University of Nottingham
UK | CHINA | MALAYSIA

The interaction of glycosylated Major Outer Membrane Protein of *Campylobacter jejuni* with natural and synthetic ligands

Jennifer Chigozie Okoye

Supervisors:

Professor Panos Soultanas

Dr. Neil Oldfield

A thesis submitted to The University of Nottingham for the award of Doctor of Philosophy

May 2021

Declaration

I hereby declare that the work presented in this thesis has not been submitted for any other degree or professional qualification, and that it is the result of my own independent work.

Signed: 

Date: 17/05/2021

Abstract

According to the WHO fluoroquinolone- resistant *Campylobacter jejuni (C. jejuni)* is number six on the list of antibiotic-resistant "priority pathogens" – a list of 12 families of bacteria that pose the greatest threat to human health[1]*. C. jejuni* is one of the four key global causes of gastrointestinal diseases. Currently, the cost of illness is $1.3billion in the US and €2.4billion in the EU27[2]. There are a large variety of host sources of *C. jejuni*, but poultry is thought be the major risk factor, as it is the main animal reservoir[3]. Sustainable and innovative solutions to combat emerging infectious disease through alternative interventions must be encouraged. For the protection of the public, control of zoonosis (a disease which can be transmitted from animals to humans) is important. Processes should include avoiding bacterial colonisation of poultry, hygienic processing of meat and the protection and control of private drinking water supplies. The overarching goal is to reduce the load of *C. jejuni* in farm poultry to a level with low probability of causing illness to humans.

The aim of this project was to study the *C. jejuni* Major Outer Membrane Protein (MOMP) and its ability to bind to carbohydrate based blood group antigens (BgAgs) and novel compounds in its glycosylated and non-glycosylated forms. Novel compounds include a class of iron (III) compounds. Characterising the porin functionality of MOMP (the role it plays in bacteria and the size of molecules that can be exchanged through this protein) is important to optimise the interactions between glycosylated MOMP and synthetic compounds which inhibit bacterial attachment to surfaces and biofilm formation.

We cloned the wild type MOMP encoding gene from *C. jejuni* NCTC11168 into *Escherichia coli* (*E. coli)* to create a novel recombinant protein for purification and subsequent analysis. However, difficulties with MOMP purification from the heterologous host, led to native MOMP being purified directly from *C. jejuni*. Studies of the *C. jejuni* 11168 derived wtMOMP showed that the glycan consists of at least 3 monosaccharides with a close similarity to *N-galactosamine-galactose-ribose.* The EC_{50} of the binding of Le^b with wtMOMP was measured to be 23 nM, conferring to high potency of the BgAgs for the glycosylated protein. Iron (III) compound - QPLEX, which has previously been shown to cause reduction in colonisation of *C. jejuni* in poultry, favoured binding to glycosylated MOMP over non-glycosylated MOMP *in vitro* thus inhibiting association of MOMP to Le^b*.* We showed that QPLEX (and to some extent Le^b) causes a conformational change of the wtMOMP protein in the transverse region of the protein from the extracellular region extending to the underside of the porin, a conformational change which was further confirmed in protein ligand

docking. Binding of QPLEX to the loop region of the porin and within the constriction zone of the porin was highly favourable with CHEMGAUSS scores of -8.5 and -8, respectfully. Finally, localisation of specific binding between wtMOMP and human intestinal gut tissue sections was demonstrated *ex vivo* with binding occurring especially on the corpus. This was further confirmed with observation of specific binding of wtMOMP on transgenic mouse stomach tissue positive for Le^b.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Panos Soultanas, for allowing me the opportunity to undertake this research project and for trusting me to direct its progression. His unwavering support, encouragement and belief in my abilities have allowed me to reach my true potential during the course of the research (even during a global pandemic!), and his students are lucky to be able to work within his group. I now have an armada of scientific skills and knowledge that I will take with me for life.

I would also like to thank my co-supervisor, Dr. Neil Oldfield, who welcomed me into his lab during my first year, for the microbial aspect of my research. He has been a constant source of support and guidance during my studies and his patience has been second to none. Further, I want to thank Dr. Jafar Mahdavi, who's extraordinary intellect and scientific knowledge in the area of *Campylobacter* biology and beyond pushed me further every day. In my first few weeks, Dr. Luana Ferrara - of St. Andrews University - kindly passed on her knowledge and expertise of bacterial outer membrane purification to me. Many thanks to Dr. Jedd Bellamy-Carter and Prof. Neil Oldham, whose expertise in the applications of protein mass spectrometry were so invaluable to my project. Many thanks to the School of Life Sciences Imaging (SLIM) team who helped with acquisition of slide images by microscopy during strict social distancing restrictions. A special thanks to my industrial contacts, Akeso Biomedical and Dr. Simon Williams (CEO), for partly funding this research, without whom this work would not have been possible. This research was also partly funded by the EPSRC Centre for Doctoral Training (CDT) in Sustainable Chemistry, whose manager and director, Perislava Williams and Prof. Pete Licence, were the driving force in my pursuit of lifelong skills obtained outside the lab, such as scientific outreach to the general public. The CDT and School of Chemistry has given me many opportunities to present my research which has ultimately made me a better scientist and communicator.

I want to extend my gratitude to all past and present members of lab A20 of the Biodiscovery Institute. Namely, Dr. Paschalis Vasilis, Dr. Osman Dufailu, Dr. Faadil Fawzy, Dr. Matthaios Pitoulias, Dr. Eleyna Martin, Dr. Alex Holland, Lenny Ferreira, Parisa Hamzavinejad, Rowan Earlam, Hannah Betts, Andrew Raslan, Isobel Holden and Jin Pan. Their presence in the lab was much appreciated, as well as the important opportunities they provided to socialise in more relaxed settings.

Last but certainly not least, I want to thank my friends and family. To name every friend that has made a positive impact during the last four years I would need to write a new chapter in this thesis, but they know who they are and I truly appreciate each and every one of them! My mother and father, Njideka and Emeka Okoye, have been an endless stream of love, support, care and advice and I truly could not have done this without them. My younger siblings, Emeka and Nneamaka, are wise beyond their years, and I learn so much from them every day. My maternal grandma peacefully passed away in the week leading up to the submission of my thesis, so I dedicate this work to her, and only hope I can emulate her light and strength in my future endeavours.

Table of contents

Declaration ..... **i**

Abstract ..... **ii**

Acknowledgements ..... **iv**

Table of contents ..... **vi**

List of figures ..... **xi**

List of tables ..... **xvi**

Chapter 1: Introduction **1**

1.1 Aims and objectives 1

Chapter 2: Literature review **2**

Introduction 2

2.1 Campylobacter in humans 2

2.2 Incidence of disease 3

2.3 Cost burden 4

2.4 Risk factors/ Sources and routes of transmission for human infection 4

2.5 Human vs chicken response to infection/colonisation 5

2.6 Campylobacter biology and pathogenicity 8

2.7 Structural diversity and biological importance of ABO, H, Lewis and secretor BgAg carbohydrates 14

2.8 *Helicobacter pylori* and *C. jejuni* adhesins 17

2.9 Major Outer Membrane Protein (MOMP) 19

2.10 Role of protein glycosylation on MOMP functionality 21

2.11 Transmission of antibiotic resistant *Campylobacter* from animal to humans 24

2.12 Alternative treatments to reduce *Campylobacter* colonisation in chickens 24

2.12.1 Vaccine therapy 25

2.12.2 Bacteriocins 26

2.12.3 Bacteriophages....27
2.12.4 QPLEX....29
2.13 Conclusion....31
Chapter 3: MOMP purification....32
3.1 Introduction....32
3.2 Cloning of MOMP NCTC1168 and introduction into *E. coli*....32
3.2.1 Purification of recombinant MOMP (rMOMP) protein....33
3.3 Purification of native MOMP from *C. jejuni*....36
3.3.1 Large scale growth and cell collection....36
3.3.2 Western blot of MOMP....41
3.4 Structural characterisation of the MOMP proteins....42
3.4.1 MOMP glycosylation....42
3.4.2 Tertiary structure....43
3.4.3 Characterisation of the secondary structure of purified native MOMP....45
3.5 Conclusion....46
Chapter 4: Characterisation of MOMP glycosylation....48
4.1 Introduction....48
4.2 Mass spectrometry of MOMP....49
4.3 Searching for other MOMP PTMs....54
4.4 Conclusion to this chapter....66
Chapter 5: Interaction between MOMP, synthetic and biologic ligands....67
5.1 Introduction....67
5.2 Binding of BgAgs to MOMP....67
5.3 Inhibition of MOMP binding to BgAgs by QPLEX....76
5.4 Locating the interaction site of ligands with MOMP....78
5.4.1 Footprinting with carbene labelling....79
5.4.2 Footprinting efficiency of aryldiazirine in detergents....80

5.4.3 Optimising protein coverage using multiple proteases 82
5.4.4 Footprinting of MOMP with QPLEX 89
5.4.5 Footprinting of MOMP with Le^b 99
5.5 Protein ligand docking 102
5.5.1 Blind docking of MOMP 107
5.5.2 Ligand docking on the extracellular region of MOMP 111
5.5.3 Docking QPLEX to the inside of the wtMOMP porin 115
5.5.4 Docking QPLEX to intercellular MOMP 122
5.5.5 Docking Le^b to MOMP 124
5.6 Conclusion 126
Chapter 6: Binding of MOMP to intestinal tissue (*ex vivo*) 128
6.1.1 Introduction 128
6.1.2 Tagging of wtMOMP with fluorescein 130
6.1.3 Unstained gut sections 132
6.1.4 Reducing background fluorescence of gut tissue 137
6.1.5 Immunofluorescence binding of wtMOMP to gut tissue 146
6.2 wtMOMP binding to Le^b +ve mouse tissue 153
6.3 Conclusion 160
Chapter 7: Methodology 161
7.1 Purification of recombinant MOMP 161
7.1.1 Codon optimisation of *momp* NCTC1168 for *E. coli* 161
7.1.2 Plasmid DNA recovery 163
7.1.3 Transformation of plasmid DNA into *E. coli* strain DH5a 163
7.1.4 Recovery of plasmid DNA from bacterial culture 164
7.1.5 Cloning *momp* gene into pTAMAHisTEV 164
7.1.6 Transformation of pTamAHisTEVmomp into storage *E. coli* TOP10F One Shot 165
7.1.7 Agarose gel electrophoresis 166

7.1.8 Primers 166
7.1.9 Transformation of *E. coli* expression cells by electroporation 167
7.1.10 Small scale expression test from *E. coli* Lemo21 DE3 cells 167
7.1.11 SDS PAGE Electrophoresis 168
7.1.12 Large scale growth and protein purification of recombinant protein 169
7.2 Purification of MOMP from native *C. jejuni* 170
7.2.1 Strain confirmation 170
7.2.2 Cell collection and purification of native MOMP 172
7.3 Protein confirmation 174
7.3.1 Western blot 174
7.3.2 Semi-dry transfer 174
7.3.3 Wet transfer 175
7.3.4 Immunoblotting 175
7.3.5 Circular Dichroism Spectropolarimeter 176
7.3.6 Maldi TOF mass spec 176
7.4 Binding of wtMOMP to Le^b, Le^a, Type 1, Type 2 177
7.5 Inhibition of wtMOMP/Le^b with QPLEX 178
7.6 Carbene footprinting 179
7.6.1 LC-MS and -MS/MS analysis of MOMP peptides 181
7.7 Protein-ligand docking 182
7.7.1 Receptor 3D structure 183
7.7.2 Ligand 3D model 183
7.7.3 Preparing the receptor 183
7.7.4 Docking 185
7.8 Immunohistochemistry of wtMOMP to gut tissue sections 186
7.8.1 Deparaffinisation and rehydration of paraffin sections 187
7.8.2 Reduction of background fluorescence 187

7.8.3 Conjugation of fluorescent probe to wtMOMP 187
7.8.4 Staining of gut sections with protein and Le^b 188
7.8.5 Confocal Microscopy 189
7.9 Statistical analysis 190
Chapter 8: Conclusion 192
8.1 Introduction 192
8.2 MOMP purification 192
8.3 Characterisation of MOMP glycan 194
8.4 Interaction of MOMP with QPLEX and Le^b 195
8.5 MOMP interaction with the gut 197
Chapter 9: Appendix 198
References 199

List of figures

2.1	Human vs chicken response to *C. jejuni*	7
2.2	Structure of MOMP	20
2.3	Structure of QPLEX	29
2.4	Siderophore uptake pathway	30
3.1	Digestion of pJO1 with Nco1 and HindIII	33
3.2	Small scale expression of rMOMP	33
3.3	Large scale expression of rMOMP	34
3.4	rMOMP purification profile	35
3.5	Biomass of native *C. jejuni* MOMP	37
3.6	Anion exchange elution profile	38
3.7	SDS PAGE analysis of anion exchange purification	39
3.8	Size exclusion chromatography profile from the purification of MOMP	40
3.9	SDS PAGE analysis of fractions from the major elution peak of the size exclusion chromatography.	40
3.10	Western blot of MOMP strains and variants	42
3.11	6% SDS PAGE gel of MOMP	43
3.12	Native gradient PAGE	44
3.13	CD spectra of wtMOMP and $MOMP^{T/G268}$	46
4.1	MALDI TOF mass spectrum of $MOMP^{T/G268}$ and wtMOMP	50
4.2	LCMS of trypsinated wtMOMP	52
4.3	LCMS of trypsinated wtMOMP	52
4.4	LCMS of trypsinated $MOMP^{T/G268}$	53
4.5	Trypsin and chymotrypsin cleavage on MOMP	58
4.6	Sequence of wtMOMP with data from trypsin and chymotrypsin digests combined	63
4.7	wtMOMP protease coverage	64
4.8	Typical nomenclature for peptide fragments	65
4.9	Typical nomenclature for oligosaccharide fragments.	65
5.1	Blood group carbohydrate structures (Le^b, Le^a, Type 1 and Type 2)	69

5.2	Biochemical pathways of ABO, Lewis and Secretor BgAg carbohydrates in the digestive tract from type 1 precursor oligosaccharide	70
5.3	Binding of wtMOMP to Le^b, Le^a, Type 1, Type 2	71
5.4	Binding of increasing concentration of Le^b to wtMOMP (1pM).	73
5.5	Binding of increasing concentration of Le^b in the absence of wtMOMP.	74
5.6	Binding of 1 pM of wtMOMP, $MOMP^{T/G268}$ and $MOMP^{\Delta pseD}$ to 1 µM of Le^b.	75
5.7	Binding of 1pM of wtMOMP, $MOMP^{T/G268}$ and $MOMP^{\Delta pseD}$ to ELISA wells adsorbed with 1 µM of Le^b in the presence of increasing concentrations of QPLEX	77
5.8	Labelling efficiency by aryl diazirine in the presence of non-ionic detergents.	81
5.9	Footprinting of wtMOMP protein with carbene probe	83
5.10	wtMOMP with associated tryptic peptides	84
5.11	Amino acid sequence of wtMOMP with associated chymotryptic peptides	86
5.12	wtMOMP digest peptides	89
5.13	Labelling of wtMOMP with carbene in the presence of QPLEX	91
5.14	Fractional modification of chymotryptic wtMOMP peptides after carbene labelling in the absence and presence of 10uM QPLEX	93
5.15	Resultant carbene footprinting of wtMOMP with the aryl diazirine probe after incubation with QPLEX	95
5.16	Charge distribution on wtMOMP	97
5.17	Fractional modification of tryptic wtMOMP peptides after carbene labelling in the absence and presence of 35 µM Le^b	99
5.18	Resultant carbene footprinting of wtMOMP with the aryl diazirine probe after incubation with Le^b	101
5.19	Schematic of the FRED docking process	104
5.20	Docking of top two QPLEX poses to MOMP after initial round of blind box searching	108
5.21	QPLEX-MOMP docking report of pose 1.1.	109
5.22	QPLEX-MOMP docking report of pose 2.1.	110
5.23	QPLEX docking pose (1.1) on MOMP	111
5.24	QPLEX-MOMP docking report of top ranked pose on extracellular site of MOMP	112

5.25	Docking result of highest ranked pose of QPLEX on extracellular surface of MOMP	113
5.26	Residue fingerprint and positioning of amino acids that QPLEX interacts with on MOMP	114
5.27	Analogous residue fingerprint of QPLEX on wtMOMP	115
5.28	Protein-ligand docking result of the two highest ranked poses of QPLEX within the constriction zone of the MOMP porin	116
5.29	QPLEX-MOMP docking report of the best two ranked poses within the MOMP porin constriction zone	116
5.30	Residue fingerprint along the transverse of the constriction zone	117
5.31	The resultant residue fingerprint of the amino acids involved in the interaction extending from pose 1 of QPLEX to the beginning of the constriction zone of MOMP 85H (5LDT) has been mapped on to the analogous residues in the homology model of wtMOMP 11168	119
5.32	The resultant residue fingerprint of the amino acids involved in the interaction extending from pose 2 of QPLEX to the beginning of the constriction zone of MOMP 85H (5LDT) has been mapped on to the analogous residues in the homology model of wtMOMP 11168	120
5.33	Charge map of wtMOMP on a surface diagram from Pymol.	123
6.1	General organisation of the digestive system	128
6.2	Basic structure of the gastrointestinal wall	129
6.3	Tagging of wtMOMP with NHS-Fluorescein	131
6.4	Micrograph of unstained sections of corpus, antrum, duodenum and colon	134
6.5	Fluorescence microscope image of the mucosa of the human stomach corpus	135
6.6	Fluorescence microscope image of the antrum of the gastrointestinal tract	135
6.7	Fluorescence microscope image of the duodenum which extends from the lower portion of the stomach to the upper part of the small intestine of the gastrointestinal tract.	136
6.8	Fluorescence microscope image of the colon of the large intestine of the gut.	136
6.9	Images of the human stomach corpus were captured at wavelengths in the A. blue, B. green and C. red channels during a spectral scan	138
6.10	Human corpus in the green fluorescent channel after treating the unstained tissue with a quencher	140
6.11	Fluorescence images taken of the human corpus in the red fluorescent channel after treating the unstained tissue with a quencher	141

6.12 Confocal microscope images of the human corpus glands that were first quenched with autofluorescence followed by staining with wtMOMP conjugated to a probe that fluoresces in the far-red spectrum 143

6.13 Confocal microscope images of a section of the human corpus glands that were first quenched with autofluorescence followed by staining with wtMOMP conjugated to a probe that fluoresces in the far-red spectrum 144

6.14 Confocal microscope images of a section of the human corpus glands 145

6.15 Confocal microscope images of the human colon after staining with wtMOMP conjugated to a probe that fluoresces in the far-red spectrum 147

6.16 Confocal microscope images of the lower mucosa of the human corpus after staining with wtMOMP conjugated to a probe that fluoresces in the far-red spectrum 148

6.17 Confocal microscope image of the intestinal glands of human corpus tissue after staining with wtMOMP conjugated with far-red probes. 149

6.18 Confocal microscope image of surface epithelium of human corpus tissue after staining with wtMOMP conjugated with far-red probes. 150

6.19 Confocal microscope image of surface epithelium of human corpus tissue after staining with wtMOMP conjugated with far-red probes 151

6.20 P28 normal and transgenic FVB/N mouse stomach tissue stained with Le^b specific mouse antibody 154

6.21 P28 normal and transgenic FVB/N mouse stomach tissue stained with Le^b specific mouse antibody (magnification) 155

6.22 Normal P28 FVB/N mouse stomach tissue and transgenic Le^b positive P28 FVB/N mouse stomach tissue that were stained with either $MOMP^{T/G268}$ or wtMOMP protein 156

6.23 Normal P28 FVB/N mouse stomach tissue and transgenic Le^b positive P28 FVB/N mouse stomach tissue that were stained with either $MOMP^{T/G268}$ or wtMOMP protein (magnification) 157

6.24 Leb +ve transgenic P28 FVB/N mouse stomach tissue stained with $wtMOMP^{far\text{-}red}$ and an auto-fluorescence quencher. 158

6.25 Le^b +ve transgenic P28 FVB/N mouse stomach tissue incubated with wtMOMP with increasing concentrations of soluble Le^b, in a competitive binding study 159

7.1 Gene sequence and protein sequence of codon optimised MOMP 161

7.2 A plasmid map of the pTAMAHisTev plasmid 162

7.3 Kanamycin resistance test 171

7.4 PCR products of the *momp* gene in *C. jejuni* NCTC11168 wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ strains 171

7.5 PCR products of the *pseD* gene in the *C. jejuni* NCTC11168 wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ strain 172

7.6 Synthesis of carbene probe 179

7.7 Irradiation (photolysis at near-UV laser at 349 nM) and flash freezing of the carbene has the dual benefit of halting enzyme turn-over and reducing diffusion of the probe 181

7.8 A series of boxes were created for exploration of the 'active site' that should be probed for ligand docking 185

9.1 Normal and transgenic Le^b positive P28 FVB/N mouse stomach tissue stained with MOMP$^{T/G268}$ protein 198

9.2 Normal and transgenic Le^b positive P28 FVB/N mouse stomach tissue stained with wtMOMP protein, 198

List of tables

4.1	Theoretical mass of the peptide of interest by itself, with *N*-acetyl-galactosamine and with *N*-acetyl-galactosamine plus unknown carbohydrate residues	51
4.2	List of peptides observed in LCMS from digestion of wtMOMP with trypsin	59
4.3	List of peptides observed in LC MS from digest of wtMOMP with chymotrypsin	61
5.1	Ranking of QPLEX poses after blind-box docking to whole surface of MOMP	108
5.2	High resolution docking of pose 1	109
5.3	High resolution docking of pose 2	110
5.4	Docking score of extracellular site of MOMP	112
5.5	Ligand-protein docking score of QPLEX and inner porin of MOMP	115
5.6	Ligand-protein docking score of QPLEX and intracellular portion of MOMP	123
5.7	Ligand-protein docking score of Le^b to various regions of the MOMP protein	124
7.1	Le^b, Le^a, Type 1, Type 2 BgAgs and BSA were adsorbed to the surface of amino-linked ELISA plates.	178

Chapter 1: **Introduction**

1.1 Aims and objectives

The aim of this thesis is to explore the *C. jejuni* virulence factor, MOMP, and its contribution to the interaction of *C. jejuni* with biological and non-biological ligands the results of which are significant for understanding the way *C. jejuni* behaves in avian and human gut and in the external environment. This thesis investigates the glycosylation of the MOMP protein and how this affects *C. jejuni* interactions with ligands such as BgAgs and inhibitors such as QPLEX. A further aim of this thesis is to explore where on the MOMP protein these ligands and inhibitors bind in order to understand the molecular details that underpin these interactions. Lastly, the thesis explores how MOMP binding patterns occur *ex vivo* in the human intestinal gut. The aims are completed by three main objectives; purification of the glycosylated and non-glycosylated MOMP, verification and characterization of the putative glycan that exists on MOMP, and characterization and localization of the MOMP interactions with BgAgs, iron(III)-chelates and gut sections.

Chapter 2: **Literature review**

Introduction

This literature review explores *C. jejuni* and one of its main virulence factors – the Major Outer Membrane Protein (MOMP), the attachment of *C. jejuni* to molecules such as histo-blood group antigens (BgAgs), and the role that glycosylation plays on MOMP protein functionality.

2.1 Campylobacter in humans

The World Health Organisation (WHO) recently published its first list of antibiotic-resistant "priority pathogens" – a catalogue of twelve families of bacteria that pose the greatest threat to human health. Fluoroquinolone- resistant *C. jejuni* was number six on this list[1]. It is one of the four key global causes of gastrointestinal diseases. Currently, the cost of illness is $1.3 billion in the US[4] and €2.4 billion in the EU27[2]. There are a large variety of host sources of *C. jejuni*, but poultry is thought be the major risk factor, as it is the main animal reservoir.

The incidence of *Campylobacter* infection has increased over the past 10 years in both the developing and developed nations[5]. In humans, *C. jejuni*, then *Campylobacter coli (C. coli)*, is the leading pathogen in campylobacteriosis globally, and cases of campylobacteriosis by other *Campylobacter* species such as *Campylobacter lari (C. lari), Campylobacter upsaliensis (C. upsaliensis)* and *Campylobacter fetus (C. fetus*) are a small fraction, in comparison[6]. This thesis focuses on *C. jejuni* and *Campylobacter* refers to the *C. jejuni* only. *C. jejuni* is the leading cause of gastroenteritis, an infection of the gut that can lead to diarrhea and vomiting. Other gastrointestinal manifestations from *Campylobacter* infection of the gastrointestinal tract include inflammatory bowel disease[7], oesophageal disease[8] , periodontal disease[9] , functional

gastrointestinal disorders[10], colorectal disease[11] , celiac disorder[12] and cholecystitis[13] . *Campylobacter* infection can cause extra-gastrointestinal manifestations by way of a local isolated infection, a systemic manifestation after an episode of enteritis, or a post infectious immune disorder[5]. These diseases can include Guillain Barré Syndrome (GBS)[14] , Miller Fischer Syndrome[15] , bacteraemia and septicaemia[16], cardiovascular complications [17, 18], meningitis[19], extraoral complications[20], reactive arthritis[21], and complications of the reproductive system[22].

2.2 Incidence of disease

After a marked increase in incidence of *Campylobacter* -induced disease in the 1980s and 1990s, the prevalence in human populations stabilized in the early 21st century but is increasing again as of recent. North America, Europe, and Australia have presented an alarming increase, and data show that in parts of Africa, Asia, and the Middle East, campylobacteriosis has become endemic especially in children[5]. According to "The Community Zoonoses Reports" of the European Food Safety Authority (EFSA) and the European Centre for Disease Prevention and Control (ECDC), in the past 5 years campylobacteriosis has been the most commonly reported zoonosis in the EU followed by salmonellosis and yersiniosis[3, 23]. In North and Central America, in 2012, cases of campylobacteriosis were up 14% from the period 2006-2008 whereas the incidences of *Cryptosporidium, Listeria, Salmonella, Shigella,* Shiga-toxigenic *Escherichia coli* (STEC) O157, and *Yersinia* infections decreased over the same period[24]. The Centre for Disease Control estimated that *Campylobacter* spp. accounted for 800,000 cases of food-related illnesses, which made up 9% of all foodborne diseases in the US that year caused by bacteria, viruses and parasites[25]. Clinical cases of campylobacteriosis are under-reported in the EU 27 member states where "there may be not less than 2 million and possibly as high as 20 million cases of clinical campylobacteriosis per year"[3]. Within the UK, *Campylobacter* is the most common cause of food poisoning - in 2012,

there were 72,562 laboratory confirmed cases of campylobacteriosis, which due to under-reporting, was estimated to represent around 460,000 actual community cases, with more than 15,000 hospitalizations and 76 deaths [26, 27]. Further investigations into pathogen-associated diarrhea saw that prevalence of *Campylobacter* species has not decreased for the past 15 years, compared to a decrease in prevalence of *E. coli*, *Salmonella,* and *Yersinia enterocolitica*[28, 29]. In most cases, campylobacteriosis is self-limiting and symptoms disappear within 7 days, requiring only maintenance of hydration and electrolyte balance and thus the use of antibiotics to treat the disease in humans is not recommended and rather restricted to those that are immunocompromised or suffering from severe illness or dysentery.

2.3 Cost burden

Although only a small number of infections lead to symptomatic illness, there is a high cost burden associated with infection[30, 31]. In 2011 in the Journal of Food Protection, Hoffmann and colleagues estimated the annual cost of campylobacteriosis in the United States to be $1.7 billion per year[32]. More recently, estimated total costs of Campylobacter-associated GBS ($0.2–$1.8 billion) were added to previously estimated costs of campylobacteriosis ($1.3– $6.2 billion) for a total annual cost from *Campylobacter* of $1.5–$8.0 billion in the US[4]. In the EU27, the total annual costs are estimated to be €2.4 billion.[33] Campylobacteriosis is a disease of importance due to the costs resulting from treatment, as well as from the loss of human resources.

2.4 Risk factors/ Sources and routes of transmission for human infection

Humans are highly susceptible to *Campylobacter* infection due to their association with risk factors such as ingestion of undercooked chicken, international travelling, environmental exposure and

direct contact with animals. It was revealed that *C. jejuni* was the cause of 70% of gastrointestinal infections acquired during international travel in the period 2006-2007[34]. Hosts include farm animals (cattle, sheep, poultry and pigs) and wild animals (birds and mammals). Especially in the avian and mammalian gut, the bacterium thrives at 37-42°C but survives longest *ex vivo* in cold, dark, moist environments - thus handling and eating of raw or undercooked contaminated poultry frequently leads to infection. In 2014/15 the UK Food Standards Agency reported that around 80% of fresh whole chicken in supermarket stores were carrying *Campylobacter*, 20% of those were measured to be carrying more than 1000 colony forming units (cfu) per gram of chicken skin (confers a high contamination rate)[26]. Due to the high carry rate of *Campylobacter* in domesticated and imported broiler chickens, *Campylobacter* species are found in abundance in poultry farms, the surrounding environment, soil and water sources. Unlike salmonella which has been successfully controlled in the European region, *Campylobacter* is still at large in sources exposed to humans.[5] In Switzerland, it was estimated that 71% of campylobacteriosis cases from 2001 to 2002 had chickens at the source of infection[35]. The literature points towards an overwhelming number of cases with chicken at the source of *Campylobacter* infection; the European Food Safety Authority panel on Biological Hazards estimate that the transmission of 50 to 80% of all human cases of campylobacteriosis can be traced back to chicken reservoirs. More specifically, 20 to 30% of cases may be accounted for from handling, preparing and consumption of broiler chicken (chickens reared and prepared for meat consumption)[33].

2.5 Human vs chicken response to infection/colonisation

A *Campylobacter* infection dose of as low as 500-800 bacteria in humans commonly results in acute gastroenteritis with symptoms lasting for up to two weeks such as inflammation, abdominal pain, fever and diarrhea[36] whilst the disease itself can be resolved within a week. Socio-economic

status can also affect the disease outcome; in the developed parts of the world, *Campylobacter* causes diarrhea with mucus and is self-limiting. In the developing parts of the world, *Campylobacter* is usually limited to infection of children and causes watery diarrhea, but the host may gain natural immunity against infection as adults[37]. To establish infection, *C. jejuni* has to overcome the mechanical and immunological barriers of the GI tract. The corkscrew morphology and motility of *C. jejuni* and the short lipooligosaccharide helps to reduce nonspecific binding to the mucin glycoproteins and aid penetration of the barrier[38]. Once *C. jejuni* has passed through the mucus layer it can interact with epithelial cells – which induces a host response of cytokine release, such as IL-8, a proinflammatory cytokine[39]. Cell-mediated (or TH1) immunity is the primary response that is used to clear *C. jejuni*, involving dendritic cells (DCs) and macrophages, as opposed to antibody-mediated immunity[40]. Infections in children younger than 6 months old result in low levels of specific IgA, IgG and IgM, possibly owing to the presence of maternal antibodies[41]. This was further proved by Ruiz- Palacios who studied maternal milk which contains α1,2-linked fucosyloligosaccharides in high percentage which appears to serve as a soluble ligand that competes with intestinal epithelial cell surface receptors for binding to *Campylobacter* and other pathogens that target α1,2 ligands[42].

There is a greater understanding of *Campylobacter* colonization of chickens which does not lead to the same response as with human infection, i.e. no symptoms or pathological inflammatory response (**figure 2.1**)[43]. *C. jejuni* can colonize chickens by up to 10^9 colony-forming units per gram (cfu/g) of infected intestine, especially in the depths of the caecum where *C. jejuni* is found in the mucus of the epithelial cells, with no detrimental effects on their health[44, 45]. Within the caeca, there is an abundance of lactic acid and short chain fatty acids[43]. Nachamkin's analysis of chicken colonisation found a slight inhibition of human epithelial cell invasion by *C. jejuni* in the presence of chicken intestinal mucus which points to its role in the asymptomatic nature of chick infection[46].

Differences in body temperature between chicks and humans may also contribute to the different outcomes of infection with *C. jejuni*. Avian body temperature is 42-45°C as opposed to 37°C for humans, and this temperature difference has been shown to shift transcription profiles from cultures grown at the two different temperatures. There appears to be alterations in membrane structure by the upregulation of genes for transport and binding proteins, as well as cell wall and envelope constituents[47].

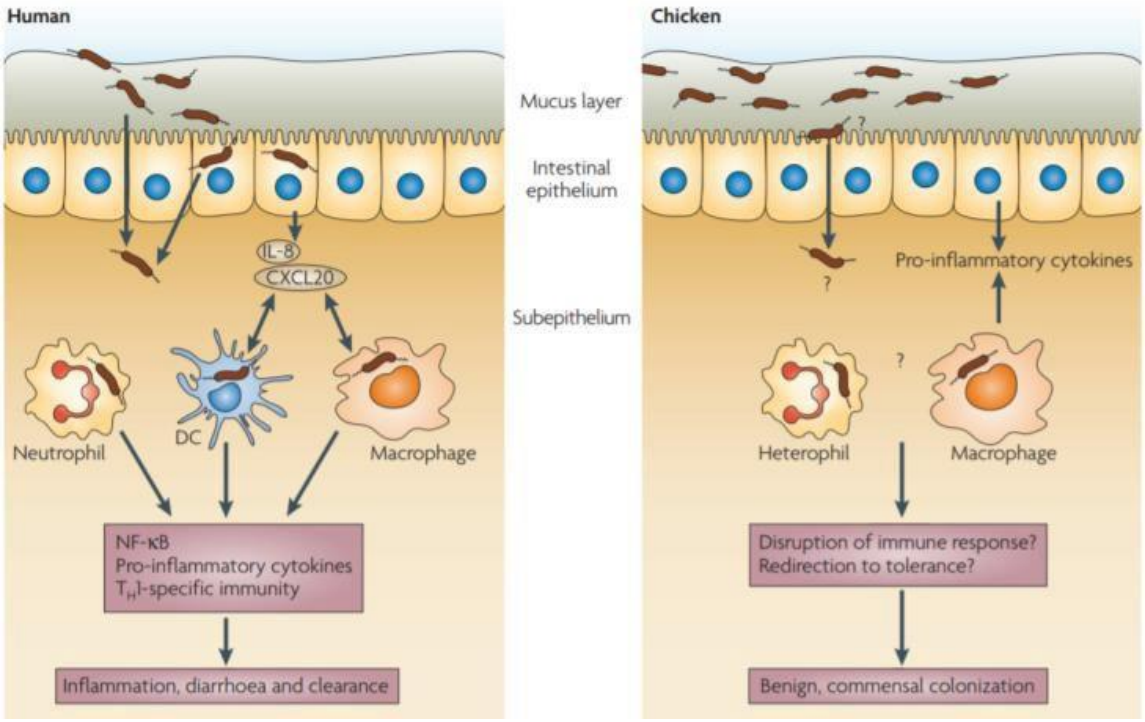


Figure 2.1: Molecular and cellular features of the innate immune response to *C. jejuni* in humans and chickens. *C. jejuni* can pass the mucus layer in humans to interact with the intestinal epithelial cells. This leads to production of interleukin (IL)-8. This leads to recruitment of dendritic cells (DC), macrophages and neutrophils, which interact with *C. jejuni* resulting in a massive pro-inflammatory response and increases in the corresponding cytokines. This is contrary to the response of chicken to *C. jejuni* where it resides primarily in the mucosal layer in chicken intestines. *In vitro* evidence shows that *C. jejuni* can stimulate the production of IL-1β, IL-6 and intracellular nitric oxide synthase from epithelial cells and macrophages, but the host response is not one of inflammatory diarrhea in chickens[48]. Image taken from Young *et al.,* Nature Reviews Microbiology 2007.

2.6 Campylobacter biology and pathogenicity

Most Campylobacter species grow under microaerophilic conditions with a respiratory metabolism. They are non-spore forming species with an average area of 0.1 to 0.9 by 0.5-5 µm. They come under chemoorganotrophs which oxidise sugars, fats and proteins as their source of energy[49]. Campylobacter species are Gram-negative bacteria as they do not retain the crystal violet staining used in the gram staining method to differentiate between bacteria. Their cell envelope comprises an inner membrane and outer membrane, which are separated by the periplasmic layer. The outer membrane is also tethered to peptidoglycans within the periplasmic layer. The bacteria can be rod or curved shaped with either single polar flagellum, bipolar flagella or no flagellum at all depending on the species[50]. As disease organisms, gram-negative bacteria can be more virulent due to the capsule or slime layer which hides the surface antigens of the cell and thus camouflages the bacterium. The host then finds it difficult to phagocytose and opsonise the invader[51]. In addition to this virulence factor, the gram negative bacteria outer membrane contains lipopolysaccharide which is endotoxic and increases the severity of inflammation[52].

The lipooligosaccharide and capsule

The lipooligosaccharide (LOS) is found in the outer leaflet of the outer membrane of some Gram-negative bacteria. LOS of *C. jejuni* consists of lipid A linked to a oligosaccharide, but lacks the O-specific polysaccharide of the LPS that is found in other Gram-negative bacteria. In *C. jejuni*, LOS is associated with immune avoidance and is highly variable[53]. An example is the molecular mimicry between a variety of *C. jejuni* LOS structures and human gangliosides. 1 in 1000 cases of campylobacteriosis leads to the autoimmune disorders of Guillain-Barré Syndrome (GBS), and Miller Fisher Syndrome – a variant of GBS[54]. Mutations found in genes that involve LOS biosynthesis affect serum resistance, adherence to and invasion of INT407 cells[52].

High Molecular Weight LPS is now known to be a highly variable capsular polysaccharide, rather than an LPS. The structure of the capsule of *C. jejuni 11168* has been found to include 6-methyl-d-glycero-α-l-glucoheptose, β-d-glucouronic acid modified with 2-amino-2-deoxyglycerol, β-d-GalfNAc and β-d-ribose[55], and contains a novel modification on the GalfNAc[56]. The *C. jejuni* capsule is also important for chick colonization[57].

Adherence mechanisms

C. jejuni has a lack of identifiable adherence organelles, which present as pili in other gram negative and positive species, but several of its proteins do contribute to *C. jejuni* adherence to eukaryotic cells. CadF is a channel forming protein which binds specifically to fibronectin, found basolaterally on epithelial cells. CadF has been shown to be required for maximal binding and invasion, especially during chick colonisation where CadF mutants fail to colonize to the same extent. CadF is similar to E. coli OmpA and forms membrane channels, but the role of this activity, if any, has not been established[58–60].

JlpA is another *C. jejuni* adhesin, a surface exposed lipoprotein that is important for Hep-2 cell binding. Surface-exposed heat shock protein 90alpha (Hsp90α) is surface localized in the hep-2 cells and binding to JlpA activates NF-κB and p38 mitogen-activated protein (MAP) kinase, both of which contribute to proinflammatory responses. Inflammation that is caused by *C. jejuni* infection may be down to JlpA-dependent adherence[61, 62]. A second lipoprotein in *C. jejuni* that has been characterised as an adhesin is CapA, an auto transporter, the mutation of which leads to decreased adherence to caco-2 cells and decreased colonisation and persistence in chicks[63].

On the other hand, there are adhesins of *C. jejuni* that are located in the periplasm. Peb1 is one such adhesin. It shares homology to the periplasmic-binding proteins of amino acid ATP-binding cassette (ABC) transporters. Peb1 binds to both aspartate and glutamate and Peb1

mutants cannot grow if they are the major carbon source. Further, mutants that lack peb1 colonize mice poorly, but this could be attributed to the loss of either the adhesion or the amino-acid-transport functions, or both. Peb1 is also crucial for adherence to HeLa cells[64, 65]. The mechanism by which these periplasmic proteins contribute to host-cell adherence by *C. jejuni* is still unclear.

Secretion systems

The fundamentals of secretion by *C. jejuni* are poorly characterised in comparison to other bacterial pathogens. One of the proteins secreted by *C. jejuni* is CiaB, required for invasion of cultured epithelial cells. An explanation of how CiaB is secreted can be attributed to the model of the Type III Secretion System (T3SS), where effectors are directly injected into the host cell. But in *C. jejuni*, the secretion of CiaB is not by T3SS but by the flagellar export apparatus. Yersinia species also display a similar method of flagellar export[66, 67].

There is a high level of genetic diversity between strains of *C. jejuni*. These differences are caused by intragenomic events such as rearrangements, point mutations, deletions, duplications and inversions but intra- and interspecies events can also cause variation via horizontal transfer of genetic information[68, 69]. Such natural transformation requires homologues of type 2 and type 4 secretion system (T2SS and T4SS) proteins. The T2SS is important for targeted delivery of toxins, proteases, cellulases and lipases. The CtS family of proteins found in *C. jejuni* are similar to proteins from T2SS system from other bacterial families[70]. T4SS proteins are also involved in horizontal DNA transfer, DNA uptake from pr into the extracellular substance, toxin secretion and the injection of bacterial substrates into the cytosol of target cells in various bacteria. In *C. jejuni*, the *cjp* family of genes are found in several open reading frames (ORFs) on the pVir plasmid. pVir has been shown to contribute to the ability of 81-176 to invade intestinal cells in vitro[71]. Although pVir lacking *C. jejuni* strains are still competent, a modest defect in natural competence was

observed with a *cjp3* mutation[72]. So pVir T4SS does not function as the primary DNA uptake system, although adapting competence levels[73].

T6SS is a secretion system that is extended through the bacterial double membrane thus substrates are conveyed straight from the cytoplasm of the bacterium into a target cell or to the extracellular space. T6SS consists of three main complexes: proteins in the inner membrane that are T4SS component-like, the baseplate complex, and the tail complex, which are formed by components evolutionarily related to contractile bacteriophage tails[74]. Recently, the existence of a novel type VI secretion system (T6SS) in *C. jejuni* strains isolated from the UK, Spain and South-East Asia has emerged as a key element with the potential to confer bacterial survival and host pathogenesis[75]. Studies showed that hypervirulent *C. jejuni* isolate that encodes a functional T6SS has a greater ability to invade and create characteristic "attaching and effacing" lesions in human INT407 compared to primary chicken embryo intestinal cells (CEICs). Additionally, higher bacterial invasion in human INT407 triggered higher levels of expression of major proinflammatory cytokines, such as IL1β and IL-6[76]. These results obtained by Singh and Mallick (2018) may hold the key to the observed differences in pathogenesis between human and chicken cells.

More than one-third of the bacterial proteome that is synthesized on cytoplasmic ribosomes is exported to extra-cytoplasmic locations, where the polypeptides acquire their native functional states. These proteins are either membrane-embedded or are secreted to the *trans* side of the plasma membrane; for example, to the periplasm. Typically, export temporally delays protein folding and alters the cellular topology of this process, as most proteins are exported in unfolded states. Bacteria have evolved several sophisticated export mechanisms for cell envelope and membrane biogenesis but only the general secretory (Sec) pathway is essential for viability and is ubiquitous in all domains of life. The central component of the Sec system is the transmembrane SecYEG channel/translocon which translocates proteins into, or across, the

plasma membrane. The ATPase motor SecA assembles with SecYEG at the plasma membrane to form the translocase holoenzyme, recognizes exported proteins and mediates chemo–mechanical conversion during transmembrane crossing[77]. Proteins that are located outside the cytoplasm are normally synthesized with amino-terminal signal peptides that target them to either the Sec or the Tat (twin-arginine translocation) protein export pathway, both of which are present in *C. jejuni*. The difference between these two export systems is that the Sec apparatus translocates polypeptides in an unstructured state, whereas the Tat pathway transports proteins that have already folded. Proteins are targeted to the Tat machinery by signal peptides that contain a conserved twin-arginine motif[78].

Autotransporters are the largest family of secreted proteins in gram negative bacteria, and their transport is categorised by the type 5 secretion system (T5SS). Autotransport denotes the ability of surface-localized outer membrane proteins to cross the outer membrane autonomously. These proteins contain an autotransporter domain, a structural domain, which is permanently located at the C-terminal end of the protein, forming a beta-barrel structure. The barrel is oriented in the membrane such that the N-terminal portion of the protein, the passenger domain, is presented on the cell surface. Usually, these proteins are virulence factors, associated with infection or virulence in pathogenic bacteria by way of adhesion, toxigenicity and intracellular spread, due to the family's highly diverse N-terminal functional domains, despite the common mode of secretion. Wooldridge and Oldfield screened *C. jejuni* NCTC 11168, discovering two putative autotransporters; CapA and CapB. A role for CapA as an adhesin was then identified in vitro using the Caco-2 cell model for campylobacter interaction with human enterocytes[79]. Further, CapA plays a role in association and invasion of human cells in vitro and is required for efficient chicken colonization by strain NCTC11168[63].

Flagella

C. jejuni flagella and flagellar-mediated motility are vital for host colonization, virulence in ferret models, secretion and host-cell invasion. Thus, the regulation of flagella biogenesis and motility (via chemotaxis) is an active area of research[48]. Chemotaxis is the movement of an organism towards or away from a chemical stimulus and the chemotactic behaviour of *C. jejuni* toward L-fucose, a constituent of both bile and mucin, may be an important factor in the affinity of the organism for the gallbladder and intestinal tract[80].Chemotaxis also plays an important role in protection of H. pylori from luminal acidity[81]. Polar flagella allows motility which enables bacteria, such as H. pylori, a cause of gastric ulcers, to reach the gastric epithelium, then adhere to it with several adhesins, and colonize the epithelium[82]. Studies by Kim et al. with a non-motile fliD mutant showed that FliD and thus a functional flagellum is required for the colonization of mice[83]. Studies by Wassenaar et al. showed that flagella were important for invasion of *C. jejuni* into INT-407 cells, but a direct adhesive function of flagellum was not observed. *C. jejuni* has two highly identical flagellin protein, flaA and flaB, but only flaA is required for flagella-mediated motility and adhesion[84]. Clyne et al. constructed flagellin mutants and a flagellar regulator mutant and they observed that all mutants still adhered to gastric cells, signaling that there are other factors involved in adhesion[85]. The flagella plays an important role for biofilm formation and results indicate that flagella reach crevices, grasp to improve bacterial adhesion, and are able to penetrate substructures not accessible to the bacterial cells[86].

Cytolethal distending toxin

C. jejuni releases a toxin called cytolethal distending toxin (CDT), which causes cell cycle arrest by triggering a G2 cell cycle checkpoint that is initiated during DNA replication and that is independent of DNA damage thus causing death in sensitive eukaryotic cells[87]. It is not yet clear

exactly how CDT contributes to *C. jejuni* pathogenesis, though it acts on a variety of mammalian cells such as T cells. Because CDT genes have been found in a diverse range of bacterial pathogens including enterohepatic Helicobacter spp., E. coli and Shigella spp. it may have a role in disease progression and immunosuppression[55]. CDT is encoded by adjacent genes, cdtA, cdtB, and cdtC, for which expression of all is required to produce active CDT[88]. CdtB has been shown to make double stranded cuts of host cell DNA and thus is similar to a nuclease. Less is known about CdtA and CdtC, which are thought to bind to an unknown receptor on the host cell surface but recent studies have shown that the three proteins are present in the CDT holotoxin in a 1:1:1 stoichiometry[55]. In conjunction with Hela cells, Lee et al used purified *C. jejuni* Cdt subunits to determine the activity of reconstituted holotoxin and identified regions within CdtA and CdtC that seem to be critical for holotoxin formation[89]. It is intriguing that *C. jejuni*, which establishes long term and asymptomatic relationship with a variety of hosts, has retained the CDT toxin. This may flag up a role for CDT in asymptomatic commensal infections, which provides a way to avoid a host immune response. For example, the persistent *C. jejuni* colonization of wild-type mice, but not mice that are deficient for nuclear factor (NF)-κB, requires CdtB, which suggests that CDT might allow *C. jejuni* to escape immune surveillance in an NF-κB-dependent manner[90]. In chickens, CDT is expressed by bacteria in the heavily colonized caeca, but chickens do not generate CDT-neutralizing antibodies, whilst in human infection, anti-CDT neutralizing antibodies are found[91].

2.7 Structural diversity and biological importance of ABO, H, Lewis and secretor BgAg carbohydrates

This chapter considers the role and functions of BgAgs, which may help to determine differences in individuals and animals responding to infection and may further answer questions as to why *C. jejuni* is pathogenic in humans and not in poultry.

Although this thesis focuses mainly on the carbohydrate based BgAgs, it is also important to recognise BgAgs based on protein structures. Blood group antigens are either sugars or proteins, and they are attached to various components such as the red blood cell membrane. At present, there are 34 recognized human blood group systems with more than 300 known variants. These are all classified by the antigens, found on the surface of red blood cells, which are most often proteins but also carbohydrates. For example, the antigens of the Rh blood group are proteins. A person's DNA holds the information for producing the protein antigens. The RhD gene encodes the D antigen, which is a large protein on the red blood cell membrane. Some people have a version of the gene that does not produce D antigen, and therefore the RhD protein is absent from their red blood cells. On the other hand, the antigens of the ABO blood group are sugars. They are produced by a series of reactions in which enzymes catalyze the transfer of sugar units. A person's DNA determines the type of enzymes they have, and, therefore, the type of sugar antigens that end up on their red blood cells[92].

The differences in blood group antigen expression can increase or decrease host susceptibility to many infections. Studies have shown that there is an important relationship between microorganisms that colonise the body surface and the carbohydrates expressed on the host's epithelial cells, gastrointestinal tract, the respiratory pathway and the genitourinary systems[93]. Since the discovery of the ABO blood group, there has been an ongoing interest in the potential role of blood groups in infectious disease. The vast structural diversity of the ABO (encoded by the ABO, gene), H (encoded by the FUT1 gene), secretor (encoded by the FUT2 gene) and Lewis

(encoded by the FUT3 gene) histo-blood system genes control the expression of most of the carbohydrate structures present in areas of the body occupied by microorganisms[94]. The presentation of these carbohydrates on cells act as potential receptors for pathogenic and non-pathogenic microorganisms which will influence both the potential of and resistance to infection and resulting illness[95]. ABO, H, secretor and Lewis BgAg carbohydrates are not primary gene products but are synthesized by specific glycosyltransferases, encoded by the ABO, FUT1,FUT2 and FUT3 genes, that incorporate monosaccharide units in sequence to linear or branched precursor oligosaccharide chains, thus modifying and creating new antigenic specificities [96, 97].

Blood groups are frequent targets in epidemiological investigations since they are genetically determined traits with known polymorphic expression among individuals and populations. Frequencies of the different BgAg phenotypes vary across different ethnicities and populations. The O phenotype is common among Africans and South American natives while A and B are common in North European and Asian countries. Although H-deficient phenotypes are rare in all populations, they are more frequent in India. 80% of Caucasians present with the secretor phenotype. The Le (a+b−) phenotype is found in more than 20% of Caucasians and Black people but it is rare in Asians. Le(a−b+) is frequent in all populations but Le(a+b+) is more common in Asians and Polynesians than populations from the Western world, whilst Le(a−b−) is rare in Caucasians but common among Blacks. It is believed that selective pressure imposed by disease-causing microorganisms contributed to the variation seen amongst populations[97–99]. The variability resulting from these BgAg systems has important implications in susceptibility to infections, innate and adaptive immune responses, cancer and solid organ transplantation. For example, individuals with the O blood group are more prone to gastroduodenal diseases such as gastritis and peptic ulcers because *Helicobacter pylori* strains expressing the blood group binding adhesin (BabA) are able to bind to the Le^b carbohydrate that is highly expressed in gastric epithelial cells related to O

and secretor BgAgs. It was also observed that the South American specialist strains of *H. pylori* were more prone to bind to H type 1 and Le^b carbohydrates than general strains which correlates with the predominance of the O blood group in Amerindians[100, 101]. Another example of the relationship between BgAg and bacteria is the ability of H carbohydrate in human milk to protect infants against *C. jejuni*, the most common cause of bacterial diarrhea and thus infant mortality. By binding to specific ligands, the H type 2 carbohydrate inhibits the attachment of these microorganisms to intestinal cells thereby protecting breastfed babies from *Campylobacter* [42]. Many blood group antigens are receptors for toxins, parasites, and bacteria, where they can facilitate colonization or invasion or evade host clearance mechanisms[102, 103]. Blood group antigens can also serve as false receptors, preventing binding to target tissue. The human gastric mucosa and epithelial lining are heavily glycosylated, and bacteria are able to interact with these glycans quite well. At the heart of many of the bacteria-host glycan interactions are bacterial lectins which act as adhesins and mediate attachment to host cells, initiate colonization and define bacterial cell and tissue tropism[104].

2.8 *Helicobacter pylori* and *C. jejuni* adhesins

Adherence of a bacterial pathogen to host cells is a key initial event for successful colonization and subsequent disease development. Typically, bacterial adhesion is mediated through a specific interaction between a bacterial outer membrane protein, called adhesin, and an oligosaccharide (glycan) expressed on the surface of host cells that serves as a cognate adhesin receptor. The gastric mucosa of over half the global human population are colonized with *Helicobacter pylori*, and 1-3% of those people develop gastric cancer, the incidence of which has led the WHO to classify the bacterium as a carcinogen[105]. *H. pylori* uses two adhesins to bind host glycosylated structures - blood group antigen-binding adhesin BabA, and sialic acid (also known as neuraminic

acid)-binding adhesin SabA. BabA is specific for fucosylated glycoconjugates such as Le^b, which is expressed by gastric epithelial cells and found on mucins within the mucosa[106]. Adherence mediated by BabA is driven by pH dependent binding and is most efficient in the higher pH environment of the gastric mucosa[106]. SabA is specific for the Lewis blood group antigens sialyl Lewis X (sLeX) and sialyl Lewis A which are expressed at low levels in the gastric lining of healthy individuals but after colonization with *H. pylori*, the expression of sialic acid expression is increased in the stomach lining demonstrating that *H. pylori* influences the host gastric glycome[107]. Although the majority of *H. pylori* colonizes the gastric mucosa, only ~20% of the bacterial population is directly attached to the gastric epithelium, suggesting that the infected mucous layer acts as a reservoir for persistent infection of the gastric epithelium. Recently, a third adhesin of *H. pylori*, LabA, was observed to have glycan-binding activity. LabA was found to bind GalNAcβ1-4GlcNAc structures conjugated to MUC5AC mucins in the gastric mucosal layer as opposed to on the gastric epithelial cells[108].

Several proteins on *C. jejuni* contribute to its adherence to eukaryotic cells, one of which is MOMP. Ruiz Palacios et al found that for successful infection, *Campylobacter* tethers itself to the intestinal mucosa by using epithelial cell surface H blood group (fucosyl α1,2) epitopes as ligands to attach to and invade host cells[42]. Further studies by the group confirmed the extent of *Campylobacter* binding to BgAgs - some strains bound to several related fucose-containing BgAgs but bound most strongly to H ligands.[109] Using Elisa binding studies, Mahdavi et al further confirmed *C. jejuni* binding to BgAg core-I, core-II, H-II, Le^b , Ley and Lex BgAgs. Retagging was used for the identification of *C. jejuni* BgAg-binding adhesins, which were found to be MOMP and the flagellar protein, FlaA[110].

2.9 Major Outer Membrane Protein (MOMP)

Gram negative bacterial outer membrane proteins interact with pattern recognition receptors[111], act as targets for neutralizing bactericidal antibodies[112] and consist of epitopes presented to T cells[113]. Porins are usually the most abundant outer membrane proteins and serve as molecular channels that mediate solute diffusion across membranes whilst also playing a role in bacterial pathogenesis, such as adherence, invasion, and serum resistance[114]. Furthermore, porins can induce protective immune responses against bacterial infections by eliciting bactericidal antibodies and T cell responses[115]. So far, three channel forming proteins have been identified in *C. jejuni*; the major outer membrane protein (MOMP), Omp50 and CadF [116–119]. Unlike *E. coli*, which has two major porins, OmpC and OmpF, *C. jejuni* only has the one major porin (MOMP) through which nutrients and antibiotics transit and thus this report focuses primarily on MOMP. MOMP, a 135 kDa protein has the tertiary structure typical of a β-barrel porin, as suggested initially by its primary sequence homology comparisons and confirmed by CD spectroscopy showing the predominance of β-strands, and more recently by a crystal structure (see below). It is a trimeric 18-stranded antiparallel β-barrel porin with an elliptical shape common to other 18-stranded porins. The protein has a central pore size and conductivity intermediate between *E. coli* OmpC and OmpF.[120] Experiments by Moser *et al.* demonstrated the role of MOMP in adhesion of *C. jejuni* to INT-407 cells. Using a ligand-binding immunoblotting assay, the flagellin, the major outer membrane protein and a 59-kDa protein were detected to be involved in adhesion to INT-407 cells and the extracellular matrix protein fibronectin.[121] These observations make MOMP an attractive ligand target for developing new antimicrobials, as MOMP blockers could potentially impair the ability of *C. jejuni* to establish the initial stages of infection. Structural and biophysical data are highly desired and the structure of MOMP from *C. jejuni* strain 85H (a human isolate) has recently been successfully determined by Ferrara *et al.* (**figure 2.2**)[120].

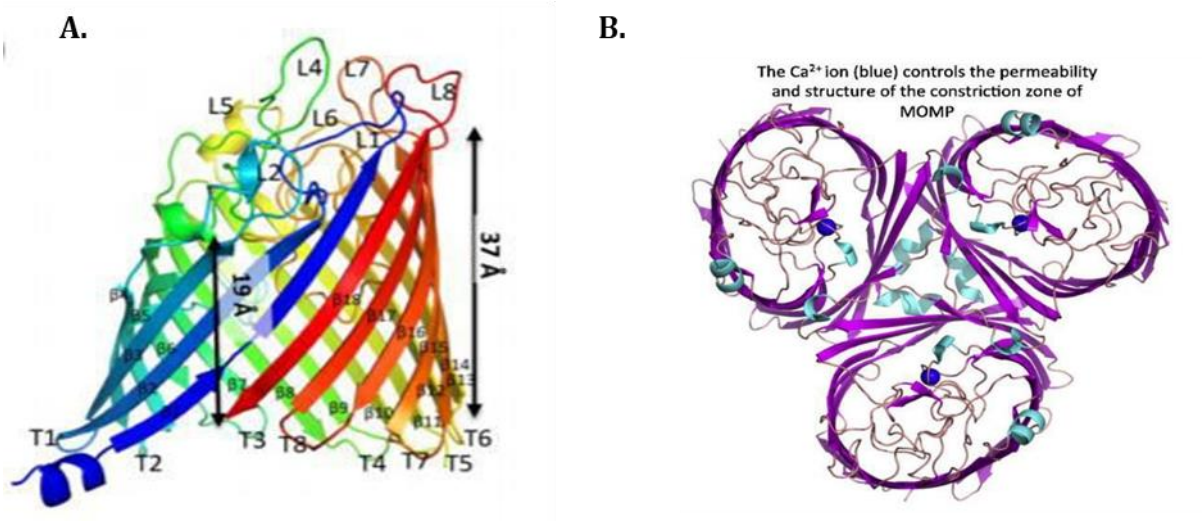


Figure 2.2: A. The structure of the MOMP monomer viewed from the side, perpendicular to the membrane. MOMP is an 18-stranded antiparallel β-barrel porin with an elliptical shape which is typical of 18- stranded porins. Extracellular loops connecting the strands are prefixed by L, whereas periplasmic loop connections are prefixed by T. **B.** The structure has been rotated 90° so that it is viewed from the outside of the cell (looking in)[120].

By analysis with MolAxis, Ferrara *et al.* determined the size of the constriction zone (or eyelet) of *C. jejuni* MOMP to be 2.6 Å[120]. This is an important consideration to take into account when designing inhibitory compounds that will target the MOMP porin.

The structure of MOMP is similar to that of *H. pylori* BabA binding adhesin. It contains two domains: an *N*-terminal extracellular host-binding domain and a C-terminal outer membrane–spanning domain predicted to also form a β-barrel structure similar to that of porins[122]. This trend is observed amongst outer membrane protein of other bacteria thar are also involved in adhesion to host cells. OmpX , OmpA and PagP of Uropathogenic *E. coli* (UPEC) have a minimal number of eight strands[123].

Conserved multi-subunit molecular machines fold and insert proteins such as MOMP into the outer membrane. One subunit of the machines is itself a β-barrel protein that has a central role in folding other β-barrels. In Gram-negative bacteria, the β-barrel assembly machine (BAM) consists

of the β-barrel protein BamA, and four lipoproteins which accelerate folding without using exogenous energy (for example, ATP)[124]. To prevent misfolding, the outer membrane protein precursor interacts with periplasmic chaperones to avoid aggregation of the protein in the aqueous periplasmic environment. The solubility of the chaperones allows the nascent protein to be escorted to the outer membrane where the BAM complex inserts the nascent polypeptide chain into the outer membrane[73].

2.10 Role of protein glycosylation on MOMP functionality

Glycosylation is the reaction in which a carbohydrate (also known as a glycan) is attached to a hydroxyl or other functional group of another molecule by enzyme transferases. Glycosylation is a posttranslational modification that is found on about 50% of all proteins, rendering it the most abundant polypeptide chain modification in nature. In particular, the modification occurs on secreted and transmembrane proteins of eukaryotes, archaea and to a lesser extent in prokaryotes[125]. Glycosylation has a number of important roles to play such as in protein folding and stability, pharmacokinetics, bioactivity, secretion, in-vivo clearance, solubility, recognition and antigenicity to name a few. The two main modes of glycosylation are *N*-linked glycosylation and *O*-linked glycosylation. *N*-linked glycans are attached to a nitrogen of asparagine or arginine side-chains whilst *O*-linked glycans are attached to the hydroxyl oxygen of serine, threonine and tyrosine side-chains. Although the *N*-linked glycosylation process occurs very rarely in bacteria, the first bacterial *N*-linked protein glycosylation pathway (Pgl) was described in *C. jejuni* [126], and more than sixty-five *C. jejuni* proteins of varying functions have been shown to be *N*-glycosylated [127]. Glycosylation increases diversity in the proteome, because almost every aspect of glycosylation can be modified, including the site of glycan linkage (glycosidic bond), the types of sugars that are linked to a given protein (glycan composition), the structure of the glycan

(branched or unbranched sugar chains) and the length of the glycan (can be short- or long-chain oligosaccharides).

Glycosylation of proteins, especially outer membrane proteins have a number of effects on the protein structure and functionality. Solá and Griebenow performed *in silico* structural and energetic analysis on a series of chemically glycosylated α-chymotrypsin conjugates that showed that glycosylation increases protein solubility due to an increase in the number of possible interactions between the glycoprotein surface and the surrounding solvent molecules due to an overall greater molecular solvent accessible surface area (SASA) caused by the presence of the glycans[128]. When proteins are exposed to extremes of pH, such as in the stomach, it can result in loss of structure by disruption of both internal electrostatic forces and charge-charge interactions. Masárová through the glycation of penicillin G acylase, showed that the half-life for the glycated version of this protein was increased 13-fold at pH 3 and 7-fold at pH 10 when compared to the non-glycated protein[129]. Again, Solá and Griebenow demonstrated that this type of stabilization occurs due to an increase in the internal electrostatic interactions of the protein as a result of glycosylation[128]. Further, results suggest a mechanism in which protein aggregation is prevented due to an increase in steric repulsions between aggregation-prone protein species due to the presence of the glycans on the protein surface[130].

In *C. jejuni,* glycosylation increases diversity of the proteome using both *N*-linked and *O*-linked glycosylation. With *O*-linked glycosylation, glycans can constitute up to 10% of the protein mass [131]. Most of the *O*-glycans attached to the *Campylobacter* flagellum are derivatives of pseudaminic acid (Pse) or legionaminic acid (Leg), which are C9 sugars that are related to sialic acids[132]. While it is estimated that more than two-thirds of all eukaryotic proteins would be glycosylated, no such estimate is available for prokaryotic glycoproteins, the study of which is lagging behind, mainly due to the enormous variability of their glycan structures and variations in the underlying

glycosylation processes[133]. They display many commonalities as well as unexpected variations. Sequence analysis, targeted mutation and chemical analysis have elucidated proteins involved in *O*-linked glycosylation system and their respective biochemical functions. The flagellin of *C. jejuni* strain 81-176 is glycosylated with pseudaminic acid at around 19 sites, which accounts for 10% of its observed mass. Thibault et al showed that flagellins from three strains of *C. jejuni* and one strain of *C. coli* were shown to be extensively modified by glycosyl residues, due to a 6000-Da shift from the molecular mass of the protein [131]. A specific recognition sequence for *O*-linked glycosylation has not been identified, and the addition of the glycan is thought to require surface exposure and hydrophobicity.[131] *O*-glycosylation of flagella is important for motility, adherence and invasion of host cells, but it is not yet known whether *O*-linked glycosylation has any role in immune avoidance or the host-cell interaction.[134] Recently, MOMP in *C. jejuni* was shown to be *O*-glycosylated, and the glycosylation is important for binding human BgAg and biofilm formation. Mahdavi et al showed that a novel *O*-linked glycan modulates *C. jejuni* major outer membrane protein (MOMP)-mediated adhesion to human BgAgs and chicken colonization. Calculations show that the glycosylated protein has more favourable van der Waals interactions compared with the non-glycosylated protein contributing to the effects auto-agglutination, biofilm formation, adherence to Caco-2 cells and chicken colonization. MS/MS analysis by Mahdavi et al showed that a peptide corresponding to amino acids 268–278 of NCTC11168 MOMP protein was glycosylated; specifically, the amino acid Thr268MOMP. Lectin binding was used to characterise the *O*-glycan attached to NCTC11168 MOMP. MOMP showed significant binding to Jacalin, a lectin that exhibits specificity for human tumour- specific Thomsen-Friedenreich disaccharide (T-antigen; Gal(b1– 3)-GalNAca1-Ser/Thr)[110]. Inactivation of the oligosaccharyltransferase-encoding gene, *pglB*, had no impact on *C. jejuni* binding to the aforementioned BgAgs indicating that the *N*- glycosylation system does not influence BgAg binding. On the other hand, mutation of *pseD*, which encodes a putative PseAm transferase, resulted in a significant reduction in the binding of *C. jejuni*

NCTC11168 to the BgAgs, signaling that *O*-glycosylation of BgAg ligands is required for optimal BgAg binding. This signals that *O*-glycosylation of *C. jejuni* is required for optimal binding to BgAg[110].

2.11 Transmission of antibiotic resistant *Campylobacter* from animal to humans

Campylobacter is a zoonotic agent and there is overwhelming evidence that fresh raw meat such as poultry is the main source of infection. Studying the transmission of antibiotic resistance from animal to human is complex due to the complex chain of transmission. However, some studies show that the number of macrolide-resistant and fluoroquinolone-resistant isolates from humans is influenced by factors including veterinary use of macrolides (which are approved for use as antimicrobial growth promoters or therapeutic drugs) and fluoroquinolones (only approved for use as therapeutic drugs).[135] Other studies suggest that resistance in *C. jejuni* is driven by human use of antibiotics to treat infections rather than by veterinary use of antibiotics. Fluoroquinolones were widely used in human medicine in the Netherlands and the US before they were approved for animal use. Prior to animal use there was no emergence of quinolone resistance in *Campylobacter* in humans. While human macrolide and fluoroquinolone use contributes to the increase in resistance in humans, their relative contribution to increase in resistance compared to the use of these agents in husbandry appears to be small[49].

2.12 Alternative treatments to reduce *Campylobacter* colonisation in chickens

While global efforts to control the transmission of enteric pathogens have been effective at reducing the incidence of a number of major foodborne pathogens, the prevalence of *Campylobacter* infection has continued to increase across most developed nations. The increased

use of antibiotics is rendering infection control increasingly inefficient and the rise of antibiotic resistance signals towards a need in investment for the development of other types of intervention. Novel antibiotics directed against antibiotic-resistant *Campylobacter* can be developed when extensive funds are committed for research but the pathogens ultimately become resistant to such drugs over time. To break this vicious cycle, it will be necessary to adopt chemotherapy-independent remedial strategies to combat bacterial infections. The key areas of the strategy should be to increase resistance of chickens to carrying campylobacter, reducing environmental exposure to *Campylobacter* and reducing/removing *Campylobacter* colonisation of chickens. Below I will review a number of strategies that are already in development.

2.12.1 Vaccine therapy

Despite the fact that MOMP is immunogenic and the recombinant MOMP (rMOMP) can be readily produced in *E. coli*, the nature of the antibody response to MOMP during *in vivo* infection is not well understood which has led to lack of research for a vaccination programme against *Campylobacter* in poultry. For designing vaccines for human use, recombinant MOMP (rMOMP) free of LOS is preferred, because the LOS antigen in *Campylobacter* could trigger an autoimmune response leading to the Guillain–Barré syndrome [136]. Although the sequence and structural features of *Campylobacter* MOMP have been well characterized, the nature of antibody responses to MOMP induced during natural infection is unknown and needs to be examined to facilitate the design of MOMP-based vaccines. Studies by Huang *et al.* (2007) found that sera from naturally or experimentally infected chickens reacted weakly with denatured rMOMP, but strongly with rMOMP reconstituted in proteoliposome, suggesting that the chicken antibody response to MOMP is predominantly directed against conformational epitopes and that proteoliposomes could be an effective strategy for anti-*Campylobacter* vaccination[115].

2.12.2 Bacteriocins

Bacteriocins (BCN) are low molecular weight peptides that are produced in bacterial ribosomes and possess antimicrobial properties. The initial focus of Stevoch and Sterns cooperative studies were to identify and use non-pathogenic enteric bacteria that would compete with or preferably inhibit the colonisation of *C. jejuni* in the gut of commercial broiler chicken. They failed to identify bacteria (acting as probiotics) that could successfully compete with *C. jejuni*, instead they found that by culturing selected microbial antagonists isolated from broilers , they could produce specific BCN that were highly effective *in vitro* against *Campylobacter* spp.[137, 138]. This was proved when purified BCN were provided in feed or water and led to the elimination or significant reduction of *C. jejuni* from infected chickens.

BCN are peptides of cationic, hydrophobic, or amphiphilic nature, with molecular weights of 5 to 6 kDa. Furthermore, mature peptides usually have 20 to 60 amino acids. BCN are their own defined set of antimicrobials and should not be confused with antibiotics. They have several characteristics that set them apart from antibiotics (some of which give them an advantage over the use of antibiotics). BCN are produced on the surface of ribosomes of microbial cells, whereas antibiotics are secondary metabolites of the cell; producers of BCN can tolerate the presence of the peptides that they have produced as they are not susceptible to the bactericidal effects; BCN can attach to any part of the cell as they do not need specific receptors on the cell walls of the microbes; the mechanism of action of BCN are varied but associated with the inhibition of pore formation in the outer cell membrane. This occurs by binding with bacteria on cell walls of susceptible microbes, interacting with the outer membrane, and thus pore generation. Inorganic ions leak through the pores, thereby resulting in the death of the target cell. Similar to the development of bacterial antibiotic resistance which is determined by changes in genetic determinants that are passed on

from one cell generation to the next, resistance of a bacterial target cell to a BCN develops when the target alters its cell membrane chemical composition [139, 140].

