Protein Glycosylation in the Gram-Negative Gammaproteobacterium, *Photorhabdus luminescens*

Thesis submitted for the degree of

Doctor of Philosophy

by

Mary Fox, B. Sc.

Supervised by

Michael O'Connell, B. Sc., Ph.D.

and

Brendan O'Connor, B. Sc., Ph.D.

School of Biotechnology

Dublin City University

Ireland

September 2011

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Degree of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____   ID No.:_____53014274_____

Date: _____

Acknowledgements

I would like to say a huge thank you to all of the members of my lab that have helped me so much during the last 4 years. Without you I would never have got through it! I want to give a special mention to three people in particular, Damien, Ruth and Vinny. They were with me from the very beginning. Damien, thanks for being a fountain of knowledge. I owe you big time for all of the advice you gave me and I learned so much from you. Thanks Vinny for taking me under your wing and teaching me so much and especially for making the lab so much fun. Ruth, it was great to have you there with me throughout it all from beginning to end. Thank you for being such a good friend to me. I'm so glad we're getting to work together for at least another two years too!

I want to thank my brilliant supervisors Mick and Brendan. I doubt there are two nicer people that I could have had to guide me along the way. Your advice and support has been invaluable.

I want to thank my friends and my family especially Mam and Da. It makes me so happy to know I have done you proud! Finally I want say thank you to my unbelievable boyfriend Stuart. When I was feeling down and felt like giving up you always gave me support and encouragement and believed in me. You're the best and I seriously would not have been able to do this without you.

Abbreviations

2-D	2-dimensional
AAL	*Aleuria aurantia* lectin
Ala	alanine
Amp^R	ampicillin resistance
APS	ammonium persulphate
Asn	asparagine
Asp	aspartic acid
Bac	bacillosamine
BCA	bicinchoninic acid
bp	base pair
BSA	bovine serum albumin
CCC	covalently closed circular
CDG	congenital disorders of glycosylation
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHCA	α-cyano-4-hydroxycinnamic acid
Cm^R	chloramphenicol resistance
Con A	concanavalin A
CP	capsular polysaccharide
DATDH	2,4-diacetamido-2,4,6-trideoxyhexose
DBA	*Dolichos biflorus* agglutinin
dH$_2$O	distilled H$_2$O
DHB	2,5-dihydroxy benzoic acid
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DSL	*Datura stramonium* lectin
DTT	dithiothreitol
ECL	*Erythrina cristagalli* lectin
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme linked immunosorbent assay
ELLA	enzyme linked lectin assay

Endo H_f	Endoglycosidase H_f
ER	endoplasmic reticulum
ESI	electrospray ionization
Fuc	fucose
FucNAc	*N*-acetylfucosamine
Gal	galactose
GalNAc	*N*-acetylgalactosamine
GalNAcT	*N*-acetylgalactosyl transferase
GGP	general glycosylation pathway
Glc	glucose
GlcA	glucuronic acid
GlcNAc	*N*-acetylglucosamine
Glu	glutamic acid
Gm^R	gentamicin resistance
GNL	*Galanthus nivalis* lectin
GPI	glycosylphosphatidylinositol
GSL I(I)	*Griffonia simplicifolia* lectin I(I)
GT	glycosyltransferase
HexNAc	*N*-acetylhexosamine
HMW1	high-molecular-weight adhesin 1
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
IdoA	iduronic acid
IEF	isoelectric focusing
Ile	isoleucine
IMAC	immobilised metal affinity chromatography
IPG	immobilised pH gradient
IPTG	isopropyl-β-D-1-thiogalactopyranoside
JCVI	J. Craig Venter Institute
(k)b	(kilo)base
(k)Da	(kilo)Dalton
Kan^R	kanamycin resistance
KDO	3-deoxy-oct-2-ulosonic-acid
LAC	lectin affinity chromatography

LB	Lysogeny broth
LCA	*Lens culinaris* agglutinin
Leu	leucine
LFA	*Limax flavus* agglutinin
LOS	lipooligosaccharide
LPS	lipopolysaccharide
MAL I(I)	*Maackia amurensis* lectin I(I)
MALDI	matrix assisted laser desorption ionization
Man	mannose
MCP	microchannel plate
MCS	multiple cloning site
Met	methionine
MHC	major histocompatibility complex
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MT	mannosyltransferase
NeuNAc	*N*-acetylneuraminic acid (sialic acid)
Ni-NTA	nickel-nitrilotriacetic acid
NMR	nuclear magnetic resonance
NPL	*Narcissus pseudonarcissus* lectin
O.D.	optical density
OC	open circular
ORF	open reading frame
OT	oligosaccharyltransferase
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pgl	protein glycosylation
PIPES	piperazine-N,N′-bis(2-ethanesulfonic acid)
PMF	peptide mass fingerprinting
PNA	peanut agglutinin
PNGaseF	*N*-glycosidase F
PTM	post translational modification

Q-TOF	quadrupole orthogonal acceleration time of flight
RCA	*Ricinus communis* agglutinin
RF	radio frequency
Rif^R	rifampicin resistance
RNase	ribonuclease
rpm	revolutions per minute
SBA	soybean agglutinin
SDS	sodium dodecyl sulphate
Ser	serine
S-layer	surface-layer
SNA	*Sambucus nigra* agglutinin
sp.	species
spp.	species (plural)
STT3	staurosporine- and temperature sensitive yeast protein 3
subsp.	subspecies
TBS(T)	tris buffered saline (with triton)
TDC	time to digital converter
TEMED	N,N,N′,N′-tetramethyl ethylenediamine
Tet^R	tetracycline resistance
TFA	trifluoroacetic acid
Thr	threonine
T_m	melting temperature
TMB	3,3′,5,5′-tetramethylbenzidine
Tris	tris (hyroxymethyl) amino methane
UDP	undecaprenyl pyrophosphate
UEA I	*Ulex europaeus* agglutinin I
UV	ultraviolet
v/v	volume per volume
Val	valine
w/v	weight per volume
WGA	wheat germ agglutinin
Xyl	xylose

Table of contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of tables

Chapter 5

Abstract

The objective of this research was to investigate the possibility that *Photorhabdus luminescens* produces glycoproteins and thus contains a protein glycosylation system. *P. luminescens* is a pathogen of insects and a symbiont of soil nematodes. Adhesion and invasion are very important in the life cycle of the organism and it is speculated that the bacteria may produce glycoproteins to facilitate infection of the host.

Proteins from *P. luminescens* were analysed using lectins for the presence of glycoproteins. Many potential glycoproteins were isolated using lectin affinity chromatography (LAC) and one such protein was identified conclusively using mass spectrometry to be an outer membrane porin, OmpN. No glycoproteins have previously been identified in *P. luminescens.* The *ompN* gene was cloned and expressed in both *Escherichia coli* and *P. luminescens*. It was found that OmpN purified from *P. luminescens* was capable of binding the lectins WGA and GSL I, while OmpN purified from *E. coli* was not.

Campylobacter jejuni is the most extensively studied bacteria in terms of protein glycosylation. It contains a *pgl* locus encoding enzymes involved in the glycosylation of many of its proteins. Orthologues of some Pgl proteins were discovered in *P. luminescens*. The gene that encodes one of these proteins, *wblK*, was mutated. LAC was used to examine the effect of this mutation on the glycoproteins produced by the organism. It was found that the glycosylation of at least two proteins was affected by the *wblK* mutation. These proteins were isolated by LAC using WGA agarose from the *wblK* mutant strain but not from the wildtype strain. One of the altered proteins was identified as plu3611. It was found to exist in at least two different glycoforms. The results provided evidence that *P. luminescens* is capable of glycosylating OmpN and plu3611 and so contains a protein glycosylation system.

Chapter 1

Introduction

1.1 Glycobiology and glycosylation

1.1.1 Glycobiology

Glycobiology is the study of the structure, biosynthesis and biology of carbohydrates i.e. sugar chains or glycans. This includes the study of glycoconjugates, glycan biosynthesis enzymes and glycan recognition proteins such as lectins. Glycobiology is one of the most rapidly growing fields in biological science. However, to date the current understanding of the area is relatively limited.

The complete set of sugars in an organism, whether they are free or present in more complex forms is known as its glycome. The glycome of an organism can be extremely complex compared to its proteome. Proteins are synthesised as an identical copy from an mRNA template encoded in the genome. However, glycan synthesis is not a template driven process but instead glycans are assembled through a series of individual enzymatic reactions linking together different monosaccharide units. Not only can different monosaccharides combine together to form different glycans but they can also be combined in various ways through different glycosidic linkages. Glycans can also be assembled in branched structures leading to many possible combinations of just a few monosaccharides (Taylor and Drickamer 2006). For example the number of theoretical combinations of three monosaccharides to produce a trisaccharide is 38,016 compared to just 64 possible permutations of four nucleotides in a three base codon (Laine 1997). Glycans therefore have an inherent capacity and potential for encoding enormous amounts of biological information. Another important point to note is that a glycoprotein can be modified with heterogeneous glycans. Glycoproteins with the same peptide chain but different glycan moieties are called glycoforms. Each glycoform can have different and distinct biological properties and characteristics (Taylor and Drickamer 2006, Campbell and Yarema 2005).

This degree of complexity makes it very difficult to assign functional roles to glycans. Many functions of glycans can be found by observing what occurs when glycans are no longer synthesised correctly by an organism. It is known that glycans are an essential

component of all living things. Incorrect glycan formation plays a role in many diseases in humans such as rheumatoid arthritis, Alzheimer's disease and certain cancers (Taylor and Drickamer 2006, Ambrosi et al. 2005). Diseases belonging to a group known as congenital disorders of glycosylation (CDG) are known to be directly caused by defects in the synthesis of glycans and are reviewed by Grünewald et al. (2002). The disorders are caused by a deficiency in enzymes involved in the glycan synthesis pathway. The disorders most commonly begin in infancy. The symptoms of CDG vary from stroke to psychomotor retardation to the fragile skin seen in progeroid syndrome. Often individuals affected by CDG do not survive infancy (Aebi and Hennet 2001). Congenital muscular dystrophy, a progressive disease that causes muscle degeneration, is a CDG where the protein α-dystroglycan which is normally modified with *O*-linked mannose is not correctly glycosylated (Nakamura et al. 2010).

Other known functions of glycans include stabilisation and solubilisation of proteins. They also play a vital role in many biological processes such as cell recognition and adhesion and cell signalling. They are known to provide structural components such as cell walls. (Taylor and Drickamer 2006).

1.1.2 Glycosylation

Glycosylation is a term used to describe the addition of a glycan to a protein or a lipid via a covalent glycosidic linkage. The sugar groups are synthesised in the organism by various sugar biosynthesis enzymes and then transferred to the growing sugar chain by enzymes called glycosyltransferases (GTs). When the desired glycan has been synthesised it is then further processed by various processing enzymes such as flippases, polymerases and chain length determinant enzymes to yield the functional glycolipid or glycoprotein (Samuel and Reeves 2003).

In eukaryotes both glycoproteins and glycolipids are found in the plasma membrane of the cell at its surface. The extracellular matrix surrounding eukaryotic cells comprises glycoproteins and they are also secreted in biological fluids such as serum (Taylor and Drickamer 2006). Around 70% of all proteins in eukaryotes are believed to be glycosylated (Apweiler et al. 1999, Dell et al. 2010). The majority of genes involved in

protein glycosylation in humans have been identified and account for about 2% of the total human genome (Campbell and Yarema 2005).
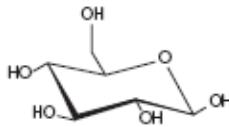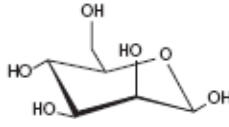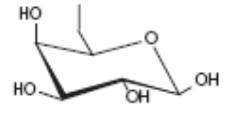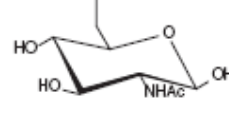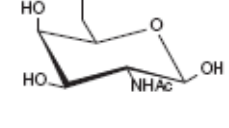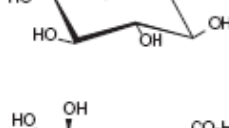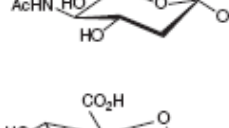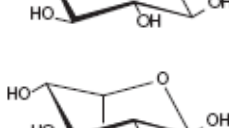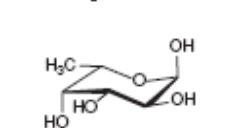
In bacterial systems, glycosylation occurs in lipopolysaccharide (LPS), lipooligosaccharide (LOS) and capsular polysaccharide (CP) biosynthesis pathways. It is now well known that many bacteria are also capable of producing glycoproteins and this will be discussed further in section 1.3.

There are five different classes of glycosylation, *N*-linked, *O*-linked, *P*-linked, *C*-linked and *G*-linked. All five classes occur in eukaryotes but so far only *N*-linked and *O*-linked protein glycosylation has been discovered in prokaryotes. In *P*-linked glycosylation phospho-glycans are linked to a protein via the phosphate of a phosphoserine. *C*-linked glycosylation occurs when glycans are attached to a carbon on a tryptophan side chain. *G*-linked glycosylation occurs in glycosylphosphatidylinositol (GPI)-anchored proteins. GPI is a glycolipid which acts as an anchor between cell surface proteins and the phospholipid bilayer in a eukaryotic cell membrane. A *G*-linkage is found between the glycan and the lipid in GPI (Pandhal and Wright 2010, Böhme and Cross 2002). *N*-linked and *O*-linked glycosylation occur most commonly and will now be discussed briefly in eukaryotic systems and in more depth in prokaryotic systems.

1.2 Eukaryotic protein glycosylation

Eukaryotic glycoprotein synthesis takes place in the rough endoplasmic reticulum (ER) and Golgi apparatus of the cell (Taylor and Drickamer 2006, Weerapana and Imperiali 2006). It is a highly complex process involving step by step enzymatic conversions and hundreds of gene products. The process is sensitive to the availability of enzymes, nutrients and nucleotide sugar donors, therefore the glycoproteins produced are highly dependent on the overall metabolic state of the cell (Taylor and Drickamer 2006, Marth and Grewal 2008). Despite this complexity, well defined pathways have been identified for the synthesis of *N*- and *O*-linked glycoproteins. Table 1.1 shows the most common monosaccharides found in eukaryotic *N*- and *O*-linked glycoproteins.

Table 1.1 Monosaccharides commonly found in eukaryotic glycoproteins

Monosaccharide	Abbreviation	3-D Chair Projection
β-D-Glucose	Glc	
β-D-Mannose	Man	
β-D-Galactose	Gal	
β-D-*N*-Acetylglucosamine	GlcNAc	
β-D-*N*-Acetylgalactosamine	GalNAc	
β-D-Xylose	Xyl	
α-*N*-Acetylneuraminic acid (sialic acid)	NeuNAc	
β-D-Glucuronic acid	GlcA	
α-L-Iduronic acid	IdoA	
α-L-Fucose	Fuc	

1.2.1 *N*-linked eukaryotic protein glycosylation

The *N*-linked protein glycosylation pathway in eukaryotes is the best understood glycosylation system. Eukaryotic *N*-linked glycans are synthesised through a common biosynthetic pathway and therefore all consist of a common pentasaccharide core structure of three mannose residues and two GlcNAc residues ($Man_3GlcNAc_2$). The glycan is attached to the protein via a β-glycosidic linkage between GlcNAc and the amide nitrogen (*N*-linked) of an asparagine residue within an Asn-X-Ser/Thr consensus sequence. The X in the Asn-X-Ser/Thr consensus sequence can be any residue apart from proline (Taylor and Drickamer 2006, Weerapana and Imperiali 2006, Geyer and Geyer 2006).

N-glycans can be differentiated into three groups according to the monosaccharides attached to the pentasaccharide core structure. The three groups are high mannose, hybrid and complex glycans (Geyer and Geyer 2006). Figure 1.1 shows an example of each.
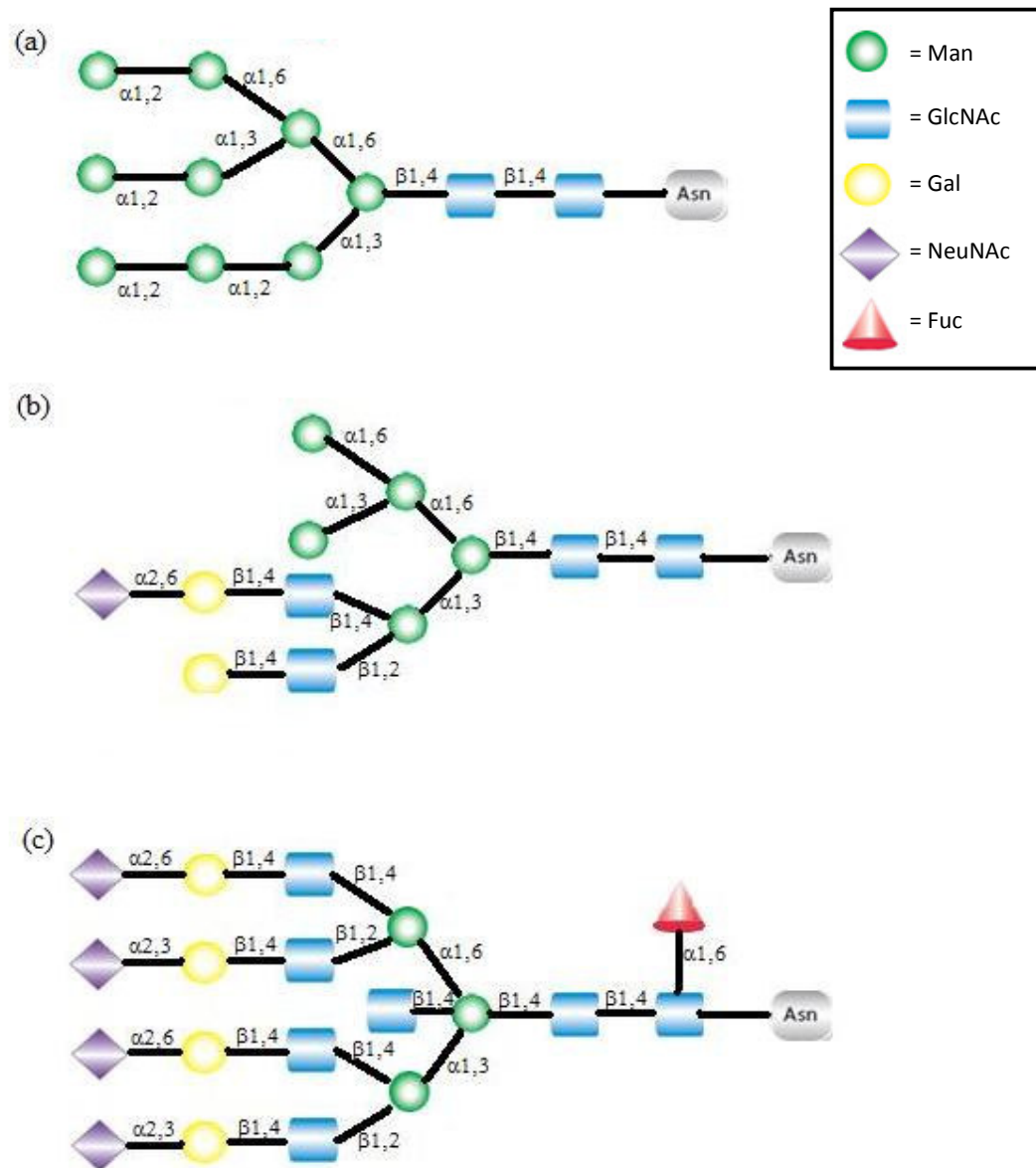
Figure 1.1 The three eukaryotic *N*-Linked glycan component classes

Examples of core mannose (a), hybrid (b) and complex (c) *N*-linked glycans (Taylor and Drickamer 2006, Geyer and Geyer 2006)**.**

Figure 1.1 shows that all eukaryotic *N*-linked glycans contain the $Man_3GlcNAc_2$ core pentasaccharide. Core mannose glycans typically contain between five and nine mannose residues. They contain unsubstituted terminal mannose residues (figure 1.1 (a)). Hybrid glycans contain both unsubstituted mannose residues and substituted mannose residues with GlcNAc (figure 1.1 (b)). In complex glycans, all terminal residues of a high mannose glycan have been replaced with GlcNAc. In complex glycans mannose residues are only present in the core pentasaccharide (figure 1.1 (c)).

7

GlcNAc structures added to the glycan core are called "antennae" therefore figure 1.1 (b) is a biantennary glycan and figure 1.1 (c) is a tetraantennary glycan. In both complex and hybrid glycans other monosaccharides can be transferred to the terminal GlcNAc of the growing glycan structure. Complex glycans usually terminate with sialic acid. Complex glycans can be further modified with a bisecting GlcNAc or a core fucose residue can be attached to the innermost GlcNAc residue (Taylor and Drickamer 2006, Geyer and Geyer 2006).

N-linked glycosylation in eukaryotes is a co-translational process that occurs in three stages. Firstly a precursor oligosaccharide is formed on a lipid carrier. Next the oligosaccharide is transferred to the polypeptide and finally the glycan is processed i.e. residues are removed and new ones added to yield the final glycoprotein (Taylor and Drickamer 2006).

The lipid carrier on which the precursor glycan is assembled is called dolichol and is found in the membrane of the ER. Seven sugar residues ($Man_5GlcNAc_2$) are attached to dolichol at the cytoplasmic side of the membrane of the ER. Single mannose and glucose residues are also bound to other dolichol molecules in the ER. The oligosaccharide/dolichol and monosaccharide/dolichol structures are flipped into the lumen of the ER. In the lumen, GTs transfer the single sugar residues from the dolichol molecule to the growing oligosaccharide. A tetradecasaccharide, $Glc_3Man_9GlcNAc_2$ (figure 1.2) is synthesised on dolichol. The synthesis of this dolichol linked tetradecasaccharide is known as the dolichol pathway (Weerapana and Imperiali 2006).

Glc₃Man₉GlcNAc₂β1,N-Asn
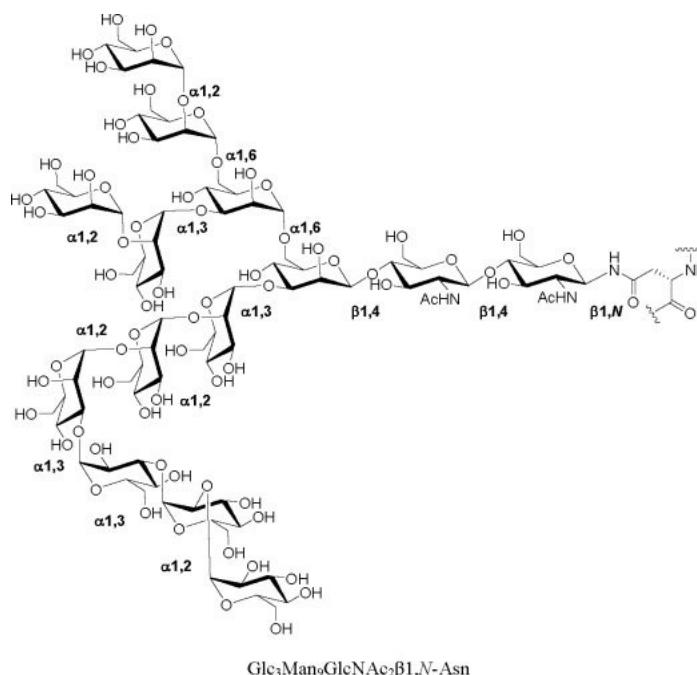
Figure 1.2 The tetradecasaccharide Glc$_3$Man$_9$GlcNAc$_2$

Structure of Glc$_3$Man$_9$GlcNAc$_2$, the glycan transferred to a polypeptide in eukaryotic *N*-linked glycosylation in the lumen of the ER. The glycan can then be further processed in the ER and Golgi apparatus to yield the final glycoprotein (Weerapana and Imperiali 2006).

Glc$_3$Man$_9$GlcNAc$_2$ is then transferred by an oligosaccharyltransferase (OT) to the polypeptide via a GlcNAc-Asn linkage. The protein is still being translated by the ribosome in the ER during this process. Released dolichol molecules are flipped back to face the cytoplasm (Taylor and Drickamer 2006, Weerapana and Imperiali 2006).

Still in the ER, the three glucose residues and one of the mannose residues are removed by glycosidases and the protein undergoes correct folding. The Man$_8$GlcNAc$_2$ bound protein is transferred to the Golgi apparatus. Here, through the action of various glycosidases and GTs the glycan is further processed to yield glycans similar to those shown in figure 1.1.

1.2.2 *O*-linked eukaryotic protein glycosylation

The most common type of *O*-linked glycosylation is where the glycan is attached to the protein through the oxygen atom (*O*-linked) of the hydroxyl group of a serine or threonine residue. There is no consensus sequence required for *O*-linked glycosylation. The most prevalent *O*-linked glycans in vertebrates are the mucin-type glycans. They contain an initial GalNAc residue α-glycosidically linked to serine or threonine (Taylor and Drickamer 2006, Geyer and Geyer 2006). Mucins are glycoproteins expressed in high abundance in epithelial cells such as the surface of the digestive and genital tracts and in the respiratory system. These surfaces are not sealed by moisture impermeable layers as is the case with the skin. Mucins serve to retain moisture, lubricate and protect from the invasion of microorganisms. They contain large numbers of clustered *O*-glycans especially sialylated glycans resulting in regions of strong negative charge allowing mucins to bind large amounts of water (Taylor and Drickamer 2006, Varki 1999). Mucin type glycans contain one of eight core structures. The four most common of these core structures are shown in figure 1.3.
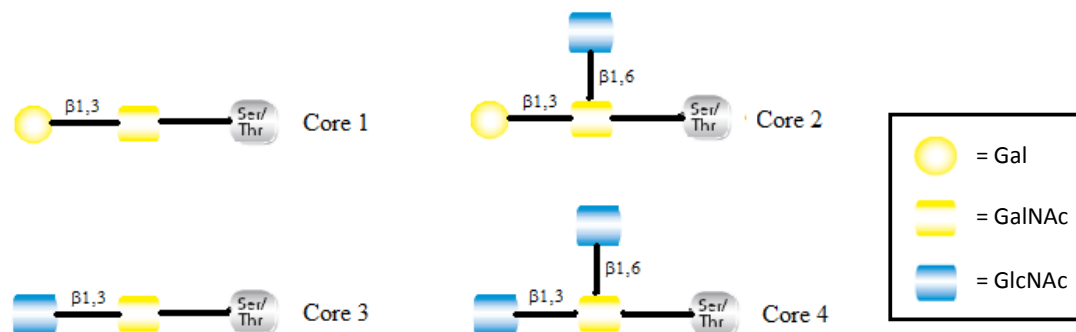
Figure 1.3 Common core structures of eukaryotic mucin-type *O*-linked glycans
The core structures all contain GalNAc linked to serine or threonine and additional Gal and/or GlcNAc residues.

Most mucin type *O*-glycans contain the core 1 or core 2 structure. Mucin type *O*-glycans vary dramatically in size and structure. They can contain merely a single residue bound to the polypeptide or the core structure can be extended to produce long chain glycans with variable termini that can be similar to the termini of *N*-glycans

(Lamblin et al. 2001). *O*-glycans are generally not as branched as *N*-glycans, with the most common structures being biantennary (Varki 1999).

Although mucins are the most common *O*-glycosylated proteins in vertebrates other *O*-glycans do occur. For example many nuclear and cytoplasmic proteins contain *O*-linked GlcNAc (Varki 1999). *O*-linked fucose and glucose are found in epidermal growth factor-like domains (Haltiwanger and Stanley 2002). *O*-linked mannose is found in dystroglycan and other nervous system glycoproteins (Endo 1999). Galactose *O*-linked to hydroxlysine is found in collagen (Anttinen and Hulkko 1980). Many proteins in yeast are *O*-glycosylated with mannose (Geyer and Geyer 2006, Varki 1999).

Because of the lack of a common protein-glycan linkage in *O*-linked glycosylation, unlike *N*-linked glycosylation, a common biosynthetic pathway can not be defined. As previously stated there are some terminal glycan structures common to both *N*-linked and *O*-linked glycans suggesting that there may be some enzymes common to both biosynthesis processes (Taylor and Drickamer 2006).

In contrast to *N*-linked glycosylation, *O*-linked glycosylation is a post-translational event that occurs in the Golgi apparatus. There is no precursor oligosaccharide transferred to the protein, instead all of the sugar units are directly covalently added, one at a time to the protein molecule. In mucin glycosylation this process begins with the addition of a GalNAc residue to a serine or threonine residue. The GalNAc is initially linked to an undecaprenyl pyrophosphate (UDP) lipid carrier and is transferred to the protein by UDP-*N*-acetylgalactosyl transferase (GalNAcT). In *N*-linked glycosylation the glycan is attached to the protein by a single OT, however in *O*-linked glycosylation numerous GalNAcT enzymes exist (there are at least twelve mammalian GalNAcT enzymes) that can transfer the first GalNAc. The enzyme used depends on the amino acid sequence of the protein. The glycan is subsequently elongated by several other GTs (Taylor and Drickamer 2006, Lamblin et al. 2001).

1.3 Prokaryotic protein glycosylation

Up until the 1970s many scientists believed that protein glycosylation was restricted to eukaryotic systems. This was believed to be the case because prokaryotes lack the cellular organelles required in eukaryotic protein glycosylation (the ER and Golgi apparatus) and also because of the relatively short life span of prokaryotes (Upreti et al. 2003). But now there is overwhelming evidence to prove that protein glycosylation does occur in prokaryotes and in fact due to the diversity of prokaryotic glycoproteins discovered in recent years, protein glycosylation in prokaryotes is now considered the norm rather than the exception. Protein glycosylation in prokaryotes appears to be even more complex than in eukaryotes. Prokaryotes are capable of synthesising a wider variety of sugars compared to eukaryotes and therefore the variations in prokaryotic glycoproteins exceed those found in eukaryotes (Benz and Schmidt 2002, Szymanski and Wren 2005, Schmidt et al. 2003).

So far only *N*-linked and *O*-linked protein glycosylation has been observed in prokaryotes. As in eukaryotes, with *N*-linked glycosylation the glycan is most commonly attached to an asparagine residue and in *O*-linked glycosylation the glycan is attached to a serine, threonine or tyrosine residue (Szymanski and Wren 2005, Hitchen and Dell 2006). In some bacterial species a distinct consensus sequence is required for *N*-glycosylation to occur (Nita-Lazar et al. 2005) which will be discussed later. Four distinct major glycosylation pathways have so far been discovered in prokaryotes. These are (i) OT mediated *N*-linked glycosylation, (ii) stepwise cytoplasmic *N*-linked glycosylation, (iii) OT mediated *O*-linked glycosylation and (iv) stepwise *O*-linked glycosylation (Dell et al. 2010). Type (i) is similar to the normal *N*-linked glycosylation found in eukaryotes (see section 1.2.1). This type of glycosylation is abundant in archaea but has only been discovered in certain bacterial species. Types (ii) and (iii) have so far only been found in the bacterial domain. Type (iv) is similar to the standard *O*-linked glycosylation pathway found in eukaryotes (section 1.2.2). This type of glycosylation is common in bacteria but very little of this type of glycosylation has been discovered in the archaeal domain (Dell et al. 2010).

Protein glycosylation has been most extensively studied in pathogenic bacteria (Szymanski and Wren 2005, Power and Jennings 2003). Many proteins associated with pathogenesis such as pili and flagella have been shown to be glycosylated. In recent years protein glycosylation has also been discovered in *Bacteroides* spp., one of the most abundant genera of commensals found in the human colon (Dell et al. 2010)(Schmidt et al. 2003). Although huge strides have been made in recent years in research of prokaryotic protein glycosylation, it is believed that the area has been largely unexplored. It is still not very well understood in many organisms and details of the glycosylation processes, the roles of glycosylation and the genes involved still remain to be discovered.

1.3.1 Glycosylation of prokaryotic S-layer proteins – a novel discovery

The earliest example of protein glycosylation in prokaryotes was found by Strominger and co-workers in the Gram-negative halophile *Halobacterium salinarum*, a member of the domain archaea (Upreti et al. 2003, Mescher et al. 1974, Mescher and Strominger 1976). A 194 kDa glycosylated surface-layer (S-layer) protein accounting for 40 to 50% of the total cell envelope proteins was found in this organism (Benz and Schmidt 2002, Mescher et al. 1974). Around the same time protein glycosylation was observed in the Gram-positive thermophiles *Clostridum thermosaccharolyticum* and *Clostridium thermohydrosulfuricum*, the first instance of protein glycosylation reported in eubacteria. They both contain a glycosylated S-layer protein with a molecular weight of 140 kDa (Sleytr and Thorne 1976).

Glycosylated S-layer proteins are commonly found in archaea and in Gram-positive bacteria. According to Schäffer and Messner (2004), up to 20% of total protein synthesis in these organisms may be devoted to the production of S-layer glycoproteins. To date only about 40 different S-layer glycoprotein structures have been elucidated but already the diversity of S-layer glycan structures and glycosidic linkages observed exceeds those in eukaryotic glycans (Weerapana and Imperiali 2006). In bacteria, linear or branched glycans with 20-50 identical repeating units are generally observed while in archaea the glycans tend to be short and do not contain repeating units. S-layer glycoproteins in Gram-positive bacteria are analogous to LPS in Gram-negative bacteria

with regard to their overall structure and constituent sugars. Some rare sugars found in LPS are also observed in S-layer glycoproteins in Gram-positive bacteria. The major difference between LPS and S-layer glycoproteins is that the lipid A in LPS is replaced with a protein in S-layer glycoproteins. *O*-linked glycans have been observed on both archaeal and bacterial S-layer glycoproteins while *N*-linked S-layer glycoproteins have only been seen in archaea (Schaffer and Messner 2004).

S-layer glycoproteins have a unique feature in that they form two-dimensional crystalline arrays on the cell wall of the organism. The cell surface is completely covered with S-layer glycoproteins and the carbohydrate moiety of the protein projects from the cell surface, as is also the case with LPS (Schaffer and Messner 2004, Whitfield 1995).

1.3.2 Protein glycosylation in bacteria

Since the discovery of glycosylated S-layer glycoproteins, many other glycosylated proteins have been discovered in important medically relevant pathogenic bacteria such as *Campylobacter* spp., *Neisseria* spp. and *Pseudomonas* spp. to name a few and recently in *Bacteroides* spp. (Dell et al. 2010). As previously stated four glycosylation pathways have been discovered in bacteria. These will now be discussed with examples representing each pathway.

