB_{SPA}$_377 has shed light on the functional residues and unique features of O-antigen acetyltransferases which had not previously been identified by *in silico* approaches. These findings have provided support towards, and further refinement of, the working model for O-antigen acetylation by OafA and OafB which will be discussed in further depth in (Chapter 6).

Chapter 6 Discussion and future directions

Acylation of carbohydrates is a common strategy in the microbial world to modify host and environmental interactions (Section 1.3). Although acylation of carbohydrates by AT3 domain-containing proteins is a ubiquitous trait across the domains of life and is of clinical and industrial interest, there is very little understanding of the catalytic mechanism by which these proteins are able to transfer acyl groups onto extra-cytoplasmic carbohydrate molecules. OafA and OafB represent two O-antigen acetyltransferases that offer a model system for mechanistic characterisation of AT3 domain acyltransferases. These proteins contain the same defined functional domains (**Figure 3.1**) but they specifically acetylate different monosaccharides within the *Salmonella* O-antigen. As the activity of OafA and OafB can be confirmed through relatively simple biological assays (Section 4.2), the above properties provide the potential for a range of hypotheses to be tested in two independent systems. Therefore, the aim of this research was to investigate the mechanism of action of O-acetylation by AT3 domain containing O-antigen acetyl transferases OafA and OafB so that this knowledge could be applied across the diverse protein family.

6.1 A refined mechanistic model of O-antigen acetylation by AT3 domain-containing O-acetyltransferases.

Prior to this research project, a proposed mechanism of action for O-antigen acetylation by OafA or OafB had not been defined in the literature. The only AT3 domain containing acetyltransferase with a clear proposed mechanism was the peptidoglycan acetyltransferase Oat (Moynihan *et al.*, 2014). However, this only made broad speculation that the N-terminal domain was responsible for delivering acetate to the periplasm and the C-terminal domain transferred this group to the acceptor substrate, and it hypothesised that this transfer involved the SGNH catalytic triad. No mechanistic roles of residues within the AT3 domain had been highlighted and the mechanism of acetate delivery to the SGNH domain had not been speculated. Some *in silico* and mutagenesis analysis of the functional residues of the AT3 only O-antigen acetyltransferase of *S. flexneri* had also been published prior to this research (Thanweer and Verma, 2012). Residues identified as critical were mostly assigned to maintaining protein production and structure, rather than playing catalytic roles. It and it was

unknown whether the proposed functional residues were shared between AT3-only and AT3-SGNH fused proteins.

This project has built on previous research into AT3 domain containing proteins, through *in silico* analysis of experimentally confirmed O-antigen acetyltransferases, to inform a working model for O-antigen acetylation by OafA and OafB. This model suggested that an acetyl group from an, as yet unconfirmed, donor in the cytoplasm would be transported across the inner membrane by the AT3 domain, before being subsequently transferred to the O-antigen acceptor sugar by the SGNH domain (**Figure 3.1**). This model is in agreement with the predicted mechanism of periplasmic peptidoglycan acetylation by the AT3-SGNH fused acetyltransferase OatA (Sychantha *et al.*, 2017). However, neither of these predicted mechanisms assigned clear roles to functional residues within the protein of interest or suggested how the AT3 and SGNH domains interact.

Optimisation of the *in situ* functional assay for OafA (Chapter 4), allowed the significance of functional residues, proposed following *in silico* analysis (Chapter 3), to be tested. This, combined with *in vitro* structural and functional characterisation of OafA and OafB (Chapter 5), has allowed a refined mechanistic model for O-antigen acetylation to be proposed. We propose that cytoplasmic acetyl-CoA is the acetyl group donor in this system. The acetyl-CoA donor is coordinated, by conserved arginine and histidine residues in TMH 1 of the AT3 domain, to present the acetyl group to the periplasmic side of the inner membrane. Conserved phenylalanine, aspartic acid and serine residues on the periplasmic side of the inner membrane then mediate transfer of the acetyl group to the O-antigen by the SGNH domain, either through interaction with the SGNH domain or the O-antigen acceptor, to orient the catalytic triad of the SGNH domain with the acceptor substrate, for the transferase reaction to occur (**Figure 6.1**).

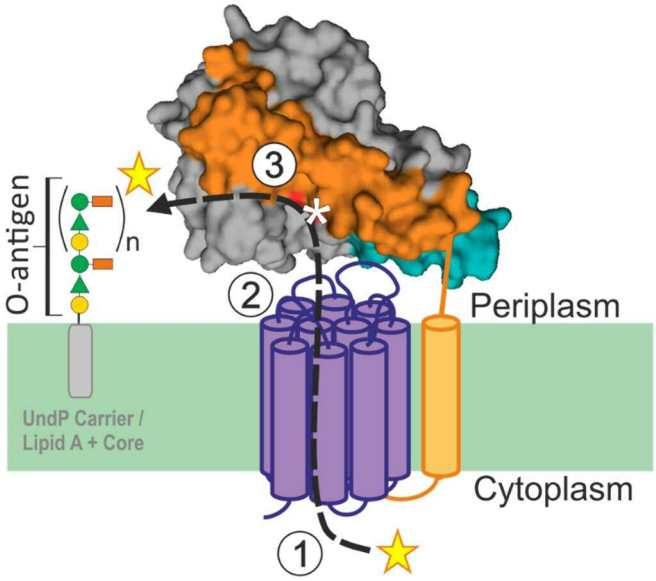


Figure 6.1| Refined model of AT3-SGNH fused O-antigen acetyltransferases. Periplasmic Ext^{SNGH} (Orange) is structured, therefore positioning the SGNH domain (Grey) close to the AT3 domain (Purple), this orients the additional α8 helix (Teal) in close proximity to the AT3 domain with potential for interaction between the two domains. These predictions result in the current working model: 1) Cytoplasmic acetyl group donor (hypothesised to be acetyl Co-A) interacts with conserved Arg in TMH1, the acetyl group is presented on the periplasmic side of the inner membrane and this process involves catalytic His residue of TMH1. 2) Conserved Phe, Asp and Ser mediate transfer of acetate to the SGNH domain. 3) SGNH domain catalyses addition of the acetate to specific O-antigen monosaccharide. The active site of the SGNH domain is highlighted by an asterisk.

The experimental observations that informed this refined mechanistic model address the questions posed in the introduction: they have highlighted essential residues in the membrane-bound AT3 domain which give clues in their role in acetyl transfer (Chapter 4), they have provided insight into the juxtaposition of the AT3 and SGNH domains through their orientations being constrained by the structured Ext^{SGNH} (Section 5.4), and they have supported the hypothesised role for the SGNH domain in the final step of the transferase reaction and shown that this domain can function independently from the AT3 domain (Section 5.5).

Although this research has shed further light on the intricacies of extra cytoplasmic carbohydrate acylation by AT3 domain-containing proteins, there is still scope to build on this knowledge and address the further questions which remain. Some of the key

questions that still remain or that this research has raised will be discussed in the following sections.

6.2 Is acetyl-CoA the only possible donor substrate?

Previous papers have speculated that acetyl-CoA is the donor substrate in transmembrane carbohydrate acetylation (Moynihan and Clarke, 2010; Moynihan *et al.*, 2014; Sychantha *et al.*, 2017). The evidence presented from this project also offers an argument for a CoA as the acyl group donor in this role. The highly conserved and essential TMH1 arginine and histidine residues, highlighted in this analysis, are proposed to co-ordinate the CoA molecule within the transmembrane domain to present the acyl group on the periplasmic side (Chapter 4). The CoA coordinating roles of arginine and histidine residues have been demonstrated in other acyltransferase proteins (Wu and Hersh, 1995; Jogl *et al.*, 2004)(Discussed further in Chapter 4). Therefore, these residues provide a feasible conserved mechanism for periplasmic localisation of a range of acyl groups linked to CoA for their transfer to periplasmic acceptor substrates.

An alternative to this hypothesis is that the acyl group donor could be lipid linked and already present on the periplasmic side. If this was the case, it is most likely that this lipid linked acyl group would be recognised by the membrane bound AT3 domain, which would allow the SGNH domain to catalyse transfer from the lipid linked donor to the O-antigen. This mechanism draws parallels with the mechanism of O-antigen glycosylation by the Gtr genes, where GtrC, an integral membrane protein with multiple transmembrane domains, transfers glucose from undecaprenyl phosphate carrier on the outer leaflet of the inner membrane to the O-antigen (Allison and Verma, 2000). In this system, GtrC functions alongside two other proteins GtrA and GtrB, which synthesise the lipid linked glucose on the inner leaflet of the inner membrane (GtrB) and flip it to the periplasmic side (GtrA). The genes for these proteins are found in operons and function together to allow synthesis of the periplasmic glucose donor and transfer to the O-antigen (Allison and Verma, 2000; Davies *et al.*, 2013). If this mechanism was used by AT3 proteins we may expect to find surrounding genes involved in synthesis and periplasmic delivery of the lipid linked acetyl group donor, which are not observed.

OafA and other O-antigen acetyltransferases do not appear to require association with any other system in order to perform the acetylation modification. A key example being

OafB, which is prophage associated, carried on the BTP1 phage genome, and has shown to be sufficient to cause O-antigen acetylation (Kintz *et al.*, 2015). Additionally, as there are many AT3 proteins across a variety of organisms that transfer diverse acyl groups such as butyryl/succinyl/propionyl (Leigh *et al.*, 1987; Hara and Hutchinson, 1992; Cong and Piepersberg, 2007), diverse sets of synthesis machinery would be required to synthesise these lipid-linked acyl chains and flip them to the periplasmic side of the inner membrane. These systems are yet to be found in association with bacterial AT3 acyltransferases.

A second alternative acyl group donor is acyl carrier protein (ACP). ACPs are universal carriers of acyl intermediates that are functionalised by addition of a phosphopantetheine group from CoA (Byers and Gong, 2007) (**Figure 6.2**).

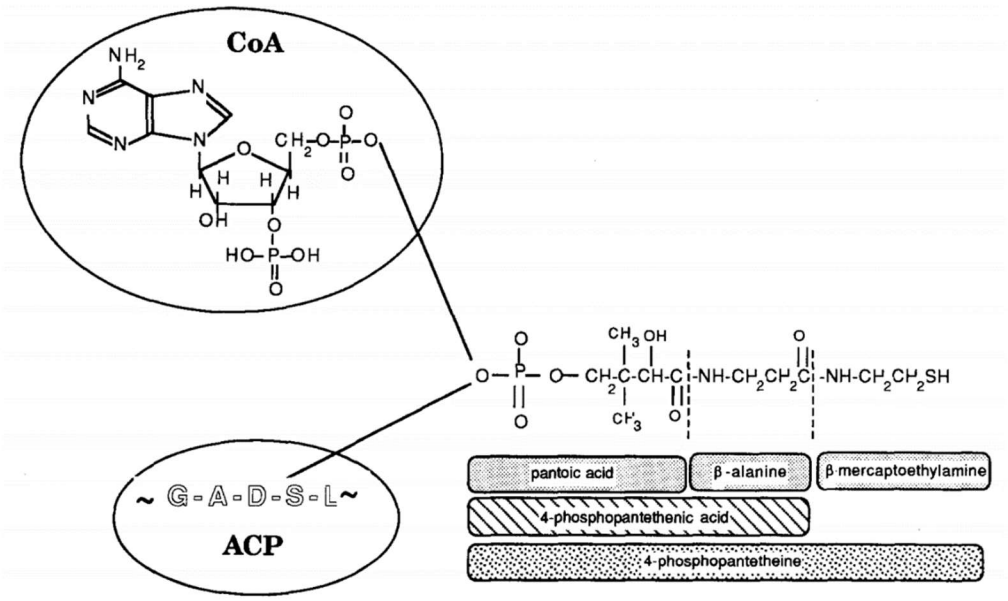


Figure 6.2| Phosphopantetheine group shared by coenzyme A and acyl carrier protein. Figure taken from (Slabas and Fawcett, 1992)

No common ACP binding motif is known for ACP partner enzymes, however, acidic amino acids in ACP alpha helix II have been implicated in binding ACP interacting partners through electrostatic interaction (Byers and Gong, 2007). Therefore, it is possible that the critical arginine's on the cytoplasmic side of the AT3 domain are involved in ACP interaction rather than CoA for the phosphopantetheine group to insert into the membrane domain and interact with catalytic histidine. Unfortunately, as is the case for CoA, ACP is essential for bacterial survival (Vagelos, 1973). Therefore, depletion

of either of these acyl donors in the cell to rule them out of the mechanism is not feasible.

A final alternative candidate for acyl group donor is acyl-phosphate. Acetyl-phosphate is a high energy intermediate between acetate and acetyl-CoA and has an evolving role in the literature as a dual phosphate and acetyl donor (Verdin and Ott, 2013). It may be possible to test the role of this acetyl donor in O-antigen acetyl transfer as it is possible, at least in *E. coli*, to create acyl-phosphate deficient cells (Seepersaud *et al.*, 2017). Therefore, acetyl-phosphate deficient *Salmonella* mutants could be created to test this acyl donor by knockout of acetyl-phosphate synthesis genes.

A potential approach to determine which acyl donors can be processed by OafA and OafB would be to develop an *in vitro* assay for acetyltransferase activity involving the whole protein. This would require expression, purification and reconstitution of OafA or OafB into a membrane mimetic and design of an experiment to show whether an acetyl group donor has been processed by the reconstituted protein. Careful design would also be required to ensure that any acetyltransferase activity would be detectable and that this could be specifically attributed to the processing of the provided acetyl group donor. A potential approach for this would be to use radiolabelled acetyl-CoA. The use of ^{14}C labelled acetyl-CoA is an established method to record direct transfer of acetate from acetyl-CoA to histones *in vitro* (Racey and Byvoet, 1971). However, the added complexity of the acetyl group acceptor being a carbohydrate molecule which is not readily commercially available in purified form poses difficulties for design of this assay.

An alternative solution to determine the acyl group donor could also be to solve the crystal structure of the AT3 domain. There are currently no published crystal structures of AT3 domains. Crystallisation of the full-length protein or AT3 domain was not pursued within this research project due to the foreseeable difficulties of expression purification and crystallisation of membrane proteins (Chapter 5). However, a crystal structure for a membrane bound O-acetyltransferase (MBOAT), a multiple membrane spanning O-acetyltransferase protein that is distinct from AT3 proteins, was solved during the course of this project (Ma *et al.*, 2018). Therefore, it may be possible to achieve a crystal structure of OafA or OafB using similar methods. Indeed, both OafA and OafB have been expressed and purified as full-length proteins previously (Reyme Herman, University of York), showing potential for set up of crystallisation trials. The crystal structure would

provide information for *in silico* substrate docking analysis, to determine the most likely acyl donor. An approach used previously to determine acyl-CoA or ACP interactions in crystal structures (Keatinge-Clay *et al.*, 2003; Galaz *et al.*, 2013).

Biophysical analysis of purified full length proteins or just the AT3 domain of OafA and OafB might also be an option to determine which acetyl-donor molecules these proteins can interact with. Ligand screening by STD-NMR is an established method to probe transient protein-ligand interactions by analysing intramolecular transfer of magnetisation from the protein to the interacting protons of the test ligand (Viegas *et al.*, 2011; Venkitakrishnan *et al.*, 2012). Although an initial trial of this technique with OafB was inconclusive (Andrew Brentnall, University of York), testing the purified proteins solubilised using different detergents may give more clearly interpretable results and is worth re-visiting if purified full length protein or AT3 domains can be obtained.

6.3 Do AT3-only proteins follow the same mechanism as AT3-SGNH fused proteins?

This research has demonstrated that the SGNH domains of OafA and OafB are able to catalyse the transfer of acetyl groups to the O-antigen (**Figure 5.18**, **Figure 5.19**). Not yet speculated in this thesis, is the catalytic mechanism of transfer of his acetyl group. The main two mechanisms by which acyl transfer reactions take place are 'ping-pong' and 'one step catalysis'. Ping-pong reactions involve a stable acyl-enzyme intermediate being formed with a catalytic residue in the enzyme active site, before the acyl group is transferred to the acceptor substrate. One step catalysis involves direct transfer of the acyl group from the donor to the acceptor when a ternary complex is formed between the enzyme, donor and acceptor molecules (**Figure 6.3**).

The ping-pong mechanism has been posed for peptidoglycan acetylation by OatA (Sychantha and Clarke, 2018) and is the suggested mechanism for the soluble SGNH protein that acts in a two component peptidoglycan acetylation system (PatB) (Moynihan and Clarke, 2014a). Therefore, it is likely that in OafA and OafB the transfer of the acetyl group occurs by a ping-pong reaction, were a stable acetyl-enzyme intermediate is formed with the catalytic serine of the SGNH domain, before the acetate is then transferred to the acceptor sugar.

The Ping-Pong mechanism is well established in other acyl transferase reactions. For example, in human arylamine acetyltransferases (NAT-1 and NAT-2) a covalent acyl-cysteine intermediate is formed before the acetyl group from acetyl-CoA is transferred to the acceptor substrate (Zhou *et al.*, 2013). The acceptor and donor binding sites in the NAT proteins overlap and this is consistent with enzymes that perform ping-pong catalytic reactions (Matthews, 1999). These proteins bear close resemblance to the N-hydroxyarylamine O-acetylransferase OAT of *S. enterica* which also functions through a catalytic cysteine (Watanabe *et al.*, 1994). Additionally, the eukaryotic enzyme CasD1, a multiple membrane spanning protein with N-terminal globular domain that has a Ser-Asp-His catalytic triad, also acts through a covalent enzyme intermediate with serine in the catalytic triad of the globular domain (Baumann *et al.*, 2015).

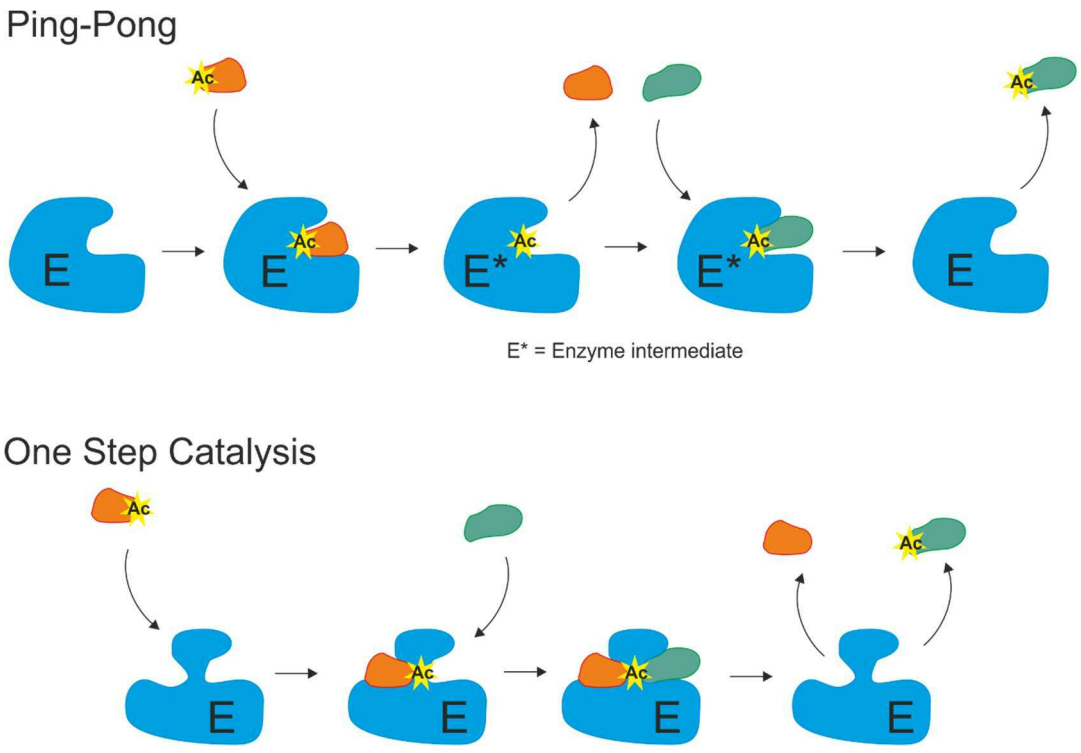


Figure 6.3| Mechanisms of enzymatic acetyl transfer. Ping-pong reaction involves the binding of the acetyl donor, formation of a stable acyl-enzyme intermediate and release of the donor. Followed by, binding of the acceptor to the stable acyl-enzyme intermediate, transfer of the acetyl group to the accepter and its subsequent release. Alternatively, one step catalysis occurs when a ternary complex is formed between the enzyme, donor and acceptor then the acetyl group is directly transferred from the donor to the acceptor with no acyl-enzyme intermediate.

The proposed catalytic mechanism of carbohydrate acetylation, in AT3-SGNH fused proteins by the SGNH domain, raises the question of how AT3-only acetyltransferases catalyse the same reaction. One possibility is that AT3-only proteins do function in the

same way as AT3-SGNH fused proteins but they function with an, as yet unidentified, un-fused partner protein. This mechanism would reflect that proposed for MBOAT proteins, involved in peptidoglycan, alginate and cellulose acetylation (**Figure 6.4**). Peptidoglycan acetylation by MBOAT and SGNH proteins PatA and PatB is introduced in Chapter 3 (Section 3.1). Alginate and cellulose are key exopolysaccharides involved in biofilm formation in *Pseudomonas aeruginosa* and *Pseudomonas fluorescens,* respectively, and they are proposed to have very similar pathways of acetylation (Riley *et al.*, 2013). In alginate acetylation, an MBOAT protein AlgI is proposed to transfer acetyl groups to the SGNH like AlgX and AlgJ proteins which subsequently transfer these acetyl groups to alginate (**Figure 6.4 A**). The respective equivalent proteins are termed WssH, WssF and WssI in cellulose acetylation (**Figure 6.4 B**) (Riley *et al.*, 2013).

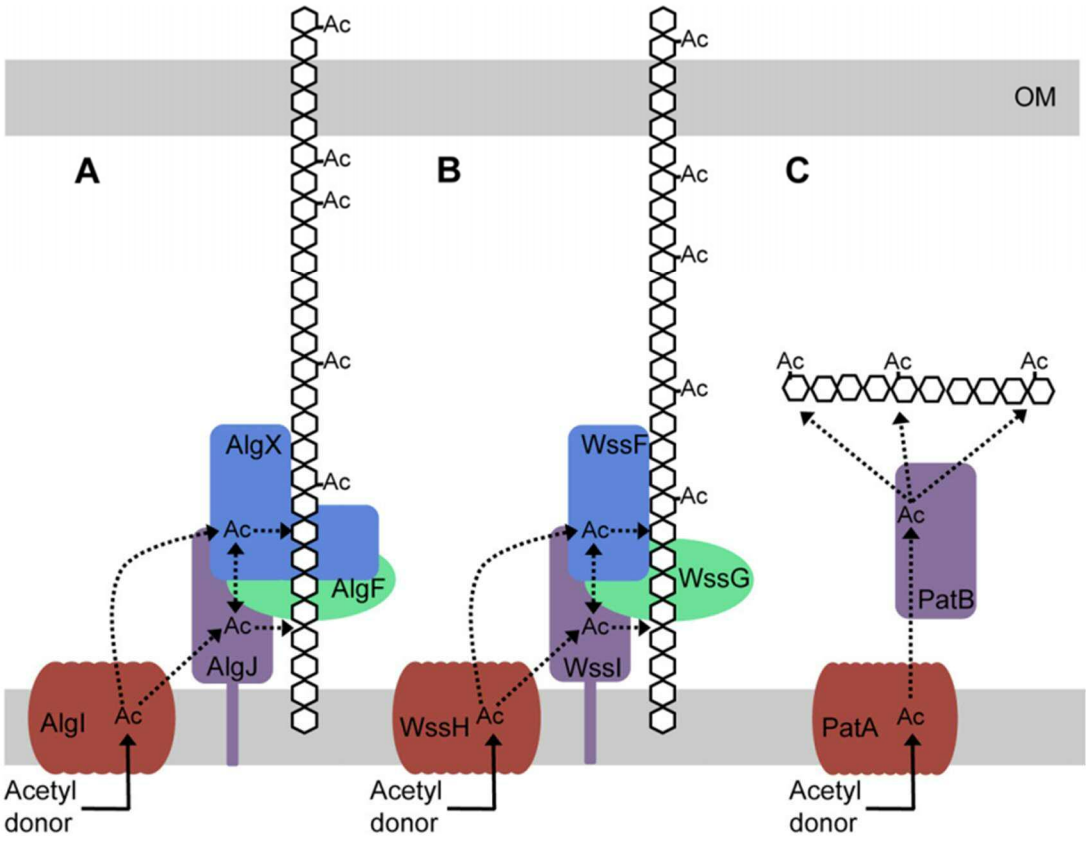


Figure 6.4| Proposed pathways for acetylation of alginate (A), cellulose (B) and peptidoglycan (C) involving MBOAT proteins (red) and various SGNH domain containing acetyltransferases (purple or blue). Figure taken from (Riley *et al.*, 2013)

If AT3-only acetyltransferases did function in a similar way to MBOAT proteins with SGNH partners, it would be highly likely that these partner proteins would be genetically linked with their cognate AT3 protein, unless they functioned with a generalist acetyltransferase protein present in the periplasm (discussed further below). Indeed,

the genes for alginate and peptidoglycan acetylation are located in close proximity in bacterial genomes (Franklin *et al.*, 2011; Sychantha, Brott, *et al.*, 2018) but as yet, no candidate SGNH partner proteins have been identified which are genetically linked with AT3-only acetyltransferases.

As discussed in Chapter 3, it may be difficult to identify these putative partner proteins through protein sequence homology as SGNH domains tend to have high structural homology but low sequence identity. Also, there are key examples of AT3-only proteins which are encoded on bacteriophage genomes with no linked SGNH partner, that are necessary to cause carbohydrate acetylation, such as *S. flexneri* Oac (Section 3.3). An additional complication is added by the fact that even the AlgX and AlgJ proteins, which are structurally similar to the SGNH domain of OafA and OafB are not characterised within the same InterPro family and have their own defined ALGX/ALGJ_SGNH-like family (IPR034655), therefore, it may not even be feasible to search bacterial genomes for proteins that contain similar protein signatures.

There is evidence that SGNH like proteins can play promiscuous roles within bacterial cells. For example, the TAP/TesA SGNH protein was identified to act as a thioesterase, esterase, arylesterase, protease and lysophospholipase (Lee *et al.*, 2006). Thus, a promiscuous extra-cytoplasmic SGNH-like protein characterised to perform an alternative function much also support AT3-only proteins in their acyltransferase activity. A potential approach to identify proteins which interact with AT3-only acetyltransferases is crosslinking. Chemical crosslinking followed by pull down of the target protein by immunoprecipitation then mass spectrometry analysis of the proteins which are chemically linked is a well-documented approach to identify interacting protein partners in living cells (Sinz, 2010). This approach may be able to identify periplasmic partners for AT3-only proteins and also inform whether these proteins interact with carbohydrate biosynthesis machinery.

An alternative to AT3-only proteins functioning with an unidentified partner is that the AT3 domain functions independently to catalyse both acyl group transport from the cytoplasm and transfer onto the acceptor substrate. If the AT3 domain performed this catalytic activity via a ping-pong mechanism, we would expect to see an invariant cysteine or serine residue towards the periplasmic side of the inner membrane as a site for the acyl-enzyme intermediate to be formed. *In silico* analysis did not identify

invariant cysteine residues within the AT3 domain of characterised bacterial acetyltransferases and the only invariant serine residues were buried within the transmembrane helices (**Figure 3.5**). This alignment did contain AT3-SGNH fused proteins as well as AT3-only but even if AT3-SGNH fused sequences were removed and the protein sequences re-aligned no candidate invariant catalytic residues on the periplasmic side were identified (data not shown).

Considering the published structure of an MBOAT protein (Ma *et al.*, 2018), it is possible that a residue that appears to be buried within the transmembrane domains may be exposed to the acceptor substrate. MBOAT's were originally predicted to contain 10 trans membrane spanning helices with a catalytic histidine residue in TMH8 (Resh, 2006; Chang and Magee, 2009) . Structural characterisation revealed that this histidine residue was actually accessible from the extracellular side via a funnel in the transmembrane domain (Ma *et al.*, 2018) . This provides an additional argument to pursue structural characterisation of the AT3 domain of OafA or OafB to determine whether the 10TMH prediction is correct and inform the orientation of the transmembrane helices within the membrane.

An alternative possibility is that the AT3 domain actually functions via one step catalysis rather than a ping-pong mechanism. Chloramphenicol O-acetyltransferase, which inactivates chloramphenicol to prevent its inhibition of protein synthesis in bacterial cells, uses one step catalysis to acetylate its target (Murray *et al.*, 1991; Day *et al.*, 1992). This soluble cytoplasmic protein forms a trimer and acetyl-CoA enters from one end of the trimer and chloramphenicol enters from the other. The acetyl group of CoA is positioned in close contact with the hydroxyl group of chloramphenicol by conserved tyrosine and phenylalanine residues and the catalytic histidine residue abstracts a proton from the hydroxyl group of chloramphenicol. This promotes nucleophilic attack of the thioester bond of acetyl-CoA which ultimately results in the transfer of the acetyl group from acetyl-CoA to the chloramphenicol hydroxyl group (Murray *et al.*, 1991; Day *et al.*, 1992).

There is potential for a similar mechanism in the AT3 domain, with the highly conserved and critical histidine residue towards the periplasmic side of the inner membrane (Chapter 4) that could interact with hydroxyl groups of the carbohydrate acceptor. Phenylalanine residues were identified as conserved in TMH2 of the AT3 domain

through alignments in this project (**Figure 3.5**), and phenylalanine and tyrosine's in TMH2 and TMH9 have been highlighted as potentially involved in the function of the AT3-only *S. flexneri* O-antigen acetyltransferase Oac (Thanweer and Verma, 2012). These residues could be involved in co-ordination of acetyl-CoA for one step catalysis.

The proposition that the AT3 domain is solely responsible for acetyl group transport from the cytoplasm and transfer onto the extra-cytoplasmic donor presents the question of how these integral membrane proteins with limited periplasmic loop regions interact with their carbohydrate acceptors to coordinate the acetylation modification. It also poses the question of the need for and role of the fused SGNH domain in AT3-SGNH fused proteins. Few conserved residues have been highlighted as critical in the periplasmic loops of AT3 domains, therefore, without further characterisation of AT3-only proteins it is not possible to conclusively determine the most feasible mechanism of acyl group transfer.

6.4 How does the nature of the acyl donor affect the need for a fused SGNH domain?

The research presented in this thesis specifically focused on proteins that perform acetylation modifications of carbohydrates. AT3 domain containing proteins have also been characterised to transfer diverse acyl substituents to a range of carbohydrate acceptors (Appendix I). Intriguingly, none of the proteins experimentally confirmed to transfer acyl groups other than acetate possess a fused SGNH domain. This has prompted the question of whether the presence of an SGNH domain is a specific adaptation towards transfer of acetyl groups rather than alternative acyl substrates.

During characterisation of acyl substrates, it has been observed that longer or branched acyl chains are less susceptible to spontaneous hydrolysis by water molecules (Miceli *et al.*, 2019). It has also been observed that the rate of prodrug degradation has a direct relationship with acyl side chain length (Jarvinen *et al.*, 2005). These observations are suggested to be due to short linear acyl chains such as acetate leaving the ester group exposed to water, whereas branched or longer acyl chains create steric hindrance of water from the thioester bond. Thus, the SGNH domain might provide increased protection of the acetyl group from spontaneous hydrolysis of the thioester bond of acetyl-CoA to prevent the futile loss of this valuable metabolite.

This observation also fits with the hypothesis that the AT3 domain, as well as the periplasmic domain, in AT3-SGNH fused proteins are involved in acceptor substrate specificity. These proteins may have initially evolved as AT3-only proteins which interacted with their carbohydrate acceptor through the periplasmic loops of the AT3 domain. Then fusion with an SGNH domain allowed for increased efficiency of acetyl group transfer and decreased loss of acetyl-CoA. This metabolically favourable system could then have been selected for through bacterial evolution. A hypothesis supported by the observation that SGNH fused AT3 domains appear to be evolutionarily more closely related, regardless of the acceptor substrate that they acetylate (**Figure 3.4**).

An alternative mechanism to prevent futile loss of acyl donors by spontaneous hydrolysis could be that a regulatory mechanism within the AT3 domain requires the acceptor carbohydrate to be bound before the acyl group can dock into the AT3 domain and be presented at the periplasmic side. A mechanism similar to this is proposed for MBOAT proteins, where a small horizontal helix contains a conserved tryptophan residue which closes the tunnel between the cytoplasmic side and the catalytic histidine residue and is proposed to act as a gate while the MBOAT is in not bound with its acceptor (Ma *et al.*, 2018). It would be very difficult to identify a similar channel gating domain in AT3 proteins without a crystal structure for this domain, therefore, this suggestion cannot yet be supported or discounted.

6.5 How do AT3-SGNH proteins determine specificity of their modification?

Prior to this research, it was unclear whether the AT3 and SGNH domains of fused acyltransferases function as a single unit in close proximity, for direct transport and transfer of the acetyl group to the acceptor sugar, or whether the two domains function independently, connected by a flexible linker, to transport the acetyl group to the periplasm then subsequently transfer it to the carbohydrate acceptor. This research has shown that the periplasmic portion of the linking region, termed the SGNH extension (Ext^{SGNH}) is structured and likely to orient the SGNH domain in close proximity to the AT3 domain (Chapter 5). We have also shown that this domain can influence the acceptor substrate specificity of the SGNH domain *in vitro* (Chapter 5).

If the Ext^{SGNH} is responsible for acceptor substrate specificity in AT3-SGNH fused proteins, it raises the question of how AT3-only proteins prevent promiscuous acetylation of nonspecific targets. It is clear that AT3-only proteins are indeed specific for the monosaccharide that they acetylate, as evidenced by the GumF and GumG proteins of *Xanthomonas campestris* which are specific to the proximal and distal mannose residues of the Man-GlcA-Man side chains of the glucose backbone of Xylose (Katzen *et al.*, 1998). Therefore, unless these proteins function with a periplasmic partner that provides the specificity of the modification, the periplasmic loop regions of the AT3 domain must play some role in acceptor substrate interaction and specificity. The observation that few conserved residues are found in the periplasmic loop regions of AT3 proteins could also support this hypothesis. Theses loops may be specifically adapted to their carbohydrate acceptor, therefore, conservation of residues interacting with diverse acyl group acceptors is unlikely.

It may be that the Ext^{SGNH} does add an extra degree of substrate specificity in AT3-SGNH fused proteins but *in situ* both the AT3 and periplasmic domains are involved in the specific acceptor substrate interaction. Preliminary experiments to determine whether the AT3 domains of OafA and OafB could be swapped to create a chimeric protein with acceptor substrate specificity determined by the periplasmic Ext^{SGNH} and SGNH domain have not yet yielded a functional protein, even though full length chimeric protein expression was demonstrated by His-tag Western blot (Rebekah Eastwood, University of York, data not shown). This observation supports the idea that the periplasmic loops of the AT3 domain coordinate specific interaction with the Ext^{SGNH} and SGNH domain, which together maintain acceptor substrate specificity. It may be possible to produce a functional chimeric protein by swapping the corresponding periplasmic loop regions. Conducting these swaps in a systematic way, until a functional chimeric protein is achieved, could also help to identify the region of the AT3 domain which infers the acceptor substrate specificity. Building from this, it may then be possible to determine whether the same periplasmic loop regions of AT3-only proteins are responsible for acceptor substrate specificity. The caveat currently being that a simple functional assay for GumF-GumG specificity or an alternative model system is not yet defined.

6.6 Conclusions

AT3 domain-containing bacterial proteins are involved in the modification of a wide range of polysaccharides and influence many host-pathogen interactions. The structural and functional insights gained from this research could be applied to biotechnologically relevant AT3 proteins, including Nod factor modifying proteins, important for plant microbe symbiosis (Davis *et al.*, 1988; Firmin *et al.*, 1993b), and anti-tumour and antibiotic modifying proteins (Arisawa *et al.*, 1994; Cong and Piepersberg, 2007). As well as acquiring broadly applicable mechanistic understanding of AT3 domain-containing acetyltransferases through the investigation of OafA and OafB, gaining further understanding of surface antigen modifications of *Salmonella* serovars is of high importance, particularly in the focus of disease control, and it could help to inform vaccination strategies and be used to create specifically acetylated *Salmonella* vaccines. Indeed, the advantage of acetylated carbohydrates in vaccine development has already been suggested (Ravenscroft *et al.*, 2015; Berti *et al.*, 2018; Hitri *et al.*, 2019).

Furthermore, this work could be expanded in future to inform studies in eukaryotic systems. There is evidence that AT3 domain-containing proteins are involved in regulation of the lifespan of *Caenorhabditis elegans* (Vora *et al.*, 2013) and in *Drosophila* development (Dzitoyeva *et al.*, 2003). Plants in the *Arabidopsis* genus also possesses a 10TMH spanning protein proposed to transport acetyl-coA to the Golgi for transfer onto plant cell wall polysaccharides by an SGNH like protein with conserved catalytic residues (Schultink *et al.*, 2015).

It is clear that many questions remain about the intricacies of O-acetylation of carbohydrates by AT3 domain-containing acetyltransferases. With strong links to industrial, clinical and pathogenic processes, further characterisation of this diverse protein family is of high significance

Appendix I Experimentally confirmed AT3 domain-containing O-acetyltransferases

Protein Name	Protein Length (aa)	SGNH Fused?	Species	Gram Stain	Carbohydrate Acceptor	Acyl substrate	Modified sugar	Accession Number	First Reference
OafA_STM	609	Y	*Salmonella* Typhimurium	Neg	LPS O-antigen	Acetyl-	C2 Abequose	WP_000639473	(Slauch *et al.*, 1996)
OafB_STM (F2GtrC)	640	Y	*Salmonella* Typhimurium D23580	Neg	LPS O-antigen	Acetyl-	C2/3 Rhamnose	SIU02679	(Kintz *et al.*, 2015)
OafA_Hi	622	Y	*Haemophilus influenzae*	Neg	LPS inner-core	Acetyl-	C2/3 Heptose	AAX87447	(Fox *et al.*, 2005)
OatA_Lm (LMO129)	622	Y	*Listeria monocytogenes*	Pos	Peptidoglycan	Acetyl-	MurNAc	CAC99369	(Aubry *et al.*, 2011)
Lot3 (PurB)	622	Y	*Neisseria meningitidis*	Neg	LPS Core Oligosaccharides	Acetyl-	GlcNAc	WP_002245844	(Kahler *et al.*, 2006)
OatA_Lp (lp_0856)	660	Y	*Lactobacillus plantarum WCFS1*	Pos	Peptidoglycan	Acetyl-	MurNAc	WP_011101182	(Bernard *et al.*, 2011)
OatB_Lp (lp_0925)	615	Y	*Lactobacillus plantarum WCFS1*	Pos	Peptidoglycan	Acetyl-	GlcNAc	YP_004888877	(Bernard *et al.*, 2011)
OatA_Sa	603	Y	*Staphylococcus aureus*	Pos	Peptidoglycan	Acetyl-	MurNAc C6	WP_000379821	(Bera *et al.*, 2005)
OatA_Ll (yvhB)	605	Y	*Lactococcus lactis*	Pos	Peptidoglycan	Acetyl-	MurNAc C6	WP_021723064	(Veiga *et al.*, 2007)

OatA_Sp (Adr)	605	Y	*Streptococcus pneumoniae*	Pos	Peptidoglycan	Acetyl-	MurNAc	WP_001220853	(Crisóstomo *et al.*, 2006)
PglI	624	Y	*Neisseria gonorrhoeae*	Neg	Pillin Glycan	Acetyl-	DATDH	WP_003687310.1	(Aas *et al.*, 2007; Anonsen *et al.*, 2017)
OatB_Ba (BAS5308)	343	N	*Bacillus anthracis* str. Sterne	Pos	Peptidoglycan	Acetyl-	MurNac	YP_031545	(Laaberki *et al.*, 2011)
ExoZ	318	N	*Rhizobium meliloti*	Neg	Succinoglycan (EPS I)	Acetyl-	C6 Glucose VII	WP_010975904	(Buendia *et al.*, 1991)
GumF	364	N	*Xanthomonas campestris*	Neg	Xanthan	Acetyl-	Proximal Mannose	WP_011037591	(Katzen *et al.*, 1998)
GumG	379	N	*Xanthomonas campestris*	Neg	Xanthan	Acetyl-	Distal Mannose	AAA86375	(Katzen *et al.*, 1998)
Lag1	357	N	*Legionella pneumophila*	Neg	LPS O-antigen	Acetyl-	Legionaminic acid C6	AAA75102	(Zou *et al.*, 1999)
NodX	367	N	*Rhizobium leguminosarum bv.viciae strain TOM Plasmid sym pRL5J*	Neg	Nod factor NodRIv-V	Acetyl-	C6 reducing GlcNAc	P08888	(Davis *et al.*, 1988; Firmin *et al.*, 1993a)
NolL	373	N	*Rhizobium leguminosarum bv.viciae*	Neg	Fucose of lipo-chitin oligosaccharide	Acetyl-	GlcNAc V	Q52778	(Pacios Bras *et al.*, 2000)
OacA_Sf	333	N	*Shigella phage SF6*	Neg	LPS O-antigen	Acetyl-	Rhamnose C2	P23214	(Verma *et al.*, 1991)

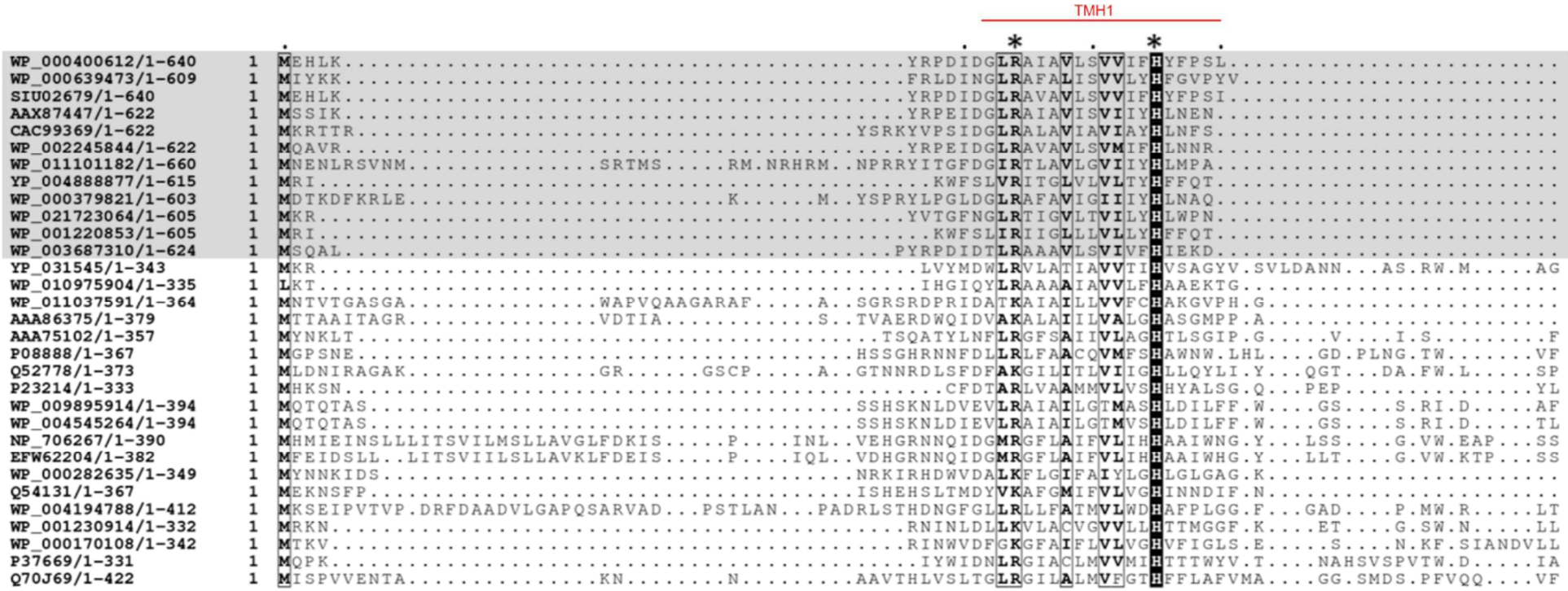
OacA _Bt	394	N	*Burkholderia thailandiensis*	Neg	LPS O-antigen	Acetyl-	6-deoxy-α-l-talopyranose C4	WP_009895914	(Brett *et al.*, 2011)
OacA_Bp	394	N	*Burkholderia pseudomallei (strain K96243)*	Neg	LPS O-antigen	Acetyl-	6-deoxy-α-l-talopyranose C4	WP_004545264	(Brett *et al.*, 2011)
OacB_Sf	390	N	*Shigella flexneri str. Sf301*	Neg	LPS O-antigen	Acetyl-	Rhamnose C3/4	NP_706267	(Wang *et al.*, 2014)
OacC_Sf	382	N	*Shigella flexneri*	Neg	LPS O-antigen	Acetyl-	Rhamnose C3/4	EFW62204	(Knirel *et al.*, 2014)
OacD_Sf	349	N	*Bacteriophage SfII Shigella flexneri*	Neg	LPS O-antigen	Acetyl-	GlcNAc C6	WP_000282635	(Sun *et al.*, 2014)
WbaK	367	N	*Salmonella enterica ser.Anatum*	Neg	LPS O-antigen	Acetyl-	Galactose C6	Q54131 (uniprot entry for Salmonella choleraesuis)	(Hong *et al.*, 2013)
WbiA	412	N	*Burkholderia pseudomallei (strain 1026b)*	Neg	LPS O-antigen	Acetyl-	6-deoxy-α-l-talopyranose C2	WP_004194788	(Brett *et al.*, 2003)
WciG	332	N	*Streptococcus pneumoniae*	Pos	Capsular Polysaccharide	Acetyl-	-	WP_001230914	(Geno, Saad, *et al.*, 2017)
WcjE	342	N	*Streptococcus pneumoniae*	Pos	Capsular polysaccharide	Acetyl-	1-phosphoglycerol	WP_000170108	(Calix and Nahm, 2010)