BCN have a few clear advantages over antibiotics, they are highly susceptible to proteases and should degrade in the host, meaning that they are not toxic to humans and animals as they cannot accumulate in the host. Secondly, BCN are equally effective against antibiotic resistant pathogens which is very important for stopping the spread of bacterial drug resistance, especially within hospital acquired infections.[141] Stevoch and Stern have carried out experimental studies of the effect of the bacteriocin L1077, produced by the bacteria *Lactobacillus salivarus* strain L-1077 isolated from broiler chickens. The treatment of chickens with this bacteriocin resulted in decrease of *C. jejuni* counts of >x4log compared with the control birds. This finding in association with other similar studies points towards bacteriocins as promising alternatives to antibiotics [137, 138, 142].

2.12.3 Bacteriophages

Bacteriophages are also being developed as an alternative method for biocontrol of Campylobacter. Bacteriophages are viruses that can lyse bacterial cells, a function which renders them a potential alternative therapy for bacterial infections. In the 1920s Félix d'Herelle co-discovered phage therapy but it was dismissed by western countries in favour of antibiotics in the 1940s.[143] There remained reluctancy to use phage therapy due to the unreliable early trials, but still the technique received attention in the USA. William Smith *et al.* reported the successful use of phages against *E. coli* in mice. In recent times, bacterial multidrug resistance has motivated the scientific community to revaluate phage therapy for bacterial infections that are incurable by conventional chemotherapy[144]. The effect of phage therapy in the control of *C. jejuni* colonization in young broilers as a preventive or therapeutic measure was tested by Wagenaar *et al*.[145] Bacteriophages are highly specific to a particular bacterial species or strain, and multiply at the expense of the cell, eventually reducing the number of viable bacterial cells without affecting

other microbial populations due to specificity. The use of bacteriophages to reduce *Campylobacter* load on chicken skin after slaughter was investigated but it was found that activity of bacteriophages is maximal at the optimal growth temperature of their host which is 40 °C in the case of Campylobacter. An important observation in the first set of experiments was the initial 3xlog reduction in bacterial load directly after introducing the phage. This indicates that careful timing of phage administration one day before slaughter might result in maximum effect by the time the birds are harvested [145]. The optimal phage dose that is most effective for reduction of *Campylobacter* counts needs to be determined, but nevertheless, this was the first report of phage therapy to decrease the colonization levels of an apparent commensal *C. jejuni*, in chickens[146].

Similar results were seen in another study by Fischer where commercial chickens were inoculated with 104 CFU of a *C. jejuni* field strain[146]. Three groups of birds (n=88 per group) were treated with either a single phage, a four-phage cocktail or solvent only. Treatment with the single phage and the phage cocktail permanently reduced *Campylobacter* with a significant reduction being observed between 1 and 4 weeks post treatment. Initial resistance rates of up to 43% in those treated with the single phage were observed, whereas a resistance rate of 24% was observed in birds that had received the cocktail. These resistance levels later stabilized at low levels[146]. There are distinct advantages of using bacteriophages over antibiotics: (1) Bacteriophage use is effective against multidrug-resistant pathogenic bacteria because of the mechanisms by which it induces bacteriolysis; (2) it has high specificity for targeting specific bacteria spp.; (3) it can respond rapidly to the appearance of phage-resistant mutants because the phages are able to mutate and bypass resistance; (4) the cost of developing a phage system is cheaper than that of developing a new antibiotic[147]. Overall, more work will be needed to improve the efficiency of phage therapy, to explore the large scale propagation of phages on pathogenic hosts and to test the robustness of the method with chicken management systems and with Campylobacter strain variation.

2.12.4 QPLEX

One technology proposed by Soultanas *et al.* is a method for inhibiting, reducing, or preventing biofilm formation or build-up on a surface. This work has led to the development of the QPLEX/TYPLEX compounds and associated iron chelates, which do not have bactericidal or bacteriostatic activities, but are potent inhibitors of biofilm formation in different bacterial species. These compounds may be used in the treatment, inhibition of growth, and prevention of colonization by bacteria, both in biological and non-biological environments; disinfecting surfaces, potentiating the effects of antibiotics and other anti-microbial agents, and increasing the sensitivity of bacteria and other microorganisms, to antimicrobial agents. Furthermore, this technology can induce changes in bacteria morphology by way of changing their metabolism. This technology is patent protected and is focused on improving animal growth by ingesting effective amounts of iron chelates, as illustrated in **figure 2.3**[148].

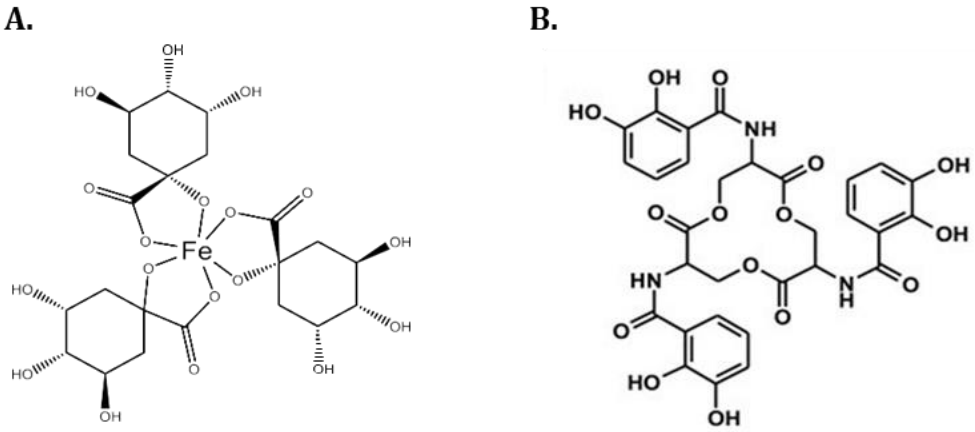


Figure 2.3: A. The structure of the synthesized QPLEX, a hexadentate octahedral that can interfere with bacterial colonisation[148]. **B.** The structure of the siderophore, a small, high-affinity iron-chelating compound that is secreted by bacteria and serves primarily to transport iron across cell membranes. There is a clear structural similarity between siderophores and QPLEX, the potency of which has been elevated by incorporation of the iron-binding functional group of the siderophore.

The structure of QPLEX is based on a similar structure to siderophores. Bacteria secrete siderophores, which are high-affinity extracellular ferric chelators to solubilise iron prior to

uptake.[149] The solubility of ferric iron is 10^{-17} M at pH 7 providing only 6×10^{6} iron atoms per litre whilst bacteria require 10^{-7} to 10^{-5} M iron to achieve optimal growth. Bacteria can close the 'concentration gap' between ferric iron solubility and iron requirement by solubilizing ferric oxides by way of chelation and reduction. Within gram-negative bacteria, uptake of ferri-siderophores occurs via specific outer membrane receptors in a process that is driven by the cytosolic membrane potential and mediated by the energy-transducing TonB-ExbB-ExbD system (**figure 2.4**). Periplasmic binding proteins then shuttle ferri-siderophores from the outer membrane receptors to cytosolic membrane bound ATP- binding cassette (ABC) transporters that, in turn, deliver the ferri-siderophores to the cytosol where the complexes are dissociated by reduction[149]. The design of QPLEX considers the high affinity, siderophore-mediated iron uptake system of bacteria.

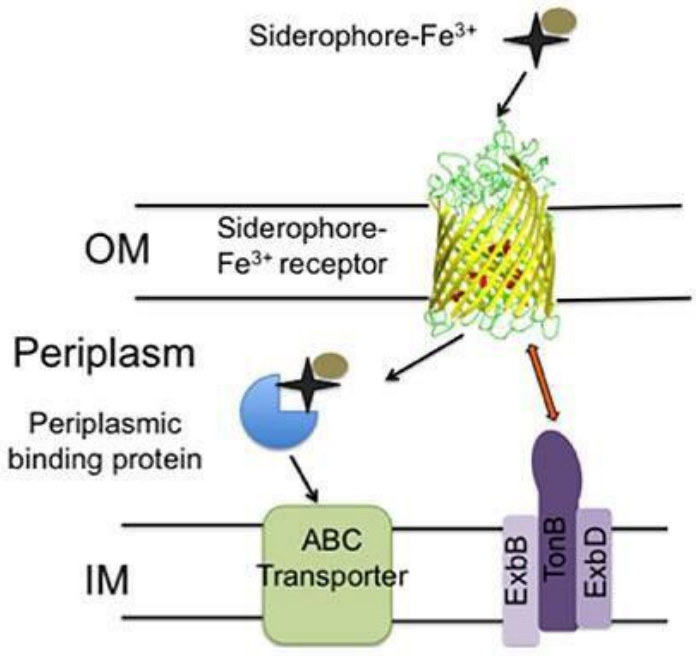


Figure 2.4: Structural basis for the uptake of siderophore-antibiotic conjugates in Gram-negative bacteria by Ferrara[150].

2.13 Conclusion

Campylobacter is the most common bacterial cause of human gastroenteritis in the world for both developed and developing nations. Infections are generally mild, but can be fatal among very young children, elderly, and immunosuppressed individuals. Poultry is the largest contributor of the disease where the commensality of the gut colonisation in avian birds has no major outcome, but has contributed to the increasing infection in humans, where symptoms persist. *C. jejuni* can further form attachments with human BgAgs which can lead to biofilm formation and this is, at least in part, caused by one of its virulence factors, MOMP. Protein glycosylation of MOMP has an increased effect on virulence and research has begun into how to reduce the incidence of colonisation in the animal system citing BCN, bacteriophage, vaccine and iron compounds such as QPLEX as some clear candidates.

Chapter 3: **MOMP purification**

3.1 Introduction

Attempts were made to purify four variants of the MOMP protein, of which three were used for further studies. In this chapter the cloning, expression, purification, and conformational structure analysis are described for each protein.

3.2 Cloning of MOMP NCTC1168 and introduction into *E. coli*

To facilitate the purification of high purity and yield MOMP protein, a recombinant gene was created for heterologous expression and purification from *E. coli*. A gene construct encoding MOMP from *C. jejuni* NCTC11168 was synthesised with codons optimised for heterologous *E. coli* expression, enabling expression of a 6-His tagged MOMP (methodology). The fragment carried unique NcoI and HindIII cleavage sites at the 5'and 3'-ends, respectively. Codon optimised *momp* gene of *C. jejuni* NCTC11168 was ligated into the pTAMAHisTEV vector to create a plasmid of 6.8 Kb. After transformation into *E. coli* TOP10F cells, the DNA from transformants was purified and digested. The digests were run on a 0.8% w/v agarose gel alongside digested but empty pTAMAHisTEV and a 1 kb ladder (NEB) to check that ligation was successful. DNA sequencing confirmed the correct insertion of *momp* gene in one of the clones, which was named pJO1.

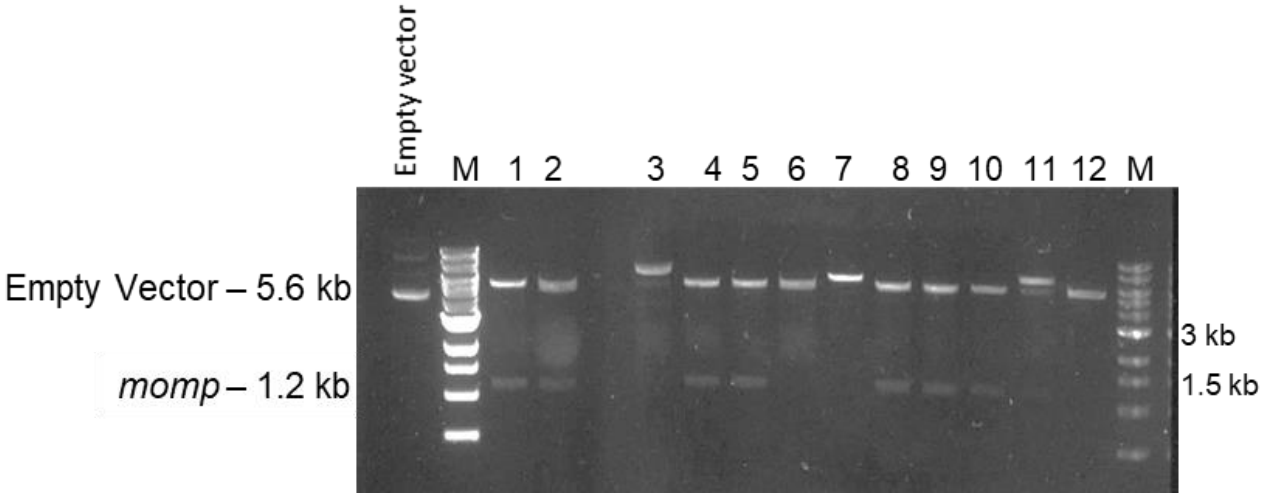


Figure 3.1: Agarose gel showing DNA bands after digestion of putative pJO1 transformants with NcoI and HindIII restriction enzymes. Twelve colonies were isolated, miniprepped and digested. DNA fragments of 5.6 kb and 1.2 kb were present, indicating the empty vector and the *momp* gene insert, respectively. Empty vector DNA was loaded as a reference. NEB 1 kb DNA ladder is shown in the lanes marked M. Two prominent size markers of 3 and 1.5 kb are shown on the right-hand sight of the gel for clarity.

3.2.1 Purification of recombinant MOMP (rMOMP) protein

The pJO1 plasmid was transformed into the Lemo21 (DE3) *E. coli* expression strain which was used to harvest the rMOMP protein. **Figure 3.2** shows that after growth of cells in LB to 0.6 OD_{600} and induction of protein expression with 1 mM IPTG, overnight growth at 25°C yielded rMOMP protein in the soluble fraction.

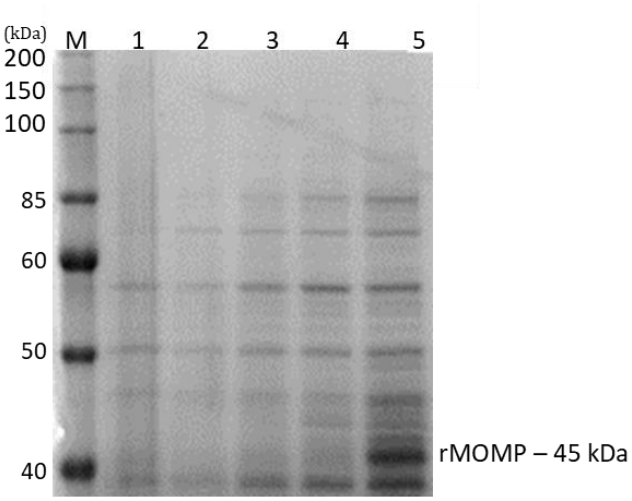


Figure 3.2: SDS PAGE gel loaded with soluble fraction of cell lysate of Lemo21 (DE3) transformed with pTAMAHisTEV MOMP after small scale expression. Lane 1 – uninduced, lane 2 – 1 hr post-induction, lane 3 – 2 hr post-induction, lane 4 – 3 hr post-induction, lane 5- 16 hr post-induction. Overexpression of rMOMP (MW 45 kDa)

is evident in lane 5- 16 post-induction hr as indicated by the label next to lane 5. Lane M indicates MW marker (NEB Unstained Protein Standard, Broad Range, 10-200 kDa).

rMOMP was purified from the soluble bacterial extract using a His-trap column. However, when the same expression conditions were scaled up for a larger scale expression, MOMP was found to be predominantly in the insoluble fraction (**figure 3.3**). This may be a result of formation of inclusion bodies which in bacterial hosts poses a major challenge for large scale recovery of bioactive proteins. Studies, by Singh *et al*., of purification of protein from inclusion bodies show that there is a high chance of recovery and refolding of protein from when using mild solubilization agents such as Elugent, Octyl -POE and DMM[151].

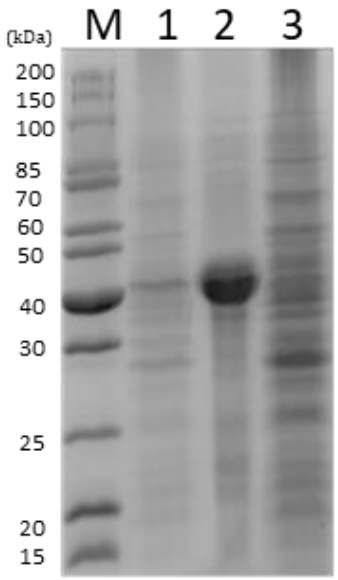


Figure 3.3: Large scale over-expression of rMOMP. Lane 1, whole cell lysate; Lane 2, insoluble fraction; Lane 3, soluble fraction. During large scale over-expression, rMOMP was expressed in the insoluble fraction, as can be seen in lane 2. Lane M indicates MW marker (NEB Unstained Protein Standard, Broad Range (10-200 kDa)).

The subsequent use of Elugent during the purification of the protein allowed solubilisation of the protein from the insoluble fraction. **Figure 3.4** represents a side-by-side comparison of the purification profile of MOMP in the absence (lanes 1 to 3) and presence (lanes 4 to 7) of mild detergent. Firstly, the cell pellet was sonicated to release the component of the cell into a lysate and the cellular debris was removed by removal of the supernatant after centrifugation at

10,000*g*. The lanes in **figure 3.4** show the yield of MOMP after the initial centrifugation step. In lane 1 and 4 the whole cell pellet was solubilised with UREA (for the sake of visualisation only) to show the presence of MOMP in the in the insoluble fraction. Lanes 2 and 3 show the absence of MOMP in the supernatant in subsequent ultracentifugation steps without detergent. Lanes 5 and 6 show an increasing concentration of MOMP in the supernatant during subsequent ultracentrifugation steps after solubilisation of the insoluble fraction with 0.45% v/v Octyl-POE detergent.

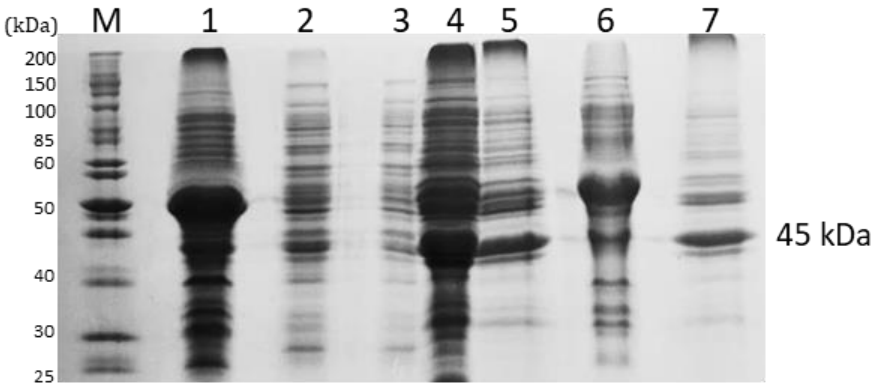


Figure 3.4: SDS PAGE gel of rMOMP protein purification profile from the insoluble fraction with and without the use of detergent**.** Without detergent) Lane 1- protein pellet post centrifugation shows that the protein is still in the insoluble pellet, lane 2- supernatant after ultracentrifugation 1, lane 3 - supernatant after ultracentrifugation 2. With detergent) lane 4- protein pellet post centrifugation shows that the protein is still in the insoluble pellet, lane 5 – protein solution after overnight solubilisation of insoluble fraction, lane 6 – supernatant from ultracentrifugation of the solubilised fraction, lane 7 - supernatant from 2nd ultracentrifugation contains solubilised protein (highlighted). Lane M indicates MW marker (NEB Unstained Protein Standard, Broad Range, 10-200 kDa).

Solubilised rMOMP was obtained after the second round of centrifugation and purification was attempted using a His-trap column (**figure 3.4**). However, rMOMP appeared mostly in the flow-through suggesting that it was not binding to the His-trap column despite the presence of an *N*-terminal His tag. There may be several reasons for the inability of His tagged rMOMP to bind to the His-trap column. The presence of imidazole in the binding buffer, albeit at low concentration, may outcompete rMOMP binding to the column, especially as rMOMP is a very hydrophobic protein and it's His tag may not be well exposed to interact with the matrix of the His-trap column.

It may also be that detergents used for the solubilisation of the protein and to keep the protein in a micelle environment may prevent binding of the His tag to the His-trap column.

Despite very low concentrations of rMOMP in the flow-through fractions, these fractions were pooled and an attempt was made to purify rMOMP by size exclusion chromatography, producing rMOMP protein of very low purity and yield. The purification of functional proteins from insoluble bacterial fractions and inclusion bodies often leads to difficulties with refolding and does not result in full recovery of biological activity. Given these difficulties we have encountered in purifying rMOMP from heterologous expression in *E. coli*, we decided to purify native MOMP directly from *C. jejuni*.

3.3 Purification of native MOMP from *C. jejuni*

The three strains of *C. jejuni* NCTC11168 that were used in this project are wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ which were kindly gifted by Jafar Mahdavi[110]. Experimental details for the creation of these three *C. jejuni* NCTC11168 strains are given in reference 64.

3.3.1 Large scale growth and cell collection

After confirmation of the *C. jejuni* NCTC11168 wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ strains we proceeded to grow bacteria on a large scale to obtain sufficient biomass for native MOMP protein purification. Glycerol stocks were plated from a single colony and expanded to a lawn, using the cotton swab and lawn techniques. This was expanded to fifteen further plates and then to five hundred plates in total. Five hundred plates were used per strain and every 10 plates yielded approximately 1 g of cell pellet (biomass).

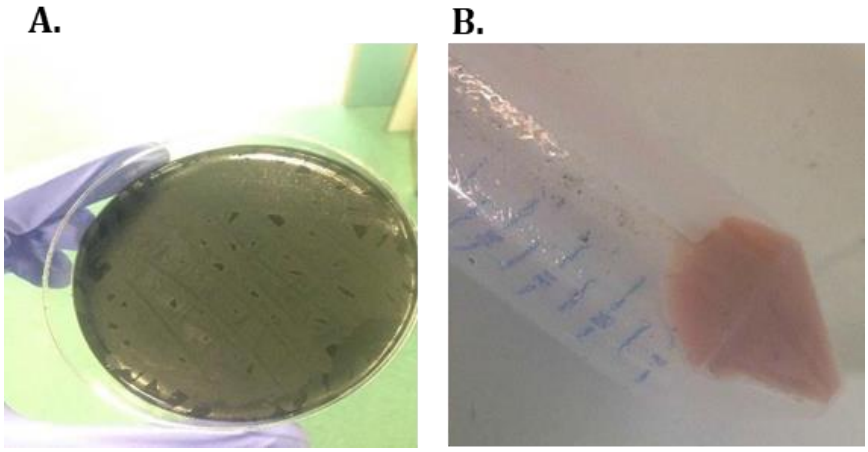


Figure 3.5: Generation of *C. jejuni* biomass for native MOMP extraction and purification. **A.** An Oxoid *Campylobacter* selective blood free agar (CCDA) plate was streaked with 10 µL of *C. jejuni* 11168 glycerol stock using a sterile loop and cultured under microaerophilic conditions for 48 hr at 42° C. A single colony was then isolated and expanded further. **B.** After a series of centrifugation steps (more detail in the methods section) the cell pellet was collected and stored at -80 °C.

Soluble MOMP was extracted from the cell pellet by sonication followed by removal of the cell debris and centrifugation to form a proteinous pellet. This protein pellet was solubilized with sodium sarcosyl and subsequent isolation of the outer membrane was performed by multiple rounds of ultracentrifugation of the solubilized pellet to separate the outer membrane from the inner membrane of the supernatant at 100,000*xg.* Soluble MOMP was identified by SDS PAGE and loaded onto a MonoQ HR anion exchange column (GE Healthcare) equilibrated in buffer A (30 mM Na_2HPO4, 10 mM NaCl, and 0.6% v/v Octyl-POE). The column was washed with five times column volume of buffer A and then bound proteins were eluted with a stepwise (5, 12, 20, 70, and 100%) gradient of buffer A supplemented with 1 M NaCl (Buffer B). Chromatography was carried out using an ÄKTA Explorer 10 system (GE Healthcare) (**figure 3.6**).

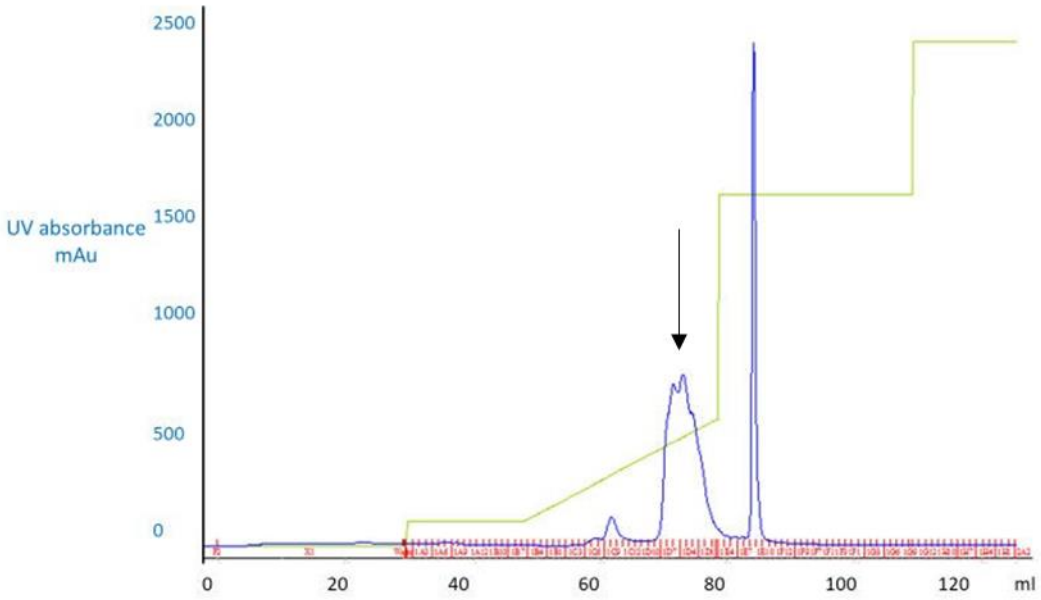


Figure 3.6: Anion exchange elution profile. Bound proteins were eluted stepwise with 5, 12, 20, 70, and 100% of buffer B supplemented with 1 M NaCl. Protein concentration in the effluent is relative to absorbance at wavelength A280 nm (blue). Buffer composition during the binding and elution phases is shown as percentage of buffer B (green). The fractions at which protein is eluted and the elution volumes are shown on the X-axis. Arrow points out major peak 1 which was pooled for further purification of MOMP after confirmation with SDS PAGE.

Samples from the fractions of the major peaks from anion exchange, peak 1 and 2, were examined by SDS PAGE to determine which peak MOMP was residing in. SDS PAGE revealed that MOMP was present in peak 1. Furthermore, SDS PAGE of peak 1 revealed that MOMP co-purified with another protein. **Figure 3.7** shows that after visualization of the eluted fractions with SDS PAGE, there were two bands, one at 45 kDa and one at 50 kDa. *C. jejuni* has a minor outer membrane protein, a single monomer of 50 Da[152], which corresponds to the size of the contaminating protein in **figure 3.7**.

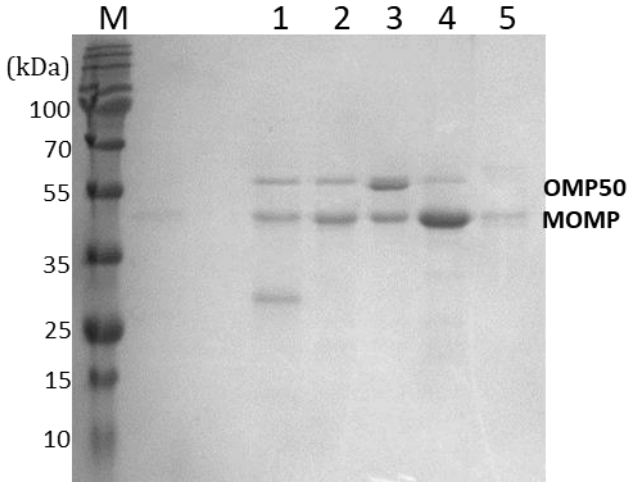


Figure 3.7: SDS PAGE analysis of samples from the major peak 1 from the anion exchange purification step. Labels 1-5 contain eluent that was eventually pooled for further purification with gel filtration. Lanes shows that MOMP was co purified with a contaminant, thought to be OMP50. Lane M denotes molecular weight marker, ThermoFisher Prestained PageRuler Plus Stained Protein Ladder.

Fractions containing MOMP were pooled but, because of the OMP 50 contaminant, further purification by size exclusion chromatography was necessary. Fractions containing MOMP were pooled, concentrated to 5 mL using a Vivaspin® 20 spin concentrator with a molecular weight cut-off of 10,000 kDa and loaded onto a Superdex 200 16/60 GL (GE Healthcare) gel filtration column (**figure 3.8**) equilibrated in 10 mM Tris HCl pH 8.0, 150 mM NaCl, and 0.45% w/v C8E4.

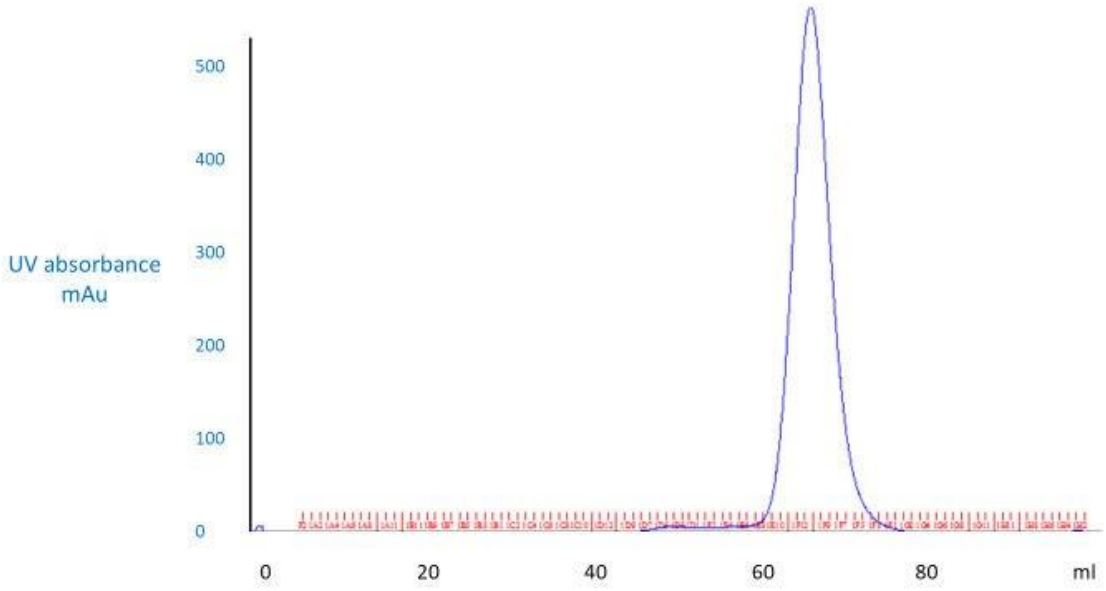


Figure 3.8: Size exclusion chromatography profile from the purification of MOMP. The collected fractions alongside the elution volume are shown on the X-axis and protein absorbance at 280 nm is shown on the Y-axis. Protein was eluted from seven fractions of 4 mL volume containing the eluent of the major peak and run on an SDS PAGE gel (**figure 3.9**).

Finally, fractions from the size exclusion chromatography containing MOMP were pooled and concentrated to 8mg/mL by use of a Vivaspin® 20 spin concentrator, followed by measuring absorbance at 280 nM by spectrophotometry. Using the following formula to calculate the molar concentration of the purified sample, where molar concentration = A280 x (dilution factor) / ϵ280, where MOMP ϵ280 = 189780 $M^{-1}cm^{-1}$, the molarity of the stock solution of the sample was 30 µM.

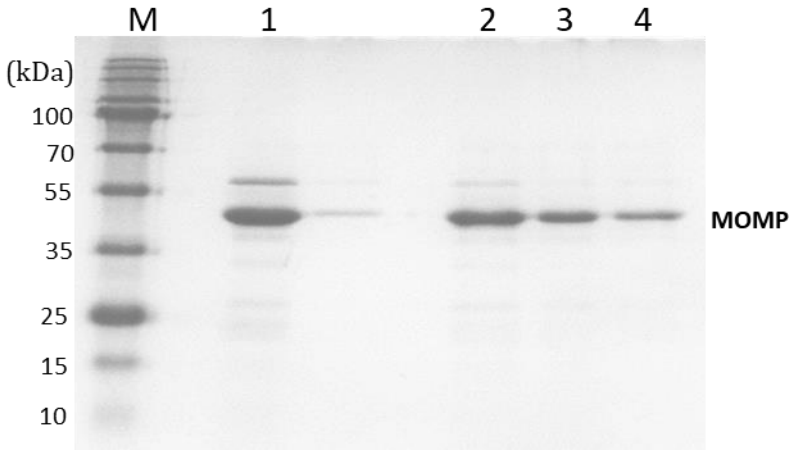


Figure 3.9: SDS PAGE analysis of fractions from the major elution peak of the size exclusion chromatography. Lane 1 shows sample eluted from height of the peak. Lanes 2-4 show eluted protein from tail of the peak to ascertain where purity of MOMP may be 100%. Although lane 4 looks to be pure, it contains lower concentration of protein, making it more difficult to visualize contaminating

proteins. Lane M denotes molecular weight marker, ThermoFisher Prestained PageRuler Plus Stained Protein Ladder.

Once the protocol was optimized for purification of the native MOMP protein, the same protocol was used for the purification of the MOMP$^{T/G268}$ protein giving a final yield of 10 mL of 25 µM protein whilst MOMP$^{\Delta pseD}$ was purified to a final volume of 8 mL at 20 µM.

3.3.2 Western blot of MOMP

Final confirmation of the identity of the purified proteins was performed by use of primary antibodies raised against MOMP. Further, as the use of antibodies were part of the assay design in downstream experiments such as in ELISA and the immunohistochemistry studies, we wanted to confirm whether or not all purified variants of MOMP (wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$) would still be recognized in downstream assays using the primary antibodies that were originally raised against wtMOMP (donated by Dr. Jafar Mahdavi). After SDS PAGE, protein was transferred from the gel to a nitrocellulose membrane for a western blot (**figure 3.10**). After blocking of the nitrocellulose membrane, the nitrocellulose membrane was incubated in the rabbit anti-MOMP primary, followed by incubation with a secondary antibody. BSA was used as negative control. The results in **figure 3.10** show that anti-MOMP primary antibody recognizes MOMP regardless of its glycosylation status. For wtMOMP that was conjugated to far-red (lane 5) and fluorescein (lane 6) fluorescent dyes, the same result could be seen where the primary antibody can still detect the protein regardless of conjugation to dyes. BSA protein was also present on the nitrocellulose membrane but no band appeared after incubation with the primary antibody, showing that the antibody is specific to MOMP variants.

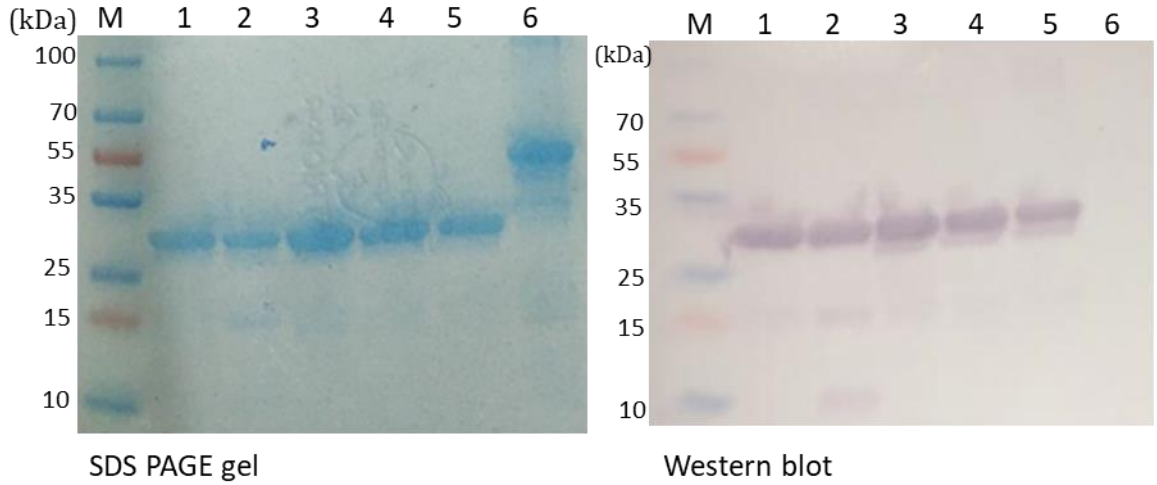


Figure 3.10: Western blot analysis of the wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ proteins. Proteins were resolved through an SDS PAGE gel (left panel) and blotted onto nitrocellulose for a western blot (right panel). Lane M - ThermoFisher Prestained PageRuler Plus Stained Protein Ladder, lane 1- wtMOMP, lane 2 – MOMP$^{T/G268}$, lane 3- MOMP$^{\Delta pseD}$. Further, wtMOMP conjugated to fluorescent dyes (for use in chapter 6) were also included in the gel; lane 4- MOMP^{flu} (wtMOMP comjugated to fluorescein dye), lane 5- MOMP^{farred} (wtMOMP conjugated to far-red dye) lane 6 – Control BSA.

3.4 Structural characterisation of the MOMP proteins

After purification of wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$, the proteins were characterized to confirm glycan decoration, correct folding, protein stability and secondary/tertiary/quaternary structure.

3.4.1 MOMP glycosylation

The presence of glycosylation was confirmed by SDS PAGE due to differences in electrophoretic mobility of glycosylated and non-glycosylated proteins. By determining the presence of a size difference between the glycosylated and non-glycosylated MOMP the presence of the glycosylation on the wild type was confirmed. Mahdavi *et al.* suggested that the MOMP glycan is likely to be the *O*-linked trimeric form of the T-antigen: Gal(β1–3)-GalNAc(β1–4)-GalNAc(β1–4)-GalNAcα1-Thr268 (MW 843.79 g/mol)[110]. The purified native wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ proteins were

resolved with a 6% SDS PAGE gel (**figure 3.11**). Such a low concentration of poly acrylamide gel was used to separate a small difference mass estimated to be less than 1 kDa.

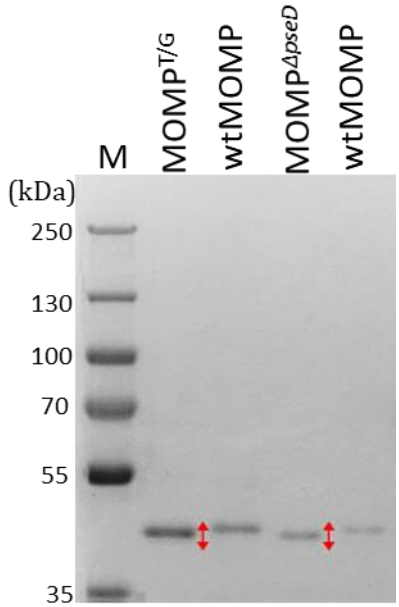


Figure 3.11: SDS PAGE of the wtMOMP, MOMP$^{T/G268}$ and MOMPΔ*pseD* native proteins resolved through an 6% polyacrylamide gel. Visible migration differences are observed between the wtMOMP and the MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ proteins as shown by the double-headed red arrows. Lane M denotes molecular weight marker, ThermoFisher Prestained PageRuler Plus Stained Protein Ladder.

SDS PAGE of purified native wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ proteins showed differences in their electrophoretic mobilities. The wtMOMP protein run visibly higher up the gel indicating a higher mass than the MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ proteins. This can be attributed to a glycan moiety present in the wtMOMP but not in the MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ proteins and is consistent with a previous suggestion that the MOMP glycan is to be the *O*-linked trimeric form of T-antigen: Gal(β1–3)-GalNAc(β1–4)-GalNAc(β1–4)-GalNAcα1-Thr268 thus conferring an increase in the mass of wtMOMP of 843.79 Da[110].

3.4.2 Tertiary structure

Confirming the tertiary structure of the purified wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ proteins is important because refolding during the purification process can cause mis-folding thus affecting

downstream experiments. One simple preliminary method of examining correct protein folding is native PAGE, where a protein's electrophoretic mobility depends on charge and shape[153].

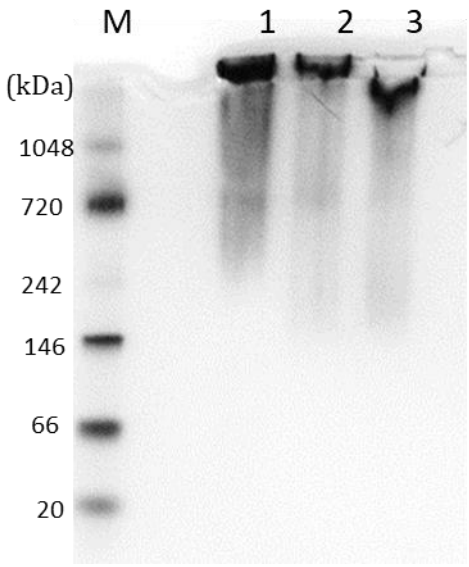


Figure 3.12: Native gradient PAGE of purified MOMP protein. Proteins were resolved through gradient PAGE gel 4-20% acrylamide. Lane 1 - wtMOMP, lane 2 - MOMP$^{T/G268}$ and lane 3 - MOMP$^{\Delta pseD}$. Lane M denotes molecular weight marker, ThermoFisher NativeMark Unstained Protein Standard.

Native PAGE can provide a non-disruptive, preliminary method for investigating the tertiary structure of small amounts of protein. But in the case of membrane proteins, native PAGE has typically been less reliable[154], as is the case for MOMP (**figure 3.12**). The protein bands can be seen at the top of the gel at the interface between the resolving gel and the wells. This indicates that the protein failed to travel through the pores of the gel, instead residing at the top of the gel. Due to its hydrophobicity, the protein may be aggregating at the top of the gel in the absence of SDS and detergent. Instead of native PAGE, a new method for visualizing proteins in their native state has been developed by Pollock *et al.* where a nano-encapsulation method based on styrene maleic acid lipid particles (SMALPs) can be used in combination with native gel electrophoresis to separate membrane protein complexes in their native state[154].

3.4.3 Characterisation of the secondary structure of purified native MOMP

Circular dichroism (CD) is a popular method for defining the secondary structure of proteins[155]. Here, CD spectrometry was used to compare the secondary structures of the wtMOMP and MOMP$^{T/G268}$ proteins. At the time of MOMP$^{\Delta pseD}$ purification, the CD spectrometer was unfortunately out of action so no CD spectroscopy data was collected for this protein. Because the spectra of proteins are dependent on their conformation, CD spectrometry can also be used to monitor conformational changes due to temperature, mutations, heat, denaturants or binding interactions[155]. CD was used to confirm correct protein folding and stability. A drawback with CD is that it is low-resolution and cannot provide high resolution structural information as protein NMR and X-ray crystallography can do. Briefly, CD relies on the unequal absorption of left-handed and right-handed circularly polarized light by proteins. When asymmetric molecules, like proteins, interact with light they may absorb right and left handed polarized light to different extents (hence the term CD)[155].

To compare the secondary structures of the wtMOMP and MOMP$^{T/G268}$, CD was carried out in two spectral regions; from 185 to 200 nm and from 200 to 250 nm which are sensitive to secondary structural perturbations.

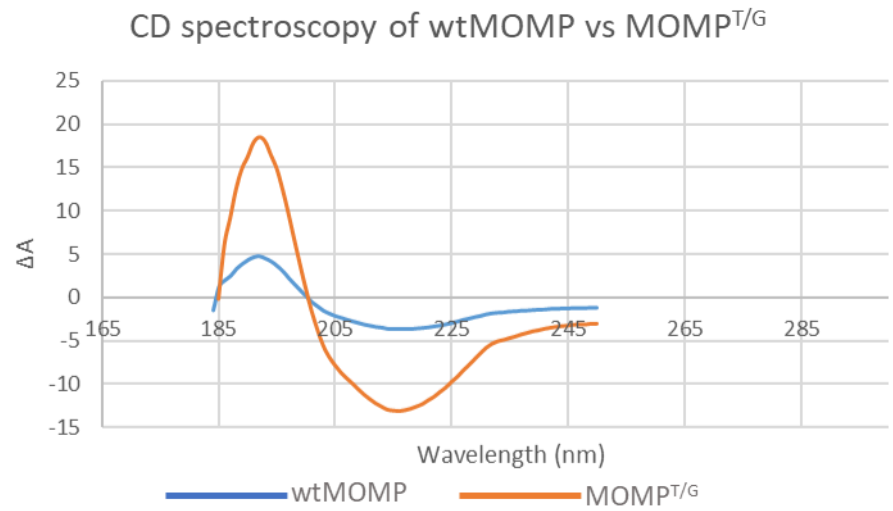


Figure 3.13: The CD spectra of wtMOMP and MOMP$^{T/G268}$ were obtained in their storage buffer (20 mM NaP, 0.45% v/v OctylPOE). The results were adjusted to take into account the CD of the storage buffer. Measurements were taken across wavelengths of 180 to 250 nm. The maximum ΔA of MOMP$^{T/G268}$ is 18 at 193 nm and the minimum is -13 at 215nm. The maximum ΔA of wtMOMP is 4.7 at 191 nm and minimum ΔA is -3.7 at 216nm. Protein concentration is 5 µM and path length is 1cm. ΔA (delta Absorbance) is the difference between absorbance of left circularly polarized (LCP) and right circularly polarized (RCP) light.

In the far UV-CD region the trends for both proteins show that of a typical beta sheet secondary structure. Proteins with well-defined antiparallel β-pleated sheets (β-helices) have negative bands at 218 nm and positive bands at 195 nm[155]. The difference in absorbance is greater for the MOMP$^{T/G}$, or, the difference in absorbance is decreased for the wtMOMP. These differences between the wtMOMP and MOMP$^{T/G268}$ proteins could be due to differences in masking of the protein, as the UV absorbance of saccharides should be negligible.

3.5 Conclusion

A recombinant gene was created for heterologous expression and purification of MOMP from *E. coli* which would reduce purification time compared to the purification of native MOMP directly from *C. jejuni.* Due to problems of insolubility of the protein caused by expression into inclusion bodies, an alternative purification pathway was used which would require less modification of the protein from the native state. wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ protein were successfully

obtained to high purity from native *C. jejuni*, after solubilisation with detergent. Following purification, the conformation of the proteins was confirmed using a variety of techniques to to confirm glycan decoration, protein folding, stability and secondary/tertiary structure.

Chapter 4: **Characterisation of MOMP glycosylation**

4.1 Introduction

Glycosylation is the process by which one or more monosaccharide or oligosaccharide units are linked to proteins or lipids. It is the most common post-translational modification (PTM)[156] and nearly half of all proteins in the SWISSPROT protein database are glycosylated[125]. When linked to a specific site on a protein backbone, glycans can modify the structural properties of the protein, affecting its solubility and conformation[157].

Glycoproteins often exist as heterogeneous mixtures of closely related structures, called glycoforms, which vary in their glycan structures and glycan attachment sites[158]. Molecularly, eight different amino acids are known to be involved in glycoprotein linkages with more than 13 different monosaccharides using more than 41 different chemical bonds[159]. Different glycan structures of equal weight (isobaric) can be linked in diverse structural isomers through a combination of multiple linkage positions, anomeric centers (i.e. α or β) and ring structures (i.e. furanose or pyranose)[159]. There are estimated to be about 10,000 different glycan substructures[160]. Although the site for *O*-glycosylation of MOMP has been identified as Thr268[110], the type of glycan is still to be fully confirmed, as *O*-glycans don't have a consensus sequence nor a consistent core structure and are extremely diverse, although it is known that *O*-glycosylation begins with the addition of a single *N*-acetyl-galactosamine (GalNAc) residue to the -OH of serine or threonine via an *O*-glycosidic bond by a glycosyltransferase, and the structures are named mucin *O*-glycans or *O*-GalNAc glycans[161].

4.2 Mass spectrometry of MOMP

The population of glycans occurring at a given site is often not homogeneous; a particular site of *N*- and *O*-glycosylation may be occupied by a number of structurally distinct glycans - a condition referred to as microheterogeneity or site heterogeneity[160]. This may also be true of the glycosylated MOMP and so the aim of this chapter was to explore the site of MOMP glycosylation and the units of saccharides that make up the glycosylation. Analysis of glycans or oligosaccharides has previously been performed by nuclear magnetic resonance whilst X-ray crystallography of glycans is relatively uncommon and limited to very simple structures. The major limitations of both methods are that they require pure samples and large amounts of material and so matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) has become a widespread analytical tool for peptides, proteins and most other biomolecules[160]. In MALDI analysis, the analyte is first co-crystallized with a large molar excess of a matrix compound, usually an ultraviolet (UV)-absorbing weak organic acid, after which laser radiation of this analyte–matrix mixture results in the vaporization of the matrix which carries the analyte with it. The matrix therefore plays a key role by strongly absorbing the laser light energy and causing, indirectly, the analyte to vaporize. The matrix also serves as a proton donor and receptor, acting to ionize the analyte in both positive and negative ionization modes, respectively. The efficient and directed energy transfer during a matrix-assisted laser-induced desorption event provides high ion yields of the intact analyte and allows for the measurement of compounds with high accuracy and subpicomole sensitivity. The ability to generate such accurate information can be extremely useful for proteoglycan identification and characterization[162]. The standard approach to determine site-specific glycosylation are to employ a combination of specific enzymatic proteolysis (usually with trypsin), fractionation of glycopeptides (most often by liquid chromatography or affinity chromatography)

and glycopeptide analysis by MS[160]. First MALDI-MS, a denaturing technique, was performed on wtMOMP and MOMP$^{T/G268}$.

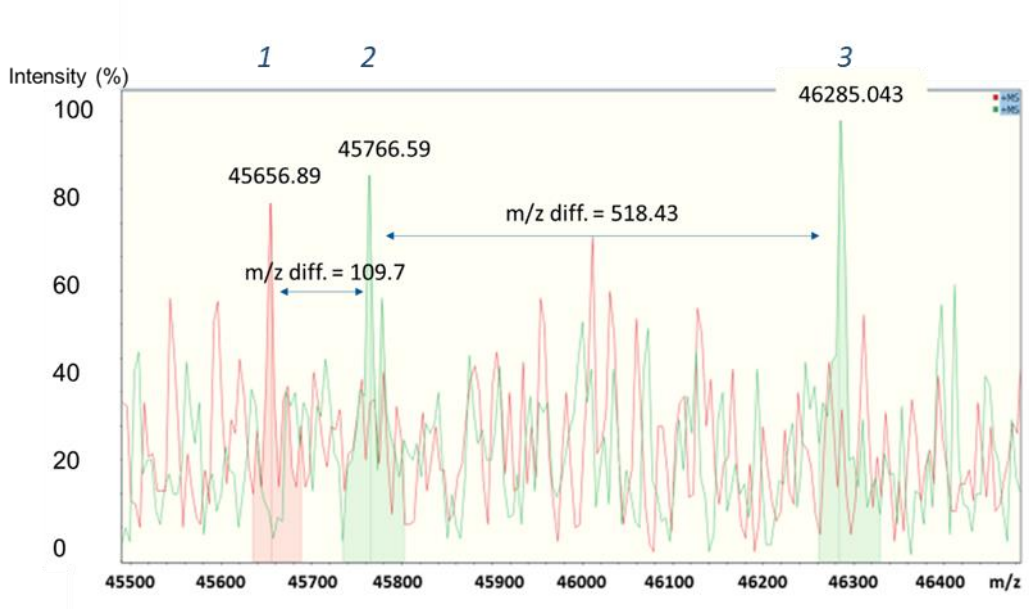


Figure 4.1: MALDI TOF mass spectrum of MOMP$^{T/G268}$ (red) and wtMOMP (green) overlaid over each other. Peak 1) m/z = 45656.89, peak 2) m/z = 45766.590, peak 3) m/z = 46285.043.

Peak 1 and 3 indicate that MOMP$^{T/G268}$ and wtMOMP have a m/z difference of 628.153, suggesting a glycan size of 628.153 Da is present in wtMOMP but not in MOMP$^{T/G268}$. Peak 2 and peak 3, both from wtMOMP, have a m/z difference of 518.43, this is likely due to hydrolysis of the glycan during preparation of the matrix, and could provide a more accurate mass of the glycan. Peaks 1 and 2 indicate that MOMP$^{T/G268}$ and wtMOMP without the glycan have a m/z difference of 109.7. After considering that substitution of threonine with glycine will further reduce the mass by 45 g/mol, there is still 64.7 Da to be accounted for.

The first monosaccharide attached in the synthesis of *O*-linked glycans is *N*-acetyl-galactosamine, which has a mass of 221.2, after this, several different pathways are possible. To search for post translational modification on the peptide, it was assumed that the m/z difference of 518.43 (the difference between peak 2 and 3) accounts for *N*-acetyl-galactosamine and unknown sugar units. Thus, the mass spectrum of trypsin digested wtMOMP was searched against trypsin digested MOMP$^{T/G268}$ to determine the presence of a peptidoglycan.

Table 4.1. Theoretical mass of the peptide of interest by itself, with *N*-acetyl-galactosamine and with *N*-acetyl-galactosamine plus unknown carbohydrate residues.

Precursor peptide	Charged peptide	Charge state
1185.64	1186.64	1
	593.82	2
	396.21	3
	297.41	4

Condensed N-acetyl-galactosamine	Precursor peptide	Charged peptide	Peptide and N-acetyl-galactosamine	Charge state
203.20	1185.64	1186.64	1389.84	1
		593.82	695.42	2
		396.21	463.95	3
		297.41	348.21	4

N-acetyl-galactosamine and unknown glycan	Precursor peptide	Charged peptide	Peptide, N-acetyl-galactosamine and unknown glycan	Charge state
518.43	1185.64	1186.64	1705.07	1
		593.82	853.04	2
		396.21	569.02	3
		297.41	427.02	4

Table 4.1: The amino acid residue of interest is Thr268, which is part of peptide THANGNLFALK with m/z of 1185.64. The mass of the *N*-acetyl-galactosamine was condensed to 203.2, which is the mass it would be if it were attached to another sugar unit as there is loss of a water molecule. Top table lists the charge states of the peptide if it were not attached to any other residue. Middle table is a list of m/z masses for finding the peptide plus *N*-acetyl-galactosamine whilst the bottom table lists masses for finding the peptide attached to *N*-acetyl-galactosamine plus the unknown carbohydrate residues. Red shading highlights the mass that was observed in both MOMP$^{T/G268}$ and wtMOMP spectra and green highlights the mass observed in the sample from trypsinated wtMOMP.

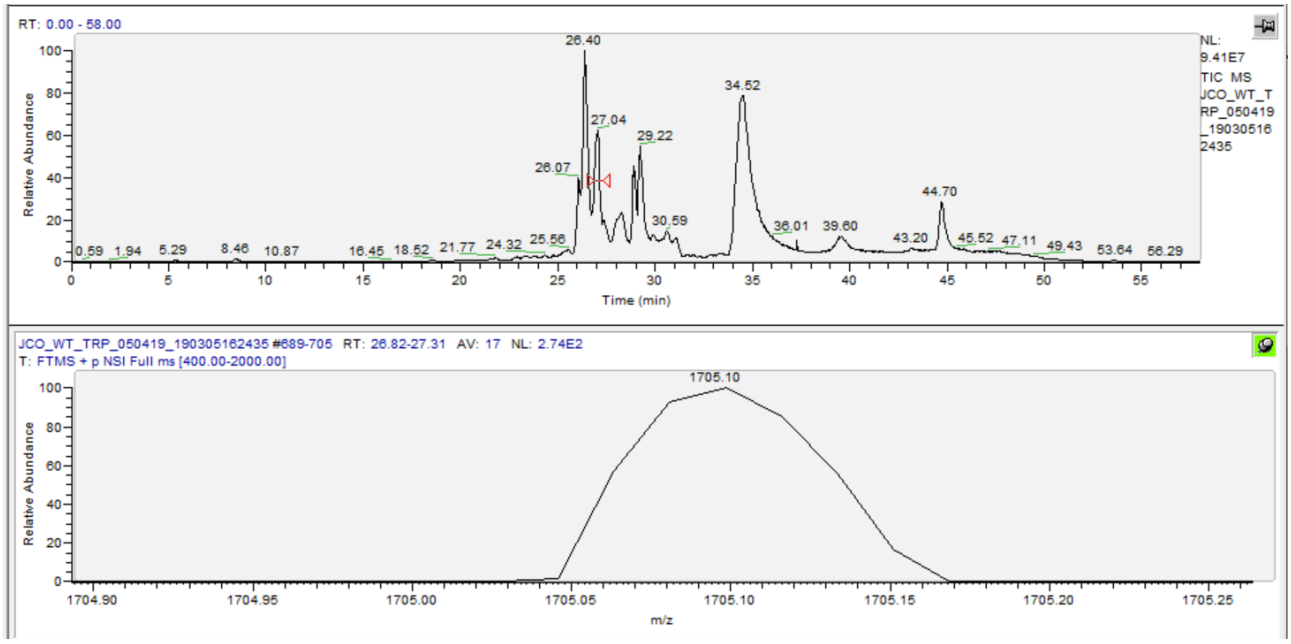


Figure 4.2: LCMS of trypsinated wtMOMP. Top panel displays the abundance of total peptides from the digest over time whilst the bottom panel displays the relative abundance of peptide mass of interest in a particular time frame. A mass of 1705.10 was observed at 26.22 min, which when allowing for mass error of +/- 20ppm fits our calculated mass of singly charged THANGNLFALK peptide in association with a mass of 518.43

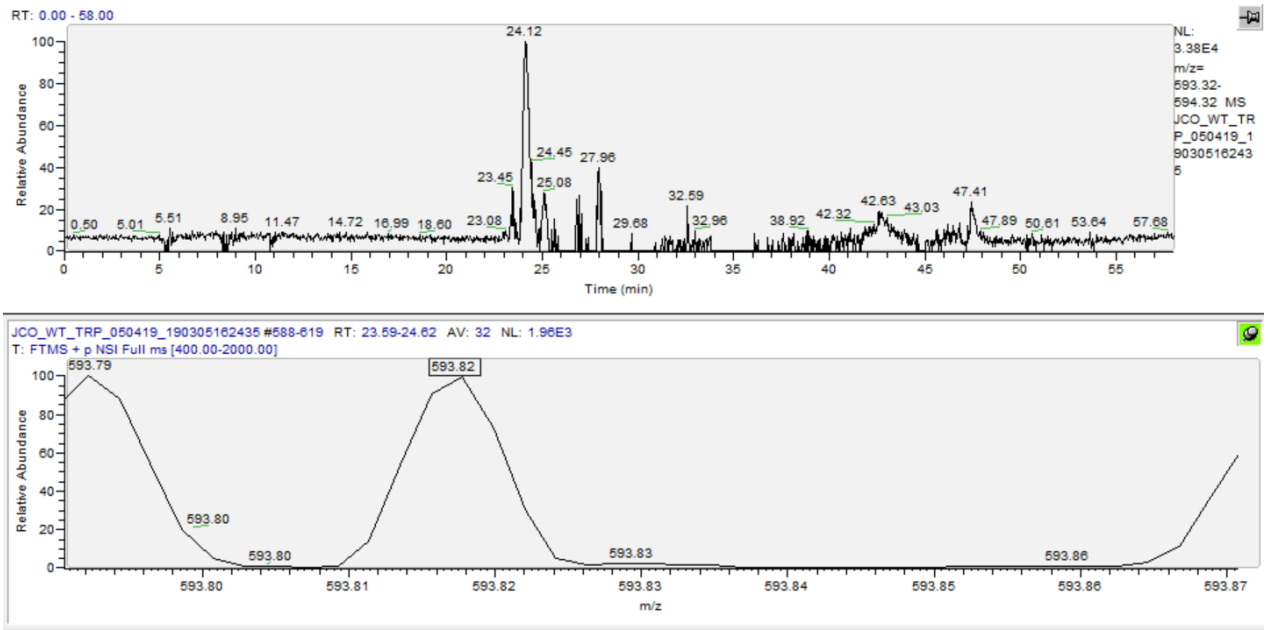


Figure 4.3: LCMS of trypsinated wtMOMP. Top panel displays the abundance of total peptides from the digest over time whilst the bottom panel displays the relative abundance of peptide mass of interest in a particular time frame. A mass of 593.82 was observed at 23.59 to 24.62 min. This mass corresponds to the calculated mass of the doubly charged unglycosylated THANGNLFALK.

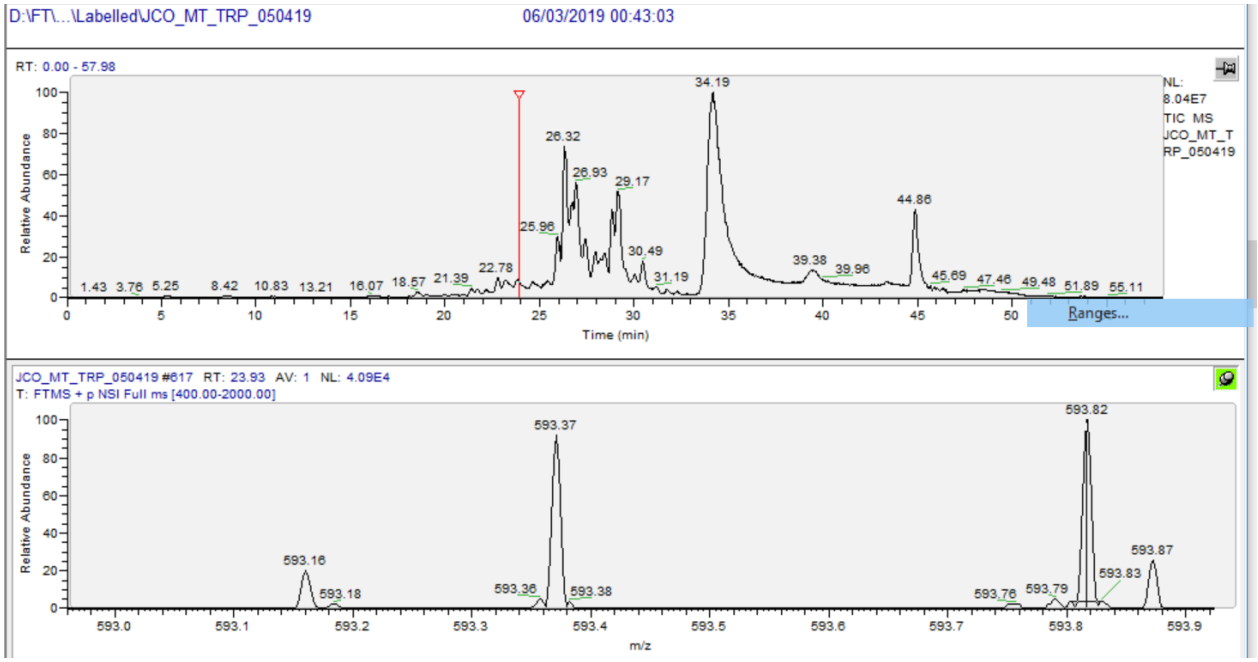


Figure 4.4: LCMS of trypsinated MOMP$^{T/G268}$. Top panel displays the abundance of total peptides from the digest over time whilst the bottom panel displays the relative abundance of peptide mass of interest in a particular time frame. A mass of 593.82 was observed at 23.59 to 24.62 min. This mass corresponds to the calculated mass of the doubly charged non-glycosylated THANGNLFALK.

Trypsination of wtMOMP resulted in a peptide that had a mass of 1705.10 m/z which accounts for the peptide THANGNLFALK linked to *N*-acetyl-galactosamine and an unknown carbohydrate residue and confirms the presence of a peptidoglycan unit that has a precursor mass of 518.43 (**figure 4.2**). Using the size of 518.43 as a guide, it means that it is likely that 2 sugars are attached to the *N*-acetyl-galactosamine as it leaves 315.43 (after condensation). For example, the attachment of a galactose, (MW = 180.16) and a ribose (MW= 150.13), allowing for removal of another water to account for a fourth carbohydrate in the glycan amounts to a m/z of 312, which may answer the question of the glycan sequence, leaving the terminal carbohydrate to be elucidated. In the case of MOMP$^{T/G268}$ only the GHANGNLFALK peptide was observed, as expected (**figure 4.4**).

There may be more carbohydrate residues attached to the peptide but they may be in low abundance, which are consequently missed from identification in the mass spectrum. Of all the

PTMs, glycosylation is one of the most challenging because of the variability in the attached glycans and glycan fragmentation during MS. Furthermore, because of glycan heterogeneity, there is no single mass shift associated with glycosylation, unlike other modifications[163].

4.3 Searching for other MOMP PTMs

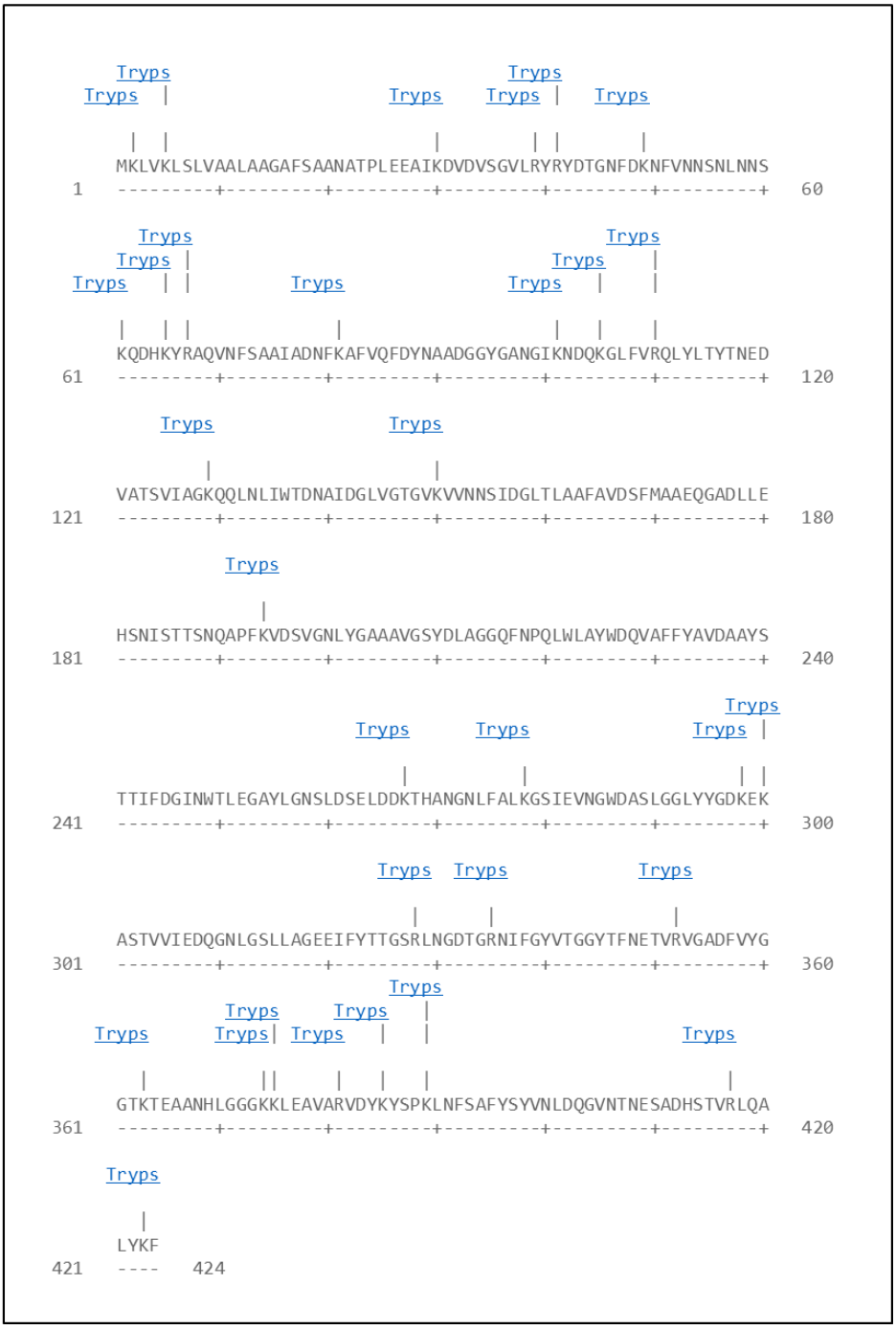
It is also important to be able to search other sites within the protein for any other potential sites of glycosylation. Analysis of PTMs by mass spectrometry can be difficult and the level of difficulty is dependent on (i) the mass shift in the peptide molecular weight, (ii) the overall abundance of the modified peptide, (iii) the stability of the modification during MS and MS/MS analysis, and (iv) the effect of the modification on the peptide's ionization efficiency (and therefore the sensitivity). Locating the exact site of a modification requires detection of the relevant peptide and in most cases MS/MS sequencing of the *specific* modified peptide[163].

The first choice of enzyme for digesting proteins into peptides for MS/MS based analysis is trypsin but the limitations of mass spectrometers often make it difficult to identify extremely short or long peptides. Based on an assumption that peptides with a length of 6-25 amino acid residues can be identified easily, only 60% of the proteome is amenable to identification, meaning that only 60% of glycoproteins will be identifiable by current MS capabilities[164]. This creates a systematic bias and creates false negatives in the acquired datasets even when they are very large. Higher coverage can be obtained by using multiple proteases for proteomic and PTM analysis in parallel experiments. In this case, chymotrypsin was used as a second enzyme. The majority of PTM sites are identified from bottom-up proteomic approaches by searching tandem mass spectra against theoretical protein databases using database search algorithms. The components of such computational pipelines dramatically affect the identification of PTM sites in the analysis. Currently, the identification of post-translationally modified peptides by MS mostly relies on

analysis of real mass spectra against theoretical mass spectra generated from selected protein databases so it was important to determine the degree of peptide digestion, not just theoretically but experimentally, and whether resultant peptides were within the required range of MW for separation on the C18 LC column before being captured by MS. Two-dimensional (2D) gel separation was followed by enzymatic digestion, matrix-assisted laser desorption/ionization -time-of-flight (MALDI-TOF) mass measurement and identification by peptide mass fingerprinting (PMF). Trypsin is used almost exclusively in MS techniques to generate peptide fragments due to its extremely high cleavage specificity on the C-terminal side of basic containing amino acids, Arg and

Lys[165]. Using the EXPASY *in silico* digest tool, a theoretical view of where trypsin should cut wtMOMP is shown in **figure 4.5A** (assuming there are no missed cleavages).

A.



B.