1.3.2.1 Type (i): Oligosaccharyltransferase mediated *N*-linked glycosylation

As previously stated, it was discovered in 1976 that the S-layer protein from the archaeal prokaryote, *H. salinarum* contained asparagine linked glycans. *N*-glycosylation was then found to be common in the S-layers of many archaea. In the early 2000s much research was being carried out on the human pathogen *Campylobacter jejuni* and it was found that it was also capable of *N*-glycosylation (Young et al. 2002, Wacker et al. 2002). The pathway that *C. jejuni* uses to do this is very similar to those that eukaryotes and archaea use for *N*-linked protein glycosylation.

- ***N*-linked protein glycosylation in *C. jejuni***

C. jejuni is the best studied bacterium in terms of *N*-linked protein glycosylation. *C. jejuni* is a human-gut mucosal pathogen that is involved in enterocolitis, the symptoms of which are acute abdominal pain and inflammatory diarrhoea (Weerapana and Imperiali 2006, Ketley 1997). Because this organism is so problematic there has been a lot of interest in studying its pathogenic mechanisms.

Szymanski et al. (1999) provided evidence that a general protein glycosylation system exits in *C. jejuni*. In 2002, this general protein glycosylation system was shown to be involved in glycosylation of two glycoproteins, PEB3 and CgpA, The lectin Soybean agglutinin (SBA), which has affinity for GalNAc was used to purify putative glycoproteins from *C. jejuni* and, following trypsin digestion the two glycoproteins were identified using mass spectrometry (Linton et al. 2002). Shortly afterwards, Young et al. (2002) proved that the PEB3 glycan was *N*-linked. This was the first confirmed report of an *N*-linked protein glycosylation system in a bacterial system. To date over 65 proteins have been identified in *C. jejuni* that are glycosylated by the general *N*-linked glycosylation system (Nothaft and Szymanski 2010). Many of these proteins were also isolated using SBA (Power and Jennings 2003, Young et al. 2002, Feldman et al. 2005). The *N*-linked glycan structure modifying these proteins was determined by mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy to be a heptasaccharide consisting of GalNAc-α1,4-[Glcβ1,3-]Gal-NAc-α1,4-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-β1,*N*-Asn-X, where Bac is bacillosamine (2,4-diacetamido-2,4,6-trideoxyglucopyranose) (Young et al. 2002).

In *C. jejuni* it appears that a specific consensus sequence is needed for *N*-glycosylation to occur. Like in eukaryotic systems the glycan is attached to an asparagine from the sequence Asn-X-Ser/Thr (Hitchen and Dell 2006, Nita-Lazar et al. 2005, Young et al. 2002). Recently it has been discovered that the *N*-glycosylation consensus sequence in *C. jejuni* also requires a glutamic acid or aspartic acid residue upstream of the Asn-X-Ser/Thr residues making the extended consensus sequence Glu/Asp-Z-Asn-X-Ser/Thr where neither Z nor X can be proline (Kowarik et al. 2006b).

The genes involved in this *C. jejuni N*-glycosylation pathway are found in a 17 kb gene locus called the '*pgl* gene cluster'. It contains twelve genes called *pglA*, *pglB* etc. to *pglK and galE* (Power and Jennings 2003, Szymanski et al. 2003, Linton et al. 2005).

These genes are significantly homologous to the genes encoding enzymes involved in bacterial LPS and CP biosynthesis and were originally thought to be involved in the biosynthesis of LPS and CP. However, mutagenesis of these genes did not affect LPS or CP biosynthesis suggesting that the proteins encoded by the *pgl* gene cluster functioned independently from the known LPS and CP biosynthetic pathways (Weerapana and Imperiali 2006, Szymanski et al. 1999). The *pgl* gene locus is highly conserved among strains of *C. jejuni* and *Campylobacter coli* (Power and Jennings 2003).

According to Szymanski (2003) the key enzyme in the *pgl* gene locus is PglB. This enzyme is very similar to the staurosporine- and temperature sensitive yeast protein 3 (STT3) subunit of the *N*-linked OT complex of *Saccharomyces cerevisiae*. STT3 is important for recognition and/or catalysis in the OT complex. There are many orthologues to STT3 in eukaryotes and archaea. In archaea the orthologue is called AglB (Abu-Qarn et al. 2008). STT3 proteins are large proteins (700-970 amino acids), contain 10-12 membrane-spanning regions. STT3 orthologues have a conserved C-terminal catalytic motif (WWDYG) of which the second and third amino acids are essential. These properties are seen in PglB and AglB suggesting they act as OTs. It is believed to be responsible for the transfer of the oligosaccharide to the Asn residue. *Campylobacter* was the first bacterial genus to be shown to have an OT (Szymanski and Wren 2005, Nita-Lazar et al. 2005, Power and Jennings 2003, Szymanski et al. 2003, Wacker et al. 2006).

A proposed model for the biosynthesis of *N*-linked glycoproteins in *C. jejuni* is outlined in Szymanski (2003) and shown in figure 1.4.
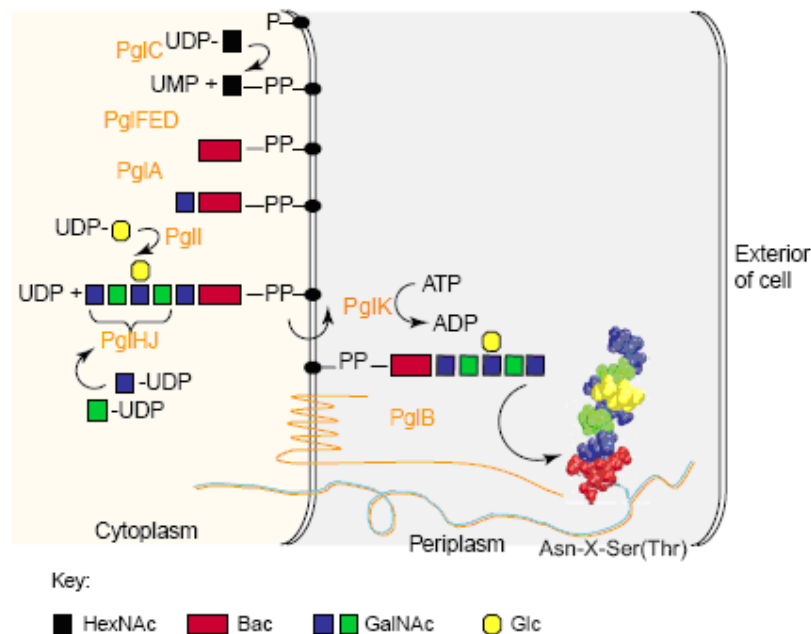
Figure 1.4 The *N*-linked protein glycosylation system in *C. jejuni*

A proposed model for the biosynthesis of *N*-linked glycoproteins in *C. jejuni* (Szymanski et al. 2003).

In this model it is proposed that nucleotide activated sugars are sequentially assembled on a lipid carrier, UDP, the same lipid carrier that is used in eukaryotic *O*-linked protein glycosylation. This occurs at the cytoplasmic side of the membrane (Feldman et al. 2005). The first residue, GlcNAc is attached to the lipid carrier by PglC. It is then sequentially modified by PglF (a dehydratase), PglE (an aminotransferase) and PglD (an acetyltransferase) to form bacillosamine. PglA is a GT that is believed to add the α-1,3-linked GalNAc to bacillosamine. PglH and PglJ are GTs that are both believed to be involved in transferring the next four α-1,4-linked GalNAc residues. The final GT, PglI adds the only branched residue, glucose. It is unsure whether this addition of the glucose occurs in sequence or after all five GalNAc residues have been added. The UDP-linked heptasaccharide is shown in figure 1.5. Once the lipid-linked heptasaccharide is assembled, it is flipped across the inner membrane into the periplasm by PglK, an ABC transporter. PglB, the most important enzyme in the process, the OT, transfers the glycan to the asparagine residue of the Glu/Asp-Z-Asn-X-Ser/Thr consensus sequence (Upreti et al. 2003, Szymanski et al. 2003, Weerapana et al. 2005, Karlyshev et al. 2005a).
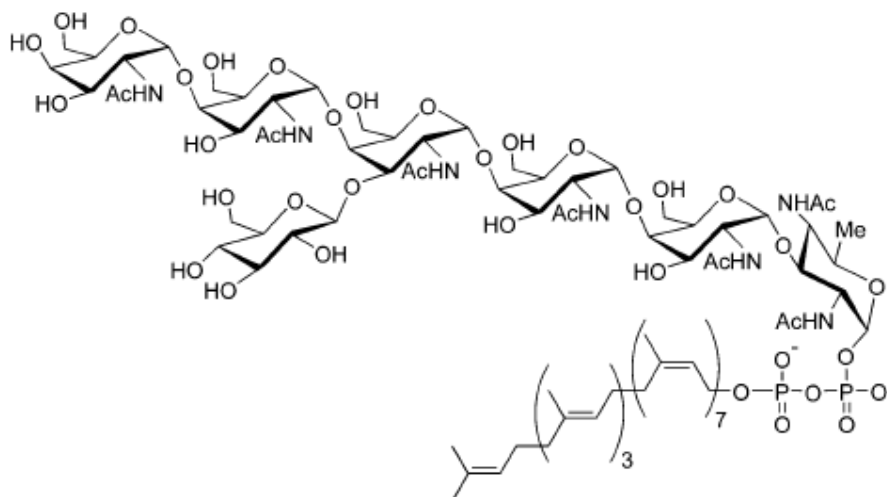
Figure 1.5 The *C. jejuni* UPD-linked heptasaccharide

The UDP-linked heptasaccharide, GalNAc-α1,4-[Glcβ1,3-]Gal-NAc-α1,4-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-α1-UDP. The heptasaccharide is the glycan donor in *N*-linked protein glycosylation in *C. jejuni* catalyzed by the *pgl* gene cluster (Weerapana et al. 2005).

There are some key differences between type (i) OT mediated *N*-linked glycosylation in prokaryotes compared to eukaryotes. In prokaryotes transfer of the oligosaccharide by OT takes place in the periplasm (in bacteria) or on the cell surface (in archaea). Because of this the protein must be fully folded, since protein folding occurs in the cytoplasm. In contrast, in eukaryotes oligosaccharides are transferred to nascent proteins which have not yet been folded. Another difference is that as yet there is no evidence to suggest that the oligosaccharide sequence seen in prokaryotes is conserved as is the conserved $Glc_3Man_9GlcNAc_2$ sequence found in all higher eukaryotes. Finally, unlike in eukaryotes, prokaryotic glycans do not appear to be further processed following flipping by PglK (Dell et al. 2010).

- *N*-linked protein glycosylation in other Epsilonproteobacteria

OT genes are ubiquitous in the eukaryotic and archaeal domains and consequently *N*-linked protein glycosylation is very common in both of these domains. However, very few bacteria actually contain OT genes and so only a limited number of bacterial species undergo *N*-linked protein glycosylation (Nothaft and Szymanski 2010). When *pglB*, the first known bacterial OT gene, was discovered in *C. jejuni*, it was thought that

its occurrence was unique in the bacterial domain. Scientists postulated that *C. jejuni* acquired this gene through lateral gene transfer from either archaea or eukarya. However, recent searches of bacterial genomes have uncovered that other bacteria from the epsilon subdivision of proteobacteria (of which the *Campylobacter* genus is a member) also contain *pglB* genes. Some of these bacteria include *Wolinella succinogenes* DSM 1740, *Sulfurovum* sp. NBC37-1, *Nitratirutor* sp. SB155-2, *Caminibacter mediatlanticus* TB-2, *Helicobacter pullorum* and *Desulfovibrio desulfuricans* subsp. *desulfuricans* G20 (Dell et al. 2010, Nothaft and Szymanski 2010). The structures of the *N*-linked glycans modifying *H. pullorum* and *W. succinogenes* glycoproteins have been elucidated using mass spectrometry and are shown in figure 1.6 along with the *C. jejuni N*-linked heptasaccharide.
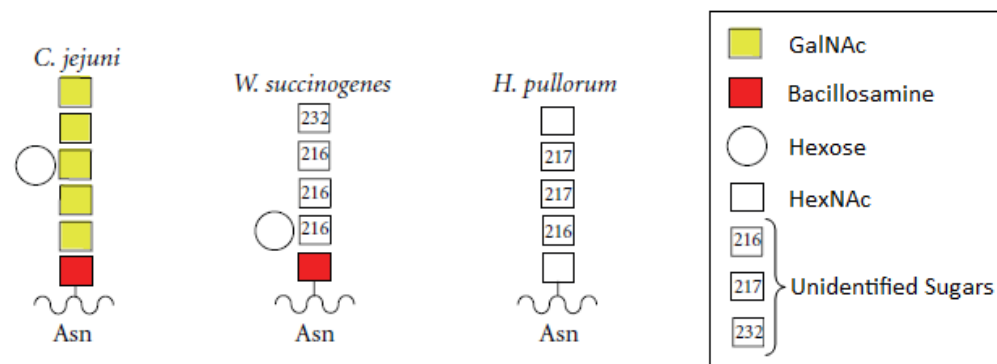
Figure 1.6 Structural representation of bacterial *N*-linked glycans

Like in *C. jejuni*, the glycan modifying *W. succinogenes N*-linked glycoproteins contains bacillosamine as its first residue. It is a hexasaccharide containing one branched hexose. In *H. pullorum* the first sugar of the pentasaccharide is an *N*-Acetylhexosamine (HexNAc). It contains no branched sugars. Some of the sugar units belonging to the *W. succinogenes* and *H. pullorum* glycans have not yet been identified (Dell et al. 2010).

Because the machinery for *N*-linked protein glycosylation in bacteria has so far only been found in Epsilonproteobacteria, some scientists now believe that the OT gene was transferred to them from archaea. This is because in primordial deep sea vents, where many archaea are found, the majority of bacteria living there are Epsilonproteobacteria (Dell et al. 2010).

1.3.2.2 Type (ii): Stepwise cytoplasmic *N*-linked glycosylation

In 2008 a novel *N*-linked protein glycosylation system was discovered by Gross and co-workers in the Gram-negative Gammaproteobacterium *Haemophilus influenzae* that involved sequential transfer of sugars to the glycoprotein rather than the block transfer of an oligosaccharide seen in type (i) (see section 1.3.2.1) (Gross et al. 2008). The modification is the only one of this kind to be discovered so far and occurs on the high-molecular-weight adhesin 1 (HMW1) protein. HMW1 adhesin is modified at 31 Asn residues. All but one of the Asn residues that are modified are part of the consensus sequence Asn-X-Ser/Thr. The modifications that occur are either monohexoses or dihexoses with the hexoses being Gal or Glc. The enzyme responsible for stepwise *N*-linked glycosylation in this organism is called HMW1C. This appears to be the only enzyme necessary for this process. It acts as an OT, transferring Glc/Gal to asparagine residues but also has a GT function as it generates the hexose-hexose bonds when the modification is a dihexose. As yet the mechanism for this pathway has not been defined. Orthologues to this protein have been found in other bacterial species and so it is likely the *H. influenzae* is not the only bacterium capable of stepwise cytoplasmic *N*-linked glycosylation (Dell et al. 2010, Nothaft and Szymanski 2010, Gross et al. 2008).

1.3.2.3 Type (iii): OT mediated *O*-linked glycosylation

In recent years a general *O*-linked protein glycosylation system has been discovered in some bacterial species. This system is very similar to the type (i) pathway mentioned earlier (section 1.3.2.1). It was first discovered in pathogenic bacteria. The pili of *Neisseria* spp. and *Pseudomonas aeruginosa* were shown to be *O*-glycosylated by this mechanism. However, even more recently this type of glycosylation has been observed in *Bacteroides* spp. This type of glycosylation has only been discovered in Gram-negative bacteria. This does not come as a surprise since the pathway is analogous to LPS biosynthesis (Dell et al. 2010).

- **OT mediated *O*-linked glycosylation of *Neisseria* proteins**

Neisseria bacteria are members of the class Betaproteobacteria and are Gram-negative diplococci. Two very important species of *Neisseria* are *Neisseria. gonorrhoeae* and *Neisseria meningitidis*. Both of these organisms are pathogenic. *N. meningitidis* is one of the most common causes of bacterial meningitis and also causes meningococcal septicaemia. *N. gonorrhoeae* is the causative agent of the human disease gonorrhoea (Power and Jennings 2003). It has been shown that these species are capable of glycosylating their pili with *O*-linked glycans. Pili are found in many Gram-negative pathogenic bacteria. They are fibrous polymeric proteins composed primarily of thousands of pilin subunits (Power et al. 2000). Pili protrude from the bacterial surface and are utilised by the organism for adhesion to and colonisation of the host cell (Power et al. 2000, Virji et al. 1991).

The structure and nature of pilin glycosylation in *N. meningitidis* was first determined in strain C311. It was shown that the pilin was glycosylated at serine 63. The glycan molecule was an unusual molecule, *O*-linked Gal(β1,4)-Gal(α1,3)-2,4-diacetamido-2,4,6-trideoxyhexose (Gal(β1,4)-Gal(α1,3)-DATDH) (Power et al. 2000, Stimson et al. 1995). *N. gonorrhoeae* strain MS11 is also glycosylated at the same region of its pilin molecules, however the modification is not the same. In *N. gonorrhoeae* the glycan is an *O*-linked disaccharide Gal(α1,3)-GlcNAc (Power et al. 2000, Parge et al. 1995). It has also been reported that this disaccharide rather than the trisaccharide may also be found in certain strains of *N. meningitidis* (Power et al. 2000, Marceau et al. 1998). Figures 1.7 and 1.8 show the structures of these glycans.
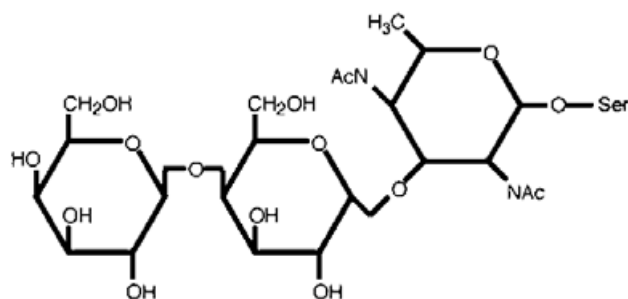
Figure 1.7 Gal(β1,4)-Gal(α1,3)-DATDH

The structure of the glycan Gal(β1,4)-Gal(α1,3)-DATDH which modifies *N. meningitidis* strain C311 pili at serine 63 of the pilin subunit (Power and Jennings 2003).
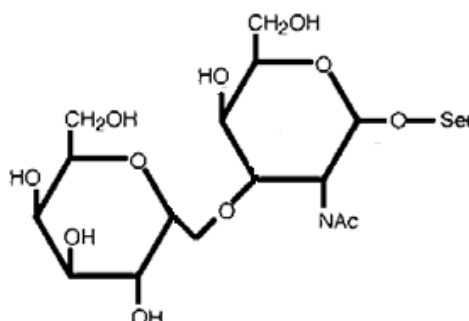
Figure 1.8 Gal(α1,3)GlcNAc

The structure of the glycan Gal(α1,3)GlcNAc which modifies *N. gonorrhoeae* strain MS11 pili at serine 63 of the pilin subunit (Power and Jennings 2003).

Jennings et al. (1998) identified the first gene involved in pilin glycosylation in *N. meningitidis*, called *pglA*. PglA is homologous to type 1 GTs and is proposed to transfer galactose to DATDH via the α1,3 linkage. PglA also catalyses the addition of the Gal(α1,3) residue in the disaccharide found in *N. gonorrhoeae* strain MS11 (Power and Jennings 2003). After the discovery of *pglA*, more *pgl* genes that were responsible for *O*-linked pilin glycosylation in *Neisseria* were found and it was shown that they were clustered together on a locus called the *pgl* gene locus. The *pgl* pilin glycosylation locus has been studied in several strains of *N. meningitidis,* for example strain NMB (Kahler et al. 2001) and strain C311#3 (Power et al. 2003a). The locus is polymorphic, containing between four and nine genes depending on the strain (Power and Jennings 2003). The locus in strain C311#3 contains four genes. Included in these four genes is a gene encoding PglF and three genes encoding proteins involved in the biosynthesis and transfer of DATDH, PglC (an aminotransferase), PglD (a dehydratase) and PglB

(believed to be bifunctional with acetyltransferase and GT domains) (Power et al. 2003a). PglF is a membrane spanning protein and is believed to be a flippase whose function is to flip the UDP-linked sugar from the cytoplasmic to the periplasmic side of the inner membrane (Power et al. 2006).

Polymorphisms that have been found include a 2 kb insertion that has been added to the locus. The insertion contains two additional genes called *pglG* and *pglH* that are homologous to GTs. An insertion-deletion in *pglB* is another polymorphism that has been found in some strains (Power and Jennings 2003, Kahler et al. 2001). There have also been two further genes found that are involved in the biosynthesis of the trisaccharide. However, these genes are not found in the glycosylation locus. PglI is believed to be an acetyltransferase involved in biosynthesis of DATDH and PglE is a type 2 GT that transfers the galactose to the trisaccharide. *pglE* contains many copies of a heptanucleotide repeat CAAACAA which mediates phase variation allowing the switching between expression of the disaccharide and trisaccharide structures in the organism at different stages of its life cycle (Power et al. 2003a).

Some of the proteins encoded by genes from the *pgl* locus in *N. meningitidis* display significant homology to Pgl proteins involved in the *N*-linked protein glycosylation pathway in *C. jejuni*. For example PglA from *N. meningitidis* displays 52% sequence similarity with PglA from *C. jejuni* subsp. jejuni 81-176 over the entire protein sequence. PglB from *N. meningitidis* displays 68% sequence similarity with PglC from *C. jejuni* over 47% of the protein sequence. PglC from *N. meningitidis* displays 52% sequence similarity with PglE from *C. jejuni* over 70% of the protein sequence. PglD from *N. meningitidis* displays 54% sequence similarity with PglF from *C. jejuni* over the entire protein sequence. Homology searches were carried out using BLASTP from NCBI (see section 2.12). Research has shown that the OT mediated *O*-linked glycosylation pathway in *Neisseria* has similarities to the OT mediated *N*-linked protein glycosylation pathway in *C. jejuni*. In each case the glycan is assembled on a lipid carried, flipped across the cytoplasmic membrane and then transferred to the protein by an OT enzyme (Abu-Qarn et al. 2008).

Bioinformatic and site-directed mutagenesis studies led to the discovery of the OT gene in *Neisseria* spp. It is called *pglL* in *N. meningitidis* and *pglO* in *N. gonorrhoeae*

23

(Nothaft and Szymanski 2010, Power et al. 2006, Faridmoayer et al. 2007). PglL/PglO OTs are responsible for the glycosylation of type IV pili, not just in *Neisseria* species but in a wide range of pathogenic bacteria. In 2009 it was discovered that in *Neisseria* this general *O*-linked protein glycosylation system was utilised to not only decorate pilin proteins with glycans but also a broad range of other proteins produced by the organism (Vik et al. 2009).

- **OT mediated *O*-linked glycosylation of *P. aeruginosa* pili**

P. aeruginosa is a Gram-negative opportunistic pathogen of humans and plants. It is the leading cause of nosocomial pneumonia in immuno-compromised individuals. The organism produces pili to aid in infection by allowing it to adhere to and colonize the cells and they also mediate motility across surfaces (Horzempa et al. 2006, Mattick 2002). *P. aeruginosa* strain 1244 pili are modified by glycosylation. The modification is an attachment of the trisaccharide α5NβOHC$_4$7NFmPse (β2,4)-Xyl (β1,3)-FucNAc (β1,3)- to the N-terminal serine residue (*O*-linked glycosylation) (Power and Jennings 2003, Castric 1995). Comer et al. (2002) demonstrated that the pilin glycosylation site was residue 148 of the protein. Figure 1.9 shows the structure of this trisaccharide.
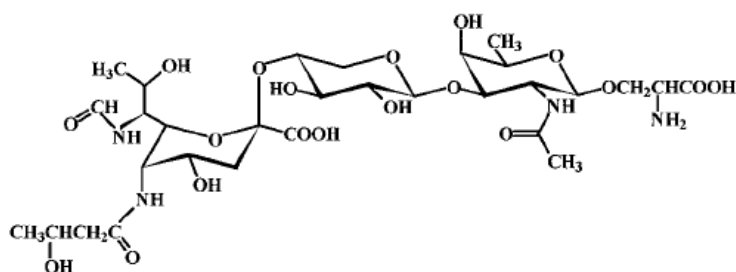
Figure 1.9 α5NβOHC$_4$7NFmPse (β2,4)-Xyl (β1,3)-FucNAc (β1,3)-
Structure of the glycan α5NβOHC$_4$7NFmPse (β2,4)-Xyl (β 1,3)-FucNAc (β 1,3)-, which glycosylates *P. aeruginosa* strain 1244 pili (Castric et al. 2001).

This trisaccharide is identical to the O-antigen repeating unit found in the LPS produced by this strain. This suggests that there is a common origin for the pathways involved in the biosynthesis of O-antigen and the glycosylation of pili in this organism (Castric et al. 2001). This has been proved by the fact that in mutants of strain 1244, where O-

antigen biosynthesis is defective, glycosylated pili are also not made (DiGiandomenico et al. 2002, Smedley III et al. 2005).

Castric (1995) identified the genes involved in the synthesis of the trisaccharide and also another gene, *pilO,* which is essential for pilin glycosylation. PilO is a membrane protein and the transmembrane profile of PilO shows that it contains 13 membrane spanning regions. PilO is homologous to PglL/PglO in *Neisseria* species. PilO is not present in *P. aeruginosa* strains that do not produce glycosylated pili (Power and Jennings 2003). The *pilO* gene is found in the same operon as the pilin structural gene, *pilA*. It encodes an OT that transfers the O-antigen repeating unit to the pilin (Smedley III et al. 2005). When this operon was cloned into a *P. aeruginosa* strain that normally does not glycosylate its pili, it produced glycosylated pili. Also when *pilO* is mutated in strain 1244, it results in non-glycosylated pili proving that PilO is the only other glycosylation factor required for pilin glycosylation besides those genes also involved in the O-antigen biosynthesis pathway (DiGiandomenico et al. 2002, Smedley III et al. 2005).

- **General OT mediated *O*-linked glycosylation in *Bacteroides* spp.**

Bacteroides are the most abundant genus of commensal bacteria found in the human intestine. Members of this genus produce a huge array of glycan structures including different CPs, some of which contain terminal fucose residues (Krinos et al. 2001). Further investigation into fucose utilization by these bacteria showed that it uses exogeneous L-fucose from the host and incorporates it not only into its CP but also into oligosaccharides that are bound to proteins (Nothaft and Szymanski 2010, Krinos et al. 2001).

A consensus sequence for *O*-linked glycosylation was determined in *Bacteroides fragilis* as Asp-Ser/Thr-Ala/Ile/Leu/Val/Met/Thr where the final amino acid required a methyl group. The gene locus responsible for *O*-linked glycosylation in *Bacteroides fragilis* was discovered and was termed *lfg* (Fletcher et al. 2009). It contained a putative flippase gene (*wzx*) and five putative GT genes. However no candidate gene has been identified so far to be the putative OT. Other *Bacteroides* spp. were shown to contain

similar genetic loci and also produce *O*-linked glycoproteins (Fletcher et al. 2009). A pathway for *O*-linked glycosylation in *Bacteroides* has not yet been defined but it is proposed that the GTs are responsible for assembling the oligosaccharide on a lipid carrier on the cytoplasmic face of the inner membrane and Wzx flips the lipid linked oligosaccharide into the periplasm where the oligosaccharide is then attached to the protein at its consensus sequence (Dell et al. 2010).

O-fucosylated fimbriae have been discovered in *Porphyromonas gingivalis*. This is not a human commensal but in fact an oral mucosal pathogen. However, it is a member of the order Bacteroidales, the same order to which *Bacteroides* spp. belong. So far the glycan sequences of its glycoproteins have not been fully determined but initial analysis has shown that they are likely to be quite complex. As well as fucose, a range of other monosaccharides having already been detected (Zeituni et al. 2010).

1.3.2.4 Type (iv): Stepwise *O*-linked glycosylation

In eukaryotes all *O*-linked glycosylation is stepwise or progressive. The first step in the process is the attachment of the initial monosaccharide to the protein via a serine or threonine residue. Stepwise *O*-linked glycosylation has also been well characterised in some bacterial species especially the flagellar glycosylation systems of *Campylobacter* spp., *Clostridium* spp. and *Pseudomonas* spp.

- ### *C. jejuni* flagella glycosylation

One of the earliest reports of protein glycosylation in *C. jejuni* was described by Doig et al. (1996). They concluded that *C. jejuni* flagella were glycosylated because they were sensitive to periodate oxidation (a test for carbohydrates) and they bound to the lectin, *Limax flavus* agglutinin (LFA), which has affinity for terminal sialic acid. The modification was believed to be *O*-linked as flagellin glycosylation occurred even in the presence of an *N*-linked glycosylation inhibitor. The study showed that *C. jejuni* flagella are extensively modified with glycans. Approximately 10% of the total mass of the flagellin glycoprotein is attributed to carbohydrate. Studies following this on *C. jejuni*

strain 81-176 revealed that the terminal residue in the glycan is not sialic acid but instead a nine-carbon related sugar of sialic acid, 5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-nonulosonic acid (pseudaminic acid, Pse5Ac7Ac) or one of its derivatives such as 7-acetamido-5-acetimidoyl-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-nonulosonic acid (legionaminic acid, Pse5Am7Ac). (Benz and Schmidt 2002, Power and Jennings 2003, Szymanski et al. 2003, Logan et al. 2002, Thibault et al. 2001). Pseudaminic acid and legionaminic acid are shown in figure 1.10.
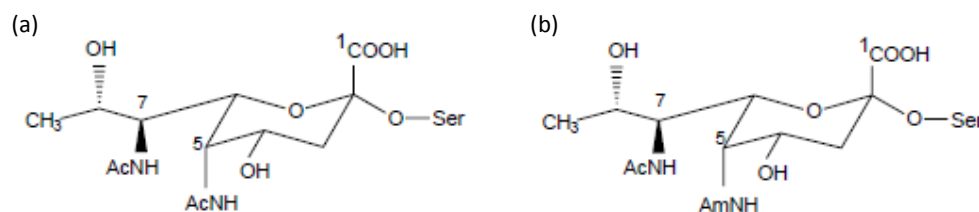
Figure 1.10 (a) Pseudaminic acid and (b) Legionaminic acid

The structure of (a) pseudaminic acid, a nine carbon related sugar of sialic acid and its derivative (b) legionaminic acid which modify *C. jejuni* flagella (Schmidt et al. 2003).

C. jejuni flagella are modified by pseudaminic acid or its derivatives at 19 sites. The modification occurs on serine/threonine residues in the central domain of the protein that is surface exposed (Thibault et al. 2001). Studies on *C. jejuni* strain NCTC 11168 (Szymanski et al. 2003) revealed that it contains a flagellar glycosylation gene locus containing about 50 genes. The gene cluster contains genes involved in the biosynthesis of pseudaminic acid and pseudaminic acid derivatives. Some of the genes encode the flagellin structural proteins FlaA and FlaB. Other genes encoded in the *O*-linked glycosylation locus of *C. jejuni* strain NCTC 11168 include hypothetical genes that are thought to encode proteins of two paralogous gene families, the 1318 motility accessory factor (1318/*maf*) family and the 617 family. These two gene families account for about 50% of the genes in the locus. There have been no known orthologues of the 617 family found in any other bacterial species and the functions of the proteins encoded by these genes are unknown (Szymanski et al. 2003). At this time the mechanism for *O*-glycosylation of flagella in *C. jejuni* is unknown. A very surprising characteristic of the flagellar glycosylation locus is that there have been no characterised GTs found in the locus (Szymanski and Wren 2005). Perhaps the organism uses GT genes from other pathways to transfer the sugar units to its flagella.

Table 1.2 summarizes the differences between the key features of the *O*-linked and *N*-linked protein glycosylation pathways in *C. jejuni* that have been discussed.

Table 1.2 Summary of the key differences between the *O*- and *N*-linked protein glycosylation systems in *C. jejuni* (adapted from Szymanski et al. (2003))

Feature	*N*-linked	*O*-linked
Nature of glycan	Heptasaccharide	Monosaccharide or disaccharide
Glycan structure	GalNAc$_2$(Glc)GalNAc$_3$Bac-	Pseudaminic acid-
Target proteins	>65 secreted proteins	FlaA and possibly others
Size of gene loci	17 kb	Variable, ~50 genes
Target sequence	Glu/Asp-Z-Asn-X-Ser/Thr	Ser/Thr

- *P. aeruginosa* **flagella glycosylation**

P. aeruginosa is also capable of glycosylating its flagella (Arora et al. 2001). There are two major flagellin types found in *P. aeruginosa*, type A and type B flagella. Both types are known to be glycosylated. *P. aeruginosa* flagella consist primarily of a FilC subunit. A gene locus adjacent to the flagellin biosynthesis genes was found by Arora et al. (2001). The locus contains up to 14 genes required for the *O*-glycosylation of type A flagella. Experiments were carried out by Arora et al. where the first and last genes in the locus, *orfA* and *orfN* were inactivated. Inactivation of either of these genes completely abolished flagellin glycosylation.

OrfA is homologous to PglC from *N. meningitidis* and PglE form *C. jejuni* (both aminotransferases), while OrfN is homologous to PglE from *N. meningitidis* and PglI form *C. jejuni* (both GTs) (Power and Jennings 2003, Arora et al. 2001). The glycosylation moiety attached to A-type flagella in two *P. aeruginosa* strains (PAK and JJ692) have been determined by Schirm et al. (2004). Their flagella are *O*-glycosylated at two sites. In both strains the glycosyl moiety is linked through a rhamnose residue. In strain JJ692 the flagella are glycosylated with only a single rhamnose but in strain PAK the glycan can contain up to 10 additional monosaccharides (Schirm et al. 2004). The B

type flagellin in *P. aeruginosa* PAO1 is modified at two nearby serine residues with a single deoxyhexose residue that can be linked to a unique modification with a mass of 209 Da that is known to include a phosphate moiety (Verma et al. 2006).

- *Clostridium* **spp. flagella glycosylation**

Several *Clostridia* spp. have been shown to contain an *O*-linked protein glycosylation system responsible for the glycosylation of its flagella (Twine et al. 2009). In *Clostridium botulinum* the flagella are modified with a derivative of legionaminic acid called 7-acetamido-5-(*N*-methyl-glutam-4-yl)-amino-3,5,7,9-tetradeoxy-D-*glycero*-α-D-*galacto*-nonulosonic acid (αLeg5GlunMe7Ac) and with di-*N*-acetylhexuronic acid. Seven Ser/Thr sites in each protein subunit are modified (Twine et al. 2008).

