Investigating the role of *N*-linked glycosylation system in bacteria

Sherif Abouelhadid

Supervisor: Professor Brendan Wren

Co-supervisor: Dr Jon Cuccui

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of London

March 2019

Department of Pathogen Molecular biology

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by: No funding received

Research group affiliation: Brendan Wren's group

Declaration

I, Sherif Abouelhadid, declare that the work presented here in this thesis is my own work. I confirm that I have acknowledged all results from published work from others. I confirm that I acknowledge the contribution of others in this thesis.

Name: **Date:**

Dedication

I dedicate this work to my wonderful mother. Words fail to comprehend how much I am grateful for everything thing you did for me. Mom… you are a legend, literally.

Acknowledgments

I would like to express my appreciation to my PhD supervisor Prof Brendan W Wren for his support and guidance. I am sincerely grateful to Dr Jon Cuccui for his incredible help, insightful discussions, inspiring dedication, outstanding molecular biology training without which this work would not have been possible. I owe you a lot, Jon. I would like also to thank the Glycogroup for the fruitful discussions, remarks and suggestions that always keep me on track. Simon North for the enjoyable Friday discussions.

Big thanks to my school friends for bearing up all these years with my sense of humour. Special thanks to the BioKMT members for keeping me motivated and inspired. I am very grateful to those who believed in me, even before I believed in myself. Mr Ezz El Deen Abdel Razek for the poetry and verses that I will never forget. My brothers and sisters who took to the streets on 25th January 2011 to mark a new beginning for Egypt, your struggle against injustice and corruption is inspirational, I keep your memory in my heart "*Al ya's kheyana*". I am thankful to Dr Amr Safwat and Dr Ihab for fixing my back so that I can walk again with no pain.

My brother, who plays a big part in my life, teaching me how to give generously, work hard, think insightfully, and most importantly, how to be a man. My father who passed away young, leaving behind a great reputation that I aspire to meet. My mother to whom I owe every good thing in my life. My beautiful grandmother who taught me how to love endlessly.

To those who told me that I would never be able to make it, thank you for making me more determined.

It would be another 100 pages to thank everyone who helped me to reach where I am now.

Abstract

Glycosylation is the most abundant protein post-translational modification found in nature. The attachment of glycans to proteins has been shown to play a central role in modulating protein folding, stability, and signalling. In recent years, it has become evident that glycosylation systems are found in all domains of life. Advances in genomics and mass spectrometry have revealed several types of glycosylation systems in bacteria. However, how and why bacterial proteins are modified remains poorly defined. The aim of this study is to a) investigate the role of general *N*-linked glycosylation in a major food poisoning bacterium; *Campylobacter jejuni* b) functionally analyse other *N*-linked glycosylation systems in deep-sea vent bacteria c) Interrogate the role of other forms of *N*-linked glycosylation systems in bacteria. This study evaluates the differential protein expression in glycosylation deficient *C. jejuni* compared to its wildtype counterpart. Thus, enriching our understanding of the role of general *N*-linked glycans and the pleotropic effects caused by knocking out this post-translational modification. Isobaric labelling mass spectrometry results indicate that protein quality control machineries are more abundant in glycosylation deficient *C. jejuni.* Also, the major multidrug efflux pump; CmeABC and nitrate reduction assembly; NapAB were shown to be impaired in *N*-linked glycosylation null *C. jejuni*. Furthermore, examination of the role of *N*-linked glycans in stabilising CmeABC indicated that *N*-linked glycans modulate protein folding, reduce protein unfolding rate and enhance protein-protein interaction. Computational and functional analysis

confirmed the presence and the activity of the oligosaccharyltransferase PglB, of *N*-linked glycosylation system from deep-sea vent bacteria. Investigating other forms of *N*-linked glycosylation in the pig pathogen bacterium; *Actinobacillus pleuropneumoniae,* showed its importance in cell adhesion and pathogenesis. This study provides deeper insights into the roles of *N*-glycans at the molecular level and enriches our understanding of microbial glycoproteome.

List of abbreviations

AIDA	autotransporter adhesin involved in diffuse adherence
App	*Actinobacillus pleuropneumoniae*
BN-PAGE	Blue-Native Polyacrylamide gel Electrophoresis
BCC	*Burkholderia cepacia* complex
CD	Circular dichroism
Cfu	Colony forming unit
Cme	Campylobacter multidrug efflux
CPS	Capsular polysaccharides
diBacNAc	2,4-diacetamido-2,4,6-trideoxy-α-D- glucose
EDC	(1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride)
EtBr	Ethidium bromide
Gal	Galactose
Glc	Glucose
Hex	Hexose
HexNAc	N-acetylhexose
HMW1	High molecular weight adhesin
iGT	Initiating glycosyltransferase
IMAC	Immobilised metal affinity Chromatography
iTRAC	Isobaric tags for relative and absolute quantitation
LC	Liquid chromatography
MDR	Multidrug resistance
MS	Mass spectrometry
NGT	N-glycosyltransferase
NMR	Nuclear magnetic resonance
NTHi	Non-typable *Hemophilus influenzae*
OST	Oligosaccharyltransferase
PCR	Polymerase chain reaction
Pgl	Protein glycosylation locus
RND	Resistance-nodulation-division
Rpm	Revolution per minute
SILAC	Stable isotope labelling by aminoacids in cell cultre
SRRPs	Serine-rich repeat proteins
TMT	Tandem mass tags
TLR	Tolll like receptor
UDP	Uridine diphosphate
Und-pp	Undecaprenolpyrophosphate

CONTENTS

Chapter 1

INTRODUCTION

1.1 What is protein glycosylation?

Glycosylation is the enzymatically catalysed covalent attachment of glycan moieties to certain amino acids in the polypeptide chain (Jayaprakash and Surolia, 2017). It is considered the most prevalent protein post-translational modifications across all kingdoms of life. Owing to its heterogeneity and stereochemical complexity, glycosylation is known to expand the repertoire and functionality of the proteome. In fact, more than half of the eukaryotic proteins are expected to be glycosylated at the nitrogen of the asparagine side chain; *N*-linked glycosylation, oxygen of threonine or serine; *O*-linked glycosylation or both (Apweiler, Hermjakob and Sharon, 1999).

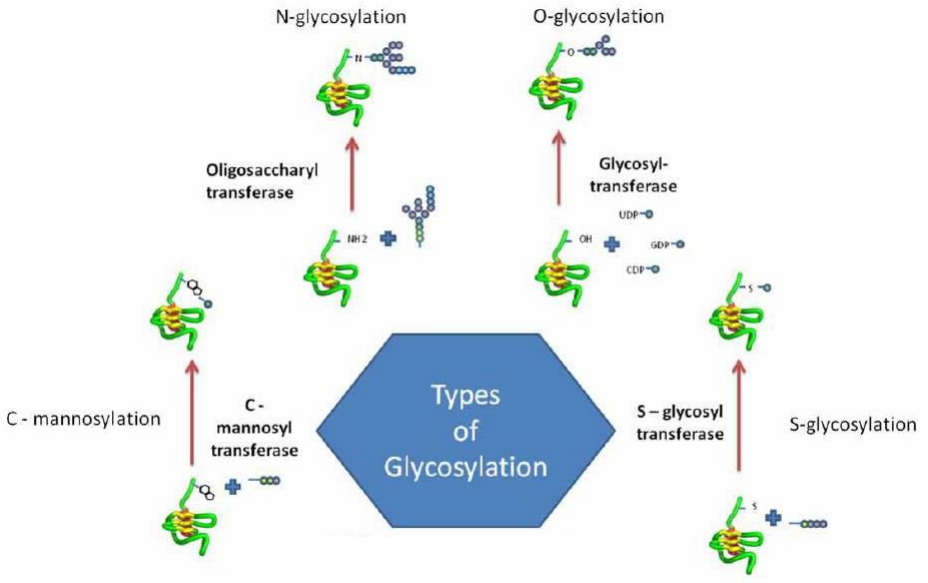


Fig 1 Schematic diagram explain the different types of glycosylation. The diagram also illustrates glycan(s) attachment sites and oligosaccharyltransferase (OST) or glycosyltransferase (GT) involved in this protein post translational modification. The figure is modified and adapted from (Jayaprakash and Surolia, 2017)

Other less common forms of glycosylation can be found in nature whereby glycans are covalently bound to tryptophan or tyrosine; *C*-glycosylation, cysteine or methionine; *S*-linked glycosylation and C2 position of tryptophan through the formation of C-C linkage; *C*-mannosylation **Fig 1** (Jayaprakash and Surolia, 2017).

1.2 A brief history of glycosylation

The study of how and why glycans are covalently attached to certain amino acids in polypeptide chain is the major pillar in the field of glycobiology. Etymologically, glycobiology is derived from the Greek “glyco” which means sugar and “bios” which means alive. Historically, early work on mucins secreted from animals and humans mark the emergence of glycobiology in mid 18thcentury. However, it was not until 1865 that the Russian physician E. Eichwald provided the first evidence of a chemical attachment of glycans to proteins by describing mucins as “a conjugate single compound consisting of a moiety with all genuine properties of a protein and a moiety released under certain conditions as a sugar” (Cabezas, 1994). The first half of the 20th century witnessed a paradigm shift in the understanding of glycan structures. Work pioneered by Emil Fischer, Norman Haworth and Phoebus Levene began to reveal the complex nature of carbohydrates (Varki and Sharon, 2016). Since then, major discoveries such as reporting the carbohydrates components in egg albumin (Neuberger, 1938), the discovery of carbohydrate determinants of ABO blood group (Watkins, 1966), investigating sialic acid as a receptor for influenza virus (Gottschalk, 1958), isolation and the characterisation of bacterial endotoxins structure (Simmons, Luderitz and Westphal, 1965)

highlighted the pivotal role played by carbohydrates in affecting our health and promotion of disease. With the advent of analytical chromatography and separation techniques, mass spectrometry (MS) and nuclear magnetic resonance (NMR), along with DNA cloning and sequencing, the unravelling of glycosylation pathways and systems have helped to give a detailed account on how and why glycosylation occurs.

1.3 Universal pathways in protein glycosylation

Despite the difference in cellular compartments and cell morphologies, recent advances demonstrated that glycosylation pathways share homologous processes across all domains of life. In *N*-linked glycosylation, the enzymatic *en bloc* transfer of glycans to be covalently linked to asparagine residue in the glycosylation sequon N-X-S/T in eukaryotes and archaea and D/E-X-N-Y-S/T in bacteria (where X and Y are any amino acids except proline) can be elaborated in three steps (Schwarz and Aebi, 2011):

1- Generation of lipid linked oligosaccharide (LLO); whereby glycans are assembled on lipid linked anchor by the means of initiating glycosyltransferase (iGTs). Then glycosyltransferases catalyse the transfer of glycans from nucleotide activated sugar to nascent LLO. Chemical analysis of LLO showed that the lipid anchor consists of isoprene unit polymer.

2- Reorientation “Flipping” of LLO; whereby LLO is flipped to the lumen of endoplasmic reticulum (ER) membrane in eukaryotes – to be transferred or further extended- or periplasmic space in bacteria.

3- The catalysis of the *en bloc* transfer of the oligosaccharides, by the means of oligosaccharyltransferase (OST), to the acceptor protein at the

canonical glycosylation sequon in eukaryotes or non-canonical glycosylation sequon in bacteria. **Fig 2**

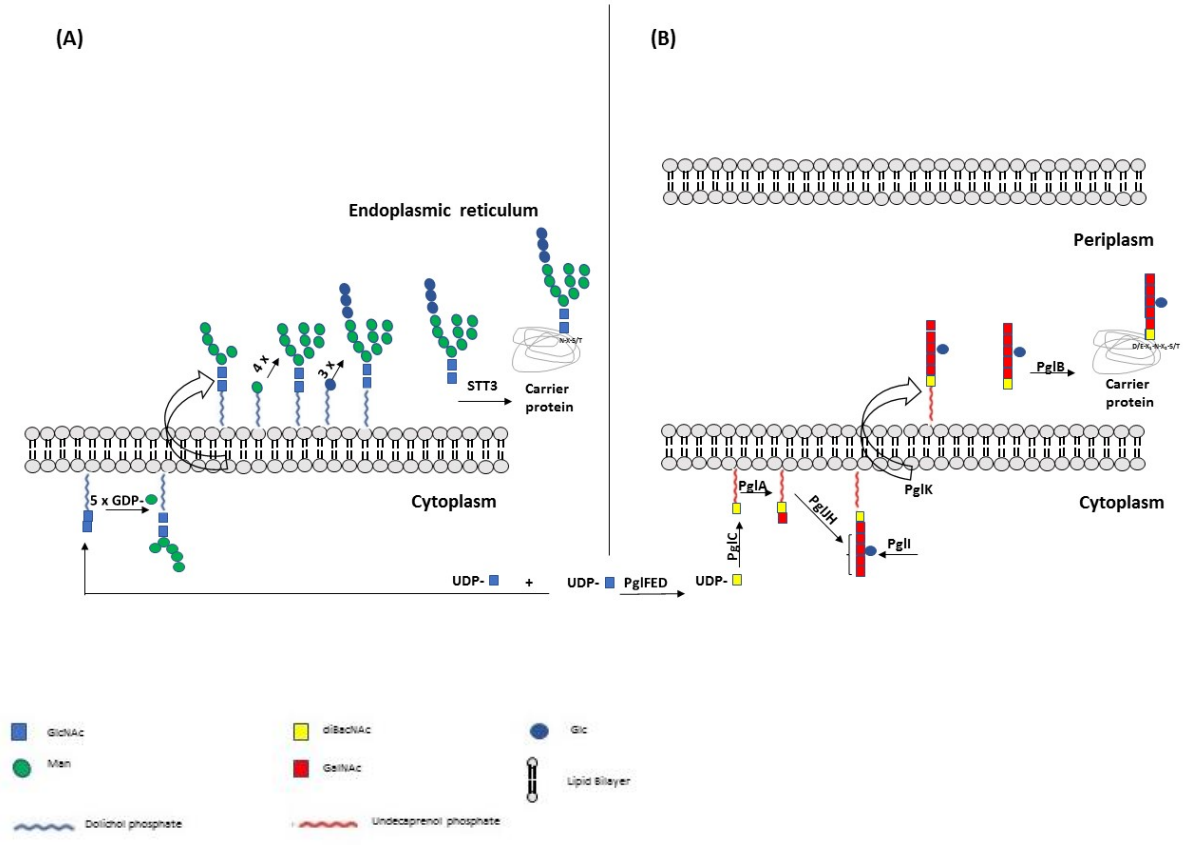


Fig 2 Key differences in *N*-glycoyslation between eukaryotes and prokaryotes.

In *O*-linked glycosylation, apart from bacterial *en bloc* *O*-linked glycosylation, the transfer of glycans to the hydroxyl group of serine or threonine has been demonstrated to occur in the Golgi apparatus in eukaryotes or cytoplasm in prokaryotes by the means of GTs. The sequential nature of the process does not involve the generation of LLO but rather the transfer of glycans directly from the nucleotide activated sugar to the polypeptide chain (Nothaft and Szymanski, 2010).

1.3.1 Diversity in glycans functions

The exploration of the biological roles of glycans shed the light on the pivotal role played by glycans in expanding the proteome repertoire. The complex and diverse nature of glycans exaggerate protein(s) biophysical characteristics and biological properties. Structural-function relationship of glycans to proteins liberated the science community from viewing glycans as just protein decoration. Various functions and roles can be attributed to glycans; however, it is hard to ascertain a general function exhibited by glycans in glycoprotein. Owing to the presence of multiple glycosylation sites, glycans can play different roles within the same polypeptide backbone. Contrary to Proteins and DNA, glycans are products of complex biosynthetic biochemical reactions that cannot be predicted directly from the DNA sequence (Lauc, Krištić and Zoldoš, 2014). Unlike in prokaryotes, the elucidation of the role of eukaryotic glycans is well studied. To encompass the biological role of glycans, state of the art techniques in genomics are required to unravel pathways as well as different proteomics strategies must be employed. Proteomics research has been proven fundamental to characterize and quantify glycoproteins. Advances in mass spectrometry and nuclear magnetic resonance revolutionised the study of the inherently complex nature of glycans in a certain host or organism. Thus, the function of glycans could be categorised into; structural or modular functions (intrinsic effects) and specific (extrinsic). Whilst the former studies the nature of glycans influence protein folding, stability and dynamics and the latter is more concerned with the interaction of the glycans with self/other partner proteins (Varki and Sharon, 2016).

1.4 Intrinsic effects of glycans on protein folding and stabilisation

1.4.1 Effect of glycosylation on protein folding

Although more than half of the proteins are expected to be glycosylated in eukaryotes, only less than 5% of the solved X-ray crystallography structure in protein data bank (PDB) carry either *N*- or *O*-linked glycosylation (Apweiler, Hermjakob and Sharon, 1999; Lee, Qi and Im, 2015). Glycans are notoriously known for being heterogenous, flexible and reduce favourable crystals contact. In addition, the expression of proteins in glycosylation null *E. coli* strains led to this reduction in the ratio of glycosylated/non-glycosylated protein structures found in the PDB. This led to limiting the glycan-protein structure relationship driven research.

Computational studies evaluating the role of glycosylation in influencing protein folding helped in generating hypotheses about a general role of glycans (Xin and Radivojac, 2012; Lee, Qi and Im, 2015). In these studies, available PDB entries were analysed systematically to investigate thestructural changes occurring as a result of protein glycosylation. Xin *et al* suggested a shift in the conformation of the lowest valley of protein landscape; protein folding funnel. Glycosylation seemed to help to assure minimal low energy requirement, improving the energy cost and the spontaneity of the protein folding process. In another study, glycans were found to stabilise proteins by decreasing protein dynamics rather than changing protein conformation (Lee, Qi and Im, 2015). This decrease in protein dynamics was explained by the bulky nature of *N*-linked glycans that act as a “molecular glue” holding amino acids around the glycosylation site

together.

Experimentally, the study of the folding energetics of mono-*N*-glycosylated adhesion domain of the human immune cell receptor cluster of differentiation 2 (hCD2ad), have demonstrated the effect of the glycans on the protein folding landscape. By employing circular dichroism spectroscopy (CD), hCD2ad variants (non-glycosylated, glycosylated with hybrid, complex or oligomannose) were compared. The *N*-linked glycans triose core ($ManGlcNAc_2$) is sufficient to exert a stabilising effect on the protein, ¾ of this stabilisation effect was due to slowing down the unfolding rate by 50 fold and the remaining ¼ was demonstrated to accelerate the folding of the protein by a factor of 4. Additionally, the glycan attachment stabilised the native state by 3.1 kcal/mol, which is a 200 fold increase in the fold equilibrium constant, relative to the non-glycosylated counterpart (Hanson *et al.*, 2009). This study presented experimental proof of the intrinsic role of *N*-linked glycans in both the rate of folding and unfolding.

Effect of glycosylation on protein stability

Biophysical studies demonstrated a reduction in thermostability in deglycosylated proteins when compared to its glycosylated counterparts (van Zuylen, Kamerling and Vliegenthart, 1997; Biswas and Chattopadhyaya, 2016). The thermal stability of the native α subunit of human chorionic gonadotropin (αhCG) and its partially deglycosylated form, was investigated by nuclear magnetic resonance (NMR) upon enzymatic deglycosylation. The deglycosylated form which still carried a GlcNAc monomer at ^{52}N and ^{78}N, showed a small yet significant reduction in

thermostability compared to the native form. This finding was verified in another experimental study using CD spectroscopy, where thermal melts of *Curcuma longa* rhizome lectin (CLA) variants were comparted. Native and hypermannosylated forms of CLA showed a higher T_m compared with glycosylation mutants and deglycosylated forms (Biswas and Chattopadhyaya, 2016). This study suggests that *N*-linked glycans increased protein solubility in denaturated or partially denaturated forms by decreasing aggregate formation.

The stabilisation of glycoproteins was found to be dependent on glycan structure and residual context of the protein. The adhesion domain of human CD2 carries a single *N*-linked glycosylation site at ^{65}N that is glycosylated with a high mannose glycan content $Man_{5-8}GlcNAc_2$. Trimming down the *N*-linked glycan to a single GlcNAc at this site was found to reduce the stability of human CD2. However, the complete removal of the glycans resulted in protein unfolding and aggregation (Wyss *et al.*, 1995) Glycans decorating this site were found to be affect the adhesion properties of the protein by directly stabilising the protein fold.

The covalent attachment of glycan structures to protein enhances the protein kinetics and thermostability (Jayaprakash and Surolia, 2017). By decreasing the free energy of folding, bulky hydrophilic glycans play roles in conformational stabilisation and reduction of thermal unfolding. The molecular mechanism by which glycans stabilise polypeptide chains was investigated in the proto-oncogene tyrosine protein-kinase SRC homology 3 protein; SH3 domain. It is note worthy that SH3 is a relatively small

protein (56 amino acids) and not found to be glycosylated in nature. The folding of 63 engineered SH3 domain variants glycosylated with either 5 glycan rings or 11 branched glycan rings were tested using computational tools (Shental- Bechor and Levy, 2008). Six glycosylation sites were chosen that are located on loops and/or coils, sites were also selected in location predicted to prevent any protein instability upon glycosylation. The results indicated that highly glycosylated variants were more thermo and kinetically stable. Mechanistically, glycans were indicated to play a role in controlling protein unfolding by modulating entropy, shifting the thermodynamic and kinetic properties of protein folding (Shental-Bechor and Levy, 2008).

1.5 Extrinsic effect of glycans on protein-protein interactions

Glycans play an important role in deciding the fate of proteins by recruiting carbohydrate binding proteins and/or enzyme to direct protein folding and trafficking. Due to the important role played by *N*-linked glycans in protein folding, impaired glycosylation triggers proteins degradation (Hebert *et al.*, 2014). The striking finding that the core *N*-linked glycans are conserved in all eukaryotes, across millions of evolution years, indicates a critical role for these glycans in protein folding and sorting. Functional studies revealed the role of *N*-linked glycans in favouring protein folding versus protein degradation. Immediately after the *en bloc* transfer of *N*-linked glycans, glucosidases; α-glucosidase I and glucosidase II are recruited to sequentially remove α-1,2 Glc and α-1,3 Glc, respectively, generating a monoglucosylated polymannose incompletely folded

polypeptide chain. ER lectin chaperones, calnexin (CNX) and calreticulin (CRT) then recognise the monoglucosylated polymannose species to facilitate its folding and prevent it from exiting the ER. Terminal glucose are then removed (deglucosylated) by glucosidase II. This step acts as a signal for proteins to be released form CNX/CRT complex. Proteins are then transferred to the Golgi by coat protein II (COPII) mediated vesicles. Once they reached the Golgi, glycosyltransferases sequentially elaborate the core *N*-linked glycans. In case of incomplete folding, a UDP-Glc:glycoprotein glycotransferase will add a glucose residue to the *N*-linked glycan. CNX/CRT will then bind to reglucosylated proteins to assist its folding (Caramelo and Parodi, 2008; Freeze, Esko and Parodi, 2009)

Alteration in glycosylation profile is often associated with abnormal cellular function. MS and NMR has allowed the comparative studies of glycosylation profile between healthy and patients thus unravelling the role of *N*-linked glycans in diseases establishment and progression (Jayaprakash and Surolia, 2017). Glycomics studies revealed a missing link between aberrant glycosylation and biological implications such as progression of carcinoma and prion protein pathogenesis. In gastric carcinoma, unusual glycosylation of epithelial cadherin (E-cadherin), a calcium dependant cell-cell adhesion molecule, with β1,6 GlcNAc branched *N*-linked glycans by GnT-V glycosyltransfase was shown to abolish its function (Carvalho *et al.*, 2016). This aberrant glycosylation was found to occur in the most distal domain from the membrane, that might be involved in cell-cell interaction, leading to physiological complications.

Integrins are a main class of surface proteins that are attached to extracellular matrix (ECM) that play a role in the cell interaction with the surrounding environment (Rippa *et al.*, 2013). The protein heterodimers consist of α and β subunit, their interaction is important in conferring a wide ligand specificity, however, glycan moieties were proposed to be leading in the interaction between the protein and cellular milieu. Deglycosylation of purified integrin resulted in abolishment of its interaction with other proteins such as fibronectin. Since α5β1 integrins have more than 26 potential glycosylation sites, assigning the role of each site might be challenging. However, studies have shown that attachment of *N*-linked glycans to α5s5 subunit is the needed for its expression on the cell surface (Isaji *et al.*, 2006). Surprisingly, mutating glycosylation site at β1 subunit led to a decrease of its cellular expression and reduction in heterodimer formation, thus affecting its biological role in cell spreading (Isaji *et al.*, 2009).

The prion protein (PrP) is an amyloid forming protein and is usually associated with spongiform encephalopathies. Disease is established when the soluble normal form of PrP (PrP^{C}) undergoes conformational alteration to insoluble proteases resistant form (PrP^{Sc}) (Kupfer, Hinrichs and Groschup, 2009). PrP^{C} is variably glycosylated at two different sites. Studies demonstrated that aberrant glycosylation of PrP^{C} promoted its conversion to PrP^{Sc} leading to change of conformation of the former. Notably, heterologous expression of unglycosylated form of PrP^{C} in yeast demonstrated similar biochemical characteristics as PrP^{Sc} (Ma and

Lindquist, 1999). Most importantly, glycosylation was shown to play a role in hindering the conversion of PrP protease sensitive to pathogenic PrP protease resistant (Rudd *et al.*, 2002). The work to unravel the role of glycosylation in stabilising PrP is still on going. The wealth of knowledge about the role of *N*-linked glycans in cell development, congenital diseases, cancer progression and infection biology, is vastly expanding.

1.6 Emergence of bacterial glycosylation: Of biology and DNA sequencing technology

Despite the burst of knowledge after the molecular biology revolution, glycobiology is fundamentally lagging behind genomic and proteomic research, possibly owing to the complexity and variability in structure. There is a gap of 48 years from when the first report appeared describing the chemical composition of an eukaryotic glycoprotein and the first discovery of a bacterial glycoprotein (Okuda and Weinbaum, 1968). Bacterial glycans are not synthesized in a template-based manner, and linkages and sequences vary enormously. The availability of whole genome sequencing, as well as structural and chemical analysis of macromolecules, paved the way for glycobiology research to progress. Developments in DNA sequencing also allowed researchers to gain an awareness into the widespread presence and diversity of glycan biosynthetic genes. The revolution in merging statistical modules with biophysical and biochemical analyses, provided further evidence for significant structural heterogeneity and function. Reports demonstrating that bacteria possess glycosylation machinery have increased exponentially over the last decade. The first general bacterial glycosylation system was identified in

1999 (Szymanski *et al.*, 1999; and charcterised in 2002 Linton *et al.*, 2002). Since then, more than 30 glycosylation pathways have been reported and functionally characterized (Schäffer and Messner, 2017). The rise in bacterial glycosylation reports follows the introduction of second and third generation DNA sequencing technologies that the science community witnesses during the last two decades. The accessibility of DNA sequencing platforms coupled with bioinformatics fuelled bacterial glycobiology research, allowing researchers to interrogate various bacterial genomes, from deep-sea vent organisms such as *Nitratiruptor tergarcus* (Mills *et al.*, 2016) to the human microbiota, *Bacteriodes fragilis* (Fletcher *et al.*, 2009).

1.7 Bacterial glycosylation

The increasing body of evidence indicating bacterial possession of glycosylation systems underscores the importance of this post-translational modification. Bacterial protein glycosylation was previously thought to be limited to surface appendages such as S-layer, pilin or flagellar proteins (Stimson *et al.*, 1995; Arnold *et al.*, 1998; Schäffer and Messner, 2017). Sequencing the bacterial genome of the epsilon proteobacterium *Campylobacter jejuni* NCTC11168 revealed a gene, in the middle of a locus, encoding a protein orthologous to the STT3 subunit of the eukaryotic OST. Further mutational and biochemical studies identified a "*p*rotein *g*lycosylation *l*ocus" (Pgl) and PglB was observed to be the key central enzyme in the covalent attachment of general *N*-linked glycans to the D/E- X-N-Y-S/T (where X and Y are any amino acid other than

proline) in the acceptor protein in *C. jejuni* (Linton *et al.*, 2002; Nothaft and Szymanski, 2010). This discovery revolutionised the bacterial glycobiology field on two parallel fronts, firstly it opened the question of how and why general glycosylation systems evolved in bacteria. Secondly, it provided a novel biological conjugation technology which is being exploited to assemble either glycoconjugate vaccines or humanised glycoproteins in *E. coli* (Wacker *et al.*, 2002; Feldman *et al.*, 2005; Terra *et al.*, 2012; Valderrama- Rincon *et al.*, 2012; Cuccui and Wren, 2015). The latter has developed rapidly, boosting the biotechnological industry with start-up companies harnessing this novel biotechnological platform (to name a few; Glycovaxyn, now Limmatech, Glycobia, Malcisbo); several of these bioconjugate vaccines are being tested in clinical trials (https://clinicaltrials.gov/ct2/show/NCT02388009).

Apart from the *C. jejuni* general glycosylation system, most basic biology studies did not report beyond the characterization of general glycosylation system and/or identification of glycoproteins. At the time, genetic disruption of glycosylation systems failed to offer a detailed account on cellular alterations. (Linton *et al.*, 2005; Fletcher *et al.*, 2009; Scott *et al.*, 2010; Iwashkiw *et al.*, 2012). However, the wealth of knowledge of the types of bacterial glycosylation systems is expanding constantly. The diversity of bacterial glycosylation system offers an archaic view into the evolution of post translational modification in higher forms of life. Bacterial glycosylation can be divided into two main categories: periplasmic *en bloc* (OST based) transfer of glycans or cytoplasmic sequential glycosylation (GT based). *N*-

and *O*- linked glycosylation systems have now been discovered and characterised that fit into both categories (Nothaft and Szymanski, 2010; Cuccui and Wren, 2015).

1.8 Periplasmic (OST mediated) glycosylation

Nature offers diverse arrays of oligosaccharyltransferases that catalyse the covalent attachment of glycans to the nitrogen in asparagine; *N*-glycosylation or oxygen in serine or threonine *O*-glycosylation. Whilst the former is restricted only to general glycosylation systems found in epsilon proteobacteria such as *C. jejuni* and *Helicobacter pullorum*, the latter could be found as general glycosylation systems such as *Burkholderia cepacia* complex, *Francisella spp* and *Neisseria spp* or specific systems decorating surface appendages such as flagella in *Aeromonas hydrophilia,* Pilin in *Pseudomonas aeruginosa* or S-layer as in *Tannerella forsythia* (Linton *et al.*, 2002; Horzempa *et al.*, 2006; A. Vik *et al.*, 2009; Jervis *et al.*, 2010; Posch *et al.*, 2011; Champasa *et al.*, 2013; Lithgow *et al.*, 2014)

1.8.1 *N*-glycosylation system

Among prokaryotes, *C. jejuni* still has one of the most well elaborated *en* bloc *N*-glycosylation system (Nothaft and Szymanski, 2013). More than 50 proteins are glycosylated with *C. jejuni* heptasaccharide**;** GalNAc-α1,4-GalNAc-[Glcβ1,3-]GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-diBacNAc-β1; where GalNAc is *N*-acetylgalactosamine; Glc is glucose; diBacNAc is 2,4-diacetamido-2,4,6-trideoxyglucopyranose, at the D/E-X-N-Y-S/T site in acceptor proteins whereby X and Y are any amino acid other than proline (Scott *et al.*, 2010). For *N*-glycosylation to happen, two major steps should

take place in different compartments; LLO biosynthesis in the cytoplasm and the transfer of glycans by the central OST in the periplasm.

LLO biosynthesis starts by the generation of diBacNAc; in this primary step, PglF, a NADH dependent 4,6 dehydratase, dehydrates UDP-GlcNAc to UDP-2-acetamido-2,6-dideoxy-α-d-xylo-hexos-4-ulose (Schoenhofen *et al.*, 2006; Riegert *et al.*, 2017). Then PglE, a UDP-4-keto-6-deoxy-GlcNAc specific aminotransferase that uses pyridoxal phosphate as cofactor and glutamate as an amino donor, transfers amino group to the keto sugar to form UDP-2-acetamido-4-amin-2,4,6-trideoxy-α-D-glucose which is then gets acetylated at the C4 position by PglD, using acetyl CoA as an acetyl donor, to form UDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucose; UDP-diBacNAc (Olivier *et al.*, 2006; Schoenhofen *et al.*, 2006) **Fig 2.**

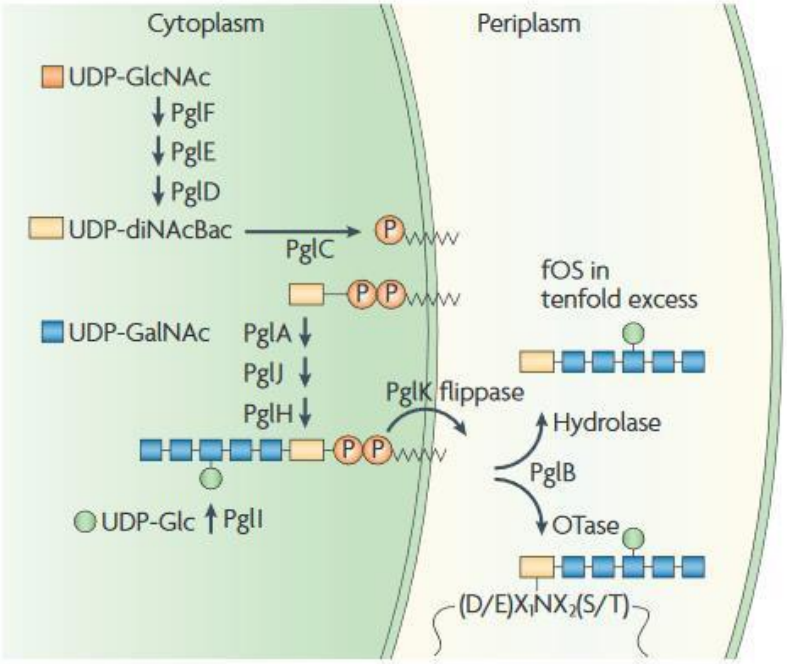


Fig 3 Schematic diagram illustrates the generation, translocation and *en bloc* transfer of *N*-linked glycans in *C. jejuni.* Figure adapted from (Nothaft and Szymanski, 2010)

The transfer of UDP-diBacNAc to Und-P is catalysed by PglC a polyprenol phosphate phosphoglycosyl transferase (PGT). PglA then transfers the first GalNAc to diBacNAc-PP-Und to form a GalNAc-α-1,3-diBacNAc-PP-Und. PglJ then transfers another GalNAc residue onto the disaccharide to form GalNAc-α-1,4-GalNAc-α-1,3-diBacNAc-PP-Und which is then elaborated by the sequential addition of three α-1,4-GalNAc molecules by the help of PglH (Weerapana *et al.*, 2005; Troutman and Imperiali, 2009).

The glycan biosynthesis step terminates by adding a Glc residue to GalNAc-α-1,4 GalNAc-α-1,4-GalNAc-α-1,4GalNAc-α-1,4-GalNAc-α-1,3-diBacNAc-PP-Und hexasaccharide by PglI (Troutman and Imperiali, 2009). After the generation of the LLO, the ABC transporter PglK, translocates “flips” the LLO to the periplasm where it serves as a glycan donor to be transferred to the acceptor protein by PglB.

1.8.2 Biological role of *N*-glycosylation

Despite its relatively small genome, 1,641,481 base pairs, *C. jejuni* dedicates a considerable part of its genome for production of diverse forms of glycans such as; LOS, CPS, *N*-linked and *O*-linked glycans. Whilst the role of LOS and CPS are well studied in *C. jejuni* and other model organisms, the function of *N*-linked glycans is still under studied. Mutagenesis studies on the *N*-glycosylation pathway resulted in pleiotropic effects (Nothaft and Szymanski, 2013). One logical explanation for these effects is that any aberration of the native post translational modification of the 50+ proteins

decorated by *N*-linked glycans will result in a global proteins alteration in *C. jejuni* leading to impairment of protein functions that would be translated into phenotypic changes. Indeed, disrupting *N*-glycosylation leads to a decrease of colonisation of 2-week-old Light Sussex chickens when compared to other *C. jejuni* variants (Jones *et al.*, 2004). Inactivation of PglH also showed a reduction in colonisation in 1-day old chicks as well (Karlyshev, 2004). It is noteworthy that in *C. jejuni pglH*::*aphA,* PglB is still functional and glycosylates proteins with a truncated *N*-linked heptasaccharide; GalNAc-α-1,4-GalNAc-α-1,3-diBacNAc trisaccharide. A 100 to 1000-fold reduction in colonisation of 12-36 hours old White leghorn chicks in was also observed in *pglE, pglF, pglH* inactivated *C. jejuni* variants (Hendrixson and DiRita, 2004). Investigating the effect of abolishing *N*-glycosylation showed a reduction in adherence and invasion of INT407 cells and mouse colonisation, interestingly, no growth differences were noticed between the wild type and glycosylation deficient strains (Szymanski, Burr and Guerry, 2002).

1.8.3 Biological and biophysical role of *N*-linked glycans

The global effects resulted from disrupting *N*-glycosylation pathway pointed at a certain biological role(s) played by *N*-linked glycan in *C. jejuni.* Attempts to study the modulation of the post translational modification to protein function were controversial (Larsen, Szymanski and Guerry, 2004; Davis, Kakuda and DiRita, 2009; Scott *et al.*, 2009). Cj0143c (ZnuA), a periplasmic metallochaperone of the zinc transport system ZnuABC, that is reported to be important in chicken colonisation assay (Davis, Kakuda and DiRita, 2009). Cj0143c is glycosylated at one site in *C. jejuni*, mutational

studies of amino acid alteration of this glycosylation site showed no difference between the protein function when compared to the wildtype (Davis, Kakuda and DiRita, 2009). A similar finding was reported in the study of *C. jejuni* surface glycoprotein JlpA. The outer membrane protein was found to be glycosylated at two glycosylation sites in all of *C. jejuni* strains except *C. jejuni* 11168. Two dimensional SDS-PAGE of the outer membrane of *C. jejuni* strains showed a minor presence of non-glycosylated JlpA when compared to the other two glycosylated forms, however conventional SDS-PAGE of extracted membrane fraction did not show much difference between the glycosylated and non-glycosylated forms of JlpA (Scott *et al.*, 2009). Surprisingly, in the same study glycosylated and non-glycoyslated forms of JlpA were recognised by patient's serum. The authors concluded that JlpA glycosylation is required for localisation and not for antigenicity. The study did not fully elaborate the possibility of glycopeptide presentation by the immune cells which might lead to generation of antibodies against both the glycans and peptides (Avci *et al.*, 2011). The role of *N*-linked glycans in localisation and protein function was further investigated in VirB10 protein in *C. jejuni*. VirB10 is a glycoprotein found as a component of type IV secretion system encoding (TFSS) plasmid; pVir found in *C. jejuni* 81-176. Inactivation of VirB10 showed a significant reduction in bacterial competence; DNA uptake. The same phenotype was observed in glycosylation deficient *C. jejuni* (Larsen, Szymanski and Guerry, 2004). Western blot analysis showed VirB10 was not correctly compartmentally localised in glycosylation deficient *C. jejuni.* Alteration of glycosylation sites showed the importance of only one glycosylation site that is required for

protein activity; ^{97}N. This was observed when DNA uptake efficiency was compared in VirB10 wild type along with its non-glycosylated variants; VirB10 N97A, VirB10 N32A and VirB10 N97A N32A variants. The study shows the role of *N*-linked glycans in protein function.

The immunological role *C. jejuni N*-linked glycans in modulating interaction with host macrophage galactose lectin (MGL) has been proposed. To that end, the *N*-linked glycan biosynthetic pathway was expressed in *E. coli* and showed a localisation of MGL when visualised by immunogold electron microscopy (van Sorge *et al.*, 2009). Interestingly, MGL was found to be interacting with other terminal GalNAc containing glycostructures from *C. jejuni* such as LOS. Whether the interaction of MGL is specific to *N*-linked glycans or to any GalNAc containing glycostructure it is still unknown.

The biophysical role of eukaryotic *N*-linked glycans is well established. Acting as a bulky hydrophilic macromolecule, *N*-linked glycans confer protein folding and thermodynamic stability (Shental-Bechor and Levy, 2008; Hanson *et al.*, 2009; Hebert *et al.*, 2014; Biswas and Chattopadhyaya, 2016). Investigating the influence of the *C. jejuni* heptasaccharide on protein stability could unveil the reason why this post translational modification is critical to the bacterium. PEB3, a class II periplasmic binding protein, is one of the first reported glycoproteins in *C. jejuni* (Linton *et al.*, 2002). The structural elucidation of PEB3 showed that glycans can be attached to the protein without influencing a local structure rearrangement at the glycosylation site (Rangarajan *et al.*, 2007). Thermal denaturation of PEB3 variants using a SYPRO orange binding assay showed that the melting

temperature T_m of glycosylated PEB3 was 4.7 °C higher than the non-glycosylated variant of PEB3 (Rangarajan *et al.*, 2007). However, in this study, comparison between wildtype and non-glycosylated PEB3 was not employed, but rather a glycosylated variant of PEB3 (PEB3 K135E) and its non-glycosylated form. However, the study corroborates the findings that *N*-linked glycans confer protein stability that was found in eukaryotes.

1.8.4 *N*-linked glycosylation prevalence

Glycomics analysis of the *N*-linked glycan structure of the *Campylobacter* genus revealed a surprising finding. Among 16 different structures in the same genus, the first two glycans are absolutely conserved at the reducing end (GalNAc-diBacNAc) and variation occurs when glycans are elaborated at the non-reducing end terminus (Jervis *et al.*, 2012; Nothaft *et al.*, 2012). This finding could explain the degree of diversity in glycosylation site occurrence within the same protein among the *Campylobacter* genus. Understanding the difference of glycan structure and monosaccharide specificity could provide insights into the evolution of the general glycosylation system in the *Campylobacter* genus. The advent of next generation sequencing drove the momentum of microbial glycobiology. The analysis of genome sequences available enriched our understanding of the prevalence of PglB orthologues not only among epsilon proteobacteria, but also in other bacterial families (Mills *et al.*, 2016). Surprisingly, some bacterial genomes were found to code for more than one form of PglB, functional studies suggests that each PglB might have a particular glycan or peptide specificity (Jervis *et al.*, 2018). Heterologous expression of PglB orthologues from deep sea vent bacteria and members of delta proteobacteria, deferribacteres and aquificae

proved its functionality in *E. coli* (Ollis *et al.*, 2015; Mills *et al.*, 2016). Due to the difficult conditions to culture most of these bacteria, it is still uncertain why and how these bacteria glycosylate their proteins.

1.9 Bacterial general *O*-linked glycosylation

Historically, the first glycan component of a glycoprotein reported in bacteria is thought to be an *O*-linked glycan from S-layer proteins in *C. thermosaccharolyticum* and *C. thermohydrosulfuricum* (Sleytr and Thorne, 1976). From this discovery, it was thought that *O*-linked glycosylation evolved in bacteria specifically to modify cellular appendages such as flagella, pilin and S-layers. The ground-breaking finding that bacteria possessed general *O*-glycosylation systems raised the question of a common evolutionary theme of glycosylation systems in prokaryotes and eukaryotes (A. Vik *et al.*, 2009).

1.9.1 General *O*-glycosylation

Similar to *N*-glycosylation, general *O*-glycosylation occurs in the periplasm after building up the *O*-linked glycan on a lipid linked anchor; undecaprenolpyrophosphate in the cytoplasm. Difficulty in characterising an *O*-glycosylation system is exemplified by two main reasons; *O*-glycosylation OST is normally wrongly annotated as WaaL due to the presence of a WzyC signature domain in both proteins, and lack of specific acceptor sequon for *O*-glycosylation to scan for glycoproteins in the bacterial proteome (Schäffer and Messner, 2017).

1.9.2 *O*-glycosylation in *Acinetobacter baumanii* 17978

The cell surface of bacteria is usually decorated with a wide set of different glycans; capsular polysaccharides (CPS), lipopolysaccharides (LPS) and glycans covalently attached to proteins. Cross talk between glycan biosynthetic pathways of these glycans is not uncommon (Cuccui and Wren, 2013). In the nosocomial opportunistic bacterium, *Acinetobacter baumanii* 17978, the initiating glycosyltransferase $PglC_{Ab}$, was found to be involved in both CPS and *O*-linked glycans pathways (Lees-Miller *et al.*, 2013). The involvement of $PglC_{Ab}$ is proposed to be at the early stage of LLO biosynthesis, whereby $PglC_{Ab}$ catalyses the transfer of GalNAc to the lipid anchor Und-p then other four GTs sequentially build LLO in the cytoplasm. LLO is then translocated to the periplasm and faces a crossroad, either to be elongated by Wzy polymerase or transferred to the acceptor protein by the *O*-glycosyation OST PglL (Lees-Miller *et al.*, 2013). $PglL_{Ab}$ decorates 26 different proteins with β-GlcNAc3NAcA4OAc-4-(β-GlcNAc-6-)-α-Gal-6-β-Glc-3-β-GalNAc pentasaccharide in *A. baumanii* 17978. However, glycan structure diversity and microheterogeneity has been reported among *A. baumanii* strains (Scott *et al.*, 2014). Disruption of $PglL_{AB}$ caused a reduction in virulence in amoeba and the *Galleria mellonella* infection models as well as decreased bacterial fitness in BALB/C mice. Interestingly, *O*-glycosylation was found to be directly linked with biofilm formation, an important virulence determinant in *A. baumanii* that is associated with survival, antibiotic resistance and pathogenesis (Iwashkiw *et al.*, 2012).

1.9.3 *O*-glycosylation in *Burkholderia cepacia* complex

The *Burkholderia cepacia* complex (Bcc) is a group of phenotypically similar but genetically distinct bacteria that infect humans, animals and plants. In humans, *B. cenocepacia* is considered the most common causative agent of cystic fibrosis (CF) (Mahenthiralingam, Urban and Goldberg, 2005). Investigating *B. cenocepacia* revealed the presence of a general *O*-glycosylation system, where the *bcl0960* gene codes for the central *O*-OST; $PglL_{Bc}$. Bioinformatic studies revealed that $PglL_{Bc}$ possess a Wzy_C and a DUF3366 domains. Heterologous expression of $PglL_{Bc}$ in *E. coli* showed its relaxed specificity in transferring *C. jejuni* heptasaccharide onto acceptor protein; DsbA1. PgL_{Bc} glycosylates 23 different proteins in *B. cenocepacia* with the trisaccharide Hex-HexNAc-HexNAc whilst other proteins carry an extra 100 Da mass at the non-reducing end of the glycan. Abolishing of *O*-glycosylation by disrupting $PglL_{Bc}$ reduced bacterial virulence and motility (Lithgow *et al.*, 2014). This finding suggests a pleotropic effect caused by altering the glycoprofile of proteins in *B. cenocepacia* that directly or indirectly affected cellular activities and virulence.

1.9.4 *O*-glycosylation in *Francisella* spp

The intracellular pathogen *Francisella tularensis* is causative agent of tularemia. Due to the ease of its transmission whereas infection could be established by inhaling as few as 25 CFU, *F. tularensis* has been designated a category A agent in bioterrorism (Oyston, Sjostedt and Titball, 2004). In a study to explore the glycoproteome of *F. tularensis* different analytical methods were employed such as hydrazide labelling,

lectin blotting and lectin affinity chromatography (Balonova *et al.*, 2010). More than 15 proteins were found to be *O*-glycosylated with a hexasaccharide glycan; HexNAc-X-HexNAc-Hex-Hex-HexNAc, where X is an unknown glycan (Balonova *et al.*, 2010; Thomas *et al.*, 2011). Mutagenesis studies revealed that PglA, a OST with a degree of similarity to PilO from *P. aeruginosa*, that is involved in pilin glycosylation is responsible for both pilin glycosylation and general *O*-glycosylation (Balonova *et al.*, 2012). The reconstitution of PglA mediated glycosylation of PilA in *E. coli* expressing the glycan donor diBacNAc, demonstrated the relaxed specificity of PglA (Egge-Jacobsen *et al.*, 2011). Further analysis of the *O*-glycosylation pathway, revealed a cross talk between O-antigen glycan biosynthesis and *O*-linked glycan assembly (Balonova *et al.*, 2012). Mouse model studies did not show difference in infection level between *pglA* mutant and the wild type, neither any difference was exhibited in 80% serum killing in both strains. Thus, indicating that *O*-glycosylation does not play a role in the virulence of *F. tularensis* (Dankova *et al.*, 2016). Interestingly, mutation in *FTS_1402* gene that codes for a putative flippase and is responsible for translocating diverse glycan structures such as LPS and CPS, showed a decrease in *O*-linked glycan attachment and a degree of attenuation in *F. tularensis.* This result indicated a critical role played by glycoconjugates in this pathogenic bacterium.

1.9.5 *O*-glycosylation in *Neisseria* spp.

The most extensively studied O-glycosylation system among prokaryotes is that found in *Neisseria spp*. Research into this system was initiated more than three decades ago, when *Neisseria meningitidis* was

found to be capable of decorating its pilin with the *O*-linked glycans trisaccharide; Gal-α-1,4-Gal-α-1,4-diBacNAc (Stimson *et al.*, 1995). However, glycan variation was detected between *N. meningitidis* and the closely related species *N. gonorrhoeae*. Investigating the role of *O*-linked glycans demonstrated their capacity to increase the solubility of pilin and ruled out their involvement in piliation (Marceau *et al.*, 1998). These findings directed attention towards a specific target protein for *O*-glycosylation in *Neisseria spp.,* however, a decade later, 11 proteins were found to be glycosylated with the same glycan, demonstrating the presence of a general *O*-glycosylation system (A. Vik *et al.*, 2009). Interestingly, pilin IV is found not to be glycosylated in *N. elongate* subsp *glycolytica* despite its possession of a functioning *O*-glycosylation system. Dissecting the locus encoding *O*-glycosylation system revealed a degree of homology and function between *Neisseria spp O*-linked system and *C. jejuni N*-linked glycosylation (Hartley *et al.*, 2011).

In brief, *O*-glycosylation consists of three major steps; generation of LLO, translocation of the LLO from the cytoplasm across to the periplasm, and *en bloc* transfer of *O*-linked glycans mediated by OST; PglO. PglD and PglC catalyse the dehydration and amination of UDP-Glc using NAD^+ as a cofactor to generate UDP-2-acetamido-4,amino-2,4,6-trideoxy-α-D-glucose which is then get acetylated at the C4 position. This step is followed the transfer of UDP-monosaccharide to the lipid anchor by the bifunctional enzyme PglB (not to be confused with $PglB_{Cj}$) **Fig 3**. Interestingly, half of the clinical isolates of *N. meningitidis* express the *pglB2* allele which encodes the bifunctional enzyme PglB2 that transfers a

glycerol instead of acetyl group to C4 position, and generates UDP-2-acetamido-4,amino-2,4,6-trideoxy-α-D-glucose to create 4-glyceramido-2- acetamido-2,4,6-trideoxy-α-d-hexose (GATDH) (Chamot-Rooke *et al.*, 2007). Other GTs then elaborate on the lipid linked glycan. PglH or its variant PglH2 transfers Glc or GalNAc, respectively. Alternatively, PglA and PlgD successively transfer Gal to the lipid linked glycan.

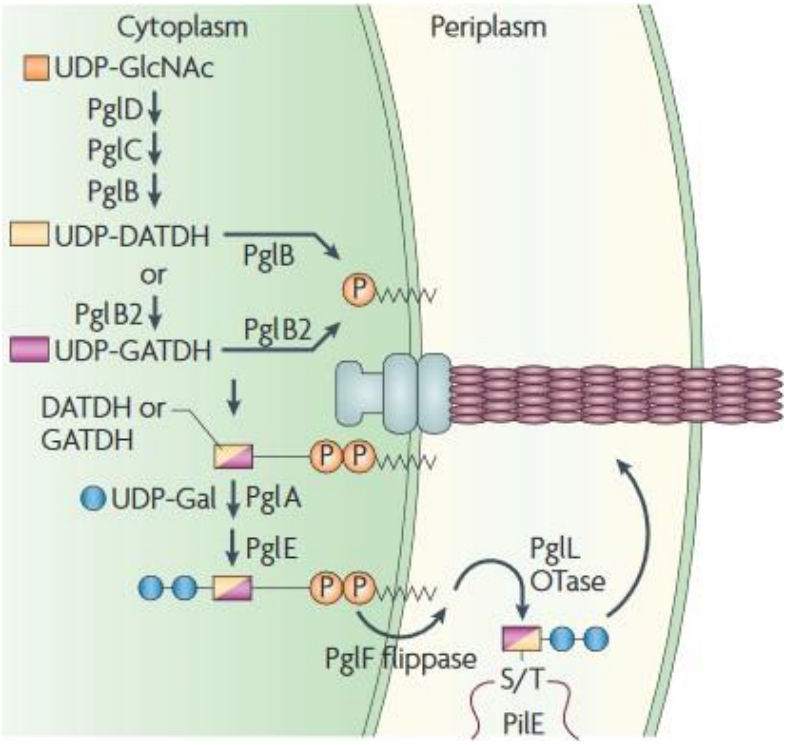


Fig 4 Schematic diagram illustrates the generation, translocation and *en bloc* transfer of *O*-linked glycans in *N. meningitidis.* Figure was adapted from (Nothaft and Szymanski, 2010)

Strains expressing both PglAD and PglH expresses microheterogenous glycoforms which later gets acetylated by Pgll (Borud *et al.*, 2011). After the translocation of LLO across to the periplasm. PglL catalyse the *en bloc* transfer of *O*-linked glycans to S/T in the acceptor protein. The association of *O*-linked glycosylation with the structural context of protein has been

investigated. Notably, low complexity regions (LCR) of the acceptor protein have been found to be targeted by PglL (A. Vik *et al.*, 2009)

Understanding the role of *O*-linked glycans is still poorly studied. However, structural analysis of AniA, a surface exposed nitrite reductase enzyme that is glycosylated at the C-terminus, suggested that *O*-linked glycans might play a role in shielding protein domains from immune system recognition (Ku *et al.*, 2009).

1.10 Protein specific *en bloc O*-glycosylation

The compelling studies that followed the first bacterial glycoprotein discovered in bacteria, suggested that glycosylation is specific to certain surface proteins (Okuda and Weinbaum, 1968; Nothaft and Szymanski, 2010; Schäffer and Messner, 2017). This dogma was later challenged in the beginning of this century after the discovery of general *N*- and *O*-glycosylation systems in bacteria (Linton *et al.*, 2002; Ashild Vik *et al.*, 2009). The discovery of OST mediated protein specific *O*-glycosylation is an important milestone in the understanding protein post translational modification. Glycomic analysis of glycans demonstrated diverse glycan structures, different linkages, microheterogeneity among the same genus and uncommon glycan moieties. (Horzempa *et al.*, 2006; Posch *et al.*, 2011; Bouché *et al.*, 2016; Schäffer and Messner, 2017).

1.10.1 Flagellin *O*-glycosylation in *Aeromonas hydrophilia*

The exciting findings supporting the notion of two potential *O*-glycosylation systems in *A. hydrophilia* provide insights on how and why bacterial glycosylation has emerged. *A. hydrophilia* is a mesophilic water-

borne bacteria that expresses two forms of flagellin which are involved in adherence to biotic and abiotic surfaces as well as biofilm formation (Kirov, Castrisios and Shaw, 2004). Both flagellins are *O*-glycosylated by two different glycans however share one same glycan; a 376 Da glycan that is a derivative of pseudomanic acid (Wilhelms *et al.*, 2012). In the same study, polar flagella were found to be *O*-glycosylated with a more complex glycan structure composed of a heptasaccharide. MS analysis of the heptasaccharide revealed a variable phosphorylation and methylation of glycans and two unknown monosaccharides of 376 Da and 102 Da in sequence of 102 Da-HexNAc-HexNAc-HexNAc-Hex-Hex-376 Da (Wilhelms *et al.*, 2012). Mutagenesis studies in *wecX* that codes for putative lipid linking glycan transferase, abolished glycosylation and flagellin production in *A. hydrophilia* AH-3. Thus suggesting an OST mediated *O*-glycosylation in the polar flagellin (Merino *et al.*, 2014). Interestingly, mutation in *wecP,* which encodes GalNAc transferase to lipid anchor showed no effect on flagellin glycan structure but rather a decrease in the intensity of polar flagellin glycosylation, suggesting no cross talk between O-antigen biosynthesis and *O*-linked heptasaccharide assembly (Merino *et al.*, 2014). Impairment of polar flagellin glycosylation was found to abolish bacterial motility and flagellin production, as well as reduce recognition by Toll-like receptor 5 (TLR5) (Merino, Wilhelms and Tomás, 2014). More work is needed to investigate the conservation of these two glycosylation systems in the *Aeromonas* genus as well as to reveal the OST responsible for the polar flagellin glycosylation.

1.11 Cytoplasmic protein glycosylation (GT mediated)

Breakthroughs in sequencing and analytical chemistries provided evidence of the presence of sequential protein glycosylation in bacteria. Occurring in the cytoplasm, sequential protein glycosylation is a form of post translational modification whereby glycans are transferred directly from nucleotide activated monosaccharides to the polypeptide chain with the means of glycosyltransferases (Szymanski *et al.*, 2003; Nothaft and Szymanski, 2010; Cuccui and Wren, 2013; Schäffer and Messner, 2017). Unlike OST mediated glycosylation, which can be general or specific post translational modification, sequential protein glycosylation has been found to be restricted only to surface proteins including bacterial adhesins. In most cases, glycosyltransferases responsible for protein glycosylation are found in the same operon. Due to limitations in bioinformatic models, most of the protein glycosylation glycosyltransferases are wrongly annotated. Functional and structural studies help to give a better picture on the activity of these GTs. GT mediated protein glycosylation is found in both gram-positive and gram-negative bacteria, pathogenic or symbiotic bacteria. Like OST mediated glycosylation, GT mediated glycosylation can be divided to *N*-glycosylation found in members of *Pasteurellaceae* family such as *Actinobacillus pleuropneumoniae* or non-typable *Haemophilus influenzae* or *O*-glycosylation found in Gram positive bacteria such as *Clostridium difficile* and *Streptococcus spp* (Grass *et al.*, 2010; Boysen *et al.*, 2016; Valiente *et al.*, 2016; J. Cuccui *et al.*, 2017; Avilés-Reyes *et al.*, 2018).

1.11.1 High molecular weight adhesin *N*-glycosylation

The discovery of high molecular weight adhesins; HMW1A and HMW2A, in non-typeable *H. influenzae* (NTHi) expanded the *N*-glycosylation repertoire beyond *en bloc* glycan transfer. Around 80% of NTHi express HMW1 and HMW2 adhesins that play a critical role in establishing infection by promoting bacterial adherence to epithelial respiratory cells (St Geme, Falkow and Barenkamp, 1993; St. Geme, 1994; St Geme, 2002). HMW1 is found in an operon coding for its transporter HMW1B and the *N*-linked glycosyltransferase, HMW1C which is the key enzyme in HMW1A glycosylation. Accounting for approximately 5 % of protein molecular weight, *N*-glycans were found to be functionally involved in tethering HMW1 and HMW2 onto the bacterial cell surface (Barenkamp and St Geme 3rd, 1994). The establishment of *N*-glycosylation of HMW1A or HMW2A is carried out in the cytoplasm of the cell, whereby, HMW1C binds to the adhesin initiating the transfer of hexose from nucleotide activated sugar donor directly to N-X-S/T, whereby X is any amino acid other than proline. Additionally, HMW1C was found to be able to generate hexose-hexose bonds (Grass *et al.*, 2010). Proteomic analysis of HMW1A identified that the 31 sites where occupied with 47 hexose or dihexose (Gross *et al.*, 2008) **Fig 5**. Bioinformatic analysis showed a 65% identity and 85% similarity between HMW1C and ApHMW1C from *A. pleuropneumonia.* Unlike *hmw1ABC* operon, *aphmw1C* was not found to be located in the same operon with its adhesin but rather in the middle of locus consisting of *rimO,* coding for methylthiotransferase and another glycosyltransferase called *agt.*

Transcriptional analysis demonstrated that the three genes function as an operon (J. Cuccui *et al.*, 2017). Complementation studies showed that ApHMW1C can complement the HMW1C function and glycosylate HMW1A thus promoting bacterial adherence to epithelial cells. (Choi *et al.*, 2010).

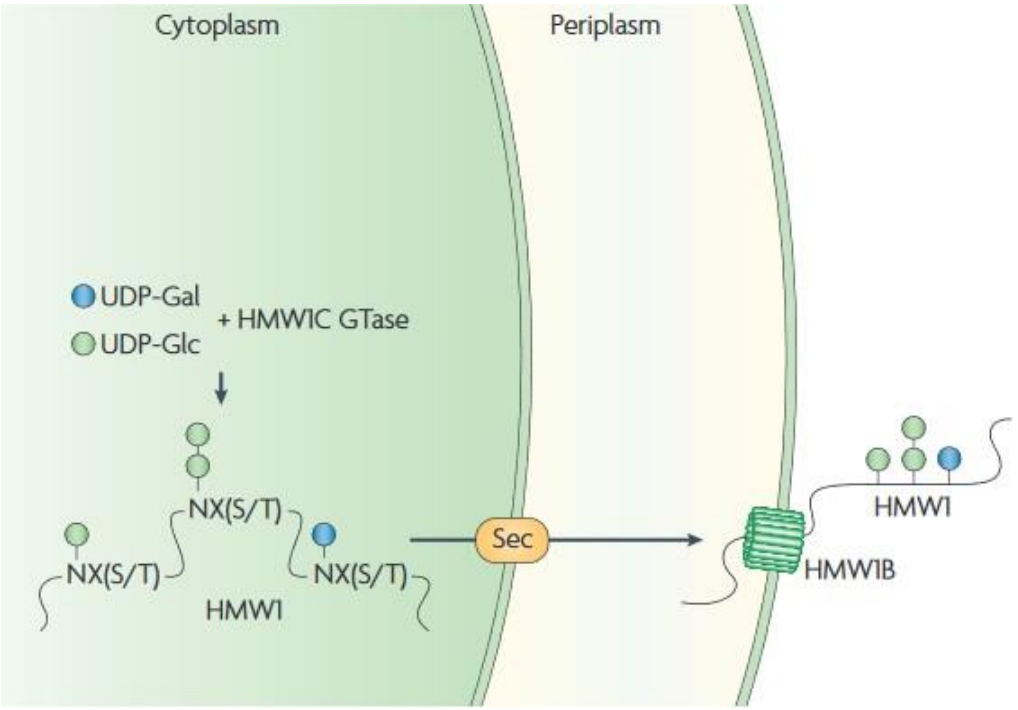


Fig 5 Schematic diagram illustrates cytoplasmic *N*-glycosylation in *H. influenzae.* Figure is adapted from (Nothaft and Szymanski, 2010)

Functional analysis experiments showed that *N*-linked glucose established by the activity of ApHMW1C could be elongated by Agt α6GT polymerising activity (Schwarz *et al.*, 2011; J. Cuccui *et al.*, 2017). ApHWM1C and/or Agt mutants exhibited reduction in adhesion to A549 epithelial cells, thus indicating that, *N*-linked glycans play a direct or indirect role in establishing App infection in host animal (J. Cuccui *et al.*, 2017). It is noteworthy that orthologues of *hmw1C* were found in other

organisms such as *Yersinia enterocolitica, Yersinia pseudotuberculosis,* enterotoxgenic *E. coli* and *Burkhoderia* spp. This type of glycosyltransferase was coined the term *N*-glycosyltransferase (NGT) (Schwarz *et al.*, 2011). The substrate specificity to high molecular weight adhesins and the efficiency of glycosylation are still to be unravelled. However, a possible explanation that both NGT and AGT work in concert thus ensuring protein quality control for such an important antigenic determinant.

1.11.2 Cytoplasmic *O*-glycosylation

Cytoplasmic mediated *O*-glycosylation is a type of glycosylation that is present in both eukaryotes and prokaryotes. Despite the advancement of genome sequencing and the exponential increase of sequenced bacterial genomes, it is still difficult to identify *O*-glycosyltransferases. *O*-glycoproteins are glycosylated with diverse glycan structures (either inter or intraspecies variation) such as α-1,5 arabinofuranose in *P. aeruginosa* PA5196, 7-acetamido-5-(*N*-methyl-glutam-4-yl)-amino-3,5,7,9-tetradeoxy-D-glycero-α-galactononulosonic acid or di-*N*-acetylhexuronic acid such as in *C. botulinum,* pseudomanic or legionaminic acid such as in *C. jejuni.* Thus exhibiting an inherent complexity to *O*-glycosylation systems (Twine *et al.*, 2008; Ewing, Andreishcheva and Guerry, 2009; Nothaft and Szymanski, 2010).

1.11.3 Bacterial autotransporters heptosyltransferase

The first study reporting the presence of a glycoprotein in bacteria was carried out in *E. coli* (Okuda and Weinbaum, 1968). That was even before the discovery of bacterial autotransporter adhesins in 1987 by Thomas

Meyer's group (Pohlner *et al.*, 1987). Since then, the number of autotransporters identified has been increasing exponentially. A subfamily of autotransporters that shares the same virulence attributes and structural profiles were discovered in *E. coli*. Interestingly, members of this family have been demonstrated to be *O*-glycosylated and named self-associating autotransporters (SAAT). They include, adhesins involved in diffuse adherence (AIDA-I) from diffusely adhering *E. coli* (DEAC), TibA from enterotoxigenic *E. coli* strain H10407 and Ag43 from uropathogenic *E. coli* (Lindenthal and Elsinghorst, 1999; Benz and Schmidt, 2001; Sherlock *et al.*, 2006). Genetic and biochemical studies identified the bacterial autotransporter heptosyltransferase (BAHT) to be the key enzyme in *O*-glycosylation of AIDA-I, TibA and Ag43 (Charbonneau *et al.*, 2012). BAHT catalyse the transfer of heptose sugar; glycerol-manno-heptose to the imperfect repetition of a 19 amino acid sequence in the extracellular domain of the SAAT. This domain adapts a right handed β-helix that is extensively modified on serine, but not threonine, by BAHT rendering a spectacular array of glycans on the edges of the domain (Lu *et al.*, 2014). The BAHT family consists of Aah and TibC that glycosylates AIDA-I and TibA respectively. The glycosyltransferase for Ag43 has not been assigned. Complementation studies demonstrated that TibC can substitute the function of Aah in an *aah* mutant (Moormann, Benz and Schmidt, 2002). Bioinformatic analysis of Aah indicates the presence of this type *O*-glycosyltransferase in *Burkholderia spp, Citrobacter rodentium, Cronobacter sakazakii*, *Shigella spp* and other members of alpha, beta and gamma proteobacteria (Charbonneau *et al.*, 2012). Functional analysis in

C. rodentium identified the target adhesin *Citrobacter* autotransporter, required for colonisation (CARC) to be the target for the BAHT enzyme. Mutations in BAHT decreased the bacterial colonisation in mice, indicating a major role player by *O*-heptosyltransferase in establishment of infection (Lu *et al.*, 2014). This type of adhesin/*O*-glycosyltransferase has a great potential in development of conjugate vaccines against bacterial major antigenic determinants.

1.11.4 Glycosylation of Serine-rich repeats proteins (SRRPs)

Serine-rich repeats proteins are family of adhesins that are found to be glycosylated in Gram-positive bacteria. They play mediate microbe-host interaction and biofilm formation. Pneumococcal SRRPs are involved in establishment of serious diseases such as pneumonia, endocarditis and meningitis (Lizcano, Sanchez and Orihuela, 2012). Genes coding for SRRPs, their translocase and their glycosyltransferase are usually found in an operon that varies in size according to the strain. Two GTs were found to be critical in the initiation of this protein specific post translational modification; Gtf1 and Gtf2 (Zhou *et al.*, 2010). This first step is mediated by GtfA and GtfB heterotetrameric complex by recognising the low complexity region in SRRP catalyse the formation of αGlcNAc-Ser/Thr. In some cases, GTs are then recruited to elongate the newly formed glycan-polypepide by sequential build-up of the glycan structure on αGlcNAc side, such observation were found in the formation of Rha-1-3Glc1-(Glc1-3GlcNAc1)-2,6-Glc1-6GlcNAc in *S. parasanguinis* and the tetrasaccharide as in *S. pneumoniae* (Zhou *et al.*, 2010; Chen *et al.*, 2016; Zhu *et al.*, 2016;

Jiangm *et al.*, 2017). The specific role of the *O*-linked glycans is still unknown, however, it has been suggested that *O*-linked glycans are critical for protein stability (Lizcano, Sanchez and Orihuela, 2012). Disruption of Gtf1 led to a reduction in biofilm formation in *S. parasanguinis* (Zhu *et al.*, 2016). Homologous protein glycosylation systems were found in *Lactobacillus planturum* and *Lactobacillus reuteri* TMW1.106, whereby *O*-glycosylation has been involved in biofilm formation, cell aggregation and mouse colonisation (Walter *et al.*, 2008; Lee *et al.*, 2014). Successful reconstruction of the GtfA/GtfB glycosylation system suggest a potential biotechnological application in vaccine development and production of glycoconjugates (Zhu *et al.*, 2016). The prevalence of this protein post translational modification in gram-positive bacteria indicates a certain evolutionary pressure to adapt the critical step of αGlcNAc covalent addition. Future studies will help to understand the reason why certain bacterium prefer a specific type of glycosylation.

1.12 Investigating the biological role of *N*-linked glycans in bacteria

The field of bacterial glycobiology has been established for more than half a century encompassing an unparalleled wealth of bacterial complex glycostructures mainly, LPS and CPS. However, the study of the role of bacterial *N*-linked glycans is still in its infancy compared to eukaryotic glycobiology. Major efforts have been made to harness the potential of bacterial glycosylation systems to develop glycoconjugate vaccines, however, basic understanding regarding the biological role of *N*-linked glycans remains elusive. One of the major hurdles to initiate such

investigations is the ability to characterise a phenotypic change that could be linked directly to the loss of *N*-linked glycans. To solve this conundrum, the role of *N*-linked glycans are investigated in this study on two levels; the proteome level and the individual protein level. Whilst the former is concerned with monitoring the change in protein abundance and function of bacterial glycoproteins, the latter focuses on the role of *N*-linked glycans in conferring protein stability and function. Both approaches complement each other and help to give a clearer picture of the role exhibited by *N*-linked glycans in increasing the proteome repertoire.