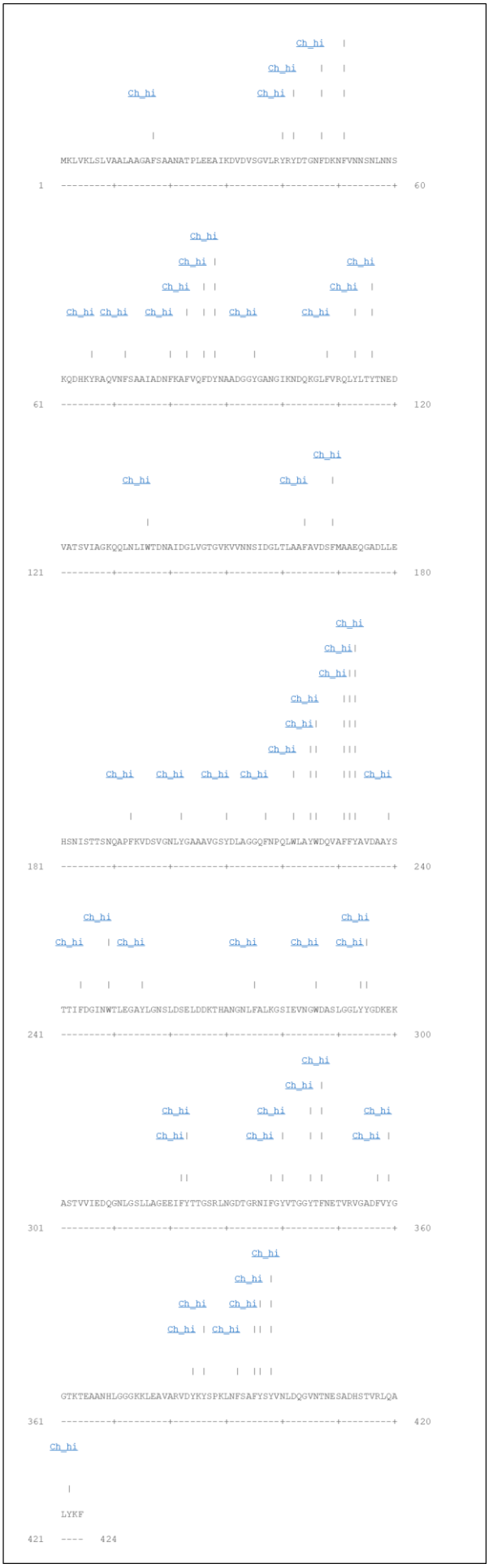

MKLVKLSLVAALAAGAFSAANATPLEEAIKDVDVSGVLRYRYDTGNFDKNFVNNSNLNNS
1 60
KQDHKYRAQVNFSAAIADNFKAFVQFDYNAADGGYGANGIKNDQKGLFVRQLYLTYTNED
61 120
VATSVIAGKQQLNLIWTDNAIDGLVGTGVKVVNNSIDGLTLAAFAVDSFMAAEQGADLLE
121 180
HSNISTTSNQAPFKVDSVGNLYGAAAVGSYDLAGGQFNPQLWLAYWDQVAFFYAVDAAYS
181 240
TTIFDGINWTLEGAYLGNSLDSELDDKTHANGNLFALKGSIEVNGWDASLGGLYYGDKEK
241 300
ASTVVIEDQGNLGSLLAGEEIFYTTGSRLNGDTGRNIFGYVTGGYTFNETVRVGADFVYG
301 360
GTKTEAANHLGGGKKLEAVARVDYKYSPKLNFSAFYSYVNLDQGVNTNESADHSTVRLQA
361 420
LYKF
421 424

Figure 4.5: The cleavage sites of trypsin and chymotrypsin mapped onto the MOMP protein sequence. Cleavage occurs at the right side (C-terminal direction) of the marked amino acid. **A.** Trypsin preferentially cleaves at Arg and Lys in position P1 with higher rates for Arg. Pro usually blocks the action when found in position P1', but not when Lys is in position P1 and Trp is in position P2 at the same time. This blocking of cleavage exerted by Pro in position P1' is also negligible when Arg is in position P1 and Met is in position P2 at the same time. **B.** Chymotrypsin preferentially cleaves at Trp, Tyr and Phe in position P1(high specificity) and to a lesser extent (taken into account when dealing with low specificity chymotrypsin) at Leu, Met and His in position P1 (Keil, 1992). Exceptions to these rules are the following: When Trp is found in position P1, the cleavage is blocked when Met or Pro are found in position P1'at the same time. Furthermore, Pro in position P1' nearly fully blocks the cleavage independent of the amino acids found in position P1. When Met is found in position P1, the cleavage is blocked by the presence of Tyr in position P1'. Finally, when His is located in position P1, the presence of Asp, Met or Trp also blocks the cleavage.

The first step was to digest the native wtMOMP protein. **Table 4.2** shows the peptides that were produced after digestion with trypsin alongside the number of cleavages that were missed by the protease in order to obtain the peptide, the precursor mass of the peptide before charging, the charge sate of the resultant peptide and whether any tandem mass spectrometry (Tandem MS or MSMS) data are available for the peptide. Tandem MS is a technique which breaks down precursor peptides from initial MS into smaller fragments or product ions. These fragments can reveal aspects of the chemical structure of the precursor ion. MSMS data are important for future studies when a more in-depth picture of the peptides is needed to determine which residues in particular are involved in any PTMs. **Tables 4.2 and 4.3** and **figures 4.3 and 4.4** highlight which peptides within wtMOMP have MSMS data available for further exploration.

Trypsin digest MS/MS data available

Sequence	Missed Cleavages	Precursor mass	Charge	MSMS fragmentation data available?
TPLEEAIKDVDVSGVLR	1	921.0	2+	Y
LEEAIKDVDVSGVLR	1	548.3	3+	Y
YDTGNFDKNFVNNSNLNNSK	1	1153.03	2+	N
YDTGNFDKNFVNNSNLNNSK	1	769.35	3+	Y
AQVNFSAAIADNFK	0	748.38	2+	Y
FSAAIADNFK	0	542.28	2+	Y
AFVQFDYNAADGGYGANGIK	0	693.33	3+	N
AFVQFDYNAADGGYGANGIK	0	1039.98	2+	Y
AFVQFDYNAADGGYGANGIKNDQK	1	855.39	3+	N
FDYNAADGGYGANGIKNDQK	1	1059.97	2+	N
QLYLTYTNEDVATSVIAGK	0	1035.52	2+	Y
QLYLTYTNEDVATSVIAGK	0	1043.54	2+	Y
QLYLTYTNEDVATSVIAGK	0	1035.03	2+	Y
TNEDVATSVIAGK	0	652.84	2+	Y
QQLNLIWTDNAIDGLVGTGVK	0	1128.1	2+	N
QQLNLIWTDNAIDGLVGTGVK	0	1119.59	2+	N
QQLNLIWTDNAIDGLVGTGVK	0	1120.09	2+	Y
QQLNLIWTDNAIDGLVGTGVK	0	1128.6	2+	Y
NLIWTDNAIDGLVGTGVK	0	943.5	2+	N
VVNNSIDGLTLAAF	0	717.39	2+	Y
VVNNSIDGLTLAAFAVDSFMAAEQGADLLEHSNISTTSNQAPFK	0	1149.32	4+	N
AVDSFMAAEQGADLLEHSNISTTSNQAPFK	0	1060.5	3+	1060.47
STTIFDGINWTLEGAYLGNSLDSELDDK	0	1025.49	3+	N
STTIFDGINWTLEGAYLGNSLDSELDDKTHANGNLFALK	1	1061.27	4+	N
THANGNLFALK	**0**	**593.81**	**2+**	**Y**
GSIEVNGWDASLGGLYYGDK	0	1051.49	2+	Y
GSIEVNGWDASLGGLYYGDK	0	701.0	3+	N
GSIEVNGWDASLGGLYYGDKEK	1	1180.06	2+	Y
ASTVVIEDQGNLGSLLAGEEIFYTT	0	876.44	3+	N
ASTVVIEDQGNLGSLLAGEEIFYTTGSR	0	976.49	3+	N
ASTVVIEDQGNLGSLLAGEEIFYTTGSR	0	976.83	3+	Y
NIFGYVTGGYTFNETVR	0	969.97	2+	Y
NIFGYVTGGYTFNETVR	0	969.97	2+	Y
NIFGYVTGGYTFNETVR	0	969.47	2+	Y
VTGGYTFNETVR	0	672.33	2+	Y
VGADFVYGGTK	0	557.28	2+	Y
LNFSAFYSYVNLDQGVNTNESADHSTVR	0	1050.49	3+	N
LNFSAFYSYVNLDQGVNTNESADHSTVR	0	1050.16	3+	Y
FSAFYSYVNLDQGVNTNESADHSTVR	0	974.45	3+	N
SAFYSYVNLDQGVNTNESADHSTVR	0	925.43	3+	N
YSYVNLDQGVNTNESADHSTVR	0	824.05	3+	Y
YSYVNLDQGVNTNESADHSTVR	0	823.71	3+	N
SYVNLDQGVNTNESADHSTVR	0	769.36	3+	769.35
VNLDQGVNTNESADHSTVR	0	685.99	3+	N

Table 4.2: List of peptides observed in LCMS from digestion of wtMOMP with trypsin. List of peptides that were produced after digestion alongside the number of cleavages that were missed by the protease in order to obtain the peptide (≤2), the precursor mass of the peptide before charging, the

charge sate of the resultant peptide and whether any tandem mass spectrometry (MSMS) data are available for the peptide. Some peptides were deamidated, this PTM is shown in red. Green highlighting shows peptides where the precursor m/z of the peptide was not found to be targeted although peptide shaker has data for it. Pink highlighting shows peptides that do not have any MSMS data for either trypsin or chymotrypsin. Precursor peptides (shaded in grey) that have no MSMS data will be targeted in another round of experiments in targeted DDA. The peptide which contains our residue of interest, T268, is highlighted in bold.

Extensive characterisation of a protein requires the observation of as many amino acids of that protein as possible, and so the use of multiple proteases increases protein coverage. Chymotrypsin was used in separate parallel protein digest steps to trypsin to generate alternative peptide fragments. Chymotrypsin cleaves peptide bonds selectively on the carboxyl terminal side of large hydrophobic amino acids such as tryptophan, tyrosine, phenylalanine, and methionine. A theoretical view of where chymotrypsin should cut is shown in **figure 4.2B** (assuming there are no missed cleavages). **Table 4.3** displays the results of digestion of wtMOMP with chymotrypsin.

Chymotrypsin digest MS/MS data available

Peptide digest	Precursor mass	Charge state	MSMS fragmentation data available?
TPLEEAIKDVDVSGVL	842.95	2+	842.96
TPLEEAIKDVDVSGVLRY			
LEEAIKDVDVSGVL	743.9	2+	Y
LEEAIKDVDVSGVLRY	602.66	3+	602.66
RYDTGNFDKNF	688.82	2+	
RYDTGNFDKNFVNNSNL	1009.47		
TGNFDKNFVNNSNL	792.38	2+	Y
DKNFVNNSNLNNSKQDHKY	570.53	4+	570.78
SAAIADNFKAF	577.8	2+	Y
VQFDYNAADGGY	660.28	2+	660.28
DYNAADGGYGANGIKNDQKGLF	763.36		
GGYGANGIKNDQKGLF	820.41	2+	Y
GANGIKNDQKGLF	681.85	2+	Y
NGIKNDQKGLF	617.82	2+	Y
GIKNDQKGLF	560.31	2+	Y
VRQLYLTY	528.3	2+	Y
LTYTNEDVATSVIAGKQQL	1026.04	2+	
TNEDVATSVIAGKQQL	837.44	2+	Y
TNEDVATSVIAGKQQLN	894.46	2+	Y
TNEDVATSVIAGKQQLNL	951.0	2+	Y
LLEHSNISTTSNQAPF	879.94	2+	879.93
EHSNISTTSNQAPF	766.85	2+	Y
KVDSVGNLYGAAAVGSY	835.92	2+	Y
KVDSVGNLYGAAAVGSY	836.42	2+	Y
DLAGGQFNPQLW	673.33	2+	Y
LAYWDQVAF	556.77	2+	Y
YAVDAAYSTTIFDGINW	935.95	2+	
AVDAAYSTTIFDGINW	872.91	2+	
AVDAAYSTTIFDGINW	872.41	2+	
SELDDKTHANGNLF	**781.36**	**2+**	**Y**
ALKGSIEVNGWD	645.32	2+	Y
YGDKEKASTVVIED	777.39	2+	Y
TTGSRLNGDTGRNIF	805.4	2+	Y
GSRLNGDTGRNIF	704.35	2+	Y
TFNETVRVGADF	678.33	2+	Y
TFNETVRVGADFVY	809.4	2+	Y
SYVNLDQGVNTNESAD	863.38	2+	Y
SYVNLDQGVNTNESADH	931.91	2+	Y
SYVNLDQGVNTNESADHSTVR	769.36	3+	Y
SYVNLDQGVNTNESADHSTVRL	807.05	3+	
SYVNLDQGVNTNESADHSTVRLQAL	911.11	3+	Y
VNLDQGVNTNESADHSTVRLQALY	882.1	3+	

Table 4.3: List of peptides observed in LC MS from digest of wtMOMP with chymotrypsin. List of peptides that were produced after digestion alongside the number of cleavages that were missed by the protease in order to obtain the peptide (≤2), the precursor mass of the peptide before charging, the charge sate of the resultant peptide and whether any tandem mass spectrometry (MSMS) data are available for the peptide. Some peptides were deamidated, this post PTM is shown in red. Green highlighting shows peptides where the precursor m/z of the peptide was not found to be targeted although peptide shaker has data for it. Pink highlighting shows peptides that do not have any MSMS

data in either trypsin or chymotrypsin digest data. Precursor peptides (shaded in grey) that have no MSMS data will be targeted in another round of experiments in targeted DDA. The peptide which contains our residue of interest, T268, is highlighted in bold.

It is important to note that peptide mass fingerprinting involves the comparison of experimentally determined peptide masses to molecular weights derived from an *in-silico* digest of a sequence database using programs such as EXPASY. Several phenomena influence the matching of the experimental data to the theoretical data. The reliability of a peptide match is dependent on the mass accuracy of the data, the validity of the sequence databases searched and the PTM patterns. One such PTM is deamidation which is a chemical reaction where an amide functional group in the side chain of the asparagine or glutamine is removed or converted to aspartic acid or glutamic acid. This PTM is important in the protein as it may alter its structure, stability or function and may lead to protein degradation. The net chemical change is the addition of a water group and removal of an ammonia group, which corresponds to a +1 (0.98402) Da mass increase. Further problems arise by the frequency with which missed cleavages by proteases occur as they fail to fully digest proteins to their complete peptides. Even trypsin, which is the go-to for MS experiments because of its high cleavage specificity, availability and cost, although not widely reported, is commonly subject to missing tryptic peptide bonds. However, other residues in close proximity to the cleavage site, in addition to factors such as local conformation, tertiary structure and experimental conditions are likely to affect the true specificity of any protease. It is important that downstream processes allow for missed cleavages when studying protein regions.

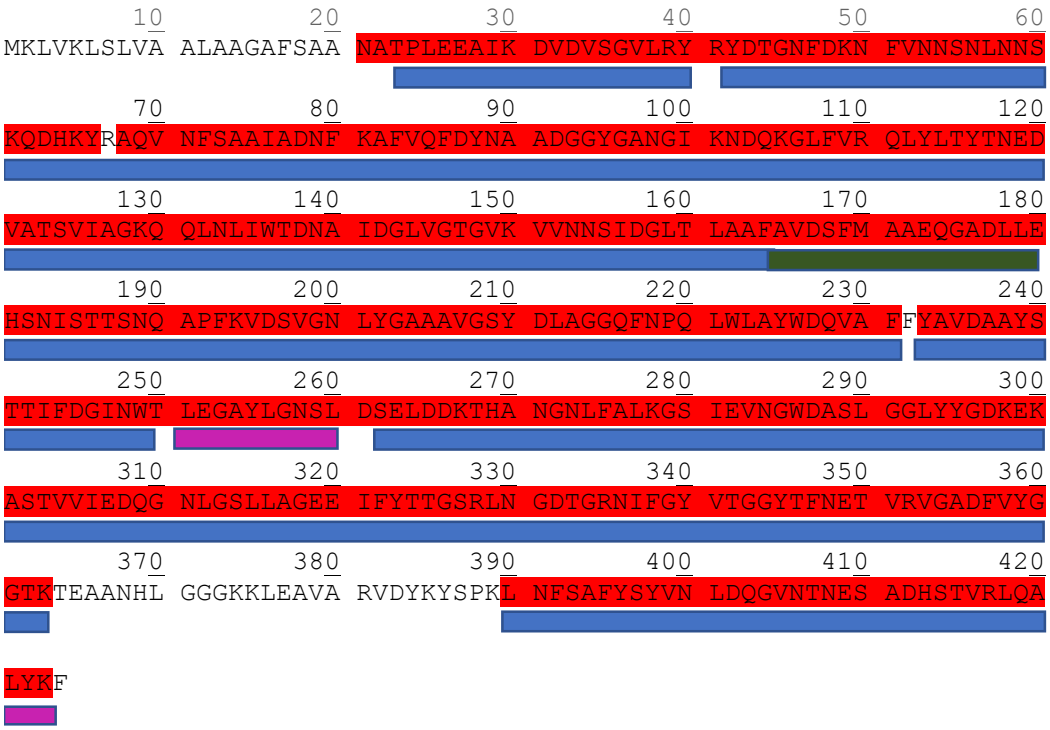


Figure 4.6: Sequence of wtMOMP with data from trypsin and chymotrypsin digests combined. Once samples are ionized to generate a mixture of ions, precursor ions of a specific m/z are selected (MS1) and then fragmented (MS2) to generate product ions for detection**.** Red highlight shows the peptides from trypsin and chymotrypsin combined whilst the blue blocks show the regions of wtMOMP that that have MSMS data. Pink block highlights the region of wtMOMP where there were no MSMS data available, thus targeted MS will be performed on peptides in this region. Green highlighting shows peptides where the precursor m/z of the peptide was not found to be targeted although peptide shaker has data for it. A list of the peptides can be found in the supporting information.

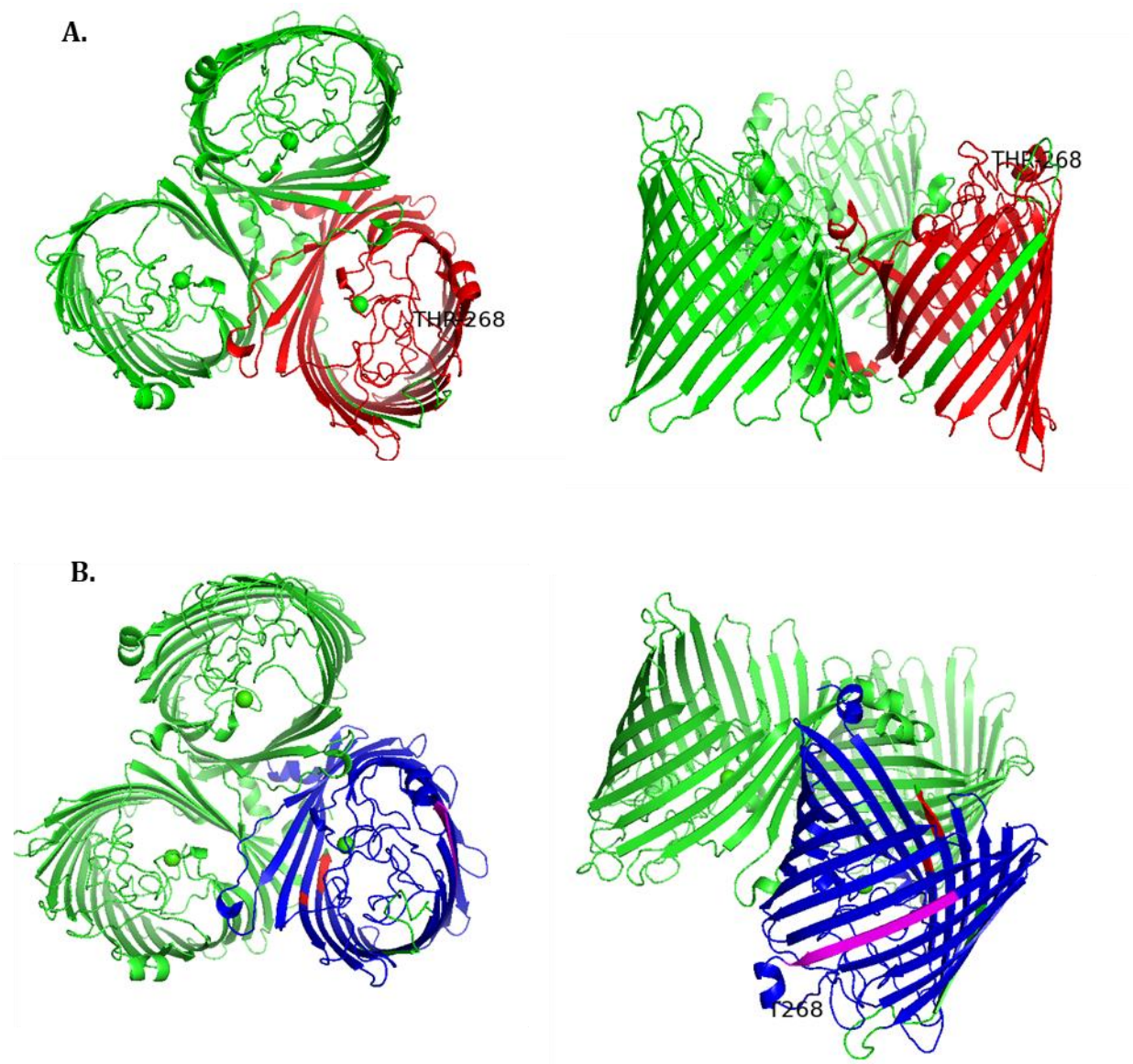


Figure 4.7: A. wtMOMP *C. jejuni* strain NCTC11168 sequence was modelled onto the crystal structure of the MOMP *C. jejuni* strain 85H obtained by Ferrara *et al.* (PDB accession number 5LDT)[120]. The peptide coverage from **figure 3** has been mapped onto a monomer unit in the structural model. The residue of interest T268 is labelled. Here the peptides from trypsin and chymotrypsin combined are highlighted in red. **B.** The regions of wtMOMP for which there is MSMS data available have been mapped in blue onto a monomer unit in the structural model. Pink highlights the region of wtMOMP where there was no MSMS data available. Red region highlights where there was no MS or MSSM data available. Residue of interest T268 is labelled.

Importance of MSMS data

Glycoproteins often have multiple sites of glycosylation, which greatly complicates analysis of the intact protein. 94% of the wtMOMP protein has MSMS data available which can be probed to reveal aspects of the chemical structure of the precursor peptides, such as attachment sites of

glycans as well as the composition of glycans. To characterise the structure of glycans, they are released from glycoproteins, purified, labelled, and subjected to MS analysis. As low abundance ions are missed from sequencing identification, the conditions for collision-induced dissociation have to be optimized to obtain diagnostic ions that maximize glycan and peptide information. The speed of acquisition of MS/MS with no loss of signal intensity is very important for successful proteomics analysis but data-independent analysis can be carried out to acquire all possible fragment ions including those from low abundance ions.

For structural elucidation, not only the monosaccharides (and their stereochemistry) need to be identified, but also how they link together to form the polymers needs to be determined. This can be determined by analysing the sequence of fragments that are released during MS/MS collisions. Peptides and oligosaccharides (including glycolipids) follow different systems of nomenclature for their fragment ions.

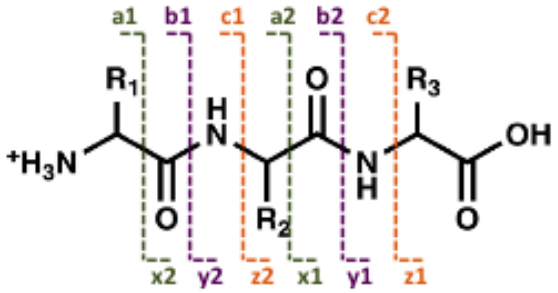


Figure 4.8: Typical nomenclature for peptide fragments. Fragments containing the *N*-terminus are labeled a, b, or c, depending on the site of the cleavage, whereas fragments containing the C-terminus are labeled x, y, or z. The numbers indicate the number of amino acid residues in the fragment ion[166].

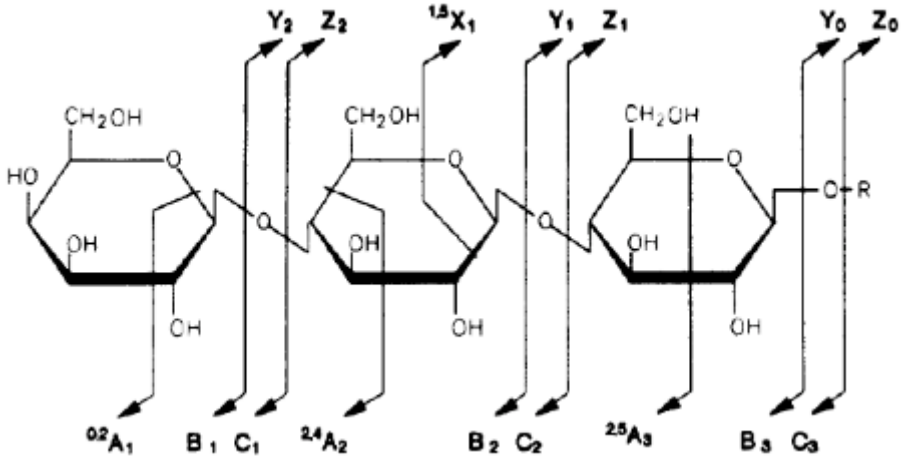


Figure 4.9: Typical nomenclature for oligosaccharide fragments. For oligosaccharides, fragments containing the reducing end (reducing end is on the right-hand side in the figure) are labeled X, Y, or Z, depending on the site of the cleavage, whereas fragments containing the other end are labelled A, B, or C. The numbers indicate the site of the sugar residue: y, z, b, and c ions are fragments due to glycosidic

cleavages (cutting glycosidic bonds holding two adjacent sugar residues), whereas a and x ions result from cross-ring cleavage[167].

4.4 Conclusion to this chapter

Glycans attached to specific sites on a protein backbone generate structurally and chemically diverse glycoproteins and the variability of the glycan on a particular site further increases the diversity of structures. Combined with results from previous studies where MS/MS analysis suggests a Hex-(Hex*N*-acetylamine)$_3$ glycan, lectin-binding analysis indicates that the MOMP glycan in NCTC11168 is likely to be the *O*-linked trimeric form of T-antigen: Gal(β1–3)-GalNAc(β1–4)-GalNAc(β1–4)-GalNAcα1-Thr[268], the studies in this chapter confirm that there is at least a 518 Da size shift between the non-glycosylated and glycosylated variants which would account for the occurrence of *N*-acetyl-galactosamine followed by 2 sugars such as the attachment of a galactose, (MW = 180.16 Da) and a ribose (MW= 150.13 Da). The character of the subsequent sugars would be further determined upon inspection with tandem MS by the aforementioned protocol.

Chapter 5: **Interaction between MOMP, synthetic and biologic ligands**

5.1 Introduction

Membrane proteins serve many critical roles in the cell including enzymatic functions, transportation, signal transduction, cell recognition, intercellular joining and attachment, as well as the transmission of information from the external environment to the inside of the cell[168, 169]. One way to make sense of the external environment is by the interaction of the membrane proteins with ligands that act as agonists, inverse agonists, antagonists, allosteric agents, or substrates. The properties of ligands are diverse and can range from ions to small molecules, peptides, or proteins. Synthetic compounds such as drugs can also bind to target proteins, leading to the inhibition of protein–ligand interactions[170, 171]. The nature of interactions between ligands and membrane proteins and the impact that ligand binding has on membrane protein structure are areas that still require more detailed insight[172]. Studies of such interactions are important because the interactions of small molecules with proteins mediate various biological phenomena including signal transduction, protein transcription and translation, and metabolism and ligand-binding interactions are used to switch proteins among states of different functions. Protein-ligand and protein-protein interactions are at the core of many areas of research including molecular docking, structure-based design, virtual screening of molecular fragments, small molecules, and other type of compounds, clustering of complexes, and structural interpretation of activity cliffs[173].

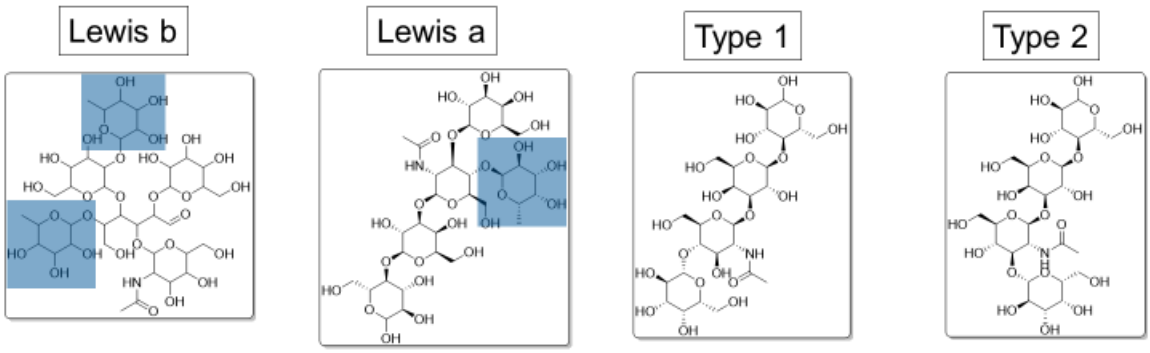
5.2 Binding of BgAgs to MOMP

Campylobacter strains are capable of strong adhesion which may be the underlying cause for the rapid cross-contamination or re-contamination of food products and may be the most significant

mode of survival for *Campylobacter* in the food chain. Sulemann *et al*. compared differences in adhesion of *C. coli and C. jejuni* and found that, unlike *C. jejuni*, no strain of *C. coli* was strongly adherent, and statistically, *C. coli* adhered less to an inert surface than *C. jejuni*[174]. Recently, in-feed chelated iron (III) complexes have shown to be effective against *Campylobacter* and other pathogenic bacteria in chicken that is bred and raised specifically for meat production, both in artificially challenged and naturally imposed infectious conditions[175, 176]. The novel feed additive used in these studies was ferric tyrosine (TYPLEX®, Akeso Biomedical Inc., USA), a complex containing tyrosine and iron. Another novel compound based on iron, developed by the same group, is QPLEX, an iron (III) quinic acid complex. The group are performing on going field trials to test the efficacy of QPLEX in reducing *Campylobacter* colonisation. Alongside these experiments *in vivo* I wanted to design experiments to test the efficacy of QPLEX *in vitro* on a molecular level.

To determine whether iron chelates such as QPLEX could interact with and disrupt MOMP binding to its natural binding ligands *in vivo*, binding studies were first carried out to identify MOMP binding partners whose interactions could subsequently be probed in the presence and absence of QPLEX. Previous studies have found that fucosylated sugar components in human breast milk could inhibit *C. jejuni* binding to HEp-2 cells *in vitro* and to human intestinal mucosa *ex vivo*, and also prevent *C. jejuni* infection of mice *in vivo*[42]. Other studies have shown that the *H. pylori* BabA (blood group antigen-binding) and SabA (sialic acid-binding adhesin) proteins are ABO/Le^b and sialyl-Lewis x (sLex) antigen-binding adhesins and predictors for the development of overt gastric disease[8, 9]. For *C. jejuni*, Mahdavi *et al.* identified the surface proteins FlaA and MOMP as BgAg-binding adhesins. Using competition binding assays of *C. jejuni* whole cells and purified MOMP to the H-II glycoconjugate, they identified MOMP as a blood group binding adhesin[10]. To build upon this work and to acquire a basis for binding studies of MOMP with QPLEX, I confirmed the binding

of the BgAgs to the MOMP protein in direct interaction studies using an ELISA based experimental design with BgAgs fucosylated to varying degrees.



Blood group sugar antigen	Structure
Lewis b	Fuc(α1-2)Gal(β1-3)[Fuc(α1-4)]GlcNAc(β1-3)Gal(β1-4)Glc
Lewis a	Gal(β1-3)[Fuc(α1-4)]GlcNAc(β1-3)Gal(β1-4)Glc
Type 1 chain	Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc
Type 2 chain	Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc

Figure 5.1: The blood group carbohydrate structures (Le^b, Le^a, Type 1 and Type 2) used in this chapter to investigate binding of BgAgs to MOMP. The fucose units are highlighted.

BgAg carbohydrates are important receptors for gram negative bacteria such as *H. pylori*. ABO and Lewis BgAg phenotypes express different levels of carbohydrate antigens into exocrine secretions and gastrointestinal tract depending on the presence, absence or combination of FUTII, FUTIII, GTA and GTB glycosyltransferases in the host[177]. These enzymes work to add monosaccharide units to precursor oligosaccharides and the various enzyme interactions lead to the structural diversity of BgAg carbohydrates (**figure 5.2**).

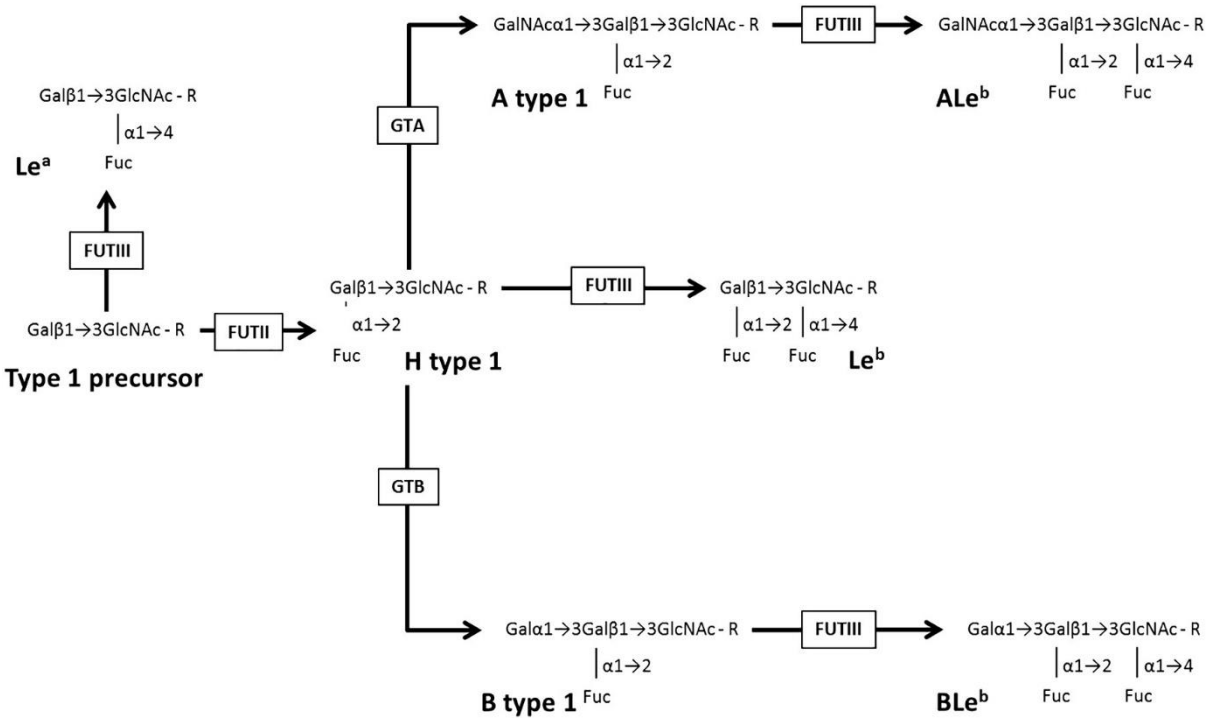


Figure 5.2: Biochemical pathways of ABO, Lewis and Secretor BgAg carbohydrates in the digestive tract from type 1 precursor oligosaccharide. FUTII, FUTIII, GTA and GTB are glycosyltransferases coded by FUT2 (Secretor), FUT3 (Lewis) and ABO (ABO) genes. Type 2 precursor is not shown but can be included with type 1 precursor as they are positional isomers. Figure from Mattos *et al.,* 2017[178].

Fucosylation is a biological process broadly observed in vertebrates, invertebrates, plants, bacteria and fungi. Fucose moieties on cell-surface glycans are increasingly recognized as critical to many cell-cell interaction and signalling processes, and are involved in a wide variety of physiological and pathological processes[179]. The carbohydrate BgAgs used in this study were chosen for the varying number of fucosyl carbohydrate groups present on their carbohydrate backbone structures to determine if fucosylation has an effect on the binding to their cognate proteins in the glycosylated and non-glycosylated states (**figure 5.1**). Type 1 and Type 2 precursors do not contain any fucosyl units on their backbone but are substrates of the FUTII enzyme (α1 → 2-l-fucosyltransferase) where the enzyme adds a molecule of fucose to the type 1 precursor disaccharide (Galβ1 → 3GlcNAcβ1 → 3-R) making the H type 1 antigen (Fucα1 → 2Galβ1 → 3GlcNAcβ1 → 3-R). The FUTIII enzyme (α-3/4-l-fucosyltransferase) is able to fucosylate different precursors to construct new

carbohydrate structures. The type 1 precursor disaccharide is converted into the singly fucosylated Lea (Galβ1 → 3[Fucα1 → 4]GlcNAcβ1 → 3-R) antigen by the FUTIII enzyme. This enzyme can also add another molecule of fucose to the monofucosylated H type 1 structure making the difucosylated Le^b (Fucα1 → 2Galβ1 → 3[Fucα1 → 4]GlcNAcβ1 → 3-R) antigen. The terminal structures of the BgAg carbohydrates fucosylated by FUTIII cannot be further glycosylated[178]. Here I tested the binding of wtMOMP to four types of soluble BgAgs to determine which BgAg bound with most efficacy to the protein.

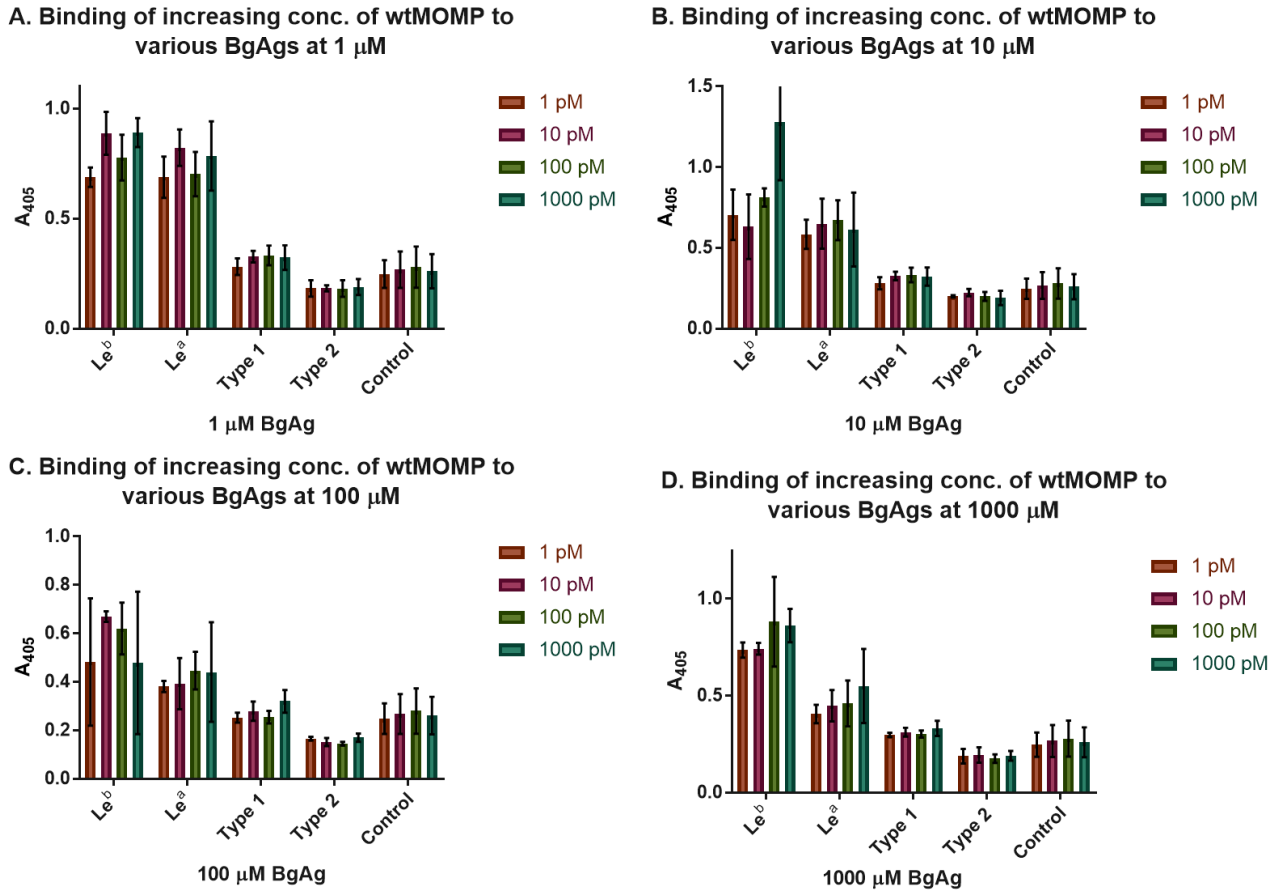


Figure 5.3: Binding of wtMOMP to Le^b, Le^a, Type 1, Type 2 and control substance. Increasing concentrations of wtMOMP (1, 10, 100 and 1,000 pM) in the picomolar region were tested for their binding ability to concentrations of BgAgs at 1, 10 100 and 1000 μM (Panels A-D, respectively. Binding of wtMOMP was measured by incubating the protein in amino linked ELISA wells coated with Le^b, Le^a, Type 1, Type 2 and control substance. This was followed by coating with primary antibody against MOMP and alkaline phosphatase (AP) conjugated secondary antibody. The Tecan was used to measure absorbance at 405 nm by the AP substrate. Error bars of standard deviation about the mean are included. The data represents four independent experiments with ≥3 replicates per well. A plate map can be found in chapter 7.4.

The order of preference for binding of wtMOMP to the various BgAgs is Le^b, Le^a, type 1 and type 2 (**figure 5.3**) suggesting a preference for the more fucosylated molecules. Carbohydrate – receptor interactions are an integral part of biological events therefore it is not surprising that carbohydrates can interact with a receptor by using several types of intermolecular interactions. The most well-known and generally considered interaction of such molecules is hydrogen bonding between the hydroxyl residues of carbohydrates and the polar amino acid residues[180], although there are other binding interactions at play. Protein – carbohydrate interactions are often facilitated by metal ions, such as calcium, zinc or magnesium. The ions form positive bridges between oxygen atoms of the carbohydrate hydroxyl group and negatively charged protein residue[181]. This could be true of wtMOMP as the crystal structure resolved by Ferrara *et al.* included a Ca^+ ion in the porin, pointing towards the presence of negatively charged residues in the porin binding site[120]. If these residues reside at the extracellular portion of the porin then they could form positive bridges to oxygen atoms of the carbohydrate hydroxyl groups of the BgAgs. The high incidence of crystal structures where non-polar carbohydrate faces are in association with aromatic amino acid residues indicate another type of van der Waals force - a London dispersion force[182]. This interaction occurs in many different carbohydrate processing enzymes including glycosidases, transglycosidases, glycosyltransferases and carbohydrate-recognizing proteins, including lectins, immunoglobulins and glycosaminoglycan[183].

Because of wtMOMP higher affinity for Le^b, than for the other carbohydrate srrutures, this sugar was chosen to be studied in further downstream experiments as the representative carbohydrate for probing the interaction mechanism between MOMP and carbohydrate molecules. **Figure 5.3** shows response of wtMOMP binding at higher concentration of Le^b. To get a more insightful dose-response curve, the binding response was measured at lower concentration range of Le^b. The concentration range was expanded to measure binding of 1 pM wtMOMP to Le^b of 1 fM to 1mM.

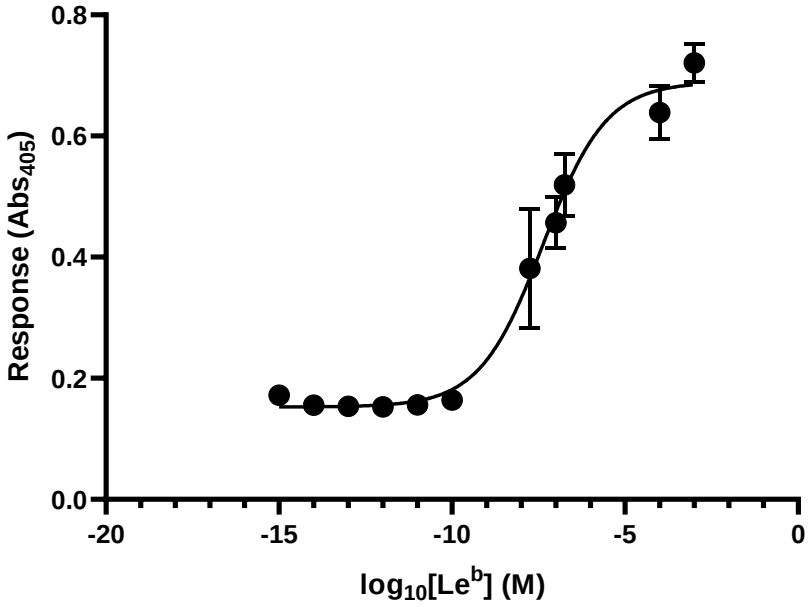


Figure 5.4: Binding of increasing concentration of Le^b to wtMOMP (1pM). Log_{10} of concentration of Le^b (M) is plotted against absorbance. Binding of wtMOMP was measured by incubating the protein in amino linked ELISA wells coated with increasing concentrations of Le^b where n≥3. This was followed by coating with primary antibody against MOMP and alkaline phosphatase (AP) conjugated secondary antibody. The Tecan was used to measure absorbance at 405 nm by the AP substrate. Graph was plotted on GraphPad Prism using with log(agonist) vs. response with variable slope (four parameters). The mean of the replicates are plotted with error bars that represent the standard deviation. Log EC50 is 2.361e-008 M whilst Hill slope is 0.45.

The results in **figure 5.4** show that there is a dose-response relationship between wtMOMP and Le^{b,} due to the sigmoidal nature of the curve. The EC50 is the ligand concentration at which the ligand is half-maximally effective. For the semi-log plot, the EC50 is the midpoint or inflection point of the curve. Here EC50 = 23 nM. The Hill slope of 0.45 is reasonable, pointing towards a genuine response of wtMOMP to Le^b, as opposed to non-specific binding. The inclusion of 0.1% BSA further reduces effects of non-specific binding.

Most antibodies have K_d values in the low micromolar (10^{-6}) to nanomolar (10^{-7} to 10^{-9}) range. High affinity antibodies generally considered to be in the low nanomolar range (10^{-9}) with very high affinity antibodies being in the picomolar (10^{-12}) range[184]. Although not an antibody interaction, it would be useful to measure the affinity for MOMP to Le^b, and to see if it is comparable and lies in the low nanomolar range so it should be considered that there is high affinity for the protein-sugar

interaction between wtMOMP and Le^b. The K_d value of this interaction can be determined with further studies using surface plasmon resonance (SPR) or isothermal titration calorimetry (ITC). An experimental advantage of high affinity binding means that less resources and sample can be used in downstream experiments where ELISA based experiments may usually require high amounts of samples for titration, binding and inhibition studies.

In addition, a control experiment was performed to measure the response or interference, if any, of the glycoconjugates in the absence of wtMOMP.

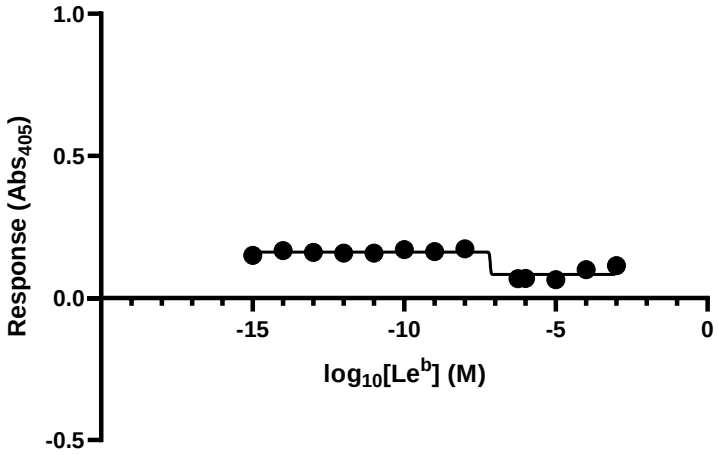


Figure 5.5: Binding of increasing concentration of Le^b in the absence of wtMOMP. Log_{10} of concentration of Le^b (M) is plotted against absorbance. Binding of was measured by incubating the glyconjugates in amino linked ELISA wells coated with increasing concentrations of Le^b where n≥3. This was followed by coating with primary antibody against MOMP and alkaline phosphatase (AP) conjugated secondary antibody. The Tecan was used to measure absorbance at 405 nm by the AP substrate. Graph was plotted on GraphPad Prism using with log(agonist) vs. response with variable slope (four parameters). The mean of the replicates are plotted with error bars that represent the standard deviation.

The lack of absorbance response seen in **figure 5.5** shows that neither the antibody nor the sugars are able to elicit a response using this technology, in the absence of the binding protein, and thus

a response that is seen in the presence of wtMOMP is an effect of the bound protein interacting with the BgAgs, as opposed to other factors.

The mutant strains of the wild type protein MOMP$^{\Delta pseD}$ and MOMP$^{T/G268}$ were used as controls in the experiment to determine whether the high binding of Le^b was specific to the glycosylated MOMP protein and the difference in binding to non-glycosylated variants of MOMP.

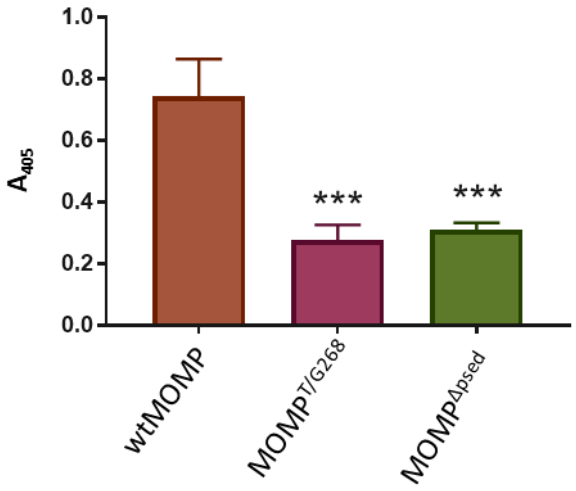


Figure 5.6: Binding of 1 pM of wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ to 1 µM of Le^b. Binding was measured by incubating in amino linked ELISA wells coated with Le^b. This was followed by incubating with primary antibody against MOMP and alkaline phosphatase (AP) conjugated secondary antibody. The Tecan was used to measure absorbance at 405 nm by the AP substrate. The results show the maximum absorbance obtained during a 16 hr assay. BSA was used as a control for background binding and the result of BSA absorbance has been taken away from total absorbance of binding of each protein to Le^b. The graph denotes the mean absorbance in each condition with standard error bars of the mean of triplicate data from 3 separate experiments. The statistical significant difference between Le^b binding to wtMOMP and Leb binding to MOMP$^{T/G268}$ and that of Leb binding to wtMOMP and Le^b binding to MOMP$^{\Delta pseD}$ was analysed with ANOVA statistical analysis. *p < 0.05; **p<0.01 ***p < 0.001.

The results in **figure 5.6** show that at the maxima, the peak absorbance was 3 times greater in wtMOMP than in the mutant strains. The mutant variants of the protein both lack the glycosylation on the protein, pointing towards a preference of the Le^b for the glycosylated protein. Future experiments may include competition studied between the glycosylated and non-

glycosylated MOMP variants to further determine the degree to which the BgAgs favours the glycosylated protein. This supports work by Mahdavi et al. where binding studies of the glycosylated and non-glycosylated mutant bacterial strains of NCTC1116 *C. jejuni* to BgAgs were performed *in vitro*. The results showed that the binding of 11168ΔpseD was significantly reduced compared with the wild-type (***$p < 0.001$)[110]. With regards to *in vivo* studies, they found that there was colonization of chickens by NCTC11168-O and 11168-O MOMP$^{268T/G}$ mutant 7 days post-infection. The mean colonization levels in caeca were significantly reduced ($p < 0.0001$) and below the limit of detection (10^2 cfu g^{-1}) in 8/10 birds inoculated with the mutant 11168-O MOMP$^{268T/G}$ [110].

5.3 Inhibition of MOMP binding to BgAgs by QPLEX

It has been shown that small compounds comprising an iron core complexed to 2-3 ligands such as tyrosine or quinic acid have the ability to inhibit biofilm formation and thus are useful as selective biofilm inhibiting compounds. In preliminary studies where QPLEX had been used as an anti-infective in poultry against *Campylobacter* it showed promising results for reducing the colonisation of *C. jejuni* in poultry farms[185]. We wanted to determine whether a small compound such as QPLEX – an iron (III) compound complexed to QPLEX could disrupt the interaction between the outer membrane protein and the carbohydrate BgAgs *in vitro*.

To assay the ability of QPLEX to inhibit binding between the protein and the blood group antigen the assay design from the previous study was used. In this round of experiments the MOMP was incubated with QPLEX before both were added to the ELISA wells for binding assay with Le^b.

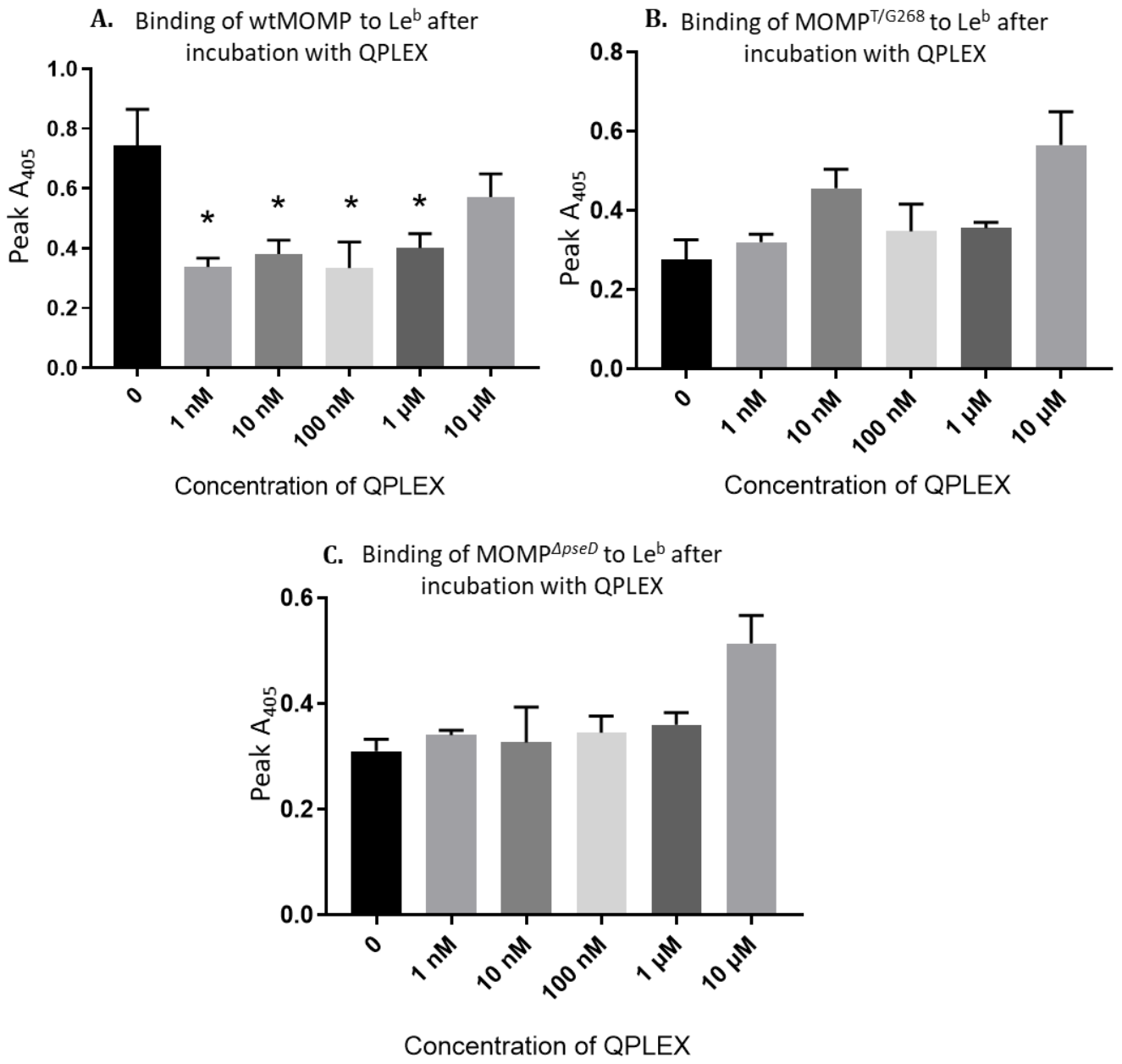


Figure 5.7: Binding of 1pM of wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ to ELISA wells adsorbed with 1 µM of Le^b in the presence of increasing concentrations of QPLEX (0 to 10 µM). This was followed by incubating with primary antibody against MOMP and alkaline phosphatase (AP) conjugated secondary antibody. The Tecan was used to measure absorbance at 405 nm by the AP substrate. The results show the maximum absorbance obtained during a 16 hr assay. Standard error bar about the means from triplicate experiments are included. The significant difference between the binding of Le^b to wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ protein at 0 concentration of QPLEX and increasing concentrations of QPLEX was calculated using ANOVA statical test, where *$p < 0.05$; **$p<0.01$ ***$p < 0.001$.

The results in **figure 5.7** show that between concentration of 1 nM and 1 µM, the QPELX reduced the maximum binding absorbance to Le^b by at least 50% in the case of wtMOMP. For the non-glycosylated protein variants – $MOMP^{T/G268}$ and $MOMP^{\Delta pseD}$, there was no significant difference in the binding of the protein to Le^b after addition of QPLEX. The lack of inhibition of the non-glycosylated MOMP strains by QPLEX shows that the glycan plays a role in increasing the affinity of the QPLEX compound for the protein. To explain why such a discriminatory binding interaction was taking place locating the site of binding of the QPLEX to the protein was important.

5.4 Locating the interaction site of ligands with MOMP

To determine the way that a protein interacts with another protein or ligand, it is imperative to consider structure. The vast majority of atomic-level membrane protein structures have been determined by X-ray crystallography[186], but some have come from utilising nuclear magnetic resonance (NMR) spectroscopy to reveal atomic resolution structures, and for lower resolution, small angle X-ray scattering and more recently, high resolution cryo-EM have been employed.[187] While transmembrane proteins such as MOMP make up 20–30% of all proteins in both prokaryotic and eukaryotic organisms[188], they represent only under 2% of known atomic-level structures, with their soluble counterparts making up the majority of resolved structures. Despite the low percentage, the majority of unique membrane proteins were reported in the past 10 years with half coming in in the last 5 years.[187] But resolved membrane proteins still remain very underrepresented and in turn, there is less information available on the interaction site of membrane proteins with ligands. Membrane proteins are more challenging than their soluble counterparts to resolve due to factors including but not limited to the recombinant expression of membrane proteins, especially for those from eukaryotes; biochemical purification and characterization is more challenging for membrane proteins than for natively soluble proteins; and

many membrane proteins are intrinsically flexible. A technical consideration is that membrane proteins must be isolated by detergent extraction from cellular lipid bilayers for purification then are reconstituted into lipidic environments for biochemical characterization or structural analysis, which further complicates structural analysis. Despite lack of information caused by these obstacles, membrane bound proteins are the molecular targets of around 60% of all FDA-approved drugs. G-protein-coupled receptors (GPCRs), ATP-binding cassette (ABC) transporters, and ion channels are targeted by drugs involved in a wide range of therapeutic areas from antibiotics to cancer[170].

5.4.1 Footprinting with carbene labelling

As well as the aforementioned techniques, structural mass spectrometry (MS) is another development in the field of proteomics that can help to identify interaction sites in proteins. Although resulting resolution is relatively low, the advantage of using MS is that it is less time-consuming and requires relatively small amounts of material compared to NMR and X-ray crystallography-based techniques.[189] MS is well established for studying covalent complexes but during the last two decades it has increasingly become accepted as a method to study protein ligand interactions and their non-covalent interfaces where contact between proteins and their ligands are primarily dynamic[190]. In this chapter, methods recently developed in MS for quantitative proteomics[191], where labelling of peptides are used to compare ligand binding with a protein, were utilised. Such data can be obtained from differential chemical labelling experiments. One such labelling chemistry involves the carbene generated from the photolysis of a diazirine functional group, which is otherwise chemically inert (see methods). Carbenes are appealing as chemical labelling reagents, as they appear to probe all amino acids, providing nonselective chemical modification rapidly and irreversibly, independent of the chemical properties of the

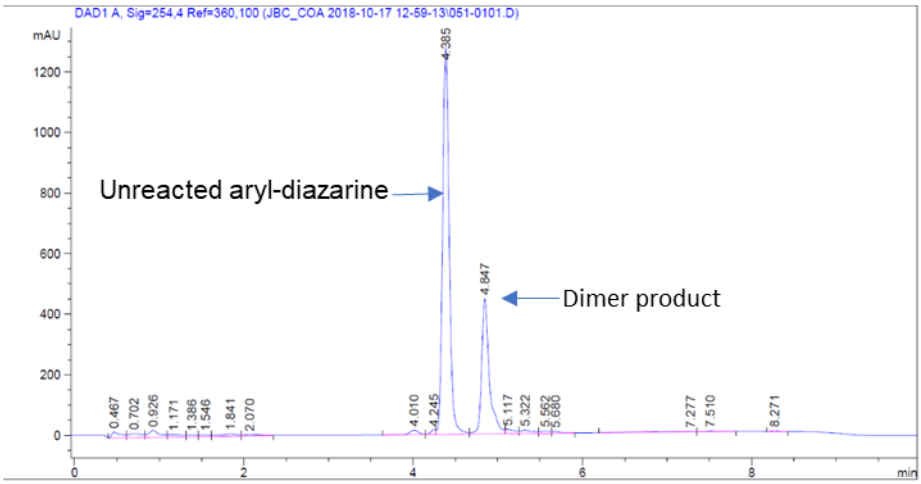
accessible residues. In this chapter, a carbene footprinting approach developed by Manzi *et al*.[192, 193] was used to map the interaction site between MOMP and QPLEX and MOMP and Le^b. It is also nonreactive toward the protein prior to activation as reactivity is only generated by use of near-ultraviolet wavelengths (E350 nm)–outside the absorbance window of amino acids. These wavelengths are generated by compact low-cost laser systems. Furthermore, protein surface labelling is independent of protein concentration within a range, tolerant of varying solution conditions and imparts clear mass shifts that are easily resolved. The technique has already been used by the group to successfully map the lysozyme substrate-binding site accurately at high resolution, and to identify changes in intramolecular interactions associated with binding[192]. In the case of an integral membrane protein, the application of diazirine was successful in footprinting the *E. coli* outer-membrane porin protein OmpF[193].

5.4.2 Footprinting efficiency of aryldiazirine in detergents

Manzi *et al.* reported the use of photoactivatable aryl diazirine for the efficient footprinting of soluble proteins. Given the amphipathic nature of the probe, they hypothesized that it may interact with detergent molecules and footprint outer transmembrane regions of micelle-solubilised protein [193]. To explore the extent to which detergent was being labelled, and thus which detergent would be better for downstream footprinting experiments, a comparative analysis was carried out between poly(ethylene glycol) octyl ether (C_8E_n (n = 2 to 9)) (octyl-POE) detergent and Octyl-beta-Glucoside (OG) detergent (**figure 5.8**). Both detergents are non-ionic detergents that contain molecules with head groups that are uncharged. Detergents that are based on polyoxyethylene and contain a neutral, polar head group and the tails are composed of hydrophobic oxyethylene chains, or, in some cases, ethylene glycoether chains. Detergents with a glycosidic base tend to use a sugar as the head group, such as glucose or maltose, and contain an

alkyl polymer tail. Non-ionic detergents have a limited ability to break protein-protein interactions. Because of this, they are considered to be non-denaturing. These detergents are effective at isolating active membrane proteins.

A.



B.

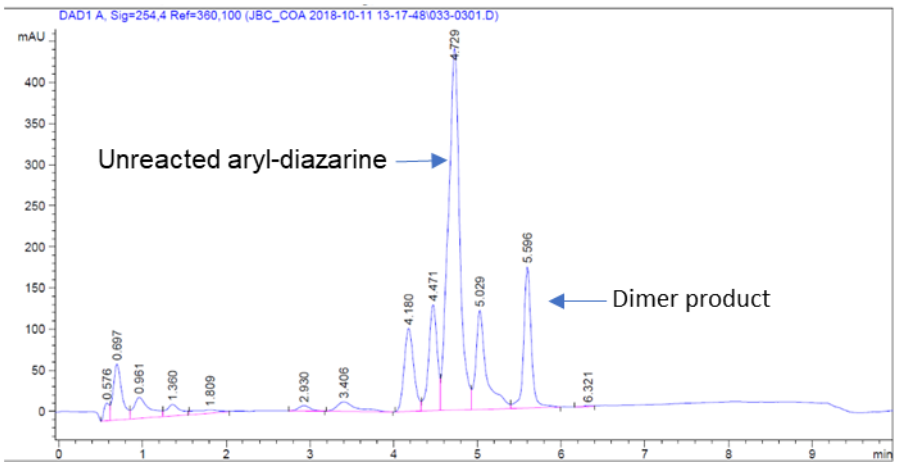


Figure 5.8: Labelling efficiency by aryl diazirine in the presence of non-ionic detergents. A. 50 mM 4-(3-Trifluoromethyl)-3H-diazirin-3-yl)benzoic acid (TDBA) was mixed with 0.45% v/v *O*-POE then irradiated for16 s before being analysed in an analytical LC (wavelengths were modified to observe probe). **B.** 50 mM TDBA was mixed with 0.45% v/v OG then irradiated for 16 s before being analysed in an analytical LC (wavelengths were modified to observe probe).

In the absence of protein, the extent to which dimerization occurs can be used as a measure of reaction efficiency. In this experiment the product is the formation of a probe dimer. The result shows that in the presence of *O*-POE, there is 25% product yield, whilst in the presence of OG, the product yield is 20% in addition to other by products. Thus, for increased labelling of outer membrane proteins such as MOMP, *O*-POE should be used as the detergent for footprint experiments.

5.4.3 Optimising protein coverage using multiple proteases

After determining that *O*-POE should be used as the detergent, labelling of MOMP with diazarine probe was optimised to get as high labelling coverage as possible. Optimization involved testing differing ratios of MOMP:probe concentration, using a combination of proteases and changing the irradiation time to obtain maximum labelling across the surface of the protein. Peptides were analysed with PepFoot (software developed by Jedediah Bellamy-Carter, 2019) to discern whether identified peptides were labelled with probe or not without having to look through the raw data for the expected peaks. To calculate the resultant fractional modification of carbene labelling, the individual data files were batch processed in PepFoot. Screenshots of this process have been included in the methods. **Figure 5.9** shows that a number of peptides of wtMOMP were significantly labelled by the carbene (>10% labelling is considered significant) whilst others did not show significant signs of labelling (<10% carbene labelling on these peptides) after digestion with trypsin.

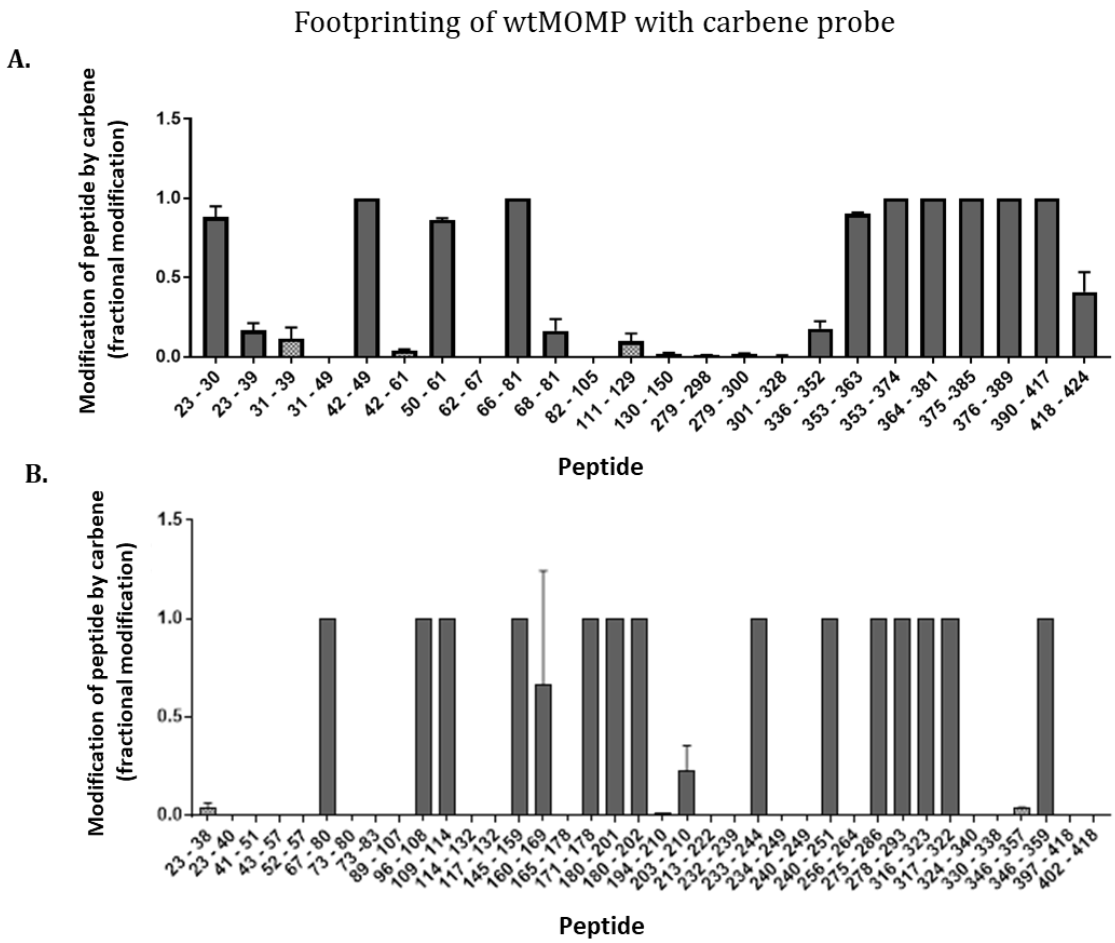


Figure 5.9: Footprinting of wtMOMP protein with carbene probe. **A.** Fractional modification of wtMOMP peptides after labelling with carbene and digestion with trypsin. Dark grey bars represent peptides from trypsin digests that were significantly modified with probe (>10% modification). Light grey bars represent peptides that were not significantly modified (0-10% modification). Each peptide was incubated with 5, 10, 25 and 50 mM of probe. Error bars are included to show the difference in masking between each concentration of probe. N = 4. **B**. Fractional modification of wtMOMP peptides after labelling with carbene and digestion with chymotrypsin. Dark grey bars represent peptides from trypsin digests that were significantly modified with probe (>10% modification). Light grey bars represent peptides that were not significantly modified (0-10% modification). Each peptide was incubated with 5, 10, 25 and 50 mM of probe. Error bars are included to show the difference in masking between each concentration of probe. N = 4. Fractional modification (fm) is determined as the ratio of chromatographic peak areas for a singly labelled peptide to the sum of non-labelled and singly-labelled peptides.