Clostridium difficile 630 flagella are also modified at seven sites but the modification is a HexNAc moiety. Mutants lacking this HexNAc moiety showed defected flagellum formation and loss of motility. Hypervirulent strains of *C. difficile* that have been recently isolated from outbreaks have been shown to contain *O*-glycans on their flagella containing HexNAc-linked oligosaccharides with at least five sugars in the chain including HexNAc, deoxyhexose, methylated deoxyhexose and heptose residues. These are the most complex flagellin-linked *O*-glycans that have been discovered so far. Analysis of the genome of these organisms have shown that the locus involved in flagellin glycosylation varies greatly to the *C. difficile* 630 strain (Dell et al. 2010, Nothaft and Szymanski 2010, Twine et al. 2009).

- *O*-**mannosylation in Gram-positive Actinomycetes**

A recent review by Lommel (2009) discusses how protein *O*-mannosylation is a process that is conserved in all three domains of life. *O*-mannosyl glycans are very common in fungi and yeast while only a limited number of proteins are *O*-mannosylated in mammals. *O*-mannosyl glycans similar to those found in yeast and fungi have been discovered in bacteria, namely Gram-positive Actinomycetes including *Mycobacterium* spp. and *Streptomyces* spp. (Lommel and Strahl 2009). The first glycoprotein of this

kind to be characterised was in *Mycobacterium tuberculosis*. A 45 kDa secreted protein, MPT 32, was shown to be glycosylated at threonine residues with a single α-D-mannose, mannobiose or mannotriose (with α1-2 linkages between mannose residues) (Dobos et al. 1996). The mannosyltransferase (MT) responsible for this glycosylation has been discovered and has structural similarity to eukaryotic protein MTs (Dell et al. 2010).

In *Streptomyces coelicolor*, its phosphate binding protein PstS is *O*-glycosylated. The glycan is a trihexose and the MT responsible is a membrane bound lipoprotein, Pmt (Wehmeier et al. 2009). *O*-mannosylation appears to be a general pathway common to all Actinomycetes as orthologues to MT genes have been found in many other genera. As well as requiring MT enzymes, a lipid carrier is also needed similar to what is seen in *Neisseria* and *Pseudomonas* pilin glycosylation but as Actinomycetes are Gram-positive, they do not have a periplasm, therefore the lipid-linked oligosaccharide would be flipped to the surface of the outer membrane with the transfer of the glycan to the protein occurring on the outer surface of the cell (Nothaft and Szymanski 2010).

- ***O*-Linked heptoses in Gram-negative *Escherichia coli***

Autotransporter proteins are the most common type of secreted proteins in Gram-negative bacteria. In pathogenic *E. coli* strains, three glycosylated autotransporter proteins have been discovered, namely, AIDA-I, TibA and Ag43. These proteins are associated with virulence phenotypes such as biofilm and aggregate formation. The modification is an *O*-linked heptose on the proteins' passenger domains (Benz and Schmidt 2001, Lindenthal and Elsinghorst 1999, Sherlock et al. 2006). These glycans are derived from ADP-linked glycans recruited from the LPS pathway in the organism. Glycosylation of these proteins appears to aid in pathogenesis by enhancing the attachment of the bacterium to human cells. Glycosylation is also required for activity of these proteins (Dell et al. 2010, Schmidt et al. 2003).

The examples discussed in sections 1.3.1 and 1.3.2 is by no means a complete overview of all known incidences of prokaryotic protein glycosylation but it does include the best understood pathways of prokaryotic protein glycosylation to date. Table 1.3 shows a

summary of some of the properties of bacterial proteins that are known to be glycosylated, including some from organisms that have not been discussed here.

Table 1.3 Overview of some characterised bacterial glycoproteins

Organism	Protein	Linked glycan	# Sugar residues/ glycan	*N-/O-* linked	# Modification sites/protein	Transfer of sugars	Biological effect of modification	Reference
C. jejuni subsp.	Multiple, secreted	GalNAc$_2$(Glc)GalNAc$_3$Bac	7	*N*	1-3	Block	competence, animal colonisation, adherence/ invasion, protein antigenicity	Young et al. (2002), Szymanski et al. (1999)
C. coli VC167	FlaA flagellin	Pseudaminic acid, acetamidino, deoxypentose and Pse/PseAm	1-2	*O*	16	Sequential	Flagella assembly, bacterial motility, protein antigenicity	Logan et al. (2002)
C. jejuni 81-176	FlaA flagellin	Pseudaminic acid, acetamidino, *O*-acetylacetamidino, dihydroxyproprionyl	1	*O*	19	Sequential	Flagella assembly, bacterial motility, protein antigenicity	Thibault et al. (2001)
Helicobacter pylori	FlaA, FlaB	Pseudaminic acid	1	*O*	7, 10	Sequential	Flagella assembly, bacterial motility	Schirm et al. (2003)
N. meningitidis	Pilin	Gal(β1,4)-Gal(α1,3)-DATDH	2-3	*O*	1	Block	Complement mediated killing	Stimson et al. (1995)
N. gonorrhoeae	Pilin	Gal(α1,3)-GlcNAc	2-3	*O*	1	Block	Correlation with disease	Parge et al. (1995)
P. aeruginosa PAK	Flagellin	Up to 11 sugars linked through rhamnose including 4-amino 4,6-dideoxyhexose viosamine	Up to 11	*O*	2	Sequential	Role in virulence-adhesion/invasion/ colonisation	Schirm et al. (2004)
P. aeruginosa JJ692	Flagellin	Rhamnose	1	*O*	2	Sequential	Role in virulence-adhesion/invasion/ colonisation	Schirm et al. (2004)

Organism	Protein	Linked glycan	# Sugar residues/ glycan	N-/O- linked	# Modification sites/protein	Transfer of sugars	Biological effect of modification	Reference
P. aeruginosa 1244	Pilin	Hydroxybutyryl-formyl-pseudaminic acid with xylose and *N*-acetylfucosamine	3	*O*	1	Block	Adherence/Virulence/ Colonisation	Castric et al. (2001), Smedley III et al. (2005)
P. aeruginosa	LecB	_	_	*N*	1	_	Localisation of the protein in the outer membrane	Bartels et al. (2011)
E. coli H10407	TibA adhesin	Heptoses	_	*O*	_	Sequential	Adherence/Invasion	Lindenthal and Elsinghorst (1999)
E. coli 2787	AIDA-1[1] adhesin	Heptoses	1	*O*	19	Sequential	Adherence	Benz and Schmidt (2001)
M. tuberculosis	MPT 32	α(1-2) linked D-mannose/ mannobiose/mannotriose	1-3	*O*	4	Sequential	Pathogenesis	Dobos et al. (1996)
Mycobacterium bovis	MBP83 antigen	(1,3)-linked mannose	3	*O*	2	Sequential	Pathogenesis	Michell et al. (2003)
C. difficile 630	Flagellin	HexNAc	1	*O*	7	Sequential	Formation of flagellin/ motility	Twine et al. (2009)
H. influenzae	HMW1[2] adhesin	Gal, Glc	1-2	*N*	31	Sequential	Protein stability/adherence	Grass et al. (2003)

[1] Adhesin involved in diffuse adherence
[2] High molecular weight protein adhesin

1.3.3 The use of glycan-specific genes in multiple pathways

Capsular polysaccharide (CP) is found on the cell surface of many bacterial species. It consists of a think mucous-like layer of polysaccharide and functions in adherence and infection and prevents the bacteria from drying out. CP is composed of repeating monosaccharides joined by glycosidic linkages. They polymers can be homo- or heterologous. CP is very diverse. It differs from organism to organism with regard to the sugar units found in the molecule and also the linkages between these sugars (Roberts 1996).

Lipopolysaccharide (LPS) and lipooligosaccharide (LOS) are both major constituents of the outer surface of Gram-negative bacteria. As the name suggests LPS consists of a polysaccharide covalently linked to a lipid and LOS contains an oligosaccharide linked to a lipid. LPS contains three regions, the lipid, an oligosaccharide core and a sequence of repetitive subunits called the O-antigen (Caroff and Karibian 2003). The core consists of sugars such as glucosamine, glucose, galactose, L-*glycero*-D-*manno*-heptose or 3-deoxy-oct-2-ulosonic-acid (KDO). In LOS there is no third region, only the core oligosaccharide consisting of the same sugars found in LPS bound to the lipid. LPS containing O-antigen is sometimes called smooth LPS and LOS is sometimes called rough LPS (Szymanski and Wren 2005, Caroff and Karibian 2003).

O-antigens are highly variable and occur in multiple forms. There is considerable variation of O-antigens between and even within different species (Samuel and Reeves 2003). The variation occurs in the nature, order and linkage of the different sugars. For example there has been 186 different O-antigens reported in *E. coli* strains and 54 in *Salmonella enterica* and only three of these are identical between the two species (Samuel and Reeves 2003).

Figure 1.11 shows a schematic representation of a typical enterobacterial Gram-negative cell envelope and the structures of LOS and LPS.
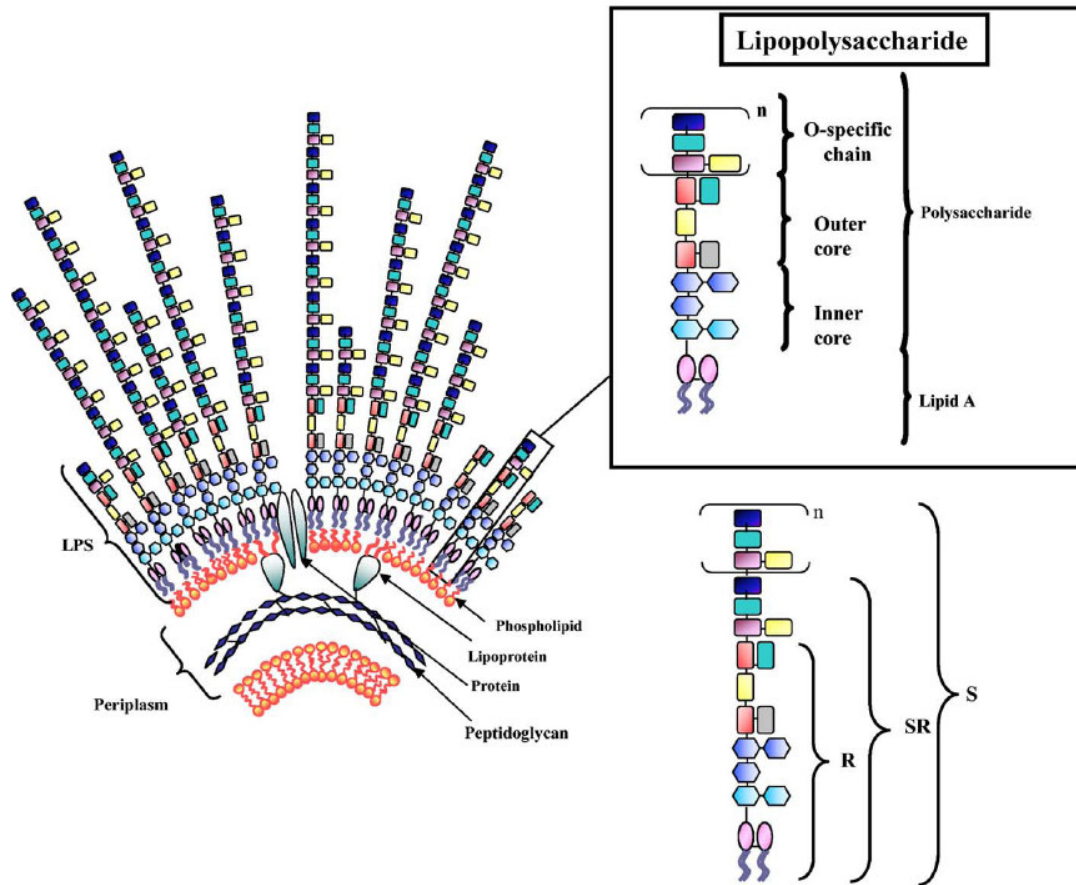
Figure 1.11 The typical structure of the Gram-negative cell envelope, LPS and LOS
LOS (rough LPS) contains only the lipid bound to the sugar core. Semi-rough LPS contains one O-antigen subunit. Smooth LPS contains multiple O-antigen subunits (Caroff and Karibian 2003).

The genes involved in O-antigen biosynthesis are usually found clustered together on the chromosome in a gene locus (Samuel and Reeves 2003). There are three major groups of genes found in the gene cluster. The first group is genes that encode proteins involved in the biosynthesis of the nucleotide precursors of the O-antigen. The second group encodes GTs. These transfer the synthesised precursor sugars to the lipid carrier, UDP. The UDP-linked oligosaccharide is positioned at the cytoplasmic side of the inner membrane. The final group of genes encode O-antigen processing proteins. These include flippases, that translocate the O-antigen across the membrane, polymerases and chain length determining enzymes (Samuel and Reeves 2003). Figure 1.12 outlines the synthesis of O-antigens and shows the three different groups of proteins involved.
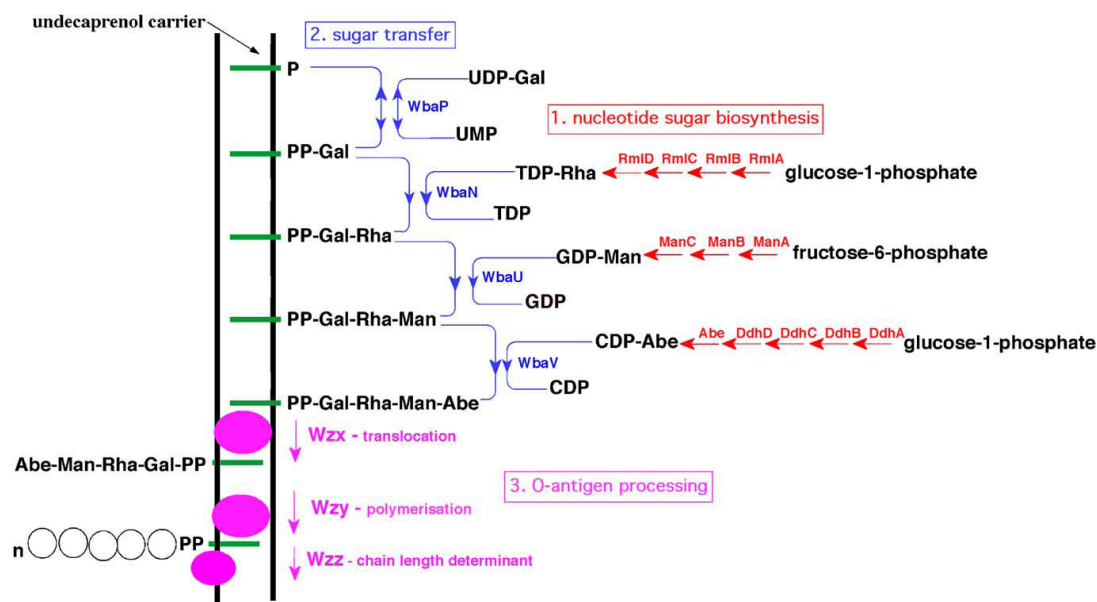
Figure 1.12 Representation of typical O-antigen biosynthesis

Nucleotide sugar precursor biosynthesis genes are shown in red, GT genes are in blue and antigen processing genes are in pink. Sugar biosynthesis and sequential transfer to the growing sugar chain occurs in the cytoplasm and processing and polymerisation of O-antigens occurs at the periplasmic face of the inner membrane (Samuel and Reeves 2003).

Many bacteria have evolved to use the same carbohydrate biosynthetic genes for different pathways i.e. bacteria that produce glycoproteins, LPS, LOS or CP often use the same carbohydrate specific genes for more than one of these pathways. Because of this bacterial genomes are extremely compact and usually do not contain any redundant genes (Szymanski and Wren 2005). It can be seen that the genes involved in the biosynthesis of LPS, LOS and CP are very similar to those involved in protein glycosylation. The three groups of enzymes used for O-antigen biosynthesis described above are the same types of enzymes that are used for protein glycosylation in many bacteria.

Some examples of prokaryotes using carbohydrate biosynthesis or GT genes in more than one pathway have already been discussed. For example, common enzymes are used in *P. aeruginosa* strain 1244 for both pilin glycosylation and O-antigen production (section 1.3.2.3). Mutation of the genes *wbpM* or *wbpL* inhibits production of both O-antigens and glycosylated pili (Horzempa et al. 2006, DiGiandomenico et al. 2002).

Another example of bacteria using a carbohydrate specific gene for more than one pathway can be seen in *C. jejuni*. It produces ADP-heptose to modify its LOS and GDP-heptose to modify its CP. A common phosphatase enzyme is used in both pathways (Szymanski and Wren 2005, Karlyshev et al. 2005). *C. jejuni* also produces the protein GalE. This is a bifunctional epimerase that converts glucose to galactose and GlcNAc to GalNAc (Bernatchez et al. 2005). When the *galE* enzyme is inactivated it affects three important pathways. It results in truncation of LOS and lack of modification of CP and *N*-linked proteins by GalNAc (Szymanski and Wren 2005, Bernatchez et al. 2005, Fry et al. 2000). When the same enzyme (*galE*) is mutated in *N. meningitidis*, this causes truncation of LOS and the pilin trisaccharide Gal(β1,4)-Gal(α1,3)-DATDH (Szymanski and Wren 2005, Stimson et al. 1995). In *H. influenzae*, the enzyme phosphoglucomutase is involved in both LOS production and glycosylation of HMW1 (Szymanski and Wren 2005, Grass et al. 2003).

It has also been found that in some cases genes involved in protein glycosylation may be very similar to genes in different organisms that are involved in LPS, LOS or CP production. Power et al. (2006) showed that pilin glycosylation in *N. meningitidis* occurs by a very similar pathway to wzy-dependent O-antigen biosynthesis in *E. coli*. The *N*-glycosylation system that has been proposed in *C. jejuni* has a lot of similarities to the O-antigen or O-polysaccharide polymerase dependent synthesis pathway seen in many Gram-negative bacteria (Feldman et al. 2005, Bugg and Brandish 1994).

1.3.4 Significance of bacterial glycosylation

Glycosylation has been demonstrated to alter the function of both eukaryotic and prokaryotic proteins. Many studies have been conducted on eukaryotic proteins and the specific effects of many the modifications on the functions of the proteins have been determined (Upreti et al. 2003). Some of these functions include maintenance of protein conformation and stability, cell adhesion, modulation of viscosity, solubility and surface charge, proteolytic processing, mediation of biological activity and embryonic development (Upreti et al. 2003, Paulson 1989).

At present there is not a great deal known about the specific function of the glycosylation of many bacterial glycoproteins but it is believed to have similar functions to eukaryotic glycosylation (Upreti et al. 2003). This area requires much further exploration. For some prokaryotic glycoproteins the function of the modification has been characterised. For example a lot of knowledge has been obtained from studies on S-layer glycoproteins and more recently, glycosylated surface appendages such as flagella and pili. Because many glycosylated proteins are surface exposed in the organism this suggests that glycosylation plays an important role in pathogenicity (Szymanski and Wren 2005).

Glycoproteins have been proven to directly function in pathogenicity in some organisms. For example in *Mycobacterium smegmatis* glycosylation of proteins with an acyl or glucosyl moiety inhibits the major histocompatibility complex (MHC) from presenting peptides to T-cells (Upreti et al. 2003, Baumeister and Lembcke 1992). Glycoproteins on the cell surfaces of bacteria are important for pathogenicity in terms of cell-cell interactions between the bacteria and host cell. They can be important for facilitating adhesion of the bacteria and invasion into the host cell. Glycosylation of cell surface structures may also stabilize attachment of the bacteria to the host cell containing the proper receptor for the glycan (Upreti et al. 2003).

Glycosylation often functions in pathogenesis by aiding in immune evasion. Many examples of this are seen in pathogenic bacteria. One such example is in *Neisseria* spp. It produces an *O*-glycosylated outer membrane protein called AniA, a nitrate reductase. Glycosylation of this protein protects the organism against killing by the human immune response. The protein is glycosylated at its carboxyl terminus. As the protein is an outer membrane protein it is surface exposed and immunogenic. The carbohydrate modification shields the substrate recognition domain of AniA and therefore protects it from immune recognition (Ku et al. 2009).

Some examples have been discussed already where bacteria *O*-glycosylate their flagella. This modification seems not only to be important for cell-cell interactions but also for facilitating correct formation of the flagellar filament and motility of the organism as seen in *H. pylori* (Schirm et al. 2003) and *C. jejuni*. (Logan et al. 2002, Thibault et al. 2001). Mutation of flagellar glycosylation genes in *C. jejuni* results in non-motile

phenotypes with no flagellar filament so *O*-linked glycosylation in *C. jejuni* is required for complex protein assembly (Szymanski and Wren 2005, Goon et al. 2003).

The exact function of *N*-linked protein glycosylation in *C. jejuni* is not fully understood. It is known that *N*-linked glycosylation functions in adherence and invasion of the host. Disruption of the *N*-linked glycosylation pathway produces cells that have reduced ability to adhere and invade *in vitro* eukaryotic cell cultures and show reduced colonisation in the intestines of chicken and mice (Szymanski et al. 2002).

Glycoproteins on cell surfaces can contribute to the structural integrity of the cell and thus aid in cell adhesion. An example of this is the glycoproteins on the surface of *Halobacterium* spp. They function in maintaining the rod shape of the bacterial cell. When addition of the glycan to the protein is inhibited the cells are changed to spherical cells (Upreti et al. 2003). Another example can be seen in *Thermoplasma acidophilum* where carbohydrates attached to the membrane proteins trap water molecules and immobilize them to form a mesh like network which gives the plasma membrane rigidity and stability (Upreti et al. 2003).

Other important biological functions have also been attributed to glycosylation. Glycosylation of S-layer proteins can facilitate the organism to survive in hostile environments. The glycans can be found clustered together in domains on the protein, restricting the folding of the protein and leading to an extended conformation. If the glycans are added at β-turns in the protein, the turn conformation is masked and the protein can be protected from proteolysis. Removal of these glycans exposes protease cleavage sites in increases the sensitivity of the protein to proteolytic cleavage (Upreti et al. 2003, Beeley 1977).

Many bacterial enzymes are glycosylated. It is believed that the glycosylation does not affect the catalytic activity of the enzyme but may function in recognition and binding of the enzyme to the substrate, allowing the enzyme to bind with higher affinity and thereby increase the effectiveness of the enzyme (Kawamura and Shockman 1983). Glycosylation can also increase the thermostability of the enzyme. It has been shown in *Bacillus amyloliquefaciens* and *Bacillus macerans* that glycosylated enzymes had a

longer half life at higher temperatures compared to their non-glycosylated forms (Upreti et al. 2003).

Glycosylation of proteins in *Bacteroides* spp. by fucose residues is believed to aid in the survival of the organism in the mammalian intestine. Because many mammalian proteins are glycosylated with fucose *Bacteroides* are able to disguise themselves and become immunologically inert (Fletcher et al. 2009).

1.4 Identification and characterisation of glycoproteins – glycoproteomics

1.4.1 Gel electrophoresis

There are a wide variety of techniques available for the identification and characterisation of glycoproteins. A combination of the available techniques is needed to adequately identify the carbohydrate moiety attached to the protein. Often the first step for the isolation and analysis of proteins is gel electrophoresis for example sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), isoelectric focusing (IEF) or 2-dimensional (2-D) gel electrophoresis, which is a combination of both SDS-PAGE and IEF. Glycoproteins sometimes exhibit aberrant migration in a gel compared to their non glycosylated counterpart (Hitchen and Dell 2006). The protein bands or spots obtained by these gels can be very broad because the glycoprotein is usually a heterogeneous mix of different glycoforms so complete separation of the different glycoforms can be difficult. Another drawback of electrophoresis is that often membrane glycoproteins are difficult to detect. This is because these proteins are hydrophobic and so do not solubilise well in the non-ionic detergents used in IEF. Another problem is that they do not stain very well with conventional dyes because of their high carbohydrate content. So other approaches are used for the isolation of membrane glycoproteins such as blue native electrophoresis or SDS-PAGE in conjunction with nano-high performance liquid chromatography (HPLC). Capillary electrophoresis has a very high resolving potential for different glycoprotein isoforms

but does not give any information on the nature of the attached glycan (Geyer and Geyer 2006).

A useful way to distinguish between glycoproteins and non-glycosylated proteins separated by gel electrophoresis is by the periodic acid-Schiff stain. This stains carbohydrates a pink colour but does not stain proteins. The periodic acid oxidises the carbohydrate to form an aldehyde, which undergoes a condensation reaction with the Schiff base and a chromagen or indicator enzyme to form a pink colour (Benz and Schmidt 2002). It is important to confirm that the carbohydrate is actually covalently linked to the protein and not just associated with it. This can be confirmed by genetic, chemical or enzymatic means. For example treatment of the glycoprotein with trifluoromethanesulphonic acid breaks the covalent bond and so the migration of the protein in a gel will be altered (Benz and Schmidt 2002, Sojar et al. 1987). Treatment of glycoproteins with the enzyme PNGaseF completely removes all *N*-linked glycoproteins. Treatment of glycoproteins using the β-elimination reagent completely removes all *O*-linked glycans.

1.4.2 Lectins

One of the most useful methods of detecting and identifying glycoproteins is the use of lectins. Lectins are defined as carbohydrate binding proteins which lack enzymatic activity and are not produced as a result of an immune response i.e. they are not antibodies (Ambrosi et al. 2005, Barondes 1988, Gabius et al. 2011). Lectins react non-covalently with carbohydrates. They bind mono- and oligosaccharides. The reaction is reversible and relatively specific. An important property of lectins is that they exhibit di- or polyvalency, meaning that they contain two or more carbohydrate binding sites. This property means that when they react with sugars on the surface of cells such as erythrocytes or on glycoproteins they cause cross-linking of the cells or molecules which leads to their agglutination or precipitation. This property of lectins is used for their detection and characterisation. The agglutination and precipitation reactions of lectins are inhibited by the sugars for which the lectin is specific (Lis and Sharon 1998).

Lectins are found in most living organisms from viruses to animals (Ambrosi et al. 2005). The first demonstration of lectin activity was shown in around 1860 by the use of rattlesnake venom to agglutinate red blood cells (Ambrosi et al. 2005, Kilpatrick 2002). Sumner isolated the first pure lectin, Concanavalin A (Con A) from jack beans in 1919 (Ambrosi et al. 2005, Sumner 1919). He also demonstrated the sugar specificity of the lectin (Sumner and Howell 1936). Nowadays lectins are readily purified from their source using techniques such as affinity chromatography with immobilised carbohydrates. Recombinant DNA technology is also used to produce pure lectins (Lis and Sharon 1998).

1.4.2.1 Carbohydrate specificity

Lectins are classified into five main groups based on the monosaccharide for which they have highest affinity. These groups are mannose, galactose/GalNAc, GlcNAc, fucose and sialic acid. The structures of these monosaccharides were shown previously in table 1.1.

This list ignores some important monosaccharides but the six monosaccharides listed are the most common monosaccharides found in eukaryotic glycans. These sugars are all in the D-configuration apart from fucose which is in the L-configuration (Lis and Sharon 1998). Although the affinity of lectins is very weak for monosaccharides the specificity is very high (Gabius et al. 2011). This means that in general lectins specific for galactose do not react with glucose or mannose even though the structures are so similar (galactose is the 4 epimer of glucose and mannose is the 2 epimer of glucose). Those specific for mannose do not bind galactose. Also except for wheat germ agglutinin (WGA), lectins specific for GlcNAc do not bind GalNAc and vice versa. This high selectivity is not always the case. For example many lectins tolerate variations at C-2 of the pyranose ring and so for this reason many lectins that bind galactose also bind GalNAc and so they are put in the same specificity group (Ambrosi et al. 2005). In some cases lectins bind preferentially to a particular anomer of a glycan. Sometimes the properties of the aglycan (the non-sugar part of a molecule) can have an effect on the binding of the lectin. For example Con A binds to aromatic glycosides more strongly

than aliphatic glycosides. This is because of the hydrophobic region that is in close proximity to the carbohydrate binding site (Poretz and Goldstein 1971).

Nowadays this method for classifying lectins is becoming obsolete because of the differences in specificities of lectins in a single category and also many lectins do not have a high specificity to simple saccharides (Ambrosi et al. 2005). Many lectins can be up to 1000 times more specific for di- tri- and tetrasaccharides than they are for monosaccharides and some lectins only interact with oligosaccharides. This is because oligosaccharides are the most likely ligands of lectins to be found in nature (Lis and Sharon 1998).

1.4.2.2 Structural features

Now lectins tend to be grouped according to their sequences and structural features. There are three main categories for defining lectins in this way. These are simple, mosaic (or multidomain) and macromolecular assemblies (Lis and Sharon 1998). Simple lectins have several subunits, each below ~40 kDa in size which are not necessarily identical. Each subunit contains a carbohydrate binding site. Simple lectins include most known plant lectins and most of the galectins. Galectins (formerly known as S-lectins) are animal lectins specific for β-galactoside. The legumes family is the largest and most comprehensively studied family of simple lectins of which Con A is a member. In 1998 there were over 100 characterised legume lectins. Almost all of these lectins have been isolated form the seeds of the plants (Lis and Sharon 1998, Sharon and Lis 1990, Konami et al. 1995).

Mosaic lectins are composite molecules consisting of several domains. Only one domain contains a carbohydrate binding site. Many of these lectins are monovalent but act like multivalent lectins because they are embedded in membranes. Included in the mosaic lectin group are diverse proteins from a range of sources such as viral hemagglutinins and C-, P- and I- type animal lectins (Lis and Sharon 1998).

Macromolecular assemblies are lectins commonly found on bacterial surfaces. These lectins are filamentous and heteropolymeric, 3-7 nm in diameter and 100-200 nm in

length. They consist of several subunits called pilins that are helically arranged and assembled in a well defined order (Gaastra and Svennerholm 1996).

Lectins may be used in a number of ways to isolate and characterise glycoproteins. They may be used to probe glycoproteins that are immobilised on the surface of a chip or a plate. The lectins are often labelled with a molecule such as biotin. Anti-biotin antibodies can then be used to detect lectins that have bound to glycoproteins. They may be used in the same way in western blots. Lectins can be covalently bound to matrices like sepharose or agarose (in lectin affinity columns), membranes or magnetic beads to allow the separation, fractionation and characterisation of glycoproteins. Depending on the requirement, lectins with broad specificities or very specific lectins can be used. Using broad range lectins an entire pool of glycoproteins can be isolated but glycoforms can be differentiated in greater detail using specific lectins or specific eluting sugars (Geyer and Geyer 2006). Recently, multi-lectin affinity chromatography columns have been developed which simultaneously allow an almost complete enrichment of glycoproteins from the sample as well as their fractionated displacement from the column (Geyer and Geyer 2006, Wang et al. 2006). Lectin microcolumns have also been developed which allow high pressures to be used and can be coupled online to mass spectrometry equipment allowing semi-automated, sensitive profiling of glycoproteins (Madera et al. 2006). Lectin arrays have been developed containing over 35 different lectins which provides a fast, sensitive, high throughput detection method for glycoproteins (Kuno et al. 2005). All of these techniques are valuable tools for the future to gain a better understanding of the glycoprotein structures found in prokaryotes.

1.4.3 Mass spectrometry

Once gel electrophoresis or lectin-based methods have been used to isolate glycoproteins analytical methods such as mass spectrometry (MS) are extremely useful for identifying the protein and determining the carbohydrate structure attached to it. MS can analyse large numbers of proteins in complex mixtures with very high sensitivity and accuracy. In MS the sample is converted to gas-phase ions. The ions are accelerated out of the ionization source into a mass analyser. Inside the mass analyser the ions are

separated according to their mass to charge ratio and detected to produce a mass spectrum (Hitchen and Dell 2006).

Matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) MS are considered the best methods available for protein identification using MS. MALDI or ESI describe the method that is used to generate the gas-phase ions. These methods can be used to analyse large biological molecules and do not degrade them. A technique known as peptide mass fingerprinting (PMF) is commonly used for protein identification. This involves cleaving the protein into peptides using an enzyme such as trypsin, which cleaves the protein only at specific residues. The mass of each peptide is determined by the mass spectrometer (mass mapping) and the information is inputted into a search engine that uses an algorithm to identify what protein the peptides belong to (Henzel et al. 1993). MALDI-MS is an unsurpassed method for this type of analysis. This type of analysis for the identification of proteins is called "bottom-up" proteomics. One disadvantage of PMF is that it cannot be used for complex samples with more than two or three proteins present. It requires relatively pure protein samples.

PMF alone is sometimes not enough to identify a protein conclusively. ESI-MS can be used to determine the exact molecular weight of the intact protein. This will confirm the identification of the protein. MS is also very useful for peptide or protein sequencing. For this type of analysis tandem mass analysers (MS/MS) are needed. The first analyser traps the peptide or molecular ions of interest. The ions are fragmented by collision with an inert gas and broken up into fragments or individual amino acids. The second analyser separates the fragmented ions which can then be detected. A program is used that takes the mass data from the fragments and uses it to deduce the sequence of the original precursor ion.

MS/MS may also be used in the same way to determine the sequence of monosaccharides in a glycan on a glycopeptide. It cannot give information on the conformation of the monosaccharides or the type of linkage between sugar units, only their mass. Glycans can also be analysed separately from a protein by first releasing the glycan from the protein by chemical or enzymatic means. The cleaved glycans can be easily separated from the much larger protein using centrifugal filters with a lower

molecular weight cut off point than the molecular mass of the protein (North et al. 2009).

The quadropole orthogonal acceleration time of flight (Q-TOF) mass spectrometer is the most common type of MS/MS instrument. It has the ability to perform MS/MS experiments on line in real time (Hitchen and Dell 2006). Nowadays tandem mass spectrometer configurations such as MALDI-Q-TOF are commonly used. This type of instrument combines the advantages of both MALDI and Q-TOF i.e. it can carry out mass mapping of the protein as well as being able to sequence it (Medzihradszky et al. 2000). Protein and glycan sequencing done is this way is quite time consuming and so this method is not suitable for high-throughput analysis.