1.13 Research aims and objectives

This thesis focuses on the biological role of *N*-linked protein glycosylation in prokaryotes. To achieve this, I followed different approaches to:

i) Interrogate the alteration in the proteome following disruption of *N*-glycosylation. Using *C. jejuni* as a model organism, I explored the changes in the proteome of a glycosylation deficient strain against its wild type counterpart. I assessed glycoprotein activity in *C. jejuni* variants as well as phenotypic changes resulting from *N*-glycosylation knock out. Finally, I evaluated bacterial fitness in colonising chickens compared to the wild type strain.

ii) Study the biophysical properties of *N*-linked glycans and their intrinsic and extrinsic roles. CmeA is one of the commonly used carrier proteins, it is also involved in the major multidrug efflux pump assembly, CmeABC. I investigated the protein folding/unfolding dynamics of glycosylated CmeA against its non-glycosylated counterpart as well as the role of *N*-linked glycans in promoting protein-protein interaction with CmeA binding partner; CmeC.

iii) Explore different substrate specificity of *N*-OST from deep sea vent bacteria. I analysed PglB orthologues using bioinformatic tools to gain more insights on how PglB orthologues are diverse from one another.

iv) Functionally analyse the role of cytoplasmic of *N*-glycosylation. I functionally constructed NGT-AGT in *E. coli*, to be enable functional analysis of cytoplasmic *N*-glycosylation as well as develop a platform for glycoconjugates production on a non-PglB-based biotechnology.

The objective of these studies is to provide insights on the critical role of protein glycosylation as well as develop a platform to investigate other bacterial glycosylation pathways.

1.14 Methodological approaches

1.14.1 Interrogating the alteration in the glycosylation deficient *C. jejuni*

Quantitative proteomics evaluate the difference in abundance of proteins in a high-through put manner. Measuring abundance can be provide a detailed picture of alteration in cellular proteome a result of genetic mutations or environmental assault. A variety of quantitative proteomics approaches have been developed notably isotopic labelling or label free method **Fig 6** (Bakalarski and Kirkpatrick, 2016)**.**

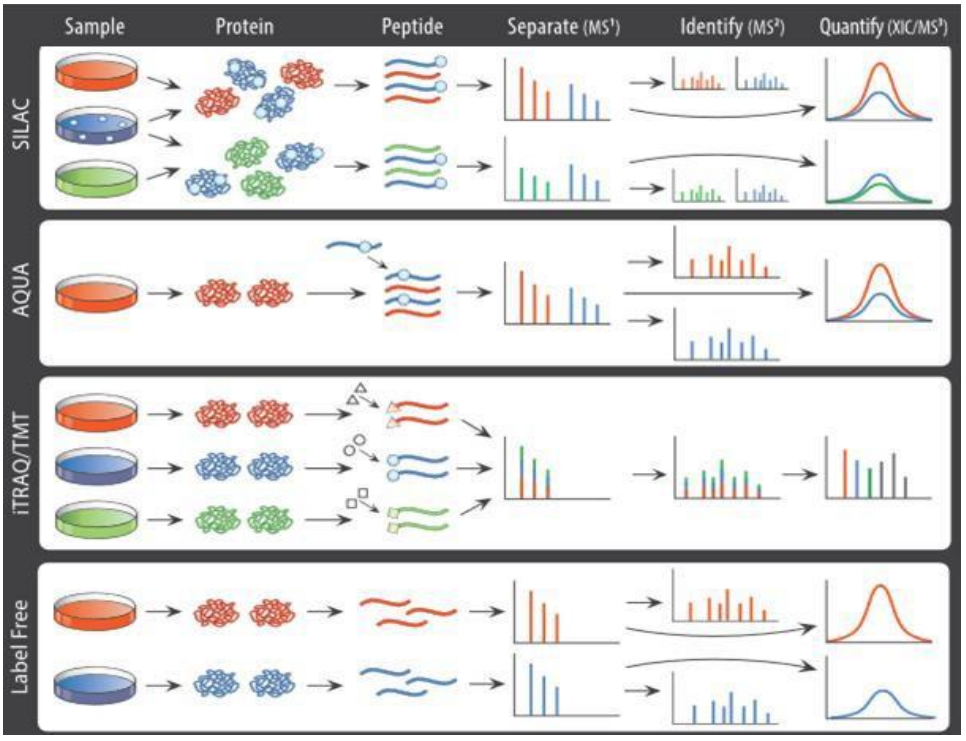


Fig 6 Schematic diagram illustrating quantitative proteomics methods. The diagram highlights differences in labelling and peptide quantification in each method. The diagram is modified and adapted from (Bakalarski and Kirkpatrick, 2016)

Isotopic labelling includes stable isotope labelling by amino acids in cell culture (SILAC), isobaric tag for relative and absolute quantification (iTRAQ), tandem mass tags (TMT) and isotopically labelled synthesised peptide (AQUA). Label free labelling quantitative proteomics is based on correlating protein abundance with either mass spectrometric of signal intensities of peptides, number of MS/MS spectra matched to peptides or proteins (Silva, 2005; Li *et al.*, 2012). Whilst isobaric labelling is based on metabolic labelling; SILAC or chemical labelling TMT and iTRAQ. Each quantitative proteomic approach has it is own advantage and disadvantage. Label free quantification approach allows identification and quantification of proteins without introducing costly reagents that might be both time consuming and laborious. However, quantification of samples are done separately and on individual basis which affects the accuracy and reproducibility of the method (Li *et al.*, 2012).

Isobaric labelling quantification allows samples multiplexing hence enhancing the accuracy, sensitivity and robustness of the quantification as well as eliminates variations according to amounts of samples injected and run-to-run variations (Kito and Ito, 2008). However, this method suffers from the drawbacks of lower proteome coverage and time consuming labelling process (Silva, 2005). Isobaric tags consists of three regions; a reactive group, which reacts with the *N*-terminal amine and ε-amine group of lysine specifically, a balance region, which normalises the mass difference among reporter ions, and a reporter ion, which is associated in measuring the relative abundance of the peptide labelled **Fig 7** (Rauniyar and Yates, 2014)

Peptides labelled with different tags are indistinguishable in full scan (MS^1) however, upon another round of fragmentation (MS/MS) the reporter ions are dissociated from the peptide. The mass of the reporter ion and its intensity determines the relative abundance of the labelled peptide with this particular reporter ion (Li *et al.*, 2012).

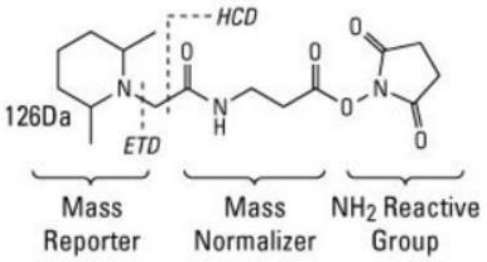
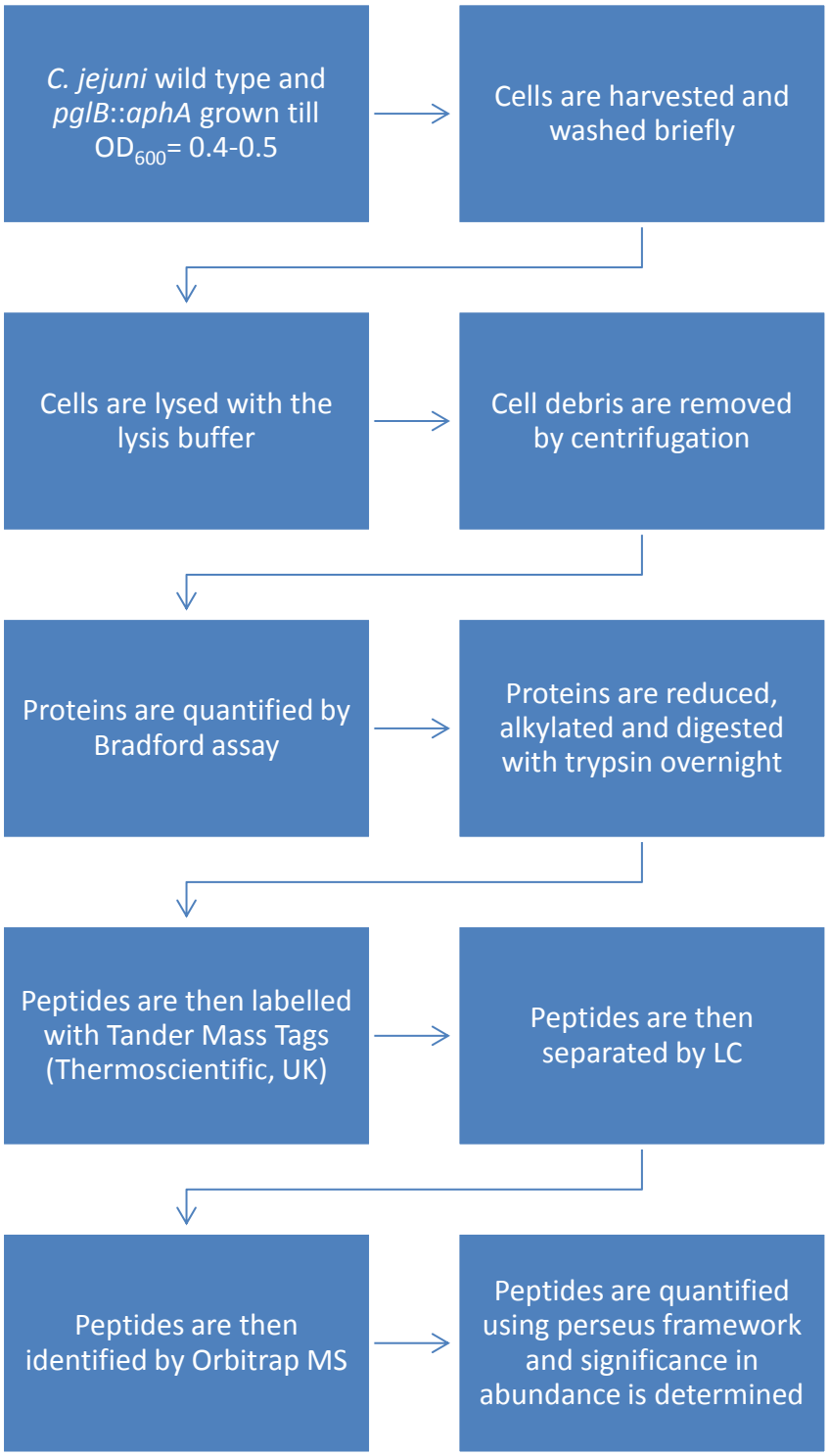


Fig 7 Chemical structure of TMT reagent highlighting different reactive groups and MS/MS fragmentation sites using high collision dissociation (HCD) or electron transfer dissociation (ETD). Image adapted from thermofisher scientific.

In this study, I will use TMT 6plex to determine protein abundance differences in *C. jejuni* strains. This allowed the comparison of three biological replicates from *C. jejuni* wildtype and glycosylation deficient strain grown at the same condition and same time. The simultaneous comparison of protein abundance with the means of TMT 6plex provided confidence in the quantification method. Various bioassays were carried out in parallel to ascertain the proteomic quantification results. The bioassays provided an insight on the protein function and activity either found in high or low abundance in glycosylation deficient strain. Furthermore, chicken colonisation was examined to provide an holistic view on bacterial fitness and host-microbe interaction.

The following scheme simplifies the workflow for quantitative proteomics:



1.14.2 Evaluating intrinsic and extrinsic roles of *N*-linked glycans

Asparagine glycosylation has been linked to promote protein folding, enhance protein thermostability, reduce protein unfolding rate and promote protein-protein interaction in eukaryotes (Hanson *et al.*, 2009; Hebert *et al.*, 2014; Biswas and Chattopadhyaya, 2016; Jayaprakash and Surolia, 2017). However, the role of *N*-linked glycans was less understood in prokaryotes, this could be due to i) first detailed general *N*-linked glycosylation system was reported only a decade and half ago (Linton *et al.*, 2002) ii) the excitement about utilising the newly discovered bacterial *N*-OTase in glycoconjugate vaccine development took precedence over the question as to why bacteria glycosylate their proteins iii) lack of means to investigate the role of *N*-linked glycans since the majority of reported glycoproteins found in bacteria are of unknown function (Scott *et al.*, 2010; Lithgow *et al.*, 2014). To solve this conundrum, I designed complementary experiments to investigate the intrinsic and extrinsic role of *N*-linked glycans in conferring CmeA stability and modulating CmeA interaction with its partner protein; CmeC. The periplasmic fusion glycoprotein; CmeA and a member of the resistance-nodulation-division (RND) multidrug efflux pump which consists other glycoproteins such as CmeB; an inner membrane multidrug transport protein and CmeC; an outer membrane associated channel (Lin, Overbye Michel and Zhang, 2002). CmeA is usually used as an acceptor protein to study not only *N*-OTase substrate specificity but also NGT glycosylation potential (Wacker *et al.*, 2002; Garcia-Quintanilla *et al.*, 2014; Cuccui and Wren, 2015; Jon Cuccui *et al.*,

2017). First to assure the role of *N*-linked glycans in modulating CmeABC multidrug extrusion activity, amino acid substitution was carried out to construct a glycosylation null CmeABC in glycosylation efficient *C. jejuni.* CmeABC activity was monitored by the means of ethidium bromide accumulation. To assess the role of *C. jejuni N*-linked glycans in conferring protein stability, CD spectroscopy was used. CD spectroscopy is a valuable and relatively fast technique to examine protein structural changes in solution. CD measurements in the far UV region can give an estimate on the protein secondary structure in solution (Kelly, Jess and Price, 2005). With new instrumentations developed, CD spectroscopy can be used to monitor, with high accuracy, the structural changes in proteins upon incrementally raising the solution temperature. CmeA variants; glycosylated CmeA and non-glycosylated CmeA, difference in secondary structure were investigated using CD spectroscopy as well as their thermostability.

The extrinsic role of *N*-linked glycans in modulating protein-protein interaction was assessed by surface plasmon resonance (SPR). SPR is an optical technique commonly used to provide a real time quantitative assessment of protein-protein interactions kinetics. It has the advantage of being label free and highly sensitive (Douzi, 2017). SPR has provided detailed insights on the complex assembly and binding of CmeABC orthologue; AcrAB-TolC from *E. coli* (Tikhonova *et al.*, 2009). CmeA variants were immobilised on a CM5 chip and Ni chip, randomly oriented and unidirectionally oriented respectively, and binding kinetics was evaluated.

1.14.3 Bioinformatic analysis of *N*-OTases

The advent of next generation sequencing technologies have revealed the presence of *N*-OTase orthologues in members of alpha and gamma proteobacteria as well as deferribacteres and aquificae (Mills *et al.*, 2016). This finding demonstrates the prevalence of *N*-glycosylation outside epsilon proteobacteria, notably, *Campylobacter spp.* and *H. pullorum.* I used phylogenetic analysis tools and 2D visualisation methods to compare different *N*-OTases from various bacterial families, thus to allowing us to explore the structural heterogeneity and similarities between the oligosaccharyltransferases.

1.14.4 Functional analysis of cytoplasmic *N*-glycosylation

The exciting finding of the discovery of cytoplasmic *N*-glycosylation opened the question of a possibility of divergent evolution of this system along with cytoplasmic *O*-glycosylation. Unravelling the key enzyme responsible for the transfer of glucose or galactose to the acceptor protein sequon N-X-S/T whereby X is any amino acid other than proline, helped to understand the nature of this post translational modification (Schwarz and Aebi, 2011). The pig pathogen *App* has been demonstrated to possess a peculiar type of genetic arrangement of cytoplasmic *N*- glycosylation operon. Unlike *H. influenzae, aphmw1C* (NGT) and α*6gt* (AGT) are not localised in the same operon as their cognate adhesin (Jon Cuccui *et al.*, 2017). The first step to systematically analyse the function of NGT and AGT is to reconstruct the system in an easily genetically manipulated bacteria such as *E. coli.* The reconstruction of NGT-AGT system allowed us to study this post translational modification, explore NGT-AGT protein-protein interaction and explore the system as a non- PglB based

bioconjugation method. As a proof of concept, I used the system to glycosylate CmeA with an α-1,6-glucose polymer. According to our knowledge, this is the first demonstration of an *in vivo* assembly and the transfer of a dextran polymer to a carrier protein using *N*-glycosyltransferase enzyme.

Chapter 2

MATERIALS AND METHODS

2 MATERIALS AND METHODS

2.1 Bacterial strains and conditions

Campylobacter jejuni 11168H and its derivatives; *C. jejuni pglB*::*aphA, C. jejuni napA*::*aphA, C. jejuni cmeB*::*aphA, C. jejuni cmeD*::*cat*, *C. jejuni cmeD*::*cat wtcmeABC* and *C. jejuni cmeD*::*cat* g0*cmeABC* were used in this study. *C. jejuni* 11168H was grown on Columbia based agar or Muller Hinton based agar supplemented with 5% horse blood according to manufacturer's instructions. Strains were grown at 37^{o}C in a variable atmospheric incubator (VAIN) cabinet (Don Whitely, UK) maintaining microaerophilic conditions of: 85% Nitrogen, 5% Oxygen and 10% carbon dioxide. All of the cloning experiments were done in *Escherichia coli* DH10Beta (New England Biolabs, USA). *E. coli* DH10B was used in expression of CmeA and cloning and expression of CmeC whilst gCmeA was expressed in *E. coli* SBD1. *E. coli* strains were grown on either Luria-Bertani Broth or Luria-Bertani Agar and antibiotics were added when necessary.

2.2 Protein preparation and digestion

Bacterial growth, cell lysis, protein quantification were undertaken at London School of Hygiene & Tropical Medicine, whilst the sample preparation peptide labelling, LC-MS/MS were done at the centre of excellence for mass spectrometery, King's College, London using the following protocol. *C. jejuni* were grown overnight in Brucella broth under microaerophilic conditions at 37 ^{o}C under shaking conditions at 75 rpm. The following day, 3.5 ml were withdrawn to inoculate 30 ml culture to a

final OD_{600} =0.1. The cultures were then incubated as above till OD_{600}= 0.4-0.5. Cells were harvested by centrifugation at 4000 xg for 45 minutes. Supernatant was discarded and cells were resuspended in lysis buffer (7 M urea, 2 M thiourea, 5 mM DTT and 2% (wt/vol) CHAPS) ultrasonicated following the manufacturers procedure for extraction of native protein samples from yeast cells (Covaris, US). Protein concentration was then measured using Bradford assay (Sigma, UK) following the manufacturer's protocol then normalised to a total protein amount of 25 µg for enzymatic digestion. Prior to enzymatic digestion and labelling, samples were loaded into a stack gel for lysis buffer clean-up to eliminate any chemical interference at the labelling stage and to compress the whole proteome into a single band. Sample volumes were dried by half in a SpeedVac (Thermo Fisher Scientific, UK) with the volume replaced by Laemmli buffer (2x) and boiled for 10 minutes at 96 ^{o}C. Reduced samples were loaded on to a 10% BisTris NuPAGE gel (Invitrogen, UK) and resolved for 10 minutes (100 volts; 59 mA; 6 watts) to 'stack' the whole sample in to a single band. Protein bands were visualised using Imperial protein stain (Thermo Fisher Scientific, UK). Gel bands were then excised, chopped, washed with acetonitrile (ACN), dried and reduced by rehydration in 10 mM DTT for 30 minutes at 56 ^{o}C. Gel slices were then washed with ACN, dried and alkylated by incubation in 55 mM IAA at ambient temperature for 20 mins in the dark followed by washing in 100 mM TEAB, drying and incubation in 40 µl trypsin (1.25 µg) at 37 ^{o}C overnight. Peptides were extracted by collecting reaction supernatant.

2.3 Peptide TMT labelling

Each sample was treated individually with labels added at a 1:1 ratio**.** TMT6plex labels (Thermo Fisher Scientific, UK) were resuspended in 41 µl ACN, vortexed and centrifuged for 1 minute at 14000 rpm. Each vial of TMT reagent was added to the appropriate sample, vortexed and briefly centrifuged at 14000 rpm before incubation at room temperature for 1 hour. Reaction was stopped by adding 8 µl of 5% hydroxylamine (Sigma, UK) to each sample and incubated for a further 15 minutes at room temperature. Samples were then combined for each TMT6plex and incubated at room temperature for further 15 minutes followed by freezing at –80 ^{o}C and drying in SpeedVac.

2.4 LC-MS/MS tandem mass spectrometry

The combined TMT labelled peptide sample was resuspended in a solution containing water:acetonitrile:trifluoroacetic acid (98%:2%:0.05%) and analysed by LC-MS/MS. Chromatographic separations were performed using an Ultimate 3000 UHPLC system (ThermoFisherScientific, UK). A 3 µl injection of peptides (equivalent of 15 µg peptides) was resolved by reversed phase chromatography on a 75 µm C18 column (50 cm) using a three-step linear gradient of acetonitrile in 0.1% formic acid. The gradient was delivered to elute the peptides at a flow rate of 250 nL/min over 120 min. The eluate was ionised by electrospray ionisation using an Orbitrap Fusion Lumos (ThermoFisherScientific, UK) operating under Xcalibur v4.1. The instrument was programmed to acquire in automated data-dependent switching mode, selecting precursor ions based on their intensity for

sequencing by Higher-energy C-trap dissociation (HCD) for peptide identification and reporter ion fragmentation. Selection of precursor ions based on their intensity for sequencing by HCD in a TopN method. The MS/MS analyses were conducted using higher than normal collision energy profiles that were chosen based on the mass-to-charge ratio (m/z) and the charge state of the peptide. To increase fragmented peptide coverage and reporter ion intensities, a further Synchronous Precursor Scan (SPS) of the Top 5 most intense peaks using MS3 was performed.

2.5 Construction of *C. jejuni cmeB*::*aphA* and *C. jejuni napA*::*aphA*

Inactivation of *cmeB* in *C. jejuni* 11168H was achieved using the method described by Lilian Pumbwe *et al.* 2001 (Pumbwe and Piddock, 2002). Briefly, 1 kbp of *cmeB* was amplified with Phusion polymerase (New England Biolabs, UK) by primers FWDcmeB and REVcmeB using *C. jejuni* 11168H genomic DNA as template. The amplicon was cloned into pJET1.2 to give pRSA. In order to disrupt the gene, the kanamycin resistant gene *aphA*, was cloned in the middle of the *cmeB* fragment to give pRSF. Construction of *C. jejuni* 11168H *napA*::*aphA* was achieved by amplifying the *napA* gene with Phusion polymerase (New England Biolabs, UK) using primers FWDnapA and REVnapA using *C. jejuni* 11168H genomic DNA as a template. The amplicon was then ligated into pGEM T-easy to give pATT3 which was transformed to *E. coli dam*$^{-}$*/dcm*$^{-}$. To disrupt *napA, aphA* cut by BamHI from pJWK30 and blunt ended by CloneJET PCR cloning kit blunting enzyme (Thermoscientific, UK) and then ligated to pATT3 that was previously cut by ClaI and blunt ended to give pATT3F. pRSF and

pATT3F were electroporated in *C. jejuni* 11168H and colonies were selected on CBA supplemented with kanamycin (30 µg/ml).

2.6 Inactivation of *cmeD* and generation of *C. jejuni cmeD*::*cat*, *C. jejuni cmeD*::*cat* wt*cmeABC* and *C. jejuni cmeD*::*cat* g0*cmeABC*.

The nucleotide sequence of *cmeD* gene was commercially synthesized (Clonetech, USA) to also carry a chloramphenicol resistant gene; *cat* was inserted in the middle of *cmeD* to disrupt the gene. The DNA was then released by restriction digestion with EcoRV and cloned in pJET1.2 - following manufacturer's instructions- to give pATN. Cloning of *cmeABC-aphA* was achieved by the following; *cmeABC* locus was amplified by primer FWDCmeA and primerREVCmeC with Phusion polymerase (New England biolabs, UK) using *C. jejuni* 11168H genomic DNA as a template, 6Xhis tag was added at the C-terminus of the CmeC to track its expression. The PCR amplicon was cloned in pJET1.2 following the manufacture's instructions to give pMH3 that was then cut by BamHI to introduce the kanamycin resistant gene *aphA*, to be used as an antibiotic selection marker after homologous recombination in *C. jejuni* 11168H to give pMHT. To add homologous recombination arms for *cmeABC-aphA*, pMH3 was cut by SaCII to ligate *cj0364* at the 3' end of *aphA* to give pMHTF. For g0*cmeABC-aphA,* each asparagine in the non-canonical glycosylation sequon (D/E-X-N-Y-S/T, where X and Y are any amino acid except proline) was altered to glutamine *in-silico* and nucleotide sequence of g0*cmeABC* was synthesized by (Clonetech, USA) DNA was then treated as above to generate pATKH.

To generate *C. jejuni cmeD*::*cat*, electroporation of pATN into *C. jejuni* 11168H was carried out as previously described (Karlyshev, 2004). The transformants were selected on CBA plates supplemented with 10 µg/ml chloramphenicol and the double cross over event was confirmed by PCR, this strain was then used as parent strain to generate other mutants. Plasmids pMHT and pATK were electroporated into *C. jejuni cmeD*::*cat* to generate *C. jejuni cmeD*::*cat cmeC*::*cmeC-aphA* and *C. jejuni cmeD*::*cat cmeABC*::*cmeABC*-(N->Q)-*aphA,* respectively. Transformants were selected on CBA plates supplemented with 10 µg/ml chloramphenicol and 30 µg/ml kanamycin and the double cross over event was confirmed by PCR.

2.7 Determination of nitrite concentration in the culture supernatants

Determining of nitrite concentration in the culture supernatants was done according to Pitman *et al* (2007)(Pittman *et al.*, 2007) with slight modification. Briefly, an overnight culture of *C. jejuni* grown in brucella broth was diluted to OD_{600} 0.1 in oxygen limited condition. Samples were drawn every hour and centrifuged at 12,000 xg for one minute. Supernatant was then diluted 1:5 with deionized water. 50 µl of diluted culture supernatant was mixed with 850 µl of 1% (w/v) sulphanilamide dissolved in 1 M HCl and 100 µl of 0.02% (w/v) naphthylethylenediamine. After 15 min, the absorbance at 540 nm was measured using plate reader (Molecular Devices M3 plate reader, USA) and nitrite concentrations were determined by reference to a standard curve.

2.8 Chicken colonisation experiments

Chicken colonisation experiments were undertaken by Mark Steven's group at Roslin Institute, University of Edinburgh. All procedures were conducted under Home Office project licence PPL 60/4420, according to the requirements of the Animal (Scientific Procedures) Act 1986, with the approval of the Ethical Review Committee of the Moredun Research Institute. A total of 40 White Leghorn chickens, obtained on the day of hatch from a Home Office licensed breeding establishment were used. Chickens were housed in groups of 20 in colony cages under specified-pathogen free conditions. Groups were of mixed sex and were wing-tagged for individual identification. Water and sterile irradiated feed based on vegetable protein (DBM Ltd., UK) was provided *ad libitum*. Fresh faecal samples collected from the cages were plated onto mCCDA plates to confirm the absence of *C. jejuni* in the chickens. At two week of age, chickens were challenged with 10^6 colony-forming units (CFU) of *C. jejuni* 11168H or 11168H *pglB::aphA* in a volume of 100 µl. At one and two weeks after challenge, post-mortem examinations of 10 birds from each group were performed. Contents of each ceacum from each bird were weighed and serially diluted in phosphate-buffered saline (PBS) then plated on mCCDA plates to enumerate viable *Campylobacter* per gram of caecal contents. Statistical analysis was performed using the Mann- Whitney test.

2.9 Scanning electron microscopy

Sample preparation was undertaken at London School of Hygiene & Tropical Medicine, whilst SEM was done at the Centre of ultrastructural

imagine, King's College, London. *C. jejuni* cells were resuspended in PBS after being scraped from an overnight culture grown on Columbia based agar plate supplemented with blood and Skirrow supplement. The bacterial suspension was then mixed with fixing buffer (0.1 M sodium cacodylate buffer containing 4% paraformaldehyde, pH 7.4) and fixed on polylysine coated coverslips for overnight at 4 ^{o}C. Samples were then washed twice for 10 minutes in 0.15 M sodium cacodylate buffer pH 7.3 then osmicated for 1 hour at room temperature with 1% osmium tetroxide in 0.15 M cacodylate phosphate buffer. Samples were dehydrated in an ethanol series consisting of: 10 minutes in 10% ethanol, 30 minutes in 70% ethanol and 3 times in 100% ethanol (20 minutes each) followed by transferring the coverslips to 100% ethanol in critical point drying holders then critical point dried using carbon dioxide in a Polaron E3000 critical point dryer with >3 flushes over 45 minutes soaking period. Coverslips were mounted on stubs (TAAB, UK) with conductive carbon adhesive discs (TAAB) and samples were sputter coated with gold in a Emitech K550 sputter coater (45 mA current, 2 minutes) cells were viewed and images were recorded using a FEI Quanta 200 Field emission scanning electron microscope, (FEI UK Ltd.) operated at 10-20 kV in high vacuum mode.

2.10 Antibiotic sensitivity test (E-test)

C. jejuni 11168H were grown in suspension in Mueller-Hinton broth equivalent to 1.0 MacFarland's standard and 100 µl aliquots were spread plated on dry Mueller-Hinton agar plates supplemented with 5 % Sheep blood (Oxoid, UK), the plates were left for 5 -10 minutes to dry before the

antibiotic strip (Oxoid, UK) was added. Plates were incubated at 37^{o}C overnight. The minimum inhibitory concentration (MIC) was read directly from the strip at the point where the zone of inhibition of bacterial growth intersected with the antibiotic concentration on the strip.

2.11 Ethidium bromide accumulation assay

Eithidium bromide accumulation assay allows monitoring the efficiency of multidrug efflux pump in real time manner. Bacterial RND multidrug efflux pump extrudes different non-related compounds, ethidium bromide that is not efficiently extruded by multidrug efflux pump is quantified spectrophotometerically. Efficiency of the multidrug pump is inversely proportion to the amount of ethidium bromide bound to the genomic DNA. Thus, more ethidium bromide bound to the cell gDNA reflects a decrease in multidrug efflux pump efficiency.

Bacterial cells were grown to mid log phase (OD_{600} 0.4-0.5). Cells were harvested, washed and resuspended in 0.1M sodium phosphate buffer pH 7 (previously incubated in the VAIN) to OD_{600} 0.2. Cells were then incubated in the VAIN for 15 mins at 37^{o}C o before a 100µl aliquot was withdrawn to indicate time zero. Ethidium bromide (Sigma, UK) was added to final concentration 2 µg/ml and fluorescence was measured at 530 nm excitation and 600 nm emission using a plate reader (Molecular Devices M3 plate reader, USA).

2.12 Expression of CmeA and gCmeA

Protein expression was carried out in *E. coli* strains unless stated otherwise. CmeA and CmeC were expressed in *E. coli* DH10B carrying pMH5 plasmid and pAT3, respectively, whilst gCmeA was expressed in *E. coli* SDB1 carrying pGVXN114, pWA2 and pACYC*(pgl*). Initiating cultures

were grown overnight in LB broth supplemented with appropriate antibiotics at 37 ^{o}C under shaking condition. The following day, 10 ml of culture was withdrawn from the shake flask to inoculate 400 ml LB broth supplemented with appropriate antibiotics. To achieve optimal glycosylation of CmeA, PglB was expressed from pGVXN114 by the addition of 0.5 mM ITPG at OD $_{600}$ 0.5-0.6. Cultures were incubated at 37^{o}C for 24 hours with shaking. Cultures were centrifuged and cell pellets. washed with binding buffer (300 mM NaCl, 50 mM NaH_2PO_4 with 25 mM imidazole) and passed twice through a high-pressure cell homogeniser (Stanstead works, UK). Cell debris was removed by centrifugation at 20,000 xg for 45 minutes. Supernatant was collected and incubated with

0.2 ml Ni-NTA for 1 hour at 4 ^{o}C then washed with 50 ml binding buffer and eluted four times in 0.5 ml elution buffer (300 mM NaCl, 50 mM NaH_2PO_4 with 250 mM imidazole).

2.13 Cloning and expression of CmeC

To express CmeC in *E. coli*, *cmeC* lacking signal peptide sequence was amplified by PCR with CmeCFwd1 and CmeCRev using *C. jejuni* 11168H genomic DNA as a template. The amplicon was then cut by NheI and SalI and cloned into pEC415 downstream of the DsbA signal peptide sequence to give pCMECDSBA. *E. coli* carrying pCMECDSBA was grown in LB media supplemented with ampicillin (100 µg/ml) overnight at 37 ^{o}C under shaking condition. On the following day, 10 ml were withdrawn from the overnight culture to inoculate 400 ml LB media. Cells were grown to OD $_{600}$ 0.5-0.6 and 0.2% L-arabinose was added to induce the expression of CmeC. Cultures were incubated at 37 ^{o}C for 24 hours with shaking at 180 rpm. Cultures were centrifuged and cell pellets washed with binding buffer

(300 mM NaCl, 50 mM NaH_2PO_4 with 25mM imidazole) and passed twice through cell homogeniser (Stanstead works, UK). Cells debris was removed by centrifugation at 20,000 xg for 45 minutes and then collected and incubated in binding buffer with 2 % DDM for 3 hours at 4 ^oC. The mixture was then centrifuged at 15,000 xg for 10 minutes. The supernatant was collected, diluted with binding buffer and incubated with 0.2 ml Ni-NTA for 1 hour at 4 ^oC then washed with 50 ml binding buffer and eluted four times in 0.5 ml elution buffer (300 mM NaCl, 50 mM NaH_2PO_4 with 250 mM imidazole).

2.13 CD Spectroscopy

Sample preparation was undertaken at London School of Hygiene & Tropical Medicine, whilst CD spectroscopy was done with the help of Tam Bui of King's College, London. All CD spectra of gCmeA and CmeA were acquired in 0.5mm rectangular cell pathlength using Chirascan spectrometer (Applied Biophysics, UK) equipped with Quantum NorthWest TC125 Peltier unit. Temperature dependent confirmation changes were monitored at wavelength 260-195 nm for gCmeA (0.2 mg/ml) and CmeA (0.2 mg/ml) in 10 mM Sodium phosphate, 75 mM Sodium chloride, 10 % glycerol buffer (pH=8.0) during stepwise increase in temperature from 6^oC to 94^oC. Temperatures were measured directly with a thermocouple probe in the sample solution. Melting temperatures were determined from the derivative CD-Temperature spectra and fitted using a Levenberg–Marquardt algorithm (LMA) on the *van*'t *Hoff isochore*. (Global 3, Global Analysis for T-ramp Version 1.2 built 1786, Applied Photophysics Ltd, 2007-2012). For Conformation Reversibility Study, far-UV CD spectra were recorded at 20^oC, raised to T_m and re-cooled to 20^oC. The

temperature at each elevated T_m was kept constant for 5 minutes and the CD spectrum was recorded to assess the rate of protein unfolding process.

2.14 Surface Plasmon Resonance

For coupling of CmeA and gCmeA to the CM5 sensor chip, carboxyl groups on the surface were activated by injecting a 1:1 mixture of 0.4 M 1- ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) and 0.1M N-hydroxysuccinimide (NHS) for 7 minutes at 5 µl/min. CmeA and gCmeA were diluted to 10-20 µg/ml in 0.1 M acetate pH 5.5 and immobilised at 5 µl/min. Immobilisation was stopped when the required RU was achieved. This was followed by injecting 1 M ethanolamine pH 8.5 (7 minutes at 5 µl/min) to inactivate excess reactive groups. To account for non-specific binding, a control flow cell was generated using the same method described minus the protein immobilisation step. For coupling of CmeA and gCmeA to a NTA chip, the chip was cleaned and loaded with $NiCl_2$ (0.5 mM). The flow cells were then activated as above and CmeA and gCmeA (10 ug/ml in HBSP buffer) were loaded into appropriate flow cells until appropriate RU were achieved. Subsequently the flow cells were treated with ethanolamine as above to block remaining activated sites. Cmec at various concentrations (3 nM- 0.2 µM) was analysed at a constant temperature of 25°C under continuous flow of HBS-PE buffer (10mM HEPES pH 7.4, 3 mM EDTA, 0.005% (w/v) Surfactant P20 (GE Healthcare) at 30 µl/min (sufficient to prevent mass transfer effects) at pH 7.4 for 3 minutes association and a dissociation time of 5 minutes. Experiments at pH 6.0 were performed with 10 mM MES pH 6.0, 3 mM EDTA, 0.005% (w/v) Surfactant P20 (GE Healthcare) The surface chip was regenerated by injecting 0.1 M triethanolamine pH 11.5. Data was

analysed using the BIAevaluation software version 4.1.1 (Biacore, GE Healthcare, Amersham). Blank flow cell controls were subtracted. The k_d was defined between 10s after the end of the sample injection and 300 sec later.

Chapter 3

RE-EVALUATION OF THE ROLE OF THE GENERAL *N*- LINKED GLYCOSYLATION SYSTEM IN *CAMPYLOBACTER JEJUNI* – A MODEL TO STUDY PROTEIN GLYCOSYLATION IN BACTERIA

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Sherif Abouelhadid
Principal Supervisor	Brendan W Wren
Thesis Title	Investigating the biological role of *N*-linked glycan in bacteria

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Nature communications
Please list the paper's authors in the intended authorship order:	Sherif Abouelhadid, Simon North, Paul Hitchen, Prerna Vohra, Cosmin Chintoan-Uta, Mark Stevens, Anne Dell, Jon Cuccui, Brendan W Wren
Stage of publication	Submitted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	The candidate designed the experiments, carried out most of the experiments, analysed the data and wrote the manuscript.

Student Signature: **Date:**

Supervisor Signature: [redacted] **Date: 14/Sept/2018**

Re-evaluation of the role of the general *N*-linked glycosylation system in *Campylobacter jejuni* – a model to study protein glycosylation in bacteria

Sherif Abouelhadid[1], Simon North[2], Paul Hitchen[2], Prerna Vohra[3], Cosmin Chintoan-Uta[3], Mark Stevens[3], Anne Dell[2], Jon Cuccui[1], Brendan W Wren[1]

[1]Department of Pathogen Biology, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

[2]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

[3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, EH25 9RG, UK

Abstract

In eukaryotes glycosylation plays a role in proteome stability, protein quality control and modulating protein function, however, similar studies in bacteria are lacking. Here, we investigate the role of general protein glycosylation systems in bacteria using the enteropathogen *Campylobacter jejuni* as a well-defined example. By using a quantitative proteomics strategy, we were able to monitor changes in the *C. jejuni* proteome when glycosylation is disrupted. We demonstrate that in *C. jejuni, N*-glycosylation is essential to maintain proteome stability and protein quality control. These findings guided us to investigate the role of *N*-glycosylation in modulating bacterial cellular activities. In glycosylation deficient *C. jejuni,* the multidrug efflux pump and electron transport pathways were significantly impaired. *In vivo*, we demonstrate that fully glycosylation deficient *C. jejuni* were unable to colonise its natural avian host. These results provide the first evidence of a link between proteome stability and complex functions via a bacterial general glycosylation system.

Introduction

Glycosylation is a prevalent protein post-translational modification found in nature. In eukaryotes, the attachment of glycans to proteins has been shown to play a central role in modulating protein folding, stability and signalling[1]. Advances in genomics and mass spectrometry have revealed several types of glycosylation systems in bacteria. Among prokaryotes, general glycosylation systems have been reported in pathogenic bacteria including *Campylobacter jejuni*[2], *Neisseria gonorrhoeae*[3], *Burkholderia cepacia*[4] as well as the commensal bacteria such as *Bacteroides fragilis*[5], indicating their abundance across the kingdom. However, the reasons why bacterial proteins are specifically modified remains poorly defined, particularly among general protein glycosylation systems.

Genetic and biochemical studies have revealed that in *C. jejuni,* the *N*-oligosaccharyltransferase PglB is responsible for decorating at least 50 proteins in the *C. jejuni* cell with the heptasaccharide; GalNAc-α1,4-GalNAc-[Glcβ1,3-]GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-β1, where GalNAc is *N*-acetylgalactosamine, Glc is glucose, Bac is 2,4-diacetamido -2,4,6-trideoxyglucopyranose. This glycostructure is linked to the asparagine residue in the acceptor sequon D/E-X-N-Y-S/T where X and Y can be any amino acid except proline[6]. Comparative genomics studies have revealed the conservation of general *N*-glycosylation gene clusters across the *Campylobacter* genus[7]. Mutagenesis studies demonstrated that disruption of *N*-glycosylation pathway led to pleiotropic effects in *C. jejuni*. These effects include a substantial decrease in adherence and invasion capability of INT 407 intestinal

epithelial cell[8], a decrease in chicken colonization[9], and a reduction in competence[10]. Similar findings have been also reported in other bacteria where inactivation of their corresponding general glycosylation systems resulted in a disruption in cellular physiology and a reduction in virulence[4,5,9]. These combined consequences were anticipated to be due to direct or indirect disruptions of several cellular processes in glycosylation deficient strains. However, these studies did not address what perturbations may be occurring to the bacterial cell when glycosylation is no longer possible. To answer this question, we developed a platform to investigate the role of general *N*-linked glycosylation systems in prokaryotes, using *C. jejuni* as a model example.

We employed tandem mass tags (TMT) isobaric 6plex labelling quantitative proteomics on wildtype and glycosylation deficient strains. Analysis of a glycosylation deficient strain, revealed chaperones and stress related proteins are more abundant, indicating a central alteration in bacterial cellular activities. Thus, contrary to current understanding the loss of glycosylation predominantly has a mass effect on the *C. jejuni* proteome, rather than being restricted exclusively to proteins modified by glycosylation. Phenotypically this resulted in the reduction of multidrug efflux pump activity as well as an impairment in nitrate reductase activity. Furthermore, we demonstrate that general *N*-glycosylation plays an important role in host-microbe interactions in the chicken colonisation model. This study provides deeper insights on the role of a general *N*-glycosylation system in bacteria demonstrating that protein post translation modification is critical to a plethora of bacterial cellular activities.

Results

N-glycosylation is essential for proteome stability

Quantitative proteomics serves as a tool to understand the alteration in proteomes resulting from gene(s) disruption. To provide deeper insights into the role of general glycosylation system in prokaryotes, we employed a mass spectrometry-based quantitative proteomics strategy to monitor differential protein abundance in glycosylation deficient *C. jejuni* and compared this to wild-type bacteria. Proteins from both *C. jejuni* strains were reduced, alkylated, digested then labelled by Tandem Mass Tags (TMT) 6plex isobaric labels to allow multiplex quantification for multiple samples**.** In total, 1052 proteins were identified by matching peptides to the *C. jejuni* 11168H curated proteome. Proteins were quantitatively compared across the samples (6plex) then filtered at 95% CI. After significance filtering, 502 proteins were found to be statistically differentially abundant among the samples, out of which 120 proteins were more abundant and 382 proteins were found in lower abundance in *C. jejuni pglB*::*aphA* lacking the heptasaccharide transferase **Fig 1, A-C** . Surprisingly, chaperones and proteases were found in significantly higher abundance in the glycosylation deficient *C. jejuni* **Fig 1, D**. Among these chaperones, we detected GroL, a periplasmic chaperone that binds to unfolded and partially folded proteins promoting folding[13], HtpG, a chaperone involved in binding to aggregated proteins and protecting the cell against environmental stresses[14] and Cj0694; a peptidyl-prolyl-*cis*/*trans* isomerase (PPIase) that interacts with the Sec translocon and promotes folding of polypeptides emerging from the translocon[15,16]. Protein degradation pathway proteins were also significantly highly abundant, notably HtrA, which is a multifunctional protein that plays a central role in proteolytically cleaving misfolded proteins as well as preventing aggregation of

non-native proteins[15,17]. This observation correlates with a significant reduction in abundance of HspR in the same strain. HspR is a negative transcriptional regulator of chaperone-encoding genes such as *clpB, dnaK* and *groLS*. Disruption of *hspR* enhanced thermotolerance in *C. jejuni*[18]*.* Cytoplasmic chaperones and proteases such as ClpB, Lon, Trigger factor (Tig) and HslU that are reported to be involved in various arrays of protein quality mechanisms from assisting protein folding to protein degradation were also found to be differentially, but not statistically significantly, abundant in *C. jejuni pglB*::*aphA*[19] **Fig 1, D***.* These results indicate that disrupting *N*-linked glycosylation leads to a global alteration in the abundance of protein quality control machineries. The high abundance of chaperones and proteases suggests a potential misfolding of proteins and a rise in protein aggregation in glycosylation deficient *C. jejuni.*

Alteration in glycoproteins abundance in glycosylation deficient *C. jejuni*

Biochemical and biophysical studies of glycoproteins have revealed that *N*-linked glycans play a role in stabilising proteins[20], promoting protein folding[21] and preventing protein aggregation[22]. Enrichment analysis of the glycoproteome of *C. jejuni* has revealed 53 proteins to be decorated by the *C. jejuni* heptasaccharide; GalNAc-α1,4-GalNAc-[Glcβ1,3-]GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-diBacNAc covalently linked to the asparagine residue in D/E-X-N-Y-S/T where X and Y can be any amino acid except proline[6,23] . Our quantitative proteomic approach managed to detect and correctly evaluate 35 glycoproteins across the samples. Out of the 35 glycoproteins detected, only 8 (22.8%) were more abundant in glycosylation deficient *C. jejuni* and 6 proteins (17.1%) were found to be unchanged. Interestingly, 21 proteins (60%) were found to be less abundant in *C. jejuni pglB*::*aphA.* It is unclear whether the low abundance of these proteins is due to

misfolding/degradation of the proteins or is a consequence of global defects resulting from disruption of the *N*-linked glycosylation pathway. To gain more insights into the role of the *C. jejuni* heptasaccharide in promoting protein stability, we sought to study the functional activity of glycoproteins in *C. jejuni.*

_N_-glycosylation modulate multidrug efflux pump activity

A difficulty in studying the role of general glycosylation systems in bacteria is that the modified proteins often have no known function. Using a genetically tractable bacterium that modifies proteins of predictable phenotypic could help to assign more definitively the role of glycosylation. Only two protein assemblies out of the 53 detected glycoproteins in our study on *C. jejuni* have known phenotypes[6,24,25]. CmeABC shares amino acid sequence homology and similar molecular architecture with other tripartite resistance-nodulation-division (RND) type multidrug efflux pumps in other Gram-negative bacteria[26]. Acting as the major multidrug efflux pump in *C. jejuni,* CmeABC plays an important role in antibiotic resistance and host colonisation[27]. The tripartite RND of *C. jejuni* is a molecular assembly of CmeB, an inner membrane multidrug transport protein, CmeA, a periplasmic fusion protein and CmeC, an outer membrane associated protein channel[24], where complex components were found to be glycosylated[6]. In *C. jejuni pglB*::*aphA,* CmeA was found to be 0.8 fold less abundant, whilst no change was seen in the level of CmeB and CmeC when compared to the wild-type Fig 2, A. To investigate the role of the heptasaccharide in multidrug resistance, we sought to assess the antibiotic resistance profile of *C. jejuni* before and after disruption of CmeABC glycosylation. A *C. jejuni cmeB*::*aphA* mutant was constructed and showed a substantial increase in antibiotic sensitivity. In glycosylation deficient *C. jejuni,* a 100% increase in sensitivity against ampicillin, erythromycin, tetracycline and ciprofloxacin was

observed, compared to the wild-type Supp Table 2. The results may be due to a defect in CmeABC multidrug efflux leading to an accumulation of antibiotics. In order to gain further insights on the effect of glycosylation on CmeABC activity, we used the fluorescent intercalating agent ethidium bromide in an accumulation assay that enables the monitoring of efflux pump activity in real-time. Ethidium bromide accumulation in *C. jejuni cmeB::aphA* was 15% higher (*P*=0.0016) than the wild-type strain. These results are in agreement with the antibiotic susceptibility tests. The ethidium bromide accumulation in *C. jejuni pglB*::*aphA* was 27.5% higher compared to wild-type (*P*<0.0001) Fig 2, B and C. The results show that the effect of glycosylation on *C. jejuni* ethidium bromide accumulation kinetics is more detrimental than inactivation of the major multidrug efflux complex. Our results demonstrate the pivotal role of the heptasaccharide in efflux activity and subsequent multidrug resistance in *C. jejuni.*

Nitrite production impairment in glycosylation deficient *C. jejuni*

To analyse whether the loss of glycosylation affects other complex assemblies in *C. jejuni*, we interrogated the activity of NapAB for nitrite production. NapAB is a two-subunit periplasmic nitrate reductase enzyme that is responsible for the reduction of nitrate to nitrite. NapAB consists of a twin arginine translocation substrate; NapA, is a 90 kDa catalytic *bis*-molybdenum guanosine dinucleoside (MGD) cofactor binding enzyme whilst NapB is a 16 kDa di-haem *C*-type cytochrome glycoprotein[6,25]. Inactivation of *napA* in *C. jejuni* results in reduction in chicken colonization[28]. Mass spectrometric analysis of the *C. jejuni* glycoproteome demonstrated that NapB is glycosylated at ^{50}N however, there is no experimental evidence that NapA is a glycoprotein[6]. Our quantitative proteomic analysis unexpectedly revealed a 1.7±0.2 fold increase in both NapA and NapB levels in *C.*

jejuni pglB::*aphA* compared to the wild type **Fig 3, A**. Interestingly, there was no change in abundance in twin arginine translocase (TatA) or NapA accompanying chaperone; NapD. To discern the role of *N*-glycosylation in modulating NapAB complex activity, we sought to investigate the ability to reduce nitrate in *C. jejuni* and *C. jejuni pglB*::*aphA* and a newly constructed *C. jejuni napA*::*aphA* mutant. The *C. jejuni* strains were grown in oxygen limited conditions for 8 hours. A biphasic pattern of nitrite production was observed whereby, nitrite production increases until 6 hours, then starts to decline Fig 3, B. After 1 hour, a difference in nitrite production between the wild-type and its glycosylation deficient counterpart was observed, whilst nitrite production in the *C. jejuni napA*::*aphA* mutant could not be detected, (below the limit of detection). At 6 hours, *C. jejuni pglB::aphA* produced 63.7% (*P*=0.0018) less nitrite than the wild-type Fig 3, C. Our results indicate a potential correlation between *N*-glycosylation and nitrate reductase activity in *C. jejuni.* It also suggests that despite the increase in NapAB abundance in glycosylation deficient strain, the complex activity was severely impaired compared of its glycosylated counterpart. That indicates that *N*-glycosylation play a role in enhancing NapAB complex efficiency in nitrite production.

Role of *N*-glycosylation in host-pathogen interaction

To determine the impact of protein glycosylation in host-pathogen interaction, we carried out chicken colonisation experiments. Owing to its prevalence in chicken and other avian species, many reports have used chicken models of infection to test the effect of mutations in *C. jejuni*[9,29]*.* We have previously reported that disruption of *pglH* substantially reduced colonisation in 2-week-old Light Sussex chickens[9]. The glycosyltransferase PglH catalyses the addition of three α1,4- GalNAc to GalNAc-α1,4-GalNAc-α1,3-diBacNAc-β1, an important step in the production of the

heptasaccharide. Proteins from *C. jejuni pglH*::*aphA* can still be glycosylated by the truncated heptasaccharide (GalNAc-α1,4-GalNAc-α1,3-Bac-β1). To understand the role of glycosylation in chicken colonisation we sought to investigate the impact of a *C. jejuni pglB* mutant devoid of the ability to fully glycoyslate proteins and compare it to the wild-type. We separately inoculated groups of twenty 2-week-old *Campylobacter*-free White Leghorn chickens with 10^6 CFU *C. jejuni* wild-type or *pglB*::*aph* mutant Fig 4, A. Ten chickens from each group were sacrificed at days 6 or 13 post-inoculation and viable *C. jejuni* in the caeca (a key site of persistence) enumerated following post mortem. Glycosylation deficient *pglB*::*aph C. jejuni* showed a six–$\log_{10}$ reduction in caecal colonization when compared to the wild-type at day 6 (P<0.0001) and day 13 (P<0.0001) Fig 4, B and C. This reduction in colonisation is in agreement with our previous work in *C. jejuni pglH*::*aphA*. However, whilst glycan truncation in *C. jejuni pglH*::*aphA* reduced chicken colonisation [9], the complete abolishment of glycosylation in *C. jejuni pglB*::*aphA* had a more severe impact on chicken colonisation. The data highlights the importance of glycosylation in *Campylobacter* survival in the avian reservoir primarily responsible for zoonotic infections.

Discussion

Our results provide in-depth analyses on the alteration of the *C. jejuni* proteome as a consequence of disrupting the *N*-linked protein glycosylation pathway. Loss of glycosylation predominantly has a mass effect on the *C. jejuni* proteome, rather than being restricted exclusively to glyco-modified proteins. Previous studies have reported that loss of glycosylation affects bacterial motility and virulence[4,30]. However, none of these studies investigated the relative quantitation of individual protein(s) and their function(s). Our TMT LC-MS/MS strategy shows that disruption of *N*-linked glycosylation correlates with a marked increase in the abundance of not

only periplasmic but also cytoplasmic chaperones and proteases likely to be involved in protein folding, protein disaggregation and protein degradation. These results suggest a critical defect in protein folding that potentially impairs protein cellular function. **Suppl 1**. The reason(s) as to why the cell count was reduced at this OD_{600} in glycosylation deficient *C. jejuni* strains remains unclear. One explanation may involve the observed reduction in the abundance of the cell division protein FtsZ (*P*=0.006) **Fig 1, A**. This was accompanied with a change in cell morphology as determined by scanning electron microscopy **Suppl 2.** suggesting major cellular alterations upon abolishing *N*-linked glycosylation. Functional analysis demonstrates that molecular assemblies and protein functions were statistically impaired in glycosylation deficient *C. jejuni.* We postulate that the difference in extrusion kinetics of ethidium bromide in *C. jejuni* strains were dependent on the *C. jejuni* heptasaccharide. Whilst the abundance of CmeABC was similar in *C. jejuni* and *C. jejuni pglB*::*aphA,* we observed a significant impairment in ethidium bromide extrusion in the glycosylation deficient strain, even when compared to the *C. jejuni cmeB*::*aphA*. This might be due to a central role of glycans in stabilizing not only CmeABC as the major multidrug efflux pump, but also CmeDEF as a secondary pump[31]. Indeed, CmeE a periplasmic fusion glycoprotein in *C. jejuni* [6], was significantly reduced in *C. jejuni pglB*::*aphA* (*P* <0.0001). The low abundance of CmeE suggests that CmeDEF is inefficiently assembled, and therefore functionally impaired. By testing other molecular assembly functions, we noticed a general role played by the *C. jejuni* heptasaccharide. Notably, the high abundance in the protein

level of NapAB in *C. jejuni pglB*::*aphA* was not accompanied by an increase in the nitrate reduction function of the complex. Indicating a severe impairment in the function of NapAB resulted from the loss of *N*-linked glycans. This effect is extended beyond cellular activities and the assembly of protein complexes. Our results show a dramatic decrease in chicken colonisation in *C. jejuni pglB*::*aphA*, greater than that previously reported in *C. jejuni pglH*::*aphA*[9]. This provides a deeper insight into the essential role of the first two or three glycans (GalNAc-α1,4-GalNAc-α1,3-diBacNAc) at the reducing end (which are conserved among *Campylobacter* genus,[7,32]) in host-microbe interaction. This study facilitates the understanding of the biological role of this general *N*-linked glycosylation system. We present a platform that enables the study of glycan-protein structural and functional relationship and could be applied to other prokaryotic protein glycosylation systems. Efforts to couple transcriptional and proteomic analysis to interrogate molecular changes resulting from altering or abolishing bacterial general glycosylation pathways will enrich our understanding of the evolution and emergence of protein post translational modification.

Materials and methods

Bacterial strains and growth conditions

Campylobacter jejuni 11168H[33], and its derivatives; *C. jejuni pglB*::*aphA*, *C. jejuni cmeB*::*aphA* and *C. jejuni napA::aphA* were used in this study. *C. jejuni* 11168H was grown on Columbia based agar or Mueller Hinton based agar supplemented with 5% horse blood according to manufacturer's instructions. *Campylobacter* strains were grown at 37 ^{o}C in a variable atmospheric incubator (VAIN) cabinet (Don Whitley, UK) maintaining microaerophilic conditions of 85% Nitrogen, 5% Oxygen and 10% Carbon dioxide.

Protein preparation and digestion

C. jejuni were grown overnight in Brucella broth under microaerophilic conditions at 37 ^{o}C under shaking conditions at 75 rpm. The following day, 3.5 ml were withdrawn to inoculate 30 ml culture to a final OD_{600} =0.1. The cultures were then incubated as above till OD_{600}= 0.4-0.5. Cells were harvested by centrifugation at 4000 xg for 45 minutes. Supernatant was discarded and cells were resuspended in lysis buffer (7 M urea, 2M thiourea, 5 mM DTT and 2% (wt/vol) CHAPS) ultrasonicated following the manufacturers procedure for extraction of native protein samples from yeast cells (Covaris, US). Protein concentration was then measured using Bradford assay (Sigma, UK) following the manufacturer's protocol then normalised to a total protein amount of 25 µg for enzymatic digestion. Prior to enzymatic digestion and labelling, samples were loaded into a stack gel for lysis buffer clean-up to eliminate any chemical interference at the labelling stage and to compress the whole proteome into a single band. Sample volumes were dried by half in a SpeedVac (Thermo Fisher Scientific, UK) with the volume replaced by Laemmli buffer (2x) and boiled for 10 minutes at 96 ^{o}C. Reduced samples were loaded on to a 10% BisTris NuPAGE gel (Invitrogen, UK) and resolved for 10 minutes (100 volts; 59 mA; 6 watts) to 'stack' the whole sample in to a single band. Protein bands were visualised using Imperial protein stain (Thermo Fisher Scientific, UK). Gel bands were then excised, chopped, washed with acetonitrile (ACN), dried and reduced by rehydration in 10 mM DTT for 30 minutes at 56 ^{o}C. Gel slices were then washed with ACN, dried and alkylated by incubation in 55 mM IAA at ambient temperature for 20 mins in the dark followed by washing in 100 mM TEAB, drying and incubation in 40 µl trypsin (1.25 µg) at 37 ^{o}C overnight. Peptides were extracted by collecting reaction supernatant.

Peptide TMT labelling

Each sample was treated individually with labels added at a 1:1 ratio**.** TMT6plex labels (Thermo Fisher Scientific, UK) were resuspended in 41 µl ACN, vortexed and centrifuged for 1 minute at 14000 rpm. Each vial of TMT reagent was added to the appropriate sample, vortexed and briefly centrifuged at 14000 rpm before incubation at room temperature for 1 hour. Reaction was stopped by adding 8 µl of 5% hydroxylamine (Sigma, UK) to each sample and incubated for a further 15 minutes at room temperature. Samples were then combined for each TMT6plex and incubated at room temperature for further 15 minutes followed by freezing at –80 ^{o}C and drying in SpeedVac.

LC-MS/MS tandem mass spectrometry

The combined TMT labelled peptide sample was resuspended in a solution containing water:acetonitrile:trifluoroacetic acid (98%:2%:0.05%) and analysed by LC-MS/MS. Chromatographic separations were performed using an Ultimate 3000 UHPLC system (ThermoFisherScientific, UK). A 3 µl injection of peptides (equivalent of 15 µg peptides) was resolved by reversed phase chromatography on a 75 µm C18 column (50 cm) using a three-step linear gradient of acetonitrile in 0.1% formic acid. The gradient was delivered to elute the peptides at a flow rate of 250 nL/min over 120 min. The eluate was ionised by electrospray ionisation using an Orbitrap Fusion Lumos (ThermoFisherScientific, UK) operating under Xcalibur v4.1. The instrument was programmed to acquire in automated data-dependent switching mode, selecting precursor ions based on their intensity for sequencing by Higher-energy C- trap dissociation (HCD) for peptide identification and reporter ion fragmentation. Selection of precursor ions based on their intensity for sequencing by HCD in a TopN method. The MS/MS analyses were conducted using higher than normal collision

energy profiles that were chosen based on the mass-to-charge ratio (m/z) and the charge state of the peptide. To increase fragmented peptide coverage and reporter ion intensities, a further Synchronous Precursor Scan (SPS) of the Top 5 most intense peaks using MS3 was performed.

Construction of *C. jejuni cmeB*::*aphA* and *C. jejuni napA*::*aphA*

Inactivation of *cmeB* in *C. jejuni* 11168H was achieved using the method described by Lilian Pumbwe *et al.* 2001[26]. Briefly, 1 kbp of *cmeB* was amplified with Phusion polymerase (New England Biolabs, UK) by primers FWDcmeB and REVcmeB using *C. jejuni* 11168H genomic DNA as template. The amplicon was cloned into pJET1.2 to give pRSA. In order to disrupt the gene, the kanamycin resistant gene *aphA*, was cloned in the middle of the *cmeB* fragment to give pRSF. Construction of *C. jejuni* 11168H *napA*::*aphA* was achieved by amplifying the *napA* gene with Phusion polymerase (New England Biolabs, UK) using primers FWDnapA and REVnapA using *C. jejuni* 11168H genomic DNA as a template. The amplicon was then ligated into pGEM T-easy to give pATT3 which was transformed to *E. coli dam*$^{-}$*/dcm*$^{-}$. To disrupt *napA, aphA* cut by BamHI from pJWK30 and blunt ended by CloneJET PCR cloning kit blunting enzyme (Thermoscientific, UK) and then ligated to pATT3 that was previously cut by ClaI and blunt ended to give pATT3F. pRSF and pATT3F were electroporated in *C. jejuni* 11168H and colonies were selected on CBA supplemented with kanamycin (30 µg/ml).

Antibiotic sensitivity

C. jejuni 11168H was grown in suspension in Mueller-Hinton broth equivalent to 1.0 MacFarland's standard and 100 µl aliquots were spread plated on dry Mueller-Hinton agar plates supplemented with 5 % Sheep blood (Oxoid, UK), the plates were left for 5 -10 minutes to dry before the antibiotic strip (Oxoid, UK) was added. Plates were

incubated at 37 °C overnight. The minimum inhibitory concentration (MIC) was read directly from the strip at the point where the zone of inhibition of bacterial growth intersected with the antibiotic concentration on the strip.

Ethidium bromide accumulation assay

Bacterial cells were grown to mid log phase (OD_{600} 0.4-0.5). Cells were harvested, washed and resuspended in 0.1 M sodium phosphate buffer pH 7 (previously incubated in the VAIN) to OD_{600} 0.2. Cells were then incubated in the VAIN for 15 mins at 37 °C before a 100 µl aliquot was withdrawn to indicate time zero. Ethidium bromide (Sigma, UK) was added to final concentration of 2 µg/ml and fluorescence was measured at 530 nm excitation and 600 nm emission using a plate reader (Molecular Devices M3 plate reader, USA).

Determination of nitrite concentration in the culture supernatants

Determining of nitrite concentration in the culture supernatants was done according to Pitman *et al* (2007)[25] with slight modification. Briefly, an overnight culture of *C. jejuni* grown in brucella broth was diluted to OD_{600} 0.1 in oxygen limited condition. Samples were drawn every hour and centrifuged at 12,000 xg for one minute. Supernatant was then diluted 1:5 with deionized water. 50 µl of diluted culture supernatant was mixed with 850 µl of 1% (w/v) sulphanilamide dissolved in 1 M HCl and 100 µl of 0.02% (w/v) naphthylethylenediamine. After 15 min, the absorbance at 540 nm was measured using plate reader (Molecular Devices M3 plate reader, USA) and nitrite concentrations were determined by reference to a standard curve.

Chicken colonisation experiments

All procedures were conducted under Home Office project licence PPL 60/4420, according to the requirements of the Animal (Scientific Procedures) Act 1986, with the approval of the Ethical Review Committee of the Moredun Research Institute. A

total of 40 White Leghorn chickens, obtained on the day of hatch from a Home Office licensed breeding establishment were used. Chickens were housed in groups of 20 in colony cages under specified-pathogen free conditions. Groups were of mixed sex and were wing-tagged for individual identification. Water and sterile irradiated feed based on vegetable protein (DBM Ltd., UK) was provided *ad libitum*. Fresh faecal samples collected from the cages were plated onto mCCDA plates to confirm the absence of *C. jejuni* in the chickens. At two week of age, chickens were challenged with 10^6 colony-forming units (CFU) of *C. jejuni* 11168H or 11168H *pglB::aphA* in a volume of 100 µl. At one and two weeks after challenge, post-mortem examinations of 10 birds from each group were performed. Contents of each ceacum from each bird were weighed and serially diluted in phosphate-buffered saline (PBS) then plated on mCCDA plates to enumerate viable *Campylobacter* per gram of caecal contents. Statistical analysis was performed using the Mann-Whitney test.

Scanning electron microscopy

C. jejuni cells were resuspended in PBS after being scraped from an overnight culture grown on Columbia based agar plate supplemented with blood and Skirrow supplement. The bacterial suspension was then mixed with fixing buffer (0.1 M sodium cacodylate buffer containing 4% paraformaldehyde, pH 7.4) and fixed on polylysine coated coverslips for overnight at 4 ^{o}C. Samples were then washed twice for 10 minutes in 0.15 M sodium cacodylate buffer pH 7.3 then osmicated for 1 hour at room temperature with 1% osmium tetroxide in 0.15 M cacodylate phosphate buffer. Samples were dehydrated in an ethanol series consisting of: 10 minutes in 10% ethanol, 30 minutes in 70% ethanol and 3 times in 100% ethanol (20 minutes each) followed by transferring the coverslips to 100% ethanol in critical point drying holders then critical point dried using carbon dioxide in a Polaron E3000 critical point

dryer with >3 flushes over 45 minutes soaking period. Coverslips were mounted on stubs (TAAB, UK) with conductive carbon adhesive discs (TAAB) and samples were sputter coated with gold in a Emitech K550 sputter coater (45 mA current, 2 minutes) cells were viewed and images were recorded using a FEI Quanta 200 Field emission scanning electron microscope, (FEI UK Ltd.) operated at 10-20 kV in high vacuum mode.

Acknowledgements

We acknowledge the Wellcome Trust grant 102978/Z/13/Z and Biotechnology & Biological Sciences Research Council grant BB/N001591/1 for funding. We thank Dr Steven Lynham (King's College, London) for his help with isobaric labelling Mass spectrometry and Dr Gema Vizcay-Barrena (King's College, London) for assistance with scanning electron microscopy.

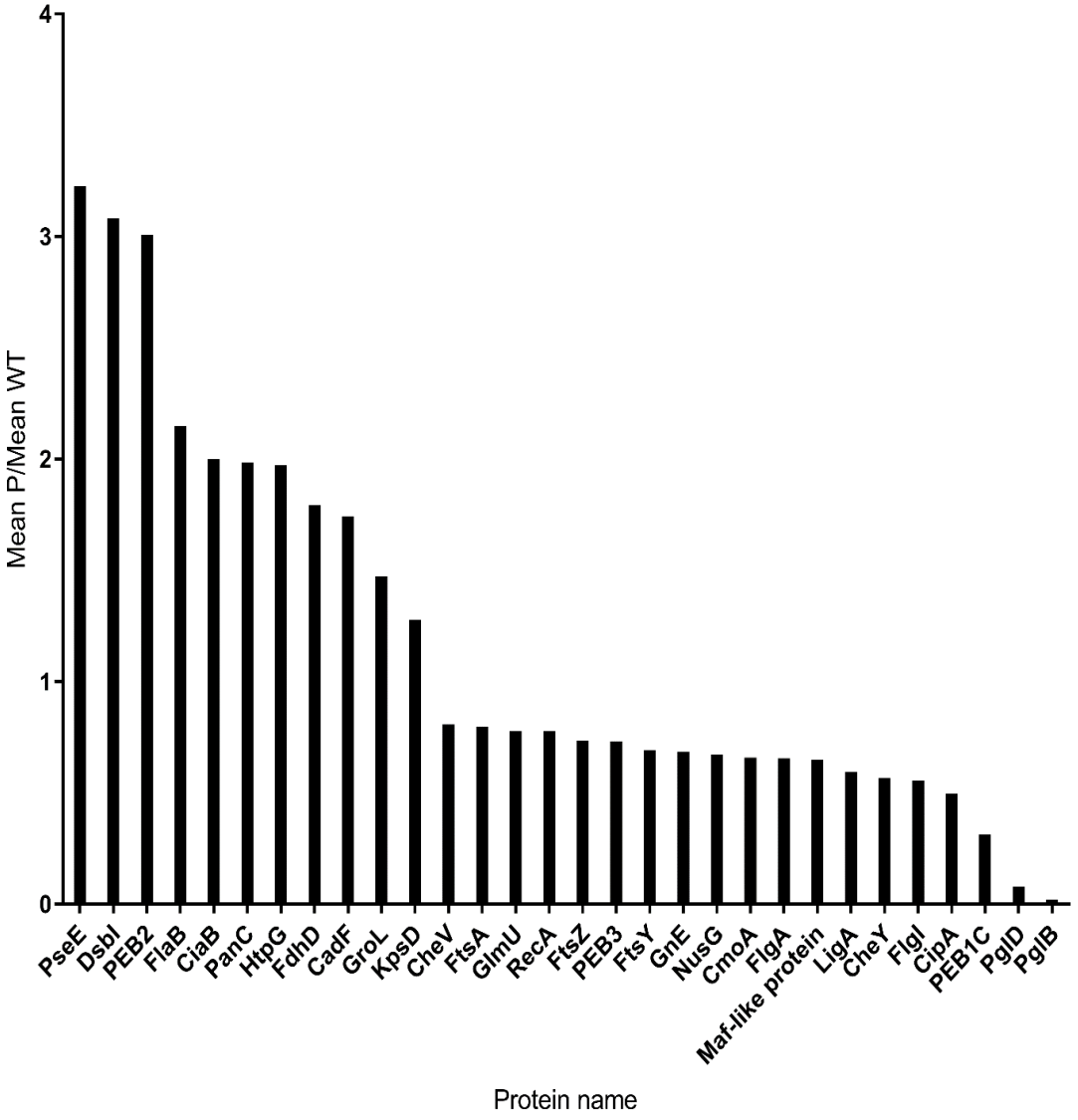
Authors Contributions

S.A. designed the experiments, carried out most of the experiments, data analysis and wrote the manuscript. P.H. and S.N. assisted with the quantitative proteomics. P.V. and C.U. carried out the chicken colonisation experiment and data analysis. M.S., A.D., J.C. and B.W. revised the manuscript. B.W. directed the research.

