Putative *O*-glycosylation Modification Sites of Tsr Are Involved in *E. coli* Motility

by

Bill Chris Vouronikos

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Physiology, Cell, and Developmental Biology

Department of Biological Sciences
University of Alberta

Abstract

Post-translational modifications (PTMs) regulate a variety of normal cellular processes, many of which are conserved across eukaryotes and prokaryotes. One PTM, protein glycosylation with the reversible modification of *O*-GlcNAc, is well studied in eukaryotes. However, evidence for its existence in prokaryotes remains controversial. Investigating how post-translational modifications work in bacteria will help us understand how bacteria adapt to and survive in changing environments.

Bacterial swarming/swimming motility is mediated by the flagella, but it is the methyl-accepting chemotaxis proteins (MCPs) that control chemotaxis by sensing changes in environmental nutrient and/or waste concentrations. One of the major MCPs in *E. coli* is Tsr, which has four putative *O*-GlcNAc modification sites that are directly adjacent to five known methylation sites which are critical for function. Our hypothesis was that the four putative modification sites in Tsr affect bacterial motility. We have shown that two of the four putative modification sites are necessary for motility, as site-directed mutants have significantly reduced motility. We also showed that addition of glucose causes changes in global proteins glycosylation as well as motility. Our work has provided a foundation to investigate the existence of *O*-GlcNAc in prokaryotes.

Preface

This thesis is an original work by Bill Chris Vouronikos. No part of this thesis had been previously published.

List of previous Presentations/Conferences

ImmunNet Research Day- August 21, 2020
Title: "*O*-GlcNAc on Target Proteins in *E. coli*."

RE Peter 2022- March 9, 2022
Title: "*O*-GlcNAc Modification & Its Effects on The Methylation Status of *E. coli* Protein Tsr."

GlycoNet Canadian Glycomics Symposium- May 12, 2022
Title: "*O*-GlcNAc Modification & Its Effects on The Methylation Status of *E. coli* Protein Tsr."

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my spectacular supervisor Dr. Lisa Willis for her support, assistance, encouragement, and most importantly her patience. Thank you for giving me the opportunity to work in your lab. I have learned so much under your supervision and I am extremely proud of the work we have accomplished together. Additionally, I would like to express my gratitude to my supervisory committee member Dr. Brian Lanoil for his time, insight, and advice during my master's program.

I would like to thank my amazing and brilliant lab members from the Willis Glycobiology lab: Dr. Carmanah Hunter, Tahlia Derksen, Muhammad Idrees, Sogand Makhsous, Amir Reza Hematyar Naghneh, Ronak Patel, David Yao, and Pedro Ribeiro Lopes Pinto de Almeida. I would like to thank all past and current lab members for their invaluable support, advice, assistance, encouragement and of course the countless coffees and laughs. I would also like extend my thanks to Dr. Warren Wakarchuk and the entire Wakarchuk Glycobiology lab for their advice and assistance. My thanks and appreciation to my TA lab coordinators Mark Wolansky and Christine Williams for their assistance and support during my graduate TA-ship.

Finally, a special thanks to Billy "The Piano Man" Joel for the motivation and inspiration you have given me these past few months.

Table of Contents

List of Tables

List of Figures

Abbreviations

GH- Glycoside Hydrolase

GT- GlycosylTransferase

IPTG: Isopropyl β-D-thiogalactoside

MET: Methylation modification sites

MCP: Methyl-accepting chemotaxis protein

MOE: Molecular oligosaccharide engineering

OG: Putative *O*-GlcNAc modification sites

OGA: *O*-GlcNAcase (hydrolase enzyme)

OGT: *O*-GlcNAc transferase

O-GlcNAc: *O*-linked β-*N*-acetylglucosamine

PTM: Post-translational modification

UDP-GlcNAc: Uridine diphosphate *N*-acetylglucosamine

WGA: Wheat germ agglutinin

Chapter 1: Introduction

1.1 O-GlcNAcylation in eukaryotes

O-linked β-*N*-acetylglucosamine (*O*-GlcNAc) is an intracellular post-translational modification that is found predominantly in eukaryotes (Bond & Hanover 2015). *O*-GlcNAc modifications are uncharged acetylated hexosamine sugars that are attached through a β-glycosidic bond to hydroxyl-containing amino acids (Bond & Hanover 2015). The cycling of *O*-GlcNAc in eukaryotes is a reversible and dynamic process that is controlled by two distinct intracellular enzymes (Fig. 1) (Bond & Hanover 2015).

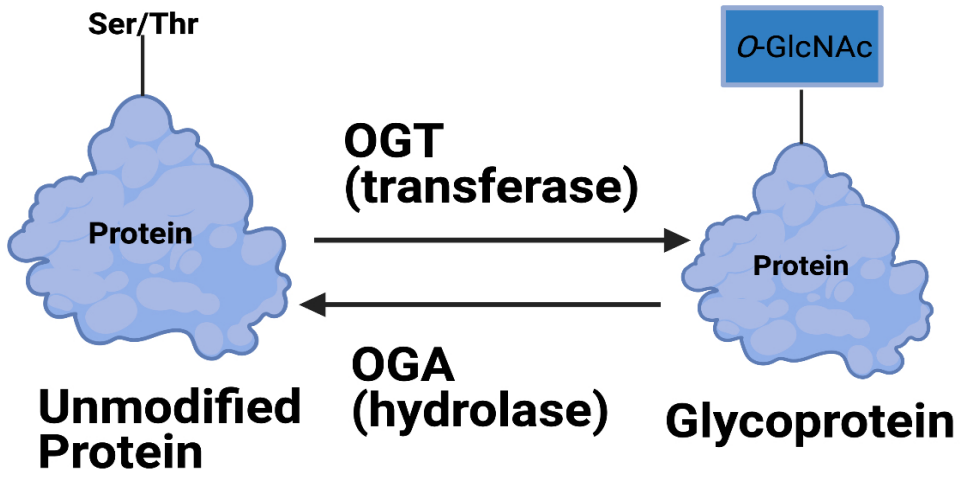

Figure 1: *O*-GlcNAc cycling system seen in eukaryotes. Eukaryotes only contain two enzymes that control *O*-GlcNAc modification cycling. *O*-GlcNAc transferase (OGT) enzyme is used to modify proteins with *O*-GlcNAc moiety onto either serine and/or threonine residues of the target protein. *O*-GlcNAcase hydrolase removes *O*-GlcNAc modifications of the target protein. Created with BioRender.com

O-GlcNAc transferase (OGT) and the hydrolase *O*-GlcNAcase (OGA) catalyze the addition and removal of *O*-GlcNAc respectively (Bond & Hanover 2015). *O*-GlcNAc transferase uses the nucleotide sugar substrate uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) which is derived from the hexosamine biosynthetic pathway to catalyze the modification of *O*-GlcNAc onto target proteins (Bond & Hanover 2015). *O*-GlcNAc modifications are different from other

protein glycosylation in that it is reversibly modified to serine and/or threonine residues predominantly on nuclear, cytoplasmic, and mitochondrial proteins (Bond & Hanover 2015).

In eukaryotes, *O*-GlcNAc modifications largely regulate important cellular processes (Harwood & Hanover 2014). For example, *O*-GlcNAc cycling is involved in the regulation of RNA polymerase II regulatory dynamics during transcription (Resto et al. 2016). *O*-GlcNAc also acts as an environmental sensor where *O*-GlcNAc is used as a signaling mechanism to respond to stressors (Hart 2019). In response to external stressors, *O*-GlcNAc levels are increased on a variety of nucleocytoplasmic proteins, which can make cells more resistant to high temperatures, DNA damage, and other harmful toxins (Zachara et al. 2004). *O*-GlcNAc functions similarly to phosphorylation in that it can regulate protein activity and half-lives of intracellular proteins (Harwood & Hanover 2014, Bond & Hanover 2015). For example, increased *O*-GlcNAcylation at serine residues decreases the level of phosphorylation at adjacent threonine residues of the tumor suppressor protein p53, which controls apoptosis (Yang et al. 2006). The *O*-GlcNAcylation of p53 prevents ubiquitin-mediated proteolysis and allows it to regulate cell growth by inducing apoptosis when necessary (Yang et al. 2006). Based on the results of this study and many others, there is a dynamic interplay between *O*-GlcNAc and phosphorylation modifications. Therefore, the attachment of *O*-GlcNAc modifications can directly affect target protein function by providing a binding site for interacting proteins or indirectly affect target protein function through altering nearby post-translational modifications (Yang et al. 2006).

1.2 *O*-GlcNAcylation in bacteria

1.2.1 Lack of evidence for O-GlcNAc moiety

The existence of *O*-GlcNAc modifications in bacteria remains a contentious topic. If bacteria possess an *O*-GlcNAc cycling system like eukaryotes, they should encode both OGT and OGA enzymes in their genome and be able to glycosylate and de-glycosylate cytosolic proteins in a cyclic manner dependent on their environments. For decades, *O*-GlcNAc modifications were thought to only occur in eukaryotes (Balonova et al. 2009, Schirm et al. 2005, Upreti et al. 2003). This was the generally accepted belief due to the undetectable levels of *O*-GlcNAcylated proteins in bacterial cell extractions when measured using the anti-*O*-GlcNAc antibody (Balonova et al. 2009, Schirm et al. 2005, Upreti et al. 2003).

Antibodies against *O*-GlcNAc moieties have been used as a tool to detect *O*-GlcNAcylated proteins (Ma & Hart, 2014). However, low immunogenicity of the neutral *O*-GlcNAc sugar has made it difficult to produce highly specific antibodies against *O*-GlcNAcylated proteins (Thompson et al. 2018). Antibodies frequently detect specific protein sequences or structural motifs and therefore can only detect a subset of *O*-GlcNAc proteins (Ma & Hart 2014). *O*-GlcNAc antibodies can also lead to conflicting results as they can bind other possible sugars, like *N*-linked glycans, in addition to *O*-GlcNAc moieties (Isono 2011). In Alzheimer's disease brain tissues, different studies using different *O*-GlcNAc antibodies showed increases and/or decreases in total *O*-GlcNAcylation, indicating a discrepancy when using antibodies against *O*-GlcNAcylated proteins (Förster et al. 2014, Liu et al. 2009). The *O*-GlcNAc antibodies used in bacterial studies are the same ones used in human studies (Ma & Hart 2014).

The use of human anti-*O*-GlcNAc antibodies in immunoblotting with different bacteria show varying degrees of signal. For example, the use of anti-*O*-GlcNAc antibodies in *E. coli* lysates showed no detectable levels of *O*- GlcNAcyated proteins (Gao et al. 2018), while other studies have found faint signals (Riu et al 2008). It should be noted, however, that most *E. coli* lysates examined are cloning strains, which may have lost the ability to *O*-GlcNAcylate proteins due to decades of culturing under “ideal” condtions. Other studies using thermophilic bacterium have shown detection of *O*-GlcNAcylated proteins with *O*-GlcNAc-specific antibodies (Ostrowski et al 2015). In a study of recombinant human OGT expression in *E. coli*, lysates immunoblotted with an anti-*O*-GlcNAc antibody show strong signal corresponding to the auto-*O*-GlcNAcylation of OGT (Riu et al 2008). However, co-expression with the human OGT adaptor protein Sp1 led to O-GlcNAcylation of multiple *E. coli* proteins, as detected by immunoblotting (Riu et al 2008). It is clear from this study that the epitopes recognized by the OGT and anti-O-GlcNAc antibodies exist in bacteria. However, whether bacteria *O*-GlcNAcylate slightly different epitopes that partially cross-react with the anti-*O*-GlcNAc antibody (corresponding to the faint signals observed in the *E. coli* lysates) or just do not make *O*-GlcNAc is unclear. It is also possible that the faint signals observed in the immunoblot of *E. coli* lysate correspond to non-specific binding of the antibody to non-*O*-GlcNAc epitopes.

Newer techniques like metabolic oligosaccharide engineering (MOE), in which sugar analogs are taken up by cells, metabolized, and incorporated into *O*-GlcNAcylated proteins via the cell's machinery, have been used to investigate *O*-GlcNAc in eukaryotes (Zaro et al. 2011).

Indeed, hundreds of glycoproteins have been identified using this method. However, due to defense barriers, such as cell walls in bacteria, as well as the size and hydrophobicity of the sugar analogs, these sugar analogs do not enter bacterial cells as easily as they do eukaryotic cells (Guianvarc'h et al. 2021). Additionally, this method potentially suffers from sensitivity issues from low incorporation of sugar analogs due to competition with natural sugars (Thompson et al. 2018).

1.2.2 Presence of O-GlcNAc cycling machinery in bacteria remains unclear

The existence of OGT and OGA cycling components in bacteria has also been controversial. Proteins that remove *O*-GlcNAc from bacterial target proteins have been identified in some bacterial species. Based on sequence similarities with eukaryotic OGA in glycoside hydrolase family GH84, bacteria have putative *O*-GlcNAc hydrolases (Henrissat 1991). BtGH84 of *Bacteroides thetaiotaomicron* has been characterized as part of the GH84 family showing 40% sequence similarity to the catalytic domain of eukaryotic OGA (Dennis et al. 2006). The enzyme can remove *O*-GlcNAc from eukaryotic *O*-GlcNAcylated proteins and is inhibited by known OGA inhibitor PUGNAc (Dennis et al. 2006). Additionally, NagJ from *Clostridium perfringens* has also been characterized in the GH84 family with 51% sequence similarity with eukaryotic OGA (Rao et al. 2006). The enzyme also possesses OGA activity against eukaryotic *O*-GlcNAcylated proteins and is inhibited by PUGNAc (Rao et al. 2006). Unfortunately, the *O*-GlcNAc transferases required for an *O*-GlcNAc cycling system are not present in either of these bacteria so these findings provide only partial evidence that bacteria may have the capabilities for an *O*-GlcNAc cycling system as seen in eukaryotes.

Eukaryotic OGT is part of the glycosyltransferase family GT41 (Ma et al. 2022). There are bacterial members in GT41, however there are also potential transferases that make protein-linked *O*-GlcNAc in other families (Ma et al. 2022). Bacterial OGTs have been classified in five different families to date: GT2, GT4, GT8, GT41, and GTNC [non-classified] (Ma et al. 2022). GmaR in *Listeria monocytogenes* has been characterized in the GT2 family, exhibiting OGT activity by *O*-GlcNAcylating the flagellin protein FlaA (Shen et al. 2006). Once flagellar proteins are exported, they can no longer be substrates for OGA and OGT and so are not part of a cycling system at this point. However, glycosylation is essential for flagella formation and assembly, and could function as an environmental sensor under conditions where the flagella is

required for motility (Fredriksen et al. 2012, Schirm et al. 2004, Schirm et al. 2005). A bacterial homologue of OGT from *Xanthomonas campestris* was classified in the GT41 family with 36% sequence similarity with eukaryotic OGT and, while there was no OGT activity on bacterial and human proteins, it was able to *O*-GlcNAcylate a single plant protein, showing its potential for OGT activity (Clarke et al. 2008). Several bacteria do not have identified OGT orthologues, including *E. coli* (Ma et al. 2022).