A new technique known as "top-down" proteomics is emerging as an alternative method for protein and glycoprotein identification using MS. For this analysis usually ESI is used. There is no requirement for cleaving the protein into peptides prior to analysis, instead the ionised protein is subjected to gas-phase fragmentation in the mass analyser. The sizes of the fragments generated are determined and the protein sequence can be obtained. Top-down proteomics is very useful because it enables post-translational modifications on a protein to be located and characterised (Siuti and Kelleher 2007). For example, different forms of a protein can be isolated from different cell states, with the proteins having the same amino acid sequence but varying post translational modifications (PTMs). Top-down proteomic analysis allows variations in protein size to be detected and hence the type of modifications attached to the protein can be elucidated. Because the protein is fragmented into single amino acids, the location of the PTMs can also be found.

1.4.4 Nuclear magnetic resonance (NMR) spectroscopy for glycan analysis

NMR spectroscopy is an extremely useful technique for obtaining detailed information about the glycan structures on a glycoprotein. It can determine the conformation of the monosaccharides and the types of linkages between monosaccharides. Certain atomic nuclei have magnetic properties and in a magnetic field they absorb and re-emit electromagnetic energy. The energy is at a specific resonance frequency which is

determined by the strength of the magnetic field applied and the magnetic properties of the nucleus of the atom. NMR spectroscopy uses the resonance frequency values to determine the physical and chemical properties of that molecule. NMR spectroscopy is a very powerful method of analysis but requires relatively large amounts of pure glycan (10-100 ng). HLPC coupled to NMR spectroscopy can be used in order to separate out different glycans first. The equipment required for NMR spectroscopy is very expensive and highly trained operators are required to use it (Evans 1995, Keeler 2005).

1.5 Exploitation of prokaryotic protein glycosylation systems for industrial applications

Over 70% of human therapeutic proteins are modified with *N*-linked glycans (Pandhal and Wright 2010). Correct glycosylation of these products is essential to ensure that these products exhibit desirable pharmacokinetic properties and to prevent unwanted immune responses to the therapeutic. Currently the majority of complex glycoprotein based therapeutics are expressed in mammalian cells for example Chinese hamster ovary cells. This is because unlike yeast and bacterial cells, mammalian cells add human like glycans to the glycoprotein product (Hamilton and Gerngross 2007, Walsh and Jefferis 2006). However many technical difficulties can arise with the use of mammalian cell culture for the expression of glycoproteins. As previously stated, glycosylation is linked to the overall metabolic state of the cell and so can be influenced by a number of different factors such as nutrient availability, media composition, pH and oxygen levels and other conditions within the fermentation tank. If the correct conditions are not optimised and maintained this can lead to aberrant glycosylation of a significant percentage of the glycoprotein product. Incorrect glycosylation can significantly alter the properties of the glycoprotein, such as stability, solubility and half life, leading to a great deal of wasted product and increased production costs (Walsh and Jefferis 2006, Werner et al. 2007).

Expression systems that are based in bacterial cells is a far more desirable option due to the relative ease with which bacteria can be grown and maintained compared to animal cells. Expression in bacteria results in much lower costs and higher yields in a shorter time compared to animal cells. It is also much easier to genetically modify a bacterial

cell compared to an animal cell. Since the discovery of *N*-linked protein glycosylation systems in bacteria, glycoengineering within bacterial cells for the production of human therapeutics has become a real possibility for the future (Pandhal and Wright 2010, Langdon et al. 2009).

A significant breakthrough in this area came in 2002 when Wacker and co-workers successfully transferred the *pgl* gene locus from *C. jejuni* responsible for *N*-linked protein glycosylation (section 1.3.2.1) into *E. coli* and were able to produce *N*-glycosylated proteins in *E. coli* (Wacker et al. 2002). PglB, the OT enzyme from the *C. jejuni N*-linked system shows relaxed substrate specificity compared to its eukaryotic counterpart and it was shown that various glycan structures could be added to proteins in *E. coli* instead of just the normal heptasaccharide (Linton et al. 2005). In 2005, purified Pgl enzymes (C, A, J, H and I) from the *C. jejuni N*-linked protein glycosylation system were used to synthesise the heptasaccharide that modifies *C. jejuni* proteins *in vitro* in a single reaction, on a the lipid carrier, UDP (Glover et al. 2005). In *C. jejuni N*-linked protein glycosylation occurs independently of protein translation (Kowarik et al. 2006a). This means that glycans synthesised *in vitro* could be transferred onto fully folded proteins using purified OT enzyme, PglB.

The OT mediated *O*-glycosylation system (Langdon et al. 2009) that modifies *Neisseria* and *P. aeruginosa* pili (section 1.3.2.3) has also been expressed in *E. coli* (Faridmoayer et al. 2007). Like PglB in *C. jejuni*, the OT enzymes (PglL in *Neisseria* and PilO in *P. aeruginosa*) show relaxed glycan specificity. PilO can only transfer short lipid-linked oligosaccharides, however, PglL can transfer very diverse lipid-linked polysaccharides. In fact it was able to transfer sugars found in peptidoglycan onto *E. coli* pili (Faridmoayer et al. 2007). This shows the potential for these *O*-linked systems to be used for glycoengineering purposes.

1.6 Project aims and objectives

The primary objective of this research project was to investigate the possibility that the Gram-negative Gammaproteobacterium, *Photorhabdus luminescens* produces glycoproteins and thus contains a protein glycosylation system. An introduction into the

life cycle of *P. luminescens* is provided in section 1.7. Using various bioinformatic tools a number of orthologues to genes known to be involved in protein glycosylation in other organisms such as *C. jejuni* were identified in *P. luminescens.*

In order to investigate whether *P. luminescens* produces glycoproteins, lectins were employed in various ways, for example in enzyme linked lectin assays, western blots and bound to affinity matrices in order to isolate and characterise glycoproteins from the organism. Isolated potential glycoproteins were then identified using MS techniques. Potential putative genes involved in protein glycosylation were mutated by insertional inactivation and the changes in the glycome of the mutant compared to the wildtype were examined.

1.7 *P. luminescens* – life cycle

P. luminescens is a motile, Gram-negative, Gammaproteobacteria belonging to the family Enterobacteriaceae. This bacterium has an extremely interesting and complex life cycle. It is a symbiont of soil nematodes belonging to the family *Heterorhabditidae* (Forst et al. 1997, Forst and Nealson 1996). *P. luminescens* is found in the gut of nematodes during its non-feeding infective stage and is an insect pathogen. The nematode enters the digestive tract of the insect in its larval stage and penetrates the hemocoel of the insect and releases the bacteria into the heomolymph. After infection death of the insect usually occurs within about 48 hours.

Within the hemocoel of the insect many generations of both bacteria and nematode are produced. The bacteria reassosiates with the nematode near the final stages of its development inside the insect. After about a week the insect cadaver bursts open and releases hundreds of thousands of nematodes ready to complete the life cycle all over again. Figure 1.13 shows an example of an insect cadaver that has burst open releasing hundreds of thousands of nematodes.
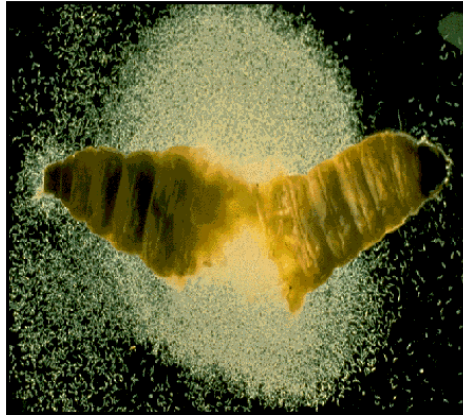
Figure 1.13 An insect cadaver one week after infection with *P. luminescens*

Thousands of nematodes containing *P. luminescens* in their gut emerge from the insect cadaver (USDA Agricultural Research Service).

To date, free living forms of *P. luminescens* have not been isolated from soil or water sources suggesting that the bacteria may not be able to survive in the soil outside of the nematode (Waterfield et al. 2009). In return, the bacteria are required for the killing of the insect larva which provides nutrients for the nematode in order for it to complete its life cycle (Forst et al. 1997, Kaya and Gaugler 1993).

P. luminescens is capable of undergoing phase variation. There are two distinct phases in the life cycle of the organism, phase I and phase II. Phase I is the normal form of the bacterial cells when the bacteria associates with the infective juvenile nematode. Inside the hemolymph of the insect, *P. luminescens* switches to phase II of its life cycle. In phase II many genes are expressed and proteins are produced that are not present in phase I. It secretes extracellular products including lipases, phospholipases, proteases and broad spectrum antibiotics during its stationary phase. The enzymes break down the macromolecules of the insect cadaver to provide nutrients and the antibiotics prevent the growth of other microorganisms on the cadaver and so prevent putrefaction of the cadaver (Forst et al. 1997).

The complete genome sequence of *P. luminescens* strain TTO1 was published in 2003. The genome is 5,688,987 base pairs in size and contains 4,839 predicted protein coding genes. Other proteins encoded in the genome of *P. luminescens* include adhesins, toxins and hemolysins which all play a roll in adhesion of the bacteria to the host cell, infection and killing of the insect (Duchaud et al. 2003). One of the toxins produced by

50

P. luminescens is toxin complex A (TcaA). The tobacco horn worm *Manduca sexta* is generally resistant to bacterial infection but this toxin shows very high oral toxicity for the insect. Figure 1.14 shows an image of *Manduca sexta* infected with luminescent *P. luminescens*.
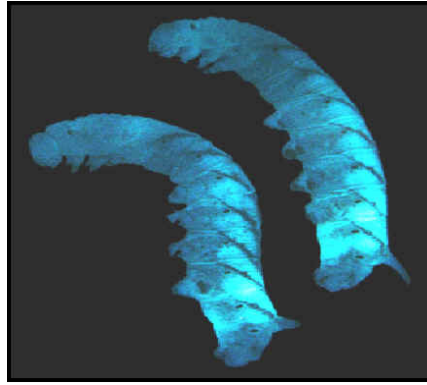
Figure 1.14 *Manduca sexta*

The tobacco horn worm Manduca sexta infected with luminescent *P. luminescens* (University of Bath, 2004).

P. luminescens strain TTO1 encodes many fimbrial genes. According to Duchaud et al. (2003) there are 11 fimbrial gene loci in the organism. Some of the fimbrial gene loci encode proteins similar to type IV pili from other organisms. This large repertoire of pili is likely to be very important for colonisation in nematode gut and invasion of the insect. The organism also produces proteins important in host cell interactions including lectins (Duchaud et al. 2003). However, to date, the cell surface properties of *P. luminescens* have not been well studied.

Homology searches of the genome of *P. luminescens* have identified a number of orthologues to genes involved in protein glycosylation in other organisms, namely *C. jejuni* subsp. This will be discussed further in chapter 5. Protein glycosylation has been shown to play an important role in pathogenesis of other organisms. Because adhesion to the nematode and insect is so important in the life cycle of *P. luminescens*, it is likely that it produces glycoproteins to aid in cell recognition and adhesion. No proteins from *P. luminescens* have previously been shown to be glycosylated.

Chapter 2

Materials and methods

2.1 Bacterial strains, plasmids and primer sequences

Table 2.1 Bacterial strains

Strain	Genotype	Features	Source
Escherichia coli			
XL1-Blue	*end*A1, *rec*A1, *rel*A1, *gyr*A96(nal^R), *thi*-1, *lac*, *gln*V44, *hsd*R17(r$_k^-$, m$_k^+$), *sup*E44, F′ [*pro*AB$^+$ *lac*I^{q}ZΔM15 ::Tn10(Tet^R)]	Cloning strain, Tet^R	Stratagene
XL10-Gold	Δ(*mcr*A)183, Δ(*mcr*CB-*hsd*SMR-mrr)173, *end*A1, *rec*A1, *rel*A1, *gln*V44, *gyr*A96, *sup*E44, *thi*-1 *lac*, F′ [*pro*AB *lac*I^{q}ZΔM15 ::Tn10(Tet^R Amy Cm^R)]	High transformation efficiency, Tet^R, Cm^R, Expression host	Stratagene
Photorhabdus luminescens			
TTO1	Wildtype	Rif^R	Dr. David Clarke (UCC)
wblK Mutant	TTO1 *wblK*::*kan*	Rif^R, Kan^R	This Study
Campylobacter jejuni			
NCTC 11168	Wildtype		Dr. Cyril Carroll (N.U.I. Galway)

Table 2.2 Plasmids

Plasmid	Description	Source/Reference
pBR322	Tet^R, Amp^R, *rop*, pMB1 origin	Bolivar et al. (1977)
pBBR1MCS-5	Broad host range cloning vector, Gm^R, *mob* (Figure 2.1)	Kovach et al. (1995)
pRK600	Provides transfer functions, Cm^R	Finan et al. (1986)
pJQ200sk+	Suicide vector, Gm^R, *mob*, *sacB* from *Bacillus subtilis*, p15a origin (Figure 2.2)	Quandt and Hynes (1993)
pUK4K	Source of Kan^R cassette, Amp^R (Figure 2.3)	Amersham Pharmacia
pBBR1MCS-5 derived plasmids		
pMF1.2 C B/H	pBBR1MCS-5 encoding C-terminally histidine tagged OmpN for expression	This Study
pJQ200sk+ derived plasmids		
pMF2.0 B/P	pJQ200sk+ vector containing 2 kb B/P fragment encoding 5′ end of *wblK* for mutagenesis	This Study
pMF2.1 P/X	pMF2.0 B/P containing 2 kb P/X fragment encoding 3′ end of *wblK* for mutagenesis	This Study
pMF2.2 P	Kan^R cassette insertion in pMF2.1 P/X	This Study
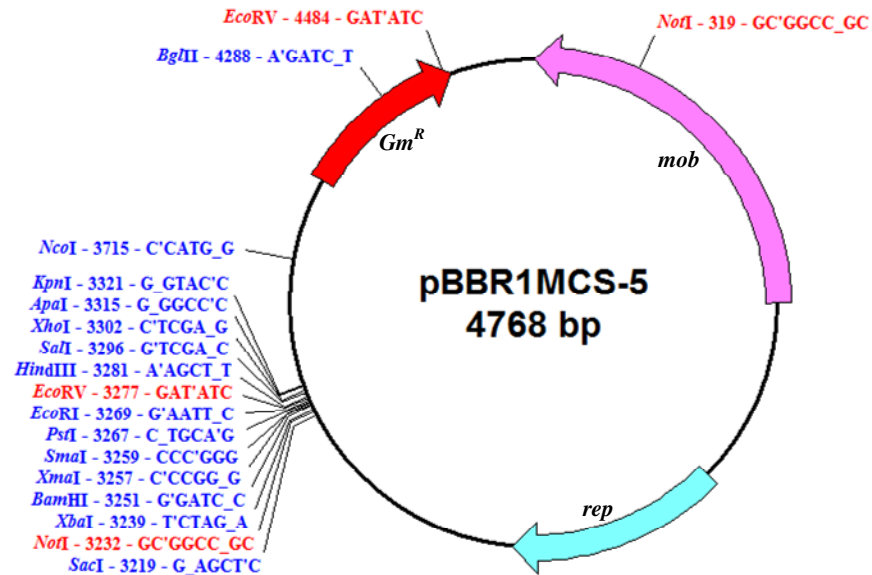
Figure 2.1 pBBR1MCS-5 vector (Kovach et al. 1995)

The pBBR1MCS-5 broad host range plasmid, showing the multiple cloning site (MCS) with restriction enzyme sites (single restriction sites in blue, and double restriction sites in red), the gentamicin resistance gene in red, the origin of replication in light blue and the mobilization site in pink.
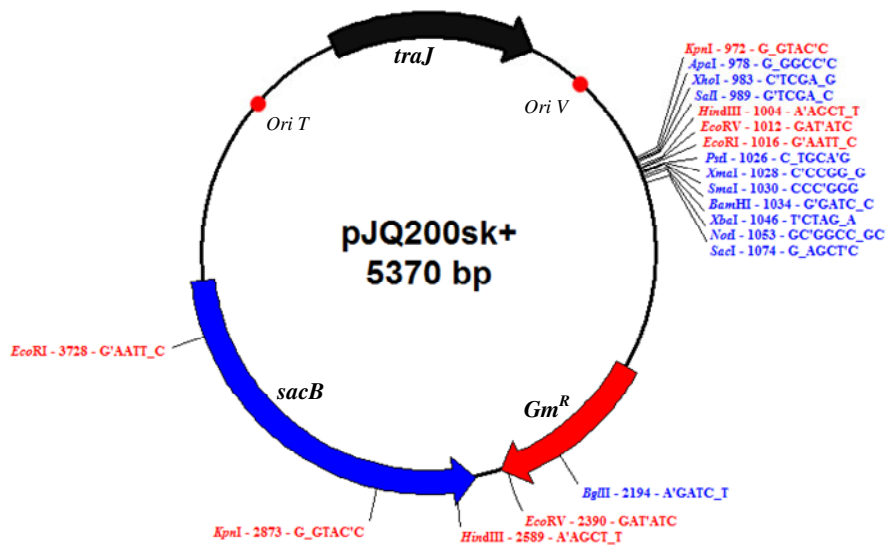
Figure 2.2 pJQ200sk+ vector (Quandt and Hynes 1993)

The suicide vector, pJQ200sk+, showing the gentamicin resistance gene in red and the *sacB* gene in blue. The mobilization site (OriT) and the origin of replication (OriV) are shown. The multiple cloning site is located between bases 972 and 1074. Selected restriction enzyme sites are indicated (unique sites are highlighted in blue while enzymes cutting twice are highlighted in red).

55
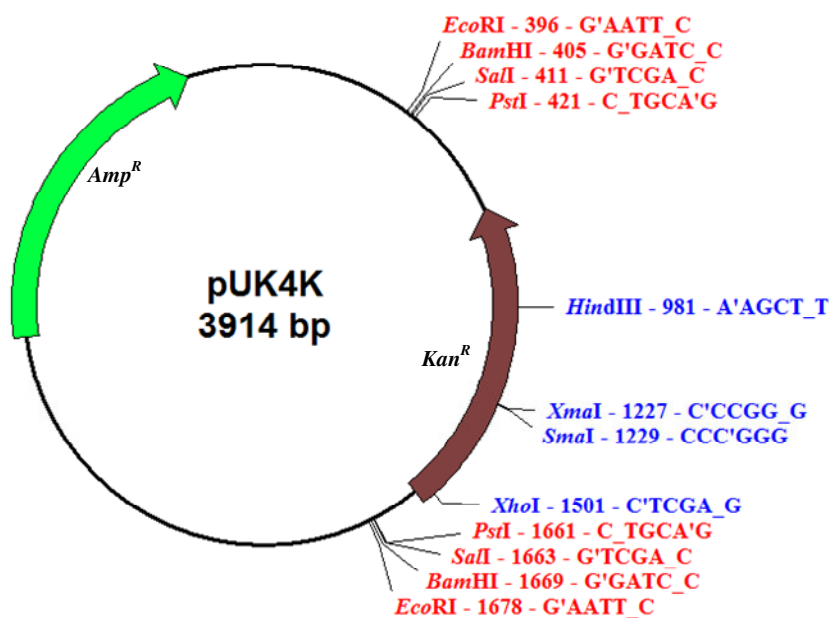
Figure 2.3 pUK4K vector (Amersham Pharmacia)

The pUK4K vector is the source of the kanamycin resistance cassette for mutagenesis. The kanamycin resistance gene is shown in brown. The gene is flanked by common restriction sites for excision of the cassette. The ampicillin resistance gene is highlighted in green.

Table 2.3 Primer sequences (synthesised by Sigma-Aldrich, UK)

Name	Primer sequence (5′→3′)	T$_m$ (°C)
Primers for cloning		
ompN_F1	**GTCAGTAAGCTT**ATGAGGGTAATAATAATGATGAAA	60
ompN_R2	**GTCAGTGGATCCTTAGTGATGGTGATGGTGATG**GAATT GGTAAATCATACCTACG	60
Primers for mutagenesis		
wblK_F1	**ACGTGACATGGATCC**TTCCATGCTTCCTAAAGTAC	56
wblK_R1	**ACGTGACATCTGCAG**ATATTGCGCAAAATCTTGTTC	56
wblK_F2	**ACGTGACATCTGCAG**TTAGGTTCCGATGTTGAAGC	58
wblK_R2	**ACGTGACATCTCGAG**TAGCGTTCTATAGATAAATCTTG	60

Notes: Non-binding bases are in bold type

Restriction sites are underlined and in red

Melting temperatures were calculated using the formula $T_m = 4(G+C) + 2(A+T)$

2.2 Microbiological media

Bacteriological agar, tryptone and yeast extract were supplied by Lab M Ltd. All Oxoid media were supplied by Fannin Healthcare. BBL Brucella broth was supplied by Becton Dickinson UK Ltd. All other chemicals were supplied by Sigma-Aldrich Co. All media were sterilised by autoclaving at 121°C and 15 lb/in^2 for 20 minutes.

- **Lysogeny broth/agar (LB)**

Used for the culturing of *E. coli* strains and *P. luminescens* strains

Tryptone	10 g/L
NaCl	10 g/L
Yeast extract	5 g/L
pH	7.0

For LB agar, 15 g/L bacteriological agar was included.

- **Nutrient broth/agar**

Used for the culturing of *P. luminescens* strains

Oxoid nutrient broth	13 g/L
Oxoid nutrient agar	28 g/L

- ***Campylobacter* selective agar**

Used for the culturing of *C. jejuni*

Oxoid *Campylobacter* Blood-Free Selective Agar Base	45.5 g/L

Once autoclaved and cooled to 55°C, one vial of Oxoid CCDA Selective Supplement that had been reconstituted in 1 ml of sterilised distilled water was aseptically added per 500 ml of agar.

- **BBL brucella broth**

Used for the culturing of *C. jejuni*

     BBL brucella broth   28 g/L

- **SB broth**

Tryptone	30 g/L
Yeast extract	20 g/L
MOPS	10 g/L
pH	7.0

Once sterilised and cooled to 55°C, 20% (w/v) glucose and 1 M $MgCl_2$ solutions that had been filter sterilised were added to final concentrations of 0.4% and 10 mM respectively.

- **SOB medium**

Tryptone	20 g/L
Yeast extract	5 g/L
NaCl	0.5 g/L
KCl	2.5 mM
pH	7.0

The solution was sterilised and allowed to cool to 55°C. Filter sterilised solutions of 1 M $MgCl_2$ and 1 M $MgSO_4$ were added to final concentrations of 10 mM each.

- **SOC medium**

After making SOB medium as described above, filter sterilised 50% glucose was added to a final concentration of 20 mM.

2.3 Solution and buffers

- **TE buffer**

Tris-HCl	10 mM
Na_2-EDTA	10 mM
pH	8.0

- **TAE buffer (50X)**

Tris	242 g/L
Glacial acetic acid	5.71% (v/v)
EDTA	50 mM (from 0.5 M stock)
pH	8.0

The solution was diluted to 1X with dH_2O before use

- **Solutions for 1, 2, 3 method of plasmid preparation (Birnboim and Doly 1979)**

Solution 1

Glucose	50 mM
Na_2-EDTA	10 mM (from 0.5 M stock)
Tris-HCl	25 mM (from 1 M stock)
pH	8.0

Solution 2

NaOH	200 mM (from 1 M stock)
SDS	1% (w/v)

Solution 3

     Potassium acetate   3M

     pH   4.8

To 60 ml of 5 M potassium acetate, 11.5 ml of glacial acetic acid and 28.5 ml of dH_2O was added. The resulting solution was 3 M with respect to potassium and 5 M with respect to acetate.

- **Agarose gel loading dye (6X)**

     Bromophenol blue   0.25% (w/v)

     Xylene cyanol 0.25% (w/v)

     Ficoll (Type 400)   15% (w/v)

Made in dH_2O and sterilised by autoclaving.

- **Ethidium bromide stain**

A 10 mg/ml solution of ethidium bromide was stored at 4°C in the dark. For the staining of agarose gels, 100 µl of this stock was mixed with 1 L of dH_2O. The solution was kept in a plastic tray and covered to protect against the light. Fresh stain was made every 1-2 weeks. Used ethidium bromide stain was collected and ethidium bromide was removed by mixing with a de-staining bag (GeneChoice) overnight. The clear liquid was disposed of routinely, while the ethidium waste was incinerated.

- **TB buffer for competent cells**

     KCl   250 mM

     $CaCl_2$ 15 mM

     PIPES   10 mM

     pH   6.7

The pH of the solution was adjusted with KOH. $MnCl_2$ was then added to a final concentration of 55 mM. The solution was then sterilised through a 0.45 µm sterile filter and stored at 4°C.

- **Sodium phosphate resuspension buffer for the water lysis method (Ward et al. 2000)**

A 0.5 M solution of di-sodium hydrogen orthophosphate (Na_2HPO_4) and a 0.5 M solution of sodium di-hydrogen orthophosphate (NaH_2PO_4) were prepared. The 0.5 M Na_2HPO_4 solution was buffered to pH 7.2 using the 0.5 M NaH_2PO_4 solution. This solution was diluted to 0.1 M with dH_2O prior to use in the water lysis method. Mercaptoethanol was added on day of use to a final concentration of 1 mM.

- **Buffers for Ni-NTA purification of histidine tagged proteins**

Ni-NTA wash buffer

NaH_2PO_4	50 mM
NaCl	300 mM
Imidazole	10 mM
pH	8.0

Ni-NTA elution buffer

NaH_2PO_4	50 mM
NaCl	300 mM
Imidazole	300 mM
pH	8.0

- **SDS-PAGE sample (Laemmli) buffer (5X)** (Laemmli 1970)

Glycerol	50% (v/v)
2-mercaptoethanol	5% (v/v)
SDS	2% (w/v)
Bromophenol blue	0.1% (w/v)
Tris-HCl, pH 6.8	62.5 mM

- **SDS-PAGE running buffer (5X)**

Tris	125 mM
Glycine	960 mM
SDS	0.5% (w/v)
pH	8.3

- **Coomassie blue stain for SDS-PAGE gels**

Methanol	45% (v/v)
dH$_2$O	45% (v/v)
Glacial acetic acid	10% (v/v)
Coomassie blue	0.25% (w/v)

To make destaining solution for coomassie blue stained gels, the coomassie blue was omitted.

- **Schiff solution for the periodic acid-Schiff stain for glycoproteins** (McGuckin and McKenzie 1958)

8 g of potassium metabisulphite was dissolved in 10.5 ml concentrated HCl and the volume was adjusted to 1 litre with dH$_2$O. Pararosaniline (4 g) was added and the solution was stirred for 2 hours, before adding 1 g G-60 Grade charcoal and filtering the solution through Whatman No. 1 filter paper. Schiff reagent is light sensitive and so was stored in a dark bottle, at 4°C.

- **IPG gel strip rehydration stock**

Urea	7 M
Thiourea	2 M
CHAPS	2% (w/v)
IPG buffer 3-10	0.5% (v/v)
Bromophenol blue	0.002% (w/v)

The solution was prepared in dH$_2$O, warmed slightly to dissolve the urea and then stored in 500 µl aliquots at -20°C. When required for use, an aliquot was thawed and to it DL-Dithiothreitol (DTT) was added to a final concentration of 20 mM before use.

- **IPG gel strip SDS equilibration buffer**

Tris	50 mM
Urea	6 M
Glycerol	30% (w/v)
SDS	2% (w/v)
Bromophenol blue	0.002% (w/v)
pH	8.8

The buffer was prepared in dH$_2$O, warmed slightly to dissolve the urea and then 10 ml aliquots were stored at -20°C. When required for use, an aliquot was thawed and to it 100 mg of DTT was added immediately before use for a final concentration of 10 mg/ml.

- **Tris buffered saline (TBS)**

Tris-HCl	20 mM
NaCl	150 mM
CaCl$_2$	1 mM (from 1 M stock)
MgCl$_2$	1 mM (from 1 M stock)
MnCl$_2$	1 mM (from 1 M stock)
pH	7.6

For TBST, the detergent Triton X-100 was added to a final concentration of 0.1% (v/v).

- **Phosphate buffered saline (PBS)**

NaCl	137 mM
NaH$_2$PO$_4$	4.3 mM
KCl	2.7 mM
K$_2$HPO$_4$	1.5 mM
pH	7.4

- **Western blot transfer buffer**

Tris	60 mM
Glycine	149 mM
SDS	0.055% (w/v)

Directly before use, Methanol was added to the buffer in the ratio of 1:4.

- **Lectin affinity chromatography (LAC) equilibration buffer**

Tris-HCl	20 mM
NaCl	150 mM
$CaCl_2$	1 mM (from 1 M stock)
$MgCl_2$	1 mM (from 1 M stock)
$MnCl_2$	1 mM (from 1 M stock)
pH	5.5

- **LAC elution buffer**

Tris-HCl	20 mM
NaCl	500 mM
pH	7.5

For the lectin WGA, *N*-acetylglucosamine was also added to a final concentration of 500 mM before pH adjustment. For the lectin Con A, 200 mM α-methyl mannoside and 200 mM α-methyl glucoside were added to the above list. Finally for agarose bound *Griffonia simplicifolia* lectin I (GSL I) the elution buffer also contained 200 mM galactose and 200 mM *N*-acetylgalactosamine.

- **Trypsin buffer for the in-gel trypsin digestion of proteins (Shevchenko et al. 2007)**

Trypsin	13 ng/µl
Ammonium bicarbonate	10 mM
Acetonitrile	10% (v/v)

2.4 Antibiotics

Table 2.4 Antibiotics

Antibiotic	Stock conc.	Prepared in	Working conc. (solid media)	Working conc. (liquid media)
Ampicillin	100 mg/ml	dH$_2$O	100 µg/ml	100 µg/ml
Rifampicin	20 mg/ml	Methanol	100 µg/ml	50 µg/ml
Tetracycline	10 mg/ml	50% Ethanol	10 µg/ml	10 µg/ml
Chloramphenicol	20 mg/nl	100% Ethanol	20 µg/ml	20 µg/ml
Gentamicin	100 mg/ml	dH$_2$O	20 µg/ml	20 µg/ml
Kanamycin	100 mg/ml	dH$_2$O	30 µg/ml	30 µg/ml

Note: The working concentrations stated in table 2.4 were used for both *E. coli* and *P. luminescens* strains.

2.5 Storing and culturing of bacteria

2.5.1 *E. coli* and *P. luminescens* strains

Glycerol stocks were prepared for each strain. Multiple stocks of the same *P. luminescens* strain were prepared at a time as the bacteria lose viability very quickly if they are repeatedly frozen and thawed. A 0.5 ml aliquot of an exponentially growing culture was added to 0.5 ml of sterile 80% (v/v) glycerol in a cryogenic tube. Stocks were stored at -80°C.

In order to culture *E. coli* and *P. luminescens* an inoculating loop was used to transfer culture from a glycerol stock to an LB or nutrient agar plate (section 2.2). *E. coli* plates were incubated at 37°C overnight and *P. luminescens* plates were incubated at 30°C for 24-48 hours. Working stocks were stored on agar plates at 4°C for up to one week for *E. coli* and up to three days for *P. luminescens*.

For liquid cultures one single colony from a plate was transferred to 5 ml of LB or nutrient broth in a 25 ml tube and tubes were incubated at 37°C for *E. coli* and 30°C for *P. luminescens* in a shaker incubator at 200 rpm overnight. For larger cultures, 1 ml of a 5 ml overnight culture was added to 100 ml of broth in a 250 ml flask or 5 ml of overnight culture was added to 500ml of broth in a 1 L flask and incubated in the same way as 5 ml cultures. Where hosts were harbouring plasmids, the appropriate antibiotic was added to the growth medium (section 2.4).

2.5.2 *C. jejuni* NCTC 11168

Glycerol stocks were prepared by adding a 0.5 ml aliquot of stationary phase culture to 0.5 ml of sterile 40% (v/v) glycerol in a cryogenic tube. Stocks were stored at -80°C.

In order to culture *C. jejuni*, sterile cotton tipped swabs were used to transfer bacteria from a glycerol stock to *Campylobacter* selective agar plates (section 2.2). Plates were immediately transferred to an anaerobic jar, lid side down. An Oxoid gas generating kit was added to the jar. This consisted of a foil sachet with tablets of sodium borohydride, sodium bicarbonate and tartaric acid in a porous membrane container. A 10 ml aliquot of sterilised dH_2O was added to the sachet, activating the system, causing hydrogen and carbon dioxide to be produced. The gas generating kit created an atmosphere of 5-6% oxygen, 10% carbon dioxide and 84-85% nitrogen. The anaerobic jar was quickly sealed and the plates were incubated at 37°C for 48 hours. Working stocks of *C. jejuni* were stored on agar plates at 4°C for no more than two days.

In order to prepare liquid cultures, sterile cotton tipped swabs were used to transfer a large amount of culture from an agar plate to 5 ml of BBL brucella broth (section 2.2) in a 5 ml tube (very little headspace was left at the top of the tube to keep the oxygen content low). The lid of the tube was tightly secured and the cultures were grown at 44°C for 24-48 hours with no shaking. For larger scale cultures, 1 ml of the stationary phase 5 ml culture was added to a 100 ml flask containing 100 ml of BBL brucella broth and grown the same way as the 5 ml cultures.

Because *C. jejuni* grows at very low densities, the cultures are very susceptible to contamination. New glycerol stocks were prepared regularly. *C. jejuni* has a very distinctive colour, smell and morphology so cultures were continually checked to ensure contamination was not present.

2.6 Preparation of total genomic DNA using the Wizard® genomic DNA kit

The kit was used according to the manufacturer's instructions (Promega) with some modifications. An overnight culture of the appropriate organism was prepared and a 1 ml aliquot of this culture was pelleted by centrifugation at 13,000 rpm for 2 minutes. The supernatant was removed and the pellet was resuspended in 600 µl of nuclei lysis solution. The suspension was incubated at 80°C for 5 minutes to lyse the cells and then cooled to room temperature. To the cell lysate 3 µl of RNase solution was added. The tube was inverted several times to mix and incubated at 37°C for 15-60 minutes. To remove the protein in the sample, 200 µl of protein precipitation solution was added and vortexed vigorously for 20 seconds. The sample was incubated on ice for 5 minutes then the precipitated protein was pelleted by centrifugation at 13,000 rpm for 3 minutes. The supernatant was transferred to a fresh tube and 600 µl of phenol/chloroform/isoamylalcohol (25:24:1) was added. The suspension was mixed slowly by inversion. The aqueous and organic layers were separated by centrifugation at 13,000 rpm for 5 minutes. The aqueous (top) layer was transferred to a fresh tube and the phenol/chloroform/isoamylalcohol (25:24:1) purification step was repeated. The aqueous layer was removed and added to a fresh tube containing 600 µl of room temperature isopropanol. The sample was mixed gently by inversion until thread-like strands of DNA formed a visible mass. The genomic DNA was pelleted by centrifugation at 13,000 rpm for 2 minutes. The pellet was washed twice with 100 µl of 70% ethanol, and air dried for 10-15 minutes. The dried pellet was resuspended in 100 µl of DNA rehydration solution overnight at 4°C and stored at 4°C.