The sequence map of wtMOMP (**figure 5.10**) shows that after labelling with carbene and digestion with trypsin, significantly labelled tryptic peptides corresponded to 35% of the protein tertiary structure whilst insignificantly labelled (or unlabelled) but identified peptides amounted to 32%. The total coverage of labelled and non-labelled peptides that were identified is 67%. The data shows that there is a lack of labelled peptides between residues 82 – 335. Peptides that were not

labelled or labelled insignificantly could be due to inaccessibility of the probe to the region of the protein that corresponds to those peptides thus they could not be labelled. After taking a closer look at the data, the sequence map shows that peptides between residues 151 - 278 were not identified (or were missing) during LCMS.

A.

```
        10         20         30         40         50         60
xxxxxxxxxx xxxxxxxxxx xxTPLEEAIK DVDVSGVLRY RYDTGNFDKN FVNNSNLNNS

        70         80         90        100        110        120
KQDHKYRAQV NFSAAIADNF KAFVQFDYNA ADGGYGANGI KNDQKGLFVR QLYLTYTNED

       130        140        150        160        170        180
VATSVIAGKQ QLNLIWTDNA IDGLVGTGVK VVNNSIDGLT LAAFAVDSFM AAEQGADLLE

       190        200        210        220        230        240
HSNISTTSNQ APFKVDSVGN LYGAAAVGSY DLAGGQFNPQ LWLAYWDQVA FFYAVDAAYS

       250        260        270        280        290        300
TTIFDGINWT LEGAYLGNSL DSELDDKTHA NGNLFALKGS IEVNGWDASL GGLYYGDKEK

       310        320        330        340        350        360
ASTVVIEDQG NLGSLLAGEE IFYTTGSRLN GDTGRNIFGY VTGGYTFNET VRVGADFVYG

       370        380        390        400        410        420
GTKTEAANHL GGGKKLEAVA RVDYKYSPKL NFSAFYSYVN LDQGVNTNES ADHSTVRLQA

LYKF
```

B.



Figure 5.10: Amino acid sequence of wtMOMP with associated tryptic peptides. A. Green areas highlight the peptides within the sequence that are labelled by the carbene whilst yellow areas highlight peptides that were identified by MS but were not identified to be labelled. None highlighted areas are non identifies peptides**. B.** Mapping of peptide labelling from **figure 5.10A**. Carbene labelled identified peptides (in green) from trypsin digest of wtMOMP have been mapped onto the it's trimer structure, whose sequence was modelled onto the crystal structure obtained by Ferrara et al. (PDB accession number 5LDT). Residue of interest threonine 268 is labelled. Unlabelled identified peptides are labelled yellow and unidentified peptides are highlighted in grey.

It is difficult to determine why an expected peptide may be missing but Kim et al. have described common errors in mass spectroscopy-based analysis of proteins and their peptides.[159] A peptide may be missing because it is not present in the sample, it may be present but at a concentration below instrument detection limit, or may be present and not detected due to an unknown effect such as a post translational modification (PTM). Modifications can be introduced *in vivo* (e.g. phosphorylation, ubiquitination), *in vitro* during sample preparation (e.g. carbamidomethylation, carbamylation) or both (e.g. deamidation, acetylation, methylation). Nielsen et al. have estimated that for every unmodified peptide there are around 10 modified versions that are typically less abundant.[194] Another cause may be a lack of arginine and lysine residues in a particular region which would cause peptides that are too long and cannot be properly identified after running on the LC column causing column fouling. A peptide's chemical or physical structure may also cause column trapping. Conversely, non-identified peptides may occur when peptides are too short due to high frequency of trypsin cut sites. Missing peptides are common in MS-based proteomic data and it is common to have 20 – 40% of all attempted intensity measures missing. Extensive characterisation of a protein demands the observation of as many amino acids of that protein as possible, so to overcome gaps caused by missing peptides, the use of multiple proteases in parallel experiments increases protein coverage. Chymotrypsin was used in separate parallel protein digest steps alongside trypsin to generate alternative labelled peptide fragments (**figure 5.11**). The fractional modification of peptides post chymotrypsin digestion (**figure 5.11**) was analysed by PepFoot.

A.

```
         10         20         30         40         50         60
xxxxxxxxxx xxxxxxxxxx xxTPLEEAIK DVDVSGVLRY RYDTGNFDKN FVNNSNLNNS

         70         80         90        100        110        120
KQDHKYRAQV NFSAAIADNF KAFVQFDYNA ADGGYGANGI KNDQKGLFVR QLYLTYTNED

        130        140        150        160        170        180
VATSVIAGKQ QLNLIWTDNA IDGLVGTGVK VVNNSIDGLT LAAFAVDSFM AAEQGADLLE

        190        200        210        220        230        240
HSNISTTSNQ APFKVDSVGN LYGAAAVGSY DLAGGQFNPQ LWLAYWDQVA FFYAVDAAYS

        250        260        270        280        290        300
TTIFDGINWT LEGAYLGNSL DSELDDKTHA NGNLFALKGS IEVNGWDASL GGLYYGDKEK

        310        320        330        340        350        360
ASTVVIEDQG NLGSLLAGEE IFYTTGSRLN GDTGRNIFGY VTGGYTFNET VRVGADFVYG

        370        380        390        400        410        420
GTKTEAANHL GGGKKLEAVA RVDYKYSPKL NFSAFYSYVN LDQGVNTNES ADHSTVRLQA

LYKF
```

B.

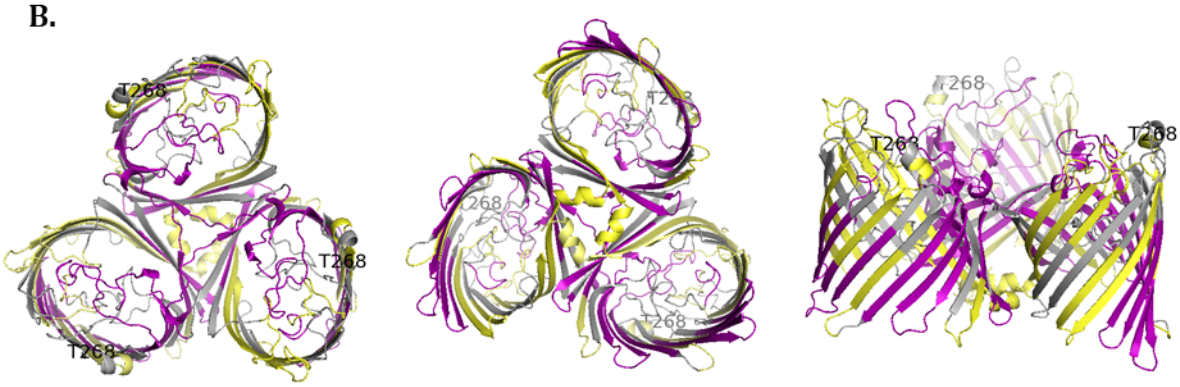


Figure 5.11: Amino acid sequence of wtMOMP with associated chymotryptic peptides. **A.** Purple areas highlight the peptides within the sequence that are labelled by the carbene whilst yellow areas highlight peptides that were identified by MS but were not identified to be labelled. No highlighted areas are non-identified peptides**. B.** Mapping of peptide labelling from panel A. Carbene labelled identified peptides (in purple) from chymotrypsin digest of wtMOMP have been mapped onto the trimer structure, whose sequence was modelled onto the crystal structure obtained by Ferrara et al. (PDB accession number 5LDT). Residue of interest threonine 268 is labelled. Unlabelled identified peptides are highlighted in yellow and unidentified peptides are highlighted in grey.

Figure 5.11 shows that after digestion with chymotrypsin, 36% of the protein exhibited labelled peptides in MS. Likewise to trypsin, there are regions of the protein that are identified and do not display labelling which amount to 32% of the protein. The total protein coverage (labelled and unlabelled) was 68%.

Separate digestion of MOMP with trypsin and chymotrypsin does not remove the areas of non-identified peptides but in combination in parallel experiments these gaps are overcome. Combining data of peptides that were labelled and digested with trypsin and chymotrypsin in parallel experimental set ups resulted in protein labelling coverage of 69% (**figure 5.12B**) and 55% coverage for total identified peptides that were not labelled. Total identified peptides of either labelled or unlabelled across both chymotrypsin and trypsin was 95% (**figure 5.12C**). In **figure 5.12**, the total labelled peptides from both digests on the wtMOMP sequence has been mapped onto the model of wtMOMP to locate the region of labelled peptides. **Figure 5.9** shows that although some areas may be labelled, not all the labelling is significant (>10%). The graph shows that residues 146 to 170 and residues 361 to 392 are not significantly labelled. There is a large gap in trypsin cleaved labelled peptides (residues 82 – 335). This gap is compensated for by chymotrypsin cleaved labelled peptides. Even when both sets of cleaved proteins are overlaid as in **figure 5.12B** there is still a lack of labelled peptides between 115 – 145, 211-231, 251-274, 294-314.

From first look it would seem as if MOMP footprinting by diazirine at the peptide level was found to be sufficient to map the buried and surface-accessible regions of the porin in Octyl -POE (**figure 5. 12B**), however further information can be obtained into the location of modification within the labelled peptides (especially for sites that are labelled by less than 100%) by secondary mass spectroscopy of these specific peptides. **Figure 5.12B** shows that the interface of the trimer is labelled whilst some outward facing regions are still to be labelled. MSMS would identify which particular amino acids of each peptide is labelled, be they amino acids with side groups that are inward facing outward facing of the porin.

A.

```
          10         20         30         40         50         60
xxxxxxxxxx xxxxxxxxxx xxTPLEEAIK DVDVSGVLRY RYDTGNFDKN FVNNSNLNNS

          70         80         90        100        110        120
KQDHKYRAQV NFSAAIADNF KAFVQFDYNA ADGGYGANGI KNDQKGLFVR QLYLTYTNED

         130        140        150        160        170        180
VATSVIAGKQ QLNLIWTDNA IDGLVGTGVK VVNNSIDGLT LAAFAVDSFM AAEQGADLLE

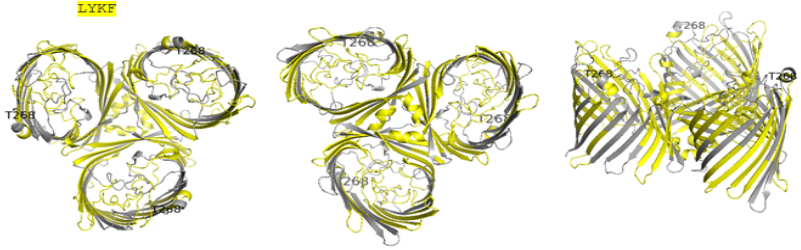
         190        200        210        220        230        240
HSNISTTSNQ APFKVDSVGN LYGAAAVGSY DLAGGQFNPQ LWLAYWDQVA FFYAVDAAYS

         250        260        270        280        290        300
TTIFDGINWT LEGAYLGNSL DSELDDKTHA NGNLFALKGS IEVNGWDASL GGLYYGDKEK

         310        320        330        340        350        360
ASTVVIEDQG NLGSLLAGEE IFYTTGSRLN GDTGRNIFGY VTGGYTFNET VRVGADFVYG

         370        380        390        400        410        420
GTKTEAANHL GGGKKLEAVA RVDYKYSPKL NFSAFYSYVN LDQGVNTNES ADHSTVRLQA
```

LYKF



B.

```
          10         20         30         40         50         60
xxxxxxxxxx xxxxxxxxxx xxTPLEEAIK DVDVSGVLRY RYDTGNFDKN FVNNSNLNNS

          70         80         90        100        110        120
KQDHKYRAQV NFSAAIADNF KAFVQFDYNA ADGGYGANGI KNDQKGLFVR QLYLTYTNED

         130        140        150        160        170        180
VATSVIAGKQ QLNLIWTDNA IDGLVGTGVK VVNNSIDGLT LAAFAVDSFM AAEQGADLLE

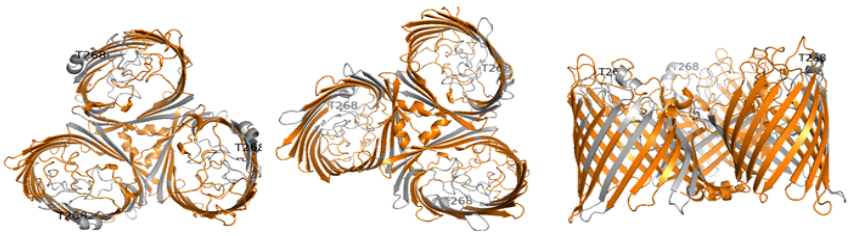
         190        200        210        220        230        240
HSNISTTSNQ APFKVDSVGN LYGAAAVGSY DLAGGQFNPQ LWLAYWDQVA FFYAVDAAYS

         250        260        270        280        290        300
TTIFDGINWT LEGAYLGNSL DSELDDKTHA NGNLFALKGS IEVNGWDASL GGLYYGDKEK

         310        320        330        340        350        360
ASTVVIEDQG NLGSLLAGEE IFYTTGSRLN GDTGRNIFGY VTGGYTFNET VRVGADFVYG

         370        380        390        400        410        420
GTKTEAANHL GGGKKLEAVA RVDYKYSPKL NFSAFYSYVN LDQGVNTNES ADHSTVRLQA
```

LYKF



C.

```
          10         20         30         40         50         60
xxxxxxxxxx xxxxxxxxxx xxTPLEEAIK DVDVSGVLRY RYDTGNFDKN FVNNSNLNNS

          70         80         90        100        110        120
KQDHKYRAQV NFSAAIADNF KAFVQFDYNA ADGGYGANGI KNDQKGLFVR QLYLTYTNED

         130        140        150        160        170        180
VATSVIAGKQ QLNLIWTDNA IDGLVGTGVK VVNNSIDGLT LAAFAVDSFM AAEQGADLLE

         190        200        210        220        230        240
HSNISTTSNQ APFKVDSVGN LYGAAAVGSY DLAGGQFNPQ LWLAYWDQVA FFYAVDAAYS

         250        260        270        280        290        300
TTIFDGINWT LEGAYLGNSL DSELDDKTHA NGNLFALKGS IEVNGWDASL GGLYYGDKEK

         310        320        330        340        350        360
ASTVVIEDQG NLGSLLAGEE IFYTTGSRLN GDTGRNIFGY VTGGYTFNET VRVGADFVYG

         370        380        390        400        410        420
GTKTEAANHL GGGKKLEAVA RVDYKYSPKL NFSAFYSYVN LDQGVNTNES ADHSTVRLQA
```

LYKF



Figure 5.12: Amino acid sequence and tertiary structure of wtMOMP. **A.** Protease peptides from trypsin and chymotrypsin that were identified but not carbene labelled have been highlighted yellow and mapped onto the structure of wtMOMP**. B.** Peptides from trypsin and chymotrypsin digestion that were found to be carbene labelled have been highlighted orange. **C.** All identified peptides whether labelled or unlabelled from both chymotrypsin and trypsin have been highlighted in red. wtMOMP sequence was modelled onto the crystal structure obtained by Ferrara et al. (PDB accession number 5LDT). Residue of interest threonine 268 is labelled.

Although the initial goal was to increase the extent of identified peptides labelled with the carbene (**figure 5.12B**), it is just as important to consider all identified peptides (**figure 5.12C**), including those that did not undergo labelling (**figure 5.12A**), as changes in protein conformational change in ligand binding experiments may cause increased labelling of these peptides. Thus, it is important to be able to track the identifiable but non labelled peptides as well. The same is true for the opposite where a change in the conformation of the protein may cause previously exposed areas that were labelled with the carbene to become concealed and thus there will be a decrease in labelling in that region of the protein. In **figure 5.12C,** all identifiable peptides have been highlighted which accounts for 95% of the protein. In fact, of the only 2 regions that were missing, one region is the area of the protein that contains our residue of interest - threonine 268 (THR-268). THR-268, was not observed as either a labelled or unlabelled peptide. This was expected because the peptide in its wild type form does not present in the non -glycosylated state so the correct parameters need to be selected for in order to observe whether this peptide is labelled or not. See chapter 4 for more work on this. Once the size of that peptide has been determined then we might be able to scan the spectrum for the glycosylated peptide in its non-labelled and labelled form. Further, the peptide in region 222-232 was also non-identified, this could point to another region of PTM on wtMOMP such as deamidation, as previously discussed (chapter 4.3).

5.4.4 Footprinting of MOMP with QPLEX

Following assessment of the extent of labelled and non-labelled peptides from both trypsin and chymotrypsin digest with the diazirine based carbene, differential experiments to compare the

fractional modification of the peptides between different treatments were performed to identify the ligand binding site on the protein. Binding studies were performed between wtMOMP and small synthetic and biological ligands. Studies (see chapter 5) had revealed significant efficacy between wtMOMP and QPLEX and wtMOMP and Le^b and thus these pairings were explored for locating ligand interaction sites on the wtMOMP trimer protein. The results from the previous experiment to obtain protease and carbene fingerprinting patterns were used as negative controls against carbene labelling of the peptides in the presence of the ligands.

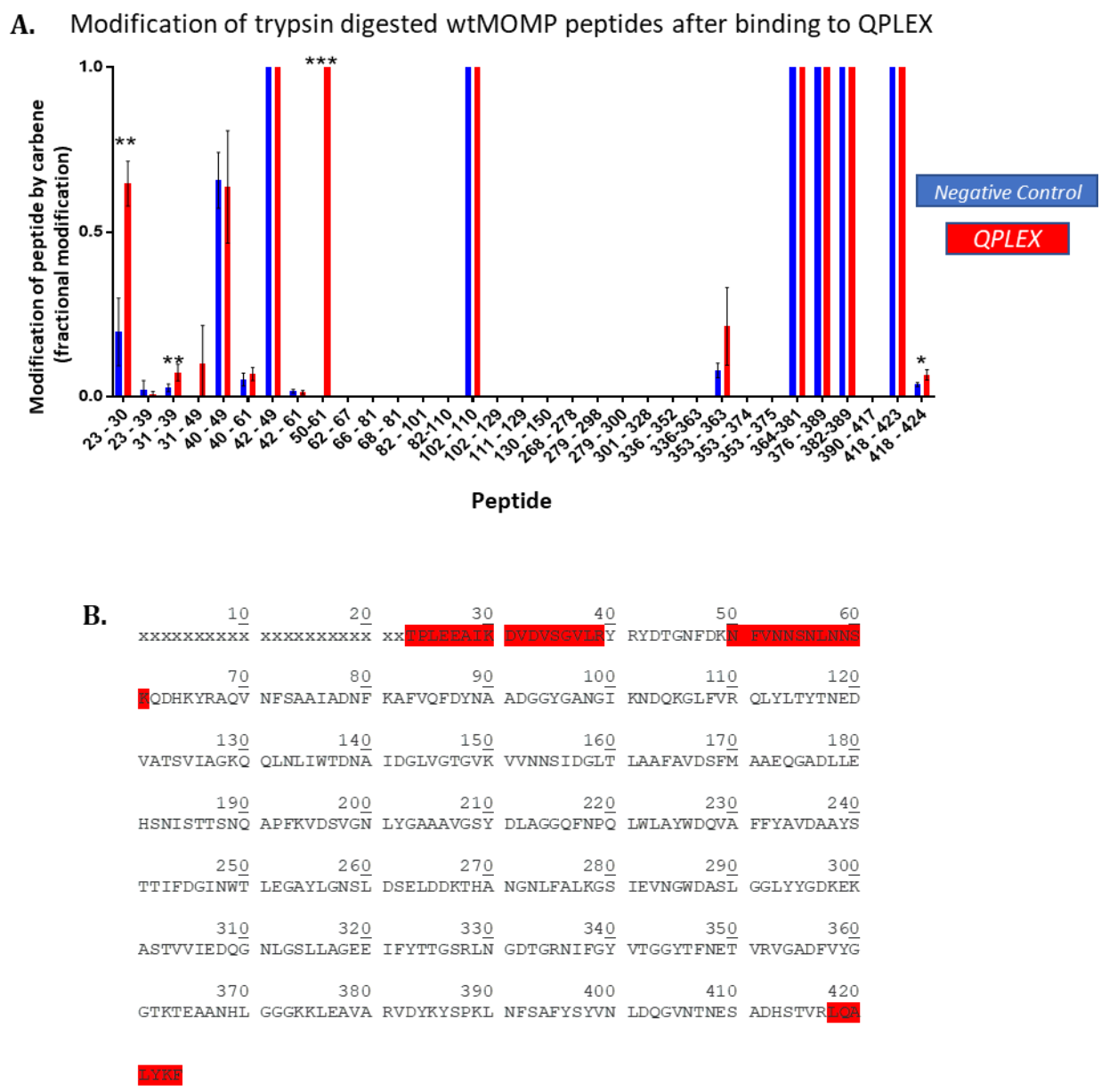


Figure 5.13: Labelling of wtMOMP with carbene in the presence of QPLEX reveals increased labelling of some peptides A. Fractional modification of tryptic wtMOMP peptides after carbene labelling in the absence and presence of 10 µM QPLEX. Each peptide has been given a peptide ID consistent with the position of the amino acid residues in the protein monomer. Blue bars indicate fractional modification of the native peptide (control) whilst red bars indicate fractional modification of the native peptides after addition of the small compound, QPLEX – an established ligand. All experiments were performed in triplicate and the mean peptide fractional modification is included with standard error bar about the mean. The significance level of the statistical difference between the control peptide and peptide in presence of the ligand is shown by asterisks (*); * = $p \leq 0.05$, ** =$p \leq 0.01$,*** = $p \leq 0.001$, analysed by paired student t-test. **B.** Peptides where a significant change in labelling occurs have been mapped onto the sequence of wtMOMP in red.

In **figure 5.13** it can be seen that QPLEX caused a significant increase in labelling at four peptides;

23 -30, 31-39, 50-61 and 418-424. Fractional modification of carbene labelling of peptide 23-30

increased from 0.2 to 0.65 in the presence of QPLEX, carbene labelling of peptide 31-39 increased from 0.03 to 0.07, carbene labelling of peptide 50-61 increased from 0% fractional modification to 100% fractional modification. Finally, fractional modification of peptide 418 to 424 increased from 0.04 to 0.07. To further probe areas that were not identified using tryptic peptides, chymotrypsin was also used.

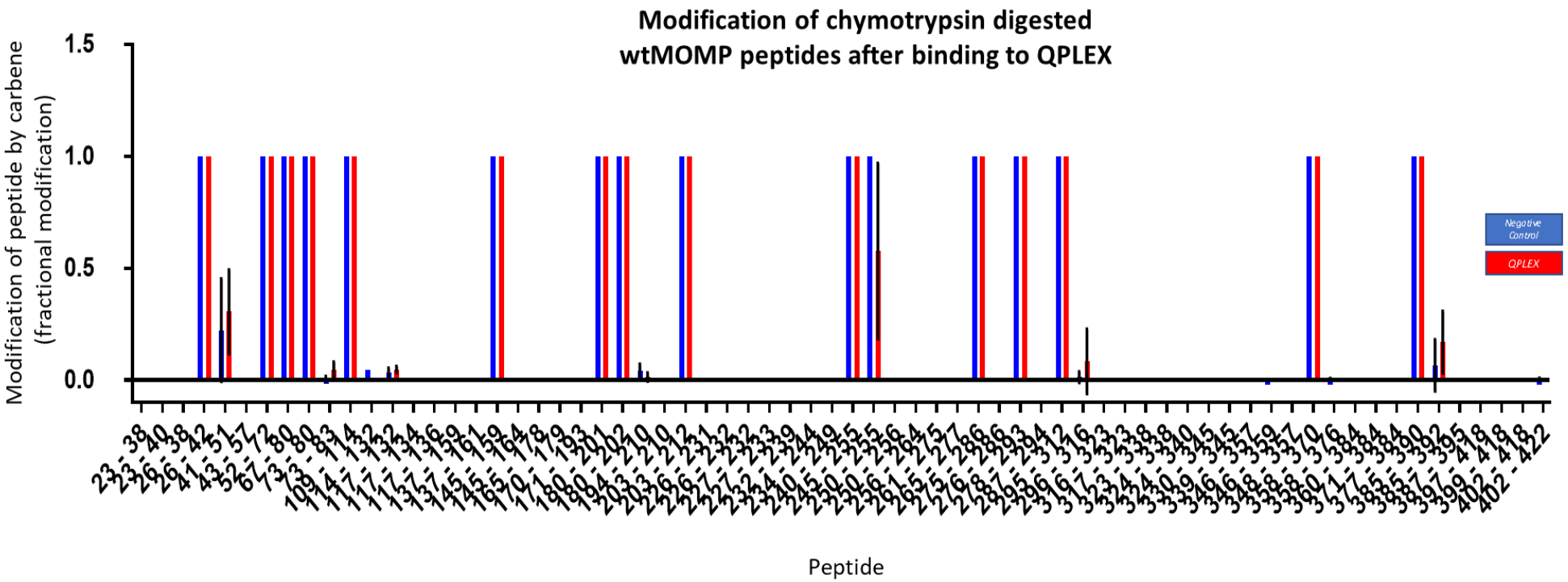


Figure 5.14: Fractional modification of chymotryptic wtMOMP peptides after carbene labelling in the absence and presence of 10 μM QPLEX. Each peptide has been given a peptide ID consistent with the position of the amino acid residues in the protein monomer. Blue bars indicate fractional modification of the native peptide (control) whilst red bars indicate fractional modification of the native peptides after addition of the small compound, QPLEX – an established ligand. All experiments were performed in triplicate and the mean peptide fractional modification is included with standard error bar about the mean.

Upon further inspection of data obtained from the use of chymotrypsin (**figure 5.14**), it was decided that labelled chymotryptic peptides would not be carried forward to make conclusions about ligand binding. Unlike the tryptic peptides which resulted in labelled peptides with continuous data (**figure 5.13**), chymotryptic peptides resulted in discretely labelled peptides (either 0 or 100% labelled), which makes it more difficult to draw reliable conclusions about differential extent of masking. This could be caused by multiple reasons. A critique of PepFoot is that if a peptide in the master sample that is being analysed is 0% or 100% carbene labelled then the other replicates are also classified as such, and it would require a lot of user time to analyse each triplicate sample separately. Furthermore, chymotrypsin preferentially cleaves protein chains at the carboxyl side of aromatic amino acids. This preference for aromatic amino acids may be interfering with the aromatic structure of the aryl based carbene labelling and further causing issues leading to discrete labelling. So only the tryptic peptides in the absence and presence of QPLEX were used to map where on the protein structure where the ligand may interact (**figure 5.15**).

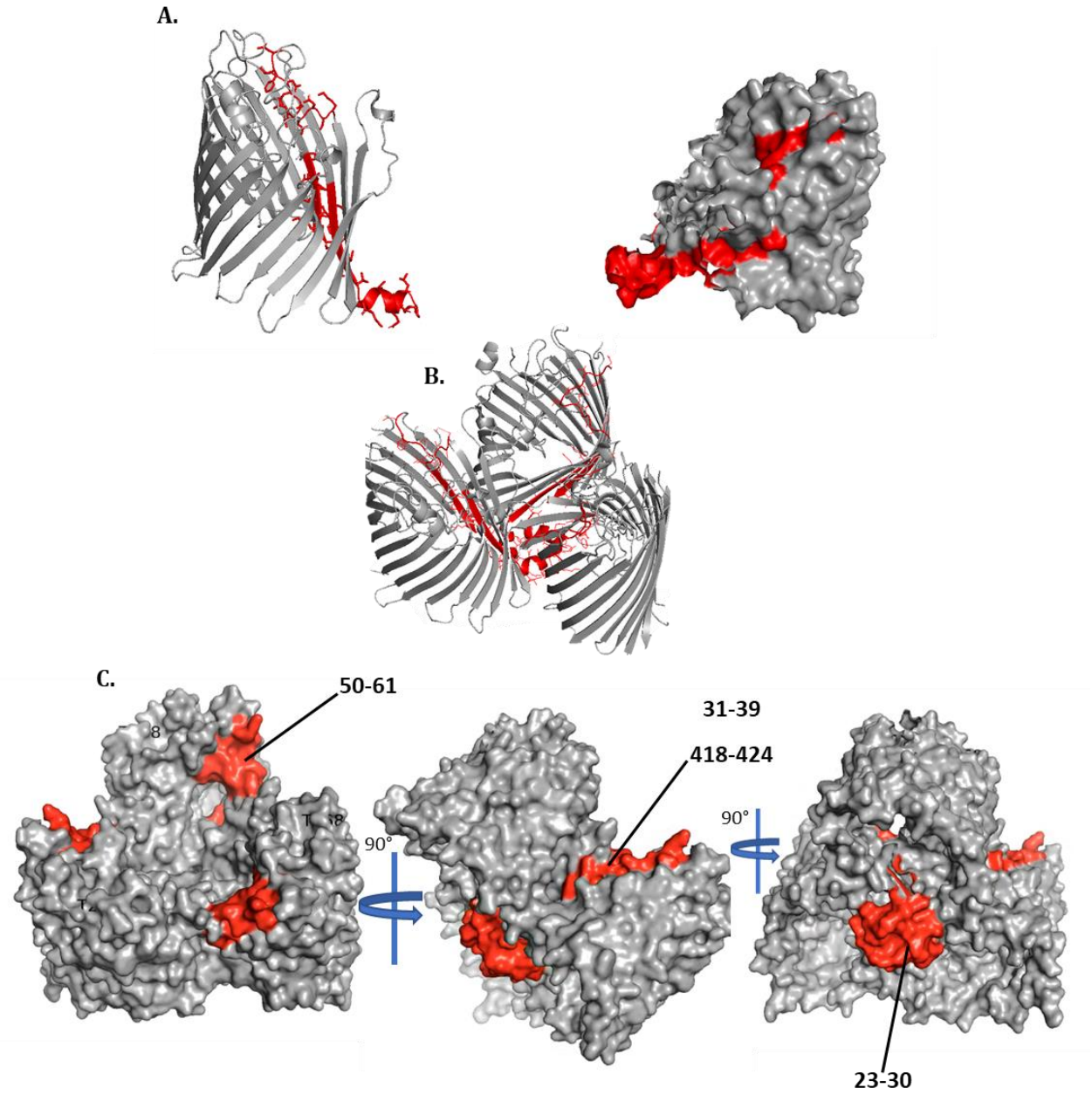


Figure 5.15: Resultant carbene footprinting of wtMOMP with the aryl diazirine probe after incubation with QPLEX. Regions in in red show areas of significant change in labelling ($p \leq 0.05$) after addition of the QPLEX ligand whilst grey areas represent show where there was no significant change in labelling after incubation with QPLEX. **A.** wtMOMP in the monomer state to show specific regions on the protein (represented in cartoon and surface style). **B.** wtMOMP in its trimer form represented in a cartoon style **C.** wtMOMP in its trimer form represented by a surface diagram to show areas where there is a significant change in carbene labelling. From left to right: extracellular region of MOMP, interface region of MOMP monomers, underside of MOMP protein. The numbers represent the amino acid residues (or peptide ID). The surface diagrams reveal that there was an increase in carbene labelling on the outer surface of the protein within cleft like regions.

The differential footprinting of wtMOMP in the presence and absence of QPLEX has revealed possible interaction sites of the ligand leading to conformational change of the protein. The results show that after carbene labelling in the presence of QPLEX, carbene labelling on the peptides shown in **figure 5.9** were significantly reduced ($p<0.05$). This result may seem counter intuitive at first as it is expected that the introduction of a ligand would decrease the amount of carbene labelling on the protein, rather than increase. An explanation of this may be that the presentation of the ligand to the protein changes the conformation of the wtMOMP, thus revealing areas that were initially inaccessible to the carbene and therefore increasing labelling efficiency at these newly revealed sites.

As can be seen in **figure 5.15**, the extent of carbene labelling significantly increased at peptide residue 23-30, 50-61 and 31-39 which is attached to residue 418 to 424. Residues 31-39 and 418 to 424 correspond with the interface of the protein - where the wtMOMP monomer units form a trimer via interaction of the hydrophobic face of the monomer. The reduction of labelling in this region points towards a change in conformation once QPLEX is presented to the protein. The most obvious conclusion of this change would be that, after binding of the ligand to the extracellular side of the protein where the flexible loops lie, the QPLEX compound travels further into the porin and into the constriction zone of the monomer where the hydrophilic amino acids lie. This results in exposing more of the interface region of the protein to the carbene thus increasing the masking after addition of QPLEX. MS/MS studies will further confirm this conclusion by revealing which individual residues are labelled (in addition to whole peptides) by fragmenting the peptide into its component labelled or non-labelled amino acid residues.

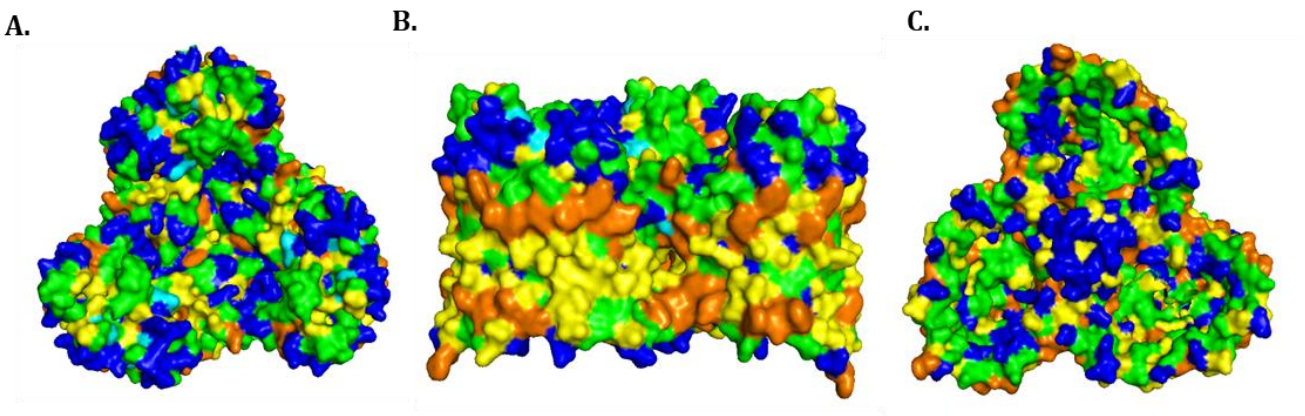


Figure 5.16: Hydrophilic regions are shown as green and blue (wet, watery colours) whilst hydrophobic regions shown as yellow and orange patches (dry desert colours) on the protein surface of wtMOMP. Colour code: Orange = Tyr+Phe+Trp, yellow = Leu+Ile+Val+Met+Pro+Ala+Cys, green = Asn+Gln+Ser+Thr+Gly, cyan = His, blue = Arg+Lys+Asp+Glu. Protein orientation: **A.** Birds eye view of extracellular side of wtMOMP, **B.** Side view of protein from within the bilayer, **C.** View of underside of the protein from the intracellular side.

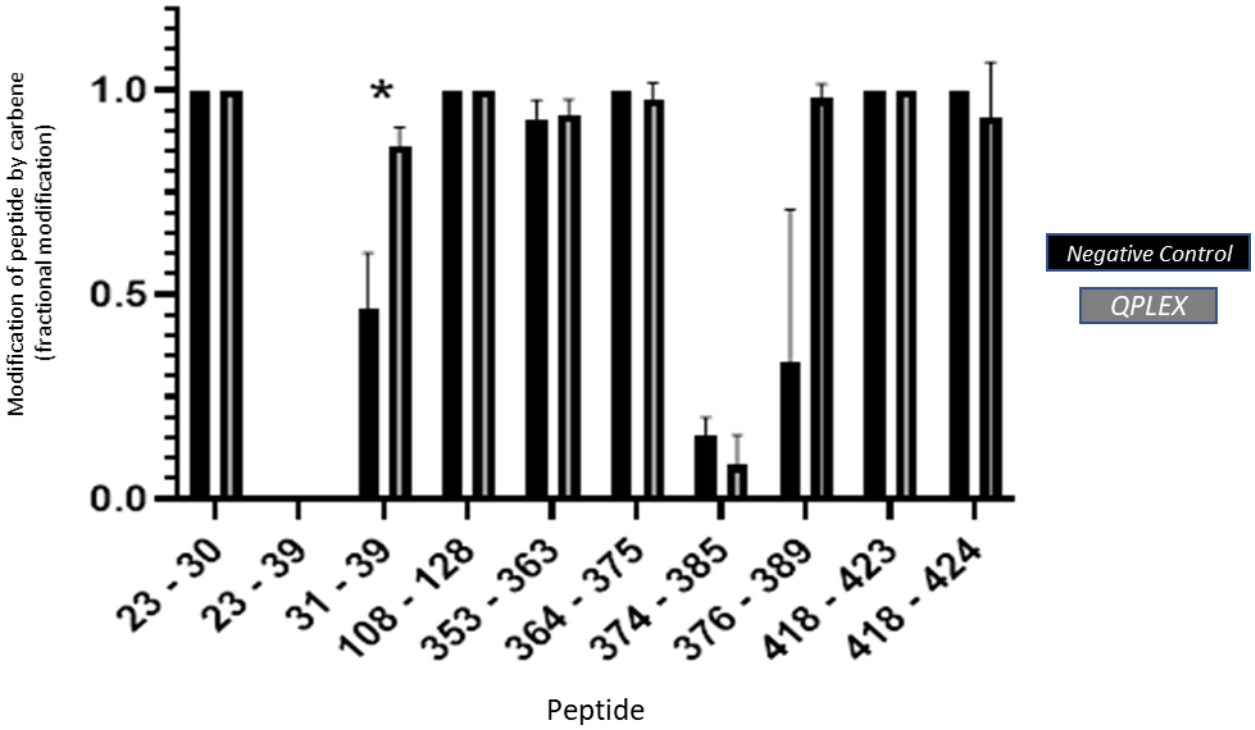
Residues 23-30 lie on the underside (the intracellular region) of the wtMOMP protein and is on the *N*-terminus on the periplasmic face which has a strand, followed by an *N*-terminal α-helix that points away from the barrel wall. This region within the bacterial cell plasma membrane would not be accessible to any small compounds or drugs from external sources as it lies on the intracellular side of the protein in the periplasm. **Figure 5.15** suggests that the compound travels through the porin and then the final stop for the ligand before being released into the intracellular region is around residues 23-30. Further study of the region shows that it is a highly hydrophilic region (**figure 5.16**) with high charge where ligands such as QPLEX are likely to make interactions. So, an increase in carbene labelling in this region suggests that the addition of QPLEX to the protein is causing binding in one area that results in conformational change in residues 23-30.

Residues 50-61 are found in the surface exposed hydrophilic loops on the extracellular side of the protein. Loops are structures of protein that do not have a defined secondary structure, unlike beta sheets or alpha helices. Rather, loops connect these secondary structures within the protein; for example, beta sheets within wtMOMP are connected via loops to form barrels. wtMOMP has

four extracellular loops L1, L3, L4, and L6 which fold inside of the barrel, with L3, L4, and L6 forming the constriction zone[120]. wtMOMP has a similar structure to OmpA, a ß-barrel membrane protein from *E. coli.* Some of the first studies found that the loops in OmpA were absent of topogenic signals (via loop size or charge distribution), the shortening of the loops didn't reduce thermal stability of the protein and that the loops did not serve as a receptor of the OmpA bacteriophage. But the same study found that all loops were necessary for the OmpA protein to function in the stabilization of mating aggregates during F conjugation.[195] Loops have been found to be on binding sites to mediate binding of antigens to immunoglobulins, toxins to protein receptors, metal ions to proteins, DNA to DNA-binding proteins and protein substrates to serine proteases.[196] Furthermore, the Davin-Regli group in 2017 determined that cell-surface-exposed loops L5 and L7 have a role in membrane physiology in *Providencia stuartii.*[197] L5 and L7 interact tightly at the cell surface in a coordinated manner and variability in this region can lead to antibiotic resistant strains, such as in antibiotic resistant *Enterobacteriaceae* induced by a structural modification of the major porin at L3[198]. The variability of the amino acids in the loop region of wtMOMP, including those in residues 50-61, may contribute to the flexibility of exposed charges, thus regulating electrostatic interactions with other membrane components like the lipooligosaccharide (LOS), cell to cell interchange within the biofilm or external molecules, such as QPLEX, interacting with the bacterial surface. An increase in carbene labelling of the peptide 50-61 in the presence of the QPLEX compound signals that some part of the loops become exposed after the addition of QPLEX. To determine which particular residues are being revealed, MS/MS studies should be performed to further which individual residues are labelled.

5.4.5 Footprinting of MOMP with Le^b

A. Modification of trypsin digested wtMOMP peptides after binding to Le^b



B.

```
          10         20         30         40         50         60
xxxxxxxxxx xxxxxxxxxx xxTPLEEAIK DVDVSGVLRY RYDTGNFDKN FVNNSNLNNS

          70         80         90        100        110        120
KQDHKYRAQV NFSAAIADNF KAFVQFDYNA ADGGYGANGI KNDQKGLFVR QLYLTYTNED

         130        140        150        160        170        180
VATSVIAGKQ QLNLIWTDNA IDGLVGTGVK VVNNSIDGLT LAAFAVDSFM AAEQGADLLE

         190        200        210        220        230        240
HSNISTTSNQ APFKVDSVGN LYGAAAVGSY DLAGGQFNPQ LWLAYWDQVA FFYAVDAAYS

         250        260        270        280        290        300
TTIFDGINWT LEGAYLGNSL DSELDDKTHA NGNLFALKGS IEVNGWDASL GGLYYGDKEK

         310        320        330        340        350        360
ASTVVIEDQG NLGSLLAGEE IFYTTGSRLN GDTGRNIFGY VTGGYTFNET VRVGADFVYG

         370        380        390        400        410        420
GTKTEAANHL GGGKKLEAVA RVDYKYSPKL NFSAFYSYVN LDQGVNTNES ADHSTVRLQA


LYKF
```

Figure 5.17: A. Fractional modification of tryptic wtMOMP peptides after carbene labelling in the absence and presence of 35 μM Le^b. Each peptide has been given a peptide ID consistent with the position of the amino acid residues in the protein monomer. Black bars indicate fractional modification of the native peptide (control) whilst grey bars indicate fractional modification of the native peptides after addition of the sugar antigen, Le^b. All experiments were performed in triplicate and the mean

peptide fractional modification is included with standard deviation bar (n ≥ 3). The significance level of the statistical difference between the control peptide and peptide in presence of the ligand is shown by asterisks (*); * = p ≤ 0.05. **B.** Peptides where a significant change in labelling occurs have been mapped onto the sequence of wtMOMP in red.

The graph of fractional modification of wtMOMP after incubation with Le^b shows that only one are of the protein is modified. In **figure 5.17**, Le^b caused a significant increase in carbene labelling at peptide 31-39. Fractional modification of carbene labelling of peptide 31-39 increased from 0.47 to 0.86 in the presence of Le^b. The peptide sequence in this region is DVDVSGVLR.

It was expected that there would be more regions of wtMOMP that resulted in a significant change in peptide carbene masking after incubation with Le^b, especially because Le^b is a larger compound than QPLEX. The lack of such a result may be due to the change in protocol that was used in the LCMS. Usually, a Linear Ion Trap Mass spectrometer, LTQ, combined with a Fourier Transform (FT) Ion Cyclotron Resonance Mass Spectrometer would be used to detect peptide species, due to the high resolving power (can be up to 2,000,000 FWHM for FT-LCMS). But due to a faulty FT in the latest round of experiments, to measure the extent of modification of wtMOMP in the presence of Le^b, the FT was removed and the resolving power of the LCMS instrument was significantly reduced to up to 60,000 FWHM. Removal of the FT also meant that the scan speed, which refers to the scans that can be performed per second, was also altered. With the attachment of the FT, the scan only sweeps through 1 scan per second but for LC-MS a high number of scans (around 500) per peak is required to perform correct peak picking and mass spectral deconvolution of full scan mass spectra. Both of these alterations result in acquiring more data, due to more scans, but with lower resolution, which undoubtedly affected the result of the second round of foot printing experiments with Le^b. Thus, future experiments should rerun the trypsin digested MOMP/Le^b sample with a complete FT-LCMS unit.

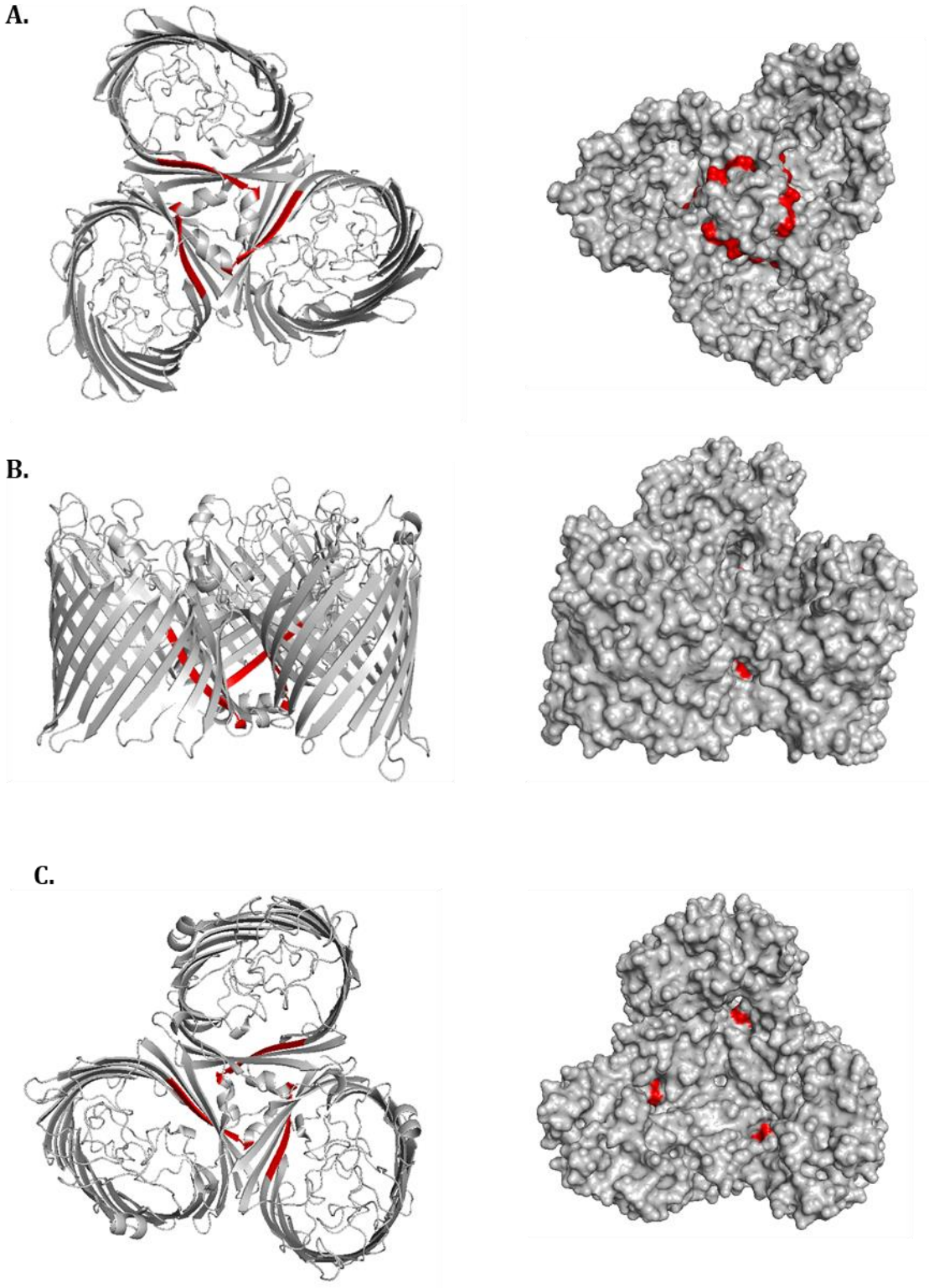


Figure 5.18: Resultant carbene footprinting of wtMOMP with the aryl diazirine probe after incubation with Le^b. Peptide regions represented in red (from **figure 5.18**) show areas of significant change in labelling ($p \leq 0.05$) after addition of Le^b. **A.** View of underside of the protein from the intracellular side. **B.** Side view of protein from within the bilayer. **C.** Birds eye view of extracellular side of wtMOMP. The wtMOMP is depicted in the trimer form to show localisation of the ligand binding region about the whole protein. Grey regions represent areas of the protein where there was no

significant change in labelling after incubation with QPLEX. The ribbon (left) and surface (right) diagrams reveal that there is an increase in carbene labelling on the underside of the protein.

Similar to the change seen in QPLEX, there is a change in residue 31-39 which corresponds with the interface of the protein - where the wtMOMP monomer units form a trimer via interaction of the hydrophobic face of the monomer. The reduction of labelling in this region points towards a change in conformation once Le^b is presented to the protein. The most obvious conclusion of this change would be that the Le^b compound is binding in the constriction zone of the monomer where the hydrophilic amino acids lie and thus revealing the interface region of the protein to the carbene. MS/MS studies will further confirm this hypothesis by revealing which individual residues are labelled (in addition to whole peptides) by fragmenting the peptide into its component labelled or non-labelled amino acid residues.

But the areas where there is a significant loss, rather than gain, of carbene labelling due to the occurrence of ligand binding in that site are still to be determined so *in silico* studies can help to further verify areas of ligand binding.

5.5 Protein ligand docking

Docking is a computational technique that places a small molecule such as a ligand in the binding site (pose) of its macromolecular target, the receptor, and estimates its binding affinity to the receptor. Determining the location of the ligand binding site is a hallmark of targeted drug design.

Here protein-ligand docking was performed via FRED RECEPTOR 3.5.0.4 software (OpenEye Scientific Software, Inc., Santa Fe, NM, USA; www.eyesopen.com). To be effective at hit discovery and lead optimization, a docking program must be able to correctly place a ligand within the

receptor site and estimate its binding affinity. FRED performs an exhaustive, systematic and non-stochastic examination of all possible protein-ligand poses. Exhaustive searching systematically samples all possible rotations and translations of the ligand to a given resolution rather than randomly sampling poses within the site in a stochastic manner. Although both the receptor and ligand are rigid during docking, ligand flexibility is implied by docking the various ligand conformers. FRED filters for shape complementarity and chemical feature alignment before selecting and optimizing poses using the Chemgauss4 scoring function, which produces the ΔG value. The Chemgauss4 functions that were selected for were the steric score, clash score, protein desolvation score, ligand desolvation score, and the score of HB interactions between the ligand and solvent and hydrogen bond interactions between ligand and protein. The higher (and more negative) the magnitude of the Chemgauss-4 score, the tighter the binding of ligand with receptor through non-covalent forces of interactions. [199]

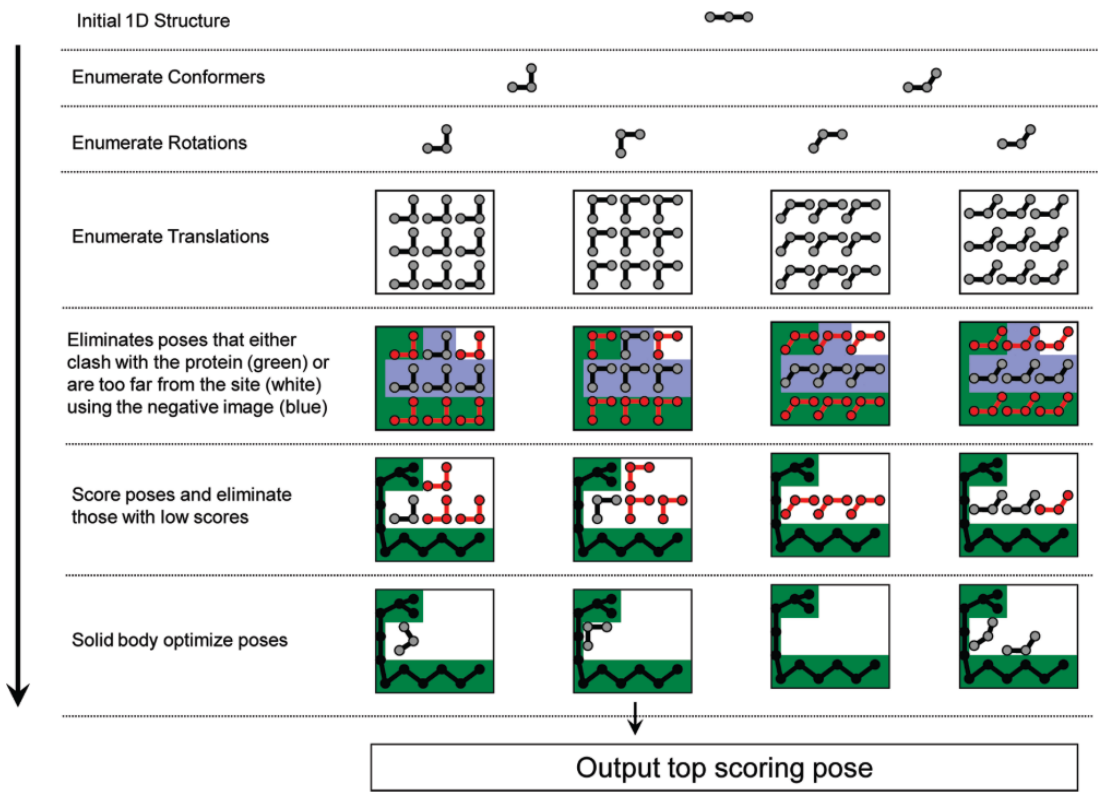


Figure 5.19: Schematic of the FRED docking process. FRED docks molecules using an exhaustive search algorithm that systematically searches rotations and translations of each conformer of the ligand within the active site at a specified resolution. During the exhaustive search unrealistic poses are filtered, and the surviving poses are scored and ranked.

The Chemgauss4 scoring function uses Gaussian smoothed potentials to measure the complementarity of ligand poses within the active site. For a ligand docked into the receptor, Chemgauss4 recognizes, hydrogen bonding between ligand and protein, the interactions of shape and hydrogen bonding interactions with implicit solvent.

To calculate the hydrogen bonding score for a ligand-protein hydrogen bond two distances are measured; 1) How far the donor heavy atom is from the position the acceptor atom would consider to be an ideal for a hydrogen bonding to form and 2) how far the acceptor heavy atom is from the position the donor atom would consider to be ideal for a hydrogen bonding interaction to occur. The score for the hydrogen bond interaction is a product of two Gaussian functions of these distances scaled by the strength of the hydrogen bonding groups involved in the interaction where HBscore = strength*g(distance1)*g(distance2). To compute the total hydrogen bonding score for the ligand-protein complex the individual pairwise scores are calculated for all protein-ligand donor-acceptor. Individual HB interactions are then eliminated if either the donor or acceptor exceeds the maximum number of interactions allowed (e.g., a hydroxyl with one hydrogen is not allowed to make more than one donor interaction), with the lowest scoring interactions eliminated first. The final hydrogen bond score is then calculated by summing the scores of the remained individual acceptor-donor interactions. When analysing the docking score, the hydrogen bond interactions are represented by the hydrogen bond between ligand and solvent and ligand and protein scores.

Shape interactions in Chemgauss are based on a united atom model (i.e. only heavy atoms are relevant to the shape calculation). Each ligand heavy atom is assigned a fixed clash penalty

score if the distance between it and a protein heavy atom is less than the sum of the VdW radii, otherwise it is assigned a score proportional to the count of the number of protein heavy atoms within 1.25 and 2.5 times the sum of the VdW radii (atoms within 2.5 count one tenth as much as those within 1.25). From this score, a penalty equal to two close protein atom contacts is subtracted to represent the VdW interactions with solvent water that are lost when the ligand docks. This score is pre-computed at grid points throughout the active site and the resulting grid is then smoothed. When analysing the docking score, the shape interactions are represented by the steric and clash scores.

Solvent contributions to changes in free binding energy arise from reorganisation of solvent-solvent and solvent-solute interactions upon binding. This mainly takes the form of desolvation of both the ligand and protein to allow their interaction. Polar ligands or side-chains on the protein can form direct hydrogen bond interactions with water. In contrast hydrophobic ligands or protein side-chains do not provide this hydrogen bonding opportunity. In computational chemistry, a solvent model allows for consideration of solvated condensed spaces for more accurate representation of thermodynamic calculations as well as other computational simulations of reactions and processes that take place in solution. The two main models in use are explicit and implicit models. OpenEye docking uses implicit models which are generally more computationally efficient but don't really account for fluctuations in solvent density around a molecule. Implicit models assume that solvent molecules can be replaced by a homogeneously polarizable medium as long as this medium, to a good approximation, gives equivalent properties [200, 201]. Broken ligand-solvent hydrogen bonds are accounted for by calculating desolvation positions around each hydrogen-bonding group on the ligand that represent the positions water could occupy when making a hydrogen bonding interaction with the protein. A penalty is then assessed that is proportional to the number of desolvation positions that can no longer be

occupied by water because the water in these positions would clash with the protein. This potential is placed on a grid and smoothed. Broken protein-solvent hydrogen bonds are accounted for by calculating how many hydrogen bonds water can make with the protein at the position of each heavy atom of the docked ligand, and a penalty score is assigned which is proportional to the number of hydrogen bonds. Hydrogen bonds with solvent that break when the ligand docks into the active site are penalized by the Chemgauss scoring function. The protein desolvation and ligand desolvation scores represent hydrogen bonding interactions with implicit solvent. The ligand desolvation score should stay the same in every docking experiment whilst the protein desolvation score will change dependent on protein region that is being investigated.

The common source of receptor structures for docking is X-ray crystallography and NMR but there is an increasing gap between the sequence of a known protein and the availability of its structure. As is the case currently, there is no available crystal structure of wtMOMP 11168. Although docking is increasingly being performed on receptor structures that have been obtained from homology modelling, I was unable to use the homology model of wtMOMP 11168 based on the crystal structure of MOMP 85H as the source of the receptor in FRED and so instead I used the original crystal structure of MOMP 85H, which has 94% sequence homology. The X-ray crystallography structure carries more geometric information than the homology model and at high sequence similarity of 94% there is limited loss of information. Fan *et al.* carried out a study comparing the quality of docking using such models[202]. Not surprisingly they found that comparative models significantly did better than random selections. Furthermore, the holo crystal structure templates gave the best enrichments, but the modelled structures were often competitive[202].

5.5.1 Blind docking of MOMP

One method employed during protein-ligand docking is blind docking. During a blind dock the whole surface of the protein, rather than specific areas of the protein, is selected for searching ligand binding sites. The result is an atomic resolution complex structure of putative binding pockets of the ligand. Blind docking is useful in the cases where the identity of the ligand is known. Usually, more than one area on the protein is identified as a docking site so the primary functional binding pocket can be selected for by comparing the binding free energy (ΔG) values derived from the interaction with the ligand at that site[203]. To investigate where the programme would predict the binding pockets initially, with no guidance from previous studies such as foot printing, the first blind box was performed around the whole surface of the protein (see methods). This shot in the dark method would give an idea of binding sites albeit at lower resolution. It is not known whether water molecules are actually involved in the binding of QPLEX to the wtMOMP binding site. In most cases of protein ligand binding, it has been shown that water molecules are not involved in binding mechanisms, so they are usually deleted before docking to make computations easier. So as not to confound the result of sites within the protein that can form favourable bonds with the ligand, and unnecessarily add penalties from ligand and protein desolvation, docking searches were performed without the water and 'native' ligands that were originally included in the crystal structure of 5LDT. The docked QPLEX resulted in several poses that satisfied the custom constraints. **Table 5.1** ranks these poses according to their Chemgauss4 score.

Table 1. Ranking of QPLEX poses after blind-box docking to whole surface of MOMP

Pose	*FRED Chemgauss4 score*	*Steric (van der Waals)*	*Clash*	*Protein Desolvation*	*Ligand Desolvation*	*HB interactions between ligand and solvent*	*HB interactions between ligand and protein*
1	*-5.58237*	*-11.634465*	*1.171104*	*1.404268*	*7.039471*	*-1.192132*	*-2.370615*
2	*-5.031992*	*-12.104366*	*1.144632*	*0.960783*	*6.81463*	*-0.376294*	*-1.471377*
3	-4.976299	-10.167148	1.377603	1.522394	5.931564	-1.315694	-2.325018
4	-4.976299	-10.167148	1.377603	1.522394	5.931564	-1.315694	-2.325018
5	-4.970019	-10.900041	1.321704	1.310175	6.419443	-0.941406	-2.179894
6	-4.970016	-10.900043	1.321704	1.310176	6.419445	-0.941406	-2.179893
7	-4.901844	-11.625813	1.108287	0.929846	6.718321	-0.403952	-1.628534
8	-4.790005	-11.816316	1.155532	0.976604	6.817947	-0.40604	-1.517731
9	-4.777973	-9.997993	1.341503	1.334801	5.950019	-1.25519	-2.151112
10	-4.768976	-10.059254	1.419015	1.59959	6.062191	-1.479733	-2.310786

Table 5.1: The poses are ranked in order of FRED Chemgauss4 score which is a function of the steric score, clash score, protein desolvation score, ligand desolvation score, the score from hydrogen bonding interactions between ligand and solvent and the score from hydrogen bonding interactions between ligand and protein. Ten poses were streamlined by FRED, of which the top two poses are highlighted and were studied in further detail.

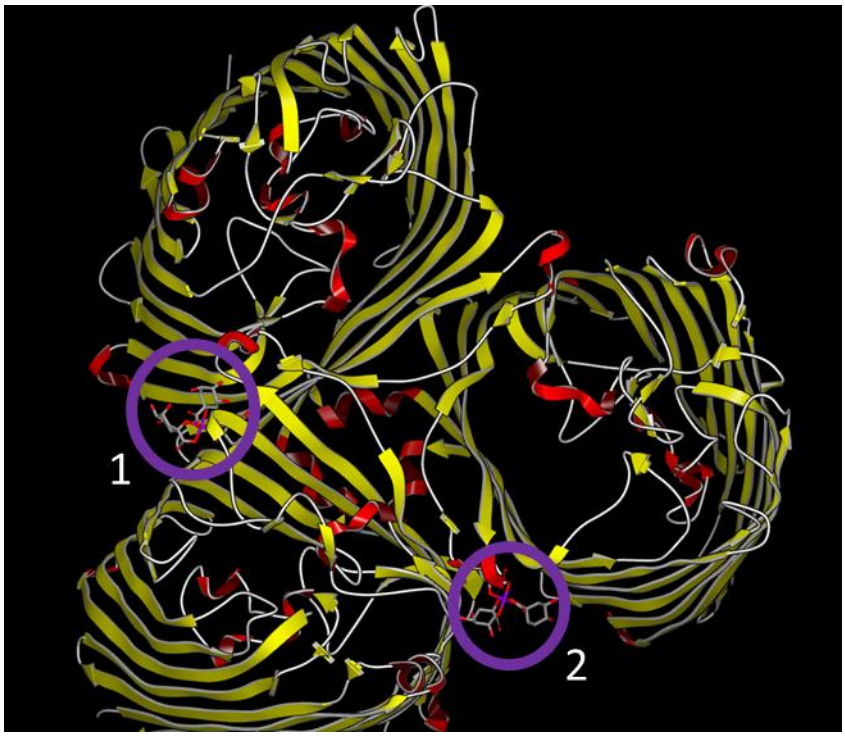


Figure 5.20: Docking of top two QPLEX poses to MOMP after initial round of blind box searching.

It was seen that pose 2-10 were in one translation whilst pose 1 was in another translation about the protein. Highly clustered results such as for poses 2 - 10 indicate that the conformational search procedure is exhaustive enough to ensure coverage of the accessible conformational space. The two poses are highlighted in **figure 5.20**. Re-docking was performed with in the interaction sites of pose 1 and pose 2 for higher resolution studies within these translations (**tables 5.2-5.3**).

High resolution docking of pose 1

Pose	*FRED Chemgauss4 score*	*Steric (van der Waals)*	*Clash*	*Protein Desolvation*	*Ligand Desolvation*	*HB interactions between ligand and solvent*	*HB interactions between ligand and protein*
1.1	*-5.796907*	*-11.70514*	*1.106775*	*1.0012*	*7.108759*	*-0.973851*	*-2.334651*

Table 5.2: Only the top-ranking pose (*pose 1.1*) out of ten poses is shown here as there is no significant difference between the top 10 poses. The poses are ranked in order of FRED Chemgauss4 score which is a function of the steric score, clash score, protein desolvation score, ligand desolvation score, the score from hydrogen bonding interactions between ligand and solvent and the score from hydrogen bonding interactions between ligand and protein.

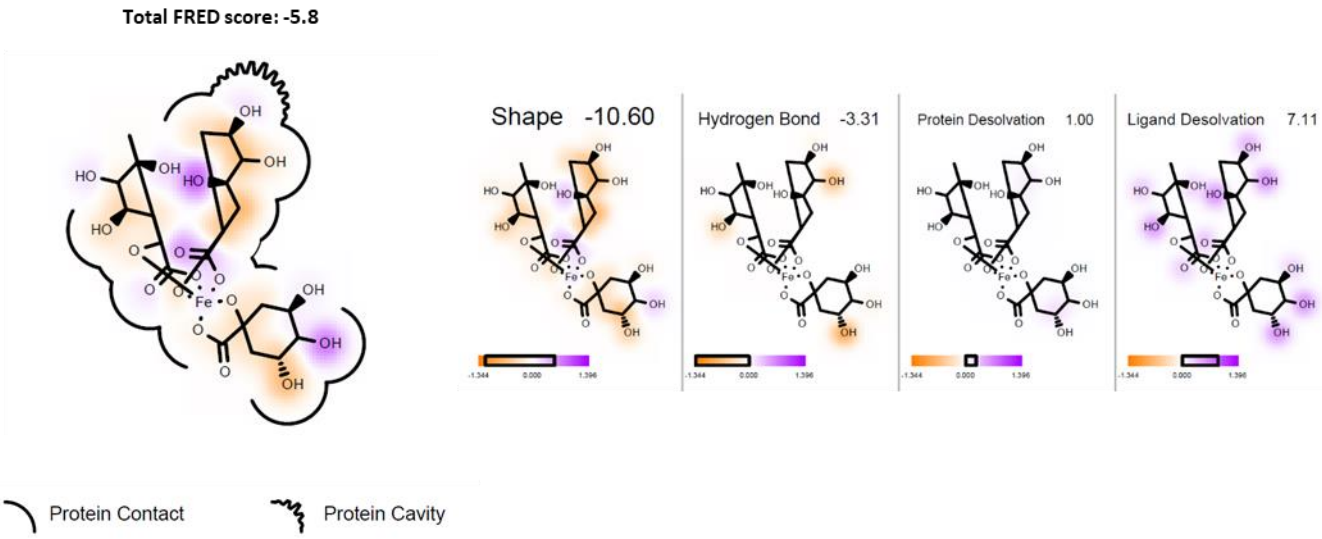


Figure 5.21: QPLEX-MOMP docking report of pose 1.1. On the left is the schematic representation of the ligand environment and the site characteristics. On the right is a breakdown of the components that contribute to the overall FRED score (shape score, hydrogen bond score, protein desolvation score and ligand desolvation score), ordered from most favourable score to least favourable score. The shape score is a function of the steric (van der Waals) and clash score. The hydrogen bond schematic shows the hydrogen bonds between the ligand and protein/ligand and solvent. Protein/ligand desolvation show the penalty from the interaction of water in that pose/translation. The degree of interaction is represented by an orange to purple scale bar where orange tends to a stronger interaction and purple tends to a weaker interaction.

High resolution docking of pose 2

Pose	*FRED Chemgauss4 score*	*Steric (van der Waals)*	*Clash*	*Protein Desolvation*	*Ligand Desolvation*	*HB interactions between ligand and solvent*	*HB interactions between ligand and protein*
2.1	*-5.0878*	*-12.2946*	*1.18775*	*1.028484*	*6.889263*	*-0.35754*	*-1.54113*

Table 5.3: Only the top-ranking pose (*pose 2.1*) out of ten poses is shown here as there is no significant difference between the top 10 poses. The poses are ranked in order of FRED Chemgauss4 score which is a function of the steric score, clash score, protein desolvation score, ligand desolvation score, the score from hydrogen bonding interactions between ligand and solvent and the score from hydrogen bonding interactions between ligand and protein.

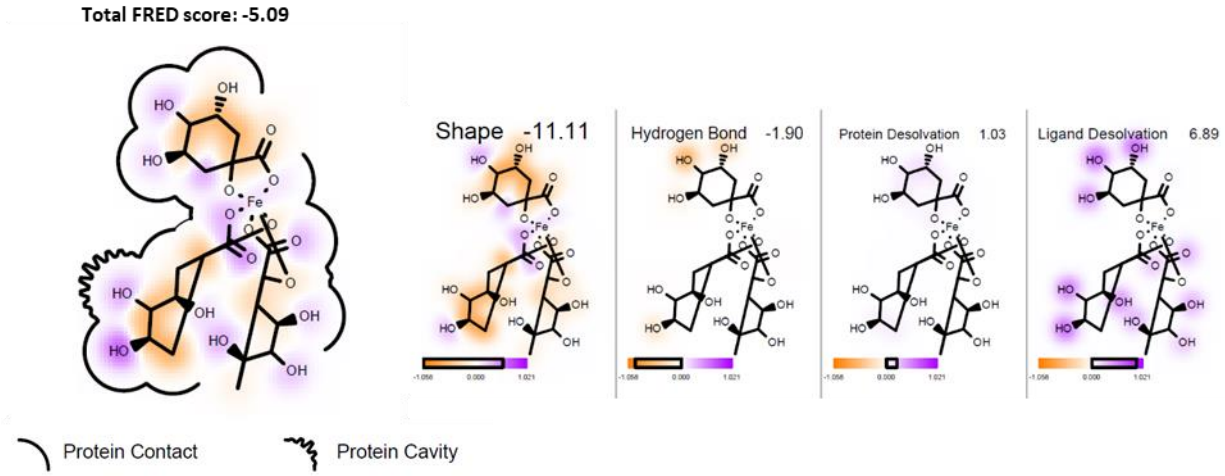


Figure 5.22: QPLEX-MOMP docking report of pose 2.1. On the left is the schematic representation of the ligand environment and the site characteristics. On the right is a breakdown of the components that contribute to the overall FRED score (shape score, hydrogen score, protein desolvation score and ligand desolvation score), ordered from most favourable score to least favourable score. The shape score is a function of the steric (van der Waals) and clash score. The hydrogen bond schematic highlights the hydrogen bonds between the ligand and protein or ligand and solvent. Protein/ligand desolvation show the penalty from the interaction of water in that pose/translation. The degree of interaction is represented by an orange to purple scale bar where orange tends to a stronger interaction and purple tends to a weaker interaction.