If an *O*-GlcNAc cycling system is present in bacteria as is seen in its eukaryotic counterpart, we would expect to find both OGT and OGA enzymes and they would predominantly modify cytoplasmic proteins. *Thermobaculum terrenum* is the only bacteria so far that has been shown to possess both OGT and OGA enzymes (Ostrowski et al. 2015). The sequences of the *T. terrenum* OGT N and C-terminal catalytic domains were 25% and 29% respectively identical to human OGT and the *T. terrenum* OGA had 36% sequence similarity as compared to human OGA (Ostrowski et al. 2015). Both *T. terrenum* OGT and OGA were found to be intracellular proteins which partially supports the possibility that an intracellular *O*-GlcNAc cycling system could exist in bacteria (Ostrowski et al. 2015). *T. terrenum* OGA could remove *O*-GlcNAc from eukaryotic proteins however, there was no detectable activity of *T. terrenum* OGT against any proteins (Ostrowski et al. 2015). It is plausible as shown in previous studies that bacterial OGT does not fall into the category of its eukaryotic counterparts, and that there may be other possible enzymes that transfer *O*-GlcNAc to serine and/or threonine residues but target different proteins. To date, no OGT and OGA activity have been found in one bacterial species, however, all the searching has relied on comparison to eukaryotic enzymes. As of now, there may be enzymes unidentified that perform these enzymatic activities.

1.2.3 Bacteria can make the donor sugar UDP-GlcNAc

Bacteria have the necessary capabilities to create the donor sugar UDP-GlcNAc and potentially produce *O*-GlcNAcylated proteins (Riu et al 2008, Harwood & Hanover 2014, Bond & Hanover 2015.) As in eukaryotes, bacterial UDP-GlcNAc is made from fructose-6-phosphatase, however, the enzymes involved in bacterial UDP-GlcNAc synthesis, GlmM, GlmS, and GlmU, are different from what is found in eukaryotes (Barreateau et al. 2008). In a study using reverse-phase high-pressure chromatography for separation of UDP-GlcNAc from *E. coli* intracellular extracts, cytosolic concentrations of UDP-GlcNAc were determined to be

approximately 100 μM (Mengin-Lecreulx et al. 1983). According to another study, overexpressing human OGT with GlmM and GlmU in the UDP-GlcNAc pathway of *E. coli* can increase endogenous *O*-GlcNAcylation of recombinant eukaryotic proteins, as detected using an anti-*O*-GlcNAc antibody. (Gao et al. 2018). Bacteria clearly the capabilities to make the donor sugar required for *O*-GlcNAcylation.

1.2.4 Evidence for O-GlcNAcylated proteins in bacteria

Numerous putative cytosolic *O*-GlcNAcylated proteins have been identified in bacteria through three seminal proteomic experiments. In the first study, *Helicobacter pylori* were fed the unnatural azide-containing sugar Ac_4GlcNAz, an analog of the metabolic precursor GlcNAc, to label glycoproteins with GlcNAz (Champasa et al. 2013). To identify the glycans, glycoproteins were analyzed by mass spectrometry and then subjected to beta-elimination to demonstrate that the metabolic label was incorporated into *O*-linked glycans (Champasa et al. 2013). Unlike *E. coli*, *H. pylori* were able to take up the metabolic label and incorporate it into glycoproteins. To identify the glycoproteins, Ac_4GlcNAz-treated cells were lysed and analyzed by mass spectrometry. Numerous peptides containing HexNAz residues (likely GlcNAz) were observed. They subsequently subjected the lysates to beta-elimination, which removes *O*-linked glycans, to demonstrate that the metabolic label was incorporated into *O*-linkages, consistent with *O*-GlcNAc (Champasa et al. 2013). This study provides strong evidence that *H. pylori* glycoproteins are modified with *O*-GlcNAc. Metabolic labelling can be an effective tool in characterizing bacterial glycans, however labelling depends on sugar uptake and integration into bacterial glycoproteins, and some sugar analogs may not be incorporated into the proteins of interest since they compete with natural sugars (Thompson et al. 2018).

Subsequently in the second study, a new technique termed "BEMAP" (β-elimination of *O*-linked carbohydrate modifications, Michael addition of 2-Aminoethyl phosphonic acid) was developed to map *O*-linked glycoproteins of *E. coli* (Boysen et al. 2016). BEMAP works by substituting *O*-linked moieties with a phosphopeptide group that is enriched by titanium dioxide affinity (Boysen et al. 2016). With this technique, hundreds of glycosylated serine and/or threonine residues were characterized that corresponded to over 150 glycoproteins in *E. coli*. Intriguingly, more proteins were identified in a pathogenic strain of *E. coli* than a laboratory strain (Boysen et al. 2016). While BEMAP identifies a variety of glycoproteins, it does not

provide any evidence regarding the actual sugar moiety. However, the substantial overlap between the proteins found in *E. coli* using the BEMAP technique and *H. pylori* using the metabolic labeling strategy helps to further support our hypothesis that bacteria may have numerous cytosolic *O*-linked glycoproteins.

Lastly, the use of lectin enrichment with wheat germ agglutinin (WGA) has also been used for the characterization of *O*-linked glycoproteins in *Lactobacillus plantarum*. WGA binds to glucose, GlcNAc and sialic acid (Peters et al., 1979, Monsigny et al. 1980). Sialic acid is not found in *L. plantarum* and would not be relevant in this study. A proteomic analysis of the glycoproteome of *L. plantarum* identified 11 glycoproteins with HexNAc moieties (and others with Hex moieties), including four proteins with well-known cytosolic functions (Fredriksen et al. 2013). While lectin enrichment can be used to detect global bacterial glycoprotein profiles, it does not disclose the details of the glycan structure or where they attach to proteins (Dube et al. 2010). The binding of lectin to free *O*-GlcNAc has been shown to be extremely weak (Leickt et al 1997). It is possible that the low affinity of WGA for sugars is why fewer proteins were detected and there may be more present in this bacterium.

The combined information from the three studies described above demonstrate strong points to the existence of dozens of cytosolic *O*-GlcNAcylated proteins in Gram-positive and Gram-negative bacteria (Boysen et al. 2016, Champasa et al. 2013, Fredriksen et al. 2013). These articles utilized complementary methodologies to detect and characterize dozens of unidentified bacterial glycoproteins. From the glycoproteins characterized between all the studies, at least six were found to be common among *E. coli*, *H. pylori,* and *L. plantarum.* These same six glycoproteins identified in bacteria were also known to be *O*-GlcNAcylated in eukaryotes (Table 1) (Wulff-Fuentes et al. 2021). Therefore, *O*-GlcNAcylation may be present in bacteria and could potentially be a general post-translational modification that regulates cellular processes.

Table 1: Cytosolic Protein Glycosylation in Bacteria & Eukaryotes

Glycoproteins	*H. pylori*	*E. coli*	*L. plantarum*	Eukaryotes
Heat Shock Protein 60	✓	✓	✓	✓
Heat Shock Protein 70	✓	✓	✓	✓

Enolase	✓	✓	✓	✓
ATP synthase	✓	✓	✓	✓
Serine hydroxymethyltransferase	✓	✓	✓	✓
RNA Polymerase	✓	✓	✓	✓
Methyl-accepting chemotaxis protein (Tsr)	✓	✓	N/A	N/A

1.3 Bacterial motility as a phenotype for glycosylation

One of the glycoproteins identified in both motile bacteria, *H. pylori* and *E. coli,* was the methyl-accepting chemotaxis protein Tsr (Table 1). Tsr is a transmembrane dimeric chemoreceptor that senses changes in environmental nutrients and/or waste concentrations to control chemotaxis (movement in response to chemical stimuli) causing the cell to move towards stimuli or away from repellents (Kim et al. 1999, Parkinson et al. 2015). Tsr ligands (e.g., serine, and related amino acids like cysteine, alanine) bind directly to the ligand-binding domain of the Tsr receptor in the periplasmic domain and transmits signal to the intracellular signalling domain via a cytoplasmic adapter domain (Fig. 2) (Kim et al. 1999). The structure of the intracellular signalling domain of Tsr is a coiled-coil of two antiparallel α-helices connected by a U-turn (Kim et al. 1999). The cytoplasmic domain of the receptor dimer forms a four helical bundle, and this same region is involved in interactions with receptor dimers that result in the formation of a trimer of dimers in response to ligand binding (Kim et al. 1999). Cross-linking of receptor trimers forms receptor clustering or lattices, which activates Tsr intracellular signaling components called Che proteins that act as key regulators in the signal transduction pathway in the cytoplasm to control flagellar rotation and sensory adaptation (Kim et al. 1999).

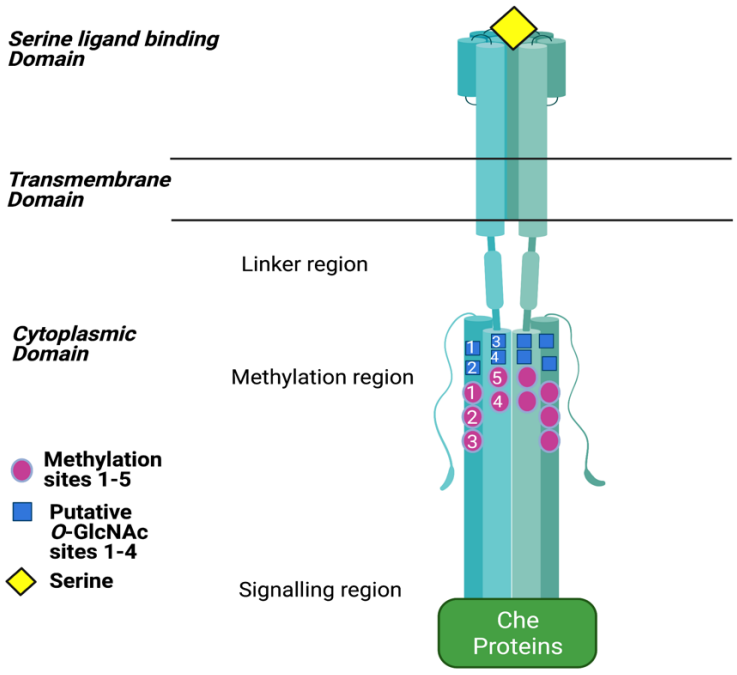

Figure 2: The methyl-accepting chemotaxis protein receptor Tsr and the corresponding putative *O*-GlcNAc and methylation sites. We predict that the four putative modification sites in Tsr will affect bacterial motility. By increasing *O*-GlcNAc modifications at putative *O*-GlcNAc sites, steric hindrance created by the bulky *O*-GlcNAc structures will decrease the methylation site occupancy. With fewer methylation modifications, this will decrease the motor response of the Tsr flagella and decrease *E. coli* motility capabilities. Created with BioRender.com.

1.3.1 Key signalling cascade of Tsr

In response to ligand binding, Tsr forms a stable signalling complex with two cytoplasmic proteins: CheA is a histidine autokinase and CheW couples CheA to receptor control (Parkinson 2003, Parkinson et al. 2015). Together the CheA/CheW complex can generate receptor signals that activate the two main pathways of the chemotaxis signalling cascade (Fig. 3) (Parkinson 2003, Parkinson et al. 2015). The first part of the pathway controls the flagellar motor, which influences the cell's movement. Counterclockwise (CCW) runs move cells into favourable directions and clockwise (CW) tumbles cause random directional changes (Parkinson 2003, Parkinson et al. 2015). The second part of the pathway controls sensory adaptation, mediated by five key methylation sites in the Tsr adapter domain (Challah et al. 2005, Parkinson 2003, Parkinson et al. 2015).

When bacteria encounter increases in attractants or decreases in repellents as sensed by the MCPs, they suppress CW tumbles, increasing the length of runs that lead them into favorable directions. *E. coli* move towards attractants and away from harmful compounds (Parkinson 2003, Parkinson et al. 2015). Changes in ligand binding to the periplasmic ligand-binding domain of the MCP alters MCP dimerization and activation of the chemoreceptor (Fig. 3) (Kim et al. 1999). MCP dimerization causes the CheA/CheW complex to forme and the signalling cascade utilizes protein phosphorylation and dephosphorylation reactions to control the cell's flagellar motors in response to environmental changes (Parkinson 2003).

When the concentration of attractant binding to the periplasmic ligand-binding domain of the MCP decreases, the receptor shifts towards the kinase-on state (Parkinson 2003, Parkinson et al. 2015). The CheA/CheW complex is formed and CheA auto-phosphorylates itself on a histidine residue, which then activates CheY by phosphorylation (Parkinson 2003, Parkinson et al. 2015). Phosphorylated CheY binds to flagellar motor switch proteins in the flagellar motors to signal the basal body of the flagellum to rotate in a CW and causes the cell to tumble (Parkinson 2003, Parkinson et al. 2015). However, the activated phosphorylated CheY signal is short-lived by phosphatase CheZ, which enhances dephosphorylation of CheY (Parkinson 2003, Parkinson et al. 2015). Likewise, when attractants like serine bind this shifts the receptor towards the kinase-off state, CheA is deactivated, which results in low levels of activated CheY and the cell remains in the default CCW signalling state allowing the cell to continue to run (Parkinson 2003, Parkinson et al. 2015).