2.7 Plasmid preparation by the 1, 2, 3 method

The method described is adapted from the procedure described by Birnboim and Doly (1979). In this method the high molecular weight chromosomal DNA undergoes selective alkaline denaturation, while the covalently closed circular plasmid DNA remains double stranded. A 1.5 ml volume of overnight bacterial culture was centrifuged at 13,000 rpm to pellet the cells. The supernatant was discarded and the pellet thoroughly resuspended in 200 µl of solution 1 (section 2.3). The tube was left at room temperature for 5 minutes. A 200 µl aliquot of solution 2 (section 2.3) was added and mixed by inversion several times. The tube was left on ice for 5 minutes. A 200 µl aliquot of solution 3 (section 2.3) was added to the tube and inverted several times. The tube was left on ice for 10 minutes. A clot of chromosomal DNA was formed and this was pelleted by centrifugation at 13,000 rpm for 10 minutes. The supernatant was transferred to a fresh tube containing 450 µl of phenol/chloroform/isoamylalcohol (25:24:1), mixed by vigorous vortexing and centrifuged for 10 minutes at 13,000 rpm to separate the aqueous and organic layers. The aqueous layer was transferred to a fresh tube and an equal volume of isopropanol was added. The contents of the tube were mixed by inversion and left on ice for 10 minutes and then centrifuged for 10 minutes at 13,000 rpm to pellet the plasmid DNA. The pellet was washed with 100 µl of 70% ethanol and centrifuged at 13,000 rpm for 5 minutes. The ethanol was removed and the pellet was dried in a vacuum dryer for 5 minutes. The dried pellet of plasmid DNA was resuspended in 50 µl of TE buffer and stored at 4°C.

2.8 Plasmid DNA isolation using the GenElute™ HP plasmid miniprep kit

The kit was used according to the manufacturer's instructions (Sigma-Aldrich). A 1.5 ml aliquot of an overnight bacterial culture was pelleted by centrifugation at 13,000 rpm for 5 minutes. The cell pellet was resuspended in 200 µl of resuspension solution. 200 µl of cell lysis solution was added to the suspension to lyse the cells. This was allowed to stand at room temperature for 5 minutes. A 350 µl aliquot of neutralisation/binding buffer was added to the tube and mixed by inversion to precipitate the cell debris and

chromosomal DNA. The precipitate was pelleted by centrifugation at 13,000 rpm for 10 minutes. In the meantime a spin column in a tube was prepared for use by the addition of 500 µl of column preparation solution and centrifugation at 13,000 rpm for 1 minute. The flow-through was discarded. The supernatant was transferred to the column and centrifuged at 13,000 rpm for 1 minute to bind the plasmid DNA to the column. The flow through was discarded. A 750 µl aliquot of wash solution was added to the column and centrifuged at 13,000 rpm for 1 minute. The flow through was discarded and the column was dried by centrifugation at 13,000 rpm for 2 minutes. The column was transferred to a fresh tube and 100 µl TE buffer was added. This was allowed to stand for 2 minutes before elution of the plasmid DNA by centrifugation at 13,000 rpm for 1 minute. Plasmid DNA was stored at 4°C.

2.9 Agarose gel electrophoresis

DNA was analysed by agarose gel electrophoresis. Typically 0.7% agarose gels prepared in 1X TAE buffer (section 2.3) were used. The agarose was dissolved in the buffer by boiling and poured into plastic trays with a plastic comb fitted to create sample wells and allowed to set. The gels were run in a BioRad horizontal gel apparatus using 1X TAE as the running buffer. The DNA samples were mixed with loading dye (section 2.3) and loaded into the wells. The loading dye aided loading and visualisation of the migration process during electrophoresis. Gels were run at 120 V for 20-40 minutes depending on the gel length. Staining was achieved by immersion in ethidium bromide staining solution (section 2.3) for 15 minutes. DNA was visualised on a UV transilluminator. 1 kb DNA ladder (Invitrogen) was run as a molecular size marker in every gel.
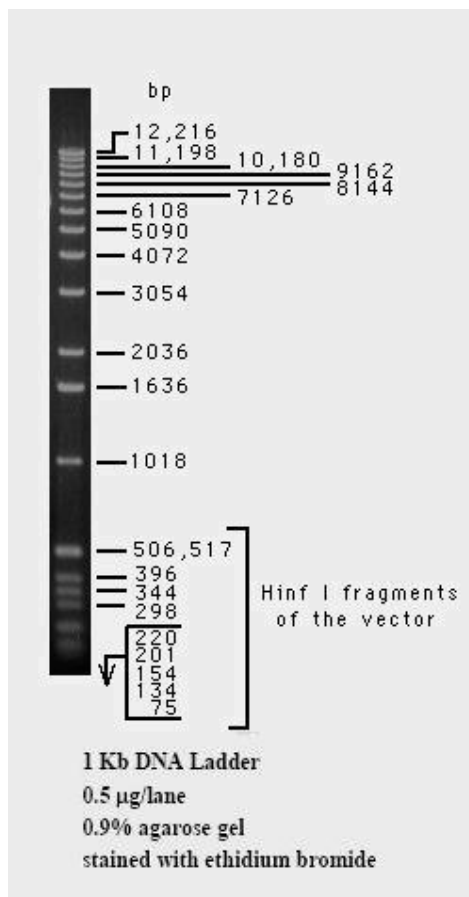
Figure 2.4 1 kb DNA ladder, Invitrogen.

A representative image of the DNA ladder used for agarose gel electrophoresis in this study.

2.10 Isolation of DNA from agarose gels

DNA that had been separated and visualised by agarose gel electrophoresis could be excised and purified. To purify DNA from agarose gels the HiYield™ Gel/PCR DNA extraction kit (Real Biotech Corporation) was used according to the manufacturer's instructions. The DNA to be purified was run as normal on an agarose gel that contained 1 ng/ml SYBR® Safe DNA gel stain (Invitrogen). The DNA was visualised using a blue light transilluminator (Invitrogen) and the bands were excised using a clean scalpel. The gel slice was placed in a microfuge tube and 500 µl of DF buffer was added. The gel slice was dissolved in the buffer by incubating at 55°C for 10-15 minutes. The solution was added to a DF column in a collection tube and centrifuged at 13,000 rpm for 30 seconds. The flow through was discarded and 750 µl of wash buffer

was added to the column. The column was centrifuged as before. The flow through was discarded and the column was centrifuged at 13,000 rpm for 2 minutes to dry the column. The dried column was transferred to a new tube and 30 µl of elution buffer was added to the centre of the column matrix allowed to stand for 2 minutes to absorb into the matrix. The purified DNA was eluted by centrifugation for 2 minutes at 13,000 rpm. Purified DNA was stored at 4°C.

2.11 Competent cells

2.11.1 Preparation of electrocompetent cells

A glycerol stock of an *E. coli* strain was streaked on LB agar and incubated at 37°C overnight. A 5 ml aliquot of SB broth (section 2.2) was inoculated with a single colony from this plate and incubated at 37°C overnight. A 1 ml inoculum of this overnight culture was added to two 200 ml flasks of pre-warmed SB broth. These cultures were grown at 37°C and 220 rpm until an O.D.$_{600 \text{ nm}}$ of $0.7 - 0.8$ was reached. The flasks were cooled on ice for 15 minutes. From this point on, all steps were preformed on ice or in a 4°C cold room. The cultures were transferred to two cooled, sterile 250 ml centrifuge tubes. The cells were pelleted in a Beckman JA-21 centrifuge at 4,000 rpm at 4°C for 15 minutes. The cell pellets were resuspended in 200 ml of pre-cooled, sterilised 10% glycerol and centrifuged as before. The cell pellets were then resuspended in 100 ml of pre-cooled, sterilised 10% glycerol. The two suspensions were transferred to one centrifuge tube and the suspension was centrifuged as before. The cell pellet was then resuspended in 25 ml of pre-cooled, sterilised 10% glycerol, transferred to a smaller centrifuge tube and centrifuged at 3,000 rpm at 4°C for 15 minutes. Finally the cell pellet was resuspended in 0.5 ml of pre-cooled, sterilised 10% glycerol. The cell suspension was dispensed into 50 µl aliquots in sterile cooled microfuge tubes. These were flash frozen in -80°C ethanol and stored at -80°C. The electrocompetent cells were then ready for electroporation. Cells prepared using this method frequently gave transformation efficiencies of 10^9 transformants/µg DNA.

2.11.2 Electroporation of electrocompetent cells

To 50 µl of electrocompetent cells that had been thawed on ice, 1 µl of plasmid DNA or a ligation mixture was added. An electroporation cuvette was also cooled on ice. The BioRad Gene Pulser electroporator was set to 2,500 volts, 200 ohms and 25 µFaraday when transforming *E. coli* cells. The cells were transferred to the cooled electroporation cuvette and any air pockets between the two metal plates in the cuvette were removed by tapping the cuvette on a hard surface. The cuvette was placed in the electroporator and the cells were shocked for approximately 5 seconds. Immediately after electroporation, 800 µl of SOC broth (section 2.2) was added to the cuvette. This suspension was then transferred to a sterile microfuge tube and then incubated in a 37°C water bath for 60 minutes. A 100 µl aliquot of the resulting transformation mixture was then plated on the appropriate selective media and incubated at 37°C overnight.

2.11.3 Preparation of chemical competent cells by the TB method

This method was performed as described by Inoue et al. (1990). A glycerol stock of an *E. coli* strain was streaked on LB agar and incubated at 37°C overnight. A 5 ml aliquot of LB broth was inoculated with a single colony from this fresh plate and incubated at 37°C overnight while shaking at 200 rpm. 1 ml of this overnight culture was used to inoculate 250 ml of SOB medium (Section 2.2) in a 1 L flask. The culture was grown at 37°C, shaking at 200 rpm until an $O.D._{600 \text{ nm}}$ of 0.6 was reached. The flask was then placed on ice for 10 minutes. All subsequent steps were carried out at 4°C. The culture was transferred to a sterile 250 ml centrifuge tube and centrifuged in a Beckman JA-21 centrifuge at 3,000 rpm for 5 minutes. The supernatant was discarded and the pellet was resuspended in 80 ml of ice-cold TB buffer (Section 2.3). The suspension was placed on ice for 10 minutes and centrifuged as before. The resulting cell pellet was gently resuspended in 20 ml of ice-cold TB buffer and DMSO was added to a final concentration of 7%. After incubation in an ice bath for 10 minutes the cell suspension was dispensed into 200 µl aliquots in microfuge tubes. The cells were then flash frozen in -80°C ethanol and stored at -80°C. The competent cells were then ready for transformation. Cells prepared using this method frequently gave transformation efficiencies of 10^8 transformants/µg DNA.

2.11.4 Transformation of chemical competent cells

A 2 µl aliquot of plasmid DNA or a ligation mixture was mixed gently with 200 µl of competent cells that had been thawed on ice. The mixture was kept on ice for 30 minutes. The cells were heat shocked at 42°C for 30 seconds and then quickly placed back on ice for 2 minutes. An 800 µl aliquot of LB broth was added to the cells and the mixture incubated in a 37°C water bath for 1 hour. A 100 µl aliquot of the resulting transformation mixture was then plated on the appropriate selective media and incubated at 37°C overnight.

2.11.5 Determination of competent cell efficiency

Competent cell efficiency is defined in terms of the number of colony forming units obtained per µg of transformed plasmid DNA. A 25 ng/µl stock of pBR322 plasmid DNA (section 2.1) was diluted to 250 pg/µl, 25pg/µl and 2.5 pg/µl. Each dilution was transformed as described in sections 2.11.2 and 2.11.4. The cell efficiency was calculated from the number of colonies obtained, taking into account the dilution factor and the fraction of culture transferred to the spread plate.

2.12 *In silico* analysis of DNA and protein sequences

- The *P. luminescens* genome sequence was obtained from the sequencing study, *Photorhabdus luminescens* subsp. laumondii TTO1 (NC_005126) and is accessible at http://cmr.tigr.org/tigr-scripts/CMR/GenomePage.cgi?org=ntpl01
- The *C. jejuni* subsp. jejuni 81-176 genome sequence was obtained from the study *Campylobacter jejuni* subsp. jejuni 81-176 (NC_008787) and is accessible at http://cmr.jcvi.org/tigr-scripts/CMR/GenomePage.cgi?org=gcj81176
- BLAST programs at NCBI, www.ncbi.nlm.nih.gov were used to identify homologous sequences in GenBank.
- DNA and protein sequences were aligned using the MultAlin programme http://prodes.toulouse.inra.fr/multalin/multalin.html and the Genedoc program, available to download from http://www.nrbsc.org/

- DNA sequences were analysed for restriction enzyme sites using the Webcutter 2.0 tool at http://rna.lundberg.gu.se/cutter2/
- DNA sequences analysed using the pDRAW32 program downloaded at http://www.acaclone.com/
- Protein sequences were analysed online at ExPASy Molecular Biology Server http://us.expasy.org/
- The MASCOT search engine available at http://www.matrixscience.com/ was used to identify proteins using peptide mass fingerprinting

2.13 Bacterial conjugation by tri-parental mating

Two *E. coli* cultures, one containing the donor plasmid and another carrying the mobilising plasmid, pRK600 (section 2.1), were grown to stationary phase in 5 ml of LB broth containing appropriate antibiotics. A 0.75 ml aliquot of each of the cultures was taken, mixed together and pelleted by centrifugation at 13,000 rpm for 1 minute. The cell pellet was resuspended in 100 µl of fresh LB broth and then spotted onto the centre of an LB agar plate with no antibiotics and incubated overnight at 37°C. The following day the bacteria were removed from the plate and resuspended in 3 ml of fresh LB broth. This mixture of *E. coli* was known as the intermediate. Meanwhile a 5 ml culture of *P. luminescens* (the recipient) had been grown to stationary phase in LB broth. A 0.75 ml aliquot of the intermediate was added to 0.75 ml of the recipient culture and pelleted as before. The cells were resuspended in 100 µl of fresh LB broth, spotted onto the centre of an LB agar plate and incubated at 30°C overnight. As negative controls, the intermediate and recipient strains were also spotted on agar plates separately and carried through the procedure. The bacteria from each plate were resuspended in 2 ml of LB broth, dilutions of the suspensions were made and 100 µl of each dilution was spread on selective media (section 2.4). Plates were incubated at the appropriate temperatures and transconjugates were visible within two days.

2.14 Protein expression

Sterile 250 ml conical flasks containing 50 ml of LB or nutrient broth were inoculated with 1 ml of a 5 ml overnight saturated stationary phase culture harbouring the expression plasmid. Selective antibiotics and sterile glycerol to a final concentration of 20 mM were added to the flask. The culture was incubated at 200 rpm and 37°C for *E. coli* or 30°C for *P. luminescens* until an O.D.$_{600\ nm}$ of 0.4-0.6 was reached. At this point expression was induced with the addition of 1 mM IPTG. Cultures were then incubated overnight.

2.15 Extracellular, periplasmic, cytoplasmic and membrane protein isolation using the water lysis method

The water lysis method was preformed as described by Ward et al. (2000). The method uses osmotic pressure to lyse the cells. A 50 ml culture of the desired bacteria was prepared as described in section 2.5 or section 2.14. The culture was centrifuged in a Beckman JA-21 centrifuge at 6,500 rpm for 8 minutes at 10°C. The supernatant (extracellular fraction) was decanted and stored at -20°C. The pellet was resuspended in 10 ml of 0.2 M Tris-HCl, pH 8.0 and left stirring for 20 minutes at room temperature. At this point 9.6 ml of dH$_2$O was aliquoted in anticipation of the water lysis step. At time zero, 4.85 ml of 1 M sucrose/ 0.2 M Tris, pH 8.0/ 1 mM EDTA was added. At time 1.5 minutes, 65 µl of 10 mg/ml lysozyme solution prepared in 1 M sucrose/ 0.2 M Tris, pH 8.0/ 1 mM EDTA buffer was added. At time 2 minutes, the prepared 9.6 ml aliquot of dH$_2$O was added. The suspension was left to stir for 20 minutes. Spheroplasts were formed and sedimented by centrifugation at 18,000 rpm for 20 minutes at 4°C. The supernatant was retained and stored at -20°C (periplasmic fraction) and the spheroplasts were completely resuspended in 15 ml of dH$_2$O using a 15 ml homogeniser. This was left stirring for 30 minutes at room temperature. The suspension was sedimented at 18,000 rpm for 20 minutes at 4°C. The supernatant was retained (cytoplasmic fraction) and stored at -20°C. The preparation was then washed by resuspending the pellet in 30 ml of sodium phosphate resuspension buffer (section 2.3) using a homogeniser followed by centrifugation at 18,000 rpm for 20 minutes at 4°C. The wash step was repeated. The

final pellet was completely resuspended in 500 µl ice cold resuspension buffer using a 1 ml homogeniser. These samples were kept on ice and stored at -80°C (membrane fraction). The protein concentration in each fraction was determined using the BCA Assay (section 2.17).

2.16 Immobilised metal affinity chromatography (IMAC)

2.16.1 Protein purification using IMAC with Ni-NTA resin

A 1 ml aliquot of nickel-nitrilotriacetic acid resin (Ni-NTA, Invitrogen) was mixed with 5 ml of periplasmic protein sample (section 2.15) at 4°C overnight. The mixture was then poured into a 0.7 x 10 cm column and the resin was allowed to settle. The resin was drained and the flow through collected. The column was then washed with 10 ml of Ni-NTA wash buffer (section 2.3) and the wash was collected. The resin was subsequently washed with wash buffer supplemented with 20, 40, 60, 80 and 100 mM imidazole using 5 ml of buffer each time. Each wash was collected. Finally the bound protein was eluted using 10 ml of Ni-NTA elution buffer (section 2.3). Imidazole was then removed and the proteins were concentrated using vivaspin 15 centrifugal ultrafiltration devices (Sartorius Stedim Biotech). The sample was transferred to the unit and centrifuged in a Hettich Rotanta 460R centrifuge fitted with a swing-out rotor (model number 5624) at 3,800 rpm for 20 minutes. The protein was washed twice in PBS, and concentrated to a final volume of 500 µl. Concentrated protein samples were stored at -20°C.

2.16.2 Recharging of Ni-NTA resin

The Ni-NTA resin was routinely recharged prior to re-use. The used resin was poured into a column and washed with 2 column volumes of dH_2O, followed by 2 column volumes of 50% (v/v) ethanol. The resin was then stripped with 2 column volumes of 100 mM EDTA, pH 8.0. Remaining impurities were removed with 2 column volumes of 200 mM NaCl, followed by 2 column volumes of dH_2O. Hydrophobically bound

proteins and lipoproteins were removed by washing with 10 column volumes of 30% isopropanol for 30 minutes, followed by 10 column volumes of water. The resin was then recharged by adding 2 column volumes of 100 mM $NiSO_4$. The resin was washed again with 10 column volumes of dH_2O.

2.17 Protein quantification by the BCA assay

The bicinchoninic acid (BCA) assay was preformed to determine total protein content in a sample using the QuantiPro™ BCA Assay Kit (Sigma). The assay gives a linear response from 0.5 to 30 µg/ml. A range of protein standards from 0 to 30 µg/ml were prepared by diluting the protein standard solution provided (bovine serum albumin (BSA)) in PBS (section 2.3). A range of dilutions for the unknown samples were also prepared in PBS. A 150 µl aliquot of each standard/sample was added to a 96 well plate in triplicate. BCA reagent was prepared by mixing QuantiPro Buffer QA, QuantiPro Buffer QB and 4% Copper (II) Sulphate pentahydrate solution QC in the ratio of 25:25:1. To each sample 150 µl BCA reagent was added. The samples were incubated at 37°C for 2 hours. The absorbance of each sample was read at 570 nm. A standard curve was constructed using the BSA standards and the protein concentrations of the unknown samples were determined from this.

2.18 SDS-PAGE

Protein samples were analysed by SDS-PAGE.

2.18.1 Gel preparation

Table 2.5 outlines the recipe used to prepare a 15% acrylamide resolving gel and 4% acrylamide stacking gel.

Table 2.5 SDS-PAGE gel recipes

Solution	15% Resolving gel	4% Stacking gel
dH$_2$O	1.56 ml	1.538 ml
1.5 M Tris-HCl, pH 8.8	1.625 ml	-
0.5 M Tris-HCl, pH 6.8	-	625 µl
20% (w/v) SDS	32.5 µl	12.5 µl
Acrylamide/bis-acrylamide 30%/0.8% (w/v)	3.25 µl	335 µl
10% Ammonium persulphate (APS)	32.5 µl	12.5 µl
TEMED	3.25 µl	2.5 µl

The reagents were added in the order listed in the table. For use in 2-D electrophoresis, 0.002% bromophenol blue was also added to the stacking gel after the addition of dH$_2$O. APS and TEMED were always added last to the mix as they polymerise the acrylamide. Gels were cast using the ATTO vertical mini electrophoresis system. The resolving gel was poured immediately after the addition of APS and TEMED, leaving about 1.5 cm of space at the top and overlayed with a layer of room temperature isopropanol. Once the gel was polymerised, the isopropanol was removed and the interface was rinsed with dH$_2$O. The stacking gel was poured on top of the resolving gel. A comb was immediately inserted into the top of the gel to create wells. The stacking gel was allowed to set.

2.18.2 Sample preparation and application

In a microfuge tube 8 µl of sample was added to 2 µl of 5X sample buffer (section 2.3). For each gel that was run, one lane was used for a relative molecular weight marker. A 10 µl aliquot of broad range protein marker (2-212 kDa) from NEB was used for coomassie or silver stained gels and 12 µl broad range prestained protein marker (7-175 kDa) from NEB was used for Schiff stained gels and gels to be used for western blotting. A representative image of the protein markers used in this study is shown in figure 2.5.
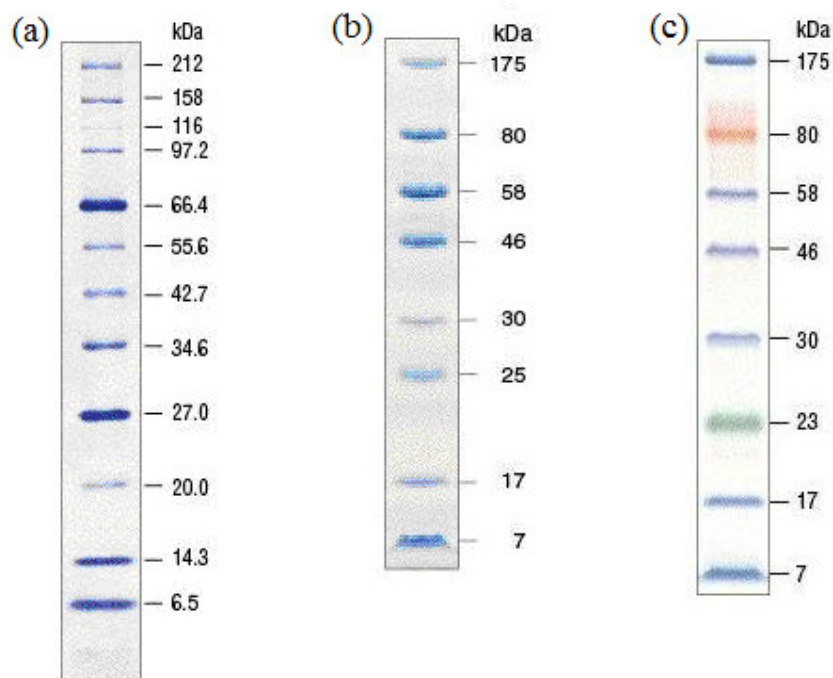
Figure 2.5 (a) Broad range protein marker (2-212 kDa), (b) Prestained protein marker, broad range (7-175 kDa), (c) ColorPlus prestained protein marker, broad range (7-175 kDa), New England Biolabs.

Representative images of protein markers used in this study for SDS-PAGE. (a) 10 µl of broad range marker on a 15% SDS-PAGE gel stained with coomassie blue. The bands at 66.4 kDa and 27.0 kDa have double the intensity to serve as reference bands. (b) 12 µl of prestained protein marker on a 15% SDS-PAGE gel. (c) 12 µl of ColorPlus prestained protein marker on a 15% SDS-PAGE gel. There is one orange band at 80 kDa and a green band at 23 kDa to serve as reference bands.

Before loading on a gel, samples were heated at 100°C in a heating block for 5 minutes. The gel was positioned in the electrophoresis chamber with 1X SDS-PAGE running buffer, the comb was removed and the wells were flushed with buffer to remove any unpolymerised acrylamide. Once cooled to room temperature the samples were loaded into the wells and the gel was run at 30 mA for 80 minutes at room temperature.

2.19 2-Dimensional gel electrophoresis

For a more comprehensive separation of proteins, 2-D gel electrophoresis was carried out. For the first step in 2-D gel electrophoresis, the isoelectric focusing (IEF) step, immobilised pH gradient (IPG) dehydrated gel strips were used. Immobiline^TM DryStrip

IPG strips, 7 cm in length with a linear pH gradient of 3-10 were purchased from GE Healthcare. IPG strips were stored at -20°C. Rehydration of IPG strips was carried out using the IPGbox kit and IEF was performed using the Ettan IPGphor 3 instrument, both manufactured by GE Healthcare.

2.19.1 Rehydration of IPG strips and addition of sample

The protein sample to be separated was prepared by adjusting the concentration to 1-3 µg/µl in dH_2O. Protein sample was loaded onto the IPG strip using the rehydration loading technique. A 10 µl aliquot of the prepared protein sample was added to 115 µl of IPG gel strip rehydration stock (section 2.3) for a total volume of 125 µl. The IPGbox was placed on a level surface and a reswell tray was positioned in the IPGbox. The solution was pipetted into a channel of the reswell tray and spread out over approximately 7 cm. Holding the IPG strip with a forceps, the protective plastic backing was removed and the strip was placed gel-side down in the solution. The strip was gently slid back and forth a few times to distribute the solution evenly over its surface also ensuring that no air bubbles were trapped under the strip. The lid of the IPGbox was closed and the strips were allowed to rehydrate at room temperature for 10-24 hours.

2.19.2 Preparation of manifold

A ceramic manifold was used to hold the IPG strips during the IEF run. The manifold contained 12 channels for the simultaneous separation of 12 different samples. The manifold was cleaned using manifold cleaning detergent and non-abrasive cloths, rinsed with dH_2O and allowed to air dry. The Ettan IPGphor platform was cleaned and the manifold was positioned on it. A spirit level was used to ensure the manifold was level. Exactly 108 ml of Immobiline^{TM} DryStrip cover fluid was measured out and distributed evenly over the 12 manifold channels (9 ml per channel).

A rehydrated IPG strip was positioned gel side up under the cover fluid in a manifold channel so that the anodic (+) end of the strip was directly above the anode of the

instrument and vice versa. Two electrode wicks were moistened with 150 µl of dH$_2$O each and were placed on the strip so that each wick overlapped the end of the gel on the IPG strip. With the electrode cams in the open position, the electrode assembly was placed on top of the wicks, ensuring that electrodes were touching the part of the wicks that were in contact with the gel. The cams were swivelled into the closed position under the external lip of the manifold and the lid of the Ettan IPGphor unit was closed.

2.19.3 1st dimension IEF

The Ettan IPGphor unit was controlled by the Ettan IPGphor 3 control software. The PC and IPGphor unit were powered on and communication between them was established. A protocol detailing the run conditions was created and the number of strips in use was entered. The protocol was transferred to the instrument and initiated. The typical run conditions used for 7 cm, pH 3-10 strips are outlined in table 2.7.

Table 2.6 Run conditions for isoelectric focusing of 7 cm, pH 3-10 IPG strips

	Step voltage mode	Voltage (V)	Time (h:min)	kVh
1	Step and Hold	300	2:30	750
2	Gradient	1000	0:30	325
3	Gradient	5000	1:20	4000
4	Step and Hold	5000	0:25	2083
Total			4:45	7158

Following IEF, the strips were removed from the manifold and if the 2nd dimension was not to be carried out immediately they were stored in separate tubes at -80°C.

2.19.4 2nd dimension SDS-PAGE

A 15% SDS-PAGE resolving gel was prepared as outlined in section 2.18.1. An IPG strip was removed from the freezer and 10 ml of IPG gel strip SDS equilibration buffer (section 2.3) was added to the tube. The strip was equilibrated for 30 minutes at room

temperature on a rocker. The isopropanol was washed from the surface of the polymerised resolving gel and, using a forceps, the IPG strip was removed from the equilibration buffer and placed carefully on top of the separating gel so that it was in contact with the gel surface and no air bubbles were present. The gel strip was pushed over to one side to leave space for a molecular weight marker on the other. A 4% stacking gel was prepared (section 2.18.1) and poured on top of the IPG strip. A 12 well comb was inserted and the gel was allowed to polymerise. A molecular weight marker sample was prepared and applied to the gel and the gel was run as outlined in section 2.18.2. Staining was performed using the silver stain method (section 2.20).

2.20 SDS-PAGE and 2-D gel staining

SDS-PAGE and 2-D gels were removed from the electrophoresis apparatus and washed with dH_2O.

- **Coomassie blue staining**

Gels were immersed in coomassie blue staining solution (section 2.3) and left stirring overnight at room temperature. The staining solution was removed and gels were destained overnight in destain solution (section 2.3). Destained gels were rinsed in dH_2O.

- **Silver staining**

When a higher degree of sensitivity was required the silver staining method was used. This method is outlined in table 2.6.

Table 2.7 Silver staining of SDS-PAGE gels

Step	Reagent (50 ml per gel)	Duration
Fix	30% (v/v) ethanol, 10% (v/v) acetic acid	1 hour to overnight
Rinse	20% (v/v) ethanol	15 minutes
Rinse	dH_2O	3 x 5 minutes
Sensitise	0.01% (w/v) $Na_2S_2O_3$	1 minute
Rinse	dH_2O	2 x 20 seconds
Silver stain	0.1% (w/v) $AgSO_4$, 0.026% (v/v) formaldehyde	30 minutes
Rinse	dH_2O	2 x 20 seconds
Develop	3% (w/v) Na_2CO_3, 0.019% (v/v) formaldehyde, 0.002% (w/v) $Na_2S_2O_3$	Until bands appear
Stop	5% (w/v) Tris, 2.5% (v/v) acetic acid	1 minute

- **Periodic acid-Schiff stain for glycoproteins**

If required, gels were first stained with coomassie blue to stain all proteins. Destained gels were completely immersed in methanol/acetic acid/dH_2O (40:5:55) containing 1% (w/v) periodic acid for 1 hour, followed by overnight incubation in methanol/acetic acid/dH_2O (40:5:55) solution. The gel was subsequently immersed in Schiff solution (section 2.3) and incubated with gentle agitation for 50 minutes in the dark at 4°C. The gel was washed three times for 30 minutes each time with 0.5% (w/v) potassium metabisulphite solution, using fresh solution for each wash. Excess stain was removed by washing with dH_2O. Glycoproteins were evident as dark pink bands against a pale pink background. Non-glycoprotein bands remained blue.

2.21 Preparation of polyclonal antibody against *P. luminescens* whole cells

P. luminescens whole cell antigen was prepared by growing a 500 ml culture of wildtype *P. luminescens* in LB broth as described in section 2.5.1. This culture was grown at 30°C overnight. The following day the culture was separated into two sterile 250 ml centrifuge tubes and the cells were pelleted in a Beckman JA-21 centrifuge at

6,500 rpm for 10 minutes. The cells in each tube were resuspended in 25 ml of PBS and then mixed together in a 50 ml centrifuge tube and pelleted using the Hettich Rotanta 460R centrifuge fitted with a swing-out rotor (model number 5624). The cell pellet was washed 3 more times using 25 ml of PBS to remove any traces of broth remaining. The final cell pellet was resuspended in 10 ml of PBS. The cells were heat killed by heating them in an autoclave to 70°C and holding at this temperature for 20 minutes. The antigen sample was sent to GenScript Corporation (New Jersey) where a polyclonal antibody was raised to it. The antibody was produced in a rabbit and was purified using Protein A affinity chromatography. The antibody was shipped in lyophilised form and resuspended to a final concentration of 5.5 mg/ml in PBS with 0.02% sodium azide and stored in 20 µl aliquots at -20°C.

2.22 Western blotting

A 1-D or 2-D gel was run as outlined in sections 2.18 and 2.19. Four pieces of blotting paper and a piece of nitrocellulose with a pore size of 0.45 µm were cut to exactly the same dimensions as the gel. The gel, blotting paper and nitrocellulose were soaked in western blot transfer buffer (section 2.3) for 15 minutes and assembled on a semi-dry electroblotter. Two pieces of blotting paper were placed first, overlaid by the nitrocellulose membrane, then the gel and finally the last two sheets of blotting paper. Any air bubbles were removed. Transfer occurred at 10 V for 30 minutes.

To detect glycoproteins the membrane was blocked with Carbo-Free blocking solution (Vector Laboratories), diluted to the correct concentration in PBS (section 2.3). The membrane was washed four times for 5 minutes in TBST then incubated for 1 hour with a 1 µg/ml solution of a biotin labelled lectin of choice in TBST. The membrane was washed again then incubated with 1:10,000 horseradish peroxidase (HRP) labelled murine anti-biotin antibody in TBS. After a final washing the membrane was developed using 15 ml of dH$_2$O containing a SIGMA*FAST*™ 3,3′-diaminobenzidine tetrahydrochloride tablet. The developed blot was washed with dH$_2$O.

To detect *P. luminescens* proteins, the membrane was blocked and washed as above and then incubated with rabbit anti-P.l. antibody (section 2.21), diluted 1:5000 in TBS. The

membrane was washed again and then incubated with a secondary antibody, HRP labelled goat anti-rabbit IgG, diluted 1:10000 in TBS. After a final washing, the blot was developed as outlined above.

2.23 Enzyme linked lectin assay (ELLA)

Solutions containing 10 µg/ml of glycoproteins were prepared in PBS (section 2.3). A 50 µl aliquot of each glycoprotein was added in triplicate to a 96 well Nunc MaxiSorp™ ELISA plate. The plate was incubated overnight at 4°C to immobilise the glycoproteins. Any unbound glycoprotein was removed by inverting the plate. The wells were blocked with 150 µl of 2.5% BSA in TBS for one hour. Excess blocking solution was removed and the wells were washed three times with TBST. A 50 µl aliquot of 5 µg/ml biotin labelled lectin in TBST was added to each well and left to incubate at room temperature for one hour. Unbound lectin was removed by inversion and the plate washed as before. 50 µl of 1:10,000 HRP labelled murine anti-biotin antibody diluted in 1% BSA-TBS was added to each well and incubated for one hour at room temperature. Unbound antibody was removed by inversion and washed as before.