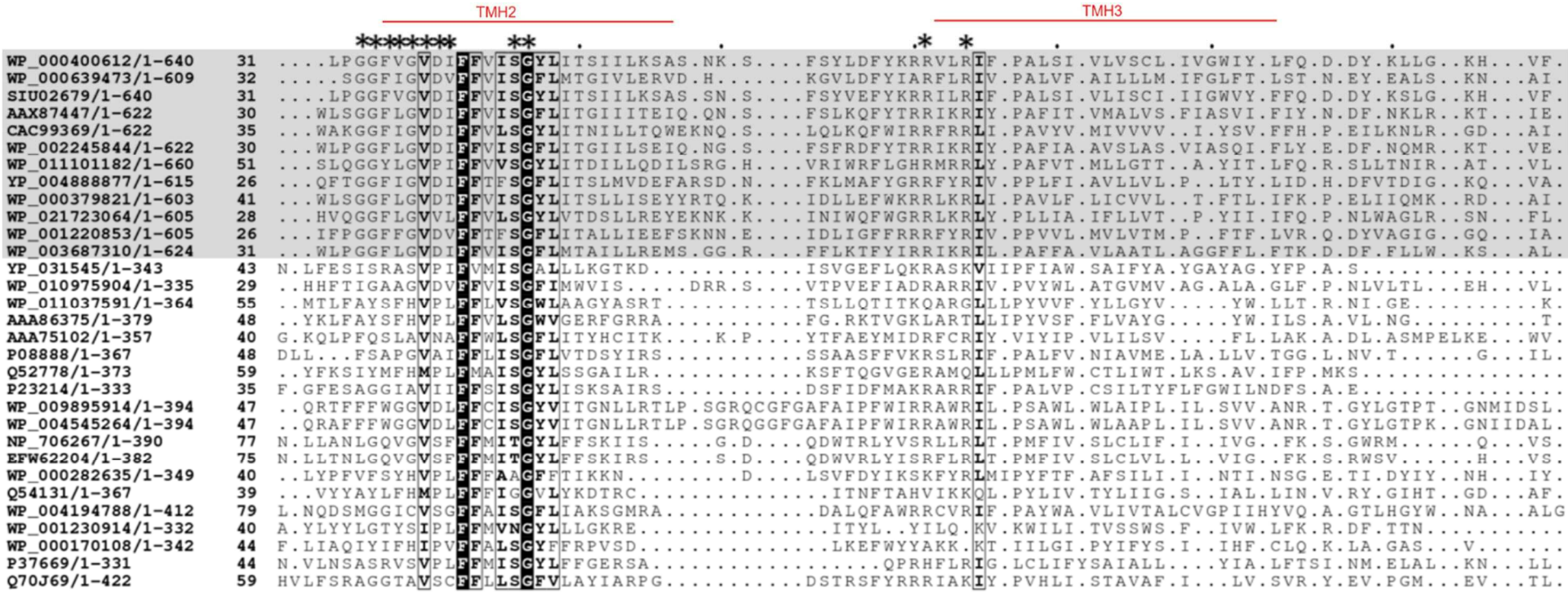
WecH (YiaH)	331	N	*Escherichia coli K-12*	Neg	Enterobacterial Common Antigen	Acetyl-	GlcNAc C6	P37669	(Kajimura *et al.*, 2006)
CmmA	422	N	*Streptomyces griseus subsp. griseus*	Pos	Chromomycin A3(antitumour drug)	Acetyl- (Isobutyryl- Propionyl-)	Oliose and Chromose B	Q70J69	(Menéndez *et al.*, 2004a)
MdmB	387	N	*Streptomyces mycarofaciens*	Pos	Macrolide antibiotics	Acetyl- Propionyl-	Lactone ring C3 in 16-member macrolide antibiotics like midecamycin and spiramycin	Q00718	(Hara and Hutchinson, 1992)
ExoH	370	N	*Sinorhizobium meliloti 1021*	Neg	Succinoglycan (EPS I)	Succinyl-	C6 Glucose III	NP_437619	(Leigh *et al.*, 1987)
CarE	388	N	*Streptomyces thermotolerans*	Pos	Carbomycin (Macrolide antibiotics)	Isovaleryl-	C4 Mycarose	P21542	(Epp *et al.*, 1989; Arisawa *et al.*, 1993)
AcyA	389	N	*Streptomyces thermotolerans*	Pos	Macrolide antibiotics	Acetyl- Propionyl-	Lactone ring C3 in 16-member macrolide antibiotics	BAA06421	(Arisawa *et al.*, 1994)
Mpt	388	N	*Streptomyces mycarofaciens*	Pos	Macrolide antibiotics	Propionyl-	C4 Spiramycin	BAA09815	(Gu *et al.*, 1996; Zhang and Wang, 1996)

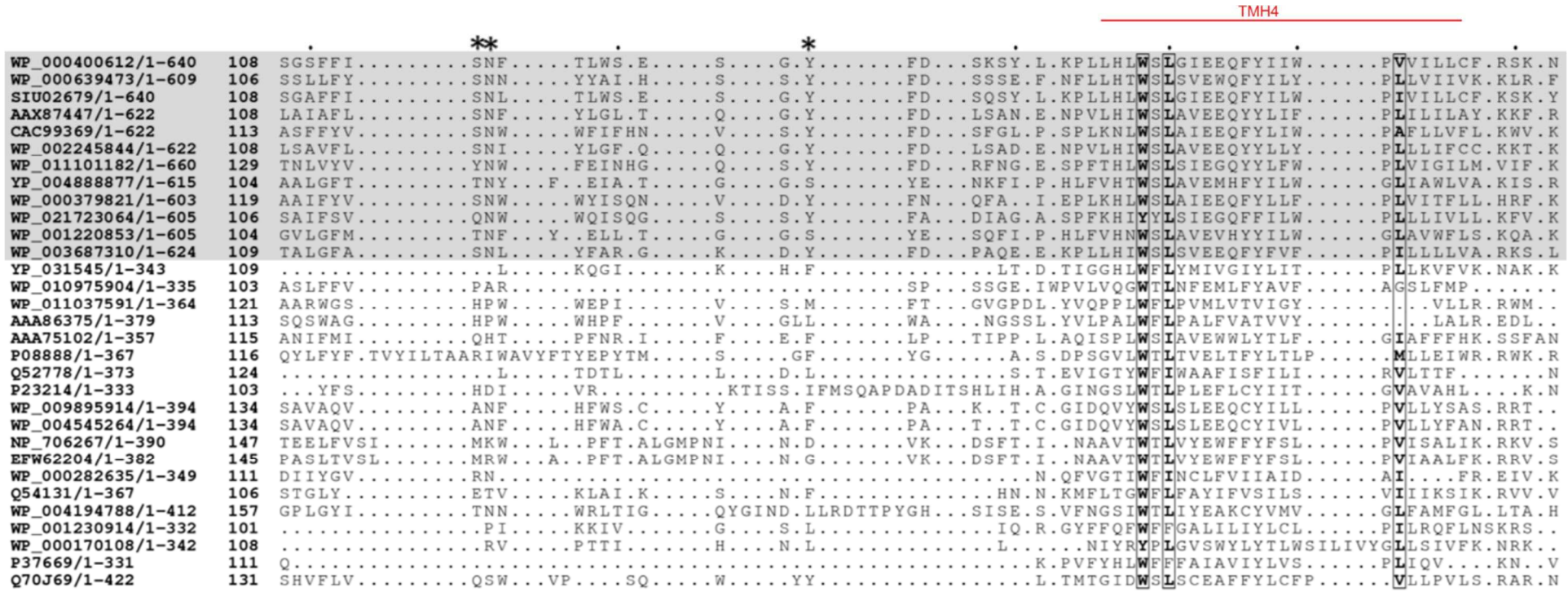
MidE	389	N	*Streptomyces mycarofaciens*	Pos	Midecamycin (Macrolide antibiotic)	Propionyl-	C4 Midemycin	ABG67709	(Cong and Piepersberg, 2007)
Asm19	378	N	*Actinosynnema pretiosum*	Pos	Ansamitocin precursor (antitumour drug)	acetyl-propionyl-isobutyryl-butyryl-isoveleryl	C3 N-desmethyl-4,5-desepoxymaytansinol	Q8KUG7	(Moss *et al.*, 2002)
MdoC	385	N	*Escherichia coli K-12*	Neg	Periplasmic glucans	Succinyl-	Glucose	WP_001070375	(Lacroix *et al.*, 1999)
OpgC (RS)	399	N	*Rhodobacter sphaeroides*	Neg	Periplasmic glucans	Succinyl-	Glucose	Q3J1N1	(Cogez *et al.*, 2002)
OpgC (BA)	393	N	*Brucella abortus*	Neg	Periplasmic glucans	Succinyl-	Glucose	Q6Q7W5	(Roset *et al.*, 2006)
Opgc (DD)	422	N	*Dickeya dadantii*	Neg	Periplasmic glucans	Succinyl-	Glucose	E0SF04	(Bontemps-Gallo *et al.*, 2016)

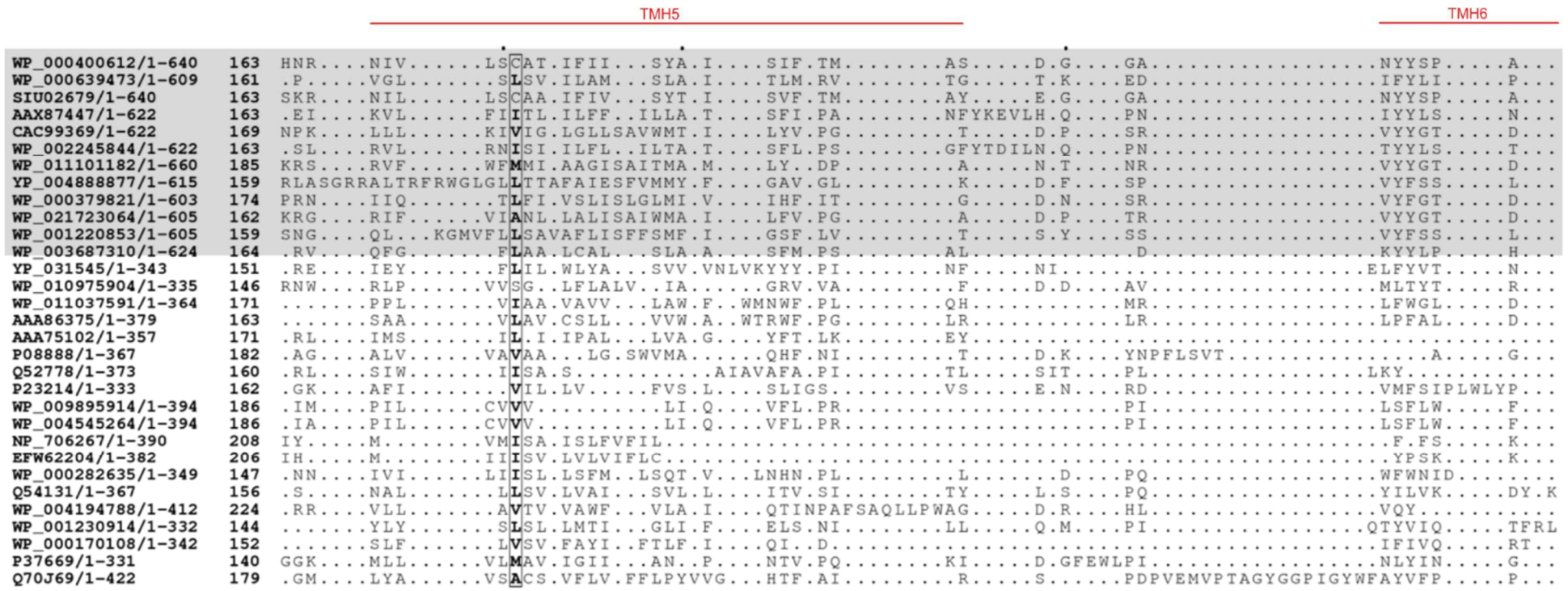
Appendix II Alignment of characterised AT3 domain-containing O-acetyltransferases

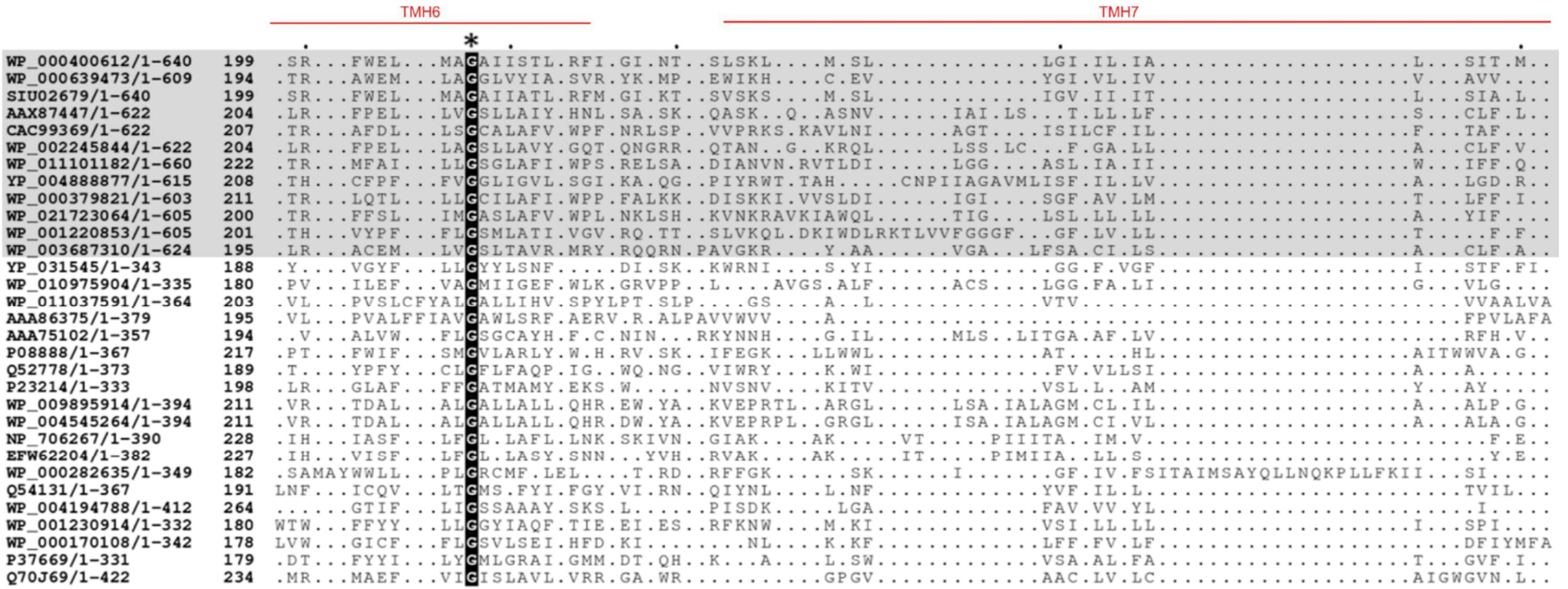
Alignment of characterised AT3 acetyltransferases. Protein sequences are in the same order as **Appendix I** after *Salmonella* ser. Paratyphi A OafB WP_00400612. SGNH fused acetyltransferases are indicated by a grey box. Asterisk marks residues selected for mutation from this alignment. Transmembrane helices predicted for OafA are indicated with a red line.

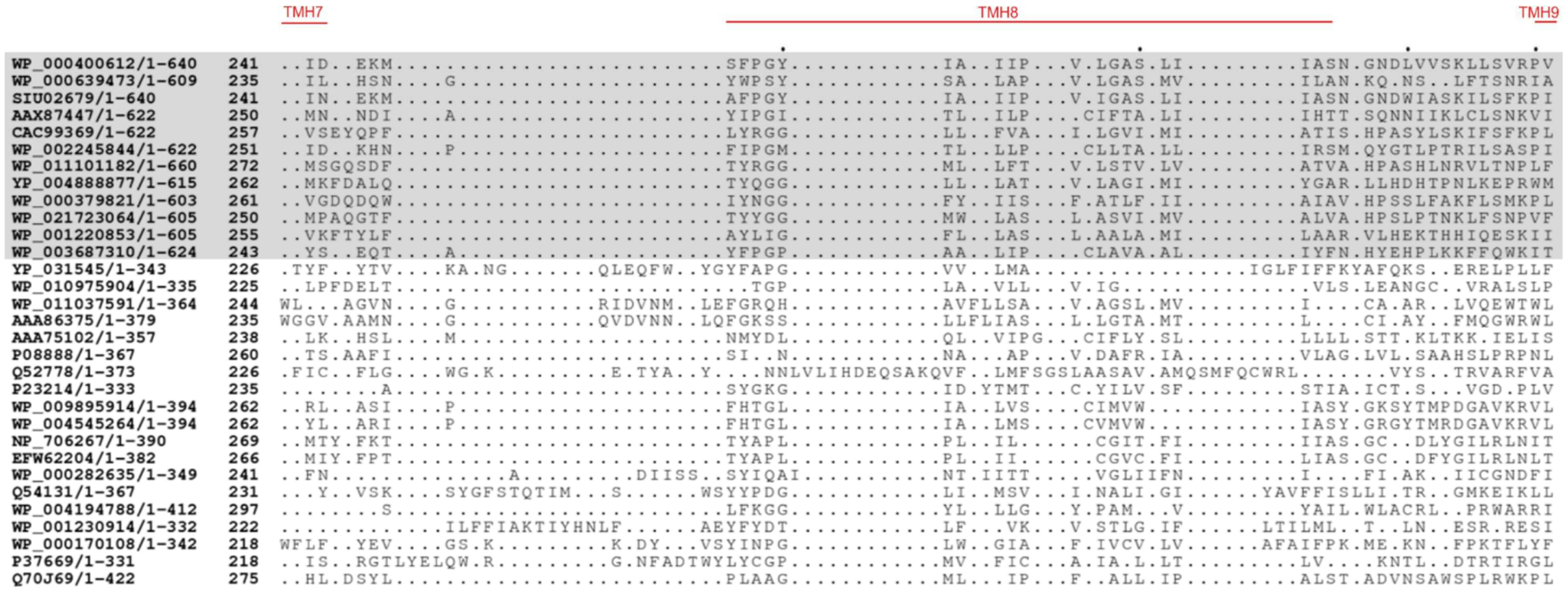


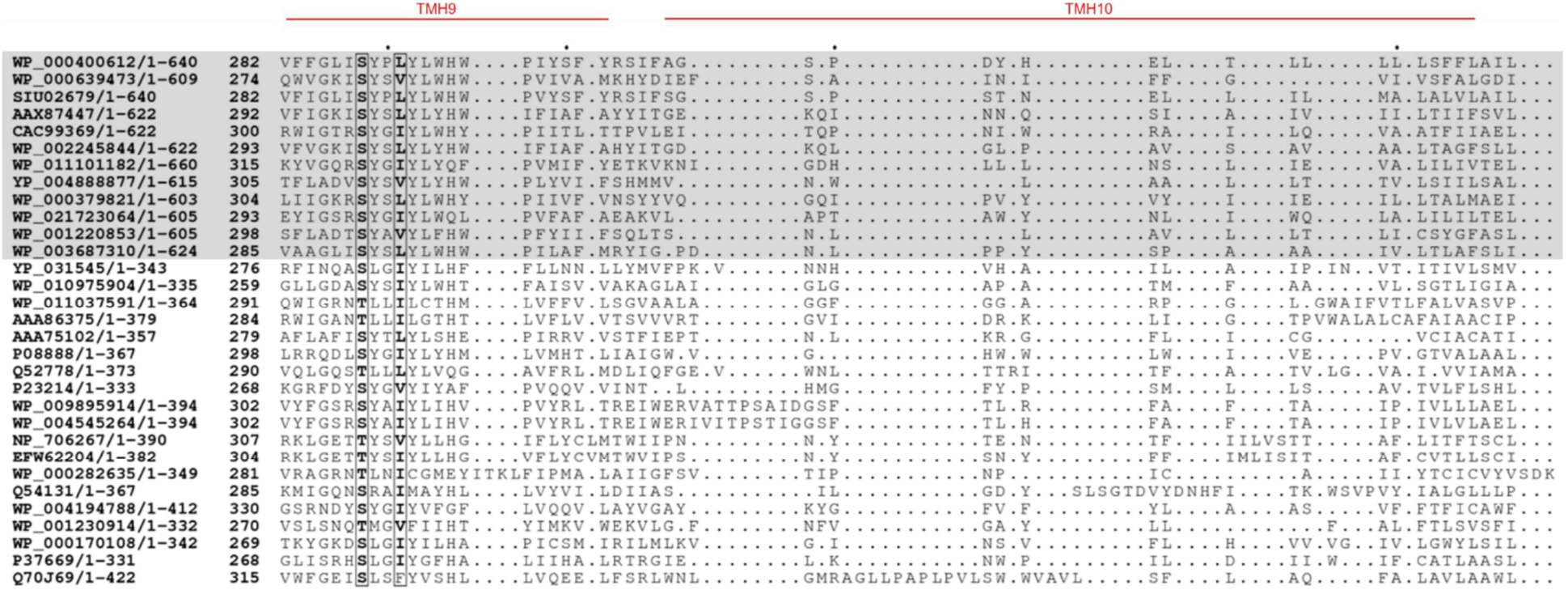

TMH2
TMH3
WP_000400612/1-640 31 ....LPGGFVGVDIFFVISGYLITSIILKSAS.NK.S....FSYLDFYKRRVLRIF.PALSI.VLVSCL.IVGWIY.LFQ.D.DY.KLLG..KH...VF.
WP_000639473/1-609 32 .....SGGFIGVDVFFVISGFLMTGIVLERVD.H.......KGVLDFYIARFLRIV.PALVF.AILLLM.IFGLFT.LST.N.EY.EALS..KN...AI.
SIU02679/1-640 31 ....LPGGFVGVDIFFVISGYLITSIILKSAS.SN.S....FSYVEFYKRRILRIF.PALSI.VLISCI.IIGWVY.FFQ.D.DY.KSLG..KH...VF.
AAX87447/1-622 30 ...WLSGGFLGVDIFFVISGFLITGIIITEIQ.QN.S....FSLKQFYTRRIKRIY.PAFIT.VMALVS.FIASVI.FIY.N.DF.NKLR..KT...IE.
CAC99369/1-622 35 ...WAKGGFIGVDIFFVLSGYLITNILLTQWEKNQ.S....LQLKQFWIRRFRRLI.PAVYV.MIVVVV..I.YSV.FFH.P.EILKNLR..GD...AI.
WP_002245844/1-622 30 ...WLPGGFLGVDIFFVISGFLITGIILSEIQ.NG.S....FSFRDFYTRRIKRIY.PAFIA.AVSLAS.VIASQI.FLY.E.DF.NQMR..KT...VE.
WP_011101182/1-660 51 ...SLQGGYLGVPIFFVVSGYLITDILLQDILSRG.H....VRIWRFLGHRMRRLY.PAFVT.MLLGTT..A.YIT.LFQ.R.SLLTNIR..AT...VL.
YP_004888877/1-615 26 ...QFTGGFIGVDIFFTFSGFLITSLMVDEFARSD.N....FKLMAFYGRRFYRIV.PPLFI.AVLLVL.P..LTY.LID.H.DFVTDIG..KQ...VA.
WP_000379821/1-603 41 ...WLSGGFLGVDTFFVISGYLITSLLISEYYRTQ.K....IDLLEFWKRRLKRLI.PAVLF.LICVVL..T.FTL.IFK.P.ELIIQMK..RD...AI.
WP_021723064/1-605 28 ...HVQGGFLGVVLFFVLSGYLVTDSLLREYEKNK.K....INIWQFWGRRLKRLY.PLLIA.IFLLVT..P.YII.IFQ.P.NLWAGLR..SN...FL.
WP_001220853/1-605 26 ...IFPGGFFGVDVFFTFSGFLITALLIEEFSKNN.E....IDLIGFFRRRFYRIV.PPVVL.MVLVTM.P..FTF.LVR.Q.DYVAGIG..GQ...IA.
WP_003687310/1-624 31 ...WLPGGFLGVDIFFVISGFLMTAILLREMS.GG.R....FFLKTFYIRRIKRIL.PAFFA.VLAATL.AGGFFL.FTK.D.DF.FLLW..KS...AL.
YP_031545/1-343 43 N.LFESISRASVPIFVMISGALLLKGTKD.....ISVGEFLQKRASKVIIPFIAW.SAIFYA.YGAYAG.YFP.A.S.....
WP_010975904/1-335 29 ..HHFTIGAAGVDVFFVISGFIMWVIS.....DRR.S....VTPVEFIADRARRIV.PVYWL.ATGVMV.AG.ALA.GLF.P.NLVLTL...EH...VL.
WP_011037591/1-364 55 ..MTLFAYSFHVPLFFLVSGWLAAGYASRT.....TSLLQTITKQARGLLLPYVVF.YLLGYV.....YW.LLT.R.NI.GE.....K.
AAA86375/1-379 48 ..YKLFAYSFHVPLFFVLSGWVGERFGRRA.....FG.RKTVGKLARTLLIPYVSF.FLVAYG.....YW.ILS.A.VL.NG.....T.
AAA75102/1-357 40 G.KQLPFQSLAVNAFFWLSGFLITYHCITK....K.P....YTFAEYMIDRFCRIY.VIYIP.VLILSV.....FL.LAK.A.DL.ASMPELKE...WV.
P08888/1-367 48 DLL...FSAPGVAIFFLISGFLVTDSYIRS.....SSAASFFVKRSLRIF.PALFV.NIAVME.LA.LLV.TGG.L.NV.T.....G...IL.
Q52778/1-373 59 ..YFKSIYMFHMPLFMAISGYLSSGAILR.....KSFTQGVGERAMQLLLPMLFW.CTLIWT.LKS.AV.IFP.MKS.....
P23214/1-333 35 F.GFESAGGIAVIIFFSISGYLISKSAIRS.....DSFIDFMAKRARRIF.PALVP.CSILTYFLFGWILNDFS.A.E.....
WP_009895914/1-394 47 ..QRTFFFWGGVDLFFCISGYVITGNLLRTLP.SGRQCGFGAFAIPFWIRRAWRIL.PSAWL.WLAIPL.IL.SVV.ANR.T.GYLGTPT..GNMIDSL.
WP_004545264/1-394 47 ..QRAFFFWGGVDLFFCISGYVITGNLLRTLP.SGRQGGFGAFAIPFWIRRAWRIL.PSAWL.WLAAPL.IL.SVV.ANR.T.GYLGTPK..GNIIDAL.
NP_706267/1-390 77 N.LLANLGQVGVSFFFMITGYLFFSKIIS.....G.D....QDWTRLYVSRLLRLT.PMFIV.SLCLIF.I..IVG..FK.S.GWRM.....Q...VS.
EFW62204/1-382 75 N.LLTNLGQVGVSFFFMITGYLFFSKIRS.....S.D....QDWVRLYISRFLRLT.PMFIV.SLCLVL.L..VIG..FK.S.RWSV.....H...VS.
WP_000282635/1-349 40 ..LYPFVFSYHVPLFFFAAGFFTIKKN.....D....LSVFDYIKSKFYRLMIPYFTF.AFSILI.I..NTI.NSG.E.TI.DYIY..NH...IY.
Q54131/1-367 39 ...VYYAYLFHMPLFFFIGGVLYKDTRC.....ITNFTAHVIKKQL.PYLIV.TYLIIG.S..IAL.LIN.V.RY.GIHT..GD...AF.
WP_004194788/1-412 79 L.NQDSMGGICVSGFFAISGFLIAKSGMRA.....DALQFAWRRCVRIF.PAYWA.VLIVTALCVGPIIHYVQ.A.GTLHGYW..NA...ALG
WP_001230914/1-332 40 A.YLYYLGTYSIPLFFMVNGYLLLGKRE.....ITYL..YILQ.KV.KWILI.TVSSWS.F..IVW.LFK.R.DF.TTN.....
WP_000170108/1-342 44 F.LIAQIYIFHIPVFFALSGYFFRPVSD.....LKEFWYYAKK.KT.IILGI.PYIFYS.I..IHF.CLQ.K.LA.GAS...V.....
P37669/1-331 44 N.VLNSASRVSVPLFFMISGYLFFGERSA.....QPRHFLRIG.LCLIFYSAIALL....YIA.LFTSI.NM.ELAL..KN...LL.
Q70J69/1-422 59 HVLFSRAGGTAVSCFFLLSGFVLAYIARPG.....DSTRSFYRRRIAKIY.PVHLI.STAVAF.I...LV.SVR.Y.EV.PGM...EV...TL.

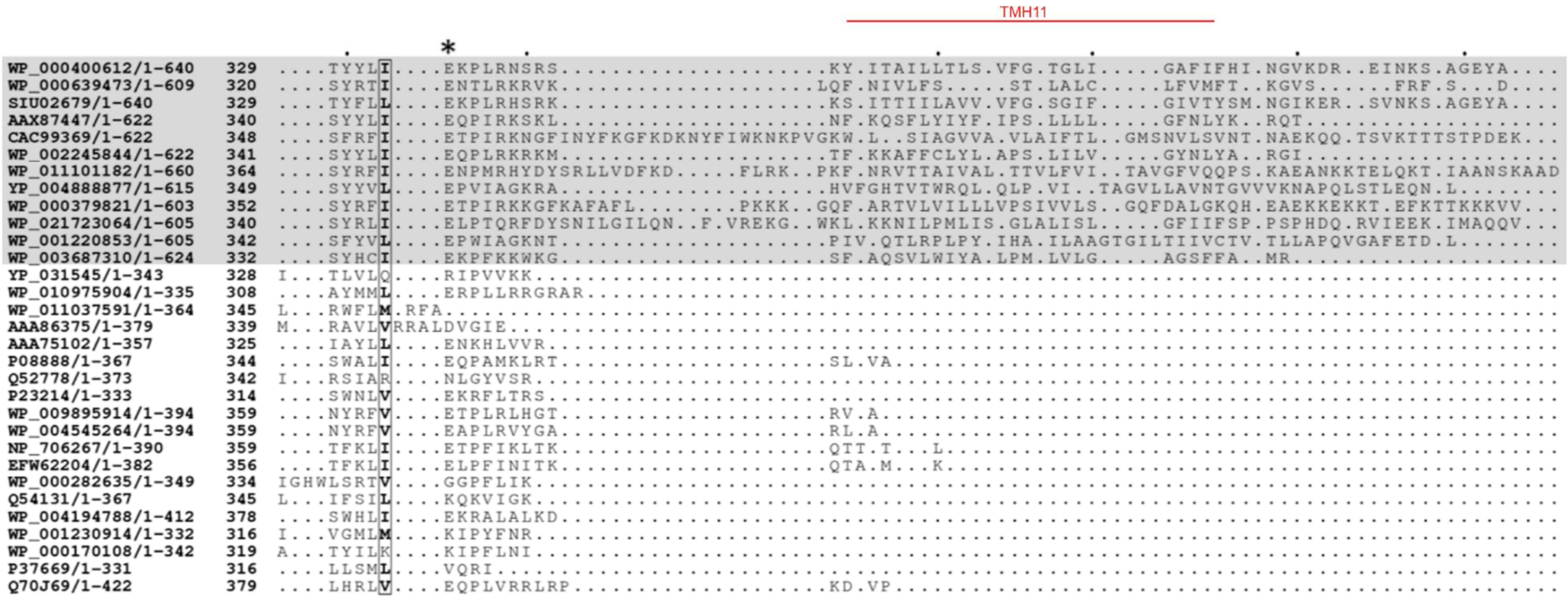

TMH4
WP_000400612/1-640 108 SGSFFI.....SNF.....TLWS.E.....S....G.Y.....FD...SKSY.L.KPLLHLWSLGIEEQFYIIW.....PVVILLCF.RSK.N
WP_000639473/1-609 106 SSLLFY.....SNN.....YYAI.H.....S....S.Y.....FD...SSSE.F.NFLLHTWSLSVEWQFYILY.....PLLVIIVK.KLR.F
SIU02679/1-640 108 SGAFFI.....SNL.....TLWS.E.....S....G.Y.....FD...SQSY.L.KPLLHLWSLGIEEQFYILW.....PIVILLCF.KSK.Y
AAX87447/1-622 108 LAIAFL.....SNF.....YLGL.T.....Q....G.Y.....FD...LSAN.E.NPVLHIWSLAVEEQYYLIF.....PLILILAY.KKF.R
CAC99369/1-622 113 ASFFYV.....SNW.....WFIFHN.....V....S.Y.....FD...SFGL.P.SPLKNLWSLAIEEQFYLIW.....PAFLLVFL.KWV.K
WP_002245844/1-622 108 LSAVFL.....SNI.....YLGF.Q.....Q....G.Y.....FD...LSAD.E.NPVLHIWSLAVEEQYYLLY.....PLLLIFCC.KKT.K
WP_011101182/1-660 129 TNLVYV.....YNW.....FEINHG.....Q....S.Y.....FD...RFNG.E.SPFTHLWSLSIEGQYYLFW.....PLVIGILM.VIF.K
YP_004888877/1-615 104 AALGFT.....TNY...F..EIA.T.....G....G.S.....YE...NKFI.P.HLFVHTWSLAVEMHFYILW.....GLIAWLVA.KIS.R
WP_000379821/1-603 119 AAIFYV.....SNW.....WYISQN.....V....D.Y.....FN...QFA..I.EPLKHLWSLAIEEQFYLLF.....PLVITFLL.HRF.K
WP_021723064/1-605 106 SAIFSV.....QNW.....WQISQG.....S....S.Y.....FA...DIAG.A.SPFKHIYYLSIEGQFFILW.....PLLLIVLL.KFV.K
WP_001220853/1-605 104 GVLGFM.....TNF...Y..ELL.T.....G....G.S.....YE...SQFI.P.HLFVHNWSLAVEVHYYILW.....GLAVWFLS.KQA.K
WP_003687310/1-624 109 TALGFA.....SNL.....YFAR.G.....K....D.Y.....FD...PAQE.E.KPLLHIWSLSVEEQFYFVF.....PILLLLVA.RKS.L
YP_031545/1-343 109 .....L.....KQGI.....K....H.F.....LT.D.TIGGHLWFLYMIVGIYLIT.....PLLKVFVK.NAK.K
WP_010975904/1-335 103 ASLFFV.....PAR.....SP...SSGE.IWPVLVQGWTLNFEMLFYAVF.....AGSLFMP.....
WP_011037591/1-364 121 AARWGS.....HPW.....WEPI.....V....S.M.....FT...GVGPDL.YVQPPLWFLPVMLVTVIGY.....VLLR.RWM..
AAA86375/1-379 113 SQSWAG.....HPW.....WHPF.....V....GLL.....WA....NGSSL.YVLPALWFLPALFVATVVY.....LALR.EDL..
AAA75102/1-357 115 ANIFMI.....QHT.....PFNR.I.....F....E.F.....LP...TIPP.L.AQISPLWSIAVEWWLYTLF.....GIAFFFHK.SSFAN
P08888/1-367 116 QYLFYF.TVYILTAARIWAVYFTYEPYTM.....S.....GF.....YG.....A.S.DPSGVLWTLTVELTFYLTLP.....MLLEIWR.RWK.R
Q52778/1-373 124 .....L.....TDTL.....L....D.L.....S.T.EVIGTYWFIWAAFISFILI.....RVLTTF.....N
P23214/1-333 103 ...YFS.....HDI.....VR.....KTISS.IFMSQAPDADITSHLIH.A.GINGSLWTLPLEFLCYIIT.....GVAVAHL....K.N
WP_009895914/1-394 134 SAVAQV.....ANF.....HFWS.C.....Y....A.F.....PA...K..T.C.GIDQVYWSLSLEEQCYILL.....PVLLYSAS.RRT..
WP_004545264/1-394 134 SAVAQV.....ANF.....HFWA.C.....Y....A.F.....PA...T..T.C.GIDQVYWSLSLEEQCYIVL.....PVLLYFAN.RRT..
NP_706267/1-390 147 TEELFVSI.....MKW...L..PFT.ALGMPNI....N.D.....VK...DSFT.I..NAAVTWTLVYEWFFYFSL.....PVISALIK.RKV.S
EFW62204/1-382 145 PASLTVSL.....MRW...A..PFT.ALGMPNI....N.G.....VK...DSFT.I..NAAVTWTLVYEWFFYFSL.....PVIAALFK.RRV.S
WP_000282635/1-349 111 DIIYGV.....RN.....N.QFVGTIWFINCLFVIIAID.....AI....FR.EIV.K
Q54131/1-367 106 STGLY.....ETV.....KLAI.K.....S....N.F.....HN.N.KMFLTGWFLFAYIFVSILS.....VIIIKSIK.RVV.V
WP_004194788/1-412 157 GPLGYI.....TNN.....WRLTIG.....QYGIND.LLRDTTPYGH...SISE.S.VFNGSIWTLIYEAKCYVMV.....GLFAMFGL.LTA.H
WP_001230914/1-332 101 .....PI.....KKIV.....G....S.L.....IQ.R.GYFFQFWFFGALILIYLCL.....PILRQFLNSKRS..
WP_000170108/1-342 108 .....RV.....PTTI.....H....N.L.....L.....NIYRYPLGVSWYLYTLWSILIVYGLLSIVFK.NRK..
P37669/1-331 111 Q.....K.PVFYHLWFFFAIAVIYLVS.....PLIQV....KN..V
Q70J69/1-422 131 SHVFLV.....QSW...VP....SQ.....W.....YY.....L.TMTGIDWSLSCEAFFYLCFP.....VLLPVLS.RAR.N

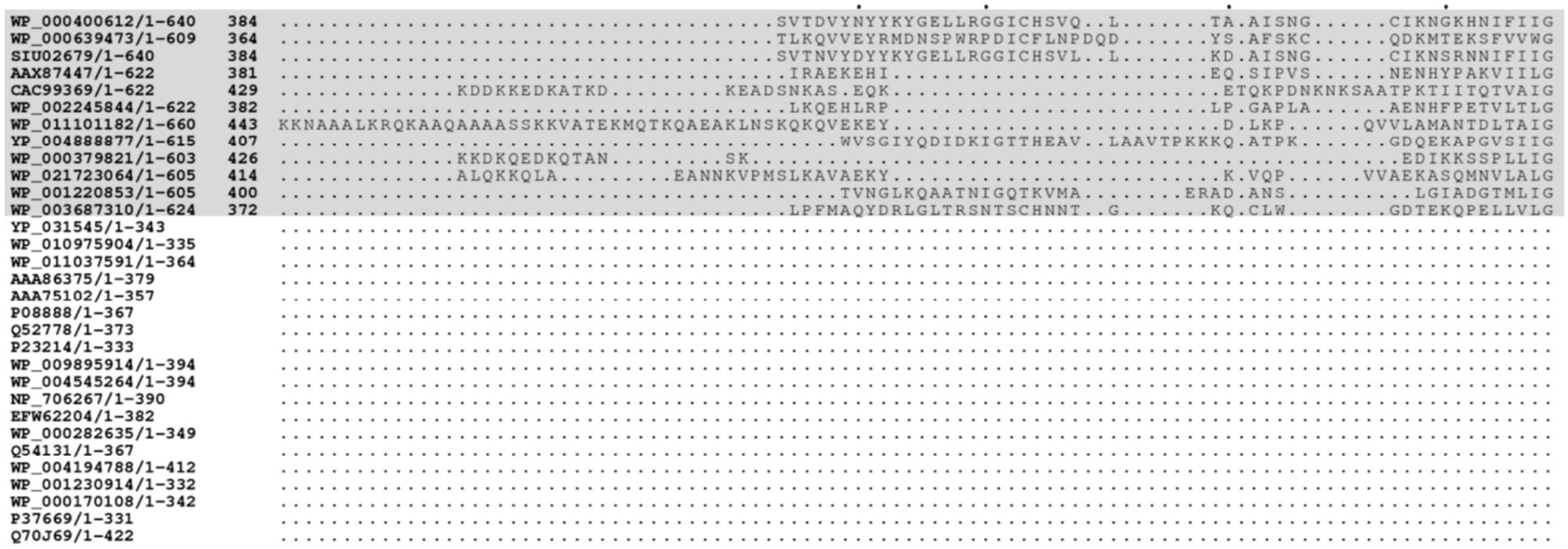

TMH5
TMH6
WP_000400612/1-640 163 HNR....NIV.....LSCAT.IFII...SYA.I....SIF.TM.....AS.....D.G....GA.....NYYSP.....A...
WP_000639473/1-609 161 .P.....VGL.....SLSV.ILAM...SLA.I....TLM.RV.....TG.....T.K....ED.....IFYLI.....P...
SIU02679/1-640 163 SKR....NIL.....LSCAA.IFIV...SYT.I....SVF.TM.....AY.....E.G....GA.....NYYSP.....A...
AAX87447/1-622 163 .EI....KVL.....FITTL.ILFF..ILLA.T....SFI.PA.....NFYKEVLH.Q....PN.....IYYLS.....N...
CAC99369/1-622 169 NPK....LLL.....KIVIG.LGLLSAVWMT.I....LYV.PG.....T.....D.P....SR.....VYYGT.....D...
WP_002245844/1-622 163 .SL....RVL.....RNISI.ILFL..ILTA.T....SFL.PS.....GFYTDILN.Q....PN.....TYYLS.....T...
WP_011101182/1-660 185 KRS....RVF.....WFMMI.AAGISAITMA.M....LY..DP.....A.....N.T....NR.....VYYGT.....D...
YP_004888877/1-615 159 RLASGRRALTRFRWGLGLLTTAFAIESFVMMY.F....GAV.GL.....K.....D.F....SP.....VYFSS.....L...
WP_000379821/1-603 174 PRN....IIQ.....TLFI.VSLISLGLMI.V....IHF.IT.....G.....D.N....SR.....VYFGT.....D...
WP_021723064/1-605 162 KRG....RIF.....VIANL.LALISAIWMA.I....LFV.PG.....A.....D.P....TR.....VYYGT.....D...
WP_001220853/1-605 159 SNG....QL...KGMVFLLSAVAFLISFFSMF.I....GSF.LV.....T.....S.Y....SS.....VYFSS.....L...
WP_003687310/1-624 164 .RV....QFG.....FLAAA.LCAL...SLA.A....SFM.PS.....AL.....D.....KYYLP.....H...
YP_031545/1-343 151 .RE....IEY.....FLIL.WLYA...SVV.VNLVKYYY.PI.....NF.....NI.....ELFYVT.....N...
WP_010975904/1-335 146 RNW....RLP.....VVSG..LFLALV..IA.....GRV.VA.....F.....D.D....AV.....MLTYT.....R...
WP_011037591/1-364 171 .....PPL.....VIAA.VAVV...LAW.F..WMNWF.PL.....QH.....MR.....LFWGL.....D...
AAA86375/1-379 163 .....SAA.....VLAV.CSLL...VVW.A..WTRWF.PG.....LR.....LR.....LPFAL.....D...
AAA75102/1-357 171 .RL....IMS.....IL.I.IPAL...LVA.G....YFT.LK.....EY.....
P08888/1-367 182 .AG....ALV.....VAVAA...LG.SWVMA.....QHF.NI.....T.....D.K....YNPFLSVT.....A.....G...
Q52778/1-373 160 .RL....SIW.....IISA.S.....AIAVAFA.PI.....TL.....SIT....PL.....LKY.....
P23214/1-333 162 .GK....AFI.....VIL.LV.....FVS.L....SLIGS.....VS.....E.N....RD.....VMFSIPLWLYP...
WP_009895914/1-394 186 .IM....PIL.....CVVV.....LI.Q....VFL.PR.....PI.....LSFLW.....F...
WP_004545264/1-394 186 .IA....PIL.....CVVV.....LI.Q....VFL.PR.....PI.....LSFLW.....F...
NP_706267/1-390 208 IY.....M.....VMISA.ISLFVFIL.....F.FS.....K...
EFW62204/1-382 206 IH.....M.....IIISV.LVLVIFLC.....YPSK.....K...
WP_000282635/1-349 147 .NN....IVI.....LIISL.LSFM..LSQT.V...LNHN.PL.....L.....D....PQ.....WFWNID.....
Q54131/1-367 156 .S.....NAL.....LLSV.LVAI...SVL.L....ITV.SI.....TY.....L.S....PQ.....YILVK.....DY.K
WP_004194788/1-412 224 .RR....VLL.....AVTV.VAWF...VLA.I....QTINPAFSAQLLPWAG.....D.R....HL.....VQY.....
WP_001230914/1-332 144 .....YLY.....SLSL.LMTI...GLI.F....ELS.NI.....LL.....Q.M....PI.....QTYVIQ.....TFRL
WP_000170108/1-342 152 .....SLF.....LVSV.FAYI..FTLF.I....QI..D.....IFIVQ.....RT..
P37669/1-331 140 GGK....MLL.....VLMAV.IGII...AN..P....NTV.PQ.....KI.....D.GFEWLPI.....NLYIN.....G...
Q70J69/1-422 179 .GM....LYA.....VSACS.VFLV.FFLPYVVG...HTF.AI.....R.....S.....PDPVEMVPTAGYGGPIGYWFAYVFP.....P...


TMH6
TMH7
WP_000400612/1-640 199 .SR...FWEL...MAGAIISTL.RFI.GI.NT..SLSKL....M.SL.....LGI.IL.IA.....L...SIT.M..
WP_000639473/1-609 194 .TR...AWEM...LAGGLVYIA.SVR.YK.MP..EWIKH....C.EV.....YGI.VL.IV.....V...AVV....
SIU02679/1-640 199 .SR...FWEL...MAGAIIATL.RFM.GI.KT..SVSKS....M.SL.....IGV.II.IT.....L...SIA.L..
AAX87447/1-622 204 .LR...FPEL...LVGSLLAIY.HNL.SA.SK..QASK..Q..ASNV.....IAI.LS...T.LL.LF.....S...CLF.L..
CAC99369/1-622 207 .TR...AFDL...LSGCALAFV.WPF.NRLSP..VVPRKS.KAVLNI.....AGT....ISILCF.IL.....F...TAF....
WP_002245844/1-622 204 .LR...FPEL...LAGSLLAVY.GQT.QNGRR..QTAN..G..KRQL.....LSS.LC...F.GA.LL.....A...CLF.V..
WP_011101182/1-660 222 .TR...MFAI...LLGSGLAFI.WPS.RELSA..DIANVN.RVTLDI.....LGG....ASL.IA.II.....W...IFF.Q..
YP_004888877/1-615 208 .TH...CFPF...FVGGLIGVL.SGI.KA.QG..PIYRWT.TAH.....CNPIIAGAVMLISF.IL.LV.....A...LGD.R..
WP_000379821/1-603 211 .TR...LQTL...LLGCILAFI.WPP.FALKK..DISKKI.VVSLDI.....IGI....SGF.AV.LM.....T...LFF.I..
WP_021723064/1-605 200 .TR...FFSL...IMGASLAFV.WPL.NKLSH..KVNKRAVKIAWQL.....TIG....LSL.LL.LL.....A...YIF....
WP_001220853/1-605 201 .TH...VYPF...FLGSMLATI.VGV.RQ.TT..SLVKQL.DKIWDLRKTLVVFGGGF...GF.LV.LL.....T.....F.F..
WP_003687310/1-624 195 .LR...ACEM...LVGSLTAVR.MRY.RQQRN.PAVGKR....Y.AA.....VGA...LFSA.CI.LS.....A...CLF.A..
YP_031545/1-343 188 .Y....VGYF...LLGYYLSNF.....DI.SK..KWRNI....S.YI.....GG.F.VGF.....I...STF.FI.
WP_010975904/1-335 180 .PV...ILEF...VAGMIIGEF.WLK.GRVPP..L....AVGS.ALF.....ACS....LGG.FA.LI.....G...VLG....
WP_011037591/1-364 203 .VL...PVSLCFYALGALLIHV.SPYLPT.SLP....GS....A..L.....VTV.....VVAALVA
AAA86375/1-379 195 .VL...PVALFFIAVGAWLSRF.AERV.R.ALPAVVWVV....A.....FPVLAFA
AAA75102/1-357 194 ..V...ALVW...FLGSGCAYH.F.C.NIN...RKYNNH....G.IL.....MLS..LITGA.AF.LV.....RFH.V..
P08888/1-367 217 .PT...FWIF...SMGVLARLY.W.H.RV.SK..IFEGK...LLWWL.....AT.....HL.....AITWWVA.G..
Q52778/1-373 189 .T....YPFY...CLGFLFAQP.IG..WQ.NG..VIWRY....K.WI.....FV.VLLSI.....A...A.....
P23214/1-333 198 .LR...GLAF...FFGATMAMY.EKS.W.....NVSNV....KITV.....VSL.L..AM.....Y...AY.....
WP_009895914/1-394 211 .VR...TDAL...ALGALLALL.QHR.EW.YA..KVEPRTL..ARGL.....LSA.IALAGM.CL.IL.....A...ALP.G..
WP_004545264/1-394 211 .VR...TDAL...ALGALLALL.QHR.DW.YA..KVEPRPL..GRGL.....ISA.IALAGM.CI.VL.....A...ALA.G..
NP_706267/1-390 228 .IH...IASF...LFGL.LAFL.LNK.SKIVN..GIAK....AK.....VT.....PIIITA..IM.V.....F.E..
EFW62204/1-382 227 .IH...VISF...LFGL.LASY.SNN...YVH..RVAK....AK.....IT.....PIMIIA..LL.S.....Y.E..
WP_000282635/1-349 182 .SAMAYWWLL...PLGRCMF.LEL....T.RD..RFFGK.....SK.....I.....GF.IV.FSITAIMSAYQLLNQKPLLFKII...SI.....
Q54131/1-367 191 LNF...ICQV...LTGMS.FYI.FGY.VI.RN..QIYNL....L.NF.....YVF.IL.L.....TVIL...
WP_004194788/1-412 264 ......GTIF...LIGSSAAAY.SKS.L.....PISDK.....LGA.....FAV.VV.YL.....I.....
WP_001230914/1-332 180 WTW...FFYY...LLGGYIAQF.TIE.EI.ES..RFKNW....M.KI.....VSI.LL.LL.....I...SPI....
WP_000170108/1-342 178 LVW...GICF...FLGSVLSEI.HFD.KI.....NL....K.KF.....LFF.FV.LF.....DFIYMFA
P37669/1-331 179 .DT...FYYI...LYGMLGRAI.GMM.DT.QH..K...A....L.SW.....VSA.AL.FA.....T...GVF.I..
Q70J69/1-422 234 .MR...MAEF...VIGISLAVL.VRR.GA.WR.....GPGV.....AAC.LV.LC.....AIGWGVN.L..