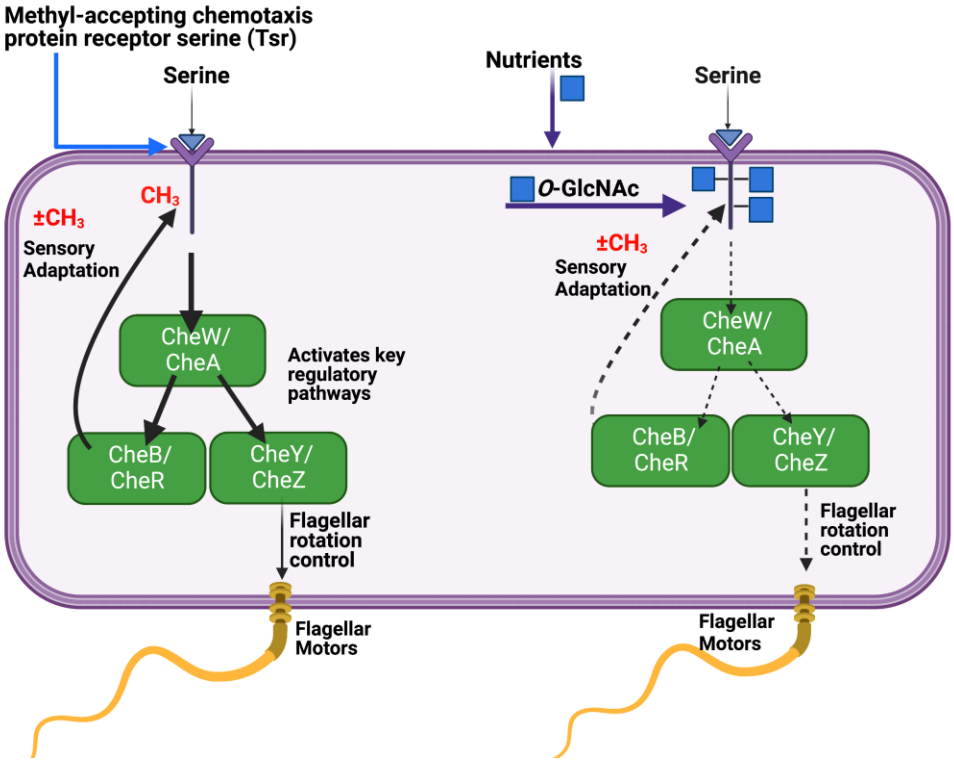

Figure 3: *E. coli* chemotaxis signaling transduction pathway of methyl-accepting chemotaxis protein (MCP) Tsr. Tsr records activation of its chemoreceptor in the form of reversible methylation at specific glutamate residues in the cytoplasmic signalling domain. Serine, a chemo-attractant, binds to Tsr, which initiates a signaling pathway that causes the chemotaxis pathway to activate intracellular signaling components called Che proteins. Activated Che proteins cause methylation of Tsr's intracellular domain, resulting in activation of the receptor and stimulation of the flagellar motors, resulting in flagellar rotation and *E. coli* motility. Sensory adaptation of the Tsr receptor changes the methylation status to adapt to current environmental conditions. We predict that increasing the putative *O*-GlcNAc modifications will change methylation sites from being modified with methyl groups and therefore decrease *E. coli* motility. Created with BioRender.com

During the signalling cascade, the cell undergoes sensory adaptation as part of the pathway, which is controlled by the CheB and CheR proteins (Parkinson 2003, Parkinson et al. 2015). Tsr receptor needs to readjust its ligand sensitivity to allow the cell to operate over a wide concentration gradient (Parkinson 2003, Parkinson et al. 2015). It accomplishes this

through reversible methylation of five glutamate residues (E297, E304, E311, E493 and E502) in the Tsr adapter domain (Challah et al. 2005). Methylation of the adaptor domain increases the ability of Tsr to activate CheA, therefore more methylation on the receptor results in more tumble movements, and less methylation results in longer runs (Parkinson 2003, Parkinson et al. 2015). CheB and CheR are the methyltransferase and methylesterase respectively (Parkinson 2003, Parkinson et al. 2015).

The signalling cascade of Tsr functions as follows. CheR is constantly methylating under all conditions (Parkinson 2003, Parkinson et al. 2015). In response to a decrease in attractant binding to Tsr, CheA phosphorylates the methylesterase CheB, in addition to CheY. While CheY causes tumbles on a short timescale, CheB activation on a longer timescale causes a decrease in methylation, which decreases the ability of Tsr to activate CheA, leading to longer runs. Sensory adaptation allows changes in receptor methylation status to shift signals to restore balances between run and tumble signal outputs depending on the conditions in the environment (Parkinson 2003, Parkinson et al. 2015).

The putative *O*-GlcNAc modification sites in Tsr occur at residues S279, S282, T484 and S494 (Boysen et al. 2016). The known methylation sites 1-3 (E297, E304, and E311) are located on one side of the receptor and sites 4-5 (E493 and E502) are located on the other side of the receptor (Chalah et al. 2005). Consequently, the putative *O*-GlcNAc modification sites are directly adjacent to the five known methylation sites, which are critical for Tsr function (Fig 2.) (Chalah et al. 2005). As previously mentioned, the modification of *O*-GlcNAc on p53 at serine residues decreased phosphorylation of p53 at adjacent threonine residues (Yang et al. 2006). It is possible that the potential *O*-GlcNAc modifications at the putative *O*-GlcNAc sites could possibly change the methylation status of Tsr and therefore can affect the motor response.

1.4 Hypothesis

We hypothesize that the four putative *O*-GlcNAc modification sites in Tsr play a role in bacterial motility. The four putative residues will be mutated, and we will determine if these mutations have any effect on the motility of *E. coli*. Because motility is a phenotype that can be easily observed and measured accordingly, we chose to use putative Tsr *O*-GlcNAcylation as a model system for investigating *O*-GlcNAcylation in bacteria. Additionally, we will develop

methodology to make the Tsr protein soluble and purify it so that future studies can determine the number of *O*-GlcNAc post-translational modifications via mass spectrometry analysis.

We will also evaluate the role of glucose in affecting glycosylation and motility. In eukaryotes, increased cellular glucose causes flux through the hexosamine biosynthetic pathway and causes increased production of UDP-GlcNAc (Harwood & Hanover 2014, Bond & Hanover 2015). *O*-GlcNAc is regulated by the hexosamine biosynthetic pathway, where UDP-GlcNAc is created and is the donor sugar of *O*-GlcNAc (Harwood & Hanover 2014, Bond & Hanover 2015). When there is an increased flux of glucose, there is an increased production of UDP-GlcNAc and therefore increased *O*-GlcNAc levels (Harwood & Hanover 2014, Bond & Hanover 2015). *E. coli* are known to flux glucose into the hexosamine biosynthetic pathway so we will be adding glucose to see if protein glycosylation and motility will be impacted (Mengin-Lecreulx et al. 1983, Barreateau et al. 2008).

Investigating how *O*-GlcNAc post-translational modifications work in bacteria will help us understand how bacteria adapt to and survive in changing environments. This work will lay down the foundation needed to study the regulation of *O*-GlcNAcylation in bacteria and provide much needed insight into how *O*-GlcNAc modifications contribute to bacterial homeostasis in their given environments.

Chapter 2: Materials & Methods

2.1 Bacterial Strains & Plasmids: Strains and plasmids used are shown in Table 2. All plasmids have an AP-01 plasmid backbone (kind gift of Dr. Warren Wakarchuk), which is a low copy number plasmid that contains chloramphenicol antibiotic resistance marker, a lac operator, and a tac promoter. A Tsr plasmid was commercially purchased and encodes the full length Tsr gene which includes the putative *O*-GlcNAc sites (S279, S282, T484, S494) and methylation sites (E297, E304, E311, E493, E502).

Table 2: Bacterial Strains & Plasmids

Strains (*E. coli* K12)	Relevant Phenotype
Wild type (MG1655 genotype: F^-, $lambda^-$, *rph-1*)	Wild type for motility
Δ*tsr*	Knockout *tsr* gene (Baba et al. 2006)
AP-01	*E. coli* BL21(DE3) transformed with AP-01 plasmid: lac operator, tac promoter, low copy number, chloramphenicol resistance marker
TSR01	Constructed *tsr* gene in AP-01 plasmid

Δ*tsr*+*tsr*	Complemented knockout Δ*tsr* with TSR01
TSR26	Full length *tsr* with N terminus 6x His-tag
OG1	Δ*tsr* containing *tsr* gene with site-directed alanine mutations at *O*-GlcNAc site (S279A)
OG2	Δ*tsr* containing *tsr* gene with site-directed alanine mutations at *O*-GlcNAc site (S282A)
OG3	Δ*tsr* containing *tsr* gene with site-directed alanine mutations at *O*-GlcNAc site (T484A)
OG4	Δ*tsr* containing *tsr* gene with site-directed alanine mutations at *O*-GlcNAc site (S494A)
MET1	Δ*tsr* containing *tsr* gene with site-directed alanine mutations at methylation site (E297A)
MET2	Δ*tsr* containing *tsr* gene with site-directed alanine mutations at methylation site (E304A)
MET3	Δ*tsr* containing *tsr* gene with site-directed alanine mutations at methylation site (E311A)
MET4	Δ*tsr* containing *tsr* gene with site-directed alanine mutations at methylation site (E493A)
MET5	Δ*tsr* containing *tsr* gene with site-directed alanine mutations at methylation site (E502A)

2.2 Subcloning: Tsr from the commercially purchased plasmid was subcloned into an AP-01 plasmid backbone (kind gift of Dr. Warren Wakarchuk) to create TSR01. Restriction enzymes SalI (#R3138S, NEB) and NdeI (#R0111S, NEB) were used to excise our Tsr insert of interest. Tsr insert was purified using a gel band extraction kit (#K0691, Thermo Scientific). The AP-01 plasmid vector was digested with restriction enzymes SalI (#R3138S, NEB) and NdeI (#R0111S, NEB) and purified using gel band extraction. The Tsr insert and destination AP-01 vector were mixed and ligated together using NEB T4 DNA ligation (#M0202S, NEB) to create TSR01. Amplification of the plasmid product was performed by transformation. NEB BL21(DE3) (#C2527H, NEB) chemically competent *E. coli* cells were used for the high efficiency transformation procedure. Calcium competent *E. coli* cells and heat-shock transformation methods were used to make Δ*tsr* competent cells for complementation of Tsr plasmid variations (Chang et al. 2017). After cloning was completed, colony PCR (#M0273L, NEB) was used to screen for colonies following transformation to confirm the desired plasmid construct by amplifying the insert. Constructs were confirmed by standard Sanger DNA sequencing (Molecular Biology Facility (MBSU), University of Alberta). All primers used for all experiments are listed below in Table 3.

Table 3: List of Primers

Primer Name	Primer Sequence	Function
lw504	CCGGTGCCAGCGAAATCGCCACCGGCAATAAC	Forward primer for making S279A in Tsr
lw505	CATAGATGGCATTGGCCCCGTTGCGCA	Reverse primer for making S279A in Tsr
lw506	CCGAAATCGCCACCGGCAATAACGAT	Forward primer for making S282A in Tsr
lw507	CGGCACCGCTATAGATGGCATTGGCCC	Reverse primer for making S282A in Tsr
lw508	CTCAACAGAACGCCGCGCTGGTGGAAGA	Forward primer for making T484A in Tsr
lw509	CTACCCGGTCCATCTCAGCAACCGCTAAGCCAAC	Reverse primer for making T484A in Tsr
lw510	TGCCGCTGCCGCCGCCGCGCTGGAAGAGCA	Forward primer for making S494A in Tsr
lw511	GCCTCTTCCACCAGCGCGGCGTTCTGTTGAG	Reverse primer for making S494A in Tsr
lw513	GGTCGTGTCGACCTATTAAAATGTTTCCCAGTTCTCC	Reverse primer for cloning Tsr using SalI
lw570	GCGCACCGAGGCACAGGCCGCTT	Forward primer for making E297A in Tsr
lw571	GAAGAGAGATCGTTATTGCCGGTG	Reverse primer for making E297A in Tsr
lw572	TTCGCTGGAAGCGACGGCAGCCA	Forward primer for making E304A in Tsr
lw573	GCGGCCTGTTGCTCGGTG	Reverse primer for making E304A in Tsr
lw574	CAGCATGGAGGCACTGACCGCAAC	Forward primer for making E311A in Tsr

lw575	GCTGCCGTCTCTTCCAGC	Reverse primer for making E311A in Tsr
lw576	GCTGGTGGAAGCGTCTGCCGC	Forward primer for making E493A in Tsr
lw577	GCGGCGTTCTGTTGAGTTAC	Reverse primer for making E493A in Tsr
lw578	CGCCGCGCTGGCAGAGCAGGCCA	Forward primer for making E502A in Tsr
lw579	GCGGCAGCGGCAGACTCTTCC	Reverse primer for making E502A in Tsr
lw580	GCACCATATGCACCATCACCATCACCATTTAAAACG TATCAAAATTGTG	Forward primer to clone Tsr with N-terminal His-tag

2.3 PCR cloning: TSR26 plasmid was created using PCR (#E0530S, NEB) to add on the 6x polyhistidine-tag (His-tag). TSR26 plasmid contains a full length Tsr gene with an N-terminus His-tag. Primers lw580 and lw513 were created and purchased to add on the His-tag onto corresponding plasmid construct. Purification using PCR cleanup protocols were performed on the PCR products. Afterwards, restriction digestion using SalI (#R3138S, NEB) and NdeI (#R0111S, NEB) was performed. PCR cleanup (#NA1020, Sigma-Aldrich) was again performed on restriction digestion mixture. The insert of interest and AP-01 vector were mixed and ligated together using T4 DNA ligation (#M0202S, NEB). Amplification of the plasmid product was performed by transformation. NEB BL21(DE3) (#C2527H, NEB) chemically competent *E. coli* cells were used for the high efficiency transformation procedure. Calcium competent *E. coli* cells and heat-shock transformation method were again used to make Δ*tsr* competent cells for complementation of Tsr plasmid variations (Chang et al. 2017). After cloning was completed, colony PCR (#M0273L, NEB) was used to screen for colonies following transformation to confirm the desired plasmid construct by amplifying the insert. Constructs were confirmed by standard Sanger DNA sequencing (MBSU).

2.4 Mutagenesis: *O*-GlcNAc mutants (OG1-4) and methylation mutants (MET1-5) were created using Q5 Site-Directed Mutagenesis (#E0554S, NEB). Q5 mutagenesis was used to create site-directed alanine mutations of our forementioned Tsr plasmid DNA. Mutant variants

were created by using mutagenic primers containing single base pair changes at specific OG and MET sites followed by intramolecular ligation and transformation into NEB high-efficiency chemically competent cells (#E0554S, NEB). After cloning was completed, colony PCR (#M0273L, NEB) was used to screen for colonies following transformation to confirm the desired plasmid construct by amplifying the insert. All mutations and other cloned regions amplified by PCR (#E0530S, NEB) were confirmed by standard Sanger DNA sequencing (MBSU).

2.5 Motility Assays: Media used in *E. coli* swarming assays consisted of 1% Luria-Bertani (LB) broth (#185288, Fisher BioReagents), 0.5% NaCl, and 0.35% Eiken swarming agar (E-MJ00, Eiken Chemical Company). When required 25 µg/mL chloramphenicol was added for the propagation of plasmids and 0.5 M isopropyl β-D-thiogalactoside (IPTG) (#15529019, Invitrogen) was added for inducing protein expression. After autoclaving 2% glucose (#BP350500, Fisher BioReagents) was added to plates when examining the effects of glucose on swarming. Swarming agar was allowed to dry at room temperature for 1 hour before use. Cells were inoculated at the centre of the swarm plates with the 5 µL of culture (same number of cells on each plate) and plates were allowed to dry and absorb the culture for roughly ten minutes. Afterwards plates were para-filmed and incubated at 37°C overnight. Three biological replicates with four technical replicates were performed for every motility assay conducted. The diameter of the swarms was measured by taking four measurements of each plate diameter and then the average diameter for each plate was recorded. The average relative distances of swarming for the various strains were then compared to our Δ*tsr*+*tsr* strain as our site directed mutants are plasmid expressed and so was our complemented Δ*tsr*+*tsr* strain.