Both poses 1 and 2 show that there is a possible interaction in the interfacial region of MOMP from the QPLEX ligand. This is more likely to occur *in silico* and *in vitro* than *in vivo*. By the location of the site, the probability of any ligand binding in this area *in vivo* - the transmembrane region area external to the outer membrane protein - is very low. Interaction is more expected to occur *in vitro* where the protein has been purified and solubilised out of the cell membrane layer into solution and thus faces of the protein that would not usually be exposed are now accessible to ligands. Furthermore, the region is very hydrophobic, thus the positive and unfavourable protein

desolvation score – it is not favourable for waters or polar compounds to interact with the protein in this region. There is a low magnitude of hydrogen bonding in this region due to lack of formation of such bonds. The most favourable contributing force is from complementary shape forces. **Figures 5.21 and 5.22** show that most of the interaction that would be made between the ligand and protein surface in this area would be via shape and favourable van der Waals interactions.

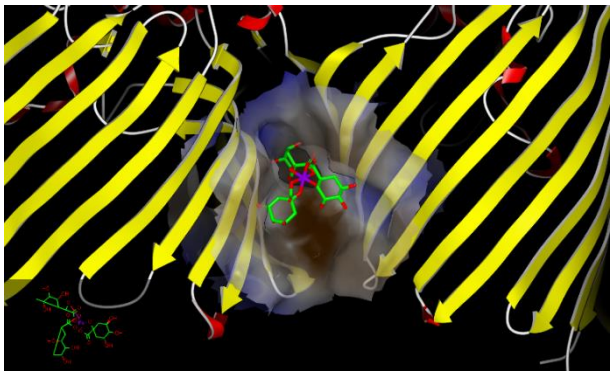


Figure 5.23: QPLEX docking pose (1.1) on MOMP. The ligand is buried in the interface site between monomer peptide chains A and B. The surface of the protein has been included to highlight the high complementary shape fit.

In this case, blind box docking did not find the location of the primary binding pocket but was a useful approach in the detection of subsidiary binding pockets. To search for binding pockets or sites that would be bound *in vivo*, the areas that were highlighted in footprinting studies as undergoing a significant change in labelling in the presence of QPLEX were docked *in silico*.

5.5.2 Ligand docking on the extracellular region of MOMP

The loop region on the extracellular side of the protein was docked to see whether we would find a greater FRED binding score than compared to the interactions found in previous rounds of blind box docking.

Docking score of extracellular site of MOMP

FRED Chemgauss4 score	*Steric (van der Waals)*	*Clash*	*Protein Desolvation*	*Ligand Desolvation*	*HB interactions between ligand and solvent*	*HB interactions between ligand and protein*
-8.37	*-10.15*	*0.80*	*8.28*	*7.01*	*-2.40*	*-11.91*

Table 5.4: Only the top ranked pose of ten poses is shown. The poses are ranked in order of FRED Chemgauss4 score which is a function of the steric score, clash score, protein desolvation score, ligand desolvation score, the score from hydrogen bonding interactions between ligand and solvent and the score from hydrogen bonding interactions between ligand and protein.

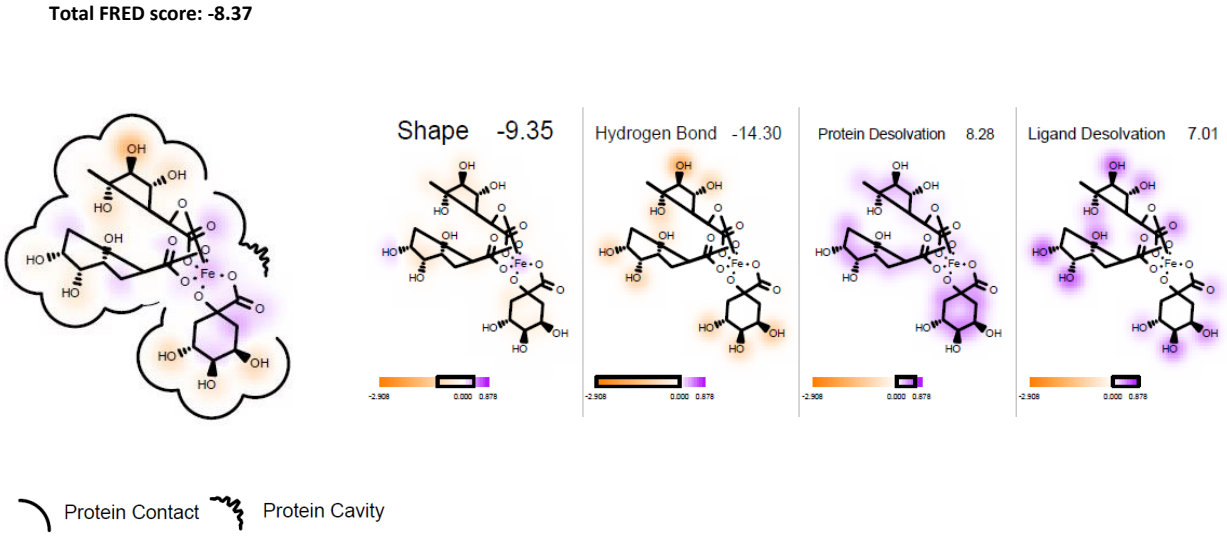


Figure 5.24: QPLEX-MOMP docking report of top ranked pose on extracellular site of MOMP. On the left is the schematic representation of the ligand environment and the site characteristics. On the right is a breakdown of the components that contribute to the overall FRED score (shape score, hydrogen score, protein desolvation score and ligand desolvation score). The shape score is a function of the steric (van der Waals) and clash score. The hydrogen bond schematic highlights the hydrogen bonds between the ligand and protein or ligand and solvent. Protein/ligand desolvation show the penalty from the interaction of water in that pose/translation. The degree of interaction is represented by an orange to purple scale bar where orange tends to a stronger interaction and purple tends to a weaker interaction.

Of the ten resultant poses, six of the poses were found in the same translation and conformation of the ligand in the top pose within this loop region (**figure 5.24**). The magnitude of the FRED score of this pose was 50% higher than for the poses found in the interface region in previous docking rounds. This indicates that the ligand is more likely to bind to the protein in this site and goes some way in proving that the previous results should be limited to scenarios occurring *in silico* or *in vitro*. This increase in the magnitude of the FRED score can be attributed to the increase in

hydrogen bond interactions formed between the ligand and protein in this site. In the loop, the electrostatic force potential (via hydrogen bonds) contributed to a score of -11.91, compared to a score of just -1.5 in the interface. The shape complimentary score is also favourable and steric or van der Waals forces contribute a score of -9.35 to the overall FRED score but this is counter balanced by the penalty from protein desolvation where water will ultimately have to be forced out of such a favourable binding pocket. There is an entropic penalty that must be paid in more solvent accessible regions where immobilization of water or solvent may be involved.

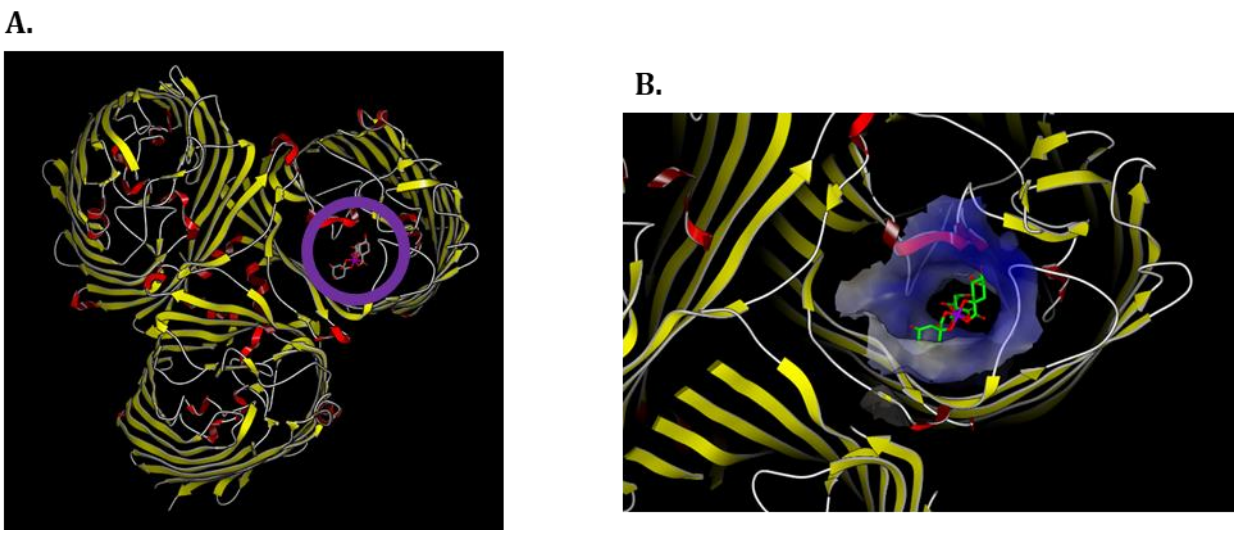


Figure 5.25: A. Docking result of highest ranked pose of QPLEX on extracellular surface of MOMP. **B.** The secondary structure of proteins in this site primarily inhabit a random loop structure so the surface of the protein has been included for easier definition of the site.

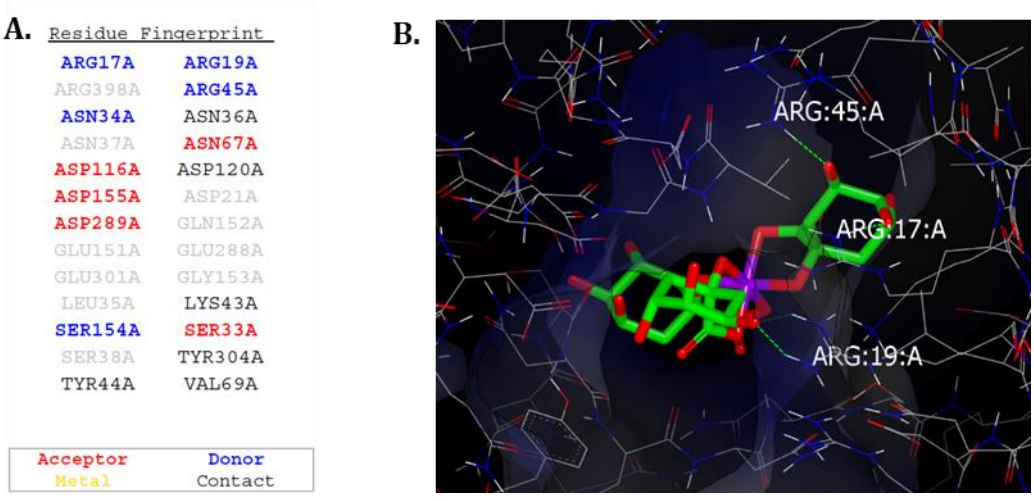


Figure 5.26: A. Residue fingerprint which highlights residues in the MOMP extracellular site that QPLEX is interacting with. Red denotes amino acid residues that accept hydrogen bonds from the ligand, blue denotes hydrogen bonds that are donated by the protein, black denotes residues on the protein that the ligand contacts and denotes other amino acids found in the region. **B.** Amino acids within the protein that donate hydrogen bonds to the ligand have been highlighted.

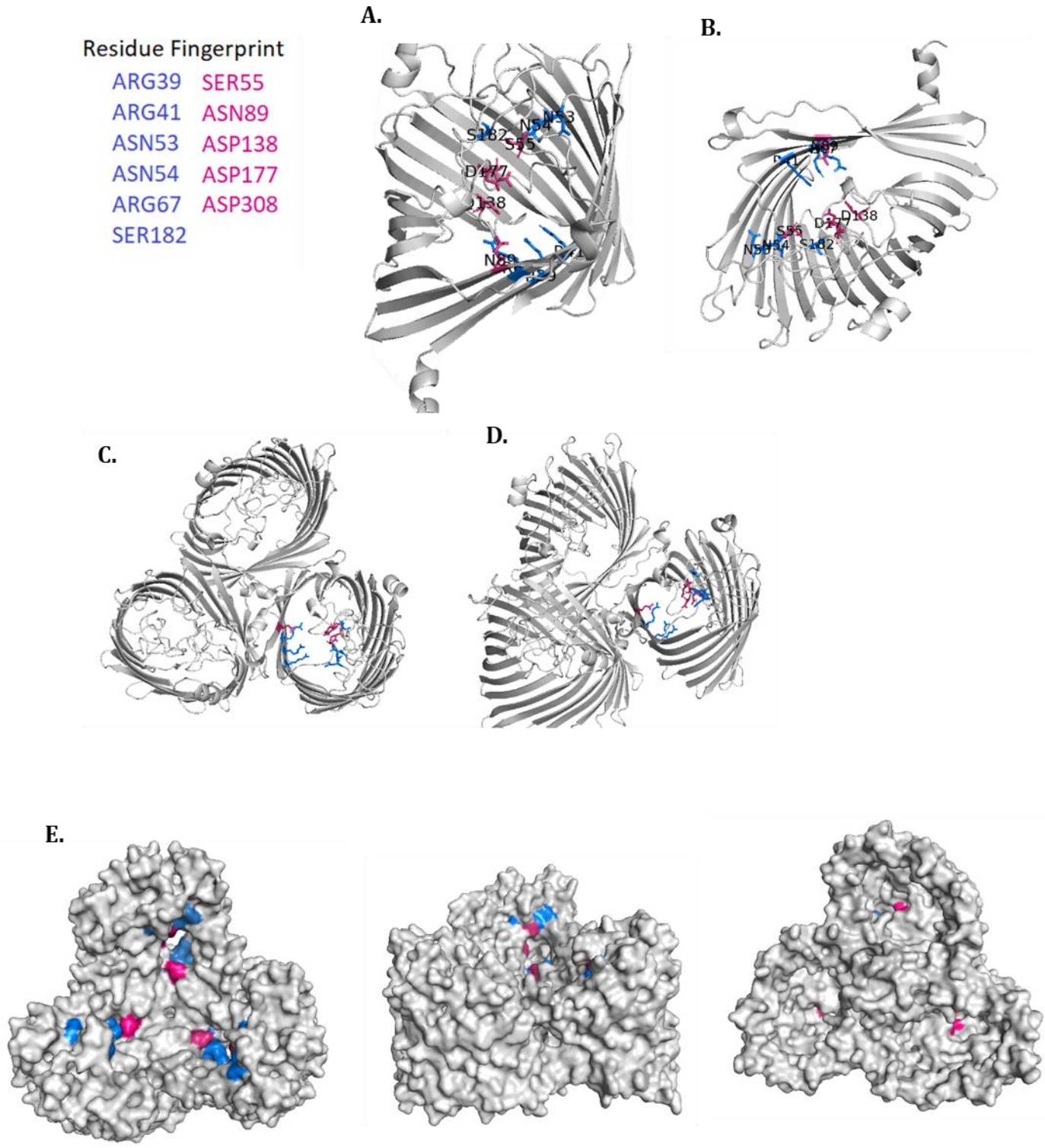


Figure 5.27: The resultant residue fingerprint of the amino acids in the extracellular loop of MOMP 85H (5LDT) that interact with QPLEX has been mapped onto the analogous residues in the homology model of our wtMOMP variant. In blue are amino acid residues that are hydrogen bond donators whilst in pink are amino acid residues that would act as hydrogen bond acceptors from QPLEX. **A.** bird's eye view of wtMOMP monomer in cartoon depiction, **B.** perpendicular rotation of A. to view hidden residues, **C.** wtMOMP trimer in cartoon depiction, **D.** rotation of C., **E.** (from left to right: bird's eye view, interfacial region and underside of) wtMOMP trimer in surface depiction. The structures are depicted in grey.

Figure 5.27 shows that QPLEX interacts with amino acid residues found in the extracellular loop region of MOMP. The residues that were calculated to donate hydrogen bonds to QPLEX in this region are Arg39, Arg41, Asn53, Asn54, Arg67 and Ser182. Amino acid residues that would accept hydrogen bonds from QPLEX are Ser55, Asn89, Asp138, Asp199 and Asp308. The hydrogen bond interactions between the ligand and these amino acid residues are the largest contributors to the favourable binding in the extracellular loop region, contributing a score of minus 11.91 **(Table 5.4)**.

5.5.3 Docking QPLEX to the inside of the wtMOMP porin

Results from carbene footprinting show that there was some conformational change of the protein within the site of the porin. Blind docking already highlighted the outside of the porin for docking and showed that although there was high shape specificity, the lack of hydrogen bond interactions in this region means that there is unlikely interaction of the ligand to the outside of the porin. More docking was performed to highlight the docking of the compound to the inner side of the porin, where a box as created along the length of the porin excluding the inter- and extracellular regions (see methods).

Ligand-protein docking score of QPLEX and inner porin of MOMP

Pose	*FRED Chemgauss4 score*	*Steric (van der Waals)*	*Clash*	*Protein Desolvation*	*Ligand Desolvation*	*HB interactions between ligand and solvent*	*HB interactions between ligand and protein*
1	*-7.98*	*-9.74*	*0.58*	*5.30*	*6.45*	*-1.86*	*-8.71*
2	*-7.91*	*-8.25*	*0.30*	*3.82*	*5.95*	*-1.91*	*-7.82*

Table 5.5: QPLEX was docked to the inner porin of MOMP in triplicate experiments. Only the top two ranking poses out of ten poses are shown here. The poses are ranked in order of FRED Chemgauss4 score which is a function of the steric score, clash score, protein desolvation score, ligand desolvation

score, the score from hydrogen bonding interactions between ligand and solvent and the score from hydrogen bonding interactions between ligand and protein.

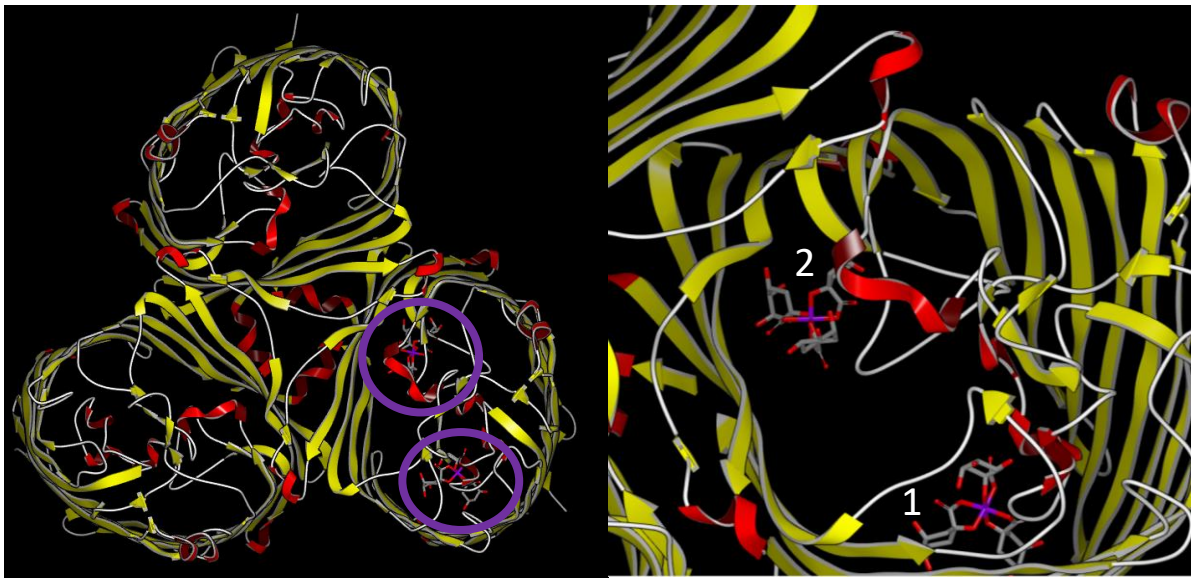


Figure 5.28: Visualization of protein-ligand docking result of the two highest ranked poses of QPLEX within the constriction zone of the MOMP porin. Pose 1 is closer to the surface of the protein whilst pose 2 is buried deeper within the porin towards the periplasm.

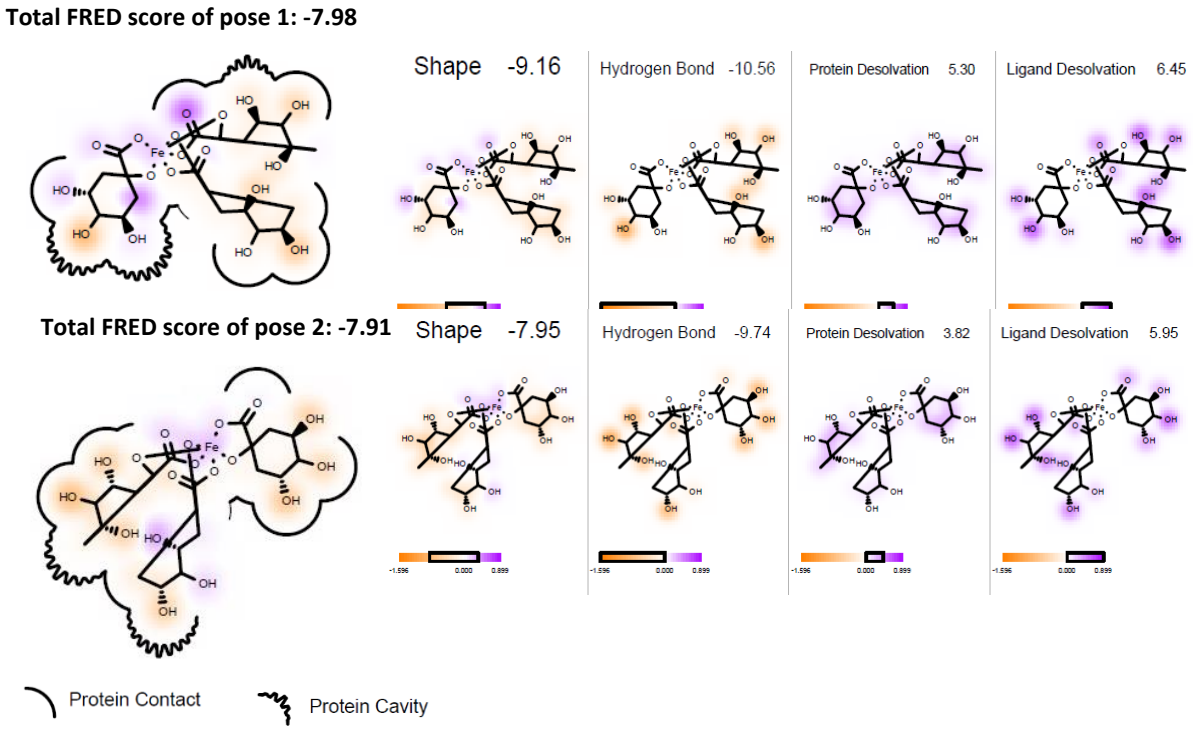


Figure 5.29: QPLEX-MOMP docking report of the best two ranked poses within the MOMP porin constriction zone. On the left is the schematic representation of the ligand environment and the site characteristics. On the right is a breakdown of the components that contribute to the overall FRED score (shape score, hydrogen bond score, protein desolvation score and ligand desolvation score). The

shape score is a function of the steric (van der Waals) and clash score. The hydrogen bond schematic highlights the hydrogen bonds between the ligand and protein or ligand and solvent. Protein/ligand desolvation show the penalty from the interaction of water in that pose/translation. The degree of interaction is represented by an orange to purple scale bar where orange tends to a stronger interaction and purple tends to a weaker interaction.

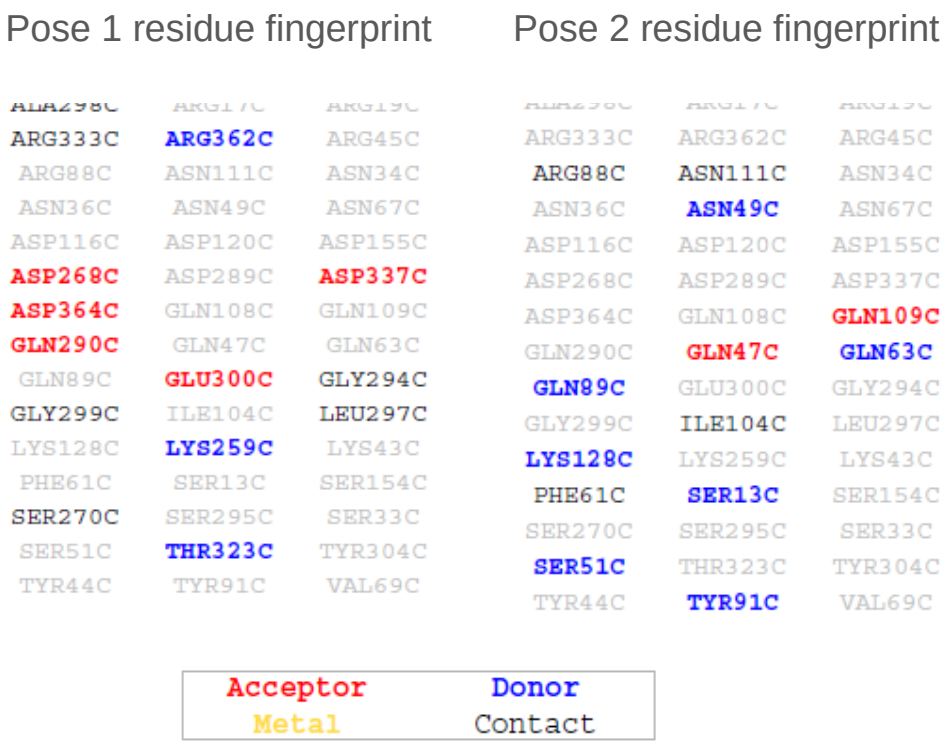


Figure 5.30: Residue fingerprint along the transverse of the constriction zone. Residues in the MOMP 85H porin site that QPLEX interacts with are highlighted. Red denotes amino acid residues that accept hydrogen bonds from the ligand, blue denotes hydrogen bonds that are donated by the protein, black denotes residues on the protein that the ligand contacts and grey denotes other amino acids found in the region. Amino acid interactions with pose 1 of the ligand within the top half of the constriction zone and amino acid interactions with pose 2 of the ligand with the bottom half of the constriction zone are shown.

Residue fingerprint pose 1

ASP236	LYS278
GLU252	THR342
ASP287	ARG381
GLY310	
GLU319	
ASP356	

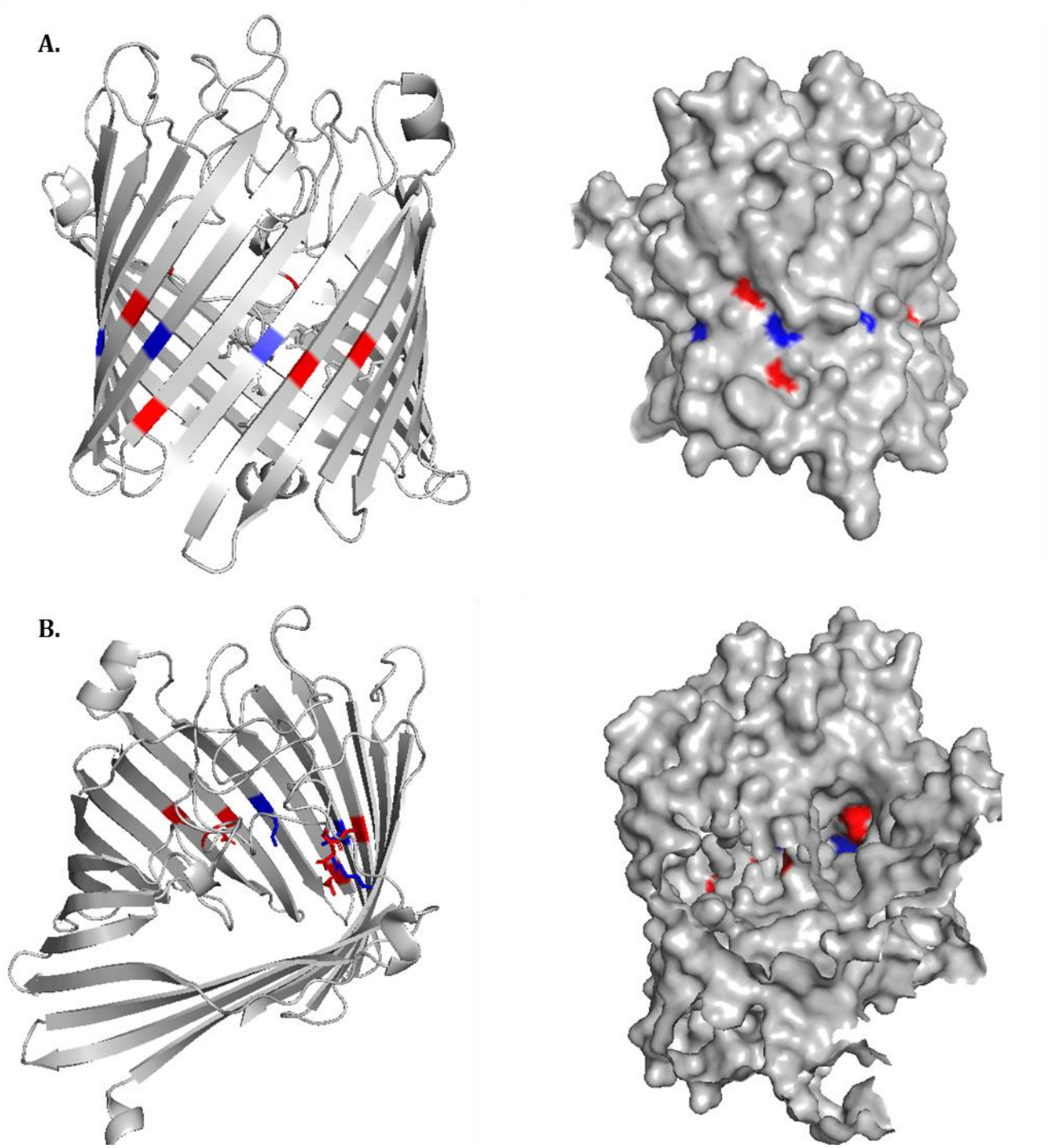


Figure 5.31: The resultant residue fingerprint of the amino acids involved in the interaction extending from pose 1 of QPLEX to the beginning of the constriction zone of MOMP 85H (5LDT) has been mapped on to the analogous residues in the homology model of wtMOMP 11168. This area is surrounded by more negatively charged amino acids. Red denotes amino acid residues that accept hydrogen bonds from the ligand, blue denotes hydrogen bonds that are donated by the protein. Cartoon (left) and corresponding surface visualisations (right) of wtMOMP monomer porin are shown from the interface of the protein monomer **(A.),** and the bird's eye view looking down into porin from the extracellular side **(B.)**.

Residue fingerprint pose 2

GLN69	SER35
GLN131	ASN71
	SER73
	GLN85
	GLN111
	TYR113
	LYS150

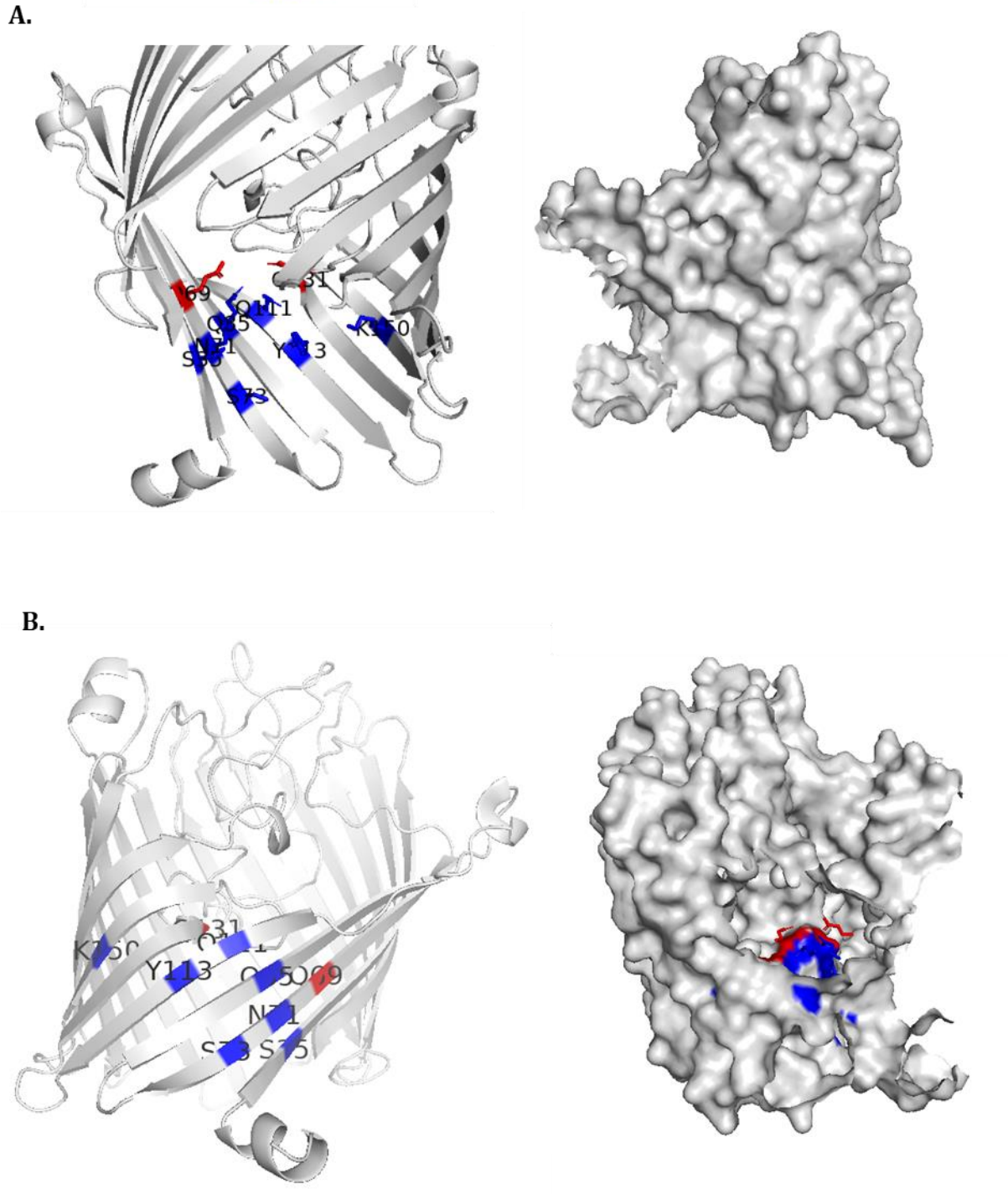


Figure 5.32: The resultant residue fingerprint of the amino acids involved in the interaction extending from pose 2 of QPLEX to the beginning of the constriction zone of MOMP 85H (5LDT) has been mapped on to the analogous residues in the homology model of wtMOMP 11168. This area is

surrounded by more positively charged amino acids. Red denotes amino acid residues that accept hydrogen bonds from the ligand, blue denotes hydrogen bonds that are donated by the protein. Cartoon (left) and corresponding surface visualisations (right) of wtMOMP monomer porin are shown from the interface of the protein monomer **(A.),** and the bird's eye view looking down into porin from the extracellular side **(B.)**.

The overall docking score (FRED Chemgauss4 score) shows that the binding between QPLEX and the porin is favourable in this region. The score is similar to that of docking the ligand to the extracellular region, owing to the continuity of transient and weak hydrogen bonds that are formed as the compound travels down through the porin. Both the shape score and the hydrogen bond score between the protein and ligand contribute significantly to the magnitude of the negative overall score. It would be safe to conclude that the compound moves into the porin after binding to the extracellular loops. **Figure 5.28** shows that in the docked configuration, the QPLEX molecule moves along the pore axis and interacts both with the acidic and basic clusters of the constriction zone. There are two main poses that are located along different points of the y-axis of the porin. Pose 1 is located more towards the extracellular surface of the porin whilst pose 2 is buried deeper within the porin. **Figure 5.30** displays more detail about the amino acids in the site of these regions and shows that there is a clear split of type of hydrogen bond that is formed; at the top half of the porin the compound is mainly donating hydrogen bonds to the surrounding amino acids then as it travels towards the bottom of the porin it begins to accept hydrogen bonds from its surrounding amino acids. This is due to the distribution of charge in the porin and its effects on the binding pattern whereby pose 1 makes contact in the top half of the porin where it is surrounded mainly by negatively charged amino acids (**figure 5.31**) whilst pose 2 makes contact in the bottom half of the porin where it interacts mainly with positively charged amino acids (**figure 5.32**). This arrangement gives rise to a dipole across the short axis of the elliptical constriction zone, termed the transverse electric field. In previous studies, this transverse field has been shown to be important for the translocation of polar molecules from the environment into

the gram-negative bacterial cell. It has also been shown to be an important characteristic involved in drug design where zwitterionic compounds such as penicillin and ciprofloxacin are able to use the polarity of the porin to pass through the constriction zone[204]. To increase selectivity of small molecules that pass through the porin, in MOMP it was shown that there is a Ca^{2+} ion in the constriction zone that can work to reverse the transverse field[120]. It is not known whether such an ion exists in our variant of wtMOMP so Ca^{2+} was not included in docking studies, but if there was the presence of an ion, it would not have a large effect on the entrance of the QPLEX compound into the constriction zone due the non-ionic nature of QPLEX, making it a powerful tool for entering the porin. When we align and blast the amino acid fingerprint onto our homology model of wtMOMP, the amino acids of interaction in the top half of the porin are Asp236, Glu252, Asp287, Gly310, Glu319, Asp356, Lys278, Thr342, Arg381. Once the ligand moves further down into the constriction zone the residue fingerprint that the ligand interacts with are Gln69, Gln131, Ser35, Asn71, Ser73, Gln85, Gln111, Tyr113, Lys150. This complements more recent studies that are moving towards a suggestion that drug design should be centred around the molecule's electrical dipole moment strength as opposed to a traditional conclusion that zwitterionic molecules be favoured [204]. Zwitterionic molecules can also cause sterically hindered rotations, something that was not seen in the protein-ligand docking result where the clash score was only slightly unfavourable in both poses (**Table 5.5** and **figure 5.29**).

5.5.4 Docking QPLEX to intercellular MOMP

The results of carbene footprinting in the absence and presence of QPLEX show that in the presence of QPLEX, there was an increase in carbene labelling at the underside of the protein that faces the periplasmic space. *In silico* protein ligand docking was used to study whether or not the ligand bound to part of this region of the protein. As can be seen from **figure 5.29**, the intracellular

side of the protein is very hydrophilic. Further, **figure 5.33** shows that there is a highly negatively charged region here.

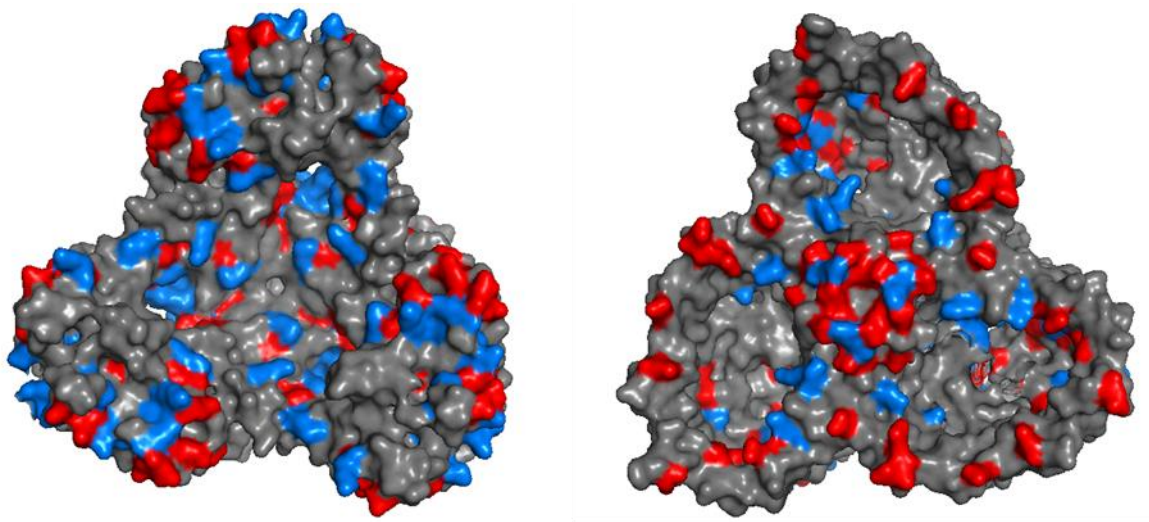


Figure 5.33: Charge map of wtMOMP on a surface diagram from Pymol. Birds eye view of extracellular side of the protein (left) and view of the underside (right) of the protein. Red regions represent negatively charged amino acids (glutamate and aspartate) whilst blue amino acids represent positively charged amino acids (arginine, lysine and histidine).

Ligand-protein docking score of QPLEX and intracellular portion of MOMP

FRED Chemgauss4 score	*Steric (van der Waals)*	*Clash*	*Protein Desolvation*	*Ligand Desolvation*	*HB interactions between ligand and solvent*	*HB interactions between ligand and protein*
1.50	0.61	0.98	3.72	4.42	-1.17	-7.08

Table 5.6: QPLEX was docked to the intracellular portion of MOMP in triplicate experiments. Only the top ranking pose out of ten poses is shown here. The poses were ranked in order of FRED Chemgauss4 score which is a function of the steric score, clash score, protein desolvation score, ligand desolvation score, the score from hydrogen bonding interactions between ligand and solvent and the score from hydrogen bonding interactions between ligand and protein.

These characteristics are usually favourable for binding of polar molecules via dipole dipole interactions but the results seen in **table 5.6** show that binding of QPLEX to this region is unfavourable with a FRED Chemgauss score of 1.5. A large part of the score was due to the lack of steric interactions in this site which meant there was less instance of hydrogen bonding to be able to overcome the protein and ligand desolvation score due to a highly hydrophilic environment in the periplasm. Thus, the result points towards the translocation of the QPLEX through the porin

changing the shape of the intracellular region rather than binding to it as it does in the extracellular part of the porin.

5.5.5 Docking Le^b to MOMP

In previous chapters it was shown that several BgAgs bind to wtMOMP more favourably than the non-glycosylated mutant variants that we created. Now that interaction was established, we wanted to locate where such interaction was taking place on the protein. Our hypothesis was that because the ELISA binding studies showed low binding to the non-glycosylated mutants, we would see limited binding in protein-ligand docking studies to Le^b too. The MOMP protein model we used for ligand docking is non-glycosylated and thus we did not expect favourable binding to take place in this round of protein-ligand experiments. Results confirmed this hypothesis.

Ligand-protein docking score of Le^b to various regions of the MOMP protein

Region	*FRED Chemgauss4 score*	*Steric (van der Waals)*	*Clash*	*Protein Desolvation*	*Ligand Desolvation*	*HB interactions between ligand and solvent*	*HB interactions between ligand and protein*
Monomer interface	4.35	-8.27	2.56	2.74	15.41	-4.04	-4.04
Extracellular	-1.34	-10.12	2.01	11.13	12.65	-3.53	-13.48
Porin	-0.81	-10.40	2.44	11.86	13.17	-3.97	-13.90
Underside	15.99	8.21	2.99	5.26	10.62	-4.04	-7.05

Table 5.7: Le^b was docked to various areas of MOMP in triplicate experiments. Only the top ranking pose out of ten poses is shown for every region. The poses were ranked in order of FRED Chemgauss4 score which is a function of the steric score, clash score, protein desolvation score, ligand desolvation score, the score from hydrogen bonding interactions between ligand and solvent and the score from hydrogen bonding interactions between ligand and protein.

The scores show that the two most unfavourable regions for binding of the sugar to the protein are the protein monomer interface and the underside of the protein. We are using the soluble BgAgs to represent similar structures that are found on the surfaces of erythrocytes but are also highly expressed in the oro-gastro-intestinal (OGI) epithelium, endothelial cells, and in secretions, such as milk, saliva and tears. Therefore, we are not expecting the soluble version of the

carbohydrate sugar to pass through the porin to the underside of the protein, less so to make interaction within the outer membrane where the protein forms as a trimer (interfacial region). Rather there is expectation that the sugars might bind only on the extracellular surface or loop region of the protein as studies have showed that fucosylated sugar components in human breast milk could inhibit *C. jejuni* binding to HEp-2 cells *in vitro* and to human intestinal mucosa *ex vivo*, and also prevent *C. jejuni* infection of mice *in vivo* [109]. The docking results show near neutral but more favourable binding of the sugar in the extracellular side of the protein followed by the porin where there are more hydrogen bond interactions between ligand and protein.

Although the docking scores show favourable binding between the BgAgs and the extracellular/porin region, the magnitude is not as high as was for the protein-ligand docking of QPLEX. Firstly, we are not able to replicate the effect of the *O*-glycosylation on the wtMOMP *in silico* and thus are effectively using a non-glycosylated version of the protein in the protein-ligand docking studies for which we expect lower binding to the carbohydrate sugar than for the glycosylated wild type. Furthermore, there are some particular difficulties that are encountered during docking experiments of protein-carbohydrate interactions over protein-ligand and protein-peptide interactions. Firstly, many carbohydrates, as well as Le^b are very flexible and carry a large amount of hydroxyl groups, which can lead to an extensive hydrogen bonding network, not just with the protein, but with water and solvents too which requires greater energy to overcome in order to make favourable interactions with the protein in that site. Secondly, in addition to hydrogen-bonding, a distinctive but under researched feature of carbohydrate-protein interactions is the formation of CH-π contacts between aromatic side chains of the protein and the C–H bonds of the carbohydrates. These interactions are accounted for differently in docking

programmes to how they would actually interact in the case of protein and carbohydrates, due to the limited amount of experimental data [205, 206].

5.6 Conclusion

Fucose moieties on cell-surface glycans are increasingly recognized as critical to many cell-cell interactions and signalling processes, and are involved in a wide variety of physiological and pathological processes[179]. The results show that wtMOMP binds better to Le^b (doubly fucosylated) and Le^a (singly fucosylated) BgAgs than to the non-fucosylated BgAgs such as type 1 and type 2 BgAgs. The results of the binding studies shows that Le^b bound 3 times better to wtMOMP than $MOMP^{T/G268}$ and $MOMP^{\Delta pseD}$, pointing towards a higher affinity for glycosylated protein over non-glycosylated protein. Dose-response studies between Le^b ligand and wtMOMP target protein resulted in an EC50 of 23 nM. To see whether the binding between the glycosylated protein and the BgAg could be disrupted, inhibition studies were performed with QPLEX. The results show that between concentration of 1 nM and 1 µM, the QPELX brought down the maximum binding absorbance to Le^b by at least 50% in the case of wtMOMP. For the non-glycosylated protein variants – $MOMP^{T/G268}$ and $MOMP^{\Delta pseD}$, there was no significant difference in the binding of the protein to Le^b after addition of QPLEX, again pointing towards preference of the compound for interaction with the glycosylated version of the protein than the non-glycosylated protein. Experiments were carried out to determine the site of interaction of the ligands and the BgAgs on the MOMP protein. Although the glycosylated version of the protein could not be input into the protein-ligand docking software, the combination of experiments performed together with the glycosylated protein in carbene footprinting studies goes some way to localising the area for interaction. Carbene footprinting allows us to determine the region in which the glycosylated MOMP is masked by QPLEX and Le^b and showed that several areas, instead of reducing the

carbene masking in that area, increased carbene masking in the presence of ligands, showing a conformational change after addition of ligand. Protein ligand studies were used to further explore possible sites of interaction and to give more detail on the bonds that are formed between the protein and ligands in these sites.

Chapter 6: **Binding of MOMP to intestinal tissue (*ex vivo*)**

6.1.1 Introduction

Previous studies have confirmed that *C. jejuni* can bind to BgAgs, which is likely to contribute to the ability of this intestinal bacterium to interact with epithelial cells lining the gastrointestinal tract [110]. We wanted to explore this relationship further and probe whether interaction of *C. jejuni* on the gastrointestinal epithelium could be replicated by the soluble MOMP protein of *C. jejuni*. Similar studies with *H. pylori* have shown that its blood group antigen binding adhesin (BabA) is capable of binding BgAg carbohydrates expressed in the mucosa[207]. Previous ELISA binding studies have already established that wtMOMP binding to BgAgs can take place *in vitro* using soluble BgAgs, but we also wanted to confirm that such an interaction is conserved *ex vivo* whereby sections of tissue from GI tract are visualised in the absence and presence of wtMOMP. The sections we were provided with were parts of the stomach – the corpus and antrum, the duodenum of the small intestine and the colon of the large intestine (**figures 6.1-6.2**).

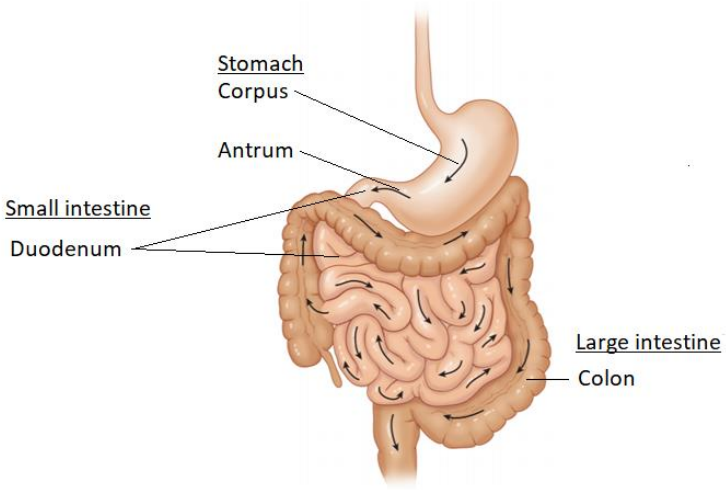


Figure 6.1: General organisation of the digestive system. Swallowed food enters the gastrointestinal tract (GI tract) consisting of oesophagus, stomach, small intestine, and large intestine. Just below the diaphragm, the oesophagus ends at the stomach of which the corpus (or central body) is the largest of the stomach's three sections which continues to the lower antrum. The opening between the stomach and the small intestine is guarded by the pyloric valve. This thickened band of smooth muscle relaxes to allow only small amounts of chyme into the small intestine at any one time. To make sure that the

intestine is not overwhelmed with more than it can digest or absorb, integrated signals and feedback loops between the intestine and stomach regulate the rate at which chyme enters the duodenum. Nearly all digestion takes place in the small intestine where most of the digested nutrients and secreted fluids are absorbed, leaving about 1.5 litres of chyme per day to pass into the large intestine. In the large intestine ions, minerals, vitamins, water and small organic molecules made by gut bacteria are absorbed out of the watery chyme resulting in semi-solid faeces[208].

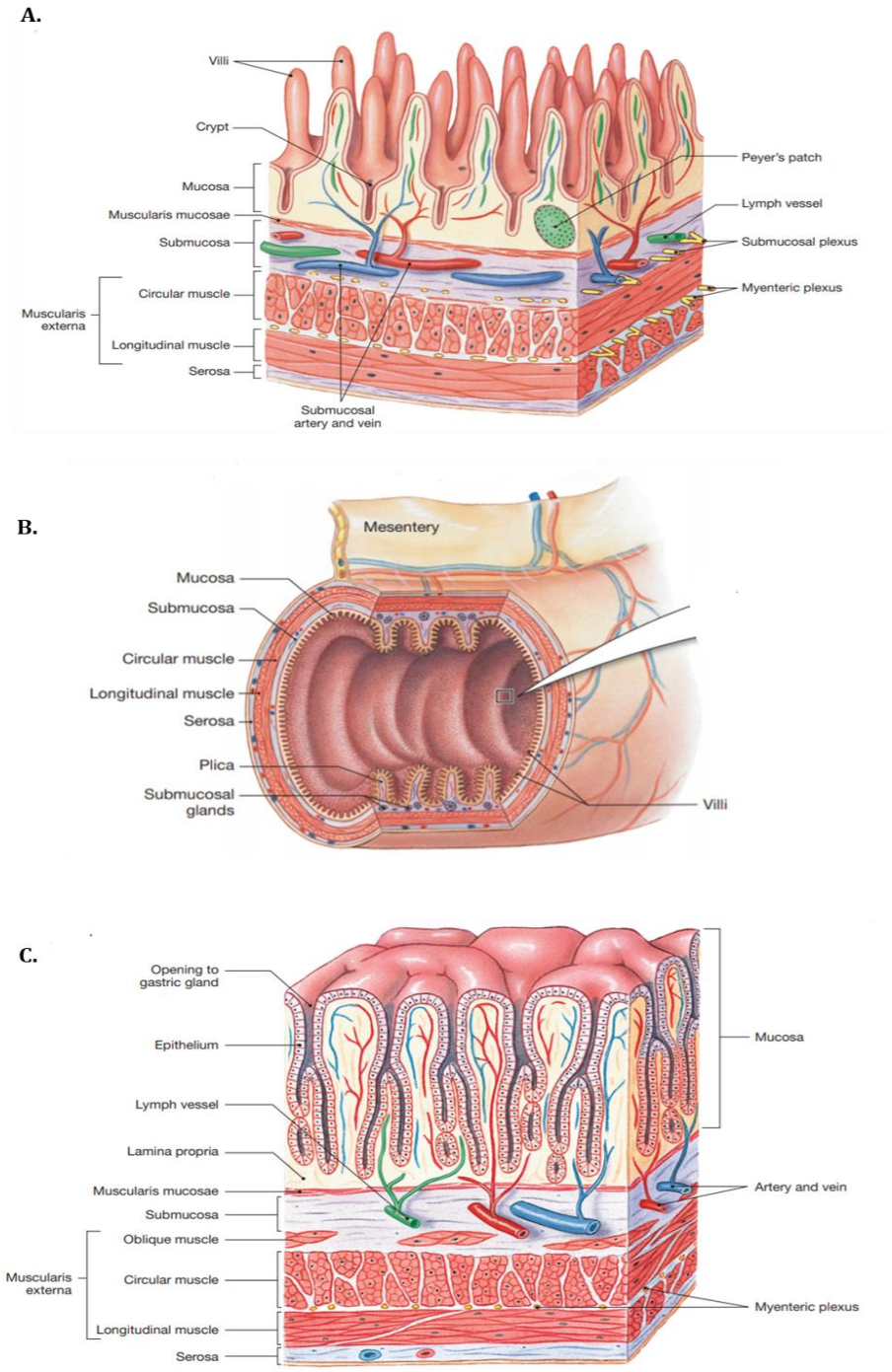


Figure 6.2: Basic structure of the gastrointestinal wall **A.** structural view of the small intestine **B.** sectional view of the small intestine **C.** sectional view of the stomach. The basic structure of the gastrointestinal wall is similar in the stomach and intestines, although there are some small differences from one section of the GI tract to another. The gut wall is crumpled into folds to increase its surface area. These folds are called rugae in the stomach and plicae in the small intestine. The

intestinal mucosa also projects into the lumen in small finger like extensions known as villi. Additional surface area is added by tubular invaginations of the surface that extend down into the supporting connective tissue called gastric glands in the stomach and crypts in the intestine. The deepest invaginations form secretory submucosal glands that open into the lumen through ducts. The gut wall consists of four layers which is consistent across the sections: (1) an inner mucosa facing the lumen, (2) a layer known as the submucosa, (3) layers of smooth muscle known collectively as the muscularis externa, and (4) a covering of connective tissue called the serosa.[208]

The gut tissue sections were collected during biopsy of patients that were being examined for gut cancer. None of the patients suffered from *C. jejuni* infection as determined by standard histochemistry and microbiology.

6.1.2 Tagging of wtMOMP with fluorescein

MOMP binding to the tissue was to be determined by immunofluorescence/direct fluorescence staining and visualisation with confocal microscopy. In direct detection of wtMOMP staining on the tissue, wtMOMP was tagged with fluorescein. Fluorescein is a dark orange/red powder that is slightly soluble in water and alcohol and a commonly used tracer. Fluorescein has an absorption maximum at 494 nm and emission maximum of 512 nm which emits in the green spectrum under UV light. The *N*-Hydroxysuccinimide (NHS)-ester labelled fluorescin (NHS-Fluorescein) was used to tag and label wtMOMP protein. NHS esters react efficiently with primary amino groups (-NH2) in pH 7-9 buffers to form stable amide bonds. The reaction results in the release of *N*-hydroxysuccinimide and a fluorescein tagged protein. Proteins, including wtMOMP, generally have several primary amines in the side chain of lysine residues that are available as targets for NHS-ester reagents.

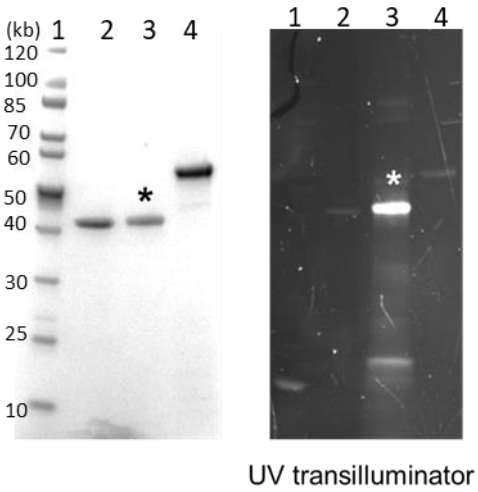


Figure 6.3: Tagging of wtMOMP with NHS-Fluorescein. 5 µM protein was tagged with 15 mmol NHS Fluorescin. Lanes: 1 - SDS page unstained ladder, 2 - wtMOMP (negative control), 3 - Fluorescein labelled wtMOMP, 4 - GFP (positive control). The left gel is an image of Coomassie blue stains captured with white light. The right image is an image of the UV bands that can be seen with the UV transilluminator. The asterisk highlights the visualisation of wtMOMP^{Flu} band under both lights.

Successful labelling of wtMOMP with fluorescein was determined by running the dialysed sample on SDS PAGE and visualising the gel under UV light (**figure 6.3**). There is a high amount of fluorescence from the fluorescein tagged wtMOMP (lane 3) that is captured by the gel imager. MOMP-fluor was used to stain the sections of the gastrointestinal tract to see if any binding would occur *ex vivo*, possibly via the lewis BgAgs that exist naturally on the gut epithelial cells.

After SDS PAGE, protein was transferred from the gel to a nitrocellulose membrane for a western blot (**figure 3.11**). Rabbit anti-MOMP antibody is the primary antibody used to identify the protein. BSA was loaded as a negative control for primary antibody binding. Anti MOMP antibody recognizes MOMP protein even after conjugation to fluorescein and far-red tags.

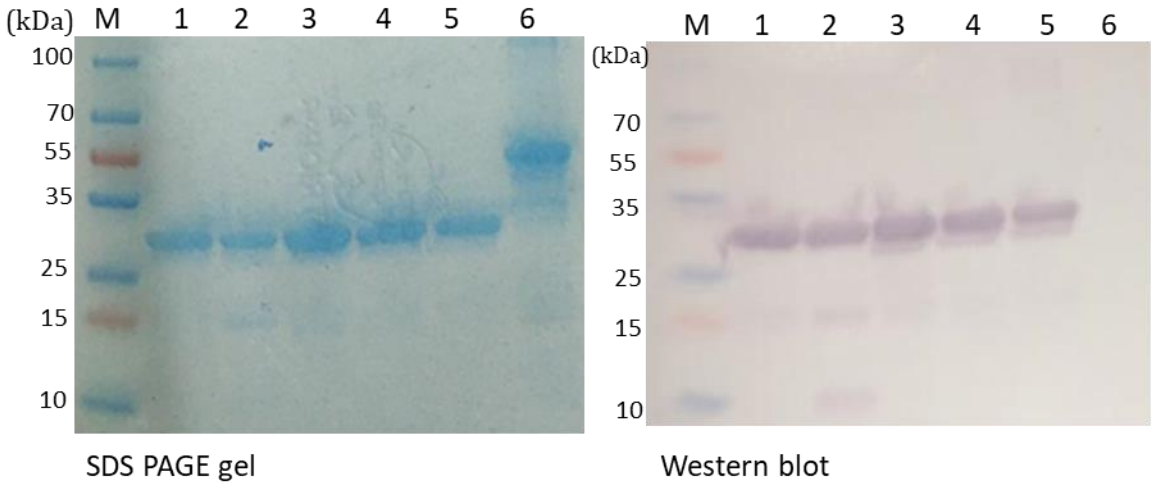


Figure 3.11: Western blot imaged after incubation of SDS PAGE gel with alkaline phosphatase substrate. Marker- Page ruler plus prestained protein ladder, lane 1- wtMOMP, lane 2 – MOMP$^{T/G268}$, lane 3- MOMP$^{\Delta pseD}$, lane 4- MOMP^{flu}, lane 5- MOMP^{farred} lane 6 – Control BSA.

6.1.3 Unstained gut sections

Unstained tissue sections of the gut were used as negative controls for the IF experiments. Firstly, we had to determine the amount of background fluorescence already emitting from the control tissues. The micrographs show that although no staining had taken place, there was still fluorescent material emitting from the tissue at 494 nm (**figure 6.4**) and thus we could use these images to observe the structure of the tissues that existed on our sections.

Two sections shown here make up part of the stomach; the corpus (upper stomach) and antrum (lower stomach). In the corpus we mainly see the thick mucosa lined with simple columnar epithelium. The surface epithelium invaginates into gastric pits into which the fundus glands open. In the antrum, the micrograph shows that it has the same general structure as the corpus of the stomach with epithelial lined villous folds that invaginate into gastric pits. The cells lining the villous folds are surface mucus cells that produce alkaline mucus to protect the gastric mucosa from the acidic content of the stomach (**figures 6.5-6.6**)

In the section of the duodenum the autofluorescence highlights intestinal glands (aka crypts of Lieberkeun) which are lined with numerous relatively undifferentiated columnar cells that usually undergo two rounds of mitosis before differentiating into either absorptive cells or goblet cells. The cells that line the intestinal glands are enterocytes, goblet cells, paneth cells and enteroendocrine cells. Usually, eosinophilic granules of paneth cells lie within the apical cytoplasm of the intestinal glands as they recognise antibacterial enzymes secreted from the paneth cells. At the edge of the image is the muscularis mucosae, a thin layer of smooth muscle that marks the bottom end of the mucosa (**figure 6.7**).

The mucosa of the colon highlights its simple columnar epithelium on the edge of the lumen where the surface columnar epithelium and the cells lining the crypts are enterocytes, with an oval basal nucleus and apical brush border, the microscopic representation of microvilli. Simple tubular intestinal glands (crypts of Lieberkuhn) extend through the entire thickness of the mucosa. There are also numerous mucous secreting goblet cells recognized by their content of a large mucous globule. The lamina propria with connective tissue and inflammatory cells surround the crypts (**figure 6.8**).

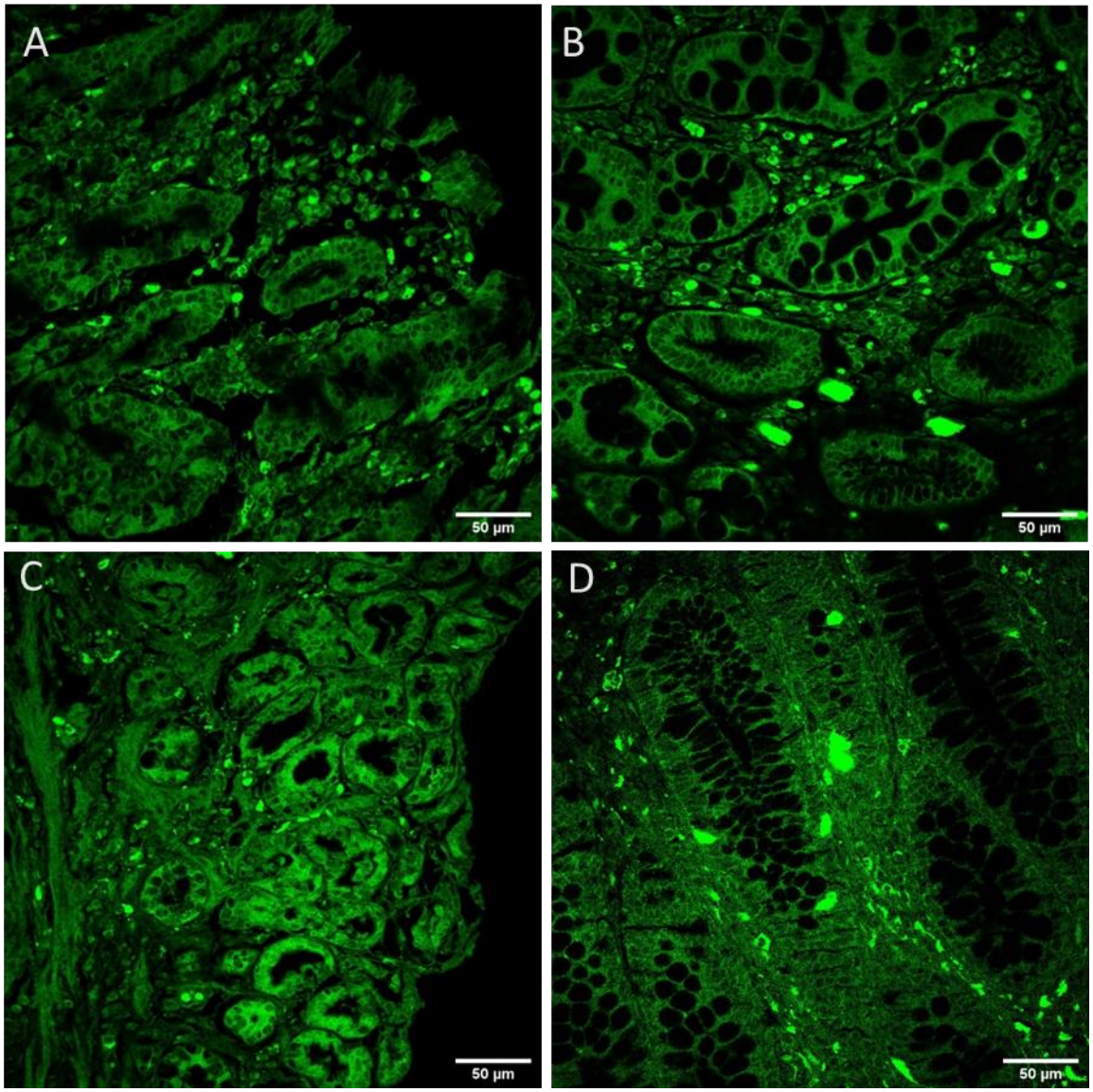


Figure 6.4: Micrograph of unstained sections of **A.** corpus, **B.** antrum, **C.** duodenum and **D.** colon taken at 487 nm. Brightness and contrast have been increased to show structural detail. Scale bar shows 50 µm length. See detailed morphological explanations in the following pages.

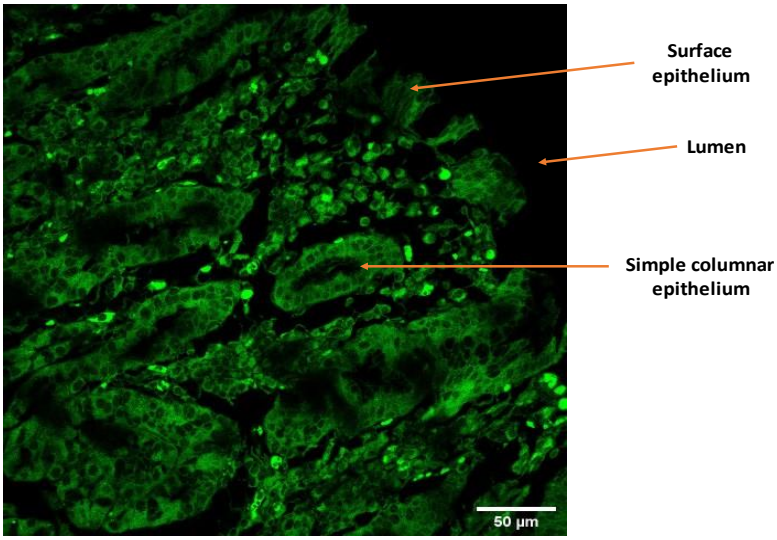


Figure 6.5: Fluorescence microscope image of the mucosa of the human stomach corpus. The thick mucosa is lined with the surface epithelium beneath which sits the simple columnar epithelium. The lumen is also shown. The image was taken at 488 nm. Scale bar shows 50 µm length.

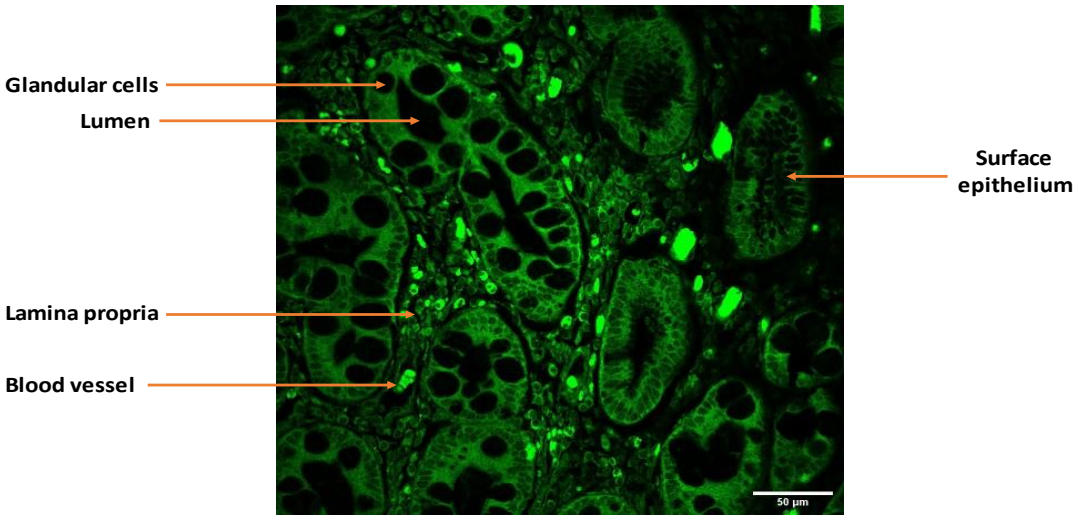


Figure 6.6: Fluorescence microscope image of the antrum of the gastrointestinal tract taken from human source. Image encompasses the mucosa in which the surface epithelium lines the outermost edge of the tissue. The glandular cells are straight glands that extend from the lowest portion of the mucosa to their opening in the bottom of the gastric pits. The mucosa is lined by lamina propria and blood vessels that run through the section. The image was taken at 487 nm. Scale bar shows 50 µm length.