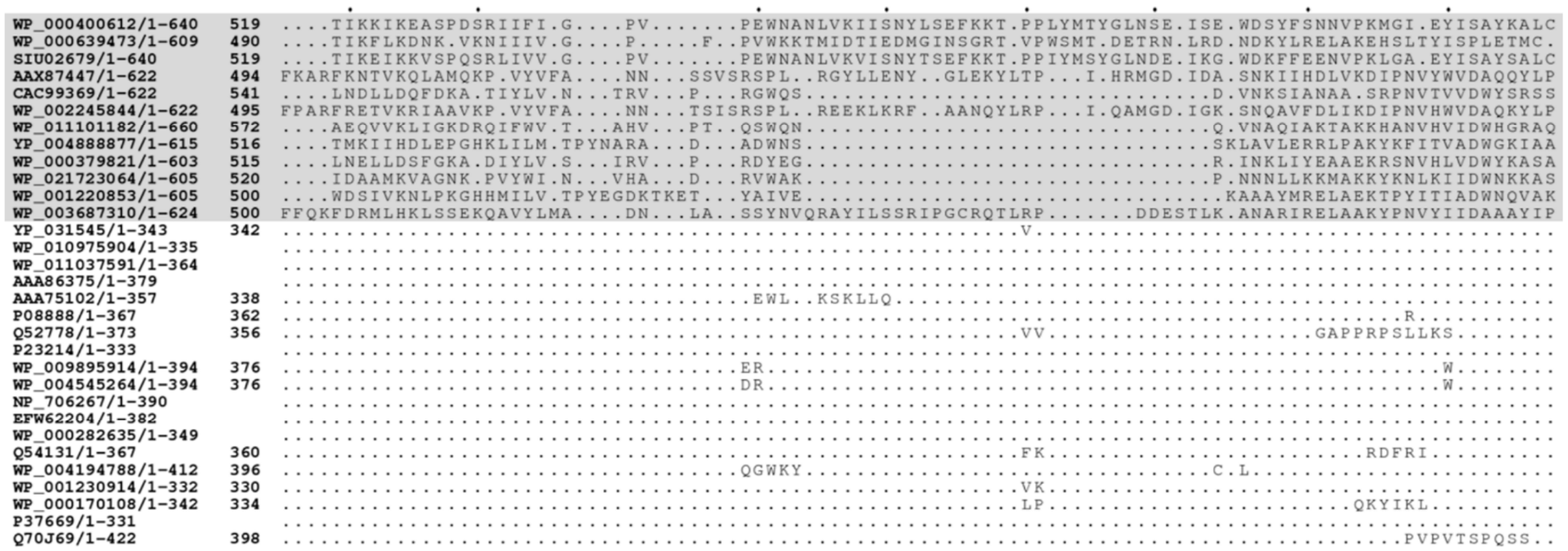

TMH7
TMH8
TMH9
WP_000400612/1-640 241 ..ID..EKM.....SFPGY.....IA..IIP...V.LGAS.LI.....IASN.GNDLVVSKLLSVRPV
WP_000639473/1-609 235 ..IL..HSN....G.....YWPSY.....SA..LAP...V.LGAS.MV.....ILAN.KQ.NS..LFTSNRIA
SIU02679/1-640 241 ..IN..EKM.....AFPGY.....IA..IIP...V.IGAS.LI.....IASN.GNDWIASKILSFKPI
AAX87447/1-622 250 ..MN..NDI....A.....YIPGI.....TL..ILP....CIFTA.LI.....IHTT.SQNNIIKLCLSNKVI
CAC99369/1-622 257 ..VSEYQPF.....LYRGG.....LL..FVA...I.LGVI.MI.....ATIS.HPASYLSKIFSFKPL
WP_002245844/1-622 251 ..ID..KHN....P.....FIPGM.....TL..LLP....CLLTA.LL.....IRSM.QYGTLPTRILSASPI
WP_011101182/1-660 272 ..MSGQSDF.....TYRGG.....ML..LFT...V.LSTV.LV.....ATVA.HPASHLNRVLTNPLF
YP_004888877/1-615 262 ..MKFDALQ.....TYQGG.....LL..LAT...V.LAGI.MI.....YGAR.LLHDHTPNLKEPRWM
WP_000379821/1-603 261 ..VGDQDQW.....IYNGG.....FY..IIS...F.ATLF.II.....AIAV.HPSSLFAKFLSMKPL
WP_021723064/1-605 250 ..MPAQGTF.....TYYGG.....MW..LAS...L.ASVI.MV.....ALVA.HPSLPTNKLFSNPVF
WP_001220853/1-605 255 ..VKFTYLF.....AYLIG.....FL..LAS...L.AALA.MI.....LAAR.VLHEKTHHIQESKII
WP_003687310/1-624 243 ..YS..EQT....A.....YFPGP.....AA..LIP....CLAVA.AL.....IYFN.HYEHPLKKFFQWKIT
YP_031545/1-343 226 .TYF..YTV....KA.NG.....QLEQFW..YGYFAPG.....VV..LMA.....IGLFIFFKYAFQKS..ERELPLLF
WP_010975904/1-335 225 ..LPFDELT.....TGP.....LA..VLL...V.IG.....VLS.LEANGC..VRALSLP
WP_011037591/1-364 244 WL...AGVN....G.....RIDVNM..LEFGRQH.....AVFLLSA...V.AGSL.MV.....I....CA.AR..LVQEWTWL
AAA86375/1-379 235 WGGV.AAMN....G.....QVDVNN..LQFGKSS.....LLFLIAS...L.LGTA.MT.....L....CI.AY..FMQGWRWL
AAA75102/1-357 238 ..LK..HSL....M.....NMYDL.....QL..VIPG...CIFLY.SL.....LLLL.STT.KLTKK.IELIS
P08888/1-367 260 ..TS.AAFI.....SI..N.....NA...AP...V.DAFR.IA.....VLAG.LVL.SAAHSLPRPNL
Q52778/1-373 226 .FIC..FLG....WG.K.....E.TYA..Y....NNLVLIHDEQSAKQVF..LMFSGSLAASAV.AMQSMFQCWRL.....VYS..TRVARFVA
P23214/1-333 235 ........A.....SYGKG.....ID.YTMT...C.YILV.SF.....STIA.ICT.S...VGD.PLV
WP_009895914/1-394 262 ..RL..ASI....P.....FHTGL.....IA..LVS....CIMVW.....IASY.GKSYTMPDGAVKRVL
WP_004545264/1-394 262 ..YL..ARI....P.....FHTGL.....IA..LMS....CVMVW.....IASY.GRGYTMRDGAVKRVL
NP_706267/1-390 269 ..MTY.FKT.....TYAPL.....PL..IL.....CGIT.FI.....IIAS.GC..DLYGILRLNIT
EFW62204/1-382 266 ..MIY.FPT.....TYAPL.....PL..II.....CGVC.FI.....LIAS.GC..DFYGILRLNLT
WP_000282635/1-349 241 ..FN.....A.....DIISS..SYIQAI.....NT.IITT.....VGLIIFN.....I....FI.AK..IICGNDFI
Q54131/1-367 231 ...Y..VSK....SYGFSTQTIM...S.....WSYYPDG.....LI..MSV...I.NALI.GI.....YAVFFISLLI.TR..GMKEIKLL
WP_004194788/1-412 297 ........S.....LFKGG.....YL..LLG...Y.PAM...V.....YAIL.WLACRL..PRWARRI
WP_001230914/1-332 222 .....ILFFIAKTIYHNLF.....AEYFYDT.....LF...VK...V.STLG.IF.....LTILML..T..LN..ESR.RESI
WP_000170108/1-342 218 WFLF..YEV....GS.K.....K.DY...VSYINPG.....LW..GIA...F.IVCV.LV.....AFAIFPK.ME.KN..FPKTFLYF
P37669/1-331 218 ..IS..RGTLYELQW.R.....G.NFADTWYLYCGP.....MV..FIC...A.IA.L.LT.....LV....KNTL..DTRTIRGL
Q70J69/1-422 275 ..HL.DSYL.....PLAAG.....ML...IP...F..ALL.IP.....ALST.ADVNSAWSPLRWKPL

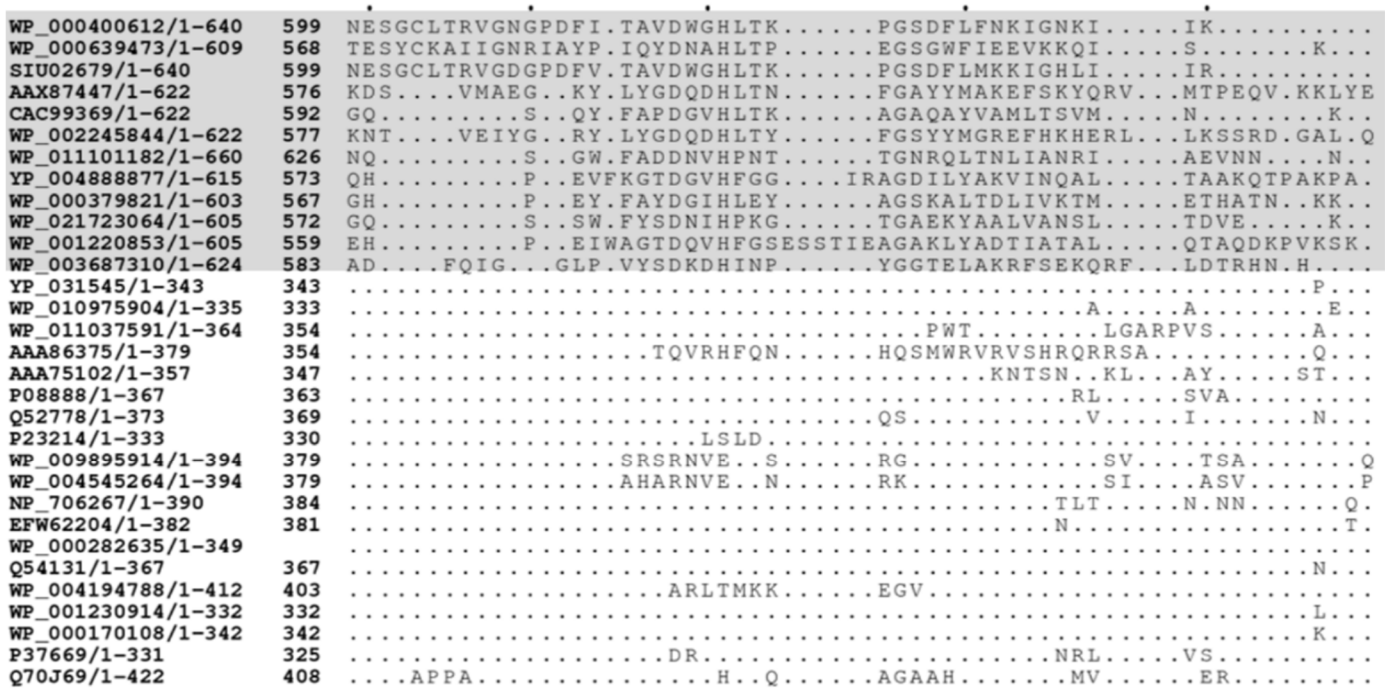

TMH9
TMH10
WP_000400612/1-640 282 VFFGLISYPLYLWHW....PIYSF.YRSIFAG.....S.P.....DY.H.....EL....T....LL.....LL.LSFFLAIL...
WP_000639473/1-609 274 QWVGKISYSVYLWHW....PVIVA.MKHYDIEF.....S.A.....IN.I.....FF....G.....VI.VSFALGDI...
SIU02679/1-640 282 VFIGLISYPLYLWHW....PVYSF.YRSIFSG.....S.P.....ST.N.....EL....L....IL.....MA.LALVLAIL...
AAX87447/1-622 292 VFIGKISYSLYLYHW....IFIAF.AYYITGE.....KQI.....NN.Q.....SI....A....IV.....II.LTIIFSVL...
CAC99369/1-622 300 RWIGTRSYGIYLWHY....PIITL.TTPVLEI.....TQP.....NI.W.....RA....I....LQ.....VA.ATFIIAEL...
WP_002245844/1-622 293 VFVGKISYSLYLYHW....IFIAF.AHYITGD.....KQL.....GL.P.....AV....S....AV.....AA.LTAGFSLL...
WP_011101182/1-660 315 KYVGQRSYGIYLYQF....PVMIF.YETKVKNI.....GDH.....LL.L.....NS....L....IE.....VA.LILIVTEL...
YP_004888877/1-615 305 TFLADVSYSVYLYHW....PLYVI.FSHMMV.....N.W.....L.....AA....L....LT.....TV.LSIILSAL...
WP_000379821/1-603 304 LIIGKRSYSLYLWHY....PIIVF.VNSYYVQ.....GQI.....PV.Y.....VY....I....IE.....IL.LTALMAEI...
WP_021723064/1-605 293 EYIGSRSYGIYLWQL....PVFAF.AEAKVL.....APT.....AW.Y.....NL....I....WQ.....LA.LILILTEL...
WP_001220853/1-605 298 SFLADTSYAVYLFHW....PFYII.FSQLTS.....N.L.....L.....AV....L....LT.....LI.CSYGFASL...
WP_003687310/1-624 285 VAAGLISYSLYLWHW....PILAF.MRYIG.PD.....N.L.....PP.Y.....SP....A....AA.....IV.LTLAFSLI...
YP_031545/1-343 276 RFINQASLGIYILHF....FLLNN.LLYMVFPK.V.....NNH.....VH.A.....IL....A....IP.IN..VT.ITIVLSMV...
WP_010975904/1-335 259 GLLGDASYSIYLWHT....FAISV.VAKAGLAI.....GLG.....AP.A.....TM....F....AA.....VL.SGTLIGIA...
WP_011037591/1-364 291 QWIGRNTLLILCTHM....LVFFV.LSGVAALA.....GGF.....GG.A.....RP....G....L.GWAIFVTLFALVASVP...
AAA86375/1-379 284 RWIGANTLLILGTHT....LVFLV.VTSVVVRT.....GVI.....DR.K.....LI....G....TPVWALALCAFAIAACIP...
AAA75102/1-357 279 AFLAFISYTLYLSHE....PIRRV.VSTFIEPT.....N.L.....KR.G.....FL....I....CG.....VCIACATI...
P08888/1-367 298 LRRQDLSYGIYLYHM....LVMHT.LIAIGW.V.....G.....HW.W.....LW....I....VE.....PV.GTVALAAL...
Q52778/1-373 290 VQLGQSTLLLYLVQG....AVFRL.MDLIQFGE.V.....WNL.....TTRI.....TF....A....TV.LG..VA.I.VVIAMA...
P23214/1-333 268 KGRFDYSYGVYIYAF....PVQQV.VINT..L.....HMG.....FY.P.....SM....L....LS.....AV.TVLFLSHL...
WP_009895914/1-394 302 VYFGSRSYAIYLIHV....PVYRL.TREIWERVATTPSAIDGSF.....TL.R.....FA....F....TA.....IP.IVLLLAEL...
WP_004545264/1-394 302 VYFGSRSYAIYLIHV....PVYRL.TREIWERIVITPSTIGGSF.....TL.H.....FA....F....TA.....IP.IVLVLAEL...
NP_706267/1-390 307 RKLGETTYSVYLLHG....IFLYCLMTWIIPN.....N.Y.....TE.N.....TF....IILVSTT.....AF.LITFTSCL...
EFW62204/1-382 304 RKLGETTYSIYLLHG....VFLYCVMTWVIPS.....N.Y.....SN.Y.....FF....IMLISIT.....AF.CVTLLSCI...
WP_000282635/1-349 281 VRAGRNTLNICGMEYITKLFIPMA.LAIIGFSV.....TIP.....NP.....IC.....A.....II.YTCICVYVSDK
Q54131/1-367 285 KMIGQNSRAIMAYHL....LVYVI.LDIIAS.....IL.....GD.Y...SLSGTDVYDNHFI....TK.WSVPVY.IALGLLLP...
WP_004194788/1-412 330 GSRNDYSYGIYVFGF....LVQQV.LAYVGAY.....KYG.....FV.F.....YL....A....AS.....VF.FTFICAWF...
WP_001230914/1-332 270 VSLSNQTMGVFIIHT....YIMKV.WEKVLG.F.....NFV.....GA.Y.....LL.....F...AL.FTLSVSFI...
WP_000170108/1-342 269 TKYGKDSLGIYILHA....PICSM.IRILMLKV.....G.I.....NS.V.....FL....H....VV.VG..IV.LGWYLSIL...
P37669/1-331 268 GLISRHSLGIYGFHA....LIIHA.LRTRGIE.....L.K.....NW.P.....IL....D....II.W...IF.CATLAASL...
Q70J69/1-422 315 VWFGEISLSFYVSHL....LVQEE.LFSRLWNL.....GMRAGLLPAPLPVLSW.WVAVL.....SF....L....AQ.....FA.LAVLAAWL...


TMH11


WP_000400612/1-640 384 .....SVTDVYNYYKYGELLRGGICHSVQ..L.....TA.AISNG.....CIKNGKHNIFIIG
WP_000639473/1-609 364 .....TLKQVVEYRMDNSPWRPDICFLNPDQD.....YS.AFSKC.....QDKMTEKSFVVWG
SIU02679/1-640 384 .....SVTNVYDYYKYGELLRGGICHSVL..L.....KD.AISNG.....CIKNSRNNIFIIG
AAX87447/1-622 381 .....IRAEKEHI.....EQ.SIPVS.....NENHYPAKVIILG
CAC99369/1-622 429 .....KDDKKEDKATKD.....KEADSNKAS.EQK.....ETQKPDNKNKSAATPKTIITQTVAIG
WP_002245844/1-622 382 .....LKQEHLRP.....LP.GAPLA.....AENHFPETVLTLG
WP_011101182/1-660 443 KKNAAALKRQKAAQAAAASSKKVATEKMQTKQAEAKLNSKQKQVEKEY.....D.LKP.....QVVLAMANTDLTAIG
YP_004888877/1-615 407 .....WVSGIYQDIDKIGTTHEAV..LAAVTPKKKQ.ATPK.....GDQEKAPGVSIIG
WP_000379821/1-603 426 .....KKDKQEDKQTAN.....SK.....EDIKKSSPLLIG
WP_021723064/1-605 414 .....ALQKKQLA.....EANNKVPMSLKAVAEKY.....K.VQP.....VVAEKASQMNVLALG
WP_001220853/1-605 400 .....TVNGLKQAATNIGQTKVMA.....ERAD.ANS.....LGIADGTMLIG
WP_003687310/1-624 372 .....LPFMAQYDRLGLTRSNTSCHNNT..G.....KQ.CLW.....GDTEKQPELLVLG
YP_031545/1-343
WP_010975904/1-335
WP_011037591/1-364
AAA86375/1-379
AAA75102/1-357
P08888/1-367
Q52778/1-373
P23214/1-333
WP_009895914/1-394
WP_004545264/1-394
NP_706267/1-390
EFW62204/1-382
WP_000282635/1-349
Q54131/1-367
WP_004194788/1-412
WP_001230914/1-332
WP_000170108/1-342
P37669/1-331
Q70J69/1-422

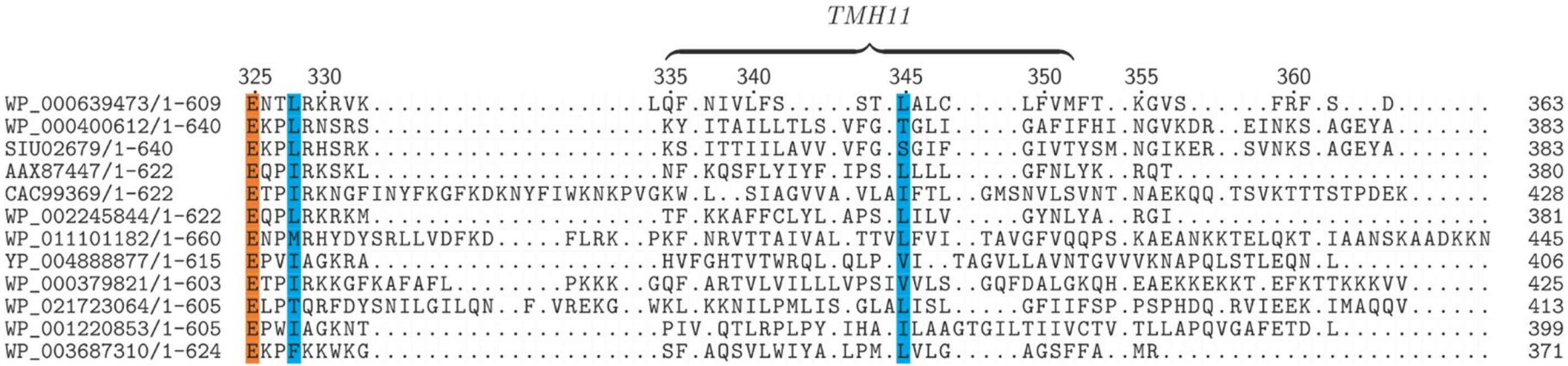

WP_000400612/1-640 429 DSYAAALFNGLSHYIDNKGSDYIIS.QMTD.GNAPPLFVD...GKDDLQRSVITLNNNRINEIKRVQPEVVLLT..WSVRGTNGVH..D.KKLAIDALSL
WP_000639473/1-609 411 DSHAAHLMPGLKSVFGNSL...NIT.QRTA.SLCPPIIGL...QKDDRP.YCKDINDMVAKEISDNKPTTVLMSALWPV...YP.........MRDYLPE
SIU02679/1-640 429 DSYAAALYNGLSSYIKNNNEKYVIS.QMTD.GNAPPLFVS...GKDDLQRDVGSINADRIKEIGIVKPEIVLLT..WSVRGSNGVH..D.KKLAIEALSL
AAX87447/1-622 409 DSHSSHLEAFLNYV.GNKE.GWKAD.IFKDKFECSFIVNE...QYQLDP.NCQSVWQKDSQY.KAIFISAFYDL.RMGG...QPVPRFR.PETFIEP..D
CAC99369/1-622 479 DSVMLDIEPYLKEAVPNITIDGLVGRQLRD...AITTATG....YKK.................FNSENSSVILE..L...GTNGPF..T.E.......DQ
WP_002245844/1-622 410 DSHAGHLRGFLDYV.GSRE.GWKAK.ILSLDSECLVWVDE...KLADNP.LCRKYRDEVEKA.EAVFIAQFYDL.RMGG...QPVPRFE.AQSFLIP..G
WP_011101182/1-660 510 DSVLLDVSSDLQDVIPGTVVQGRVGRQVTE...VPGIINS....LKS.................QGQLAHNVLLN..I...GTNGTI..T.D.......DQ
YP_004888877/1-615 453 DSVTLGTRSYLGDHVANSNIDAEGDRTMNL...AYKVMMN....QQR.................SHTLRQYVVIC..I...GTNALD..DYE.......EQ
WP_000379821/1-603 452 DSVMVDIGNVFTKKIPNAQIDGKVGRQLVD...ATPIVKS...QYKD.................YAKKGQKVVVE..L...GTNGAF..T.K.......DQ
WP_021723064/1-605 458 DSVMVAASTNLQEVFPHMYIDAAVGRQAES...LDTDLTN....AKS.................KMPNPDAYLIG..L...GTNGTI..K.E.......GE
WP_001220853/1-605 437 DSVALRANTALQTALPGAQINAQVSVTTKT...ANEIMLN....NSQ.................NKFLPKTVVIA..T...GVNNPE..NYK.......DD
WP_003687310/1-624 414 DSHADHYKTFFDAV.GKKE.KWSAT.MVSA.DACAYVEGYASRVFQNWA.ACRAVYRYAEEH.LPRYPKVVLAM.RWGS...QMPE.NS.RSLAYDA..G
YP_031545/1-343 341 ....L.
WP_010975904/1-335 324 ....RVTAGLAGR.
WP_011037591/1-364
AAA86375/1-379
AAA75102/1-357
P08888/1-367
Q52778/1-373 355 ....I.
P23214/1-333 327 ....SPK.
WP_009895914/1-394
WP_004545264/1-394
NP_706267/1-390 378 ....VKELIP.
EFW62204/1-382 375 ....IRGIIK.
WP_000282635/1-349
Q54131/1-367 358 .........................I.K.
WP_004194788/1-412 392 ....WGPG.
WP_001230914/1-332 329 ....I.
WP_000170108/1-342 332 .......................V.L.
P37669/1-331
Q70J69/1-422

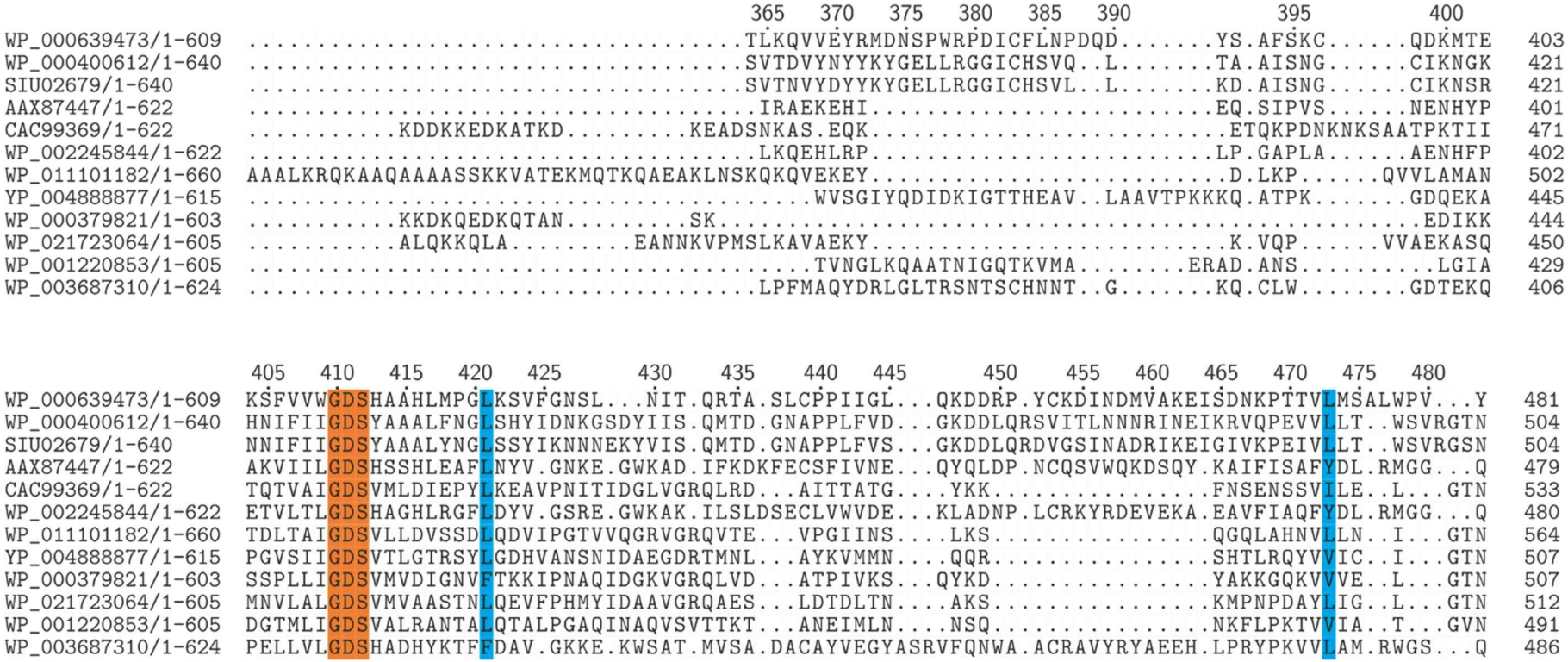

WP_000400612/1-640 519 ....TIKKIKEASPDSRIIFI.G....PV.......PEWNANLVKIISNYLSEFKKT.PPLYMTYGLNSE.ISE.WDSYFSNNVPKMGI.EYISAYKALC
WP_000639473/1-609 490 ....TIKFLKDNK.VKNIIIV.G....P.....F..PVWKKTMIDTIEDMGINSGRT.VPWSMT.DETRN.LRD.NDKYLRELAKEHSLTYISPLETMC.
SIU02679/1-640 519 ....TIKEIKKVSPQSRLIVV.G....PV.......PEWNANLVKVISNYTSEFKKT.PPIYMSYGLNDE.IKG.WDKFFEENVPKLGA.EYISAYSALC
AAX87447/1-622 494 FKARFKNTVKQLAMQKP.VYVFA....NN...SSVSRSPL..RGYLLENY..GLEKYLTP...I.HRMGD.IDA.SNKIIHDLVKDIPNVYWVDAQQYLP
CAC99369/1-622 541 ....LNDLLDQFDKA.TIYLV.N...TRV...P...RGWQS.....D.VNKSIANAA.SRPNVTVVDWYSRSS
WP_002245844/1-622 495 FPARFRETVKRIAAVKP.VYVFA....NN...TSISRSPL..REEKLKRF..AANQYLRP...I.QAMGD.IGK.SNQAVFDLIKDIPNVHWVDAQKYLP
WP_011101182/1-660 572 ....AEQVVKLIGKDRQIFWV.T...AHV...PT..QSWQN.....Q.VNAQIAKTAKKHANVHVIDWHGRAQ
YP_004888877/1-615 516 ....TMKIIHDLEPGHKLILM.TPYNARA...D...ADWNS.....SKLAVLERRLPAKYKFITVADWGKIAA
WP_000379821/1-603 515 ....LNELLDSFGKA.DIYLV.S...IRV...P...RDYEG.....R.INKLIYEAAEKRSNVHLVDWYKASA
WP_021723064/1-605 520 ....IDAAMKVAGNK.PVYWI.N...VHA...D...RVWAK.....P.NNNLLKKMAKKYKNLKIIDWNKKAS
WP_001220853/1-605 500 ....WDSIVKNLPKGHHMILV.TPYEGDKTKET...YAIVE.....KAAAYMRELAEKTPYITIADWNQVAK
WP_003687310/1-624 500 FFQKFDRMLHKLSSEKQAVYLMA....DN...LA..SSYNVQRAYILSSRIPGCRQTLRP.....DDESTLK.ANARIRELAAKYPNVYIIDAAAYIP
YP_031545/1-343 342 .....V.....
WP_010975904/1-335 .....
WP_011037591/1-364 .....
AAA86375/1-379 .....
AAA75102/1-357 338 .....EWL..KSKLLQ.....
P08888/1-367 362 .....R.....
Q52778/1-373 356 .....VV.....GAPPRPSLLKS.....
P23214/1-333 .....
WP_009895914/1-394 376 .....ER.....W.....
WP_004545264/1-394 376 .....DR.....W.....
NP_706267/1-390 .....
EFW62204/1-382 .....
WP_000282635/1-349 .....
Q54131/1-367 360 .....FK.....RDFRI.....
WP_004194788/1-412 396 .....QGWKY.....C.L.....
WP_001230914/1-332 330 .....VK.....
WP_000170108/1-342 334 .....LP.....QKYIKL.....
P37669/1-331 .....
Q70J69/1-422 398 .....PVPVTSPQSS..

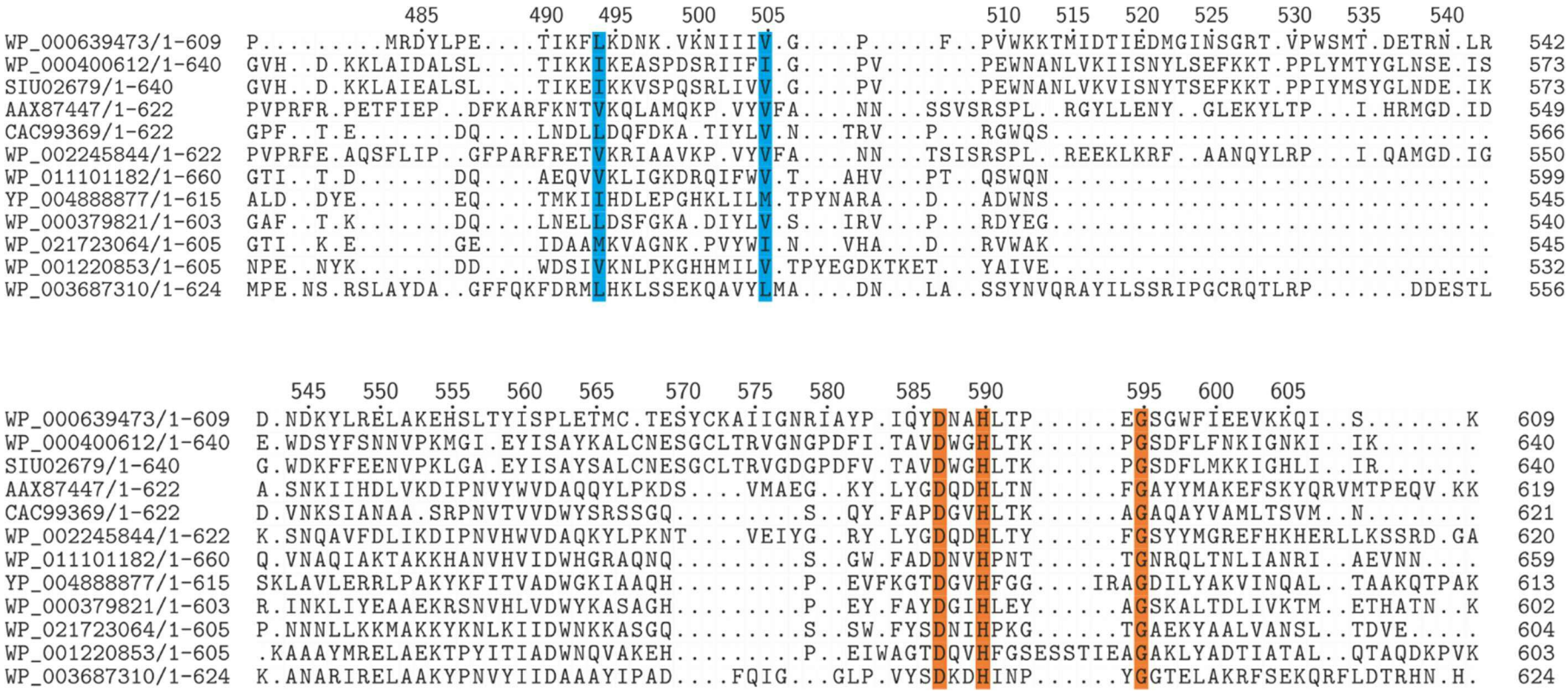

WP_000400612/1-640 599 NESGCLTRVGNGPDFI.TAVDWGHLTK......PGSDFLFNKIGNKI.....IK.....
WP_000639473/1-609 568 TESYCKAIIGNRIAYP.IQYDNAHLTP......EGSGWFIEEVKKQI.....S.......K...
SIU02679/1-640 599 NESGCLTRVGDGPDFV.TAVDWGHLTK......PGSDFLMKKIGHLI.....IR.....
AAX87447/1-622 576 KDS....VMAEG..KY.LYGDQDHLTN......FGAYYMAKEFSKYQRV...MTPEQV.KKLYE
CAC99369/1-622 592 GQ.........S..QY.FAPDGVHLTK......AGAQAYVAMLTSVM.....N.........K..
WP_002245844/1-622 577 KNT....VEIYG..RY.LYGDQDHLTY......FGSYYMGREFHKHERL...LKSSRD.GAL.Q
WP_011101182/1-660 626 NQ.........S..GW.FADDNVHPNT......TGNRQLTNLIANRI.....AEVNN....N..
YP_004888877/1-615 573 QH.........P..EVFKGTDGVHFGG....IRAGDILYAKVINQAL.....TAAKQTPAKPA.
WP_000379821/1-603 567 GH.........P..EY.FAYDGIHLEY......AGSKALTDLIVKTM.....ETHATN..KK..
WP_021723064/1-605 572 GQ.........S..SW.FYSDNIHPKG......TGAEKYAALVANSL.....TDVE.....K..
WP_001220853/1-605 559 EH.........P..EIWAGTDQVHFGSESSTIEAGAKLYADTIATAL.....QTAQDKPVKSK.
WP_003687310/1-624 583 AD....FQIG...GLP.VYSDKDHINP......YGGTELAKRFSEKQRF...LDTRHN.H....
YP_031545/1-343 343 ....P...
WP_010975904/1-335 333 ..........A.....A.........E..
WP_011037591/1-364 354 .....PWT.........LGARPVS.......A...
AAA86375/1-379 354 ...................TQVRHFQN......HQSMWRVRVSHRQRRSA..........Q...
AAA75102/1-357 347 ..........KNTSN..KL...AY.....ST...
P08888/1-367 363 ..........RL.....SVA.....
Q52778/1-373 369 .....QS..........V.....I.......N...
P23214/1-333 330 ....LSLD.....
WP_009895914/1-394 379 ..................SRSRNVE..S......RG.............SV....TSA.......Q
WP_004545264/1-394 379 ..................AHARNVE..N......RK.............SI....ASV.......P
NP_706267/1-390 384 ..........TLT.....N.NN.......Q.
EFW62204/1-382 381 ..........N.................T.
WP_000282635/1-349 .....
Q54131/1-367 367 ....N...
WP_004194788/1-412 403 ....................ARLTMKK......EGV.....
WP_001230914/1-332 332 ....L...
WP_000170108/1-342 342 ....K...
P37669/1-331 325 ....................DR.....................NRL.....VS.....
Q70J69/1-422 408 ....APPA.............H..Q......AGAAH.......MV......ER.....

Appendix III Conserved SGNH domain residues of characterised AT3 domain-containing O-acetyltransferases

Alignment as per Appendix II with sequences of AT3-Only proteins removed and subsequent empty columns deleted. Protein sequences are in the same order as Appendix I after *Salmonella* ser. Paratyphi A OafB (WP_00400612). 100% identical residues are coloured orange, similar residues conserved in > 90% of sequences are coloured blue.

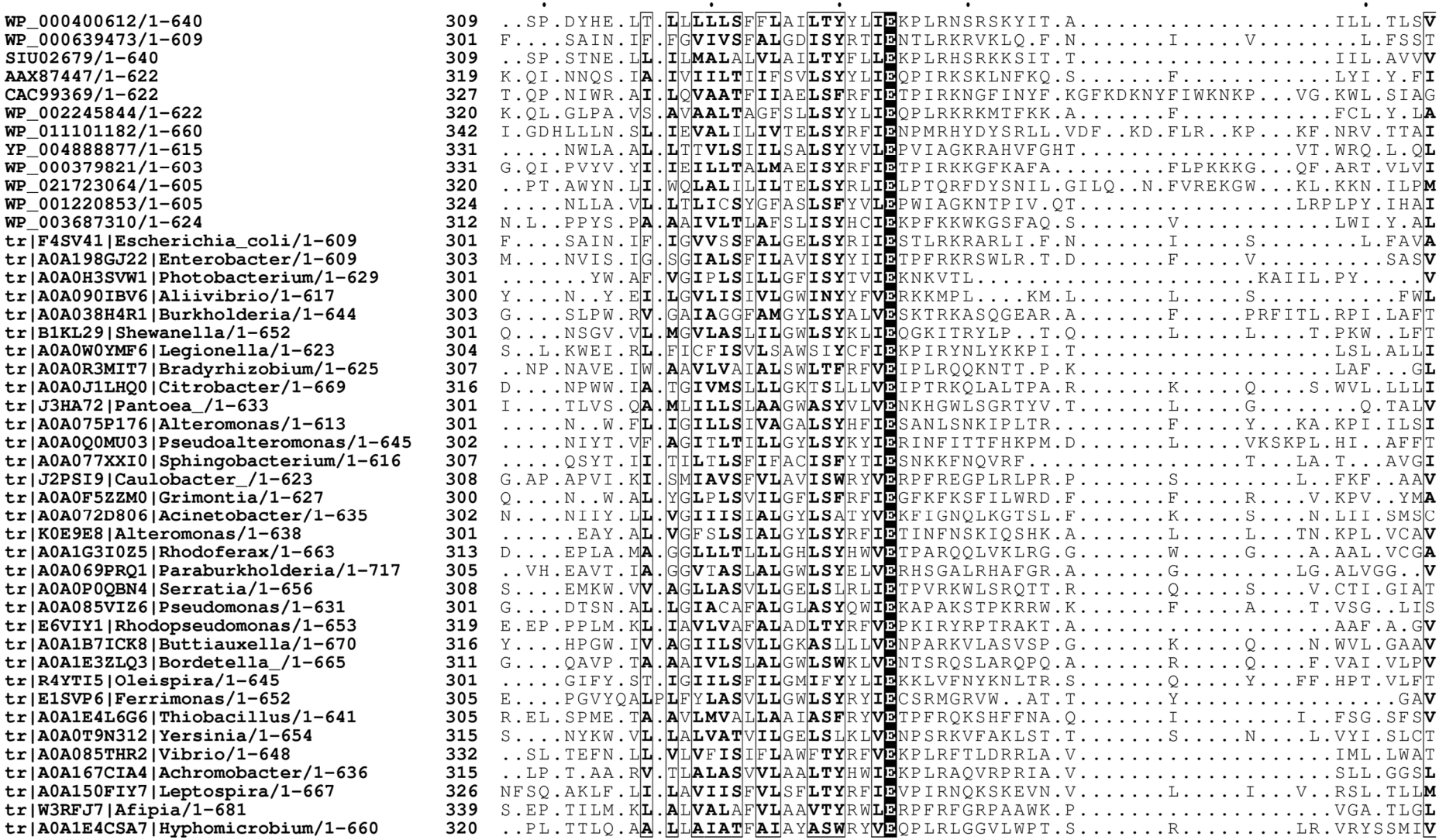


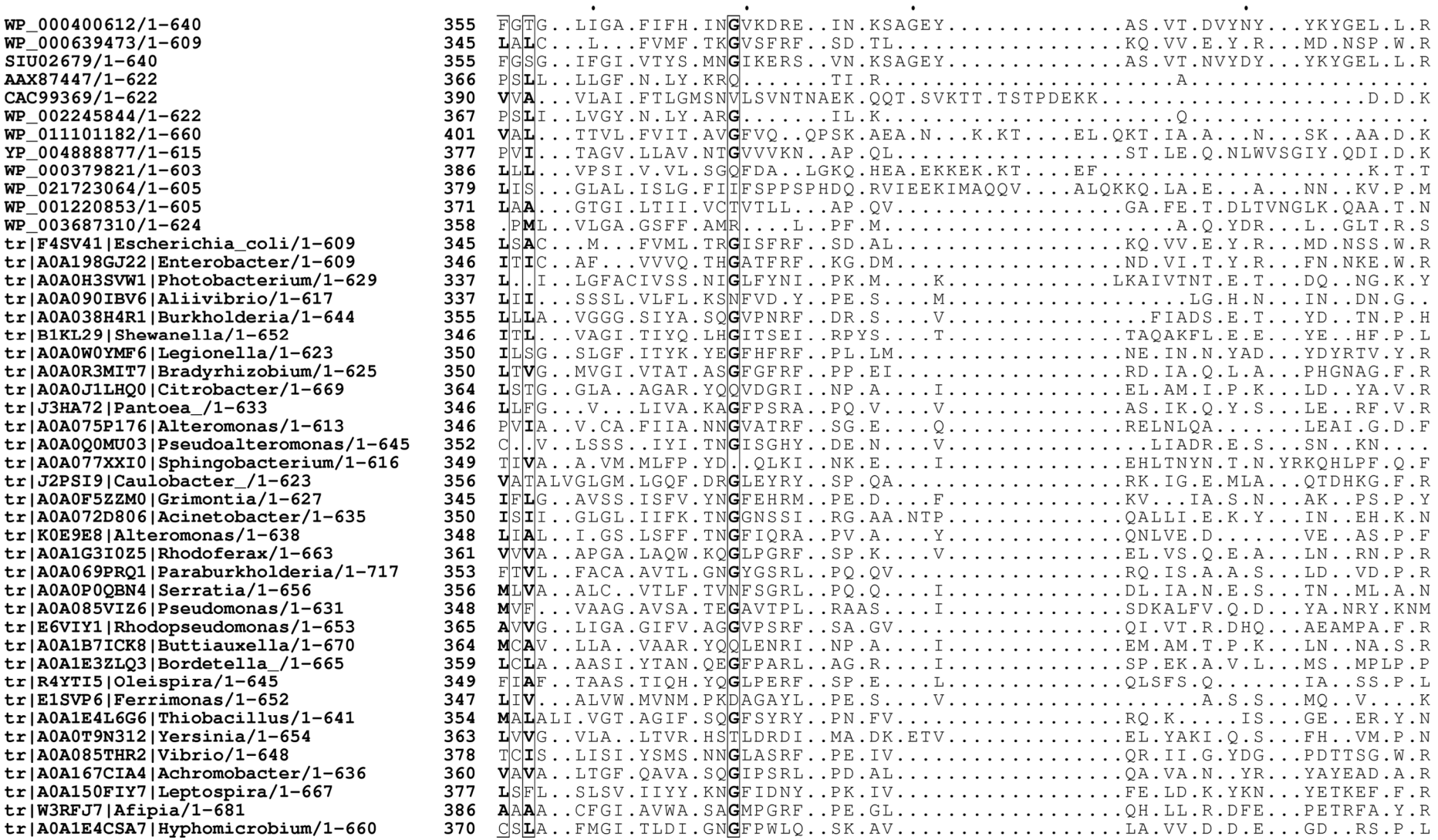

365 370 375 380 385 390 395 400
WP_000639473/1-609 TLKQVVEYRMDNSPWRPDICFLNPDQD YS.AFSKC QDKMTE 403
WP_000400612/1-640 SVTDVYNYYKYGELLRGGICHSVQ..L TA.AISNG CIKNGK 421
SIU02679/1-640 SVTNVYDYYKYGELLRGGICHSVL..L KD.AISNG CIKNSR 421
AAX87447/1-622 IRAEKEHI EQ.SIPVS NENHYP 401
CAC99369/1-622 KDDKKEDKATKD KEADSNKAS.EQK ETQKPDNKNKSAATPKTII 471
WP_002245844/1-622 LKQEHLRP LP.GAPLA AENHFP 402
WP_011101182/1-660 AAALKRQKAAQAAAASSKKVATEKMQTKQAEAKLNSKQKQVEKEY D.LKP QVVLAMAN 502
YP_004888877/1-615 WVSGIYQDIDKIGTTHEAV..LAAVTPKKKQ.ATPK GDQEKA 445
WP_000379821/1-603 KKDKQEDKQTAN SK EDIKK 444
WP_021723064/1-605 ALQKKQLA EANNKVPMSLKAVAEKY K.VQP VVAEKASQ 450
WP_001220853/1-605 TVNGLKQAATNIGQTKVMA ERAD.ANS LGIA 429
WP_003687310/1-624 LPFMAQYDRLGLTRSNTSCHNNT..G KQ.CLW GDTEKQ 406
405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480
WP_000639473/1-609 KSFVVWGDSHAAHLMPGLKSVFGNSL...NIT.QRTA.SLCPPIIGL...QKDDRP.YCKDINDMVAKEISDNKPTTVLMSALWPV...Y 481
WP_000400612/1-640 HNIFIIGDSYAAALFNGLSHYIDNKGSDYIIS.QMTD.GNAPPLFVD...GKDDLQRSVITLNNNRINEIKRVQPEVVLLT..WSVRGTN 504
SIU02679/1-640 NNIFIIGDSYAAALYNGLSSYIKNNNEKYVIS.QMTD.GNAPPLFVS...GKDDLQRDVGSINADRIKEIGIVKPEIVLLT..WSVRGSN 504
AAX87447/1-622 AKVIILGDSHSSHLEAFLNYV.GNKE.GWKAD.IFKDKFECSFIVNE...QYQLDP.NCQSVWQKDSQY.KAIFISAFYDL.RMGG...Q 479
CAC99369/1-622 TQTVAIGDSVMLDIEPYLKEAVPNITIDGLVGRQLRD...AITTATG....YKK.....FNSENSSVILE..L...GTN 533
WP_002245844/1-622 ETVLTLGDSHAGHLRGFLDYV.GSRE.GWKAK.ILSLDSECLVWVDE...KLADNP.LCRKYRDEVEKA.EAVFIAQFYDL.RMGG...Q 480
WP_011101182/1-660 TDLTAIGDSVLLDVSSDLQDVIPGTVVQGRVGRQVTE...VPGIINS....LKS.....QGQLAHNVLLN..I...GTN 564
YP_004888877/1-615 PGVSIIGDSVTLGTRSYLGDHVANSNIDAEGDRTMNL...AYKVMMN....QQR.....SHTLRQYVVIC..I...GTN 507
WP_000379821/1-603 SSPLLIGDSVMVDIGNVFTKKIPNAQIDGKVGRQLVD...ATPIVKS...QYKD.....YAKKGQKVVVE..L...GTN 507
WP_021723064/1-605 MNVLALGDSVMVAASTNLQEVFPHMYIDAAVGRQAES...LDTDLTN....AKS.....KMPNPDAYLIG..L...GTN 512
WP_001220853/1-605 DGTMLIGDSVALRANTALQTALPGAQINAQVSVTTKT...ANEIMLN....NSQ.....NKFLPKTVVIA..T...GVN 491
WP_003687310/1-624 PELLVLGDSHADHYKTFFDAV.GKKE.KWSAT.MVSA.DACAYVEGYASRVFQNWA.ACRAVYRYAEEH.LPRYPKVVLAM.RWGS...Q 486