2.6 Protein Extraction & Purification: 6 mL of overnight culture, 600µL of chloramphenicol and 0.5M IPTG was added to a 600mL flask of LB to be incubated at 37°C shaking for six hours. Afterwards proteins were harvested by centrifugation at 15,000g for 10 minutes and a pellet was collected and weighed. Pellets were lysed by the pestle and mortar technique. Pellet and celite was added 1:1 ratio gram along with 1 pellet of protease inhibitor, 1 µL benzonase and lysis buffer A. Cell lysate was spun at 3,000g centrifugation at 4°C for 10 minutes. Supernatant was taken out and spun again at 3,000g centrifugation at 4°C for 10 minutes to remove any remaining celite. The supernatant was collected and centrifuged at 15,000g centrifugation at 4°C for 30 minutes. Pellets were collected and resuspended in 0.1%

triton and left on ice for 10 minutes. A nickel affinity chromatography was performed to purify our His-tag Tsr protein. Buffers A and B were prepared: Buffer A wash buffer contained 500 mM NaCl, 20 mM Tris HCL pH 8, 5 mM imidazole, 10% glycerol and 10 mM β-Mercaptoethanol. Buffer B elution buffer contained 500 mM NaCl, 20 mM Tris-HCL pH 8, 300 mM imidazole, 10% glycerol and 10 mM β-Mercaptoethanol. Supernatant, flow through, wash and elution fractions were collected from the column.

2.7 Western Blotting: SDS-PAGE gel electrophoresis was performed at 200V for 40 minutes using 10% SDS-PAGE gels. Afterwards a membrane transfer using PVDF (polyvinylidene difluoride) membrane at 60V for 60 minutes was performed. Next the membrane was blocked with 15 mL of 5% BSA blocking buffer in PBS-Tween for 1 hour with shaking. Afterwards an anti-His primary antibody conjugated with a Horseradish Peroxidase (HRP) was used to detect the His-tag. The antibody was diluted 1 in 4000 and incubated shaking for 1 hour. The PVDF membrane was washed three times each time for 10 minutes with PBS-Tween. Enhanced chemiluminescence ((#32106, Thermo Scientific) was used for detection of HRP enzyme activity. A 1:1 ratio of Reagent I and Reagent II of the enhanced chemiluminescence solutions (total volume 2 mL) were mixed and added to the surface of the membrane and incubated for two minutes at room temperature. Afterwards the membrane was then imaged.

2.8 Pro-Q Emerald 300 Glycoprotein Staining: Pro-Q Emerald 300 staining method (#P20495, Invitrogen) was used for the specific detection of glycoproteins in 10% SDS-PAGE gels. This method depends on fixation (denaturing of proteins) and washing to remove SDS from proteins. Periodate oxidation is performed and washing after oxidation is done to remove excess periodate which interferes with staining. The Pro-Q 300 stain reacts with periodate-oxidized carbohydrate groups and a bright green, fluorescent signal is seen on present glycoproteins in the gel. A 10% SDS-PAGE gel was run with our protein and samples of interest. Afterwards the gel was washed with distilled water for five minutes in a glass container to remove excess SDS. The gel was fixed with 100 mL of fix solution (50% methanol and 5% acetic acid in distilled H_2O) and shaking for 30 minutes at room temperature. This step was repeated once more. The gel was washed by incubating it in 100 mL of wash solution (3% acetic acid in distilled H_2O) and shaking for 10-20 minutes. This step was repeated once more. The carbohydrates were oxidized by adding 25 mL of oxidizing solution (3% acetic acid to 2.5g of periodic acid) and shaking for 30 minutes. The gel was then washed again by adding 100 mL of wash solution and shaking for

15 minutes. This step was repeated two more times. The Pro-Q Emerald 300 stock solution (6ml of DMF-Dimethylformamide added to Pro-Q Emerald 300 reagent vial) solution was prepared by diluting 500 µL in 25mL of Pro-Q staining buffer solution creating a 50-fold dilution. The gel is incubated in the dark by wrapping the container in aluminum foil and shaking for 90 minutes. The gel was washed for 15 minutes shaking and this step was repeated once more. The gel was then imaged using a clean and wet transilluminator surface with deionized water to prevent any speckling.

2.9 Bacterial Growth Curves of Various Strains: Bacterial growth curves of wild type, Δ*tsr,* Δ*tsr*+*tsr*, OG1, OG2, OG3, and OG4 mutants were performed. Overnight cultures of each strain were incubated at 37°C shaking. Overnight cultures where then put into individual flasks of 200mL of LB (200µL of chloramphenicol added to plasmids with chloramphenicol resistance markers: Δ*tsr*+*tsr*, OG1, OG2, OG3, and OG4) and incubated at 37°C shaking for a total of eight hours. One set of flasks contained 10mL of 2% glucose (#BP350500, Fisher BioReagents) and another set of flasks did not have glucose. The optical density (OD600nm) of each set of flasks were measured at each hourly time point starting at zero hours up until eight hours total.

Chapter 3: Results

3.1 Temperature as an environmental sensor doesn't affect protein glycosylation

O-GlcNAc in eukaryotes is used as an environmental sensor and temperature is an example of a sensor that cells respond to in their environment (Harwood & Hanover 2014, Bond & Hanover 2015, Zachara et al. 2004). To determine if temperature affects protein glycosylation, we subjected *E. coli* to heat shock and assessed protein glycosylation via the Pro-Q Emerald glycoprotein staining technique. Liquid growth samples in LB broth (which promotes swimming motility) were heat shocked at 42°C and grown at 37°C (control) for either 10 or 30 minutes to assess short time points. Lysates were analyzed by a 10% SDS-PAGE gel and stained using the Pro-Q Emerald glycoprotein staining technique (Fig. 4). In Lanes 2-5 we do not see a change in protein glycosylation under the different temperature and time length conditions. This experiment shows us that temperature as an environmental sensor does not drastically affect protein glycosylation under the conditions that were tested.

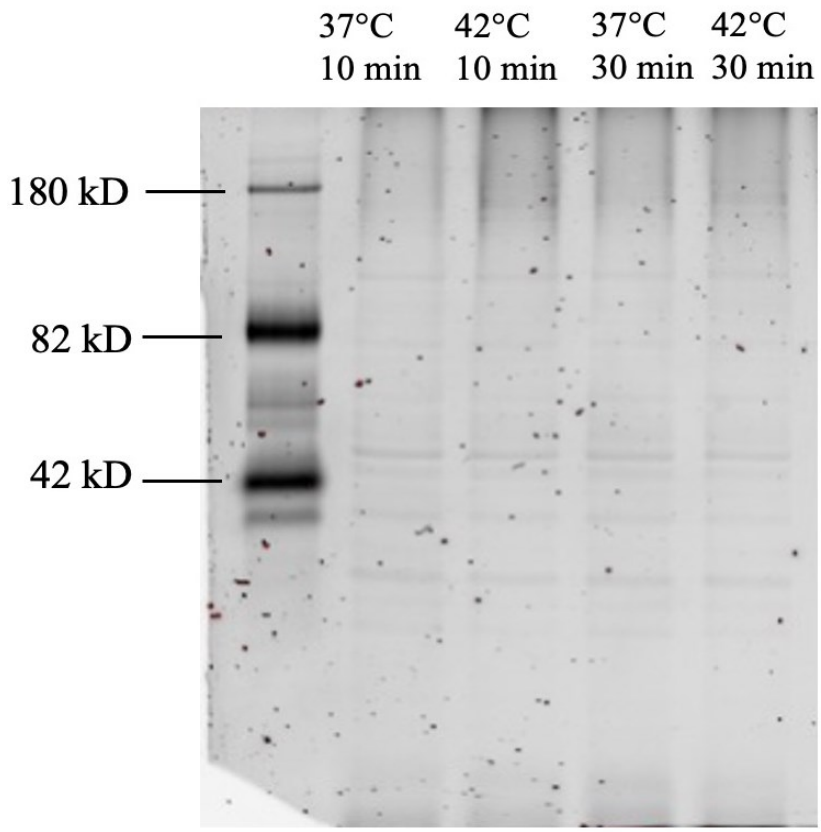

Figure 4: Temperature as an environmental sensor does not affect protein glycosylation. Samples were prepared as liquid growth samples in LB broth which promotes swimming motility. Liquid growth samples were heat shocked at various temperature and time conditions. Samples were harvested and lysed and run on a 10% SDS-PAGE gel. The gel was stained for glycoproteins using the Pro-Q Emerald 300 staining technique. Lane 1 contains the CandyCane glycoprotein molecular weight standard. Lanes 2-3 contain liquid growth samples heat shocked at 37°C and 42°C respectively for 10 minutes. Lanes 4-5 contain liquid growth samples being heat shocked at 37°C and 42°C respectively for 30 minutes.

3.2 Various proteins are glycosylated under different agar and nutrient conditions

The transportation of glucose into cells causes a change in metabolic processes. One of these changes is increased flux through the hexosamine biosynthetic pathway. In eukaryotes, flux through the pathway causes changes in *O*-GlcNAc modifications through an increase in the donor sugar UDP-GlcNAc (Harwood & Hanover 2014). To determine if glucose changes protein glycosylation in *E. coli*, we grew *E. coli* in the presence of glucose and stained it for glycoproteins via the ProQ Emerald glycoprotein staining technique (Fig 5). We examined three

different growth conditions based on different agar concentrations that affect motility and nutrient availability. Agar concentrations affect bacterial motility; 0.35% agar promotes swarming, 1.5% agar inhibits swarming, and 0% agar promotes swimming motility. The same conditions also affect nutrient availability. Plate grown cells have more oxygen than cells grown in liquid samples. Swarming samples have a nutrient gradient, non-motile samples can only consume nutrients in proximity, and swimming samples have no gradient because shaking incubation causes nutrients to be evenly distributed and therefore have no chemotactic abilities.

The patterns are different in all lanes when glucose is added regardless of the growth conditions. It is more of a drastic effect in the swarming condition. When cells are swarming in Lane 2, we see glycoproteins are present. However, when glucose is added in Lane 3, some of the same proteins remain glycosylated (e.g., blue star) but other glycoproteins appear and or disappear on the gel (e.g., yellow star) suggesting that glucose influences protein glycosylation (Fig. 5). We would have expected to see an increase in *O*-GlcNAc because of the potential increase in UDP-GlcNAc, however, it is important to note that this is global protein glycosylation and not specifically *O*-GlcNAc. We do not yet have a method to specifically detect *O*-GlcNAc in bacteria. This experiment shows us that different proteins are being glycosylated under different agar and nutrient concentrations.

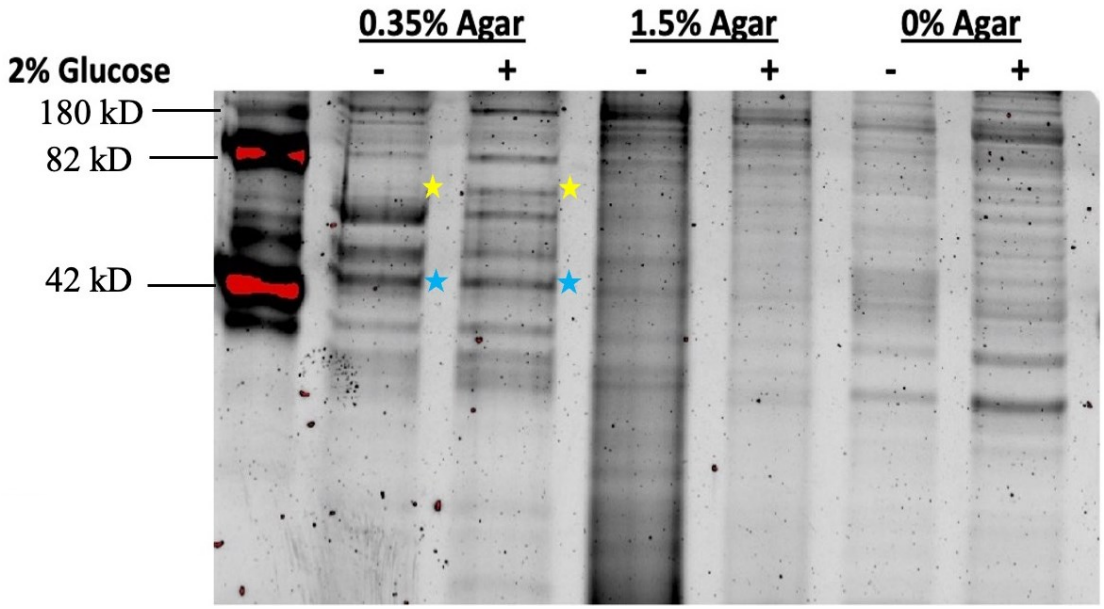

Figure 5: Different proteins are glycosylated under different agar and nutrient concentrations. Swarming motility assays cells grown on plates 0.35% (swarming) or 1.5% (non-motile) in

liquid growth media LB. To each condition we tested the addition of 2% glucose. Cells were induced for six hours at 37°C, harvested and lysed. All samples were run on a 10% SDS-PAGE gel and the gel was stained for glycoproteins using the Pro-Q Emerald 300 staining technique. Equivalent amounts of protein were loaded into each lane. Coloured stars indicated the glycoproteins bands that don not change (blue) and the glycoproteins bands that do change (yellow) in the absence and presence of glucose.

3.3 Developed an assay to examine flagella-based motility

Based on the BEMAP paper, Tsr contains four putative O-GlcNAc modification sites (Boysen et al 2016). To see whether the putative *O*-GlcNAc sites affected motility, we developed a motility assay system. The first thing that was done was we set up an *E. coli* system where *tsr* was knocked out (*Δtsr*) and then complemented with a plasmid that contained our full length *tsr* (*Δtsr+tsr*). We chose to use a low copy plasmid so it does not produce too much protein and is less likely to affect the function within our system since high protein amounts may not integrate in the motility protein system. We used the MG1655 *E. coli* strain, as it is the strain that was used in the BEMAP paper, and genetically is related to the Keio collection, from which the *Δtsr* knockout was obtained (Baba et al. 2006).

Swarming assays were developed with 0.35% agar to promote swarming motility. Swarming assays of our wild type, *Δtsr* knockout and *Δtsr+tsr* strain were performed in plates without and with 2% glucose (Fig. 6). In the absence of glucose, we observe what we expected which is the wild type swarming in a uniform halo pattern, the *Δtsr* knockout having decreased swarming capabilities and the complementation restoring swarming in the *Δtsr* knockout strain (Fig. 6A). This shows us that our plasmid construct is working and restoring swarming capabilities in our complemented strain.

Due to glucose's effects on global protein glycosylation, particularly under swarming conditions, we wanted to determine if glucose also affect motility. When glucose was added we saw a drastic change in swarming pattern in all three strains (Fig. 6B). A dendritic branched swarming pattern was unexpectedly seen in all three strains. In the *Δtsr* knockout we expected the strain to have a decrease in swarming motility like what was seen in the absence of glucose. However, in the presence of glucose, *Δtsr* knockout would swarm the most compared to the wild

type and *Δtsr+tsr* strain. We do not fully understand the mechanism and or reasoning behind the change of swarming pattern when glucose is added.