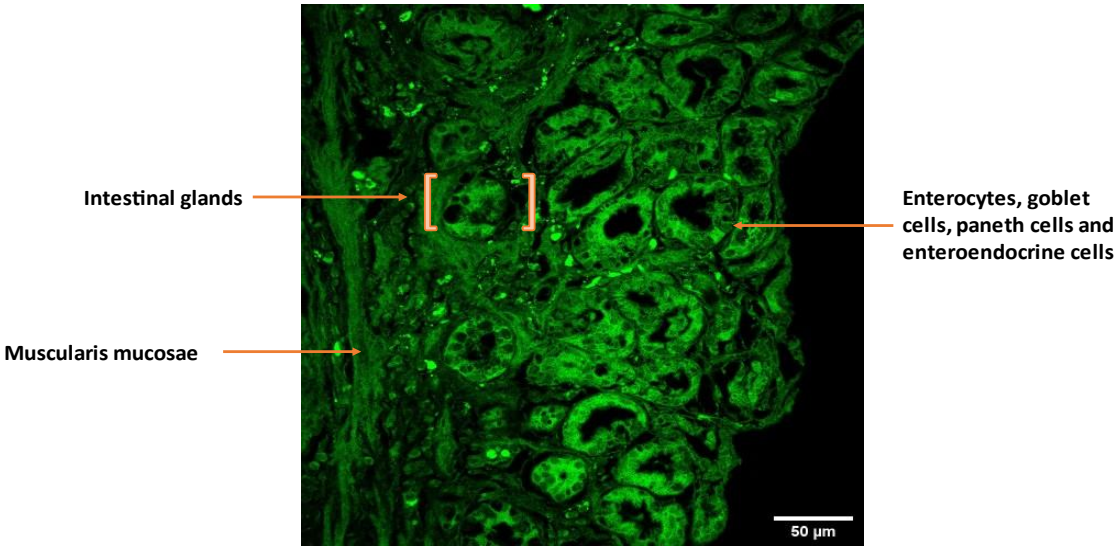


Figure 6.7: Fluorescence microscope image of the duodenum which extends from the lower portion of the stomach to the upper part of the small intestine of the gastrointestinal tract. The image is of a section of the mucosa of the duodenum that underlies the villi. Intestinal glands are lined by either absorptive (enterocytes) or goblet cells. At the left side of the image is the junction between the mucosa and the muscularis mucosae. The image was taken at 487 nm. Scale bar shows 50 µm length.

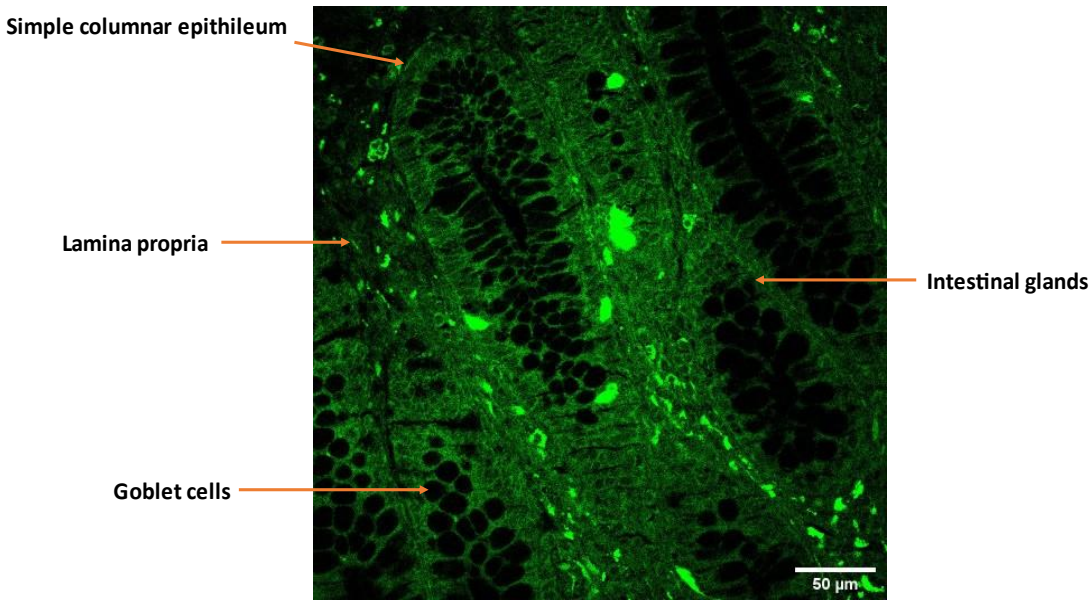


Figure 6.8: Fluorescence microscope image of the colon of the large intestine of the gut. The intestinal glands extend through the whole section of the mucosa. Enterocytes are found within the simple columnar epithelium. The lamina propria has connective tissue and inflammatory cells that surrounds the crypts. There are also numerous mucous secreting goblet cells recognized by their content of a large mucous globule. The image was taken at 487 nm. Scale bar shows 50 µm length.

6.1.4 Reducing background fluorescence of gut tissue

As can be seen in **figures 36** to **40,** even in the absence of staining with the fluorescent wtMOMP^{flu} protein, there was background fluorescence from the tissue in the green spectrum. Although useful for structural analysis of the sections of the gastrointestinal gut, the background fluorescence would make it difficult to ascertain the amount of signal generated from the immunofluorescence assay using fluorescein once wtMOMP^{flu} was added. Tissue components such as red blood cells (RBCs) and collagen are strongly fluorescent, making it difficult to differentiate between relevant signal and background noise. In addition, the fixation of the tissue by formalin, which is commonly used to immobilize antigens while retaining cellular structure, introduces a significant amount of fluorescence[209]. A spectral scan was performed to see if background fluorescence was only present in the green spectrum or if fluorescence was also emitted in other channels (blue and red). This would allow to choose the correct probe to use for future studies of these particular tissue sections.

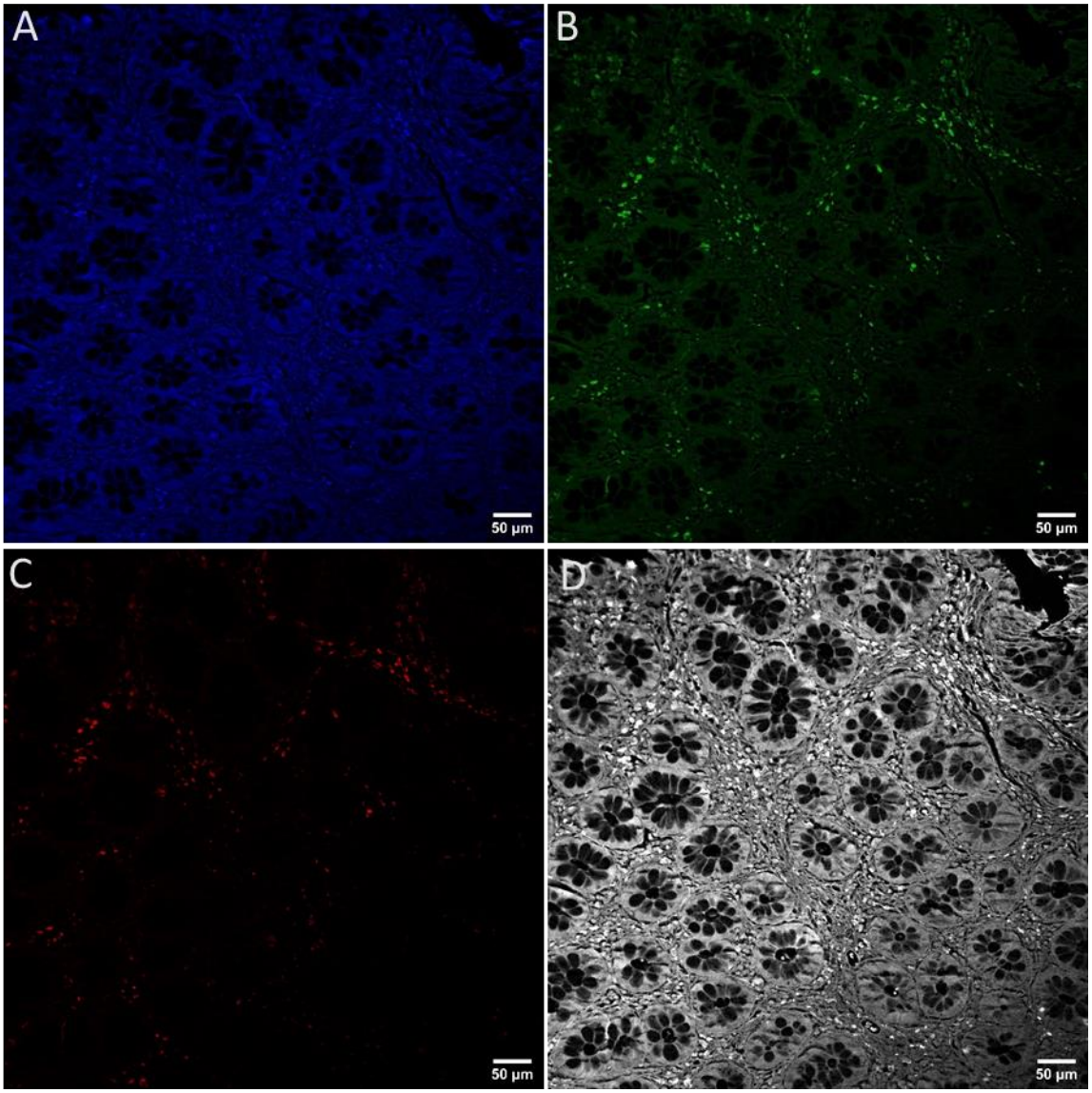


Figure 6.9: Images of the human stomach corpus were captured at wavelengths in the **A.** blue, **B.** green and **C.** far-red channels during a spectral scan. **D.** Digitally inverted white light image for structural detail is also shown. The minimum and maximum brightness and contrast levels are consistent to portray different levels of fluorescence from the same tissue section. Scale bar shows 50 µm length. Histogram visualisation settings: Minimum 274, maximum 6185.

The results from spectral scanning (**figure 6.9**) shows that there is some background fluorescence in every major channel; blue, green and far-red (in order of decreasing intensity). One method to counteract the presence of autofluorescence, when fluorescent staining, is to use a fluorescent tag that corresponds to the channel that shows the least background fluorescence. **Figure 6.9** shows that the far-red channel emits the least light and thus would interfere the least with the

immunohistochemistry experiments. To further eliminate high background signal potential, the probe was switched from fluorescein to far-red. Thus, a far-red linker was used to tag wtMOMP (**figure 6.9**).

To continue with the use of fluorescently tagged protein the auto fluorescence in each channel needed to be reduced. Usually, Sudan Black would be used to treat the cells to reduce their background fluorescence, but it is less efficient at reducing autofluorescence from aldehyde fixation, connective tissue elements, and RBCs than it is at reducing background fluorescence caused by lipofuscin granules. Other treatments such as sodium borohydride and photobleaching have been shown to be effective in diminishing tissue autofluorescence, however there are limitations and problems associated with each of these methods. Instead, a novel technique for autofluorescence removal was used to eliminate native background signals so as not to interfere with the signal after wtMOMP^{flu} staining. Vector® TrueVIEWTM Autofluorescence Quenching Kit was used to reduce tissue fluorescence. It is a non-fluorescent molecule that also binds effectively to formalin-fixed tissue. It involves the treatment of tissue with an aqueous solution of a hydrophilic molecule that binds electrostatically to collagen, elastin, and RBCs.[210]

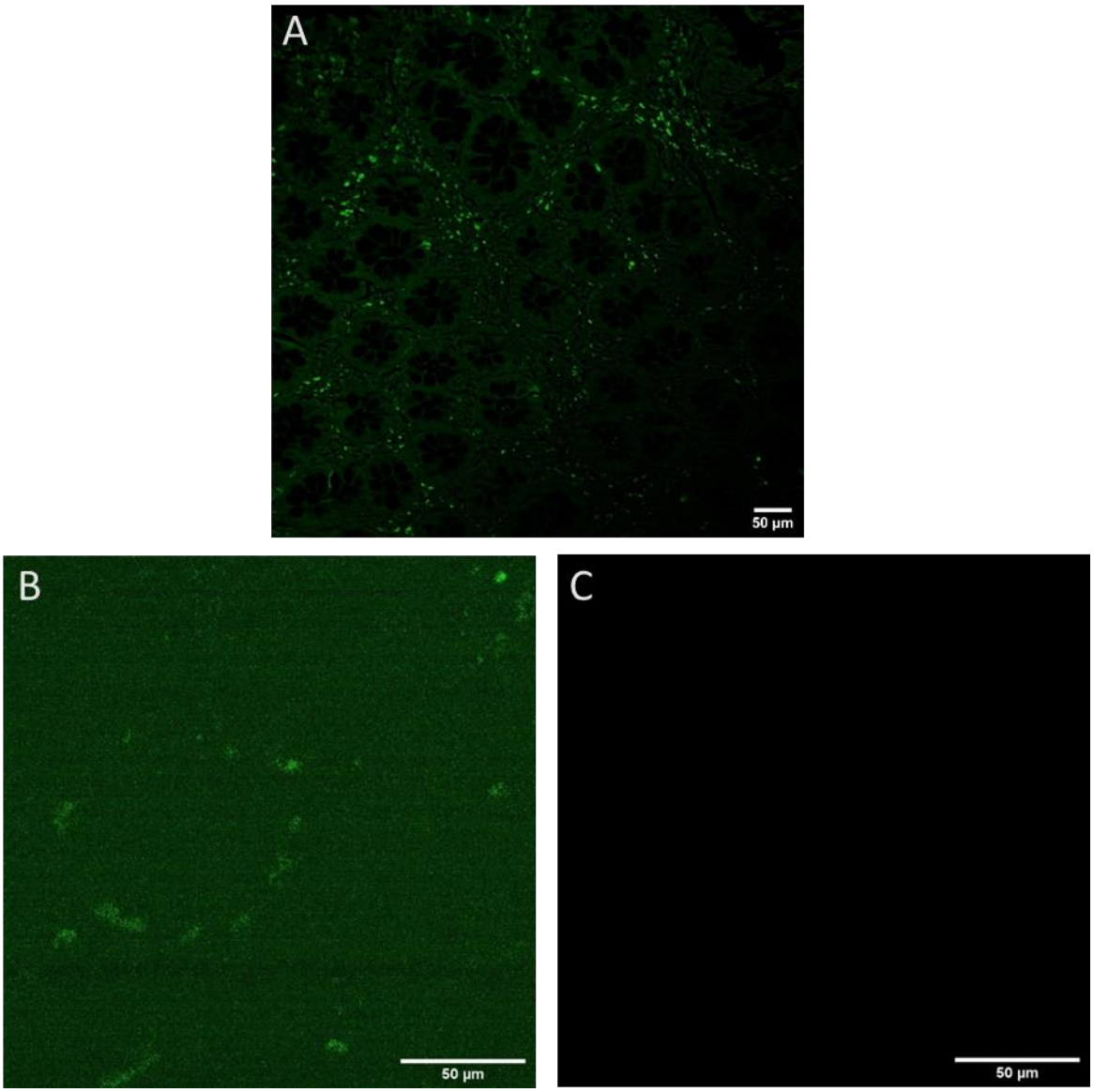


Figure 6.10: Image of the human corpus in the green fluorescent channel after treating the unstained tissue with a quencher. 16 bit-raw images of **A.** and **C**. were visualised on a scale of 274 low and 6185 high values in the Brightness Contrast tool of Fiji to show presence of autofluorescence before **(A.)** and after **(C.)** quenching. **B**. has been 20x magnified and brightness increased to show structural detail. Scale bar is 50 µm length.

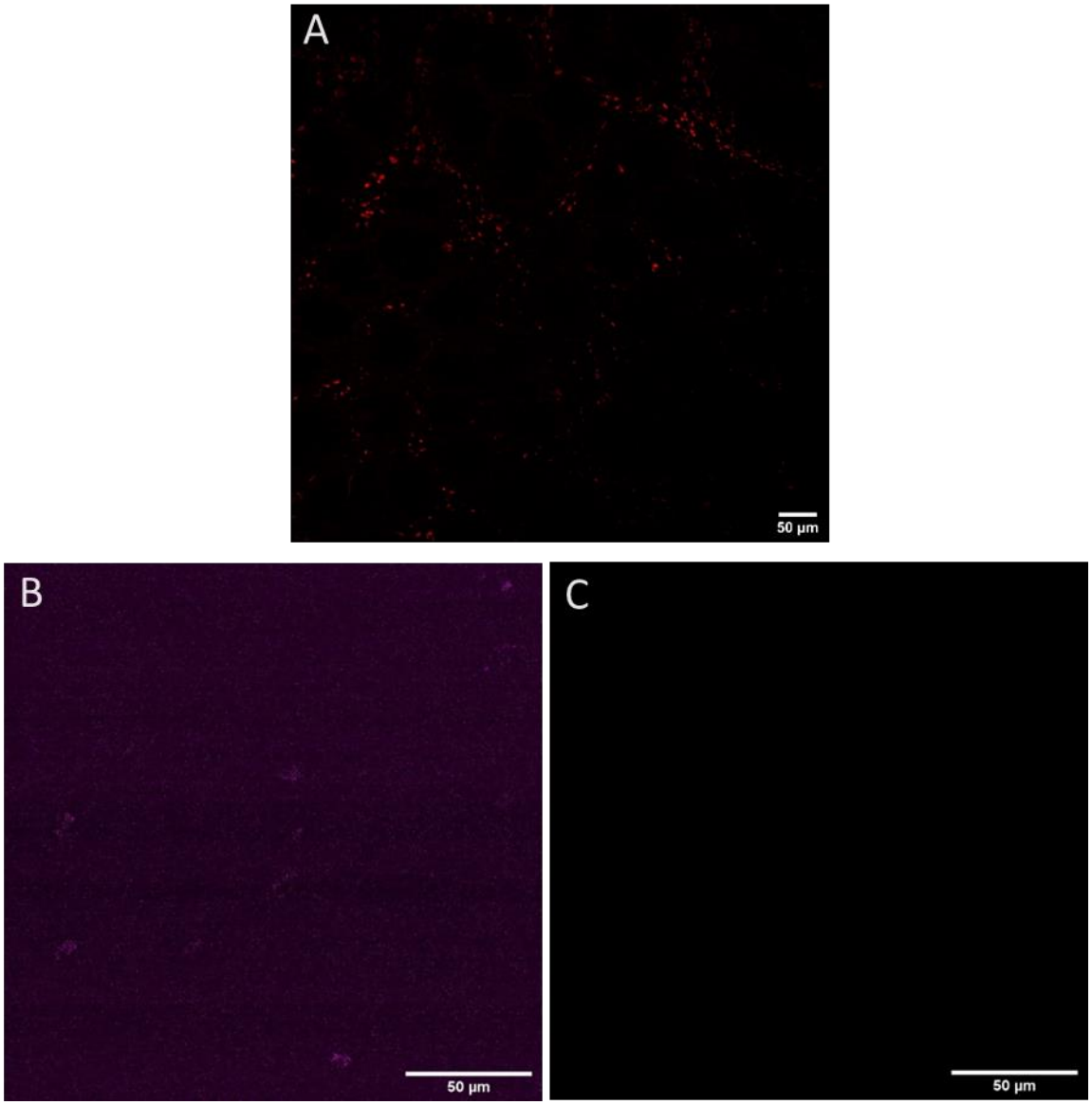


Figure 6.11: Fluorescence images taken of the human corpus in the far-red fluorescent channel after treating the unstained tissue with a quencher**. A.** Fluorescence from tissue before quenching **B.** Fluorescence from tissue after quenching where the brightness and contrast has been increased to show the structural outlines of the cells. **C.** The same image as **B** but here the minimum and maximum contrast and brightness values have been set to that of A to show reduction in autofluorescence after use of quencher. **B** .and **C.** are 20x magnified. Scale bar shows 50 µm length.

Results of quenching show that the there is a significant reduction in autofluorescence after the use of the quencher both in the green and in the far-red channel. The structure of the tissue can no longer be seen at pre-quench levels of brightness and contrast (**figure 6.10C, figure 6.11C**) unless the exposure of fluorescence is significantly increased (**figure 6.10B, 6.11B**). This shows that quenching has worked to reduce the native tissue fluorescence.

To determine whether some of the staining that was observed was specific to the cells an image that comprised mainly of the glass slide was viewed in confocal microscopy. In **figure 6.12** it can be seen that as well as staining of the human corpus tissue with wtMOMP, there are also signals emitting from where there is no tissue. The fluorescence signal emitting from the glass slide and the signal emitting from the tissue differ in intensity and concentration whereby the signal from the tissue is more concentrated. Optimizing the incubation and washing steps reduced the amount of overloading that was occurring in some instances.

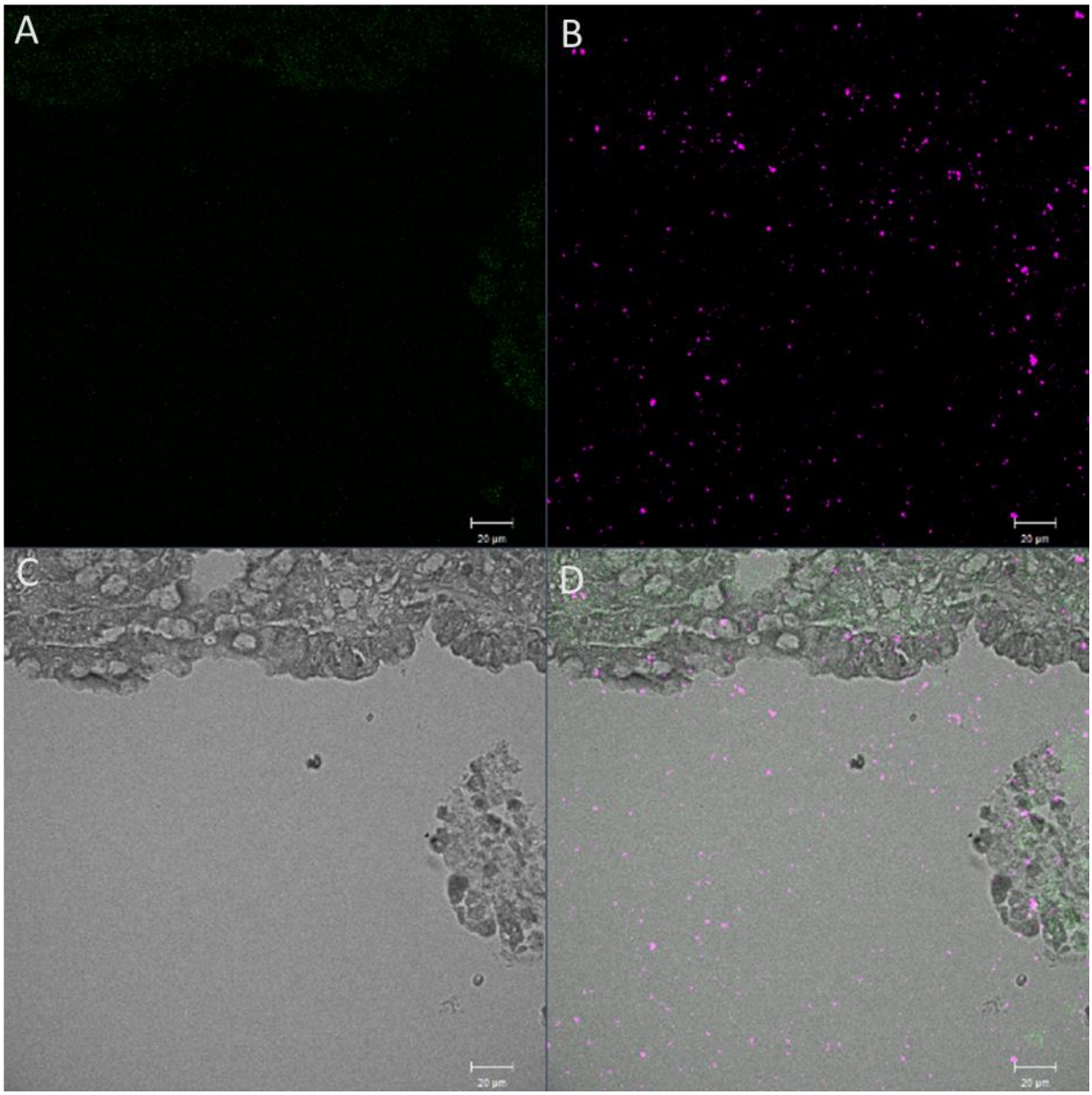


Figure 6.12: Confocal microscope images of the human corpus glands that were first quenched with autofluorescence followed by staining with wtMOMP conjugated to a probe that fluoresces in the far-red spectrum. **A.** Fluorescence of the stained tissue in the green channel **B.** Fluorescence of the stained tissue in the far-red channel. **C.** Light phase image of the tissue. **D.** Overlay of image **B** on the light phase image. The glass slide is prominent in this image. Scale bar shows 20 μm length.

Next, we wanted to optimise the amount of signal that could be obtained from the fluorescent probe in the presence of the quencher. The TrueView kit states that the quencher can be applied before or after the standard immunofluorescence procedure. The effect of adding the quencher before and after staining the tissue with wtMOMP was examined. The kit requires the addition of an aqueous solution of a hydrophilic molecule that binds electrostatically to collagen, RBCs and elastin. We wanted to test a) whether this hydrophilic molecule would disrupt interaction of the tissue antigens with the wtMOMP$^{far\text{-}red}$ if the quencher was used first followed by IF staining and b) if the fluorescence of the probe was dampened when the quencher was added post staining with wtMOMP$^{far\text{-}red}$.

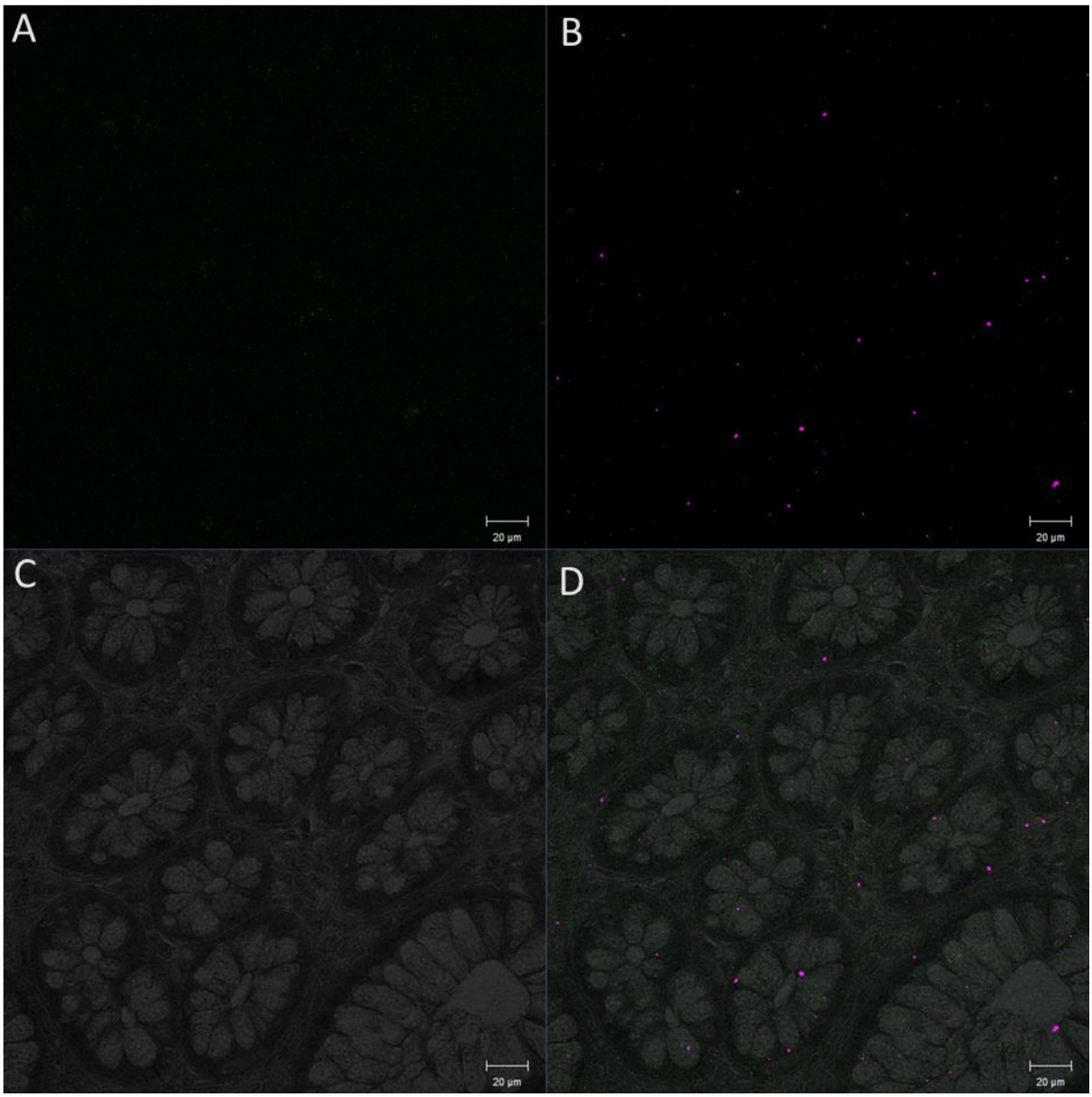


Figure 6.13: Confocal microscope images of a section of the human corpus glands that were first quenched with autofluorescence followed by staining with wtMOMP conjugated to a probe that fluoresces in the far-red spectrum. **A.** Fluorescence of the stained tissue in the green channel to check for residual autofluorescence after quenching. **B.** Fluorescence of the stained tissue in the far-red channel. **C.** Light phase image of the tissue. **D**. Overlay of image B on the light phase image. Far-red signal has been processed as magenta to help with visualisation. Scale bar shows 20 µm length.

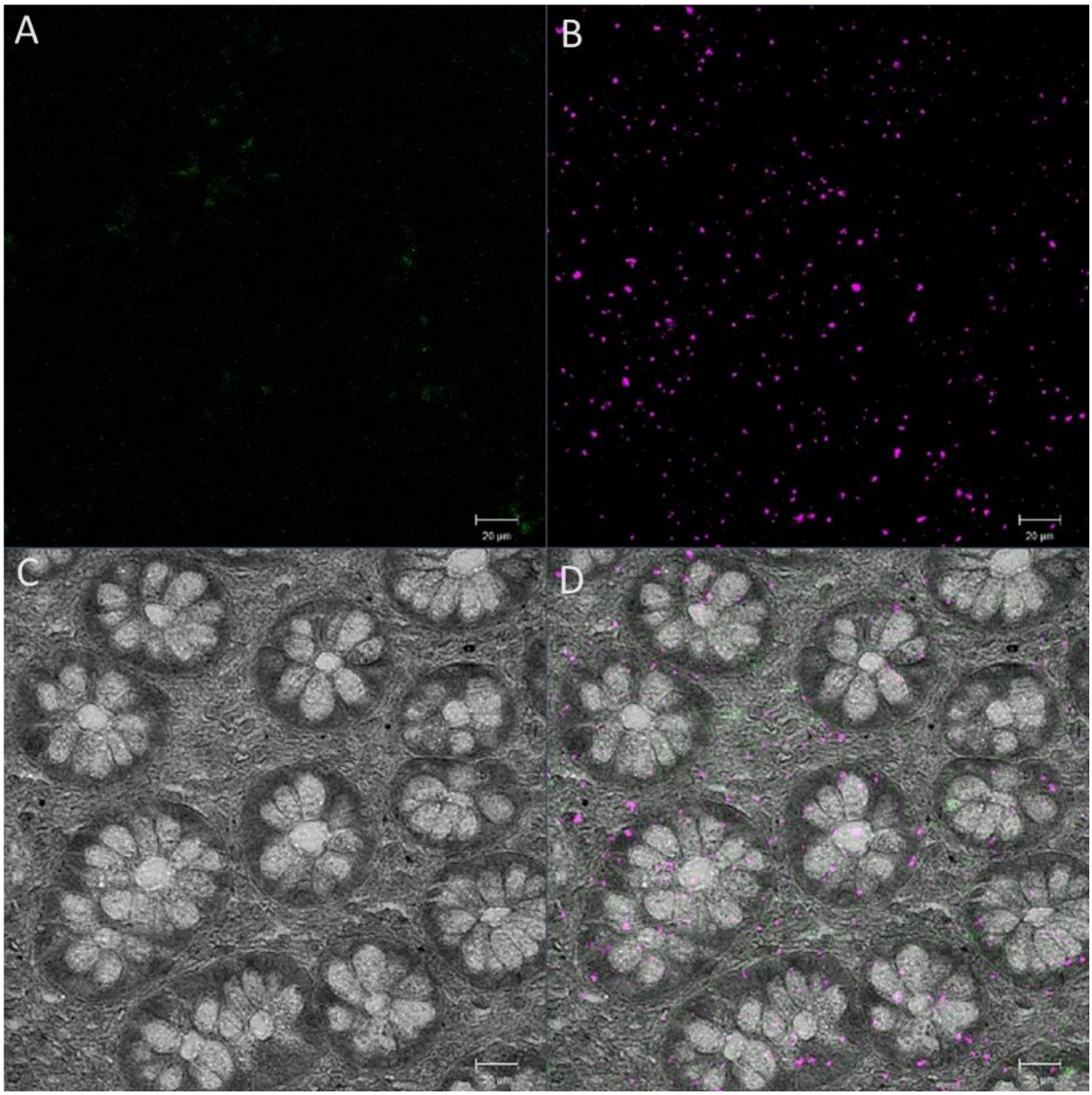


Figure 6.14: Confocal microscope images of a section of the human corpus glands. The section was first stained with wtMOMP conjugated to a probe that fluoresces in the far-red spectrum then followed with treatment with an autofluorescence quencher. **A.** Fluorescence of the stained tissue in the green channel **B.** Fluorescence of the stained tissue in the far-red channel. **C.** Light phase image of the tissue. **D.** Overlay of image B on the light phase image. Far-red signal has been processed as magenta to help with visualisation. Scale bar shows 20 µm length.

In **figure 6.13**, quenching the tissue before going on to stain with the fluorescently tagged protein shows a reduction in autofluorescence of the tissue in the green channel. Following quenching, the addition of the immunofluoresecent tag wtMOMP$^{far\text{-}red}$ is not affected as there is signal in the far-red channel after normalisation of the contrast and brightness. This shows that the application of the quencher is not affecting the binding of the protein to the antigens on the tissue. When we

reversed the order of the application of the treatment where the quencher was added after staining the tissue slide with the fluorescently tagged protein, there was a large presence of the signal of the far-red tag in the red channel, so the quencher did not dampen the emission of the fluorescence from the probe (**figure 6.14**). We concluded that using any order of treatment on the tissue did not affect the signal that was observed from the tissue. Although there is signal from both protocols, there is a cluster of the wtMOMP-far-red localised around the gland cells when the tissue slide is first stained with the fluorescent protein. This may be because there is more access to the tissue before quenching with the kit and the addition of the hydrophilic substrate. So future experiments were performed as in **figure 6.13.**

6.1.5 Immunofluorescence binding of wtMOMP to gut tissue

Now that background fluorescence was reduced and specificity of protein binding to the tissues could be considered, the next step was to localise the binding of the MOMP protein to the gut tissues.

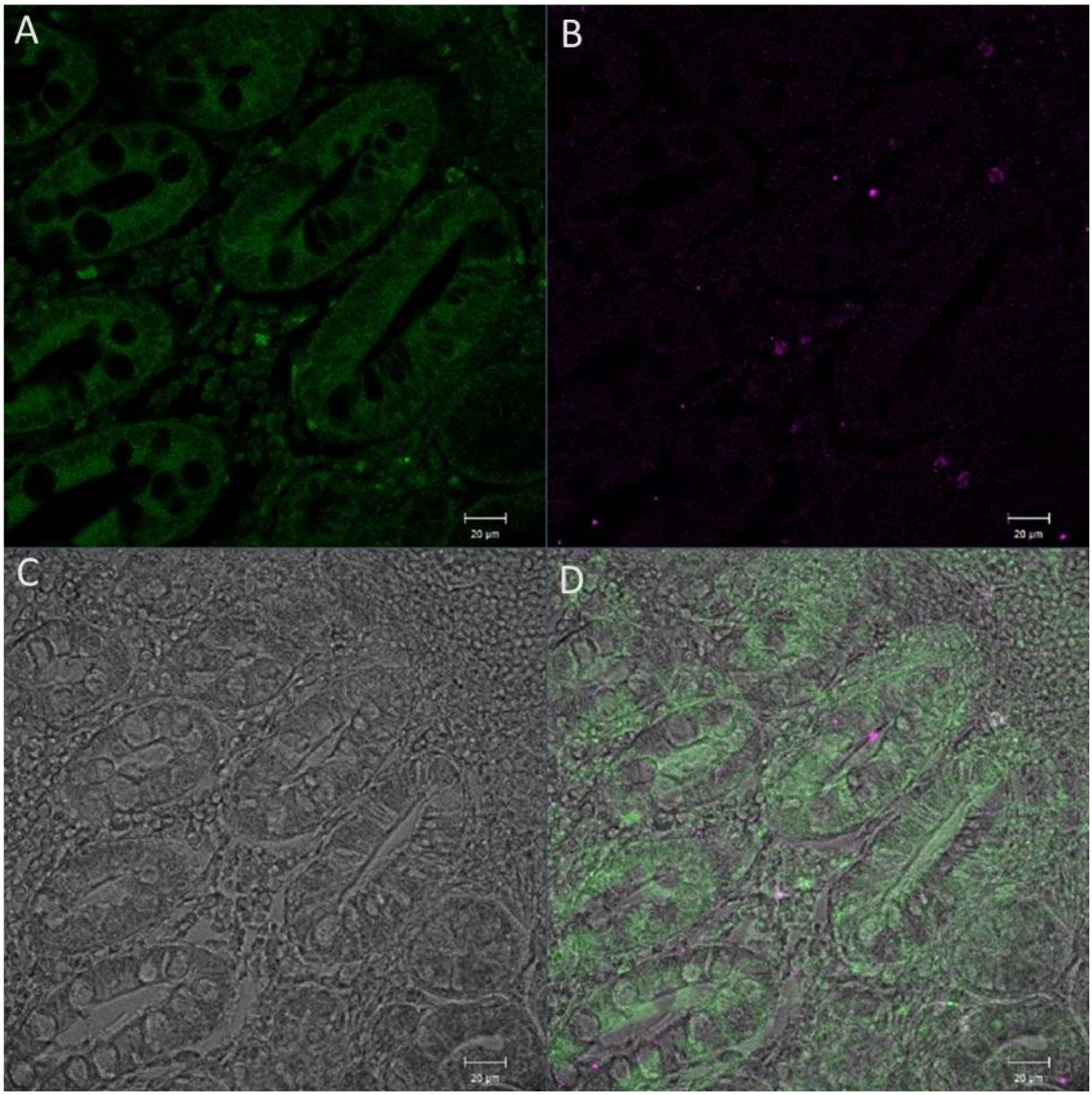


Figure 6.15: Confocal microscope images of the human colon after staining with wtMOMP conjugated to a probe that fluoresces in the far-red spectrum. **A.** Fluorescence of the stained tissue in the green channel **B.** Fluorescence of the stained tissue in the red channel. **C.** Light phase image of the tissue. **D.** Overlay of image B on the light phase image. Scale bar shows 20 μm length.

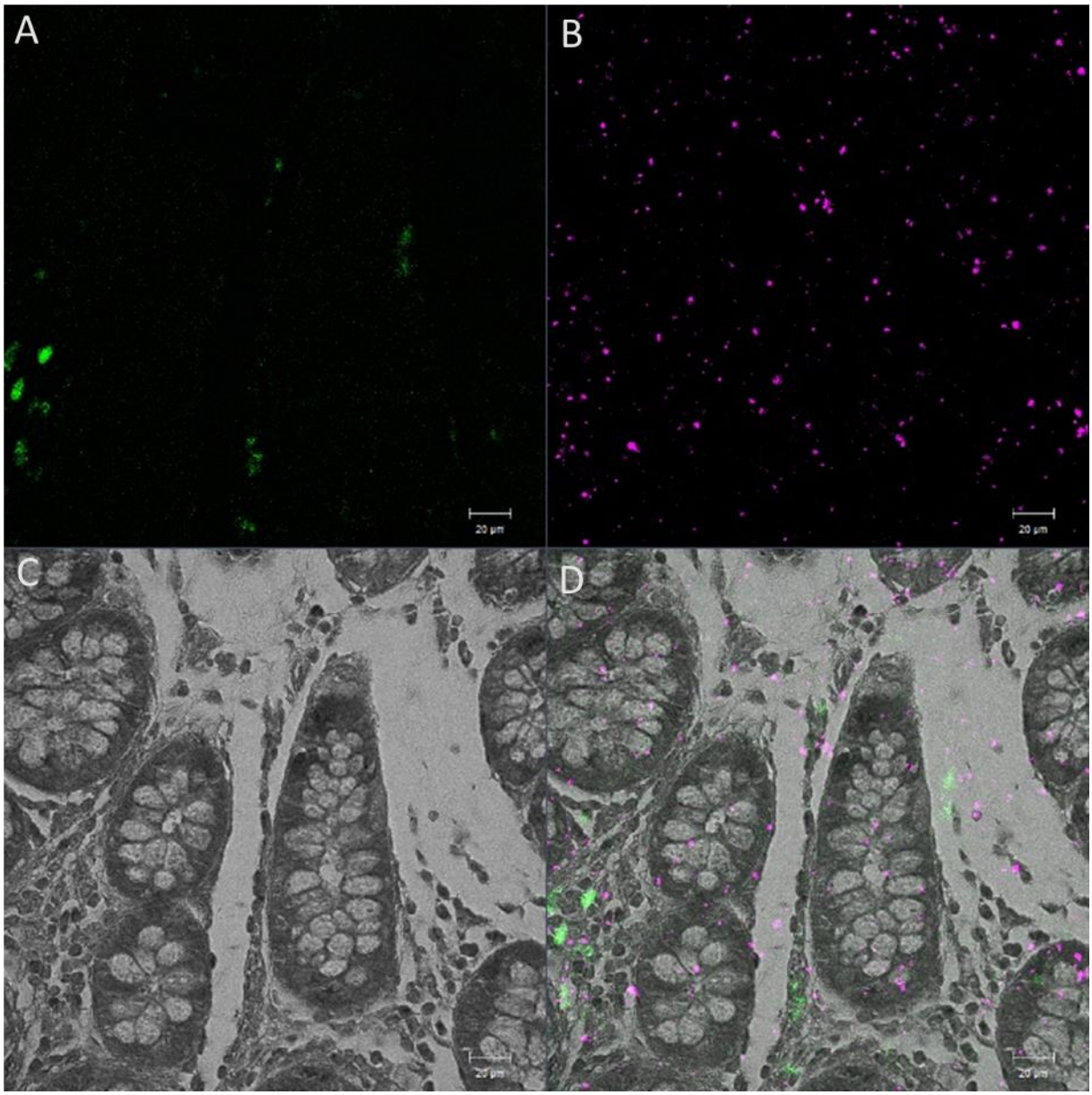


Figure 6.16: Confocal microscope images of the lower mucosa of the human corpus after staining with wtMOMP conjugated to a probe that fluoresces in the far-red spectrum. **A.** Fluorescence of the stained tissue in the green channel. **B.** Fluorescence of the stained tissue in the far-red channel, **C.** Light phase image of the tissue, **D.** Overlay of image **B** on the light phase image. Scale bar shows 20 μm length.

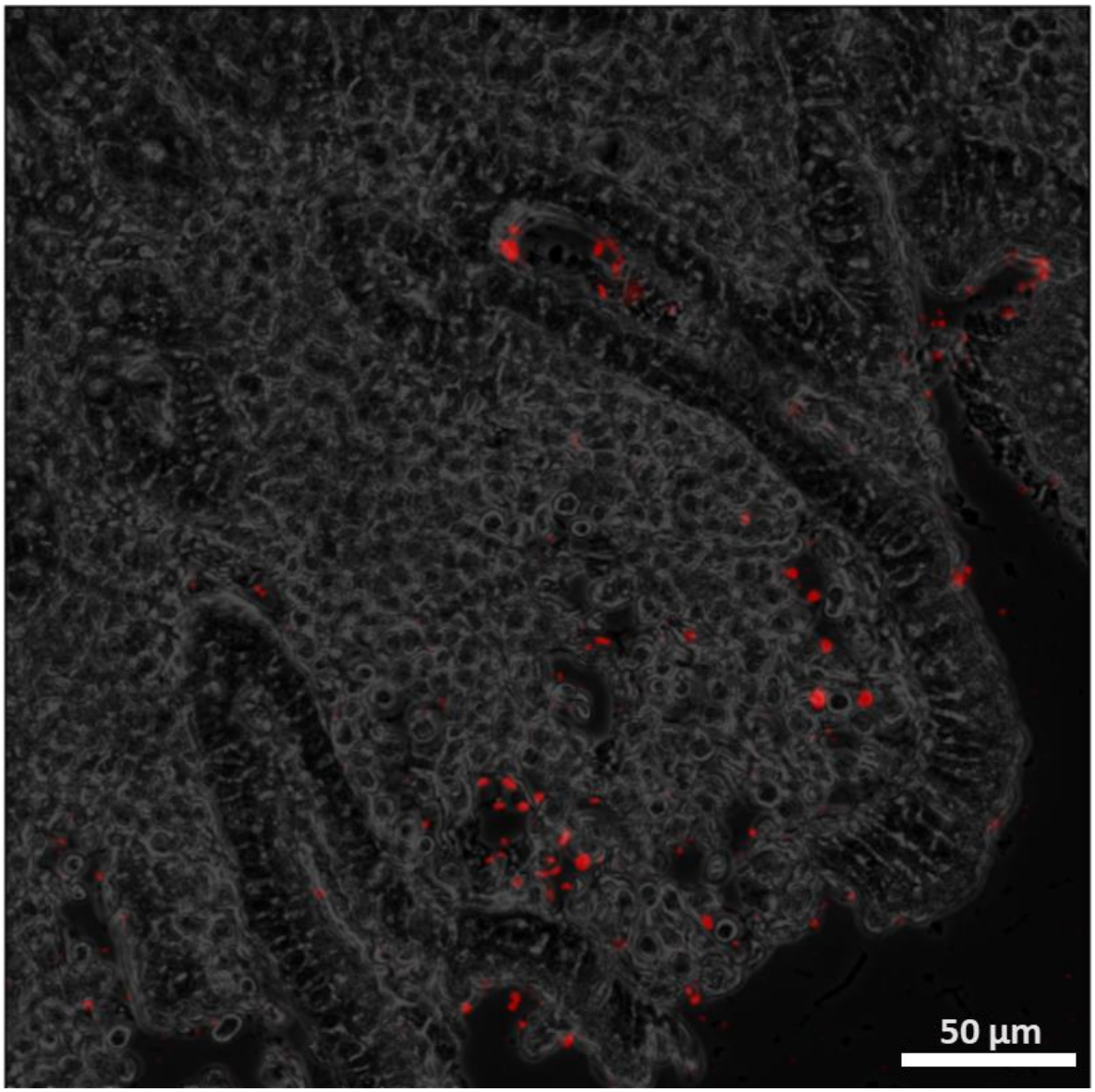


Figure 6.17: Confocal microscope image of the simple columnar epithelium of human corpus tissue which forms gastric pits. Image acquired after staining with wtMOMP conjugated with far-red probes. Images of the light phase and far-red fluorescence filter are overlayed over each other. 16-bit raw images of the far red were visualised on a scale of 274 low and 6185 high values in the Brightness Contrast tool of Fiji. Scale bar shows 50 µm length.

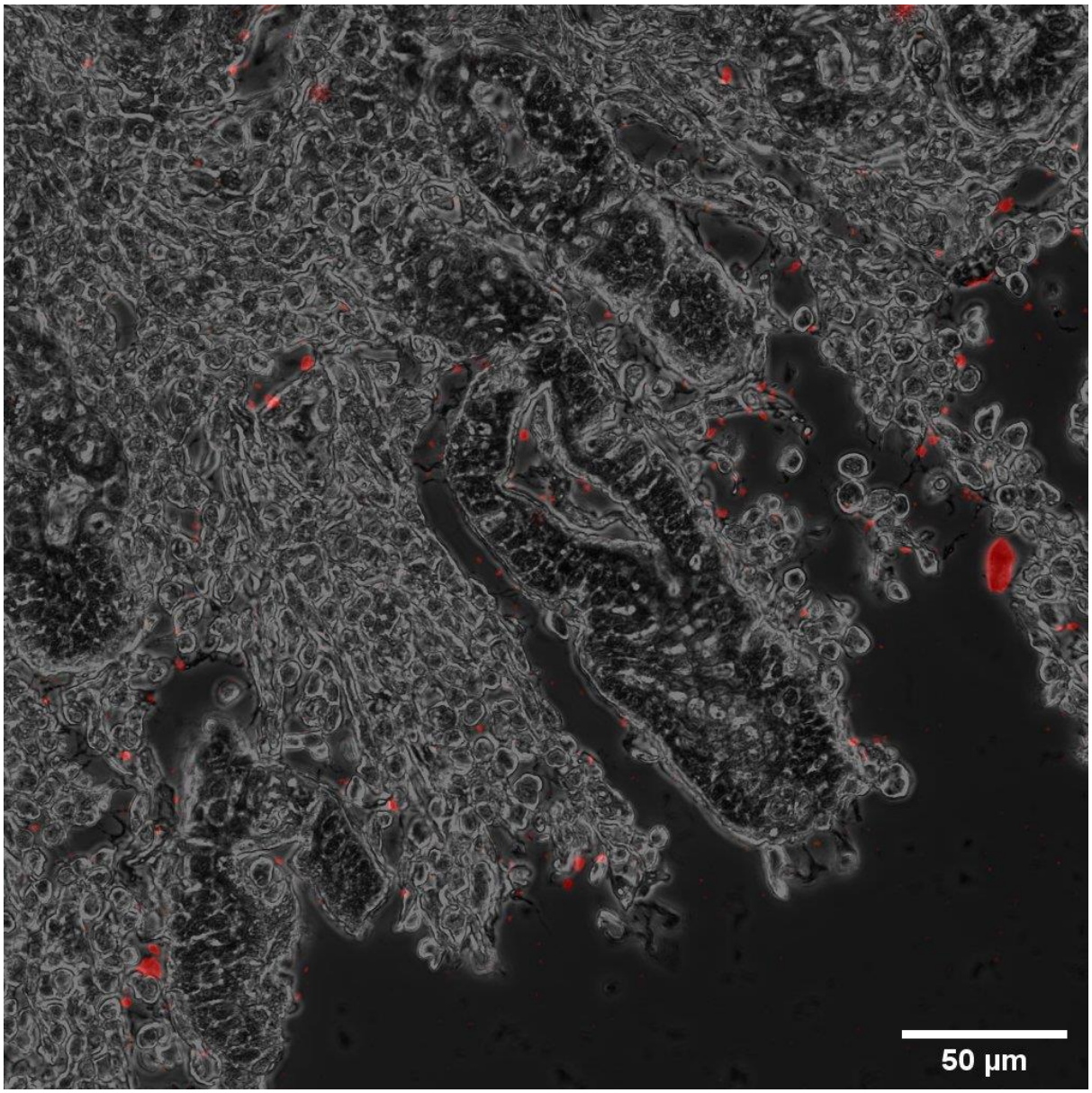


Figure 6.18: Confocal microscope image of the simple columnar epithelium and surface epithelium of human corpus tissue after staining with wtMOMP conjugated with far-red probes. Images of the light phase and far-red fluorescence filter are overlayed over each other. 16-bit raw images of the far red were visualised on a scale of 274 low and 6185 high values in the Brightness Contrast tool of Fiji. Scale bar shows 50 µm length.

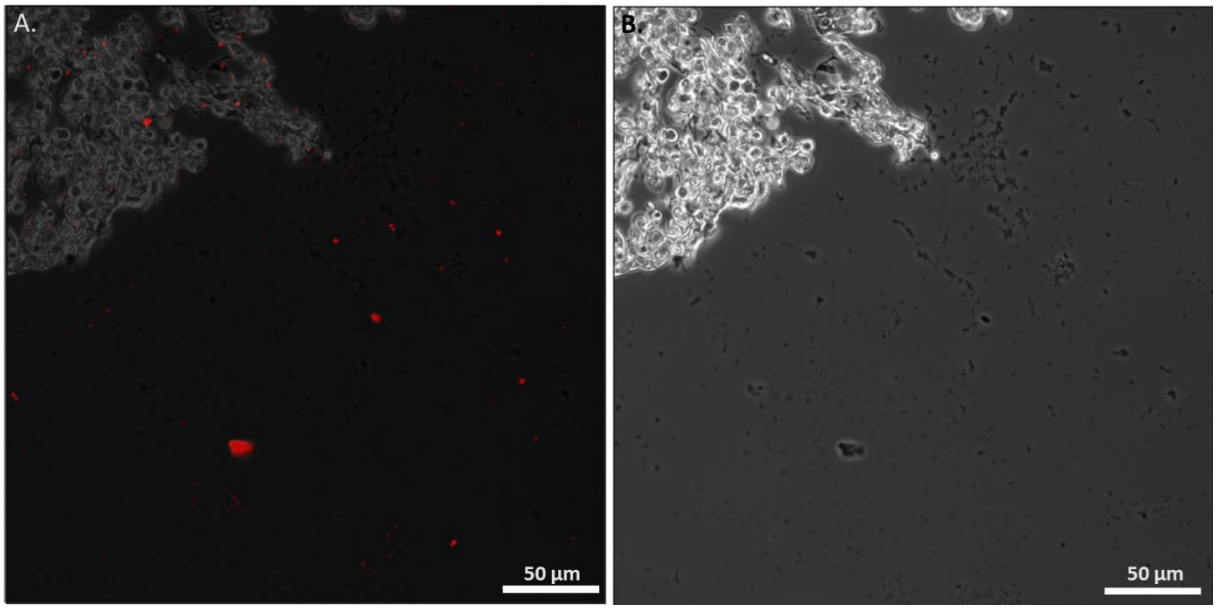


Figure 6.19: Confocal microscope image of surface epithelium of human corpus tissue after staining with wtMOMP conjugated with far-red probes. The glass slide is prominent in the image. **A.** Images of the light phase and far-red fluorescence filter are overlayed over each other.16-bit raw images of the far red were visualised on a scale of 274 low and 6185 high values in the Brightness Contrast tool of Fiji. **B.** Light phase image only, to show contamination on the glass slide which may be contributing to adhesion of protein on the glass. Scale bar shows 50 µm length.

It can be seen that there is a low level of binding of wtMOMP$^{far\text{-}red}$ on the cells of the colon (**figure 6.15**) when compared to binding of wtMOMP$^{far\text{-}red}$ on the corpus (**figure 6.17, 6.18**). This is consistent with the fact that most blood group substances are absent from the distal portion of adult colon (but are expressed throughout fetal colon). Studies show that it is only in cancers of the adult distal colon where there is re-expression of the BgAgs, suggesting that these antigens behave as oncofetal antigens and biomarkers for colon cancer[211]. **Figure 6.19** shows that there are areas outside of the gut tissue that wtMOMP$^{far\text{-}red}$ is binding to such as the glass slide. The light phase image shows that the binding is consistent with contamination on the slide. **Figure 6.18** also

shows areas of glass contamination but wtMOMP$^{far\text{-}red}$ has not bound to these contaminants, which provides increased reassurance that the binding that is seen on the tissue is highly specific.

The results in **figure 6.13** to **figure 6.16** of the colon and the lower mucosa of the corpus show that after treatment of the tissue with the fluorescence blocker, signal could be seen from staining of the tissue with wtMOMP$^{far\text{-}red}$. The localisation of the signal is random and of a punctate nature. In **figure 6.17** and **figure 6.18**, the binding of the protein to the surface epithelium of the corpus tissue cells are more specific and localised to the edge of the tissue cells. The protein is localised to the areas of the simple columnar epithelium, surface epithelium and crypts of the GI tract within the corpus. More specific binding can be seen in the surface of the corpus than on its glandular cells, as seen in **figure 6.14**, where the nature of the interaction with wtMOMP and the tissue seems more random and less localised to specific areas of the cells. We considered the possibility of unspecific binding of the protein to the tissue or the glass slide. **Figure 6.19** shows that there is less binding on the glass and any interaction on the glass may be as a result of some contamination (**figure 6.19B**). From the ELISA binding studies it was shown that wtMOMP has preferential binding to Le^b and Le^a BgAgs over type 1 and type 2. The binding pattern that would be seen in these sets of tissue sections would depend on the blood group of the individuals that the tissue sections have been sourced from, their blood group phenotype and the presence of certain enzymes. There is variation in the biosynthesis of BgAg carbohydrates in the exocrine secretions and gastrointestinal mucosa depending on the presence, absence and combination of FUTII, FUTIII, GTA and GTB are glycosyltransferases coded by FUT2 (Secretor), FUT3 (Lewis) and ABO (ABO) genes. Henry and Samuelsson classified that the presence of FUTII defines the Secretor phenotype and allows the expression of a variety of BgAg carbohydrates whilst its absence leads to the non-Secretor phenotype and reduces the variability of BgAg carbohydrates in exocrine

secretions. FUTIII has a role where its presence means individuals are Lewis positive and allows the expression of a variety of difucosylated BgAg carbohydrate antigens in combination with FUTII. The presence of FUTIII only reduces the variability of BgAg carbohydrates expressed. Its absence classifies individuals as Lewis negative and no difucosylated BgAg carbohydrates antigens can be expressed[94]. If O blood group individuals are Secretors then they will express the monofucosylated H type 1 antigen which will be converted into high levels of difucosylated Le^b antigens. If these individuals are non-Secretors, they will express only high levels of monofucosylated and elongated Lea antigen in the presence of FUTIII. Regarding A, B, and AB blood groups, secretors from these groups express monofucosylated A type 1 and/or B type 1 antigens which will be converted into high levels of difucosylated ALe^b and/or BLe^b, respectively. The non Secretors and Lewis positives in A, B, and AB blood groups express high levels of monofucosylated Lea antigen but if they are Lewis negative they will express high levels of type 1 oligosaccharide precursor which will remain unmodified due to the absence of FUTII and FUTIII[94].

6.2 wtMOMP binding to Le^b +ve mouse tissue

The experiments performed in the previous sub chapter were performed on human tissue, but it is not known whether the patients were Le^b positive (Le^b +ve) or negative. To get a clearer picture and to further study the gene products involved in the bacteria's homing to the gut epithelium via MOMP and Le^b expressing tissue cells, stomach tissue from a genetically well-defined transgenic mouse model (adult FVB/N mice) was sourced to provide further evidence to strengthen the observed binding of wtMOMP to the intestinal tissue. Several assays were performed on this new set of stomach tissue which were either wild type in nature or genetically modified to express Le^b (adult FVB/N mice do not produce Le^b in any of their stomach or intestinal epithelial cell lineages)

from mice that were sacrificed at 28 days old (P28)[212]. The sections were cut from several parts of the whole stomach, which includes the forestomach, corpus, pyloric part and the initial part of the duodenum.

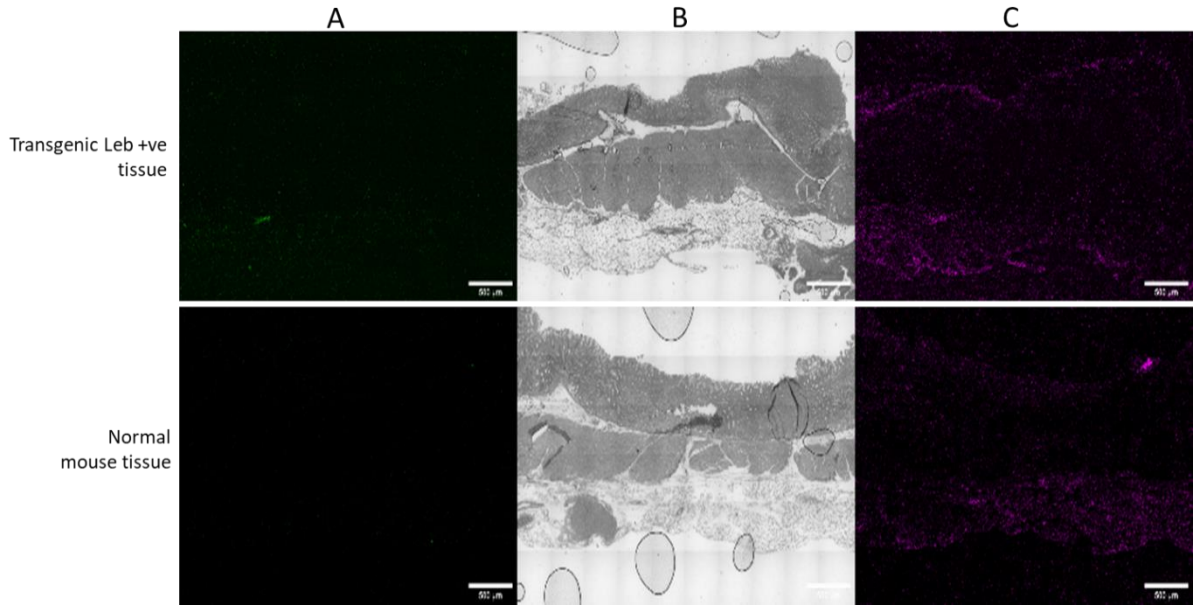


Figure 6.20: P28 transgenic and normal FVB/N mouse stomach tissue stained with Le^b specific mouse antibody followed by staining with far-red conjugated secondary antibody and an autofluorescence quencher. **A.** Images taken in green fluorescent channel, to observe presence of any remaining autofluorescence. **B.** Light phase images at optimum display to display tissue structures. **C.** Images taken in the red fluorescent channel, which has been adapted to magenta for easier viewing. For the transgenic tissue, Image A and C are visualised on a scale of 54 low and 6185 high values in the Brightness Contrast tool of Fiji. For the normal tissue, image A and C are visualised on a scale of 54 low and 4691 high values in the Brightness Contrast tool of Fiji. Scale bar represents distance of 500 µm.

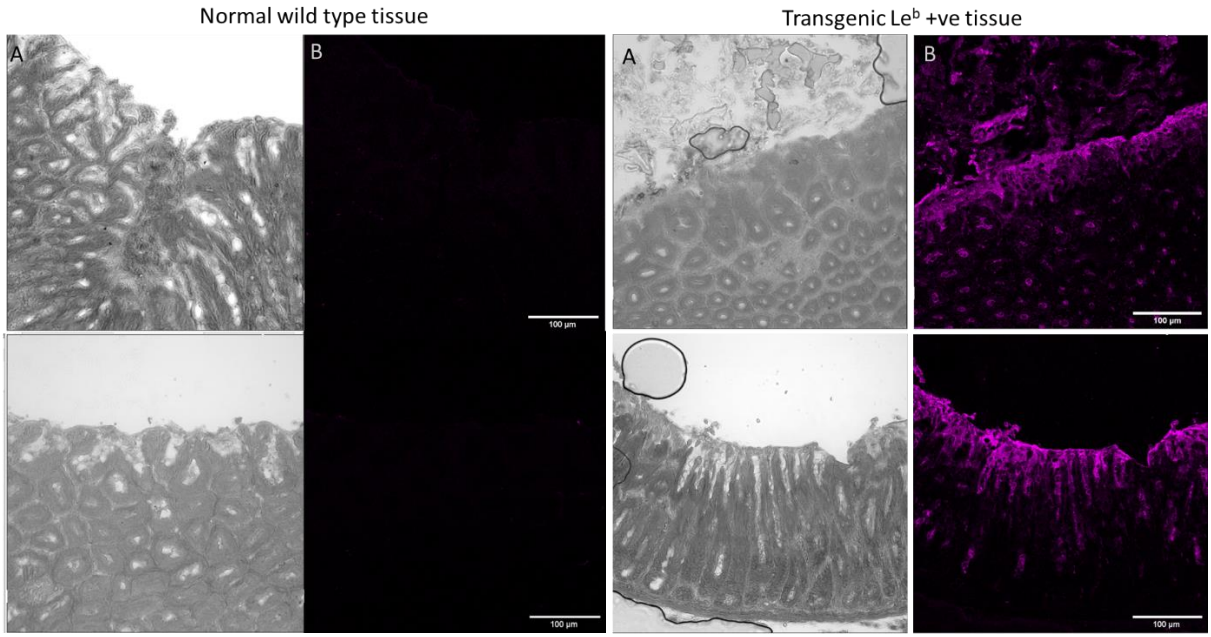


Figure 6.21: P28 FVB/N mouse stomach normal transgenic Le^b +ve tissue stained with Le^b specific mouse antibody followed by staining with far-red secondary antibody and an auto-fluorescence quencher. Images are magnified 20x to focus on different parts of the tissue sections. Panel A shows light phase images at optimum contrast to display corresponding tissue structure. Panel B, which is visualised on a scale of 54 low and 6185 high values in the Brightness Contrast tool of Fiji, shows images taken in the far-red fluorescent channel, which have been modified to magenta for easier viewing. Scale bar represents distance of 100 µm.

The first assay performed on the tissue was to ascertain the extent to which the Le^b +ve tissue express the Le^b antigen, in comparison to normal tissue, by staining with anti-Le^b mouse antibody. It was not possible to obtain anti-Le^b antibody from non-mouse sera. Once stained, a secondary antibody followed, which was goat raised anti-mouse antibody, thus the mouse-on-mouse staining protocol was employed, to reduce the high background signal that may be observed by using secondary mouse antibody on mouse tissue. At first glance from low magnification it seems as if the transgenic and normal tissue are expressing Le^b to the same level (**figure 6.20**) but a lot of the signal is attributed to background noise caused by using mouse antibody on mouse tissue. Although the epithileal surface of the transgenic tissue does show more defined signal than the same region on the normal tissue. Furthermore, magnified images of the same tissue sections (**figure 6.21**) show that the red fluorescence signal is greatly reduced in the normal tissue

compared to the Le^b +ve transgenic tissue, confirming that the transgenic tissue are expressing Le^b antigens.

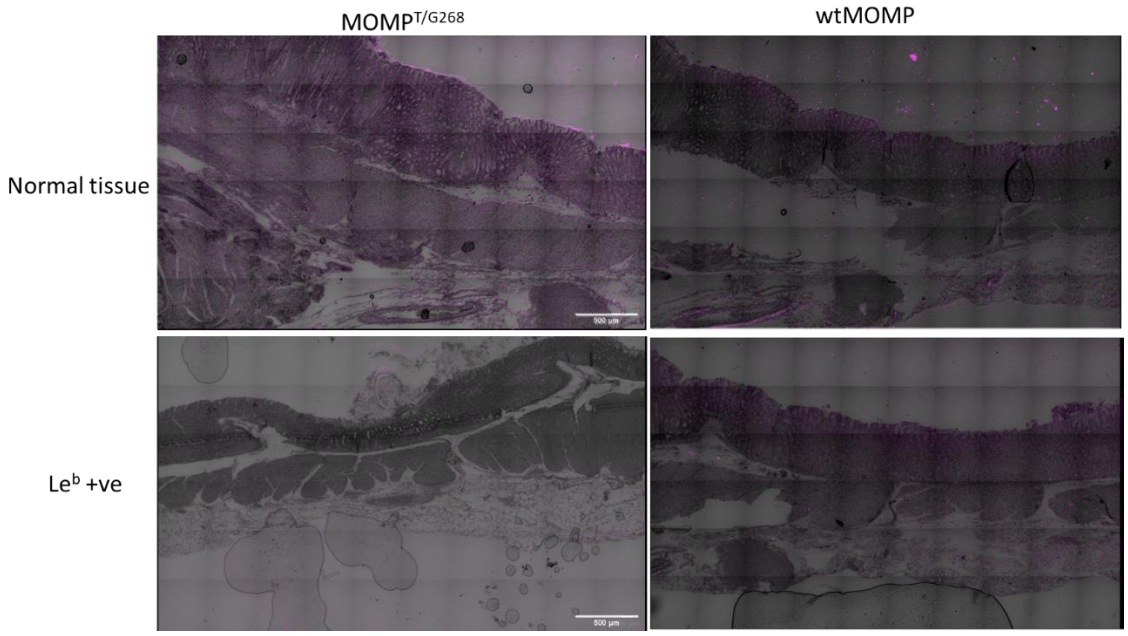


Figure 6.22: The grid shows normal P28 FVB/N mouse stomach tissue and transgenic Le^b +ve P28 FVB/N mouse stomach tissue that were stained with either $MOMP^{T/G268}$ or wtMOMP protein. This was followed by anti-MOMP rabbit antibody counterstained with far-red tagged secondary antibody and an auto-fluorescence quencher. Light phase images have been overlaid with the fluorescent far-red channel images. All images visualised on a scale of 0 low and 65279 high values in the Brightness Contrast tool of Fiji in far-red channel and visualized in magenta. Scale bar represents 500 µm. Further related images supplied in appendix.

To establish the specificity of the binding of the glycosylated wtMOMP to Le^b +ve stomach tissue, I compared the labelling affinity of wtMOMP and the labelling affinity of $MOMP^{T/G268}$ to normal and transgenic tissue. **Figure 6.22** shows that although there was high binding to normal tissue of both wild type and mutant MOMP protein, there was a prominent decrease in binding of $MOMP^{T/G268}$ to transgenic tissue. In the case of wtMOMP, although there is also a decrease in binding to transgenic tissue compared to normal tissue, the difference is not as drastic but the localization of the signal to Le^b +ve tissue is more specific as can be seen by reduced background noise and

contaminating signals on the slide surface. One explanation for high levels of signal when wtMOMP is bound to the normal tissue although there is a lack of Le^b expression, is that glycosylated MOMP would also be binding nonspecifically to other components found in the gut, so optimizing washing/blocking steps during tissue incubation with the protein would decrease such instances of non-specific binding.