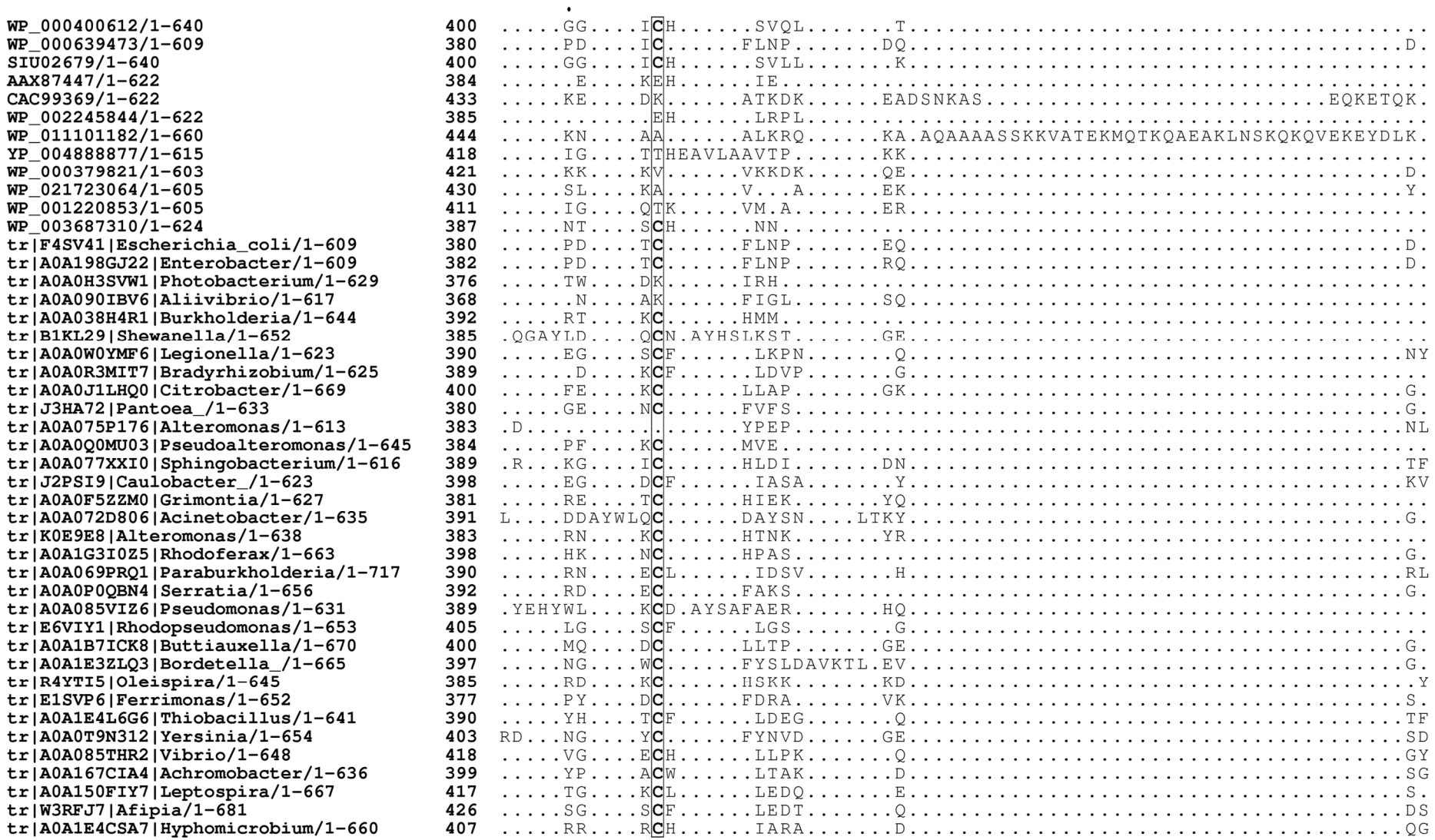

485 490 495 500 505 510 515 520 525 530 535 540
WP_000639473/1-609 542
WP_000400612/1-640 573
SIU02679/1-640 573
AAX87447/1-622 549
CAC99369/1-622 566
WP_002245844/1-622 550
WP_011101182/1-660 599
YP_004888877/1-615 545
WP_000379821/1-603 540
WP_021723064/1-605 545
WP_001220853/1-605 532
WP_003687310/1-624 556
545 550 555 560 565 570 575 580 585 590 595 600 605
WP_000639473/1-609 609
WP_000400612/1-640 640
SIU02679/1-640 640
AAX87447/1-622 619
CAC99369/1-622 621
WP_002245844/1-622 620
WP_011101182/1-660 659
YP_004888877/1-615 613
WP_000379821/1-603 602
WP_021723064/1-605 604
WP_001220853/1-605 603
WP_003687310/1-624 624

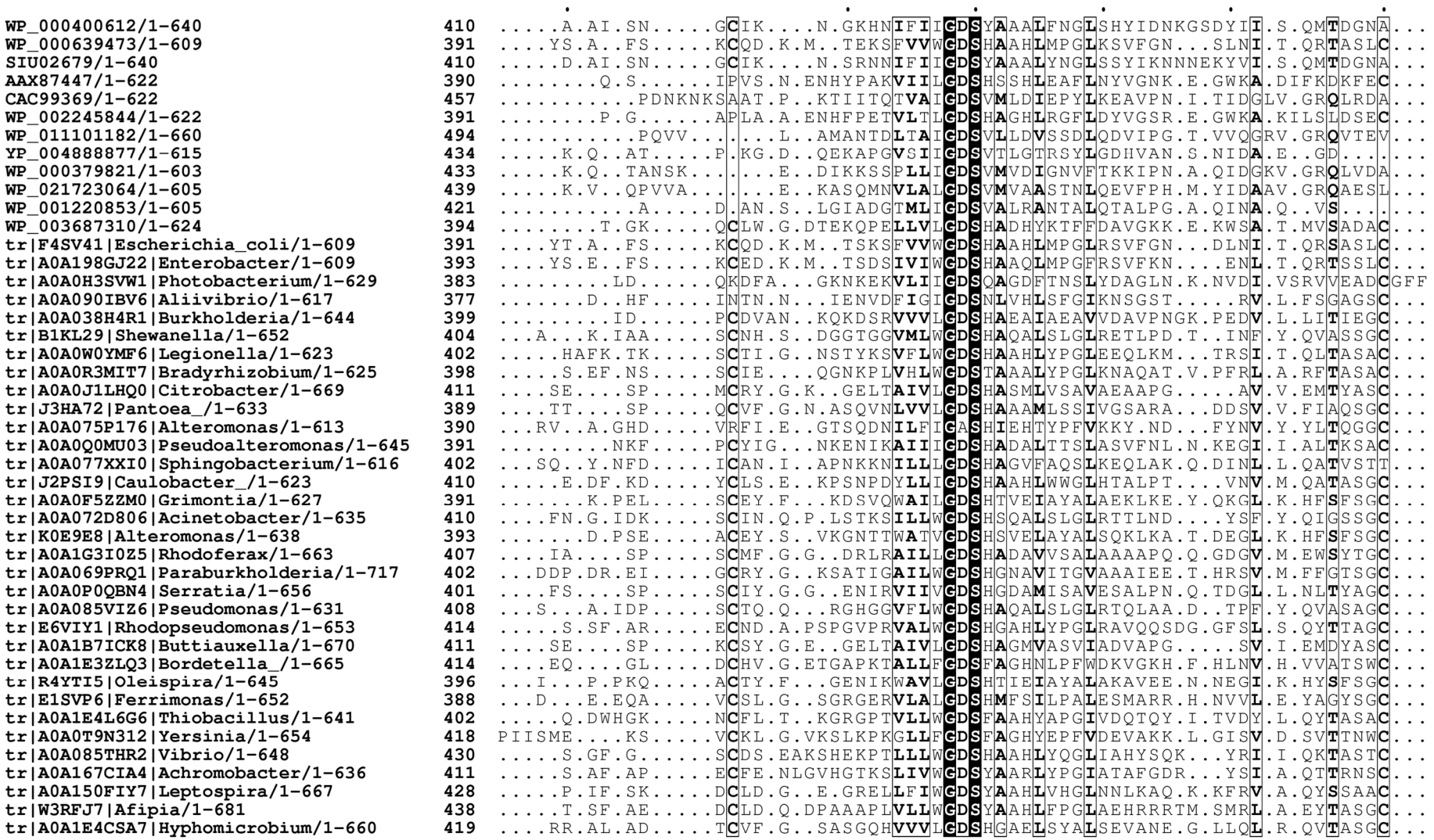
Appendix IV Alignment of diverse proteins with homology to O-antigen acetyltransferase OafA

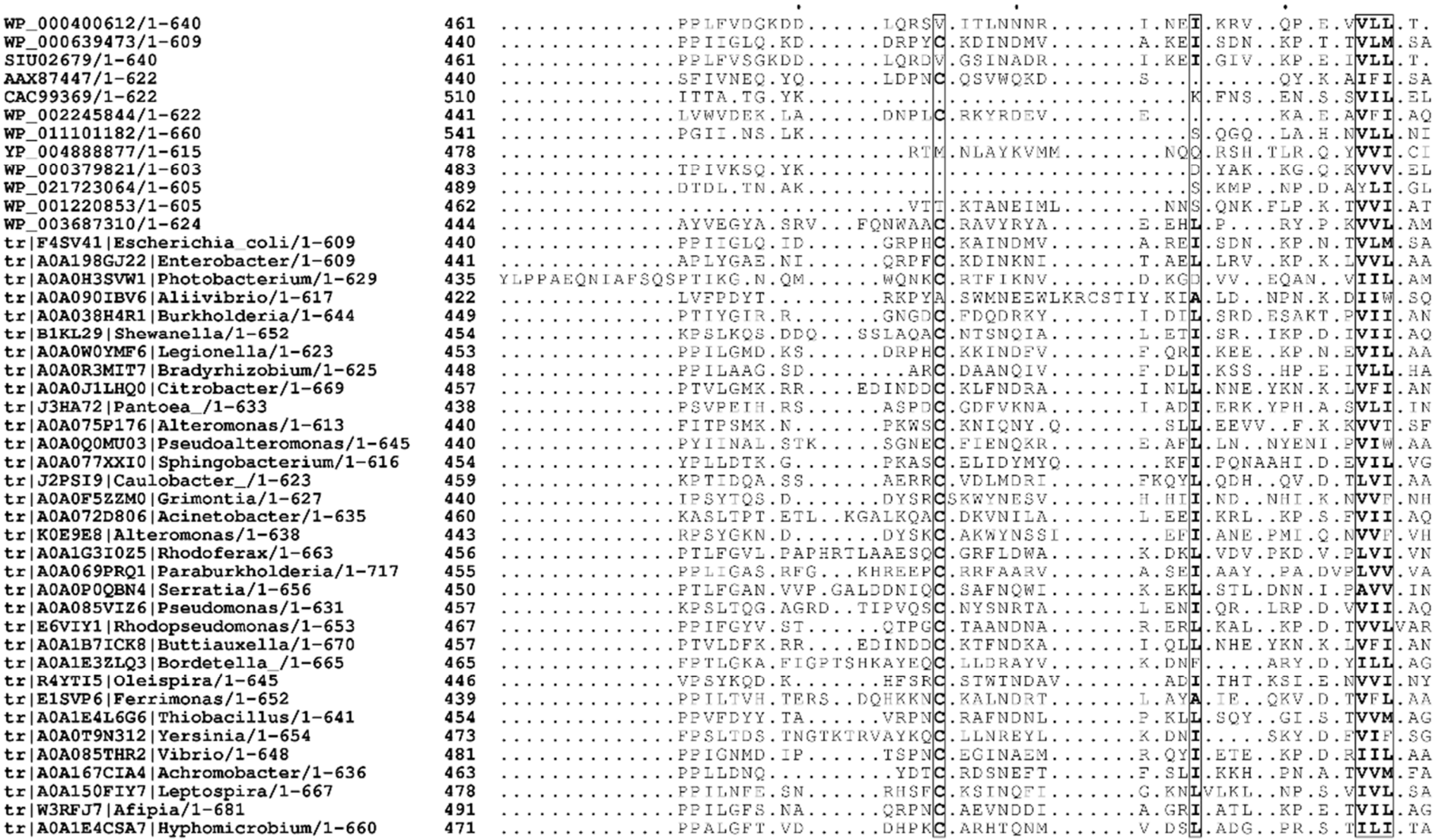
TCoffee alignment of AT3-SGNH fused proteins listed in Appendix I and the top 33 Uniref 50 sequences after blast searching the Uniref50 database against $OafA_{STM}$. The C-terminal half of the alignment is presented using ESPript coloured using black and white colour scheme by % Equivalent (global score 0.7).

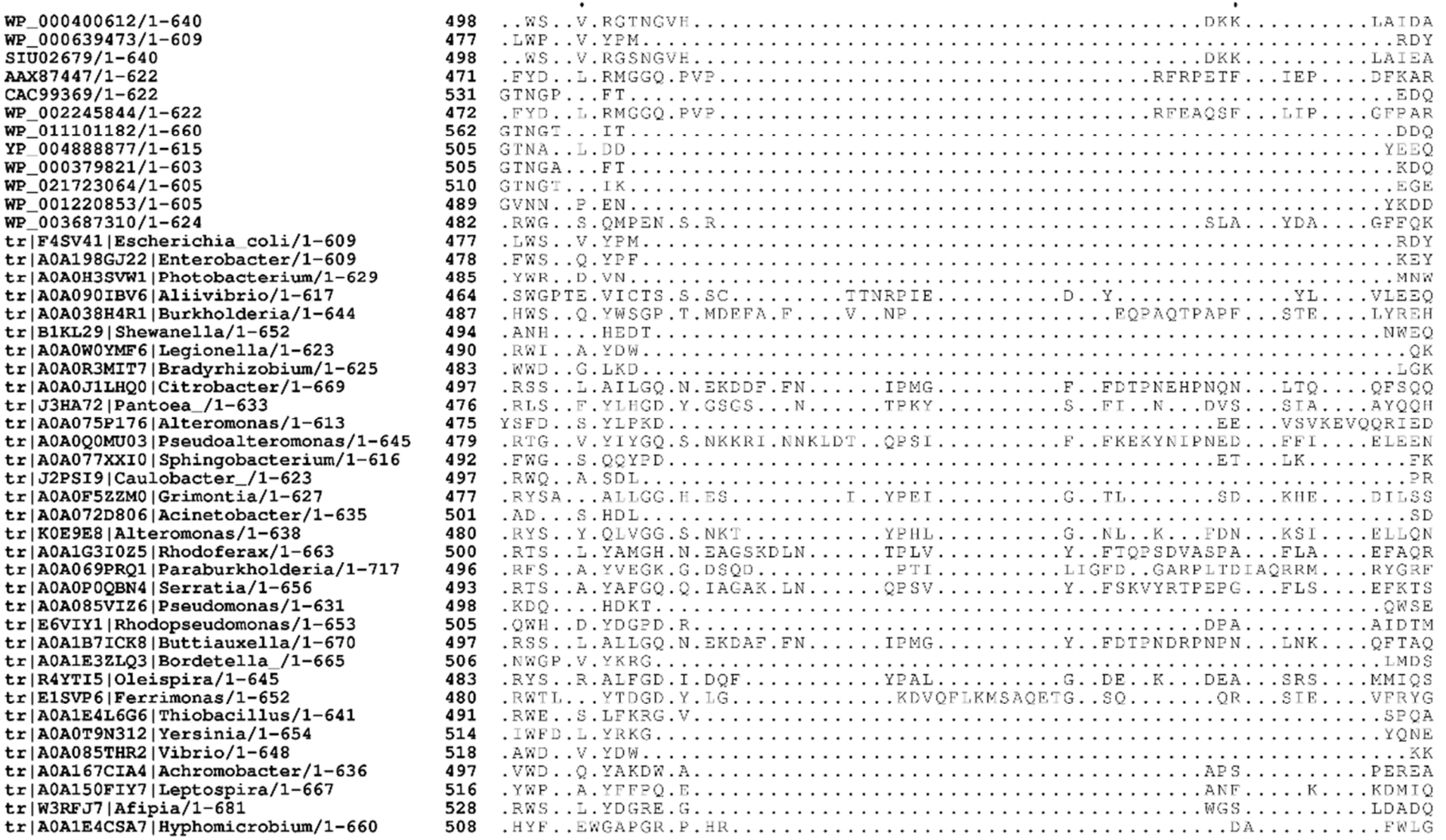

WP_000400612/1-640 309 ..SP.DYHE.LT.LLLLLSFFLAILTYYLIEKPLRNSRSKYIT.A.....ILL.TLSV
WP_000639473/1-609 301 F....SAIN.IF.FGVIVSFALGDISYRTIENTLRKRVKLQ.F.N.....I.....V.....L.FSST
SIU02679/1-640 309 ..SP.STNE.LL.ILMALALVLAILTYFLLEKPLRHSRKKSIT.T.....IIL.AVVV
AAX87447/1-622 319 K.QI.NNQS.IA.IVIILTIIFSVLSYYLIEQPIRKSKLNFKQ.S.....F.....LYI.Y.FI
CAC99369/1-622 327 T.QP.NIWR.AI.LQVAATFIIAELSFRFIETPIRKNGFINYF.KGFKDKNYFIWKNKP...VG.KWL.SIAG
WP_002245844/1-622 320 K.QL.GLPA.VS.AVAALTAGFSLLSYYLIEQPLRKRKMTFKK.A.....F.....FCL.Y.LA
WP_011101182/1-660 342 I.GDHLLLN.SL.IEVALILIVTELSYRFIENPMRHYDYSRLL.VDF..KD.FLR..KP...KF.NRV.TTAI
YP_004888877/1-615 331 .....NWLA.AL.LTTVLSIILSALSYYVLEPVIAGKRAHVFGHT.....VT.WRQ.L.QL
WP_000379821/1-603 331 G.QI.PVYV.YI.IEILLTALMAEISYRFIETPIRKKGFKAFA.....FLPKKKG...QF.ART.VLVI
WP_021723064/1-605 320 ..PT.AWYN.LI.WQLALILILTELSYRLIELPTQRFDYSNIL.GILQ..N.FVREKGW...KL.KKN.ILPM
WP_001220853/1-605 324 .....NLLA.VL.LTLICSYGFASLSFYVLEPWIAGKNTPIV.QT.....LRPLPY.IHAI
WP_003687310/1-624 312 N.L..PPYS.PA.AAIVLTLAFSLISYHCIEKPFKKWKGSFAQ.S.....V.....LWI.Y.AL
tr|F4SV41|Escherichia_coli/1-609 301 F....SAIN.IF.IGVVSSFALGELSYRIIESTLRKRARLI.F.N.....I.....S.....L.FAVA
tr|A0A198GJ22|Enterobacter/1-609 303 M....NVIS.IG.SGIALSFILAVISYYIIETPFRKRSWLR.T.D.....F.....V.....SASV
tr|A0A0H3SVW1|Photobacterium/1-629 301 .....YW.AF.VGIPLSILLGFISYTVIEKNKVTL.....KAIIL.PY.....V
tr|A0A090IBV6|Aliivibrio/1-617 300 Y....N..Y.EI.LGVLISIVLGWINYYFYVERKKMPL....KM.L.....L.....S.....FWL
tr|A0A038H4R1|Burkholderia/1-644 303 G....SLPW.RV.GAIAGGFAMGYLSYALVESKTRKASQGEAR.A.....F.....PRFITL.RPI.LAFT
tr|B1KL29|Shewanella/1-652 301 Q....NSGV.VL.MGVLASLILGWLSYKLIEQGKITRYLP..T.Q.....L.....L....T.PKW..LFT
tr|A0A0W0YMF6|Legionella/1-623 304 S..L.KWEI.RL.FICFISVLSAWSIYCFIEKPIRYNLYKKPI.T.....LSL.ALLI
tr|A0A0R3MIT7|Bradyrhizobium/1-625 307 ..NP.NAVE.IW.AAVLVAIALSWLTFRFVEIPLRQQKNTT.P.K.....LAF...GL
tr|A0A0J1LHQ0|Citrobacter/1-669 316 D....NPWW.IA.TGIVMSLLLGKTSLLLVEIPTRKQLALTPA.R.....K.....Q....S.WVL.LLLI
tr|J3HA72|Pantoea_/1-633 301 I....TLVS.QA.MLILLSLAAGWASYVLVENKHGWLSGRTYV.T.....L.....G.....Q.TALV
tr|A0A075P176|Alteromonas/1-613 301 .....N..W.FL.IGILLSIVAGALSYHFIESANLSNKIPLTR.....F.....Y...KA.KPI.ILSI
tr|A0A0Q0MU03|Pseudoalteromonas/1-645 302 .....NIYT.VF.AGITLTILLGYLSYKYIERINFITTFHKPM.D.....L.....VKSKPL.HI..AFFT
tr|A0A077XXI0|Sphingobacterium/1-616 307 .....QSYT.II.TILTLSFIFACISFYTIESNKKFNQVRF.....T...LA.T...AVGI
tr|J2PSI9|Caulobacter_/1-623 308 G.AP.APVI.KI.SMIAVSFVLAVISWRYVERPFREGPLRLPR.P.....S.....L..FKF..AAV
tr|A0A0F5ZZM0|Grimontia/1-627 300 Q....N..W.AL.YGLPLSVILGFLSFRFIEGFKFKSFILWRD.F.....F.....R....V.KPV..YMA
tr|A0A072D806|Acinetobacter/1-635 302 N....NIIY.LL.VGIIISIALGYLSATYVEKFIGNQLKGTSL.F.....K.....S....N.LII.SMSC
tr|K0E9E8|Alteromonas/1-638 301 .....EAY.AL.VGFSLSIALGYLSYRFIETINFNSKIQSHK.A.....L.....L...TN.KPL.VCAV
tr|A0A1G3I0Z5|Rhodoferax/1-663 313 D....EPLA.MA.GGLLLTLLLGHLSYHWVETPARQQLVKLRG.G.....W.....G....A.AAL.VCGA
tr|A0A069PRQ1|Paraburkholderia/1-717 305 ..VH.EAVT.IA.GGVTASLALGWLSYELVERHSGALRHAFGR.A.....G.....LG.ALVGG..V
tr|A0A0P0QBN4|Serratia/1-656 308 S....EMKW.VV.AGLLASVLLGELSLRLIETPVRKWLSRQTT.R.....Q.....S....V.CTI.GIAT
tr|A0A085VIZ6|Pseudomonas/1-631 301 G....DTSN.AL.LGIACAFALGLASYQWIEKAPAKSTPKRRW.K.....F.....A....T.VSG..LIS
tr|E6VIY1|Rhodopseudomonas/1-653 319 E.EP.PPLM.KL.IAVLVAFALADLTYRFVEPKIRYRPTRAKT.A.....AAF.A.GV
tr|A0A1B7ICK8|Buttiauxella/1-670 316 Y....HPGW.IV.AGIILSVLLGKASLLLVENPARKVLASVSP.G.....K.....Q....N.WVL.GAAV
tr|A0A1E3ZLQ3|Bordetella_/1-665 311 G....QAVP.TA.AAIVLSLALGWLSWKLVENTSRQSLARQPQ.A.....R.....Q....F.VAI.VLPV
tr|R4YTI5|Oleispira/1-645 301 .....GIFY.ST.IGIILSFILGMIFYYLIEKKLVFNYKNLTR.S.....Q.....Y...FF.HPT.VLFT
tr|E1SVP6|Ferrimonas/1-652 305 E....PGVYQALPLFYLASVLLGWLSYRYIECSRMGRVW..AT.T.....Y.....GAV
tr|A0A1E4L6G6|Thiobacillus/1-641 305 R.EL.SPME.TA.AVLMVALLAAIASFRYVETPFRQKSHFFNA.Q.....I.....I..FSG.SFSV
tr|A0A0T9N312|Yersinia/1-654 315 S....NYKW.VL.LALVATVILGELSLKLVENPSRKVFAKLST.T.....S.....N....L.VYI.SLCT
tr|A0A085THR2|Vibrio/1-648 332 ..SL.TEFN.LL.VLVFISIFLAWFTYRFVEKPLRFTLDRRLA.V.....IML.LWAT
tr|A0A167CIA4|Achromobacter/1-636 315 ..LP.T.AA.RV.TLALASVVLAALTYHWIEKPLRAQVRPRIA.V.....SLL.GGSL
tr|A0A150FIY7|Leptospira/1-667 326 NFSQ.AKLF.LI.LAVIISFVLSFLTYRFIEVPIRNQKSKEVN.V.....L.....I...V..RSL.TLLM
tr|W3RFJ7|Afipia/1-681 339 S.EP.TILM.KL.ALVALAFVLAAVTYRWLERPFRFGRPAAWK.P.....VGA.TLGL
tr|A0A1E4CSA7|Hyphomicrobium/1-660 320 ..PL.TTLQ.AA.LLAIATFAIAYASWRYVEQPLRLGGVLWPT.S.....R.....LR.VRYSSMIV

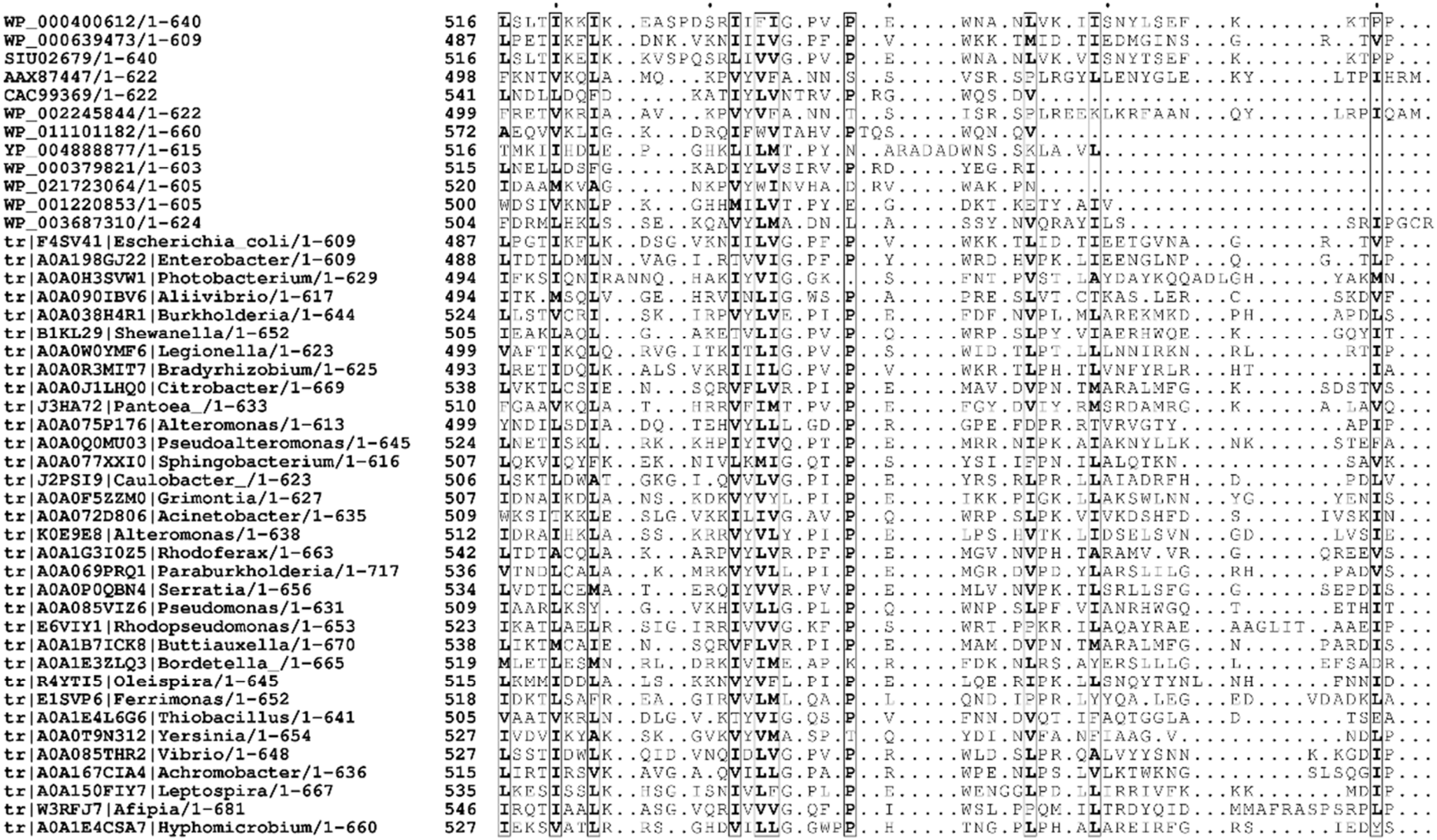

WP_000400612/1-640 355 FGTG..LIGA.FIFH.INGVKDRE..IN.KSAGEY.....AS.VT.DVYNY...YKYGEL.L.R
WP_000639473/1-609 345 LALC...L...FVMF.TKGVSFRF..SD.TL.....KQ.VV.E.Y.R...MD.NSP.W.R
SIU02679/1-640 355 FGSG..IFGI.VTYS.MNGIKERS..VN.KSAGEY.....AS.VT.NVYDY...YKYGEL.L.R
AAX87447/1-622 366 PSLL..LLGF.N.LY.KRQ.....TI.R.....A.....
CAC99369/1-622 390 VVA...VLAI.FTLGMSNVLSVNTNAEK.QQT.SVKTT.TSTPDEKK.....D.D.K
WP_002245844/1-622 367 PSLI..LVGY.N.LY.ARG.....IL.K.....Q.....
WP_011101182/1-660 401 VALL...TTVL.FVIT.AVGFVQ..QPSK.AEA.N...K.KT....EL.QKT.IA.A...N...SK..AA.D.K
YP_004888877/1-615 377 PVIL...TAGV.LLAV.NTGVVVKN..AP.QL.....ST.LE.Q.NLWVSGIY.QDI.D.K
WP_000379821/1-603 386 LLL...VPSI.V.VL.SGQFDA..LGKQ.HEA.EKKEK.KT....EF.....K.T.T
WP_021723064/1-605 379 LIS...GLAL.ISLG.FIIFSPPSPHDQ.RVIEEKIMAQQV....ALQKKQ.LA.E...A...NN..KV.P.M
WP_001220853/1-605 371 LAA...GTGI.LTII.VCTVTLL...AP.QV.....GA.FE.T.DLTVNGLK.QAA.T.N
WP_003687310/1-624 358 .PML..VLGA.GSFF.AMR....L..PF.M.....A.Q.YDR...L..GLT.R.S
tr|F4SV41|Escherichia_coli/1-609 345 LSAC...M...FVML.TRGISFRF..SD.AL.....KQ.VV.E.Y.R...MD.NSS.W.R
tr|A0A198GJ22|Enterobacter/1-609 346 ITIC..AF...VVVQ.THGATFRF..KG.DM.....ND.VI.T.Y.R...FN.NKE.W.R
tr|A0A0H3SVW1|Photobacterium/1-629 337 L..I..LGFACIVSS.NIGLFYNI..PK.M....K.....LKAIVTNT.E.T...DQ..NG.K.Y
tr|A0A090IBV6|Aliivibrio/1-617 337 LII...SSSL.VLFL.KSNFVD.Y..PE.S....M.....LG.H.N...IN..DN.G..
tr|A0A038H4R1|Burkholderia/1-644 355 LLLA..VGGG.SIYA.SQGVPNRF..DR.S....V.....FIADS.E.T...YD..TN.P.H
tr|B1KL29|Shewanella/1-652 346 ITL...VAGI.TIYQ.LHGITSEI..RPYS....T.....TAQAKFL.E.E...YE..HF.P.L
tr|A0A0W0YMF6|Legionella/1-623 350 ILSG..SLGF.ITYK.YEGFHFRF..PL.LM.....NE.IN.N.YAD...YDYRTV.Y.R
tr|A0A0R3MIT7|Bradyrhizobium/1-625 350 LTVG..MVGI.VTAT.ASGFGFRF..PP.EI.....RD.IA.Q.L.A...PHGNAG.F.R
tr|A0A0J1LHQ0|Citrobacter/1-669 364 LSTG..GLA..AGAR.YQQVDGRI..NP.A....I.....EL.AM.I.P.K...LD..YA.V.R
tr|J3HA72|Pantoea_/1-633 346 LLFG...V...LIVA.KAGFPSRA..PQ.V....V.....AS.IK.Q.Y.S...LE..RF.V.R
tr|A0A075P176|Alteromonas/1-613 346 PVIA..V.CA.FIIA.NNGVATRF..SG.E....Q.....RELNLQA.....LEAI.G.D.F
tr|A0A0Q0MU03|Pseudoalteromonas/1-645 352 C..V...LSSS..IYI.TNGISGHY..DE.N....V.....LIADR.E.S...SN..KN....
tr|A0A077XXI0|Sphingobacterium/1-616 349 TIVA..A.VM.MLFP.YD..QLKI..NK.E....I.....EHLTNYN.T.N.YRKQHLPF.Q.F
tr|J2PSI9|Caulobacter_/1-623 356 VATALVGLGM.LGQF.DRGLEYRY..SP.QA.....RK.IG.E.MLA...QTDHKG.F.R
tr|A0A0F5ZZM0|Grimontia/1-627 345 IFLG..AVSS.ISFV.YNGFEHRM..PE.D....F.....KV...IA.S.N...AK..PS.P.Y
tr|A0A072D806|Acinetobacter/1-635 350 ISII..GLGL.IIFK.TNGGNSSI..RG.AA.NTP.....QALLI.E.K.Y...IN..EH.K.N
tr|K0E9E8|Alteromonas/1-638 348 LIAL..I.GS.LSFF.TNGFIQRA..PV.A....Y.....QNLVE.D.....VE..AS.P.F
tr|A0A1G3I0Z5|Rhodoferax/1-663 361 VVVA..APGA.LAQW.KQGLPGRF..SP.K....V.....EL.VS.Q.E.A...LN..RN.P.R
tr|A0A069PRQ1|Paraburkholderia/1-717 353 FTVL..FACA.AVTL.GNGYGSRL..PQ.QV.....RQ.IS.A.A.S...LD..VD.P.R
tr|A0A0P0QBN4|Serratia/1-656 356 MLVA..ALC..VTLF.TVNFSGRL..PQ.Q....I.....DL.IA.N.E.S...TN..ML.A.N
tr|A0A085VIZ6|Pseudomonas/1-631 348 MVF...VAAG.AVSA.TEGAVTPL..RAAS....I.....SDKALFV.Q.D...YA.NRY.KNM
tr|E6VIY1|Rhodopseudomonas/1-653 365 AVVG..LIGA.GIFV.AGGVPSRF..SA.GV.....QI.VT.R.DHQ...AEAMPA.F.R
tr|A0A1B7ICK8|Buttiauxella/1-670 364 MCAV..LLA..VAAR.YQQLENRI..NP.A....I.....EM.AM.T.P.K...LN..NA.S.R
tr|A0A1E3ZLQ3|Bordetella_/1-665 359 LCLA..AASI.YTAN.QEGFPARL..AG.R....I.....SP.EK.A.V.L...MS..MPLP.P
tr|R4YTI5|Oleispira/1-645 349 FIAF..TAAS.TIQH.YQGLPERF..SP.E....L.....QLSFS.Q.....IA..SS.P.L
tr|E1SVP6|Ferrimonas/1-652 347 LIV...ALVW.MVNM.PKDAGAYL..PE.S....V.....A.S.S...MQ..V....K
tr|A0A1E4L6G6|Thiobacillus/1-641 354 MALALI.VGT.AGIF.SQGFSYRY..PN.FV.....RQ.K.....IS...GE..ER.Y.N
tr|A0A0T9N312|Yersinia/1-654 363 LVVG..VLA..LTVR.HSTLDRDI..MA.DK.ETV.....EL.YAKI.Q.S...FH..VM.P.N
tr|A0A085THR2|Vibrio/1-648 378 TCIS..LISI.YSMS.NNGLASRF..PE.IV.....QR.II.G.YDG...PDTTSG.W.R
tr|A0A167CIA4|Achromobacter/1-636 360 VAVA..LTGF.QAVA.SQGIPSRL..PD.AL.....QA.VA.N..YR...YAYEAD.A.R
tr|A0A150FIY7|Leptospira/1-667 377 LSFL..SLSV.IIYY.KNGFIDNY..PK.IV.....FE.LD.K.YKN...YETKEF.F.R
tr|W3RFJ7|Afipia/1-681 386 AAAA..CFGI.AVWA.SAGMPGRF..PE.GL.....QH.LL.R.DFE...PETRFA.Y.R
tr|A0A1E4CSA7|Hyphomicrobium/1-660 370 CSLA..FMGI.TLDI.GNGFPWLQ..SK.AV.....LA.VV.D.D.E...GD..RS.P.L

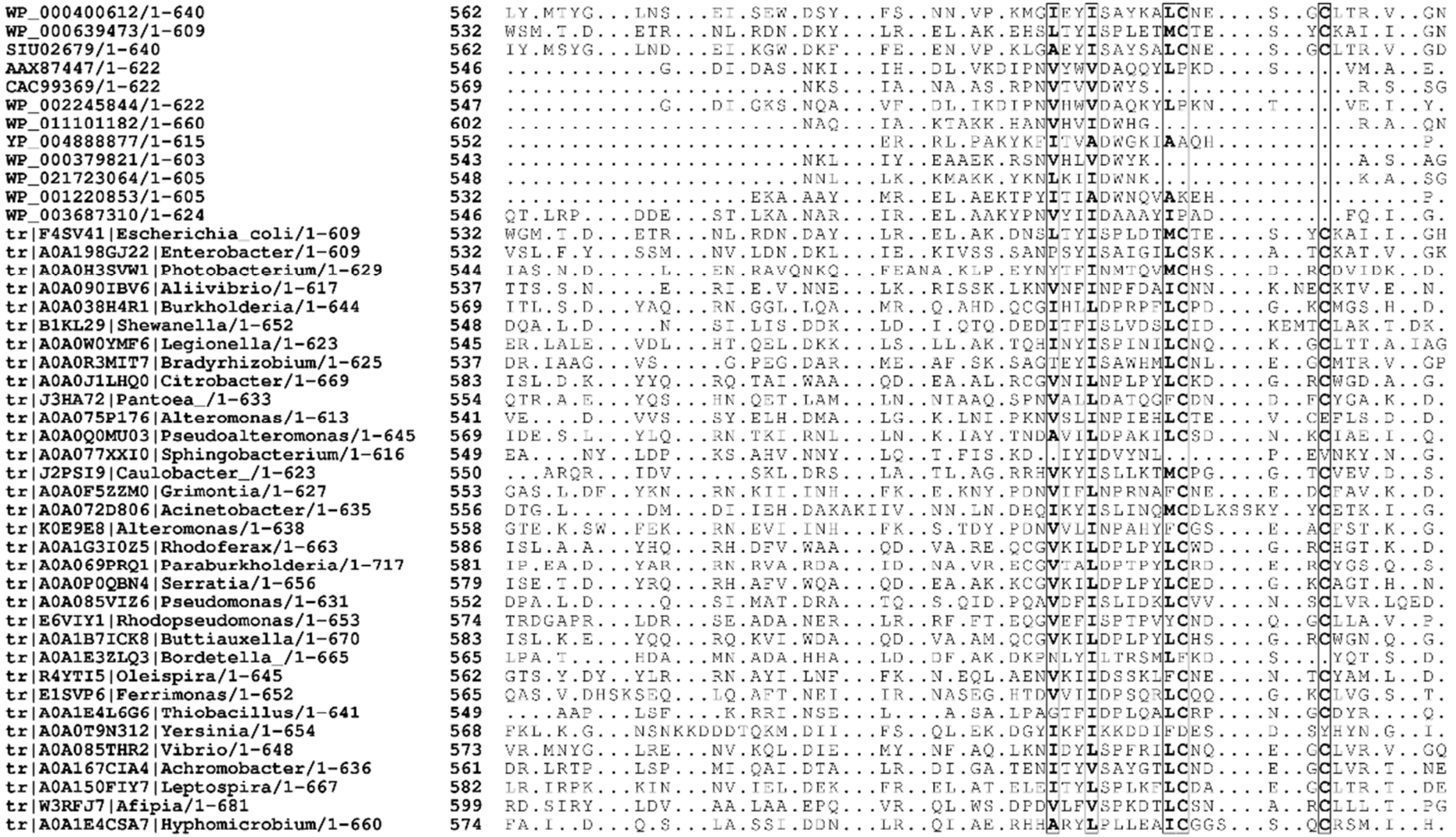

WP_000400612/1-640 400 .....GG....ICH......SVQL.......T.....
WP_000639473/1-609 380 .....PD....IC......FLNP.......DQ.....D.
SIU02679/1-640 400 .....GG....ICH......SVLL.......K.....
AAX87447/1-622 384 ......E....KEH......IE.....
CAC99369/1-622 433 .....KE....DK......ATKDK......EADSNKAS.....EQKETQK.
WP_002245844/1-622 385 ..........EH......LRPL.....
WP_011101182/1-660 444 .....KN....AA......ALKRQ......KA.AQAAAASSKKVATEKMQTKQAEAKLNSKQKQVEKEYDLK.
YP_004888877/1-615 418 .....IG....TTHEAVLAAVTP.....KK.....
WP_000379821/1-603 421 .....KK....KV......VKKDK......QE.....D.
WP_021723064/1-605 430 .....SL....KA......V...A......EK.....Y.
WP_001220853/1-605 411 .....IG....QTK.....VM.A.....ER.....
WP_003687310/1-624 387 .....NT....SCH......NN.....
tr|F4SV41|Escherichia_coli/1-609 380 .....PD....TC......FLNP.......EQ.....D.
tr|A0A198GJ22|Enterobacter/1-609 382 .....PD....TC......FLNP.......RQ.....D.
tr|A0A0H3SVW1|Photobacterium/1-629 376 .....TW....DK......IRH.....
tr|A0A090IBV6|Aliivibrio/1-617 368 ......N....AK......FIGL.....SQ.....
tr|A0A038H4R1|Burkholderia/1-644 392 .....RT....KC......HMM.....
tr|B1KL29|Shewanella/1-652 385 .QGAYLD....QCN.AYHSLKST.....GE.....
tr|A0A0W0YMF6|Legionella/1-623 390 .....EG....SCF......LKPN.....Q.....NY
tr|A0A0R3MIT7|Bradyrhizobium/1-625 389 ......D....KCF......LDVP.....G.....
tr|A0A0J1LHQ0|Citrobacter/1-669 400 .....FE....KC......LLAP.....GK.....G.
tr|J3HA72|Pantoea_/1-633 380 .....GE....NC......FVFS.....G.
tr|A0A075P176|Alteromonas/1-613 383 .D.....YPEP.....NL
tr|A0A0Q0MU03|Pseudoalteromonas/1-645 384 .....PF....KC......MVE.....
tr|A0A077XXI0|Sphingobacterium/1-616 389 .R...KG....IC......HLDI.....DN.....TF
tr|J2PSI9|Caulobacter_/1-623 398 .....EG....DCF......IASA.....Y.....KV
tr|A0A0F5ZZM0|Grimontia/1-627 381 .....RE....TC......HIEK.....YQ.....
tr|A0A072D806|Acinetobacter/1-635 391 L....DDAYWLQC......DAYSN....LTKY.....G.
tr|K0E9E8|Alteromonas/1-638 383 .....RN....KC......HTNK.....YR.....
tr|A0A1G3I0Z5|Rhodoferax/1-663 398 .....HK....NC......HPAS.....G.
tr|A0A069PRQ1|Paraburkholderia/1-717 390 .....RN....ECL......IDSV.....H.....RL
tr|A0A0P0QBN4|Serratia/1-656 392 .....RD....EC......FAKS.....G.
tr|A0A085VIZ6|Pseudomonas/1-631 389 .YEHYWL....KCD.AYSAFAER.....HQ.....
tr|E6VIY1|Rhodopseudomonas/1-653 405 .....LG....SCF......LGS.....G.....
tr|A0A1B7ICK8|Buttiauxella/1-670 400 .....MQ....DC......LLTP.....GE.....G.
tr|A0A1E3ZLQ3|Bordetella_/1-665 397 .....NG....WC......FYSLDAVKTL.EV.....G.
tr|R4YTI5|Oleispira/1-645 385 .....RD....KC......HSKK.....KD.....Y
tr|E1SVP6|Ferrimonas/1-652 377 .....PY....DC......FDRA.....VK.....S.
tr|A0A1E4L6G6|Thiobacillus/1-641 390 .....YH....TCF......LDEG.....Q.....TF
tr|A0A0T9N312|Yersinia/1-654 403 RD...NG....YC......FYNVD.....GE.....SD
tr|A0A085THR2|Vibrio/1-648 418 .....VG....ECH......LLPK.....Q.....GY
tr|A0A167CIA4|Achromobacter/1-636 399 .....YP....ACW......LTAK.....D.....SG
tr|A0A150FIY7|Leptospira/1-667 417 .....TG....KCL......LEDQ.....E.....S.
tr|W3RFJ7|Afipia/1-681 426 .....SG....SCF......LEDT.....Q.....DS
tr|A0A1E4CSA7|Hyphomicrobium/1-660 407 .....RR....RCH......IARA.....D.....QG