3,3′,5,5′-Tetramethylbenzidine (TMB) substrate was prepared by dissolving 2 mg of TMB in 200 µl of dimethyl sulphoxide (DMSO). This was added to 9.8 ml of citrate buffer, pH 5.5 which had been prepared by mixing 5 ml of dH$_2$O with 1.37 ml of 0.1 M citric acid and 3.63 ml of 0.1 M sodium citrate. Directly before use 3 µl of 30% (v/v) H$_2$O$_2$ was added to the TMB solution. A 100 µl aliquot of TMB substrate was added to each well. After 5-10 minutes the reaction was stopped by the addition of 50 µl of dilute H$_2$SO$_4$. The absorbance was read at 450 nm.

2.24 Lectin affinity chromatography (LAC)

A 0.5 ml aliquot of agarose bound lectin resin (Vector Laboratories) was poured into a column and washed with 10 ml of LAC equilibration buffer (section 2.3). Proteins prepared by the water lysis method (section 2.15) or a solution of a commercial glycoprotein was added to the column and allowed to bind to the resin. Unbound

material was collected as flow though followed by a 20 ml wash using equilibration buffer collected in 1 ml fractions. Glycoprotein elution was facilitated by the addition of 10 x 1 ml of LAC elution buffer (section 2.3) specific for the lectin being used. All flow through, wash and elution fractions were analysed for the presence of protein by reading their absorbance at 280 nm in a quartz cuvette and comparing the values to a blank run.

2.25 Enzymatic reactions

2.25.1 Enzymes and buffers

All enzymes and their relevant buffers were obtained from Promega Corporation, Invitrogen Life Technologies®, New England BioLabs® or Sigma-Aldrich Corporation and were used according to the manufacturer's instructions.

2.25.2 Standard Phusion PCR reaction mixture

Sterile dH$_2$O	32 µl
5X Buffer	10 µl
dNTPs (10 mM each)	3 µl
Template DNA	2 µl
Primers (10 mM each)	1 µl of each
Phusion DNA polymerase (1:3 dilution)	1 µl

2.25.3 Standard PCR program cycle

Stage 1	Step 1: 98°C for 5 minutes

Stage 2   Step 1: 98°C for 30 seconds
Step 2: Annealing temperature for 30 seconds
Step 3: 72°C for 30 seconds for every kb to be
         synthesised

} 30 cycles

Stage 3   Step 1: 72°C for 10 minutes

2.25.4 T4 DNA ligase reaction

PCR product	5 µl
Plasmid DNA	3 µl
10X ligase buffer	1 µl
T4 DNA ligase	1 µl

2.25.5 In-gel trypsin digestion of proteins

The in-gel trypsin digestion of proteins was performed as described by (Shevchenko et al. 2007). All reagents used in this procedure were of mass spectrometry (MS) grade and supplied by Sigma-Aldrich Chemical Co. A protein from a band from an SDS-PAGE gel was digested using trypsin and the tryptic peptides analysed by MS in order to identify the protein in a technique known as peptide mass fingerprinting.

An SDS-PAGE gel was run as described in section 2.18.2 and stained using silver staining (section 2.20). The entire gel was rinsed with dH$_2$O for at least two hours. Using a clean scalpel the protein band of interest was excised, cut into 1 x 1 mm cubes and placed in a microfuge tube. A 500 µl aliquot of acetonitrile was added to the tube and the gel pieces were incubated for 10 minutes until they shrank, became opaque and

stuck together. The gel pieces were spun down and all liquid was removed. A 50 µl aliquot of 10 mM DTT/100 mM ammonium bicarbonate was added to completely cover the gel pieces. The gel pieces were incubated for 30 minutes at 56°C in an air circulation incubator. The gel pieces were cooled to room temperature, 500 µl of acetonitrile was added and the gel pieces shrunk as before. All liquid was removed. A 50 µl aliquot of 55 mM iodoacetamide/100 mM ammonium bicarbonate was added to cover the gel pieces and they were incubated at room temperature in the dark for 20 minutes. The gel pieces were shrunk as before and all liquid was removed.

A 50 µl aliquot of trypsin buffer (section 2.3) was added to cover the gel pieces and the tube was placed on ice. If after 30 minutes the buffer had been completely absorbed more trypsin buffer was added and the tube was left on ice for a further 90 minutes. A 10 µl aliquot of 100 mM ammonium bicarbonate was added to the tube and the digestion reaction was left to proceed overnight at 37°C in an air circulation incubator. Extraction buffer was prepared by mixing one part 5% (v/v) formic acid with two parts acetonitrile. Extraction buffer was added to the digest such that the approximate digest to extraction buffer ratio was 1:2. The tube was incubated in a shaker at 37°C for 15 minutes. The mixture was briefly spun down in a centrifuge and the supernatant was withdrawn into a new microfuge tube. The liquid was dried down completely in a vacuum dryer and then reconstituted in 10 µl of 0.1% trifluoroacetic acid (TFA).

The sample was desalted using Supel-Tips C18 Micropipette Tips (Sigma-Aldrich). The tip was first prepared by attaching it to a 10 µl pipette. Trapped air was displaced and the sorbent bed was conditioned by aspirating and dispensing 10 µl of methanol, followed by 10 µl of 50% methanol and finally 2 x 10 µl of dH_2O. The sample was allowed to bind to the medium by aspirating and dispensing it 10-20 times. Salts and detergents were removed by washing twice with 10 µl dH_2O. Finally the bound peptides were eluted in 3 x 5 µl aliquots of 70% methanol containing 0.1% TFA. The three aliquots were combined and the peptides were then ready for analysis using the mass spectrometer (section 2.27).

2.25.6 Glycosidase treatment of glycoproteins

• **Endo H$_f$/PNGaseF**

Denaturation reaction

Glycoprotein	20-100 µg
10X glycoprotein denaturing buffer	1 µl
dH$_2$O	to 10 µl

Glycoproteins were denatured by heating the reaction mixture at 100°C for 10 minutes followed by cooling to room temperature.

Glycosidase reaction

Denaturation reaction mixture	10 µl
10X reaction buffer	2 µl
(10% NP40 (PNGaseF only)	2 µl
Enzyme	2 µl
dH$_2$O	to 20 µl

Reaction was incubated in a 37°C water bath overnight.

• **Neuraminidase/β-*N*-Acetyl-Hexosaminidase$_f$**

Glycoprotein	20-100 µg
10X reaction buffer	1 µl
Enzyme	2 µl
dH$_2$O	to 10 µl

Reaction was incubated in a 37°C water bath overnight.

- **α1-3, 6 Galactosidase/α-*N*-Acetyl-Galactosaminidase**

Glycoprotein	20-100 µg
10X reaction buffer	1 µl
10X BSA	1 µl
Enzyme	2 µl
dH$_2$O	to 10 µl

Reaction was incubated in a 37°C water bath overnight.

2.26 β-Elimination reaction to remove *O*-linked glycans from glycoproteins

The β-elimination reaction was performed using the Glycoprofile^{TM} β-Elimination Kit from Sigma-Aldrich. The glycoprotein was dissolved in dH$_2$O to a concentration of 1-10 µg/µl. The glycoprotein solution was heated at 100°C for 30 minutes to completely denature the protein and then allowed to cool to room temperature. β-elimination reagent mixture was prepared just before use by mixing β-elimination reagent with 5 M NaOH in the ratio 94:6. To the denatured protein sample, β-elimination reagent mixture was added equal to 20% of the protein sample volume i.e. to 100 µl of protein sample, 20 µl of β-elimination reagent mixture was added. The mixture was incubated at 4°C overnight and was then neutralised by adjusting the pH to 6-8 with 1 M HCl.

A centrifugal filter unit was assembled and the filter was washed using 500 µl of dH$_2$O by centrifugation at 14,000 g for 30 minutes. The water wash was discarded and the wash step was repeated. The neutralised protein sample was transferred to the sample reservoir and centrifuged as before until most of the protein sample had passed through (about 30 minutes per 500 µl), being careful not to let the filter dry out. The protein solution was washed 2-3 times with 200 µl of dH$_2$O by centrifugation for 10 minutes. The protein sample was recovered by inverting the filter unit in a new collection tube and centrifuging for 3 minutes at 1,000 g. The sample reservoir was washed twice with 200 µl of dH$_2$O by pipetting up and down and the washes were combined with the rest of the deglycosylated protein sample.

2.27 Matrix assisted laser desorption ionization mass spectrometry (MALDI-MS)

The matrix used for MALDI analysis was a 1:1 (w/w) mixture of α-cyano-4-hydroxycinnamic acid and 2,5-dihydroxy benzoic acid (CHCA:DHB). The matrix was prepared by dissolving it in 70% acetonitrile (mass spec grade) to a concentration of 20 mg/ml. TFA was then added to a final concentration of 0.1%. Samples prepared as described in section 2.25.5 were mixed 1:1 with the prepared matrix and spotted on a stainless steel MALDI target plate and air dried.

MALDI analysis was carried out using a quadrupole orthogonal acceleration time of flight (Q-TOF) mass spectrometer. Two models were available for use, the Q-TOF Ultima™ Global and the Q-TOF Premier (Micromass MS Techologies, Manchester, UK). The Q-TOF Premier is an upgraded version of the Q-TOF Ultima™ Global. Both instruments used a nitrogen laser, operating at 337 nm but the laser had a hit frequency of 10 Hz for the Q-TOF Ultima™ Global and 200 Hz for the Q-TOF Premier. All other parameters used were identical. The laser was fired at the dried matrix/analyte mixture on the MALDI target plate. Ions were generated and focused into a beam by a radio frequency (RF) lens. The ions were transmitted through a hexapole collision cell, with a collision energy of 10 V then pulsed into the TOF analyser. Samples were analysed in positive ion mode with the TOF analyser set at 9.10 kV. The instrument was operated in V reflectron mode and the reflectron was set to 35.6 V. The pusher was set to 974 µs. Data was acquired using a time-to-digital converter (TDC) operating at 4 GHz. Mass spectra were recorded using a microchannel plate (MCP) detector. A mass range (*m/z*) of 500 – 5000 was scanned over 2.4 seconds with an inter scan delay time of 0.1 second.

The instrument was calibrated using a range of standard peptides with molecular weights ranging from 756.4 Da to 3493.7 Da, the range that most tryptic digestion fragments will lie between. Data was analysed using the software MassLynx version 4.1 (Waters, Manchester, UK).

Chapter 3

Detection & isolation of glycoproteins from *P. luminescens*

3.1 Introduction

As previously stated, the primary objective of this research is to investigate the possibility that *P. luminescens* produces glycoproteins and thus contains a protein glycosylation system. In order to do this, glycoproteins have to be isolated from the large mix of proteins normally found in the cell and characterised. The use of lectins is one of the most effective means of detecting, purifying and isolating glycoproteins. Lectins are carbohydrate binding proteins that bind to mono or oligosaccharides. The affinity of lectins is relatively weak but the specificity of the binding is very high. Lectins can be used in enzyme linked lectin assays (ELLAs), in western blots and bound to affinity matrices to detect, characterise and isolate glycoproteins.

In order to further characterise the isolated glycoproteins, glycosidase enzymes can be used to cleave sugar units from the linked glycan of a glycoprotein. The resulting glycoprotein may be retested with the same lectin in order to investigate how the interactions between the carbohydrate and the lectin have changed. In this way more knowledge can be gained about the types of sugar moieties that are present on a glycoprotein and the way that they are linked together.

Lectins and glycosidases have been utilised previously to successfully detect glycoproteins in *C. jejuni* (Linton et al. 2002), which has an *N*-linked protein glycosylation system that is already well characterised. In this chapter certain lectins and glycosidases were employed to detect, isolate and characterise glycoproteins from the various cell fractions of *P. luminescens*. Comparable experiments were carried out using *C. jejuni* proteins to act as a control for the *P. luminescens* study.

3.2 *P. luminescens* growth curve

In order to examine the growth pattern of *P. luminescens*, a growth curve was constructed. A 500 ml culture of *P. luminescens* was grown at 30°C (section 2.5.1) for 30 hours and samples were taken at regular time intervals. The absorbance readings of

the samples were read at 600 nm and a growth curve was constructed as shown in figure 3.1.
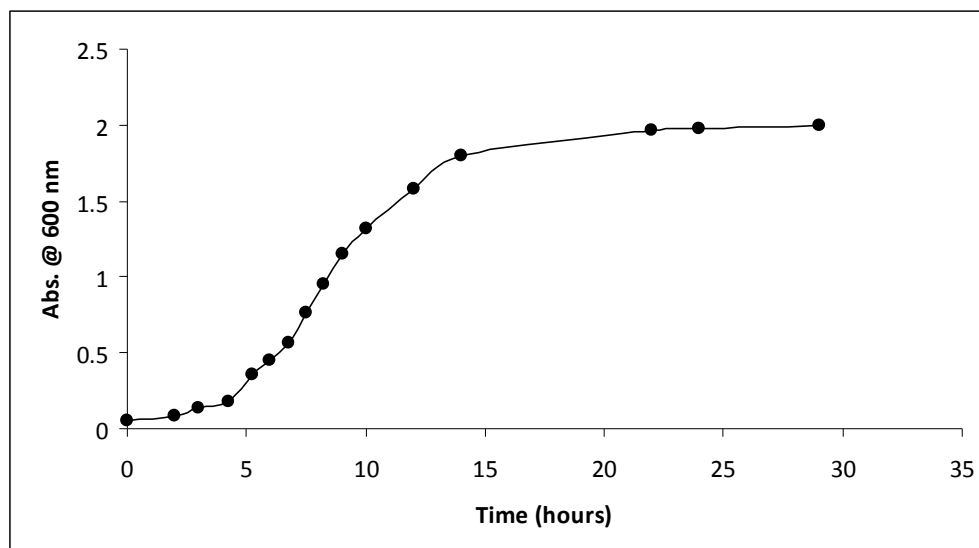
Figure 3.1 *P. luminescens* growth curve

A 500 ml culture *P. luminescens* culture was grown at 30°C over 30 hours. Samples were taken approximately every two hours and their absorbance readings at 600 nm were plotted against time.

The result is the characteristic S-shape growth curve. The lag phase lasts ~4 hours, followed by the exponential phase for ~8 hours after which the growth of the culture enters the stationary phase.

3.3 Extracellular, periplasmic, cytoplasmic and membrane protein isolation from *P. luminescens* using the water lysis method

In order to get maximum protein yield two 100 ml cultures of *P. luminescens* were grown for 24 and 48 hours as described in section 2.5.1. Using the water lysis method described in section 2.15, extracellular proteins and proteins from the periplasm, cytoplasm and membrane were isolated. The protein concentration in each fraction was determined using a BCA assay (section 2.17). The results are shown in figure 3.2 and table 3.1.
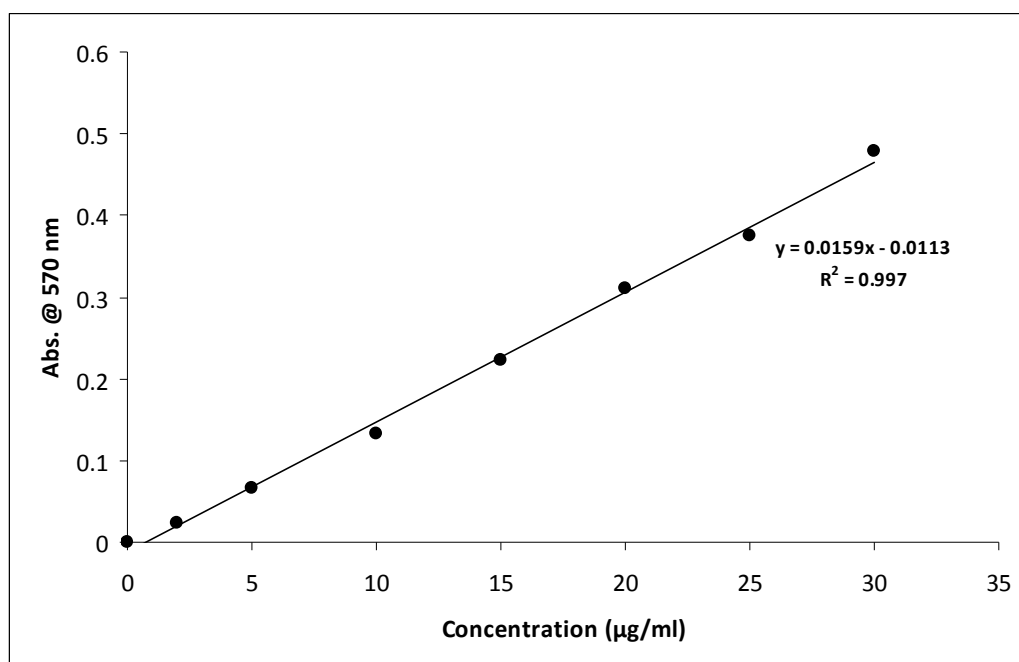
Figure 3.2 BCA assay standard curve used to quantify protein concentration in *P. luminescens* water lysis samples

The equation of the standard curve y = 0.0159 x – 0.0113 was used to calculate x, the concentration of the unknown samples. The calculated concentration values are shown in table 3.1.

Table 3.1 Concentrations of water lysis samples from *P. luminescens*

Sample	Concentration	Total protein
Extracellular (24 h)	1.62 mg/ml	81 mg
Extracellular (48 h)	1.56 mg/ml	78 mg
Periplasmic (24 h)	236 µg/ml	6.1 mg
Periplasmic (48 h)	173 µg/ml	4.5 mg
Cytoplasmic (24 h)	1.11 mg/ml	17.8 mg
Cytoplasmic (48 h)	1.54 mg/ml	24.6 mg
Membrane (24 h)	9.68 mg/ml	14.5 mg
Membrane (48 h)	6.34 mg/ml	9.5 mg

Using the calculated concentration values, 30 µg of each sample was loaded on a 15% SDS-PAGE gel and proteins were separated (section 2.18) followed by staining with coomassie blue (section 2.20). Figure 3.3 shows a scanned image of this gel.
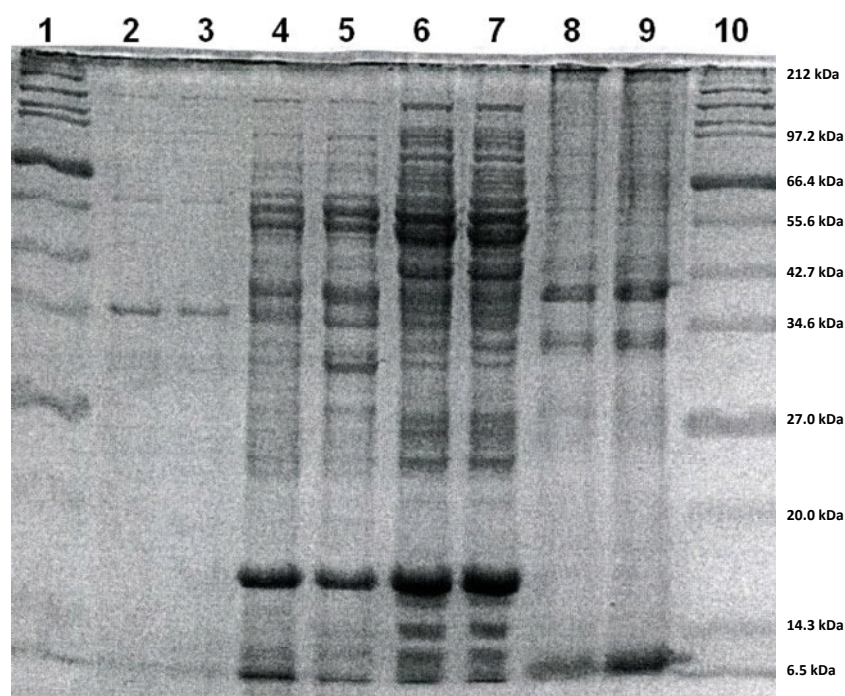
Figure 3.3 Isolated protein samples from *P. luminescens* released by the water lysis method
30 µl of each protein sample was separated on a 15% SDS-PAGE gel which was then stained with coomassie blue. Lane 1 = NEB broad range protein marker, lane 2 = extracellular (24 h), lane 3 = extracellular (48 h), lane 4 = periplasmic (24 h), lane 5 = periplasmic (48 h), lane 6 = cytoplasmic (24 h), lane 7 = cytoplasmic (48 h), lane 8 = membrane (24 h), lane 9 = membrane (48 h) and lane 10 = NEB broad range protein marker.

A distinct protein pattern is visible in each fraction suggesting that the protein fractions have been well separated. There appears to be very little difference in protein expression between the two samples. Both cultures would have been in stationary phase of the growth cycle (figure 3.1).

Figure 3.4 shows the same SDS-PAGE gel which has been stained using the periodic acid-Schiff stain for carbohydrate (section 2.20) following the coomassie blue staining. Coomassie blue stains all proteins a blue colour while the Schiff stain can distinguish between glycosylated and non-glycosylated proteins. The carbohydrate part of the glycoprotein is stained pink with the Schiff stain while non-glycosylated proteins remain blue from the coomassie blue stain.
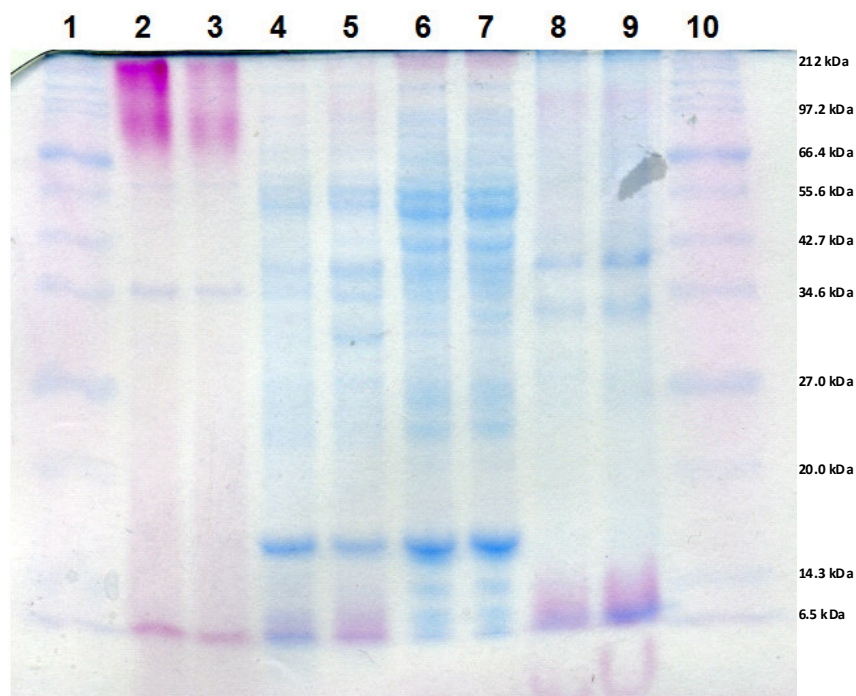
Figure 3.4 Isolated protein samples from *P. luminescens* released by the water lysis method

Following coomassie staining of the water lysis protein samples shown in figure 3.3, the gel was subsequently stained with periodic acid-Schiff stain for the presence of carbohydrate. Lane 1 = NEB broad range protein marker, lane 2 = extracellular (24 h), lane 3 = extracellular (48 h), lane 4 = periplasmic (24 h), lane 5 = periplasmic (48 h), lane 6 = cytoplasmic (24 h), lane 7 = cytoplasmic (48 h), lane 8 = membrane (24 h), lane 9 = membrane (48 h) and lane 10 = NEB broad range protein marker.

Figure 3.4 demonstrates that glycoproteins could possibly be present in each fraction. There is some very high molecular weight carbohydrate present in the extracellular fraction, however, it is difficult to see distinct bands. This was not seen with the coomassie stain (figure 3.3) so this suggests that it is not glycoproteins but is probably some very high molecular weight carbohydrates that were present in the media. Some bands are present in the periplasmic fractions at ~200 kDa and ~10 kDa, in the cytoplasmic fraction at ~200 kDa and membrane proteins at ~100 kDa. There is also distinct pink area visible in the low molecular weight region of the extracellular fractions which could be low molecular weight sugars that were present in the media. There is also a pink area in the low molecular weight region of the membrane proteins between about 5 and 15 kDa which could be lipopolysaccharide. Figure 3.4 shows that it is probable that the glycoproteins are present in the periplasmic, cytoplasmic and membrane protein fractions.

3.4 Analysis of isolated protein fractions using ELLAs and western blots

A number of ELLAs were carried out on the four different protein fractions using a variety of commercially available plant lectins. Table 3.2 outlines the lectins used in this analysis, the concentration used in the assays and the glycans they have affinity for.

Table 3.2 Commercial plant lectins used in this study and their corresponding affinities

Lectin	Working Conc.	Affinity
AAL	1 µg/ml	α-1,6-Fucose
UEA I	5 µg/ml	α-1,2-Fucose
Con A	1 µg/ml	core mannose of biantennary complex glycan
LCA	5 µg/ml	fucosylated core mannose of complex glycan
NPL	5 µg/ml	terminal mannose / high mannose
GNL	5 µg/ml	terminal mannose / high mannose
RCA	5 µg/ml	terminal β-gal / lactosamine
ECL	5 µg/ml	terminal β-gal / lactosamine
PNA	5 µg/ml	terminal β-gal / T-antigen
GSL I	5 µg/ml	terminal α-gal
Jacalin	5 µg/ml	β-gal / T-antigen
SBA	5 µg/ml	terminal GalNAc
DBA	5 µg/ml	terminal GalNAc
DSL	5 µg/ml	GlcNAc / lactosamine
GSL II	5 µg/ml	terminal GlcNAc
WGA	5 µg/ml	GlcNAc, NeuNAc
MAL I	5 µg/ml	Lactosamine / tolerates α-2,3 NeuNAc
MAL II	10 µg/ml	α-2,3 NeuNAc
SNA	10 µg/ml	α-2,6 NeuNAc

ELLAs were performed as per the procedure in section 2.23. The protein samples were cleaned up using vivaspin 500 µl centrifugal ultrafiltration devices (Sartorius Stedim Biotech) to remove any interfering sugars. Vivaspin columns with a molecular weight

cut off of 10,000 Da were used. The protein samples were added to the columns which were then centrifuged at 13,000 rpm for 10 minutes. The flow through was removed and PBS was added to the concentrated sample and the centrifugation step was repeated. The protein concentration of each sample was found using a BCA assay and the samples were diluted down to 10 µg/ml for the ELLA analysis. Figure 3.5 shows the results of the ELLA analysis.

a)
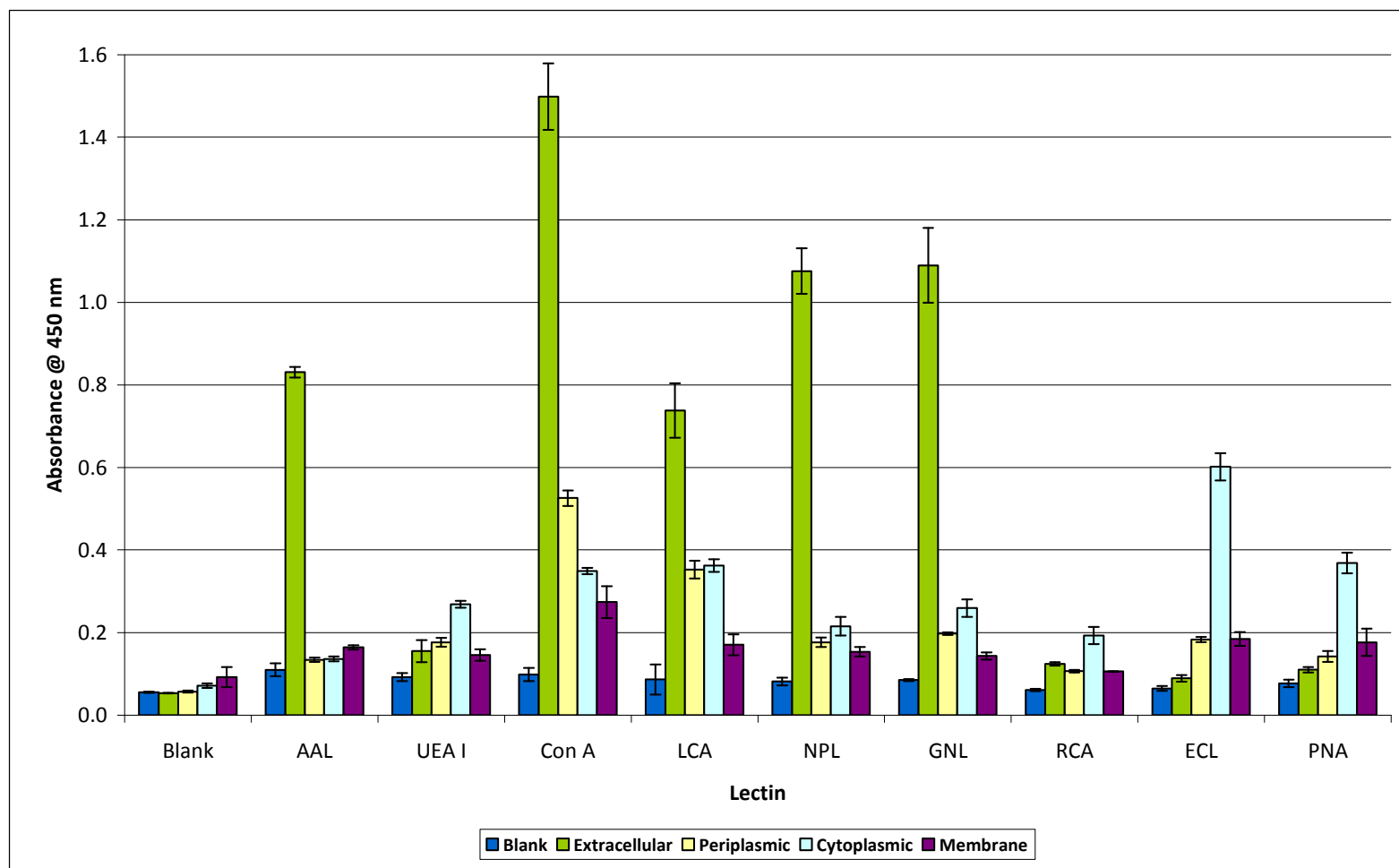
Figure 3.5 ELLAs profiling the interactions between *P. luminescens* extracellular, periplasmic, cytoplasmic and membrane proteins with various commercial lectins

b)
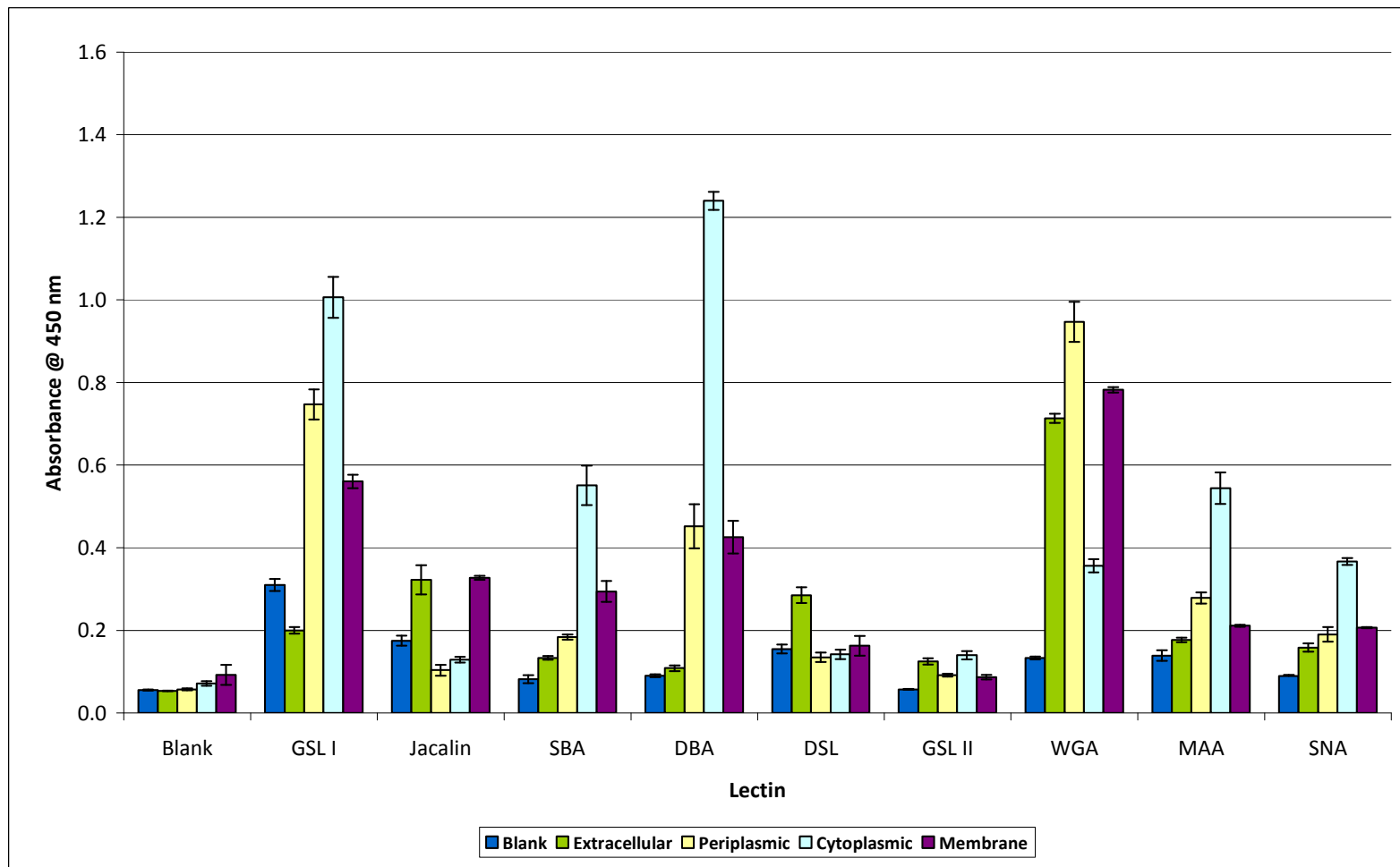
Figure 3.5 continued... ELLAs profiling the interactions between *P. luminescens* extracellular, periplasmic, cytoplasmic and membrane proteins with various commercial lectins

Because of the specificity of the linkage between a glycan and lectin, as outlined in section 1.4.2, ELLAs enable us to get some insight into the nature of the carbohydrate moieties of glycoproteins that might be present. The results in figure 3.5 show that there is a huge diversity between the four fractions in terms of the glycoproteins that are possibly present. More than one lectin reacted quite strongly with each fraction. The extracellular proteins are binding quite strongly to the mannose binding lectins and WGA and so the results indicate that they may contain high mannose and GlcNAc. The cytoplasmic proteins are binding strongly to galactose, GalNAc and GlcNAc binding. Finally the periplasmic and membrane proteins are both binding GSL I, SBA, DBA and WGA most strongly and so they may also contain terminal galactose, GalNAc or GlcNAc in their glycan moieties.