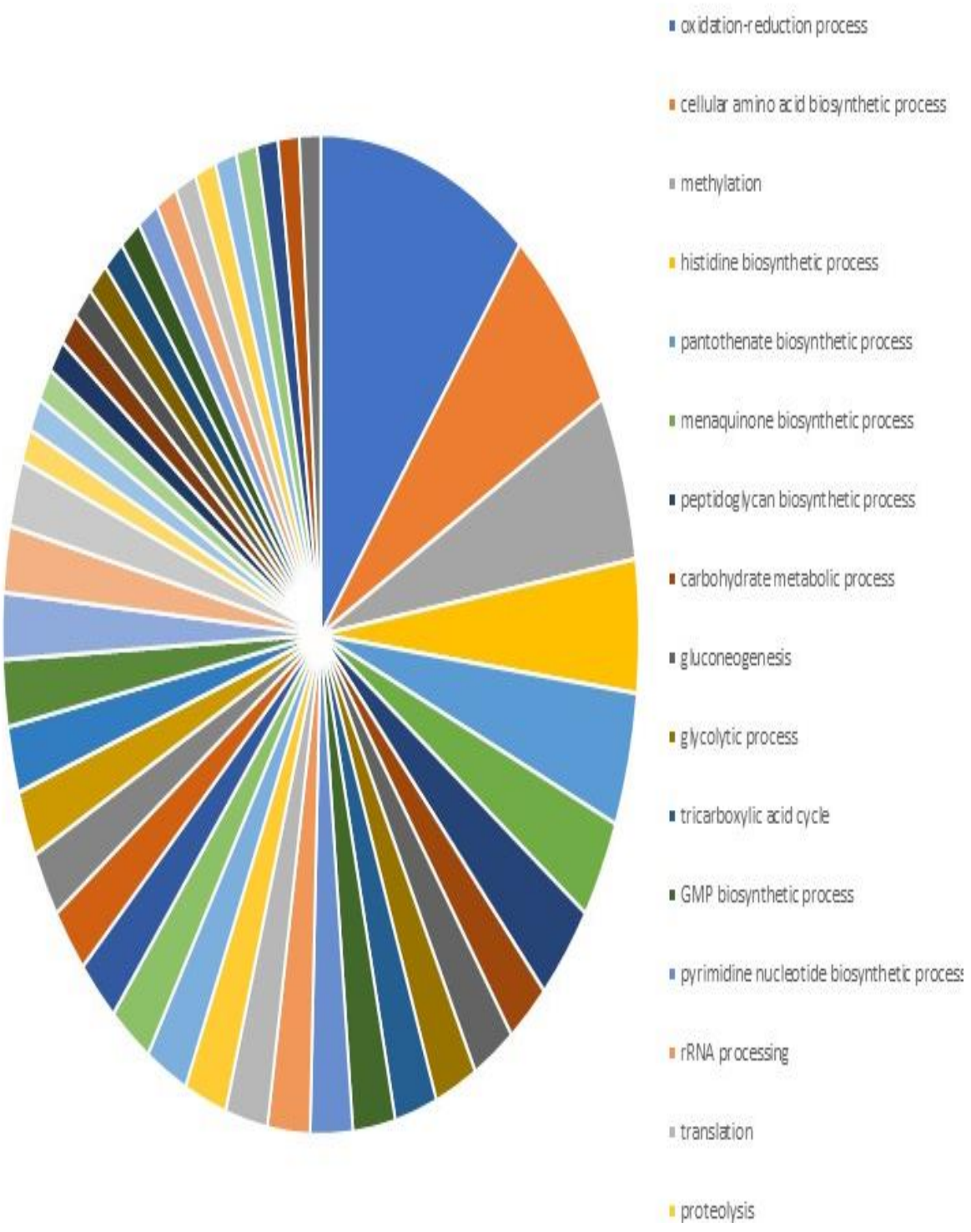
Figures

Fig 1

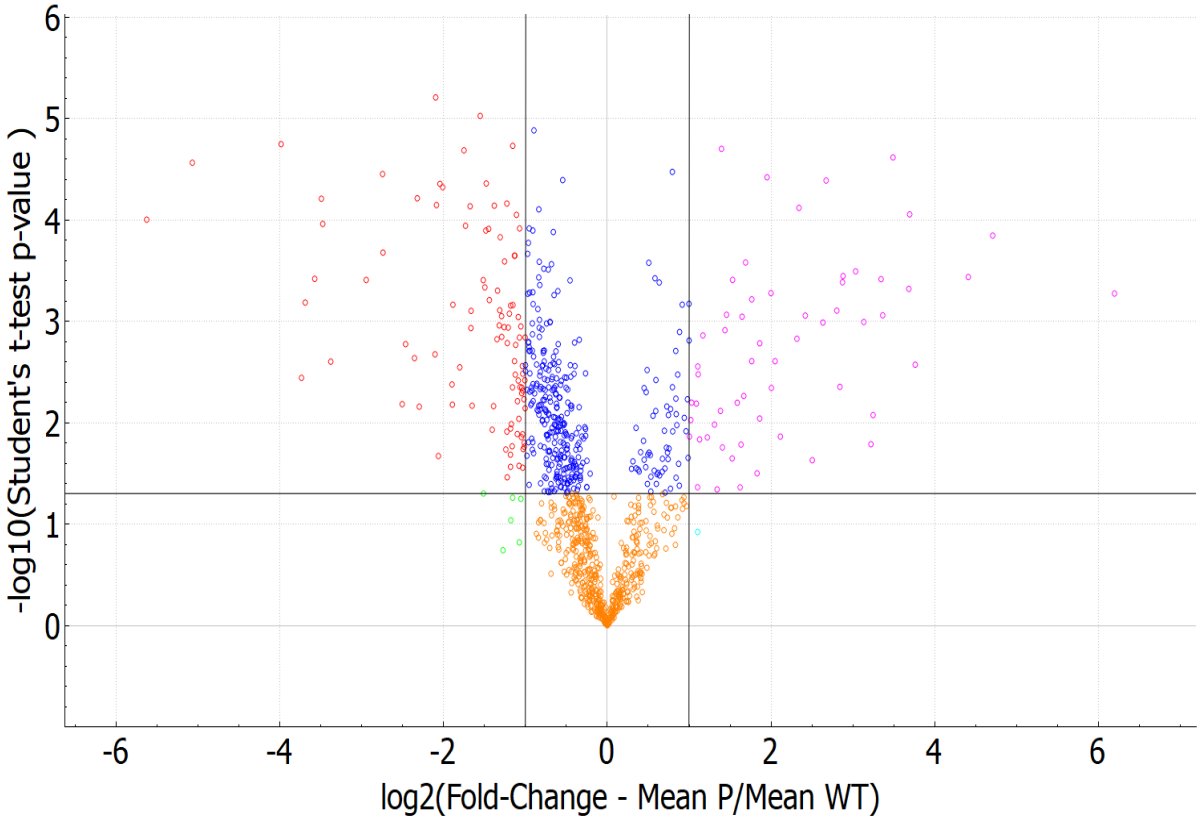
(A)



(B)


Upregulated proteins
oxidation-reduction process
cellular amino acid biosynthetic process
methylation
histidine biosynthetic process
pantothenate biosynthetic process
menaquinone biosynthetic process
peptidoglycan biosynthetic process
carbohydrate metabolic process
gluconeogenesis
glycolytic process
tricarboxylic acid cycle
GMP biosynthetic process
pyrimidine nucleotide biosynthetic process
rRNA processing
translation
proteolysis

(C)


-log10(Student's t-test p-value)
6
5
4
3
2
1
0
-6
-4
-2
0
2
4
6
log2(Fold-Change - Mean P/Mean WT)

(D)

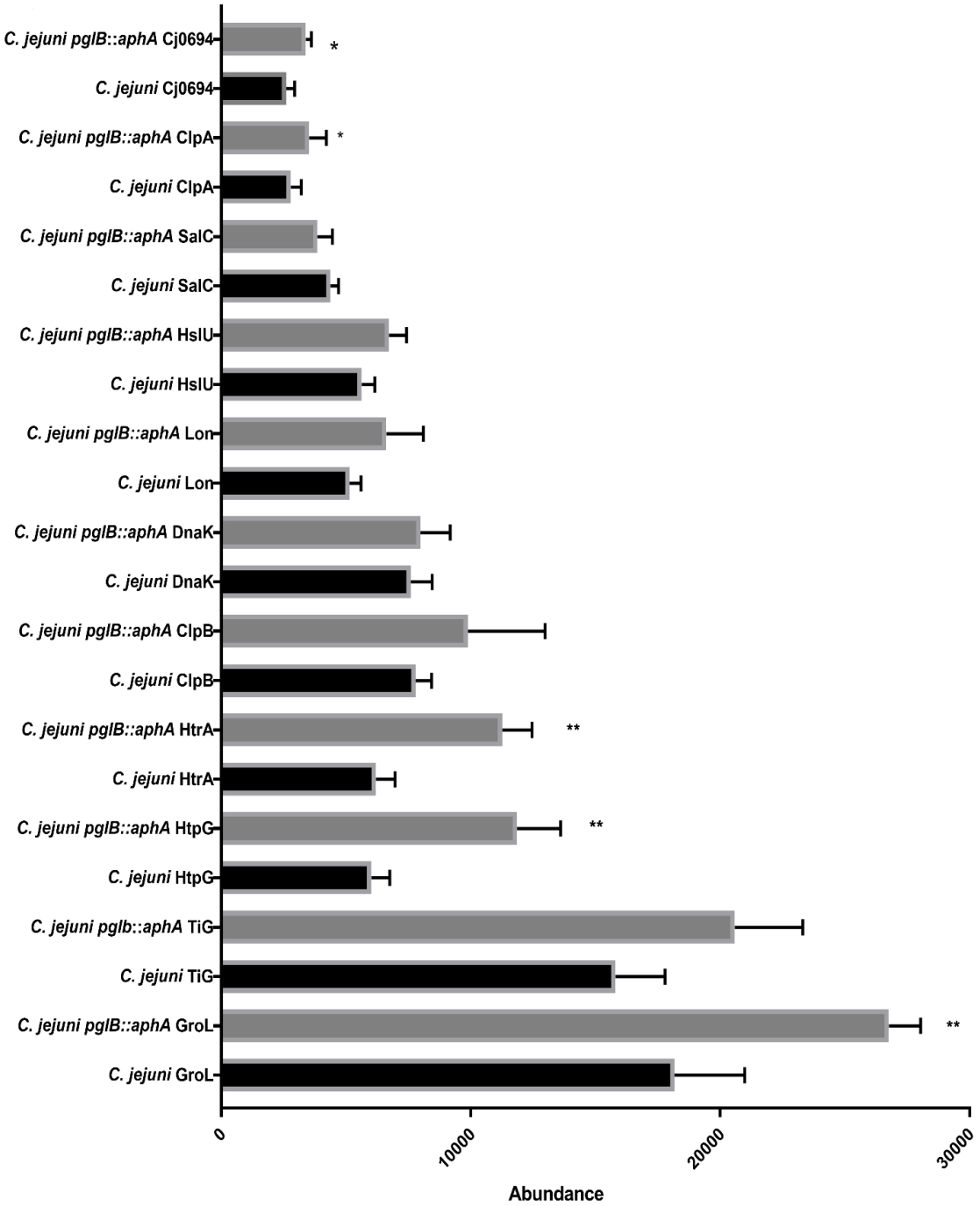
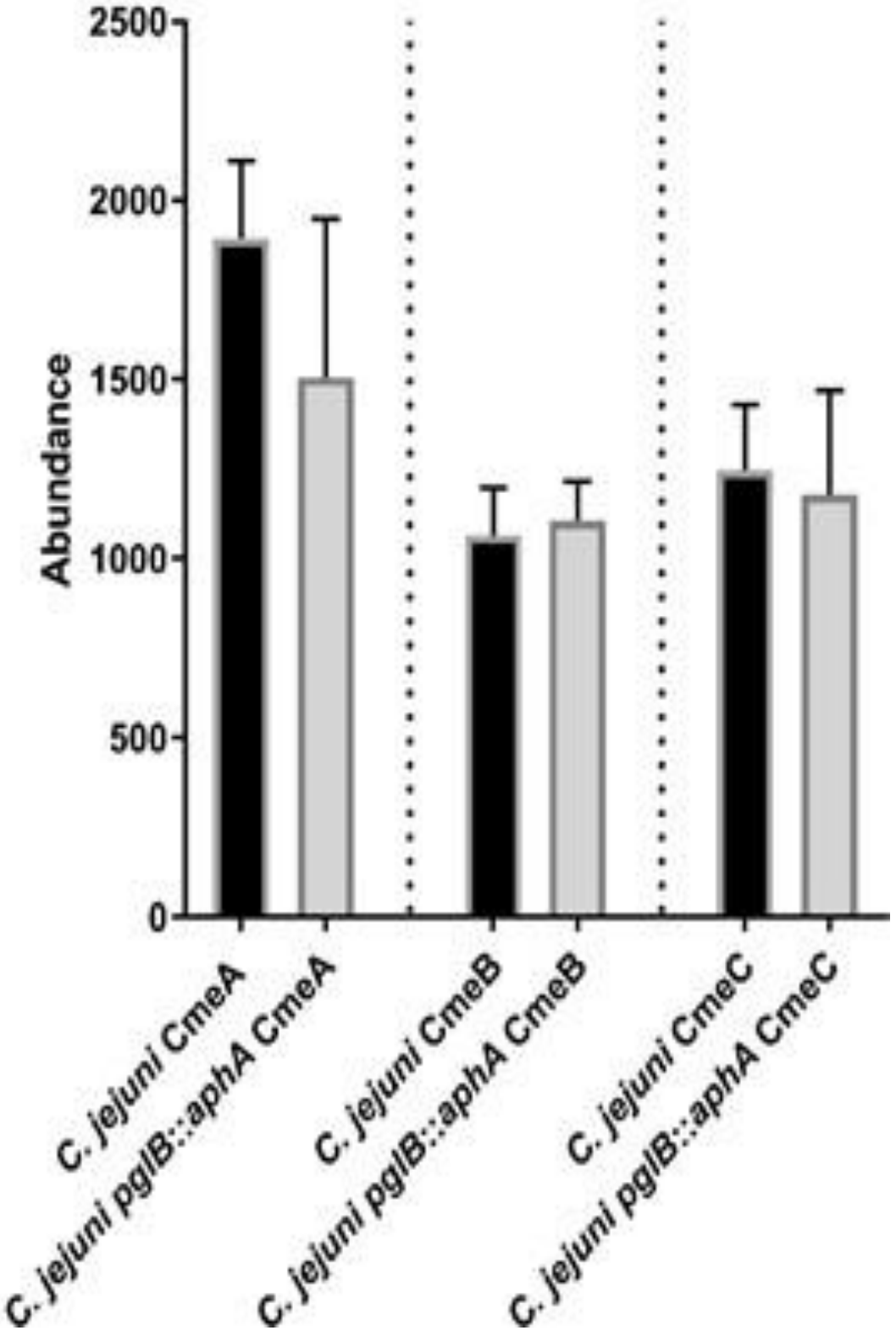

C. jejuni pglB::aphA Cj0694
C. jejuni Cj0694
C. jejuni pglB::aphA ClpA
C. jejuni ClpA
C. jejuni pglB::aphA SalC
C. jejuni SalC
C. jejuni pglB::aphA HslU
C. jejuni HslU
C. jejuni pglB::aphA Lon
C. jejuni Lon
C. jejuni pglB::aphA DnaK
C. jejuni DnaK
C. jejuni pglB::aphA ClpB
C. jejuni ClpB
C. jejuni pglB::aphA HtrA
C. jejuni HtrA
C. jejuni pglB::aphA HtpG
C. jejuni HtpG
C. jejuni pglb::aphA TiG
C. jejuni TiG
C. jejuni pglB::aphA GroL
C. jejuni GroL
*
*
**
**
**
0
10000
20000
30000
Abundance

Fig (1) Quantitative comparison of *C. jejuni* and *C. jejuni pglB::aphA* proteomes. (A) Selected proteins that are statistically differentially expressed between of *C. jejuni* and *C. jejuni pglB::aphA* proteomes. Proteins were analysed by Perseus and presented according to their fold of change with 95% CI. (B) GO analysis of statistically differentially upregulated proteins in *C. jejuni pglB::aphA.* Upregulated proteins were grouped according to their biological function (https://www.ebi.ac.uk/QuickGO)[9] and presented in pie chart according to their percentages. (C) Volcano plot analysis of significantly abundant proteins. The $\log_{10}$ (student's *t*-test) is plotted against $\log_2$ mean fold change of *C. jejuni pglB::aphA,*P *C. jejuni,* WT. The non-axial horizontal line denotes *p*=0.05 while the non-axial vertical lines denote 0.5 and 2-fold change, respectively. The volcano plot was generated using PANDA-view (Chang C, *et al* 2018)[10] (D) Differential expression of chaperons and proteases in *C. jejuni* and *C. jejuni pglB*::*aphA.* Data represents three replicates, error bars represent standard deviation and analysed by student's *t*-test, ***P*<0.05

(A)


2500
2000
1500
1000
500
0
Abundance
C. jejuni CmeA
C. jejuni pglB::aphA CmeA
C. jejuni CmeB
C. jejuni pglB::aphA CmeB
C. jejuni CmeC
C. jejuni pglB::aphA CmeC

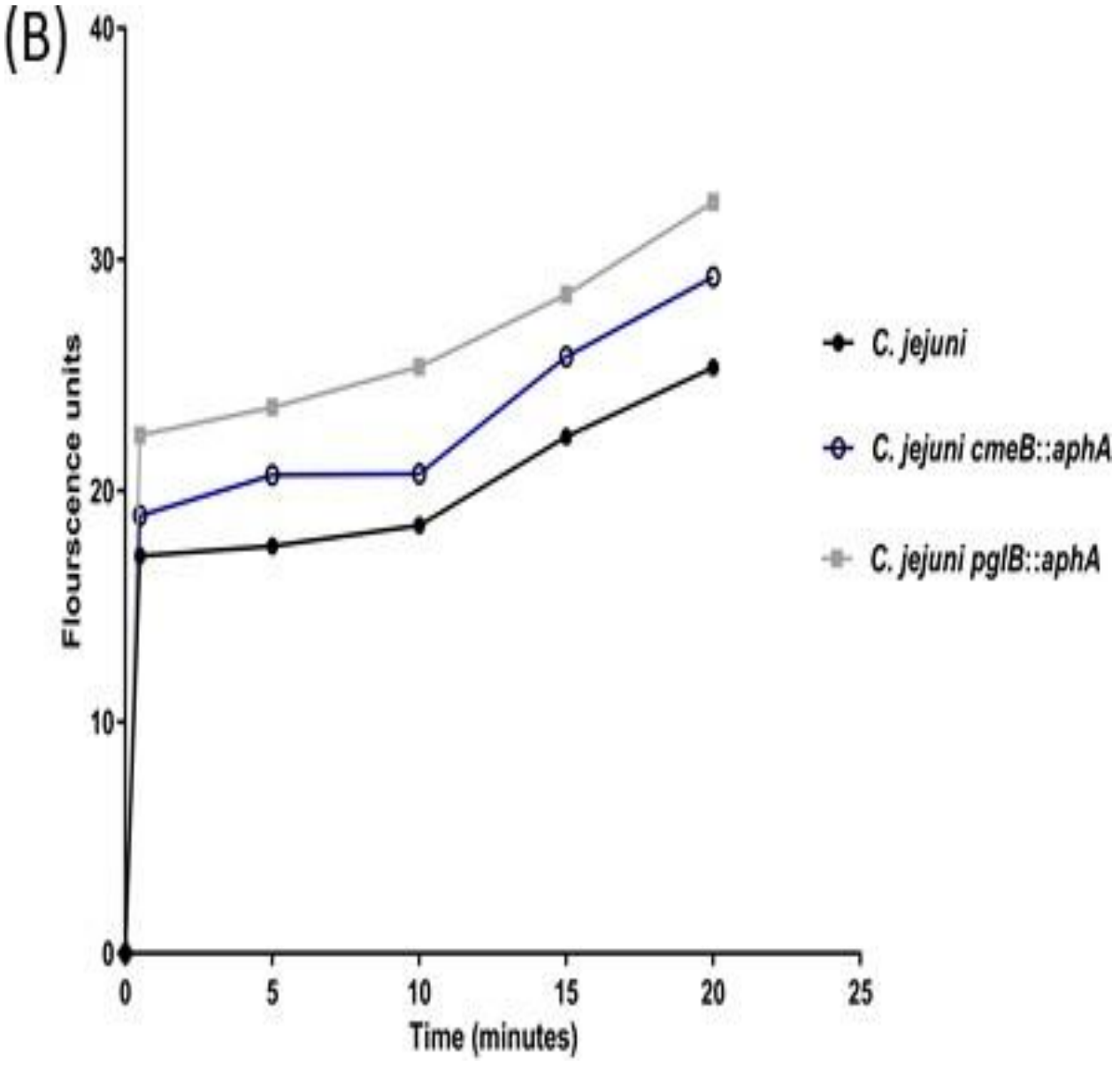

(B)
Flourscence units
Time (minutes)
0
10
20
30
40
0
5
10
15
20
25
C. jejuni
C. jejuni cmeB::aphA
C. jejuni pglB::aphA

Fig (2)

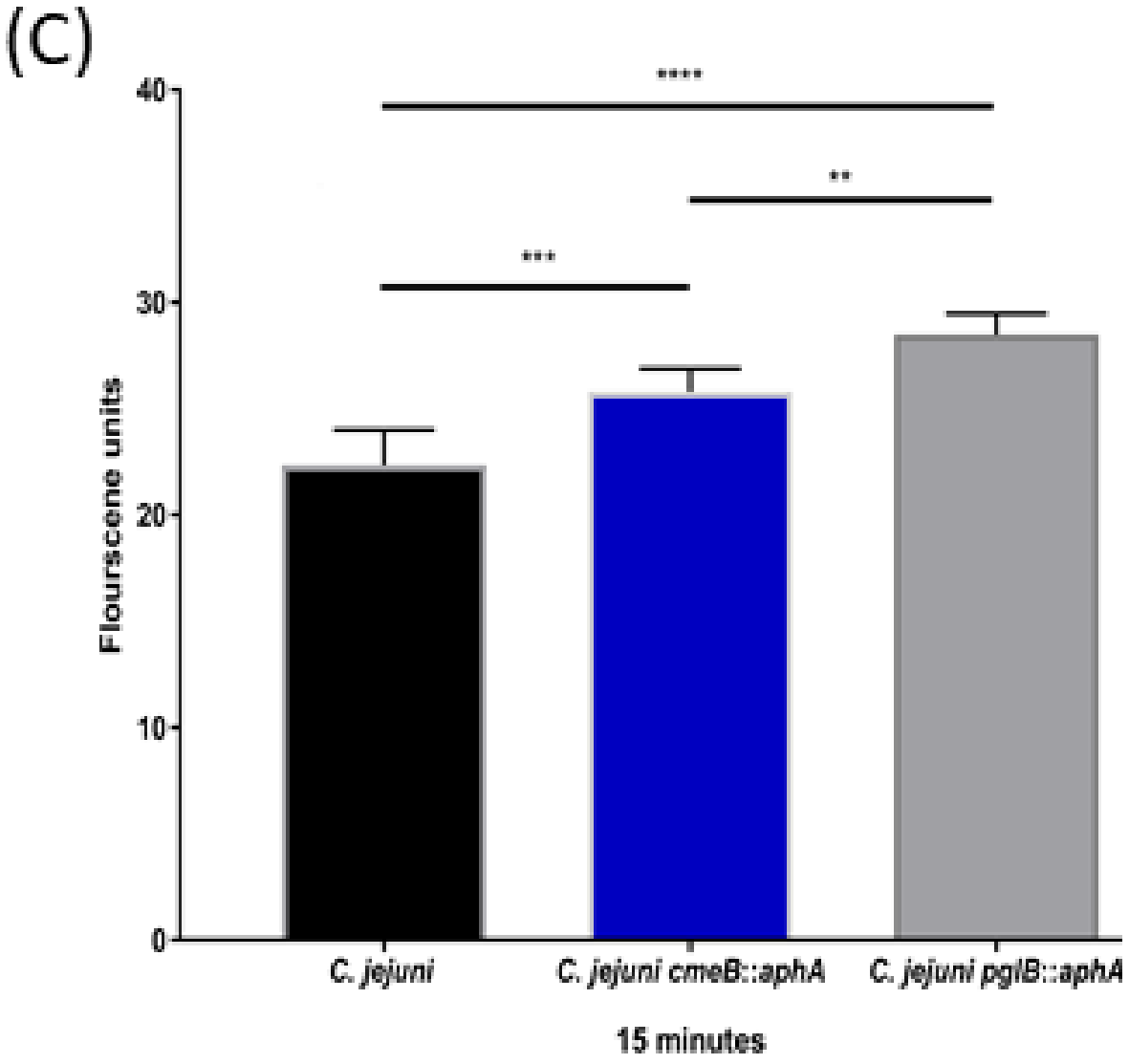
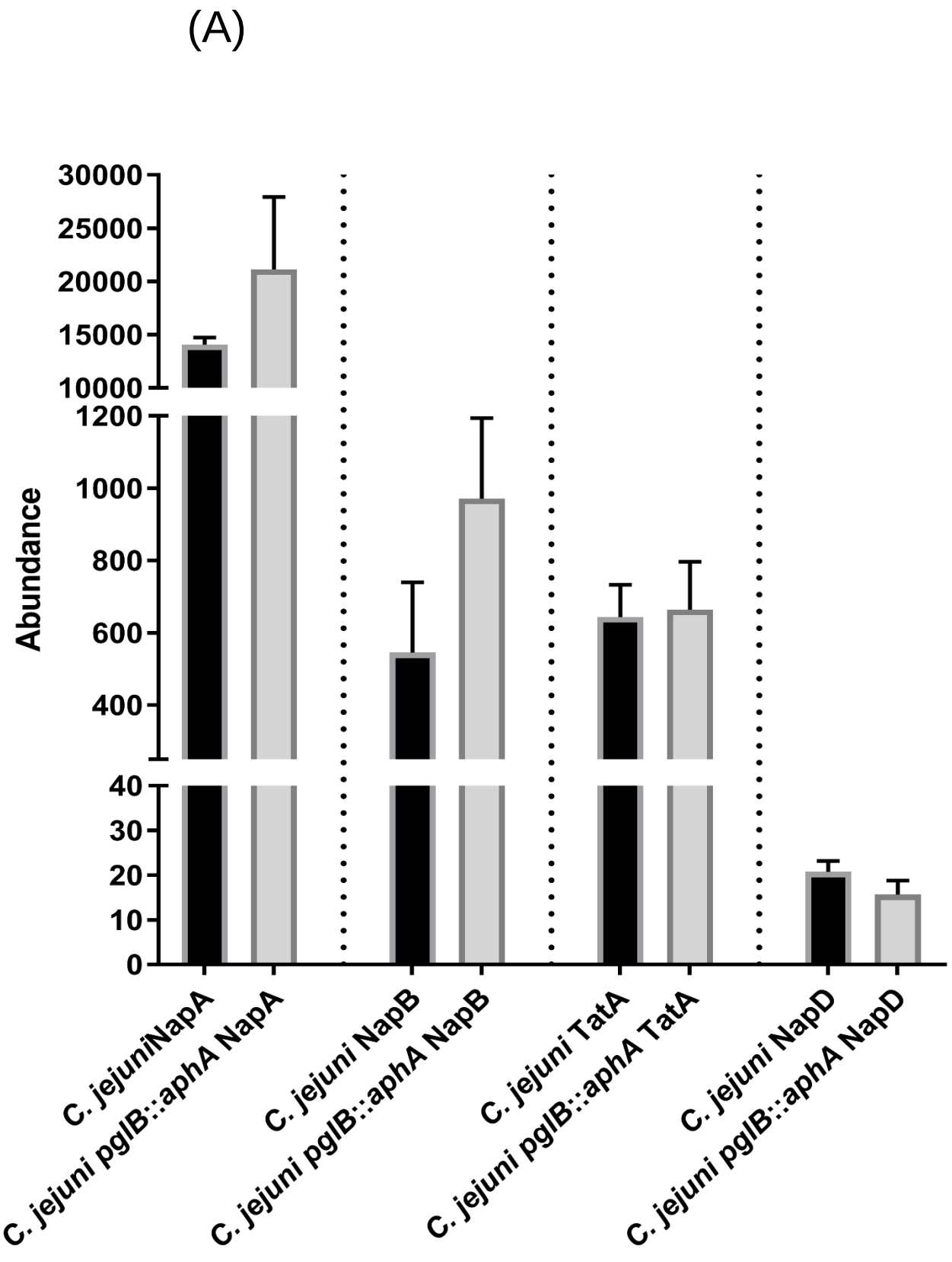


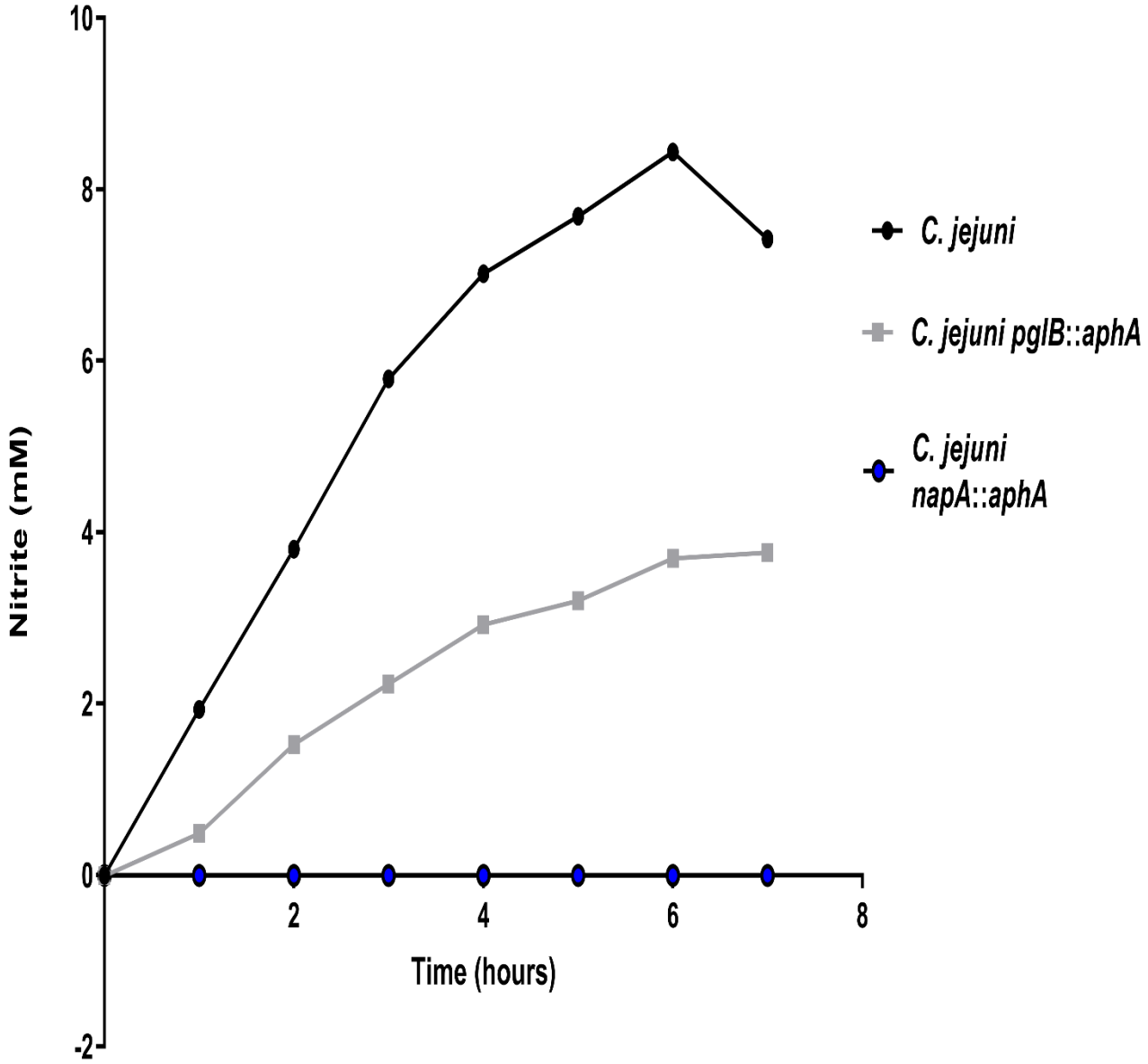
Fig (2) Differential expression and functional analysis of CmeABC in *C. jejuni.* *(*A) Differential expression of CmeABC in *C. jejuni.* (B) Ethidium bromide accumulation test in *C. jejuni* strains. 30 ml Brucella broth was separately inoculated with overnight culture of *C. jejuni* (•); *C. jejuni cmeB*::*aphA* ( ); *C. jejuni pglB*::*aphA* ( ) to OD_{600} 0.1. Cells were grown until OD_{600} 0.4-0.5 was reached, then spun down, washed and resuspended to $O.D_{600}$ 0.2 in 10 mM sodium phosphate buffer pH 7. Cells were then incubated in a VAIN for 15 minutes at 37°c then ethidium bromide was added to final concentration of 0.2 mg/ml. Flourescence

was read at excitation and emission for 20 minutes at 37^{o}c. (A) Ethidium bromide accumulation in *C. jejuni* strains throughout 20 minutes (C) Ethidium bromide accumulation in *C. jejuni* strains at 15 minutes. The data represents the mean of three biological replicates and two technical replicates, error bars represent standard deviation. Significance was calculated using one-way ANOVA test with multiple comparison. **** P <0.0001

Fig (3)


(A)
Abundance
30000
25000
20000
15000
10000
1200
1000
800
600
400
40
30
20
10
0
C. jejuniNapA
C. jejuni pglB::aphA NapA
C. jejuni NapB
C. jejuni pglB::aphA NapB
C. jejuni TatA
C. jejuni pglB::aphA TatA
C. jejuni NapD
C. jejuni pglB::aphA NapD

(B)


10
8
6
4
2
0
-2
Nitrite (mM)
2
4
6
8
Time (hours)
C. jejuni
C. jejuni pglB::aphA
C. jejuni napA::aphA

(C)

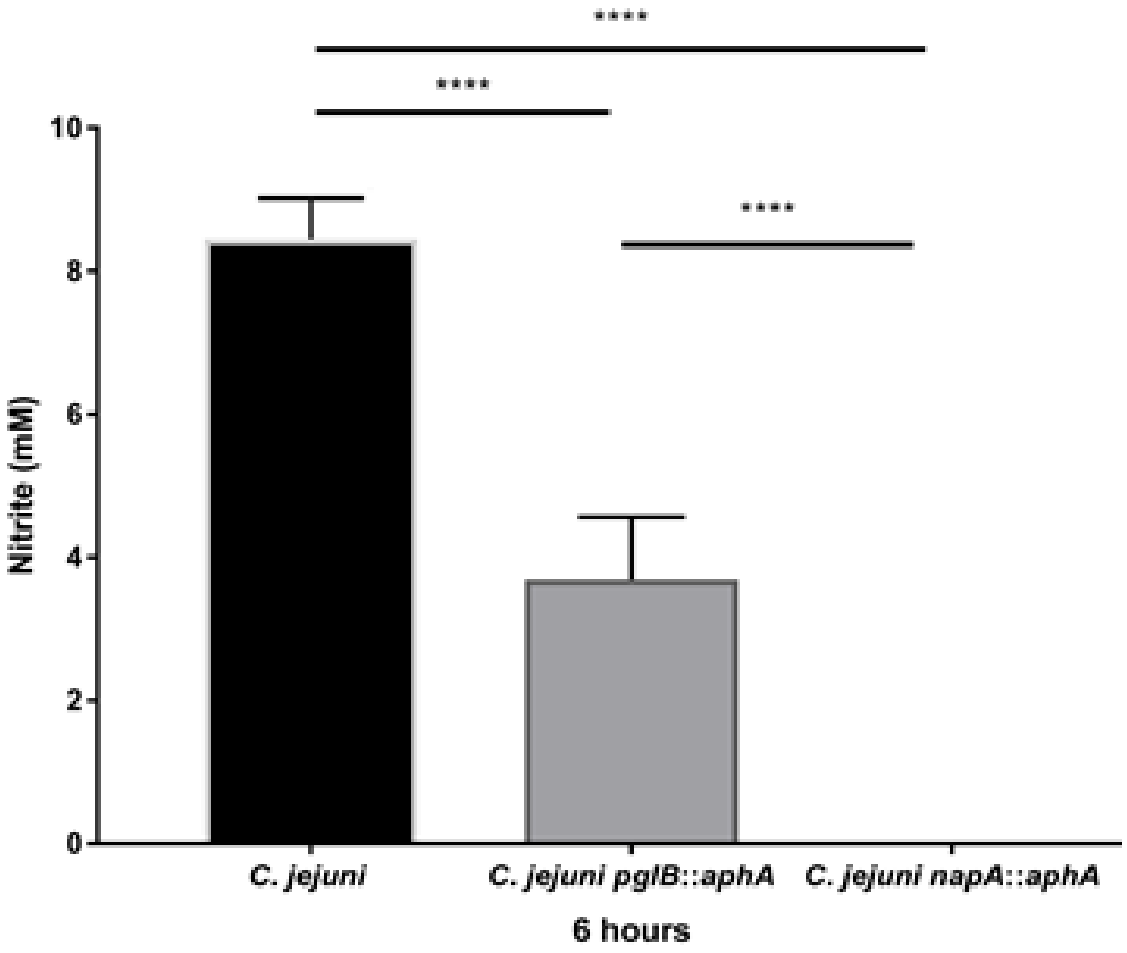


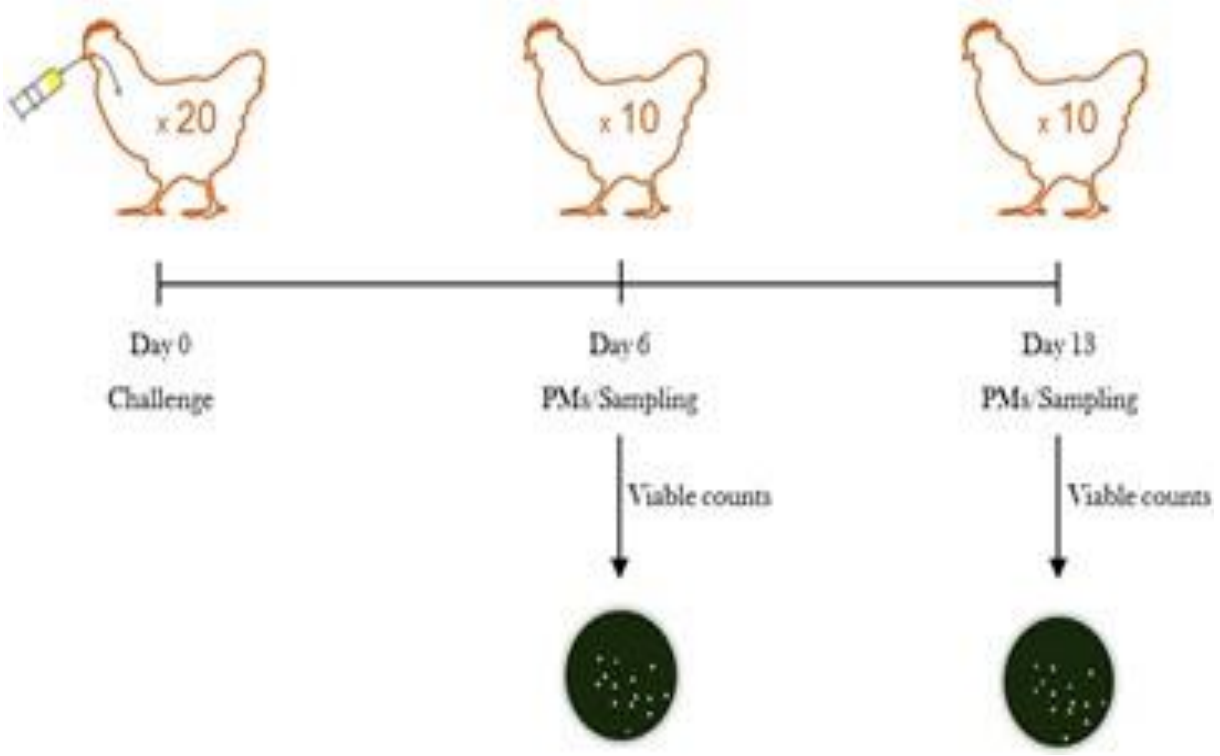
Fig (3) Differential expression and functional analysis of NapAB complex in *C. jejuni*

(A) Differential expression of NapABD and TatA proteins in *C. jejuni* and *C. jejuni pglB*::*aphA (*B) Nitrate reduction in *C. jejuni* strains. 90 ml of Brucella broth supplemented with 200 mM nitrate were inoculated separately with *C. jejuni* (•); *C. jejuni pglB*::*aphA* ( ); *C. jejuni* napA::*aphA* (•) to OD_{600} 0.1 from an overnight culture. Cultures were incubated statically in a VAIN at 37 ^{o}c. Samples were withdrawn every hour, *C. jejuni* cells were removed by centrifugation and nitrate reduction was measured straight from the supernatant against nitrite standards. (A) Nitrite accumulation in the supernatant over 7 hours (B) Nitrite accumulation in the

supernatant at 6 hours. The data represents the mean of three biological replicates and two technical replicates, error bars represent standard deviation. Significance was calculated using one-way ANOVA test with multiple comparison. **** P <0.0001

Fig (4)

(A)


x 20
x 10
x 10
Day 0
Challenge
Day 6
PMs Sampling
Day 13
PMs Sampling
Viable counts
Viable counts

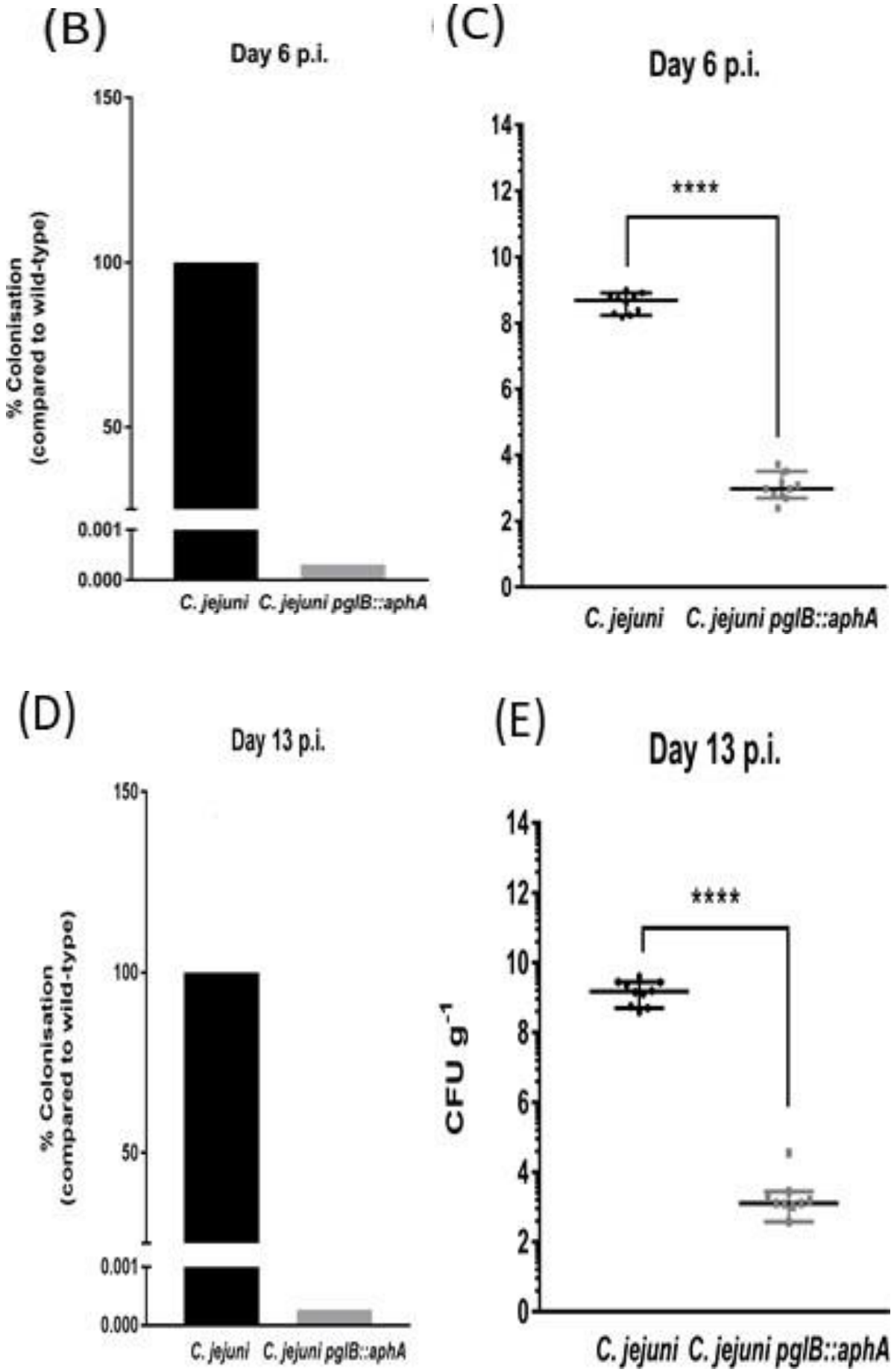

(B)
Day 6 p.i.
% Colonisation (compared to wild-type)
150
100
50
0.001
0.000
C. jejuni
C. jejuni pglB::aphA
(C)
Day 6 p.i.
****
14
12
10
8
6
4
2
0
C. jejuni
C. jejuni pglB::aphA
(D)
Day 13 p.i.
% Colonisation (compared to wild-type)
150
100
50
0.001
0.000
C. jejuni
C. jejuni pglB::aphA
(E)
Day 13 p.i.
****
CFU g-1
14
12
10
8
6
4
2
0
C. jejuni
C. jejuni pglB::aphA

Fig (4) Chicken colonisation of *C. jejuni* and *C. jejuni pglB*::*aphA* in 17 days White Leghorn chickens. (A) Schematic diagram of the chicken colonisation experiment. 17 days old chicken were inoculated with 100 µl of 10^6 CFU of *C. jejuni* (•) or *C. jejuni pglB*::*aphA* ( ). Chickens were then sacrificed at day 6 and day 13 and post-mortem examination was carried out by directly enumerating *C. jejuni* on CCDA plates (B) CFU counts of *C. jejuni* strains on CCDA plates day 6 post inoculation (C) Percentage of *C. jejuni* strains colonisation at day 6 post inoculation (D) CFU counts of *C. jejuni* strains on CCDA plates day 13 post inoculation (E) Percentage of *C. jejuni* strains colonisation at day 13 post inoculation. **** $P<0.0001$ Statistical significance was calculated using Mann-Whitney test.

Supplementary figures/tables

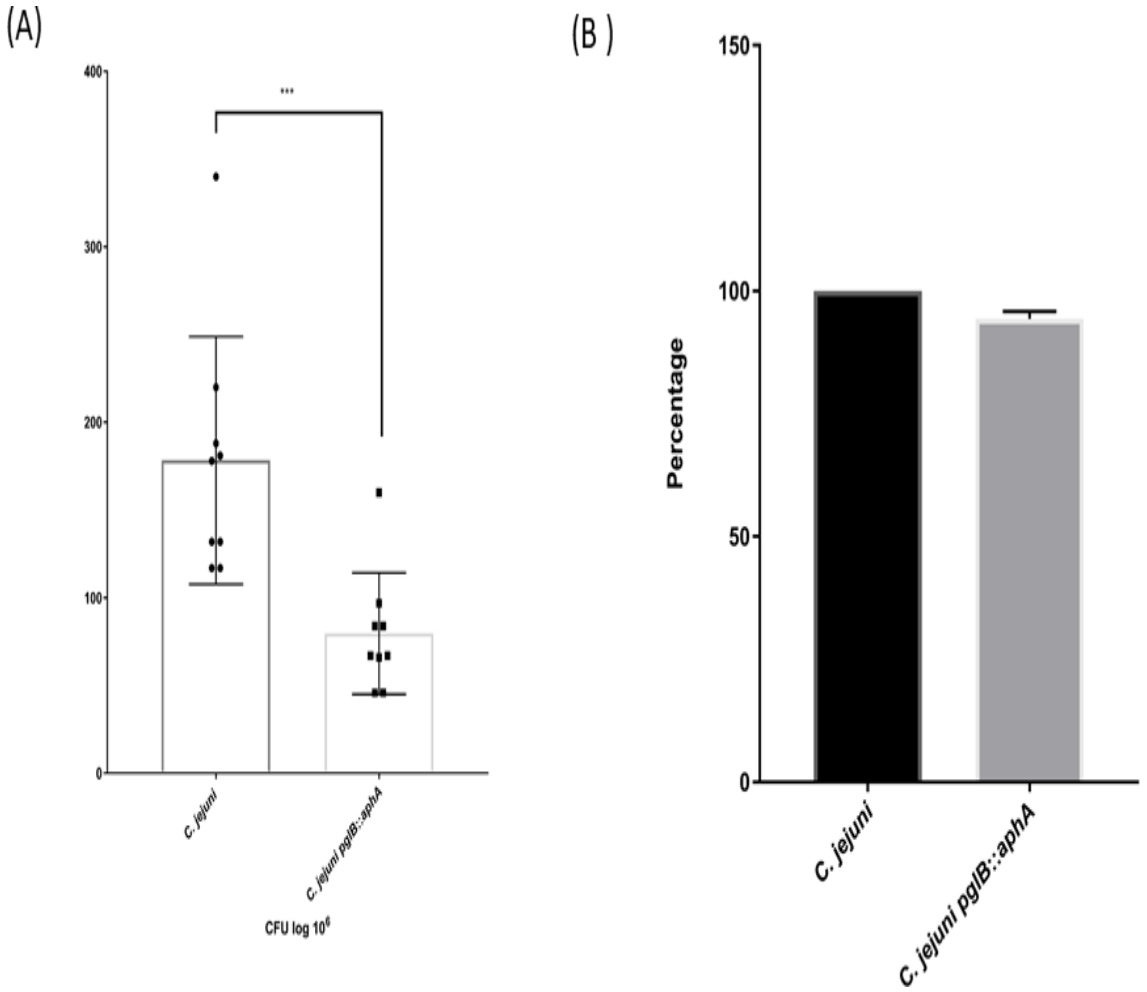
Strain/Plasmid/Primers	Description	Reference
E. coli DH10B	F$^-$ mcrA Δ(*mrr-hsd*RMS-*mcr*BC) Φ80d*lac*ZΔM15 Δ*lac*X74 *end*A1 *rec*A1 *deo*R Δ(*ara*,*leu*)7697 *ara*D139 *gal*U *gal*K *nup*G *rps*L λ$^-$	New England Biolabs, UK
E. coli dam$^-$/dcm$^-$	*ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10)* Tet^S *endA1 rspL136* (Str^R) *dam13::Tn9* (Cam^R) *xylA-5 mtl-1 thi-1 mcrB1 hsdR2*	New England Biolabs, UK
C. jejuni 11168H	Hypermotile variant of *C. jejuni* 11168	9
C. jejuni 11168H *cmeB*::*aphA*	*C. jejuni* 11168H *cmeB* is inactivated by *aphA* encoding kanamycin resistance cassette	This study
C. jejuni 11168H *napA*::*aphA*	*C. jejuni* 11168H *napA* is inactivated by *aphA* encoding kanamycin resistance cassette	This study
pJMK30	*aphA* gene cloned in BamHI restriction site	34
pRSA	*cmeB* amplicon cloned in pJET1.2	This study
pRSF	*aphA* gene cloned in BamHI to inactivate *cmeB* in pRSA	This study
pATT3	*napA* cloned in pJET1.2	This study
pATT3F	*aphA* cloned in ClaI site inactivate *napA* in pATT3	This study
FWDcmeB	GACGTAATGAAGGAGAGCCA	
REVcmeB	CTGATCCACTCCAAGCTATG	
FWDnapA	ACCGCTATTGCAAGTGCTGCTAG	
REVnapA	GAAAGCGGACAAGTCGCATCC	

Supplementary Table 1

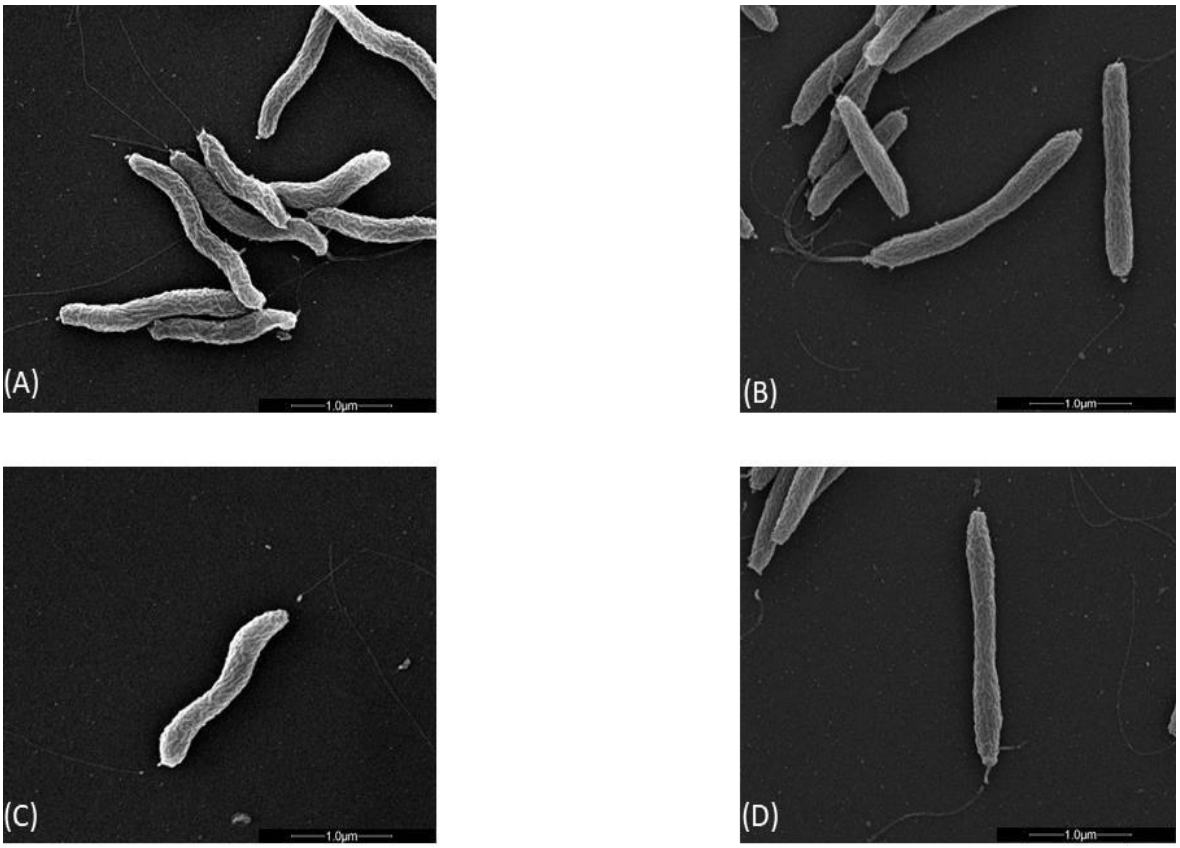
Supp Table (2) Antimicrobial resistance of *C. jejuni* strains

Antimicrobial agent	Agar dilution range (µg/ml)	*C. jejuni* 11168H	*C. jejuni pglB::aphA*	Fold change	*C. jejuni cmeB::aphA*
Ampicillin	0.015-256	1	0.5	2	0.015
Ciprofloxacin	0.002-32	0.12	0.06	2	0.04
Erythromycin	0.015-256	0.25	0.12	2	0.015
Tetracycline	0.015-256	0.12	0.03	4	0.015

Supp Table (2) Antimicrobial resistance of *C. jejuni* strains. Etest strips were used to assess minimum inhibitory concentration (MIC) of *C. jejuni* strains. The data represents the mean from three biological replicates, two technical replicates each.



Supp figure (1) Bacterial cell count in *C. jejuni* at OD_{600}=0.4-0.5. A) Cell enumeration at log 10^6 of *C. jejuni* and *C. jejuni pglB*::*aphA.* B) Dry cell weight of *C. jejuni* and *C. jejuni pglB*::*aphA.* Data shows the mean of three biological replicates two technical replicates each, error bars represent standard deviation. Significance was calculated using student's *t*-test. *P*<0.0005



Supp figure (2) Scanning electron micrographs of *C. jejuni* strains. A.

Scanning electron micrographs of *C. jejuni* 11168H wild-type; B. *C. jejuni*

11168H *pglB::aphA*;

C. A single bacterial cell of Scanning electron micrographs of *C. jejuni*

11168H wild- type; D. A single bacterial cell of *C. jejuni* 11168H *pglB*::*aphA*.

The bar represents 1.0 µm.

References

1. Varki, A. Biological roles of glycans. *Glycobiology* (2017). doi:10.1093/glycob/cww086

2. Linton, D., Allan, E., Karlyshev, A. V., Cronshaw, A. D. & Wren, B. W. Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in Campylobacter jejuni. *Mol. Microbiol.* (2002). doi:10.1046/j.1365-2958.2002.02762.x

3. Vik, A. *et al.* Broad spectrum O-linked protein glycosylation in the human pathogen Neisseria gonorrhoeae. *Proc. Natl. Acad. Sci. U. S. A.* (2009). doi:10.1073/pnas.0809504106

4. Lithgow, K. V. *et al.* A general protein O-glycosylation system within the Burkholderia cepacia complex is involved in motility and virulence. *Mol. Microbiol.* (2014). doi:10.1111/mmi.12540

5. Fletcher, C. M., Coyne, M. J., Villa, O. F., Chatzidaki-Livanis, M. & Comstock, L. E. A General O-Glycosylation System Important to the Physiology of a Major Human Intestinal Symbiont. *Cell* (2009). doi:10.1016/j.cell.2009.02.041

6. Scott, N. E. *et al.* Simultaneous glycan-peptide characterization using hydrophilic interaction 2 chromatography and parallel fragmentation by CID, HCD and ETD-MS applied to the N- 3 linked glycoproteome of Campylobacter jejuni. *Am. Soc. Biochem. Mol. Biol.* (2010). doi:M000031-MCP201 [pii]\n10.1074/mcp.M000031-MCP201

7. Nothaft, H. *et al.* Diversity in the protein N-glycosylation pathways within the Campylobacter genus. *Mol. Cell. Proteomics* (2012). doi:10.1074/mcp.M112.021519

8. Szymanski, C. M., Burr, D. H. & Guerry, P. Campylobacter protein glycosylation affects host cell interactions. *Infect. Immun.* (2002). doi:10.1128/IAI.70.4.2242-2244.2002

9. Jones, M. A. *et al.* Adaptation of Campylobacter jejuni NCTC11168 to high-level colonization of the avian gastrointestinal tract. *Infect. Immun.* (2004). doi:10.1128/IAI.72.7.3769-3776.2004

10. Larsen, J. C., Szymanski, C. & Guerry, P. N-linked protein glycosylation is required for full competence in Campylobacter jejuni 81-176. *J. Bacteriol.* (2004). doi:10.1128/JB.186.19.6508-6514.2004

11. Binns, D. *et al.* QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp536

12. Chang, C. *et al.* PANDA-view: An easy-to-use tool for statistical analysis and visualization of quantitative proteomics data. *Syst. Biol. (Stevenage).* (2018). doi:10.1093/bioinformatics/bty408/5001389

13. Fenton, W. A. & Horwich, A. L. Chaperonin-mediated protein folding: Fate of substrate polypeptide. *Quarterly Reviews of Biophysics* (2003). doi:10.1017/S0033583503003883

14. Nakamoto *et al.* Physical interaction between bacterial heat shock protein (Hsp) 90 and Hsp70 chaperones mediates their cooperative action to refold denatured proteins. *J. Biol. Chem.* (2014). doi:10.1074/jbc.M113.524801

15. Merdanovic, M., Clausen, T., Kaiser, M., Huber, R. & Ehrmann, M. Protein Quality Control in the Bacterial Periplasm. *Annu. Rev. Microbiol.* (2011).

doi:10.1146/annurev-micro-090110-102925

16. Taylor, A. J., Zakai, S. A. I. & Kelly, D. J. The periplasmic chaperone network of Campylobacter jejuni: Evidence that SalC (Cj1289) and PpiD (Cj0694) are involved in maintaining outer membrane integrity. *Front. Microbiol.* (2017). doi:10.3389/fmicb.2017.00531

17. Boehm, M., Lind, J., Backert, S. & Tegtmeyer, N. Campylobacter jejuni serine protease HtrA plays an important role in heat tolerance, oxygen resistance, host cell adhesion, invasion, and transmigration. *Eur. J. Microbiol. Immunol. (Bp).* (2015). doi:10.1556/EUJMI-D-15-00003

18. Andersen, M. T. *et al.* Diverse roles of HspR in Campylobacter jejuni revealed by the proteome, transcriptome and phenotypic characterization of an hspR mutant. *Microbiology* (2005). doi:10.1099/mic.0.27513-0

19. Butler, S. M., Festa, R. A., Pearce, M. J. & Darwin, K. H. Self-compartmentalized bacterial proteases and pathogenesis. *Molecular Microbiology* (2006). doi:10.1111/j.1365-2958.2006.05128.x

20. Jayaprakash, N. G. & Surolia, A. Role of glycosylation in nucleating protein folding and stability. *Biochem. J.* (2017). doi:10.1042/BCJ20170111

21. Hanson, S. R. *et al.* The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc. Natl. Acad. Sci. U. S. A.* (2009). doi:10.1073/pnas.0810318105

22. Biswas, H. & Chattopadhyaya, R. Stability of curcuma longa rhizome lectin: Role of n-linked glycosylation. *Glycobiology* (2016). doi:10.1093/glycob/cwv110

23. Linton, D. *et al.* Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway. *Mol. Microbiol.* (2005). doi:10.1111/j.1365-

2958.2005.04519.x

24. Lin, J., Overbye Michel, L. & Zhang, Q. CmeABC functions as a multidrug efflux system in Campylobacter jejuni. *Antimicrob. Agents Chemother.* (2002). doi:10.1128/AAC.46.7.2124-2131.2002

25. Pittman, M. S. *et al.* Growth of Campylobacter jejuni on nitrate and nitrite: Electron transport to NapA and NrfA via NrfH and distinct roles for NrfA and the globin Cgb in protection against nitrosative stress. *Mol. Microbiol.* (2007). doi:10.1111/j.1365-2958.2006.05532.x

26. Pumbwe, L. & Piddock, L. J. Identification and molecular characterisation of CmeB, a Campylobacter jejuni multidrug efflux pump. *FEMS Microbiol. Lett.* (2002). doi:10.1111/j.1574-6968.2002.tb11007.x

27. Lin, J., Sahin, O., Michel, L. O. & Zhang, Q. Critical role of multidrug efflux pump CmeABC in bile resistance and in vivo colonization of Campylobacter jejuni. *Infect. Immun.* (2003). doi:10.1128/IAI.71.8.4250-4259.2003

28. Weingarten, R. A., Grimes, J. L. & Olson, J. W. Role of Campylobacter jejuni respiratory oxidases and reductases in host colonization. *Appl. Environ. Microbiol.* (2008). doi:10.1128/AEM.02261-07

29. De Vries, S. P. W. *et al.* Genome-wide fitness analyses of the foodborne pathogen Campylobacter jejuni in in vitro and in vivo models. *Sci. Rep.* (2017). doi:10.1038/s41598-017-01133-4

30. Iwashkiw, J. A. *et al.* Identification of a general O-linked protein glycosylation system in Acinetobacter baumannii and its role in virulence and biofilm formation. *PLoS Pathog.* (2012). doi:10.1371/journal.ppat.1002758

31. Akiba, M., Lin, J., Barton, Y. W. & Zhang, Q. Interaction of CmeABC and CmeDEF

in conferring antimicrobial resistance and maintaining cell viability in Campylobacter jejuni. *J. Antimicrob. Chemother.* (2006). doi:10.1093/jac/dki419

32. Jervis, A. J. *et al.* Characterization of the structurally diverse N-linked glycans of Campylobacter species. *J. Bacteriol.* (2012). doi:10.1128/JB.00042-12

33. Karlyshev, A. V., Linton, D., Gregson, N. A. & Wren, B. W. A novel paralogous gene family involved in phase-variable flagella-mediated motility in Campylobacter jejuni. *Microbiology* (2002). doi:10.1099/00221287-148-2-473

34. Van Vliet, A. H. M., Wooldridge, K. G. & Ketley, J. M. Iron-responsive gene regulation in a Campylobacter jejuni fur mutant. *J. Bacteriol.* (1998).

Chapter 4

N-LINKED GLYCANS CONFER PROTEIN STABILITY AND MODULATE MULTIDRUG EFFLUX PUMP ASSEMBLY

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Sherif Abouelhadid
Principal Supervisor	Brendan W Wren
Thesis Title	Investigating the biological role of *N*-linked glycan in bacteria

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Cell chemical biology
Please list the paper's authors in the intended authorship order:	Sherif Abouelhadid, John Raynes, Matthew Colledge, Tam Bui, Andy Miles, Bonnie Wallace, Jon Cuccui, Brendan W Wren
Stage of publication	To be submitted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	The candidate designed the experiments, carried out most of the experiments, analysed the data and wrote the manuscript.

Student Signature: **Date:**

Supervisor Signatur [redacted] **Date: 14/Sept/2018**

_N_-linked glycans confer protein stability and modulate multidrug efflux pump assembly

Sherif Abouelhadid[1], John Raynes[1], Matthew Colledge[2], Tam Bui[3], Andy Miles,[2]

Bonnie Wallace[2], Jon Cuccui[1], Brendan W Wren[1]

[1]Department of Pathogen Biology, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

[2]Institute of Structural and Molecular Biology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

[3]Biomolecular Spectroscopy Centre, King's College London, The Wolfson Wing, Hodgkin, Building, SE1 1UL, UK

Abstract

It is now apparent that nearly all bacteria have at least a single glycosylation system, but the direct effect of these modifications on glycoprotein are unresolved. In this study we used the general *N*-linked glycosylation pathway from *Campylobacter jejuni* to investigate the biophysical roles of protein modification on the CmeABC multidrug efflux pump complex. The study reveals the multifunctional role of *N*-linked glycans in enhancing protein thermostability, stabilising protein complexes and the promotion of protein-protein interaction. Our findings demonstrate, for the first time, that regardless of glycan diversification among domains of life, *N*-linked glycans confer a common evolutionary intrinsic role.

Introduction

Glycosylation is a prevalent protein post translational modification in all domains of life. The attachment of a carbohydrate moiety to certain amino acid side chains in proteins was traditionally thought to be exclusive to eukaryotes and archaea. In the last decade, protein glycosylation, both *N* and *O*-linked, has been increasingly reported in pathogenic bacteria such as well as commensal bacteria[1]. Functional analysis of the role of *N*-glycans in eukaryotes shows multiple roles. Intrinsically, glycans confer protein stabilisation through a plethora of mechanisms such as, promoting secondary structure formation, accelerating protein folding, slowing the unfolding rate[2], enhancing protein thermostability[3] and reducing aggregation[4]. Extrinsically, *N*-linked glycosylation modulates protein-protein interactions and protein targeting[5]. Notably, prokaryotes and eukaryotes share similar biosynthetic pathways of general *N*-linked glycosylation, indicating a common evolutionary pathway[6]. However, the role of *N*-linked glycans remains poorly studied in prokaryotes.

The first discovery of an *N*-linked glycosylation in a bacterium was in the enteric pathogen *Campylobacter jejuni* which now has one of the most studied of all prokaryotic glycosylation pathways[7]. Genomic and Proteomic studies have demonstrated that, PglB is an oligosaccharyltransferase responsible for the covalent attachment of *N*-linked glycan (GalNAc -α1,4-GalNAc-α1,4-GalNAc-[Glcβ1,3-]GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-β1; where GalNAc is *N*-acetylgalactosamine; Glc is glucose; diBacNAc is 2,4-diacetamido -2,4,6-trideoxyglucopyronose) to the asparagine residue in the acceptor sequon D/E-X-N-Y-S/T where X and Y are any amino acid except proline[7]. In depth studies on *N*-linked glycans diversity have revealed that there are 16 different glycans sequences present in *Campylobacter* genus. Surprisingly, the first two monosaccharides

(diBacNAc and HexNAc) were conserved in all species[8,9]. Whilst the role of *N*-linked glycan remains to be elucidated, disruption of the *N*-linked glycosylation pathway resulted in pleotropic effects such as decreased chicken colonisation[10], reduced adherence to intestinal cells[11] as well as impaired bacterial competence[12]. These studies did not ascertain the direct role of *N*-linked glycans in the *C. jejuni* glycoproteome. To address this, we sought to investigate the biophysical role of *N*-linked glycosylation on a representative glycoprotein. We demonstrate a critical role for glycosylation by focusing on a resistance-nodulation-division type (RND) multidrug efflux pump denoted *Campylobacter* multidrug efflux; CmeABC.

CmeABC is a tripartite molecular assembly of glycoproteins, CmeB; an inner membrane multidrug transport protein, CmeA; a periplasmic fusion protein and CmeC; an outer membrane associated channel. It plays a pivotal role in chicken colonization[13], as well as being responsible for multidrug resistance (MDR)[14]. Previously, we demonstrated that disrupting *pglB* in *C. jejuni* impaired the efflux activity of CmeABC resulting in significantly higher ethidium bromide accumulation when compared to the wildtype (Abouelhadid S, *et al* submitted manuscript). The absence of glycosylation on the CmeABC locus within *C. jejuni* was also shown to reduce resistance to four different antibiotic classes. Here in, we show that the loss of *N*-linked glycans in CmeABC is the sole reason for this phenotype and not a pleiotropic effect caused by *pglB* disruption. We also unravel the intrinsic role of *N*-linked glycans in a) modulating global protein structure b) enhancing glycosylated CmeA; g2CmeA thermostability and c) significantly slowing the unfolding rate of g2CmeA. Finally, we evaluate the extrinsic role of *N*-linked glycans in the molecular assembly of CmeABC to discern the difference in binding kinetics of CmeA variants

to CmeC. The study highlights the multifunctional role of *N*-linked glycans in enhancing protein thermostability, stabilising protein complexes and the promotion of protein-protein interaction. Our findings demonstrate, for the first time, that regardless glycans diversification among domains of life, *N*-linked glycans seem to confer a common evolutionary intrinsic role. Here we present a model with tractable phenotype to be used in studying glycan evolution, function and diversity.