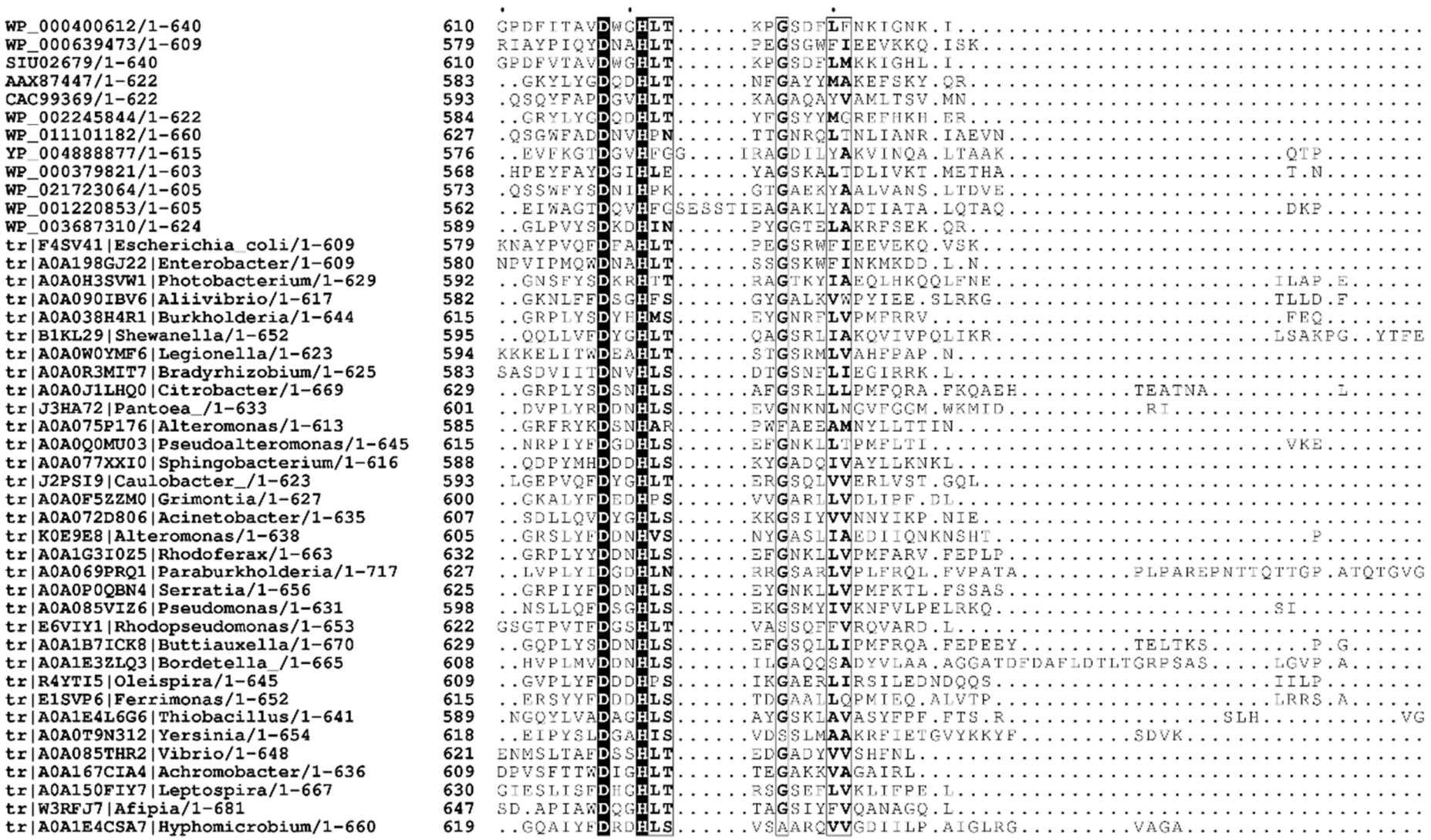

WP_000400612/1-640 410 .....A.AI.SN.....GCIK....N.GKHNIFIIGDSYAAALFNGLSHYIDNKGSDYII.S.QMTDGNA...
WP_000639473/1-609 391 ....YS.A..FS.....KCQD.K.M..TEKSFVVWGDSHAAHLMPGLKSVFGN...SLNI.T.QRTASLC...
SIU02679/1-640 410 .....D.AI.SN.....GCIK....N.SRNNIFIIGDSYAAALYNGLSSYIKNNNEKYVI.S.QMTDGNA...
AAX87447/1-622 390 .........Q.S......IPVS.N.ENHYPAKVIILGDSHSSHLEAFLNYVGNK.E.GWKA.DIFKDKFEC...
CAC99369/1-622 457 ..........PDNKNKSAAT.P..KTIITQTVAIGDSVMLDIEPYLKEAVPN.I.TIDGLV.GRQLRDA...
WP_002245844/1-622 391 .........P.G......APLA.A.ENHFPETVLTLGDSHAGHLRGFLDYVGSR.E.GWKA.KILSLDSEC...
WP_011101182/1-660 494 ....PQVV.....|....L..AMANTDLTAIGDSVLLDVSSDLQDVIPG.T.VVQGRV.GRQVTEV...
YP_004888877/1-615 434 .....K.Q..AT.....P.KG.D..QEKAPGVSIIGDSVTLGTRSYLGDHVAN.S.NIDA.E..GD.....|...
WP_000379821/1-603 433 .....K.Q..TANSK.....|....E..DIKKSSPLLIGDSVMVDIGNVFTKKIPN.A.QIDGKV.GRQLVDA...
WP_021723064/1-605 439 .....K.V..QPVVA.....|....E..KASQMNVLALGDSVMVAASTNLQEVFPH.M.YIDAAV.GRQAESL...
WP_001220853/1-605 421 ..........A.....D.AN.S..LGIADGTMLIGDSVALRANTALQTALPG.A.QINA.Q..VS.....|...
WP_003687310/1-624 394 .........T.GK.....QCLW.G.DTEKQPELLVLGDSHADHYKTFFDAVGKK.E.KWSA.T.MVSADAC...
tr|F4SV41|Escherichia_coli/1-609 391 ....YT.A..FS.....KCQD.K.M..TSKSFVVWGDSHAAHLMPGLRSVFGN...DLNI.T.QRSASLC...
tr|A0A198GJ22|Enterobacter/1-609 393 ....YS.E..FS.....KCED.K.M..TSDSIVIWGDSHAAQLMPGFRSVFKN....ENL.T.QRTSSLC...
tr|A0A0H3SVW1|Photobacterium/1-629 383 .........LD......QKDFA...GKNKEKVLIIGDSQAGDFTNSLYDAGLN.K.NVDI.VSRVVEADCGFF
tr|A0A090IBV6|Aliivibrio/1-617 377 ........D..HF.....INTN.N...IENVDFIGIGDSNLVHLSFGIKNSGST.....RV.L.FSGAGSC...
tr|A0A038H4R1|Burkholderia/1-644 399 .........ID......PCDVAN..KQKDSRVVVLGDSHAEAIAEAVVDAVPNGK.PED.V.L.LITIEGC...
tr|B1KL29|Shewanella/1-652 404 ...A...K.IAA.....SCNH.S..DGGTGGVMLWGDSHAQALSLGLRETLPD.T..INF.Y.QVASSGC...
tr|A0A0W0YMF6|Legionella/1-623 402 .....HAFK.TK.....SCTI.G..NSTYKSVFLWGDSHAAHLYPGLEEQLKM...TRSI.T.QLTASAC...
tr|A0A0R3MIT7|Bradyrhizobium/1-625 398 .....S.EF.NS.....SCIE....QGNKPLVHLWGDSTAAALYPGLKNAQAT.V.PFRL.A.RFTASAC...
tr|A0A0J1LHQ0|Citrobacter/1-669 411 ....SE....SP.....MCRY.G.K..GELTAIVLGDSHASMLVSAVAEAAPG.....AV.V.EMTYASC...
tr|J3HA72|Pantoea_/1-633 389 ....TT....SP.....QCVF.G.N.ASQVNLVVLGDSHAAAMLSSIVGSARA...DDSV.V.FIAQSGC...
tr|A0A075P176|Alteromonas/1-613 390 ...RV..A.GHD.....VRFI.E..GTSQDNILFIGASHIEHTYPFVKKY.ND...FYNV.Y.YLTQGGC...
tr|A0A0Q0MU03|Pseudoalteromonas/1-645 391 .........NKF.....PCYIG...NKENIKAIIIGDSHADALTTSLASVFNL.N.KEGI.I.ALTKSAC...
tr|A0A077XXI0|Sphingobacterium/1-616 402 ...SQ..Y.NFD.....ICAN.I..APNKKNILLLGDSHAGVFAQSLKEQLAK.Q.DINL.L.QATVSTT...
tr|J2PSI9|Caulobacter_/1-623 410 .....E.DF.KD.....YCLS.E..KPSNPDYLLIGDSHAAHLWWGLHTALPT....VNV.M.QATASGC...
tr|A0A0F5ZZM0|Grimontia/1-627 391 .......K.PEL.....SCEY.F...KDSVQWAILGDSHTVEIAYALAEKLKE.Y.QKGL.K.HFSFSGC...
tr|A0A072D806|Acinetobacter/1-635 410 ....FN.G.IDK.....SCIN.Q.P.LSTKSILLWGDSHSQALSLGLRTTLND....YSF.Y.QIGSSGC...
tr|K0E9E8|Alteromonas/1-638 393 .......D.PSE.....ACEY.S..VKGNTTWATVGDSHSVELAYALSQKLKA.T.DEGL.K.HFSFSGC...
tr|A0A1G3I0Z5|Rhodoferax/1-663 407 ....IA....SP.....SCMF.G.G..DRLRAILLGDSHADAVVSALAAAAPQ.Q.GDGV.M.EWSYTGC...
tr|A0A069PRQ1|Paraburkholderia/1-717 402 ...DDP.DR.EI.....GCRY.G..KSATIGAILWGDSHGNAVITGVAAAIEE.T.HRSV.M.FFGTSGC...
tr|A0A0P0QBN4|Serratia/1-656 401 ....FS....SP.....SCIY.G.G..ENIRVIIVGDSHGDAMISAVESALPN.Q.TDGL.L.NLTYAGC...
tr|A0A085VIZ6|Pseudomonas/1-631 408 ...S...A.IDP.....SCTQ.Q...RGHGGVFLWGDSHAQALSLGLRTQLAA.D..TPF.Y.QVASAGC...
tr|E6VIY1|Rhodopseudomonas/1-653 414 .....S.SF.AR.....ECND.A.PSPGVPRVALWGDSHGAHLYPGLRAVQQSDG.GFSL.S.QYTTAGC...
tr|A0A1B7ICK8|Buttiauxella/1-670 411 ....SE....SP.....KCSY.G.E..GELTAIVLGDSHAGMVASVIADVAPG.....SV.I.EMDYASC...
tr|A0A1E3ZLQ3|Bordetella_/1-665 414 ....EQ....GL.....DCHV.G.ETGAPKTALLFGDSFAGHNLPFWDKVGKH.F.HLNV.H.VVATSWC...
tr|R4YTI5|Oleispira/1-645 396 ...I...P.PKQ.....ACTY.F...GENIKWAVLGDSHTIEIAYALAKAVEE.N.NEGI.K.HYSFSGC...
tr|E1SVP6|Ferrimonas/1-652 388 ...D...E.EQA.....VCSL.G..SGRGERVLALGDSHMFSILPALESMARR.H.NVVL.E.YAGYSGC...
tr|A0A1E4L6G6|Thiobacillus/1-641 402 .....Q.DWHGK.....NCFL.T..KGRGPTVLLWGDSFAAHYAPGIVDQTQY.I.TVDY.L.QYTASAC...
tr|A0A0T9N312|Yersinia/1-654 418 PIISME....KS.....VCKL.G.VKSLKPKGLLFGDSFAGHYEPFVDEVAKK.L.GISV.D.SVTTNWC...
tr|A0A085THR2|Vibrio/1-648 430 .....S.GF.G.....SCDS.EAKSHEKPTLLLWGDSHAAHLYQGLIAHYSQK...YRI.I.QKTASTC...
tr|A0A167CIA4|Achromobacter/1-636 411 .....S.AF.AP.....ECFE.NLGVHGTKSLIVWGDSYAARLYPGIATAFGDR...YSI.A.QTTRNSC...
tr|A0A150FIY7|Leptospira/1-667 428 .....P.IF.SK.....DCLD.G..E.GRELLFIWGDSYAAHLVHGLNNLKAQ.K.KFRV.A.QYSSAAC...
tr|W3RFJ7|Afipia/1-681 438 .....T.SF.AE.....DCLD.Q.DPAAAPLVLLWGDSYAAHLFPGLAEHRRRTM.SMRL.A.EYTASGC...
tr|A0A1E4CSA7|Hyphomicrobium/1-660 419 ....RR.AL.AD.....TCVF.G..SASGQHVVVLGDSHGAELSYALSEVANE.G.LLQL.R.QVTASGC...

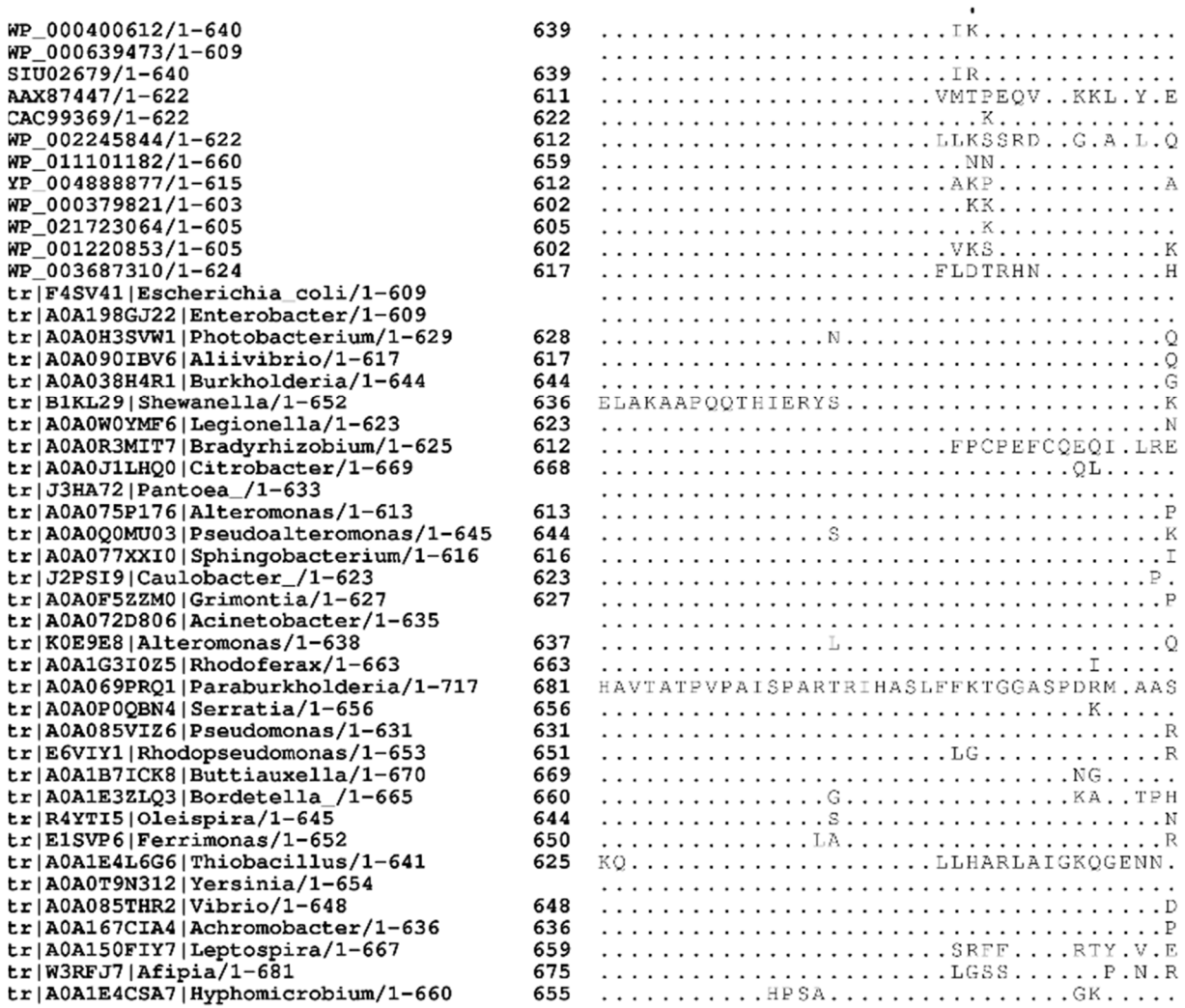

WP_000400612/1-640 461 .....PPLFVDGKDD.....LQRSV.ITLNNNR.....I.NFI.KRV..QP.E.VVLL.T.
WP_000639473/1-609 440 .....PPIIGLQ.KD.....DRPYC.KDINDMV.....A.KEI.SDN..KP.T.TVLM.SA
SIU02679/1-640 461 .....PPLFVSGKDD.....LQRDV.GSINADR.....I.KEI.GIV..KP.E.IVLL.T.
AAX87447/1-622 440 .....SFIVNEQ.YQ.....LDPNC.QSVWQKD.....S.....QY.K.AIFI.SA
CAC99369/1-622 510 .....ITTA.TG.YK.....K.FNS..EN.S.SVIL.EL
WP_002245844/1-622 441 .....LVWVDEK.LA.....DNPLC.RKYRDEV.....E.....KA.E.AVFI.AQ
WP_011101182/1-660 541 .....PGII.NS.LK.....S.QGQ..LA.H.NVLL.NI
YP_004888877/1-615 478 .....RTM.NLAYKVMM.....NQQ.RSH.TLR.Q.YVVI.CI
WP_000379821/1-603 483 .....TPIVKSQ.YK.....D.YAK..KG.Q.KVVV.EL
WP_021723064/1-605 489 .....DTDL.TN.AK.....S.KMP..NP.D.AYLI.GL
WP_001220853/1-605 462 .....VTT.KTANEIML.....NNS.QNK.FLP.K.TVVI.AT
WP_003687310/1-624 444 .....AYVEGYA.SRV...FQNWAAC.RAVYRYA.....E.EHL.P....RY.P.KVVL.AM
tr|F4SV41|Escherichia_coli/1-609 440 .....PPIIGLQ.ID.....GRPHC.KAINDMV.....A.REI.SDN..KP.N.TVLM.SA
tr|A0A198GJ22|Enterobacter/1-609 441 .....APLYGAE.NI.....QRPFC.KDINKNI.....T.AEL.LRV..KP.K.LVVL.AA
tr|A0A0H3SVW1|Photobacterium/1-629 435 YLPPAEQNIAFSQSPTIKG.N.QM.....WQNKC.RTFIKNV.....D.KGD.VV..EQAN..VIIL.AM
tr|A0A090IBV6|Aliivibrio/1-617 422 .....LVFPDYT.....RKPYA.SWMNEEWLKRCSTIY.KIA.LD..NPN.K.DIIW.SQ
tr|A0A038H4R1|Burkholderia/1-644 449 .....PTIYGIR.R.....GNGDC.FDQDRKY.....I.DIL.SRD.ESAKT.PVII.AN
tr|B1KL29|Shewanella/1-652 454 .....KPSLKQS.DDQ...SSLAQAC.NTSNQIA.....L.ETI.SR..IKP.D.IVII.AQ
tr|A0A0W0YMF6|Legionella/1-623 453 .....PPILGMD.KS.....DRPHC.KKINDFV.....F.QRI.KEE..KP.N.EVIL.AA
tr|A0A0R3MIT7|Bradyrhizobium/1-625 448 .....PPILAAG.SD.....ARC.DAANQIV.....F.DLI.KSS..HP.E.IVLL.HA
tr|A0A0J1LHQ0|Citrobacter/1-669 457 .....PTVLGMK.RR....EDINDDC.KLFNDRA.....I.NLL.NNE.YKN.K.LVFI.AN
tr|J3HA72|Pantoea_/1-633 438 .....PSVPEIH.RS.....ASPDC.GDFVKNA.....I.ADI.ERK.YPH.A.SVLI.IN
tr|A0A075P176|Alteromonas/1-613 440 .....FITPSMK.N.....PKWSC.KNIQNY.Q.....SLL.EEVV..F.K.KVVI.SF
tr|A0A0Q0MU03|Pseudoalteromonas/1-645 440 .....PYIINAL.STK.....SGNEC.FIENQKR.....E.AFL.LN..NYENI.PVIW.AA
tr|A0A077XXI0|Sphingobacterium/1-616 454 .....YPLLDTK.G.....PKASC.ELIDYMYQ.....KFI.PQNAAHI.D.EVIL.VG
tr|J2PSI9|Caulobacter_/1-623 459 .....KPTIDQA.SS.....AERRC.VDLMDRI.....FKQYL.QDH..QV.D.TLVI.AA
tr|A0A0F5ZZM0|Grimontia/1-627 440 .....IPSYTQS.D.....DYSRCSKWYNESV.....H.HII.ND..NHI.K.NVVF.NH
tr|A0A072D806|Acinetobacter/1-635 460 .....KASLTPT.ETL..KGALKQAC.DKVNILA.....L.EEI.KRL..KP.S.FVII.AQ
tr|K0E9E8|Alteromonas/1-638 443 .....RPSYGKN.D.....DYSKC.AKWYNSSI.....EFI.ANE.PMI.Q.NVVF.VH
tr|A0A1G3I0Z5|Rhodoferax/1-663 456 .....PTLFGVL.PAPHRTLAAESQC.GRFLDWA.....K.DKL.VDV.PKD.V.PLVI.VN
tr|A0A069PRQ1|Paraburkholderia/1-717 455 .....PPLIGAS.RFG...KHREEPC.RRFAARV.....A.SEI.AAY..PA.DVPLVV.VA
tr|A0A0P0QBN4|Serratia/1-656 450 .....PTLFGAN.VVP.GALDDNIQC.SAFNQWI.....K.EKL.STL.DNN.I.PAVV.IN
tr|A0A085VIZ6|Pseudomonas/1-631 457 .....KPSLTQG.AGRD..TIPVQSC.NYSNRTA.....L.ENI.QR..LRP.D.VVII.AQ
tr|E6VIY1|Rhodopseudomonas/1-653 467 .....PPIFGYV.ST.....QTPGC.TAANDNA.....R.ERL.KAL..KP.D.TVVLVAR
tr|A0A1B7ICK8|Buttiauxella/1-670 457 .....PTVLDFK.RR....EDINDDC.KTFNDKA.....I.QLL.NHE.YKN.K.LVFI.AN
tr|A0A1E3ZLQ3|Bordetella_/1-665 465 .....FPTLGKA.FIGPTSHKAYEQC.LLDRAYV.....K.DNF.....ARY.D.YILL.AG
tr|R4YTI5|Oleispira/1-645 446 .....VPSYKQD.K.....HFSRC.STWINDAV.....ADI.THT.KSI.E.NVVI.NY
tr|E1SVP6|Ferrimonas/1-652 439 .....PPILTVH.TERS..DQHKKNC.KALNDRT.....L.AYA.IE..QKV.D.TVFL.AA
tr|A0A1E4L6G6|Thiobacillus/1-641 454 .....PPVFDYY.TA.....VRPNC.RAFNDNL.....P.KLL.SQY..GI.S.TVVM.AG
tr|A0A0T9N312|Yersinia/1-654 473 .....FPSLTDS.TNGTKTRVAYKQC.LLNREYL.....K.DNI.....SKY.D.FVIF.SG
tr|A0A085THR2|Vibrio/1-648 481 .....PPIGNMD.IP.....TSPNC.EGINAEM.....R.QYI.ETE..KP.D.RIIL.AA
tr|A0A167CIA4|Achromobacter/1-636 463 .....PPLLDNQ.....YDTC.RDSNEFT.....F.SLI.KKH..PN.A.TVVM.FA
tr|A0A150FIY7|Leptospira/1-667 478 .....PPILNFE.SN.....RHSFC.KSINQFI.....G.KNLVLKL..NP.S.VIVL.SA
tr|W3RFJ7|Afipia/1-681 491 .....PPILGFS.NA.....QRPNC.AEVNDDI.....A.GRI.ATL..KP.E.TVIL.AG
tr|A0A1E4CSA7|Hyphomicrobium/1-660 471 .....PPALGFT.VD.....DHPKC.ARHTQNM.....V.DSL.ADG..PR.S.TILI.TA


WP_000400612/1-640 498 ..WS..V.RGTNGVH.....DKK.....LAIDA
WP_000639473/1-609 477 .LWP..V.YPM.....RDY
SIU02679/1-640 498 ..WS..V.RGSNGVE.....DKK.....LAIEA
AAX87447/1-622 471 .FYD..L.RMGGQ.PVP.....RFRPETF...IEP....DFKAR
CAC99369/1-622 531 GTNGP...FT.....EDQ
WP_002245844/1-622 472 .FYD..L.RMGGQ.PVP.....RFEAQSF...LIP....GFPAR
WP_011101182/1-660 562 GTNGT...IT.....DDQ
YP_004888877/1-615 505 GTNA..L.DD.....YEEQ
WP_000379821/1-603 505 GTNGA...FT.....KDQ
WP_021723064/1-605 510 GTNGT...IK.....EGE
WP_001220853/1-605 489 GVNN..P.EN.....YKDD
WP_003687310/1-624 482 .RWG..S.QMPEN.S.R.....SLA...YDA....GFFQK
tr|F4SV41|Escherichia_coli/1-609 477 .LWS..V.YPM.....RDY
tr|A0A198GJ22|Enterobacter/1-609 478 .FWS..Q.YPF.....KEY
tr|A0A0H3SVW1|Photobacterium/1-629 485 .YWR..D.VN.....MNW
tr|A0A090IBV6|Aliivibrio/1-617 464 .SWGPTE.VICTS.S.SC.....TTNRPIE.....D..Y.....YL....VLEEQ
tr|A0A038H4R1|Burkholderia/1-644 487 .HWS..Q.YWSGP.T.MDEFA.F....V..NP.....EQPAQTPAPF...STE....LYREH
tr|B1KL29|Shewanella/1-652 494 .ANH....HEDT.....NWEQ
tr|A0A0W0YMF6|Legionella/1-623 490 .RWI..A.YDW.....QK
tr|A0A0R3MIT7|Bradyrhizobium/1-625 483 .WWD..G.LKD.....LGK
tr|A0A0J1LHQ0|Citrobacter/1-669 497 .RSS..L.AILGQ.N.EKDDF.FN.....IPMG.....F..FDTPNEHPNQN...LTQ....QFSQQ
tr|J3HA72|Pantoea_/1-633 476 .RLS..F.YLHGD.Y.GSGS...N.....TPKY.....S..FI..N...DVS...SIA....AYQQH
tr|A0A075P176|Alteromonas/1-613 475 YSFD..S.YLPKD.....EE...VSVKEVQQRIED
tr|A0A0Q0MU03|Pseudoalteromonas/1-645 479 .RTG..V.YIYGQ.S.NKKRI.NNKLDT..QPSI.....F..FKEKYNIPNED...FFI....ELEEN
tr|A0A077XXI0|Sphingobacterium/1-616 492 .FWG..S.QQYPD.....ET...LK.....FK
tr|J2PSI9|Caulobacter_/1-623 497 .RWQ..A.SDL.....PR
tr|A0A0F5ZZM0|Grimontia/1-627 477 .RYSA...ALLGG.H.ES.....I..YPEI.....G..TL.....SD...KHE....DILSS
tr|A0A072D806|Acinetobacter/1-635 501 .AD...S.HDL.....SD
tr|K0E9E8|Alteromonas/1-638 480 .RYS..Y.QLVGG.S.NKT.....YPHL.....G..NL..K...FDN...KSI....ELLQN
tr|A0A1G3I0Z5|Rhodoferax/1-663 500 .RTS..L.YAMGH.N.EAGSKDLN.....TPLV.....Y..FTQPSDVASPA...FLA....EFAQR
tr|A0A069PRQ1|Paraburkholderia/1-717 496 .RFS..A.YVEGK.G.DSQD.....PTI.....LIGFD..GARPLTDIAQRRM....RYGRF
tr|A0A0P0QBN4|Serratia/1-656 493 .RTS..A.YAFGQ.Q.IAGAK.LN.....QPSV.....Y..FSKVYRTPEPG...FLS....EFKTS
tr|A0A085VIZ6|Pseudomonas/1-631 498 .KDQ....HDKT.....QWSE
tr|E6VIY1|Rhodopseudomonas/1-653 505 .QWH..D.YDGPD.R.....DPA.....AIDTM
tr|A0A1B7ICK8|Buttiauxella/1-670 497 .RSS..L.ALLGQ.N.EKDAF.FN.....IPMG.....Y..FDTPNDRPNPN...LNK....QFTAQ
tr|A0A1E3ZLQ3|Bordetella_/1-665 506 .NWGP.V.YKRG.....LMDS
tr|R4YTI5|Oleispira/1-645 483 .RYS..R.ALFGD.I.DQF.....YPAL.....G..DE..K...DEA...SRS....MMIQS
tr|E1SVP6|Ferrimonas/1-652 480 .RWTL...YTDGD.Y.LG.....KDVQFLKMSAQETG..SQ.....QR...SIE....VFRYG
tr|A0A1E4L6G6|Thiobacillus/1-641 491 .RWE..S.LFKRG.V.....SPQA
tr|A0A0T9N312|Yersinia/1-654 514 .IWFD.L.YRKG.....YQNE
tr|A0A085THR2|Vibrio/1-648 518 .AWD..V.YDW.....KK
tr|A0A167CIA4|Achromobacter/1-636 497 .VWD..Q.YAKDW.A.....APS.....PEREA
tr|A0A150FIY7|Leptospira/1-667 516 .YWP..A.YFFPQ.E.....ANF.....K....KDMIQ
tr|W3RFJ7|Afipia/1-681 528 .RWS..L.YDGRE.G.....WGS.....LDADQ
tr|A0A1E4CSA7|Hyphomicrobium/1-660 508 .HYF..EWGAPGR.P.HR.....DA.....FWLG


WP_000400612/1-640 516 LSLTIKKIK..EASPDSRIIFIG.PV.P..E.....WNA.NLVK.IISNYLSEF...K.....KTPP...
WP_000639473/1-609 487 LPETIKFLK..DNK.VKNIIIVG.PF.P..V.....WKK.TMID.TIEDMGINS...G.....R..TVP...
SIU02679/1-640 516 LSLTIKEIK..KVSPQSRLIVVG.PV.P..E.....WNA.NLVK.VISNYTSEF...K.....KTPP...
AAX87447/1-622 498 FKNTVKQLA..MQ...KPVYVFA.NN.S..S.....VSR.SPLRGYLLENYGLE...KY.....LTPIHRM.
CAC99369/1-622 541 LNDLLDQFD.....KATIYLVNTRV.P.RG.....WQS.DV.....
WP_002245844/1-622 499 FRETVKRIA..AV...KPVYVFA.NN.T..S.....ISR.SPLREEKLKRFAAN...QY.....LRPIQAM.
WP_011101182/1-660 572 AEQVVKLIG..K...DRQIFWVTAHV.PTQS.....WQN.QV.....
YP_004888877/1-615 516 TMKIIHDLE..P...GHKLILMT.PY.N..ARADADWNS.SKLA.VL.....
WP_000379821/1-603 515 LNELLDSFG.....KADIYLVSIRV.P.RD.....YEG.RI.....
WP_021723064/1-605 520 IDAAMKVAG.....NKPVYWINVHA.D.RV.....WAK.PN.....
WP_001220853/1-605 500 WDSIVKNLP..K...GHHMILVT.PY.E..G.....DKT.KETY.AIV.....
WP_003687310/1-624 504 FDRMLHKLS..SE..KQAVYLMA.DN.L..A.....SSY.NVQRAYILS.....SRIPGCR
tr|F4SV41|Escherichia_coli/1-609 487 LPGTIKFLK..DSG.VKNIILVG.PF.P..V.....WKK.TLID.TIEETGVNA...G.....R..TVP...
tr|A0A198GJ22|Enterobacter/1-609 488 LTDTLDMLN..VAG.I.RTVVIG.PF.P..Y.....WRD.HVPK.LIEENGLNP...Q.....G..TLP...
tr|A0A0H3SVW1|Photobacterium/1-629 494 IFKSIQNIRANNQ..HAKIYVIG.GK.....S.....FNT.PVST.LAYDAYKQQADLGH.....YAKMN...
tr|A0A090IBV6|Aliivibrio/1-617 494 ITK.MSQLV..GE..HRVINLIG.WS.P..A.....PRE.SLVT.CTKAS.LER...C.....SKDVF...
tr|A0A038H4R1|Burkholderia/1-644 524 LLSTVCRI...SK..IRPVYLVE.PI.P..E.....FDF.NVPL.MLAREKMKD...PH.....APDLS...
tr|B1KL29|Shewanella/1-652 505 IEAKLAQL...G...AKETVLIG.PV.P..Q.....WRP.SLPY.VIAERHWQE...K.....GQYIT...
tr|A0A0W0YMF6|Legionella/1-623 499 VAFTIKQLQ..RVG.ITKITLIG.PV.P..S.....WID.TLPT.LLLNNIRKN...RL.....RTIP...
tr|A0A0R3MIT7|Bradyrhizobium/1-625 493 LRETIDQLK..ALS.VKRIIILG.PV.P..V.....WKR.TLPH.TLVNFYRLR...HT.....IA...
tr|A0A0J1LHQ0|Citrobacter/1-669 538 LVKTLCSIE..N...SQRVFLVR.PI.P..E.....MAV.DVPN.TMARALMFG...K.....SDSTVS...
tr|J3HA72|Pantoea_/1-633 510 FGAAVKQLA..T...HRRVFIMT.PV.P..E.....FGY.DVIY.RMSRDAMRG...K.....A.LAVQ...
tr|A0A075P176|Alteromonas/1-613 499 YNDILSDIA..DQ..TEHVYLLL.GD.P..R.....GPE.FDPR.RTVRVGTY.....APIP...
tr|A0A0Q0MU03|Pseudoalteromonas/1-645 524 LNETISKL...RK..KHPIYIVQ.PT.P..E.....MRR.NIPK.AIAKNYLLK...NK.....STEFA...
tr|A0A077XXI0|Sphingobacterium/1-616 507 LQKVIQYFK..EK..NIVLKMIG.QT.P..S.....YSI.IFPN.ILALQTKN.....SAVK...
tr|J2PSI9|Caulobacter_/1-623 506 LSKTLDWAT..GKG.I.QVVLVG.PI.P..E.....YRS.RLPR.LLAIADRFH...D.....PDLV...
tr|A0A0F5ZZM0|Grimontia/1-627 507 IDNAIKDLA..NS..KDKVYVYL.PI.P..E.....IKK.PIGK.LLAKSWLNN...YG.....YENIS...
tr|A0A072D806|Acinetobacter/1-635 509 WKSITKKLE..SLG.VKKILIVG.AV.P..Q.....WRP.SLPK.VIVKDSHFD...S.....IVSKIN...
tr|K0E9E8|Alteromonas/1-638 512 IDRAIHKLA..SS..KRRVYVLY.PI.P..E.....LPS.HVTK.LIDSELSVN...GD.....LVSIE...
tr|A0A1G3I0Z5|Rhodoferax/1-663 542 LTDTACQLA..K...ARPVYLVR.PF.P..E.....MGV.NVPH.TARAMV.VR...G.....QREEVS...
tr|A0A069PRQ1|Paraburkholderia/1-717 536 VTNDLCALA...K..MRKVYVLL.PI.P..E.....MGR.DVPD.YLARSLILG...RH.....PADVS...
tr|A0A0P0QBN4|Serratia/1-656 534 LVDTLCEMA..T...ERQIYVVR.PV.P..E.....MLV.NVPK.TLSRLLSFG...G.....SEPDIS...
tr|A0A085VIZ6|Pseudomonas/1-631 509 IAARLKSY...G...VKHIVLLG.PL.P..Q.....WNP.SLPF.VIANRHWGQ...T.....ETHIT...
tr|E6VIY1|Rhodopseudomonas/1-653 523 IKATLAELR..SIG.IRRIVVVG.KF.P..S.....WRT.PPKR.ILAQAYRAE...AAGLIT..AAEIP...
tr|A0A1B7ICK8|Buttiauxella/1-670 538 LIKTMCAIE..N...SQRVFLVR.PI.P..E.....MAM.DVPN.TMARALMFG...N.....PARDIS...
tr|A0A1E3ZLQ3|Bordetella_/1-665 519 MLETLESMN..RL..DRKIVIME.AP.K..R.....FDK.NLRS.AYERSLLLG...L.....EFSADR...
tr|R4YTI5|Oleispira/1-645 515 LKMMIDDLA..LS..KKNVYVFL.PI.P..E.....LQE.RIPK.LLSNQYTYNL..NH.....FNNID...
tr|E1SVP6|Ferrimonas/1-652 518 IDKTLSAFR..EA..GIRVVLML.QA.P..L.....QND.IPPR.LYYQA.LEG...ED....VDADKLA...
tr|A0A1E4L6G6|Thiobacillus/1-641 505 VAATVKRLN..DLG.V.KTYVIG.QS.P..V.....FNN.DVQT.IFAQTGGLA...D.....TSEA...
tr|A0A0T9N312|Yersinia/1-654 527 IVDVIKYAK..SK..GVKVYVMA.SP.T..Q.....YDI.NVFA.NFIAAG.V.....NDLP...
tr|A0A085THR2|Vibrio/1-648 527 LSSTIDWLK..QID.VNQIDLVG.PV.P..R.....WLD.SLPR.QALVYYSNN.....K.KGDIP...
tr|A0A167CIA4|Achromobacter/1-636 515 LIRTIRSVK..AVG.A.QVILLG.PA.P..R.....WPE.NLPS.LVLKTWKNG.....SLSQGIP...
tr|A0A150FIY7|Leptospira/1-667 535 LKESISSLK..HSG.ISNIVLFG.PL.P..E.....WENGGLPD.LLIRRIVFK...KK.....MDIP...
tr|W3RFJ7|Afipia/1-681 546 IRQTIAALK..ASG.VQRIVVVG.QF.P..I.....WSL.PPQM.ILTRDYQID...MMAFRASPSRPLP...
tr|A0A1E4CSA7|Hyphomicrobium/1-660 527 IEKSVATLR..RS..GHDVILLG.GWPP..H.....TNG.PLPH.ALAREIRFG...RS.....IEDYS...

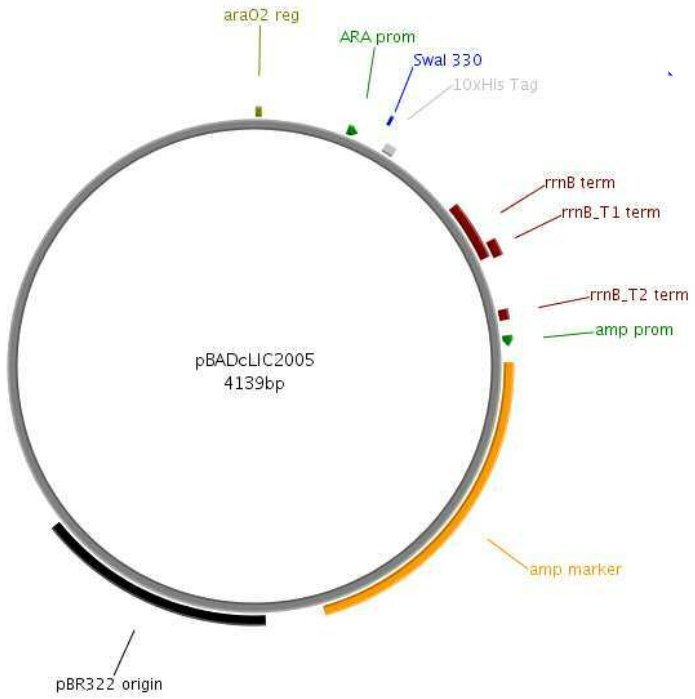

WP_000400612/1-640 562 LY.MTYG...LNS...EI.SEW.DSY...FS..NN.VP.KMGIFYISAYKALCNE....S..GCLTR.V..GN
WP_000639473/1-609 532 WSM.T.D...ETR...NL.RDN.DKY...LR..EL.AK.EHSLTYISPLETMCTE....S..YCKAI.I..GN
SIU02679/1-640 562 IY.MSYG...LND...EI.KGW.DKF...FE..EN.VP.KLGAEYISAYSALCNE....S..GCLTR.V..GD
AAX87447/1-622 546 ..........G...DI.DAS.NKI...IH..DL.VKDIPNVYWVDAQQYLPKD....S.....VM.A..E.
CAC99369/1-622 569 .....................NKS...IA..NA.AS.RPNVTVVDWYS.....................R.S..SG
WP_002245844/1-622 547 ..........G...DI.GKS.NQA...VF..DL.IKDIPNVHWVDAQKYLPKN....T.....VE.I..Y.
WP_011101182/1-660 602 .....................NAQ...IA..KTAKK.HANVHVIDWHG.....................R.A..QN
YP_004888877/1-615 552 ........................ER..RL.PAKYKFITVADWGKIAAQH.....................P.
WP_000379821/1-603 543 ....................NKL...IY..EAAEK.RSNVHLVDWYK.....................A.S..AG
WP_021723064/1-605 548 ....................NNL...LK..KMAKK.YKNLKIIDWNK.....................K.A..SG
WP_001220853/1-605 532 ...................EKA.AAY...MR..EL.AEKTPYITIADWNQVAKEH.....................P.
WP_003687310/1-624 546 QT.LRP....DDE...ST.LKA.NAR...IR..EL.AAKYPNVYIIDAAAYIPAD.........FQ.I..G.
tr|F4SV41|Escherichia_coli/1-609 532 WGM.T.D...ETR...NL.RDN.DAY...LR..EL.AK.DNSLTYISPLDTMCTE....S..YCKAI.I..GH
tr|A0A198GJ22|Enterobacter/1-609 532 VSL.F.Y...SSM...NV.LDN.DKL...IE..KIVSS.SANPSYISAIGILCSK....A..TCKAT.V..GK
tr|A0A0H3SVW1|Photobacterium/1-629 544 IAS.N.D.....L...EN.RAVQNKQ...FEANA.KLP.EYNYTFINMTQVMCHS....D..RCDVIDK..D.
tr|A0A090IBV6|Aliivibrio/1-617 537 TTS.S.N.....E...RI.E.V.NNE...LK..RISSK.LKNVNFINPFDAICNN....K.NECKTV.E..N.
tr|A0A038H4R1|Burkholderia/1-644 569 ITL.S.D...YAQ...RN.GGL.LQA...MR..Q.AHD.QCGIHLLDPRPFLCPD....G..KCMGS.H..D.
tr|B1KL29|Shewanella/1-652 548 DQA.L.D.....N...SI.LIS.DDK...LD..I.QTQ.DEDITFISLVDSLCID....KEMTCLAK.T.DK.
tr|A0A0W0YMF6|Legionella/1-623 545 ER.LALE...VDL...HT.QEL.DKK...LS..LL.AK.TQHINYISPINILCNQ....K..GCLTT.A.IAG
tr|A0A0R3MIT7|Bradyrhizobium/1-625 537 DR.IAAG...VS.....G.PEG.DAR...ME..AF.SK.SAGTEYISAWHMLCNL....E..GCMTR.V..GP
tr|A0A0J1LHQ0|Citrobacter/1-669 583 ISL.D.K...YYQ...RQ.TAI.WAA...QD..EA.AL.RCGVNILNPLPYLCKD....G..RCWGD.A..G.
tr|J3HA72|Pantoea_/1-633 554 QTR.A.E...YQS...HN.QET.LAM...LN..NIAAQ.SPNVALLDATQGFCDN....D..FCYGA.K..D.
tr|A0A075P176|Alteromonas/1-613 541 VE....D...VVS...SY.ELH.DMA...LG..K.LNI.PKNVSLINPIEHLCTE....V..CEFLS.D..D.
tr|A0A0Q0MU03|Pseudoalteromonas/1-645 569 IDE.S.L...YLQ...RN.TKI.RNL...LN..K.IAY.TNDAVILDPAKILCSD....N..KCIAE.I..Q.
tr|A0A077XXI0|Sphingobacterium/1-616 549 EA....NY..LDP...KS.AHV.NNY...LQ..T.FIS.KD..IYIDVYNL.........P..EVNKY.N..G.
tr|J2PSI9|Caulobacter_/1-623 550 ...ARQR...IDV.....SKL.DRS...LA..TL.AG.RRHVKYISLLKTMCPG....G..TCVEV.D..S.
tr|A0A0F5ZZM0|Grimontia/1-627 553 GAS.L.DF..YKN...RN.KII.INH...FK..E.KNY.PDNVIFLNPRNAFCNE....E..DCFAV.K..D.
tr|A0A072D806|Acinetobacter/1-635 556 DTG.L.....DM...DI.IEH.DAKAKIIV..NN.LN.DHQIKYISLINQMCDLKSSKY..YCETK.I..G.
tr|K0E9E8|Alteromonas/1-638 558 GTE.K.SW..FEK...RN.EVI.INH...FK..S.TDY.PDNVVLINPAHYFCGS....E..ACFST.K..G.
tr|A0A1G3I0Z5|Rhodoferax/1-663 586 ISL.A.A...YHQ...RH.DFV.WAA...QD..VA.RE.QCGVKILDPLPYLCWD....G..RCHGT.K..D.
tr|A0A069PRQ1|Paraburkholderia/1-717 581 IP.EA.D...YAR...RN.RVA.RDA...ID..NA.VR.ECGVTALDPTPYLCRD....E..RCYGS.Q..S.
tr|A0A0P0QBN4|Serratia/1-656 579 ISE.T.D...YRQ...RH.AFV.WQA...QD..EA.AK.KCGVKILDPLPYLCED....G..KCAGT.H..N.
tr|A0A085VIZ6|Pseudomonas/1-631 552 DPA.L.D.....Q...SI.MAT.DRA...TQ..S.QID.PQAVDFISLIDKLCVV....N..SCLVR.LQED.
tr|E6VIY1|Rhodopseudomonas/1-653 574 TRDGAPR...LDR...SE.ADA.NER...LR..RF.FT.EQGVEFISPTPVYCND....Q..GCLLA.V..P.
tr|A0A1B7ICK8|Buttiauxella/1-670 583 ISL.K.E...YQQ...RQ.KVI.WDA...QD..VA.AM.QCGVKILDPLPYLCHS....G..RCWGN.Q..G.
tr|A0A1E3ZLQ3|Bordetella_/1-665 565 LPA.T.....HDA...MN.ADA.HHA...LD..DF.AK.DKPNLYILTRSMLFKD....S.....YQT.S..D.
tr|R4YTI5|Oleispira/1-645 562 GTS.Y.DY..YLR...RN.AYI.LNF...FK..N.EQL.AENVKIIDSSKLFCNE....N..TCYAM.L..D.
tr|E1SVP6|Ferrimonas/1-652 565 QAS.V.DHSKSEQ...LQ.AFT.NEI...IR..NASEG.HTDVVIIDPSQRLCQQ....G..KCLVG.S..T.
tr|A0A1E4L6G6|Thiobacillus/1-641 549 ....AAP...LSF....K.RRI.NSE...L....A.SA.LPAGTFIDPLQALCRP....N..GCDYR....Q.
tr|A0A0T9N312|Yersinia/1-654 568 FKL.K.G...NSNKKDDDTQKM.DII...FS..QL.EK.DGYIKFIKKDDIFDES....D..SYHYN.G..I.
tr|A0A085THR2|Vibrio/1-648 573 VR.MNYG...LRE...NV.KQL.DIE...MY..NF.AQ.LKNIDYLSPFRILCNQ....E..CCLVR.V..GQ
tr|A0A167CIA4|Achromobacter/1-636 561 DR.LRTP...LSP...MI.QAI.DTA...LR..DI.GA.TENITYVSAYGTLCND....E..GCLVR.T..NE
tr|A0A150FIY7|Leptospira/1-667 582 LR.IRPK...KIN...NV.IEL.DEK...FR..EL.AT.ELEITYLSPLKFLCDA....E..GCLTR.T..DE
tr|W3RFJ7|Afipia/1-681 599 RD.SIRY...LDV...AA.LAA.EPQ...VR..QL.WS.DPDVLFVSPKDTLCSN....A..RCLLL.T..PG
tr|A0A1E4CSA7|Hyphomicrobium/1-660 574 FA.I..D...Q.S...LA.SSI.DDN...LR..QI.AE.RHHARYLPLLEAICGGS...S..QCRSM.I..H.