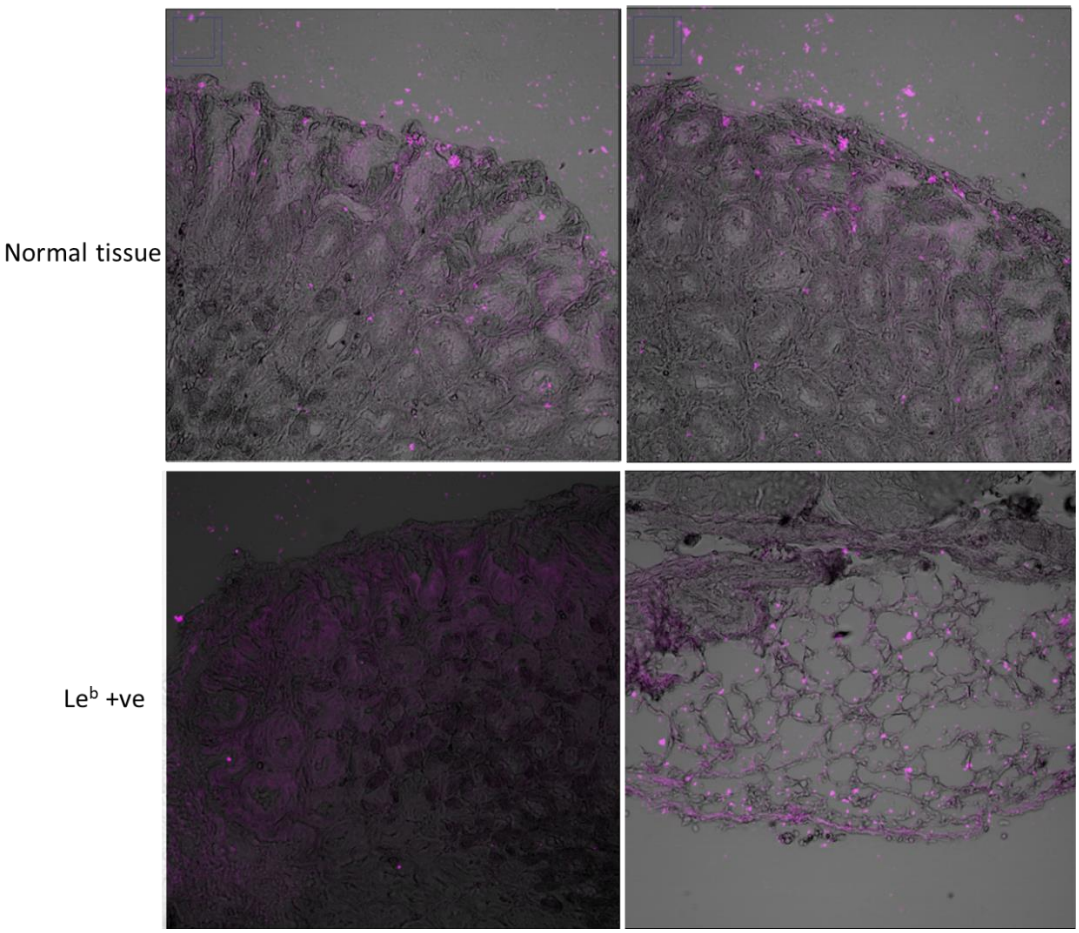


Figure 6.23: 20 x magnification of two regions of normal and Le^b +ve P28 FVB/N mouse stomach tissue stained with wtMOMP followed by staining with anti-MOMP rabbit antibody and counterstained with secondary antibody and an auto-fluorescence quencher. Light phase images have been overlaid with the fluorescent far-red channel images. Images in far-red channel visualised on a scale of 0 low and 65279 high values in the Brightness Contrast tool and pictured in magenta.

Figure 6.23 is of 20x magnification of two regions of the normal tissue and two regions of the transgenic tissue. Here it is easier to see that the fluorescence signal from wtMOMP is more specific to the transgenic tissue than to the normal tissue. When binding to normal tissue, there is

also a greater amount of signal found on the glass slide, pointing to less affinity of the wtMOMP protein to normal tissue than for Le^b +ve tissue.

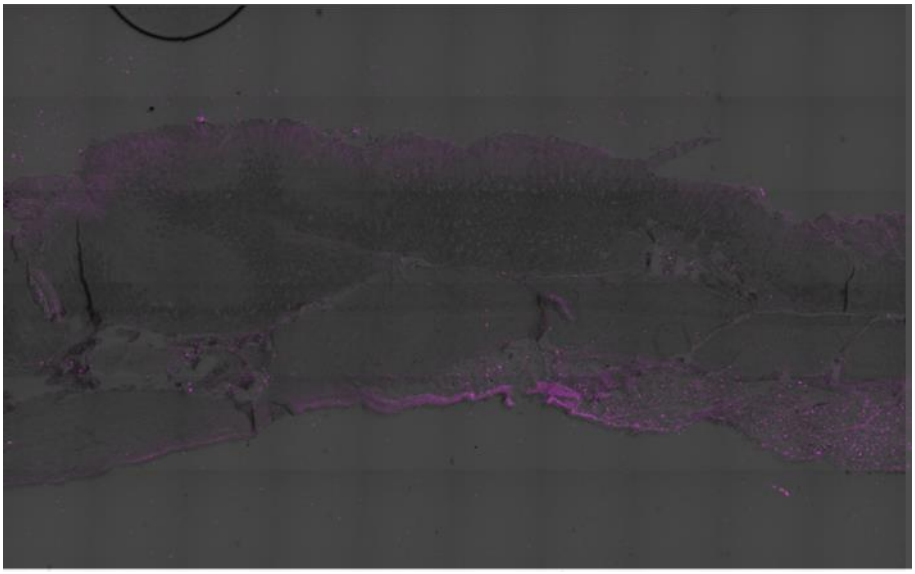


Figure 6.24: Le^b +ve transgenic P28 FVB/N mouse stomach tissue stained with $wtMOMP^{far\text{-}red}$ and an auto-fluorescence quencher. Light phase images have been overlaid with the fluorescent far-red channel images. Images in far-red channel visualised on a scale of 0 low and 65279 high values in the Brightness Contrast tool of Fiji and pictured in magenta.

To confirm the result that we see in **figures 6.23**, an alternative method of staining the Le^b +ve tissue was followed **(figure 6.24)**, to see if similar results could be obtained and to discount possibility of non-specific binding. In this method, the transgenic mice tissue were stained directly with far-red tagged wtMOMP as opposed to anti-MOMP followed by a fluorescently tagged secondary antibody. The results show that the same conclusion was still achieved, whereby binding to Le^b +ve tissue by wtMOMP is specific and localized.

Next, I studied whether the binding specificity of wtMOMP could be blocked by soluble Le^b. In this case, wtMOMP was incubated with increasing dilutions of soluble Le^b in PBS and 0.1% BSA before incubating the tissue sections in these dilutions.

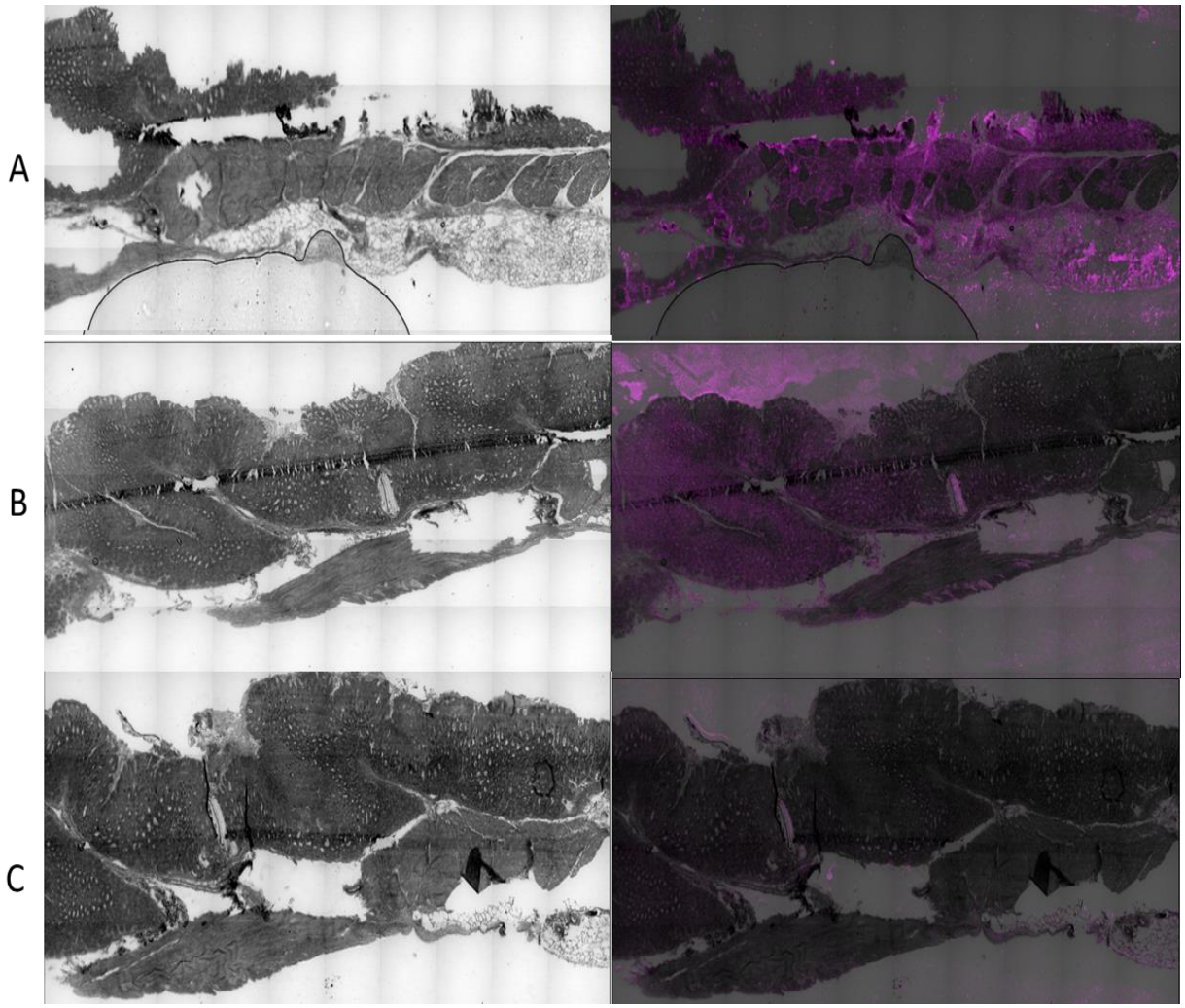


Figure 6.25: Le^b +ve transgenic P28 FVB/N mouse stomach tissue incubated with wtMOMP with increasing concentrations of soluble Le^b, in a competitive binding study. A) Tissue was stained with 0.5 µM Le^b and 5 µM wtMOMP. B) Tissue was stained with 5 µM Le^b and 5 µM wtMOMP. C) Tissue was stained with 50 µM Le^b and 5 µM wtMOMP. Tissues were then stained with anti-MOMP rabbit antibody and counterstained with far-red tagged goat anti-rabbit antibody and an auto-fluorescence quencher. Light phase images have been overlaid with the fluorescent far-red channel images. Images in far-red channel visualised on a scale of 512 low and 11520 high values in the Brightness Contrast tool of Fiji and pictured in magenta.

The results of the competition binding assay (**figure 6.25**) shows that there is a decrease in observed signal from secondary antibody bound to wtMOMP as the concentration of soluble Le^b

increases, pointing towards a competition for glycosylated wtMOMP by soluble and tissue bound Le^b carbohydrates.

6.3 Conclusion

Ex vivo work allowed us to determine how the wtMOMP may interact with gut tissue and which areas the interaction take place. wtMOMP had the most specific localisation on the surface epithelium of the corpus. Gastric tissue lacking Le^b expression did not bind wtMOMP to the same level of specificity. Furthermore, mutant MOMP did not bind Le^b expressing tissue at all. Soluble glycoproteins presenting the Le^b antigen inhibited binding of the glycosylated protein to the Le^b expressing stomach tissue. The next steps would be to determine extent to which novel compounds such as QPLEX may perform in inhibiting association of wtMOMP to BgAgs on the gut.

Chapter 7: **Methodology**

7.1 Purification of recombinant MOMP

MOMP protein was initially purified from recombinant *E. coli* because of the flexibility of protein purification from large scale liquid batches as well as the cost effectiveness of *E. coli* purification when compared to *Campylobacter* growth.

7.1.1 Codon optimisation of *momp* NCTC1168 for *E. coli*

To reflect the codon bias of the *E. coli* expression system, the *momp* gene was codon optimised. This synthetic DNA was obtained from Eurofins (**figure 7.1**).

A) MOMP NCTC11168 DNA sequence

GAAAACCTGTATTTTCAGGGCG:CCATGGCTCCACTTGAAGAAGCGATCAAAGATGTTGATGTATCAGGTGTATTAAGATACAGATACG
ATACAGGTAATTTTGATAAAAATTTCGTTAACAACTCAAATTTAAACAACAGCAAACAAGATCACAAATATAGAGCACAAGTTAACTTCA
GTGCTGCTATAGCTGATAACTTCAAAGCTTTTGTTCAATTTGACTATAATGCTGCTGATGGTGGTTATGGTGCTAATGGAATAAAAAATG
ATCAAAAAGGACTTTTTGTTCGTCAATTATACTTAACTTATACAAATGAAGATGTTGCTACAAGTGTAATCGCTGGTAAACAACAATTAA
ACCTTATCTGGACGGATAACGCTATTGATGGTTTAGTTGGCACAGGTGTTAAAGTAGTAAATAACAGCATCGATGGTTTAACTCTAGCT
GCTTTTGCTGTAGATAGCTTCATGGCTGCAGAGCAAGGTGCAGATTTATTAGAACATAGTAATATTTCAACAACATCAAATCAAGCTCCT
TTTAAAGTAGATTCAGTAGGAAATCTTTACGGTGCTGCTGCTGTAGGTTCTTATGATCTTGCTGGTGGACAATTCAACCCACAATTATGG
TTAGCTTATTGGGATCAAGTAGCATTCTTCTATGCTGTAGATGCAGCTTATAGTACAACTATCTTTGATGGAATCAACTGGACACTTGAA
GGTGCTTACTTAGGAAATAGCCTTGATAGCGAACTTGATGATAAAACACACGCTAATGGCAATTTATTTGCTTTAAAAGGTAGCATTGA
AGTAAATGGTTGGGATGCTAGCCTTGGTGGTTTATACTACGGTGATAAAGAAAAAGCTTCTACAGTTGTAATCGAAGATCAAGGTAATC
TTGGTTCTTTACTTGCAGGTGAGGAAATTTTCTATACTACTGGTTCAAGACTAAATGGTGATACTGGTAGAAATATCTTCGGTTATGTAA
CTGGTGGATATACTTTCAACGAAACAGTTCGCGTTGGTGCTGACTTCGTATATGGTGGAACAAAAACAGAAGCTGCTAATCATTTAGGT
GGTGGTAAAAAACTTGAAGCTGTTGCAAGAGTAGATTACAAATACTCTCCAAAACTTAACTTCTCAGCATTCTATTCTTATGTGAACCTA
GATCAAGGTGTAAACACTAATGAAAGTGCTGATCATAGCACTGTAAGACTTCAAGCTCTTTACAAATTCTAAA:AGCTT

B) Protein sequence

Met A P L E E A I K D V D V S G V L R Y R Y D T G N F D K N F V N N S N L N N S K Q D H K Y R A Q V N F S A A I A D N F K A F V Q
F D Y N A A D G G Y G A N G I K N D Q K G L F V R Q L Y L T Y T N E D V A T S V I A G K Q Q L N L I W T D N A I D G L V G T G V K
V V N N S I D G L T L A A F A V D S F Met A A E Q G A D L L E H S N I S T T S N Q A P F K V D S V G N L Y G A A A V G S Y D L A G
G Q F N P Q L W L A Y W D Q V A F F Y A V D A A Y S T T I F D G I N W T L E G A Y L G N S L D S E L D D K T H A N G N L F A L K G
S I E V N G W D A S L G G L Y Y G D K E K A S T V V I E D Q G N L G S L L A G E E I F Y T T G S R L N G D T G R N I F G Y V T G G Y T
F N E T V R V G A D F V Y G G T K T E A A N H L G G G K K L E A V A R V D Y K Y S P K L N F S A F Y S Y V N L D Q G V N T N E S A
D H S T V R L Q A L Y K F Stop

Figure 7.1: A. Codon optimised gene sequence of MOMP NCTC11168. NcoI and HindIII restriction enzyme starts are highlighted in green with start and stop sites highlighted in red. **B.** Translated amino acid sequence.

The codon optimised *momp* gene was cloned into the pTAMAHisTEV expression vector (**figure 7.2**). The vector was acquired from Ferrara *et al.* and was originally used during cloning of MOMP

85H. The vector includes the T7 promoter sequence and an ampicillin resistance gene. Once the gene is cloned into the vector, the encoded protein would include a TAMA signal peptide at the N terminus, followed by a histidine tag, then the tobacco etch virus (TEV) recognition sequence ENLYFQG, and lastly the protein of interest wtMOMP. His tag fusion via a 6-His-tag allows for purification of proteins by immobilized metal ion affinity chromatography (IMAC) which could be cleaved off by the TEV protease.

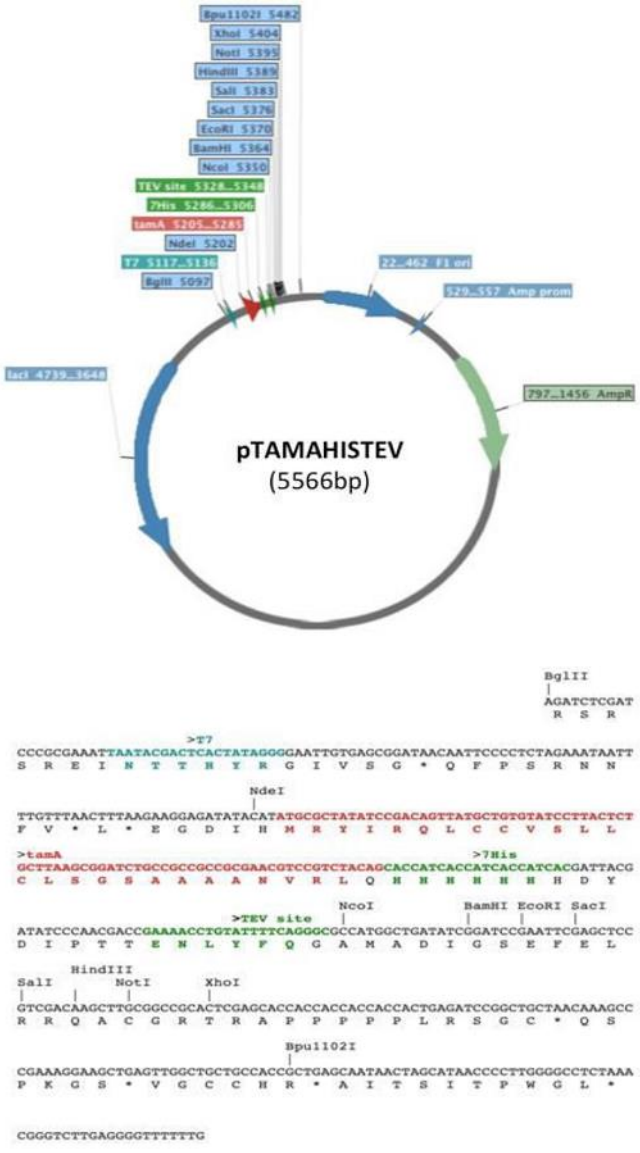


Figure 7.2: A plasmid map of the pTAMAHisTev plasmid (5566bp). Highlighted are the T7 promoter region (blue) the TAMA signal (red), the His tag (green) and the TEV cleavage site (green). *momp* was inserted via the NcoI and HindIII restriction enzyme sites.

7.1.2 Plasmid DNA recovery

Plasmid DNA was retrieved from a paper disc and recovered. The paper discs were collected in a sterilised 0.5 mL tube where the bottom of the tube was punctured. 30 µL of sterile distilled water was applied to the paper disc to wet for 1 min. The 0.5 mL tube was then transferred into a 1.5 mL eppendorf and spun for 1 min on a bench top centrifuge. The DNA collected in the 1.5 mL tube was transferred into a new sterilized tube. 5 µl of the washed DNA was transformed into *E. coli* strain DH5α.

7.1.3 Transformation of plasmid DNA into *E. coli* strain DH5a.

One tube of DH5α™ cells (Thermo Fisher) was thawed on ice. Cells were mixed gently with the pipette tip and 50 µL of cells aliquoted for each transformation into a 1.5 mL microcentrifuge tube. Any unused cells were refrozen in the dry ice/ethanol bath for 5 min before returning to the -80°C freezer. 5 µL of plasmid DNA was then added to the cells and mixed gently. For the pUC19 control (positive control), 2.5 µL (250 pg) of DNA was added to the cells and mixed gently. Tubes were incubated on ice for 30 min then heat shocked for 20 sec in a 42°C water bath without shaking and returned to ice. 950 µL of pre-warmed Super Optimal broth with Catabolite repression (SOC) medium was added to each tube. The tubes were incubated at 37°C for 1 hr at 225 rpm. Two different volumes (50 µL and 150 µL) were spread from each transformation onto pre-warmed ampicillin resistant selective plates, containing 100 µg/mL ampicillin, to ensure that at least one plate would have well-spaced colonies. For the pUC19 control, 100 µL was plated on an LB plate containing 100 µg/mL ampicillin and incubated overnight at 37°C. After overnight incubation, 4 separate colonies were taken from one plate and inoculated into 4 tubes, each containing 3 mL LB with ampicillin. They were incubated overnight at 37°C in shaker at 200 rpm.

7.1.4 Recovery of plasmid DNA from bacterial culture

An alkaline extraction procedure using the Gene Elute Plasmid Miniprep Kit (Sigma Aldrich) was used to recover plasmid DNA from overnight *E. coli* cultures. The 3 mL overnight cultures were pelleted for 1 min at 12, 000 x g and the supernatant discarded. Cells were resuspended in 200 µL Resuspension Solution and vortexed. 200 µL of Lysis Solution was added before inverting gently to mix and allowing to clear for 5 min. 350 µL of Neutralization Solution was added and the mix was inverted 4–6 times. The debris was pelleted for 10 min at 16000 x g. Meanwhile, the binding column was prepared by adding 500 µL Column Preparation Solution to binding column in a collection tube and spinning at ≥12,000 x g, 1 min to wash. The flow through was discarded and the cleared lysate was transferred to the binding column to be spun for 1 min. Again, the flow-through was discarded. To remove contaminants 750 µL Wash Solution was added to the column and spun for 1 min before discarding flow-through and spinning again for 1 min to dry the column. Elution of purified plasmid DNA was done by transferring column to a new collection tube, adding 100 µL Elution Solution and spinning for 1 min. 2 µL of each sample was analysed on the NanoDrop to read the DNA concentration.

7.1.5 Cloning *momp* gene into pTAMAHisTEV

The *momp* gene was cloned into the plasmid pTAMAHisTEV by restriction enzyme digest followed by ligation. 25ng of plasmid vector DNA was needed for 75 ng of insert DNA. For plasmid digestion, 5 µL of 10 X CutSmart buffer (NEB), 1 µL NcoI, 1µl HindIII, 4 µg pTAMAHisTEV plasmid and H_2O were mixed in a reaction volume of 50 µL. For digestion of *momp*, the restriction reaction of 15 µL contained 11.5 µl DNA, 5 µL of 10 X CutSmart buffer, 1 µL NcoI and 1 µL HindIII. Both eppendorfs were incubated at 37°C overnight. After spinning the Eppendorfs, the tube containing the *momp* digest was heated to 80°C for 20 min to stop the enzyme reaction, whilst the Monarch® PCR & DNA Cleanup Kit was used to clean up the plasmid digests. PCR clean-up is a routine method to

purify DNA using a simple and rapid PCR clean-up method that efficiently removes short primers, unincorporated dNTPs, enzymes, short-failed PCR products, and salts from PCR reactions. Finally, the cleaned-up DNA was eluted into 40 µL of H_2O.

T4 DNA Ligase (Promega) catalyzes the joining of two strands of DNA between the 5'-phosphate and the 3'-hydroxyl groups of adjacent nucleotides in either a cohesive-ended or blunt-ended configuration. For ligation, digested pTamAHisTEV plasmid DNA 100ng, *momp* insert DNA 17ng, 1 µL Ligase 10X Buffer, 1 µL T4 DNA Ligase was added to Nuclease-Free Water to final volume of 10 µL in a sterile microcentrifuge tube. The reaction was incubated at 4°C overnight. A control was set up that contained only 4 µl of the plasmid vector and no insert made up to 10 µL.

7.1.6 Transformation of pTamAHisTEVmomp into storage *E. coli* TOP10F One Shot

2 µL of ligation reaction were added into a vial of One Shot® Chemically Competent *E. coli* and mixed gently then incubated on ice for 5–30 min. A transformation control was introduced where the 3rd sample was *E. coli* only. The cells were heat shocked for 30 secs at 42°C without shaking and the tubes were immediately transferred to ice. 250 µL of room temperature SOC medium was added to the tube. The tube was shaken horizontally (200 rpm) at 37°C for 1 hr. 10–50 µL from each transformation was spread on prewarmed selective plates and incubated overnight at 37°C. To ensure even spreading of small volumes, 20 µL of SOC medium was added. Two different volumes were plated to ensure that at least one plate would have well-spaced colonies. 12 colonies were picked and patched onto a fresh selective plate, then incubated overnight. The next day the 12 patches were inoculated into 5 mL of LB and 100 µg/mL ampicillin per tube and shaken at 200 rpm at 37°C overnight. The 12 overnight cultures were then miniprepped to purify the plasmid DNA. After purifying plasmids via miniprep, the 12 plasmid samples were digested with

HindIII and NcoI in a digestion mix of 7 µL DNA (200ng DNA), 1 µL HindII, 1 µL NcoI and 1 µL Smart Buffer.

7.1.7 Agarose gel electrophoresis

The digests were run on a 0.8% w/v agarose gel alongside a 1kb ladder (NEB). 0.8 g of agarose was mixed with 100 mL 1xTAE in a microwavable flask and microwaved in pulses until the agarose was completely dissolved. The agarose solution was allowed to cool down to around 50°C then poured into the gel tray with the well comb in place and allowed to set until completely solidified. 2 µL 6 X loading buffer was added to 10 µL each of the digest samples. The loading buffer contains a visible dye and high percentage of glycerol. 1 µL of SYBR green solution (25 µL DMSO and 1 µL of 10000 x dsGreen) solution was also added to each sample to be separated on an agarose or polyacrylamide gel 15 min before running. The gel box was filled with 1x TAE until the gel is covered.The gel was run at 120 V until the dye line was approximately 75-80% of the way down the gel. A typical run time is about 1-1.5 hr, depending on the gel concentration and voltage. At the end of the run, the gel was placed in water and destain for 5 min. The Biorad UV transilluminator was used to visualise the DNA fragments.

7.1.8 Primers

The commercial T7 forward primer was used to sequence the DNA:

5’ -TAATACGACTCACTATAGGG - 3’

This primer attaches at the T7 site within the plasmid (highlighted in **figure 7.2**). Additional primers used to verify the sequence of the *momp* gene insert were:

MOMPFOR719: 5’-CCGTTCAAAGTAGATAGTGTCG-3’ and

pTAMAHISTEVreverse: 5’-TATTGCTCAGCGGTGGCAGC-‘3

Concentration of primer used for DNA sequencing was 10pmol/µL.

7.1.9 Transformation of *E. coli* expression cells by electroporation

The plasmid vector was transformed into Lemo21 expression cells by electroporation. Electroporation allows transformation of cells by temporarily creating holes in the cell membrane which allow the passage of charged molecules such as plasmid DNA. The physical mechanism involves multiple phases. Phase 1 is a <1 ms 300-400 mV pulse which creates a voltage across the membrane by inducing an ion migration from the surrounding solution. When the critical field is reached a pre-pore forms (a small conductive hydrophobic defect). Phase 2 involves rearrangements at the pore edge which lead to a hydrophilic interface. Using a sterile technique throughout, the competent cells, plasmid pTamAHisTev-based plasmids and electrocuvettes were placed on ice to defrost/chill. The electroporation machine was set to Ec2. When defrosted, 1 µL of the plasmid was transferred into 80 µL of competent cells and gently mixed up and down then left on ice for 2 min before transferring to the cuvette and inserted into the machine. The cells were pulsed and 1 mL LB was instantly added and mixed before transferring to an Eppendorf. The tube was incubated for 1 hr at 37°C and an agar plate was inoculated with 250 µL of the broth (containing 100 µg/mL ampicillin). After overnight culture, glycerol stocks were made.

7.1.10 Small scale expression test from *E. coli* Lemo21 DE3 cells

This protocol is designed to optimize expression and solubility of MOMP using *E. coli* host Lemo21 DE3 and the expression vector which has an IPTG-inducible T7 promoter.

Small-scale expression cultures are used to identify the optimum expression parameters for a given target. Thorough analysis of the total cell content and soluble fraction is used to screen out failed targets and those unlikely to succeed in large-scale purification cultures. LB was autoclaved

then allowed to cool down. 12 mL of overnight culture of Lemo + pJO1 in LB was inoculated into a 600 mL flask of LB (1 in 50 inoculation) followed by addition of 100 µg/mL ampicillin and a drop of defoamer. When the OD reached 0.3, 0.8 and 2 the culture was aliquoted into 2 x 200 mL. At this point one flask was induced with 1 mM Isopropyl β- D-1-thiogalactopyranoside (IPTG). The other flask was left uninduced. The 200 mL flasks were then incubated at temperature of 25°C and shaken at 200 rpm. At each time point of 1, 2, 3, 4, 6, 8 hr and overnight (16 hr), 10 mL of culture was taken and stored in -20°C. The bacterial growth from each time point was centrifuged at 3000 x g for 15 min to create a cell pellet and the supernatant was discarded. The cell pellets were then resuspended in 1 mL lysis buffer (20 mM Tris pH 8, 20 mM NaCl, 10% w/v glycerol) and sonicated with a microtip at 7 kHz with 5 rounds of 30 sec on and 30 sec off. The whole cell lysate was spun down at 16,000 x g for 30 min to separate the soluble (supernatant) and insoluble (pellet) fraction. The insoluble fraction was treated with 1 mL of 8 M urea, which was run on a gel alongside the soluble fraction.

7.1.11 SDS PAGE Electrophoresis

To measure the molecular weights of different proteins in a sample, an SDS PAGE gel was used. This required running buffer to fill the tank (25 mM Tris base, 190 mM glycine, 0.1% (w/v) SDS), 10% resolving buffer (30% w/v acrylamide, 1.5M Tris HCl pH 8.3, 10% w/v SDS, H_20, with 20 µL 10% v/v APS and 5 mL TEMED per gel), stacker at 4% v/v acrylamide which serves to concentrate samples before resolving (30% w/v acrylamide, 0.5M Tris HCl pH 6.8, 10% SDS, H_2O with 20 µL 10 % APS and 5 mL TEMED per gel) and 1X loading buffer which denatures the protein (4% w/v SDS, 10% v/v 2-mercaptoethanol, 20% v/v glycerol, 0.004% w/v bromophenol blue, 0.125 M Tris HCl). A lane of protein markers was also run alongside the samples (NEB Color Prestained Protein Standard, Broad Range (11–245 kDa)). Gels were run at 180 V for 5 min then stepped up to 200 V for 35 min in a Mini Protean Tetra Cell (Bio-Rad). To prepare the gel for documentation in the gel

imager, the gel was washed for 5 min in dH_2O on a low heat in the microwave. This would remove the bromophenol blue dye front from the gel. To stain protein bands, the gel was incubated with Gel Code™ Blue Safe Protein Stain Coomassie dye in the microwave for 10 min on medium heat. To visualize the bands, the background stain was removed by washing the gel overnight in dH_2O before documenting in the gel imager.

7.1.12 Large scale growth and protein purification of recombinant protein

Following the results of the small-scale expression and solubility test, a large-scale growth was set up. 1.96 L of LB was inoculated with 40 mL of overnight culture and allowed to grow till the OD reached 0.6 at 37°C. Time points were taken at 1, 2, 3 and 16 hr. Following the same protocol as previously, 2 L of culture was separated into 6 tubes of around 330 mL each and spun down at 6,000 x g to pellet. The supernatant was discarded and the pellets stored in -20°C.

One pellet of 3 g was resuspended in lysis buffer of 20 mM Tris HCl (pH8), 150 mM NaCl, 5% v/v Elugent (VWR) and 200 µM phenylmethylsulfonyl fluoride (PMSF) serine protease inhibitor. After resuspension, the cells were sonicated at 25 kHz for 10 rounds of 30 sec on and 30 sec off on ice. This was followed by ultracentrifugation at 100,000 x g (Beckman coulter Optima l-80 ultracentrifuge) for 1 hr to remove insoluble material (the pellet). The soluble fraction was run on a His trap column for purification of His-tagged recombinant MOMP proteins by immobilized metal affinity chromatography (IMAC) using precharged Ni Sepharose™ High Performance. The column was equilibrated with buffer A (2 mM imidazole, 20 mM Tris pH8, 150 mm NaCl, 5% v/v Elugent, 200 µM PMSF). 30 mL of the sample was loaded onto a HisTrap FF crude 5 mL column. To accommodate large sample volumes and to reduce manual intervention, the sample was loaded through the system pump at a rate of 3 mL/min. The column was washed with 2 column volumes (CV) of binding buffer. Buffer A was replaced by wash buffer (20 mM Tris pH 8, 30 mM imidazole,

150 mM NaCl and 0.25% v/v Elugent) and the column was washed with 2 CV of wash buffer to remove non- specifically bound protein. This was followed by elution with 6 CV of 100 % Buffer B (250 mM imidazole,150 mM NaCl, 20 mM Tris pH8) to ensure removal of all tightly bound proteins. 3 mL fractions were collected using Frac30 fraction collector. The peak fractions were pooled and stored in - 80^{o}C. Aliquots were collected to run on a gel. MOMP was further purified by size exclusion chromatography gel filtration using the 16/60 Superdex 200pg (GE Healthcare) on an ÄKTA prime purifier. Buffer exchange and desalting were also carried out via the column. The rate of flow was 0.5 mL/min with pressure of 0.6 mPa. The gel filtration buffer consisted of 10 mM Tris HCl pH8, 150 mM NaCl and 0.45% w/v C8E4.

7.2 Purification of MOMP from native *C. jejuni*

7.2.1 Strain confirmation

C. jejuni strain NCTC11168[213] was isolated from human faeces and grows optimally on Columbia blood agar, for 48 hr, at 37°C, in microaerophilic conditions. The bacteria were obtained from frozen lab stocks. Both the wild type and mutated version of this strain were provided by Dr. Jafar Mahdavi.

Three variants of 11168 were acquired; wt11168, 11168$^{T/G268}$ and 11168$^{\Delta PSED}$. Briefly, before proceeding with *Campylobacter* growth, a simple test was performed to confirm mutant and wild type stocks. MOMP$^{T/G268}$ and MOMP$^{\Delta PSED}$ contain a kanamycin resistant cassette that was inserted during mutation, so their glycerol stocks were streaked onto BHI agar plates with 50 µg/mL of kanamycin to test for resistance (**figure 7.3**). Furthermore, PCR and gene sequencing of the *momp* and/or *pseD* loci in each strain was performed to confirm their identity (**figure 7.4 and 7.5**).

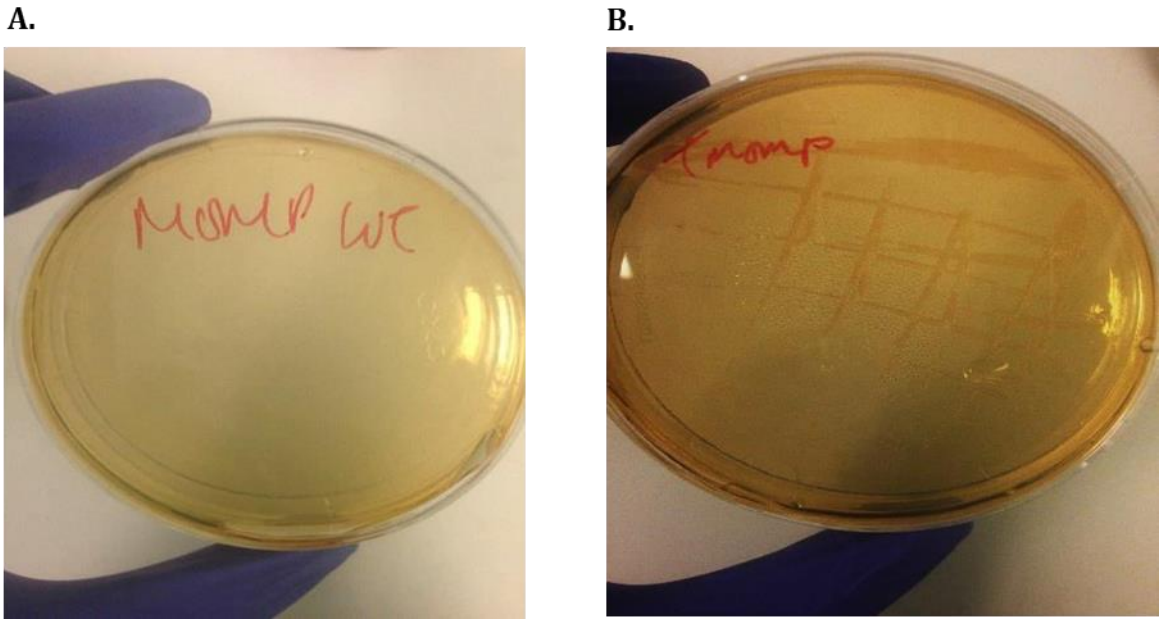


Figure 7.3: Glycerol stocks of the *C. jejuni* NCTC11168 wtMOMP, MOMP$^{T/G268}$ strains were streaked onto BHI agar plates with 50 µg/mL of kanamycin. **A.** No wtMOMP colonies grew on this plate. **B.** Colonies grew for the MOMP$^{T/G268}$ strain confirming kanamycin resistance.

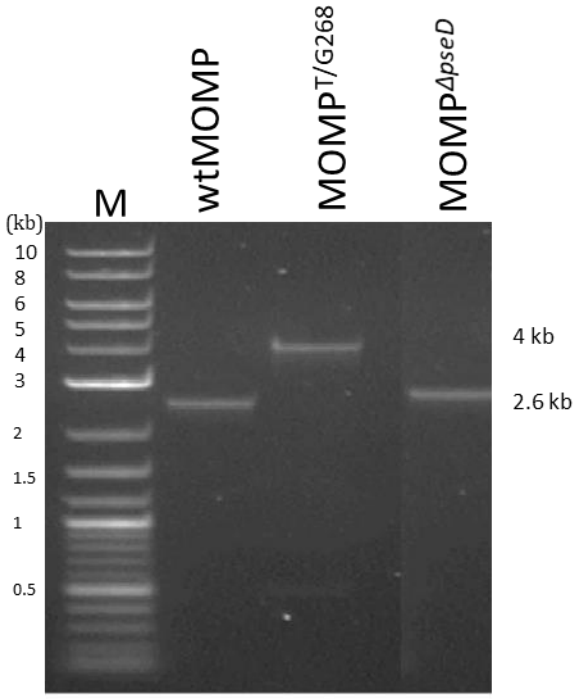


Figure 7.4: The PCR products of the *momp* gene in *C. jejuni* NCTC11168 wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ strains. MW markers are shown in lane M. MOMPF2 and MOMPR2 primers bind upstream and downstream of *momp* gene, respectively. In the MOMP$^{T/G268}$ strain, a kanamycin resistance cassette was inserted between the *momp* and *dnaJ* genes resulting in a PCR product of ~ 4 kb compared to ~ 2.6 kb for the PCR products from the wtMOMP and MOMP$^{\Delta pseD}$ strains. NEB 1 kb DNA ladder is shown in the lane marked M.

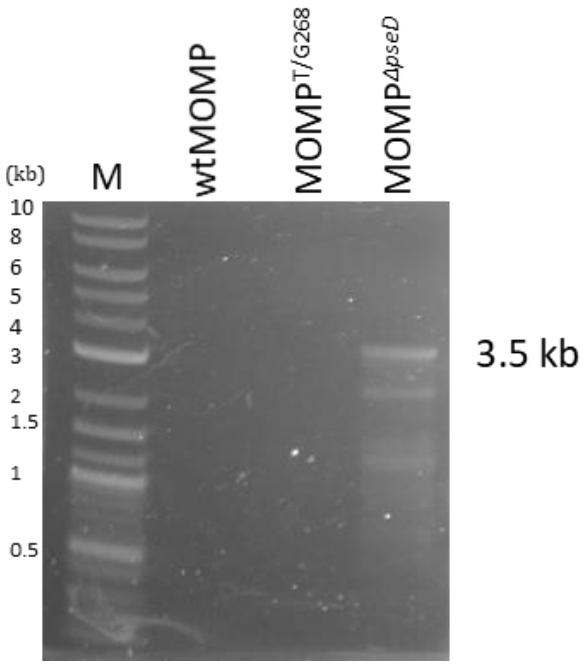


Figure 7.5: PCR products of the *pseD* gene in the *C. jejuni* NCTC11168 wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ strains using primers Cj1333_F and Cj1333_R2. The PCR product from the MOMP$^{\Delta pseD}$ strain is a 3.5 kb band because of the deletion of the 1.9 kb *pseD* gene and replacement with a 1.5kb Kan cassette. No PCR product of *pseD* gene was obtained with the wtMOMP and MOMP$^{T/G268}$ strains because a longer extension time was required to obtain the 3.9 kb PCR product. Lane M shows molecular weight marker NEB 1 kb DNA ladder.

7.2.2 Cell collection and purification of native MOMP

A chromogenic plate (Brilliance CampyCount AGAR), an Oxoid Campylobacter selective blood free agar (CCDA) plate and a Campylobacter blood agar plate were each inoculated with 10 µL of the frozen glycerol stock of *C. jejuni* NCTC 11168 in a cross-streak pattern using a sterile loop. Cells were grown under microaerophilic conditions for 48 hr at 42° C. Successful campylobacter isolation wass ensured by the appearance of dark red spots on the chromogenic plates. Once single colonies were seen, 15 new plates were streaked with one single colony each, this time using a cotton swab to create a bacterial lawn and increased coverage of the plates. Again, the plates were incubated under microaerobic conditions at 42° C for 48 hr. To grow enough bacterial cells for downstream applications, the 15 plates were expanded to 450 plates using the cotton swab and lawn technique. In total, wild type and mutant strains were grown on 450 plates each.

Bacterial cells from each plate were harvested and rinsed with 5 ml of 10 mM Tris, 1 mM EDTA (pH 7.2) for 15 min at room temperature with gentle agitation. The following steps were carried out at 4°C. After centrifugation at 10,000 x g for 30 min, cell pellets were washed in 200 mM glycine-HCl (pH 2.2) to eliminate proteases and associated outer membrane proteins and agitated for 15 min, followed by another round of centrifugation at 10,000 x g. The bacterial pellet was washed twice successively in 100 mM Tris (pH 7.2) to titrate excess HCl. Then the supernatant was discarded and cells were resuspended in 10 mM Tris HCl (pH 7.2) and was sonicated in the same buffer. The sonication was carried out at 30 kHz with 10 rounds of 30 sec on and off on ice. After sonication, cell fragments were removed by centrifugation (6,000 x g for 30 min) and the total membrane fraction (the supernatant) was recovered by ultracentrifugation at 100,000 x g for 1 hr at 4°C. The pellet was homogenized in 10 mM Tris (pH 7.4) and 0.1% (w/v) of SDS (Sigma) to solubilize the inner membrane and was left rocking for 30 min. The outer membrane was recovered by ultracentrifugation at 100,000 g for 1 hr. The supernatant containing the inner membrane protein fraction was discarded, and the pellet was homogenized with 20 mM sodium phosphate buffer (pH 7.4) and 1% (v/v) of poly(ethylene glycol) octyl ether (Octyl-POE) (Bachem AG) and was left rocking at 4 °C for 30 min. From this point on Octyl-POE was used as the detergent to keep MOMP in solution due to its hydrophobic surface. Ultracentrifugation at 100,000 g for 1 hr lead to the recovery of solubilised MOMP proteins in association with POE micelles. The protein was purified further by anion exchange chromatography, a form of ion exchange chromatography (IEX), which is used to separate molecules based on their net surface charge. Anion exchange chromatography, more specifically, uses a positively charged ion exchange resin with an affinity for molecules having net negative surface charges, such as MOMP. Protein was loaded onto a MonoQ HR anion exchange column (GE Healthcare) and was equilibrated with 5 CV of buffer A (30 mM Na_2HPO_4, 10 mM NaCl, and 0.6 % v/v Octyl-POE). The bound proteins were eluted stepwise with 5, 12, 20, 70,

and 100% of buffer A supplemented with 1 M NaCl. The chromatography was performed on ÄKTA Explorer 10 system (GE Healthcare). Each fraction was examined with SDS PAGE and Western blot with specific antibodies. Eluted fractions containing MOMP were collected and concentrated to 5 mL and injected onto a Superdex 200 16/60 GL (GE Healthcare) gel filtration column equilibrated with 2 CV of 10 mM Tris (pH 8.0), 150 mM NaCl, and 0.45% (w/v) C8E4. Fractions containing MOMP were combined and concentrated with a spin concentrator. A nanodrop was used to measure protein concentration. Yield per 150 plates was 6.8 mg of MOMP protein.

7.3 Protein confirmation

7.3.1 Western blot

Western blot was used to confirm that antibodies (donated by Dr. Jafar Mahdavi) raised against previous batches of MOMP and used in downstream experiments such as ELISA and the immunohistochemistry studies, would have affinity for all purified variants of MOMP (wtMOMP, $MOMP^{T/G268}$ and $MOMP^{\Delta pseD}$). Western blot involved transfer of the protein bands from the polyacrylamide gels to nitrocellulose membrane. Western blots were performed for wtMOMP, $MOMP^{T/G}$, $MOMP^{\Delta pseD}$ and wtMOMP conjugated with far-red and fluorescein dyes.

Two methods of gel transfer were used; semi-dry or wet transfer.

7.3.2 Semi-dry transfer

The nitrocellulose and filter paper were prepared in advance. After SDS PAGE gels had been washed of bromophenol blue, the gel was equilibrated in a small container of 1X Semi-dry transfer buffer (48 mM Tris base, 39 mM glycine, 20% v/v methanol) for 15 min. A piece of filter paper was

completely saturated by soaking in transfer buffer and placed onto the platinum anode of the Trans-Blot® SD semi-dry cell. A pipet was rolled over the surface of the paper to exclude all air bubbles. The nitrocellulose membrane was then placed on top of the filter paper, followed by the equilibrated gel, making sure to align the stack. Another piece of filter paper was soaked in semi dry transfer buffer and placed on top of the gel, carefully removing air bubbles from between the gel and filter paper. The cathode plate was placed on top the stack and the latches engaged with the guide posts without disturbing the gel/nitrocellulose stack. The power settings were 15 V for 30 min. Once complete, the membrane was removed with forceps and rinsed in deionised water.

7.3.3 Wet transfer

One side of the nitrocellulose membrane was marked with a pencil and soaked in methanol for a few min and rinsed with distilled water. The membrane, pads, filter papers (6 pieces) and transfer foam were equilibrated in 1X wet transfer buffer (25 mM Tris, 192 mM glycine, 20% v/v methanol stored in 4°C) for 5 min. The transfer sandwich was assembled in the following order: black frame (negative electrode) >> foam >> 3 pieces of filter papers >> SDS PAGE gel >> membrane >> 3 pieces of filter papers >> foam >> red frame (positive electrode). The marked side of membrane was placed facing the gel then the sandwich was transferred to the tank. The gel was transferred at 80 V for 90 min in the cold room.

7.3.4 Immunoblotting

After transfer of proteins to the membrane, the membrane was blocked for 1 hr at room temperature using blocking buffer (5% w/v milk in 20 mL PBS Tween). For recombinant MOMP protein, the membrane was then incubated overnight at 4°C in 1:2,000 dilution of primary antibody targeted against the 6-His-tag (penta-His mouse antibody) in blocking buffer. For native protein, the membrane was incubated overnight at 4°C in 1:2,000 dilution of primary antibody

targeted against MOMP (rabbit antibody) in blocking buffer. The membrane was washed in three washes of PBST, for 5 min each. The membrane was then incubated with 1:3,000 dilution of conjugated secondary antibody (anti mouse or anti rabbit alkaline phosphatase conjugated antibody) in blocking buffer at room temperature for 1 hr. The membranes were washed in three washes of PBST, for 5 min each. For signal development, the bands were visualised by adding alkaline phosphatase substrate (1-Step NBT/BCIP) to the blots which formed a purple product when cleaved by the conjugated enzyme. The gels were scanned in the BioRad Gel imager system using upper white light.

7.3.5 Circular Dichroism Spectropolarimeter

CD experiments were performed with a Chirascan Plus (Applied Photophysics). Purified wtMOMP and $MOMP^{T/G268}$ were each diluted to 5 µM in 20 mM NaP, pH 7.4. 600 µL of the sample was loaded into a 1 mm path length cuvette. Spectra were collected from 180 – 250 nm with a resolution of 1 nm by on continuous mode for 50 nm/min. The data was saved as a txt file and imported into Excel for further analysis.

7.3.6 Maldi TOF mass spec

For analysis by MALDI, the sample was first deposited on a metal target plate, along with a matrix compound, which assists in ionisation. A polished steel plate was used as the metal target plate for high MW polymers and peptides/proteins. Sample (1 mg/mL wtMOMP) and matrix solution (20 mg Sinapinic acid in 1 mL solution of 30 % v/v acetonitrile, 70% v/v water and 0.1 % v/v trifluoroacetic acid (TFA)) were made up in Eppendorf tubes, the open tops of which formed a convenient place to mix both together in preparation for spotting. The protein stock solution was

diluted 1:5 in matrix solution. 1 µL of the sample-matrix mixture was applied to the target and allowed to dry naturally until a nice distribution of needle-like white crystals appeared on the plate. The ultraFlex III MALDI-ToFMS was used to analyse the sample. flexControl is the control and acquisition software for the ultraFlex and was used to acquire data in linear mode (lower resolution than with reflecton mode but upper mass is higher at 100 kDa) and positive ion detection. The laser was positioned onto the required spot on the plate by clicking on the required position in the flexControl window and then fine tuning the position by clicking within the camera view. The optimum position was determined by trial and error and the process repeated until the matrix and sample ions were seen. The laser beam was repositioned until a sufficiently intense spectrum had been built up. This process was repeated for $MOMP^{T/G268}$ and $MOMP^{\Delta pseD}$.

7.4 Binding of wtMOMP to Le^b, Le^a, Type 1, Type 2

The minimum concentration at which binding of various blood group sugars to wtMOMP occurred was determined and applied to $MOMP^{T/G268}$ and $MOMP^{\Delta pseD}$ in order to establish the difference in binding of sugars to glycosylated and non-glycosylated protein. Initial studies used concentrations of wtMOMP from 1 pM to 100 pM alongside concentrations of the BgAgs from 1 µM to 100 µM. Soluble BgAgs were bought from IsoSep and required concentrations were diluted in 100 µL PBS with 0.5% w/v BSA and added to each well before incubating on the shaker at 4°C overnight. Following incubation, the plate was washed twice with 300 µL PBST. Required concentrations of MOMP protein diluted in 0.5% w/v BSA in PBS was added to the wells and left to incubate for 2 hrs at room temperature. Following incubation, the plate was washed twice with 300 µL PBST. Anti-MOMP antibody raised in rabbit was diluted 1 in 1,000 in 0.5% w/v BSA in PBS before adding 100 µl to each well and incubating overnight at 4 °C. After washing with PBST, an alkaline phosphatase conjugated secondary antibody, anti-rabbit antibody raised in goat, was diluted 1 in 5,000 in 0.5%

w/v BSA in PBS and added to the wells. The plate was incubated at room temperature for 2 hr. Lastly, the wells were washed with PBST and followed by addition of 100 µL of 1 mg/mL p-Nitrophenyl phosphate (pNPP) substrate in glycine buffer. pNPP produces a yellow water-soluble reaction product that absorbs light at 405 nm and the plate absorbance was measured for 16 hr in the TECAN. The same experimental design was used to measure binding of BgAgs to MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$.

	[wtMOMP]											
	1	2	3	4	5	6	7	8	9	10	11	12
A	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM
B	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM
C	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM
D	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM
E	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM
F	1 pM	10 pM	100 pM	1001 pM	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM
G	1 pM	10 pM	100 pM	1002 pM	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM
H	1 pM	10 pM	100 pM	1003 pM	1 pM	10 pM	100 pM	1000 pM	1 pM	10 pM	100 pM	1000 pM
	[BgAgs]											
	1 uM	1nm	10nm									
	10 uM	1um	10um									

Table 7.1: Le^b, Le^a, Type 1, Type 2 BgAgs and BSA were adsorbed to the surface of amino-linked ELISA plates. The concentration of protein and BgAg to be added to each well of the 96 well plate is shown. The plate layout was repeated for each of the BgAgs.

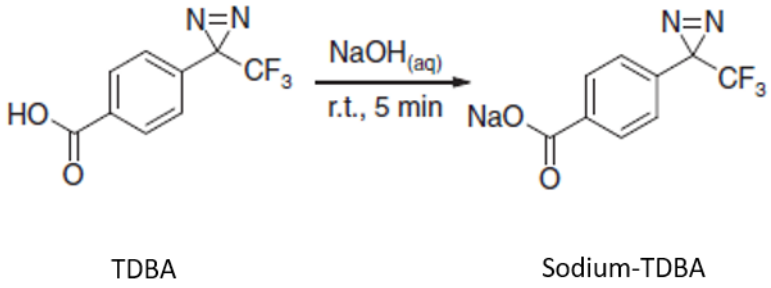
7.5 Inhibition of wtMOMP/Le^b with QPLEX

Once it was established that BgAgs favoured binding to wtMOMP over MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$, a small compound, QPLEX, was introduced into the experiment to observe whether binding could be inhibited and at which concentration the inhibition occurred. Because 1 pM of the wtMOMP protein could bind well to 1 µM of Le^b, the same concentrations were used in inhibitory studies where QPLEX was titrated into the protein-sugar solution at concentration of 1 nM to 10 µM with a control sample of 0 QPLEX. wtMOMP, Le^b and QPLEX were diluted in 100 µL PBS with 0.5% w/v BSA and added to each well before incubating on the shaker at 4°C overnight. The plate was then treated as previously described when determining binding of wtMOMP to Le^b, Le^a, Type 1, Type 2.

7.6 Carbene footprinting

To map the site of interaction of the ligand on the protein, carbene, a diazirine-based footprinting agent, was used to rapidly label the surface of the protein to show difference in areas of masking in the absence and presence of the QPLEX ligand. This footprinting methods introduces an irreversible covalent modification, unlike HDX labelling, and hence structural information is retained well throughout proteolytic and chromatographic steps. The extent of peptide modification is expressed as fractional modification (*fm*). Differential experiments compare the fractional modification of peptides between treatments to identify sites of interaction on a protein. Subpeptide-level modification can then be determined with MS/MS experiments to pinpoint key residues.

In an Eppendorf, a solution of sodium hydroxide (16 mg, 0.39 mmol) in H_2O (1 mL) was added to 4-(3-(trifluoromethyl)-3H-diazirin-3-yl) benzoic acid (TDBA) (100 mg, 0.43 mmol), and the Eppendorf inverted a few times to mix the solution and form a sodium salt of TDBA.



TDBA Sodium-TDBA

Figure 7.6: Synthesis of carbene probe, sodium 4-(3-(trifluoromethyl)-3H-diazirin-3-yl) benzoate (Sodium-TDBA). The addition of sodium hydroxide to TDBA produces sodium-TDBA, a carbene that is the precursor of its reactive state. The carbene contains an aromatic ring to provide a non-polar surface for interaction with hydrophobic patches on the protein, and an ionic group both to associate with polar amino acid side chains and to improve water solubility of the probe. The CF3 group has the benefit of reducing side reactions associated with the diazoisomer.

After synthesis of the sodium-TDBA carbene, it was ready for reacting with the protein by flash freezing and irradiation at 394 nm. For carbene labelling in the absence of the QPLEX ligand, an aqueous solution of wtMOMP (10 µM) in 1% v/v octyl-POE and PBS was mixed with an equal volume of a solution of aryldiazirine (100 mM) in 20 mM Tris/150 mM NaCl, pH 7.4, to give a final probe concentration of 50 mM and protein concentration of 5 µM.

For carbene labelling in the presence of the QPLEX ligand, an aqueous solution of wtMOMP (10 µM) in 1% v/v octyl-POE in PBS and 10 µM QPLEX in PBS was mixed with an equal volume of a solution of aryldiazirine (100 mM) in 20 mM Tris/150 mM NaCl, pH 7.4, to give a final probe concentration of 50 mM, protein concentration of 5 µM and QPLEX concentration of 5 µM in 5µL solution. For carbene labelling in the presence of the Le^b ligand, an aqueous solution of wtMOMP (10 µM) in 1% octyl-POE in PBS and 70 µM QPLEX in PBS was mixed with an equal volume of a solution of aryldiazirine (100 mM) in 20 mM Tris/150 mM NaCl, pH 7.4, to give a final probe concentration of 50 mM, protein concentration of 5 µM and Le^b concentration of 35 µM.

5 µL aliquots of the sample were placed in tapered X100 crystal clear vials (ThermoFisher Scientific, Paisley, UK) and flash-frozen with liquid nitrogen. Each sample was irradiated for 16 sec using a Spectra Physics Explorer 349 laser (an actively Q-switched Nd:YLF laser operating at 349 nm, with a repetition frequency of 1000 Hz, and a pulse energy of 125 µJ, Newport, Didcot, UK). The laser beam was directed into the open top of the sample vial using a small 45° mirror. The same laser conditions were used to irradiate a solution of MOMP in presence and absence of 1 % v/v OG using 5 mM, 10 mM, 25 mM or 50 mM concentrations of diazirine. All labelling experiments were repeated in triplicate.

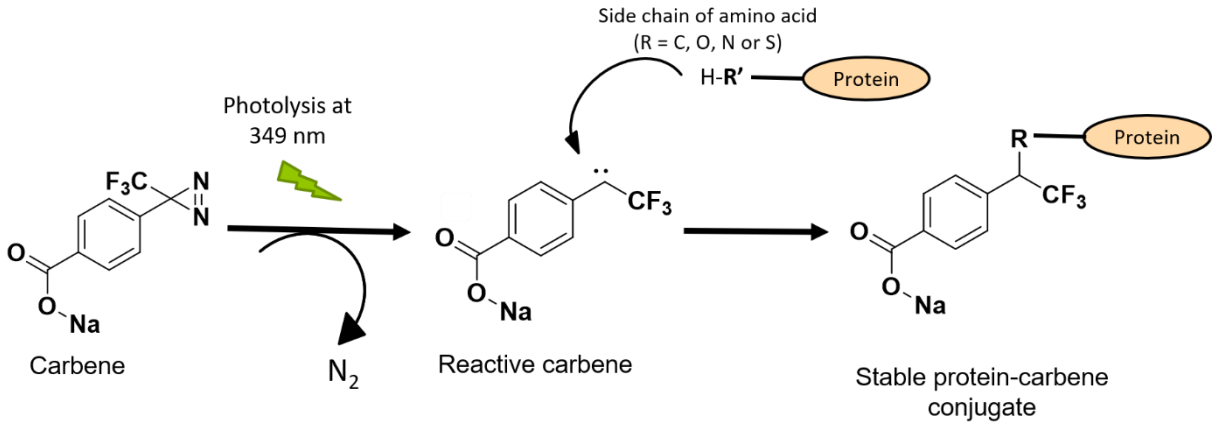


Figure 7.7: Irradiation (photolysis at near-UV laser at 349 nM) and flash freezing of the carbene has the dual benefit of halting enzyme turn-over and reducing diffusion of the probe. The removal of the azo group by UV leaves a lone pair that forms a reactive carbene with the potential to rapidly insert into any R–H bond (where R can be C, O, N, or S) within the protein's amino acid side chains.

The carbene labelled protein samples were purified and visualized by SDS PAGE. Bands corresponding to the protein were excised from the gel and followed by several rounds of destaining by ammonium bicarbonate and dehydration by acetonitrile until 1 mm^3 dehydrated white pieces of agarose gel saturated with protein remained.

The protein was incubated overnight for 16 hr by covering the gel slices with a solution of the enzyme in 50mM ammonium bicarbonate and final concentration of 20 mg/mL of trypsin or 2 mg/mL chymotrypsin. After overnight incubation, the reaction was stopped with 1% v/v formic acid and the samples were spun down at 3,000 xg for 3 min to separate gel pieces from the solution. The supernatant was decanted into new LC vials for analysis by nanoHPLC-MS.

7.6.1 LC-MS and -MS/MS analysis of MOMP peptides

Digests were analysed on a Dionex U3000 nanoLC coupled to a ThermoFisher LTQ FT Ultra mass spectrometer. Samples (2 μL) were injected in load-trapping mode. Peptides were eluted using a 30 min linear gradient of mobile phase B from 0% to 55% followed by 5 min at 90% B and 15 min column re-equilibration. The LTQ FT Ultra mass spectrometer was equipped with a

nanoelectrospray (nESI) source through which a 1.7 kV voltage was applied to the Picotip emitter. The inlet capillary of the mass spectrometer was held at 275 °C with a tube lens value of 145 V. More information on instrument set up has been published by the Oldham group [193].

For labelled peptides, in order to quantify the modification of the peptides by carbene, samples were separated and data acquired in full scan mode followed by data analysis with PepFoot software [214].

For non-labelled protein, nanoHPLC-MS was operated in data directed analysis (DDA) mode, where a low energy survey is used to identify precursors of interest by intensity, charge state and isotope pattern. When precursors of interest have been identified, MSMS is performed on these peptides to obtain fragmentation data. Precursor peptides that have no MSMS data will be targeted in another round of experiments in targeted DDA. The data were submitted to Bioworks software (Thermo Fisher Scientific) and searched against a custom database including the wtMOMP and MOMP$^{T/G268}$ sequence.

7.7 Protein-ligand docking

To be able to characterise the behaviour of small molecules in the binding site of MOMP and to elucidate the fundamental biochemical processes that take place in the site, *in silico* molecular docking was used to model the interaction between QPLEX/Le^{b} and wtMOMP. Protein ligand docking was performed via FRED RECEPTOR 3.5.0.4 software (OpenEye Scientific Software, Inc., Santa Fe, NM, USA; www.eyesopen.com)[215]. FRED is a docking program that docks ligand molecules from a multiconformer database into a receptor site using an exhaustive search algorithm. The docking process simultaneously predicts the ligand conformation as well as its

position and orientation within these sites (usually referred to as *pose*) and assesses the binding affinity.

7.7.1 Receptor 3D structure

In silico docking requires receptor and ligand coordinate files. Typically, computational biologists start with receptor coordinates from crystallography or NMR spectroscopy and with ligand coordinates generated from SMILES (Simplified Molecular-Input Line-Entry System) strings. As wtMOMP NCTC11168 does not yet have a crystal structure, homology modelling was performed with the aim of building a three-dimensional protein structure model. SwissProt Modeller enabled upload of the crystal structure of MOMP 85H to be used as a template to generate a 3D model of wtMOMP. Although the model was not used for OpenEye docking software, it was useful for emphasising areas highlighted by docking (and carbene footprinting studies). Instead, during docking experiments, the original crystal structure of MOMP85H was used for protein ligand docking as it carries more geometric information than a model and at high sequence similarity of 94% there is limited loss of information.

7.7.2 Ligand 3D model

Firstly, 2D structures of QPLEX and Le^b were constructed on ChemDraw. For docking in OpenEye FRED, the ligand molecule files were converted to a conformationally expanded 3D format such as SDF, MOL2 or PDB. This was performed by copying their ChemDraw structures as SMILES and converting to 3D structure using Openbabel file conversion software.

7.7.3 Preparing the receptor

FRED requires a single receptor to dock ligands into. Receptors are specialized files used by FRED that contain the structure of a target protein, the location and shape of the active site, and optionally the structure of a bound ligand and docking constraints. The wtMOMP target protein

was prepared as a receptor in MakeReceptor for which there are four stages before a receptor is complete; molecule, box, shape constraints.

During the 'molecule' stage the pdb file of MOMP85H was loaded into the programme. Chains A, B, and C were marked as protein whilst the crystallographic waters, bound ligands and other non-bonded solvent molecules were neither marked as protein or ligand, in this case FRED would just ignore them as extra molecules.

The goal of the 'Box' stage of the receptor setup workflow was to define a box enclosing the active site (**figure 7.8**). The box should enclose the entire region active site where heavy atoms of the docking ligand can be placed. Any docked pose with any heavy atom that lies outside the box will be rejected.

The shape of the active site is also called the negative image. It is stored as a potential grid surrounding the active site. The negative image has high potentials where ligand atoms make many contacts with atoms of the active site without clashing and in positions some ligand atoms are likely to occupy when other atoms of the ligand make good contacts with the receptor. The negative image is defined by the inner and outer contours of the receptor. Both contours are always enclosed by the box setup in the previous stage of the setup workflow. Negative image potentials were always favoured towards the protein as opposed to the solvent or being balanced. This meant that the contours would extend closer to the protein before extending out into solvent, while "Favors Solvent" will cause the reverse to happen. Balanced, the default, is reasonable in most cases. Unless the active site was very large, both contours were disabled to increase the space of poses searched for each docking ligand although docking without the outer contour increased docking time by as much as 100-fold.

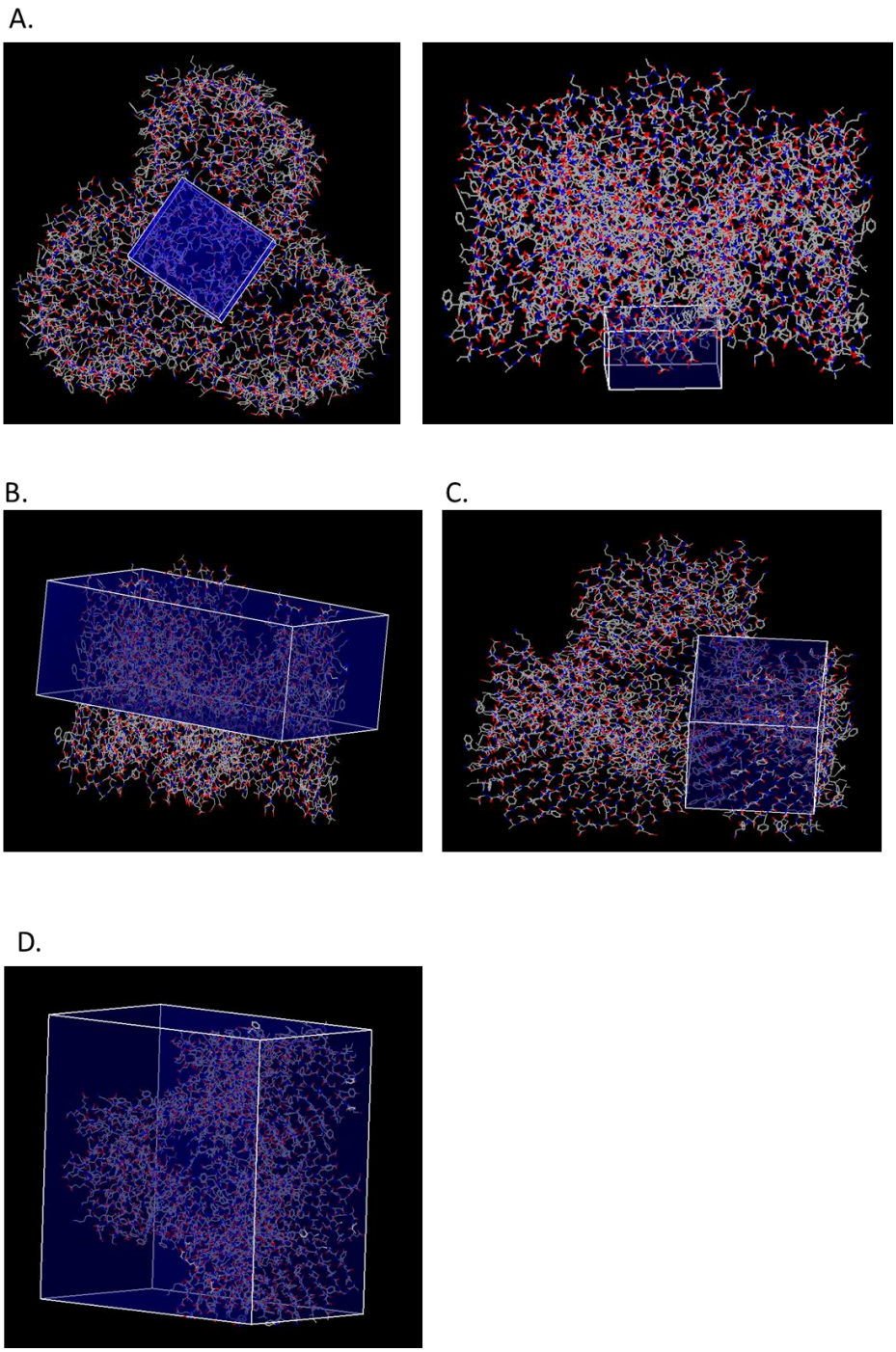


Figure 7.8: A series of boxes were created for exploration of the 'active site' that should be probed for ligand docking. The grid boxes were created on MakeReceptor. **A.** The grid box specifying the underside of the protein. **B.** Grid box specifying extracellular side of MOMP. **C.** Grid box specifying Porin of MOMP. **D.** Grid box to create blind box around the whole surface of MOMP.

7.7.4 Docking

Now that 3D ligand conformation was set up and the receptor box/shapes had been specified, the next stage was to run the docking of the ligand in the specified site of the protein. This was

performed using FRED in OpenEye's command line utility. The following command was used to dock QPLEX to the extracellular site of MOMP using specific parameters:
E:\Docking folder 1411> fred -receptor extracellular.oeb -dbase "QPLEX PDB.pdb" -param fred_settings.param. FRED required the input of the receptor and ligand (dbase) files that had been previously prepared. High resolution of the exhaustive search was used to dock the ligand where translational and rotational step size were set to 1 Å. Docked results were visualised by VIDA 4.4.0.4 where the top 10 resultant ligand poses and translations were placed in the active site of the MOMP receptor.

7.8 Immunohistochemistry of wtMOMP to gut tissue sections

Traditional Immunohistochemistry (IHC) uses antibodies to detect the location of proteins and other antigens in tissue sections. Here wtMOMP binding to gut tissue was detected by direct fluorescence where the protein-antigen interaction was visualized by fluorescent detection of a probe conjugated to the protein. Immunofluorescence staining was also used whereby primary antibodies against the target and fluorescently tagged secondary antibodies were used to visualise binding.

Microscopic analysis of cells and tissues requires the preparation of very thin, high-quality sections mounted on glass slides. Two types of tissue were used: human and mouse. Biopsy specimens of normal (i.e., non-*C. jejuni* infected, noninflamed) human corpus, antrum, duodenum and colon were collected per biopsy as part of diagnosis for gastrointestinal complaints. Regarding the mice sourced tissue, normal and Le^b transgenic stomach tissue was provided. The specimens were fixed in PBS, 4% w/v paraformaldehyde solution for 4 hr and then processed in paraffin blocks and sliced according to standard procedures. The tissue sections were kindly donated to me by Thomas Boren (Department of Medical Biochemistry and Biophysics at Umea University).

7.8.1 Deparaffinisation and rehydration of paraffin sections

The sections were deparaffinised by soaking in xylene and rehydrated by immersing the slides through serial dilutions of ethanol. Firstly, for deparaffinisation, the blocks were immersed in xylene, for three washes of 5 min each. This was followed by 2 washes of 5 min each in 100% ethanol, two washes 5 min each in 95% v/v ethanol, two washes 5 min each in 70% v/v ethanol, two washes 5 min each in 50% v/v ethanol. The last wash was in deionised water for 10 min.