Results

N-linked glycans do not affect CmeABC protein expression nor protect CmeC from native proteolytic degradation.

The genome sequencing of *C. jejuni* 11168 revealed that the bacterium possess 13 multidrug efflux transporters, which are common to most *C. jejuni* strains[15]. Genetic and biochemical testing demonstrated that *cmeABC* is located in an operon and encodes the predominant multidrug efflux pump in *C. jejuni.* In addition to its central role in extruding structurally non-related compounds such as antimicrobials, bile salts, dyes and heavy metals[14], CmeABC has been reported to function interactively with CmeDEF[16]; a secondary multidrug efflux pump in *C. jejuni.* Glycoproteomic analysis of *C. jejuni* demonstrated that the CmeABC complex is multi *N*-linked glycosylated where, CmeA is glycosylated at positions ^{121}DFNR^{124} and ^{271}DNNNS^{275}, CmeB is glycosylated at position ^{634}DRNVS^{648} and CmeC is glycosylated at position ^{47}ETNSS^{51}. Notably, CmeE has also been shown to be *N*

-glycosylated[17]. Previously, we showed that the multidrug efflux pump is impaired in glycosylation deficient *C. jejuni*. This resulted in a significant increase in ethidium bromide accumulation more than the parent strain and an reduction in antibiotics resistance (Abouelhadid S, *et al* submitted manuscript). We hypothesised that this deficiency may be due to assembly destabilisation as a consequence of glycosylation

aberration. To address this question, we sought to study the major multidrug efflux pump of *C. jejuni*, in a glycosylation deficient CmeABC complex. We conducted the experiments in a *C. jejuni cmeD*::*cat*^{r} background strain in order to avoid the interaction of CmeDEF with CmeABC that might mask the functional role of *N*-linked glycans. This parent strain was used to construct a *C. jejuni* WTCmeABC strain and a *C. jejuni* glycosylation altered strain; g0CmeABC whereby, N->Q in each glycosylation sequon (D/E-X-N-Y-S/T, where X and Y are any amino acids other than proline). Since *cmeABC* is an operon, we added a 6x His tag at the C-terminal of CmeC to monitor changes in CmeABC expression. We then grew both WTCmeABC and g0CmeABC strains to an OD_{600} =0.4, tetracycline was then added to inhibit protein synthesis. CmeC levels in both strains was monitored for 2 hours. Our data shows no difference in CmeC expression in WTCmeABC and g0CmeABC strains **Fig 1, A**. CmeC might be in an unreachable region to the proteases once inserted in the outer membrane. Therefore we conclude that *N*-linked glycans, in this case, do not protect CmeC from native proteolytic degradation We observed two bands in the WTCmeC (lanes 2, 4, 6, 8) that migrated slower than the band corresponding to non-glycosylated CmeC in g0CmeC (lane 1, 3, 5, 7). Western blot analysis confirmed that the two bands observed in WTCmeC lanes correspond to two glycospecies for the protein. **Fig 1, B**. Our bioinformatic analysis of the primary amino acid sequence shows that CmeC contains two glycosylation sequons; ^{30}EANYS^{34} and ^{47}ETNSS^{51} in *C. jejuni* 11168. Structurally both of the glycosylation sequons are located in flexible loops[18] **Fig 6, B.**

_N_-linked glycans affect multidrug efflux pump efficiency

To examine the role of *N*-linked glycans in CmeABC molecular assembly, we assessed the efficiency of the multidrug efflux pump using an ethidium bromide accumulation assay. Ethidium bromide accumulation was 22% higher in g0CmeABC when

compared to WTCmeABC. This difference was consistent at 5, 10 and 15 minutes, indicating an impairment in the extrusion of ethidium bromide from g0CmeABC **Fig 1, C**. To confirm this finding, E-test antibiotic strips were used to calculate the minimum inhibitory concentration (MIC) of four non-structurally related antibiotics that have different mechanisms of actions. In comparison to WTCmeABC, an increase in antibiotic susceptibility was noticed in g0CmeABC confirming the previous finding **Table 1**. The results indicate that *N*-linked glycans play a role in enabling the multidrug efflux pump to work efficiently in *C. jejuni*.

Generation of fully glycosylated CmeA in glycocompetent *E. coli*

Previous studies showed that UDP-N-acetylglucosamine—undecaprenyl-phosphate *N*-acetylglucosaminephosphotransferase; WecA, could interfere with the biosynthesis of heterologous expression of polysaccharides built on the undecaprenyl-phosphate lipid anchor- rendering it built on incorrect glycan at the reducing end. To circumvent this problem and ascertain that g2CmeA is glycosylated with the native *C. jejuni N*-linked glycan we used glycocompetent *E. coli* SDB1[19]. The heterologous expression of an acceptor protein with protein glycosylation locus (*pgl*) usually yielded a mix population of glycosylated and non-glycosylated protein variant, indicating a suboptimal glycosylation process[20,21]. We observed that *pglB* expression from pACYC(*pgl*) is insufficient to achieve optimal glycosylation (data not shown). To overcome this bottleneck, we sought boosting PglB expression by introducing pGXVN114 to *E. coli* SDB1 expressing CmeA and *N*-linked glycan biosynthetic pathway. **Fig 2** Lane 1 shows the optimal glycosylation of constitutively expressed CmeA and *N*-linked glycosylation pathway along with IPTG inducible PglB from pGXVN114 backbone.

Glycosylation modulates protein global structure.

The enzymatic attachment of *N*-linked glycans to amide group of asparagine in the glycosylation consensus sequon, is correlated with changes in the biophysical properties of the protein such as thermostability[3], aggregation, function[4] and structure[22] in eukaryotes. Glycans can confer interactions not only at the glycosylation site but also other regions in the protein. Serving as bulky hydrophilic groups, glycans can also favour certain conformational modifications that stabilise protein structure[22]. Bioinformatic studies suggested that glycans force the polypeptide chain to adopt a more extended conformations through restricting residual structures formation in the unfolded state[23,24]. To investigate the role that *N*-linked glycans play in modulating the biophysical properties of CmeABC, we used circular dichroism (CD) spectroscopy. This allowed us to monitor the secondary structure as well as the conformational changes upon thermal denaturation of CmeA variants. Far-UV spectra for both g0CmeA and g2CmeA in 10 mM sodium phosphate, 75 mM sodium chloride, 10% glycerol buffer (pH 8) were collected at 20 °C. The spectrum exhibited helical structure signature with two negative minima at 208 and 222 nm. It also showed a positive maximum at 196 nm suggesting the presence of beta-sheets structure **Fig 3** Superimposed CD spectra exhibit g0CmeA spectrum, shown in black, was slightly red shifted towards the β sheet. The CD spectra of both proteins were then analysed by BESTSEL[25] for secondary structure content. **Table 2** shows the secondary structure content of g0CmeA and g2CmeA at room temperature. Our results show that *N*-linked glycans confer subtle change in protein global conformation.

Glycans enhances protein thermostability

It has been established that *N*-linked glycans play a role in stabilising glycoproteins thermodynamically in eukarotes[3,22]. The intrinsic role of *N*-linked glycans in stabilising

CmeA was investigated through analysing the CD spectra recorded for g0CmeA and g2CmeA at elevated temperatures. The multi–wavelength melting profiles monitored at 260- 195 nm were recorded during the heating of g0CmeA and g2CmeA from 6^{o}C to 94^{o}C at 1^{o}C per minute rate with a 2^{o}C step size. Isodichroic points were observed in the far-UV CD spectra at **Fig 4, A and B** supports more than two-state nature of the unfolding transition. Derivative of CD spectra were used to calculate the melting temperature (T_m) of both g0CmeA and g2CmeA **Suppl 1**. The loss of CD spectra was observed upon incremental rises in temperature, melting curves measured for g0CmeA and g2CmeA show that both proteins have three transitional phases at T_{m1} 46.1^{o}C ± 0.2, T_{m2} 53.5^{o}C ± 0.4 and T_{m3} 56.7^{o}C ± 0.6 for g0CmeA and T_{m1} 43.8^{o}C ± 0.3, T_{m2} 49.3^{o}C ± 0.2 and T_{m3} 62.8^{o}C ± 0.2 for g2CmeA. This shift in final melting temperature suggests that glycans stabilise g2CmeA at elevated temperature **Fig 4, C**.

To confirm the previous findings, we examined conformational folding reversibility and unfolding rate for both g0CmeA and g2CmeA. The assay is based on three successive cycles whereby CmeA variants were cooled at 20^{o}C, heated up to the corresponding T_m for 5 minutes and then cooled again at 20^{o}C. To assess conformational folding reversibility, CD spectra that were recorded at 20^{o}C, before and after increasing the temperature to the corresponding T_m were compared. CD spectra of CmeA variants were superimposable before and after the first two cycles of heating (T_{m1} and T_{m2}) but not after heating at T_{m3}, indicating conformational changes **Fig 4, D and E**. Unfolding rate was evaluated according to changes in CD spectra with respect to time at T_{m3}. A significant reduction in CD spectra was observed when g0CmeA and g2CmeA were heated at their corresponding T_{m3} . The unfolding of CmeA was achieved in 5 minutes at its T_{m3}. Notably, the CD spectra recorded for g2CmeA at its corresponding T_{m3} kept changing for 30 minutes

indicating a slower unfolding rate. This result along with the above mentioned data highly suggest that *N*-linked glycans play a pivotal intrinsic role in protein thermodynamic stabilisation.

Glycans modulate molecular assembly and protein-protein interaction

Unlike eukaryotes, there is no evidence that *N*-linked glycans modulate protein-protein interaction or complex assembly in prokaryotes. To explore the potential role of *C. jejuni* general *N*-linked glycans in modulating the interaction of glycoproteins with their cognate partners, surface plasmon resonance (SPR) was used. SPR was previously used to investigate the interactions in orthologous membrane fusion proteins (MFP) with TolC from of *E. coli*[26]. Quantitative analysis of the binding curves showed multiple reaction events. The model used to determine binding kinetics indicated the presence of two populations of MFP proteins. The two populations exhibited different binding kinetics, notably, fast and slow dissociation rates that contributed to weak and strong interactions, respectively[26]. We employed a CM5 chip with g0CmeA and g2CmeA immobilised through amine coupling, CmeC was then injected over CmeA variants surfaces in different concentrations. In our model, g0CmeA and g2CmeA exhibited multiple events interaction with CmeC. These interaction events can be attributed to a fast and a slow association and dissociation rates. Quantitative analysis of the sensogram yielded excellent results for slow interactions however, fast interactions could not be fitted in a model to generate accurate binding kinetics. At pH 7.4, both CmeA variants exhibited similar dissociation rate constants (k_{off}) of $9e^{-4}$ s^{-1} for g0CmeA and $7.5e^{-4}$ s^{-1} for g2CmeA **Fig 5, A and B.** Interestingly, difference in association rate constant (k_{on}) was observed, g0CmeA k_{on} = $5e^{4}$ ($M^{-1}s^{-1}$) whilst g2CmeA k_{on} =$1.5e^{5}$ ($M^{-1}s^{-1}$). This difference in the k_{on} rate indicates that g2CmeA possess more binding pockets that allows slow yet high affinity interactions with CmeC compared to g0CmeA. The equilibrium dissociation

rate constants (K_D) derived from the binding kinetics analysis were $1.7e^{-8}$ (M) and $5e^{-9}$ (M) and g0CmeA and g2CmeA, respectively.

To investigate the effect of pH on modulating binding kinetics, we observed CmeA-CmeC interactions at pH=6.0 **Fig 5, C and D**. At this pH CmeA-CmeC interactions were more avid and with a greater number of sites bound**.** Similar to binding curves observed at pH 7.4, g2CmeA showed a favourable slow association and dissociation binding curves than g0CmeA. The number of sites for slow interaction were greater for g2CmeA contributing to a modestly higher affinity for interaction with CmeC compared to a weaker affinity for CmeC exhibited by g0CmeA.

To confirm that variations in binding kinetics were not due to differences in structural orientation between g0CmeA and g2CmeA, both proteins were immobilised on NTA chip using C-terminal 6xHis tag, CmeC was then passed in different concentrations. Binding kinetics indicated similar k_{on} and k_{off} for both CmeA variants, although fewer sites were available **Suppl 2**. Interestingly, g2CmeA bound more CmeC than g0CmeA, confirming the data seen with amine coupling. These results show a complex binding pattern between CmeA variants and CmeC. They also suggest an extrinsic role for *N*-linked glycans, exhibited in the variation in binding kinetics between g0CmeA and g2CmeA, where the glycosylated form of CmeA showed a greater proportion of higher affinity interaction sites than its non-glycosylated counterpart.

Discussion

Whilst the role of eukaryotic glycosylation has been thoroughly explored, a similar depth of investigation is lacking in prokaryotic glycobiology. Publications have reported the presence of different glycosylation systems in prokaryotes[2,7,11–13], however, few described their effects in virulence, adhesion and motility by creating genetic knock outs of the glycosylation machinery[32,33,10–12]. Nonetheless, these

reports have not provided in-depth studies into the role that *N*-linked glycans exert on protein function. Glycoproteomic studies revealed diBacNAc to be conserved across *Campylobacter* species. Notably, diBacNAc was found to be the reducing end of *O*-linked glycans in *Neisseria gonorrhoeae*, indicating a probable parallel evolution between *N*-linked and *O*-linked glycosylation systems in bacteria[8,9,34]. The role of *N*-linked glycans in stabilising major multidrug efflux pump in *C. jejuni* has been shown to contribute in efficiently extruding antimicrobials and ethidium bromide. Disrupting glycosylation in CmeABC resulted in higher accumulation of ethidium bromide and lowering antibiotic MIC in *C. jejuni.* These differences in activity are not due to the low abundance of CmeABC complex in g0CmeABC strain **Fig 1, A**. A protein synthesis arrest assay showed that loss of glycosylation did not promote CmeC proteolytic degradation. It also comes with an agreement with the previous finding that CmeC protein abundance were equal in *C. jejuni* and *C. jejuni pglB*::*aphA* (Abouelhadid S, *et al* paper submitted). This result suggests that *N*-linked glycans might be modulating molecular assembly. Studies on truncated *N*-linked glycans will reveal the role of each glycostructure, it will also help to understand the role of the conserved first two glycans between different *Campylobacter* strains.

Bioinformatic studies investigating protein structural changes exerted by glycans has been inconclusive. These studies rely heavily on the in-silico analysis of protein structure entries in the protein database bank (PDB). Whilst modern advances in crystallographic techniques pave the way for more structural studies, obtaining glycoprotein structure is still challenging and remains poorly represented in PDB. Xin *et al* reported that protein glycosylation causes significant yet unexpectedly subtle (up to 7%) changes in both local and global protein structure[23]. However, Hui Sun lee *et al* concluded that *N*-glycosylation causes non-significant changes in protein structure but increases protein stability likely due it a role played in reduction of

protein dynamics[24]. Experimentally, our initial CD study of CmeA variants showed that both have the same conformational fold, however they confer subtle structural differences to the protein **Fig 3, B**. A small shift has been observed in the percentage of alpha helices and beta sheets between g0CmeA and g2CmeA, ±1.4% ±0.2. It is still unclear whether the structural variations are due to local stabilisation resulting from the glycosidic bond between the asparagine side chain in the glycosylation site and *N*-linked glycans, or global structural rearrangement due to the interaction of the glycan with other distant regions in the protein backbone. A structural elucidation of CmeA in its glycosylated and non-glycosylated might provide insights on the extent of the importance of conformational changes.

It has been suggested that *N*-linked glycans might enhance protein thermostability. The glycoprotein PEB3 from *C. jejuni* was used as a scaffold to test the stabilisation effect of *N*-linked glycans. Average melting temperature of PEB3 (K135E) variants were analysed using SYPRO orange thermoflour. Interestingly, the T_m of glycosylated PEB3 was shown to be 4.7^{o}C higher than its non-glycosylated counterpart; PEB3[35]. This comes with an agreement with CD thermal melts of g0CmeA and g2CmeA. CD thermal melts showed that whilst both of the CmeA variants have the same apparent unfolding behaviour, T_m of g2CmeA was 6.4^{o}C ±0.5 higher than that of g0CmeA. The three transitional phases of both variants showed that g2CmeA seems to be responding to a rise in temperature via conformational rearrangements at 2.4^{o}C ±0.1 lower than g0CmeA **Fig 4, A**. CD spectra recorded after cooling showed that the structural rearrangements were reversible and the protein could fold again, suggesting that protein fold/unfolding T_{m1} and T_{m2} are reversible for both of the CmeA variants. Remarkably, the unfolding behaviour of g2CmeA at T_{m3} was different to g0CmeA at its correspondent T_{m3} in the conformational reversibility assay. Time taken to unfold g2CmeA was at least 5 times more than g0CmeA, thus indicating a

role played by *N*-linked glycans in conferring greater resistance to unfolding **Fig 4, C and D**. We postulate that *N*-linked glycans stabilise g2CmeA through a reduction in the unfolding rate in g2CmeA, this finding agrees with the observation that eukaryotic *N*-linked glycans stabilise the hCD2ad through slowing the unfolding rate of the protein 50 fold when compared to its non-glycosylated counterpart[2].

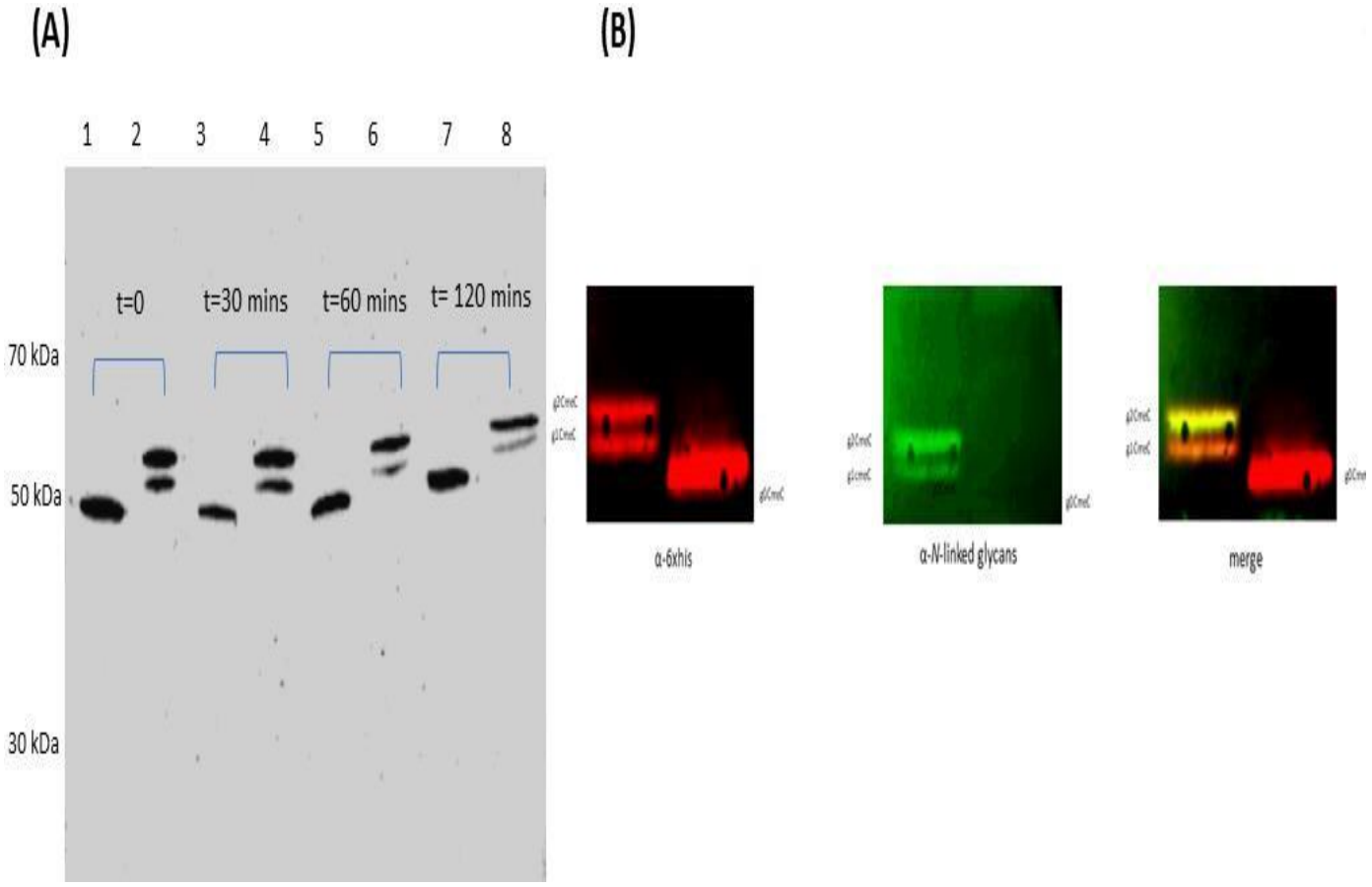
Owing to the lack of subcellular compartments, the extrinsic role of prokaryotic *N*-linked glycans in protein-protein interaction has been not fully appreciated. Despite the scarcity of glycoproteomic data, few molecular assemblies have been reported to have at least one of its component to be glycosylated [17,31]. We demonstrate a potential extrinsic role of *N*-linked glycans in CmeA interaction with CmeC. In an orthologous multidrug efflux pump; AcrAB-TolC, AcrA showed the presence of two populations, of the same protein, interacting with different kinetics to TolC. The two populations contributed to a fast weak interaction and slow strong interaction **Fig 5, A and B**. The complexity of these interactions is exaggerated in *C. jejuni* due to the presence of *N*-linked glycans, that could modulate interaction of CmeA with CmeC. Quantitative analysis for the interaction kinetics of CmeA variants with CmeC showed that, *N*-linked glycans increase the binding affinity to CmeC by 3.4-fold. That was clearly demonstrated in the difference in K_D between CmeA variants at pH= 7.4. The difference in binding affinity was confirmed when CmeA variants were immobilised with the same orientation on Ni chip. Recently, a pseudoatomic structure provided a detailed picture of interaction between AcrA and TolC. This elaborated the adaptor bridging-binding model that involved an intermesh cogwheel-like binding between AcrA and TolC[36]. The conserved binding motif Val-Gly-Leu/Thr (VGL) is located at the tip region of the coiled coil α-hairpin of the protein, serving as a site of interaction with the RXXXLXXXXXS (RLS) motif of AcrA[36]. In light of this study, our computational analysis showed that CmeC from *Campylobacter spp* does contain a truncated VGL

motif, denoted VGA, whilst we found the RLS motif to be conserved in among *C. jejuni, C. lari, C. coli* and *C. fetus* **Fig 6**, **A**. To understand whether *N*-linked glycans modulate protein-protein interaction we analysed the proximity of glycosylation sites to VGA and RLS binding site in both CmeC and CmeA, respectively. The glycosylation sites in CmeC were shown to be far from the proposed binding site **Fig 6, C** and probably closer to the transmembrane domain of the protein. Remarkably, we found that one of the glycan modified asparagine (^{123}N) is at the X_{-1} position to RLS motif and is conserved in *C. jejuni* and *C. coli,* but not in *C. lari* and *C. fetus* **Fig 6, A and C**. This strongly suggests that the localisation of *N*-linked glycan adjacent to RLS might be affecting either the local site conformation and/or promote a stronger interaction with the VGA motif in CmeC; resulting in the interaction kinetics differences between g2CmeA and g0CmeA with CmeC observed by SPR in this study. In eukaryotes, it is well established that *N*-linked glycans at different glycosylation sites in the same protein could play different roles. The roles of these *N*- linked glycans can be categorised into; (a) promoting protein folding, (b) modulating protein trafficking and localisation and (c) effecting protein functionality. In future, we will aim to understand the role played by glycans at ^{123}N and ^{273}N in CmeA.

This study provides the first detailed analysis of the role of *N*-linked glycans in prokaryotes. The role of bacterial general *N*-linked glycans has been difficult to elucidate. This led to a notion that *N*-linked glycans do not play any role in protein folding or function. This notion was based on previous inconclusive results on the role of *N*-linked glycans in modulating proteins function[37–40]. Our work refutes this widely held notion and demonstrates that *N*-linked glycans do not only play a role in slowing protein unfolding process and enhancing its thermostability but also it modulates protein interaction with its cognate partner**.** It also draws a line showing a conserved role of general *N*-linked glycans previously seen in eukaryotes. This also

suggests a common evolutionary role that led to the emergence of *N*-linked protein post translation modification, in expanding the functionality of proteome repertoire across all domains of life. Our proposed model can be used to interrogate prokaryotic general

Fig 1


(A)
(B)
1 2 3 4 5 6 7 8
t=0
t=30 mins
t=60 mins
t= 120 mins
70 kDa
50 kDa
30 kDa
α-6xhis
α-N-linked glycans
merge

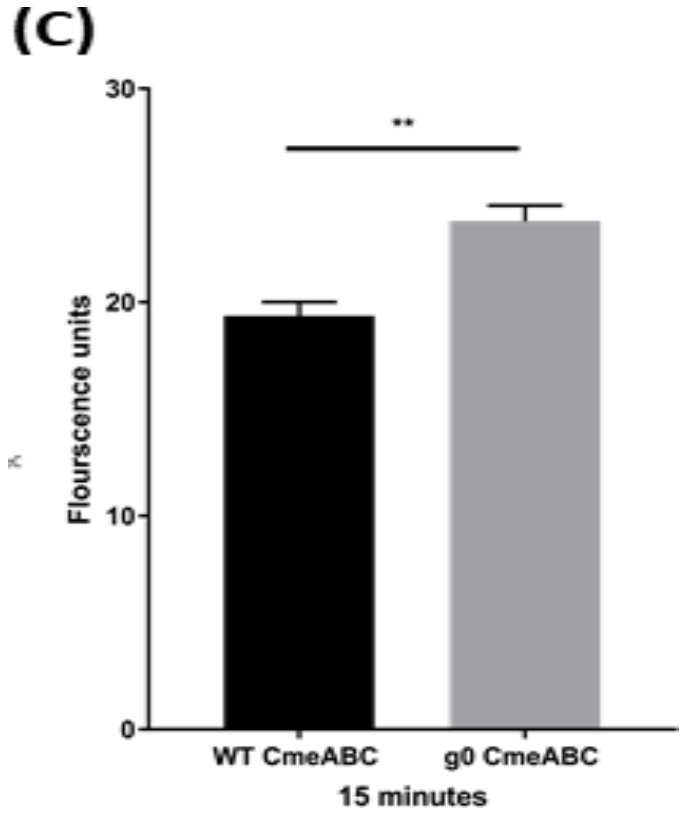


Fig 1 Functional studies and effect of glycosylation on WTCmeABC and g0CmeABC

(A) Proteolytic degradation assay of CmeC and gCmeC. Cells were grown to OD_{600} =0.4 then Chloramphenicol and Tetracycline were added. Initial sample were withdrawn and labeled t=0, then samples were taken every 30 minutes. Cells were stored on ice, centrifuged, lysed by sonication then incubated with 2% SDS and Sodium sarkosyl for 2 hours at room temperature. Cells debris were then pelleted by centrifugation and supernatants were mixed 1:1 with Laemmli loading buffer supplemented with DTT. Proteins were then separated by SDS-PAGE followed by electroblotting to PVDF membrane, 6xhis tagged CmeC was probed by 1ry anti-6xhis mouse antibody and visualized by Li-COR odyssey. Equal amount of proteins was loaded, lane 1, 3, 5, 7, *C. jejuni* g0CmeC; lane 2, 4, 6, 8 *C. jejuni* WTCmeABC (B) Western blot detection of CmeC variants, lane 1, g2CmeC; lane 2, g0CmeC. Proteins were detected with anti-6xHis antibody and *N*-linked glycans were detected with SBA

lectin-biotin. (C) Ethidium bromide accumulation test in *C. jejuni* strains. 30 ml Brucella broth was separately inoculated with overnight culture of *C. jejuni* WTcmeABC (black) and *C. jejuni* g0CmeABC (grey) to $O.D_{600}$ 0.1. Cells were grown till $O.D_{600}$ 0.4-0.5 then spun down, washed and resusupended to OD_{600} 0.2 in 10 mM sodium phosphate buffer pH 7. Cells were then incubated in VAIN for 15 minutes at 37^{o}c then Ethidium bromide was added to final concentration of 0.2 mg/ml. Fluorescence was read at excitation and emission for 20 minutes at 37^{o}c accumulation in *C. jejuni* strains at 15 minutes. The data represents the mean of three biological replicates, two technical replicates each. Significance was calculated using Mann-Whitney test. ***P*<0.005

Table 1 Minimum inhibitory concentration in *C. jejuni* strains.

Antibiotics	Concentration range (µg/ml)	*C. jejuni* WTCmeABC	*C. jejuni* g0CmeABC	Fold difference
Ampicillin	256–0.015	4	2	2
Erythromycin	256–0.015	0.25	0.12	2
Ciprofloxacin	32–0.002	0.06	0.03	2
Tetracycline	256–0.015	0.5	0.25	2

Table 1 Minimum inhibitory concentration in *C. jejuni* strains. The minimum inhibitory concentration (MIC) of *C. jejuni* WTCmeABC and *C. jejuni* g0CmeABC was read directly from the strip at the point where the zone of inhibition of bacterial growth intersected with the antibiotic concentration on the strip. The results presented are the means from three biological replicates two technical replicates each.

Fig 2

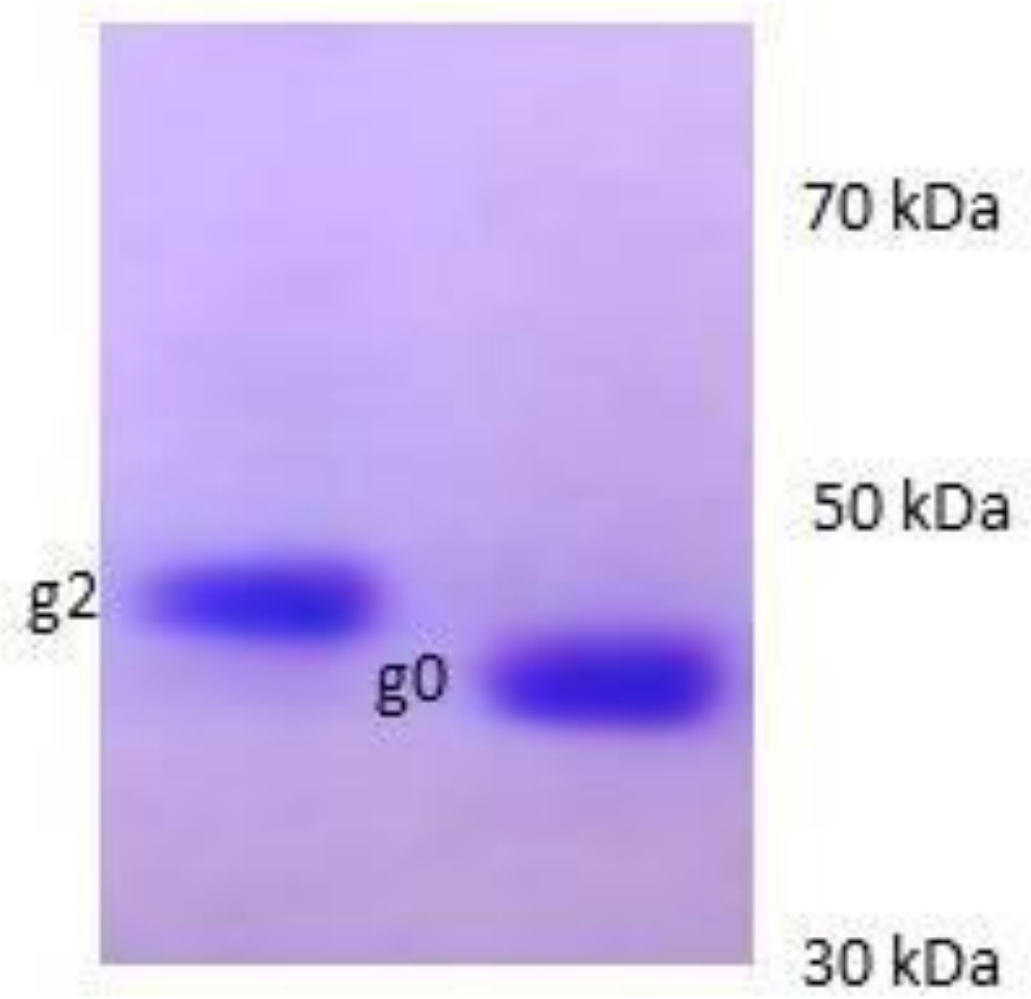


Fig 2 Generation of fully glycosylated CmeA in *E. coli* SDB1. CmeA variants were purified using IMAC followed by concentration and buffer exchange using Amicon ultra-0.5 ml centrifugal filter units. Proteins were then separated by SDS-PAGE and visualized by Coomassie blue staining.

Fig 3

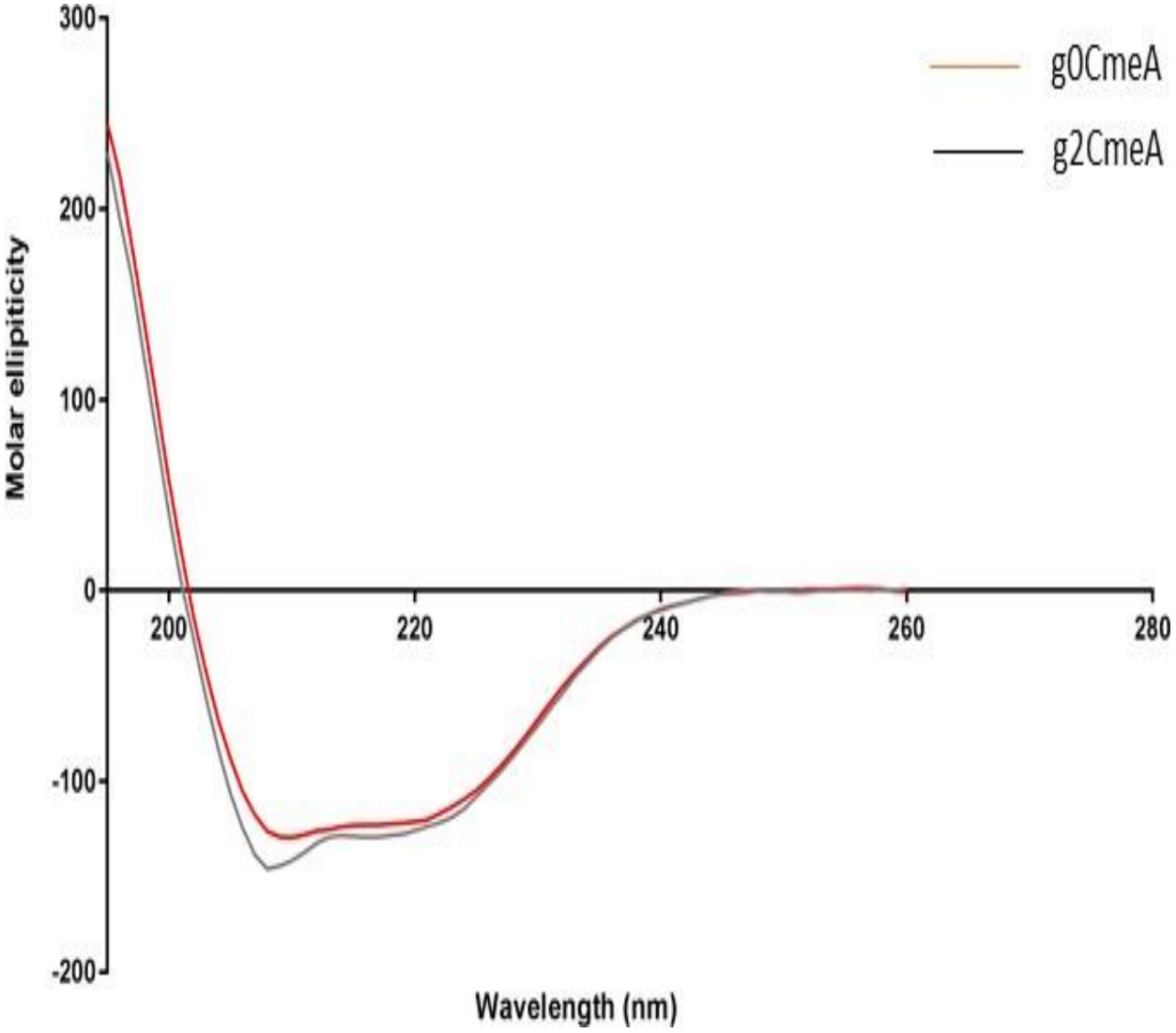


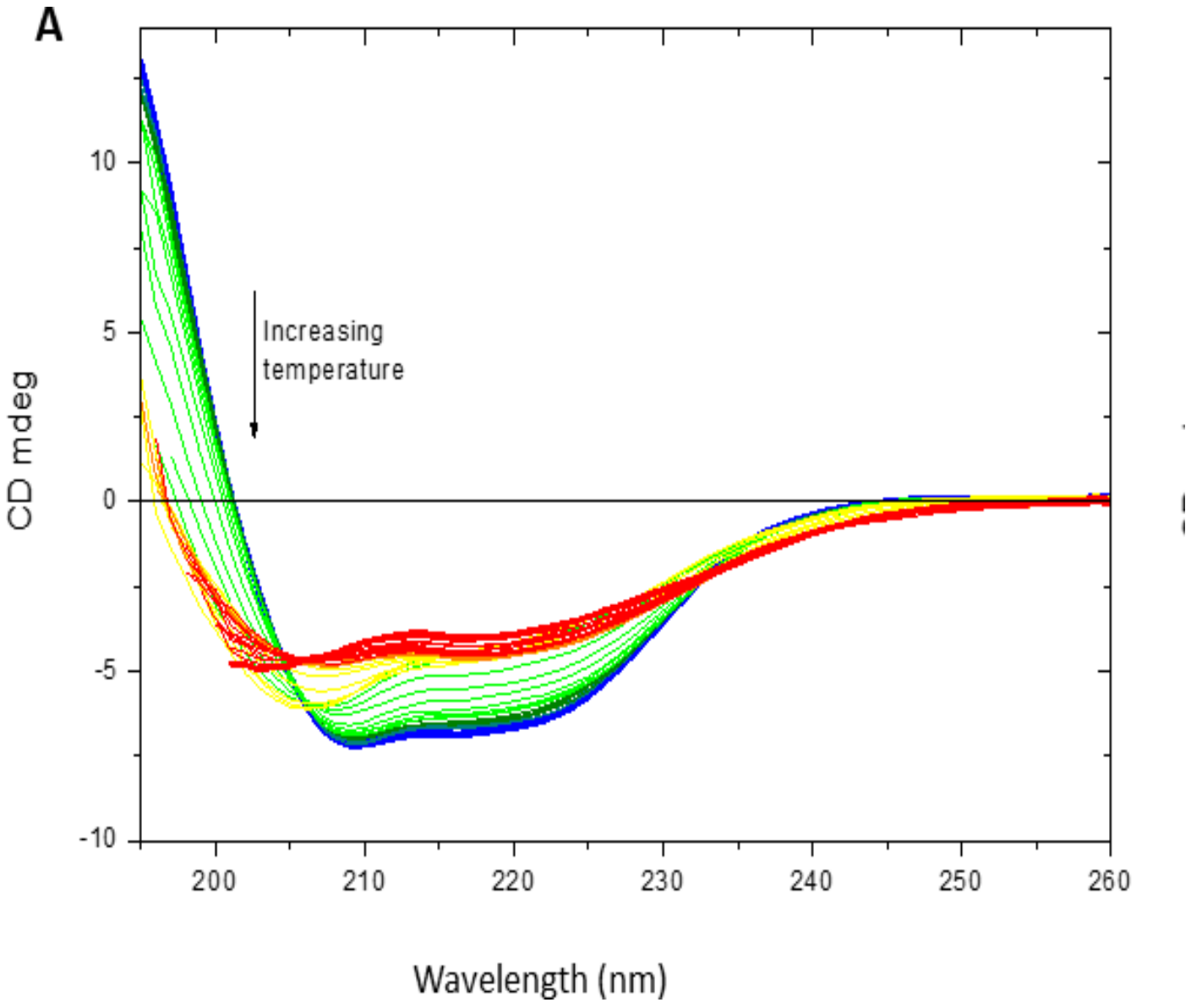
Fig 3 CD spectra of g0CmeA and g2CmeA in 10 mM sodium phosphate, 75 mM sodium chloride and 10% glycerol (pH 8.0). Far-UV CD spectra was collected for g0CmeA(0.124 mg/ml) and g2CmeA (0.174 mg/ml) variants in 0.5 mm rectangular cell pathlength. Molar ellipiticity was calculated and corrected for proteins concentration.

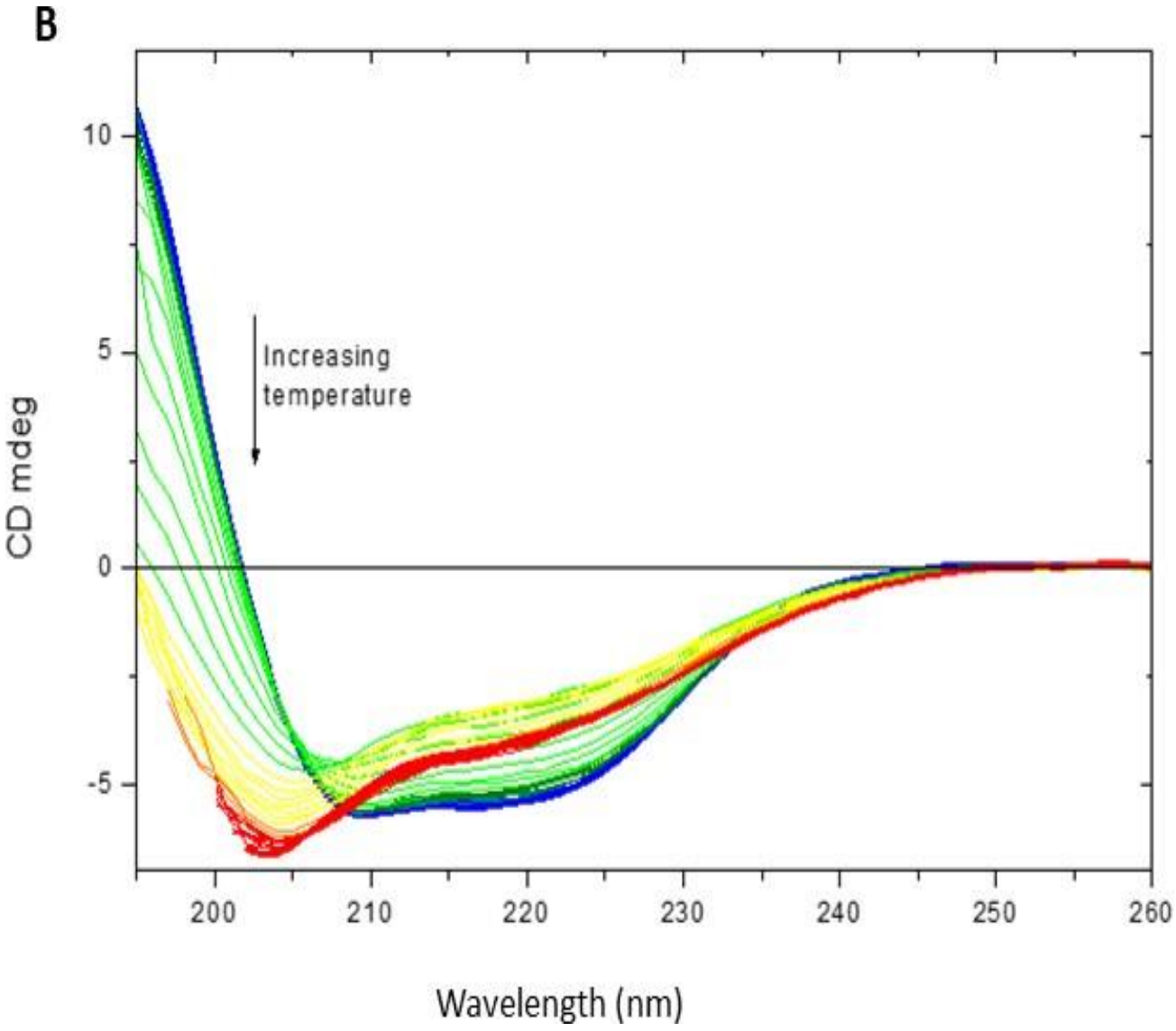
Table 2 Secondary structure calculation of g0CmeA and g2CmeA variants

	α-helix	β-sheets antiparallel	β-sheets parallel	Turn	Others
g0CmeA	28.7%	26.8%	0.0%	12.0%	32.5%
g2CmeA	27.3%	27.6%	0.0%	10.8%	34.3%

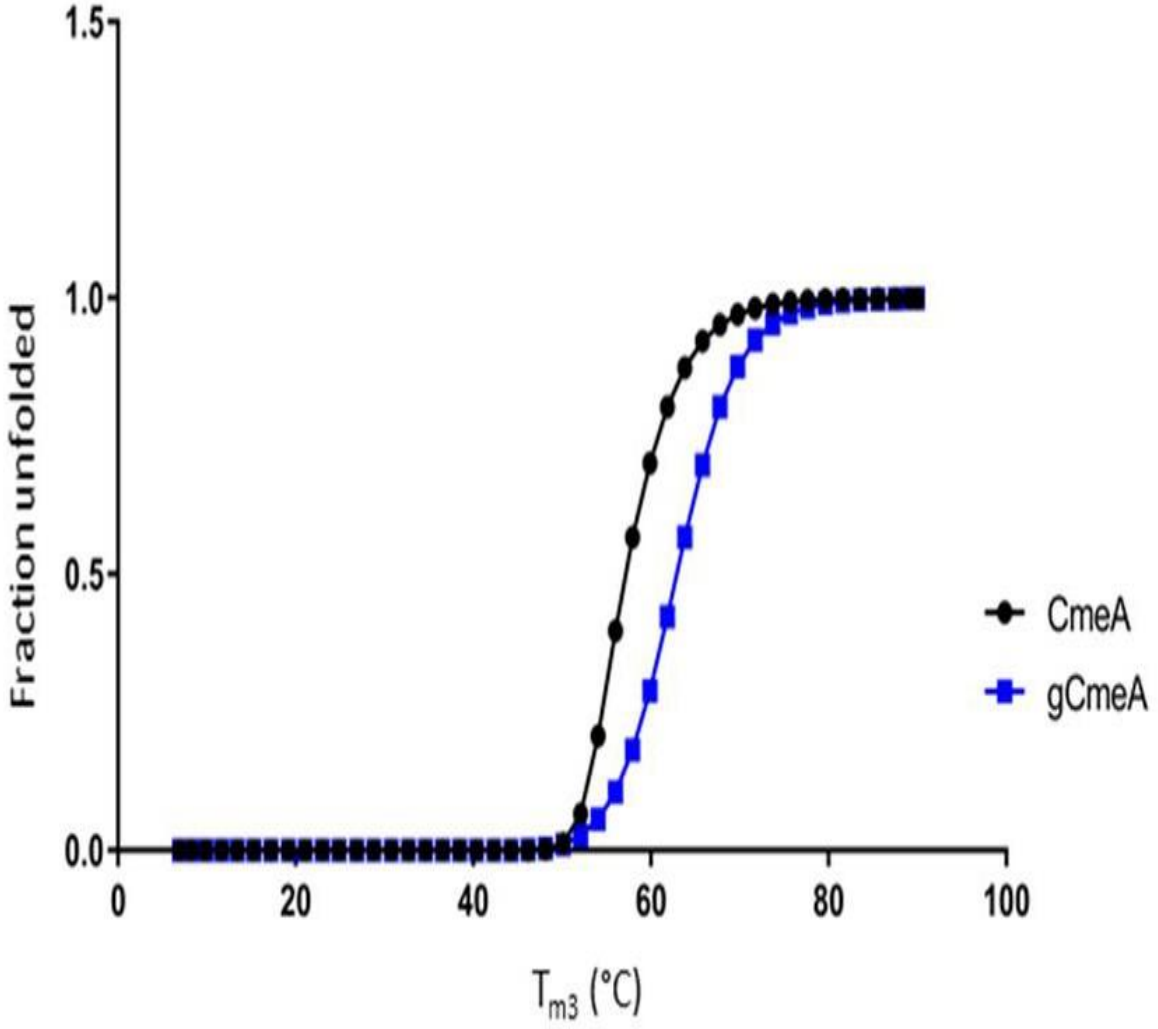
Table 2 Secondary structure calculation of g0CmeA and g2CmeA variants. CD units were converted to delta epsilon units and loaded to BESTSEL server. Although the conformations of both proteins are structurally similar there is a subtle shift in the alpha helices and beta sheets ratios between both variants.

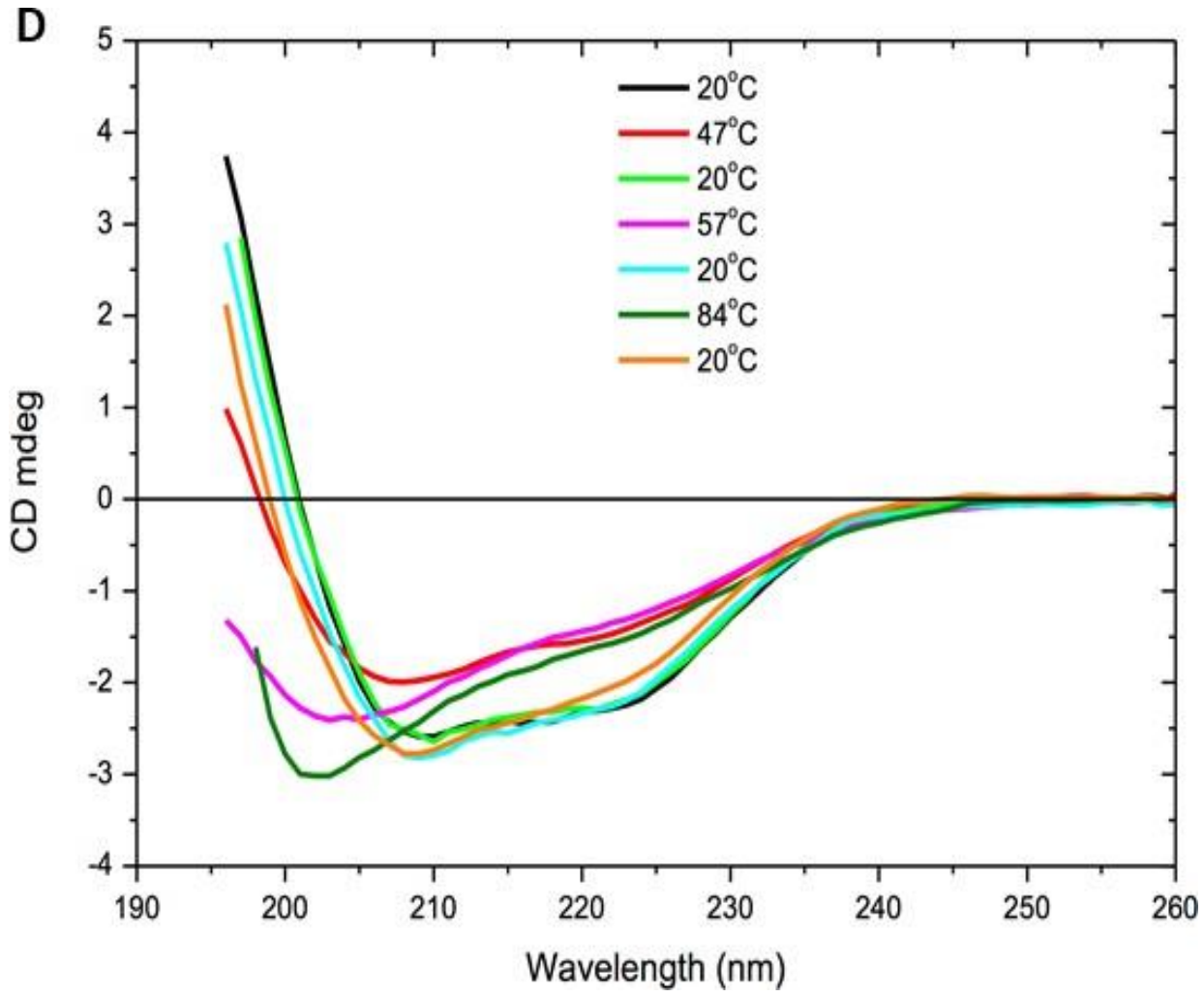
Fig 4


A
10
5
0
-5
-10
CD mdeg
Increasing
temperature
200
210
220
230
240
250
260
Wavelength (nm)


B
10
5
0
-5
CD mdeg
Increasing
temperature
200
210
220
230
240
250
260
Wavelength (nm)

C




D
CD mdeg
Wavelength (nm)
20°C
47°C
20°C
57°C
20°C
84°C
20°C
5
4
3
2
1
0
-1
-2
-3
-4
190
200
210
220
230
240
250
260

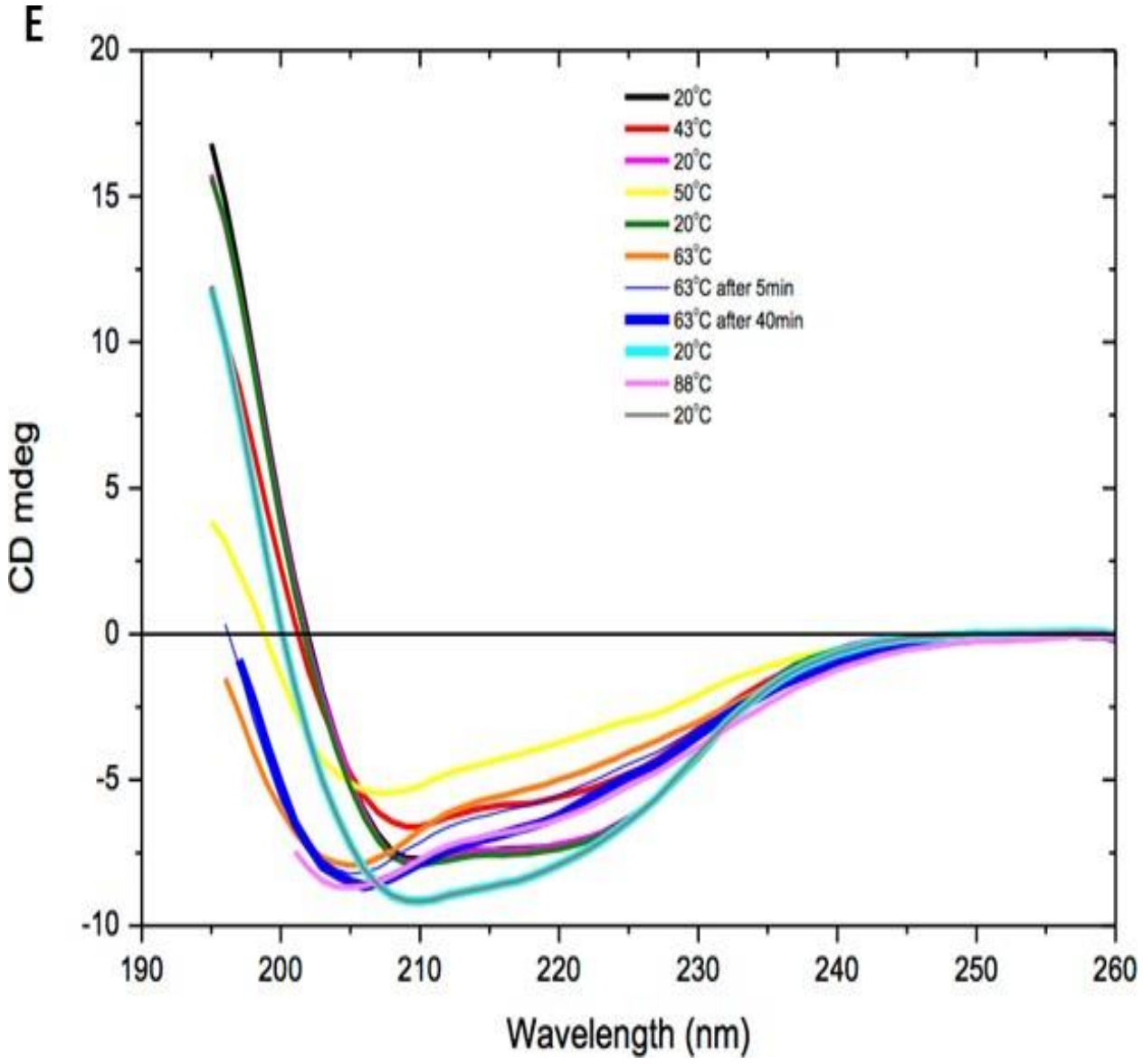
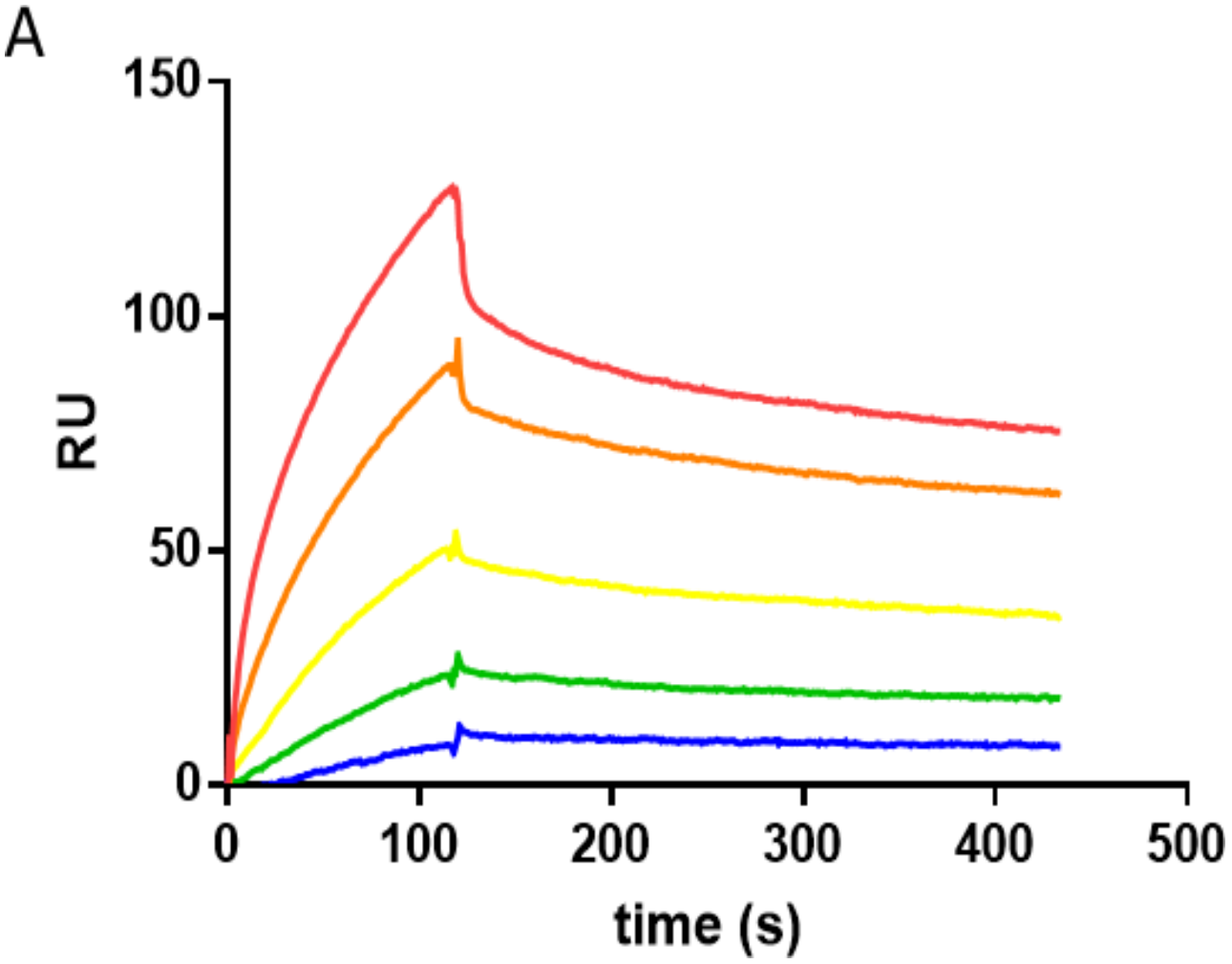
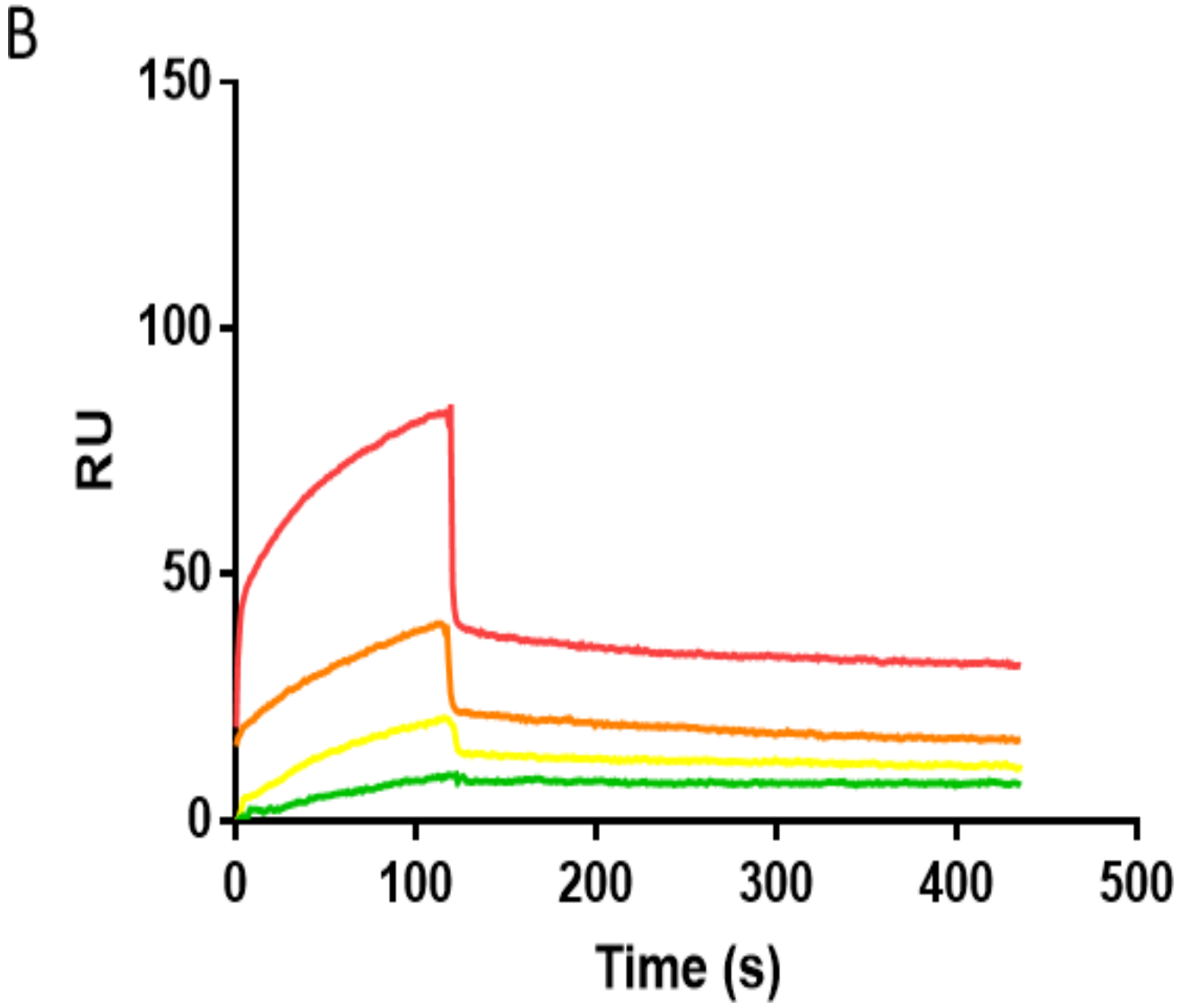


Fig 4 Thermal melts of g0CmeA and g2CmeA in 10 mM sodium phosphate, 75 mM sodium chloride and 10% glycerol (pH 8.0). Far-UV CD spectra was collected for g0CmeA(0.124 mg/ml) and g2CmeA (0.174 mg/ml) variants in 0.5 mm rectangular cell pathlength. CD mdeg were recorded as a function of temperature from blue (6 ^{o}C) to red (94 ^{o}C) for g0CmeA, (A) and g2CmeA (B). Each colour in between was obtained at rate 1 ^{o}C per minute with a 2 ^{o}C stepwise increase.CD spectra to asses the reversibility of thermal unfolding study was recorded at 20^{o}C, raised to T_m and re-cooled to 20^{o}C sequentially. CD spectra was collected for 5 minutes at each temperature interval for g0CmeA, (C); Thermal denaturation of CmeA

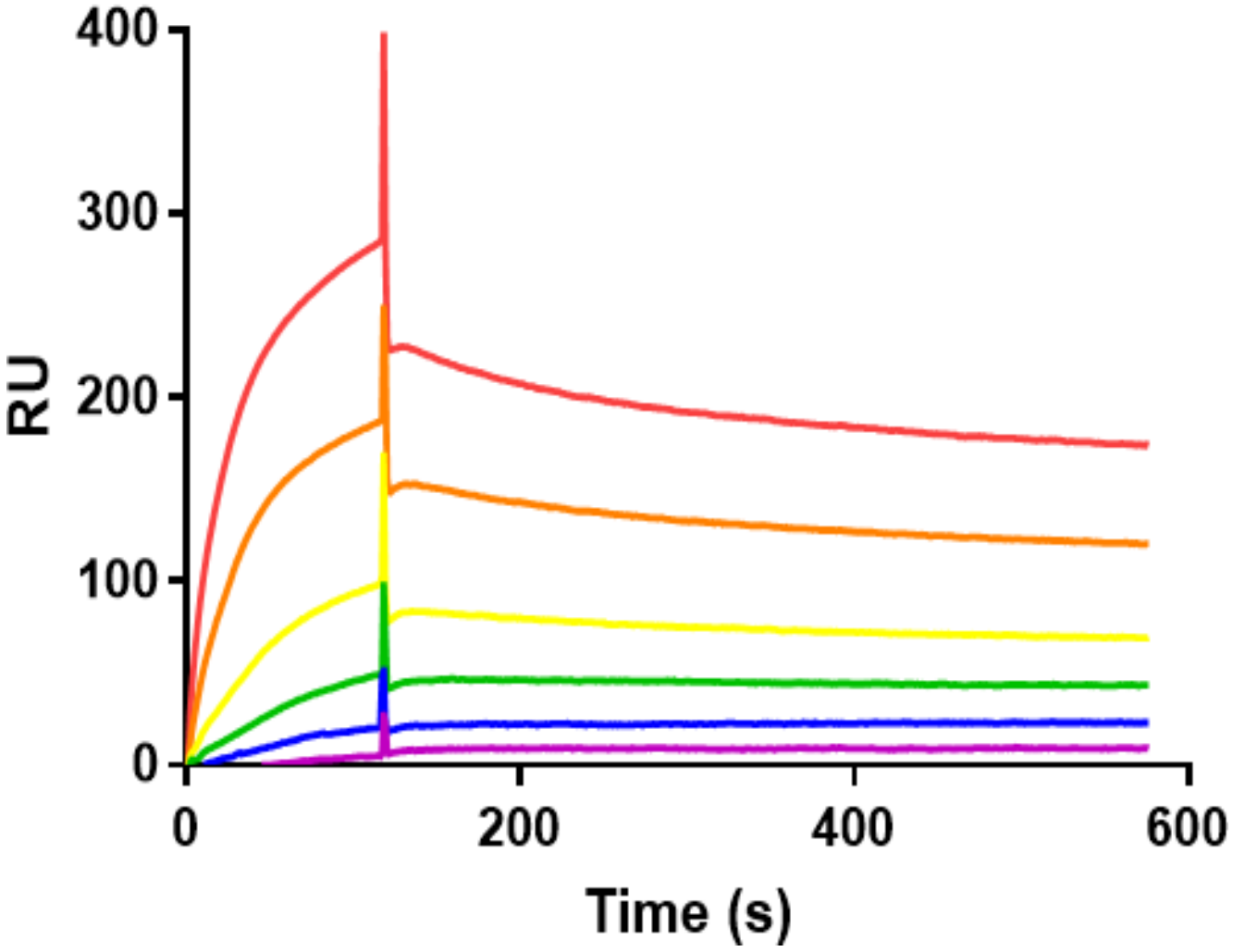
(black) and g2CmeA (blue) as change of ellipticity values upon temperature increase (D); g2CmeA (E). CD spectra of g2CmeA was stabilised after 30 minutes at T_{m3} indicating a more resilient behaviour thermal unfolding process.