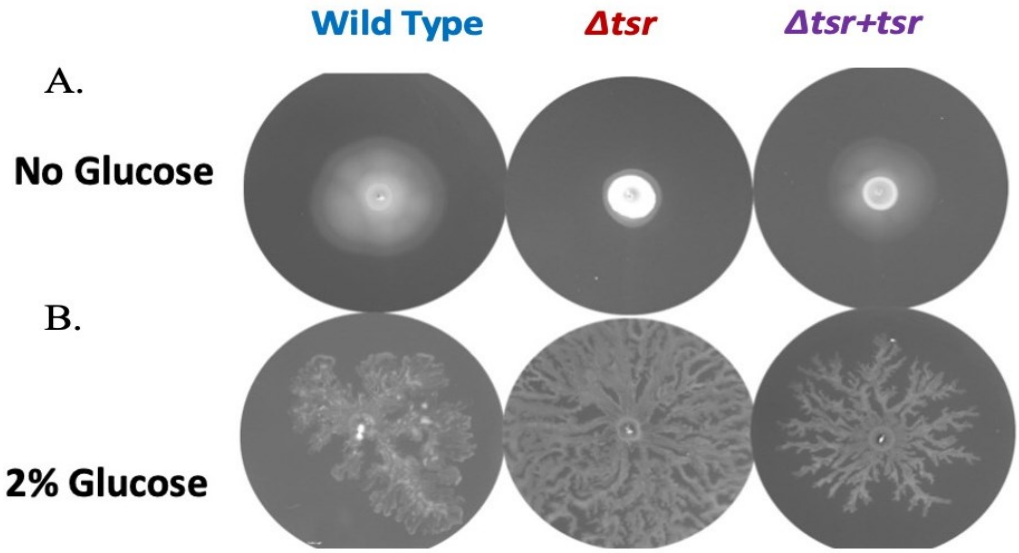

Figure 6: Swarming motility assay of three *E. coli* strains wild type, *Δtsr* knockout and complemented (*Δtsr+tsr*) strains without and with 2% glucose. Equivalent cell numbers were plated as determined by the OD600; equal number of cells inoculated on the center of plates. Swarming assays were conducted in four technical replicates per strain and three biological replicates were performed. Plates were para-filmed and incubated at 37°C overnight.
A) In the absence of glucose, wild type swarms in a uniform halo pattern, *Δtsr* knockouts have decreased swarming capabilities, and swarming is restored in the *Δtsr+tsr* strain. B) In the presence of 2% glucose, a dendritic branched swarming pattern was unexpectedly seen in all three strains.

3.4 Two of the four putative O-GlcNAc sites were important for swarming motility

There are four putative *O*-GlcNAc modification sites in Tsr: S272 (OG1), S282 (OG2), T484 (OG3) and S494 (OG4) (Boysen et. al 2016). To determine if these four putative sites affect motility, OG1-4 modification sites were individually mutated to alanine mutations via Q5 mutagenesis and transformed into *Δtsr* knockout competent cells. In the absence of glucose, all strains swarmed except for OG3 and OG4 which were completely non-motile (Fig 7). The mutants do not phenocopy the knockouts, suggesting that the mechanism of reduced swarming is not simply due to non-functional Tsr. To determine if glucose differently affected the requirement of the four putative sites, we examined the OG mutants' motility abilities in the

presence of 2% glucose using our motility swarming assay system (Fig.7) OG3 and OG4 swarmed significantly less as compared to the complemented strain but retained some motility in the presence of glucose, in contrast to what was observed in absence of glucose. These results indicate that the putative OG3 and OG4 modification sites are highly critical for the motility capabilities of E. coli under the conditions tested.

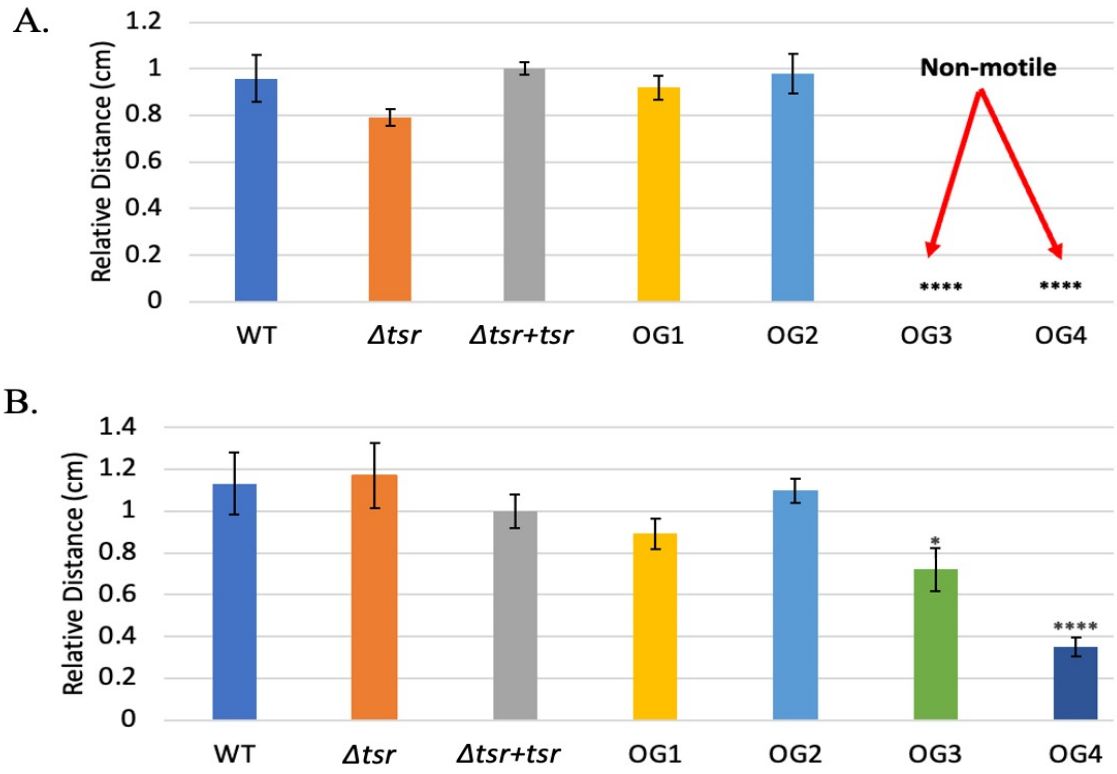

Figure 7: Putative *O*-GlcNAc modification sites 3 and 4 (OG3, OG4) are important for swarming motility. The effect of site-directed mutants of putative *O*-GlcNAc modification sites (OG1-4) on swarming motility distances, compared with wild type, *Δtsr* and *Δtsr+tsr*. A) The relative distances of OG1-4 mutants were compared to our complemented strain (*Δtsr+tsr*) in the absence of 2% glucose. All strains swarmed except for OG3 and OG4 which were significantly non-motile. B) The relative distances of OG1-4 mutants were compared to our complemented strain (*Δtsr+tsr*) in the presence of 2% glucose. OG3 and OG4 swarmed significantly less as compared to the complemented strain. This leads to the conclusion that OG3 and OG4 are highly important for motility capabilities of *E. coli.* Swarming assays were conducted in four technical replicates per strain and three biological replicates were performed. Plates were para-filmed and incubated at 37°C overnight. Error bars refer to the standard deviation. A p-value of 0.05 was used to indicate statistically significant values. p-values were calculated using a student's t-test

via Microsoft Excel. A p-value ≥ 0.05 was not statistically significant, a p-value <0.05 is statistically significant and is flagged with asterisk (*), a p-value <0.01 is flagged with two asterisks (**), a p-value <0.001 is flagged with three asterisks (***), a p-value <0.0001, is flagged with four asterisks (****).

As a control, we assessed the effects of mutating known methylation sites, E297 (MET1), E304 (MET2), E311 (MET3), E493(MET4) and E502 (MET5) on motility (Challah et. al 2005) Of these sites, MET 1-3 have been shown to be the most important sites in terms of function (Challah et. al 2005). We examined the motility of the MET mutants in the absence and presence of 2% glucose using our motility swarming assay system. The relative distances of our MET mutants were compared to our complemented strain (*Δtsr+tsr*). In the absence of glucose, MET2 was significantly less motile compared to our complemented strain (Fig. 8A). In the presence of 2% glucose, MET2 and MET3 significantly swarmed far less compared to our complemented strain (Fig. 8B). These results indicate that the MET2 modification site is highly critical for motility capabilities of *E. coli* under multiple conditions and that MET1 and MET3 regulate motility differently in the absence and presence of glucose.

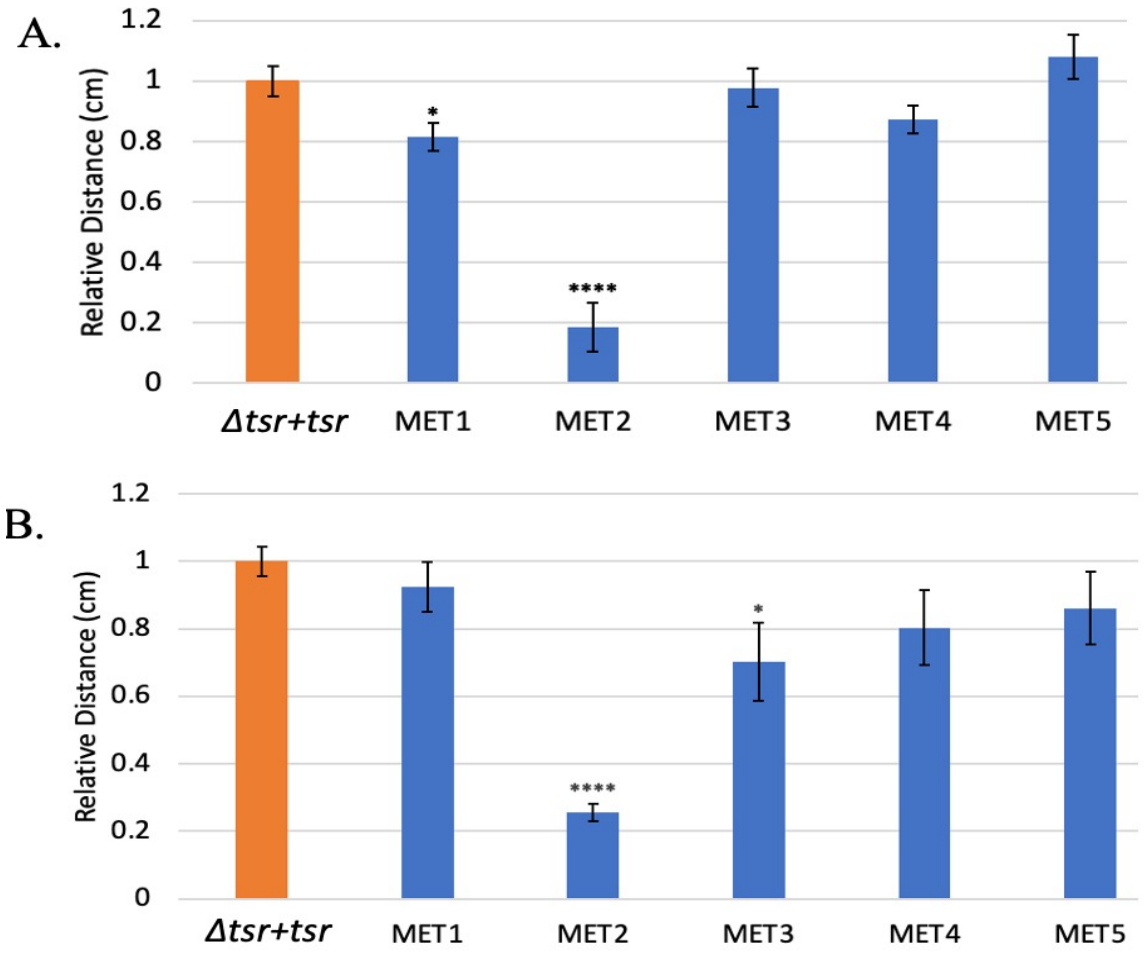

Figure 8: Known methylation modification site 2 (MET2) is important for swarming motility. The effect of site-directed mutants of known methylation modification sites (MET1-5) on swarming motility distances, compared with *Δtsr+tsr*. A) The relative distances of MET1-5 mutants were compared to our complemented strain (*Δtsr+tsr*.) in the absence of 2% glucose. MET2 significantly swarmed far less compared to our complemented strain. B) The relative distances of MET1-5 mutants were compared to our complemented strain (*Δtsr+tsr*.) in the presence of 2% glucose. MET2 and MET3 significantly swarmed far less compared to our complemented strain. MET2 is highly important for motility capabilities of *E. coli.* Swarming assays were conducted in four technical replicates per strain and three biological replicates were performed. Plates were para-filmed and incubated at 37°C overnight. Error bars refer to the standard deviation. A p-value of 0.05 was used to indicate statistically significant values. p-values were calculated using a student's t-test via Microsoft Excel. A p-value ≥ 0.05 was not statistically significant, a p-value <0.05 is statistically significant and is flagged with asterisk (*),

a p-value <0.01 is flagged with two asterisks (**), a p-value <0.001 is flagged with three asterisks (***), a p-value <0.0001, is flagged with four asterisks (****).

3.5 Site-directed mutants of putative O-GlcNAc sites does not affect growth

We next wanted to observe the growth patterns of our putative OG site-directed mutants (OG1-4) to determine if the strains expressing mutants grew normally or if growth was in any way halted. Observing the putative OG mutant growth patterns would indicate if our site-directed mutants were not swarming because they were not properly growing. In Figure 9, a bacterial growth curve experiment was set up to measure the OD600 of the wild type, *Δtsr*, *Δtsr+tsr*, OG1, OG2, OG3 and OG4 mutants in the absence and presence of 2% glucose every hour up to a total of eight hours. In Figure 9A, we have a bacterial growth curve in the absence of glucose. Looking at the growth of the putative OG mutants, we see that their growth over the course of eight hours did not significantly differ from that of the wild type strain. In Figure 9B, we have a bacterial growth curve in the presence of 2% glucose. Looking at the growth of the putative OG mutants, we see that again their growth over the course of eight hours paralleled that of the wildtype, *Δtsr* and *Δtsr+tsr* strains. These growth curves demonstrate that the putative OG mutants did not swarm because they lost their swarming abilities based on their site-directed putative OG sites and not because growth was halted or affected by site-directed mutants.