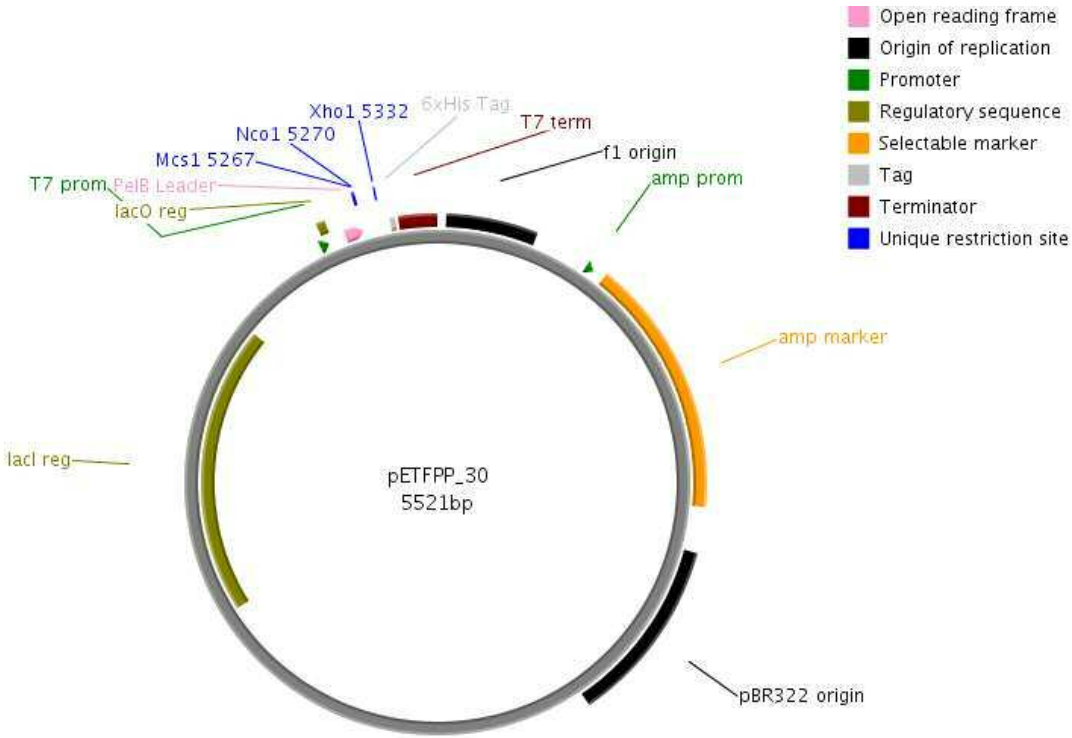

WP_000400612/1-640 610 GPDFITAVDWGHLT.....KPGSDFLFNKIGNK.I
WP_000639473/1-609 579 RIAYPIQYDNAHLT.....PEGSGWFIEEVKKQ.ISK
SIU02679/1-640 610 GPDFVTAVDWGHLT.....KPGSDFLMKKIGHL.I
AAX87447/1-622 583 ..GKYLYGDQDHLT.....NFGAYYMAKEFSKY.QR
CAC99369/1-622 593 .QSQYFAPDGVHLT.....KAGAQAYVAMLTSV.MN
WP_002245844/1-622 584 ..GRYLYGDQDHLT.....YFGSYYMGREFHKH.ER
WP_011101182/1-660 627 .QSGWFADDNVHPN.....TTGNRQLTNLIANR.IAEVN
YP_004888877/1-615 576 ..EVFKGTDGVHFGG....IRAGDILYAKVINQA.LTAAK.....QTP
WP_000379821/1-603 568 .HPEYFAYDGIHLE.....YAGSKALTDLIVKT.METHA.....T.N
WP_021723064/1-605 573 .QSSWFYSDNIHPK.....GTGAEKYAALVANS.LTDVE
WP_001220853/1-605 562 ..EIWAGTDQVHFGSESSTIEAGAKLYADTIATA.LQTAQ.....DKP
WP_003687310/1-624 589 ..GLPVYSDKDHIN.....PYGGTELAKRFSEK.QR
tr|F4SV41|Escherichia_coli/1-609 579 KNAYPVQFDFAHLT.....PEGSRWFIEEVEKQ.VSK
tr|A0A198GJ22|Enterobacter/1-609 580 NPVIPMQWDNAHLT.....SSGSKWFINKMKDD.L.N
tr|A0A0H3SVW1|Photobacterium/1-629 592 ..GNSFYSDKRHIT.....RAGTKYIAEQLHKQQLFNE.....ILAP.E
tr|A0A090IBV6|Aliivibrio/1-617 582 ..GKNLFFDSGHFS.....GYGALKVWPYIEE.SLRKG.....TLLD.F
tr|A0A038H4R1|Burkholderia/1-644 615 ..GRPLYSDYFHMS.....EYGNRFLVPMFRRV.....FEQ
tr|B1KL29|Shewanella/1-652 595 ..QQLLVFDYGHLT.....QAGSRLIAKQVIVPQLIKR.....LSAKPG..YTFE
tr|A0A0W0YMF6|Legionella/1-623 594 KKKELITWDEAHLT.....STGSRMLVAHFPAP.N
tr|A0A0R3MIT7|Bradyrhizobium/1-625 583 SASDVIITDNVHLS.....DTGSNFLIEGIRRK.L
tr|A0A0J1LHQ0|Citrobacter/1-669 629 ..GRPLYSDSNHLS.....AFGSRLLPMFQRA.FKQAEH.....TEATNA.....L
tr|J3HA72|Pantoea_/1-633 601 ..DVPLYRDDNHLS.....EVGNKNLNGVFGGM.WKMID.....RI
tr|A0A075P176|Alteromonas/1-613 585 ..GRFRYKDSNHAR.....PWFAEEAMNYLLTTIN
tr|A0A0Q0MU03|Pseudoalteromonas/1-645 615 ..NRPIYFDGDHLS.....EFGNKLLTPMFLTI.....VKE
tr|A0A077XXI0|Sphingobacterium/1-616 588 ..QDPYMHDDDHLS.....KYGADQIVAYLLKNKL
tr|J2PSI9|Caulobacter_/1-623 593 .LGEPVQFDYGHLT.....ERGSQLVVERLVST.GQL
tr|A0A0F5ZZM0|Grimontia/1-627 600 ..GKALYFDEDHPS.....VVGARLLVDLIPF.DL
tr|A0A072D806|Acinetobacter/1-635 607 ..SDLLQVDYGHLS.....KKGSIYVVNNYIKP.NIE
tr|K0E9E8|Alteromonas/1-638 605 ..GRSLYFDDNHVS.....NYGASLIAEDIIQNKNSHT.....P
tr|A0A1G3I0Z5|Rhodoferax/1-663 632 ..GRPLYYDDNHLS.....EFGNKLLVPMFARV.FEPLP
tr|A0A069PRQ1|Paraburkholderia/1-717 627 ..LVPLYIDGDHLN.....RRGSARLVPLFRQL.FVPATA.....PLPAREPNTTQTTGP.ATQTGVG
tr|A0A0P0QBN4|Serratia/1-656 625 ..GRPIYFDDNHLS.....EYGNKLLVPMFKTL.FSSAS
tr|A0A085VIZ6|Pseudomonas/1-631 598 ..NSLLQFDSGHLS.....EKGSMYIVKNFVLPELRKQ.....SI
tr|E6VIY1|Rhodopseudomonas/1-653 622 GSGTPVTFDGSHLT.....VASSQFFVRQVARD.L
tr|A0A1B7ICK8|Buttiauxella/1-670 629 ..GQPLYSDDNHLS.....EFGSQLLIPMFRQA.FEPEEY.....TELTKS.....P.G
tr|A0A1E3ZLQ3|Bordetella_/1-665 608 ..HVPLMVDDNHLS.....ILGAQQSADYVLAA.AGGATDFDAFLDTLTGRPSAS.....LGVP.A
tr|R4YTI5|Oleispira/1-645 609 ..GVPLYFDDDHPS.....IKGAERLIRSILEDNDQQS.....IILP
tr|E1SVP6|Ferrimonas/1-652 615 ..ERSYYFDDDHLS.....TDGAALLQPMIEQ.ALVTP.....LRRS.A
tr|A0A1E4L6G6|Thiobacillus/1-641 589 .NGQYLVADAGHLS.....AYGSKLAVASYFPF.FTS.R.....SLH.....VG
tr|A0A0T9N312|Yersinia/1-654 618 ..EIPYSLDGAHIS.....VDSSLMAAKRFIETGVYKKYF.....SDVK
tr|A0A085THR2|Vibrio/1-648 621 ENMSLTAFDSSHLT.....EDGADYVVSHFNL
tr|A0A167CIA4|Achromobacter/1-636 609 DPVSFTTWDIGHLT.....TEGAKKVAGAIRL
tr|A0A150FIY7|Leptospira/1-667 630 GIESLISFDHGHLT.....RSGSEFLVKLIFPE.L
tr|W3RFJ7|Afipia/1-681 647 SD.APIAWDQGHLT.....TAGSIYFVQANAGQ.L
tr|A0A1E4CSA7|Hyphomicrobium/1-660 619 ..GQAIYFDRDHLS.....VSAARQVVGDIILP.AIGLRG.....VAGA


WP_000400612/1-640 639 .....IK.....
WP_000639473/1-609 .....
SIU02679/1-640 639 .....IR.....
AAX87447/1-622 611 .....VMTPEQV..KKL.Y.E
CAC99369/1-622 622 .....K.....
WP_002245844/1-622 612 .....LLKSSRD..G.A.L.Q
WP_011101182/1-660 659 .....NN.....
YP_004888877/1-615 612 .....AKP.....A
WP_000379821/1-603 602 .....KK.....
WP_021723064/1-605 605 .....K.....
WP_001220853/1-605 602 .....VKS.....K
WP_003687310/1-624 617 .....FLDTRHN.....H
tr|F4SV41|Escherichia_coli/1-609 .....
tr|A0A198GJ22|Enterobacter/1-609 .....
tr|A0A0H3SVW1|Photobacterium/1-629 628 .....N.....Q
tr|A0A090IBV6|Aliivibrio/1-617 617 .....Q
tr|A0A038H4R1|Burkholderia/1-644 644 .....G
tr|B1KL29|Shewanella/1-652 636 ELAKAAPQQTHIERYS.....K
tr|A0A0W0YMF6|Legionella/1-623 623 .....N
tr|A0A0R3MIT7|Bradyrhizobium/1-625 612 .....FPCPEFCQEQI.LRE
tr|A0A0J1LHQ0|Citrobacter/1-669 668 .....QL.....
tr|J3HA72|Pantoea_/1-633 .....
tr|A0A075P176|Alteromonas/1-613 613 .....P
tr|A0A0Q0MU03|Pseudoalteromonas/1-645 644 .....S.....K
tr|A0A077XXI0|Sphingobacterium/1-616 616 I
tr|J2PSI9|Caulobacter_/1-623 623 .....P.
tr|A0A0F5ZZM0|Grimontia/1-627 627 .....P
tr|A0A072D806|Acinetobacter/1-635 .....
tr|K0E9E8|Alteromonas/1-638 637 .....L.....Q
tr|A0A1G3I0Z5|Rhodoferax/1-663 663 .....I.....
tr|A0A069PRQ1|Paraburkholderia/1-717 681 HAVTATPVPAISPARTRIHASLFFKTGGASPDRM.AAS
tr|A0A0P0QBN4|Serratia/1-656 656 .....K.....
tr|A0A085VIZ6|Pseudomonas/1-631 631 .....R
tr|E6VIY1|Rhodopseudomonas/1-653 651 .....LG.....R
tr|A0A1B7ICK8|Buttiauxella/1-670 669 .....NG.....
tr|A0A1E3ZLQ3|Bordetella_/1-665 660 .....G.....KA..TPH
tr|R4YTI5|Oleispira/1-645 644 .....S.....N
tr|E1SVP6|Ferrimonas/1-652 650 .....LA.....R
tr|A0A1E4L6G6|Thiobacillus/1-641 625 KQ.....LLHARLAIGKQGENN.
tr|A0A0T9N312|Yersinia/1-654 .....
tr|A0A085THR2|Vibrio/1-648 648 .....D
tr|A0A167CIA4|Achromobacter/1-636 636 .....P
tr|A0A150FIY7|Leptospira/1-667 659 .....SRFF....RTY.V.E
tr|W3RFJ7|Afipia/1-681 675 .....LGSS.....P.N.R
tr|A0A1E4CSA7|Hyphomicrobium/1-660 655 .....HPSA.....GK.....

Appendix V pBADcLIC2005 and pETFPP_30 plasmid maps

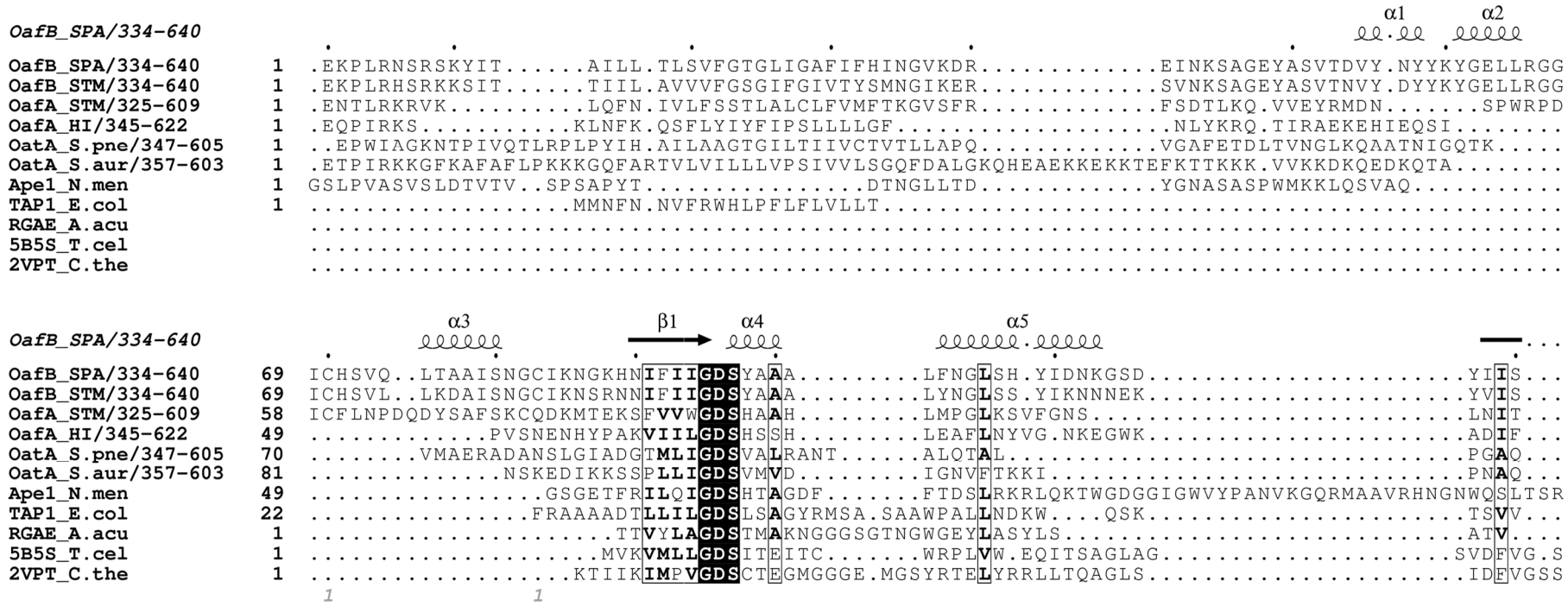
See methods **Table 2.2** for plasmid reference. Plasmid maps were produced using PlasMapper (Dong *et al.*, 2004).

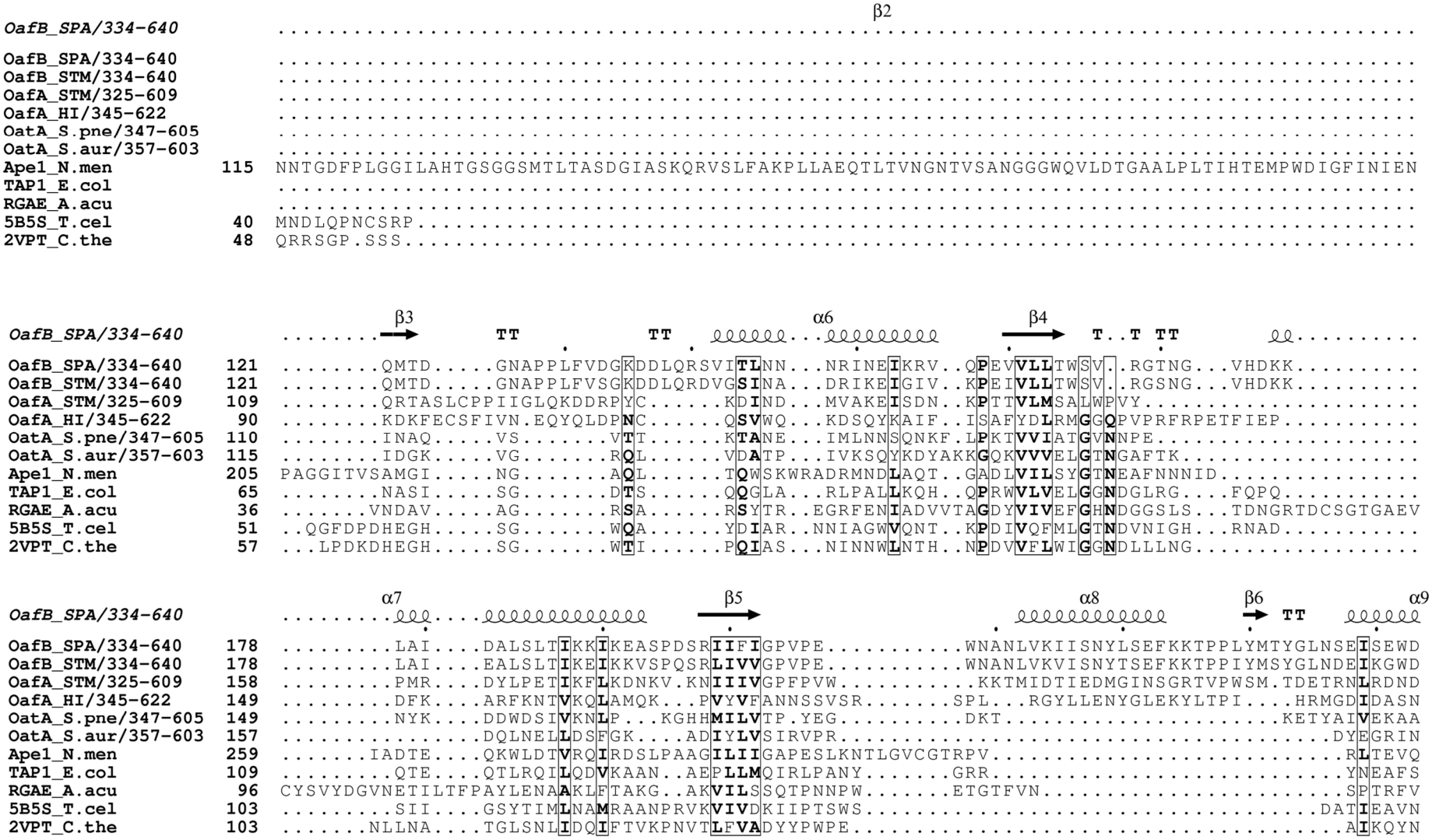


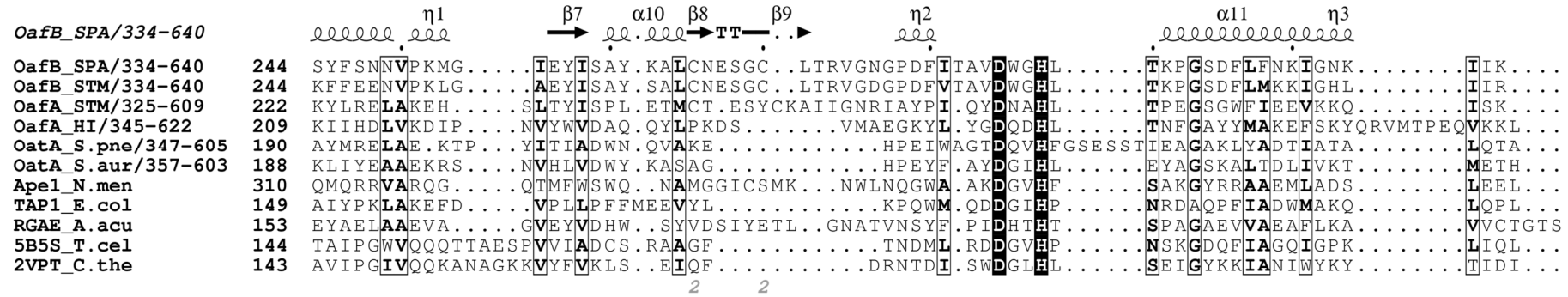


Appendix VI Structure based sequence alignment of $OafB_{SPA}$_377 and structural homologues

Structure based sequence alignment of OafB, OafA and closest structural homologues. Details of sequences are outlined in section 2.4.1. The alignment is presented using ESPript coloured using black and white colour scheme by % Equivalent (global score 0.7). Secondary structural elements of OafB_SPA are presented above the alignment and disulphide bonding pairs below.




β2
OafB_SPA/334-640
OafB_SPA/334-640
OafB_STM/334-640
OafA_STM/325-609
OafA_HI/345-622
OatA_S.pne/347-605
OatA_S.aur/357-603
Ape1_N.men 115 NNTGDFPLGGILAHTGSGGSMTLTASDGIASKQRVSLFAKPLLAEQTLTVNGNTVSANGGGWQVLDTGAALPLTIHTEMPWDIGFINIEN
TAP1_E.col
RGAE_A.acu
5B5S_T.cel 40 MNDLQPNCSRP
2VPT_C.the 48 QRRSGP.SSS
β3
α6
β4
OafB_SPA/334-640
OafB_SPA/334-640 121 .....QMTD.....GNAPPLFVDGKDDLQRSVITLNN...NRINEIKRV..QPEVVLLTWSV..RGTNG...VHDKK.....
OafB_STM/334-640 121 .....QMTD.....GNAPPLFVSGKDDLQRDVGSINA...DRIKEIGIV..KPEIVLLTWSV..RGSNG...VHDKK.....
OafA_STM/325-609 109 .....QRTASLCPPIIGLQKDDRPYC.....KDINDM...VAKEISDN..KPTTVLMSALWPVY.....
OafA_HI/345-622 90 .....KDKFECSFIVN.EQYQLDPNC.....QSVWQ...KDSQYKAIF..ISAFYDLRMGGQPVPRFRPETFIEP.....
OatA_S.pne/347-605 110 .....INAQ.....VS.....VTT.....KTANE...IMLNNSQNKF.LPKTVVIATGVNNPE.....
OatA_S.aur/357-603 115 .....IDGK.....VG.....RQL.....VDATP...IVKSQYKDYAKKGQKVVVELGTNGAFTK.....
Ape1_N.men 205 PAGGITVSAMGI.....NG.....AQL.....TQWSKWRADRMNDLAQT..GADLVILSYGTNEAFNNNID.....
TAP1_E.col 65 .....NASI.....SG.....DTS.....QQGLA...RLPALLKQH..QPRWVLVELGGNDGLRG....FQPQ.....
RGAE_A.acu 36 .....VNDAV.....AG.....RSA.....RSYTR..EGRFENIADVVTAGDYVIVEFGHNDGGSLS...TDNGRTDCSGTGAEV
5B5S_T.cel 51 ..QGFDPDHEGH.....SG.....WQA.....YDIAR..NNIAGWVQNT..KPDIVQFMLGTNDVNIGH...RNAD.....
2VPT_C.the 57 ...LPDKDHEGH.....SG.....WTI.....PQIAS...NINNWLNTH..NPDVVFLWIGGNDLLLNG.....
α7
β5
α8
β6
α9
OafB_SPA/334-640
OafB_SPA/334-640 178 .....LAI....DALSLTIKKIKEASPDSRIIFIGPVPE.....WNANLVKIISNYLSEFKKTPPLYMTYGLNSEISEWD
OafB_STM/334-640 178 .....LAI....EALSLTIKEIKKVSPQSRLIVVGPVPE.....WNANLVKVISNYTSEFKKTPPIYMSYGLNDEIKGWD
OafA_STM/325-609 158 .....PMR....DYLPETIKFLKDNKV.KNIIIVGPFPVW.....KKTMIDTIEDMGINSGRTVPWSM.TDETRNLRDND
OafA_HI/345-622 149 .....DFK....ARFKNTVKQLAMQK...PVYVFANNSSVSR.....SPL...RGYLLENYGLEKYLTPI....HRMGDIDASN
OatA_S.pne/347-605 149 .....NYK....DDWDSIVKNLP...KGHHMILVTP.YEG.....DKT.....KETYAIVEKAA
OatA_S.aur/357-603 157 .....DQLNELLDSFGK....ADIYLVSIRVPR.....DYEGRIN
Ape1_N.men 259 .....IADTE....QKWLDTVRQIRDSLPAAGILIIGAPESLKNTLGVCGTRPV.....RLTEVQ
TAP1_E.col 109 .....QTE....QTLRQILQDVKAAN..AEPLLMQIRLPANY.....GRR.....YNEAFS
RGAE_A.acu 96 CYSVYDGVNETILTFPAYLENAAKLFTAKG..AKVILSSQTPNNPW.....ETGTFVN.....SPTRFV
5B5S_T.cel 103 .....SII....GSYTIMLNAMRAANPRVKVIVDKIIPTSWS.....DATIEAVN
2VPT_C.the 103 .....NLLNA....TGLSNLIDQIFTVKPNVTLFVADYYPWPE.....AIKQYN



OafB_SPA/334-640

OafB_SPA/334-640 .....
OafB_STM/334-640 .....
OafA_STM/325-609 .....
OafA_HI/345-622 277 ....YE.....
OatA_S.pne/347-605 252 ....QDKPVKSK...
OatA_S.aur/357-603 243 ....ATNKK.....
Ape1_N.men 372 ....VRSAAIR....
TAP1_E.col 204 ....VNHDS.....
RGAE_A.acu 219 LKSVLTTTSFEGTCL
5B5S_T.cel 203 ....IKDVS.....
2VPT_C.the 202 ....LRRALAG....

Abbreviations

Abe	Abequose
Ac	Acetate
ACP	Acyl Carrier Protein
Amp	Ampicillin
APS	ammonium persulfate
AT3	Acyltransferase_3
BSA	Bovine Serum Albumin
CoA	Coenzyme A
ddH_2O	Double distilled (Milli Q) H_2O
diH_2O	Deionised H_2O
DNA	Deoxyribonucleic acid
ECA	Enterobacterial common antigen
ESI-TOF	Electrospray ionisation time of flight
FT-ICR-MS	Fourier transform ion cyclotron resonance mass spectrometry
GlcNAc	N-acetylglucosamine
GlcA	Glucuronic acid
HRP	Horseradish peroxidase
iNTS	invasive non-typhoidal salmonella
Kan	Kanamycin
KDO	3-deoxy-d-manno-2-octulosonic acid
Ko	D-glycero-D-talo-oct-2-ulosonic acid
LB	Luria Broth
LIC	Ligation independent cloning
LOS	Lipooligosaccharide
LPS	Lipopolysaccharide
LTA	Lipoteichoic acid
MALDI	Matrix Assisted Laser Desorption Ionization
MBOAT	Membrane bound O-acetyltransferase
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MurNAc	N-acetylmuramic acid
NMR	Nuclear magnetic resonance
OD	Optical density
OPG	Osmoregulated periplasmic glucans
PAGE	Poly acrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PNAG	poly-N-acetylglucosamine
pNP-Ac	p-nitrophenyl acetate

PVDF	polyvinylidene difluoride
Rha	Rhamnose
RNA	Ribonucleic acid
RPM	Revolutions per minute
SCV	Salmonella containing vacuole
SDS	Sodium dodecyl sulphate
SOC	Super Optimal broth with Catabolite repression
SPA	*Salmonella entrica* subspecies *enterica* serovar Paratyphi A
STD-NMR	Saturation transfer difference – Nuclear magnetic resonance
STM	*Salmonella entrica* subspecies *enterica* serovar Typhimurium
TEMED	NNN'N' tetramethyl-ethyldiamine
TMH	Trans membrane helices
TOF/TOF	Tandem Time-of-Flight
WT	Wild-type
WTA	Wall teichoic acid

References

Aas, F. E., Vik, A., Vedde, J., Koomey, M. and Egge-Jacobsen, W. (2007) 'Neisseria gonorrhoeae O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure.', *Molecular microbiology*. Wiley-Blackwell, 65(3), pp. 607–24. doi: 10.1111/j.1365-2958.2007.05806.x.

Abedon, S. T. and Yin, J. (2009) 'Bacteriophage Plaques: Theory and Analysis', in *Bacteriophages: Methods and Protocols*. Humana Press, pp. 161–174. doi: 10.1007/978-1-60327-164-6_17.

Abellón-Ruiz, J., Kaptan, S. S., Baslé, A., Claudi, B., Bumann, D., Kleinekathöfer, U. and van den Berg, B. (2017) 'Structural basis for maintenance of bacterial outer membrane lipid asymmetry', *Nature Microbiology*. Nature Publishing Group, 2(12), pp. 1616–1623. doi: 10.1038/s41564-017-0046-x.

Akoh, C. C., Lee, G.-C., Liaw, Y.-C., Huang, T.-H. and Shaw, J.-F. (2004) 'GDSL family of serine esterases/lipases.', *Progress in lipid research*, 43(6), pp. 534–52. doi: 10.1016/j.plipres.2004.09.002.

Allison, G. E. and Verma, N. K. (2000) 'Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*', *Trends in Microbiology*, 8(1), pp. 17–23. doi: 10.1016/S0966-842X(99)01646-7.

Álvarez-Ordóñez, A., Begley, M., Prieto, M., Messens, W., López, M., Bernardo, A. and Hill, C. (2011) 'Salmonella spp. survival strategies within the host gastrointestinal tract', *Microbiology*, pp. 3268–3281. doi: 10.1099/mic.0.050351-0.

Álvarez-Ordóñez, A., Prieto, M., Bernardo, A., Hill, C. and López, M. (2012) 'The Acid Tolerance Response of Salmonella spp.: An adaptive strategy to survive in stressful environments prevailing in foods and the host', *Food Research International*. Elsevier, 45(2), pp. 482–492. doi: 10.1016/j.foodres.2011.04.002.

Anonsen, J. H., Børud, B., Vik, Å., Viburiene, R. and Koomey, M. (2017) 'Structural and genetic analyses of glycan O-acetylation in a bacterial protein glycosylation system: evidence for differential effects on glycan chain length', *Glycobiology*, 27(9), pp. 888–899. doi: 10.1093/glycob/cwx032.

Aribam, S. D., Elsheimer-Matulova, M., Matsui, H., Hirota, J., Shiraiwa, K., Ogawa, Y., Hikono, H., Shimoji, Y. and Eguchi, M. (2015) 'Variation in antigen-antibody affinity among serotypes of Salmonella O4 serogroup, determined using specific antisera', *FEMS Microbiology Letters*, 362(21). doi: 10.1093/femsle/fnv168.

Arisawa, A., Kawamura, N., Takeda, K., Tsunekawa, H., Okamura, K. and Okamoto, R. (1994) 'Cloning of the Macrolide Antibiotic Biosynthesis Gene acyA, Which Encodes 3-0-Acyltransferase, from Streptomyces thermotolerans and Its Use for Direct Fermentative Production of a Hybrid Macrolide Antibiotic', *Applied and Environmental Microbiology*, 60(7), pp. 2657–2660.

Arisawa, A., Kawamura, N., Tsunekawa, H., Okamura, K., Tone, H. and Okamoto, R. (1993) 'Cloning and nucleotide sequences of two genes involved in the 4''-O-acylation of macrolide antibiotics from Streptomyces thermotolerans.', *Bioscience, biotechnology, and biochemistry*, 57(12), pp. 2020–5.

Atkin, K. E., Macdonald, S. J., Brentnall, A. S., Potts, J. R. and Thomas, G. H. (2014) 'A different path: Revealing the function of staphylococcal proteins in biofilm formation', *FEBS Letters*, 588(10), pp. 1869–1872. doi: 10.1016/j.febslet.2014.04.002.

Aubry, C., Goulard, C., Nahori, M. A., Cayet, N., Decalf, J., Sachse, M., Boneca, I. G., Cossart, P. and Dussurget, O. (2011) 'OatA, a peptidoglycan O-acetyltransferase involved in Listeria monocytogenes immune escape, is critical for virulence', *Journal of Infectious Diseases*. Oxford University Press, 204(5), pp. 731–740. doi: 10.1093/infdis/jir396.

Baggesen, D. L., Sørensen, G., Nielsen, E. M. and Wegener, H. C. (2010) 'Phage typing of Salmonella Typhimurium - is it still a useful tool for surveillance and outbreak investigation?', *Eurosurveillance*, 15(4), p. 1. doi: 10.2807/ese.15.04.19471-en.

Baker, J. A., Wong, W.-C., Eisenhaber, B., Warwicker, J. and Eisenhaber, F. (2017) 'Charged residues next to transmembrane regions revisited: "Positive-inside rule" is complemented by the "negative inside depletion/outside enrichment rule"', *BMC Biology*. BioMed Central, 15(1), p. 66. doi: 10.1186/s12915-017-0404-4.

Baker, P., Ricer, T., Moynihan, P. J., Kitova, E. N., Walvoort, M. T. C., Little, D. J., Whitney, J. C., Dawson, K., Weadge, J. T., Robinson, H., Ohman, D. E., Codée, J. D. C., Klassen, J. S., Clarke, A. J. and Howell, P. L. (2014) 'P. aeruginosa SGNH hydrolase-like proteins AlgJ

and AlgX have similar topology but separate and distinct roles in alginate acetylation.', *PLoS pathogens*, 10(8), p. e1004334. doi: 10.1371/journal.ppat.1004334.

Baliban, S. M., Yang, M., Ramachandran, G., Curtis, B., Shridhar, S., Laufer, R. S., Wang, J. Y., Van Druff, J., Higginson, E. E., Hegerle, N., Varney, K. M., Galen, J. E., Tennant, S. M., Lees, A., MacKerell, A. D., *et al.* (2017) 'Development of a glycoconjugate vaccine to prevent invasive Salmonella Typhimurium infections in sub-Saharan Africa', *PLoS Neglected Tropical Diseases*. Edited by M. Picardeau. Public Library of Science, 11(4), p. e0005493. doi: 10.1371/journal.pntd.0005493.

Baranwal, G., Mohammad, M., Jarneborn, A., Reddy, B. R., Golla, A., Chakravarty, S., Biswas, L., Götz, F., Shankarappa, S., Jin, T. and Biswas, R. (2017) 'Impact of cell wall peptidoglycan O-acetylation on the pathogenesis of Staphylococcus aureus in septic arthritis', *International Journal of Medical Microbiology*. Urban & Fischer, 307(7), pp. 388–397. doi: 10.1016/J.IJMM.2017.08.002.

Battisti, L., Lara, J. C. and Leigh, J. A. (1992) 'Specific oligosaccharide form of the Rhizobium meliloti exopolysaccharide promotes nodule invasion in alfalfa', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 89(12), pp. 5625–5629. doi: 10.1073/pnas.89.12.5625.

Baumann, A.-M. T., Bakkers, M. J. G., Buettner, F. F. R., Hartmann, M., Grove, M., Langereis, M. A., de Groot, R. J. and Mühlenhoff, M. (2015) '9-O-Acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate.', *Nature communications*. Nature Publishing Group, 6, p. 7673. doi: 10.1038/ncomms8673.

Bera, A., Herbert, S., Jakob, A., Vollmer, W. and Götz, F. (2005) 'Why are pathogenic staphylococci so lysozyme resistant? The peptidoglycan O-acetyltransferase OatA is the major determinant for lysozyme resistance of Staphylococcus aureus', *Molecular Microbiology*, 55(3), pp. 778–787. doi: 10.1111/j.1365-2958.2004.04446.x.

Berg, J. M., Tymoczko, J. L. and Stryer, L. (2002) 'Proteases: Facilitating a Difficult Reaction.', *Biochemistry*. 5th edn. W.H. Freeman, p. Section 9.1.

Berical, A. C., Harris, D., Dela Cruz, C. S. and Possick, J. D. (2016) 'Pneumococcal Vaccination Strategies. An Update and Perspective', *Annals of the American Thoracic Society*. American Thoracic Society, 13(6), pp. 933–944. doi:

10.1513/AnnalsATS.201511-778FR.

Berkmen, M. (2012) 'Production of disulfide-bonded proteins in Escherichia coli.', *Protein expression and purification*, 82(1), pp. 240–51. doi: 10.1016/j.pep.2011.10.009.

Bernard, E., Rolain, T., Courtin, P., Guillot, A., Langella, P., Hols, P. and Chapot-Chartier, M.-P. (2011) 'Characterization of O-acetylation of N-acetylglucosamine: a novel structural variation of bacterial peptidoglycan.', *The Journal of biological chemistry*, 286(27), pp. 23950–8. doi: 10.1074/jbc.M111.241414.

Bernard, E., Rolain, T., David, B., André, G., Dupres, V., Dufrêne, Y. F., Hallet, B., Chapot-Chartier, M.-P. P. and Hols, P. (2012) 'Dual Role for the O-Acetyltransferase OatA in Peptidoglycan Modification and Control of Cell Septation in Lactobacillus plantarum m'. Public Library of Science, 7(10), p. e47893. doi: 10.1371/journal.pone.0047893.

Berti, F., De Ricco, R. and Rappuoli, R. (2018) 'Role of o-acetylation in the immunogenicity of bacterial polysaccharide vaccines', *Molecules*. doi: 10.3390/molecules23061340.

Beuzón, C. R., Méresse, S., Unsworth, K. E., Ruíz-Albert, J., Garvis, S., Waterman, S. R., Ryder, T. A., Boucrot, E. and Holden, D. W. (2000) 'Salmonella maintains the integrity of its intracellular vacuole through the action of SifA.', *The EMBO journal*. European Molecular Biology Organization, 19(13), pp. 3235–49. doi: 10.1093/emboj/19.13.3235.

Bhagwat, A. A., Jun, W., Liu, L., Kannan, P., Dharne, M., Pheh, B., Tall, B. D., Kothary, M. H., Gross, K. C., Angle, S., Meng, J. and Smith, A. (2009) 'Osmoregulated periplasmic glucans of Salmonella enterica serovar Typhimurium are required for optimal virulence in mice', *Microbiology*. Microbiology Society, 155(1), pp. 229–237. doi: 10.1099/mic.0.023747-0.

Birch, J., Axford, D., Foadi, J., Meyer, A., Eckhardt, A., Thielmann, Y. and Moraes, I. (2018) 'The fine art of integral membrane protein crystallisation', *Methods*. Academic Press, pp. 150–162. doi: 10.1016/j.ymeth.2018.05.014.

Bjarnsholt, T. (2013) 'The role of bacterial biofilms in chronic infections', *APMIS*. John Wiley & Sons, Ltd (10.1111), 121, pp. 1–58. doi: 10.1111/apm.12099.

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.', *Molecular microbiology*, 70(5), pp. 1105–19. doi: 10.1111/j.1365-2958.2008.06461.x.

Bohin, J.-P. (2000) 'Osmoregulated periplasmic glucans in Proteobacteria.', *FEMS microbiology letters*. The Oxford University Press, 186(1), pp. 11–9. doi: 10.1111/j.1574-6968.2000.tb09075.x.

Bontemps-Gallo, S., Madec, E., Dondeyne, J., Delrue, B., Robbe-Masselot, C., Vidal, O., Prouvost, A.-F., Boussemart, G., Bohin, J.-P. and Lacroix, J.-M. (2013) 'Concentration of osmoregulated periplasmic glucans (OPGs) modulates the activation level of the RcsCD RcsB phosphorelay in the phytopathogen bacteria *Dickeya dadantii*', *Environmental Microbiology*. John Wiley & Sons, Ltd (10.1111), 15(3), pp. 881–894. doi: 10.1111/1462-2920.12054.

Bontemps-Gallo, S., Madec, E., Robbe-Masselot, C., Souche, E., Dondeyne, J. and Lacroix, J. M. (2016) 'The opgC gene is required for OPGs succinylation and is osmoregulated through RcsCDB and EnvZ/OmpR in the phytopathogen Dickeya dadantii', *Scientific Reports*, 6, p. 19619. doi: 10.1038/srep19619.

Boore, A. L., Hoekstra, R. M., Iwamoto, M., Fields, P. I., Bishop, R. D. and Swerdlow, D. L. (2015) 'Salmonella enterica Infections in the United States and Assessment of Coefficients of Variation: A Novel Approach to Identify Epidemiologic Characteristics of Individual Serotypes, 1996-2011.', *PloS one*. Public Library of Science, 10(12), p. e0145416. doi: 10.1371/journal.pone.0145416.

Brady, A. M., Calix, J. J., Yu, J., Geno, K. A., Cutter, G. R. and Nahm, M. H. (2014) 'Low invasiveness of pneumococcal serotype 11A is linked to ficolin-2 recognition of O-acetylated capsule epitopes and lectin complement pathway activation.', *The Journal of infectious diseases*. Oxford University Press, 210(7), pp. 1155–65. doi: 10.1093/infdis/jiu195.

Brandtzaeg, P., Bjerre, A., Øvstebø, R., Brusletto, B., Joø, G. B. and Kierulf, P. (2004) 'Neisseria meningitidis lipopolysaccharides in human pathology', *Journal of Endotoxin Research*, 7(6), pp. 401–420. doi: 10.1179/096805101101533016.

Breedveld, M. W. and Miller, K. J. (1994) 'Cyclic beta-glucans of members of the family

Rhizobiaceae.', *Microbiological reviews*. American Society for Microbiology (ASM), 58(2), pp. 145–61.

Brett, P. J., Burtnick, M. N., Heiss, C., Azadi, P., DeShazer, D., Woods, D. E. and Gherardini, F. C. (2011) 'Burkholderia thailandensis oacA mutants facilitate the expression of Burkholderia mallei-like O polysaccharides.', *Infection and immunity*. American Society for Microbiology (ASM), 79(2), pp. 961–9. doi: 10.1128/IAI.01023-10.

Brett, P. J., Burtnick, M. and Woods, D. (2003) 'The wbiA locus is required for the 2-O-acetylation of lipopolysaccharides expressed by Burkholderia pseudomallei and Burkholderia thailandensis', *FEMS Microbiology Letters*. Blackwell Publishing Ltd, pp. 323–328. doi: 10.1111/j.1574-6968.2003.tb11536.x.

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', *Molecular Microbiology*, 77(2), pp. 337–353. doi: 10.1111/j.1365-2958.2010.07203.x.

Broeker, N. K. and Barbirz, S. (2017) 'Not a barrier but a key: How bacteriophages exploit host's O-antigen as an essential receptor to initiate infection', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 105(3), pp. 353–357. doi: 10.1111/mmi.13729.

Brogioni, B. and Berti, F. (2014) 'Surface plasmon resonance for the characterization of bacterial polysaccharide antigens: a review', *MedChemComm*. Royal Society of Chemistry, 5(8), p. 1058. doi: 10.1039/C4MD00088A.

Brown, S., Santa Maria, J. P., Walker, S. and Walker, S. (2013) 'Wall teichoic acids of gram-positive bacteria.', *Annual review of microbiology*. NIH Public Access, 67, pp. 313–36. doi: 10.1146/annurev-micro-092412-155620.

Le Brun, A. P., Clifton, L. A., Halbert, C. E., Lin, B., Meron, M., Holden, P. J., Lakey, J. H. and Holt, S. A. (2013) 'Structural characterization of a model gram-negative bacterial surface using lipopolysaccharides from rough strains of Escherichia coli.', *Biomacromolecules*. American Chemical Society, 14(6), pp. 2014–22. doi: 10.1021/bm400356m.

Buendia, A. M., Enenkel, B., Köplin, R., Niehaus, K., Arnold, W. and Pünier, A. (1991) 'The Rhizobium meliloti exoZl exoB fragment of megaplasmid 2: ExoB functions as a UDP-

glucose 4-epimerase and ExoZ shows homology to NodX of Rhizobium leguminosarum biovar viciae strain TOM', *Molecular Microbiology*. Blackwell Publishing Ltd, 5(6), pp. 1519–1530. doi: 10.1111/j.1365-2958.1991.tb00799.x.

Byers, D. M. and Gong, H. (2007) 'Acyl carrier protein: structure–function relationships in a conserved multifunctional protein family', *Biochemistry and Cell Biology*, 85(6), pp. 649–662. doi: 10.1139/O07-109.

Cabeen, M. T. and Jacobs-Wagner, C. (2005) 'Bacterial cell shape', *Nature Reviews Microbiology*. Nature Publishing Group, 3(8), pp. 601–610. doi: 10.1038/nrmicro1205.

Calix, J. J. and Nahm, M. H. (2010) 'A new pneumococcal serotype, 11E, has a variably inactivated wcjE gene.', *The Journal of infectious diseases*. NIH Public Access, 202(1), pp. 29–38. doi: 10.1086/653123.

Calix, J. J., Oliver, M. B., Sherwood, L. K., Beall, B. W., Hollingshead, S. K. and Nahm, M. H. (2011) 'Streptococcus pneumoniae Serotype 9A Isolates Contain Diverse Mutations to wcjE That Result in Variable Expression of Serotype 9V-specific Epitope', *The Journal of Infectious Diseases*. Narnia, 204(10), pp. 1585–1595. doi: 10.1093/infdis/jir593.

Calix, J. J., Saad, J. S., Brady, A. M. and Nahm, M. H. (2012) 'Structural characterization of Streptococcus pneumoniae serotype 9A capsule polysaccharide reveals role of glycosyl 6-O-acetyltransferase wcjE in serotype 9V capsule biosynthesis and immunogenicity.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 287(17), pp. 13996–4003. doi: 10.1074/jbc.M112.346924.

Carpenter, E. P., Beis, K., Cameron, A. D. and Iwata, S. (2008) 'Overcoming the challenges of membrane protein crystallography', *Current Opinion in Structural Biology*. Elsevier, pp. 581–586. doi: 10.1016/j.sbi.2008.07.001.

Chan, W. C., Bycroft, B. W., Leyland, M. L., Lian, L. Y. and Roberts, G. C. (1993) 'A novel post-translational modification of the peptide antibiotic subtilin: isolation and characterization of a natural variant from Bacillus subtilis A.T.C.C. 6633.', *The Biochemical journal*. Portland Press Ltd, (Pt 1), pp. 23–7.

Chang, S.-C. and Magee, A. I. (2009) 'Acyltransferases for secreted signalling proteins (Review)', *Molecular Membrane Biology*. Taylor & Francis, 26(1–2), pp. 104–113. doi:

10.1080/09687680802706432.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., Richardson, D. C. and Richardson, D. C. (2010) 'MolProbity: all-atom structure validation for macromolecular crystallography.', *Acta crystallographica. Section D, Biological crystallography*. International Union of Crystallography, 66(Pt 1), pp. 12–21. doi: 10.1107/S0907444909042073.

Cheong, H. J., Lee, Y. J., Hwang, I. S., Kee, S. Y., Cheong, H. W., Song, J. Y., Kim, J. M., Park, Y. H., Jung, J.-H. and Kim, W. J. (2007) 'Characteristics of non-typhoidal Salmonella isolates from human and broiler-chickens in southwestern Seoul, Korea.', *Journal of Korean medical science*. Korean Academy of Medical Sciences, 22(5), pp. 773–8. doi: 10.3346/jkms.2007.22.5.773.

Clark, C. A., Beltrame, J. and Manning, P. A. (1991) 'The oac gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of Shigella flexneri bacteriophage Sf6', *Gene*, 107(1), pp. 43–52. doi: 10.1016/0378-1119(91)90295-M.

Cogez, V., Gak, E., Puskas, A., Kaplan, S. and Bohin, J. P. (2002) 'The opgGIH and opgC genes of Rhodobacter sphaeroides form an operon that controls backbone synthesis and succinylation of osmoregulated periplasmic glucans', *European Journal of Biochemistry*. Blackwell Science, Ltd, 269(10), pp. 2473–2484. doi: 10.1046/j.1432-1033.2002.02907.x.

Cong, L. and Piepersberg, W. (2007) 'Cloning and characterization of genes encoded in dTDP-D-mycaminose biosynthetic pathway from a midecamycin-producing strain, streptomyces mycarofaciens', *Acta Biochimica et Biophysica Sinica*, 39(3), pp. 187–193. doi: 10.1111/j.1745-7270.2007.00265.x.

Cremers, C. M. and Jakob, U. (2013) 'Oxidant sensing by reversible disulfide bond formation.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 288(37), pp. 26489–96. doi: 10.1074/jbc.R113.462929.

Crisóstomo, M. I., Vollmer, W., Kharat, A. S., Inhülsen, S., Gehre, F., Buckenmaier, S. and Tomasz, A. (2006) 'Attenuation of penicillin resistance in a peptidoglycan O-acetyl transferase mutant of Streptococcus pneumoniae.', *Molecular microbiology*, 61(6), pp. 1497–509. doi: 10.1111/j.1365-2958.2006.05340.x.

Dalrymple, B. P., Cybinski, D. H., Layton, I., McSweeney, C. S., Xue, G. P., Swadling, Y. J. and Lowry, J. B. (1997) 'Three Neocallimastix patriciarum esterases associated with the degradation of complex polysaccharides are members of a new family of hydrolases', *Microbiology*, 143(8), pp. 2605–2614. doi: 10.1099/00221287-143-8-2605.