Fig 5




B
150
100
50
0
RU
0
100
200
300
400
500
Time (s)

C



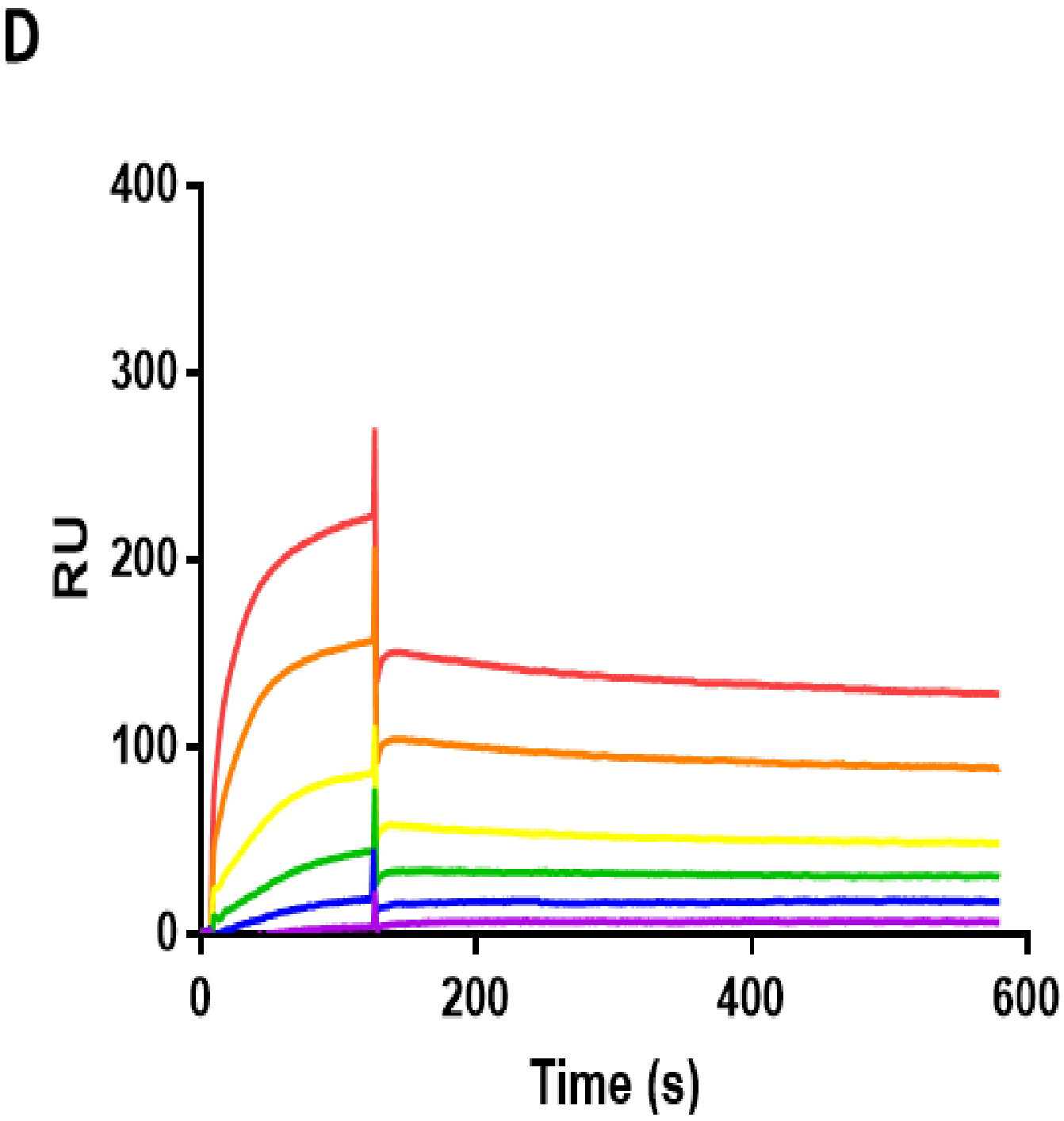
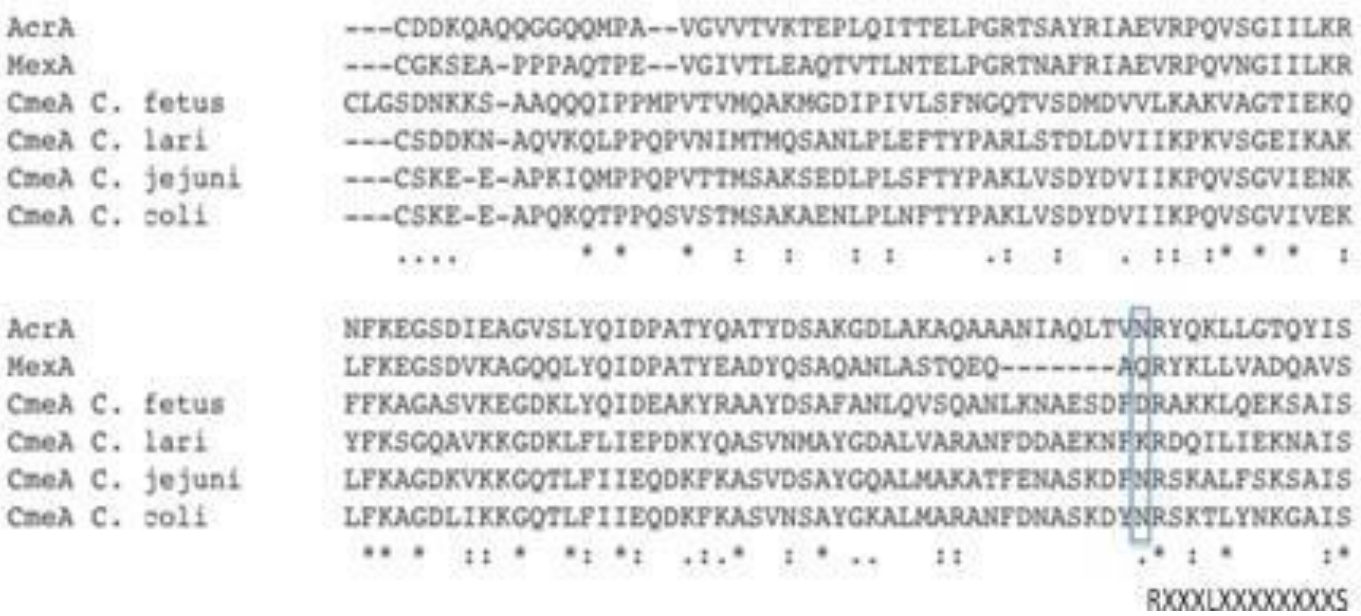


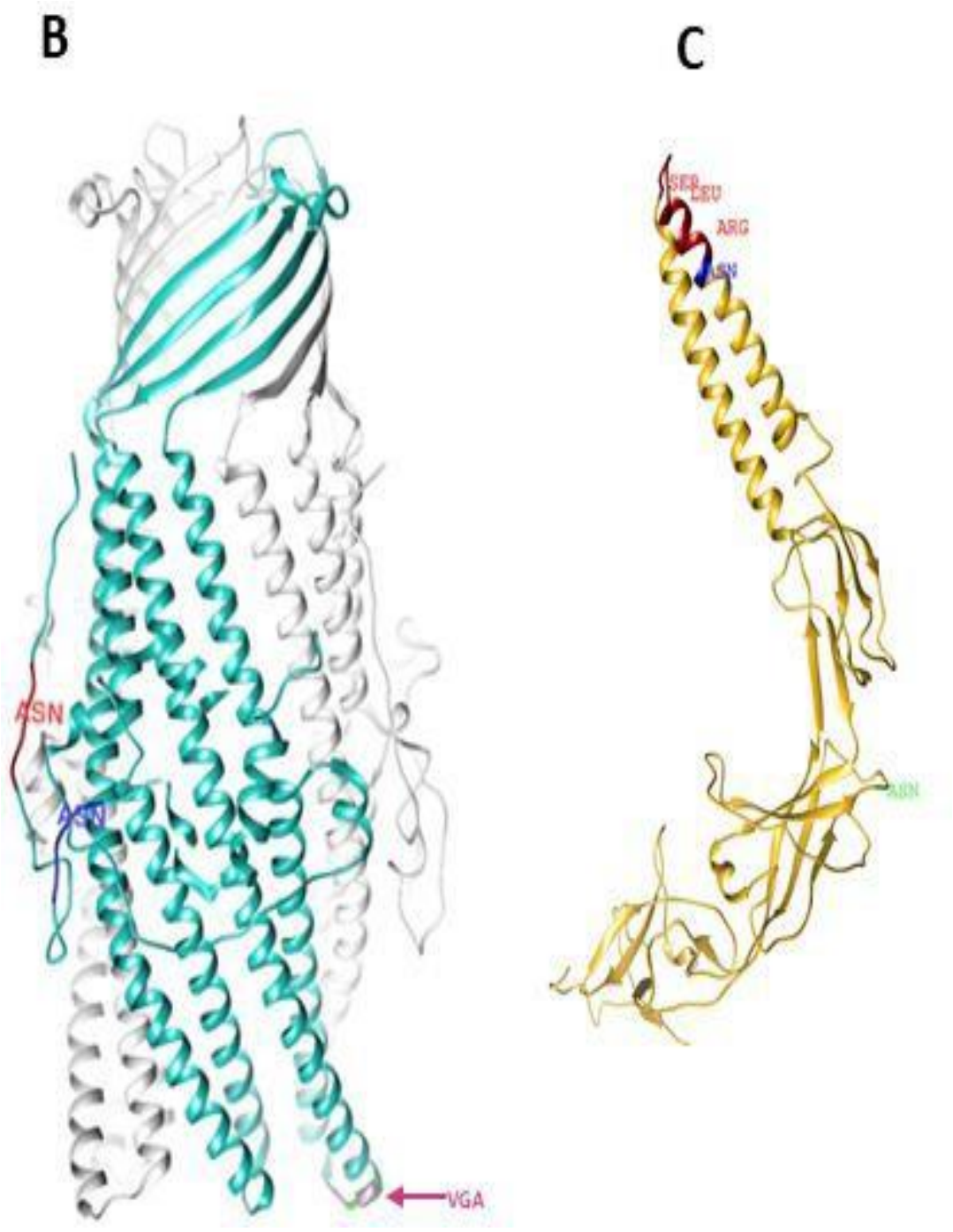
Fig 5 Glycosylation enhances interactions between CmeA variants and CmeC. SPR analysis of CM5 chip with A) 900 RU of g2cmeA immobilised and B) 1040 RU of g0cmeA immobilised. Association of CmeC at pH 7.4 was performed for 2 mins and dissociation was followed for 5 mins. Concentrations of CmeC were two-fold dilutions from $2x10^{-7}$ M (red) to $1.25\ x10^{-8}$M (blue) or $2.5\ x10^{-8}$ M (green). SPR analysis of CM5 chip at pH 6.0 with C) 900 RU of g2CmeA immobilised and D)

1040 RU of g0CmeA immobilised. Association of CmeC was performed for 2 mins and dissociation was followed for 5 mins. Concentrations of CmeC were two-fold dilutions from $2x10^{-7}$ M (red) to $0.6 x10^{-8}$M (purple)

Fig 6

A


AcrA ---CDDKQAQQGGQQMPA--VGVVTVKTEPLQITTELPGRTSAYRIAEVRPQVSGIILKR
MexA ---CGKSEA-PPPAQTPE--VGIVTLEAQTVTLNTELPGRTNAFRIAEVRPQVNGIILKR
CmeA C. fetus CLGSDNKKS-AAQQQIPPMPVTVMQAKMGDIPIVLSFNGQTVSDMDVVLKAKVAGTIEKQ
CmeA C. lari ---CSDDKN-AQVKQLPPQPVNIMTMQSANLPLEFTYPARLSTDLDVIIKPKVSGEIKAK
CmeA C. jejuni ---CSKE-E-APKIQMPPQPVTTMSAKSEDLPLSFTYPAKLVSDYDVIIKPQVSGVIENK
CmeA C. coli ---CSKE-E-APQKQTPPQSVSTMSAKAENLPLNFTYPAKLVSDYDVIIKPQVSGVIVEK
AcrA NFKEGSDIEAGVSLYQIDPATYQATYDSAKGDLAKAQAAANIAQLTVNRYQKLLGTQYIS
MexA LFKEGSDVKAGQQLYQIDPATYEADYQSAQANLASTQEQ-------AQRYKLLVADQAVS
CmeA C. fetus FFKAGASVKEGDKLYQIDEAKYRAAYDSAFANLQVSQANLKNAESDFDRAKKLQEKSAIS
CmeA C. lari YFKSGQAVKKGDKLFLIEPDKYQASVNMAYGDALVARANFDDAEKNFKRDQILIEKNAIS
CmeA C. jejuni LFKAGDKVKKGQTLFIIEQDKFKASVDSAYGQALMAKATFENASKDFNRSKALFSKSAIS
CmeA C. coli LFKAGDLIKKGQTLFIIEQDKFKASVNSAYGKALMARANFDNASKDYNRSKTLYNKGAIS
RXXXLXXXXXXXXS


B
C
ASN
ASN
VGA
SER
LEU
ARG
ASN
ASN

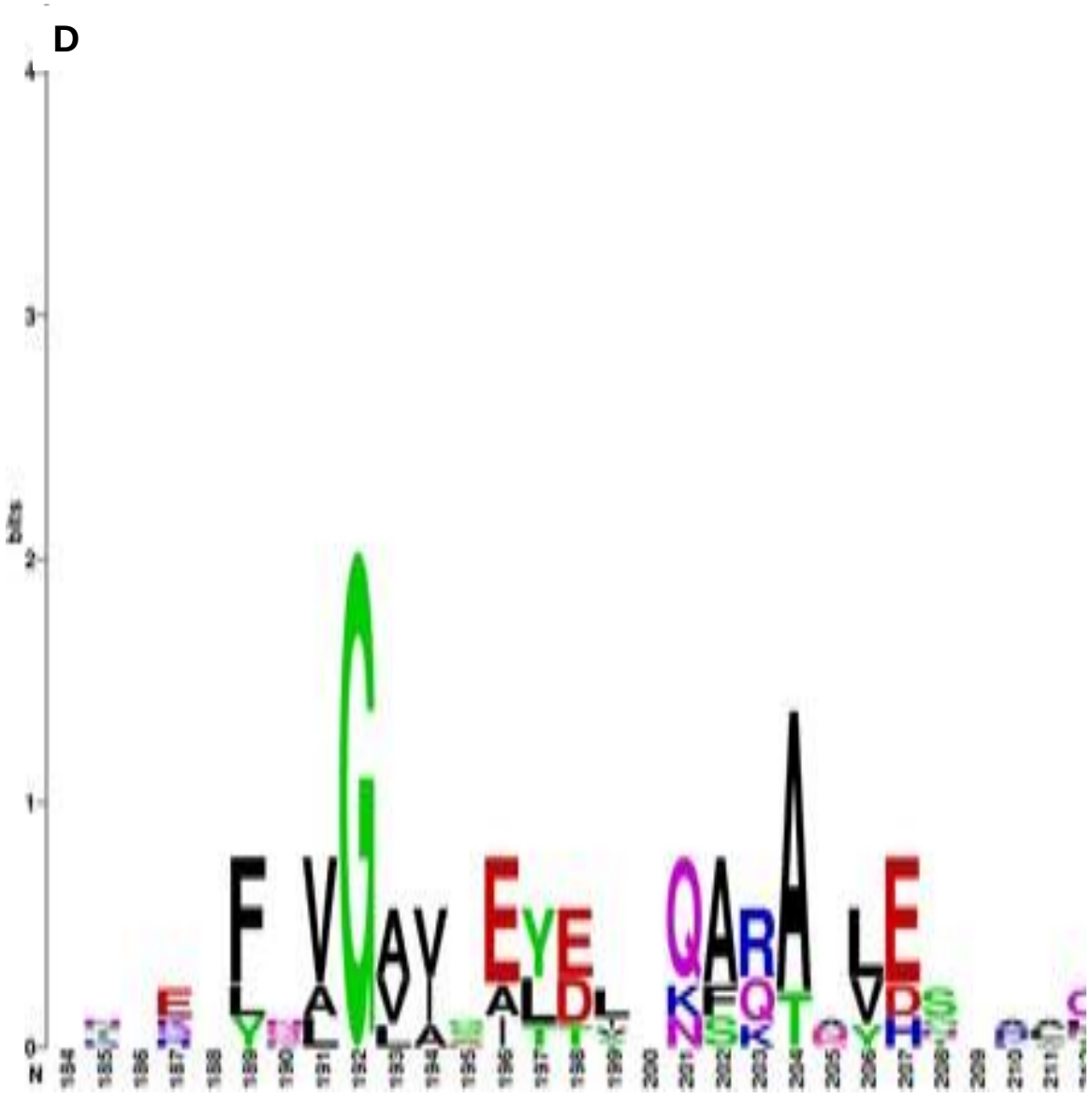


Fig 6 Analysis of binding sites in CmeA and CmeC (A) Amino acid alignment of signal peptide processed CmeA orthologues. Conserved amino acids are denoted by an asterisk, similar amino acids are denoted by colon and weak amino acid similarity is denoted by period. The amino acid sequences were retrieved from Uniprot and aligned using Clustal Omega[42]. RLS attachment site is shown to be conserved among periplasmic accessory proteins from different strains. The localisation of N is highlighted in blue box, showing the presence of ^{123}N at X_{-1} to

the conserved RLS motif in *C. jejuni* and *C. coli* but not *C. fetus* nor *C. lari*. (B) Structural representation focusing on chain A of CmeC trimer (PDB:4MT4). Chain A is highlighted in cyan, ^{32}N and ^{49}N are highlighted in red and blue respectively. The proposed attachment site

VGA motif is highlighted in magenta showing its distant from both of the glycosylation sites. (C) Structural prediction of CmeA. Signal processed amino acid sequence was deposited in I-TASSER and the best structural fit was based on MexA model(ref). RLS motif is highlighted in dark red, ^{123}N and ^{273}N are highlighted in blue and light green, respectively showing the close proximity of ^{123}N to RLS motif in CmeA. (D) Analysis of outer membrane channel; CmeC, AcrA and OrpM showing the conservation of Gly structurally located at the tip region of coiled-coil α hairpin domain among *Campylobacter* species, *E. coli* and *P. aeruginosa.*

Materials and methods

Bacterial strains and growth conditions

Campylobacter jejuni 11168[10] and its derivatives; *C. jejuni cmeD*::*cat*, *C. jejuni cmeD*::*cat wtcmeABC* and *C. jejuni cmeD*::*cat* g0*cmeABC* were used in this study. *C. jejuni* 11168H was grown on Columbia based agar or Muller Hinton based agar supplemented with 5% horse blood according to manufacturer's instructions. Strains were grown at 37^{o}C in a variable atmospheric incubator (VAIN) cabinet (Don Whitely, UK) maintaining microaerophilic conditions of: 85% Nitrogen, 5% Oxygen and 10% carbon dioxide. All of the cloning experiments were done in *Escherichia coli* DH10Beta (New England Biolabs, USA). *E. coli* DH10B was used in expression of CmeA and cloning and expression of CmeC whilst gCmeA was expressed in *E. coli* SBD1. *E. coli* strains were grown on either Luria-Bertani Broth or Luria-Bertani Agar and antibiotics were added when necessary.

Inactivation of *cmeD* and generation of *C. jejuni cmeD*::*cat*, *C. jejuni cmeD*::*cat* wt*cmeABC* and *C. jejuni cmeD*::*cat g0cmeABC.*

The nucleotide sequence of *cmeD* gene was commercially synthesized (Clonetech, USA) to also carry a chloramphenicol resistant gene; *cat* was inserted in the middle of *cmeD* to disrupt the gene. The DNA was then released by restriction digestion with EcoRV and cloned in pJET1.2 -following manufacturer's instructions- to give pATN. Cloning of *cmeABC-aphA* was achieved by the following; *cmeABC* locus was amplified by primer FWDCmeA and primerREVCmeC with Phusion polymerase (New England biolabs, UK) using *C. jejuni* 11168H genomic DNA as a template, 6Xhis tag was added

at the C-terminus of the CmeC to track its expression. The PCR amplicon was cloned in pJET1.2 following the manufacture's instructions to give pMH3 that was then cut by BamHI to introduce the kanamycin resistant gene *aphA*, to be used as an antibiotic selection marker after homologous recombination in *C. jejuni* 11168H to give pMHT. To add homologous recombination arms for *cmeABC-aphA*, pMH3 was cut by SaCII to ligate *cj0364* at the 3' end of *aphA* to give pMHTF. For g0*cmeABC-aphA,* each asparagine in the non-canonical glycosylation sequon (D/E-X-N-Y-S/T, where X and Y are any amino acid except proline) was altered to glutamine *in-silico* and nucleotide sequence of g0*cmeABC* was synthesized by (Clonetech, USA) DNA was then treated as above to generate pATKH.

To generate *C. jejuni cmeD*::*cat*, electroporation of pATN into *C. jejuni* 11168H was carried out as previously described[10]. The transformants were selected on CBA plates supplemented with 10 µg/ml chloramphenicol and the double cross over event was confirmed by PCR, this strain was then used as parent strain to generate other mutants. Plasmids pMHT and pATK were electroporated into *C. jejuni cmeD*::*cat* to generate *C. jejuni cmeD*::*cat cmeC*::*cmeC-aphA* and *C. jejuni cmeD*::*cat cmeABC*::*cmeABC*-(N->Q)-*aphA,* respectively. Transformants were selected on CBA plates supplemented with 10 µg/ml chloramphenicol and 30 µg/ml kanamycin and the double cross over event was confirmed by PCR.

Antibiotic sensitivity test (E-test)

C. jejuni 11168H were grown in suspension in Mueller-Hinton broth equivalent to 1.0 MacFarland's standard and 100 µl aliquots were spread plated on dry Mueller-Hinton agar plates supplemented with 5 % Sheep blood (Oxoid, UK), the plates were left for 5 -10 minutes to dry before the antibiotic strip (Oxoid, UK) was added. Plates were incubated at 37^{o}C overnight. The minimum inhibitory concentration (MIC) was read

directly from the strip at the oint where the zone of inhibition of bacterial growth intersected with the antibiotic concentration on the strip.

Ethidium bromide accumulation assay

Bacterial cells were grown to mid log phase (OD_{600} 0.4-0.5). Cells were harvested, washed and resuspended in 0.1M sodium phosphate buffer pH 7 (previously incubated in the VAIN) to OD_{600} 0.2. Cells were then incubated in the VAIN for 15 mins at 37°C o before a 100µl aliquot was withdrawn to indicate time zero. Ethidium bromide (Sigma, UK) was added to final concentration 2 µg/ml and fluorescence was measured at 530 nm excitation and 600 nm emission using a plate reader (Molecular Devices M3 plate reader, USA).

Expression of CmeA and gCmeA

Protein expression was carried out in *E. coli* strains unless stated otherwise. CmeA and CmeC were expressed in *E. coli* DH10B carrying pMH5 plasmid and pAT3, respectively, whilst gCmeA was expressed in *E. coli* SDB1 carrying pGVXN114, pWA2 and pACYC*(pgl*). Initiating cultures were grown overnight in LB broth supplemented with appropriate antibiotics at 37 °C under shaking condition. The following day, 10 ml of culture was withdrawn from the shake flask to inoculate 400 ml LB broth supplemented with appropriate antibiotics. To achieve optimal glycosylation of CmeA, PglB was expressed from pGVXN114 by the addition of 0.5 mM ITPG at OD_{600} 0.5-0.6. Cultures were incubated at 37°C for 24 hours with shaking. Cultures were centrifuged and cell pellets washed with binding buffer (300 mM NaCl, 50 mM NaH_2PO_4 with 25mM imidazole) and passed twice through a high pressure cell homogeniser (Stanstead works, UK). Cell debris was removed by centrifugation at 20,000 xg for 45 minutes. Supernatant was collected and incubated with 0.2 ml Ni-

NTA for 1 hour at 4 °C then washed with 50 ml binding buffer and eluted four times in 0.5 ml elution buffer (300 mM NaCl, 50 mM NaH_2PO_4 with 250mM imidazole).

Cloning and expression of CmeC

To express CmeC in *E. coli*, *cmeC* lacking signal peptide sequence was amplified by PCR with CmeCFwd1 and CmeCRev using *C. jejuni* 11168H genomic DNA as a template. The amplicon was then cut by NheI and SalI and cloned into pEC415 downstream of the DsbA signal peptide sequence to give pCMECDSBA. *E. coli* carrying pCMECDSBA was grown in LB media supplemented with ampicillin (100 µg/ml) overnight at 37 °C under shaking condition. On the following day, 10 ml were withdrawn from the overnight culture to inoculate 400 ml LB media. Cells were grown to OD_{600} 0.5-0.6 and 0.2% L-arabinose was added to induce the expression of CmeC. Cultures were incubated at 37 °C for 24 hours with shaking at 180 rpm. Cultures were centrifuged and cell pellets washed with binding buffer (300 mM NaCl, 50 mM NaH_2PO_4 with 25mM imidazole) and passed twice through cell homogeniser (Stanstead works, UK). Cells debris was removed by centrifugation at 20,000 xg for 45 minutes and then collected and incubated in binding buffer with 2 % DDM for 3 hours at 4 °C. The mixture was then centrifuged at 15,000 xg for 10 minutes. The supernatant was collected, diluted with binding buffer and incubated with 0.2 ml Ni-NTA for 1 hour at 4 °C then washed with 50 ml binding buffer and eluted four times in 0.5 ml elution buffer (300 mM NaCl, 50 mM NaH_2PO_4 with 250mM imidazole).

CD Spectroscopy

All CD spectra of gCmeA and CmeA were acquired in 0.5mm rectangular cell pathlength using Chirascan spectrometer (Applied Biophysics, UK) equipped with Quantum NorthWest TC125 Peltier unit. Temperature dependent confirmation changes were monitored at wavelength 260-195nm for gCmeA (0.2 mg/ml) and

CmeA (0.2 mg/ml) in 10 mM Sodium phosphate, 75 mM Sodium chloride, 10 % glycerol buffer (pH=8.0) during stepwise increase in temperature from 6^{o}C to 94^{o}C. Temperatures were measured directly with a thermocouple probe in the sample solution. Melting temperatures were determined from the derivative CD-Temperature spectra and fitted using a Levenberg–Marquardt algorithm (LMA) on the van*'t* Hoff isochore. (Global 3, Global Analysis for T-ramp Version 1.2 built 1786, Applied Photophysics Ltd, 2007-2012). For Conformation Reversibility Study, far-UV CD spectra were recorded at 20^{o}C, raised to T_m and re-cooled to 20^{o}C. The temperature at each elevated T_m was kept constant for 5 minutes and the CD spectrum was recorded to assess the rate of protein unfolding process.

Surface Plasmon Resonance

For coupling of CmeA and gCmeA to the CM5 sensor chip, carboxyl groups on the surface were activated by injecting a 1:1 mixture of 0.4M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) and 0.1 M N-hydroxysuccinimide (NHS) for 7 minutes at 5 µl/min. CmeA and gCmeA were diluted to 10-20 µg/ml in 0.1 M acetate pH 5.5 and immobilised at 5 µl/min. Immobilisation was stopped when the required RU was achieved. This was followed by injecting 1M ethanolamine pH 8.5 (7 minutes at 5 µl/min) to inactivate excess reactive groups. To account for non- specific binding, a control flow cell was generated using the same method described minus the protein immobilisation step. For coupling of CmeA and gCmeA to a NTA chip, the chip was cleaned and loaded with $NiCl_2$ (0.5 mM). The flow cells were then activated as above and CmeA and gCmeA (10 ug/ml in HBSP buffer) were loaded into appropriate flow cells until appropriate RU were achieved. Subsequently the flow cells were treated with ethanolamine as above to block remaining activated sites.

Cmec at various concentrations (3 nM- 0.2 µM) was analysed at a constant temperature of 25°C under continuous flow of HBS-PE buffer (10mM HEPES pH 7.4, 3 mM EDTA, 0.005% (w/v) Surfactant P20 (GE Healthcare) at 30µl/min (sufficient to prevent mass transfer effects) at pH 7.4 for 3 minutes association and a dissociation time of 5 minutes. Experiments at pH 6.0 were performed with 10 mM MES pH 6.0, 3 mM EDTA, 0.005% (w/v) Surfactant P20 (GE Healthcare) The surface chip was regenerated by injecting 0.1 M triethanolamine pH 11.5. Data was analysed using the BIAevaluation software version 4.1.1 (Biacore, GE Healthcare, Amersham). Blank flow cell controls were subtracted. The k_d was defined between 10s after the end of the sample injection and 300 sec later.

Acknowledgements

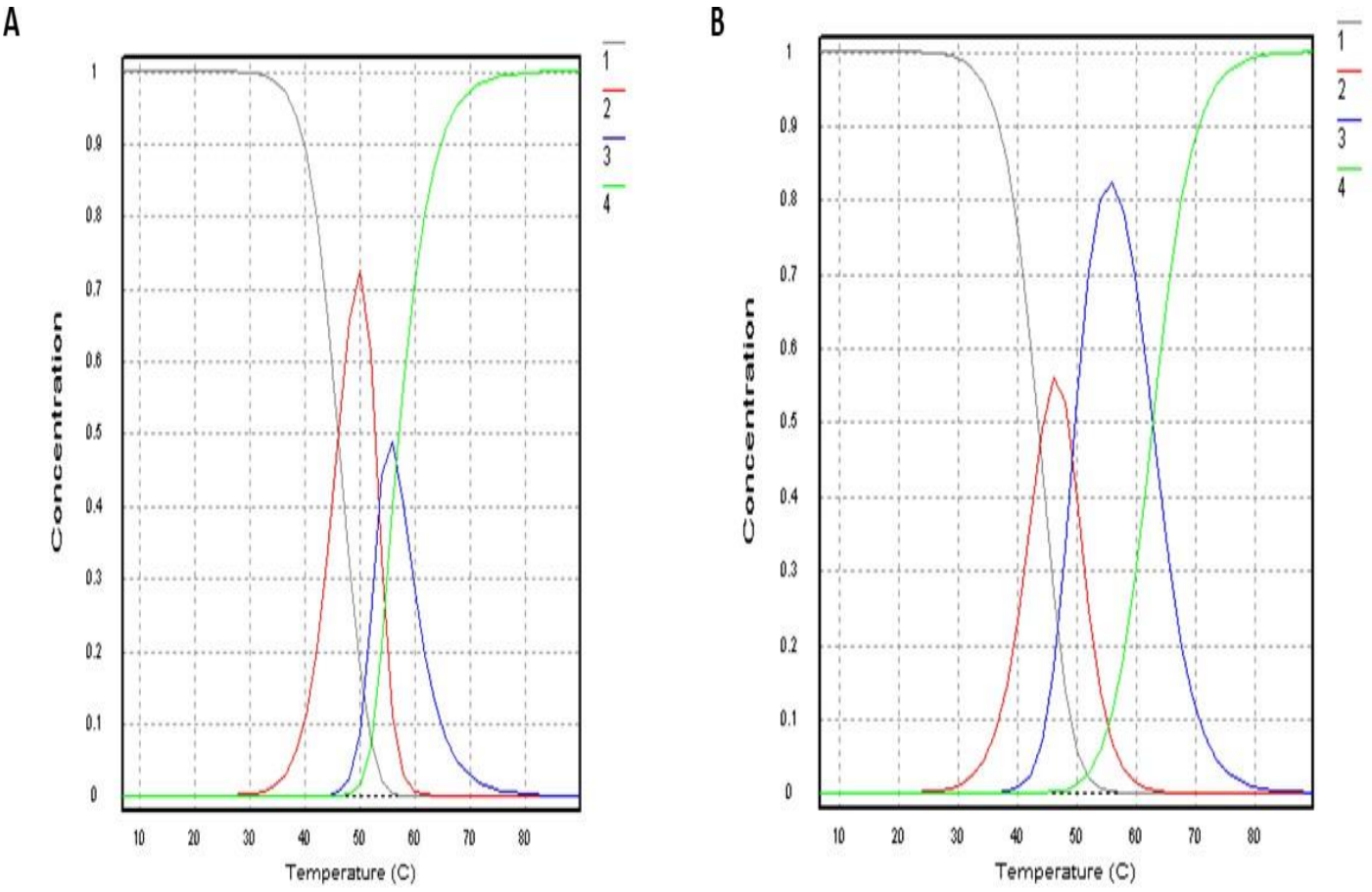
We acknowledge the Wellcome Trust grant 102978/Z/13/Z for funding.

Supplementary figures

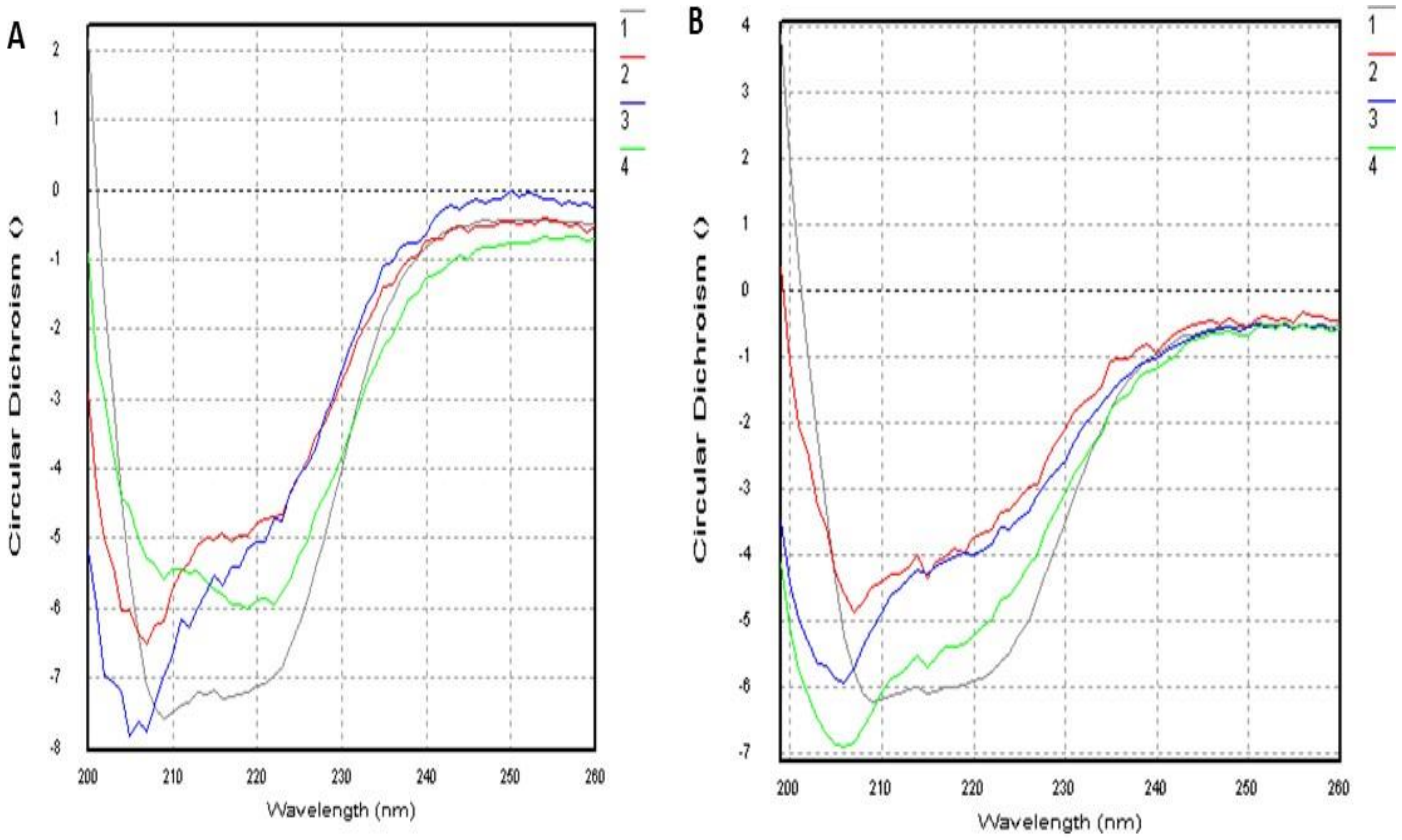
Strain/Plasmid/ Primers	Description	Reference
E. coli DH10B	F^- mcrA Δ(*mrr-hsd*RMS-*mcr*BC) Φ80d*lac*ZΔM15 Δ*lac*X74 *end*A1 *rec*A1 *deo*R Δ(*ara*,*leu*)7697 *ara*D139 *gal*U *gal*K *nup*G *rps*L λ^-	New England Biolabs, UK
E. coli SDB1	F- lambda- IN*(rrnD-rrnE)1 rph-1*, Δ*waaL* Δ*wecA*	30
C. jejuni 11168H	Hypermotile variant of *C. jejuni* 11168	21
C. jejuni 11168H *cmeD*::*cat*	*C. jejuni* 11168H *cmeD* is inactivated by chloramphenicol cassette insertion	This study
C. jejuni cmeD::*cat* wt*cmeABC*	*C. jejuni* 11168H *cmeD*::*cat, cmeC* is 6xhis tagged followed by kanamycin cassette to help for selection of CBA plate	This study
C. jejuni cmeD::*cat* g0*cmeABC*	*C. jejuni* 11168H *cmeD*::*cat, cmeABC* is glycosylation deficient by altering N->Q in *C.jejuni* glycosylation sequon (D/E-X-N-X-S/T where X is any amino acid other than proline)	This study
pGVXN114	PglB cloned in pEXT21 under *lac* promoter	41
pWA2	Soluble periplasmic 6xHis tagged CmeA under Tet promoter in pBR322	32
pMH5	Soluble periplasmic 6xHis tagged CmeA under Tet promoter in pCAYC184	32
pACYC(*pgl*)	*C. jejuni* heptasaccharide coding sequence under Tet promoter in pCAYC184	32
pJMK30	*aphA* gene cloned in BamHI resistriction site	42
pCMECSDBA	Membrane bound 10xHis tagged CmeC driven to periplasm by DsbA signal peptide under L-arabinose promoter in pEC145	This study

pATN	*cmeD*::*cat* cloned in pJET1.2	This study
pMH3	*cmeABC* locus cloned in pJET1.2	This study
pMHT	*aphA* cloned in BamHI site in pMH3	This study
pMHTF	*cj0364* cloned in SacII site in pMHT	This study
pATM	g0*cmeABC* locus cloned in pJET1.2	This study
pATMN	*aphA* cloned in BamHI site in pATM	This study
pATKH	*cj0364* cloned in SacII site in pATMN	This study
FWDCmeA	AGCGAAGTTAAAGAAATTGGAGCAC	
REVCmeC	TTTT*CCGCGG*ATT*GGATCC*CATTATGATGATGATGAT GATGATGTTCTCTAAAGACATATCT	
FWDcj0364	TTTT*CCGCGG*ATTCTCTAAATAAATTAAAAATCTTTG TCT	
REVcj0364	TTTT*CCGCGG*CATTGAACCTTTTTGGAGGGATTTTTC C	
FWDCmeC	TTTT*GCTAGC*GCCGCCCCAAATTTAAATATTCCCGAA GCAAACTATAGCATTG	
REVCmeC	TTTTT*GTCGAC*ctaATGATGATGATGATGATGATGAT GATGATG TTCTCTAAAAGACATATCTAAATTTTTTGATTC	

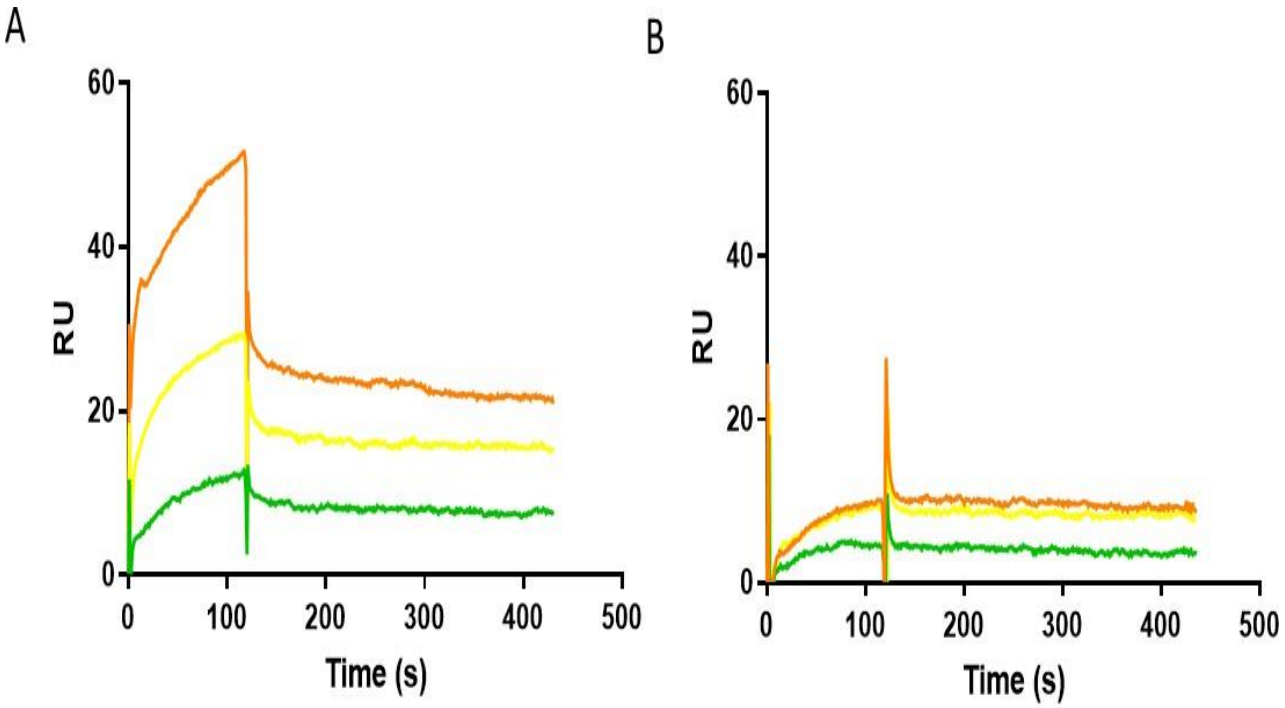
Supplementary figures



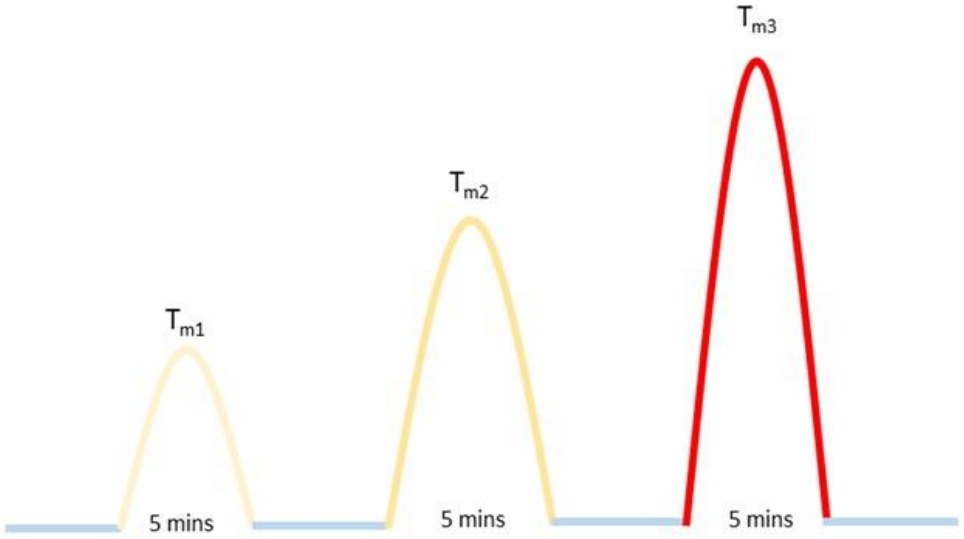
Supplementary 1. Thermal melts of g0CmeA and g2CmeA in 10 mM sodium phosphate, 75 mM sodium chloride and 10% glycerol (pH 8.0). Concentration as function to temperature representing three transition melting phases for g0CmeA ,(**A**); g2CmeA (**B**).



Supplementary 2. Thermal melts of g0CmeA and g2CmeA in 10 mM sodium phosphate, 75 mM sodium chloride and 10% glycerol (pH 8.0). CD spectra as function to temperature representing more than one melting phases for g0CmeA ,(**A**); g2CmeA, (**B**).



Supplementary 3 SPR analysis of A) 1090 RU of immobilised g2cmeA and B) 1000RU of immobilisation g0cmeA binding to cmeC offered at 1 $x10^{-7}$ M (orange); 5 $x10^{-8}$M (yellow) 2.5$x10^{-8}$ M (green) for 2 mins and 5 mins dissociation. CmeA variants were covalently associated by NHS/EDC after association through C-terminal 6Xhis-tag association with the NTA surface.



Supplementary 4 Scheme representing reversibility study CmeA variants were cooled at 20 C (blue) then heated up to T_m for 5 minutes then cooled again at 20 C the corresponding T_m are shown from golden yellow to red, CD spectra were recorded at each temperature.

References

1. Schäffer, C. & Messner, P. Emerging facets of prokaryotic glycosylation. *FEMSMicrobiology Reviews* (2017). doi:10.1093/femsre/fuw036

2. Hanson, S. R. *et al.* The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc. Natl. Acad. Sci. U. S. A.* (2009). doi:10.1073/pnas.0810318105

3. Biswas, H. & Chattopadhyaya, R. Stability of curcuma longa rhizome lectin: Role of n-linked glycosylation. *Glycobiology* (2016). doi:10.1093/glycob/cwv110

4. Mitra, N., Sinha, S., Ramya, T. N. C. & Surolia, A. N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends in Biochemical Sciences* (2006). doi:10.1016/j.tibs.2006.01.003

5. Helenius, A. & Aebi, M. Roles of N-Linked Glycans in the Endoplasmic Reticulum. *Annu. Rev. Biochem.* (2004). doi:10.1146/annurev.biochem.73.011303.073752

6. Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Current Opinion in Structural Biology* (2011). doi:10.1016/j.sbi.2011.08.005

7. Linton, D. *et al.* Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway. *Mol. Microbiol.* (2005). doi:10.1111/j.1365-2958.2005.04519.x

8. Jervis, A. J. *et al.* Characterization of the structurally diverse N-linked glycans of Campylobacter species. *J. Bacteriol.* (2012). doi:10.1128/JB.00042-12

9. Nothaft, H. *et al.* Diversity in the protein N-glycosylation pathways within the Campylobacter genus. *Mol. Cell. Proteomics* (2012).

doi:10.1074/mcp.M112.021519

10. Karlyshev, A. V. The Campylobacter jejuni general glycosylation system is

important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology* **150,** 1957–1964 (2004).

11. Szymanski, C. M., Burr, D. H. & Guerry, P. Campylobacter protein glycosylation affects host cell interactions. *Infect. Immun.* (2002). doi:10.1128/IAI.70.4.2242-2244.2002

12. Larsen, J. C., Szymanski, C. & Guerry, P. N-linked protein glycosylation is required for full competence in Campylobacter jejuni 81-176. *J. Bacteriol.* (2004). doi:10.1128/JB.186.19.6508-6514.2004

13. Lin, J., Sahin, O., Michel, L. O. & Zhang, Q. Critical role of multidrug efflux pump CmeABC in bile resistance and in vivo colonization of Campylobacter jejuni. *Infect. Immun.* (2003). doi:10.1128/IAI.71.8.4250-4259.2003

14. Lin, J., Overbye Michel, L. & Zhang, Q. CmeABC functions as a multidrug efflux system in Campylobacter jejuni. *Antimicrob. Agents Chemother.* (2002). doi:10.1128/AAC.46.7.2124-2131.2002

15. Parkhill, J. *et al.* The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. *Nature* (2000). doi:10.1038/35001088

16. Akiba, M., Lin, J., Barton, Y. W. & Zhang, Q. Interaction of CmeABC and CmeDEF in conferring antimicrobial resistance and maintaining cell viability in Campylobacter jejuni. *J. Antimicrob. Chemother.* (2006). doi:10.1093/jac/dki419

17. Scott, N. E. *et al.* Simultaneous glycan-peptide characterization using hydrophilic interaction 2 chromatography and parallel fragmentation by CID, HCD and ETD-MS applied to the N- 3 linked glycoproteome of Campylobacter jejuni. *Am. Soc. Biochem. Mol. Biol.* (2010). doi:M000031-MCP201

[pii]\n10.1074/mcp.M000031-MCP201

18. Su, C. C. *et al.* Crystal structure of the Campylobacter jejuni CmeC outer membrane channel. *Protein Sci.* (2014). doi:10.1002/pro.2478

19. Garcia-Quintanilla, F., Iwashkiw, J. A., Price, N. L., Stratilo, C. & Feldman, M. F. Production of a recombinant vaccine candidate against Burkholderia pseudomallei exploiting the bacterial N-glycosylation machinery. *Front. Microbiol.* (2014). doi:10.3389/fmicb.2014.00381

20. Silverman, J. M. & Imperiali, B. Bacterial N-glycosylation efficiency is dependent on the structural context of target sequons. *J. Biol. Chem.* (2016). doi:10.1074/jbc.M116.747121

21. Feldman, M. F. *et al.* Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proc. Natl. Acad. Sci.* (2005). doi:10.1073/pnas.0500044102

22. Jayaprakash, N. G. & Surolia, A. Role of glycosylation in nucleating protein folding and stability. *Biochem. J.* (2017). doi:10.1042/BCJ20170111

23. Xin, F. & Radivojac, P. Post-translational modifications induce significant yet not extreme changes to protein structure. *Bioinformatics* (2012). doi:10.1093/bioinformatics/bts541

24. Lee, H. S., Qi, Y. & Im, W. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Sci. Rep.* (2015). doi:10.1038/srep08926

25. Micsonai, A. *et al.* Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci.* (2015). doi:10.1073/pnas.1500851112

26. Tikhonova, E. B., Dastidar, V., Rybenkov, V. V. & Zgurskaya, H. I. Kinetic control of TolC recruitment by multidrug efflux complexes. *Proc. Natl. Acad. Sci.* (2009). doi:10.1073/pnas.0906601106

27. Thomas, R. M. *et al.* Glycosylation of DsbA in Francisella tularensis subsp. Tularensis. *J. Bacteriol.* (2011). doi:10.1128/JB.00438-11

28. Fletcher, C. M., Coyne, M. J., Villa, O. F., Chatzidaki-Livanis, M. & Comstock, L. E. A General O-Glycosylation System Important to the Physiology of a Major Human Intestinal Symbiont. *Cell* (2009). doi:10.1016/j.cell.2009.02.041

29. Fredriksen, L. *et al.* Lactobacillus plantarum WCFS1 O-linked protein glycosylation: An extended spectrum of target proteins and modification sites detected by mass spectrometry. *Glycobiology* (2013). doi:10.1093/glycob/cwt071

30. Horzempa, J., Dean, C. R., Goldberg, J. B. & Castric, P. Pseudomonas aeruginosa 1244 pilin glycosylation: Glycan substrate recognition. *J. Bacteriol.* (2006). doi:10.1128/JB.00273-06

31. Champasa, K., Longwell, S. A., Eldridge, A. M., Stemmler, E. A. & Dube, D. H. Targeted Identification of Glycosylated Proteins in the Gastric Pathogen *Helicobacter pylori* (*Hp*). *Mol. Cell. Proteomics* (2013). doi:10.1074/mcp.M113.029561

32. Lithgow, K. V. *et al.* A general protein O-glycosylation system within the Burkholderia cepacia complex is involved in motility and virulence. *Mol. Microbiol.* (2014). doi:10.1111/mmi.12540

33. Iwashkiw, J. A. *et al.* Identification of a general O-linked protein glycosylation system in Acinetobacter baumannii and its role in virulence and biofilm formation. *PLoS Pathog.* (2012). doi:10.1371/journal.ppat.1002758

34. Vik, A. *et al.* Broad spectrum O-linked protein glycosylation in the human

pathogen Neisseria gonorrhoeae. *Proc. Natl. Acad. Sci. U. S. A.* (2009). doi:10.1073/pnas.0809504106

35. Min, T. *et al.* Specificity of campylobacter jejuni adhesin PEB3 for phosphates and structural differences among its ligand complexes. *Biochemistry* (2009). doi:10.1021/bi802195d

36. Jeong, H. *et al.* Pseudoatomic Structure of the Tripartite Multidrug Efflux Pump AcrAB-TolC Reveals the Intermeshing Cogwheel-like Interaction between AcrA and TolC. *Structure* (2016). doi:10.1016/j.str.2015.12.007

37. Davis, L. M., Kakuda, T. & DiRita, V. J. A Campylobacter jejuni znuA orthologue is essential for growth in low-zinc environments and chick colonization. *J. Bacteriol.* (2009). doi:10.1128/JB.01394-08

38. Kawai, F. *et al.* Crystal structure of JlpA, a surface-exposed lipoprotein adhesin of Campylobacter jejuni. *J. Struct. Biol.* (2012). doi:10.1016/j.jsb.2012.01.001

39. Kakuda, T., Koide, Y., Sakamoto, A. & Takai, S. Characterization of two putative mechanosensitive channel proteins of Campylobacter jejuni involved in protection against osmotic downshock. *Vet. Microbiol.* (2012). doi:10.1016/j.vetmic.2012.04.044

40. Eichler, J. & Koomey, M. Sweet New Roles for Protein Glycosylation in Prokaryotes. *Trends in Microbiology* (2017). doi:10.1016/j.tim.2017.03.001

41. Ihssen, J. *et al.* Production of glycoprotein vaccines in Escherichia coli. *Microb. Cell Fact.* (2010). doi:10.1186/1475-2859-9-61

42. Van Vliet, A. H. M., Wooldridge, K. G. & Ketley, J. M. Iron-responsive gene regulation in a Campylobacter jejuni fur mutant. *J. Bacteriol.* (1998).

43. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* (2011). doi:10.1038/msb.2011.75

43

Chapter 5

FUNCTIONAL ANALYSIS OF N-LINKING OLIGOSACCHARYL TRANSFERASE ENZYMES ENCODED BY DEEP-SEA VENT PROTEOBACTERIA

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Sherif Abouelhadid
Principal Supervisor	Brendan W Wren
Thesis Title	Investigating the biological role of *N*-linked glycan in bacteria

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Glycobiology		
When was the work published?	26 November 2015		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	The candidate retrieved, bioinformatically analysed PglB orthologues, helped in designing experiments and assisted in data analysis.

Student Signature: **Date:**

Supervisor Signatur  **Date: 14/Sept/2018**

Glycobiology, 2016, vol. 26, no. 4, 398–409
doi: 10.1093/glycob/cwv111
Advance Access Publication Date: 26 November 2015
Original Article

OXFORD

Microbial Biology

Functional analysis of N-linking oligosaccharyl transferase enzymes encoded by deep-sea vent proteobacteria

Dominic C Mills[2,4], Adrian J Jervis[3], Sherif Abouelhadid[2], Laura E Yates[2,4], Jon Cuccui[2], Dennis Linton[3], and Brendan W Wren[1,2]

[2]Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, University of London, Keppel Street, London WC1E 7HT, UK, and [3]Faculty of Life Sciences, University of Manchester, Michael Smith Building, Manchester M13 9PT, UK

[1]To whom correspondence should be addressed: Tel: +44-207-927-2288; e-mail: brendan.wren@lshtm.ac.uk

[4]Present address: School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA.

Received 24 September 2015; Revised 16 November 2015; Accepted 20 November 2015

Abstract

Bacterial N-linking oligosaccharyl transferases (OTase enzymes) transfer lipid-linked glycans to selected proteins in the periplasm and were first described in the intestinal pathogen *Campylobacter jejuni*, a member of the ε-proteobacteria-subdivision of bacteria. More recently, orthologues from other ε-proteobacterial *Campylobacter* and *Helicobacter* species and a δ-proteobacterium, *Desulfovibrio desulfuricans*, have been described, suggesting that these two subdivisions of bacteria may be a source of further N-linked protein glycosylation systems. Whole-genome sequencing of both ε- and δ-proteobacteria from deep-sea vent habitats, a rich source of species from these subdivisions, revealed putative ORFs encoding OTase enzymes and associated adjacent glycosyltransferases similar to the *C. jejuni* N-linked glycosylation locus. We expressed putative OTase ORFs from the deep-sea vent species *Nitratiruptor tergarcus*, *Sulfurovum lithotrophicum* and *Deferribacter desulfuricans* in *Escherichia coli* and showed that they were able to functionally complement the *C. jejuni* OTase, CjPglB. The enzymes were shown to possess relaxed glycan specificity, transferring diverse glycan structures and demonstrated different glycosylation sequon specificities. Additionally, a permissive *D. desulfuricans* acceptor protein was identified, and we provide evidence that the N-linked glycan synthesized by *N. tergarcus* and *S. lithotrophicum* contains an acetylated sugar at the reducing end. This work demonstrates that deep-sea vent bacteria encode functional N-glycosylation machineries and are a potential source of biotechnologically important OTase enzymes.

Key words: bacterial glycobiology, deep-sea vent bacteria, glycoengineering, N-linked glycosylation

Introduction

Asparagine-linked glycosylation is a common post-translational modification of eukaryotic proteins, and is involved in many cellular functions such as quality control, protein folding and secretion (Helenius and Aebi 2004). The eukaryotic N-glycosylation machinery is located within the endoplasmic reticulum, where the glycan is assembled on the lipid carrier dolichyl pyrophosphate in the membrane by the action of several glycosyl transferases, flipped across the membrane and finally transferred to an acceptor protein at the consensus sequon N-X-S/T, where X can be any amino acid except proline (Aebi 2013). This transfer is accomplished by the action of the oligosaccharyltransferase (OTase) enzyme complex. Initially believed to be limited to eukaryotes, N-linked glycosylation was also identified in Archaea, where a surface layer (S-layer) protein was shown to be N-glycosylated in

Halobacterium salinarum (Mescher and Strominger 1976). Further investigation revealed that this type of protein modification was present in many Archaeal species, displaying a variety of *N*-glycan structures (Eichler 2013). The S-layer protein is the best characterized Archaeal glycoprotein, but other proteins have been identified such as the ABC transporter SSO1273 in *Sulfolobus solfataricus* P2 (Palmieri et al. 2013) and an unusual type IV pilus protein in *Methanococcus maripaludis* (Ng et al. 2011). It was not until 1999 that the first bacterial N-linked protein glycosylation system was identified in the intestinal pathogen *Campylobacter jejuni* (Szymanski et al. 1999). Intriguingly, while the transfer of the N-linked glycan to acceptor proteins in the eukaryotic model organism, *Saccharomyces cerevisiae* requires the action of the multi-enzyme OTase complex (Aebi 2013), a single enzyme, termed PglB, with significant levels of amino acid similarity to the STT3 subunit of the yeast OTase complex, was sufficient for glycan transfer (Wacker et al. 2002; Young et al. 2002). The functional transfer of the *C. jejuni* N-linked protein glycosylation machinery encoded by the protein glycosylation (*pgl*) locus into *Escherichia coli* allowed structural determination of the glycan and functional characterization of individual gene products involved (Wacker et al. 2002; Linton et al. 2005). Subsequent demonstration of relaxed glycan substrate specificity of the *C. jejuni* OTase PglB (CjPglB) paved the way for the development of recombinant glycoengineering, an approach by which a desired glycoconjugate can be generated entirely in *E. coli* by co-expression of a glycan-coding locus, an acceptor protein and the OTase enzyme (Feldman et al. 2005). The *C. jejuni* N-linked glycosylation acceptor sequon was shown to differ from the eukaryotic N-X-S/T with an extended motif containing a negatively charged amino acid at the −2 position D/E-**Z**-N-**X**-S/T (where **Z**, like **X**, can be any amino acid except proline) (Kowarik, Young, et al. 2006). The most recent survey of *C. jejuni* glycoproteins identified 154 glycopeptides corresponding to 53 glycoproteins, confirming the general nature of the glycosylation machinery (Scott et al. 2011).

Orthologs of the CjPglB enzyme have been found in all *Campylobacter* spp. genome sequences, a subset of *Helicobacter* species, some members of the genus *Desulfovibrio*, as well as some δ- and ε-proteobacterial species found in deep-sea vent habitats (Nothaft and Szymanski 2010). Three of these orthologues have been function- ally characterized; those from *Campylobacter lari* (ClPglB) (Schwarz et al. 2011), *Helicobacter pullorum* (HpPglB) (Jervis et al. 2010) and the δ-proteobacterium, *Desulfovibrio desulfuricans* (DdPglB) (Ielmini and Feldman 2011).

Deep-sea hydrothermal vents are light-independent environments on the sea-floor that host a large community of chemolithoautotrophic proteobacterial symbionts, predominantly ε-proteobacteria (Huber et al. 2007). The genome sequences of several bacterial species isolated from deep-sea vents have been determined (Inagaki et al. 2004; Nakagawa et al. 2005, 2007; Takaki et al. 2010) and several of the species were found to possess orthologs of the *C. jejuni pglB* gene. In the ε-proteobacteria *N. tergarcus* and *S. lithotrophicum*, the region adjacent to the putative *pglB* genes contains genes encoding proteins predicted to be involved in the generation and transfer of nucleotide-activated sugars, indicating the presence of at least a partial *pgl* operon similar to the one encoded by *C. jejuni* (Figure 1A) (Nothaft and Szymanski 2010). In contrast, the *pglB* gene in the δ-proteobacterium *D. desulfuricans* appears to be an orphan gene and not part of a *pgl* operon. The presence of these putative *pglB* genes suggests that these uncharacterized organisms may possess functional protein N-glycosylation machineries. Investigation of the putative N-glycosylation systems directly in these species is complicated by their relatively complex and unusual growth requirements, such as growth in supplemented synthetic sea water and at higher temperatures (Takai et al. 2003; Nakagawa et al. 2005). We therefore characterized the putative OTase enzymes in *E. coli* by co-expression with an acceptor protein and a lipid-linked glycan substrate, as previously reported for the OTase enzymes of *D. desulfuricans* and *H. pullorum* (Jervis et al. 2010; Ielmini and Feldman 2011).

We present the functional expression and characterization of three novel OTase enzymes from deep-sea vent bacteria, identify a possible glycoprotein encoded by one of these species and gain an insight to the nature of the native N-linked glycan structures.

Results

Identification of putative N-linking OTase enzymes from deep-sea vent bacteria

In order to identify putative orthologs of CjPglB, the amino acid sequence was used as the query against all databases of prokaryotic proteins using blastp. Numerous non-*Campylobacter* orthologs were identified, and three encoded by deep-sea vent bacteria were chosen for further analysis (Figure 1B; Supplementary data, Figure 1). These included *Nitratiruptor tergarcus* (54% amino acid similarity, 34% identity to the *C. jejuni* PglB enzyme) and *Sulfurovum lithotro- phicum* (55% similarity, 37% identity), two representatives of the ε-proteobacteria and *Deferribacter desulfuricans* (40% similarity, 24% identity), a species classified within the phylum Deferribacteres from the δ-proteobacteria.

Several amino acid residues and structural features have been identified as important for OTase activity in ClPglB (Lizak et al. 2011; Ihssen et al. 2012, 2014; Gerber et al. 2013). An amino acid alignment of the three deep-sea vent OTase enzymes with CjPglB indicates that the deep-sea vent OTase enzymes possess all the important residues for the function of ClPglB (highlighted in Supplementary data, Figure S2 and Table I), aside from residues R331 and I572 that are absent in DfdPglB.

In most *Campylobacter* species, the OTase gene is present within the locus encoding for the assembly of the N-linked oligosaccharide (Nothaft and Szymanski 2010). The genes flanking the deep-sea vent putative *pglB* genes from *N. tergarcus* (NtPglB) and *S. lithotro- phicum* (SlPglB) include genes predicted to encode an initiating undecaprenol-phosphate sugar phosphotransferase (PglC) and a number of glycosyltransferases. However, no gene encoding a "flippase" enzyme required for membrane translocation of the lipid-linked oligosaccharide (LLO) into the periplasm was identified. In contrast, there are no orthologs of *C. jejuni* N-linked glycosylation pathway genes adjacent to the *D. desulfuricans* (DfdPglB) ortholog (Figure 1A).

The three deep-sea vent OTase enzymes are functional in *E. coli*

For functional analysis, the putative OTase enzymes were tested for their ability to complement CjPglB in *E. coli*. The predicted ORFs coding for NtPglB, SlPglB and DfdPglB were codon-optimized, synthesized and cloned into inducible expression vectors of the pEXT family (see *Methods*) (Dykxhoorn et al. 1996). Initial activity assays tested the ability of the OTase enzymes to transfer the *C. jejuni* N-linked heptasaccharide to the commonly used *C. jejuni* reporter glycoproteins, AcrAand Cj0114. The three deep sea vent OTase enzymes (and CjPglB as a positive control) were co-expressed with hexa-his-tagged AcrA and the *C. jejuni pgl* gene locus with an inser- tionally inactivated *pglB* gene in *E. coli* strain CLM24 (Feldman et al. 2005). Analysis of purified AcrA by western blotting, using

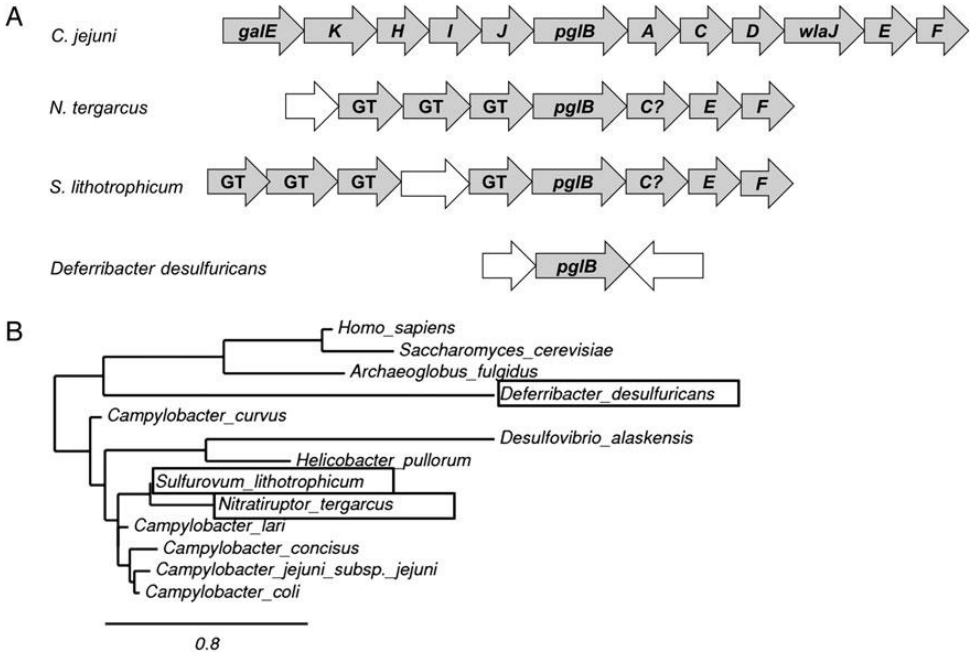


Fig. 1. Analysis of the *pgl* loci of the deep-sea vent bacteria and phylogenetic tree of PglB/AglB/STT3 OTase enzymes (A) Depiction of the genomic localization of the PglB-encoding genes of *C. jejuni*, *N. tergarcus*, *S. lithotrophicum* and *Deferribacter desulfuricans*. GT, glycosyl transferase. White arrows: genes with no involvement in sugar biosynthesis. ORFs annotated as C? encode putative initiating glycosyl transferase enzymes with low homology to *C. jejuni* PglC. (B) Subset view of the phylogenetic tree of the oligosaccharyl transferase protein sequence (PglB, AglB and STT3 subunit) from the three domains of life (Bacteria, Archaea and Eukarya, respectively) (full tree shown in S1). The tree was a CLUSTAL W (Larkin et al. 2007) generated protein sequence alignment using the neighbor-joining (NJ) method with 1000 bootstrap replicates in phylogenetic package MEGA6 (Tamura et al. 2013). The PglB orthologs investigated in this study are boxed.

Table I. Presence of critical amino acid residues in PglB orthologs from *N. tergarcus*, *S. lithotrophicum* and *Deferribacter desulfuricans*

	D56	R147	D154/156	E319	R331	R375	WWDYG	I572
NtPglB	✓	✓	✓	✓	✓	✓	✓	✓
SlPglB	✓	✓	✓	✓	✓	✓	✓	✓
DfdPglB	✓	✓	✓	✓	8	✓	✓	8

anti-hexa-his and anti-heptasaccharide antibodies demonstrated CjPglB-mediated glycosylation of AcrA at two sites as previously reported (Wacker et al. 2002), but no modification was observed by the three putative deep-sea vent OTase enzymes (Figure 2A and B).

Analysis of purified Cj0114 confirmed CjPglB-dependent glycosylation of Cj0114 at four extended N-linked glycosylation sequons as previously reported (Jervis et al. 2010) (Figure 2C and D). Two of the four Cj0114 glycoforms were observed with NtPglB and SlPglB (Figure 2C and D) demonstrating OTase activity of these two enzymes. No extra bands were observed for DfdPglB, suggesting that this enzyme was not able to transfer the *C. jejuni* heptasaccharide to the Cj0114 protein. These data confirm protein N-linked glycosylation activity of two of three deep-sea vent OTase enzymes.

The sequon is required but not sufficient for glycosylation by NtPglB

To further investigate NtPglB-mediated Cj0114 glycosylation, we employed Cj0114 mutants in which each of the asparagine residues within the four glycosylation sequons (N100, N154, N172 and N178) were replaced with glutamine (Jervis et al. 2010). Co-expression of these four variants with pACYCpglΔ*pglB* and CjPglB generated three Cj0114 glycoforms as expected (Figure 3A, left panel). Co- expression of Cj0114 N100Q and N178Q with pACYCpglΔ*pglB* and NtPglB resulted in two glycoforms as for wild-type Cj0114, however, Cj0114 N154Q and N172Q produced only a single glycoform (Figure 3A right panel). This demonstrates that asparagine residues N154 and N172, but not N100 and N178, are required for modification with the *C. jejuni* heptasaccharide by NtPglB. The sequons surrounding N154 and N172 contain an aspartic acid at the −2 position, while the sequons surrounding N100 and N178 contain glu- tamic acid, suggesting that NtPglB may display a preference towards the former. However, replacement of aspartic acid at position −2 of N172 with a glutamic acid (D170E) did not disrupt glycosylation with NtPglB (data not shown). To investigate the requirement for a negatively charged residue at the −2 position for activity of NtPglB and SlPglB, the aspartic acid at the −2 position of N172 was replaced with alanine (mutant D170A). This resulted in one less Cj0114 glycoform produced by CjPglB, NtPglB and SlPglB (Figure 3B), demonstrating that as for CjPglB both NtPglB and SlPglB require a negatively charged amino acid at the −2 position.