7.8.2 Reduction of background fluorescence

Vector TrueVIEW Autofluorescence Quenching Kit was used to remove unwanted fluorescence in tissue sections due to aldehyde fixation, red-blood cells, and structural elements such as collagen and elastin. Firstly, the extent of autofluorescence with negative control unstained sections (i.e. no detection reagents) was determined in each colour channel. The reagents were prepared as specified in the Vector® TrueVIEW® Autofluorescence Quenching Kit user manual. Excess water was drained from tissue section and a circle was drawn on the slide around the tissue with a hydrophobic barrier pen. The tissue section was then completely covered with 150 µL of the reagent and incubated for 2 - 5 min. After incubation, the sections were washed in PBS buffer for 5 min and excess buffer was drained before mounting the tissue with the coverslip. The treatment of the tissue section for autofluorescence reduction turned the tissue blue which did not interfere with visualisation.

7.8.3 Conjugation of fluorescent probe to wtMOMP

Before binding wtMOMP to the tissue, the protein was conjugated to fluorescent dyes for visualisation. Both the fluorescein (NHS-Fluorescein, Thermo Scientific) and far-red dye (Alexa Fluor™ 647 NHS Ester Thermo Scientific) form amide bonds to the protein's primary amines (R-NH2) through the dye's succinymidyl esters. Succinimidyl esters are preferred over other

amine-reactive reagents, such as isothiocyanates, for attaching fluorophores to amine-containing molecules, because the amide bonds formed in the reaction are as stable as peptide bonds. wtMOMP was diluted to 1 mg/mL in 1mL PBS. 15mmol molar excess of NHS-Fluorescein to wtMOMP was used to conjugate the dye to the protein which required 4 µL of the dye where 1 mg NHS-fluorescein was dissolved in 100 µL DMSO. To protect the reagent from moisture, allow NHS-Fluorescein to equilibrate to room temperature before opening. The solution was mixed well and incubated on ice for 2 hr. Non-reacted NHS-Fluorescein was removed by dialysis with a Slide-A-Lyzer Dialysis Cassettes, 10,000 MWCO, for volumes up to 2 mL. Successful labelling was determined by running the conjugated protein on SDS PAGE and visualizing the gel under UV light.

7.8.4 Staining of gut sections with protein and Le^b

Staining with fluorescently conjugated wtMOMP

The deparaffinised, rehydrated and autofluorescence quenched sections were incubated in 100 µL of 7 pM to 7 µM of far-red or fluorescein labelled wtMOMP protein prepared in PBS. The concentration of the protein was increased from 7 pM to 7 µM until protein staining could be seen on the plates. After washing in PBS-Tween 0.5%, 3 times for 5 mins, fluorescent staining was visualised by confocal microscopy and images were analysed by ImageJ.

Staining with wtMOMP

The deparaffinised and rehydrated sections were incubated in 100 µL of 5 µM of wtMOMP protein prepared in blocking buffer of PBS and 0.1% BSA for an hour at room temperature. The slides were then washed with PBS-Tween 0.5%, 3 times for 5 mins before incubating in 100 uL 1:500 anti-

rabbit anti-MOMP antibody in blocking buffer of PBS and 0.1% BSA for an hour at room temperature. After a second washing step, the slides were incubated in 100 uL far red conjugated goat anti-rabbit secondary antibody conjugated to Alexa fluor 633 (Thermo Fisher #A-21070). After a third washing step, fluorescent staining was visualised by confocal microscopy and images were analysed by ImageJ.

Staining with Le^b

The deparaffinised and rehydrated sections were incubated in 100 µL of 1:500 dilution of anti-Le^b (Seraclone Anti-B #801350) prepared in blocking buffer of PBS and 0.1% BSA for an hour at room temperature. The slides were then washed with PBS-Tween 0.5%, 3 times for 5 mins before incubating with 100 uL 1:500 dilution of goat anti-mouse IgG antibody conjugated to Alexa Fluor 633 (Thermo Fisher # A-21052) in blocking buffer for an hour at room temperature. After a third washing step, the slides were incubated and quenched in 100 uL Vector TrueVIEW Autofluorescence Quenching Kit and fluorescent staining was visualised by confocal microscopy and images analysed by ImageJ.

7.8.5 Confocal Microscopy

The slides were visualised with Zeiss LSM 880 confocal microscope, using 20x/0.8 NA objective. Lasers, 488nm and 633 nm respectively were used to visualise residual autofluorescence and wtMOMP staining. 633 nm laser was used to create transmitted brightfield images as third channel. Detection (emission detected) was set to 499-517 nm for the green and 640-735 nm for the far-red, respectively. Zen Black software and Image j Fiji software were used to acquired images and performs image processing and analysis. 16-bit raw images of the far-red were

visualised on a scale of low and high values as noted under each image, in the Brightness Contrast tool of Fiji.

7.9 Statistical analysis

Multiple t tests, one-way analysis of variance (ANOVA), and paired/unpaired t tests were performed on the relevant data sets throughout the thesis analysis and write up. On data sets where multiple t-tests were initially applied, upon further learnings, the statical analysis was updated and performed with ANOVA analysis instead. This sub chapter outlines the benefits of using ANOVA over multiple T tests.

ANOVA vs multiple paired t-test

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups. The null hypothesis is tested where:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

where μ = group mean and k = number of groups. If a statistically significant result is returned by the one-way ANOVA then we accept the alternative hypothesis (H_A), which is that there are at least two group means that are statistically significantly different from each other. ANOVA was applied to data obtained in chapter 5.2, to analyse the statical significant difference between binding of wtMOMP, MOMP$^{T/G268}$ and MOMP$^{\Delta pseD}$ to Le^b. Further, the same analysis was applied to measure the statistical significant difference between the binding of Le^b to protein at at 0 concentration of QPLEX and increasing concentrations of QPLEX. Alternatively, a paired t-test tests the null hypothesis that the true mean difference is zero. When comparing two groups, if we use either the t-test or ANOVA method to analyse the data set, the outcome of the statical analysis would be identical. Usually paired t tests are performed on experiments with two groups but multiple paired t tests can be performed on data with 3 or more groups, at a loss of statical power (such as a control group and several treatment groups). For increased statical power, ANOVA is used when comparing across data sets with three or more groups as ANOVA tests for

the variation between the means and within each mean. It not only compares the means (like the t-test) but examines the variation in calculating the means. Excessive variation in calculating a mean reduces the chance of determining a significant difference.

Chapter 8: **Conclusion**

8.1 Introduction

In this chapter I summarise my findings from every chapter and put them in the context of the literature whilst pointing out contributions to existing knowledge. I will also put forward suggestions for future work in the field.

8.2 MOMP purification

In chapter 3, we were able to successfully express NCTC11168 MOMP in *E. coli,* by using a codon optimised *momp* gene. This allowed purification of recombinant MOMP NCTC11168 (rMOMP) albeit, at low yield. The purification of functional proteins from insoluble bacterial fractions and inclusion bodies often leads to difficulties with refolding and does not result in full recovery of biological activity. Future studies would involve optimisation of expression and purification of recombinant MOMP.

Alternatively, MOMP obtained directly from *C. jejuni* was solubilised with some changes and optimisation of pre-existing MOMP purification protocols by the use of minimum concentration of detergent at 0.45% v/v octyl-POE, which seemed to have minimum interference with downstream processes. Detergent can have a negative effect on the result of downstream experiments due to their interference in the technology but the removal of the detergent from the protein can affect its conformation, which is important for the function of the protein, when it is removed from its hydrophobic environment. Although membrane proteins represent some of the most important molecules in life sitting between the cell and the outside world or between cellular compartments, the lack of progress in the study of membrane protein structure and function remains a significant

frustration for academics and commercial organisations. Gaps in the knowledge of outer membrane proteins exist due to the eventual removal of the membrane from around the protein which generally renders it unfolded and inactive. Only 2% of high-resolution structures are of membrane proteins[216]. The inclusion of detergent, although keeping the protein in its active form, has negative effects in downstream applications, for example, interactions between subunits within a membrane protein complex are often perturbed or abolished by the presence of the detergent and detergent solubilized membrane proteins are generally inherently unstable with very short "shelf lives". This means that comparative experiments between samples often suffer from uncertainty in terms of the specific activity of the preparation. Since the late 90's, there has been an increasing amount of research into new approaches to membrane protein solubilisation and new studies focus on the fact that successful membrane protein folding required the physically complex membrane structure which includes the lipids that are usually removed from the surrounding environment. These new methods extract between 10-100 lipids in close contact with the chosen protein, although they still require the membrane be pre-solubilised in detergent before extraction[217]. In 2009, data was published that showed that a simple organic polymer, Styrene Maleic Acid Co-polymer, SMA, could be used to directly extract proteins from membranes into self-contained styrene maleic acid lipid particles (SMALPs), without the need for a detergent. They showed that the structure is stabilised by the intercalation of the hydrophobic styrene groups between the acyl chains of the bilayer while the hydrophilic maleic acid groups face the solvent. Further, they showed that that the encapsulated bilayer retains many of the physical properties of the parent membrane including the lipid mixture, structural organisation and phase behaviour. The method can be used to extract proteins with up to 36 transmembrane helices, less than 400 kDa and with experiments carried out at above pH 6.5[154, 218, 219]. As MOMP contains 16

transmembrane helices and has a molecular weight of 135 kDa, it would be suitable for non-detergent extraction with SMA.

Future work with the use of MOMP may include optimising the extraction of MOMP with the use of such SMA or cookie cutter techniques to better sustain protein activity and increase performance in downstream experiments.

8.3 Characterisation of MOMP glycan

In chapter 4, I determined that there is an m/z difference of 518.43 between the glycosylated (wtMOMP) and non-glycosylated ($MOMP^{T/G268}$) variants which accounts for *N*-acetyl-galactosamine (due to *O*-glycosylation) and a further two unknown units of monosaccahride with a variety of possibilities, e.g. the attachment of a galactose, (MW = 180.16 Da) and a ribose (MW= 150.13 Da). The glycan unit is at least this size as monosaccharides could be hydrolysed during preparation for mass spectrometry. Of all the post-translational modifications, glycosylation is one of the most challenging to characterise because of the variability in the attached glycans, and the isobaric nature of many of these glycans as the modification can undergo fragmentation. Adding to the analytical challenge is that glycosylation is not just a single modification but is often a set of modifications in which a single glycosylated site on a peptide may have a number of glycan isoforms with different chain lengths and branches attached. Because of this heterogeneity, there is no single mass shift associated with glycosylation[220]. Of the three basic approaches to study protein glycosylation, we chose to leave the glycosylation on the protein to observe potential mass shifts using MS, as opposed to using glycosidases to separate the glycan and study it by itself. Further we have outlined peptides for which there is MSMS data available and future work would

be to fragment the peptide of interest (peptide of threonine 268) for in-depth study of the monosaccharide.

Future studies to further characterise the glycan on wtMOMP should screen for lectins that have the ability to bind to wtMOMP. This may be performed by western blotting which can be used to distinguish between lectin binding. The bound lectins are visualized by binding streptavidin-alkaline phosphatase and conversion of luminescent substrates. The concentration of lectins used must be low enough to reduce false-positive binding and signal high specificity[221]. When combined in a serial format, multiple lectins can be used in affinity chromatography to isolate glycan containing protein based on affinity for the lectin[222]. MOMP^{pseD} and MOMP$^{T/G268}$ should also be tested against lectins to remove any background lectin binding that may be present in the protein after removal of the glycan.

8.4 Interaction of MOMP with QPLEX and Le^b

Interaction of BgAgs with MOMP protein was shown by the use of ELISA binding experiments in chapter 5, and supported previous studies by Mahdavi *et al*.[110]. Le^b binding to wtMOMP resulted in a response with EC50 of 23nM, displaying tight binding affinity. Inhibition with QPLEX was also revealed by ELISA binding studies where the results show that there was a preference of the glycosylated MOMP over the non-glycosylated variants. The change in protein conformation of glycosylated MOMP upon ligand binding was shown by carbene footprinting where distinct areas were exposed after addition of ligand. The transient nature of the binding of QPLEX and Le^b to the protein may be the cause of increase in carbene masking (upon exposed areas due to conformational change) rather than reduction in carbene masking (upon binding of ligand to the protein). The nature of these interactions can be measured by techniques such as SPR to answer

the questions of how fast the molecules associate and dissociate, the affinity between molecules (affinity) and a measure of the specificity[223] between the glycosylated and non-glycosylated MOMP.

Due to a number of restrictions of lab work caused by COVID-19, SPR data could not be obtained, so computational protein-ligand docking was the next best alternative. A series of protein-ligand software was trialled (e.g. Autodock Vina, GOLD) but I concluded that OpenEye Docking would deliver more consistent results due to its simpler user interface, thus leading to more repeatable results. More importantly, OpenEye was the only programme that would allow for docking of a metal containing compound, such as the iron complexed QPLEX, to the receptor protein, without major issues. Docking studies confirmed results of footprinting of QPLEX to the protein as the regions of conformational change seen in footprinting were complementary to the areas of favourable ligand-protein docking. The results of docking in chapter 5 shows novel results of QPLEX binding within the porin as it travels from the extracellular loop region to the upper part of the porin down to the constriction zone, with decreasing intensity. In the loop region, there is a favourable overall score of -8.37 which is contributed to by the amino acids: Arg39, Arg41, Asn53, Asn54, Arg67 and Ser182. Amino acid residues that would accept hydrogen bonds from QPLEX are Ser55, Asn89, Asp138, Asp199 and Asp308. Once QPLEX travels deeper into the porin, although the bonds between the compound and the protein are still favourable, the bonds become slightly weaker, reducing the CHEMGAUSS score to -8.0. The amino acids of interaction in the top half of the porin are Asp236, Glu252, Asp287, Gly310, Glu319, Asp356, Lys278, Thr342 and Arg381. Once the ligand moves further down into the constriction zone the residue fingerprint that the ligand interacts with are Gln69, Gln131, Ser35, Asn71, Ser73, Gln85, Gln111, Tyr113 and Lys150. The ligand does not make any favourable bonds with the underside of the protein.

Future docking studies would involve expert considering the glycan artefact and how this changes the interaction of the compound in the porin for example by changing the quaternary structure for the protein.

8.5 MOMP interaction with the gut

Researchers have identified a number of *C. jejuni* proteins that bind to cultured epithelial cells. These adhesive proteins include CadF, CapA, JlpA, MOMP and PEB1[224]. Mosser *et al*. showed that MOMP was able to partially inhibit the attachment of the bacteria to INT 407 cell membranes partly, when the protein was isolated under native conditions[121]. The source of INT-407 cell lines have recently been thrown into question, as HEP-2 and INT-407 cells are now widely known to have been contaminated with HeLa cells and are easily interchangeable[225, 226]. In my studies, wtMOMP bound *ex vivo* to human gut tissue, specifically tissue from the corpus. Interaction may be caused by the ability of wtMOMP to bind Le^b and other blood group antigens (chapter 5.2), as these antigens are regularly found on the mucosa of the gut. The degree to which glycosylation is contributing to binding of wtMOMP to the intestinal tissue was determined by replicating the experiments with the non-glycosylated $MOMP^{T/G268}$ and $MOMP^{pseD}$ and locating the binding areas as well as visualising the extent of binding. The use of intestinal tissue sections for adhesion studies in chapter 6 makes for a stronger positive result than what is seen in the literature where cultured cells, as opposed to tissue sections, are used to show binding of MOMP protein to intestinal cells, as we can be sure of the source of the tissue and that there is no issue of cell line contamination as there would be with cultured cells. The next set of experiments should measure the inhibition of binding of MOMP to the intestinal tissue by compounds such as QPLEX in titration studies to measure the inhibitory concentration.

Chapter 9: **Appendix**

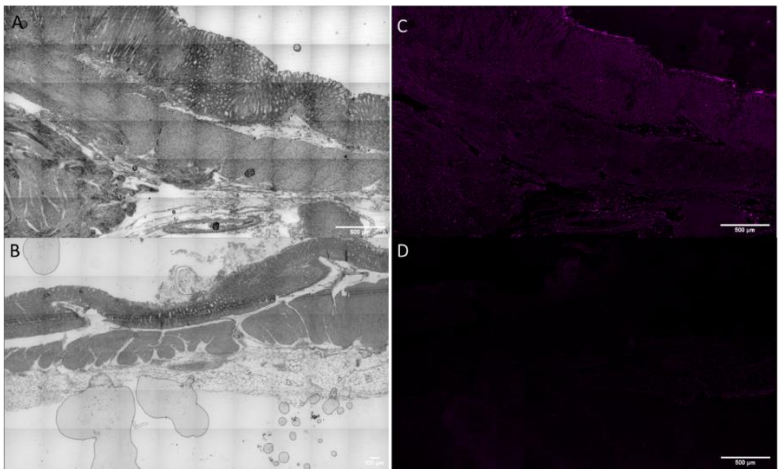


Figure 9.1: Normal (A) and transgenic Le^b +ve (B) P28 FVB/N mouse stomach tissue stained with $MOMP^{T/G268}$ protein, followed by anti-MOMP antibody raised in rabbit and counterstained with far-red tagged goat anti-rabbit antibody and an auto-fluorescence quencher. C and D) Images in far-red channel were manipulated to magenta and normalized to 0 to 65279.

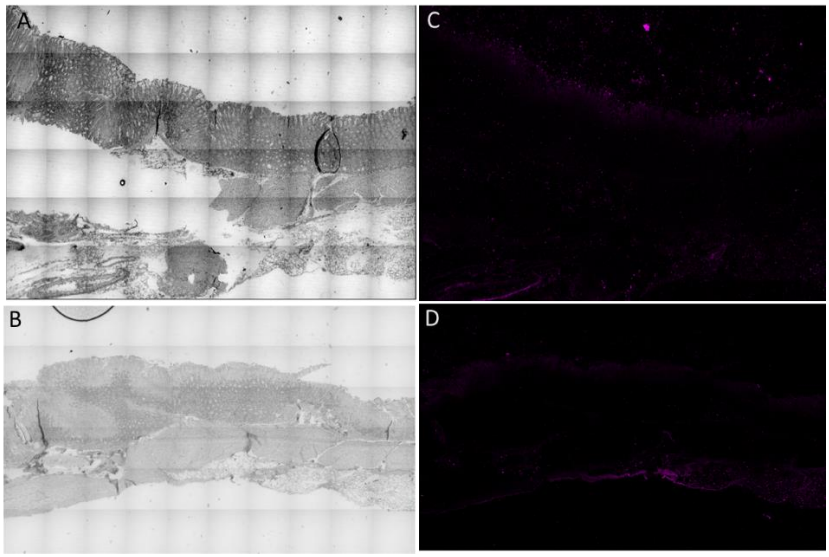


Figure 9.2: Normal (A) and transgenic Le^b +ve (B) P28 FVB/N mouse stomach tissue stained with wtMOMP protein, followed by anti-MOMP antibody raised in rabbit and counterstained with far-red tagged goat anti-rabbit antibody and an auto-fluorescence quencher. C and D) Images in far-red channel were manipulated to magenta and normalized to 0 to 65279.

References

1. Organization,W.H. (2007) Critically important antimicrobials for human medicine: categorization for the development of risk management strategies to contain antimicrobial resistance due to non-human antimicrobial use: report of the second WHO Expert Meeting, Copenhagen, 29-31 May.

2. Devleesschauwer,B., Bouwknegt,M., Mangen,M.-J.J. and Havelaar,A.H. (2017) Chapter 2 - Health and economic burden of Campylobacter. *Campylobacter*, 10.1016/B978-0-12-803623-5.00002-2.

3. Silva,J., Leite,D., Fernandes,M., Mena,C., Anthony Gibbs,P., Teixeira,P., Koutsoumanis,K. and Torriani,S. (2011) Campylobacter spp. as a foodborne pathogen: a review. 10.3389/fmicb.2011.00200.

4. Buzby,J.C., Allos,B.M. and Roberts,T. (1997) The economic burden of Campylobacter-associated Guillain-Barre syndrome. In *Journal of Infectious Diseases*. Oxford University Press, Vol. 176.

5. Kaakoush,N.O., Castaño-Rodríguez,N., Mitchell,H.M. and Man,S.M. (2015) Global epidemiology of Campylobacter infection. *Clin. Microbiol. Rev.*, **28**, 687–720.

6. Igwaran,A. and Okoh,A.I. (2019) Human campylobacteriosis: A public health concern of global importance. *Heliyon*, **5**.

7. Gradel,K.O., Nielsen,H.L., Schønheyder,H.C., Ejlertsen,T., Kristensen,B. and Nielsen,H. (2009) Increased short-and long-term risk of inflammatory bowel disease after salmonella or campylobacter gastroenteritis. *Gastroenterology*, **137**, 495–501.

8. Macfarlane,S., Furrie,E., Macfarlane,G.T. and Dillon,J.F. (2007) Microbial colonization of the upper gastrointestinal tract in patients with Barrett's esophagus. *Clin. Infect. Dis.*, **45**, 29–38.

9. ETOH,Y., DEWHIRST,F.E., PASTER,B.J., YAMAMOTO,A. and GOTO,N. (1993) Campylobacter showae sp. nov., isolated from the human oral cavity. *Int. J. Syst. Evol. Microbiol.*, **43**, 631–639.

10. Parry,S.D., Stansfield,R., Jelley,D., Gregory,W., Phillips,E., Barton,J.R. and Welfare,M.R. (2003) Does bacterial gastroenteritis predispose people to functional gastrointestinal disorders? A prospective, community-based, case–control study. *Am. J. Gastroenterol.*,

98, 1970–1975.

11. Wu,N., Yang,X., Zhang,R., Li,J., Xiao,X., Hu,Y., Chen,Y., Yang,F., Lu,N. and Wang,Z. (2013) Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microb. Ecol.*, **66**, 462–470.

12. Verdu,E.F., Mauro,M., Bourgeois,J. and Armstrong,D. (2007) Clinical onset of celiac disease after an episode of Campylobacter jejuni enteritis. *Can. J. Gastroenterol. Hepatol.*, **21**, 453–455.

13. Vaughan-Shaw,P.G., Rees,J.R., White,D. and Burgess,P. (2010) Campylobacter jejuni cholecystitis: a rare but significant clinical entity. *BMJ Case Rep.*, **2010**, bcr1020092365.

14. Baker,M.G., Kvalsvig,A., Zhang,J., Lake,R., Sears,A. and Wilson,N. (2012) Declining Guillain-Barré syndrome after campylobacteriosis control, New Zealand, 1988–2010. *Emerg Infect Dis*, **18**, 226–233.

15. Koga, áM, Gilbert, áM, Li,J., Koike,S., Takahashi,M., Furukawa,K., Hirata,K. and Yuki,N. (2005) Antecedent infections in Fisher syndrome A common pathogenesis of molecular mimicry. *Neurology*, **64**, 1605–1611.

16. Man,S.M. (2011) The clinical importance of emerging Campylobacter species. *Nat. Rev. Gastroenterol. Hepatol.*, **8**, 669–685.

17. Dinant,S., Schurink,C.A.M., Deckers,J.W. and Severin,J.A. (2011) Aortic homograft endocarditis caused by Campylobacter jejuni. *J. Clin. Microbiol.*, **49**, 4016–4017.

18. Hannu,T., Mattila,L., Rautelin,H., Siitonen,A. and Leirisalo-Repo,M. (2005) Three cases of cardiac complications associated with Campylobacter jejuni infection and review of the literature. *Eur. J. Clin. Microbiol. Infect. Dis.*, **24**, 619–622.

19. Tsoni,K., Papadopoulou,E., Michailidou,E. and Kavaliotis,I. (2013) Campylobacter jejuni meningitis in a neonate: a rare case report. *J. Neonatal. Perinatal. Med.*, **6**, 183–185.

20. Han,X.Y., Tarrand,J.J. and Rice,D.C. (2005) Oral Campylobacter species involved in extraoral abscess: a report of three cases. *J. Clin. Microbiol.*, **43**, 2513–2515.

21. Ajene,A.N., Fischer Walker,C.L. and Black,R.E. (2013) Enteric pathogens and reactive arthritis: a systematic review of Campylobacter, salmonella and Shigella-associated reactive arthritis. *J.*

Heal. Popul. Nutr., **31**, 299–307.

22. Simor,A.E., Karmali,M.A., Jadavji,T. and Roscoe,M. (1986) Abortion and perinatal sepsis associated with Campylobacter infection. *Rev. Infect. Dis.*, **8**, 397–402.

23. EFSA,E. (2011) European Food Safety Authority, European Centre for Disease Prevention and Control. *Eur. Union Summ. Rep. trends sources zoonoses, zoonotic agents food-borne outbreaks*.

24. Scallan,E., Crim,S.M., Runkle,A., Henao,O.L., Mahon,B.E., Hoekstra,R.M. and Griffin,P.M. (2015) Bacterial Enteric Infections Among Older Adults in the United States: Foodborne Diseases Active Surveillance Network, 1996–2012. *Foodborne Pathog. Dis.*, **12**, 492–499.

25. Scallan,E., Hoekstra,R.M., Angulo,F.J., Tauxe,R. V, Widdowson,M.-A., Roy,S.L., Jones,J.L. and Griffin,P.M. (2011) Foodborne illness acquired in the United States—major pathogens. *Emerg Infect Dis*, **17**.

26. Agency,F.S. and Agency,F.S. (2014) A microbiological survey of Campylobacter contamination in fresh whole UK produced chilled chickens at retail sale - interim report to cover quarters 1 - 3. Food Standards Agency, London, UK, London.

27. Agency,F.S. (2014) A UK WIDE MICROBIOLOGICAL SURVEY OF CAMPYLOBACTER CONTAMINATION IN FRESH WHOLE CHILLED CHICKENS AT RETAIL SALE.

28. Tam,C.C., O'Brien,S.J., Tompkins,D.S., Bolton,F.J., Berry,L., Dodds,J., Choudhury,D., Halstead,F., Iturriza-Gómara,M. and Mather,K. (2012) Changes in causes of acute gastroenteritis in the United Kingdom over 15 years: microbiologic findings from 2 prospective, population-based studies of infectious intestinal disease. *Clin. Infect. Dis.*

29. Tam,C.C., Rodrigues,L.C., Viviani,L., Dodds,J.P., Evans,M.R., Hunter,P.R., Gray,J.J., Letley,L.H., Rait,G. and Tompkins,D.S. (2012) Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut*, **61**, 69–77.

30. Havelaar,A.H., Ivarsson,S., Löfdahl,M. and Nauta,M.J. (2013) Estimating the true incidence of campylobacteriosis and salmonellosis in the European Union, 2009. *Epidemiol. Infect.*, **141**, 293–302.

31. Wagenaar,J.A., French,N.P. and Havelaar,A.H. (2013) Preventing campylobacter at the source: Why is it so difficult? *Clin. Infect. Dis.*, **57**, 1600–1606.

32. Hoffmann,S., Batz,M.B. and Morris Jr,J.G. (2012) Annual cost of illness and quality-adjusted life year losses in the United States due to 14 foodborne pathogens. *J. Food Prot.*, **75**, 1292–1302.

33. Panel,E.B. (2011) Scientific Opinion on Campylobacter in broiler meat production: control options and performance objectives and/or targets at different stages of the food chain. *EFSA J.*, **9**, 2015.

34. Swaminathan,A., Torresi,J., Schlagenhauf,P., Thursky,K., Wilder-Smith,A., Connor,B.A., Schwartz,E., Keystone,J. and O'Brien,D.P. (2009) A global study of pathogens and host risk factors associated with infectious gastrointestinal disease in returned international travellers. *J. Infect.*, **59**, 19–27.

35. Skarp,C.P.A., Hänninen,M.-L. and Rautelin,H.I.K. (2016) Campylobacteriosis: the role of poultry meat. *Clin. Microbiol. Infect.*, **22**, 103–109.

36. Robinson,D.A. (1981) Infective dose of Campylobacter jejuni in milk. *Br. Med. J. (Clin. Res. Ed).*, **282**, 1584.

37. Blaser,M.J., Glass,R.I., Huq,M.I., Stoll,B., Kibriya,G.M. and Alim,A.R. (1980) Isolation of Campylobacter fetus subsp. jejuni from Bangladeshi children. *J. Clin. Microbiol.*, **12**, 744–7.

38. McSweegan,E. and Walker,R.I. (1986) Identification and characterization of two Campylobacter jejuni adhesins for cellular and mucous substrates. *Infect. Immun.*, **53**, 141–8.

39. Hickey,T.E., McVeigh,A.L., Scott,D.A., Michielutti,R.E., Bixby,A., Carroll,S.A., Bourgeois,A.L. and Guerry,P. (2000) Campylobacter jejuni cytolethal distending toxin mediates release of interleukin-8 from intestinal epithelial cells. *Infect. Immun.*, **68**, 6535–41.

40. Hu,L., Bray,M.D., Osorio,M. and Kopecko,D.J. (2006) Campylobacter jejuni Induces Maturation and Cytokine Production in Human Dendritic Cells. *Infect. Immun.*, **74**, 2697–2705.

41. Kaldor,J., Pritchard,H., Serpell,A. and Metcalf,W. (1983) Serum Antibodies in Campylobacter Enteritis. *J. Clin. Microbiol.*, **18**, 1–4.

42. Ruiz-Palacios,G.M., Cervantes,L., Ramos,P., Chavez-Munguia,B. and Newburg,D.S. (2003) Campylobacter jejuni Binds Intestinal H(O) Antigen (Fucα1, 2Galβ1, 4GlcNAc), and Fucosyloligosaccharides of Human Milk Inhibit Its Binding and Infection. *J. Biol. Chem.*, **278**, 14112–14120.

43. Byrne,C.M., Clyne,M. and Bourke,B. (2007) Campylobacter jejuni adhere to and invade chicken intestinal epithelial cells in vitro. *Microbiology*, **153**, 561–569.

44. Clench,M.H. and Mathias ',J.R. (1995) THE AVIAN CECUM: A REVIEW. *Wilson Bull*, **107**, 93–121.

45. Thibodeau,A., Fravalo,P., Yergeau,E., Arsenault,J., Lahaye,L. and Letellier,A. (2015) Chicken Caecal Microbiome modifications induced by campylobacter Jejuni colonization and by a non-antibiotic feed additive. *PLoS One*, **10**, e0131978.

46. Nachamkin,I., Yang,X.H. and Stern,N.J. (1993) Role of Campylobacter jejuni flagella as colonization factors for three-day-old chicks: analysis with flagellar mutants. *Appl. Environ. Microbiol.*, **59**, 1269–73.

47. Brás,A.M., Chatterjee,S., Wren,B.W., Newell,D.G. and Ketley,J.M. (1999) A novel Campylobacter jejuni two-component regulatory system important for temperature-dependent growth and colonization. *J. Bacteriol.*, **181**, 3298–302.

48. Young,K.T., Davis,L.M. and DiRita,V.J. (2007) Campylobacter jejuni: molecular biology and pathogenesis. *Nat. Rev. Microbiol.*, **5**, 665–679.

49. Garrity,G.M., Bell,J.A. and Lilburn,T.G. (2005) Family VII. Bradyrhizobiaceae In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology.

50. Man,S.M. (2011) The clinical importance of emerging Campylobacter species. *Nat. Rev. Gastroenterol. Hepatol.*, **8**, 669–685.

51. Daffé,M. and Etienne,G. (1999) The capsule of Mycobacterium tuberculosis and its implications for pathogenicity. *Tuber. Lung Dis.*, **79**, 153–169.

52. Fry,B.N., Feng,S., Chen,Y.Y., Newell,D.G., Coloe,P.J. and Korolik,V. (2000) The galE gene of Campylobacter jejuni is involved in lipopolysaccharide synthesis and virulence. *Infect. Immun.*, **68**, 2594–601.

53. Cullen,T.W. and Trent,M.S. (2010) A link between the assembly of flagella and lipooligosaccharide of the Gram-negative bacterium Campylobacter jejuni. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 5160–5.

54. Hughes,R. (2004) Campylobacter jejuni in Guillain-Barré syndrome. *Lancet. Neurol.*, **3**, 644.

55. Lara-Tejero,M. and Galán,J.E. (2000) A bacterial toxin that controls cell cycle progression as a

deoxyribonuclease I-like protein. *Science*, **290**, 354–7.

56. Szymanski,C.M., Michael,F.S., Jarrell,H.C., Li,J., Gilbert,M., Larocque,S., Vinogradov,E. and Brisson,J.-R. (2003) Detection of conserved N-linked glycans and phase-variable lipooligosaccharides and capsules from campylobacter cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy. *J. Biol. Chem.*, **278**, 24509–20.

57. Jones,M.A., Marston,K.L., Woodall,C.A., Maskell,D.J., Linton,D., Karlyshev,A. V., Dorrell,N., Wren,B.W. and Barrow,P.A. (2004) Adaptation of Campylobacter jejuni NCTC11168 to High-Level Colonization of the Avian Gastrointestinal Tract. *Infect. Immun.*, **72**, 3769–3776.

58. ME,K., JE,C., AM,K., MR,M., JD,K. and SG,G. (2005) Identification of a fibronectin-binding domain within the Campylobacter jejuni CadF protein. *Mol. Microbiol.*, **57**, 1022–1035.

59. MR,M., JE,Y. and ME,K. (2003) Maximal adherence and invasion of INT 407 cells by Campylobacter jejuni requires the CadF outer-membrane protein and microfilament reorganization. *Microbiology*, **149**, 153–165.

60. Monteville,M.R. and Konkel,M.E. (2002) Fibronectin-Facilitated Invasion of T84 Eukaryotic Cells by Campylobacter jejuni Occurs Preferentially at the Basolateral Cell Surface. *Infect. Immun.*, **70**, 6665.

61. S,J., YC,S., A,E., PM,S. and VL,C. (2003) JlpA of Campylobacter jejuni interacts with surface-exposed heat shock protein 90alpha and triggers signalling pathways leading to the activation of NF-kappaB and p38 MAP kinase in epithelial cells. *Cell. Microbiol.*, **5**, 165–174.

62. S,J., A,J., J,L., EK,H., P,S. and VL,C. (2001) JlpA, a novel surface-exposed lipoprotein specific to Campylobacter jejuni, mediates adherence to host epithelial cells. *Mol. Microbiol.*, **39**, 1225–1236.

63. SS,A., NJ,O., KG,W., MA,J., GJ,I., DP,T. and DA,A. (2007) CapA, an autotransporter protein of Campylobacter jejuni, mediates association with human epithelial cells and colonization of the chicken gut. *J. Bacteriol.*, **189**, 1856–1865.

64. Peis,Z. and Blasersqli,M.J. (1993) PEB1, the Major Cell-binding Factor of Cumpylobucter jejuni, Is a Homolog of the Binding Component in Gram-negative Nutrient Transport Systems*. *J. BIOLOCJICAL CHBMI~TRY*, **268**, 18717–18735.

65. Z,P., C,B., B,G., S,B., XZ,H., DJ,K., AL,B., JL,F. and MJ,B. (1998) Mutation in the peb1A locus of

Campylobacter jejuni reduces interactions with epithelial cells and intestinal colonization of mice. *Infect. Immun.*, **66**, 938–943.

66. Konkel,M.E., Klena,J.D., Rivera-Amill,V., Monteville,M.R., Biswas,D., Raphael,B. and Mickelson,J. (2004) Secretion of Virulence Proteins from Campylobacter jejuni Is Dependent on a Functional Flagellar Export Apparatus. *J. Bacteriol.*, **186**, 3296.

67. ME,K., BJ,K., V,R.-A. and SG,G. (1999) Identification of proteins required for the internalization of Campylobacter jejuni into cultured mammalian cells. *Adv. Exp. Med. Biol.*, **473**, 215–224.

68. Dingle,K.E., Colles,F.M., Wareing,D.R.A., Ure,R., Fox,A.J., Bolton,F.E., Bootsma,H.J., Willems,R.J.L., Urwin,R. and Maiden,M.C.J. (2001) Multilocus sequence typing system for Campylobacter jejuni. *J. Clin. Microbiol.*, **39**, 14–23.

69. Dorrell,N., Mangan,J.A., Laing,K.G., Hinds,J., Linton,D., Al-Ghusein,H., Barrell,B.G., Parkhill,J., Stoker,N.G., Karlyshev,A. V., *et al.* (2001) Whole genome comparison of Campylobacter jejuni human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res.*, **11**, 1706–1715.

70. Wiesner,R.S., Hendrixson,D.R. and DiRita,V.J. (2003) Natural transformation of Campylobacter jejuni requires components of a type II secretion system. *J. Bacteriol.*, **185**, 5408–5418.

71. Bacon,D.J., Alm,R.A., Hu,L., Hickey,T.E., Ewing,C.P., Batchelor,R.A., Trust,T.J. and Guerry,P. (2002) DNA Sequence and Mutational Analyses of the pVir Plasmid of Campylobacter jejuni 81-176. *Infect. Immun.*, **70**, 6242.

72. Larsen,J.C., Szymanski,C. and Guerry,P. (2004) N-Linked Protein Glycosylation Is Required for Full Competence in Campylobacter jejuni 81-176. *J. Bacteriol.*, **186**, 6508.

73. Wooldridge,K. (2009) Bacterial secreted proteins Caister Academic Press (March 2, 2009).

74. Navarro-Garcia,F., Ruiz-Perez,F., Cataldi,Á. and Larzábal,M. (2019) Type VI Secretion System in Pathogenic Escherichia coli: Structure, Role in Virulence, and Acquisition. *Front. Microbiol.*, **0**, 1965.

75. Lertpiriyapong,K., Gamazon,E.R., Feng,Y., Park,D.S., Pang,J., Botka,G., Graffam,M.E., Ge,Z. and Fox,J.G. (2012) Campylobacter jejuni type VI secretion system: Roles in adaptation to deoxycholic acid, host cell adherence, invasion, and in vivo colonization. *PLoS One*, **7**.

76. Singh,A. and Mallick,A.I. (2019) Role of putative virulence traits of Campylobacter jejuni in regulating differential host immune responses. *J. Microbiol. 2019 574*, **57**, 298–309.

77. Tsirigotaki,A., De Geyter,J., Šoštaric´,N., Economou,A. and Karamanou,S. (2016) Protein export through the bacterial Sec pathway. *Nat. Rev. Microbiol. 2016 151*, **15**, 21–36.

78. Palmer,T. and Berks,B.C. (2012) The twin-arginine translocation (Tat) protein export pathway. *Nat. Rev. Microbiol. 2012 107*, **10**, 483–496.

79. KG,W., PH,W. and JM,K. (1996) Host signal transduction and endocytosis of Campylobacter jejuni. *Microb. Pathog.*, **21**, 299–305.

80. Hugdahl,M.B., Beery,J.T. and Doyle,M.P. (1988) Chemotactic behavior of Campylobacter jejuni. *Infect. Immun.*, **56**, 1560–1566.

81. Bugaytsova,J.A., Björnham,O., Chernov,Y.A., Gideonsson,P., Henriksson,S., Mendez,M., Sjöström,R., Mahdavi,J., Shevtsova,A., Ilver,D., *et al.* (2017) Helicobacter pylori Adapts to Chronic Infection and Gastric Disease via pH-Responsive BabA-Mediated Adherence. *Cell Host Microbe*, **21**, 376–389.

82. Sheu,B., Yang,H., Yeh,Y. and Wu,J. (2010) Helicobacter pylori colonization of the human gastric epithelium: a bug's first step is a novel target for us. *J. Gastroenterol. Hepatol.*, **25**, 26–32.

83. Kim,J.S., Chang,J.H., Chung,S. Il and Yum,J.S. (1999) Molecular Cloning and Characterization of theHelicobacter pylori fliD Gene, an Essential Factor in Flagellar Structure and Motility. *J. Bacteriol.*, **181**, 6969–6976.

84. Wassenaar,T.M., Bleumink-Pluym,N.M. and Van der Zeijst,B.A. (1991) Inactivation of Campylobacter jejuni flagellin genes by homologous recombination demonstrates that flaA but not flaB is required for invasion. *EMBO J.*, **10**, 2055.

85. Clyne,M., Ocroinin,T., Suerbaum,S., Josenhans,C. and Drumm,B. (2000) Adherence of Isogenic Flagellum-Negative Mutants ofHelicobacter pylori and Helicobacter mustelae to Human and Ferret Gastric Epithelial Cells. *Infect. Immun.*, **68**, 4335–4339.

86. Haiko,J. and Westerlund-Wikström,B. (2013) The role of the bacterial flagellum in adhesion and virulence. *Biology (Basel).*, **2**, 1242–67.

87. Sert,V., Cans,C., Tasca,C., Bret-Bennis,L., Oswald,E., Ducommun,B. and De Rycke,J. (1999) The

bacterial cytolethal distending toxin (CDT) triggers a G2 cell cycle checkpoint in mammalian cells without preliminary induction of DNA strand breaks. *Oncogene*, **18**, 6296–6304.

88. Lara-Tejero,M. and Galan,J.E. (2001) CdtA, CdtB, and CdtC Form a Tripartite Complex That Is Required for Cytolethal Distending Toxin Activity. *Infect. Immun.*, **69**, 4358–4365.

89. Lee,R.B., Hassane,D.C., Cottle,D.L. and Pickett,C.L. (2003) Interactions of Campylobacter jejuni cytolethal distending toxin subunits CdtA and CdtC with HeLa cells. *Infect. Immun.*, **71**, 4883–90.

90. Fox,J.G., Rogers,A.B., Whary,M.T., Ge,Z., Taylor,N.S., Xu,S., Horwitz,B.H. and Erdman,S.E. (2004) Gastroenteritis in NF-kappaB-deficient mice is produced with wild-type Camplyobacter jejuni but not with C. jejuni lacking cytolethal distending toxin despite persistent colonization with both strains. *Infect. Immun.*, **72**, 1116–25.

91. AbuOun,M., Manning,G., Cawthraw,S.A., Ridley,A., Ahmed,I.H., Wassenaar,T.M. and Newell,D.G. (2005) Cytolethal Distending Toxin (CDT)-Negative Campylobacter jejuni Strains and Anti-CDT Neutralizing Antibodies Are Induced during Human Infection but Not during Colonization in Chickens. *Infect. Immun.*, **73**, 3053–3062.

92. Dean,L. (2005) Blood group antigens are surface markers on the red blood cell membrane.

93. de Mattos,L.C. (2016) Structural diversity and biological importance of ABO, H, Lewis and secretor histo-blood group carbohydrates. *Rev. Bras. Hematol. Hemoter.*, **38**, 331–340.

94. Henry,S.M. (2001) Molecular diversity in the biosynthesis of GI tract glycoconjugates. A blood group related chart of microorganism receptors. In *Transfusion Clinique et Biologique*. Elsevier Masson SAS, Vol. 8, pp. 226–230.

95. Anstee,D.J. (2010) The relationship between blood groups and disease. *Blood*, **115**, 4635–4643.

96. King,M.-J., Henry,S. and Samuelsson,B. (2000) ABO POLYMORPHISMS AND THEIR PUTATIVE BIOLOGICAL RELATIONSHIPS WITH DISEASE. In *Human Blood Cells*. PUBLISHED BY IMPERIAL COLLEGE PRESS AND DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO., pp. 15–103.

97. Oriol,R. (1995) ABO, Hh, Lewis, and Secretion. In. Springer, Boston, MA, pp. 37–73.

98. Roberts,D.F. (1984) Blood Relations. Blood Groups and Anthropology BMJ Publishing Group.

99. Henry,S., Oriol,R. and Samuelsson,B. (1995) Lewis Histo-Blood Group System and Associated

Secretory Phenotypes. *Vox Sang.*, **69**, 166–182.

100. Aspholm-Hurtig,M., Dailide,G., Lahmann,M., Kalia,A., Ilver,D., Roche,N., Vikström,S., Sjöström,R., Lindén,S., Bäckström,A., *et al.* (2004) Functional adaptation of BabA the H. pylori ABO blood group antigen binding adhesin. *Science (80-.).*, **305**, 519–522.

101. Ilver,D., Arnqvist,A., Ögren,J., Frick,I.M., Kersulyte,D., Incecik,E.T., Berg,D.E., Covacci,A., Engstrand,L. and Borén,T. (1998) Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science (80-.).*, **279**, 373–377.

102. Moulds,J.M. and Moulds,J.J. (2000) Blood group associations with parasites, bacteria, and viruses. *Transfus. Med. Rev.*, **14**, 302–311.

103. Heggelund,J.E., Varrot,A., Imberty,A. and Krengel,U. (2017) Histo-blood group antigens as mediators of infections. *Curr. Opin. Struct. Biol.*, **44**, 190–200.

104. Cooling,L. (2015) Blood groups in infection and host susceptibility. *Clin. Microbiol. Rev.*, **28**, 801–870.

105. Dunne,C., Dolan,B. and Clyne,M. (2014) Factors that mediate colonization of the human stomach by Helicobacter pylori. *World J. Gastroenterol.*, **20**, 5610–5624.

106. Ilver,D., Arnqvist,A., Ogren,J., Frick,I.M., Kersulyte,D., Incecik,E.T., Berg,D.E., Covacci,A., Engstrand,L., Borén,T., *et al.* (1998) Helicobacter pylori Adhesin Binding Fucosylated Histo-Blood Group Antigens Revealed by Retagging. *Science (80-.).*, **279**, 373–377.

107. Magalhães,A., Marcos-Pinto,R., Nairn,A. V., dela Rosa,M., Ferreira,R.M., Junqueira-Neto,S., Freitas,D., Gomes,J., Oliveira,P., Santos,M.R., *et al.* (2015) Helicobacter pylori chronic infection and mucosal inflammation switches the human gastric glycosylation pathways. *Biochim. Biophys. Acta - Mol. Basis Dis.*, **1852**, 1928–1939.

108. Rossez,Y., Gosset,P., Boneca,I.G., Magalhães,A., Ecobichon,C., Reis,C.A., Cieniewski-Bernard,C., Curt,M.J.C., Léonard,R., Maes,E., *et al.* (2014) The lacdiNAc-specific adhesin LabA mediates adhesion of Helicobacter pylori to human gastric mucosa. *J. Infect. Dis.*, **210**, 1286–1295.

109. Ruiz-Palacios,G.M., Cervantes,L.E., Ramos,P., Chavez-Munguia,B. and Newburg,D.S. (2003) Campylobacter jejuni binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. *J. Biol. Chem.*, **278**,

14112–20.

110. Mahdavi,J., Pirinccioglu,N., Oldfield,N.J., Carlsohn,E., Stoof,J., Aslam,A., Self,T., Cawthraw,S.A., Petrovska,L. and Colborne,N. (2014) A novel O-linked glycan modulates Campylobacter jejuni major outer membrane protein-mediated adhesion to human histo-blood group antigens and chicken colonization. *Open Biol.*, **4**, 130202.

111. Jeannin,P., Bottazzi,B., Sironi,M., Doni,A., Rusnati,M., Presta,M., Maina,V., Magistrelli,G., Haeuw,J.F., Hoeffel,G., *et al.* (2005) Complexity and complementarity of outer membrane protein A recognition by cellular and humoral innate immunity receptors. *Immunity*, **22**, 551–560.

112. Humphries,H.E., Williams,J.N., Blackstone,R., Jolley,K.A., Yuen,H.M., Christodoulides,M. and Heckels,J.E. (2006) Multivalent liposome-based vaccines containing different serosubtypes of PorA protein induce cross-protective bactericidal immune responses against Neisseria meningitidis. *Vaccine*, **24**, 36–44.

113. Davenport,V., Guthrie,T., Findlow,J., Borrow,R., Williams,N.A. and Heyderman,R.S. (2003) Evidence for Naturally Acquired T Cell-Mediated Mucosal Immunity to Neisseria meningitidis . *J. Immunol.*, **171**, 4263–4270.

114. Hancock,R.E.W. (1987) Role of porins in outer membrane permeability. *J. Bacteriol.*, **169**, 929–933.

115. Huang,S., Sahin,O. and Zhang,Q. (2007) Infection-induced antibodies against the major outer membrane protein of *Campylobacter jejuni* mainly recognize conformational epitopes. *FEMS Microbiol. Lett.*, **272**, 137–143.

116. Dhanasekar,N.N., Aliouane,S., Winterhalter,M., Pagès,J.-M. and Bolla,J.-M. (2017) Peptide translocation across MOMP, the major outer membrane channel from Campylobacter jejuni. *Biochem. Biophys. Reports*, **11**, 79–83.

117. Bolla,J.M., Dé,E., Dorez,A. and Pagès,J.M. (2000) Purification, characterization and sequence analysis of Omp50,a new porin isolated from Campylobacter jejuni. *Biochem. J.*, **352 Pt 3**, 637–43.

118. Mamelli,L., Pagès,J.-M., Konkel,M.E. and Bolla,J.-M. (2006) Expression and purification of native and truncated forms of CadF, an outer membrane protein of Campylobacter. *Int. J.*

Biol. Macromol., **39**, 135–140.

119. Dedieu,L., Pagès,J.-M. and Bolla,J.-M. (2002) Environmental regulation of Campylobacter jejuni major outer membrane protein porin expression in Escherichia coli monitored by using green fluorescent protein. *Appl. Environ. Microbiol.*, **68**, 4209–15.

120. Ferrara,L.G.M., Wallat,G.D., Moynié,L., Dhanasekar,N.N., Aliouane,S., Acosta-Gutiérrez,S., Pagès,J.-M., Bolla,J.-M., Winterhalter,M. and Ceccarelli,M. (2016) MOMP from Campylobacter jejuni Is a Trimer of 18-Stranded β-Barrel Monomers with a Ca 2+ Ion Bound at the Constriction Zone. *J. Mol. Biol.*, **428**, 4528–4543.

121. Moser,I., Schroeder,W. and Salnikow,J. (1997) Campylobacter jejuni major outer membrane protein and a 59-kDa protein are involved in binding to fibronectin and INT 407 cell membranes. *FEMS Microbiol. Lett.*, **157**, 233–238.

122. Subedi,S., Moonens,K., Romaõ,E., Lo,A., Vandenbussche,G., Bugaytsova,J., Muyldermans,S., Borén,T. and Remaut,H. (2014) Expression, purification and X-ray crystallographic analysis of the Helicobacter pylori blood group antigen-binding adhesin BabA. *Acta Crystallogr. Sect. FStructural Biol. Commun.*, **70**, 1631–1635.

123. Advances in Protein and Peptide Sciences (2013) BENTHAM SCIENCE PUBLISHERS.

124. Tomasek,D., Rawson,S., Lee,J., Wzorek,J.S., Harrison,S.C., Li,Z. and Kahne,D. (2020) Structure of a nascent membrane protein as it folds on the BAM complex. *Nature*, **583**, 473–478.

125. Apweiler,R., Hermjakob,H. and Sharon,N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta - Gen. Subj.*, **1473**, 4–8.

126. Kelly,J., Jarrell,H., Millar,L., Tessier,L., Fiori,L.M., Lau,P.C., Allan,B. and Szymanski,C.M. (2006) Biosynthesis of the N-linked glycan in Campylobacter jejuni and addition onto protein through block transfer. *J. Bacteriol.*, **188**, 2427–34.

127. Szymanski,C., Nothaft,H., Amber,S. and Aebi,M. (2008) N-Linked Protein Glycosylation in Campylobacter. In *Campylobacter, Third Edition*. American Society of Microbiology, pp. 447–469.

128. Solá,R.J., Rodríguez-Martínez,J.A. and Griebenow,K. (2007) Modulation of protein biophysical properties by chemical glycosylation: Biochemical insights and biomedical implications. *Cell.*

Mol. Life Sci., **64**, 2133–2152.

129. Masárová,J., Mislovičová,D., Gemeiner,P. and Michalková,E. (2001) Stability enhancement of Escherichia coli penicillin G acylase by glycosylation with yeast mannan. *Biotechnol. Appl. Biochem.*, **34**, 127.

130. Sola,R.J. and Griebenow,K. (2009) Effects of glycosylate on the stability of protein pharmaceuticals. *J. Pharm. Sci.*, **98**, 1223–1245.

131. Thibault,P., Logan,S.M., Kelly,J.F., Brisson,J.-R., Ewing,C.P., Trust,T.J. and Guerry,P. (2001) Identification of the Carbohydrate Moieties and Glycosylation Motifs in *Campylobacter jejuni* Flagellin. *J. Biol. Chem.*, **276**, 34862–34870.

132. Nothaft,H. and Szymanski,C.M. (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.*, **8**, 765–778.

133. Schäffer,C., Graninger,M. and Messner,P. (2001) Prokaryotic glycosylation. *Proteomics*, **1**, 248–261.

134. Rossez,Y., Wolfson,E.B., Holmes,A., Gally,D.L. and Holden,N.J. (2015) Bacterial Flagella: Twist and Stick, or Dodge across the Kingdoms. *PLOS Pathog.*, **11**, e1004483.

135. Aarestrup,F.M., Nielsen,E.M., Madsen,M. and Engberg,J. (1997) Antimicrobial susceptibility patterns of thermophilic Campylobacter spp. from humans, pigs, cattle, and broilers in Denmark. *Antimicrob. Agents Chemother.*, **41**, 2244–2250.

136. Engberg,J., Aarestrup,F.M., Taylor,D.E., Gerner-Smidt,P. and Nachamkin,I. (2001) Quinolone and macrolide resistance in Campylobacter jejuni and C. coli: resistance mechanisms and trends in human isolates. *Emerg. Infect. Dis.*, **7**, 24.

137. Svetoch,E.A. and Stern,N.J. (2010) Bacteriocins to control Campylobacter spp. in poultry—a review. *Poult. Sci.*, **89**, 1763–1768.

138. Stern,N.J., Eruslanov,B. V, Pokhilenko,V.D., Kovalev,Y.N., Volodina,L.L., Perelygin,V. V, Mitsevich,E. V, Mitsevich,I.P., Borzenkov,V.N. and Levchuk,V.P. (2008) Bacteriocins reduce Campylobacter jejuni colonization while bacteria producing bacteriocins are ineffective. *Microb. Ecol. Health Dis.*, **20**, 74–79.

139. Cleveland,J., Montville,T.J., Nes,I.F. and Chikindas,M.L. (2001) Bacteriocins: safe, natural

antimicrobials for food preservation. *Int. J. Food Microbiol.*, **71**, 1–20.

140. Riley,M.A. and Wertz,J.E. (2002) Bacteriocins: evolution, ecology, and application. *Annu. Rev. Microbiol.*, **56**, 117–137.

141. Joerger,R.D. (2003) Alternatives to antibiotics: bacteriocins, antimicrobial peptides and bacteriophages. *Poult. Sci.*, **82**, 640–647.

142. Stern,N.J. and Robach,M.C. (2003) Enumeration of Campylobacter spp. in broiler feces and in corresponding processed carcasses. *J. Food Prot.*, **66**, 1557–1563.

143. van HELVOORT,T.O.N. and Summers,W.C. (2001) Félix d'Herelle and the Origins of Molecular Biology.

144. Carlton,R.M. (1999) Phage therapy: past history and future prospects. *Arch. Immunol. Ther. Exp. Ed.*, **47**, 267–274.

145. Wagenaar,J.A., Van Bergen,M.A.P., Mueller,M.A., Wassenaar,T.M. and Carlton,R.M. (2005) Phage therapy reduces Campylobacter jejuni colonization in broilers. *Vet. Microbiol.*, **109**, 275–283.

146. Fischer,S., Kittler,S., Klein,G. and Glünder,G. (2013) Impact of a single phage and a phage cocktail application in broilers on reduction of Campylobacter jejuni and development of resistance. *PLoS One*, **8**, e78543.

147. Pirisi,A. (2000) Phage therapy—advantages over antibiotics?

148. Ala'Aldeen,D., Mahdavi,J. and Soultanas,P. (2015) Antimicrobial compounds and compositions, and uses thereof.

149. Köster,W. ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12. *Res. Microbiol.*, **152**, 291–301.

150. Moynie,L., Ferrara,L., Tortajada,A. and Naismith,J. (2014) Structural studies of proteins involved in the activity of novel antibiotics. *Acta Crystallogr. Sect. A Found. Adv.*, **70**, C710–C710.

151. Singh,A., Upadhyay,V., Upadhyay,A.K., Singh,S.M. and Panda,A.K. (2015) Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microb. Cell Fact.*, **14**, 41.

152. Bolla,J.M., Dé,E., Dorez,A. and Pagès,J.M. (2000) Purification, characterization and sequence analysis of Omp50, a new porin isolated from Campylobacter jejuni. *Biochem. J.*, **352**, 637–643.

153. Wittig,I., Braun,H.P. and Schägger,H. (2006) Blue native PAGE. *Nat. Protoc.*, **1**, 418–428.

154. Pollock,N.L., Rai,M., Simon,K.S., Hesketh,S.J., Teo,A.C.K., Parmar,M., Sridhar,P., Collins,R., Lee,S.C., Stroud,Z.N., *et al.* (2019) SMA-PAGE: A new method to examine complexes of membrane proteins using SMALP nano-encapsulation and native gel electrophoresis. *Biochim. Biophys. Acta - Biomembr.*, **1861**, 1437–1445.

155. Greenfield,N.J. (2007) Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.*, **1**, 2876–2890.

156. Abou-Abbass,H., Abou-El-Hassan,H., Bahmad,H., Zibara,K., Zebian,A., Youssef,R., Ismail,J., Zhu,R., Zhou,S., Dong,X., *et al.* (2016) Glycosylation and other PTMs alterations in neurodegenerative diseases: Current status and future role in neurotrauma. *Electrophoresis*, **37**, 1549–1561.

157. Mulloy,B., Hart,G.W. and Stanley,P. (2009) Structural Analysis of Glycans Cold Spring Harbor Laboratory Press.

158. Han,L. and Costello,C.E. (2013) Mass spectrometry of glycans. *Biochem.*, **78**, 710–720.

159. Kim,M.S., Zhong,J. and Pandey,A. (2016) Common errors in mass spectrometry-based analysis of post-translational modifications. *Proteomics*, **16**, 700–714.

160. An,H.J., Froehlich,J.W. and Lebrilla,C.B. (2009) Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr. Opin. Chem. Biol.*, **13**, 421–426.

161. Brockhausen,I., Schachter,H. and Stanley,P. (2009) O-GalNAc Glycans Cold Spring Harbor Laboratory Press.

162. Carlsohn,E. and Nilsson,C.L. (2007) Proteomic Techniques for Functional Identification of Bacterial Adhesins. In *Lectins: Analytical Technologies*. Elsevier, pp. 299–325.

163. Parker,C.E., Mocanu,V., Mocanu,M., Dicheva,N. and Warren,M.R. (2010) Mass Spectrometry for Post-Translational Modifications CRC Press/Taylor & Francis.

164. Fricker,L.D. (2015) Limitations of Mass Spectrometry-Based Peptidomic Approaches. *J. Am.*

Soc. Mass Spectrom., **26**, 1981–1991.

165. Olsen,J. V., Ong,S.E. and Mann,M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics*, **3**, 608–614.

166. Tandem Mass Spectrometry (MS/MS) - MagLab.

167. Costello,C., enzymology,J.V.-M. in and 1990, undefined Tandem mass spectrometry of glycolipids. *Elsevier*.

168. Lodish,H., Berk,A., Zipursky,S.L., Matsudaira,P., Baltimore,D. and Darnell,J. (2000) Membrane Proteins.

169. Cournia,Z., Allen,T.W., Andricioaei,I., Antonny,B., Baum,D., Brannigan,G., Buchete,N.V., Deckman,J.T., Delemotte,L., del Val,C., *et al.* (2015) Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory. *J. Membr. Biol.*, **248**, 611–640.

170. Overington,J.P., Al-Lazikani,B. and Hopkins,A.L. (2006) How many drug targets are there? *Nat. Rev. Drug Discov.*, **5**, 993–996.

171. Yildirim,M.A., Goh,K. Il, Cusick,M.E., Barabási,A.L. and Vidal,M. (2007) Drug-target network. *Nat. Biotechnol.*, **25**, 1119–1126.

172. Whited,A.M. and Park,P.S.H. (2014) Atomic force microscopy: A multifaceted tool to study membrane proteins and their interactions with ligands. *Biochim. Biophys. Acta - Biomembr.*, **1838**, 56–68.

173. Dunn,M.F. (2010) Protein-Ligand Interactions: General Description. In *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd, Chichester, UK.

174. Sulaeman,S., Le Bihan,G., Rossero,A., Federighi,M., Dé,E. and Tresse,O. (2010) Comparison between the biofilm initiation of Campylobacter jejuni and Campylobacter coli strains to an inert surface using BioFilm Ring Test®. *J. Appl. Microbiol.*, **108**, 1303–1312.

175. Khattak,F., Paschalis,V., Green,M., Houdijk,G.M., Soultanas,P. and Mahdavi,J. (2018) TYPLEX® Chelate, a novel feed additive, inhibits Campylobacter jejuni biofilm formation and cecal colonization in broiler chickens. *Poult. Sci.*, **97**, 1391–1399.

176. Currie,D., Green,M., Dufailu,O.A., Matthaios,P., Soultanas,P., McCartney,E., Lester,H., Van Den Eede,L., Apajalahti,J. and Mahdavi,J. (2018) Dietary supplementation with ferric tyrosine

improves zootechnical performance and reduces caecal *Campylobacter* spp. load in broilers. *Br. Poult. Sci.*, **59**, 646–653.

177. de Mattos,L.C. (2016) Structural diversity and biological importance of ABO, H, Lewis and secretor histo-blood group carbohydrates. *Rev. Bras. Hematol. Hemoter.*, **38**, 331–340.

178. Brandão de Mattos,C.C. and de Mattos,L.C. (2017) Histo-blood group carbohydrates as facilitators for infection by Helicobacter pylori. *Infect. Genet. Evol.*, **53**, 167–174.

179. Li,J., Hsu,H.C., Mountz,J.D. and Allen,J.G. (2018) Unmasking Fucosylation: from Cell Adhesion to Immune System Regulation and Diseases. *Cell Chem. Biol.*, **25**, 499–512.

180. Quiocho,F.A. (1986) Carbohydrate-Binding Proteins: Tertiary Structures and Protein-Sugar Interactions. *Annu. Rev. Biochem.*, **55**, 287–315.

181. Wimmerová,M., Kozmon,S., Nečasová,I., Mishra,S.K., Komárek,J. and Koča,J. (2012) Stacking Interactions between Carbohydrate and Protein Quantified by Combination of Theoretical and Experimental Methods. *PLoS One*, **7**, e46032.

182. Hudson,K.L., Bartlett,G.J., Diehl,R.C., Agirre,J., Gallagher,T., Kiessling,L.L. and Woolfson,D.N. (2015) Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.*, **137**, 15152–15160.

183. Weis,W.I. and Drickamer,K. (1996) Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.*, **65**, 441–473.

184. JP,L., Y,F. and X,Z. (2012) Simultaneous measurement of 10,000 protein-ligand affinity constants using microarray-based kinetic constant assays. *Assay Drug Dev. Technol.*, **10**, 250–259.

185. United States Patent: 10793587.

186. Ilari,A. and Savino,C. (2008) Protein structure determination by X-ray crystallography. *Methods Mol. Biol.*, **452**, 63–87.

187. Hendrickson,W.A. (2016) Atomic-level analysis of membrane-protein structure. *Nat. Struct. Mol. Biol.*, **23**, 464–467.

188. Wallin,E. and Von Heijne,G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.*, **7**, 1029–1038.

189. Singh,P., Panchaud,A. and Goodlett,D.R. (2009) Chemical Cross-Linking and Mass Spectrometry As a Low-Resolution Protein Structure Determination Technique. *Mol. Cell. Proteomics*, **8**, 29.

190. Schmidt,C. and Robinson,C. V. (2014) Dynamic protein ligand interactions - Insights from MS. *FEBS J.*, **281**, 1950–1964.

191. Benesch,J.L.P. and Ruotolo,B.T. (2011) Mass spectrometry: Come of age for structural and dynamical biology. *Curr. Opin. Struct. Biol.*, **21**, 641–649.

192. Manzi,L., Barrow,A.S., Scott,D., Layfield,R., Wright,T.G., Moses,J.E. and Oldham,N.J. (2016) Carbene footprinting accurately maps binding sites in protein–ligand and protein–protein interactions. *Nat. Commun.*, **7**.

193. Manzi,L., Barrow,A.S., Hopper,J.T.S., Kaminska,R., Kleanthous,C., Robinson,C. V., Moses,J.E. and Oldham,N.J. (2017) Carbene Footprinting Reveals Binding Interfaces of a Multimeric Membrane-Spanning Protein. *Angew. Chemie Int. Ed.*, **56**, 14873–14877.

194. Nielsen,M.L., Savitski,M.M. and Zubarev,R.A. (2006) Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. *Mol. Cell. Proteomics*, **5**, 2384–2391.

195. Koebnik,R. (1999) Structural and functional roles of the surface-exposed loops of the β- barrel membrane protein OmpA from Escherichia coli. *J. Bacteriol.*, **181**, 3688–3694.

196. Shehu,A. and Kavraki,L.E. (2012) Modeling Structures and Motions of Loops in Protein Molecules. *Entropy*, **14**, 252–290.

197. Tran,Q.T., Maigre,L., D'Agostino,T., Ceccarelli,M., Winterhalter,M., Pagès,J.M. and Davin-Regli,A. (2017) Porin flexibility in Providencia stuartii: cell-surface-exposed loops L5 and L7 are markers of Providencia porin OmpPst1. *Res. Microbiol.*, **168**, 685–699.

198. Dé,E., Baslé,A., Jaquinod,M., Saint,N., Malléa,M., Molle,G. and Pagès,J.M. (2001) A new mechanism of antibiotic resistance in Enterobacteriaceae induced by a structural modification of the major porin. *Mol. Microbiol.*, **41**, 189–198.

199. McGann,M. (2011) FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.*, **51**, 578–596.

200. Skyner,R.E., McDonagh,J.L., Groom,C.R., Van Mourik,T. and Mitchell,J.B.O. (2015) A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Phys. Chem. Chem. Phys.*, **17**, 6174–6191.

201. Mcdonagh,J.L. (2015) COMPUTING THE AQUEOUS SOLUBILITY OF ORGANIC DRUG-LIKE MOLECULES AND UNDERSTANDING HYDROPHOBICITY University of St Andrews.

202. Fan,H., Irwin,J.J., Webb,B.M., Klebe,G., Shoichet,B.K. and Sali,A. (2009) Molecular docking screens using comparative models of proteins. *J. Chem. Inf. Model.*, **49**, 2512–2527.

203. Het Enyi,C. and Van Der Spoel,D. (2011) Toward prediction of functional protein pockets using blind docking and pocket search algorithms. 10.1002/pro.618.

204. Acosta-Gutierrez,S., Scorciapino,M.A., Bodrenko,I. and Ceccarelli,M. (2015) Filtering with Electric Field: The Case of E. coli Porins. *J. Phys. Chem. Lett.*, **6**, 1807–1812.

205. Agostillo,M., Jene,C., Boyle,T., Ramsland,P.A. and Yuriev,E. (2009) Molecular docking of carbohydrate ligands to antibodies: Structural validation against crystal structures. *J. Chem. Inf. Model.*, **49**, 2749–2760.

206. Laughrey,Z.R., Kiehna,S.E., Riemen,A.J. and Waters,M.L. Carbohydrate-π Interactions: What Are They Worth? 10.1021/ja803960x.

207. Kalali,B., Mejías-Luque,R., Javaheri,A. and Gerhard,M. (2014) H. pylori virulence factors: Influence on immune system and pathology. *Mediators Inflamm.*, **2014**.

208. STEGGERDA,F.R. (1955) The digestive system. *Annu. Rev. Physiol.*, **17**, 129–154.

209. Yang,J., Yang,F., Campos,L.S., Mansfield,W., Skelton,H., Hooks,Y. and Liu,P. (2017) Quenching autofluorescence in tissue immunofluorescence. *Wellcome Open Res.*, **2**, 79.

210. Karpishin,T. (2018) Reducing Tissue Autofluorescence. *Biotechniques*, **64**, 131.

211. Abegaz,S.B. (2021) Human ABO Blood Groups and Their Associations with Different Diseases. *Biomed Res. Int.*, **2021**, 6629060.

212. PG,F., L,B., J,H. and JI,G. (1995) Expression of a human alpha-1,3/4-fucosyltransferase in the pit cell lineage of FVB/N mouse stomach results in production of Leb-containing glycoconjugates: a potential transgenic mouse model for studying Helicobacter pylori infection. *Proc. Natl. Acad. Sci. U. S. A.*, **92**, 1515–1519.

213. Revez,J., Schott,T., Rossi,M. and Hänninen,M.L. (2012) Complete genome sequence of a variant of campylobacter jejuni NCTC 11168. *J. Bacteriol.*, **194**, 6298–6299.

214. Bellamy-Carter,J. and Oldham,N.J. (2019) PepFoot: A Software Package for Semiautomated Processing of Protein Footprinting Data. *J. Proteome Res.*, **18**, 2925–2930.

215. FRED Fast Rigid Exhaustive Docking (2008).

216. Arinaminpathy,Y., Khurana,E., Engelman,D.M. and Gerstein,M.B. (2009) Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov. Today*, **14**, 1130–1135.

217. Denisov,I.G., Grinkova,Y. V., Lazarides,A.A. and Sligar,S.G. (2004) Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *J. Am. Chem. Soc.*, **126**, 3477–3487.

218. Lee,S.C., Knowles,T.J., Postis,V.L.G., Jamshad,M., Parslow,R.A., Lin,Y.-P., Goldman,A., Sridhar,P., Overduin,M., Muench,S.P., *et al.* A Method for Detergent-free isolation of Membrane Protein with its Local Lipid Environment.

219. Jamshad,M., Charlton,J., Lin,Y.P., Routledge,S.J., Bawa,Z., Knowles,T.J., Overduin,M., Dekker,N., Dafforn,T.R., Bill,R.M., *et al.* (2015) G-protein coupled receptor solubilization and purification for biophysical analysis and functional studies, in the total absence of detergent. *Biosci. Rep.*, **35**.

220. Parker,C.E., Mocanu,V., Mocanu,M., Dicheva,N. and Warren,M.R. (2009) Mass spectrometry for post-translational modifications. In *Neuroproteomics*. CRC Press, pp. 93–113.

221. Cummings,R.D. and Etzler,M.E. (2009) Antibodies and Lectins in Glycan Analysis Cold Spring Harbor Laboratory Press.

222. Kobata,A. and Endo,T. (1992) Immobilized lectin columns: Useful tools for the fractionation and structural analysis of oligosaccharides. In *Journal of Chromatography*. J Chromatogr, Vol. 597, pp. 111–122.

223. Yu,H.W., Halonen,M.J. and Pepper,I.L. (2015) Immunological Methods. In *Environmental Microbiology: Third Edition*. Elsevier Inc., pp. 245–269.

224. Flanagan,R.C., Neal-McKinney,J.M., Dhillon,A.S., Miller,W.G. and Konkel,M.E. (2009)

Examination of Campylobacter jejuni putative adhesins leads to the identification of a new protein, designated FlpA, required for chicken colonization. *Infect. Immun.*, **77**, 2399–2407.

225. Horbach,S.P.J.M. and Halffman,W. (2017) The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. *PLoS One*, **12**, e0186281.

226. Contaminated cell lines throw thousands of studies into question.