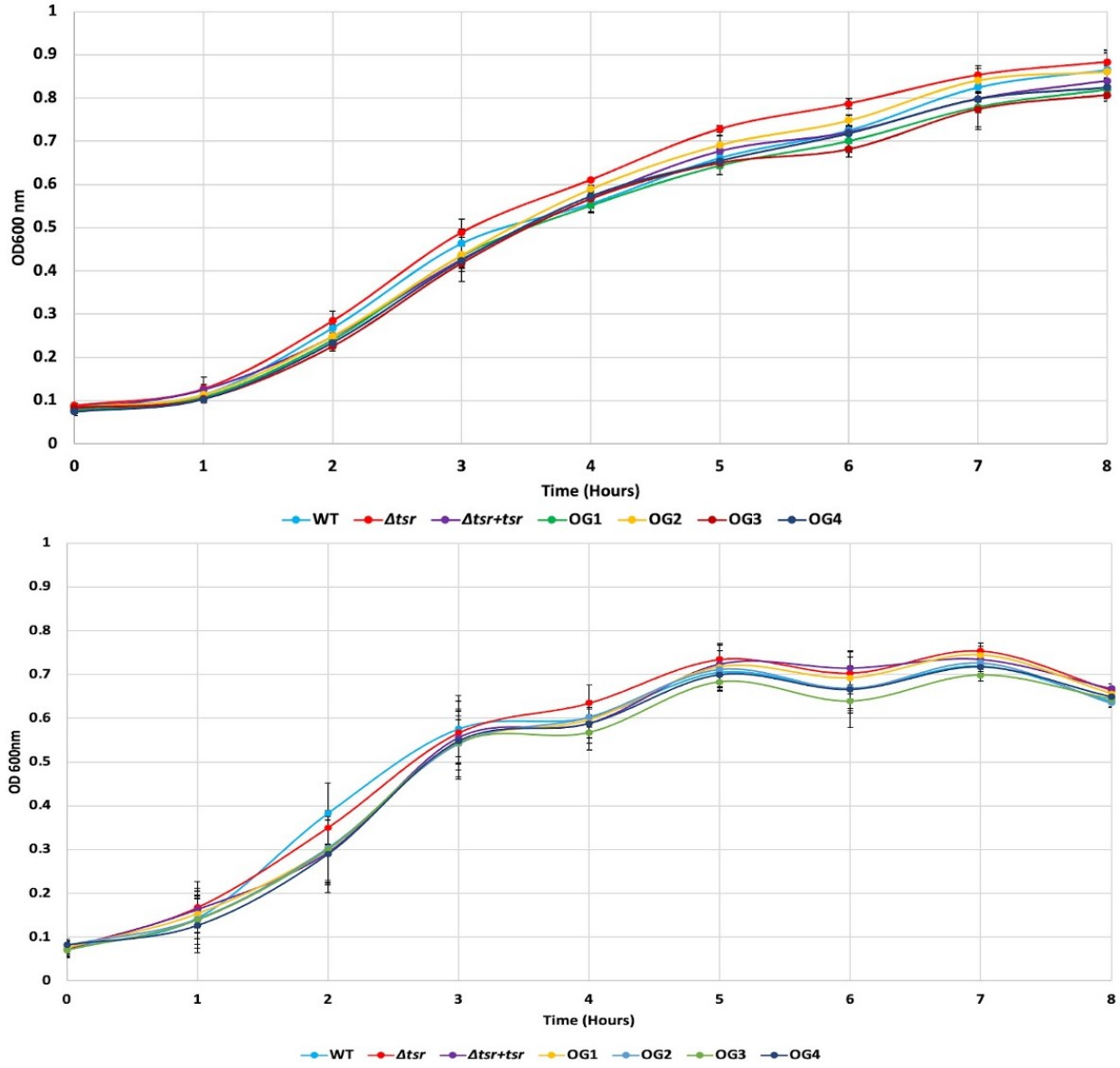

Figure 9: Site-directed mutants of the putative *O*-GlcNAc modification sites (OG1-4) do not affect bacterial growth. Bacterial growth curves of wild type, *Δtsr*, *Δtsr+tsr*, putative OG1, OG2, OG3 and OG4 mutant strains were generated in the absence and presence of 2% glucose. A) Eight-hour growth curve of strains in the absence of 2% glucose. B) Eight-hour growth curve of strains in the presence of 2% glucose. The optical density (OD600) was measured at hourly timepoints from zero up to eight hours total and incubated at 37°C. Each growth curve was done in three separate biological replicates and averaged to obtain a total average OD600 for each timepoint. Error bars refer to the standard deviation. A p-value of 0.05 was used to indicate statistically significant values. The p-values were calculated using a student's t-test via Microsoft Excel.

3.6 Solubility and purification of Tsr protein

To unequivocally show that these sites are modified, it would be necessary to detect them directly via mass-spectrometry analysis. To facilitate this analysis, a purified Tsr protein would be required. We developed a methodology to express and purify Tsr. This was done by adding a polyhistidine tag (His-Tag) with six histidine residues at the N-terminus of Tsr (TSR26). To make sure the His-Tag does not interfere with the biology of Tsr, specifically being able to have endogenous post-translational modifications, we conducted a motility assay experiment with TSR26 and our complemented strain (*Δtsr+tsr*) to see if the His-Tag affected Tsr and its ability to regulate chemotaxis. Figure 10 shows the relative distances of TSR26 compared to our complemented strain in the absence of glucose. There were no significant differences between the TSR26 swarming distance compared to that of our complemented strain. Therefore, we can conclude based on the data obtained from this motility assay that the His-tag at the N-terminus of Tsr does not impede its motility abilities.

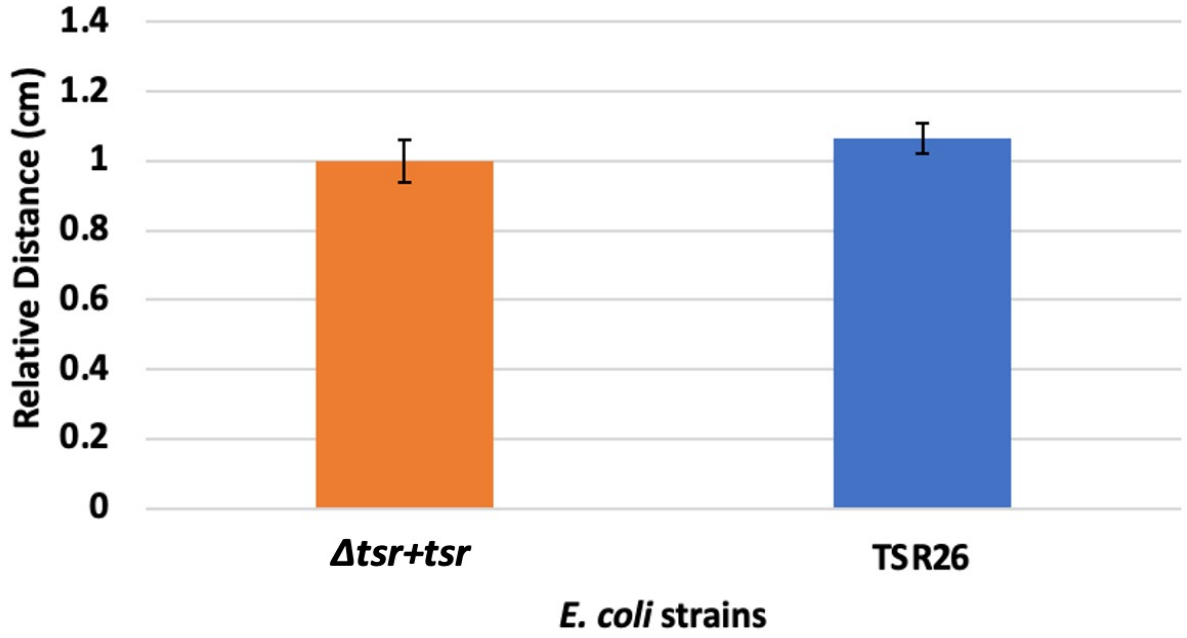

Figure 10: Adding a His-tag to Tsr does not affect its ability to swarm. A polyhistidine tag (His-tag) was added to the N terminus of Tsr (TSR26). The relative distance of TSR26 was compared to our complemented strain (*Δtsr+tsr*.) There were no significant differences of the relative distance swarmed between our complemented strain and TSR26. The His-tag did not interfere with Tsr motility capabilities. Swarming assays were conducted in four technical replicates per

strain and three biological replicates were performed. Plates were para-filmed and incubated at 37°C overnight. Error bars refer to the standard deviation. A p-value of 0.05 was used to indicate statistically significant values. The p-values were calculated using a student's t-test via Microsoft Excel.

To analyze proteins using mass spectrometry, a soluble pure protein is necessary. We assessed the solubility of TSR26 by seeing if it was soluble in the supernatant or the pellet form. TSR26 protein expression was induced with IPTG, and cells were harvested and lysed and 1mL of the lysate, supernatant and pellet (resuspended in 1mL of 0.1% SDS) were kept and the samples were analyzed on a 10% SDS-PAGE gel. Figure 11 shows a Coomassie blue stained gel and an anti-His immunoblot of our samples. In the immunoblot, we can see that a strong signal for the His-tag is visible in the lysate lane as expected and in the pellet lane but not in the supernatant lane (Fig. 11B). The results of the immunoblot demonstrates that TSR26 is not soluble in the supernatant and that all the TSR26 protein remains in pellet form.

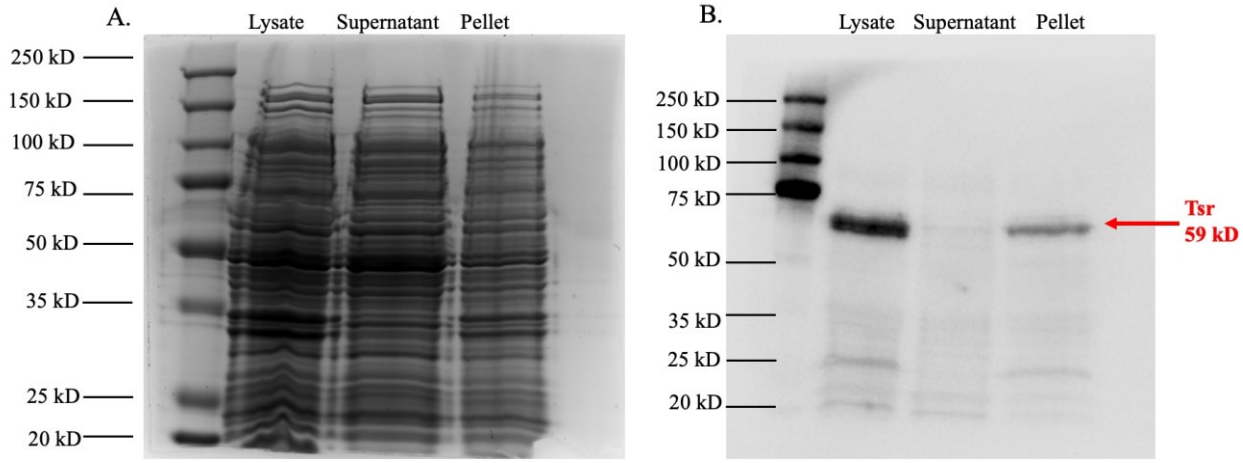

Figure 11: TSR26 protein with an N-terminal His-tag is not soluble. A) A Coomassie blue stained gel and B) An anti-His immunoblot of TSR26 cell lysate, supernatant and pellet after 21,000g spin. Anti-His signal is seen in lysate and pellet but not in the supernatant. Lane 1 is the molecular weight ladder, Lane 2 is the cell lysate, Lane 3 is the supernatant and Lane 4 is the pellet. A 100mg of TSR26 pellet was lysed in 1ml of lysis buffer, harvested and centrifuged. Pellet was resuspended in 1mL 0.1% SDS. Samples were run on a 10% SDS-PAGE gel.

As an insoluble protein cannot be purified or subjected to mass spectrometry analysis, we needed to find conditions under which TSR26 was soluble. We tested the effect of the following

detergents on protein solubility: Triton, NP-40 and SDS. Triton and NP-40 are mild non-ionic detergents that are good at solubilizing membrane proteins and isolating cytoplasmic proteins. SDS is an ionic detergent that has strong denaturing properties. TSR26 cells were harvested, lysed, spun and the pellets were resuspended in the following detergents: in 1mL 0.1% SDS, 0.1% Triton, 1.0% Triton and 0.1% NP-40 detergents. Samples were re-centrifuged; the supernatants were collected, and the pellets were resuspended 0.1% SDS. Figure 12 shows the Coomassie blue stained gel and the anti-His immunoblot of the TSR26 supernatant and pellet samples. From the anti-His immunoblot in Figure 12B, we can see that TSR26 remains in the pellet and is not soluble in 0.1% SDS. In 0.1% and 1.0% Triton along with in 0.1% NP-40, TSR26 is soluble in the supernatant, however TSR26 is the most soluble and abundant in 0.1% Triton compared to the other supernatant samples. Based on the gels, we can clearly see that TSR26 is the most soluble in 0.1% Triton.

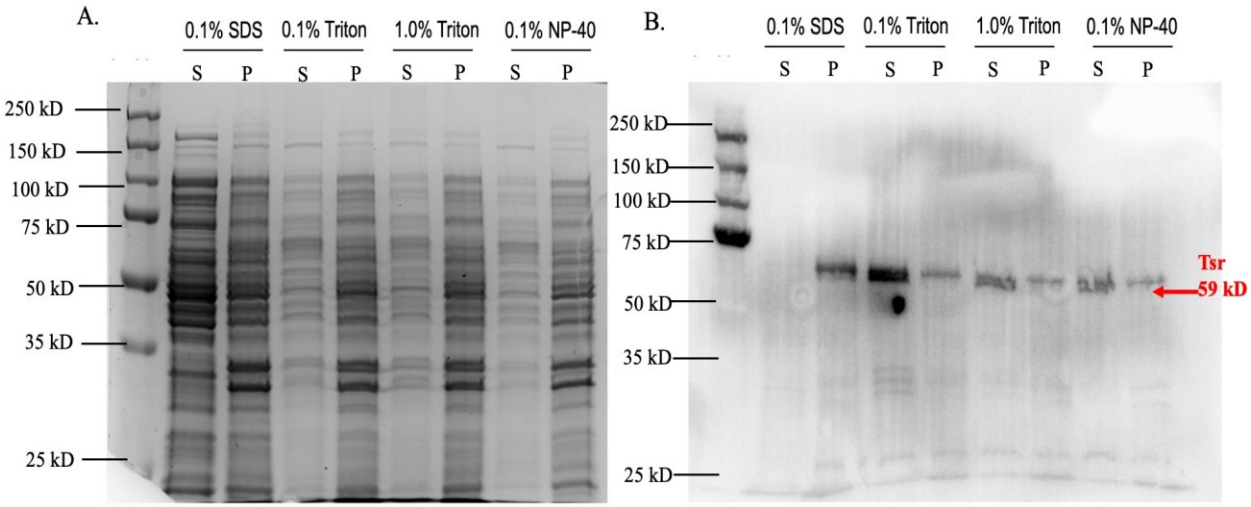

Figure 12: TSR26 protein with N-terminus His-tag is the most soluble in 0.1% Triton. A) A Coomassie blue stained gel and B) An anti-His immunoblot of TSR26 detergent supernatants and pellet samples. First lane is the molecular weight ladder. 100mg of TSR26 cells was lysed in 1mL of lysis buffer and spun at 21,000 g and then pellets were resuspended in 1mL of the following detergents: 0.1% SDS, 0.1% Triton, 1.0% Triton and 0.1% NP-40. These samples were re-centrifuged at 21,000 g. Supernatants (S) were collected and the pellets (P) were resuspended 0.1% SDS.

We wanted the lowest Tsr expression to not overwhelm the system, so we evalutaed different induction times for Tsr expression. Protein expression after induction with IPTG was measured at two, four and six hours at 37°C incubation. Cells were harvested, lysed, centrifuged and the pellets were resuspended in 0.1% Triton and supernatants were analyzed on a 10% SDS-PAGE gel. Figure 13 shows the Coomassie blue stained gel and the anti-His immunoblot of the TSR26 supernatants at their respective induction times. In Figure 13B, we see that TSR26 protein expression was only seen at the six-hour induction period. Based on this data, six hours of induction is the minimum time required to obtain protein expression of TSR26.