The deep-sea vent OTase enzymes can transfer non-campylobacter glycan structures

CjPglB possesses relaxed glycan specificity (Feldman et al. 2005). In order to assess the ability of the three deep-sea vent OTase enzymes to transfer a variety of glycan moieties, the enzymes were co-expressed

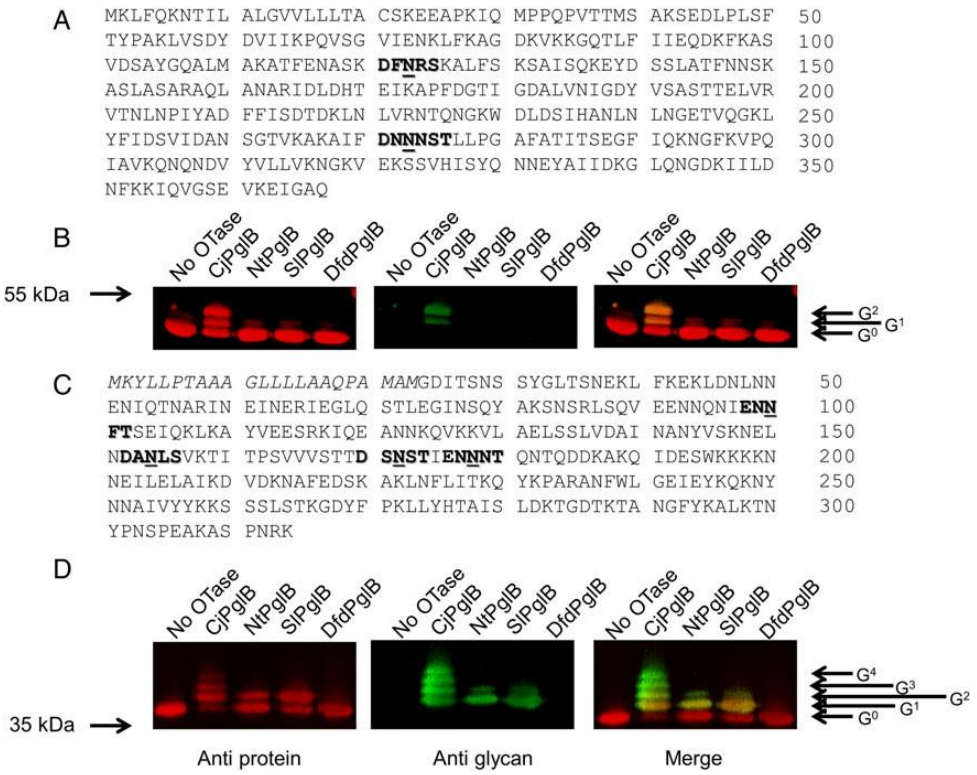


Fig. 2. Functional expression of deep-sea vent OTase enzymes. (A) Amino acid sequence of the *C. jejuni* glycoprotein AcrA. Bold sequence: Two extended bacterial N-linked glycosylation sequons. Underlined amino acids: acceptor asparagines (B) NtPglB, SlPglB and DfdPglB OTase enzymes were co-expressed in *E. coli* CLM24 with plasmids coding for the synthesis of the *C. jejuni* N-linked glycan and the affinity-tagged *C. jejuni* glycoprotein AcrA. Empty vector (no OTase) and CjPglB expressing cultures were included as controls. Following purification, the acceptor proteins were separated by SDS–PAGE, western blotted and probed with antisera against the affinity tag (red) and glycan (green). (C) Amino acid sequence of Cj0114. Italicized sequence: Engineered *pelB* leader sequence for efficient targeting of the protein to the periplasm. Bold sequence: Four extended bacterial glycosylation sequons. Underlined amino acids: acceptor asparagines (D) NtPglB, SlPglB and DfdPglB OTase enzymes were co-expressed in *E. coli* CLM24 with plasmids coding for the synthesis of the *C. jejuni* N-linked glycan and the affinity-tagged *C. jejuni* glycoprotein Cj0114. Empty vector (no OTase) and CjPglB expressing cultures were included as controls. After purification, the acceptor proteins were separated by SDS–PAGE, western blotted and probed with antisera against the affinity tag (red) and glycan (green).

with Cj0114 in *E. coli* strain E69 that synthesizes the *O*9 *O*-antigen with *N*-acetylglucosamine (GlcNAc) at the reducing end (McCallum et al. 1989), and in *E. coli* strain CLM24 producing the *F. tularensis O*-antigen with QuiNAc (2-acetamido-2,6-dideoxy-*O*-D-glucose) at the reducing end (Cuccui et al. 2013). All three OTase enzymes were able to transfer both structures to Cj0114 (Figure 4), demonstrating that similar to CjPglB, the deep-sea vent OTase enzymes possess relaxed glycan specificity. Interestingly, both NtPglB and SlPglB preferentially transferred shorter chains of the *F. tularensis O*-antigen compared with CjPglB, while the transfer efficiency of DfdPglB appeared very low.

Analysis of putative glycoproteins encoded by the deep-sea vent bacteria

As Cj0114 was glycosylated by the deep-sea vent OTase enzymes in the recombinant *E. coli* system, a BLAST search was performed to identify potential orthologs encoded by the three species. A Cj0114 ortholog was identified in each species, and designated Nt0114 (34% identity to Cj0114), Sl0114 (30% identity to Cj0114) and Dfd0114 (29% identity to Cj0114). Only Nt0114 contained an extended bacterial N-glycosylation sequon, whilst Sl0114 and Dfd0114 contained two and three eukaryotic N-X-S/T sequons, respectively (Figure 5, Supplementary data, Figures S3 and S4).

Both Nt0114 and Dfd0114 were tested for their capacity to be glycosylated through expression in N-linked glycosylation competent *E. coli* CLM24 producing either the *C. jejuni* N-linked heptasaccharide glycan or *F. tularensis O*-antigen. In this system, Nt0114 was glycosylated with the *C. jejuni* heptasaccharide and *F. tularensis O*-antigen by CjPglB but not by any of the three deep-sea OTase enzymes (Supplementary data, Figure S3). Interestingly, when Dfd0114 lacking extended bacterial N-linked glycosylation sequons was co-expressed in *E. coli* CLM24 with OTase enzymes and the *C. jejuni* heptasaccharide, a glycosylated form was detected in the presence of CjPglB, suggesting transfer of glycan to a eukaryotic sequon (Figure 5B). When Dfd0114 was co-expressed with the *F. tularensis O*-antigen and the OTase enzymes, glycoforms were observed in the presence of CjPglB and DfdPglB, but not NtPglB or SlPglB (Figure 5B). This suggested glycan transfer by both CjPglB and DfdPglB to a eukaryotic-type sequon. CjPglB was previously believed to strictly require a negatively charged residue at the −2 position of the sequon (Kowarik, Young, et al. 2006). However, recent work has demonstrated modification of asparagine residues not located within an extended sequon by CjPglB, both in the native host as well as in *E. coli* (Ollis et al. 2014; Scott et al. 2014). To identify the Dfd0114 sequon glycosylated by CjPglB, site-directed mutant constructs were generated by individually changing the asparagine residue in each of the three

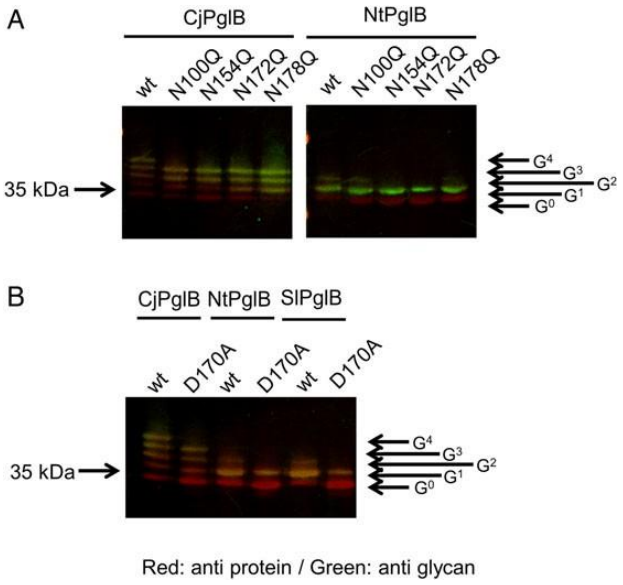


Fig. 3. Analysis of the sequon usage by NtPglB and requirement for a negative charge at the −2 position for NtPglB and SlPglB. (A) CjPglB and NtPglB were co-expressed in *E. coli* CLM24 with plasmids coding for the synthesis of the *C. jejuni* N-linked glycan and four site-directed mutant versions of affinity-tagged *C. jejuni* glycoprotein Cj0114, where the acceptor asparagine residue of each site had been replaced with a glutamine. After purification, the acceptor proteins were separated by SDS–PAGE, western blotted and probed with antisera against the affinity tag (red) and glycan (green). (B) CjPglB, NtPglB and SlPglB were co-expressed in *E. coli* CLM24 with a plasmid coding for the synthesis of the *C. jejuni* N-linked glycan and either the wild-type Cj0114 protein or a site-directed mutant version of Cj0114 where the aspartic acid residue at the −2 position of sequon 3 had been replaced with an alanine (D170A).

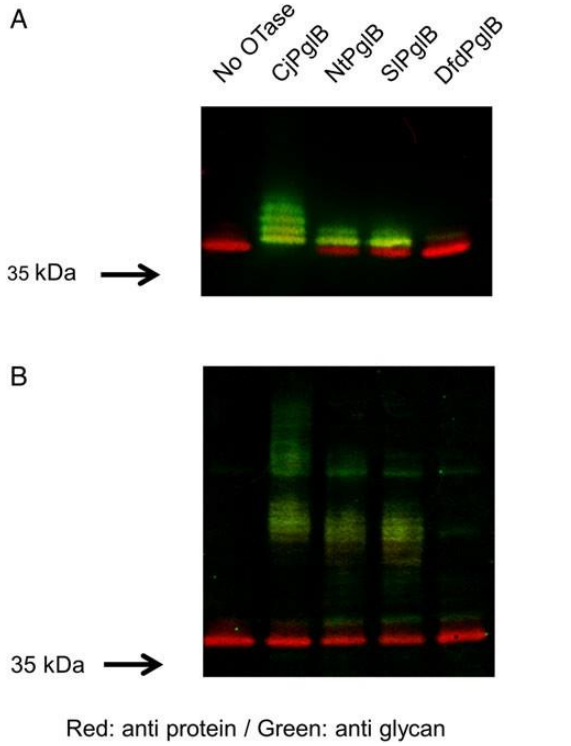


Fig. 4. NtPglB, SlPglB and DfdPglB possess relaxed glycan specificity. NtPglB, SlPglB and DfdPglB OTase enzymes were co-expressed with affinity-tagged *C. jejuni* glycoprotein Cj0114 in (A) *E. coli* E69 which synthesizes O9 *O*-antigen and (B) *E. coli* CLM24 containing a plasmid coding for the synthesis of the *F. tularensis O*-antigen. After purification, the acceptor proteins were separated by SDS–PAGE, western blotted and probed with antisera against the affinity tag (red) and glycan (green).

eukaryotic sequons to alanine. The mutant N107A was no longer glycosylated by CjPglB, while mutants N101A and N118A remained glycosylated, suggesting transfer of the glycan to asparagine N101 by CjPglB (Figure 5C).

Functional characterization of putative initiating glycosyl transferase enzymes involved in generation of the N-linked LLO in *N. tergarcus* and *S. lithotrophicum*

Unlike *C. jejuni*, the genes involved in the synthesis of the N-linked LLO in *N. tergarcus*, *S. lithotrophicum* and *Deferribacter desulfuricans* are not encoded within a single locus (Nothaft and Szymanski 2010) (Figure 1A). However, a putative undecaprenol-phosphate UDP-glycosyl transferase (*pglC*) is located immediately downstream of *pglB* in both *N. tergarcus* and *S. lithotrophicum* (Figure 1A). Additionally, orthologs of two of the three enzymes involved in the synthesis of UDP-*N*,*N*′-diacetylbacillosamine (diNAcBac) from UDP-*N*-acetylglucosamine (UDP-GlcNAc), *pglE* and *pglF*, but not *pglD* were identified downstream of *pglB* and *pglC* (Morrison and Imperiali 2014). No orthologs of *pglC*, *E* or *F* were identified from *D. desulfuricans*, suggesting that the native glycan transferred in this organism is unlikely to contain diNAcBac at the reducing end. Recombinant expression of NtPglC or SlPglC together with Cj0114 and pACYCpgl*pglC*::Km in *E. coli* strain CLM37, which lacks the initiating transferase of the *O*-antigen biosynthesis repeat unit synthesis machinery (WecA), resulted in Cj0114 glycosylation with a glycan recognized by the anti-*C. jejuni* heptasaccharide antiserum HR6 (Figure 6A). This demonstrates that both deep-sea vent PglC orthologs can functionally complement Cj PglC and are able to transfer a glycan moiety to undecaprenol-phosphate, on which the full *C. jejuni* heptasaccharide can be synthesized by enzymes encoded by the *C. jejuni pgl* locus. It has previously been shown that CjPglB can transfer heptasaccharides with either a diNAcBac or a GlcNAc residue at the reducing end (Linton et al. 2005). However, CjPglC has been shown to catalyze only the transfer of UDP-Bac to undecaprenol-phosphate (Glover et al. 2006). Elucidation of the nature of the reducing end glycan requires mass spectrometric analysis; however, the full-length Cj0114 protein was unsuitable for mass spectrometry analysis. A modified C-terminal truncation of Cj0114 termed NGRP, which displays improved performance of the N172 tryptic peptide during MS analysis has previously been generated (A. J. Jervis, unpublished). This protein was used in an analogous PglC complementation experiment, and the resultant glycoform subjected to tandem MS analysis to identify the nature of the glycan at the reducing end of the oligosaccharide (Figure 6B). This showed the presence of a 228 Da residue consistent with a diNAcBac at the reducing end of the glycan, demonstrating that this was the substrate for both NtPglC and SlPglC. Two further minor peaks were observed in the NtPglC sample. The peak at 2822 corresponds to a *C. jejuni* heptasaccharide with a HexNAc at the reducing end based on mass difference. The peak at 2832 either corresponds to a *C. jejuni* heptasaccharide with an unknown glycan at the reducing end or is the result of partial fragmentation. This result suggests that NtPglC may also be able to transfer a HexNAc residue to undecaprenol-phosphate with low efficiency. Unfortunately, the intensity of the peaks was too low to sequence the glycan and confirm this hypothesis.

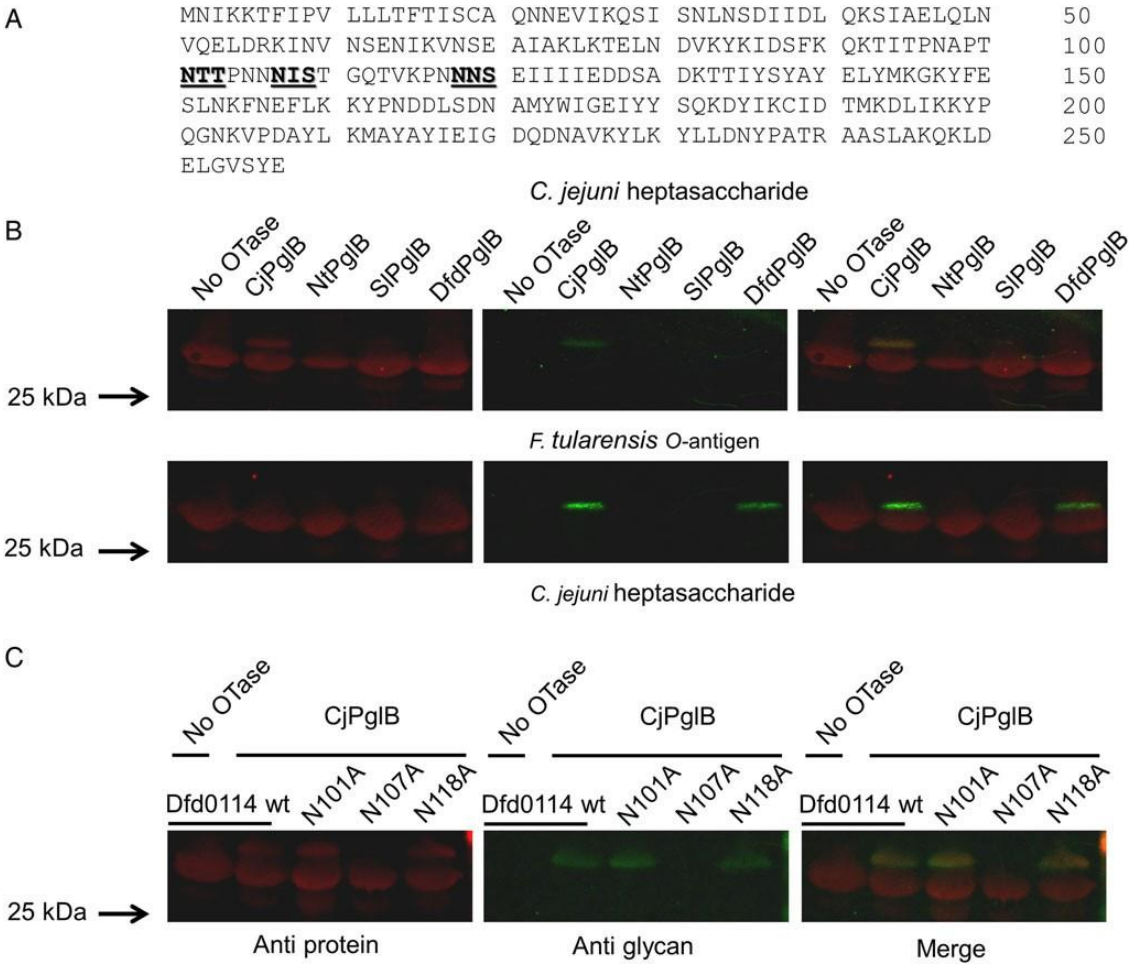


Fig. 5. A putative deep-sea vent glycoprotein encoded by *D. desulfuricans* is glycosylated by CjPglB and DfdPglB at a non-canonical asparagine. (A) Amino acid sequence of Dfd0114. (B) Affinity-tagged Dfd0114 was co-expressed with the OTase enzymes in *E. coli* CLM24 with a plasmid coding for the synthesis of the *C. jejuni* N-linked glycan (C) Affinity-tagged Dfd0114 was co-expressed with the OTase enzymes in *E. coli* CLM24 containing a plasmid coding for the synthesis of the *F. tularensis O*-antigen. (D) CjPglB was co-expressed in *E. coli* CLM24 with a plasmid coding for the synthesis of the *C. jejuni* N-linked glycan and either the wild-type or three site-directed mutant versions of affinity-tagged *C. jejuni* glycoprotein Dfd0114, where the putative acceptor asparagine residue had been replaced with an alanine.

Discussion

The discovery and functional characterization of a bacterial N-linked general protein glycosylation system in *C. jejuni* (Szymanski et al. 1999; Wacker et al. 2002) challenged the dogma that this type of post-translational protein modification was limited to the eukaryotic and archaeal kingdoms. Since then, further bacterial N-linking OTase enzymes have been functionally characterized (Jervis et al. 2010; Ielmini and Feldman 2011; Schwarz et al. 2011). Intriguingly, genome sequencing of three bacterial species from a deep-sea vent habitat identified genes encoding putative orthologs of the *C. jejuni* OTase PglB (Nakagawa et al. 2007; Nothaft and Szymanski 2010; Takaki et al. 2010). We report the functional characterization of the OTase orthologs encoded by these three species (Figure 1) using a recombinant approach in *E. coli*.

We have demonstrated that the *N. tergarcus*, *S. lithotrophicum* and *D. desulfuricans pglB* orthologs encode functional OTase enzymes that are able to transfer lipid-linked oligo- and polysaccharides to an acceptor protein. However, while it has been shown that CjPglB is able to glycosylate any N-glycosylation sequon as long as the protein is targeted to the periplasm and the glycosylation sequon is present within a flexible, accessible loop (Kowarik, Numao, et al. 2006; Kowarik, Young, et al. 2006; Fisher et al. 2011), our data suggest a more stringent acceptor protein requirement for the three deep-sea vent OTase enzymes, as only one acceptor protein, Cj0114, was glycosylated and at only two of the four possible sites. Such an acceptor protein specificity has not been reported previously for N-linking OTase enzymes, but has been demonstrated for two bac- terial O-linking OTase enzymes (Horzempa et al. 2006; Harding et al. 2015). It is therefore possible that the deep-sea vent enzymes present a new class of N-linking OTase enzymes with more stringent acceptor protein specificity.

Even within the one protein that was successfully glycosylated by the deep-sea vent OTase enzymes, a preference for particular sequons was observed. This preference was not a result of the primary amino acid sequence of the sequon, as altering the non-modified sequons did not result in glycosylation. The sequon preference was also not due to general unavailability of the sequon within the secondary structure of the protein, as CjPglB was able to glycosylate all four sequons. It has previously been shown that the *H. pullorum* PglB1 was only able to glycosylate two of the four sequons within the Cj0114 protein (Jervis

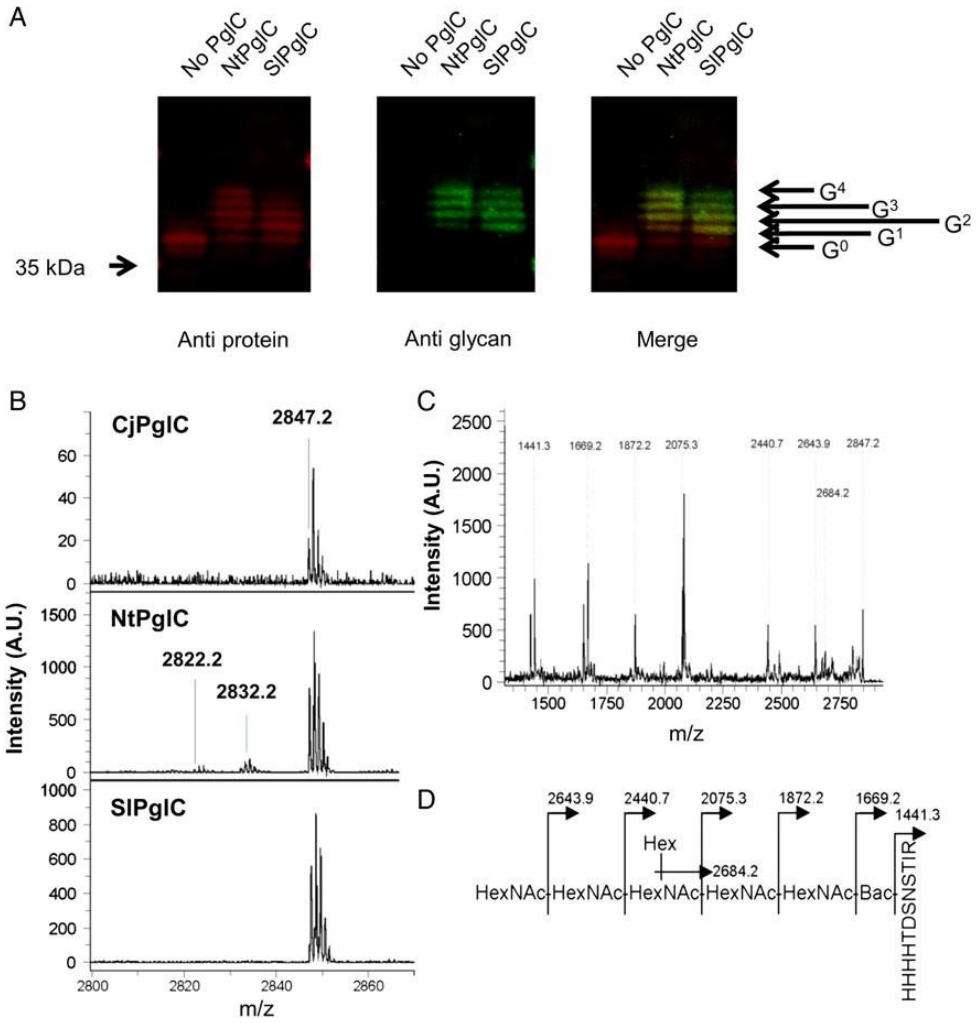


Fig. 6. Complementation of *C. jejuni* PglC by two deep-sea orthologs. (A) NtPglC and SlPglC were co-expressed in *E. coli* CLM37 (Δ*wecA*) with plasmid pACYCpgl*pglC*::Km and affinity-tagged *C. jejuni* glycoprotein Cj0114. After purification, the acceptor proteins were separated by SDS–PAGE, western blotted and probed with antisera against the affinity tag (red) and glycan (green). (B) MALDI-MS analysis of the reporter protein NGRP glycosylated by the *C. jejuni* Pgl system in *E. coli* and an isogenic *pglC* mutant complemented by PglC orthologues from deep-sea species *N. tergarcus* and *S. lithotrophicum* in *E. coli* CLM37. (B) Full mass spectra of tryptic digests in the *m/z* range containing the glycosylated peptide from reporter protein NGRP. The peak with an *m/z* of 2847.2 corresponds to the peptide modified with *C. jejuni* heptasaccharide with the reducing end sugar diNAcBac as confirmed by MALDI-LIFT-TOF/TOF MS (C) and represented in (D). The low-intensity peak with an *m/z* of 2822.2 corresponds to the predicted mass of peptide glycosylated with a heptasaccharide containing a HexNAc at the reducing end.

et al. 2010), and alternative glycosylation of the *C. jejuni* protein AcrA by the enzymes from *C. lari* and *D. desulfuricans* has also been reported (Ielmini and Feldman 2011; Schwarz et al. 2011), suggesting that sequon usage may be variable among bacterial *N*-OTase enzymes.

As both NtPglB and SlPglB possess the R331 amino acid residue (Table I) which is implicated in the interaction of ClPglB with the negatively charged residue at the −2 position of the extended glycosylation sequon (Lizak et al. 2011), it was unsurprising that glycosylation by both enzymes required this negative charge. R331 is absent in DfdPglB, similar to the ortholog encoded by *D. desulfuricans*, which is able to glycosylate sequons lacking the negatively charged residue at the −2 position (Ielmini and Feldman 2011). However, it was not possible to identify the Cj0114 sequon preference of this OTase due to low levels of activity. Similar to CjPglB (Wacker et al. 2006; Cuccui et al. 2013), all three OTase enzymes displayed relaxed glycan specificity, and were able to transfer both the *E. coli* O9 *O*-antigen and the *F. tularensis O*-antigen.

In order to support the hypothesis of a fully functional N-glycosylation machinery encoded by deep-sea vent bacteria, we sought to identify putative native glycoproteins. We identified orthologs of Cj0114 in all three bacterial species. The *S. lithotrophicum* ortholog (Sl0114) does not contain any glycosylation sequons, while the *N. tergarcus* protein (Nt0114) contains one extended sequon and the *D. desulfuricans* ortholog (Dfd0114) contains three eukaryotic sequons lacking the −2 negatively charged residues. Recombinant expression of Dfd0114 with the *C. jejuni* N-linked heptasaccharide and CjPglB resulted in transfer of the glycan to Dfd0114 at an aspara- gine residue within the sequence PNNNIS. The *C. lari* PglB has been shown to glycosylate asparagine residues not located within an ex- tended bacterial sequon both in vivo (Schwarz et al. 2011) and in vitro (Gerber et al. 2013) and recent evidence suggests that CjPglB has similar activity, both in the native host as well as in a recombinant *E. coli* system (Ollis et al. 2014; Scott et al. 2014). No glycosylation of Dfd0114 with the *C. jejuni* heptasaccharide was observed for

DfdPglB. However, low levels of Ddf0114 glycosylation by DfdPglB was observed when co-expressed with the *F. tularensis O*-antigen. This indicates that DfdPglB, similar to the enzyme from *Desulfovibrio desulfuricans* is able to glycosylate short, eukaryotic se- quons (Ielmini and Feldman 2011). Further studies in *Deferribacter desulfuricans* are required to confirm the glycosylation status of Dfd0114 in the native organism as well as to identify the native glycoproteome and investigate the presence of eukaryotic glycosylation sequons within these proteins. In contrast, Nt0114 was only glycosylated by CjPglB, and not by any of the deep-sea vent OTase enzymes, including the "native" NtPglB (Supplementary data, Figure S3). Glycosylation of Nt0114 by CjPglB demonstrated that the protein is present in the correct subcellular compartment and the sequon is in principle accessible to the OTase enzymes. This suggests that the lack of glycosylation of Nt0114 by the deep-sea vent OTase enzymes is likely due to the more stringent acceptor protein specificity of these enzymes. This does not, however, rule out the possibility that this protein may be glycosylated in *N. tergarcus*. Further studies in *N. tergarcus* are required to address this and to identify the complete native glycoproteome.

No data are available regarding N-linked glycan structure in *N. tergarcus* or *S. lithotrophicum*. To investigate the nature of the sugar residue present at the reducing end of the *N*-glycan produced by these species, a CjPglC complementation experiment using the two putative initiating glycosyl transferase enzymes encoded by *N. tergarcus* and *S. lithotrophicum* was performed. This demonstrated that both NtPglC and SlPglC are able to transfer a diNAcBac residue to undecaprenol-phosphate, on top of which the remaining heptasaccharide was assembled. In the case of NtPglC, a small amount of glycan likely containing a HexNAc at the reducing end as judged by mass difference was also observed, suggesting that NtPglC is able to transfer a HexNAc to undecaprenol-phosphate at a very low rate. Analysis of the structure of the glycans in the native organisms, or attempts to reconstitute the complete *N*-glycan biosynthesis pathway in *E. coli* are required to fully investigate the structure of the glycan synthesized by the two species.

The demonstration of functional N-linked glycosylation systems encoded by deep-sea vent bacteria raises interesting questions regarding the evolution of this post-translational modification system and the role of N-linked protein glycosylation in the biology of these species. While this type of post-translational modification is ubiquitous among higher organisms, and has also been found to be encoded by 166 of 168 archaeal genome sequences obtained to date (Kaminski, Lurie-Weinberger, et al. 2013), it has so far only been shown in a small number of bacterial species (Jervis et al. 2010; Nothaft and Szymanski 2010; Ielmini and Feldman 2011; Schwarz et al. 2011). In eukaryotes, the role of the *N*-glycan is multifunctional, ranging from protein quality control to secretion and interaction between proteins, and the modification is essential for the function of the cell (Aebi 2013). In the archaeal species *H. volcanii* and members of the genus *Methanococcus*, the N-glycosylation machinery is not essential for cell viability (Jarrell et al. 2010). However, disruption of N-glycosylation in *H. volcanii* led to decreased ability to grow in high salt concentrations (Kaminski, Naparstek, et al. 2013). Additionally, *H. volcanii* cells containing a disrupted OTase gene were unable to produce intact flagella and were non-motile (Tripepi et al. 2012). Glycosylation of the S-layer protein in *H. volcanii* was recently shown to be dependent on salinity levels, suggesting a role in survival in the relatively harsh environment (Kaminski, Guan, et al. 2013). It has been demonstrated that CjPglB can function as a hydrolase in addition to an OTase, resulting in the release of the *N*-glycan as a free oligosaccharide into the periplasm of the bacterium (Nothaft et al. 2009). This release has been shown to be influenced by altering the salt and osmolyte concentration of the environment, suggesting an adaptive function similar to that observed for archaea. It can be speculated that the OTase enzymes encoded by deep-sea species, living predominantly under harsh environmental conditions such as high temperature and high osmolarity, may contribute to the survival in those conditions by generation of free oligosaccharides to counteract the high osmotic levels of sea water. It has also been shown that a glycosylated form of the *C. jejuni* glycoprotein PEB3 is more thermostable than a non-glycosylated protein (Min et al. 2009). Therefore, protein glycosylation in the deep-sea vent bacteria may increase the overall thermostability of the glycosylated subsection of the proteome.

Studies in *C. jejuni* using cells deficient in either the OTase CjPglB or enzymes involved in synthesis of the lipid-linked heptasaccharide have suggested roles for N-linked glycosylation in chicken colonization and adhesion and invasion of epithelial cells (Szymanski et al. 2002; Karlyshev et al. 2004). However, as a total number of 53 *C. jejuni* proteins have been shown to be N-glycosylated (Scott et al. 2011), it has not been possible to identify the precise role of glycosylation in this pathogen. It has also been demonstrated that the *C. jejuni N*-glycan is recognized by the human galactose-type lectin MGL, suggesting a potential role for the glycan in modulation of the immune system (van Sorge et al. 2009). As many deep-sea vent bacterial species can exist as free-living biomass as well as symbionts on other deep-sea vent animals such as polychaetes and shrimps (Polz and Cavanaugh 1995), it can be speculated that the bacteria may employ an N-linked glycan to interact with and modulate the immune system of their symbiotic partners. However, more studies in the native organisms are needed to address the function and scope of protein glycosylation in these species.

From a glycoengineering point of view, it is interesting to note that two of the three organisms that encode the OTase enzymes characterized here require higher growth temperature than those used for the recombinant expression in the *E. coli* host. This has two potential implications. Firstly, these enzymes may possess different biophysical characteristics compared with the best studied CjPglB, such as better thermostability/"shelf life". Secondly, the activity of the recombinant OTase enzymes in the in vivo *E. coli* expression host may be different from the activity in the native host. Further studies in the native organisms, as well as adaptation of published in vitro glycosylation experiments (Kowarik, Numao, et al. 2006; Jervis et al. 2010) for testing of these OTase enzymes under different conditions are required to investigate these possibilities.

In summary, this is the first functional characterization of bacterial OTase enzymes encoded by three bacterial species from a deep-sea vent habitat and paves the way for further studies of the role of protein N-glycosylation in these specialized bacteria and other bacterial species such as *C. jejuni*. Additionally, the study provides a deeper understanding of a biotechnologically important class of enzymes.

Materials and methods

Bacterial strains, plasmids and growth conditions

Escherichia coli strains were grown in Lysogeny Broth (LB) or on LB agar at 37°C. Where required, the medium was supplemented with antibiotics at the following concentrations: 100 μg mL^{-1} ampicillin, 34 μg mL^{-1} chloramphenicol, 50 μg mL^{-1} kanamycin and 100 μg mL^{-1} spectinomycin. *Escherichia coli* DH5α library efficiency cells (Invitrogen, Carlsbad) were routinely used as a host for cloning experiments. *Escherichia coli* strain E69 was kindly provided by Chris Whitfield (University of Guelph, Canada). All strains and plasmids are listed in Table II.

Table II. Bacterial strains and plasmids

Bacterial strains	Genotype	Source
E. coli DH5α	F-φ80*lac*ZΔM15Δ (*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r^-, m^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ$^-$	Invitrogen
E. coli E69	Prototroph; O9a:K30:H12	Homonylo et al. (1988)
E. coli CLM24	W3110 derivative, Δ*waaL*	Feldman et al. (2005)
E. coli CLM37	W3110 derivative, Δ*wecA*	Linton et al. (2005)
Plasmids		
pBR322	Expression vector, constitutive expression, Amp^R	Bolivar et al. (1977)
pMLBAD	Expression vector, arabinose inducible, Tmp^R	Lefebre and Valvano (2002)
pEXT20	Expression vector, IPTG inducible, Amp^R	Dykxhoorn et al. (1996)
pEXT21	Expression vector, IPTG inducible, $Spec^R$	Dykxhoorn et al. (1996)
pEXT22	Expression vector, IPTG inducible, Kan^R	Dykxhoorn et al. (1996)
pNtPglB21	Codon-optimized *N. tergarcus pglB* cloned into pEXT21	This study
pSlPglB21	Codon-optimized *S. lithotropicum pglB* cloned into pEXT21	This study
pNtPglB22	Codon-optimized *N. tergarcus pglB* cloned into pEXT22	This study
pSlPglB22	Codon-optimized *S. lithotropicum pglB* cloned into pEXT22	This study
pDfdPglB21	Codon-optimized *D. desulfuricans pglB* cloned into pEXT21	This study
pDfdPglB22	Codon-optimized *D. desulfuricans pglB* cloned into pEXT22	This study
pGVXN114	HA-tagged *C. jejuni* PglB cloned into pEXT21	Ihssen et al. (2010)
pMAF10	HA-tagged PglB from *C. jejuni* cloned in pMLBAD, Tmp^R	Feldman et al. (2005)
pACYCpgl*pglB*Kan	*pgl* locus from *C. jejuni* cloned into pACYC184, Kan^R transposon in *pglB*	Linton et al. (2005)
pGAB2	Recombinant expression of the *F. tularensis O*-antigen	Cuccui et al. (2013)
pWA2	His-tagged soluble AcrA expressed from pBR322	Feldman et al. (2005)
pBRCj0114	His-tagged Cj0114 constitutively expressed from pBR322	Jervis et al. (2010)
pBRCj0114N100Q	Derivative of pBRCj0114 with mutated glycosylation site: $N_{100}Q$	This study
pBRCj0114N154Q	Derivative of pBRCj0114 with mutated glycosylation site: $N_{154}Q$	This study
pBRCj0114N172Q	Derivative of pBRCj0114 with mutated glycosylation site: $N_{172}Q$	This study
pBRCj0114N178Q	Derivative of pBRCj0114 with mutated glycosylation site: $N_{178}Q$	This study
pBRCj0114D172E	Derivative of pBRCj0114 with mutated glycosylation site: $D_{172}E$	This study
pBRCj0114D172A	Derivative of pBRCj0114 with mutated glycosylation site: $D_{172}A$	This study
pBRCj0114N172QN100Q	Derivative of pBRCj0114 with mutated glycosylation sites: $N_{172}QN_{100}Q$	This study
pBRCj0114N172QN178Q	Derivative of pBRCj0114 with mutated glycosylation sites: $N_{172}QN_{178}Q$	This study
pBRDfd0114	His-tagged Dfd0114 constitutively expressed from pBR322	This study
pBRDfd0114N101Q	Derivative of pBRDfd0114 with mutated glycosylation site: $N_{101}Q$	This study
pBRDfd0114N107Q	Derivative of pBRDfd0114 with mutated glycosylation site: $N_{107}Q$	This study
pBRNt0114	His-tagged Nt0114 constitutively expressed from pBR322	This study
pBRDfd0114N118Q	Derivative of pBRDfd0114 with mutated glycosylation site: $N_{118}Q$	This study
pMLBADNtPglC	Codon-optimized *N. tergarcus pglC* cloned into pMLBAD	This study
pMLBADSlPglC	Codon-optimized *S. lithotrophicum pglC* cloned into pMLBAD	This study
pNGRP	His-tagged, truncated version (amino acids 1–180) of Cj0114 in pEXT20	This study

Synthesis and subcloning of genes from deep-sea vent bacteria

The ORFs encoding NtPglB, SlPglB and DfdPglB, as well as the putative acceptor proteins were codon-optimized for expression in *E. coli*, synthesized by Celtek Genes (Celtek-genes) and delivered in vector pGH flanked by recognition sites for restriction endonucleases SacI and XbaI for the *pglB* genes and EcoRI and XbaI for the acceptor protein and *pglC* genes. The ORFs were subcloned by restriction digestion into vectors pEXT20, pEXT21, pEXT22 or pMLBAD (Table II) and sequence verified.

Functional analysis of the deep-sea vent PglB proteins

Transfer of the *C. jejuni* heptasaccharide in *E. coli*

Escherichia coli CLM24 cells were transformed with plasmids pACYCpglΔ*pglB*, along with a plasmid encoding an acceptor protein (see Table II for details) and either a control plasmid or a plasmid encoding a PglB. One colony was grown in LB broth to an OD_{600} of 0.4–0.6 at 37°C, and *pglB* expression induced with 1 mM IPTG. Cultures were grown for an additional 16 h, the cells harvested by centrifugation, lysed (see below) and acceptor proteins purified by immobilized metal ion affinity chromatography (IMAC).

Transfer of the *E. coli* O9 *O*-antigen polysaccharide

Plasmid pBRCj0114 was co-expressed with a plasmid encoding the desired PglB protein in *E. coli* strain E69 which synthesizes the O9 *O*-antigen, and induction and sample preparation were performed as above.

Transfer of the *F. tularensis O*-antigen polysaccharide

Escherichia coli CLM24 cells were transformed with plasmids pGAB2, pBRCj0114 and either a control plasmid or a plasmid encod- ing the desired PglB protein. Protein expression and sample prepar- ation were performed as above.

Purification of acceptor proteins using IMAC

Briefly, cell pellets obtained after overnight induction of the acceptor proteins were resuspended in lysis solution (500 mM NaCl, 25 mM NaH_2PO_4, 15 mM imidazole containing 1 mg/mL lysozyme, pH 7.5), and sonicated. Lysates were clarified by centrifugation, and

Table III. Antibodies used in this study

Antibody	Dilution	Source/reference
Mouse anti-hexahistidine monoclonal	1:1000	Abcam
Rabbit anti-hexahistidine polyclonal	1:2000	Abcam
Mouse anti-*F. tularensis* LPS monoclonal	1:1000	Abcam
Rabbit anti-*C. jejuni* heptasaccharide polyclonal	1:500	Schwarz et al. (2011)
Rabbit anti *E. coli* O9 LPS polyclonal	1:500	Cuthbertson et al. (2005)
IRDye® 680RD Goat-anti mouse	1:20,000	LI-COR Bioscience
IRDye® 680RD Goat-anti rabbit	1:20,000	LI-COR Bioscience
IRDye® 800 CW Goat-anti mouse	1:20,000	LI-COR Bioscience
IRDye® 800 CW Goat-anti rabbit	1:20,000	LI-COR Bioscience

Ni-NTA agarose added. After 1 h with mixing, the slurry was loaded onto a Pierce Spin cup, washed five times with wash solution (500 mM NaCl, 25 mM NaH_2PO_4, 25 mM imidazole, pH 7.5), and bound proteins eluted with elution buffer (500 mM NaCl, 25 mM NaH_2PO_4, 500 mM imidazole, pH 7.5).

Analysis of glycosylated acceptor protein

Purified acceptor proteins were separated by SDS–PAGE using NuPAGE™ Novex™ 4–12% Bis-Tris protein gels (Invitrogen), transferred to nitrocellulose membrane and analyzed by two color immunoblot using anti-hexahistidine, anti-glycan and corresponding fluorescent-labelled secondary antibodies (Table III) using an Odyssey near-infrared imager (LI-COR Biosciences).

Site-directed mutagenesis of acceptor sequons

Site-directed mutagenesis of acceptor protein sequons was performed using the QuikChange XL site-directed mutagenesis kit (Agilent) according to the manufacturer's instructions.

MALDI-MS analysis of glycopeptides

Coomassie-stained bands in SDS–PAGE gels were excised and subjected to in-gel trypsin digestion followed by cleanup on a C_{18} Zip-Tip (Millipore). MALDI-TOF MS and MALDI-LIFT-TOF/TOF MS spec- tra were acquired by laser-induced dissociation using a Bruker Ultra- flex II mass spectrometer in the positive-ion reflection mode with 2,5-dihydroxybenzoic acid (20 mg mL^{-1} in 0.1% formic acid, 30% acetonitrile) as the matrix. Data were analyzed with FlexAnalysis 3.0 software (Bruker Daltonics).

Supplementary data

Supplementary data for this article are available online at http://glycob.oxfordjournals.org/.

Funding

This work was supported by the Wellcome Trust (grant 102978/Z/13/Z) and the Biotechnology and Biological Sciences Research Council (grant BB/H017437/1).

Acknowledgements

Plasmid pGVXN114 encoding CjPglB was obtained from Glycovaxyn AG, Switzerland under MTA. We thank David Knight and the staff of the Biomolecular Analysis Core Research Facility, Faculty of Life Sciences, University of Manchester, for assistance with MS.

Conflict of interest statement

None declared.

Abbreviations

GlcNAc, *N*-acetylglucosamine; IMAC, immobilized metal ion affinity chromatography; LB, Lysogeny Broth; LLO, lipid-linked oligosaccharide; diNAcBac, N,N'-diacetylbacillosamine, OTase, oligosaccharyltransferase; UDP-GlcNAc, UDP-*N*-acetylglucosamine.

References

Aebi M. 2013. N-Linked protein glycosylation in the ER. *Biochim Biophys Acta*. 1833:2430–2437.

Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heyneker HL, Boyer HW, Crosa JH, Falkow S. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene*. 2:95–113.

Cuccui J, Thomas RM, Moule MG, D'Elia R V, Laws TR, Mills DC, Williamson D, Atkins TP, Prior JL, Wren BW. 2013. Exploitation of bacterial N-linked glycosylation to develop a novel recombinant glycoconjugate vaccine against Francisella tularensis. *Open Biol*. 3:130002.

Cuthbertson L, Powers J, Whitfield C. 2005. The C-terminal domain of the nucleotide-binding domain protein Wzt determines substrate specificity in the ATP-binding cassette transporter for the lipopolysaccharide O-antigens in Escherichia coli serotypes O8 and O9a. *J Biol Chem*. 280:30310–30319.

Dykxhoorn DM, St Pierre R, Linn T. 1996. A set of compatible tac promoter expression vectors. *Gene*. 177:133–136.

Eichler J. 2013. Extreme sweetness: protein glycosylation in archaea. *Nat Rev Microbiol*. 11:151–156.

Feldman MF, Wacker M, Hernandez M, Hitchen PG, Marolda CL, Kowarik M, Morris HR, Dell A, Valvano MA, Aebi M. 2005. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proc Natl Acad Sci USA*. 102:3016–3021.

Fisher AC, Haitjema CH, Guarino C, Çelik E, Endicott CE, Reading CA, Merritt JH, Ptak AC, Zhang S, DeLisa MP. 2011. Production of secretory and extracellular N-linked glycoproteins in Escherichia coli. Appl Environ Microbiol. 77:871–881.

Gerber S, Lizak C, Michaud G, Bucher M, Darbre T, Aebi M, Reymond J-L, Locher KP 2013. Mechanism of bacterial oligosaccharyltransferase: In vitro quantification of sequon binding and catalysis. *J Biol Chem*. 288: 8849–8861.

Glover KJ, Weerapana E, Chen MM, Imperiali B. 2006. Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the Campylobacter jejuni N-linked glycosylation pathway. *Biochemistry* 45:5343–5350.

Harding CM, Nasr MA, Kinsella RL, Scott NE, Foster LJ, Weber BS, Fiester SE, Actis LA, Tracy EN, Munson RS, et al. 2015. Acinetobacter strains carry two functional oligosaccharyltransferases, one devoted exclusively to type IV pilin, and the other one dedicated to O-glycosylation of multiple pro- teins. *Mol Microbiol*. 96:1023–1041.

Helenius A, Aebi M. 2004. Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem*. 73:1019–1049.

Homonylo MK, Wilmot SJ, Lam JS, MacDonald LA, Whitfield C. 1988. Monoclonal antibodies against the capsular K antigen of Escherichia coli (O9: K30(A):H12): Characterisation and use in analysis of K antigen organisation on the cell surface. *Can J Microbiol*. 34:1159–1165.

Horzempa J, Comer JE, Davis SA, Castric P. 2006. Glycosylation substrate specificity of Pseudomonas aeruginosa 1244 pilin. *J Biol Chem*. 281: 1128–1136.

Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML. 2007. Microbial population structures in the deep marine biosphere. *Science (New York, NY)*. 318:97–100.

Ielmini MV, Feldman MF. 2011. Desulfovibrio desulfuricans PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology*. 21:734–742.

Ihssen J, Kowarik M, Dilettoso S, Tanner C, Wacker M, Thöny-Meyer L. 2010. Production of glycoprotein vaccines in Escherichia coli. *Microbial Cell Factories*. 9:61.

Ihssen J, Kowarik M, Wiesli L, Reiss R, Wacker M, Thöny-Meyer L. 2012. Structural insights from random mutagenesis of Campylobacter jejuni oligosaccharyltransferase PglB. *BMC Biotechnol*. 12:67.

Inagaki F, Takai K, Nealson KH, Horikoshi K. 2004. Sulfurovum lithotrophicum gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the epsilon-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. *Int J Syst Evol Microbiol*. 54:1477–1482.

Jarrell KF, Jones GM, Kandiba L, Nair DB, Eichler J. 2010. S-layer glycoproteins and flagellins: Reporters of archaeal posttranslational modifications. *Archaea*. 2010. pii: 612948.

Jervis AJ, Langdon R, Hitchen P, Lawson AJ, Wood A, Fothergill JL, Morris HR, Dell A, Wren B, Linton D. 2010. Characterization of N-linked protein glycosylation in Helicobacter pullorum. *J Bacteriol*. 192:5228–5236.

Kaminski L, Guan Z, Yurist-Doutsch S, Eichler J. 2013. Two distinct N-glycosylation pathways process the Haloferax volcanii S-layer glycoprotein upon changes in environmental salinity. *mBio*. 4:e00716–13.

Kaminski L, Lurie-Weinberger MN, Allers T, Gophna U, Eichler J. 2013. Phylogenetic- and genome-derived insight into the evolution of N-glycosylation in Archaea. *Mol Phylogenet Evol*. 68:327–339.

Kaminski L, Naparstek S, Kandiba L, Cohen-Rosenzweig C, Arbiv A, Konrad Z, Eichler J. 2013. Add salt, add sugar: N-glycosylation in Halo- ferax volcanii. *Biochemical Society Transactions*. 41:432–435.

Karlyshev AV, Everest P, Linton D, Cawthraw S, Newell DG, Wren BW. 2004. The Campylobacter jejuni general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology (Reading, England)*. 150:1957–1964.

Kowarik M, Numao S, Feldman MF, Schulz BL, Callewaert N, Kiermaier E, Catrein I, Aebi M. 2006. N-Linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science (New York, NY)*. 314: 1148–1150.

Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, et al. 2006. Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J*.25:1957–1966.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*. 23:2947–2948.

Lefebre MD, Valvano MA. 2002. Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in Burkholderia cepacia complex isolates. *Appl Environ Microbiol*. 68:5956–5964.

Linton D, Dorrell N, Hitchen PG, Amber S, Karlyshev AV, Morris HR, Dell A, Valvano MA, Aebi M, Wren BW. 2005. Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway. *Mol Microbiol*. 55:1695–1703.

Lizak C, Gerber S, Numao S, Aebi M, Locher KP. 2011. X-ray structure of a bacterial oligosaccharyltransferase. *Nature*. 474:350–355.

Lizak C, Gerber S, Zinne D, Michaud G, Schubert M, Chen F, Bucher M, Darbre T, Zenobi R, Reymond J-L, et al. 2014. A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. *J Biol Chem*. 289:735–746.

McCallum KL, Laakso DH, Whitfield C. 1989. Use of a bacteriophage-encoded glycanase enzyme in the generation of lipopolysaccharide O side chain deficient mutants of Escherichia coli O9:K30 and Klebsiella O1:K20: Role of O and K antigens in resistance to complement-mediated serum killing. *Can J Microbiol*. 35:994–999.

Mescher MF, Strominger JL. 1976. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of Halobacterium salinarium. *J Biol Chem*. 251:2005–2014.

Min T, Vedadi M, Watson DC, Wasney GA, Munger C, Cygler M, Matte A, Young NM. 2009. Specificity of Campylobacter jejuni adhesin PEB3 for phosphates and structural differences among its ligand complexes. *Biochemistry*. 48:3057–3067.

Morrison MJ, Imperiali B. 2014. The renaissance of bacillosamine and its derivatives: pathway characterization and implications in pathogenicity. *Biochemistry* 53:624–638.

Nakagawa S, Takai K, Inagaki F, Horikoshi K, Sako Y. 2005. Nitratiruptor tergarcus gen. nov., sp. nov. and Nitratifractor salsuginis gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the epsilon-Proteobacteria isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough. *Int J Syst Evol Microbiol*. 55:925–933.

Nakagawa S, Takaki Y, Shimamura S, Reysenbach A-L, Takai K, Horikoshi K. 2007. Deep-sea vent epsilon-proteobacterial genomes provide insights into emergence of pathogens. *Proc Natl Acad Sci USA*. 104:12146–12150.

Ng SYM, Wu J, Nair DB, Logan SM, Robotham A, Tessier L, Kelly JF, Uchida K, Aizawa S-I, Jarrell KF. 2011. Genetic and mass spectrometry ana- lyses of the unusual type IV-like pili of the archaeon Methanococcus maripaludis. *J Bacteriol*. 193:804–814.

Nothaft H, Liu X, McNally DJ, Li J, Szymanski CM. 2009. Study of free oligosaccharides derived from the bacterial N-glycosylation pathway. *Proc Natil Acad Sci USA*. 106:15019–15024.

Nothaft H, Szymanski CM. 2010. Protein glycosylation in bacteria: Sweeter than ever. *Nat Rev Microbiol*. 8:765–778.

Ollis AA, Zhang S, Fisher AC, DeLisa MP. 2014. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat Chem Biol*. 10:816–822.

Palmieri G, Balestrieri M, Peter-Katalinić J, Pohlentz G, Rossi M, Fiume I, Pocsfalvi G. 2013. Surface-exposed glycoproteins of hyperthermophilic Sulfolobus solfataricus P2 show a common N-glycosylation profile. *J Proteome Res*. 12:2779–2790.

Polz MF, Cavanaugh CM. 1995. Dominance of one bacterial phylotype at a Mid-Atlantic Ridge hydrothermal vent site. *Proc Natl Acad Sci USA*. 92:7232–7236.

Schwarz F, Lizak C, Fan Y-Y, Fleurkens S, Kowarik M, Aebi M. 2011. Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. *Glycobiology*. 21:45–54.

Scott NE, Parker BL, Connolly AM, Paulech J, Edwards AVG, Crossett B, Falconer L, Kolarich D, Djordjevic SP, Højrup P, et al. 2011. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of Campylobact. *Mol Cell Proteomics*. 10:M000031-–MCP201.

Scott NE, Marzook NB, Cain JA, Solis N, Thaysen-Andersen M, Djordjevic SP, Packer NH, Larsen MR, Cordwell SJ. 2014. Comparative proteomics and glycoproteomics reveal increased N-linked glycosylation and relaxed sequon specificity in Campylobacter jejuni NCTC11168 O. *J Proteome Res*. 13:5136–5150.

Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P. 1999. Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol Microbiol*. 32:1022–1030.

Szymanski CM, Burr DH, Guerry P. 2002. Campylobacter protein glycosylation affects host cell interactions. *Infect Immun*. 70:2242–2244.

Takai K, Kobayashi H, Nealson KH, Horikoshi K. 2003. Deferribacter desulfuricans sp. nov., a novel sulfur-, nitrate- and arsenate-reducing thermophile isolated from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol*. 53:839–846.

Takaki Y, Shimamura S, Nakagawa S, Fukuhara Y, Horikawa H, Ankai A, Harada T, Hosoyama A, Oguchi A, Fukui S, et al. 2010. Bacterial lifestyle in a deep-sea hydrothermal vent chimney revealed by the genome sequence of the thermophilic bacterium Deferribacter desulfuricans SSM1. *DNA Res*. 17:123–137.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 30:2725–2729.

Tripepi M, You J, Temel S, Önder Ö, Brisson D, Pohlschröder M. 2012. N-Glycosylation of Haloferax volcanii flagellins requires known Agl proteins and is essential for biosynthesis of stable flagella. *J Bacteriol*. 194:4876–4887.

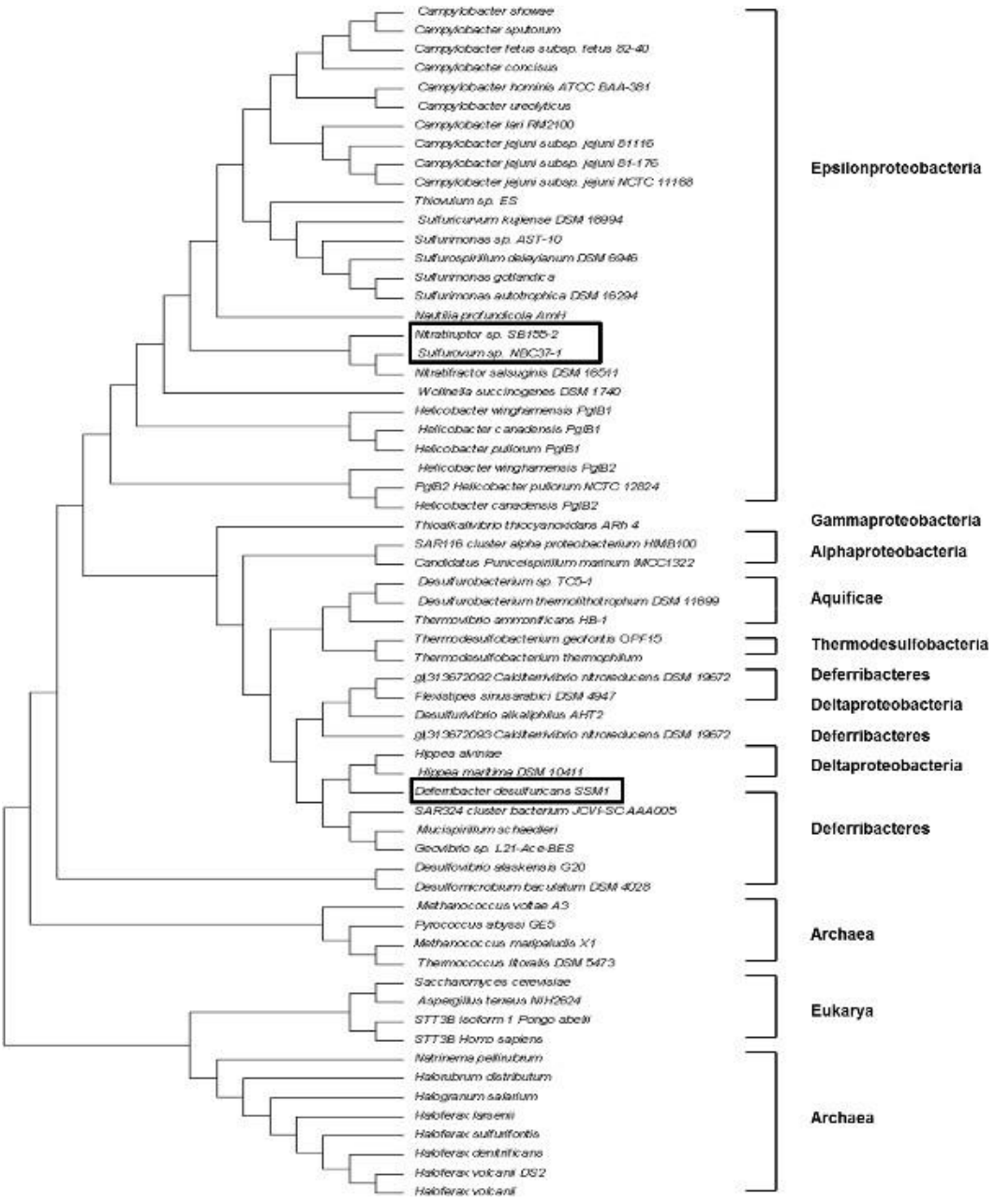
van Sorge NM, Bleumink NMC, van Vliet SJ, Saeland E, van der Pol W-L, van Kooyk Y, van Putten JPM. 2009. N-Glycosylated proteins and distinct lipooligosaccharide glycoforms of Campylobacter jejuni target the human C-type lectin receptor MGL. *Cell Microbiol*. 11: 1768–1781.

Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, et al. 2002. N-Linked glycosyla- tion in Campylobacter jejuni and its functional transfer into *E. coli*. *Science (New York, NY)*. 298:1790–1793.

Wacker M, Feldman MF, Callewaert N, Kowarik M, Clarke BR, Pohl NL, Hernandez M, Vines ED, Valvano MA, Whitfield C, et al. 2006. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc Natl Acad Sci USA*. 103:7088–7093.

Young NM, Brisson J-R, Kelly J, Watson DC, Tessier L, Lanthier PH, Jarrell HC, Cadotte N, St Michael F, Aberg E, et al. 2002. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. *J Biol Chem*. 277:42530–42539.

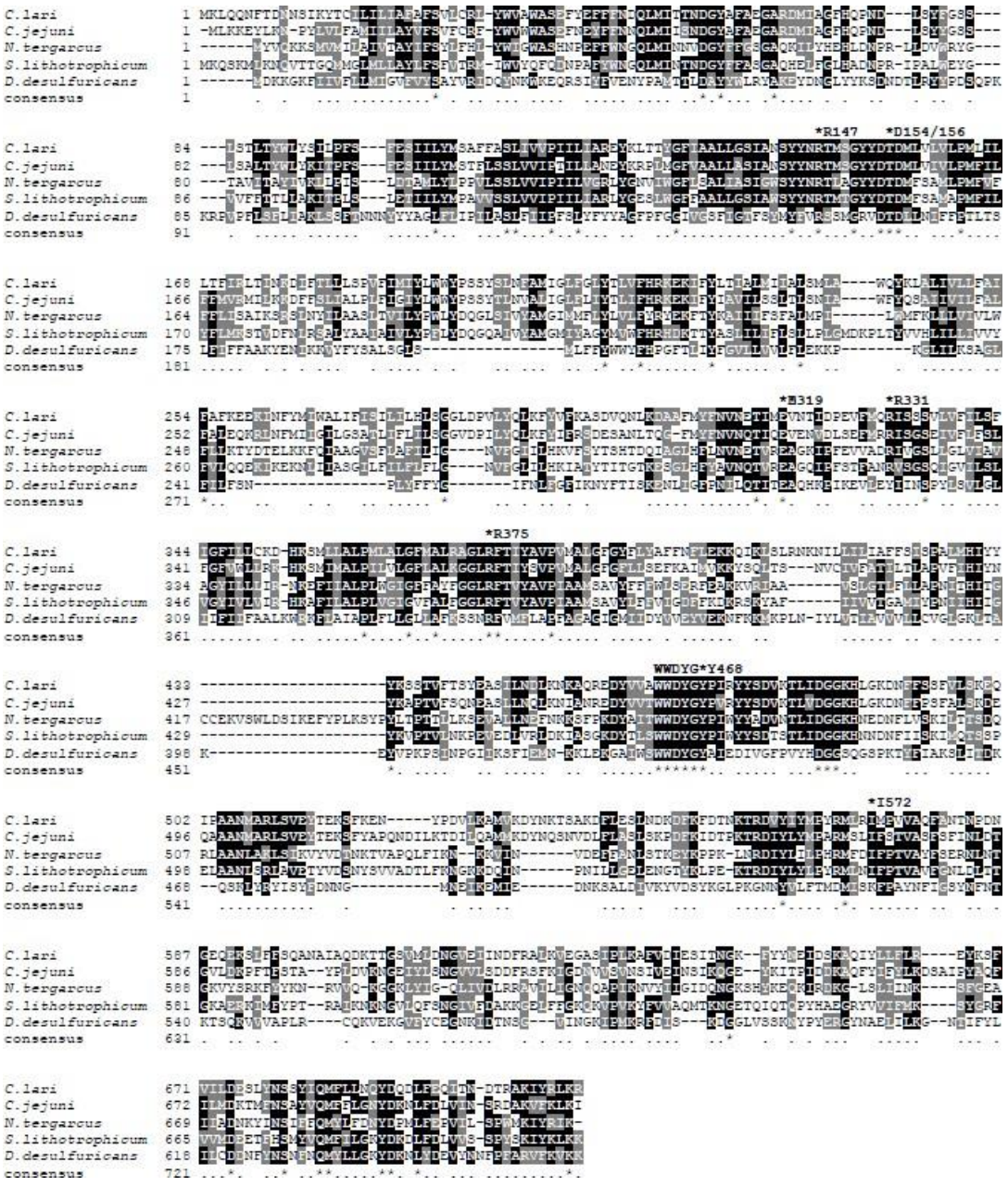
Supplementary data



Supplementary figure 1 : A phylogenetic tree of the oligosaccharyltransferase protein sequence (PglB, AglB, STT3 subunit) from the three domains of life (Bacteria, Archaea, Eukaryotes, respectively). The tree was generated from multiple protein sequence alignment using CLUSTALW[(1)] followed by neighbor-joining (NJ) method with 1000 bootstrap replicates using MEGA6[(2)]. The bar represents changes per 100 position. The PglB orthologues investigated in this study are highlighted (waiting to know which PglBs were used in the study)

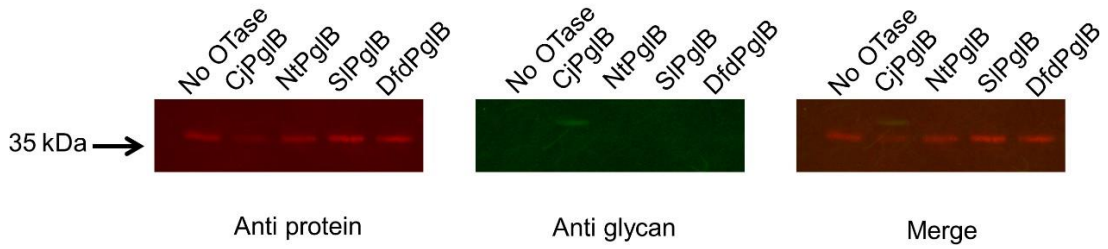
1- Higgins D., Thompson J., Gibson T. Thompson J. D., Higgins D. G., Gibson T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix. choice. *Nucleic Acids Res* 22:4673-4680. (1994)

2-Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*30: 2725-2729. (2013)



Supplementary figure 2 Protein alignment of PglB orthologues used in the study

```
MRYLVLFLAS ILMAAEPSVF EAGNLDSDQP YGLTQSEKHI LQNKKAIKRL QNSLYTVQSK   50
VERIDERVKG LESIVEGLDE NVRDLKKRLK EKSGNDEKIA QLQRELNASM NIQKENFEQI   100
KKILKELSSL IDQINSTYVT KDELKSELSK IYALIKKQNV SKKSGAQLFK EALHSFRKKE   150
YEKAKELFSY AIQKHYKPAT SNFYIAESCY YQKNYGCAVK HYKKSASLYQ KASYMPTLLL   200
HTAISLEKLG RKKEAKKFYR NLVSLYPKSK AAKIAKKKLK                        250
```



Anti protein Anti glycan Merge

Chapter 6

THE *N*-LINKING GLYCOSYLATION SYSTEM FROM *ACTINOBACILLUS PLEUROPNEUMONIAE* IS REQUIRED FOR ADHESION AND HAS POTENTIAL USE IN GLYCOENGINEERING

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Sherif Abouelhadid
Principal Supervisor	Brendan W Wren
Thesis Title	Investigating the biological role of *N*-linked glycan in bacteria

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Glycobiology		
When was the work published?	26 November 2015		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	The candidate reconstructed NGT-AGT in *E. coli,* produced CmeA glycosylated with hexose polymer and assisted in data analysis.

Student Signature **Date:**

Supervisor Signature:  **Date: 14/Sept/2018**

OPEN BIOLOGY

rsob.royalsocietypublishing.org

Research

Cite this article: Cuccui J *et al*. 2017 The *N*-linking glycosylation system from *Actinobacillus pleuropneumoniae* is required for adhesion and has potential use in glycoengineering. *Open Biol.* 7: 160212. http://dx.doi.org/10.1098/rsob.160212

Received: 18 July 2016
Accepted: 28 November 2016

Subject Area:
biochemistry/biotechnology/molecular biology/microbiology/genetics

Keywords:
N-linked glycosylation, *Actinobacillus pleuropneumoniae*, adhesion

Author for correspondence:
Brendan W. Wren
e-mail: brendan.wren@lshtm.ac.uk

†These authors contributed equally to this study.
‡Present address: Division of Infection Medicine, Department of Clinical Sciences, Lund University, Sölvegatan 19, 221 84 Lund, Sweden.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3655586.

The *N*-linking glycosylation system from *Actinobacillus pleuropneumoniae* is required for adhesion and has potential use in glycoengineering

Jon Cuccui[1,†], Vanessa S. Terra[1,†], Janine T. Bossé[2], Andreas Naegeli[3,‡], Sherif Abouelhadid[1], Yanwen Li[2], Chia-Wei Lin[3], Prerna Vohra[1], Alexander W. Tucker[4], Andrew N. Rycroft[5], Duncan J. Maskell[4], Markus Aebi[3], Paul R. Langford[2] and Brendan W. Wren[1] on behalf of the BRaDP1T Consortium

[1]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
[2]Section of Paediatrics, Department of Medicine, Imperial College London, St. Mary's Campus, London W2 1PG, UK
[3]Institute of Microbiology, ETH Zu¨rich, Wolfgang-Pauli-Strasse 10, 8093 Zu¨rich, Switzerland
[4]Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK
[5]The Royal Veterinary College, Hawkshead Campus, Hatfield, Hertfordshire AL9 7TA, UK

JC, 0000-0002-4538-4007; VST, 0000-0002-2734-4036; AN, 0000-0002-5938-9124; BWW, 0000-0002-6140-9489

Actinobacillus pleuropneumoniae is a mucosal respiratory pathogen causing contagious porcine pleuropneumonia. Pathogenesis studies have demon- strated a major role for the capsule, exotoxins and outer membrane proteins. *Actinobacillus pleuropneumoniae* can also glycosylate proteins, using acytoplasmic *N*-linked glycosylating enzyme designated NGT, but its transcriptional arrange- ment and role in virulence remains unknown. We investigated the NGT locus and demonstrated that the putative transcriptional unit consists of *rimO*, *ngt* and a glycosyltransferase termed *agt*. From this information we used the *A. pleuropneumoniae* glycosylation locus to decorate an acceptor protein, within *Escherichia coli,* with a hexose polymer that reactedwith an anti-dextran antibody. Mass spectrometry analysis of a truncated protein revealed that this operon could add up to 29 repeat units to the appropriate sequon. We demonstrated the impor- tance of NGT in virulence, by creating deletion mutants and testing them in a novel respiratory cell line adhesion model. This study demonstrates the impor- tance of the NGT glycosylation system for pathogenesis and its potential biotechnological application for glycoengineering.

1. Introduction

Actinobacillus pleuropneumoniae is a Gram-negative bacterium and the causative agent of porcine pleuropneumonia, a severe respiratory disease responsible for significant losses to the pig industry worldwide. Economically, this disease has a huge impact on the pig industry, costing an average £6.4 per fattened pig in an affected herd in Europe [1]. *Actinobacillus pleuropneumoniae* enters the lungs and colonizes tissues by binding to mucus proteins and cells of the lower respirat- ory tract, including ciliated cells of the terminal bronchioli and alveolar epithelial cells [2,3]. There are 15 established serovars that differ in capsular polysaccharide composition [4], with another proposed based on serological results [5]. Several surface structures have been identified as being involved in adhesion, including fimbriae [6] and lipopolysaccharide (LPS) [7].