Darveau, R. P. and Hancock, R. E. W. (1983) 'Procedure for isolation of bacterial lipopolysaccharides from both smooth and rough Pseudomonas aeruginosa and Salmonella typhimurium strains', *Journal of Bacteriology*, 155(2), pp. 831–838.

Datsenko, K. A. and Wanner, B. L. (2000) 'One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products', *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), pp. 6640–6645. doi: 10.1073/pnas.120163297.

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 Genetics*, 9(6), p. e1003568. doi: 10.1371/journal.pgen.1003568.

Davis, C. P. (1996) *Medical Microbiology. 4th edition. Chapter 6 Normal Flora*, *Medical Microbiology*. University of Texas Medical Branch at Galveston. doi: 10.1016/S0749-5978(03)00027-X.

Davis, E. O., Evans, I. J. and Johnston, A. W. B. (1988) 'Identification of nodX, a gene that allows Rhizobium leguminosarum biovar viciae strain TOM to nodulate Afghanistan peas', *MGG Molecular & General Genetics*. Springer-Verlag, 212(3), pp. 531–535. doi: 10.1007/BF00330860.

Day, P. J., Shaw, W. V., Gibbs, M. R. and Leslie, A. G. W. (1992) 'Acetyl Coenzyme A Binding by Chloramphenicol Acetyltransferase: Long-Range Electrostatic Determinants of Coenzyme A Recognition', *Biochemistry*, 31(17), pp. 4198–4205. doi: 10.1021/bi00132a007.

Delcour, A. H., Adler, J., Kung, C. and Martinac, B. (1992) 'Membrane-derived oligosaccharides (MDO's) promote closing of an *E. coli* porin channel', *FEBS Letters*. John Wiley & Sons, Ltd, 304(2–3), pp. 216–220. doi: 10.1016/0014-5793(92)80622-N.

Dénarié, J., Debellé, F. and Promé, J.-C. (1996) 'Rhizobium Lipo-Chitooligosaccharide

Nodulation Factors: Signaling Molecules Mediating Recognition and Morphogenesis', *Annual Review of Biochemistry*, 65(1), pp. 503–535. doi: 10.1146/annurev.bi.65.070196.002443.

Derewenda, Z. S. (2004) 'The use of recombinant methods and molecular engineering in protein crystallization', *Methods*. Academic Press, 34(3), pp. 354–363. doi: 10.1016/J.YMETH.2004.03.024.

Van Deuren, M., Brandtzaeg, P. and Van Der Meer, J. W. M. (2000) 'Update on meningococcal disease with emphasis on pathogenesis and clinical management', *Clinical Microbiology Reviews*. American Society for Microbiology (ASM), 13(1), pp. 144–166. doi: 10.1128/CMR.13.1.144-166.2000.

Díaz-Sánchez, S., Sánchez, S., Herrera-León, S., Porrero, C., Blanco, J., Dahbi, G., Blanco, J. E., Mora, A., Mateo, R., Hanning, I. and Vidal, D. (2013) 'Prevalence of Shiga toxin-producing Escherichia coli, Salmonella spp. and Campylobacter spp. in large game animals intended for consumption: Relationship with management practices and livestock influence', *Veterinary Microbiology*. Elsevier, 163(3–4), pp. 274–281. doi: 10.1016/J.VETMIC.2012.12.026.

Dong, H., Sharma, M., Zhou, H. X. and Cross, T. A. (2012) 'Glycines: Role in α-helical membrane protein structures and a potential indicator of native conformation', *Biochemistry*. NIH Public Access, pp. 4779–4789. doi: 10.1021/bi300090x.

Dong, X., Stothard, P., Forsythe, I. J. and Wishart, D. S. (2004) 'PlasMapper: a web server for drawing and auto-annotating plasmid maps.', *Nucleic acids research*. Oxford University Press, 32(Web Server issue), pp. W660-4. doi: 10.1093/nar/gkh410.

Donlan, R. M. (2002) 'Biofilms: microbial life on surfaces.', *Emerging infectious diseases*. Centers for Disease Control and Prevention, 8(9), pp. 881–90. doi: 10.3201/eid0809.020063.

Draper, J., Enot, D. P., Parker, D., Beckmann, M., Snowdon, S., Lin, W., Zubair, H., Fiehn, O., Dunn, W., Bailey, N., Johnson, H., Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., *et al.* (2009) 'Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive m/z annotation tool utilising predicted ionisation behaviour "rules"', *BMC Bioinformatics*. BioMed Central, 10(1), p. 227. doi: 10.1186/1471-2105-

10-227.

Dzitoyeva, S., Dimitrijevic, N. and Manev, H. (2003) 'Identification of a novel Drosophila gene, beltless , using injectable embryonic and adult RNA interference (RNAi)', *BMC Genomics*. BioMed Central, 4(1), p. 33. doi: 10.1186/1471-2164-4-33.

von Eiff, C., Peters, G. and Heilmann, C. (2002) 'Pathogenesis of infections due to coagulasenegative staphylococci', *The Lancet Infectious Diseases*. Elsevier, 2(11), pp. 677–685. doi: 10.1016/S1473-3099(02)00438-3.

Emsley, P., Cowtan, K. and IUCr (2004) '*Coot* : model-building tools for molecular graphics', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 60(12), pp. 2126–2132. doi: 10.1107/S0907444904019158.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) 'Features and development of Coot', *Acta Crystallographica Section D Biological Crystallography*, 66(4), pp. 486–501. doi: 10.1107/s0907444910007493.

Eng, S.-K., Pusparajah, P., Ab Mutalib, N.-S., Ser, H.-L., Chan, K.-G. and Lee, L.-H. (2015) 'Salmonella: A review on pathogenesis, epidemiology and antibiotic resistance', *Frontiers in Life Science*. Taylor & Francis, 8(3), pp. 284–293. doi: 10.1080/21553769.2015.1051243.

Epp, J. K., Huber, M. L. B., Turner, J. R., Goodson, T. and Schoner, B. E. (1989) 'Production of a hybrid macrolide antibiotic in Streptomyces ambofaciens and Streptomyces lividans by introduction of a cloned carbomycin biosynthetic gene from Streptomyces thermotolerans', *Gene*, 85(2), pp. 293–301. doi: 10.1016/0378-1119(89)90421-6.

Evans, P. R. and Murshudov, G. N. (2013) 'How good are my data and what is the resolution?', *Acta crystallographica. Section D, Biological crystallography*. International Union of Crystallography, 69(Pt 7), pp. 1204–14. doi: 10.1107/S0907444913000061.

Felsenstein, J. (1985) 'CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP', *Evolution*, 39(4), pp. 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x.

Fierer, J. and Guiney, D. G. (2001) 'Diverse virulence traits underlying different clinical outcomes of Salmonella infection.', *The Journal of clinical investigation*, 107(7), pp. 775–80. doi: 10.1172/JCI12561.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J. and Punta, M. (2014) 'Pfam: the protein families database.', *Nucleic acids research*. Oxford University Press, 42(Database issue), pp. D222-30. doi: 10.1093/nar/gkt1223.

Firmin, J. L., Wilson, K. E., Carlson, R. W., Davies, A. E. and Downie, J. A. (1993a) 'Resistance to nodulation of cv. Afghanistan peas is overcome by nodX, which mediates an O-acetylation of the Rhizobium leguminosarum lipo-oligosaccharide nodulation factor', *Molecular Microbiology*. Blackwell Publishing Ltd, 10(2), pp. 351–360. doi: 10.1111/j.1365-2958.1993.tb01961.x.

Firmin, J. L., Wilson, K. E., Carlson, R. W., Davies, A. E. and Downie, J. A. (1993b) 'Resistance to nodulation of cv. Afghanistan peas is overcome by nodX, which mediates an O-acetylation of the Rhizobium leguminosarum lipo-oligosaccharide nodulation factor', *Molecular Microbiology*. Wiley/Blackwell (10.1111), 10(2), pp. 351–360. doi: 10.1111/j.1365-2958.1993.tb01961.x.

Flemming, H.-C. and Wingender, J. (2010) 'The biofilm matrix', *Nature Reviews Microbiology*. Nature Publishing Group, 8(9), pp. 623–633. doi: 10.1038/nrmicro2415.

Fookes, M., Schroeder, G. N., Langridge, G. C., Blondel, C. J., Mammina, C., Connor, T. R., Seth-Smith, H., Vernikos, G. S., Robinson, K. S., Sanders, M., Petty, N. K., Kingsley, R. A., Bäumler, A. J., Nuccio, S.-P., Contreras, I., *et al.* (2011) 'Salmonella bongori Provides Insights into the Evolution of the Salmonellae', *PLoS Pathogens*. Edited by H. Ochman. Public Library of Science, 7(8), p. e1002191. doi: 10.1371/journal.ppat.1002191.

Fox, K. L., Yildirim, H. H., Deadman, M. E., Schweda, E. K. H., Moxon, E. R. and Hood, D. W. (2005) 'Novel lipopolysaccharide biosynthetic genes containing tetranucleotide repeats in Haemophilus influenzae, identification of a gene for adding O-acetyl groups', *Molecular Microbiology*. Blackwell Science Ltd, 58(1), pp. 207–216. doi: 10.1111/j.1365-2958.2005.04814.x.

Franklin, M. J., Nivens, D. E., Weadge, J. T. and Howell, P. L. (2011) 'Biosynthesis of the Pseudomonas aeruginosa Extracellular Polysaccharides, Alginate, Pel, and Psl.', *Frontiers in microbiology*. Frontiers Media SA, 2, p. 167. doi: 10.3389/fmicb.2011.00167.

Gaidelyte, A., Cvirkaite-Krupovic, V., Daugelavicius, R., Bamford, J. K. H. and Bamford, D.

H. (2006) ‘The entry mechanism of membrane-containing phage Bam35 infecting Bacillus thuringiensis.’, *Journal of bacteriology*. American Society for Microbiology Journals, 188(16), pp. 5925–34. doi: 10.1128/JB.00107-06.

Galaz, S., Morales-Quintana, L., Moya-Leõn, M. A. and Herrera, R. (2013) ‘Structural analysis of the alcohol acyltransferase protein family from Cucumis melo shows that enzyme activity depends on an essential solvent channel’, *FEBS Journal*, 280(5), pp. 1344–1357. doi: 10.1111/febs.12127.

Garai, P., Gnanadhas, D. P. and Chakravortty, D. (2012) ‘Salmonella enterica serovars Typhimurium and Typhi as model organisms: Revealing paradigm of host-pathogen interactions’, *Virulence*, 3(4), pp. 377–388. doi: 10.4161/viru.21087.

García, B., González-Sabín, J., Menéndez, N., Brana, A. F., Nunez, L. E., Morís, F., Salas, J. A. and Méndez, C. (2011) ‘The chromomycin CmmA acetyltransferase: A membrane-bound enzyme as a tool for increasing structural diversity of the antitumour mithramycin’, *Microbial Biotechnology*. Wiley-Blackwell, 4(2), pp. 226–238. doi: 10.1111/j.1751-7915.2010.00229.x.

Garib N. Murshudov; Alexei A. Vagin; Eleanor J. Dodson (1997) ‘Refinement of Macromolecular Structures by the Maximum-Likelihood Method’, *Acta Crystallographica Section D: Biological Crystallography*, D53, pp. 240–255. doi: 10.1107/S0907444996012255.

Gassmann, M., Grenacher, B., Rohde, B. and Vogel, J. (2009) ‘Quantifying Western blots: Pitfalls of densitometry’, *Electrophoresis*, 30(11), pp. 1845–1855. doi: 10.1002/elps.200800720.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) ‘Protein Analysis Tools on the ExPASy Server’, *The Proteomics Protocols Handbook Protein Identification and Analysis Tools on the ExPASy Server*, pp. 571–607. doi: 10.1385/1592598900.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) ‘Protein Identification and Analysis Tools on the ExPASy Server’, in *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, pp. 571–607. doi: 10.1385/1-59259-890-0:571.

Geertsma, E. R. and Poolman, B. (2007) 'High-throughput cloning and expression in recalcitrant bacteria', *Nature Methods*. Nature Publishing Group, 4(9), pp. 705–707. doi: 10.1038/nmeth1073.

Geno, K. A., Bush, C. A., Wang, M., Jin, C., Nahm, M. H. and Yang, J. (2017) 'WciG O-Acetyltransferase Functionality Differentiates Pneumococcal Serotypes 35C and 42.', *Journal of clinical microbiology*. American Society for Microbiology, 55(9), pp. 2775–2784. doi: 10.1128/JCM.00822-17.

Geno, K. A., Saad, J. S. and Nahm, M. H. (2017) 'Discovery of novel pneumococcal serotype 35D, a natural WciG-deficient variant of serotype 35B', *Journal of Clinical Microbiology*. American Society for Microbiology, 55(5), pp. 1416–1425. doi: 10.1128/JCM.00054-17.

Geurts, R. and Bisseling, T. (2013) 'American Society of Plant Biologists Rhizobium Nod Factor Perception and Signalling', *The Plant cell*. American Society of Plant Biologists, 14 Suppl(Suppl), pp. S239-49. doi: 10.1105/tpc.002451.

Geurts, R., Heidstra, R., Hadri, A. E., Downie, J. A., Franssen, H., Van Kammen, A. and Bisseling, T. (1997) 'Sym2 of pea is involved in a nodulation factor-perception mechanism that controls the infection process in the epidermis', *Plant Physiology*. American Society of Plant Biologists, 115(2), pp. 351–359. doi: 10.1104/pp.115.2.351.

Ghosh, R., Gilda, J. E. and Gomes, A. V (2014) 'The necessity of and strategies for improving confidence in the accuracy of western blots.', *Expert review of proteomics*. NIH Public Access, 11(5), pp. 549–60. doi: 10.1586/14789450.2014.939635.

Glazyrina, J., Materne, E. M., Dreher, T., Storm, D., Junne, S., Adams, T., Greller, G. and Neubauer, P. (2010) 'High cell density cultivation and recombinant protein production with Escherichia coli in a rocking-motion-type bioreactor', *Microbial Cell Factories*. BioMed Central, 9, p. 42. doi: 10.1186/1475-2859-9-42.

Gorman, J. J., Wallis, T. P. and Pitt, J. J. (2002) 'Protein disulfide bond determination by mass spectrometry', *Mass Spectrometry Reviews*. John Wiley & Sons, Ltd, 21(3), pp. 183–216. doi: 10.1002/mas.10025.

Greenfield, L. K. and Whitfield, C. (2012) 'Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways', *Carbohydrate Research*. Elsevier, 356, pp. 12–

24. doi: 10.1016/J.CARRES.2012.02.027.

Grimont, P. and Weill, F.-X. (2008) ‘Antigenic formulae of the Salmonella servovars’, *WHO Collaborating Centre for Reference and Research on Salmonella*, pp. 1–167.

Gu, H., Wang, Y., Xu, X., Wei, X. and Fong, M. (1996) ‘Increase of the expression of midecamycin 4"-hydroxyl propionyltransferase gene (mpt) by a promoter-like fragment from the midecamycin producing strain.’, *Chinese journal of biotechnology*, 12(3), pp. 147–52.

Guan, S., Bastin, D. A. and Verma, N. K. (1999) ‘Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX’, *Microbiology*. Microbiology Society, 145(5), pp. 1263–1273. doi: 10.1099/13500872-145-5-1263.

Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995) ‘Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.’, *Journal of Bacteriology*, 177(14), pp. 4121–30. doi: 0021-9193/95/$04.00+0.

Haji-Ghassemi, O., Blackler, R. J., Young, N. M. and Evans, S. V (2015) ‘Antibody recognition of carbohydrate epitopes’, *Glycobiology*, pp. 920–952. doi: 10.1093/glycob/cwv037.

Hara, O. and Hutchinson, C. R. (1992) ‘A macrolide 3-O-acyltransferase gene from the midecamycin-producing species Streptomyces mycarofaciens’, *Journal of Bacteriology*, pp. 5141–5144.

Haraga, A., Ohlson, M. B. and Miller, S. I. (2008) ‘Salmonellae interplay with host cells’, *Nature Reviews Microbiology*. Nature Publishing Group, pp. 53–66. doi: 10.1038/nrmicro1788.

Hart, G., Esko, J. D., Kinoshita, T., Cummings, R. D., Schnaar, R. L., Freeze, H. H., Etzler, M. E., Bertozzi, C. R., Lütteke, T., Dell, A., Vliegenthart, J. F., Stanley, P., Prestegard, J. J., Aebi, M., Rudd, P., *et al.* (2015) ‘Symbol Nomenclature for Graphical Representations of Glycans’, *Glycobiology*, 25(12), pp. 1323–1324. doi: 10.1093/glycob/cwv091.

Harvey, L., David, B., Arnold, B., S., Z. L., Paul, M. and Darnell, J. (1995) *Molecular Cell Biology*. Third Edit. W. H. Freeman & Co (Sd).

Hassler, R. A. and Doherty, D. H. (1990) ‘Genetic Engineering of Polysaccharide

Structure: Production of Variants of Xanthan Gum in Xanthomonas campestris', *Biotechnology Progress*, 6(3), pp. 182–187. doi: 10.1021/bp00003a003.

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', *Journal of Applied Microbiology*. Blackwell Publishing Ltd, 110(1), pp. 248–253. doi: 10.1111/j.1365-2672.2010.04877.x.

Van Der Heijden, J., Reynolds, L. A., Deng, W., Mills, A., Scholz, R., Imami, K., Foster, L. J., Duong, F., Finlay, B. B. and Miller, S. I. (2016) 'Salmonella Rapidly Regulates Membrane Permeability To Survive Oxidative Stress', 7(4), pp. 1238–1254. doi: 10.1128/mBio.01238-16.

von Heijne, G. (1986) 'The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology', *The EMBO Journal*, 5(11), pp. 3021–3027. doi: 10.1002/j.1460-2075.1986.tb04601.x.

Heilmann, C., Schweitzer, O., Gerke, C., Vanittanakom, N., Mack, D. and Götz, F. (1996) 'Molecular basis of intercellular adhesion in the biofilm-forming Staphylococcus epidermidis', *Molecular Microbiology*. Blackwell Publishing Ltd, 20(5), pp. 1083–1091. doi: 10.1111/j.1365-2958.1996.tb02548.x.

Heinig, M. and Frishman, D. (2004) 'STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins', *Nucleic Acids Research*, 32(Web Server), pp. W500–W502. doi: 10.1093/nar/gkh429.

Hellerqvist, C. G., Lindberg, B., Svensson, S., Holme, T. and Lindberg, A. A. (1969) 'Structural studies on the O-specific side-chains of the cell-wall lipopolysaccharide from Salmonella typhimurium LT2', *Carbohydrate Research*. Elsevier, 9(2), pp. 237–241. doi: 10.1016/S0008-6215(00)82139-4.

Hildebrand, P. W., Preissner, R. and Frömmel, C. (2004) 'Structural features of transmembrane helices', *FEBS Letters*. No longer published by Elsevier, 559(1–3), pp. 145–151. doi: 10.1016/S0014-5793(04)00061-4.

Hillion, M. and Antelmann, H. (2015) 'Thiol-based redox switches in prokaryotes', *Biological Chemistry*, 396(5), pp. 415–444. doi: 10.1515/hsz-2015-0102.

Hitri, K., Kuttel, M. M., De Benedetto, G., Lockyer, K., Gao, F., Hansal, P., Rudd, T. R., Beamish, E., Rijpkema, S., Ravenscroft, N. and Bolgiano, B. (2019) 'O-acetylation of typhoid capsular polysaccharide confers polysaccharide rigidity and immunodominance by masking additional epitopes', *Vaccine*. Elsevier, 37(29), pp. 3866–3875. doi: 10.1016/j.vaccine.2019.05.050.

Hiyoshi, H., Tiffany, C. R., Bronner, D. N. and Bäumler, A. J. (2018) 'Typhoidal Salmonella serovars: ecological opportunity and the evolution of a new pathovar', *FEMS microbiology reviews*. Narnia, pp. 527–541. doi: 10.1093/femsre/fuy024.

Holliday, G. L., Mitchell, J. B. O. and Thornton, J. M. (2009) 'Understanding the Functional Roles of Amino Acid Residues in Enzyme Catalysis', *Journal of Molecular Biology*, 390(3), pp. 560–577. doi: 10.1016/j.jmb.2009.05.015.

Holst, O. (2011) 'Structure of the Lipopolysaccharide Core Region', in *Bacterial Lipopolysaccharides*, pp. 21–39. doi: 10.1007/978-3-7091-0733-1_2.

Höltje, J. V (1998) 'Growth of the stress-bearing and shape-maintaining murein sacculus of Escherichia coli.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology (ASM), 62(1), pp. 181–203.

Höltje, J. V and Tomasz, A. (1975) 'Lipoteichoic acid: a specific inhibitor of autolysin activity in Pneumococcus.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 72(5), pp. 1690–4. doi: 10.1073/pnas.72.5.1690.

Hölzer, S. U., Schlumberger, M. C., Jäckel, D. and Hensel, M. (2009) 'Effect of the O-antigen length of lipopolysaccharide on the functions of Type III secretion systems in Salmonella enterica.', *Infection and immunity*. American Society for Microbiology Journals, 77(12), pp. 5458–70. doi: 10.1128/IAI.00871-09.

Hong, Y., Duda, K. A., Cunneen, M. M., Holst, O. and Reeves, P. R. (2013) 'The WbaK acetyltransferase of Salmonella enterica group E gives insights into O antigen evolution', *Microbiology (United Kingdom)*, 159(PART11), pp. 2316–2322. doi: 10.1099/mic.0.069823-0.

Horga, L. G., Halliwell, S., Castiñeiras, T. S., Wyre, C., Matos, C. F. R. O., Yovcheva, D. S., Kent, R., Morra, R., Williams, S. G., Smith, D. C. and Dixon, N. (2018) 'Tuning recombinant

protein expression to match secretion capacity', *Microbial Cell Factories*. BioMed Central, 17(1), p. 199. doi: 10.1186/s12934-018-1047-z.

Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L. and Woolfson, D. N. (2015) 'Carbohydrate-Aromatic Interactions in Proteins', *Journal of the American Chemical Society*. American Chemical Society, 137(48), pp. 15152–15160. doi: 10.1021/jacs.5b08424.

Ilg, K., Zandomeneghi, G., Rugarabamu, G., Meier, B. H. and Aebi, M. (2013) 'HR-MAS NMR reveals a pH-dependent LPS alteration by de-O-acetylation at abequose in the O-antigen of Salmonella enterica serovar Typhimurium', *Carbohydrate Research*, 382, pp. 58–64. doi: 10.1016/j.carres.2013.10.002.

Islam, S. T. and Lam, J. S. (2014) 'Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway', *Canadian Journal of Microbiology*, 60(11), pp. 697–716. doi: 10.1139/cjm-2014-0595.

Issenhuth-Jeanjean, S., Roggentin, P., Mikoleit, M., Guibourdenche, M., de Pinna, E., Nair, S., Fields, P. I. and Weill, F. X. (2014) 'Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme', *Research in Microbiology*, 165(48), pp. 526–530. doi: 10.1016/j.resmic.2014.07.004.

Jachymek, W., Niedziela, T., Petersson, C., Lugowski, C., Czaja, J. and Kenne, L. (1999) 'Structures of the O-specific polysaccharides from Yokenella regensburgei (Koserella trabulsii) strains PCM 2476, 2477, 2478, and 2494: High-resolution magic-angle spinning NMR investigation of the O-specific polysaccharides in native lipopolysaccharides a', *Biochemistry*. American Chemical Society, 38(36), pp. 11788–11795. doi: 10.1021/bi990673y.

Jacques, M. (1996) 'Role of lipo-oligosaccharides and lipopolysaccharides in bacterial adherence', *Trends in Microbiology*. Elsevier Current Trends, 4(10), pp. 408–410. doi: 10.1016/0966-842X(96)10054-8.

Jansson, P., Kenne, L. and Lindberg, B. (1975) 'Structure of the extracellular polysaccharide from xanthomonas campestris', *Carbohydrate Research*. Elsevier, 45(1), pp. 275–282. doi: 10.1016/S0008-6215(00)85885-1.

Jarvinen, T., Rautio, J., Masson, M. and Loftsson, T. (2005) 'Design and Pharmaceutical

Applications of Prodrugs', in *Drug Discovery Handbook*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 733–796. doi: 10.1002/0471728780.ch17.

Javadpour, M. M., Eilers, M., Groesbeek, M. and Smith, S. O. (1999) 'Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association.', *Biophysical journal*, 77(3), pp. 1609–1618. doi: 10.1016/S0006-3495(99)77009-8.

Jenkins, H. T. (2018) 'Fragon: rapid high-resolution structure determination from ideal protein fragments', *Acta Crystallographica Section D Structural Biology*. International Union of Crystallography, 74(3), pp. 205–214. doi: 10.1107/s2059798318002292.

Jennings, H. J., Lugowski, C. and Ashton, F. E. (1984) 'Conjugation of meningococcal lipopolysaccharide R-type oligosaccharides to tetanus toxoid as route to a potential vaccine against group B Neisseria meningitidis.', *Infection and immunity*. American Society for Microbiology, 43(1), pp. 407–12.

Jia-Xing, Y., Woolfson, M. M., Wilson, K. S. and Dodson, E. J. (2005) 'A modified ACORN to solve protein structures at resolutions of 1.7 ?? or better', *Acta Crystallographica Section D: Biological Crystallography*, 61(11), pp. 1465–1475. doi: 10.1107/S090744490502576X.

Jogl, G., Hsiao, Y. S. and Tong, L. (2004) 'Structure and function of carnitine acyltransferases', in *Annals of the New York Academy of Sciences*. John Wiley & Sons, Ltd (10.1111), pp. 17–29. doi: 10.1196/annals.1320.002.

Jones, D. T., Taylor, W. R. and Thornton, J. M. (1992) 'The rapid generation of mutation data matrices from protein sequences.', *Computer applications in the biosciences : CABIOS*, 8(3), pp. 275–82.

Jones, T. F., Ingram, L. A., Cieslak, P. R., Vugia, D. J., Tobin-D'Angelo, M., Hurd, S., Medus, C., Cronquist, A. and Angulo, F. J. (2008) 'Salmonellosis Outcomes Differ Substantially by Serotype', *The Journal of Infectious Diseases*. Narnia, 198(1), pp. 109–114. doi: 10.1086/588823.

de Jong, H. K., Parry, C. M., van der Poll, T. and Wiersinga, W. J. (2012) 'Host–Pathogen Interaction in Invasive Salmonellosis', *PLoS Pathogens*. Edited by C. E. Chitnis. Public Library of Science, 8(10), p. e1002933. doi: 10.1371/journal.ppat.1002933.

Kabsch, W. and IUCr (2010) '*XDS*', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 66(2), pp. 125–132. doi: 10.1107/S0907444909047337.

Kagambèga, A., Lienemann, T., Aulu, L., Traoré, A. S., Barro, N., Siitonen, A. and Haukka, K. (2013) 'Prevalence and characterization of Salmonella enterica from the feces of cattle, poultry, swine and hedgehogs in Burkina Faso and their comparison to human Salmonella isolates', *BMC Microbiology*. BioMed Central, 13(1), p. 253. doi: 10.1186/1471-2180-13-253.

Kahler, C. M., Lyons-Schindler, S., Choudhury, B., Glushka, J., Carlson, R. W. and Stephens, D. S. (2006) 'O-acetylation of the terminal N-acetylglucosamine of the lipooligosaccharide inner core in Neisseria meningitidis: Influence on inner core structure and assembly', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 281(29), pp. 19939–19948. doi: 10.1074/jbc.M601308200.

Kahler, C. M. and Stephens, D. S. (1998) 'Genetic basis for biosynthesis, structure, and function of meningococcal lipooligosaccharide (Endotoxin)', *Critical Reviews in Microbiology*, pp. 281–334. doi: 10.1080/10408419891294216.

Kajimura, J., Rahman, A., Hsu, J., Evans, M. R., Gardner, K. H. and Rick, P. D. (2006) 'O Acetylation of the Enterobacterial Common Antigen Polysaccharide Is Catalyzed by the Product of the yiaH Gene of Escherichia coli K-12', *Journal of Bacteriology*. American Society for Microbiology, 188(21), pp. 7542–7550. doi: 10.1128/JB.00783-06.

Kajimura, J., Rahman, A. and Rick, P. D. (2005) 'Assembly of cyclic enterobacterial common antigen in Escherichia coli K-12.', *Journal of bacteriology*. American Society for Microbiology (ASM), 187(20), pp. 6917–27. doi: 10.1128/JB.187.20.6917-6927.2005.

Kalynych, S., Morona, R. and Cygler, M. (2014) 'Progress in understanding the assembly process of bacterial O-antigen', *FEMS Microbiology Reviews*. Narnia, 38(5), pp. 1048–1065. doi: 10.1111/1574-6976.12070.

Katzen, F., Ferreiro, D. U., Oddo, C. G., Ielmini, M. V, Becker, A., Pühler, A. and Ielpi, L. (1998) 'Xanthomonas campestris pv. campestris gum mutants: effects on xanthan biosynthesis and plant virulence.', *Journal of bacteriology*. American Society for

Microbiology (ASM), 180(7), pp. 1607–17.

Ke, N., Landgraf, D., Paulsson, J. and Berkmen, M. (2016) 'Visualization of Periplasmic and Cytoplasmic Proteins with a Self-Labeling Protein Tag', *Journal of Bacteriology*. American Society for Microbiology Journals, 198(7), pp. 1035–1043. doi: 10.1128/JB.00864-15.

Keatinge-Clay, A. T., Shelat, A. A., Savage, D. F., Tsai, S.-C., Miercke, L. J. W., O'Connell, J. D., Khosla, C. and Stroud, R. M. (2003) 'Catalysis, Specificity, and ACP Docking Site of Streptomyces coelicolor Malonyl-CoA:ACP Transacylase', *Structure*. Cell Press, 11(2), pp. 147–154. doi: 10.1016/S0969-2126(03)00004-2.

Keenleyside, W. J. and Whitfield, C. (1996) 'A novel pathway for O-polysaccharide biosynthesis in Salmonella enterica serovar Borreze.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 271(45), pp. 28581–92. doi: 10.1074/jbc.271.45.28581.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. and Sternberg, M. J. E. (2015) 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature Protocols*. Nature Publishing Group, 10(6), pp. 845–858. doi: 10.1038/nprot.2015.053.

Kennedy, E. P. (1982) 'Osmotic regulation and the biosynthesis of membrane-derived oligosaccharides in Escherichia coli.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 79(4), pp. 1092–5. doi: 10.1073/pnas.79.4.1092.

Kimura, A., Patrick, C. C., Miller, E. E., Cope, L. D., McCracken, G. H., Hansen, E. J. and Hansen, E. J. (1987) 'Haemophilus influenzae type b lipooligosaccharide: stability of expression and association with virulence.', *Infection and immunity*. American Society for Microbiology (ASM), 55(9), pp. 1979–86.

Kingsley, R. A. and Baumler, A. J. (2000) 'Host adaptation and the emergence of infectious disease: the Salmonella paradigm', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 36(5), pp. 1006–1014. doi: 10.1046/j.1365-2958.2000.01907.x.

Kingsley, R. a, Msefula, C. L., Thomson, N. R., Kariuki, S., Holt, K. E., Gordon, M. a, Harris, D., Clarke, L., Whitehead, S., Sangal, V., Marsh, K., Achtman, M., Molyneux, M. E., Cormican, M., Parkhill, J., *et al.* (2009) 'Epidemic multiple drug resistant Salmonella

Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype', pp. 2279–2287. doi: 10.1101/gr.091017.109.

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.', *Molecular microbiology*. Wiley-Blackwell, 96(2), pp. 263–75. doi: 10.1111/mmi.12933.

Kintz, E., Heiss, C., Black, I., Donohue, N., Brown, N., Davies, M. R., Azadi, P., Baker, S., Kaye, P. M. and van der Woude, M. (2017) 'Salmonella Typhi Lipopolysaccharide O-antigen Modifications Impact on Serum Resistance and Antibody Recognition', *Infection and Immunity*. American Society for Microbiology, 85(4), p. IAI.01021-16. doi: 10.1128/IAI.01021-16.

Kittelberger, R. and Hilbink, F. (1993) 'Sensitive silver-staining detection of bacterial lipopolysaccharides in polyacrylamide gels', *Journal of Biochemical and Biophysical Methods*, 26(1), pp. 81–86. doi: 10.1016/0165-022X(93)90024-I.

Klein, G. and Raina, S. (2019) 'Regulated Assembly of LPS, Its Structural Alterations and Cellular Response to LPS Defects', *International Journal of Molecular Sciences*. Multidisciplinary Digital Publishing Institute (MDPI), 20(2). doi: 10.3390/IJMS20020356.

Knirel, Y. A., Bystrova, O. V., Shashkov, A. S., Lindner, B., Kocharova, N. A., Senchenkova, S. N., Moll, H., Zähringer, U., Hatano, K. and Pier, G. B. (2001) 'Structural analysis of the lipopolysaccharide core of a rough, cystic fibrosis isolate of Pseudomonas aeruginosa', *European Journal of Biochemistry*. John Wiley & Sons, Ltd (10.1111), 268(17), pp. 4708–4719. doi: 10.1046/j.1432-1327.2001.02396.x.

Knirel, Y. A., Prokhorov, N. S., Shashkov, A. S., Ovchinnikova, O. G., Zdorovenko, E. L., Liu, B., Kostryukova, E. S., Larin, A. K., Golomidova, A. K. and Letarov, A. V (2015) 'Variations in O-antigen biosynthesis and O-acetylation associated with altered phage sensitivity in Escherichia coli 4s.', *Journal of bacteriology*. American Society for Microbiology, 197(5), pp. 905–12. doi: 10.1128/JB.02398-14.

Knirel, Y. A., Rietschel, E. T., Marre, R. and Zahringer, U. (1994) 'The structure of the O-

specific chain of Legionella pneumophila serogroup 1 lipopolysaccharide', *European Journal of Biochemistry*. Blackwell Publishing Ltd, 221(1), pp. 239–245. doi: 10.1111/j.1432-1033.1994.tb18734.x.

Knirel, Y. A., Wang, J., Luo, X., Senchenkova, S. N., Lan, R., Shpirt, A. M., Du, P., Shashkov, A. S., Zhang, N., Xu, J. and Sun, Q. (2014) 'Genetic and structural identification of an O-acyltransferase gene (oacC) responsible for the 3/4-O-acetylation on rhamnose III in Shigella flexneri serotype 6.', *BMC microbiology*. BioMed Central, 14, p. 266. doi: 10.1186/s12866-014-0266-7.

Kozak, N. A., Benson, R. F., Brown, E., Alexander, N. T., Taylor, T. H., Shelton, B. G. and Fields, B. S. (2009) 'Distribution of lag-1 alleles and sequence-based types among Legionella pneumophila serogroup 1 clinical and environmental isolates in the United States.', *Journal of clinical microbiology*. American Society for Microbiology (ASM), 47(8), pp. 2525–35. doi: 10.1128/JCM.02410-08.

Krivoruchko, A., Zhang, Y., Siewers, V., Chen, Y. and Nielsen, J. (2015) 'Microbial acetyl-CoA metabolism and metabolic engineering', *Metabolic Engineering*. Academic Press, 28, pp. 28–42. doi: 10.1016/J.YMBEN.2014.11.009.

Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E. L. . (2001) 'Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes', *Journal of Molecular Biology*, 305(3), pp. 567–580. doi: 10.1006/jmbi.2000.4315.

Kuhn, H. M., Meier-Dieter, U. and Mayer, H. (1988) 'ECA, the enterobacterial common antigen', *FEMS Microbiology Letters*. Narnia, 54(3), pp. 195–222. doi: 10.1016/0378-1097(88)90002-X.

Kuhn, H. M., Neter, E. and Mayer, H. (1983) 'Modification of the lipid moiety of the enterobacterial common antigen by the "Pseudomonas factor"', *Infection and Immunity*, 40(2), pp. 696–700.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.', *Molecular biology and evolution*. Oxford University Press, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.

Kunin, C. M. (1963) 'SEPARATION, CHARACTERIZATION, AND BIOLOGICAL SIGNIFICANCE

OF A COMMON', *The Journal of experimental medicine*. The Rockefeller University Press, 118(4), pp. 565–586. doi: 10.1084/jem.118.4.565.

Laaberki, M. H., Pfeffer, J., Clarke, A. J. and Dworkin, J. (2011) 'O-acetylation of peptidoglycan is required for proper cell separation and S-layer anchoring in Bacillus anthracis', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 286(7), pp. 5278–5288. doi: 10.1074/jbc.M110.183236.

Lacroix, J. M., Lanfroy, E., Cogez, V., Lequette, Y., Bohin, A. and Bohin, J. P. (1999) 'The mdoC gene of Escherichia coli encodes a membrane protein that is required for succinylation of osmoregulated periplasmic glucans.', *Journal of bacteriology*. American Society for Microbiology, 181(12), pp. 3626–31.

Lanzilao, L., Stefanetti, G., Saul, A., MacLennan, C. A., Micoli, F. and Rondini, S. (2015) 'Strain selection for generation of O-antigen-based glycoconjugate vaccines against invasive nontyphoidal Salmonella disease', *PLoS ONE*. Edited by N. J. Mantis. Public Library of Science, 10(10), p. e0139847. doi: 10.1371/journal.pone.0139847.

Larentis, A. L., Nicolau, J. F. M. Q., Esteves, G. dos S., Vareschini, D. T., de Almeida, F. V. R., dos Reis, M. G., Galler, R. and Medeiros, M. A. (2014) 'Evaluation of pre-induction temperature, cell growth at induction and IPTG concentration on the expression of a leptospiral protein in E. coli using shaking flasks and microbioreactor.', *BMC research notes*. BioMed Central, 7(1), p. 671. doi: 10.1186/1756-0500-7-671.

Lee, L.-C., Lee, Y.-L., Leu, R.-J. and Shaw, J.-F. (2006) 'Functional role of catalytic triad and oxyanion hole-forming residues on enzyme activity of Escherichia coli thioesterase I/protease I/phospholipase L1.', *The Biochemical journal*, 397(1), pp. 69–76. doi: 10.1042/BJ20051645.

Leigh, J. A., Reed, J. W., Hanks, J. F., Hirsch, A. M. and Walker, G. C. (1987) 'Rhizobium meliloti mutants that fail to succinylate their Calcofluor-binding exopolysaccharide are defective in nodule invasion', *Cell*. Cell Press, 51(4), pp. 579–587. doi: 10.1016/0092-8674(87)90127-9.

Leigh, J. A., Signer, E. R. and Walker, G. C. (1985) 'Exopolysaccharide-deficient mutants of Rhizobium meliloti that form ineffective nodules.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences,

82(18), pp. 6231–6235. doi: 10.1073/pnas.82.18.6231.

Lerouge, I. (2001) 'O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions', *FEMS Microbiology Reviews*, 25(5), pp. 17–47. doi: 10.1016/S0168-6445(01)00070-5.

Leščić Ašler, I., Ivić, N., Kovačić, F., Schell, S., Knorr, J., Krauss, U., Wilhelm, S., Kojić-Prodić, B. and Jaeger, K.-E. (2010) 'Probing enzyme promiscuity of SGNH hydrolases.', *Chembiochem : a European journal of chemical biology*, 11(15), pp. 2158–67. doi: 10.1002/cbic.201000398.

Leščić Ašler, I., Štefanić, Z., Maršavelski, A., Vianello, R. and Kojić-Prodić, B. (2017) 'Catalytic Dyad in the SGNH Hydrolase Superfamily: In-depth Insight into Structural Parameters Tuning the Catalytic Process of Extracellular Lipase from Streptomyces rimosus', *ACS Chemical Biology*, 12(7), pp. 1928–1936. doi: 10.1021/acschembio.6b01140.

Li, L., Vorobyov, I. and Allen, T. W. (2013) 'The Different Interactions of Lysine and Arginine Side Chains with Lipid Membranes', *The Journal of Physical Chemistry B*. American Chemical Society, 117(40), pp. 11906–11920. doi: 10.1021/jp405418y.

Li, Y., Orlando, B. J. and Liao, M. (2019) 'Structural basis of lipopolysaccharide extraction by the LptB2FGC complex', *Nature*. Nature Publishing Group, 567(7749), pp. 486–490. doi: 10.1038/s41586-019-1025-6.

Liang, B. and Tamm, L. K. (2016) 'NMR as a tool to investigate the structure, dynamics and function of membrane proteins', *Nature Structural and Molecular Biology*. NIH Public Access, pp. 468–474. doi: 10.1038/nsmb.3226.

Lipman, N. S., Jackson, L. R., Trudel, L. J. and Weis-Garcia, F. (2005) 'Monoclonal Versus Polyclonal Antibodies: Distinguishing Characteristics, Applications, and Information Resources', *ILAR Journal*. Narnia, 46(3), pp. 258–268. doi: 10.1093/ilar.46.3.258.

Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, ya N., Reeves, P. R., Wang, L., Senchenkova, S. N., Reeves, P. R. and Wang, L. (2013) 'Structural diversity in Salmonella O antigens and its genetic basis', *FEMS Microbiology Reviews*, 38(1), pp. 56–89. doi: 10.1111/1574-6976.12034.

Lizak, C., Gerber, S., Numao, S., Aebi, M. and Locher, K. P. (2011) 'X-ray structure of a bacterial oligosaccharyltransferase', *Nature*, 474. doi: 10.1038/nature10151.

Lo, Y.-C., Lin, S.-C., Shaw, J.-F. and Liaw, Y.-C. (2003) 'Crystal Structure of Escherichia coli Thioesterase I/Protease I/Lysophospholipase L1: Consensus Sequence Blocks Constitute the Catalytic Center of SGNH-hydrolases through a Conserved Hydrogen Bond Network', *Journal of Molecular Biology*, 330(3), pp. 539–551. doi: 10.1016/S0022-2836(03)00637-5.

Lopes, L., Andrade, C. T., Milas, M. and Rinaudo, M. (1992) 'Role of conformation and acetylation of xanthan on xanthan-guar interaction', *Carbohydrate Polymers*. Elsevier, 17(2), pp. 121–126. doi: 10.1016/0144-8617(92)90105-Y.

Luck, P. C., Freier, T., Steudel, C., Knirel, Y. A., Luneberg, E., Zahringer, U. and Helbig, J. H. (2001) 'A point mutation in the active site of Legionella pneumophila O-acetyltransferase results in modified lipopolysaccharide but does not influence virulence', *International journal of medical microbiology : IJMM*. Urban & Fischer, 291(5), pp. 345–352. doi: 10.1078/1438-4221-00140.

Lugowski, C., Romanowska, E., Kenne, L. and Lindberg, B. (1983) 'Identification of a trisaccharide repeating-unit in the enterobacterial common-antigen', *Carbohydrate Research*. Elsevier, 118(C), pp. 173–181. doi: 10.1016/0008-6215(83)88045-8.

Luo, Q., Yang, X., Yu, S., Shi, H., Wang, K., Xiao, L., Zhu, G., Sun, C., Li, T., Li, D., Zhang, X., Zhou, M. and Huang, Y. (2017) 'Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG', *Nature Structural & Molecular Biology*. Nature Publishing Group, 24(5), pp. 469–474. doi: 10.1038/nsmb.3399.

Lynch, M. F., Blanton, E. M., Bulens, S., Polyak, C., Vojdani, J., Stevenson, J., Medalla, F., Barzilay, E., Joyce, K., Barrett, T. and Mintz, E. D. (2009) 'Typhoid Fever in the United States, 1999-2006', *JAMA*. American Medical Association, 302(8), p. 859. doi: 10.1001/jama.2009.1229.

Ma, D., Wang, Z., Merrikh, C. N., Lang, K. S., Lu, P., Li, X., Merrikh, H., Rao, Z. and Xu, W. (2018) 'Crystal structure of a membrane-bound O-acyltransferase', *Nature*. Nature Publishing Group, p. 1. doi: 10.1038/s41586-018-0568-2.

Majowicz, S. E., Musto, J., Scallan, E., Angulo, F. J., Kirk, M., O'Brien, S. J., Jones, T. F.,

Fazil, A. and Hoekstra, R. M. (2010) 'The Global Burden of Nontyphoidal Salmonella Gastroenteritis', *Clinical Infectious Diseases*, 50(6), pp. 882–889. doi: 10.1086/650733.

Maldonado, R. F., Sá-Correia, I. and Valvano, M. A. (2016) *Lipopolysaccharide modification in gram-negative bacteria during chronic infection*, *FEMS Microbiology Reviews*. Edited by C. Whitfield. Narnia. doi: 10.1093/femsre/fuw007.

Matthews, R. G. (1999) 'Bi Bi Ping Pong: Is there really such a mechanism', in Frey, P. A. and Northrop, D. B. (eds) *Enzymatic mechanisms*. IOS Press, pp. 155–161.

Mattick, J. S. (2002) 'Type IV Pili and Twitching Motility', *Annual Review of Microbiology*, 56(1), pp. 289–314. doi: 10.1146/annurev.micro.56.012302.160938.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., Read, R. J. and IUCr (2007) '*Phaser* crystallographic software', *Journal of Applied Crystallography*. International Union of Crystallography, 40(4), pp. 658–674. doi: 10.1107/S0021889807021206.

McQuiston, J. R., Fields, P. I., Tauxe, R. V. and Logsdon, J. M. (2008) 'Do Salmonella carry spare tyres?', *Trends in Microbiology*. Elsevier Current Trends, 16(4), pp. 142–148. doi: 10.1016/J.TIM.2008.01.009.

Menéndez, N., Nur-e-Alam, M., Braña, A. F., Rohr, J., Salas, J. A. and Méndez, C. (2004a) 'Biosynthesis of the Antitumor Chromomycin A3 in Streptomyces griseus: Analysis of the Gene Cluster and Rational Design of Novel Chromomycin Analogs', *Chemistry & Biology*, 11(1), pp. 21–32. doi: 10.1016/j.chembiol.2003.12.011.

Menéndez, N., Nur-e-Alam, M., Braña, A. F., Rohr, J., Salas, J. A. and Méndez, C. (2004b) 'Tailoring modification of deoxysugars during biosynthesis of the antitumour drug chromomycin A3 by Streptomyces griseus ssp. griseus', *Molecular Microbiology*. Blackwell Science Ltd, 53(3), pp. 903–915. doi: 10.1111/j.1365-2958.2004.04166.x.

Mergulhão, F. J. M., Taipa, M. A., Cabral, J. M. S. and Monteiro, G. A. (2004) 'Evaluation of bottlenecks in proinsulin secretion by Escherichia coli', *Journal of Biotechnology*. Elsevier, 109(1–2), pp. 31–43. doi: 10.1016/J.JBIOTEC.2003.10.024.

Meziane-Cherif, D., Stogios, P. J., Evdokimova, E., Egorova, O., Savchenko, A. and Courvalin, P. (2015) 'Structural and functional adaptation of vancomycin resistance vant

serine racemases', *mBio*. American Society for Microbiology, 6(4), pp. e00806-15. doi: 10.1128/mBio.00806-15.

Miceli, M., Casati, S., Ottria, R., Di Leo, S., Eberini, I., Palazzolo, L., Parravicini, C. and Ciuffreda, P. (2019) 'Set-up and validation of a high throughput screening method for human monoacylglycerol lipase (MAGL) based on a new red fluorescent probe', *Molecules*. Multidisciplinary Digital Publishing Institute (MDPI), 24(12). doi: 10.3390/molecules24122241.

Micoli, F., Ravenscroft, N., Cescutti, P., Stefanetti, G., Londero, S., Rondini, S. and Maclennan, C. a. (2013) 'Structural analysis of O-polysaccharide chains extracted from different Salmonella Typhimurium strains.', *Carbohydrate research*. Elsevier Ltd, 385C, pp. 1–8. doi: 10.1016/j.carres.2013.12.003.

Miller, K. J., Kennedy, E. P. and Reinhold, V. N. (1986) 'Osmotic adaptation by gram-negative bacteria: Possible role for periplasmic oligosaccharides', *Science*, 231(4733), pp. 48–51. doi: 10.1126/science.3941890.

Miller, S. I., Ernst, R. K. and Bader, M. W. (2005) 'LPS, TLR4 and infectious disease diversity', *Nature Reviews Microbiology*. Nature Publishing Group, 3(1), pp. 36–46. doi: 10.1038/nrmicro1068.