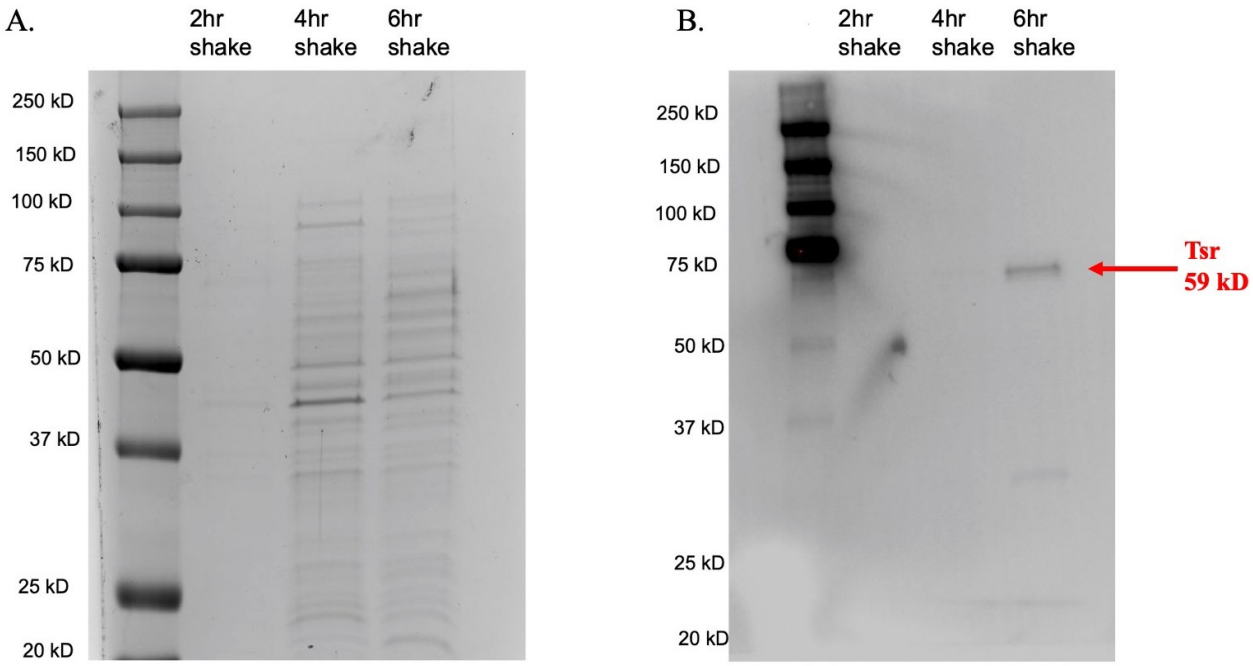

Figure 13: TSR26 protein expression requires a minimum of six hours of induction with IPTG. A) A Coomassie blue stained gel and B) An anti-His immunoblot of TSR26 at various shaking incubation time periods to determine minimum protein expression. Lane 1 is the molecular weight ladder, Lane 2 is two-hour induction, Lane 3 is four-hour induction and Lane 4 is six-hour induction. Protein expression and induction was done with IPTG, and samples were taken at two, four, and six hours of induction. Cells were lysed and the pellets were resuspended in 1mL 0.1% Triton. The supernatants were run on a 10% SDS-PAGE gel and an anti-His immunoblot was performed.

We were now able to purify TSR26 protein by using standard purification techniques. The pellet was resuspended in 0.1% Triton and the protein was purified from the resuspension. using a nickel column affinity chromatography (Fig. 14). The anti-His immunoblot shows that the purification was successful. We see in Lanes 5-10 that our purified TSR26 protein (59 kD) was successfully eluted but only partially purified (Fig. 14). This purification was reproducible for a total of three replicates. However afterwards we were unable to reproduce the purification of TSR26 for reasons we were not able to figure out. This will need to be addressed in the future to perform mass spectrometry analysis to determine if Tsr is modified.

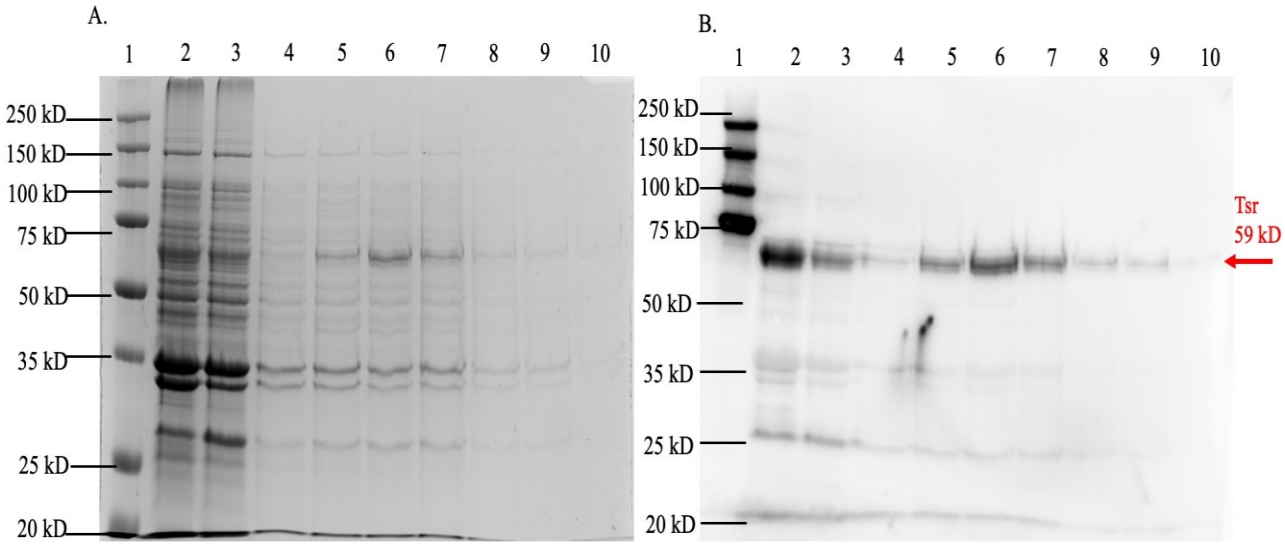

Figure 14: Partial purification of TSR26 using a nickel column affinity chromatography. Purification was reproducible and repeated three times. A) A Coomassie blue stained gel and B) An anti-His immunoblot of TSR26. Lane 1 is the molecular weight ladder, Lane 2 is the supernatant, Lane 3 is the flow through, Lane 4 is the wash, Lanes 5-10 are the elution fractions collected and purified. Protein expression and induction were done with IPTG at six hour shaking induction. Pellet was resuspended in 1mL of 0.1% Triton and resuspension was put through the nickel column affinity chromatography.

Chapter 4: Discussion & Conclusion

4.1 Putative sites OG3 & OG4 are important for motility

We have shown that the putative OG3 and OG4 modification sites are important for bacterial swarming motility, under our specific utilized conditions: agar, NaCl, nutrients (LB broth) and oxygen level requirements (Fig 7). The lack of motility with OG3 and OG4 was not

caused by growth defects (Fig 9). We only conducted swarming specifically at 37°C, ambient oxygen levels, and in rich media. We did not create an artificial gradient, but rather a natural gradient exists in swarming plates because of local consumption of nutrients by bacteria and the production of waste. It is however possible that OG1 and or OG2 may be important putative sites at lower temperatures or stronger gradient concentrations for further adaptation.

The results for the control MET mutants were consistent with previous studies (Challah et al. 2005). There are critical roles for MET1-3 sites in serine-specific adaptation, while the roles of MET4-5 sites are relatively unknown (Challah et al. 2005). We have shown that the MET2 modification site plays a significant role in swarming, with MET1 and MET3 playing minor roles. The MET1-3 modification sites are near the putative OG3 and OG4 modification sites (Fig. 15), and it is possible that the known MET and putative OG modifications may influence each other, as is in eukaryotes, or may even have independent roles of each other.

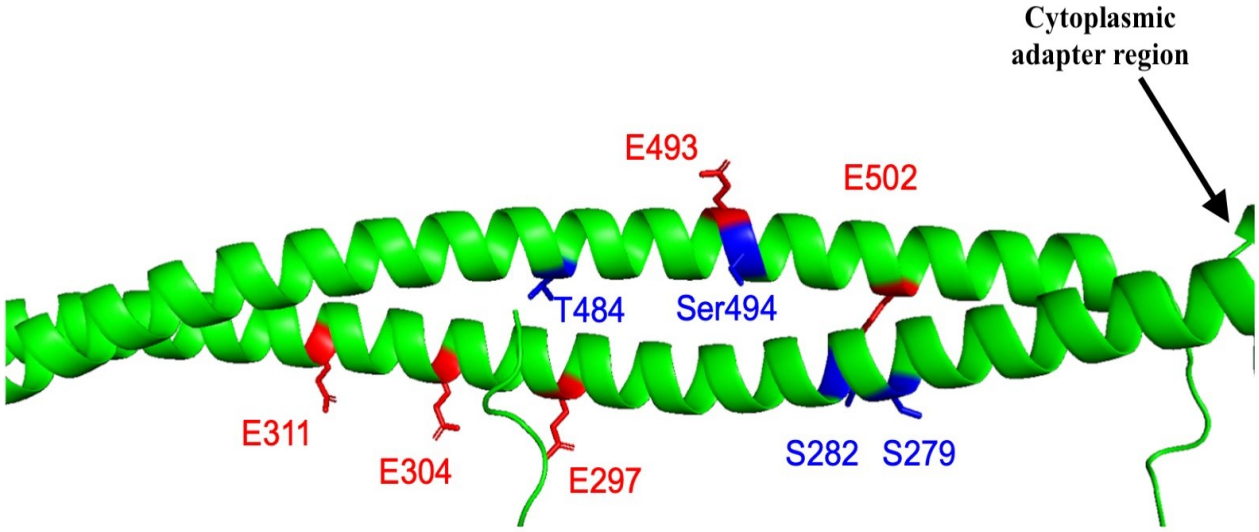

Figure 15: Putative *O*-GlcNAc modification sites are near the known methylation modification sites of the Tsr chemoreceptor. Tsr has a periplasmic and an intracellular signalling domain, which is composed of a coiled-coil of two antiparallel α-helices (shown in green) connected by a U-turn. The four putative *O*-GlcNAc modification sites are highlighted in blue: S279 (OG1), S282 (OG2), T484 (OG3), and S494 (OG4) and the five known methylation sites are highlighted in red: E297 (MET1), E304 (MET2), E311 (MET3), E493 (MET4), and E502 (MET5). This structure is predicted from the AlphaFold Protein Structure Database. Created by PyMOL.

4.2 Different modes of motility affect protein glycosylation

Different modes of motility seem to affect global protein glycosylation under the conditions tested. However, the effect of motility on protein glycosylation is difficult to directly measure as there are lots of environmental conditions that change with the growth conditions. For example, nutrient uptake is different based on the mode of motility conducted (Kearns 2010). In our experiments on swimming motility, we incubated with shaking, which homogenizes the environment and evenly distributes nutrients, preventing chemotaxis from occurring (Kearns 2010). Swarming motility has a chemotaxis gradient because of nutrient consumption as bacteria swarm further from the center of the plates towards higher areas of nutrients (Kearns 2010). The non-motile is not able to be chemotactic because of specific agar conditions (1.5% agar inhibits swarming) (Kearns 2010). Oxygen levels are also varied depending on the mode of motility: more oxygen is present on the plate assays as opposed to swimming (Hölscher et al. 2015). Results seen from swimming and swarming motility assays could be due to the nutrient and oxygen requirements, which could affect global protein glycosylation. One aspect we could change going forward would be looking at varied oxygen levels between the different modes of motility and seeing if oxygen requirements affect protein glycosylation. Introducing a chemical gradient in the assay plates (non-homogenous in terms of nutrients) could be another aspect to investigate to see if chemical stimulus plays a role in protein glycosylation.

4.3 Increased glucose changes protein glycosylation & affects motility

The addition of glucose to media changes protein glycosylation, causing differential patterns of glycosylation. Some glycoproteins appeared unaffected while additional glycoproteins appeared and/or disappeared. By increasing glucose levels and flux into cells, bacteria can switch their metabolic processes to glucose-utilizing mechanisms (Brown et al. 2008, Passalacqua et al. 2015). In oxygen limiting environments, bacteria can switch from aerobic respiration to fermentation, and this changes how cells respond to the availability of glucose (Brown et al. 2008). In the presence of oxygen, glycolysis breaks down glucose to pyruvate, shuttles it to the Krebs cycle and then into the electron transport chain (Passalacqua et al. 2015). The electron transport chain creates the proton motive force gradient, which enables ATP production (Passalacqua et al. 2015). These different processes require different proteins, which may show up as differences in our glycoprotein-stained gels.

Increasing the influx of glucose into bacterial cells affected the motility capacities of *E. coli*, possibly by affecting the transcription of the *lac* operon of *E. coli* (Brown et al. 2008). Catabolite repression is the inhibition of machinery that processes nutrients other than the preferred energy source in this case glucose (Brown et al. 2008). When intracellular glucose levels are low, levels of the signalling molecule cyclic AMP (cAMP) are high and cAMP can bind with catabolite activator protein (CAP), which binds upstream of the *lac* operon (Brown et al. 2008). This leads to enhanced RNA polymerase binding, which leads to increased transcription rates and increased protein production (Brown et al. 2008). When intracellular glucose levels are high, cAMP levels are decreased and RNA polymerase does not bind to DNA, leading to low levels of transcription (Brown et al. 2008). Therefore, less protein expression from our plasmid would occur and could potentially affect genomic expression of genes. As a result of catabolite repression, Tsr protein expression may decrease, restoring its motility capabilities to those shown in the Δtsr motility assays with glucose, which may explain why cells partially restore swarming abilities in glucose-treated cells. (Fig. 6B & 7B). Expression of mutant OG strains in the absence of glucose could cause intracellular glucose levels to be at low levels thus allowing the increased transcription rates and increased production of Tsr protein expression (Fig 7A). If we had an antibody that could detect Tsr we could possibly test for the Tsr expression in the absence or presence of glucose.

In addition to changing protein glycosylation, adding glucose to our media plates substantially changed the swarming patterns of *E. coli*. In all three strains, the swarming patterns drastically changed from a halo pattern to a branched dendritic pattern when glucose was introduced into the assay (Fig. 6B). We saw that the *Δtsr* knockouts surprisingly swarmed further as compared to the wild type strains along with some motility being restored to OG mutants (Fig 7B). A dendrite is a long, thin branch of colonization that emerges from the center of the inoculation (Kearns 2010). The formation of dendritic patterns has been observed in bacteria like *Pseudomonas aeruginosa* where multiple surfactants are secreted (Caiazza et al. 2005). Under certain conditions, dendrites can form when the rate of motility exceeds the rate of bulk population growth (Julkowska et al. 2004). Glucose could be a possible surfactant that causes the dendritic pattern to form (Kearns 2010). *E. coli* swarm when a wetting agent is present, but the wetting agent that promotes *E. coli* swarming remains unknown (Kearns 2010). An alternative type of motility could also be enabled by glucose (Burkart et al. 1998). Sliding motility can also

result in dendritic patterns in bacteria (Murray & Kazmierczak 2008). There is no active motor required for sliding motility, which relies on again on surfactants to reduce surface tension to enable cells to spread away from their origins (Kearns 2010).