Advances in DNA sequencing technologies and mass spectrometry techniques reveal that post-translational modification of proteins by glycosylation is not

THE ROYAL SOCIETY
PUBLISHING

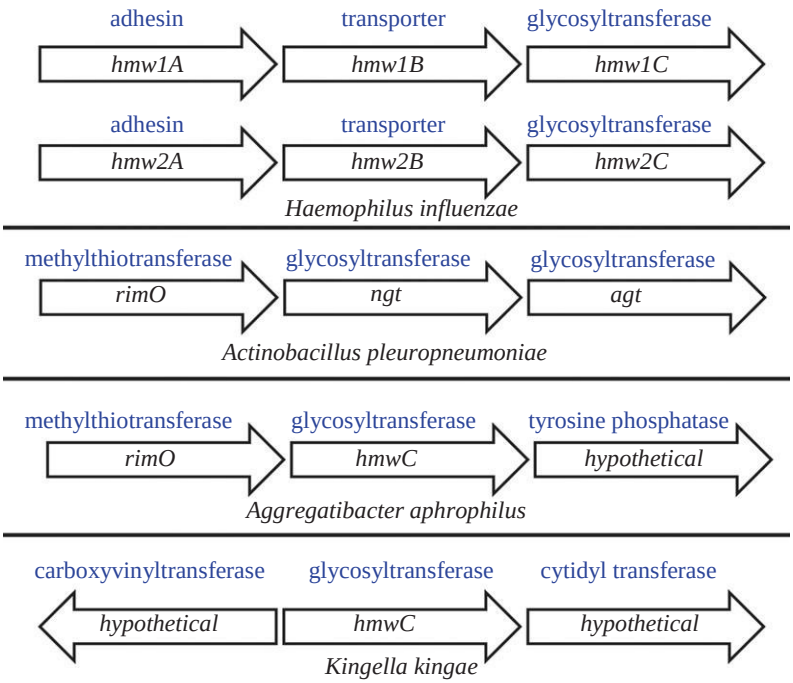


Figure 1. Genetic organization of the HMWC enzyme family. In contrast to other bacteria, NTHi has two copies of the HMW locus, and each has a gene encoding an acceptor protein. *A. pleuropneumoniae* is the only species that has a second glycosyltransferase adjacent to the *N*-linking enzyme.

restricted to a few bacterial species and is often important in pathogenesis [8,9]. Understanding the mechanisms of bacterial glycosylation and its role in pathogenesis can have practical applications such as the design of novel bioglycoconjugate vaccines, antimicrobials and diagnostics [10,11]. Bacterial protein glycosylation systems can be broadly divided into two main categories: glycans that are covalently attached to amide groups of asparagine residues (*N*-linked) or to hydroxyl groups on serine/threonine residues (*O*-linked). These categories can be further subdivided depending on the cellular compartment where protein glycosylation takes place. Oligosaccharyltransferases (OTases) function in the periplasmic compartment of a bacterial cell and catalyse the transfer of an oligosaccharide from a lipid donor to an acceptor molecule, usually a protein. The best-studied bacterial OTases are the *C. jejuni* PglB system, where *en bloc* glycosylation operates through an *N*-OTase [12,13], and the *Neisseria meningitidis O*-OTase PglL. *N*- and *O*-linked glycosylation can also occur in the cytoplasmic compartment of the bacterial cell, mediated through the action of glycosyltransferases that use nucleotide activated sugar donors as substrates for transfer onto the acceptor protein. Examples of cytoplasmic glycosylation can be found in *Clostridium difficile*, where flagellin is *O*-glycosylated [14], and non-type-able *Haemophilus influenzae* (from here on referred to as NTHi) [15], where two copies of a cytoplasmic *N*-linked glycosylation modify a high-molecular-weight adhesin with hexoses using the enzyme HMWC. In NTHi, all proteins responsible for high-molecular-weight adhesin synthesis, transport and glycosylation are encoded in the same locus.

Actinobacillus pleuropneumoniae also carries a cytoplasmic *N*-linking glycosyltransferase, known as NGT. It is a member of the HMWC-like glycosyltransferase family [16–20], but lacks an adjacent adhesin or transporter and its transcriptio- nal unit remains to be characterized. Recently, studies into the human pathogens *Kingella kingae* and *Aggregatibacter aphrophilus* [21] demonstrated a similar genetic arrangement. These 'orphan' HMWC enzymes have been found to glycosylate trimeric autotransporter adhesins, encoded in distant locations of the genome [21]. Autotransporter proteins, such as the trimeric autotransporter adhesin (TAA) Apa found in *A. pleuropneumoniae*, mediate attachment to host cells [22]. Apa is predicted to have an *N*-terminal signal peptide for secretion, a functional passenger domain containing head, neck and stalk motifs, and a conserved C-terminal translocator domain [22]. However, *A. pleuropneumoniae* has a unique chromosomal feature. Adjacent to *ngt*, there is a second ORF, which we named *agt,* coding for an accessory glycosyltransferase (figure 1).

When *agt* is heterologously expressed in *Escherichia coli* and purified, it can be used *in vitro,* to add further glucose residues to the *N*-linked glycan that NGT generates [19]. However, *agt* has never been demonstrated to function *in vivo* in conjunction with *ngt*. In addition, no virulence phenotype has been reported in *A. pleuropneumoniae* for this glycosylation locus owing to known difficulties in constructing genetic mutations in this organism.

In this study, we report the generation of *A. pleuropneumoniae ngt* and *agt* deletion mutants, and demonstrate a biological role for this *N*-linked glycosylation system using a human adenocarcinoma lung epithelial cell adhesion assay. Our results suggest that *ngt* is part of an operon that contains the upstream ORF *rimO*, encoding a methylthiotransferase, and the downstream ORF *agt*, encoding an *a*-6-glucosyltransferase (*a*6GlcT). Further- more, we were able to clone and express *ngt* and *agt* in *E. coli*, demonstrating for the first time, to the best of our knowledge, the *in vivo* assembly of *N*-linked dextran.

2. Material and methods

2.1. Bacterial strains used and culture conditions

Actinobacillus pleuropneumoniae serovar 15 reference strain, HS143, or derived mutants were grown at 378C with 5% CO_2 on BHI (Oxoid, UK) agar or broth, supplemented with

10 mg ml^{-1} nicotinamide adenine dinucleotide (NAD) and when required with kanamycin (50 mg ml^{-1}) or chloramphenicol (1 mg ml^{-1}). *Escherichia coli* TOP10 and Mu Free Donor (MFD) [23] were grown in LB broth or agar (Oxoid) supplemented, when required, with 50 mg ml^{-1} kanamycin at 37°C. *E. coli* DH10 were grown in LB broth or agar (Oxoid) at 37°C supplemented, when required, with 80 mg ml^{-1} spectinomycin and/or 100 mg ml^{-1} trimethoprim.

2.2. Genomic DNA extraction

Total genomic DNA was extracted from a 10 ml overnight culture of *A. pleuropneumoniae* HS143, using a proteinase K and phenol: chloroform: isoamyl-alcohol-based procedure as previously described by Cuccui *et al.* [24].

2.3. Construction of *Actinobacillus pleuropneumoniae* knockout mutants

The *A. pleuropneumoniae* HS143 orthologues of *apl_1634* and *apl_1635* (also known as *agt* and *ngt*, coding for α6GlcT and NGT, respectively), found in the *A. pleuropneumoniae* L20 genome [25], were deleted using our recently described unmarked mutation system [26]. Primers used to generate the *cat-sacB* insertion/deletion and the unmarked deletion constructs for each gene are shown in electronic supplementary material, table S2. Briefly, the target genes and approximately 600–900 bp of flanking sequences were amplified using CloneAmp HiFi PCR Premix (Clontech), A-tailed and cloned into pGEMT (Promega), as previously described [26]. Inverse PCR was then used to open up the clones, using the appropriate primers, removing the target sequence and adding 15 bp overhangs to allow insertion of the *cat-sacB* cassette by In-Fusion cloning (Clontech). Unmarked deletion constructs were generated by amplifying the left and right flanking sequences for each gene, using appropriate primers with added 15 bp overhangs designed to allow direct fusion by overlap-extension PCR. The unmarked deletion mutants were then obtained by two sequential rounds of natural transformation as previously described [27].

2.4. Plasmid complementation

The vector pMKExpress [28] was digested with *Eco*RI and *Sac*I (New England Biolabs, UK) and the resulting digest was gel purified using a Qiagen MinElute gel extraction kit (Qiagen, UK) according to the manufacturer's instructions, to remove the GFP coding ORF.

The *ngt* gene was PCR amplified using Accuprime Taq Hifi (Invitrogen, UK) using the forward primer ngtCOMPFWD (5′-TTTTGAATTCGTGGGTAAAACGCTTGCAGT-3′) and reverse primer ngtCOMPREV (5′-TTTTGAGCTCTTAATTTTCTTTTAGGAACGCATTT-3′). The *agt* gene was amplified using the primers agtCOMPFWD (5′-AAACTGCAGATTAAATGCGTTCCTAAAAGAAAA-3′) and agtCOMPREV (5′-TTT GCGGCCGCTTAACTCCGACTATTCTCAAG-3′).

When *agt* only complementation failed, complementation with *ngt–agt* was attempted. Both ORFs were PCR amplified using Accuprime Taq Hifi (Invitrogen) using the forward primer ngtagtCOMPFWD (5′-TTTGAATTCCGAGCAAGAAGTGAAAGTCG-3′) and reverse primer ngtagtCOMPREV (5′-TTTGCGGCCGCCACCGATAGCCGTATTTCGT-3′) with the following cycling conditions: 94°C/30 s followed by 24 cycles of 94°C/30 s, 53°C/30 s, 68°C/2 min and a final 68°C/5 min cycle. All ORFS were expressed under the control of the plasmid promoter.

The resulting *ngt* only PCR product was digested with *Eco*RI and *Sac*I, the *agt* only product was digested with *Pst*I and *Not*I, and the *ngt–agt* PCR product was digested with *Eco*RI and *Not*I before being purified using a Qiagen PCR purification kit. Digested vector and PCR products were ligated using Promega T4 DNA ligase (Promega, UK) to yield the vectors pMK*ngt*, pMK*agt* and pMK*ngt–agt*, prior to transformation of the plasmid into One Shot *E. coli* TOP10 cells (Invitrogen) according to manufacturer's instructions. Transformants were selected on LB agar supplemented with kanamycin (50 mg ml^{-1}). The complementation vectors were transformed into the mutant recipient strains by natural transformation as previously described [27].

2.5. Cell culture

The A549 cell line, adenocarcinoma human alveolar basal epithelial cells, (ATCC, CCL-185, US) was grown at 37°C, 5% CO_2 in F-12 K medium (Gibco) supplemented with 10% fetal calf serum (Sigma).

2.6. Adhesion assay using A549 cell line

The A549 cell line (ATCC, CCL-185, USA) was seeded into 12-well tissue culture plates at a concentration of 2.5×10^5 cells ml^{-1} and incubated overnight at 37°C 5% CO_2. Bacterial overnight cultures (HS143 wild-type, isogenic *ngt* and *agt* mutants and complemented mutants) were used to seed into BHI–NAD medium and grown to an OD_{600nm} of 0.6. One millilitre of the suspension was added to the A549 cells at a multiplicity of infection (MOI) of 100 : 1, and the plates incubated at 37°C 5% CO_2. After 3 h, non-adherent bacteria were removed by washing three times with 1 ml DPBS (Gibco), and adherent bacteria were released by adding 100 µl of 0.25% trypsin–EDTA (Sigma) for 5 min at 37°C. Trypsinization was stopped by the addition of 900 µl of DPBS. Serial dilutions were plated onto BHI–NAD plates for quantification of adherent bacteria. In order to determine if any of the recovered bacteria had invaded the A549 cells, controls were treated with gentamycin (Sigma) at a final concentration of 10 mg ml^{-1} for 1 h to allow for killing of adherent extracellular bacteria. The cells were then lysed by the addition of ice-cold sterile water, and serial dilutions were plated out on BHI–NAD.

2.7. Statistical analysis of adhesion assay data

The number of adherent cells was calculated by counting the colony forming units and comparing with the initial inoculum of each individual culture to determine the percentage of adherent cells. The statistical analysis was performed using a one-way analysis of variance followed by a Bonferroni's multiple comparison test. The significance level was set at 0.05 throughout. Statistical analysis was done using GraphPad Prism v. 4.00 for Windows (GraphPad Software, San Diego, CA, www.graphpad.com).

2.8. Reverse transcriptase PCR

An overnight culture of *A. pleuropneumoniae* HS143 was diluted 1 : 20 in BHI–NAD broth. At the time points of 1.5, 3.0, 5 and

24 h, RNA was extracted as previously described [29], with the minor modification that 2 mg of total RNA from each sample was treated with Ambion TURBO DNase (Invitrogen) according to the manufacturer's instructions. cDNA was generated from DNase-treated RNA using the SuperScript II kit (Invitrogen) according to the manufacturer's instructions. For each sample, 2 ml of reverse transcribed cDNA was used as a template in a 25 ml total volume PCR mixture and amplified using MyTaq master mix (Bioline, UK) using the following cycling conditions: 948C/30 s, followed by 35 cycles of 948C/10 s, 538C/10 s, 728C/10 s and a final single 728C/30 s cycle using the primers listed in electronic supplementary material, table S2.

2.9. Quantitative real-time PCR validation

Total RNA was extracted from bacteria grown in BHI– NAD broth by using Tri Reagent (Sigma, UK) as previously described. Two micrograms from each sample was treated with Ambion TURBO DNase (Invitrogen) according to manufacturer's instructions with some modifications. Total RNA was incubated for 1 h at 378C followed by the addition of another two units of DNase. The sample was then incubated for an extra 1 h before inactivation.

cDNA was generated from DNase-treated RNA using the SuperScript III kit (Invitrogen) using random hexamers (Invitrogen) according to the manufacturer's instructions. One microlitre of this material was used in a QPCR using SYBR Green dye-based PCR amplification and detection system (Applied Biosystems). The *A. pleuropneumoniae* HS143 WT, HS143*Dngt* and HS143*Dagt* were analysed for absolute quantification of cDNA using an ABI7500 Fast instrument (Applied Biosystems). Amplification was carried out using the following primers at a final concentration of 500 nM. agtfwd: 5′-GAT TGG ATA GGT GAA GGC GA-3′, agtrev: 5′-CCC TTG CTC AAA ATG ACG GA-3′, ngtfwd: 5′-AGT TTG TGA GAG CAA CGG TG-3′, ngtrev: 5′-AGT CCG AAT GTG TTG TTG CC-3′, rimOfwdv2: 5′-CGT CCG ATT GTG CAA GTG TT-3′, rimOrevv2: 5′-CAC CGT TCC AGA AAA CCG TT-3′. Samples tested were four biological replicates, each tested as three technical replicates.

For comparative qRTPCR analysis of *A. pleuropneumoniae* HS143 WT, HS143*Dngt* and HS143*Dagt*, gene induction or reduction values were calculated by comparing the normalized values of the wild-type and mutant samples, using the statistical formulation for the threshold cycle (*DD*CT) method. The threshold value of each gene was first normalized to the value of the constitutively expressed control gene *glyA* [30] (glyA primers: fwd: 5′-CAA GCG AAT GCA GCT GTT TA-3′, glyArev: 5′- CTG TGA TGC CGT AGA GGA CA-3′).

2.10. Subcloning and heterologous expression of *agt* and *ngt*

The putative NGT operon was PCR amplified using the primers *ngt-agt*fwd: 5′-TTTTGAATTCCGAGCAAGAAG TGAAAGTCG-3′ and *ngt-agt*rev: 5′-TTTTTGGTACCCACC GATAGCCGTATTTCGT-3′ using Accuprime Taq Hifi (Invitrogen) and the following cycling conditions: 948C/30 s, followed by 24 cycles of 948C/30 s, 538C/30 s, 688C/4 min and a final cycle of 688C/5 min.

The amplicon was ligated into the vector pEXT20 using T4 DNA ligase (New England Biolabs, UK) following digestion of the plasmid and the PCR product with *Eco*RI and *Kpn*I. *Escherichia coli* NEB10*b* (New England Biolabs, UK) was transformed with the ligation reaction generating the plasmid pJC78. Expression was induced by growing an *E. coli* colony in LB broth with ampicillin 100 mg ml^{-1} until an OD_{600} of 0.4 was reached. At that point 1 mM, IPTG was added, and the cultures were incubated at 378C with shaking for a further 16 h. Expression of NGT and *a*6GlcT was monitored using SDS– PAGE, Coomassie staining and western blotting.

2.11. Glycosylation of AtaC by NGT and *a*6GlcT in *Escherichia coli* cells

Escherichia coli DH10*b* cells carrying pJC78 were transformed with the construct $pMLBADAtaC_{1866–2428}$ [17], and cultured in LB broth with ampicillin 100 mg ml^{-1}, trimethoprim 20 mg ml^{-1} at 378C with shaking until an OD_{600nm} of 0.4 was reached, followed by induction with 0.2% L-arabinose and 1 mM IPTG. After 16 h incubation, AtaC was purified. The bacterial cell pellet was isolated by centrifugation at 6000*g* for 10 min and lysed using a cell homogenizer (Stansted Fluidics Ltd. SPCH-10). Any intact cell debris was thereafter pelleted by centrifugation at 10 000*g* for 30 min before purification from the supernatant using an Ni– NTA (Qiagen, UK) gravity column (Thermo Scientific, USA).

Glycosylated product was analysed by SDS– PAGE and transferred onto a nitrocellulose membrane before being analysed by immunoblot using a mouse anti-His antibody (AbCam, UK) and an IRDye 680CW goat anti-mouse conjugate secondary antibody. Detection of fluorescent signal was carried out using a LI-COR imaging system.

2.12. Mutagenesis of the *ngt* locus

The cloned locus coding for NGT and *a*6GlcT was mutated using the QuickChange XL II site-directed mutagenesis kit (Agilent Technologies, CA) using the following primers. agtt120a_antis: 5′-CAAAACAGAAGTAAACGTTTTAATC TATATTATTTTCCATAACAT AACCTTAAGAGCC-3′ and agtt120a: 5′-GGCTCTTAAGGTTATGTTATGGAAAATAAT ATAGATTAAAACGTTTACTTCTGTTTTG-3′ (underlined nucleotide denotes the change).

The *ngt* gene was mutated using the following primers: ngta1321g_a132: 5′-CGGTATAGCTTCAACCACGATGGCG CTAAATCCGTATTT CTTAGAA-3′ and ngta1321g_a132: 5′-TTCTAAGAAATACGGATTTAGCGCCATCGTGGTTGAAG CTATACCG-3′ (underlined nucleotides denotes the change). The following conditions were used: 958C/60 s followed by 18 cycles of 958C/50 s, 608C/50 s, 688C/8 min and a final 688C/7 min cycle.

Following amplification, the PCR products were *Dpn*I treated according to the manufacturer's instructions and used to transform *E. coli* XL-10 Gold cells (New England Biolabs, UK).

2.13. Western blot analysis of glycosylated AtaC

Purified AtaC from *E. coli* DH10*b* was analysed by western blotting. Unglycosylated, fully glycosylated and monoglycosylated AtaC were analysed by dot blot by placing a 3 ml drop of a 1 mg ml^{-1} solution of protein or dextran (dextran standard

MW 1000 from Leuconostoc, Sigma-Aldrich UK) onto a nitrocellulose blotting membrane (Amersham Protran, GE HealthCare, Germany) and allowing to air dry before blocking the membrane by incubating with phosphate-buffered saline, 2% milk solution for 1 h at room temperature. The membrane was then probed using a 1 : 1000 dilution of a mouse monoclonal antibody raised specifically against a tetrasaccharide of α1 – 6 linked glucose (MS α-Dextran Clone Dx1, Stem Cell Technologies, Canada). An IRDye 680CW goat anti-mouse antibody at a 1 : 10 000 dilution was used as the secondary antibody. The western blot images were visualized using a LI-COR imaging system.

2.14. Analysis of N-glycans released from AtaC

Glycans were released from 200 *m*g AtaC using the Ludger Liberate Hydrazinolysis kit, according to the manufacturer's recommendations. The released glycans were fluorescently labelled with 2-aminobenzamide (2-AB) as described previously [31]. Excess labelling reagent was removed as follows: four discs of filter paper (Whatman) were soaked in 30% acetic acid, inserted into a 1 ml plastic syringe and washed sequentially with 2 × 1 ml acetic acid, 2 × 1 ml water, 2 × 1 ml acetonitrile (ACN) and finally 2 × 1 ml 95% ACN. The labelled glycans were diluted to 500 *m*l with 95% ACN and loaded onto the column. Of 500 *m*l 95% ACN was used to rinse the labelling tube and was added onto the column as well. The column was washed with 8 × 1 ml 95% ACN and glycans were finally eluted in 50 *m*l water twice. Elution fractions were pooled and passed through a
0.45 *m*m filter (Ultrafree-MC Durapore HV filter unit, Millipore) before analysis by normal-phase HPLC (Supelcosil LC-NH2 column, 80 – 20% ACN gradient over 90 min, fluorescence detection at 320 nm excitation and 420 nm emission wavelength). Glyko 2-AB glucose homopolymer standard (Prozyme) was used as a reference. The identity of the labelled glycans was confirmed by MALDI mass spectrometry. Samples were mixed 1 : 1 with dihydroxybenzonic acid matrix (15 mg ml^{-1} in 75% ACN in water with 0.1% formic acid (FA)), and spotted onto a matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF– TOF MS) target plate. Data acquisition was performed manually on a Model 4800 Proteomics Analyser (Applied Biosystems, Framingham, MA) with an Nd : YAG laser, and 1000 shots were accumulated in the reflectron positive ion mode.

2.15. Nano-LC–ESI–MS/MS analysis of glycosylated AtaC

For structural analysis, 50 *m*g of AtaC was reduced, alkylated and digested with trypsin using the filter-aided sample preparation protocol [32]. Samples were analysed on a calibrated LTQ-Orbitrap Velos mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to an Eksigent-Nano-HPLC system (Eksigent Technologies, Dublin (CA)). Peptides were resuspended in 2.5% ACN and 0.1% FA, and loaded on a self-made tip column (75 *m*m × 80 mm) packed with reverse phase C18 material (AQ, 3 *m*m 200 Å, Bischoff GmbH, Leonberg, Germany) and eluted with a flow rate of 200 nl per min by a gradient from 3% to 30% ACN, 0.1% FA in 22 min, 50% ACN, 0.1% FA in 25 min, 97% ACN, 0.1% FA in 27 min. One scan cycle comprised a full-scan MS survey spectrum, followed by up to 20 sequential collision-induced dissociation (CID) MS/MS on the most intense signals above a threshold of 1500. Full-scan MS spectra (400 – 2000 m/z) were acquired in the FT-Orbitrap at a resolution of 60 000 at 400 m/z, whereas CID MS/MS spectra were recorded in the linear ion trap. CID was performed with a target value of 1e4 in the linear trap, collision energy at 35 V, Q-value at 0.25 and activation time at 30 min. AGC target values were 5e5 for full FTMS scans and 1e4 for ion trap MSn scans. For all experiments, dynamic exclusion was used with one repeat count, 15 s repeat duration and 60 s exclusion duration.

2.16. Database analysis and identification of modified residues

MS and MS/MS data were processed into Mascot generic format files and searched against the Swissprot database (version 201402) through the Mascot engine (v. 2.2) with the consideration of carbamidomethylation at cysteine, oxi- dation at methionine and N-hexosylation at Asparagine. The monoisotopic masses of 2 þ or more charged peptides were searched with a peptide tolerance of 10 ppm and an MS/MS tolerance of 0.6 Da for fragment ions. Only peptides with a maximum of two missed cleavage sites were allowed in database searches. Positive identification of hexosylated peptides was performed by manual inspection of spectra. Peptides modified with extended glycan chains were investigated manually, and their corresponding MS/MS spectra were annotated. Here, XCalibur v. 2.2 sp1.48 was used for data processing, and MS deconvolution was performed by XtractRaw file from Thermo Scientific.

2.17. Construction of acceptor protein JC1

Amino acid residues 23 – 163 of Cj0114 from *C. jejuni* NCTC 11168 were used as a scaffold to design a novel acceptor protein. The native signal sequence from residues 1 to 23 was removed along with the native tetratricopeptide domain encoded within residues 164– 315. Twelve NAT glycosylation sequons were added at the C-terminus of the new protein, each separated by a proline and a glycine. Finally, a hexa- histidine tag was added to the C-terminus to enable protein purification. This construct was DNA synthesized (Celtek Genes, USA) and subcloned into *Bam*HI and *Sph*I digested expression vector pACYC184.

3. Results

3.1. NGT and α6GlcT are required for adhesion of *Actinobacillus pleuropneumoniae* HS143 to A549 cell lines

Within the genome of *A. pleuropneumoniae* strain HS143, we identified two ORFs, orthologues of *apl_0104* (70% identity, BlastP) and *apl_0443* (82% identity, BlastP) in the L20 genome [25], coding for autotransporter adhesins. *In silico analysis* revealed that these adhesins have 75 and 95 N-X-(S/T) sequons, respectively (PROGLYCPROT). Naegeli *et al*. [17] car- ried out mass spectrometry analysis of *A. pleuropneumoniae*'s proteome for strain 4074 serotype 7, and the only

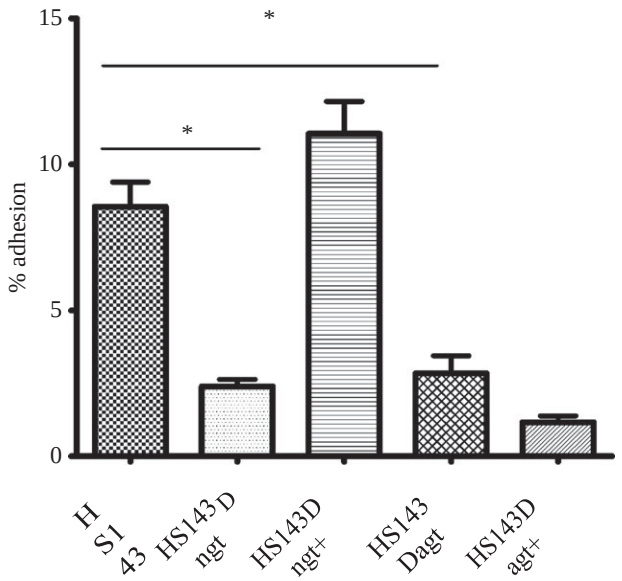


Figure 2. Percentage of adhesion of *A. pleuropneumoniae* strain HS143, isogenic mutants and complemented mutants to A549 cells infected at an MOI of 100 : 1 for 3 h prior to quantification of adherence. Horizontal bars indicate pairs of columns that are significantly different when compared with the wild-type HS143 (p ¸ 0.05).

glycopepetides identified belonged to two autotransporter adhesins [17], making these two adhesins the only native substrates for NGT identified so far. In NTHi, deletion of the *N*-linked glycosylation system results in a significantly reduced adherence phenotype [15]. In order to investigate whether this was the case for *A. pleuropneumoniae*, adhesion of WT HS143, isogenic mutants HS143*Dngt* and HS143*Dagt* and its complements to A549 human adenocarcinoma lung epithelial cells was investigated.

Actinobacillus pleuropneumoniae strain HS143, the wild-type strain, was found to have a percentage of adherent cells of 8.55 + 0.84, n ¼ 78 to A549 cells after 3 h incubation (figure 2). In order to understand the role of the cytoplasmic NGT in this adhesion phenotype, an in-frame deletion mutant of the *ngt* gene was generated in *A. pleuropneumoniae* and found to have a reduced percentage of adherent cells, 2.39+0.25 (n ¼ 78, p ¸ 0.05), when compared with the wild-type. This phenotype was restored (11.05+1.10, n ¼ 30, p ¸ 0.05) upon complementation with the *ngt* gene. Furthermore, when an *a*6GlcT deletion mutant in *A. pleuropneumoniae* was tested for adherence, it was observed that there was a decrease in adhesion (2.85+0.60, p ¸ 0.05, n ¼ 30) to the same level as the NGT mutant. However, complementation with the ORF coding for *a*6GlcT was unable to rescue this phenotype (figure 2). Gentamycin treatment of cells confirmed that the bacterial counts observed were due to adhering and not invading bacteria.

3.2. *Agt* is part of a conserved putative operon that includes *ngt* and *rimO*

Unlike NTHi, adjacent to the *A. pleuropneumoniae N*-linking transferase gene, *ngt*, the flanking genes do not encode an adhesin or a dedicated adhesin transporter (figure 1). Instead, we identified an ORF coding for a protein with amino acid similarity to 30S ribosomal protein S12 methylthiotransferase, *rimO*, upstream of *ngt*, and a second glycosyltransferase-encoding gene downstream of *ngt*. Analysis of all available *A. pleuropneumoniae* genomes demonstrated that the genetic arrangement of the locus was absolutely conserved in all published *A. pleuropneumoniae* genomes and over 180 sequenced isolates (J.T.B. 2016, personal communication). Reverse transcriptase-PCR (RT-PCR) was used to analyse the expression of this locus in serovar 15 *A. pleuropneumoniae* reference strain HS143. Primers were designed spanning intergenic regions between the three ORFs. Probe 1 tested if an mRNA transcript was generated between *rimO* and *ngt*, and probe 2 tested for the presence of an mRNA tran- script between *ngt* and *agt*. The results suggest that all three genes form an operon (figure 3*a*). Further RT-PCR analysis showed that the promoter driving *rimO* expression was independent of the ORF immediately upstream (figure 3*b*).

Messenger RNA was extracted at different time points during the growth of *A. pleuropneumoniae* and cDNA was generated by RT-PCR using a probe designed within *ngt*. This showed that *ngt* was transcribed at all-time points tested (electronic supplementary material, figure S1).

3.3. Absolute quantification of *rimO*, *ngt* and *agt* by qPCR

In order to further validate the hypothesis that *rimO*, *ngt* and *agt* are co-transcribed, an absolute quantification qPCR was performed. The results were normalized by the absolute number of copies of *rimO* within each sample assuming *rimO* is the first ORF in the operon and therefore the closest to the putative promoter identified by bioinformatics analysis. A trend was observed in all four biological replicates (n ¼ 12; 4 biologicals, 3 technical replicates) indicating a decrease in expression level from *rimO* to *agt* consistent with the genetic organization of the putative operon (*rimO* versus *ngt*, p , 0.05; *rimO* versus *agt*, p ¸ 0.001; figure 4).

3.4. Reconstruction of the NGT glycosylation operon and its functional transfer and expression in *Escherichia coli*

Following on from the RT-PCR studies indicating that both *agt* and *ngt* were co-transcribed, we amplified by PCR the two ORFs as a single amplicon and cloned them into the IPTG-inducible expression vector pEXT20 [33], to generate the plasmid pJC78. When the ORFs encoding *a*6GlcT and NGT were co-expressed with a fragment of an autotrans- porter adhesin from *A. pleuropneumoniae* (AtaC), which is a natural acceptor [17], a reduction in protein migration on SDS– PAGE was observed, indicating an increase in molecu- lar weight consistent with the addition of an oligosaccharide (figure 5).

To further understand the *in vivo* glycosylation operon, individual mutations in *ngt* or *agt* were constructed within the plasmid pJC78. NGT activity was abolished by substituting the conserved lysine residue at position 441 by alanine (K441A) [18,20], whereas *a*6GlcT activity was abolished by the replacement of the leucine codon at amino acid position 7 with a stop codon (L7*). Schwarz *et al.* [19] indicated that *in vitro*, NGT and *a*6GlcT could assemble a glucose polymer between two and six residues on an acceptor peptide. We reasoned therefore that a commercially available antibody

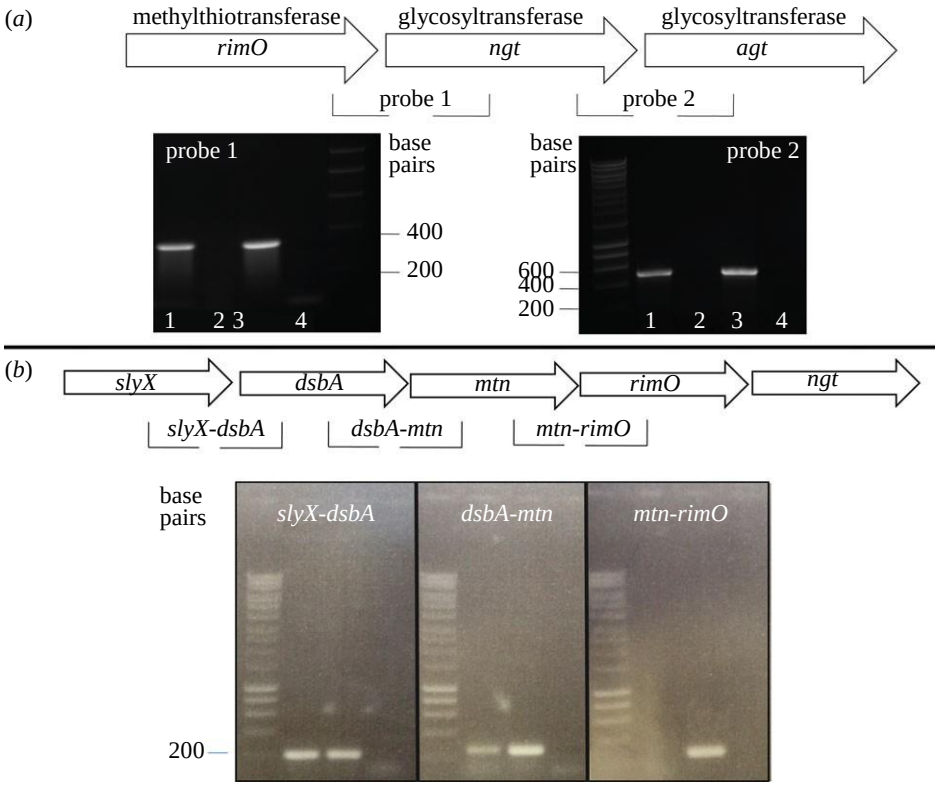


Figure 3. (*a*) Transcriptional analysis of the *A. pleuropneumoniae* NGT locus. Lane 1: cDNA as template; lane 2: RNA as template; lane 3: *A. pleuropneumoniae* HS143 genomic DNA positive control; lane 4: negative PCR control (no template). (*b*) Transcriptional analysis of the region upstream of *ngt*. Lane 1: cDNA as template; lane 2: genomic DNA positive control; lane 3: PCR control (RNA as template).

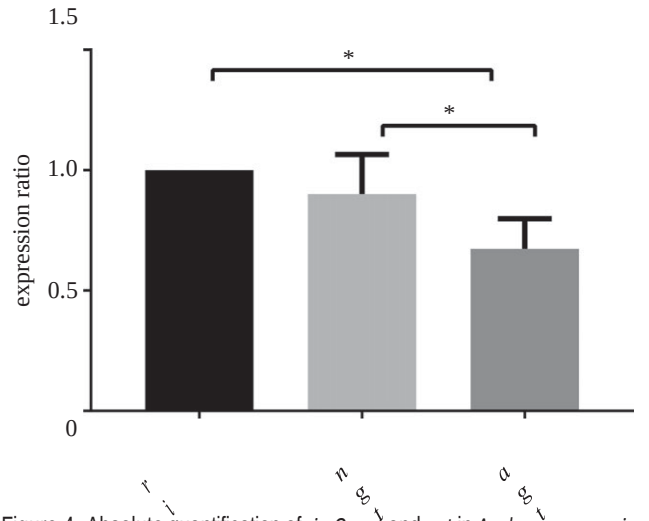


Figure 4. Absolute quantification of *rimO*, *ngt* and *agt* in *A. pleuropneumoniae* strain HS143 expressed in fold change. Horizontal bars indicate pairs of columns that are significantly different when compared with each other. Asterisk indicates $p \leq 0.05$ ($n = 12$).

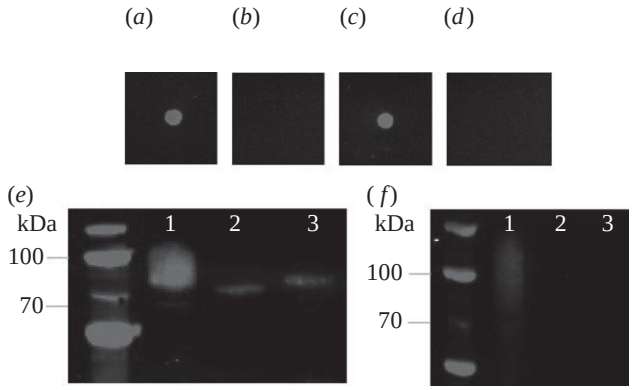


Figure 5. Glycosylation analysis of $AtaC_{1866-2428}$ alongside NGT and α6GlcT by anti-dextran and anti-HIS western blots. Top panel: anti-dextran dot blot; (a) dextran; (*b*) AtaC with NGT K441A and functional *a*6GlcT; (*c*) AtaC with functional NGT and *a*6GlcT; (*d*) AtaC with functional NGT only. Bottom panel: anti-HIS and anti-dextran Western blots (*e*,*f*, respectively); lane 1, AtaC with functional NGT and *a*6GlcT; lane 2, AtaC with NGT K441A and functional *a*6GlcT; lane 3, AtaC with functional NGT only.

specific for *a*1 – 6 linked glucose tetrasaccharide (isomaltotetraose) may be able to detect and verify the nature of the polysaccharide generated by the cloned *agt– ngt* operon and of the knockouts. Ni– NTA purified proteins from the three construct combinations were tested for expression by dot blot analysis using an anti-dextran monoclonal antibody (mAb). This showed that a recognizable epitope could only be generated when NGT and *a*6GlcT were both functional (figure 5, top panel). SDS– PAGE and western blot analysis using an anti-HIS monoclonal antibody showed a smear vis- ible above the point at which AtaC should migrate, but only when NGT and *a*6GlcT are both functional. This smearing was also detected using the anti-dextran monoclonal antibody (figure 5*f*, lane 1). AtaC glycosylated appears to migrate less than when detected by anti-HIS antibody, because the anti-dextran antibody will only recognize AtaC modified with four or more glucoses per site. Removing the function of NGT yielded an AtaC fragment that migrated to its unglycosylated location losing the epitope recognized by the anti-dextran mAb (figure 5*e*, lane 2). Finally, knocking out the function of *a*6GlcT reduced protein migration to a slightly higher level than that observed with NGT mutation alone (figure 5*e*, lane 3). This can be explained by glycosylation with a single hexose at multiple sites within the acceptor protein. Furthermore, this material was not recognized by anti-dextran mAb, suggesting that glycosylation

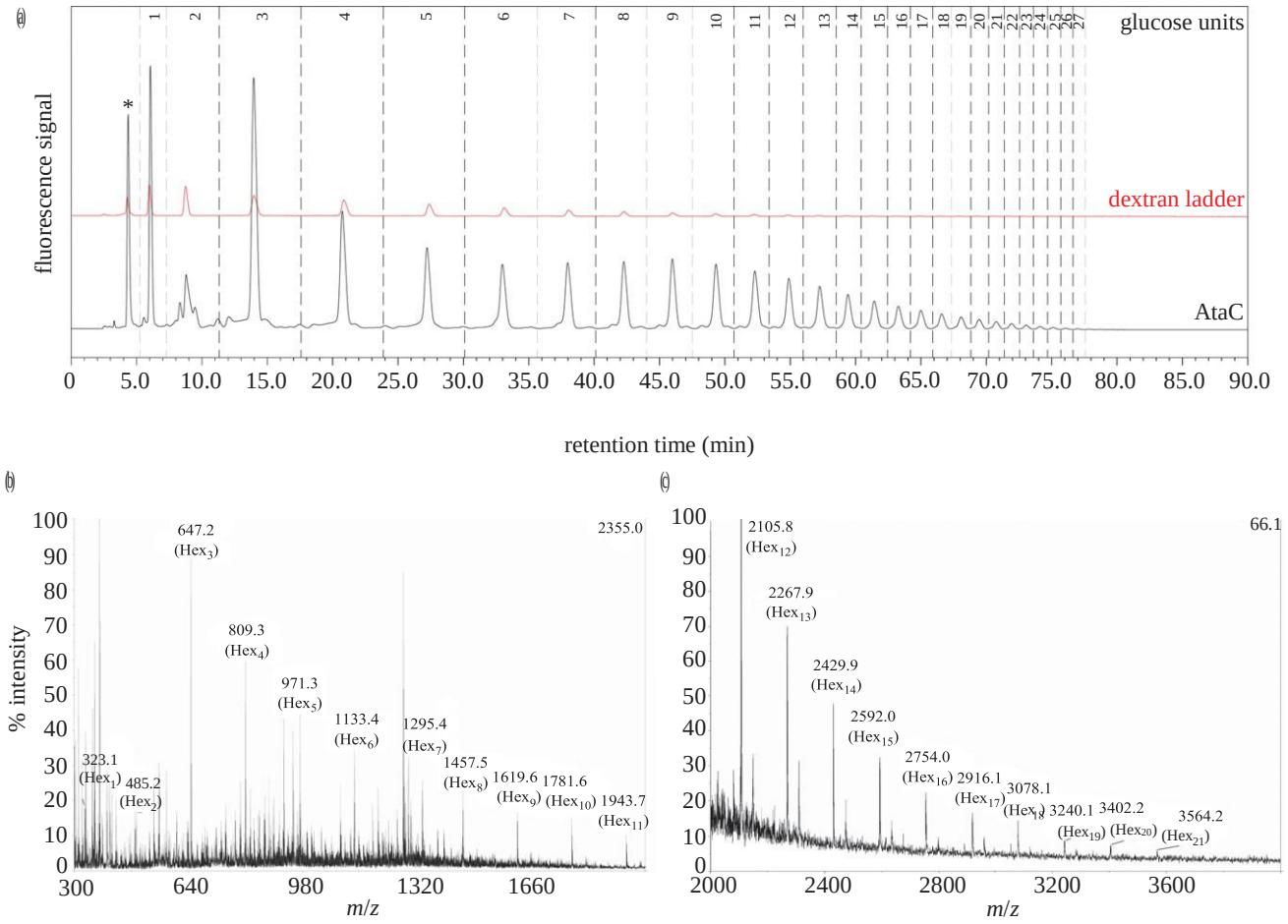


Figure 6 (*a*) NP-HPLC analysis of 2-AB labelled glycans released from purified AtaC which was co-expressed with NGT and *a*6GlcT in *E. coli* DH10*b* (black). A glucose homopolymer (dextran) ladder serves as a reference for retention time (red). The peak originating from excess 2-AB label (rt 44 min) is marked with an asterisk. (*b,c*) MALDI mass spectrometry analysis of the same glycan sample confirms the identity of the observed glycan chains as a hexose polymer.

had occurred, but that no polymer had been generated (figure 5*f*, lane 3).

3.5. Confirmation of hexose build-up on AtaC

In order to confirm the identity of the observed post-translational modification of AtaC, the glycans from purified protein were released by hydrazinolysis, fluorescently labelled with 2-aminobenzamide (2-AB) and analysed by normal-phase HPLC (figure 6*a*).

With a labelled dextran ladder used as a reference, this analysis revealed the presence of glycan chains of varying lengths (1 – 27 monosaccharide units). This was further confirmed by MALDI-MS analysis. Peaks differing in mass by 162 Da suggested potential for glucose or galactose attach- ment (figure 6*b,c*). Therefore, the results from both methods were in agreement showing a hexose polymer ranging up to at least 20 units.

To confirm that particular sites on the protein were modified with these elongated glycans, LC– ESI– MS/MS analysis of glycosylated AtaC was performed (figure 7 and table 1). This showed that previously identified glycosylation sites were occupied [17]. In total, 15 asparagine (Asn) residues were identified as being modified with glycan chains of vari- able length. For example, glycopeptide GNLSTAADVTDK could be detected modified with an *N*-linked glycan consist- ing of 1 – 29 hexose units (table 1). On other sites, only short glycan chains could be detected, whereas one peptide (NISTVVK) could only be detected as being modified with glycan chains of more than 14 hexoses. These results are summarized in table 1 and confirm western blot evidence that co-expression of NGT and *a*6GlcT leads to the formation of Asn-linked, linear hexose chains of up to 29 units in length.

3.6. The *ngt/agt* operon can be used to modify alternative substrates with dextran

Following assembly of plasmid pJC78, we began testing if the *ngt/agt* operon could be used to make *N*-linked glucose polymers on non-native substrate proteins in a similar manner to NGT alone [17]. We selected Cj0114 from the 1-proteobacterium *Campylobacter jejuni* as a scaffold for designing a new acceptor protein. The native Cj0114 tetratricopeptide domain was removed to reduce protein toxicity and simplify purification. At the *C*-terminus of the protein, the 12 added NAT glycosylation sequons were followed by a hexa-histidine tag to enable protein purification. The new protein, named JC1, was constitutively expressed from the plasmid pJC1. Combin- ing the plasmids pJC1 and pJC78 generated an epitope that could be recognized by the anti-dextran mouse mAb; this dis- appeared upon knocking out the function of *ngt* or a*gt* (figure 8). The marginally different sizes in the anti-His and anti-dextran western blot are due to the recognition epitope for the anti-dextran antibody, where only highly polymerized proteins are detected (acceptors modified with four or more glucose residues). These findings indicated that NGT and *a*6GlcT can be made to target any protein.

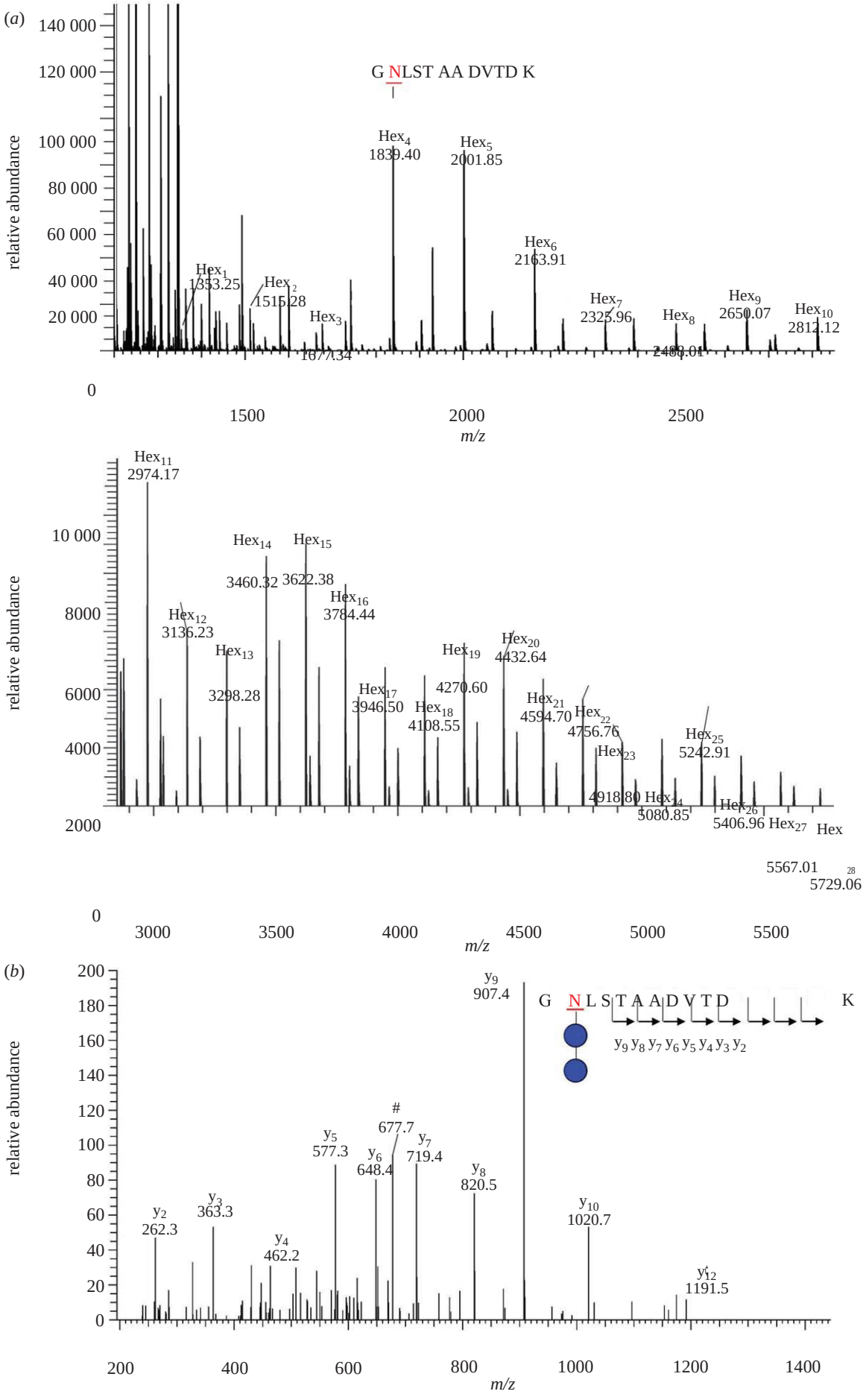


Figure 7. Deconvoluted MS spectra show that peptide GNLSTAADVTDK from AtaC is modified with a hexose polymer ranging from 1 to 28 units. (*b*) MS/MS spectrum of *m/z* 758.3497(2+) corresponding to GNLSTAADVTDK modified with two hexoses, showed continuous fragmentation ions, which confirm the peptide identity. The hash tag marks doubly charged ion with neutral loss of hexose from precursor ion. y' indicates the y ion without hexoses.

4. Discussion

Novel bacterial glycosylation systems are regularly being dis- covered as glycan analyses methodologies improve [34 – 37]. The functions of these

glycosylation systems are yet to be fully appreciated, but it is now apparent that glycosylation is a feature common to most bacteria.

In this study, we report the investigation of a cytoplasmic glycosylation system in a member of the *Pasteurellaceae* family, *A. pleuropneumoniae*. Our results demonstrate that despite similarities between NGT and its orthologue, HMW1C, in NTHi, the system described here is unique. The *A. pleuropneumoniae N*-linking locus consists of two co-transcribed glycosyl- transferases (*ngt* and *agt*) with no associated adhesin or

Table 1. Summary of glycosylation status for each site from AtaC. Underlined letters: in red, N-X-S/T sequons; in blue, asparagine residues found to be occupied but not part of an N-X-S/T consensus sequon.

tryptic peptide	peptide sequence	glycan
1	ELNETLTIK	Hex1 – 5
2	GNLSTAADVTDK	Hex1 – 29
3	VINVAAGDVNANSTDAVNGSQLYAVSEVANK	Hex1 – 14
4	GWNIQTNGNDTTNVKPGDTVNFVNGDNIAITNDGTK	Hex1 – 15
5	VGDNVSLTK	Hex3 – 19
6	GANVTQNLGK	Hex1 – 20
7	NISTVVK	Hex14 – 20
8	DGSLTIGNTTINSDQVK	Hex1 – 22
9	VSGVADGDISPNSTEAINGSQLYDANQNIANYLGGGSK	Hex1 – 13
10	VDGNTTTANNVGDAITNLNNEVVKPLTFEGDTGVASK	Hex1 – 12
11	TVNANTVNANTVK	Hex1 – 6
12	AGQADTDAVNVSQLK	Hex1 – 5

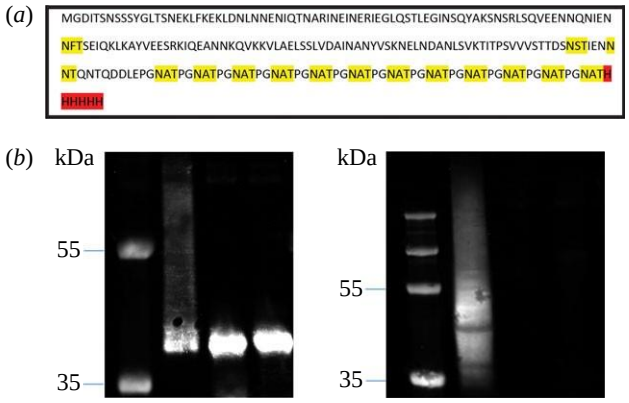


Figure 8. Glycosylation of engineered acceptor protein (JC1) by NGT and a6GlcT. (*a*) Amino acid sequence of the new target glycoprotein JC1. Highlighted in yellow are glycosylation sequons, and in red, the hexa-HIS tag used for protein purification. (*b*) Glycosylation of the acceptor protein JC1 with NGT and a6GlcT. Left panel: anti-histidine tag western blot; right panel, anti-dextran western blot. Lane 1: JC1 expressed with functional NGT and *a*6GlcT; lane 2: JC1 with NGT K441A and *a*6GlcT; lane 3: JC1 with NGT but non-functional a6GlcT.

transporter. Another significant difference between the *A. pleuropneumoniae* system and that of NTHi is that the promoter for *rimO*, upstream of *ngt*, appears to be responsible for driving transcription of *ngt* and *agt* (figure 3). Recent studies have shown that in *Aggregibacter aphrophilus* and *Haemophilus ducreyi* [16,21] *hmwC* is also located downstream of *rimO*, although a transcriptional link has yet to be proven [21]. In *E. coli*, RimO is an enzyme that catalyses the methylthiolation of ribosomal subunit S12 at the universally conserved D88 residue. Furthermore, it has been shown that knocking out *rimO* in *E. coli* leads to a growth defect [38,39]. The signifi- cance of RimO has also been reported in *Thermus thermophilus*, where residue D88 cannot be mutated [40], leading to the conclusion that although methylthiolation is not essential in every organism, RimO clearly plays an important role in main- taining bacterial fitness [39]. Our tests indicate that the *A. pleuropneumoniae rimO* promoter is active at every time point tested, suggesting that *ngt* and *agt* are constitutively expressed (electronic supplementary material, figure S1), and therefore the cytoplasmic *N*-linking glycosylation system is always available to modify substrate proteins. Furthermore, qPCR analysis of the locus indicated transcriptional levels consistent with an operonic structure, where the highest level of transcription detected was of *rimO*, followed by *ngt* and *agt*, respectively (figure 4). This is in agreement with the findings reported by Lim *et al*. [41] whereby the expression level of the genes proximal to the promoter was greater than the ones farthest from it. Bioinformatic analysis of the DNA sequence surrounding the putative *rimO/ngt/agt* operon ident- ifies a transcriptional promoter just upstream of *rimO* and a Rho-independent terminator downstream of *agt* (electronic supplementary material, figure S2) [42]. Nevertheless, to gain further insights into the regulation of the locus, other approaches such as RNAseq could be carried out. Furthermore, *in silico* analysis of all publically available genomes and over 180 others (J.T.B. 2016, personal communication) indicates that the gene order is absolutely conserved (data not shown).

In this work, we also demonstrate that NGT plays an important biological role in the ability of *A. pleuropneumoniae* to adhere to A549 human adenocarcinoma lung epithelial cells, which, although from human origin, are from biologically relevant tissue. The rationale for using A549 cells, instead of St Jude Porcine lung (SJPL) cells, which have been widely used to assess *A. pleuropneumoniae* adhesion [4,43,44], was that the SJPL cell line was found to be misclassified, and is simian in origin [45]. In order to draw absolute conclusions regarding the role of this N-linked glycosylation system in aiding *A. pleur- opneumoniae* pathogenesis in the pig, the adhesion assay data that obtained in this study could be extended to investigate other tissues such as *ex vivo* organ cultures [46] possibly primary cell cultures from pig lung epithelial cells.

Similarly to our study, a significant reduction in adherence was reported for *E. coli* expressing the cloned hmw1 locus from NTHi when the function of HMW1C (the NGT orthologue) was removed [15]. The *hmw1/hmw2* loci and the *ngt* operon differ in that the NTHi loci encode an adhesin and an adhesin transporter alongside *hmwC*, whereas the *A. pleuropneumoniae* locus encodes an *a*6GlcT polymerizing glucosyltransferase (figure 1). Surprisingly, knocking out the function of the

α6GlcT transferase also resulted in a significant reduction in adherence comparable to that detected when NGT activity was abolished. Plasmid-based complementation with *agt* only (figure 2) and *ngt*/*agt* (data not shown) proved insufficient to rescue the adhesion phenotype seen with the wild-type. Our failed attempts to rescue the phenotype suggest a need for fine transcriptional control of *agt* levels in the bacterium. It is possible that, when *ngt* is being expressed in the chromo- some and *agt* is on a plasmid, over-glycosylation of target sites occurs, resulting in incorrect adhesin structural conforma- tion. Potentially, this incorrect folding could prevent surface presentation or adhesin function.

Some of the best-studied examples of glycosylated autotransporter adhesins are *O*-linked, and found in *E. coli*. *O*-linked glycosyltransferases can glycosylate TibA, Ag43 and AIDA-I in their passenger domains [47 – 49]. Bioinformatic analysis of the autotransporter adhesin used in our study, AtaC from *A. pleuropneumoniae* AP76 (GenBank accession number ACE61172.1), indicated the presence of 72 NX-(S/T) sequons. The majority of these (59/72) are localized in the passenger domain of the adhesin, further demonstrating the similarities with the *O*-linked counterparts. Whether glycosylation is required, so that the adhesin assumes the correct conformation and is not degraded as observed in AIDA-I, or if it is so that the adhesin can adopt a conformation suited for adhesion as described in TibA [50], remains to be determined even in these well-studied proteins.

Our demonstration of glucose polymer assembly within *E. coli* cells, when the *ngt*/*agt* locus is overexpressed alongside an acceptor protein (figure 5), led us to investigate if this glycan could be detected on the surface of *A. pleuropneumoniae*. Immunofluorescence studies using an anti-dextran antibody that recognizes isomaltotetraose as a minimum epitope failed to detect any signal, even in permeabilized cells (data not shown). This suggests that although the capability exists to form α-1,6 glu- cose chains greater than four subunits heterologously, in *A. pleuropneumoniae* the necessary epitope for detection with anti-dextran antibody does not appear to be formed. Analysis of the glycosylated peptides generated within *E. coli*, as determined by peak quantification of the HPLC chromatogram (figure 6), revealed a steady decrease in abundance of the oligosaccharide with increasing chain length. It was however noteworthy that the peak corresponding to Glc_2-2AB (retention time: 8.8 min) was considerably smaller than the peaks corresponding to Glc_1-2AB and Glc_3-2AB, indicating that the addition of the first α1–6-linked glucose might be considerably slower than the subsequent transfer reactions. This suggests that the first α6GlcT-catalysed reaction is, in fact, the rate-limiting step in the biosynthesis of these extended *N*-glycan chains.

In a review of the HMWC literature, we found instances where adhesin glycopeptides with dihexose modifications have been reported, in the absence of a co-localized ORF-like *agt* [15,51]. Rempe *et al*. [21] report dihexose modifications on four glycopeptides belonging to the autotransporter adhesin of *K. kingae*. These raise several possibilities; the first is that the reported glycopeptides actually contain two individual hexose attachments and not two hexoses together. Second, it may be possible that the HmwC from *K. kingae* is able to catalyse *N*-linked attachment and subsequent polymerization. Third, one cannot rule out there may be another glycosyltransferase in the genome that is enabling dihexose assembly.

A review of the NGT-specific literature reveals an interesting disparity in the function of this enzyme when tested *in vitro* and *in vivo*. Choi *et al*. [51] reported that *in vitro*, NGT is capable of forming dihexoses. However, this study indicates that α6GlcT is essential for glycosidic bond formation and extension of the glucose polymer *in vivo*. This is in agreement with a previous study by Naegeli *et al*., which failed to detect any polymerization when NGT alone was expressed in *E. coli* to glycosylate an acceptor protein [17].

Our finding that α6GlcT function is necessary to maintain adhesion in *A. pleuropneumoniae* indicates that this enzyme must be extending glucose residues at some sites within the autotransporter adhesins. However, by transferring the *N*-linking glycosylation locus into *E. coli*, we showed that α6GlcT and NGT are unable to fully complement each other's functions. Our study also provides further evidence that 'orphan' HMWC family of enzymes that have not evolved to be co-localized with their target substrate continue to modify proteins involved in adhesion. It is noteworthy that every bac- terial species reported thus far with this genetic arrangement uses the glycosylation system to target autotransporter adhesins [16,21]. Glycosylation has been linked to protection from proteo- lytic degradation, correct protein folding and correct transport to the surface, all of which would have an effect on cell adhesion. Further studies are ongoing to ascertain the level of interaction between α6GlcT/NGT and the target protein(s).

By demonstrating how to harness the *ngt*/*agt* operon, we have shown potential for glycoengineering applications, including the generation of *N*-linked glucose-based conjugate vaccines against *A. pleuropneumoniae*. The genetic conservation of the *ngt* operon in *A. pleuropneumoniae* would favour the development of a glycoconjugate vaccine against multiple *A. pleuropneumoniae* serovars. Other potential applications include the development of dextran-based conjugates that may be useful against bacteria such as *Helicobacter pylori* [52]. Recently, such conjugates have been shown to be immunogenic, and post-immune sera from rabbits vaccinated with dextran-based conjugates exhibited activity against strains of *H. pylori* that contain α(1 – 6) glucose as part of their LPS [52].

The field of bacterial glycobiology is burgeoning and investigations into various glycosylation systems, such as the NGT/α6GlcT system reported here, help to understand their functional roles. Our results demonstrate the importance of genetic and phenotypic screens for investigating glycosylation systems, and that this data can directly benefit bacterial glycoengineering.

Data accessibility. The datasets supporting this article have been uploaded as part of the electronic supplementary material.

Authors' contributions. V.S.T. and J.C. wrote the manuscript, carried out most of the experimental work and data analysis. J.T.B. and Y.L. constructed the *agt* and *ngt* mutants. A.N. and C.-W.L., carried out Mass Spectrometry analysis of the data. S.A. and P.V. helped develop NGT/AGT expression within *E. coli*. A.W.T., A.N.R., D.J.M., M.A., P.R.L. and B.W.W. conceived the study and revised the manuscript. All authors gave final approval for publication.

Competing interests. We have no competing interests.

Funding. This work was supported by a Longer and Larger (LoLa) grant from the UK Biotechnology and Biological Sciences Research Council (grant numbers BB/G020744/1, BB/G019177/1, BBG019274/1 and BB/G003203/1) and the Wellcome Trust (grant number 102979/Z/13/Z). J.C. was supported by the Wellcome Trust's Institutional Strategic Support Fund (grant number 105609/Z/14/Z).

Acknowledgements. This work was supported by a Longer and Larger (LoLa) grant from the UK Biotechnology and Biological Sciences Research Council (grant nos BB/G020744/1, BB/G019177/1, BB/G019274/1 and BB/G003203/1) and The Wellcome Trust (grant no. 102979/Z/13/Z). The BRaDP1T Consortium comprises: Duncan

J. Maskell, Alexander W. (Dan) Tucker, Sarah E. Peters, Lucy A. Weinert, Jinhong (Tracy) Wang, Shi-Lu Luan, Roy R. Chaudhuri (University of Cambridge; present address for R. Chaudhuri is Centre for Genomic Research, University of Liverpool, Crown Street, Liverpool, L69 7ZB, UK.); Andrew N. Rycroft, Gareth A. Maglennon, Dominic Matthews (Royal Veterinary College); Brendan W. Wren, Jon Cuccui, Vanessa S. Terra (London School of Hygiene and Tropical Medicine); and Paul R. Langford, Janine T. Bossé, Yanwen Li (Imperial College London). We are grateful to Jun Wheeler for her assistance in protein identification, Dr Emily Kay, Dr Andreas Ioannis Karsisiotis and Dr Alexandra Faulds-Pain for their critical review of the manuscript.

References

1. ProHealth 2015 Production diseases: the cost to pig producers. See www.fp7-prohealth.eu/news-index/newsletter-november-2015/production-diseases-cost-pig-producers.
2. Dom P, Haesebrouck F, Ducatelle R, Charlier G. 1994 *In vivo* association of *Actinobacillus pleuropneumoniae* serotype 2 with the respiratory epithelium of pigs. *Infect. Immun*. 62, 1262–1267.
3. Boekema BK, Stockhofe-Zurwieden N, Smith HE, Kamp EM, van Putten JP, Verheijden JH. 2003 Adherence of *Actinobacillus pleuropneumoniae* to primary cultures of porcine lung epithelial cells. *Vet. Microbiol*. 93, 133 – 144. (doi:10.1016/S0378-1135(03)00020-8)
4. Auger E, Deslandes V, Ramjeet M, Contreras I, Nash JHE, Harel J, Gottschalk M, Olivier M, Jacques M. 2009 Host-pathogen interactions of *Actinobacillus pleuropneumoniae* with porcine lung and tracheal epithelial cells. *Infect. Immun*. 77, 1426–1441. (doi:10.1128/IAI.00297-08)
5. Sarkozi R, Makrai L, Fodor L. 2015 Identification of a proposed new serovar of *Actinobacillus Pleuropneumoniae*: Serovar 16. *Acta Vet. Hung*. 63, 444 – 450. (doi:10.1556/004.2015.041)
6. Van Overbeke I, Chiers K, Charlier G, Vandenberghe I, Van Beeumen J, Ducatelle R, Haesebrouck F. 2002 Characterization of the *in vitro* adhesion of *Actinobacillus pleuropneumoniae* to swine alveolar epithelial cells. *Vet. Microbiol*. 88, 59–74. (doi:10.1016/S0378-1135(02)00080-9)
7. Ramjeet M, Deslandes V, St Michael F, Cox AD, Kobisch M, Gottschalk M, Jacques M. 2005 Truncation of the lipopolysaccharide outer core affects susceptibility to antimicrobial peptides and virulence of *Actinobacillus pleuropneumoniae* serotype 1. *J. Biol. Chem*. 280, 39 104–39 114. (doi:10.1074/jbc.M502852200)
8. Howard SL *et al.* 2009 *Campylobacter jejuni* glycosylation island important in cell charge, legionaminic acid biosynthesis, and colonization of chickens. *Infect. Immun*. 77, 2544–2556. (doi:10.1128/IAI.01425-08)
9. Iwashkiw JA *et al.* 2012 Identification of a general O-linked protein glycosylation system in *Acinetobacter baumannii* and its role in virulence and biofilm formation. *PLoS Pathog*. 8, e1002758. (doi:10.1371/journal.ppat.1002758)
10. Ciocchini AE, Rey Serantes DA, Melli LJ, Iwashkiw JA, Deodato B, Wallach J, Feldman MF, Ugalde JE, Comerci DJ. 2013 Development and validation of a novel diagnostic test for human brucellosis using a glyco-engineered antigen coupled to magnetic beads. *PLoS Negl. Trop. Dis*. 7, e2048. (doi:10.1371/journal.pntd.0002048)
11. Wacker M *et al.* 2014 Prevention of *Staphylococcus aureus* infections by glycoprotein vaccines synthesized in *Escherichia coli*. *J. Infect. Dis*. 209, 1551–1561. (doi:10.1093/infdis/jit800)
12. Wacker M *et al.* 2002 N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* 298, 1790 – 1793. (doi:10.1126/science.298.5599.1790)
13. Linton D *et al.* 2005 Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol. Microbiol*. 55, 1695–1703. (doi:10.1111/j.1365-2958.2005.04519.x)
14. Twine SM, Reid CW, Aubry A, McMullin DR, Fulton KM, Austin J, Logan SM. 2009 Motility and flagellar glycosylation in *Clostridium difficile*. *J. Bacteriol*. 191, 7050–7062. (doi:10.1128/JB.00861-09)
15. Grass S, Buscher AZ, Swords WE, Apicella MA, Barenkamp SJ, Ozchlewski N, St Geme JW. 2003 The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Mol. Microbiol*. 48, 737 –751. (doi:10.1046/j.1365-2958.2003.03450.x)
16. McCann JR, St Geme III JW. 2014 The HMW1C-like glycosyltransferases–an enzyme family with a sweet tooth for simple sugars. *PLoS Pathog*. 10, e1003977. (doi:10.1371/journal.ppat.1003977)
17. Naegeli A, Neupert C, Fan YY, Lin CW, Poljak K, Papini AM, Schwarz F, Aebi M. 2014 Molecular analysis of an alternative N-glycosylation machinery by functional transfer from *Actinobacillus pleuropneumoniae* to *Escherichia coli*. *J. Biol. Chem*. 289, 2170 – 2179. (doi:10.1074/jbc.M113.524462)
18. Naegeli A, Michaud G, Schubert M, Lin CW, Lizak C, Darbre T, Reymond J-L, Aebi M. 2014 Substrate specificity of cytoplasmic N-glycosyltransferase. *J. Biol. Chem*. 289, 24 521–24 532. (doi:10.1074/jbc.M114.579326)
19. Schwarz F, Fan YY, Schubert M, Aebi M. 2011 Cytoplasmic N-glycosyltransferase of *Actinobacillus pleuropneumoniae* is an inverting enzyme and recognizes the NX(S/T) consensus sequence. *J. Biol. Chem*. 286, 35 267 – 35 274. (doi:10.1074/jbc.M111.277160)
20. Kawai F, Grass S, Kim Y, Choi K-J, St Geme JW, Yeo H-J. 2011 Structural insights into the glycosyltransferase activity of the *Actinobacillus pleuropneumoniae* HMW1C-like protein. *J. Biol. Chem*. 286, 38 546–38 557. (doi:10.1074/jbc.M111.237602)
21. Rempe KA, Spruce LA, Porsch EA, Seeholzer SH, Nørskov-Lauritsen N, St Geme JW. 2015 Unconventional N-Linked Glycosylation Promotes Trimeric Autotransporter Function in *Kingella kingae* and *Aggregatibacter aphrophilus*. *MBio* 6, e01206-15. (doi:10.1128/mBio.01206-15)
22. Xiao L *et al.* 2012 Apa is a trimeric autotransporter adhesin of *Actinobacillus pleuropneumoniae* responsible for autoagglutination and host cell adherence. *J. Basic Microbiol*. 52, 598–607. (doi:10.1002/jobm.201100365)
23. Ferrieres L, Hemery G, Nham T, Guerout AM, Mazel D, Beloin C, Ghigo J-M. 2010 Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol*. 192, 6418–6427. (doi:10.1128/JB.00621-10)
24. Cuccui J, Easton A, Chu KK, Bancroft GJ, Oyston PCF, Titball RW, Wren BW. 2007 Development of signature-tagged mutagenesis in *Burkholderia pseudomallei* to identify genes important in survival and pathogenesis. *Infect. Immun*. 75, 1186–1195. (doi:10.1128/IAI.01240-06)
25. Foote SJ, Bosse JT, Bouevitch AB, Langford PR, Young NM, Nash JH. 2008 The complete genome sequence of *Actinobacillus pleuropneumoniae* L20 (serotype 5b). *J. Bacteriol*. 190, 1495 – 1496. (doi:10.1128/JB.01845-07)
26. Bosse JT *et al.* 2014 The generation of successive unmarked mutations and chromosomal insertion of heterologous genes in *Actinobacillus pleuropneumoniae* using natural transformation. *PLoS ONE* 9, e111252. (doi:10.1371/journal.pone.0111252)
27. Bosse JT, Nash JH, Kroll JS, Langford PR. 2004 Harnessing natural transformation in *Actinobacillus pleuropneumoniae*: a simple method for allelic replacements. *FEMS Microbiol. Lett*. 233, 277–281. (doi:10.1111/j.1574-6968.2004.tb09492.x)
28. Bosse JT, Durham AL, Rycroft AN, Kroll JS, Langford PR. 2009 New plasmid tools for genetic analysis of *Actinobacillus pleuropneumoniae* and other pasteurellaceae. *Appl. Environ. Microbiol*. 75, 6124 – 6131. (doi:10.1128/AEM.00809-09)
29. Pumirat P, Cuccui J, Stabler RA, Stevens JM, Muangsombut V, Singsuksawat E, Stevens MP, Wren BW, Korbsrisate S. 2010 Global transcriptional profiling of *Burkholderia pseudomallei* under salt stress reveals differential effects on the Bsa type III secretion system. *BMC Microbiol*. 10, 171. (doi:10.1186/1471-2180-10-171)
30. Nielsen KK, Boye M. 2005 Real-time quantitative reverse transcription-PCR analysis of expression

stability of *Actinobacillus pleuropneumoniae* housekeeping genes during *in vitro* growth under iron-depleted conditions. *Appl. Environ. Microbiol*. 71, 2949 – 2954. (doi:10.1128/AEM.71.6.2949-2954.2005)

31. Bigge JC, Patel TP, Bruce JA, Goulding PN, Charles SM, Parekh RB. 1995 Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Anal. Biochem*. 230, 229–238. (doi:10.1006/abio.1995.1468)
32. Wisniewski JR, Zougman A, Nagaraj N, Mann M. 2009 Universal sample preparation method for proteome analysis. *Nat. Methods*. 6, 359 – 362. (doi:10.1038/nmeth.1322)
33. Dykxhoorn DM, St Pierre R, Linn T. 1996 A set of compatible tac promoter expression vectors. *Gene* 177, 133–136. (doi:10.1016/0378-1119(96)00289-2)
34. Young NM *et al.* 2002 Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J. Biol. Chem*. 277, 42 530 – 42 539. (doi:10.1074/jbc.M206114200)
35. Pustylnikov S, Sagar D, Jain P, Khan ZK. 2014 Targeting the C-type lectins-mediated host-pathogen interactions with dextran. *J. Pharm. Pharm. Sci*. 17, 371–392. (doi:10.18433/J3N590)
36. Cuskin F *et al.* 2015 Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* 517, 165–169. (doi:10.1038/nature13995)
37. Harding CM *et al.* 2015 Acinetobacter strains carry two functional oligosaccharyltransferases, one devoted exclusively to type IV pilin, and the other one dedicated to O-glycosylation of multiple proteins. *Mol. Microbiol.* 96, 1023–1041. (doi:10.1111/mmi.12986)
38. Kowalak JA, Walsh KA. 1996 Beta-methylthio-aspartic acid: identification of a novel posttranslational modification in ribosomal protein S12 from *Escherichia coli*. *Protein Sci*. 5, 1625 – 1632. (doi:10.1002/pro.5560050816)
39. Anton BP, Saleh L, Benner JS, Raleigh EA, Kasif S, Roberts RJ. 2008 RimO, a MiaB-like enzyme, methylthiolates the universally conserved Asp88 residue of ribosomal protein S12 in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* 105, 1826–1831. (doi:10.1073/pnas.0708608105)
40. Carr JF, Hamburg D-M, Gregory ST, Limbach PA, Dahlberg AE. 2006 Effects of streptomycin resistance mutations on posttranslational modification of ribosomal protein S12. *J. Bacteriol*. 188, 2020 – 2023. (doi:10.1128/JB.188.5.2020-2023.2006)
41. Lim HN, Lee Y, Hussein R. 2011 Fundamental relationship between operon organization and gene expression. *Proc Natl Acad. Sci. USA* 108, 10 626–10 631. (doi:10.1073/pnas.1105692108)
42. Solovyev AS. 2011 Automatic annotation of microbial genomes and metagenomic sequences. In *Metagenomics and its applications in agriculture, biomedicine and environmental studies* (ed. RW Li), pp. 61 – 78. Hauppauge, NY: Nova Science Publishers.
43. Levesque C, Provost C, Labrie J, Hernandez Reyes Y, Burciaga Nava JA, Gagnon CA, Jacques M. 2014 Actinobacillus pleuropneumoniae possesses an antiviral activity against porcine reproductive and respiratory syndrome virus. *PLoS ONE* 9, e98434. (doi:10.1371/journal.pone.0098434)
44. Liu J, Hu L, Xu Z, Tan C, Yuan F, Fu S, Cheng H, Chen H, Bei W. 2015 *Actinobacillus pleuropneumoniae* two-component system QseB/QseC regulates the transcription of PilM, an important determinant of bacterial adherence and virulence. *Vet. Microbiol*. 177, 184–192. (doi:10.1016/j.vetmic.2015.02.033)
45. Silversides DW, Music N, Jacques M, Gagnon CA, Webby R. 2010 Investigation of the species origin of the St. Jude Porcine Lung epithelial cell line (SJPL) made to researchers. *J. Virol*. 84, 5454–5455. (doi:10.1128/JVI.00042-10)
46. Dugal F, Belanger M, Jacques M. 1992 Enhanced adherence of *Pasteurella multocida* to porcine tracheal rings preinfected with *Bordetella bronchiseptica*. *Can. J. Vet. Res*. 56, 260–264.
47. Elsinghorst EA, Weitz JA. 1994 Epithelial cell invasion and adherence directed by the enterotoxigenic *Escherichia coli* tib locus is associated with a 104-kilodalton outer membrane protein. *Infect. Immun*. 62, 3463–3471.
48. Benz I, Schmidt MA. 2001 Glycosylation with heptose residues mediated by the aah gene product is essential for adherence of the AIDA-I adhesin. *Mol. Microbiol*. 40, 1403 – 1413. (doi:10.1046/j.1365-2958.2001.02487.x)
49. van der Woude MW, Henderson IR. 2008 Regulation and function of Ag43 (flu). *Annu. Rev. Microbiol*. 62, 153 – 169. (doi:10.1146/annurev.micro.62.081307.162938)
50. Cote J-P, Charbonneau M-E, Mourez M. 2013 Glycosylation of the *Escherichia coli* TibA self-associating autotransporter influences the conformation and the functionality of the protein. *PLoS ONE* 8, e80739. (doi:10.1371/journal.pone.0080739)
51. Choi K-J, Grass S, Paek S, St Geme JW3rd, Yeo H-J. 2010 The *Actinobacillus pleuropneumoniae* HMW1C-like glycosyltransferase mediates N-linked glycosylation of the *Haemophilus influenzae* HMW1 adhesin. *PLoS ONE* 5, e15888. (doi:10.1371/journal.pone.0015888)
52. Altman E, Chandan V, Harrison B. 2014 The potential of dextran-based glycoconjugates for development of *Helicobacter pylori* vaccine. *Glycoconj. J.* 31, 13–24. (doi:10.1007/s10719-013-9496-4)

Chapter 7

GENERAL DISCUSSION

7 GENERAL DISCUSSION

Campylobacter jejuni is a major cause of bacterial gastroenteritis worldwide. Studies estimate the disease burden of *C. jejuni* to be associated with 7.5 million disability-adjusted life years in the 2010 Global Burden of Disease study, exceeding *Shigella* (7.1 million) and enterotoxigenic *E. coli* (6.9 million) (Murray *et al.*, 2012). Disease presentation varies between self-limiting illness manifested by diarrhoea, fever and malaise to autoimmune conditions such as Guillian-Barre syndrome (GBS) and Millller Fischer syndrome (Kaakoush *et al.*, 2015). The bacterium *C. jejuni* possess the first reported and most studied bacterial general *N*-glycosylation system, with more than 50 proteins are glycosylated with *C. jejuni* heptasaccharide GalNAc-α1,4-GalNAc-[Glcβ1,3-]GalNAc- α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-β1, where GalNAc is *N*-acetylgalactosamine, Glc is glucose, Bac is 2,4-diacetamido -2,4,6-trideoxyglucopyranose. This glycostructure is linked to the asparagine residue in the acceptor sequon D/E-X-N-Y-S/T where X and Y can be any amino acid except proline (Scott *et al.*, 2010).