It is probable that the interactions shown in figure 3.5 are not all due to lectin / glycoprotein binding. The lectins may also be binding to proteins that are not glycosylated but are involved in the transport or biosynthesis of glycans. Whether any of the proteins that are binding to the lectins in the ELLAs are definitely glycosylated needs to be further investigated.

Some of the lectins that produced very strong ELLA responses with the different fractions were chosen for western blot analysis. Western blots were performed using the periplasmic, cytoplasmic and membrane proteins and probing them with various lectins. Western blot analysis was carried out according to section 2.22 and the results are shown in figure 3.6
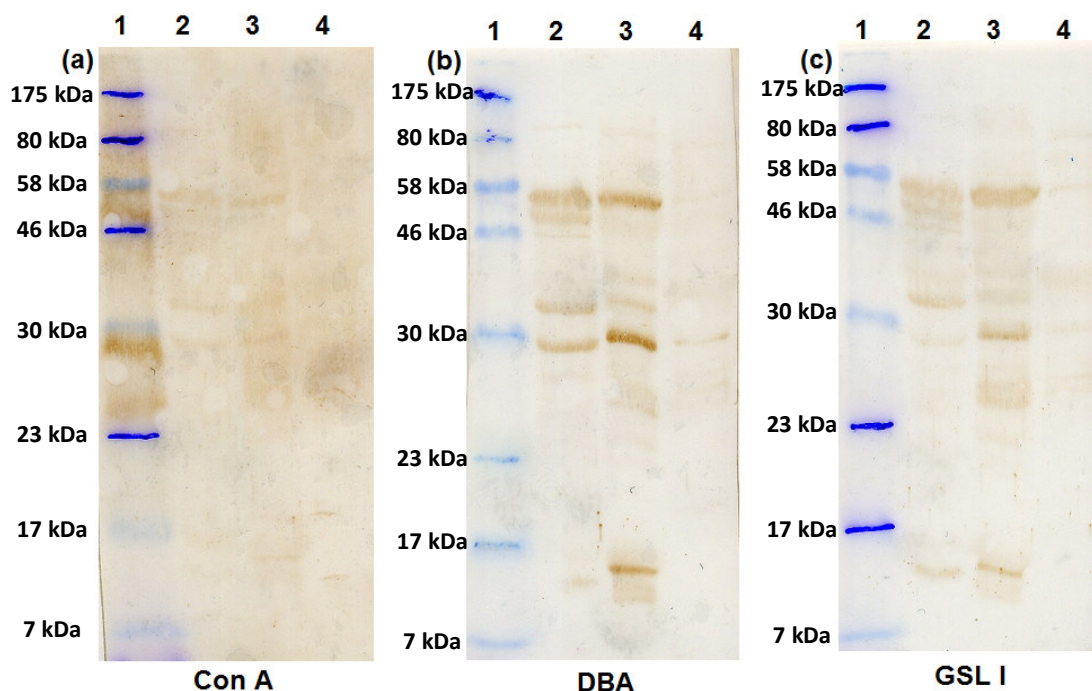
Figure 3.6 Western blot analysis showing interactions between *P. luminescens* periplasmic, cytoplasmic and membrane proteins with the lectins Con A (a), DBA (b) and GSL I (c)

Lane 1 = NEB ColorPlus prestained protein marker (7-175 kDa), lane 2 = periplasmic proteins, lane 3 = cytoplasmic proteins and lane 4 = membrane proteins.

The protein pattern visible in the Con A, DBA and GSL I blots for each of the three fractions are very similar. Strong bands are showing up at ~55, 50, 46, 34, 29 and 13 kDa for the periplasmic proteins. Bands are visible at ~55, 38, 34 29, 25, 23 and 15 kDa for the cytoplasmic proteins. For the membrane proteins bands can be seen at ~55, 36, 29 and 26 kDa in the blots.

From the result in figure 3.6 it can be deduced that the putative glycoproteins from the three fractions contain mannose, terminal galactose and terminal GalNAc. As discussed in chapter 1, *N*-linked glycans from *C. jejuni* contain terminal GalNAc. Typical *O*-linked glycans found in *Neisseria* spp. contain terminal galactose. Proteins from Gram-positive Actinomycetes have been shown to be *O*-mannosylated. Therefore, the types of glycans being recognised in figure 3.6 are consistent with glycans that have already been discovered in other bacteria.

3.5 Isolation/purification of glycoproteins from *P. luminescens* protein fractions using lectin affinity chromatography (LAC)

The next step in the analysis of glycoproteins from *P. luminescens* was to isolate them from the large mix of other proteins in the cell using LAC. If glycoproteins are present they were likely to be in very low concentrations. LAC was used to concentrate any glycoproteins present to make further analysis of them easier. Agarose bound lectin resins are commercially available for this purpose. In this study agarose bound WGA, Con A and GSL I were the resins used. They were chosen because they produced a strong signal in the ELLAs and/or western blots with all four protein fractions (figure 3.5 and figure 3.6).

LAC was carried out according to the procedure described in section 2.24. LAC was preformed with either the periplasmic fractions or the cytoplasmic fractions. A column containing 500 µl of resin was used and 2-5 ml of protein sample prepared using the water lysis method was passed over the resin. The absorbance readings (at 280 nm) of the various protein fractions were plotted. Readings were compared to a blank run containing no protein. A positive control was also used where a commercial glycoprotein that is known to bind the lectin in question was passed over the resin.

3.5.1 LAC using Wheat Germ Agglutinin (WGA) agarose

Figures 3.7 and 3.8 show the results of the LAC using WGA resin and periplasmic and cytoplasmic proteins.
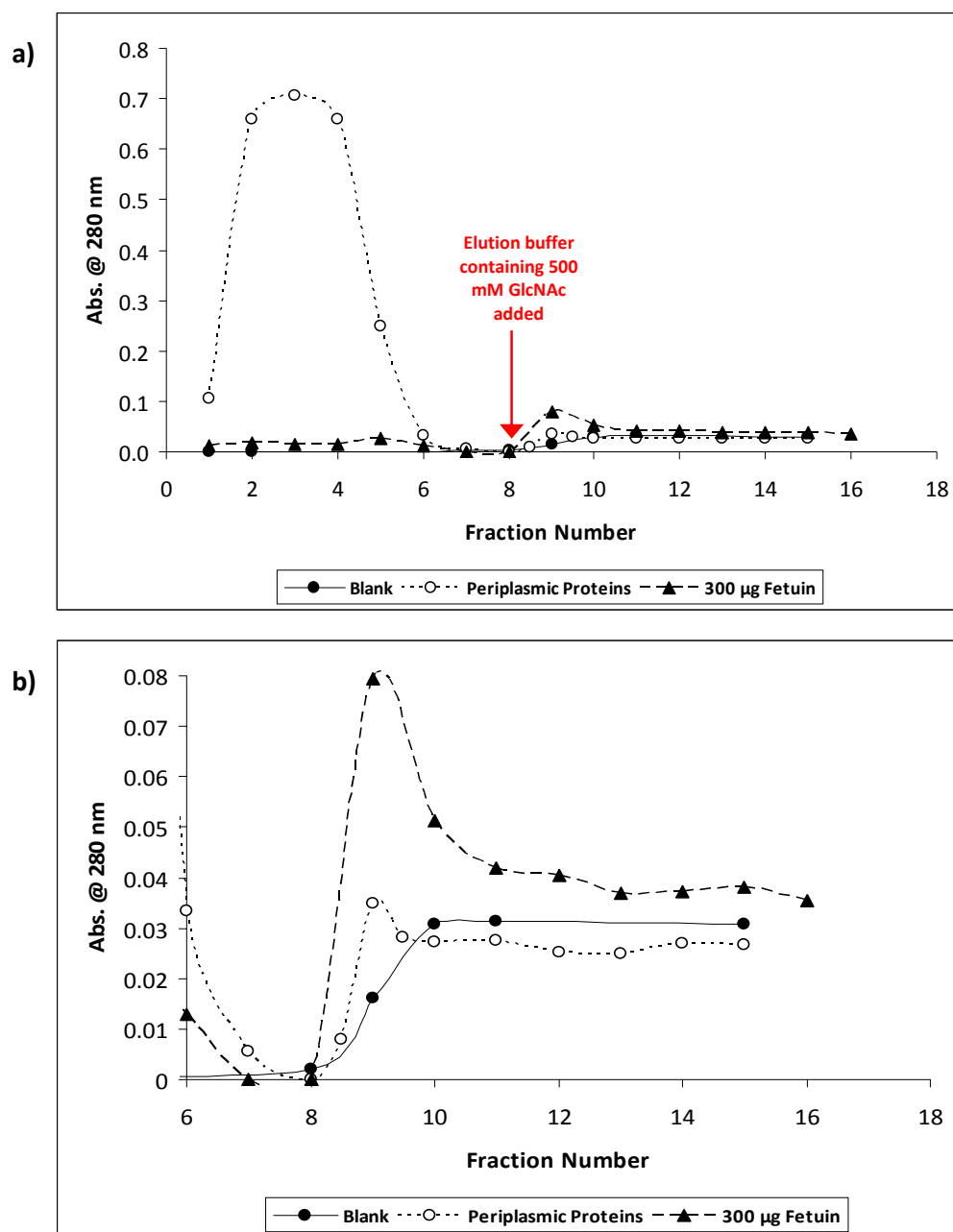
Figure 3.7 Periplasmic proteins from *P. luminescens* fractionated using LAC with WGA bound agarose

A 3 ml sample of periplasmic proteins prepared using the water lysis method was passed over the column and any unbound proteins were washed from the column (fractions 1-8). Any proteins that bound to the resin were eluted with 500 mM GlcNAc (fractions 9-16). Fetuin was used as a positive control. (b) Enlarged area of the graph where the proteins were eluted.

It can be seen that a big majority of protein from the periplasmic extract did not bind to the column and was washed off in the unbound fractions. The column was washed with

buffer until the absorbance readings reached baseline levels to completely remove any unbound proteins prior to elution of bound protein. A small amount of protein from the periplasmic extract did bind and was eluted in fractions 9 to 10.
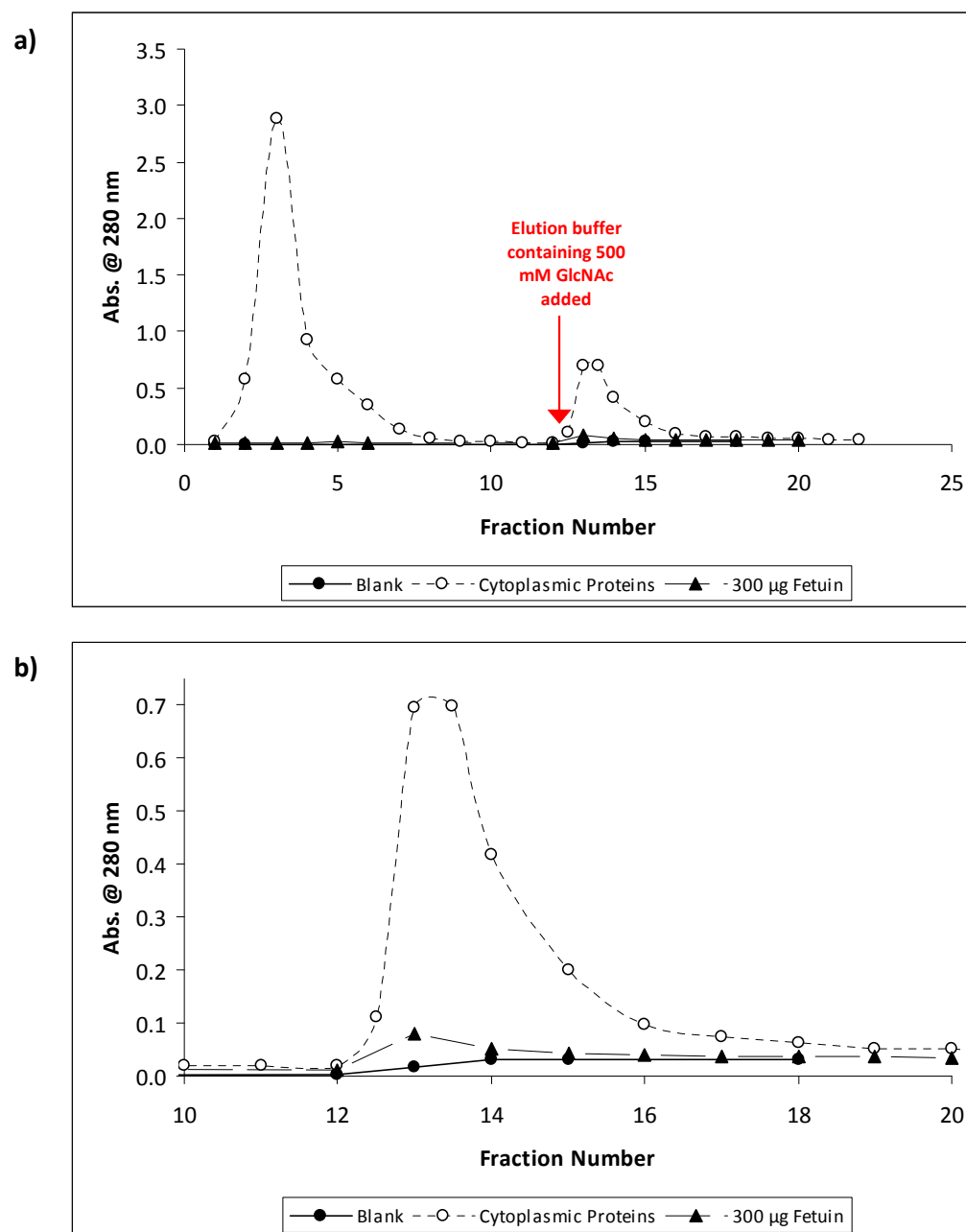
Figure 3.8 Cytoplasmic proteins from *P. luminescens* fractionated using LAC with WGA bound agarose

(a) A 2 ml sample of cytoplasmic proteins prepared using the water lysis method was passed over the column. Unbound proteins were washed from the column in fractions 1-12. Any proteins that bound to the resin were eluted and collected in fractions 13-20. Fetuin was again used as a positive control. (b) Enlarged area of the graph where the proteins were eluted.

The cytoplasmic protein extract contained a majority of protein that did not bind to the WGA resin and was washed off in the unbound fractions. A bigger percentage of protein bound to the column for the cytoplasmic protein run compared to the periplasmic protein run in figure 3.7. Fractions 13 to 18 contained eluted proteins. The absorbance readings for the eluted cytoplasmic proteins were even higher than those seen for the fetuin control which contained 300 µg of fetuin. This indicates that the cytoplasmic fraction could contain a significant quantity of glycoprotein. Fractions from the LAC of the periplasmic and cytoplasmic proteins were run on a 15% SDS-PAGE gel and stained with silver stain (section 2.20) as shown in figure 3.9.
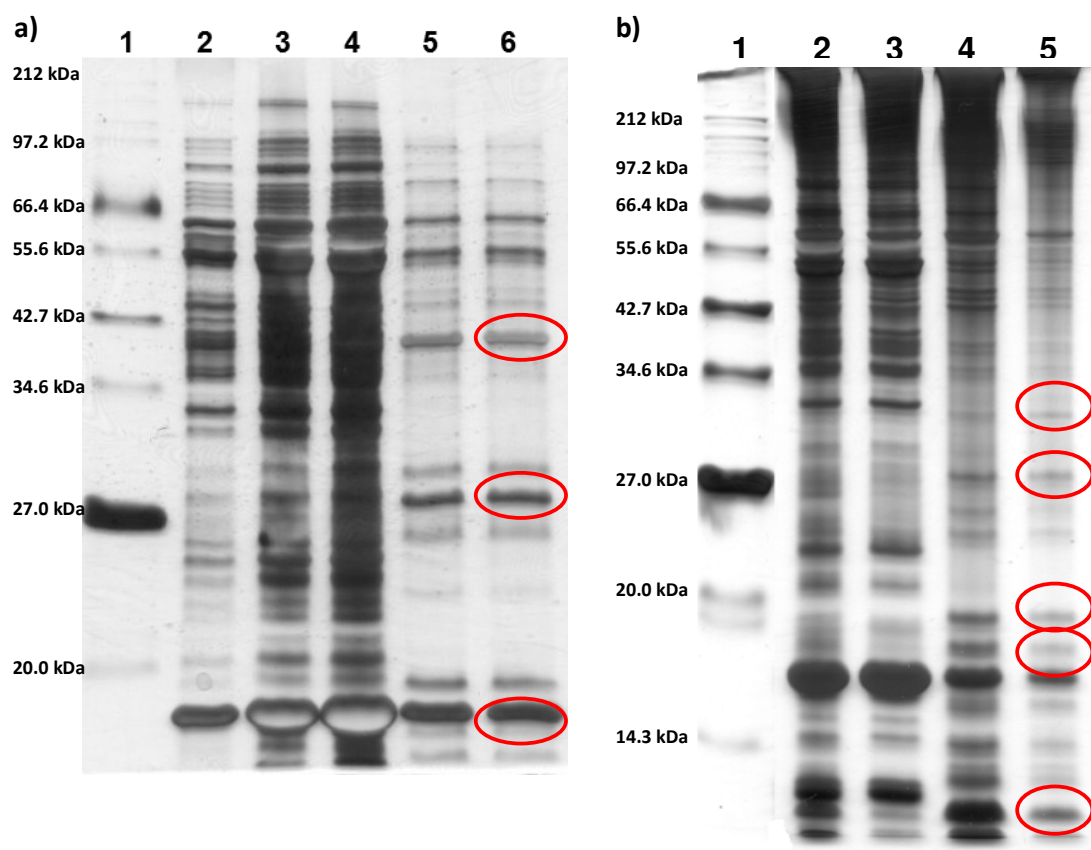
Figure 3.9 WGA LAC periplasmic protein fractions (a) and cytoplasmic protein fractions (b)
Proteins were separated on a 15% SDS-PAGE gel and stained with silver stain. (a) Lane 1 = NEB broad range protein marker, lane 2 = crude periplasmic fraction from water lysis method, lanes 3, 4 = unbound periplasmic proteins from LAC, lanes 5, 6 = eluted periplasmic proteins from LAC. (b) Lane 1 = NEB broad range protein marker, lane 2 = crude cytoplasmic fraction from water lysis method, lane 3 = unbound cytoplasmic proteins from LAC, lanes 4, 5 = eluted cytoplasmic proteins from LAC. Red ovals indicate some proteins that have been enriched.

Lanes 3 and 4 in figure 3.9 (a) contain many proteins that did not bind to the resin. Lanes 5 and 6 contain the proteins that were eluted from the resin of which the three at ~40 kDa, ~27 kDa and ~17kDa are the most abundant (circled in red). The eluted periplasmic proteins visible in lanes 5 and 6 of figure 3.9 (a) are also visible in the crude sample (lane 2) but they have been concentrated. This is especially evident with the 27 kDa protein.

For the cytoplasmic proteins (figure 3.9 (b)) again many proteins did not bind to the resin and these are visible in lane 3. Some of the eluted cytoplasmic proteins visible in lanes 4 and 5 in figure 3.9 (b) have also been substantially enriched (circled in red) compared to the crude sample in lane 2. Lanes 4 and 5 contain the eluted proteins. Many more cytoplasmic proteins bound to the resin compared to the periplasmic proteins.

Some of the eluted periplasmic and cytoplasmic proteins are also visible in the unbound protein fraction. This could be explained by the fact that lectins, although they have a very high specificity, have a low affinity for glycoproteins. Therefore the lectin may not bind all of the available glycoprotein and even some that it does bind may dissociate from the lectin again during washing.

It is very likely that the proteins that have bound to the lectin are binding through glycan-protein interactions and not through protein-protein interactions because binding was inhibited with the addition of GlcNAc in the elution buffer which is the inhibiting sugar for WGA. The results shown in figure 3.9 were reproducible over several runs.

In order to confirm that the proteins that were eluted from the LAC column (figure 3.9) were actually binding to WGA and not just binding non-specifically to the column a western blot was carried out using the samples from figure 3.9. The samples on the blot were probed with biotinylated WGA lectin, followed by an anti-biotin lectin and were then developed. The results of this western blot are shown in figure 3.10.
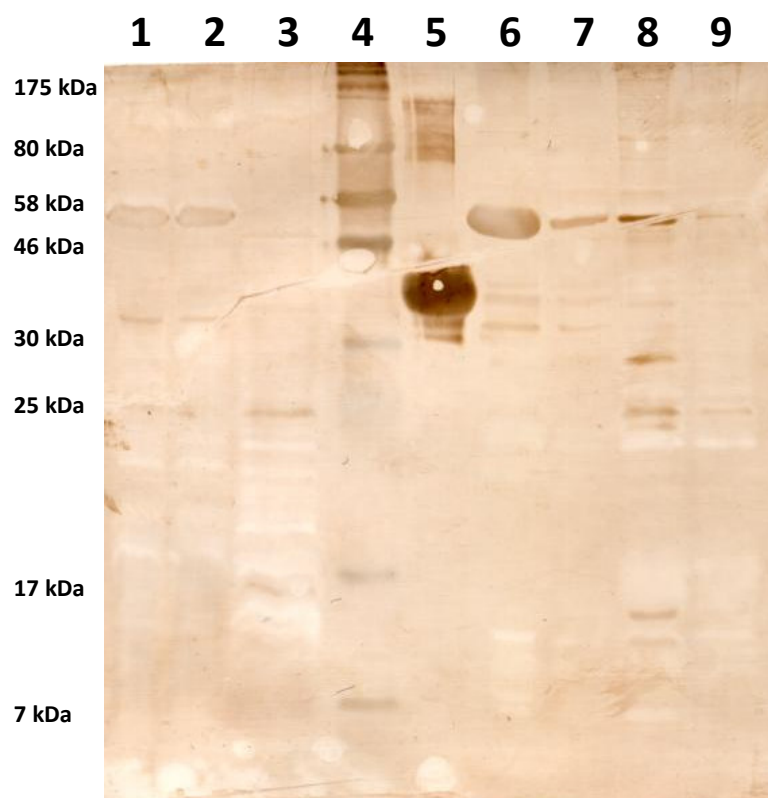
Figure 3.10 Western blot showing WGA LAC periplasmic protein fractions and cytoplasmic protein fractions

The western blot was probed with 1 µg/ml WGA. Lane 1 = crude periplasmic fraction from water lysis method, lane 2 = unbound periplasmic proteins from LAC, lane 3 = eluted periplasmic proteins from LAC, lane 4 = NEB prestained protein marker, lane 5 = ovalbumin, lane 6 = crude cytoplasmic fraction from water lysis method, lane 7 = unbound cytoplasmic proteins from LAC, lanes 8, 9 = eluted cytoplasmic proteins from LAC.

Ovalbumin, a protein known to bind to the lectin WGA was used as a control in this blot. In lane 3, which contains the periplasmic proteins that were eluted from the WGA column, there are some distinct bands showing up at approximately 50, 40, 25 and 17 kDa. The prestained protein marker is not as accurate as the broad range protein marker and so taking this into account these sizes correspond quite well to the strong bands that showed up in the eluted protein sample in figure 3.9 (a). There is a strong band showing up in the crude and unbound periplasmic protein samples (lanes 1 and 2) at ~55 kDa that does not appear to be present in the eluted sample. This may be because this protein bound very weakly to WGA in the column and dissociated from the column in the washing step. For the eluted cytoplasmic proteins (lanes 8 and 9) there are quite a few

bands showing up and again the pattern of bands corresponds quite well with what was seen in figure 3.9 (b). Strong bands are showing up at approximately 55, 35, 29, 25, 15 and 12 kDa. Again, the slight variation in sizes can be accounted for by the fact that the prestained protein marker is not extremely accurate. The fact that many of the same bands that showed up on the protein gels in figure 3.9 are showing up on the western blot provides strong evidence that these proteins are in fact binding to WGA through protein-sugar interactions.

ELLAs were also carried out to show the interactions between these glycoproteins and the lectin WGA as well as Con A and GNL (mannose binding lectins) and DSL (another GlcNAc binding lectin). A BCA assay was first performed using the crude periplasmic and cytoplasmic proteins and unbound and eluted LAC proteins. The concentration of each sample was adjusted to 10 µg/ml for the ELLAs so that results could be easily compared. The results of the ELLA analysis are shown in figures 3.11 and 3.12.
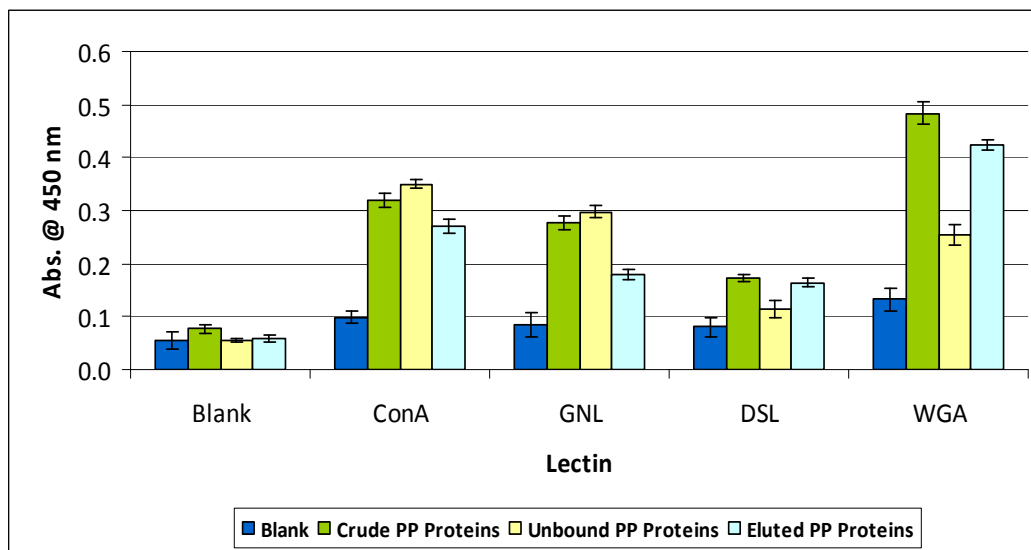
Figure 3.11 ELLA profiling the interactions of mannose and GlcNAc binding lectins with periplasmic protein samples fractionated using WGA LAC
Crude periplasmic proteins and unbound and eluted proteins from WGA LAC were analysed.

The results of this experiment indicate that the glycoproteins that are eluted from the WGA resin contain more GlcNAc than those in the unbound fraction. This result was expected as this resin should have bound glycoproteins containing GlcNAc. The unbound proteins do still contain some GlcNAc however because of the low affinity of

the lectin carbohydrate binding. The periplasmic proteins that bound to WGA contain less mannose than those that did not bind but the eluted proteins are still producing a significant signal with the mannose binding lectins. It is possible that those proteins that did not bind to WGA contain terminal mannose and those that did contain some core mannose.
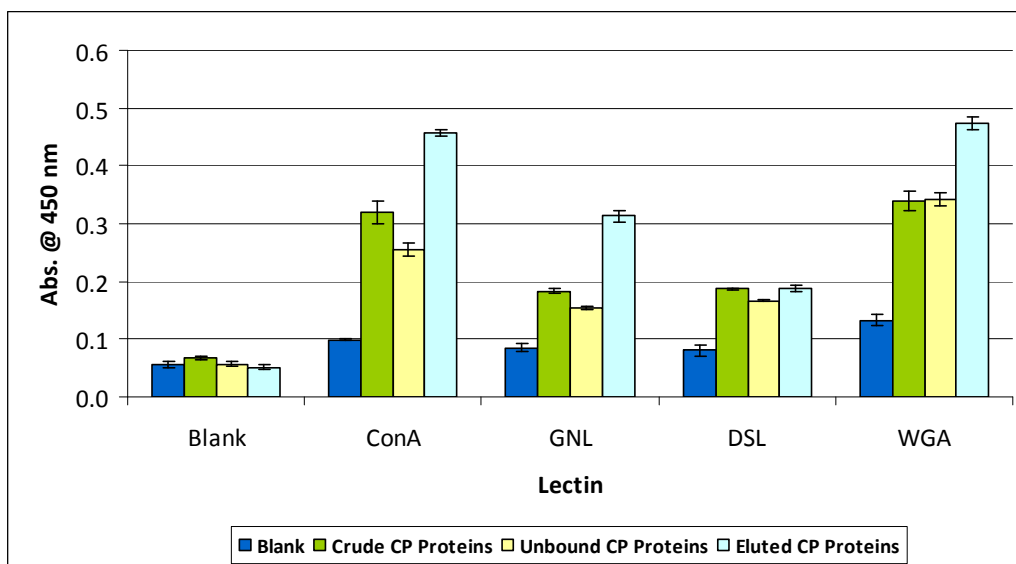
Figure 3.12 ELLA profiling the interactions of mannose and GlcNAc binding lectins with cytoplasmic protein samples fractionated using WGA LAC
Crude cytoplasmic proteins and unbound and eluted proteins from WGA LAC were analysed.

The results in figure 3.12 show that the signals for the GlcNAc binding lectins, WGA and DSL have gone up for the eluted proteins compared to the unbound proteins and the signals for the mannose binding lectins have also increased. This suggests that the glycoproteins that were isolated from the cytoplasm using WGA contain mannose as well as GlcNAc.

3.5.2 LAC using Concanavalin A (Con A) agarose

LAC was also carried out with cytoplasmic proteins using Con A agarose. A 2 ml protein sample prepared using the water lysis method was passed over the column and collected, the column was washed and bound proteins were eluted using 200 mM α-methyl mannoside and 200 mM α-methyl glucoside. As a control a run using 500 µg of

invertase (a high mannose glycoprotein) was also carried out. The O.D. $_{280\ nm}$ readings were plotted and the results are shown in figure 3.13.
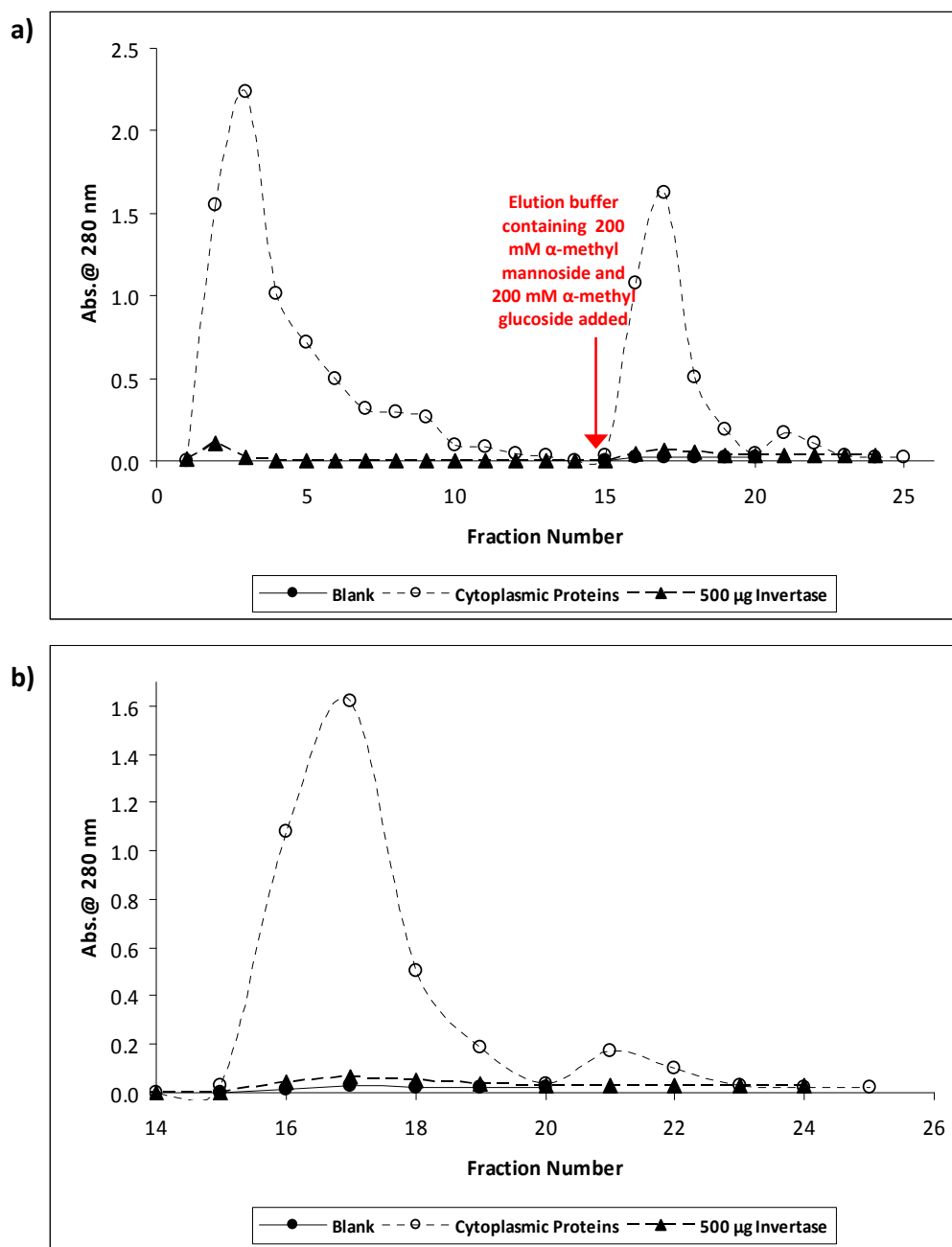
Figure 3.13 Cytoplasmic proteins from *P. luminescens* fractionated using LAC with Con A bound agarose

(a) A 2 ml sample of cytoplasmic proteins prepared using the water lysis method was passed over the column. Unbound proteins were washed from the column in fractions 1-15. Any proteins that bound to the resin were eluted and collected in fractions 16-23. Invertase was used as a positive control. (b) Enlarged area of the graph where the proteins were eluted

For the Con A LAC a significant amount of protein bound to the resin and was eluted. The proteins were eluted in two peaks, one major peak from fractions 15 to 20 and another smaller peak from fractions 20 to 23 which likely contains glycoproteins that bound more tightly to the resin and so required a larger volume of inhibiting sugar to elute them. The peak containing the eluted invertase appears very insignificant next to the eluted cytoplasmic proteins. Some of the protein fractions were run on a 15% SDS-PAGE gel and stained with silver stain as shown in figure 3.14.
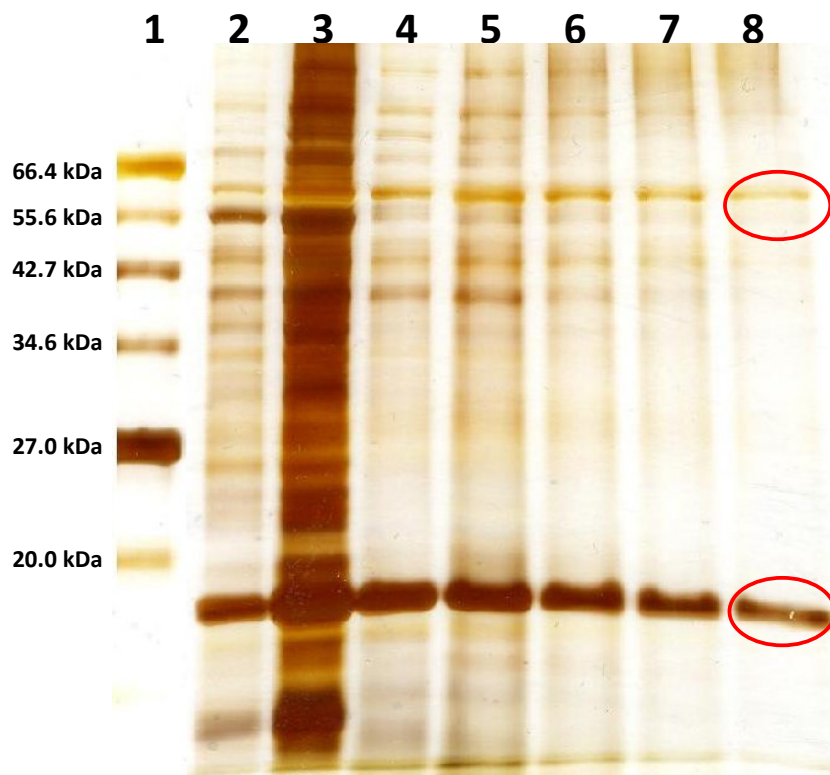
Figure 3.14 Con A LAC cytoplasmic protein fractions

Proteins were separated on a 15% SDS-PAGE gel and stained with silver stain. Lane 1 = NEB broad range protein marker, lane 2 = crude cytoplasmic proteins, lane 3 = unbound cytoplasmic proteins, lanes 4-8 eluted cytoplasmic proteins. The two most abundant proteins to be eluted from the column had molecular weights of ~60 kDa and ~17 kDa and these proteins are circled in red.