We observed that the *Δtsr* knockouts which were non-motile in non-glucose media, were partially motile in the presence of glucose. Different MCPs could possibly assist in restoring motility in *Δtsr* knockout strains. In the presence of glucose, *Δtsr* cells swarmed the furthest out of all the strains (Fig. 6B). This was not our expected result as we hypothesized that knocking out Tsr would impede its motility capabilities but when glucose was added, it restored its swarming capabilities. In line with previous literature (Harshey et al. 1994, Burkart et al. 1998), in the presence of glucose, *Δtsr* knockout cells are capable of swarming because they have other chemoreceptors and an intact Che signalling pathway. Indeed, *E. coli* mutants that lack any one of the individual four MCPs (*Δtsr, Δtar, Δtrg* or *Δtap*) still have the capabilities to swarm (Harshey et al. 1994, Burkart et al. 1998). Tar alone could potentially support swarming of *Δtsr* knockouts as Tsr and Tar are the most abundant chemotactic proteins in *E. coli* (Burkart et al. 1998). Tar could be potentially up regulated in the chemotaxis signaling cascade pathway in response to Δtsr knockouts and enable and/or restore swarming capabilities. Only when three or four MCPs are deleted or when Che proteins are mutated, is swarming completely abolished (Harshey et al. 1994). This illustrates the importance of the regulatory pathway in swarming and adaptation of bacteria to their environments. *E. coli* require at least two functional MCPs to initiate the chemotaxis signalling pathway (Harshey et al. 1994, Burkart et al. 1998). Phosphorylation of Che proteins in the signalling cascade cannot transpire without a fully functional signalling pathway present (Harshey et al. 1994, Parkinson 2003, Parkinson et al. 2015). These results could be explained by upregulation of chemotaxis proteins or other possible signalling pathways.

4.4 Future Directions

As a means of demonstrating unequivocally that OG3 and OG4 sites are modified with *O*-GlcNAc, we initiated a methodology process to purify Tsr for potential mass spectrometry analysis. The purification of Tsr needs to be addressed for future considerations. Because the N-terminal His-tag did not function as expected, changes need to be implemented to fully express and purify the Tsr protein. We may need to make a few minor adjustments to the construct that

we have made, such as moving the His-tag to the C-terminus as Tsr purification using a C-terminal His-tag proved to be functional (Lee & Kim 2009). It is possible that a completely different protein tag would even be required for the successful purification of any new construct. A different plasmid vector with a different antibiotic resistance marker could also be used. There is also the possibility of experimenting with different growth conditions. Once these issues are addressed, Tsr and any subsequent mutant Tsr constructs could be purified, and the number of post-translational modifications can be determined by mass spectrometry analysis to determine if Tsr is potentially modified by *O*-GlcNAc modifications (Escobar et al. 2020).

4.5 Conclusion

In summary, we have shown that two of the four putative modification sites, OG3 (T484) and OG4 (S494), are necessary for motility under the conditions tested, as site-directed mutants have significantly reduced motility capabilities. Our work has provided a substantial foundation to further investigate the existence of *O*-GlcNAc modifications in bacteria.

References

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology*, 2, 2006.0008.

Balonova, L., Hernychova, L., & Bilkova, Z. (2009). Bioanalytical tools for the discovery of eukaryotic glycoproteins applied to the analysis of bacterial glycoproteins. *Expert review of Proteomics*, *6*(1), 75-85.

Barreteau, H., Kovač, A., Boniface, A., Sova, M., Gobec, S., & Blanot, D. (2008). Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS microbiology reviews*, *32*(2), 168-207.

Bond, M. R., & Hanover, J. A. (2015). A little sugar goes a long way: the cell biology of O-GlcNAc. *Journal of Cell Biology*, 208(7), 869-880.

Boysen, A., Palmisano, G., Krogh, T. J., Duggin, I. G., Larsen, M. R., & Møller-Jensen, J. (2016). A novel mass spectrometric strategy "BEMAP" reveals Extensive O-linked protein glycosylation in Enterotoxigenic Escherichia coli. *Scientific reports*, *6*(1), 1-13.

Brown, W., Ralston, A. & Shaw, K. (2008) Positive transcription control: The glucose effect. *Nature* Education 1(1):202.

Burkart, M., Toguchi, A., & Harshey, R. M. (1998). The chemotaxis system, but not chemotaxis, is essential for swarming motility in Escherichia coli. *Proceedings of the National Academy of Sciences*, *95*(5), 2568-2573.

Caiazza, N. C., Shanks, R. M., & O'toole, G. A. (2005). Rhamnolipids modulate swarming motility patterns of Pseudomonas aeruginosa. *Journal of bacteriology*, 187(21), 7351-7361.

Chalah, A., & Weis, R. M. (2005). Site-specific and synergistic stimulation of methylation on the bacterial chemotaxis receptor Tsr by serine and CheW. *BMC microbiology*, *5*(1), 1-13.

Champasa, K., Longwell, S. A., Eldridge, A. M., Stemmler, E. A., & Dube, D. H. (2013). Targeted identification of glycosylated proteins in the gastric pathogen Helicobacter pylori (Hp). *Molecular & Cellular Proteomics*, *12*(9), 2568-2586.

Chang, A. Y., Chau, V., Landas, J. A., & Pang, Y. (2017). Preparation of calcium competent Escherichia coli and heat-shock transformation. *JEMI methods*, 1, 22-25.

Clarke, A. J., Hurtado-Guerrero, R., Pathak, S., Schüttelkopf, A. W., Borodkin, V., Shepherd, S. M., ... & Van Aalten, D. M. (2008). Structural insights into mechanism and specificity of O-GlcNAc transferase. *The EMBO journal*, *27*(20), 2780-2788.

Dennis, R. J., Taylor, E. J., Macauley, M. S., Stubbs, K. A., Turkenburg, J. P., Hart, S. J., ... & Davies, G. J. (2006). Structure and mechanism of a bacterial β-glucosaminidase having O-GlcNAcase activity. *Nature structural & molecular biology*, *13*(4), 365-371.

Escobar, E. E., King, D. T., Serrano-Negrón, J. E., Alteen, M. G., Vocadlo, D. J., & Brodbelt, J. S. (2020). Precision mapping of O-linked N-acetylglucosamine sites in proteins using ultraviolet photodissociation mass spectrometry. *Journal of the American Chemical Society*, 142(26), 11569-11577.

Förster, S., Welleford, A. S., Triplett, J. C., Sultana, R., Schmitz, B., & Butterfield, D. A. (2014). Increased O-GlcNAc levels correlate with decreased O-GlcNAcase levels in Alzheimer disease brain. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1842(9), 1333-1339.

Fredriksen, L., Mathiesen, G., Moen, A., Bron, P. A., Kleerebezem, M., Eijsink, V. G., & Egge-Jacobsen, W. (2012). The major autolysin Acm2 from Lactobacillus plantarum undergoes cytoplasmic O-glycosylation. *Journal of bacteriology*, *194*(2), 325-333.

Fredriksen, L., Moen, A., Adzhubei, A. A., Mathiesen, G., Eijsink, V. G., & Egge-Jacobsen, W. (2013). Lactobacillus plantarum WCFS1 O-linked protein glycosylation: an extended spectrum of target proteins and modification sites detected by mass spectrometry. *Glycobiology*, 23(12), 1439-1451.

Gao, H., Shi, M., Wang, R., Wang, C., Shao, C., Gu, Y., & Yu, W. (2018). A widely compatible expression system for the production of highly O-GlcNAcylated recombinant protein in Escherichia coli. *Glycobiology*, 28(12), 949-957.

Guianvarc'h, D., Bourdreux, Y., Biot, C., & Vauzeilles, B. (2021). Metabolic Labeling of Bacterial Glycans.

Harshey, R. M., & Matsuyama, T. (1994). Dimorphic transition in Escherichia coli and Salmonella typhimurium: surface-induced differentiation into hyperflagellate swarmer cells. *Proceedings of the National Academy of Sciences*, *91*(18), 8631-8635.

Hart, G. W. (2019). Nutrient regulation of signaling and transcription. Journal of Biological Chemistry, 294(7), 2211-2231.

Harwood, K. R., & Hanover, J. A. (2014). Nutrient-driven O-GlcNAc cycling–think globally but act locally. Journal of cell science, 127(9), 1857-1867.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical journal*, *280*(2), 309-316.

Hölscher, T., Bartels, B., Lin, Y. C., Gallegos-Monterrosa, R., Price-Whelan, A., Kolter, R., ... & Kovács, Á. T. (2015). Motility, chemotaxis and aerotaxis contribute to competitiveness during bacterial pellicle biofilm development. *Journal of molecular biology*, 427(23), 3695-3708.

Isono, T. (2011). O-GlcNAc-specific antibody CTD110. 6 cross-reacts with N-GlcNAc2-modified proteins induced under glucose deprivation. PloS one, 6(4), e18959.

Julkowska, D., Obuchowski, M., Holland, I. B., & Séror, S. J. (2004). Branched swarming patterns on a synthetic medium formed by wild-type Bacillus subtilis strain 3610: detection of different cellular morphologies and constellations of cells as the complex architecture develops. *Microbiology*, 150(6), 1839-1849.

Kearns, D. B. (2010). A field guide to bacterial swarming motility. *Nature Reviews Microbiology*, 8(9), 634-644.

Lee, J., & Kim, S. H. (2009). High-throughput T7 LIC vector for introducing C-terminal poly-histidine tags with variable lengths without extra sequences. *Protein expression and purification*, 63(1), 58-61.

Leickt, L., Bergström, M., Zopf, D., & Ohlson, S. (1997). Bioaffinity chromatography in the 10 mM range of Kd. *Analytical biochemistry*, 253(1), 135-136.

Liu, F., Shi, J., Tanimukai, H., Gu, J., Gu, J., Grundke-Iqbal, I., ... & Gong, C. X. (2009). Reduced O-GlcNAcylation links lower brain glucose metabolism and tau pathology in Alzheimer's disease. Brain, 132(7), 1820-1832.

Ma, J., & Hart, G. W. (2014). O-GlcNAc profiling: from proteins to proteomes. *Clinical proteomics*, *11*(1), 1-16.

Ma, J., Hou, C., & Wu, C. (2022). Demystifying the O-GlcNAc Code: A Systems View. *Chemical Reviews*.

Mengin-Lecreulx, D., Flouret, B., & van Heijenoort, J. (1983). Pool levels of UDP N-acetylglucosamine and UDP N-acetylglucosamine-enolpyruvate in Escherichia coli and correlation with peptidoglycan synthesis. *Journal of bacteriology*, *154*(3), 1284–1290.

Messner, P. (2004). Prokaryotic glycoproteins: unexplored but important. *Journal of bacteriology,* 186(9), 2517-2519.

Monsigny, M., Roche, A. C., Sene, C., Maget-Dana, R., & Delmotte, F. (1980). Sugar-lectin interactions: how does wheat-germ agglutinin bind sialoglycoconjugates?. *European journal of biochemistry*, *104*(1), 147–153.

Murray, T. S., & Kazmierczak, B. I. (2008). Pseudomonas aeruginosa exhibits sliding motility in the absence of type IV pili and flagella. *Journal of bacteriology*, 190(8), 2700–2708.

Ostrowski, A., Gundogdu, M., Ferenbach, A. T., Lebedev, A. A., & Van Aalten, D. M. (2015). Evidence for a functional O-linked N-acetylglucosamine (O-GlcNAc) system in the thermophilic bacterium Thermobaculum terrenum. *Journal of Biological Chemistry*, *290*(51), 30291-30305.

Parkinson, J. S. (2003). Bacterial chemotaxis: a new player in response regulator dephosphorylation. *Journal of bacteriology*, 185(5), 1492-1494.

Parkinson, J. S., Hazelbauer, G. L., & Falke, J. J. (2015). Signaling and sensory adaptation in Escherichia coli chemoreceptors: 2015 update. *Trends in microbiology*, 23(5), 257-266.

Passalacqua, K. D., Charbonneau, M. E., & O'riordan, M. X. (2016). Bacterial metabolism shapes the host–pathogen interface. Virulence Mechanisms of Bacterial Pathogens, 15-41.

Peters, B. P., Ebisu, S., Goldstein, I. J., & Flashner, M. (1979). Interaction of wheat germ agglutinin with sialic acid. *Biochemistry*, *18*(24), 5505-5511.

Rao, F. V., Dorfmueller, H. C., Villa, F., Allwood, M., Eggleston, I. M., & Van Aalten, D. M. (2006). Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. *The EMBO journal*, *25*(7), 1569-1578.

Riu, I. H., Shin, I. S., & Do, S. I. (2008). Sp1 modulates ncOGT activity to alter target recognition and enhanced thermotolerance in E. coli. Biochemical and biophysical research communications, 372(1), 203-209.

Schirm, M., Kalmokoff, M., Aubry, A., Thibault, P., Sandoz, M., & Logan, S. M. (2004). Flagellin from Listeria monocytogenes is glycosylated with β-O-linked N-acetylglucosamine. *Journal of bacteriology*, *186*(20), 6721-6727.

Schirm, M., Schoenhofen, I. C., Logan, S. M., Waldron, K. C., & Thibault, P. (2005). Identification of unusual bacterial glycosylation by tandem mass spectrometry analyses of intact proteins. *Analytical chemistry*, *77*(23), 7774-7782.

Shen, A., Kamp, H. D., Gründling, A., & Higgins, D. E. (2006). A bifunctional O-GlcNAc transferase governs flagellar motility through anti-repression. *Genes & Development*, *20*(23), 3283-3295.

Thompson, J. W., Griffin, M. E., & Hsieh-Wilson, L. C. (2018). Methods for the detection, study, and dynamic profiling of O-GlcNAc glycosylation. In *Methods in enzymology* (Vol. 598, pp. 101-135). Academic Press.

Upreti, R. K., Kumar, M., & Shankar, V. (2003). Bacterial glycoproteins: functions, biosynthesis, and applications. *Proteomics,* 3(4), 363-379.

Wulff-Fuentes, E., Berendt, R. R., Massman, L., Danner, L., Malard, F., Vora, J., ... & Stichelen, O. V. (2021). The human O-GlcNAcome database and meta-analysis. Scientific data, 8(1), 1-11.

Yang, W. H., Kim, J. E., Nam, H. W., Ju, J. W., Kim, H. S., Kim, Y. S., & Cho, J. W. (2006). Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. *Nature cell biology*, *8*(10), 1074-1083.

Zachara, N. E., O'Donnell, N., Cheung, W. D., Mercer, J. J., Marth, J. D., & Hart, G. W. (2004). Dynamic O-GlcNAc modification of nucleocytoplasmic proteins in response to stress: a survival response of mammalian cells. *Journal of Biological Chemistry*, *279*(29), 30133-30142.