Miller, S. I. and Salama, N. R. (2018) 'The gram-negative bacterial periplasm: Size matters', *PLOS Biology*. Public Library of Science, 16(1), p. e2004935. doi: 10.1371/journal.pbio.2004935.

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S.-Y., Bateman, A., *et al.* (2015) 'The InterPro protein families database: the classification resource after 15 years.', *Nucleic acids research*. Oxford University Press, 43(Database issue), pp. D213-21. doi: 10.1093/nar/gku1243.

Mølgaard, A., Kauppinen, S. and Larsen, S. (2000) 'Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases', *Structure*, 8(4), pp. 373–383. doi: 10.1016/S0969-2126(00)00118-0.

Morona, R., Daniels, C. and Van Den Bosch, L. (2003) 'Genetic modulation of Shigella flexneri 2a lipopolysaccharide O antigen modal chain length reveals that it has been

optimized for virulence', *Microbiology*, pp. 925–939. doi: 10.1099/mic.0.26141-0.

Moss, S. J., Bai, L., Toelzer, S., Carroll, B. J., Mahmud, T., Yu, T. W. and Floss, H. G. (2002) 'Identification of Asm19 as an acyltransferase attaching the biologically essential ester side chain of ansamitocins using N-desmethyl-4,5-desepoxymaytansinol, not maytansinol, as its substrate', *Journal of the American Chemical Society*. American Chemical Society, 124(23), pp. 6544–6545. doi: 10.1021/ja020214b.

Moxon, E. R. and Kroll, J. S. (1990) 'The Role of Bacterial Polysaccharide Capsules as Virulence Factors', in. Springer, Berlin, Heidelberg, pp. 65–85. doi: 10.1007/978-3-642-74694-9_4.

Moynihan, P. J. and Clarke, A. J. (2010) 'O-acetylation of peptidoglycan in gram-negative bacteria: identification and characterization of peptidoglycan O-acetyltransferase in Neisseria gonorrhoeae.', *The Journal of biological chemistry*, 285(17), pp. 13264–73. doi: 10.1074/jbc.M110.107086.

Moynihan, P. J. and Clarke, A. J. (2011) 'O-Acetylated peptidoglycan: Controlling the activity of bacterial autolysins and lytic enzymes of innate immune systems', *The International Journal of Biochemistry & Cell Biology*, 43(12), pp. 1655–1659. doi: 10.1016/j.biocel.2011.08.007.

Moynihan, P. J. and Clarke, A. J. (2013) 'Assay for peptidoglycan O-acetyltransferase: a potential new antibacterial target.', *Analytical biochemistry*, 439(2), pp. 73–9. doi: 10.1016/j.ab.2013.04.022.

Moynihan, P. J. and Clarke, A. J. (2014a) 'Mechanism of action of peptidoglycan O-acetyltransferase B involves a Ser-His-Asp catalytic triad.', *Biochemistry*. American Chemical Society, 53(39), pp. 6243–51. doi: 10.1021/bi501002d.

Moynihan, P. J. and Clarke, A. J. (2014b) 'Substrate specificity and kinetic characterization of peptidoglycan O-acetyltransferase B from Neisseria gonorrhoeae.', *The Journal of biological chemistry*, 289(24), pp. 16748–60. doi: 10.1074/jbc.M114.567388.

Moynihan, P. J., Sychantha, D. and Clarke, A. J. (2014) 'Chemical biology of peptidoglycan acetylation and deacetylation', *Bioorganic Chemistry*, pp. 44–50. doi: 10.1016/j.bioorg.2014.03.010.

Murray, G. L., Attridge, S. R. and Morona, R. (2006) 'Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of Salmonella enterica serovar Typhimurium with macrophages and complement.', *Journal of bacteriology*. American Society for Microbiology (ASM), 188(7), pp. 2735–9. doi: 10.1128/JB.188.7.2735-2739.2006.

Murray, I. A., Lewendon, A., Williams, J. A., Cullis, P. M., Shaw, W. V. and Leslie, A. G. W. (1991) 'Alternative Binding Modes for Chloramphenicol and 1-Substituted Chloramphenicol Analogues Revealed by Site-Directed Mutagenesis and X-ray Crystallography of Chloramphenicol Acetyltransferase', *Biochemistry*, 30(15), pp. 3763–3770. doi: 10.1021/bi00229a025.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., Vagin, A. A. and IUCr (2011) '*REFMAC* 5 for the refinement of macromolecular crystal structures', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 67(4), pp. 355–367. doi: 10.1107/S0907444911001314.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. and Dodson, E. J. (1999) 'Efficient anisotropic refinement of macromolecular structures using FFT', *Acta Crystallographica Section D Biological Crystallography*, 55(1), pp. 247–255. doi: 10.1107/S090744499801405X.

Navarre, W. W. and Schneewind, O. (1999) 'Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology (ASM), 63(1), pp. 174–229.

Nelson, A. L., Roche, A. M., Gould, J. M., Chim, K., Ratner, A. J. and Weiser, J. N. (2007) 'Capsule Enhances Pneumococcal Colonization by Limiting Mucus-Mediated Clearance', *Infection and Immunity*. American Society for Microbiology (ASM), 75(1), p. 83. doi: 10.1128/IAI.01475-06.

Nicholls, R. A., Long, F., Murshudov, G. N. and IUCr (2012) 'Low-resolution refinement tools in *REFMAC* 5', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 68(4), pp. 404–417. doi: 10.1107/S090744491105606X.

Nikaido, H. (1999) ‘Microdermatology: cell surface in the interaction of microbes with the external world.’, *Journal of bacteriology*. American Society for Microbiology (ASM), 181(1), pp. 4–8.

Notredame, C., Higgins, D. G. and Heringa, J. (2000) ‘T-coffee: A novel method for fast and accurate multiple sequence alignment’, *Journal of Molecular Biology*. Academic Press, 302(1), pp. 205–217. doi: 10.1006/jmbi.2000.4042.

Okuda, S., Freinkman, E. and Kahne, D. (2012) ‘Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in E. coli.’, *Science (New York, N.Y.)*. NIH Public Access, 338(6111), pp. 1214–7. doi: 10.1126/science.1228984.

Okuda, S., Sherman, D. J., Silhavy, T. J., Ruiz, N. and Kahne, D. (2016) ‘Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model.’, *Nature reviews. Microbiology*. doi: 10.1038/nrmicro.2016.25.

Ōmura, S. (2002) *Macrolide antibiotics : chemistry, biology, and practice*. Academic Press.

Osborn, M. J., Rosen, S. M., Rothfield, L., Zeleznick, L. D. and Horecker, B. L. (1964) ‘Lipopolysaccharide of the gram-negative cell wall’, *Science*. American Association for the Advancement of Science, pp. 783–789. doi: 10.1126/science.145.3634.783.

Owen, S. V, Wenner, N., Canals, R., Makumi, A., Hammarlöf, D. L., Gordon, M. A., Aertsen, A., Feasey, N. A. and Hinton, J. C. D. (2017) ‘Characterization of the Prophage Repertoire of African Salmonella Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1.’, *Frontiers in microbiology*. Frontiers Media SA, 8, p. 235. doi: 10.3389/fmicb.2017.00235.

Owens, T. W., Taylor, R. J., Pahil, K. S., Bertani, B. R., Ruiz, N., Kruse, A. C. and Kahne, D. (2019) ‘Structural basis of unidirectional export of lipopolysaccharide to the cell surface’, *Nature*. Nature Publishing Group, 567(7749), pp. 550–553. doi: 10.1038/s41586-019-1039-0.

Pacios Bras, C., Jordá, M. A., Wijfjes, a H., Harteveld, M., Stuurman, N., Thomas-Oates, J. E. and Spaink, H. P. (2000) ‘A Lotus japonicus nodulation system based on heterologous expression of the fucosyl transferase NodZ and the acetyl transferase NoIL in Rhizobium leguminosarum.’, *Molecular plant-microbe interactions : MPMI*, 13(4), pp.

475–479. doi: 10.1094/MPMI.2000.13.4.475.

Page, F., Altabe, S., Hugouvieux-Cotte-Pattat, N., Lacroix, J.-M., Robert-Baudouy, J. and Bohin, J.-P. (2001) 'Osmoregulated Periplasmic Glucan Synthesis Is Required for Erwinia chrysanthemi Pathogenicity', *Journal of Bacteriology*. American Society for Microbiology Journals, 183(10), pp. 3134–3141. doi: 10.1128/JB.183.10.3134-3141.2001.

Parisi, A., Crump, J. A., Stafford, R., Glass, K., Howden, B. P. and Kirk, M. D. (2019) 'Increasing incidence of invasive nontyphoidal Salmonella infections in Queensland, Australia, 2007-2016', *PLOS Neglected Tropical Diseases*. Edited by T. C. Darton. Public Library of Science, 13(3), p. e0007187. doi: 10.1371/journal.pntd.0007187.

Patrone, J. B. and Stein, D. C. (2007) 'Effect of gonococcal lipooligosaccharide variation on human monocytic cytokine profile', *BMC Microbiology*. BioMed Central, 7(1), p. 7. doi: 10.1186/1471-2180-7-7.

Pearson, C. R., Tindall, S. N., Herman, R., Jenkins, H. T., Thomas, G. H., Potts, J. R. and van der Woude, M. W. (2020) *Acetylation of surface carbohydrates in bacterial pathogens requires coordinated action of a two-domain membrane-bound acyltransferase. Manuscript submitted for publication*.

Peltola, H., Mäkelä, P. H., Käyhty, H., Jousimies, H., Herva, E., Hällström, K., Sivonen, A., Renkonen, O.-V., Pettay, O., Karanko, V., Ahvonen, P. and Sarna, S. (1977) 'Clinical Efficacy of Meningococcus Group A Capsular Polysaccharide Vaccine in Children Three Months to Five Years of Age', *New England Journal of Medicine*. Massachusetts Medical Society , 297(13), pp. 686–691. doi: 10.1056/NEJM197709292971302.

Percy, M. G. and Gründling, A. (2014) 'Lipoteichoic Acid Synthesis and Function in Gram-Positive Bacteria', *Annual review of Microbiology*, 68(1), pp. 81–100. doi: 10.1146/annurev-micro-091213-112949.

Perepelov, A. V, Shekht, M. E., Liu, B., Shevelev, S. D., Ledov, V. A., Senchenkova, S. N., L'vov, V. L., Shashkov, A. S., Feng, L., Aparin, P. G., Wang, L., Knirel, Y. A., Allison, G. E., Verma, N. K., Ansaruzzaman, M., *et al.* (2012) 'Shigella flexneri O-antigens revisited: final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity.', *FEMS immunology and medical microbiology*. The Oxford University Press, 66(2), pp. 201–10. doi: 10.1111/j.1574-695X.2012.01000.x.

Perrakis, A., Morris, R. and Lamzin, V. S. (1999) 'Automated protein model building combined with iterative structure refinement', *Nature Structural Biology*. Nature Publishing Group, 6(5), pp. 458–463. doi: 10.1038/8263.

Peterson, A. A., Haug, A. and McGroarty, E. J. (1986) 'Physical properties of short- and long-O-antigen-containing fractions of lipopolysaccharide from Escherichia coli 0111:B4', *Journal of Bacteriology*. American Society for Microbiology Journals, 165(1), pp. 116–122. doi: 10.1128/jb.165.1.116-122.1986.

Pfeffer, J. M., Weadge, J. T. and Clarke, A. J. (2013) 'Mechanism of action of Neisseria gonorrhoeae O-acetylpeptidoglycan esterase, an SGNH serine esterase.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 288(4), pp. 2605–13. doi: 10.1074/jbc.M112.436352.

Pollitt, S. and Zalkin, H. (1983) 'Role of primary structure and disulfide bond formation in beta-lactamase secretion.', *Journal of bacteriology*, 153(1), pp. 27–32.

Preston, A., Mandrell, R. E., Gibson, B. W. and Apicella, M. A. (1996) 'The lipooligosaccharides of pathogenic gram-negative bacteria', *Critical Reviews in Microbiology*. Taylor & Francis, pp. 139–180. doi: 10.3109/10408419609106458.

Quiocho, F. (1986) 'Carbohydrate-Binding Proteins: Tertiary Structures and Protein-Sugar Interactions', *Annual Review of Biochemistry*, 55(1), pp. 287–315. doi: 10.1146/annurev.biochem.55.1.287.

Rabsch, W. (2007) 'Salmonella Typhimurium Phage Typing for Pathogens', in. Humana Press, pp. 177–211. doi: 10.1007/978-1-59745-512-1_10.

Racey, L. A. and Byvoet, P. (1971) 'Histone acetyltransferase in chromatin: Evidence for in vitro enzymatic transfer of acetate from acetyl-coenzyme A to histones', *Experimental Cell Research*. Academic Press, 64(2), pp. 366–370. doi: 10.1016/0014-4827(71)90089-9.

Raetz, C. R. H. and Whitfield, C. (2002) 'Lipopolysaccharide Endotoxins', *Annual Review of Biochemistry*, 71(1), pp. 635–700. doi: 10.1146/annurev.biochem.71.110601.135414.

Rajagopal, S., Eis, N., Bhattacharya, M. and Nickerson, K. W. (2003) 'Membrane-derived oligosaccharides (MDOs) are essential for sodium dodecyl sulfate resistance in

Escherichia coli', *FEMS Microbiology Letters*. John Wiley & Sons, Ltd (10.1111), 223(1), pp. 25–31. doi: 10.1016/S0378-1097(03)00323-9.

Rajalingam, D., Loftis, C., Xu, J. J. and Kumar, T. K. S. (2009) 'Trichloroacetic acid-induced protein precipitation involves the reversible association of a stable partially structured intermediate.', *Protein science : a publication of the Protein Society*. Wiley-Blackwell, 18(5), pp. 980–93. doi: 10.1002/pro.108.

Ramos-Morales, F., Prieto, A. I., Beuzón, C. R., Holden, D. W. and Casadesús, J. (2003) 'Role for Salmonella enterica enterobacterial common antigen in bile resistance and virulence.', *Journal of bacteriology*. American Society for Microbiology Journals, 185(17), pp. 5328–32. doi: 10.1128/jb.185.17.5328-5332.2003.

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. and Deber, C. M. (2009) 'Detergent binding explains anomalous SDS-PAGE migration of membrane proteins', *Proceedings of the National Academy of Sciences*, 106(6), pp. 1760–1765. doi: 10.1073/pnas.0813167106.

Rauwerdink, A. and Kazlauskas, R. J. (2015) *How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: The Serine-Histidine-Aspartate Catalytic Triad of ??/??-Hydrolase Fold Enzymes*, *ACS Catalysis*. American Chemical Society. doi: 10.1021/acscatal.5b01539.

Ravenscroft, N., Cescutti, P., Gavini, M., Stefanetti, G., MacLennan, C. A. A., Martin, L. B. B. and Micoli, F. (2015) 'Structural analysis of the O-acetylated O-polysaccharide isolated from Salmonella paratyphi A and used for vaccine preparation', *Carbohydrate Research*. Elsevier, 404, pp. 108–116. doi: 10.1016/j.carres.2014.12.002.

RCSB (1999) *PDB Chemical Component COA*. Available at: http://ligand-expo.rcsb.org/reports/C/COA/index.html (Accessed: 2 August 2019).

Reddy, E. A., Shaw, A. V and Crump, J. A. (2010) 'Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis.', *The Lancet. Infectious diseases*. NIH Public Access, 10(6), pp. 417–32. doi: 10.1016/S1473-3099(10)70072-4.

Reeves, P. (1994) 'Biosynthesis and assembly of lipopolysaccharide', *New Comprehensive Biochemistry*. Elsevier, 27(C), pp. 281–317. doi: 10.1016/S0167-7306(08)60416-0.

Reinhold, B. B., Chan, S. Y., Reuber, T. L., Marra, A., Walker, G. C. and Reinhold, V. N. (1994) 'Detailed structural characterization of succinoglycan, the major exopolysaccharide of Rhizobium meliloti Rm1021.', *Journal of bacteriology*. American Society for Microbiology (ASM), 176(7), pp. 1997–2002. doi: 10.1128/jb.176.7.1997-2002.1994.

Resh, M. D. (2006) 'Palmitoylation of ligands, receptors, and intracellular signaling molecules.', *Science's STKE : signal transduction knowledge environment*, pp. 1–14. doi: 10.1126/stke.3592006re14.

Reth, M. (2013) 'Matching cellular dimensions with molecular sizes', *Nature Immunology*, pp. 765–767. doi: 10.1038/ni.2621.

Reuber, T. L. and Walker, G. C. (1993) 'The acetyl substituent of succinoglycan is not necessary for alfalfa nodule invasion by Rhizobium meliloti Rm1021.', *Journal of bacteriology*, 175(11), pp. 3653–5. doi: 10.1128/jb.175.11.3653-3655.1993.

Rezania, S., Amirmozaffari, N., Tabarraei, B., Jeddi-Tehrani, M., Zarei, O., Alizadeh, R., Masjedian, F. and Zarnani, A. H. (2011) 'Extraction, Purification and Characterization of Lipopolysaccharide from Escherichia coli and Salmonella typhi.', *Avicenna journal of medical biotechnology*. Avicenna Research Institute, 3(1), pp. 3–9.

Riley, L. M., Weadge, J. T., Baker, P., Robinson, H., Codee, J. D. C., Tipton, P. A., Ohman, D. E., Howell, P. L., Codée, J. D. C., Tipton, P. A., Ohman, D. E. and Howell, P. L. (2013) 'Structural and Functional Characterization of Pseudomonas aeruginosa AlgX: ROLE OF AlgX IN ALGINATE ACETYLATION', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 288(31), pp. 22299–22314. doi: 10.1074/jbc.M113.484931.

Roberts, I. S. (1996) 'THE BIOCHEMISTRY AND GENETICS OF CAPSULAR POLYSACCHARIDE PRODUCTION IN BACTERIA', *Annual Review of Microbiology*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 50(1), pp. 285–315. doi: 10.1146/annurev.micro.50.1.285.

Robison, A. D., Sun, S., Poyton, M. F., Johnson, G. A., Pellois, J.-P., Jungwirth, P., Vazdar, M. and Cremer, P. S. (2016) 'Polyarginine Interacts More Strongly and Cooperatively than Polylysine with Phospholipid Bilayers.', *The journal of physical chemistry. B*. NIH

Public Access, 120(35), pp. 9287–96. doi: 10.1021/acs.jpcb.6b05604.

Rohde, H., Frankenberger, S., Zähringer, U. and Mack, D. (2010) 'Structure, function and contribution of polysaccharide intercellular adhesin (PIA) to Staphylococcus epidermidis biofilm formation and pathogenesis of biomaterial-associated infections', *European Journal of Cell Biology*. Urban & Fischer, 89(1), pp. 103–111. doi: 10.1016/J.EJCB.2009.10.005.

Rondini, S., Lanzilao, L., Necchi, F., O'Shaughnessy, C. M., Micoli, F., Saul, A. and MacLennan, C. A. (2013) 'Invasive african salmonella typhimurium induces bactericidal antibodies against O-antigens', *Microbial Pathogenesis*, 63, pp. 19–23. doi: 10.1016/j.micpath.2013.05.014.

Rosenberg, H. F. (1998) 'Isolation of recombinant secretory proteins by limited induction and quantitative harvest', *BioTechniques*, 24(2), pp. 188–192. doi: 10.2144/98242bm03.

Roset, M. S., Ciocchini, A. E., Ugalde, R. A. and Inon de Iannino, N. (2006) 'The Brucella abortus Cyclic -1,2-Glucan Virulence Factor Is Substituted with O-Ester-Linked Succinyl Residues', *Journal of Bacteriology*. American Society for Microbiology, 188(14), pp. 5003–5013. doi: 10.1128/JB.00086-06.

Rybka, J. and Gamian, A. (2006) 'Determination of endotoxin by the measurement of the acetylated methyl glycoside derivative of Kdo with gas–liquid chromatography-mass spectrometry', *Journal of Microbiological Methods*. Elsevier, 64(2), pp. 171–184. doi: 10.1016/J.MIMET.2005.04.029.

Sadovskaya, I., Vinogradov, E., Flahaut, S., Kogan, G. and Jabbouri, S. S. (2005) 'Extracellular Carbohydrate-Containing Polymers of a Model Biofilm-Producing Strain, Staphylococcus epidermidis RP62A', *INFECTION AND IMMUNITY*, 73(5), pp. 3007–3017. doi: 10.1128/IAI.73.5.3007-3017.2005.

Salmonella Subcommittee of the Nomenclature Committee of the International Society for Microbiology (1934) 'The genus salmonella ligniéres, 1900: Issued by the salmonella subcommittee of the nomenclature committee of the international society for microbiology', *Journal of Hygiene*. Cambridge University Press, 34(3), pp. 333–350. doi: 10.1017/S0022172400034677.

Samuel, G. and Reeves, P. (2003) *Biosynthesis of O-antigens: Genes and pathways*

involved in nucleotide sugar precursor synthesis and O-antigen assembly, *Carbohydrate Research*. doi: 10.1016/j.carres.2003.07.009.

Santos, N. C., Silva, A. C., Castanho, M. A. R. B., Martins-Silva, J. and Saldanha, C. (2003) 'Evaluation of Lipopolysaccharide Aggregation by Light Scattering Spectroscopy', *ChemBioChem*. John Wiley & Sons, Ltd, 4(1), pp. 96–100. doi: 10.1002/cbic.200390020.

Sato, M. and Mueckler, M. (1999) 'A conserved amino acid motif (R-X-G-R-R) in the Glut1 glucose transporter is an important determinant of membrane topology', *Journal of Biological Chemistry*, 274(35), pp. 24721–24725. doi: 10.1074/jbc.274.35.24721.

Schein, C. H. (1989) 'Production of soluble recombinant proteins in bacteria', *Nature Biotechnology*, 7(11), pp. 1141–1149. doi: 10.1038/nbt1189-1141.

Schleifer, K. H. and Kandler, O. (1972) 'Peptidoglycan types of bacterial cell walls and their taxonomic implications.', *Bacteriological reviews*. American Society for Microbiology (ASM), 36(4), pp. 407–77.

Schultink, A., Naylor, D., Dama, M. and Pauly, M. (2015) 'The role of the plant-specific ALTERED XYLOGLUCAN9 protein in Arabidopsis cell wall polysaccharide O-acetylation.', *Plant physiology*. American Society of Plant Biologists, 167(4), pp. 1271–83. doi: 10.1104/pp.114.256479.

Seepersaud, R., Sychantha, D., Bensing, B. A., Clarke, A. J. and Sullam, P. M. (2017) 'O-acetylation of the serine-rich repeat glycoprotein GspB is coordinated with accessory Sec transport', *PLoS Pathogens*. Public Library of Science, 13(8), p. e1006558. doi: 10.1371/journal.ppat.1006558.

Seth-Smith, H. M. B. (2008) 'SPI-7: Salmonella's Vi-Encoding Pathogenicity Island', *The Journal of Infection in Developing Countries*, 2(04), pp. 267–271. doi: 10.3855/jidc.220.

Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. and Pace, C. N. (2002) 'The effect of net charge on the solubility, activity, and stability of ribonuclease Sa', *Protein Science*. Wiley-Blackwell, 10(6), pp. 1206–1215. doi: 10.1110/ps.440101.

Shen, H.-H., Lithgow, T. and Martin, L. (2013) 'Reconstitution of membrane proteins into model membranes: seeking better ways to retain protein activities.', *International journal of molecular sciences*. Multidisciplinary Digital Publishing Institute (MDPI), 14(1),

pp. 1589–607. doi: 10.3390/ijms14011589.

Sherman, D. J., Lazarus, M. B., Murphy, L., Liu, C., Walker, S., Ruiz, N. and Kahne, D. (2014) 'Decoupling catalytic activity from biological function of the ATPase that powers lipopolysaccharide transport.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(13), pp. 4982–7. doi: 10.1073/pnas.1323516111.

Silhavy, T. J., Kahne, D. and Walker, S. (2010) 'The bacterial cell envelope.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, p. a000414. doi: 10.1101/cshperspect.a000414.

Simpson, B. W., May, J. M., Sherman, D. J., Kahne, D. and Ruiz, N. (2015) 'Lipopolysaccharide transport to the cell surface: Biosynthesis and extraction from the inner membrane', *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2015.0029.

Sinz, A. (2010) 'Investigation of protein–protein interactions in living cells by chemical crosslinking and mass spectrometry', *Analytical and Bioanalytical Chemistry*. Springer-Verlag, 397(8), pp. 3433–3440. doi: 10.1007/s00216-009-3405-5.

Slabas, A. R. and Fawcett, T. (1992) 'The biochemistry and molecular biology of plant lipid biosynthesis', *Plant Molecular Biology*, 19(1), pp. 169–191. doi: 10.1007/BF00015613.

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.', *Journal of Bacteriology*, 178(20), pp. 5904–5909.

Slauch, J. M., Mahan, M. J., Michetti, P., Neutra, M. R. and Mekalanos, J. J. (1995) 'Acetylation (O-factor 5) affects the structural and immunological properties of Salmonella typhimurium lipopolysaccharide O antigen.', *Infection and immunity*, 63(2), pp. 437–41.

Sohlenkamp, C. and Geiger, O. (2016) 'Bacterial membrane lipids: diversity in structures and pathways', *FEMS Microbiology Reviews*. Edited by F. Narberhaus. Narnia, 40(1), pp. 133–159. doi: 10.1093/femsre/fuv008.

Sørensen, U. B. S., Henrichsen, J., Chen, H.-C. and Szu, S. C. (1990) 'Covalent linkage between the capsular polysaccharide and the cell wall peptidoglycan of Streptococcus pneumoniae revealed by immunochemical methods', *Microbial Pathogenesis*, 8(5), pp. 325–334. doi: 10.1016/0882-4010(90)90091-4.

Spencer, B. L., Saad, J. S., Shenoy, A. T., Orihuela, C. J. and Nahm, M. H. (2017) 'Position of O-acetylation within the capsular repeat unit impacts the biological properties of pneumococcal serotypes 33A and 33F', *Infection and Immunity*. American Society for Microbiology (ASM), 85(7). doi: 10.1128/IAI.00132-17.

Sperandeo, P., Martorana, A. M. and Polissi, A. (2017) 'The lipopolysaccharide transport (Lpt) machinery: A nonconventional transporter for lipopolysaccharide assembly at the outer membrane of Gram-negative bacteria', *Journal of Biological Chemistry*, 292(44), pp. 17981–17990. doi: 10.1074/jbc.R117.802512.

SSI Diagnostica (2013) *Salmonella Antisera*, *SSI Diagnostics*. Available at: https://www.ssidiagnostica.com/upload/images/webshop/products/115/downloads/85711-3-salmonella-brochre-web.pdf (Accessed: 5 June 2019).

Stanaway, J. D., Reiner, R. C., Blacker, B. F., Goldberg, E. M., Khalil, I. A., Troeger, C. E., Andrews, J. R., Bhutta, Z. A., Crump, J. A., Im, J., Marks, F., Mintz, E., Eun Park, S., M Zaidi, A. K., Abebe, Z., *et al.* (2019) 'The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017'. doi: 10.1016/S1473-3099(18)30685-6.

Stanfield, R. L., Fieser, T. M., Lerner, R. A. and Wilson, I. A. (1990) 'Crystal structures of an antibody to a peptide and its complex with peptide antigen at 2.8 Å', *Science*. American Association for the Advancement of Science, 248(4956), pp. 712–719. doi: 10.1126/science.2333521.

Stevenson, G., Kessler, A. and Reeves, P. R. (1995) *A plasmid-borne O-antigen chain length determinant and its relationship to other chain length determinants*, *FEMS Microbiology Letters*. doi: 10.1111/j.1574-6968.1995.tb07330.x.

Stimson, E., Virji, M., Makepeace, K., Dell, A., Morris, H. R., Payne, G., Saunders, J. R., Jennings, M. P., Barker, S., Panico, M., Blench, I. and Moxon, E. R. (1995) 'Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-

trideoxyhexose', *Molecular Microbiology*. Wiley/Blackwell (10.1111), 17(6), pp. 1201–1214. doi: 10.1111/j.1365-2958.1995.mmi_17061201.x.

Suckau, D. and Resemann, A. (2003) 'T3-Sequencing: Targeted Characterization of the N- and C-Termini of Undigested Proteins by Mass Spectrometry', *Analytical Chemistry*. American Chemical Society, 75(21), pp. 5817–5824. doi: 10.1021/ac034362b.

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.', *Journal of bacteriology*. American Society for Microbiology (ASM), 196(20), pp. 3656–66. doi: 10.1128/JB.02009-14.

Sun, Q., Lan, R., Wang, Y. Y., Wang, J., Xia, S., Wang, Y. Y., Zhang, J., Yu, D., Li, Z., Jing, H. and Xu, J. (2012) 'Identification of a divergent O-acetyltransferase gene oac 1b from Shigella flexneri serotype 1b strains.', *Emerging microbes & infections*. Nature Publishing Group, 1(9), p. e21. doi: 10.1038/emi.2012.22.

Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., Wu, C. H. and UniProt Consortium, the U. (2015) 'UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches.', *Bioinformatics (Oxford, England)*. Oxford University Press, 31(6), pp. 926–32. doi: 10.1093/bioinformatics/btu739.

Sweadner, K. J., Forte, M. and Nelsen, L. L. (1977) 'Filtration removal of endotoxin (pyrogens) in solution in different states of aggregation.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 34(4), pp. 382–5.

Sychantha, D., Brott, A. S., Jones, C. S. and Clarke, A. J. (2018) 'Mechanistic Pathways for Peptidoglycan O-Acetylation and De-O-Acetylation.', *Frontiers in microbiology*. Frontiers Media SA, 9, p. 2332. doi: 10.3389/fmicb.2018.02332.

Sychantha, D. and Clarke, A. J. (2018) 'Peptidoglycan Modification by the Catalytic Domain of Streptococcus pneumoniae OatA Follows a Ping-Pong Bi-Bi Mechanism of Action', *Biochemistry*, 57(16), pp. 2394–2401. doi: 10.1021/acs.biochem.8b00301.

Sychantha, D., Jones, C. S., Little, D. J., Moynihan, P. J., Robinson, H., Galley, N. F., Roper, D. I., Dowson, C. G., Howell, P. L. and Clarke, A. J. (2017) 'In vitro characterization of the antivirulence target of Gram-positive pathogens, peptidoglycan O-acetyltransferase A

(OatA)', *PLoS Pathogens*. Edited by G. Zhang. Public Library of Science, 13(10), p. e1006667. doi: 10.1371/journal.ppat.1006667.

Sychantha, D., Little, D. J., Chapman, R. N., Boons, G. J., Robinson, H., Howell, P. L. and Clarke, A. J. (2018) 'PatB1 is an O-acetyltransferase that decorates secondary cell wall polysaccharides', *Nature Chemical Biology*. Nature Publishing Group, 14(1), pp. 79–85. doi: 10.1038/nchembio.2509.

Tack, D. M., Marder, E. P., Griffin, P. M., Cieslak, P. R., Dunn, J., Hurd, S., Scallan, E., Lathrop, S., Muse, A., Ryan, P., Smith, K., Tobin-D'Angelo, M., Vugia, D. J., Holt, K. G., Wolpert, B. J., *et al.* (2019) 'Preliminary incidence and trends of infections with pathogens transmitted commonly through food — Foodborne Diseases Active Surveillance Network, 10 U.S. sites, 2015–2018', *American Journal of Transplantation*, 19(6), pp. 1859–1863. doi: 10.1111/ajt.15412.

Takamura Y, N. G. (1988) 'Changes in the intracellular concentration of acetyl-CoA and malonyl-CoA in relation to the carbon and energy metabolism of Escherichia coli K12.', *J Gen Microbiol.*, 134(8):224, pp. 2249–2253. doi: 10.1099/00221287-134-8-2249.

Tang, J., Lander, G. C., Olia, A. S., Olia, A., Li, R., Casjens, S., Prevelige, P., Cingolani, G., Baker, T. S. and Johnson, J. E. (2011) 'Peering down the barrel of a bacteriophage portal: the genome packaging and release valve in p22.', *Structure (London, England : 1993)*. NIH Public Access, 19(4), pp. 496–502. doi: 10.1016/j.str.2011.02.010.

Taroni, C., Jones, S. and Thornton, J. M. (2000) 'Analysis and prediction of carbohydrate binding sites', *Protein Engineering, Design and Selection*. Narnia, 13(2), pp. 89–98. doi: 10.1093/protein/13.2.89.

Thanweer, F., Tahiliani, V., Korres, H. and Verma, N. K. (2008) 'Topology and identification of critical residues of the O-acetyltransferase of serotype-converting bacteriophage, SF6, of Shigella flexneri', *Biochemical and Biophysical Research Communications*, 375(4), pp. 581–585. doi: 10.1016/j.bbrc.2008.08.069.

Thanweer, F. and Verma, N. K. (2012) 'Identification of critical residues of the serotype modifying O-acetyltransferase of Shigella flexneri.', *BMC biochemistry*. BioMed Central, 13, p. 13. doi: 10.1186/1471-2091-13-13.

Thiem, V. D., Lin, F.-Y. C., Canh, D. G., Son, N. H., Anh, D. D., Mao, N. D., Chu, C., Hunt, S.

W., Robbins, J. B., Schneerson, R. and Szu, S. C. (2011) 'The Vi conjugate typhoid vaccine is safe, elicits protective levels of IgG anti-Vi, and is compatible with routine infant vaccines.', *Clinical and vaccine immunology : CVI*. American Society for Microbiology (ASM), 18(5), pp. 730–5. doi: 10.1128/CVI.00532-10.

Tipton, K. A., Chin, C.-Y., Farokhyfar, M., Weiss, D. S. and Rather, P. N. (2018) 'Role of Capsule in Resistance to Disinfectants, Host Antimicrobials, and Desiccation in Acinetobacter baumannii'. doi: 10.1128/AAC.01188-18.

Uche, I. V., MacLennan, C. A. and Saul, A. (2017) 'A Systematic Review of the Incidence, Risk Factors and Case Fatality Rates of Invasive Nontyphoidal Salmonella (iNTS) Disease in Africa (1966 to 2014)', *PLOS Neglected Tropical Diseases*. Edited by S. Baker. Public Library of Science, 11(1), p. e0005118. doi: 10.1371/journal.pntd.0005118.

Unger, F. M. (1981) 'The Chemistry and Biological Significance of 3-Deoxy-d-manno-2-Octulosonic Acid (KDO)', *Advances in Carbohydrate Chemistry and Biochemistry*. Academic Press, 38, pp. 323–388. doi: 10.1016/S0065-2318(08)60313-3.

Upton, C. and Buckley, J. T. (1995) 'A new family of lipolytic enzymes?', *Trends in Biochemical Sciences*, 20(5), pp. 178–179. doi: 10.1016/S0968-0004(00)89002-7.

Vagelos, P. R. (1973) '5 Acyl Group Transfer (Acyl Carrier Protein)', *The Enzymes*. Academic Press, 8, pp. 155–199. doi: 10.1016/S1874-6047(08)60065-7.

Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. and Murshudov, G. N. (2004) '*REFMAC* 5 dictionary: organization of prior chemical knowledge and guidelines for its use', *Acta Crystallographica Section D Biological Crystallography*, 60(12), pp. 2184–2195. doi: 10.1107/S0907444904023510.

Vasina, J. A. and Baneyx, F. (1997) 'Expression of aggregation-prone recombinant proteins at low temperatures: A comparative study of the Escherichia coli cspA and tac promoter systems', *Protein Expression and Purification*, 9(2), pp. 211–218. doi: 10.1006/prep.1996.0678.

Vazquez, D. (1967) 'Macrolide antibiotics — Spiramycin, Carbomycin, Angolamycin, Methymycin and Lancamycin', in *Mechanism of Action*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 366–377. doi: 10.1007/978-3-642-46051-7_25.

Veiga, P., Bulbarela-Sampieri, C., Furlan, S., Maisons, A., Chapot-Chartier, M.-P., Erkelenz, M., Mervelet, P., Noirot, P., Frees, D., Kuipers, O. P., Kok, J., Gruss, A., Buist, G. and Kulakauskas, S. (2007) 'SpxB Regulates O-Acetylation-dependent Resistance of Lactococcus lactis Peptidoglycan to Hydrolysis', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 282(27), pp. 19342–19354. doi: 10.1074/jbc.M611308200.

Venkitakrishnan, R. P., Benard, O., Max, M., Markley, J. L. and Assadi-Porter, F. M. (2012) 'Use of NMR saturation transfer difference spectroscopy to study ligand binding to membrane proteins.', *Methods in molecular biology (Clifton, N.J.)*. NIH Public Access, 914, pp. 47–63. doi: 10.1007/978-1-62703-023-6_4.

Verdin, E. and Ott, M. (2013) 'Acetylphosphate: A Novel Link between Lysine Acetylation and Intermediary Metabolism in Bacteria', *Molecular Cell*, 51(2), pp. 132–134. doi: 10.1016/j.molcel.2013.07.006.

Verma, N. K., Brandt, J. M., Verma, D. J. and Lindberg, A. A. (1991) 'Molecular characterization of the O-acetyl transferase gene of converting bacteriophage SF6 that adds group antigen 6 to Shigella flexneri', *Molecular Microbiology*. Blackwell Publishing Ltd, 5(1), pp. 71–75. doi: 10.1111/j.1365-2958.1991.tb01827.x.

Viegas, A., Manso, J., Nobrega, F. L. and Cabrita, E. J. (2011) 'Saturation-transfer difference (STD) NMR: A simple and fast method for ligand screening and characterization of protein binding', *Journal of Chemical Education*. American Chemical Society and Division of Chemical Education, Inc., 88(7), pp. 990–994. doi: 10.1021/ed101169t.

Vollmer, W. (2008) 'Structural variation in the glycan strands of bacterial peptidoglycan', *FEMS Microbiology Reviews*. John Wiley & Sons, Ltd (10.1111), 32(2), pp. 287–306. doi: 10.1111/j.1574-6976.2007.00088.x.

Vollmer, W., Blanot, D. and De Pedro, M. A. (2008) 'Peptidoglycan structure and architecture', *FEMS Microbiology Reviews*. John Wiley & Sons, Ltd (10.1111), 32(2), pp. 149–167. doi: 10.1111/j.1574-6976.2007.00094.x.

Vora, M., Shah, M., Ostafi, S., Onken, B., Xue, J., Ni, J. Z., Gu, S. and Driscoll, M. (2013) 'Deletion of microRNA-80 Activates Dietary Restriction to Extend C. elegans Healthspan

and Lifespan', *PLoS Genetics*. Edited by S. K. Kim. Public Library of Science, 9(8), p. e1003737. doi: 10.1371/journal.pgen.1003737.

Wang, J., Knirel, Y. A., Lan, R., Senchenkova, S. N., Luo, X., Perepelov, A. V, Wang, Y., Shashkov, A. S., Xu, J. and Sun, Q. (2014) 'Identification of an O-acyltransferase gene (oacB) that mediates 3- and 4-O-acetylation of rhamnose III in Shigella flexneri O antigens.', *Journal of bacteriology*. American Society for Microbiology (ASM), 196(8), pp. 1525–31. doi: 10.1128/JB.01393-13.

Wang, L., Andrianopoulos, K., Liu, D., Popoff, M. Y. and Reeves, P. R. (2002) 'Extensive variation in the O-antigen gene cluster within one Salmonella enterica serogroup reveals an unexpected complex history.', *Journal of bacteriology*. American Society for Microbiology, 184(6), pp. 1669–77. doi: 10.1128/JB.184.6.1669-1677.2002.

Warren, M. J., Roddam, L. F., Power, P. M., Terry, T. D. and Jennings, M. P. (2004) 'Analysis of the role of pglI in pilin glycosylation of Neisseria meningitidis', *FEMS Immunology and Medical Microbiology*. Wiley/Blackwell (10.1111), 41(1), pp. 43–50. doi: 10.1016/j.femsim.2004.01.002.

Watanabe, M., Igarashi, T., Kaminuma, T., Sofuni, T., Nohmil, T. and Nohmi, T. (1994) 'N-hydroxyarylamine O-acetyltransferase of Salmonella typhimurium: proposal for a common catalytic mechanism of arylamine acetyltransferase enzymes.', *Environmental health perspectives*. National Institute of Environmental Health Science, 102 Suppl(Suppl 6), pp. 83–9. doi: 10.1289/ehp.94102s683.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. (2009) 'Jalview Version 2--a multiple sequence alignment editor and analysis workbench.', *Bioinformatics (Oxford, England)*. Oxford University Press, 25(9), pp. 1189–91. doi: 10.1093/bioinformatics/btp033.

Webster, N. S. (2014) 'Cooperation, communication, and co-evolution: grand challenges in microbial symbiosis research', *Frontiers in Microbiology*. Frontiers, 5, p. 164. doi: 10.3389/fmicb.2014.00164.

Wen, Z. and Zhang, J. R. (2014) 'Bacterial Capsules', in *Molecular Medical Microbiology: Second Edition*. Academic Press, pp. 33–53. doi: 10.1016/B978-0-12-397169-2.00003-2.

Whitney, J. C. and Howell, P. L. (2013) 'Synthase-dependent exopolysaccharide secretion

in Gram-negative bacteria.', *Trends in microbiology*. PMC Canada manuscript submission, 21(2), pp. 63–72. doi: 10.1016/j.tim.2012.10.001.

Williams, A. H., Veyrier, F. J., Bonis, M., Michaud, Y., Raynal, B., Taha, M.-K., White, S. W., Haouz, A. and Boneca, I. G. (2014) 'Visualization of a substrate-induced productive conformation of the catalytic triad of the Neisseria meningitidis peptidoglycan O -acetylesterase reveals mechanistic conservation in SGNH esterase family members', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 70(10), pp. 2631–2639. doi: 10.1107/S1399004714016770.

Wilson, J. W., Schurr, M. J., LeBlanc, C. L., Ramamurthy, R., Buchanan, K. L. and Nickerson, C. A. (2002) 'Mechanisms of bacterial pathogenicity.', *Postgraduate medical journal*. The Fellowship of Postgraduate Medicine, 78(918), pp. 216–24. doi: 10.1136/pmj.78.918.216.

Winn, M. D., Murshudov, G. N. and Papiz, M. Z. (2003) 'Macromolecular TLS Refinement in REFMAC at Moderate Resolutions', in, pp. 300–321. doi: 10.1016/S0076-6879(03)74014-2.

Winter, G. (2010) 'xia2: an expert system for macromolecular crystallography data reduction', *J. Appl. Cryst*, 43, pp. 186–190. doi: 10.1107/S0021889809045701.

Winter, S. E., Winter, M. G., Godinez, I., Yang, H.-J., Rüssmann, H., Andrews-Polymenis, H. L. and Bäumler, A. J. (2010) 'A Rapid Change in Virulence Gene Expression during the Transition from the Intestinal Lumen into Tissue Promotes Systemic Dissemination of Salmonella', *PLoS Pathogens*. Edited by J. Fierer. Public Library of Science, 6(8), p. e1001060. doi: 10.1371/journal.ppat.1001060.

Wollin, R., Stocker, B. A. D. and Lindberg, A. A. (1987) 'Lysogenic conversion of Salmonella typhimurium bacteriophages A3 and A4 consists of O-acetylation of rhamnose of the repeating unit of the O-antigenic polysaccharide chain', *Journal of Bacteriology*. American Society for Microbiology (ASM), 169(3), pp. 1003–1009. doi: 10.1128/jb.169.3.1003-1009.1987.

Wood, J. M. (2015) 'Bacterial responses to osmotic challenges.', *The Journal of general physiology*. The Rockefeller University Press, 145(5), pp. 381–8. doi: 10.1085/jgp.201411296.

Wu, D. and Hersh, L. B. (1995) 'Identification of an active site arginine in rat choline acetyltransferase by alanine scanning mutagenesis.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 270(49), pp. 29111–6. doi: 10.1074/JBC.270.49.29111.

Yadav, A. K., Espaillat, A. and Cava, F. (2018) 'Bacterial Strategies to Preserve Cell Wall Integrity Against Environmental Threats', *Frontiers in Microbiology*. Frontiers, 9, p. 2064. doi: 10.3389/fmicb.2018.02064.

Yan, B. X. and Sun Qing, Y. (1997) 'Glycine residues provide flexibility for enzyme active sites', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 272(6), pp. 3190–3194. doi: 10.1074/jbc.272.6.3190.

Yogev, R., Arditi, M., Chadwick, E. G., Amer, M. D. and Sroka, P. A. (1990) 'Haemophilus influenzae type b conjugate vaccine (meningicoccal protein conjugate): Immunogenicity and safety at various doses', *Pediatrics*, 85(4), pp. 690–693.

York, G. M. and Walker, G. C. (1998a) 'The Rhizobium meliloti ExoK and ExsH glycanases specifically depolymerize nascent succinoglycan chains.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 95(9), pp. 4912–7. doi: 10.1073/pnas.95.9.4912.

York, G. M. and Walker, G. C. (1998b) 'The succinyl and acetyl modifications of succinoglycan influence susceptibility of succinoglycan to cleavage by the Rhizobium meliloti glycanases ExoK and ExsH.', *Journal of bacteriology*. American Society for Microbiology (ASM), 180(16), pp. 4184–91.

Yu, T.-W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K. Q., Xu, J., Moss, S. J., Leistner, E. and Floss, H. G. (2002) 'The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from Actinosynnema pretiosum.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 99(12), pp. 7968–73. doi: 10.1073/pnas.092697199.

Zähringer, U., Knirel, Y. A., Lindner, B., Helbig, J. H., Sonesson, A., Marre, R. and Rietschel, E. T. (1995) 'The lipopolysaccharide of Legionella pneumophila serogroup 1 (strain Philadelphia 1): chemical structure and biological significance.', *Progress in clinical and biological research*, 392, pp. 113–39.

Zandomeneghi, G., Ilg, K., Aebi, M. and Meier, B. H. (2012) 'On-cell MAS NMR: Physiological clues from living cells', *Journal of the American Chemical Society*, 134(42), pp. 17513–17519. doi: 10.1021/ja307467p.

Zhang, G., Meredith, T. C. and Kahne, D. (2013) 'On the essentiality of lipopolysaccharide to Gram-negative bacteria.', *Current opinion in microbiology*. NIH Public Access, 16(6), pp. 779–85. doi: 10.1016/j.mib.2013.09.007.

Zhang, J. (2012) 'Protein-Protein Interactions in Salt Solutions', in *Protein-Protein Interactions - Computational and Experimental Tools*. doi: 10.5772/38056.

Zhang, X. and Wang, Y. (1996) 'Studies on midecamycin 4"-O-propionyltransferase gene structure', *Acta microbiologica Sinica*, 36(6), pp. 417–22.

Zhou, X., Ma, Z., Dong, D. and Wu, B. (2013) 'Arylamine N-acetyltransferases: A structural perspective', *British Journal of Pharmacology*. Wiley-Blackwell, pp. 748–760. doi: 10.1111/bph.12182.

Zhu, J., Sanborn, J. Z., Diekhans, M., Lowe, C. B., Pringle, T. H., Haussler, D., Tournamille, C., Colin, Y., Cartron, J., Kim, C. L. Van, Dean, M., Carrington, M., Winkler, C., Huttley, G., Smith, M., *et al.* (2007) 'Comparative Genomics Search for Losses of Long-Established Genes on the Human Lineage', *PLoS Computational Biology*. Public Library of Science, 3(12), p. e247. doi: 10.1371/journal.pcbi.0030247.

Zhu, P., Klutch, M. J. and Tsai, C.-M. (2001) 'Genetic analysis of conservation and variation of lipooligosaccharide expression in two L8-immunotype strains of *Neisseria meningitidis*', *FEMS Microbiology Letters*. John Wiley & Sons, Ltd (10.1111), 203(2), pp. 173–177. doi: 10.1111/j.1574-6968.2001.tb10837.x.

Zou, C. H., Knirel, Y. A., Helbig, J. H., Zähringer, U. and Mintz, C. S. (1999) 'Molecular cloning and characterization of a locus responsible for O acetylation of the O polysaccharide of Legionella pneumophila serogroup 1 lipopolysaccharide.', *Journal of bacteriology*. American Society for Microbiology (ASM), 181(13), pp. 4137–41.