Advances in genome sequencing and glycans analysis methodologies have expanded our knowledge of bacterial glycosylation systems. These discoveries however have not been followed by investigations aimed at answering why bacteria possess glycosylation system, how bacteria glycosylate proteins and what are the biological roles of these glycans. The journey of unravelling the role of these intriguing post-translational modifications starts by establishing platforms to interrogate the role of these glycosylation systems. Without these, research into bacterial glycosylation will be limited to reporting the presence or absence of these systems.

This thesis aimed to demonstrate the role of *N*-glycosylation in bacteria as well as the potential prevalence of these general N-glycosylation systems among bacterial families. In this chapter, the findings in the reported research papers come together to provide insights about bacterial *N*-glycosylation systems.The following lists the key findings of this thesis:

1- Bacterial *N*-glycosylation is responsible for proteome stability in *C. jejuni.*

2- Disruption of *N*-glycosylation is accompanied by an increase in chaperones and protease abundance which might indicate an impairment in proteins folding.

3- Abolition of *N*-glycoyslation led to impairment of important glycoprotein complexes activities such as major multidrug efflux pump (CmeABC) and nitrate reductase complex (NapAB) .

4- Chicken colonisation was substantially reduced upon inactivating *pgl*.

5- *N*-glycans modulates CmeA folding and improve protein thermostability *in vitro.*

6- *C. jejuni* heptasaccharide promote protein-protein interaction, this has been clearly demonstrated in the difference in binding kinetics between CmeA and gCmeA to its binding partner CmeC.

7- Deep sea vent bacteria possesss functional *N*-OSTs that differ in substrate preference. This also indicates the prevalence of *N*-glycosylation among non-epsilonproteobacteria.

8- Non-conventional *N*-glycosylation, cytoplasmic *N*-glycosylation, is required for host-pathogen interaction.

7.1 *N*-glycosylation: lessons from diversity and conservation

The combination of next generation sequencing and high-throughput proteomics enabled the scientific community to identify different bacterial glycosylation pathways (Jervis *et al.*, 2010, 2018; Nothaft *et al.*, 2012; Nothaft and Szymanski, 2013). Although the genetic organisation of *N*-glycosylation loci might vary, bioinformatic analysis demonstrated the presence of PglB orthologues in 50 bacterial species representing seven different bacterial families, with some species possessing two PglB copies. In addition, the functional activity of PglB orthologues has also been demonstrated (chapter five) (Nothaft and Szymanski, 2010; Ollis *et al.*, 2015; Mills *et al.*, 2016). The presence of an *N*-OST within so many organisms could hypothetically be responsible for maintaining bacterial proteome stability. In chapter three, the role of *N*- glycosylation was demonstrated to modulate protein complexes activity. Inactivation of *pglB* leads to abolishment of glycosylation of all glycoproteins in *C. jejuni* (Jervis *et al.*, 2012).Thus, insertional mutagenesis in *pglB* allowed us to establish a glycosylation deficient *C. jejuni* to study the role of *N*-glycosylation in *C. jejuni.* Proteome analysis of a glycosylation deficient strain exhibited a high abundance in protein quality control machineries. This might be due to the aggregation of non-natively folded proteins. The high abundance of protein quality machineries was not accompanied by an alteration in growth kinetics. When tested, *C. jejuni pglB*::*aphA* cell count was reduced two fold when compared to *C. jejuni* wildtype, however, both strains had the same dry weight (chapter three).

Abolishment of *N*-glycans also affected protein complexes formation as demonstrated in the impairment of the multidrug efflux pump and nitrate

reductase complex. Bacterial fitness was also severely compromised, resulting in a substantial reduction in chicken colonisation. The effect of abolishing *N*-glycosylation was seen to be extended beyond glycoproteins function. We speculate that the role of general *N*- glycosylation might be common among other bacterial families. Investigating the role of *N*-glycosylation in other bacterial families would confirm this hypothesis. Interestingly, most of bacteria possessing PglB orthologues are fastidious organisms with some requires specific growth conditions such as, *Deferribacter desulfuricans* that grows at 60-65°C with doubling time 40 min (Takai *et al.*, 2003), *Desulfovibrio gigas* which grows at 35°C under anaerobic conditions and *Nitratiruptor tergarcus,* grows at 55°C, 80%H_2+20% CO_2 at 350 kPa (http://bacdive.dsmz.de). Of note, the *N*-linked glycan structures within these bacteria are still to be unravelled. Notably, investigating 16 different glycostructures from *Campylobacter spp.* demonstrated that first two monosaccharides; GalNAc-α-1,4-GalNAc-α-1,3-diBacNAc are conserved (Jervis *et al.*, 2012; Nothaft and Szymanski, 2013). Conservation of *N*-linked glycans among species is a common theme in eukaryotes (Varki, 2017). In Eukaryotes, glycosylation occurs co- and post-translationally in the endoplasmic reticulum to promote polypeptide folding and protein quality control (Jayaprakash and Surolia, 2017). However, in prokaryotes protein glycosylation occurs post-translationally whereby majority of the glycosylated proteins in *C. jejuni* are translocated to the periplasm unfolded via sec pathway, where they get glycosylated by PglB (Szymanski *et al.*, 2003). The periplasmic space is speculated to be evolved as first extracytoplasmic space where proteins transport occurs as well as protein folding and oxidation, hence it is not uncommon to draw similarities between eukaryotes endoplasmic reticulum and bacterial periplasm (Miller and Salama, 2018).

Atomic structure of the yeast oligosaccharyltransferase complex revealed that, in standalone OST complexes (STT3B-containiing OST complexes), the acceptor polypeptide side faces an oxidoreductase domain, featured in the thioredoxin fold of OST6 or OST3 (Bai *et al.*, 2018; Wild *et al.*, 2018) **Fig 8 A**. Interestingly, our quantitative proteomics (chapter three) demonstrated that the bacterial periplasmic oxidoreductase DsbA, that is responsible for intrachain disulphide formation, was significantly reduced in abundance in glycosylation deficient *C. jejuni* (*P*<0.0001) **Fig 8, B** In the light of our quantitative proteomics finding, we speculate that bacterial general *N*-glycosylation is coupled to protein folding, whereby, disulphide bonds are introduced once a protein gets glycosylated in the periplasm. This might also suggest a further interaction between PglB and Sec translocation machinery in bacteria.

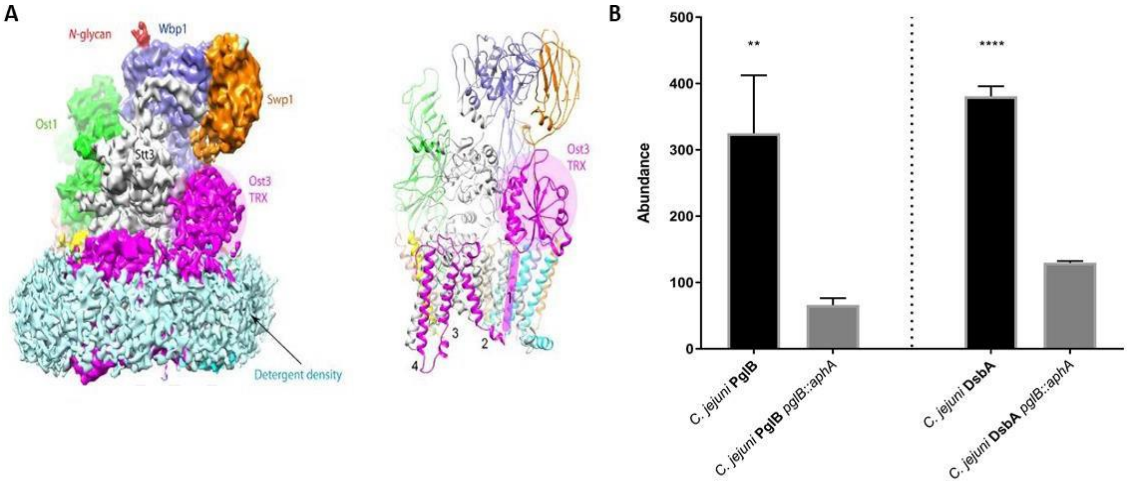


Fig 8 Oligosaccharyltransferase and oxidoreductase A) EM density of map of yeast OST complex. Thioredoxin (TRX) domain is highlighted in magenta and appears to be in close proximity to STT3 active subunit the image is adapted from Bai *et al* 2018. B) Change in protein abundance evaluated by TMT 6plex LC-MS/MS. The result represents the mean of three biological replicates, error bars represents standard deviation. Significance was calculated by student's *t*-test, ***P*<0.01 and *****P*<0.0001

Cross-linking MS and CryoEM could be used to unravel this type of interaction.This highlights a further potential similarity between bacterial and eukaryotic protein *N*-glycosylation. It also suggests a common theme in protein

quality control across all domains of life. This also might help us to understand the evolution of how *N*-glycosylation is coupled to sulphide bond formation and protein quality control.

The exciting finding that some bacteria possess two PglB copies raise the question about the functional role of each PglB copy. Whilst this finding was not common among prokaryotic glycosylation system, it is prevalent in eukaryotes. The catalytic subunits STT3A and STT3B in eukaryotic OST were found to vary in function. STT3 containing OST are associated with translocon machinery hence responsible for efficient co-translational glycosylation as the nascent polypeptide chain enters the ER, whilst STT3B are standalone OST complexes that glycosylate adjacent N-terminals of secreted polypeptide chains (Ruiz-Canada, Kelleher and Gilmore, 2009; Wild *et al.*, 2018). Indeed, Functional analysis of each of the bacterial OST two copies suggest substrate specificity (Jervis *et al.*, 2018).

The cross-talk between general glycosylation machineries and protein folding chaperones/oxidoreductase is not uncommon. Studies in the last couple of decades have uncovered a common evolutionary pathway not only among bacterial strains but also between eukaryotes and prokaryotes (Schwarz and Aebi, 2011). These similarities indicate a potential common evolutionary mechanism of protein quality control that is conserved among all kingdoms of life.

7.2 *N*-linked glycans role in modulating protein function

It is certain that eukaryotic *N*-linked glycans are associated with protein folding, stabilisation of glycoproteins and protein-protein interaction (Freeze, Esko and Parodi, 2009; Hanson *et al.*, 2009; Jayaprakash and Surolia, 2017). However, the role of bacterial *N*-linked glycans is still understudied. In chapter three, the

role of *N*-glycosylation has been assessed, in depth analysis of the role of *N*-glycans in specific protein has been discussed in chapter four, where the biophysical properties attributed by *N*-linked glycans were studied in CmeA, a member of the major multidrug efflux pump, CmeABC. CmeABC is a tripartite multidrug efflux pump consisting of an assembly that plays a pivotal role in chicken colonisation, extrusion of structurally unrelated compounds. It also plays a critical role in the clinical manifestation of multidrug resistance(Lin, Overbye Michel and Zhang, 2002). Quantitative proteomics analysis in chapter three did not show difference in abundance in CmeABC between the wildtype and the glycosylation deficient *C. jejuni*. However, a remarkable difference in multidrug efflux pump activity was observed when *N*-glycoyslation is no longer possible. Since there was no difference in abundance in CmeABC was observed, this means that impairment of *N*-glycosylation did not affect the protein quantity in the cell, therefore, reduction in activity seen in glycosylation deficient strain could be attributed to a role played by *N*-glycans. This intrigued us to investigate the role of *C. jejuni* heptasaccharide in modulating CmeA function and stability (chapter four). To systematically confirm that difference in ethidum bromide accumulation is solely due to the lack of *N*-glycans and not any other factor, we systematically designed the following strategy (chapter four). First, we inactivated a secondary multidrug efflux pump; CmeDEF, to avoid masking any difference that could be due to altering glycosylation of CmeABC. Secondly, we constructed a glycosylation deficient CmeABC in *cmeD* impaired strain. This strain has a functioning *pglB* however, only CmeABC is not glycosylated due to altering the asparagine in the glycosylation sequon to glutamine. Thirdly, we carried out ethidium accumulation assay between both strains (wtCmeABC and g0CmeABC). Our results indicated a statistically significant reduction in multidrug efflux in g0CmeABC (chapter

four). We then investigated the role of *N*-glycans in individual protein of the CmeABC complex, CmeA CmeA is glycosylated by *C. jejuni* heptassacharide at ^{123}N and ^{273}N. The thermal melts profile of CmeA variants (g0CmeA and g2CmeA) were compared, showing different unfolding profiles. Glycosylation clearly stabilised g2CmeA by 6.4°C ±0.5 when compared to its non-glycosylated counterpart. The thermostablisation effect of *N*-linked glycans was conferred by reversibility studies showing that the unfolding time of g2CmeA was reduced 6 folds when compared to the unfolding time of CmeA. These results comply with the finding that that core eukaryotic *N*-linked glycans enhance glycoproteins thermostability and decrease the unfolding time (Hanson *et al.*, 2009; Biswas and Chattopadhyaya, 2016). When the extrinsic effect of *C. jejuni N*-linked glycans was tested, g2CmeA showed a lower K_D to protein partner CmeC, than g0CmeA (chapter four). Delineation of the role of each glycosylation site could reveal different roles of *N*-linked glycans played at different acceptor sequons. Studies conducted in the tyrosinase glycoprotein showed that glycosylation at ^{94}N and ^{104}N did not enhance protein stability nor function, whilst glycosylation at ^{181}N and ^{304} N where found to be critical for protein transport and enzymatic activity, respectively (Jayaprakash and Surolia, 2017). Our bioinformatic analysis demonstrated that in g2CmeA, ^{123}N is located at X_{-1} of the RLS binding motif to CmeC. This finding was conserved in CmeA from *C. jejuni* and *C. coli*. This exciting finding indicates a direct or indirect role of *N*-linked glycans in modulating CmeABC multidrug efflux pump assembly. For CmeABC to be assembled, six molecules of CmeA bind to three molecules of each CmeB and CmeC. Thus, oligomerisation is critical in multidrug pump assembly. This interaction of *N*-glycans could be directly forcing structural arrangement of the glycosylation sequon or as a result of indirect interaction of *N*-glycans with the protein aminoacid backbone (Tikhonova *et al.*, 2009; Su *et*

al., 2014; Jeong *et al.*, 2016).

The role of *N*-linked glycans in promoting protein oligomerisation has been debatable (Mitra *et al.*, 2006; Jayaprakash and Surolia, 2017). Quercetin 2,3-dioxygenase, a copper- containing dimeric enzyme is glycosylated at five different sites, with *N*-linked glycans attributing to approximately 25% of its molecular weight. Structure analysis of this enzyme revealed that glycosylation at ^{191}N could promote stable dimerization of the enzyme. The oligosaccharide was found to modulate protein-carbohydrate interaction by packing against β turn 185-191 of the other molecule of the dimer (Fusetti *et al.*, 2002) It is still unclear whether *C. jejuni N*-linked glycans directly confer CmeA oligomerization through modulating protein-carbohydrate interaction or indirectly through causing local site structure rearrangements that enhance the formation of oligomers. To test this hypothesis, CmeA and gCmeA (used in the study in chapter four) were run on a Blue native gel (BN-native gel) to assess their oligmerisation state. By comparing the oligmerization state of CmeA variants on BN native. PAGE g2CmeA was found to exist in two forms (monomer and trimer) however, CmeA was found in three forms (monomer, dimer and trimer). Since the difference between CmeA and gCmeA is the presence of *N*-glycans (chapter four).This result might suggest that *N*-glycans helps in structurally arranging the CmeABC monomers in order to either interact with each other (**Fig 9)** and/or interact with its binding partner (surface plasmon resonance results in chapter four).Thus, ensuring that *N*-glycans might be playing different roles when attached to different sites on the same protein.

Size exclusion chromatography multi-angle light scattering (SEC-MALS) could give a definite answer for the oligomerisation state of g2CmeA and g0CmeA as well as calculating the total molecular weight of the oligomers.**Fig 9**.

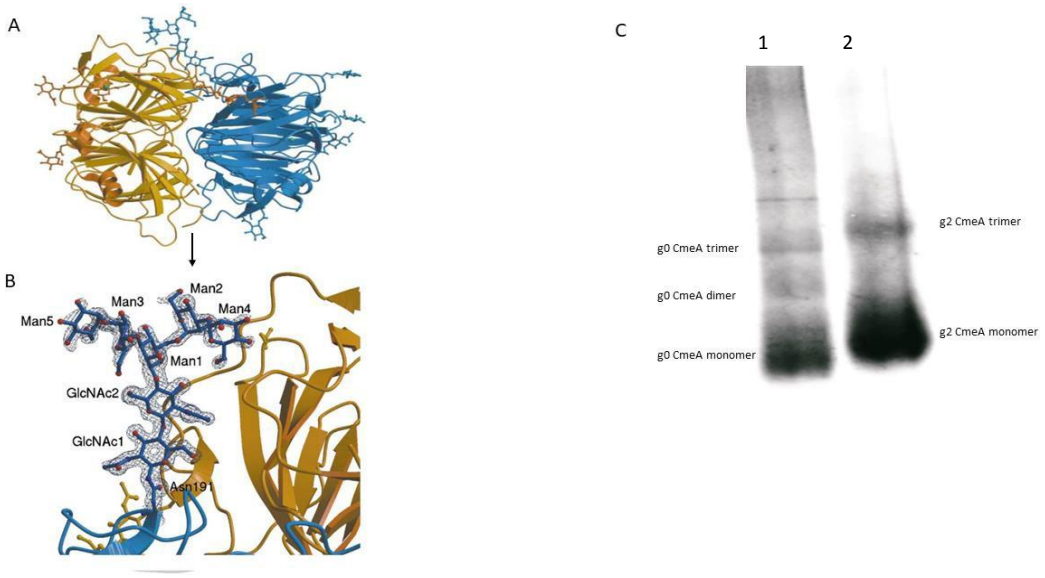


Fig 9 Role of *N*-linked glycans in promoting protein oligomerization. A) Side view of 2,3 Quercetin 2,3- dioxygenase dimer structure showing *N*-linked glycans in ball and stick, monomers are coloured in blue and gold, B) An electron density map of the *N*-linked glycans, shown in balls and sticks, presented in panel A, images are adapted from Fusetti, *et al* 2002, C) Blue Native PAGE of g0CmeA, lane 1; and g2CmeA, lane 2. Equal amount of proteins was loaded and ran for 90 minutes then visualised by Coomassie staining.

7.3 Glycosylation and host-microbe interaction

Birds, especially chickens are regarded as the natural host of *C. jejuni* (Newell and Fearnley, 2003). For years, *C. jejuni* has been considered as a part of the chicken normal flora, however, mounting evidences suggest that *C. jejuni* might cause intestinal damage in the chicken host (Awad *et al.*, 2015). Whilst the underlying mechanism by which the bacterium might be causing intestinal damage to the host remains poorly understood, chicken colonisation is still considered as one of the main factors to assess gene function and/or attenuation in *C. jejuni* (Jones *et al.*, 2004) .Previous studies demonstrated a significant reduction in chicken colonisation in *pglH::aphA C. jejuni* when compared to the wildtype. PglH is a glycosyltransferase that catalyses the addition of three α1,4- GalNAc to GalNAc- α1,4-GalNAc-α1,3-diBacNAc-β1, an

important step in the production of *C. jejuni N*-glycan. Inactivation of *pglH* results in alteration of the glycoproteome of *C. jejuni* however, does not entirely abolish protein glycosylation. Thus,unlike in *pglB*::*aphA C. jejuni,* proteins in *pglH* deficient *C. jejuni* would still be glycosylated with the truncated glycan (GalNAc-α1,4-GalNAc-α1,3-Bac-β1). In chapter three, the impact of *N*-glycosylation disruption on chicken colonisation was assessed.Twenty 2-week-old *Campylobacter*-free White Leghorn chickens with 10^6 CFU *C. jejuni* wild-type or *pglB*::*aph* mutant. Ten chickens from each group were sacrificed at days 6 or 13 post-inoculation and viable *C. jejuni* in the caeca (a key site of persistence) enumerated following post mortem. Glycosylation deficient *pglB*::*aph C. jejuni* showed a six–$\log_{10}$ reduction in caecal colonization when compared to the wild- type at day 6 (P<0.0001) and day 13 (P<0.0001).

The substantial reduction in chicken colonisation could be linked to direct and/or indirect effects emerged as a result of *N*-glycosylation disruption. Directly, disruption of *N*-glycosylation has been shown to induce stress response in *C. jejuni* and also impair protein function (chapter three). Indirectly, surface exposed *N*-glycans could play a role in interaction with the host cell receptors. This protein-glycan interaction could be involved in the establishment of chicken colonisation. During my PhD examination, Prof Ten Fiezi, raised an important question of whether the interaction between *C. jejuni N*-glycans and chicken epithelial cells has been investigated. Such interaction could be studied using glycans arrays available at Imperial College glycoscience facility, either by investigating the presence of any lectin-like proteins expressed on the surface of *C. jejuni* that binds to glycans presented on the surface of chicken gut or the opposite (lectin-like proteins expressed on the surface of the chicken gut that interacts with *C. jejuni N*-glycans. This will

allow us to understand the molecular basis of host-microbe interaction in *C. jejuni* and underly the mechanism by which *N*-glycosylation plays a pivotal role in establishmed of chicken colonisation.

7.4 Cytoplasmic glycosylation

The versatility of bacteria to modify its proteins exemplify the necessity of this post- translational modification. The direct link between adhesin tethering and cytoplasmic *N*-glycosylation has been reported in NTHi. Biochemical studies revealed an unconventional cytoplasmic *N*-glycosylation system enabled by an enzyme known as NGT (Grass *et al.*, 2010). This discovery led to further investigations demonstrating the prevalence of this system not only among members of the *Pasteurallacea* family but also other bacterial families (Schwarz *et al.*, 2011; Rempe *et al.*, 2015). The system is analogous to the intensively studied bacterial cytoplasmic *O*-glycosylation systems. Interestingly, both cytoplasmic *N*- and *O*-glycosylation decorates autotransporters (Benz and Schmidt, 2001; Grass *et al.*, 2010). In chapter three and four, the study of general *N*-glycosylation as well as the role of *N*-glycans has been investigated. To address the functional role of the specific *N*-glycosylation, the pig pathogen *App* was used as a model. In chapter six, the study of the role of *N*-glycosylation in *App* demonstrated the importance of this post translational modification in cell adhesion (J. Cuccui *et al.*, 2017). The implications of this finding increases our understanding of the functional role of *N*-glycans in protein function, it also comes in agreement with the previous work which demonstrated that abolishing of *O*-glycosylation of AIDA affected bacterial adhesion to cultured epithelial cells (Charbonneau *et al.*, 2012). This indicates that both *N*- or *O*- cytoplasmic glycosylation systems might be serving a similar specific function.

7.5 Conclusion and future work

It is now clear that the prevalence of glycosylation systems in bacteria are not coincidental. The role of *N*-linked glycans in dictating protein function and modulating its activity is proved to be a common theme among all kingdoms of life. Future studies to investigate the function and biophysical properties for proteins glycosylated with truncated glycan variants would provide a logical explanation for the role of each monosaccharide in the *C. jejuni N*-linked heptasaccharide. It would also assist in understanding the role of the reducing end glycan commonly found in both *N*- and *O*-linked glycans; diBacNAc. Disruption in glycosylation pathway, notably PglH, led to reduction in bacterial fitness as proteins were still glycosylated with a trisaccharide instead of the heptasaccharide. Systematic investigation of the first three glycans will also provide an archaic view on the emergence of *N*-glycans conservation.

Cryo-EM and 2D NMR would come in great help to evaluate protein conformational structure changes at the glycosylation site and could unravel inter and intra protein- glycan interactions. This thesis sets a platform to be used in investigating bacterial glycosylation systems. It supports the hypothesis that glycans might be playing different roles at different glycosylation sites. Understanding the basic biology of *N*- glycosylation in bacteria could provide insights into the emergence of eukaryotic *N*- glycosylation.

Summary

This thesis provides deeper insights; focusing on the role of general *N*-linked protein glycosylation. It demonstrates that *C. jejuni* general *N*-linked glycans are required to expand proteome repertoire and maintain bacterial fitness. Acting as bulky hydrophilic groups, *C. jejuni N*-linked glycans enhance protein thermostability and modulate protein-protein interaction. Mechanisms by which *C. jejuni N*-linked influence and dictate protein function are yet to be unravelled. The thesis also demonstrates the commonness of *N*-OST among bacterial families as well as their function and substrate specificity. Finally, the role of cytoplasmic *N*-glycosylation in establishing infection was studied, suggesting a common evolutionary mechanism to the analogous *N*- and *O*-glycosylation of bacterial adhesin

REFERENCES

REFERENCES

Apweiler, R., Hermjakob, H. and Sharon, N. (1999) 'On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database', *Biochimica et Biophysica Acta - General Subjects*. doi: 10.1016/S0304-4165(99)00165-8.

Arnold, F. *et al.* (1998) 'Biochemical and immunological analyses of the flagellin of *Clostridium tyrobutyricum* ATCC 25755', *Microbiology and Immunology*. doi: 10.1111/j.1348-0421.1998.tb01965.x.

Avci, F. Y. *et al.* (2011) 'A mechanism for glycoconjugate vaccine activation of the adaptive immune system and its implications for vaccine design', *Nature Medicine*. doi: 10.1038/nm.2535.

Avilés-Reyes, A. *et al.* (2018) 'Characterization of the *pgf* operon involved in the posttranslational modification of *Streptococcus mutans* surface proteins', *Scientific Reports*. doi: 10.1038/s41598-018-23170-3.

Bai, L. *et al.* (2018) 'The atomic structure of a eukaryotic oligosaccharyltransferase complex', *Nature*. doi: 10.1038/nature25755.

Bakalarski, C. E. and Kirkpatrick, D. S. (2016) 'A Biologist's Field Guide to Multiplexed Quantitative Proteomics', *Molecular & Cellular Proteomics*. doi: 10.1074/mcp.O115.056986.

Balonova, L. *et al.* (2010) 'Multimethodological approach to identification of glycoproteins from the proteome of *Francisella tularensis*, an intracellular microorganism', *Journal of Proteome Research*. doi: 10.1021/pr9011602.

Balonova, L. *et al.* (2012) 'Characterization of Protein Glycosylation in *Francisella tularensis* subsp. holarctica: IDENTIFICATION OF A NOVEL GLYCOSYLATED LIPOPROTEIN REQUIRED FOR VIRULENCE', *Molecular & Cellular Proteomics*. doi: 10.1074/mcp.M111.015016.

Barenkamp, S. J. and St Geme 3rd, J. W. (1994) 'Genes encoding high-molecular-weight

adhesion proteins of nontypeable *Haemophilus influenzae* are part of gene clusters', *Infect Immun*.

Benz, I. and Schmidt, M. A. (2001) 'Glycosylation with heptose residues mediated by the *aah* gene product is essential for adherence of the AIDA-I adhesin', *Molecular Microbiology*. doi: 10.1046/j.1365-2958.2001.02487.x.

Bien, J., Palagani, V. and Bozko, P. (2013) 'The intestinal microbiota dysbiosis and *Clostridium difficile* infection: is there a relationship with inflammatory bowel disease', *Therapeutic Advances in Gastroenterology*. doi: 10.1177/1756283X12454590.

Biswas, H. and Chattopadhyaya, R. (2016) 'Stability of *Curcuma longa* rhizome lectin: Role of n-linked glycosylation', *Glycobiology*. doi: 10.1093/glycob/cwv110.

Borud, B. *et al.* (2011) 'Genetic and molecular analyses reveal an evolutionary trajectory for glycan synthesis in a bacterial protein glycosylation system', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1103321108.

Bouché, L. *et al.* (2016) 'The type B flagellin of hypervirulent *Clostridium difficile* is modified with novel sulfonated peptidylamido-glycans', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M116.749481.

Boysen, A. *et al.* (2016) 'A novel mass spectrometric strategy "BEMAP" reveals Extensive O-linked protein glycosylation in Enterotoxigenic *Escherichia coli*.', *Scientific reports*. doi: 10.1038/srep32016.

Cabezas, J. (1994) 'The origins of glycobiology', *Biochemical Education*. doi: 10.1016/0307-4412(94)90126-0.

Caramelo, J. J. and Parodi, A. J. (2008) 'Getting in and out from calnexin/calreticulin cycles', *Journal of Biological Chemistry*. doi: 10.1074/jbc.R700048200.

Carvalho, S. *et al.* (2016) 'Preventing E-cadherin aberrant *N*-glycosylation at Asn-554 improves its critical function in gastric cancer', *Oncogene*. doi: 10.1038/onc.2015.225.

Castric, P., Cassels, F. J. and Carlson, R. W. (2001) 'Structural Characterization of the *Pseudomonas aeruginosa* 1244 Pilin Glycan', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M102685200.

Chamot-Rooke, J. *et al.* (2007) 'Alternative *Neisseria* spp. type IV pilin glycosylation with a glyceramido acetamido trideoxyhexose residue.', *Proceedings of the National Academy of Sciences of the USA*. doi: 10.1073/pnas.0705335104.

Champasa, K. *et al.* (2013) 'Targeted Identification of Glycosylated Proteins in the Gastric Pathogen *Helicobacter pylori* (*Hp*)', *Molecular & Cellular Proteomics*. doi: 10.1074/mcp.M113.029561.

Charbonneau, M.-È. *et al.* (2012) 'A structural motif is the recognition site for a new family of bacterial protein *O*-glycosyltransferases.', *Molecular microbiology*. doi: 10.1111/j.1365-2958.2012.07973.x.

Chen, Y. *et al.* (2016) 'Mechanism of a cytosolic *O*-glycosyltransferase essential for the synthesis of a bacterial adhesion protein.', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1600494113.

Choi, K. J. *et al.* (2010) 'The *Actinobacillus pleuropneumoniae* HMW1C-like glycosyltransferase mediates *N*-Linked glycosylation of the haemophilus influenzae HMW1 adhesin', *PLoS ONE*. doi: 10.1371/journal.pone.0015888.

Cuccui, J. *et al.* (2017) 'The *N*-linking glycosylation system from *Actinobacillus pleuropneumoniae* is required for adhesion and has potential use in glycoengineering', *Open Biology*. doi: 10.1098/rsob.160212.

Cuccui, J. and Wren, B. (2015) 'Hijacking bacterial glycosylation for the production of

glycoconjugates, from vaccines to humanised glycoproteins', *Journal of Pharmacy and Pharmacology*. doi: 10.1111/jphp.12321.

Cuccui, J. and Wren, B. W. (2013) 'Bacteria like sharing their sweets', *Molecular Microbiology*. doi: 10.1111/mmi.12328.

Dankova, V. *et al.* (2016) 'Inactivation of *Francisella tularensis* Gene Encoding Putative ABC Transporter Has a Pleiotropic Effect upon Production of Various Glycoconjugates', *Journal of Proteome Research*. doi: 10.1021/acs.jproteome.5b00864.

Davis, L. M., Kakuda, T. and DiRita, V. J. (2009) 'A *Campylobacter jejuni* znuA orthologue is essential for growth in low-zinc environments and chick colonization', *Journal of Bacteriology*. doi: 10.1128/JB.01394-08.

DiGiandomenico, A. *et al.* (2002) 'Glycosylation of *Pseudomonas aeruginosa* 1244 pilin: Glycan substrate specificity', *Molecular Microbiology*. doi: 10.1046/j.1365-2958.2002.03171.x.

Douzi, B. (2017) Protein–protein interactions: Surface plasmon resonance, *Methods in Molecular Biology*. doi: 10.1007/978-1-4939-7033-9_21.

Egge-Jacobsen, W. *et al.* (2011) 'O-linked glycosylation of the PilA pilin protein of *Francisella tularensis*: Identification of the endogenous protein-targeting oligosaccharyltransferase and characterization of the native oligosaccharide', *Journal of Bacteriology*. doi: 10.1128/JB.00383-11.

Ewing, C. P., Andreishcheva, E. and Guerry, P. (2009) 'Functional characterization of flagellin glycosylation in *Campylobacter jejuni* 81-176', *Journal of Bacteriology*. doi: 10.1128/JB.00378-09.

Faridmoayer, A. *et al.* (2007) 'Functional characterization of bacterial oligosaccharyltransferases involved in *O*-linked protein glycosylation', *Journal of Bacteriology*. doi: 10.1128/JB.01318-07.

Faulds-Pain, A. *et al.* (2014) 'The post-translational modification of the *Clostridium difficile*

flagellin affects motility, cell surface properties and virulence', *Molecular Microbiology*. doi: 10.1111/mmi.12755.

Feldman, M. F. *et al.* (2005) 'Engineering *N*-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli'*, *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0500044102.

Fletcher, C. M. *et al.* (2009) 'A General *O*-Glycosylation System Important to the Physiology of a Major Human Intestinal Symbiont', *Cell*. doi: 10.1016/j.cell.2009.02.041.

Freeze, H., Esko, J. and Parodi, A. J. (2009) 'Glycans in Glycoprotein Quality Control', *Essentials of Glycobiology*. doi: NBK1958 [bookaccession].

Fusetti, F. *et al.* (2002) 'Crystal Structure of the Copper-Containing Quercetin 2,3-Dioxygenase from *Aspergillus japonicus'*, *Structure*. doi: 10.1016/S0969-2126(02)00704-9.

Garcia-Quintanilla, F. *et al.* (2014) 'Production of a recombinant vaccine candidate against *Burkholderia pseudomallei* exploiting the bacterial N-glycosylation machinery', *Frontiers in Microbiology*. doi: 10.3389/fmicb.2014.00381.

Gastric, P. (1995) 'PilO, a gene required for glycosylation of *Pseudomonas aeruginosa* 1244 pilin', *Microbiology*. doi: 10.1099/13500872-141-5-1247.

St. Geme, J. W. (1994) 'The HMW1 adhesin of nontypeable *Haemophilus influenzae* recognizes sialylated glycoprotein receptors on cultured human epithelial cells', *Infection and Immunity*.

Gottschalk, A. (1958) 'The influenza virus neuraminidase', *Nature*. doi: 10.1038/181377a0.

Grass, S. *et al.* (2010) 'The *Haemophilus influenzae* HMW1c protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin', *PLoS Pathogens*. doi: 10.1371/journal.ppat.1000919.

Gross, J. *et al.* (2008) 'The *Haemophilus influenzae* HMW1 adhesin is a glycoprotein with an

unusual *N*-linked carbohydrate modification', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M801819200.

Hanson, S. R. *et al.* (2009) 'The core trisaccharide of an *N*-linked glycoprotein intrinsically accelerates folding and enhances stability.', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0810318105.

Hartley, M. D. *et al.* (2011) 'Biochemical characterization of the *O*-linked glycosylation pathway in *Neisseria gonorrhoeae* responsible for biosynthesis of protein glycans containing N, N′-diacetylbacillosamine', *Biochemistry*. doi: 10.1021/bi2003372.

Hebert, D. N. *et al.* (2014) 'The intrinsic and extrinsic effects of *N*-linked glycans on glycoproteostasis', *Nature Chemical Biology*. doi: 10.1038/nchembio.1651.

Hendrixson, D. R. and DiRita, V. J. (2004) 'Identification of *Campylobacter jejuni* genes involved in commensal colonization of the chick gastrointestinal tract', *Molecular Microbiology*. doi: 10.1111/j.1365-2958.2004.03988.x.

Horzempa, J. *et al.* (2006) '*Pseudomonas aeruginosa* 1244 pilin glycosylation: Glycan substrate recognition', *Journal of Bacteriology*. doi: 10.1128/JB.00273-06.

Isaji, T. *et al.* (2006) '*N*-glycosylation of the propeller domain of the integrin 5 subunit is essential for α5β1 heterodimerization, expression on the cell surface, and its biological function', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M607771200.

Isaji, T. *et al.* (2009) '*N*-glycosylation of the I-like domain of β1 integrin is essential for β1 integrin expression and biological function. Identification of the minimal *N*-glycosylation requirement for α5β1', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M807920200.

Iwashkiw, J. A. *et al.* (2012) 'Identification of a general *O*-linked protein glycosylation system in *Acinetobacter baumannii* and its role in virulence and biofilm formation', *PLoS Pathogens*. doi: 10.1371/journal.ppat.1002758.

Jayaprakash, N. G. and Surolia, A. (2017) 'Role of glycosylation in nucleating protein folding

and stability', *Biochemical Journal*. doi: 10.1042/BCJ20170111.

Jeong, H. *et al.* (2016) 'Pseudoatomic Structure of the Tripartite Multidrug Efflux Pump AcrAB-TolC Reveals the Intermeshing Cogwheel-like Interaction between AcrA and TolC', *Structure*. doi: 10.1016/j.str.2015.12.007.

Jervis, A. J. *et al.* (2010) 'Characterization of *N*-linked protein glycosylation in *Helicobacter pullorum'*, *Journal of Bacteriology*. doi: 10.1128/JB.00211-10.

Jervis, A. J. *et al.* (2012) 'Characterization of the structurally diverse *N*-linked glycans of *Campylobacter* species', *Journal of Bacteriology*. doi: 10.1128/JB.00042-12.

Jervis, A. J. *et al.* (2018) 'Functional analysis of the *Helicobacter pullorum N*-linked protein glycosylation system', *Glycobiology*. doi: 10.1093/glycob/cwx110.

Jiangm, Y. L. *et al.* (2017) 'Defining the enzymatic pathway for polymorphic *O*-glycosylation of the pneumococcal serine-rich repeat protein PsrP', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M116.770446.

Jones, M. A. *et al.* (2004) 'Adaptation of *Campylobacter jejuni* NCTC11168 to high-level colonization of the avian gastrointestinal tract', *Infection and Immunity*. doi: 10.1128/IAI.72.7.3769-3776.2004.

Karlyshev, A. V. (2004) 'The *Campylobacter jejuni* general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks', *Microbiology*, 150(6), pp. 1957–1964. doi: 10.1099/mic.0.26721-0.

Kelly, S. M., Jess, T. J. and Price, N. C. (2005) 'How to study proteins by circular dichroism', *Biochimica et Biophysica Acta - Proteins and Proteomics*. doi: 10.1016/j.bbapap.2005.06.005.

Kirov, S. M., Castrisios, M. and Shaw, J. G. (2004) '*Aeromonas Flagella* (Polar and Lateral) Are Enterocyte Adhesins That Contribute to Biofilm Formation on Surfaces', *Infection and Immunity*. doi: 10.1128/IAI.72.4.1939-1945.2004.

Kito, K. and Ito, T. (2008) 'Mass Spectrometry-Based Approaches Toward Absolute Quantitative Proteomics', *Current Genomics*. doi: 10.2174/138920208784533647.

Ku, S. C. *et al.* (2009) 'The pilin *O*-glycosylation pathway of pathogenic *Neisseria* is a general system that glycosylates AniA, an outer membrane nitrite reductase', *Biochemical and Biophysical Research Communications*. doi: 10.1016/j.bbrc.2008.11.025.

Kupfer, L., Hinrichs, W. and Groschup, M. H. (2009) 'Prion protein misfolding.', *Current molecular medicine*. doi: 10.2174/156652409789105543.

Kus, J. V. *et al.* (2004) 'Significant differences in type IV pilin allele distribution among Pseudomonas aeruginosa isolates from cystic fibrosis (CF) versus non-CF patients', *Microbiology*. doi: 10.1099/mic.0.26822-0.

Kus, J. V. *et al.* (2008) 'Modification of *Pseudomonas aeruginosa* Pa5196 type IV pilins at multiple sites with D-Araf by a novel GT-C family arabinosyltransferase, TfpW', *Journal of Bacteriology*. doi: 10.1128/JB.01075-08.

Larsen, J. C., Szymanski, C. and Guerry, P. (2004) '*N*-linked protein glycosylation is required for full competence in *Campylobacter jejuni* 81-176', *Journal of Bacteriology*. doi: 10.1128/JB.186.19.6508-6514.2004.

Lauc, G., Krištić, J. and Zoldoš, V. (2014) 'Glycans - the third revolution in evolution', *Frontiers in Genetics*. doi: 10.3389/fgene.2014.00145.

Lee, H. S., Qi, Y. and Im, W. (2015) 'Effects of *N*-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study', *Scientific Reports*. doi: 10.1038/srep08926.

Lee, I. C. *et al.* (2014) 'GtfA and GtfB are both required for protein *O*-glycosylation in *Lactobacillus plantarum'*, *Journal of Bacteriology*. doi: 10.1128/JB.01401-13.

Lees-Miller, R. G. *et al.* (2013) 'A common pathway for *O*-linked protein-glycosylation and synthesis of capsule in Acinetobacter baumannii', *Molecular Microbiology*. doi:

10.1111/mmi.12300.

Li, Z. *et al.* (2012) 'Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ orbitrap velos', in *Journal of Proteome Research*. doi: 10.1021/pr200748h.

Lin, J., Overbye Michel, L. and Zhang, Q. (2002) 'CmeABC functions as a multidrug efflux system in *Campylobacter jejuni*', *Antimicrobial Agents and Chemotherapy*. doi: 10.1128/AAC.46.7.2124-2131.2002.

Lindenthal, C. and Elsinghorst, E. A. (1999) 'Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*', *Infection and Immunity*.

Linton, D. *et al.* (2002) 'Identification of *N*-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*', *Molecular Microbiology*. doi: 10.1046/j.1365-2958.2002.02762.x.

Linton, D. *et al.* (2005) 'Functional analysis of the *Campylobacter jejuni N*-linked protein glycosylation pathway', *Molecular Microbiology*. doi: 10.1111/j.1365-2958.2005.04519.x.

Lithgow, K. V. *et al.* (2014) 'A general protein *O*-glycosylation system within the *Burkholderia cepacia* complex is involved in motility and virulence', *Molecular Microbiology*. doi: 10.1111/mmi.12540.

Lizcano, A., Sanchez, C. J. and Orihuela, C. J. (2012) 'A role for glycosylated serine-rich repeat proteins in Gram-positive bacterial pathogenesis', *Molecular Oral Microbiology*. doi: 10.1111/j.2041-1014.2012.00653.x.

Lu, Q. *et al.* (2014) 'An iron-containing dodecameric heptosyltransferase family modifies bacterial autotransporters in pathogenesis', *Cell Host and Microbe*. doi: 10.1016/j.chom.2014.08.008.

Ma, J. and Lindquist, S. (1999) 'De novo generation of a PrPSc-like conformation in living cells', *Nature Cell Biology*. doi: 10.1038/14053.

Mahenthiralingam, E., Urban, T. A. and Goldberg, J. B. (2005) 'The multifarious, multireplicon *Burkholderia cepacia* complex.', *Nature Reviews Microbiology*. doi: 10.1038/nrmicro1085.

Marceau, M. *et al.* (1998) 'Consequences of the loss of *O*-linked glycosylation of meningococcal type IV pilin on piliation and pilus-mediated adhesion', *Molecular Microbiology*. doi: 10.1046/j.1365-2958.1998.00706.x.

Merino, S. *et al.* (2014) '*Aeromonas hydrophila* flagella glycosylation: Involvement of a lipid carrier', *PLoS ONE*. doi: 10.1371/journal.pone.0089630.

Merino, S., Wilhelms, M. and Tomás, J. M. (2014) 'Role of *Aeromonas hydrophila* flagella glycosylation in adhesion to hep-2 cells, biofilm formation and immune stimulation', *International Journal of Molecular Sciences*. doi: 10.3390/ijms151221935.

Miller, S. I. and Salama, N. R. (2018) 'The gram-negative bacterial periplasm: Size matters', *PLoS Biology*. doi: 10.1371/journal.pbio.2004935.

Mills, D. C. *et al.* (2016) 'Functional analysis of *N*-linking oligosaccharyl transferase enzymes encoded by deep-sea vent proteobacteria', *Glycobiology*, 26(4). doi: 10.1093/glycob/cwv111.

Mitra, N. *et al.* (2006) '*N*-linked oligosaccharides as outfitters for glycoprotein folding, form and function', *Trends in Biochemical Sciences*. doi: 10.1016/j.tibs.2006.01.003.

Moormann, C., Benz, I. and Schmidt, M. A. (2002) 'Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesin heptosyltransferase of the AIDA system', *Infection and Immunity*. doi: 10.1128/IAI.70.5.2264-2270.2002.

Neuberger, a (1938) 'Carbohydrates in protein: The carbohydrate component of crystalline egg albumin.', *The Biochemical journal*. doi: 10.1042/bj0321435.

Nothaft, H. *et al.* (2012) 'Diversity in the protein *N*-glycosylation pathways within the Campylobacter genus.', *Molecular & cellular proteomics : MCP*. doi: 10.1074/mcp.M112.021519.

Nothaft, H. and Szymanski, C. M. (2010) 'Protein glycosylation in bacteria: Sweeter than ever', *Nature Reviews Microbiology*. doi: 10.1038/nrmicro2383.

Nothaft, H. and Szymanski, C. M. (2013) 'Bacterial protein *N*-glycosylation: New perspectives and applications', *Journal of Biological Chemistry*. doi: 10.1074/jbc.R112.417857.

Okuda, S. and Weinbaum, G. (1968) 'An Envelope-Specific Glycoprotein from *Escherichia coli* B', *Biochemistry*. doi: 10.1021/bi00848a018.

Olivier, N. B. *et al.* (2006) 'In vitro biosynthesis of UDP-*N,N*-diacetylbacillosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system', *Biochemistry*. doi: 10.1021/bi061456h.

Ollis, A. A. *et al.* (2015) 'Substitute sweeteners: Diverse bacterial oligosaccharyltransferases with unique *N*-glycosylation site preferences', *Scientific Reports*. doi: 10.1038/srep15237.

Oyston, P. C. F., Sjostedt, A. and Titball, R. W. (2004) 'Tularaemia: Bioterrorism defence renews interest in *Francisella tularensis'*, *Nature Reviews Microbiology*. doi: 10.1038/nrmicro1045.

Pittman, M. S. *et al.* (2007) 'Growth of *Campylobacter jejuni* on nitrate and nitrite: Electron transport to NapA and NrfA via NrfH and distinct roles for NrfA and the globin Cgb in protection against nitrosative stress', *Molecular Microbiology*. doi: 10.1111/j.1365-2958.2006.05532.x.

Pohlner, J. *et al.* (1987) 'Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease', *Nature*. doi: 10.1038/325458a0.

Posch, G. *et al.* (2011) 'Characterization and scope of S-layer protein *O*-glycosylation in *Tannerella forsythia'*, *Journal of Biological Chemistry*. doi: 10.1074/jbc.M111.284893.

Pumbwe, L. and Piddock, L. J. (2002) 'Identification and molecular characterisation of CmeB, a *Campylobacter jejuni* multidrug efflux pump', *FEMS Microbiology Letters*. doi: 10.1111/j.1574-6968.2002.tb11007.x.

Rangarajan, E. S. *et al.* (2007) 'Structural context for protein *N*-glycosylation in bacteria: The structure of PEB3, an adhesin from *Campylobacter jejuni*.', *Protein science : a publication of the Protein Society*. doi: 10.1110/ps.062737507.

Rauniyar, N. and Yates, J. R. (2014) 'Isobaric labeling-based relative quantification in shotgun proteomics', *Journal of Proteome Research*. doi: 10.1021/pr500880b.

Rempe, K. A. *et al.* (2015) 'Unconventional *N*-linked glycosylation promotes trimeric autotransporter function in *Kingella kingae* and *Aggregatibacter aphrophilus'*, *mBio*. doi: 10.1128/mBio.01206-15.

Riegert, A. S. *et al.* (2017) 'Structural and Biochemical Investigation of PglF from *Campylobacter jejuni* Reveals a New Mechanism for a Member of the Short Chain Dehydrogenase/Reductase Superfamily', *Biochemistry*. doi: 10.1021/acs.biochem.7b00910.

Rippa, A. L. *et al.* (2013) 'The role of integrins in the development and homeostasis of the epidermis and skin appendages', *Acta Naturae*.

Rudd, P. M. *et al.* (2002) 'Glycosylation and prion protein', *Current Opinion in Structural Biology*. doi: 10.1016/S0959-440X(02)00377-9.

Ruiz-Canada, C., Kelleher, D. J. and Gilmore, R. (2009) 'Cotranslational and Posttranslational *N*-Glycosylation of Polypeptides by Distinct Mammalian OST Isoforms', *Cell*. doi: 10.1016/j.cell.2008.11.047.

Sára, M. and Sleytr, U. B. (2000) 'S-layer proteins', *Journal of Bacteriology*. doi: 10.1128/JB.182.4.859-868.2000.

Schäffer, C. and Messner, P. (2017) 'Emerging facets of prokaryotic glycosylation', *FEMS Microbiology Reviews*. doi: 10.1093/femsre/fuw036.

Schoenhofen, I. C. *et al.* (2006) 'Functional characterization of dehydratase/aminotransferase pairs from *Helicobacter* and *Campylobacter*: Enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways', *Journal of Biological Chemistry*. doi:

10.1074/jbc.M511021200.

Schwarz, F. *et al.* (2011) 'Cytoplasmic *N*-glycosyltransferase of *Actinobacillus pleuropneumoniae* is an inverting enzyme and recognizes the NX(S/T) consensus sequence', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M111.277160.

Schwarz, F. and Aebi, M. (2011) 'Mechanisms and principles of *N*-linked protein glycosylation', *Current Opinion in Structural Biology*. doi: 10.1016/j.sbi.2011.08.005.

Scott, N. E. *et al.* (2009) 'Mass spectrometric characterization of the surface-associated 42 kDa lipoprotein JlpA as a glycosylated antigen in strains of *Campylobacter jejuni'*, *Journal of Proteome Research*. doi: 10.1021/pr900544x.

Scott, N. E. *et al.* (2010) 'Simultaneous glycan-peptide characterization using hydrophilic interaction 2 chromatography and parallel fragmentation by CID, HCD and ETD-MS applied to the N- 3 linked glycoproteome of *Campylobacter jejuni'*, *The American Society for Biochemistry and Molecular Biology*. doi: M000031-MCP201 [pii]\n10.1074/mcp.M000031-MCP201.

Scott, N. E. *et al.* (2014) 'Diversity Within the *O*-linked Protein Glycosylation Systems of Acinetobacter Species.', *Molecular & Cellular Proteomics*. doi: 10.1074/mcp.M114.038315.

Settem, R. P. *et al.* (2013) 'A bacterial glycan core linked to surface (S)-layer proteins modulates host immunity through Th17 suppression', *Mucosal Immunology*. doi: 10.1038/mi.2012.85.

Shental-Bechor, D. and Levy, Y. (2008) 'Effect of glycosylation on protein folding: A close look at thermodynamic stabilization', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0801340105.

Sherlock, O. *et al.* (2006) 'Glycosylation of the self-recognizing *Escherichia coli* Ag43 autotransporter protein', *Journal of Bacteriology*. doi: 10.1128/JB.188.5.1798-1807.2006.

Silva, J. C. (2005) 'Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel ms

Acquisition', *Molecular & Cellular Proteomics*. doi: 10.1074/mcp.M500230-MCP200.

Simmons, D. A., Luderitz, O. and Westphal, O. (1965) 'The immunochemistry of *Salmonella* chemotype VI O-antigens. The structure of oligosaccharides from Salmonella group G (o 13,22) lipopolysaccharides', *Biochem J*. doi: 10.1042/bj0970807.

Sleytr, U. B. and Thorne, K. J. I. (1976) 'Chemical characterization of the regularly arranged surface layers of *Clostridium thermosaccharolyticum* and *Clostridium thermohydrosulfuricum*', *Journal of Bacteriology*.

Smedley, J. G. *et al.* (2005) 'Influence of pilin glycosylation on *Pseudomonas aeruginosa* 1244 pilus function', *Infection and Immunity*. doi: 10.1128/IAI.73.12.7922-7931.2005.

van Sorge, N. M. *et al.* (2009) '*N*-glycosylated proteins and distinct lipooligosaccharide glycoforms of *Campylobacter jejuni* target the human C-type lectin receptor MGL', *Cellular Microbiology*. doi: 10.1111/j.1462-5822.2009.01370.x.

St Geme, J. W. (2002) 'Molecular and cellular determinants of non-typeable *Haemophilus influenzae* adherence and invasion', *Cellular Microbiology*. doi: 10.1046/j.1462-5822.2002.00180.x.

St Geme, J. W., Falkow, S. and Barenkamp, S. J. (1993) 'High-molecular-weight proteins of nontypable *Haemophilus influenzae* mediate attachment to human epithelial cells.', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.90.7.2875.

Stimson, E. *et al.* (1995) 'Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose', *Molecular Microbiology*. doi: 10.1111/j.1365-2958.1995.mmi_17061201.x.

Su, C. C. *et al.* (2014) 'Crystal structure of the *Campylobacter jejuni* CmeC outer membrane channel', *Protein Science*. doi: 10.1002/pro.2478.

Szymanski, C. M. *et al.* (1999) 'Evidence for a system of general protein glycosylation in

Campylobacter jejuni', *Molecular Microbiology*. doi: 10.1046/j.1365-2958.1999.01415.x.

Szymanski, C. M. *et al.* (2003) '*Campylobacter* – a tale of two protein glycosylation systems', *Trends in Microbiology*, 11(5), pp. 233–238. doi: 10.1016/S0966-842X(03)00079-9.

Szymanski, C. M., Burr, D. H. and Guerry, P. (2002) '*Campylobacter* protein glycosylation affects host cell interactions', *Infection and Immunity*. doi: 10.1128/IAI.70.4.2242-2244.2002.

Takai, K. *et al.* (2003) '*Deferribacter desulfuricans* sp. nov., a novel sulfur-, nitrate- and arsenate-reducing thermophile isolated from a deep-sea hydrothermal vent', *International Journal of Systematic and Evolutionary Microbiology*. doi: 10.1099/ijs.0.02479-0.

Terra, V. S. *et al.* (2012) 'Recent developments in bacterial protein glycan coupling technology and glycoconjugate vaccine design', *Journal of Medical Microbiology*, 61(PART7). doi: 10.1099/jmm.0.039438-0.

Thomas, R. M. *et al.* (2011) 'Glycosylation of DsbA in *Francisella tularensis* subsp. *Tularensis*', *Journal of Bacteriology*. doi: 10.1128/JB.00438-11.

Tikhonova, E. B. *et al.* (2009) 'Kinetic control of TolC recruitment by multidrug efflux complexes', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0906601106.

Troutman, J. M. and Imperiali, B. (2009) '*Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions', *Biochemistry*. doi: 10.1021/bi802284d.

Twine, S. M. *et al.* (2008) 'Flagellar glycosylation in *Clostridium botulinum*', *FEBS Journal*. doi: 10.1111/j.1742-4658.2008.06589.x.

Twine, S. M. *et al.* (2009) 'Motility and flagellar glycosylation in *Clostridium difficile'*, *Journal of Bacteriology*. doi: 10.1128/JB.00861-09.

Valderrama-Rincon, J. D. *et al.* (2012) 'An engineered eukaryotic protein glycosylation

pathway in *Escherichia coli'*, *Nature Chemical Biology*. doi: 10.1038/nchembio.921.

Valiente, E. *et al.* (2016) 'Role of glycosyltransferases modifying type B flagellin of emerging hypervirulent *Clostridium difficile* lineages and their impact on motility and biofilm formation', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M116.749523.

Varki, A. (2017) 'Biological roles of glycans', *Glycobiology*. doi: 10.1093/glycob/cww086.

Varki, A. and Sharon, N. (2016) 'Chapter 1 Historical Background and Overview', *Essentials of Glycobiology*. doi: 10.1101/glycobiology.3e.001.

Vik, A. *et al.* (2009) 'Broad spectrum *O*-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae'*, *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0809504106.

Wacker, M. *et al.* (2002) '*N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into E. coli', *Science*. doi: 10.1126/science.298.5599.1790.

Walter, J. *et al.* (2008) 'Glucosyltransferase A (GtfA) and inulosucrase (Inu) of *Lactobacillus reuteri* TMW1.106 contribute to cell aggregation, in vitro biofilm formation, and colonization of the mouse gastrointestinal tract', *Microbiology*. doi: 10.1099/mic.0.2007/010637-0.

Watkins, W. M. (1966) 'Blood-Group Substances', *Science*, 152(3719), pp. 172–181. doi: 10.1126/science.152.3719.172.

Weerapana, E. *et al.* (2005) 'Investigating bacterial *N*-linked glycosylation: Synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine', *Journal of the American Chemical Society*. doi: 10.1021/ja054265v.

Wild, R. *et al.* (2018) 'Structure of the yeast oligosaccharyltransferase complex gives insight

into eukaryotic N-glycosylation', *Science*. doi: 10.1126/science.aar5140.

Wilhelms, M. *et al.* (2012) 'Differential glycosylation of polar and lateral flagellins *in Aeromonas hydrophila* ah-3', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M112.376525.

Wyss, D. F. *et al.* (1995) 'Conformation and function of the *N*-linked glycan in the adhesion domain of human CD2.', *Science (New York, N.Y.)*. doi: 10.1126/science.7544493.

Xin, F. and Radivojac, P. (2012) 'Post-translational modifications induce significant yet not extreme changes to protein structure', *Bioinformatics*. doi: 10.1093/bioinformatics/bts541.

Zhou, M. *et al.* (2010) 'A novel glucosyltransferase is required for glycosylation of a serine-rich adhesin and biofilm formation by *Streptococcus parasanguinis'*, *Journal of Biological Chemistry*. doi: 10.1074/jbc.M109.066928.

Zhu, F. *et al.* (2016) 'Engineering and dissecting the glycosylation pathway of a streptococcal serine-rich repeat adhesin', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M116.752998.

van Zuylen, C. W., Kamerling, J. P. and Vliegenthart, J. F. (1997) 'Glycosylation beyond the Asn78-linked GlcNAc residue has a significant enhancing effect on the stability of the alpha subunit of human chorionic gonadotropin', *Biochem Biophys Res Commun*. doi: S0006-291X(97)96241-0 [pii]\r10.1006/bbrc.1997.6241.

APPENDICES

Other publication contributions

Production and efficacy of a low-cost recombinant pneumococcal protein polysaccharide conjugate vaccine.

Herbert JA, Kay EJ, Faustini SE, Richter A, **Abouelhadid S**, Cuccui J, Wren B, Mitchell TJ.

Vaccine. 2018 Jun 18;36(26):3809-3819. doi: 10.1016/j.vaccine.2018.05.036. Epub 2018 May 25.

Pancreatic amylase is an environmental signal for regulation of biofilm formation and host interaction in Campylobacter jejuni.

Jowiya W, Brunner K, **Abouelhadid S**, Hussain HA, Nair SP, Sadiq S, Williams LK, Trantham EK, Stephenson H, Wren BW, Bajaj-Elliott M, Cogan TA, Laws AP, Wade J, Dorrell N, Allan E.

Infect Immun. 2015 Dec;83(12):4884-95. doi: 10.1128/IAI.01064-15. Epub 2015 Oct 5

Bacterial Strain/Plasmid	Description	Reference
***C. jejuni* 11168H**	**Hypermotile variant of *C. jejuni* 11168**	**Jones, M. A. et al. (2004)**
C. jejuni* 11168H *cmeB::aphA	***C. jejuni* 11168H cmeB is inactivated by aphA encoding kanamycin resistance cassette**	**This study**
C. jejuni 11168H cmeB::aphA	**C. jejuni 11168H cmeB is inactivated by aphA encoding kanamycin resistance cassette**	**This study**
C. jejuni* 11168H *napA::aphA	***C. jejuni* 11168H napA is inactivated by aphA encoding kanamycin resistance cassette**	**This study**
C. jejuni cmeD*::*cat* g0*cmeABC	***C. jejuni* 11168H *cmeD*::*cat, cmeABC* is glycosylation deficient by altering N->Q in *C. jejuni* glycosylation sequon (D/E-X_1-N-X_2-S/T where X_1 and X_2 are any amino acid other than proline)**	**This study**

C. jejuni cmeD*::*cat* wt*cmeABC	***C. jejuni* 11168H *cmeD*::*cat, cmeC* is 6xhis tagged followed by kanamycin cassette to help for selection of CBA plate**	**This study**
E. coli dam-/dcm-	***ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10) TetS endA1 rspL136 (StrR) dam13::Tn9 (CamR) xylA-5 mtl-1 thi-1 mcrB1 hsdR2***	**New England Biolabs, UK**
***E. coli* DH10B**	***F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara,leu)7697 araD139 galU galK nupG rpsL λ-***	**New England Biolabs, UK**
pACYC(*pgl*)	***C. jejuni* heptasaccharide coding sequence under Tet promoter in pCAYC184**	**Feldman, M. F. et al. (2005)**
pATKH	***cj0364* cloned in SacII site in pATMN**	**This study**
pATM	**g0*cmeABC* locus cloned in pJET1.2**	**This study**
pATMN	***aphA* cloned in BamHI site in pATM**	**This study**
pATN	***cmeD*::*cat* cloned in pJET1.2**	**This study**
pATT3	***napA* cloned in pJET1.2**	**This study**
pATT3F	***aphA* cloned in ClaI site inactivate *napA* in pATT3**	**This study**
pCMECSDBA	**Membrane bound 10xHis tagged CmeC driven to periplasm by DsbA signal peptide under L-arabinose promoter in pEC145**	**This study**
pGVXN114	**PglB cloned in pEXT21 under *lac* promoter**	**Ihssen *et al.*, 2010**
pJMK30	***aphA* gene cloned in BamHI restriction site**	**Van Vliet, Wooldridge and Ketley, 1998**
pMH3	***cmeABC* locus cloned in pJET1.2**	**This study**

pMH5	**Soluble periplasmic 6xHis tagged CmeA under Tet promoter in pCAYC184**	**Feldman, M. F. et al. (2005)**
pMHT	***aphA* cloned in BamHI site in pMH3**	**This study**
pMHTF	***cj0364* cloned in SacII site in pMHT**	**This study**
pRSA	***cmeB* amplicon cloned in pJET1.2**	**This study**
pRSF	***aphA* gene cloned in BamHI to inactivate *cmeB* in pRSA**	**This study**
pWA2	**Soluble periplasmic 6xHis tagged CmeA under Tet promoter in pBR322**	**Feldman, M. F. et al. (2005)**
Primer	**DNA sequence**	
FWDcmeB	**GACGTAATGAAGGAGAGCCA**	
REVcmeB	**CTGATCCACTCCAAGCTATG**	
FWDCmeA	**AGCGAAGTTAAAGAAATTGGAGCAC**	
REVCmeC His *C.jejuni*	**TTTT*CCGCGG*ATT*GGATCC*CATTATGATGATG ATGATGATGATGTTCTCTAAAGACATATCT**	
FWDcj0364	**TTTT*CCGCGG*ATTCTCTAAATAAATTAAAAATC TTTGTCT**	
REVcj0364	**TTTT*CCGCGG*CATTGAACCTTTTTGGAGGGATT TTTCC**	
FWDCmeC	**TTTT*GCTAGC*GCCGCCCCAAATTTAAATATTCC CGAAGCAAACTATAGCATTG**	
REVCmeC	**TTTTTG*TCGAC*ctaATGATGATGATGATGATGAT GATGATGATG TTCTCTAAAAGACATATCTAAATTTTTTGATTC**	
FWDnapA	**ACCGCTATTGCAAGTGCTGCTAG**	
REVnapA	**GAAAGCGGACAAGTCGCATCC**	