The two proteins that are circled in red in figure 3.14 were relatively abundant and had molecular weights of approximately 60 and 17 kDa. For the LAC using WGA resin with cytoplasmic proteins (figure 3.9 (b)) there were also two very strong eluted bands at ~60 and 17 kDa. It is possible that these are the same proteins because from figure

3.12 it was seen that eluted proteins from the WGA LAC also gave a very strong response with Con A in an ELLA.

3.5.3 LAC using *Griffonia Simplicifolia* Lectin I (GSL I) agarose

The final lectin that was used in LAC was GSL I. This lectin was used to purify glycoproteins from *P. luminescens* periplasmic proteins. A 5 ml sample of protein prepared by the water lysis method was passed over the column and bound proteins were eluted with buffer containing 200 mM galactose and 200 mM *N*-acetylgalactosamine. For this lectin 100 µg of the commercial glycoprotein asialofetuin (from foetal calf serum) was used as a control. Asialofetuin contains terminal β-galactose. GSL I binds most strongly to α-galactose but will also tolerate some β-galactose. Therefore it is not expected that a large amount of asialofetuin will bind to the GSL I resin however this was the best commercial glycoprotein that was available at the time to act as a control.

Figure 3.15 shows a plot of the O.D. $_{280\,nm}$ readings.
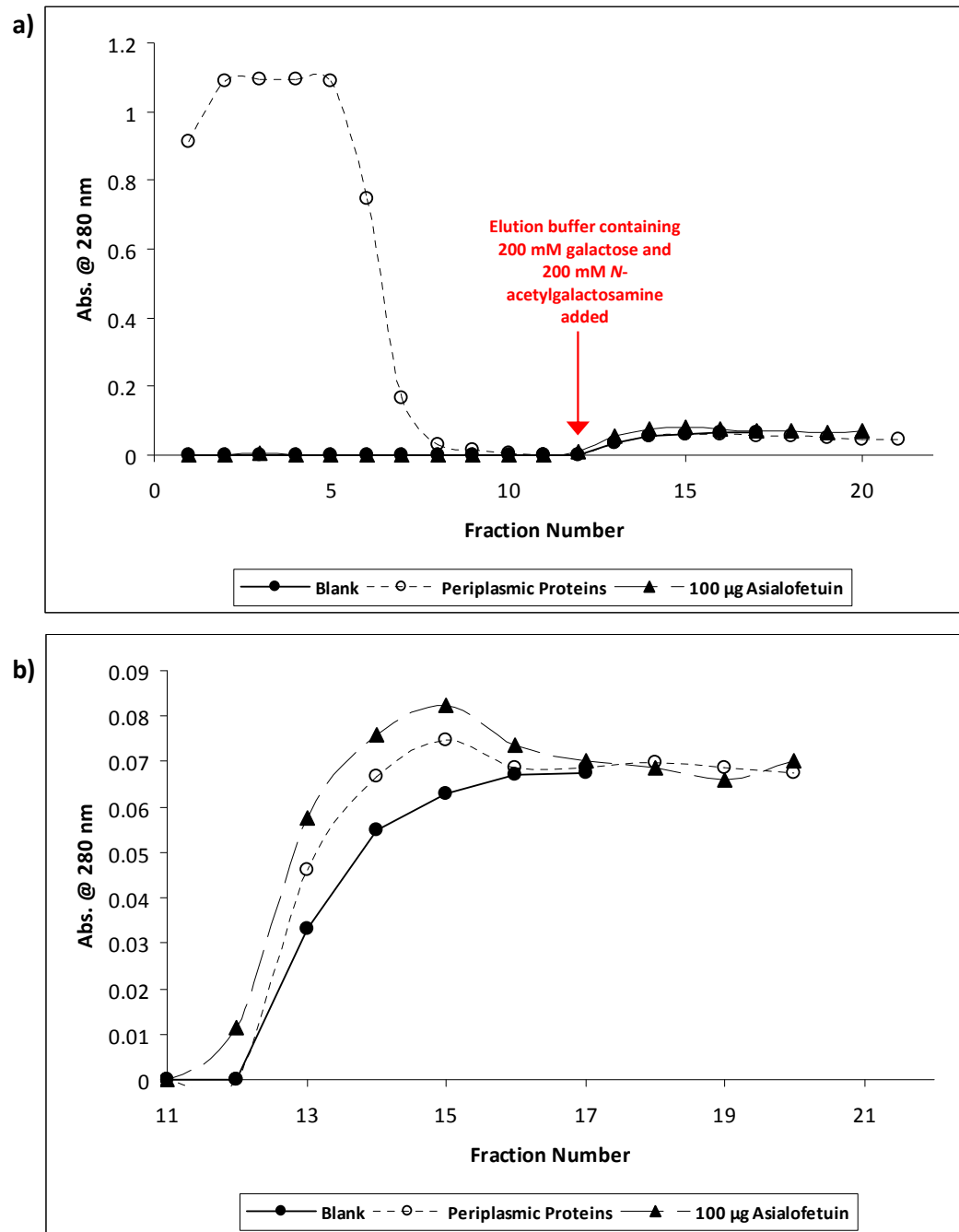
Figure 3.15 Periplasmic proteins from *P. luminescens* fractionated using LAC with GSL I bound agarose

(a) A 5 ml sample of periplasmic proteins prepared using the water lysis method was passed over the column. Unbound proteins were washed from the column in fractions 1-11. Any proteins that bound to the resin were eluted and collected in fractions 13-16. Asialofetuin was used as a positive control. (b) Enlarged area of the graph where the proteins were eluted.

It can be seen from this graph that the majority of the protein sample did not bind to the column and was washed away but a small amount of protein from the periplasm did bind to the column and was eluted in fractions 13-16. Again the fractions were run on a 15% SDS-PAGE gel which was stained with silver stain and the results are shown in figure 3.16.
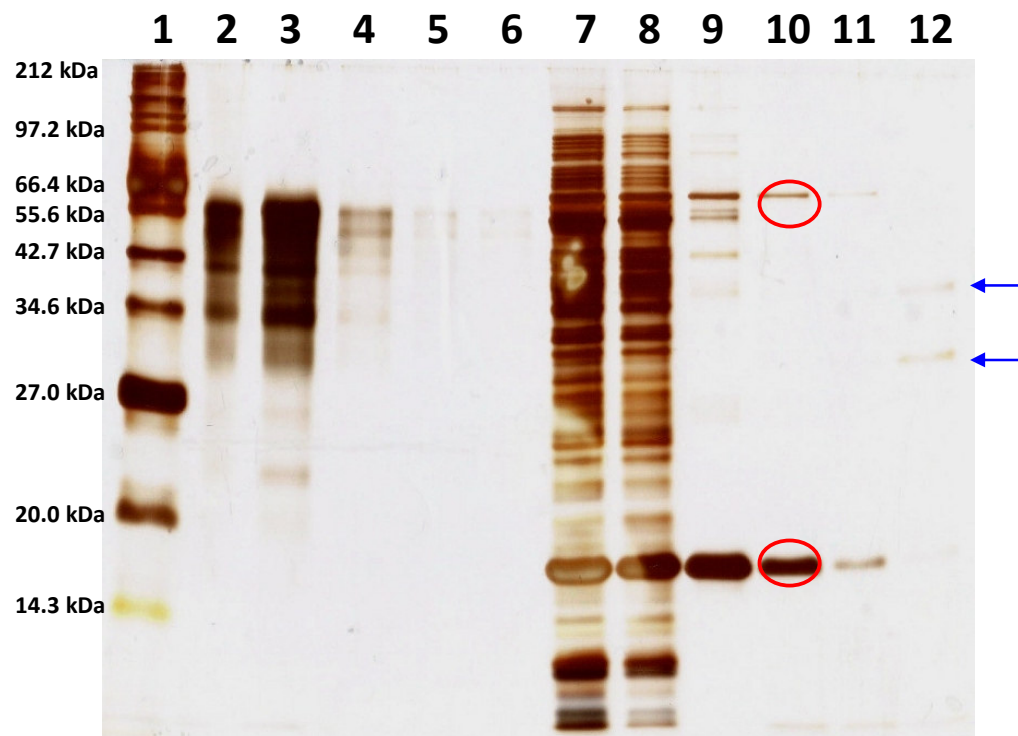
Figure 3.16 GLS I LAC periplasmic protein fractions

Proteins were separated on a 15% SDS-PAGE gel and stained with silver stain. Lane 1 = NEB broad range protein marker, lane 2 = crude asialofetuin, lane 3 = unbound asialofetuin, lanes 4-6 = eluted asialofetuin, lane 7 = crude periplasmic proteins, lane 8 = unbound periplasmic proteins, lanes 9-12 = eluted periplasmic proteins. The two most abundant proteins to be eluted from the column had molecular weights of ~60 kDa and ~17 kDa and these proteins are circled in red.

For the control glycoprotein, as expected the majority of the asialofetuin did not bind to the resin, however, a proportion of it did and it was eluted (lanes 4 and 5). Some of the proteins that were eluted from the GSL I column are similar in molecular weight to the periplasmic proteins that were eluted from the WGA column and could possibly be the same glycoproteins that could contain both galactose and GlcNAc residues. An interesting observation can be made about the proteins eluted from this column. Lane 12 contains proteins that were eluted last after the addition of 4 ml of elution buffer. The

two proteins highlighted by blue arrows are not visible in the other elution samples and must be glycoproteins that were very tightly bound to the resin and needed a large volume of elution buffer to inhibit their binding. This provides further evidence that the interaction between the lectin and the protein is carbohydrate based.

3.6 Extracellular, periplasmic, cytoplasmic and membrane protein isolation from *C. jejuni* using the water lysis method

In order to validate the study carried out using *P. luminescens* proteins, a control study was carried out using an organism that had previously been shown conclusively to contain a protein glycosylation system. The organism chosen was *C. jejuni* NCTC 11168. Many glycoproteins have been discovered in this organism and their glycan structures have been determined. Research carried out by Linton et al. (2002) discovered two glycoproteins produced by *C. jejuni* NCTC 11168 by using the lectin soybean agglutinin (SBA) to isolate them. They found that the glycoproteins contained *N*-acetylgalactosamine residues on their glycans.

In order to compare glycoproteins from the two organisms, proteins were prepared from *C. jejuni* cells in the same way as they were for *P. luminescens* i.e. using the water lysis method (section 2.15). The total concentration of protein in each fraction was determined using a BCA assay (section 2.17) and the results are shown in figure 3.17 and table 3.3.
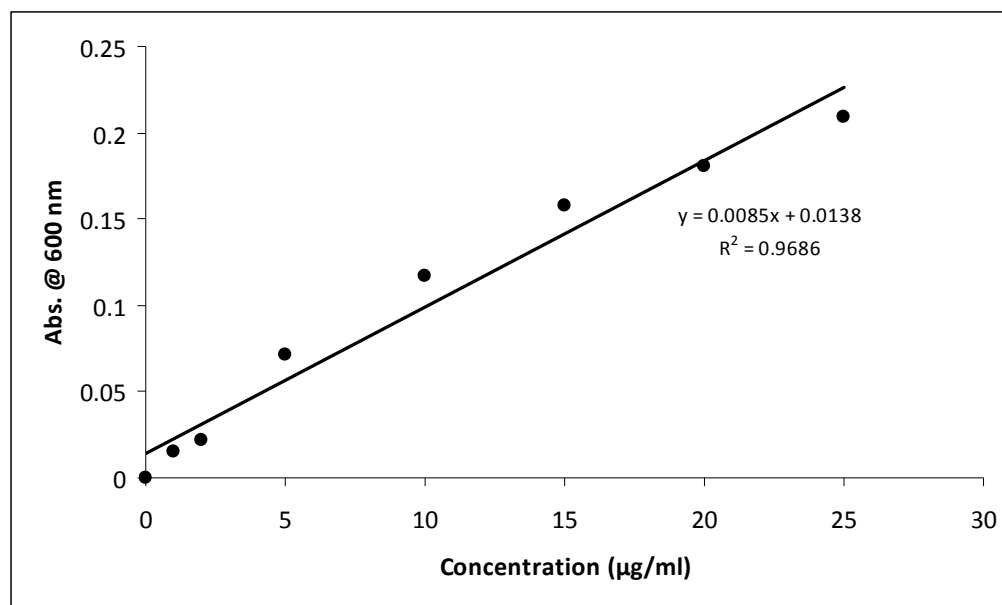
Figure 3.17 BCA assay standard curve to quantify protein concentration in *C. jejuni* water lysis samples

The equation of the standard curve y = 0.0085 x + 0.0138 was used to calculate x, the concentration of the unknown samples. The calculated concentration values are shown in table 3.3.

Table 3.3 Concentrations of *C. jejuni* water lysis samples

Sample	Concentration	Total protein
Extracellular	144 µg/ml	7.2 mg
Periplasmic	77 µg/ml	2.0 mg
Cytoplasmic	120 µg/ml	1.9 mg
Membrane	340 µg/ml	0.5 mg

It can be seen from table 3.3 that the protein concentrations were extremely low for each sample compared to those obtained for *P. luminescens* in table 3.1. This is because the cell density of *C. jejuni* in liquid cultures is very low, only reaching an O.D.$_{600 \text{ nm}}$ of approximately 0.1 at most. In order to run these samples on an SDS-PAGE gel they were first concentrated in vivaspin 500 µl centrifugal ultrafiltration devices (Sartorius Stedim Biotech). Vivaspin columns with a molecular weight cut off of 10,000 Da were used. The protein samples were added to the columns which were then centrifuged at 13,000 rpm for 10 minutes. The flow through was removed and PBS was added to the concentrated sample and the centrifugation step was repeated. Once concentrated to

~10% of their original volume the proteins were run on a 15% SDS-PAGE gel and stained with coomassie blue as shown in figure 3.18.
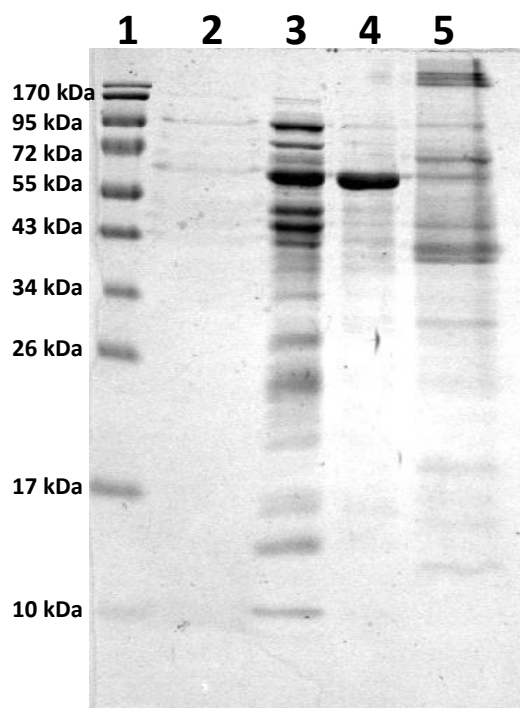
Figure 3.18 Isolated protein samples from *C. jejuni* released by the water lysis method

Protein samples were separated on a 15% SDS-PAGE gel which was then stained with coomassie blue. Lane 1 = broad range protein marker, lane 2 = extracellular proteins, lane 3 = periplasmic proteins, lane 4 = cytoplasmic proteins, lane 5 = membrane proteins.

3.7 Analysis of isolated *C. jejuni* protein fractions using western blots and ELLAs

According to the paper by Linton, et al. (2002), SBA gave the best response of all of the lectins that they tested against *C. jejuni* proteins and so a western blot that was probed with SBA using the same samples as in figure 3.18 was carried out and this is shown in figure 3.19.
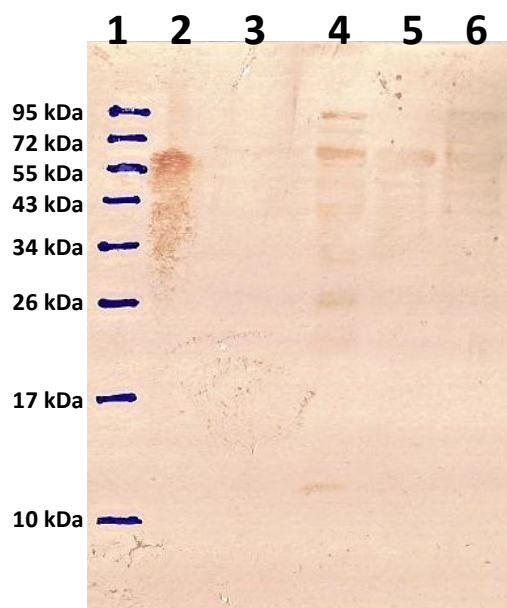
Figure 3.19 Western blot analysis showing interactions between *C. jejuni* extracellular, periplasmic, cytoplasmic and membrane proteins with the lectin SBA

Lane 1 = prestained protein marker, lane 2 = asialofetuin control, lane 3 = extracellular proteins, lane 4 = periplasmic proteins, lane 5 = cytoplasmic proteins, lane 6 = membrane proteins.

As expected, strong binding was demonstrated by the SBA blot. In order to test other lectins for which *C. jejuni* glycoproteins had affinity for an ELLA was carried out. The concentration of each sample was adjusted to 10 µg/ml with PBS and ELLAs were carried out using a range of lectins with different binding specificities. The results are shown in figure 3.20.
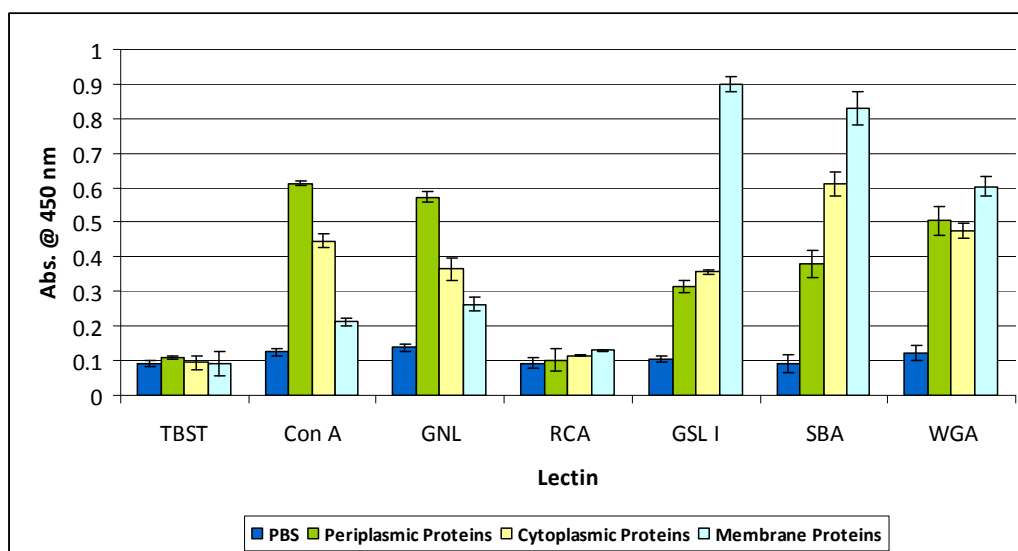
Figure 3.20 ELLA profiling interactions between *C. jejuni* protein fractions and various commercial lectins

The periplasmic and cytoplasmic proteins are producing strong signals with mannose binding lectins Con A and GNL. All proteins, especially the membrane proteins are binding strongly to GSL I, SBA and WGA. RCA, a lectin that binds β-linked galactose is not binding to the proteins in any fraction. These results correspond quite well to what was seen by Linton et al. (2002). In their analysis SBA and WGA gave the best response against whole-cell proteins and RCA gave very little response. They did not test the lectin GSL I in their analysis.

It can be seen from the ELLA analysis that the periplasmic proteins are producing a greater response with the mannose binding lectins than with the membrane proteins. On the other hand the membrane proteins are producing a stronger signal with GalNAc and α-galactose binding lectins than the periplasmic proteins. Many of the proteins found in the periplasm of a bacterial cell will ultimately end up in the membrane of the cell. The ELLA results are providing some evidence that further processing of the glycans linked to some glycoproteins may be occurring in the periplasm before they are transported to the membrane. Perhaps some of the glycans that are found on glycoproteins in the periplasm contain mannose structures initially which are then cleaved off and replaced with GalNAc or galactose residues.

A western blot of a 2-D gel was also carried out to get better separation of the proteins and a better view of what proteins are actually binding to SBA. A western blot of periplasmic proteins on a 1-D gel was compared to proteins visualised on a 2-D gel. Periplasmic proteins were chosen because they produced a strong signal with SBA on the blot in figure 3.19. The results of the 1-D analysis versus the 2-D analysis are shown in figure 3.21.
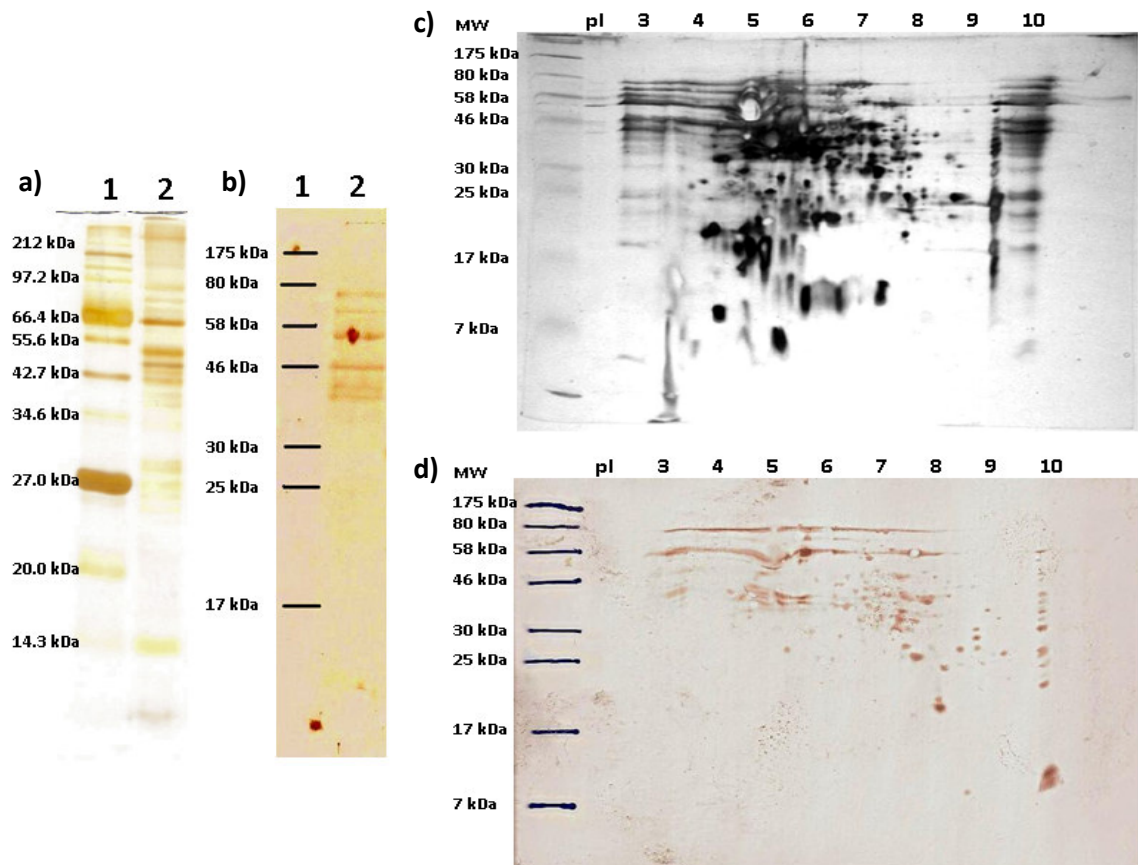
Figure 3.21 Comparison of 1-D vs. 2-D separation of *C. jejuni* periplasmic proteins

a) *C. jejuni* periplasmic proteins separated on a 15% 1-D SDS-PAGE gel and stained with silver stain. Lane 1 = NEB broad range marker, lane 2 = *C. jejuni* periplasmic proteins. b) *C. jejuni* periplasmic proteins separated on a western blot and probed with SBA lectin. Lane 1 = NEB prestained protein marker, lane 2 = periplasmic proteins. c) 2-D gel containing *C. jejuni* periplasmic proteins stained with silver stain. d) 2-D western blot containing *C. jejuni* periplasmic proteins probed with SBA.

From the 1-D analysis ((a) and (b)) it appears that almost all of the periplasmic proteins are reacting with SBA. This is because each band contains multiple proteins. However, it may be seen from (c) and (d) that in fact a lot of the periplasmic proteins are not interacting with SBA. It is mainly high molecular weight proteins with a broad range of

pI values and lower molecular weigh proteins with high pI values that are interacting with SBA. It appears that as the glycoproteins containing GalNAc decrease in molecular weight their pI increases. This can be explained by the fact that the smaller the protein, the more of an effect any modification will have on its pI. GalNAc is a negatively charged molecule and so this modification could be altering the properties of the low molecular weight glycoproteins and giving them higher pI values. It can be seen from figure 3.21 that 2-D gel electrophoresis is a much more powerful method for use in the analysis of protein glycosylation than 1-D electrophoresis.

3.8 Susceptibility of *P. luminescens* and *C. jejuni* proteins to glycosidases and β-elimination

Glycoside hydrolases or glycosidases are enzymes that catalyse the hydrolysis of glycosidic linkages. They are extremely useful molecules with which to investigate the types of glycan moieties that are present in glycoproteins. In this study *P. luminescens* and *C. jejuni* proteins were treated with various glycosidases and then the binding of treated and untreated samples to lectins were compared using ELLAs. The glycosidases used in this study and their targets are listed in table 3.4. The lectins that were used to analyse the deglycosylated proteins are also listed in table 3.4.

Table 3.4 Glycosidase enzymes used in this study and their corresponding targets

Glycosidase	Target	Lectin
Endoglycosidase H_f (Endo H_f)	The chitobiose core (GlcNAc-GlcNAc) of high mannose and hybrid *N*-linked glycoproteins	Con A
α-(1,3/6) galactosidase	α-(1,3/6) linked D-galactose	GSL I
β-*N*-acetyl-hexosaminidase$_f$	Terminal β-(1,3/4/6) linked D-*N*-acetyl-galactosamine Terminal β-(1,4) linked D-*N*-acetyl-glucosamine	SBA
α-*N*-acetyl-galactosaminidase	α-linked D-*N*-acetyl-galactosamine	SBA
Neuraminidase	α-(2,3/6/8) linked *N*-acetyl-neuraminic acid	WGA

The β-elimination reaction, a chemical means of completely removing *O*-linked glycans was also utilised in this study. This reaction will remove any *O*-linked glycan regardless of its linkage.

In order to carry out the glycosidase or β-elimination treatment of protein samples, the protein sample was concentrated to ~10 µg/µl. Glycosidase reactions were carried out as described in section 2.25.6. The β-elimination reaction was carried out as described in section 2.26. A non-treated sample was also prepared in the same way except that instead of enzyme or β-elimination reagent mixture, dH$_2$O was used. All other constituents of the reaction were identical. The same protein sample was used in treated and non-treated samples so that the signals seen in the ELLA analysis could be directly compared.

Following treatment the samples were adjusted to 10 µg/ml with the addition of PBS and ELLA analysis was carried out to see if the signal was reduced in the treated samples compared to untreated samples. In the ELLAs, lectins were used that bind to the glycan moiety for which the glycosidase is specific. So for Endo H$_f$ treatment, the lectin Con A was used. For α1-3,6 galactosidase treatment, GSL I was used. For β-*N*-acetyl-hexosaminidase$_f$ and α-*N*-acetyl-galactosaminidase, SBA was used. Finally for neuraminidase, WGA was used. WGA was also used to examine the effect of the β-elimination treatment. The results of the ELLA analysis are shown in figure 3.22.
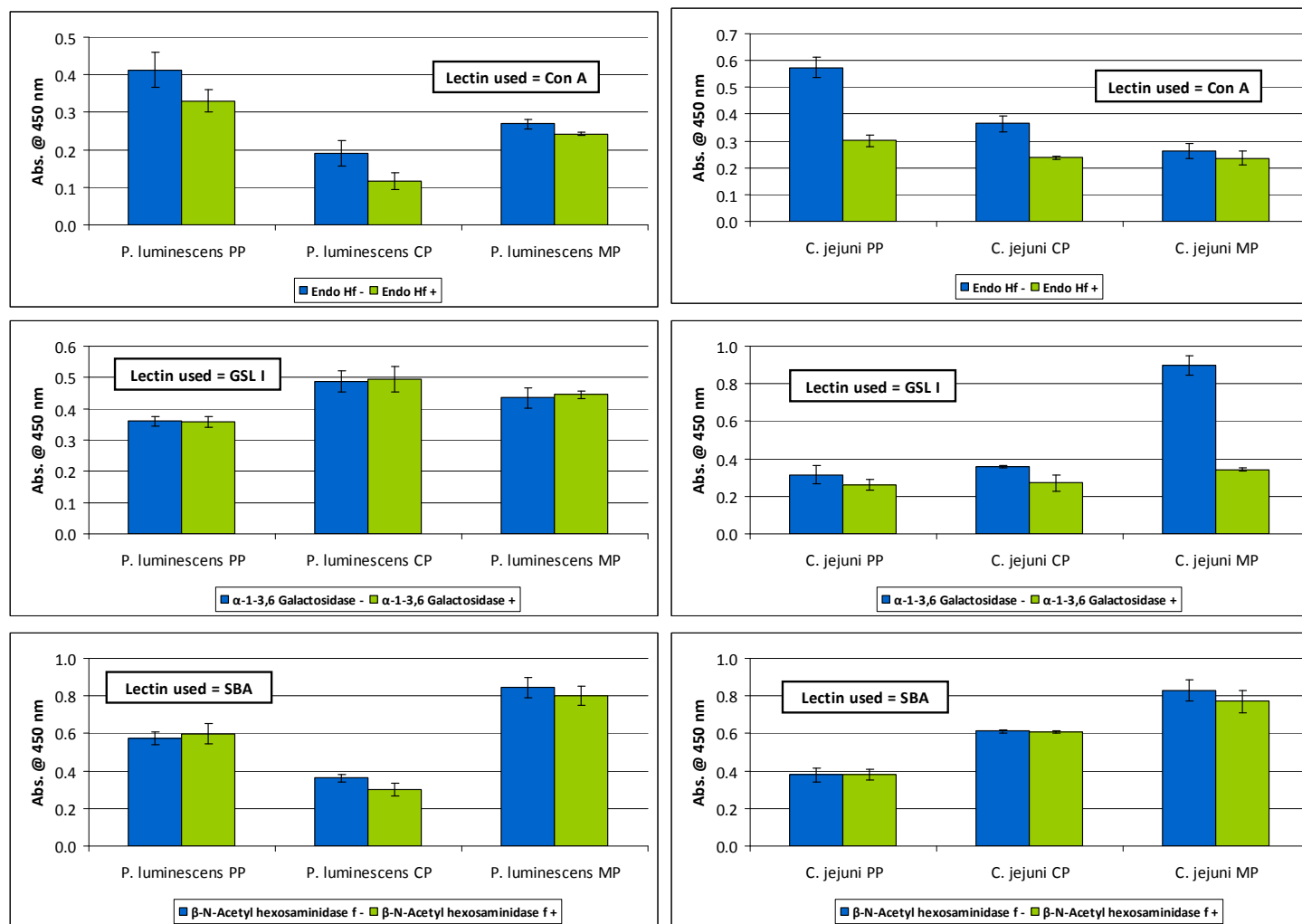
Figure 3.22 ELLA analysis of glycosidase and β-elimination treated *P. luminescens* and *C. jejuni* proteins

Untreated samples are represented by the blue bars and treated samples are represented by the green bars.
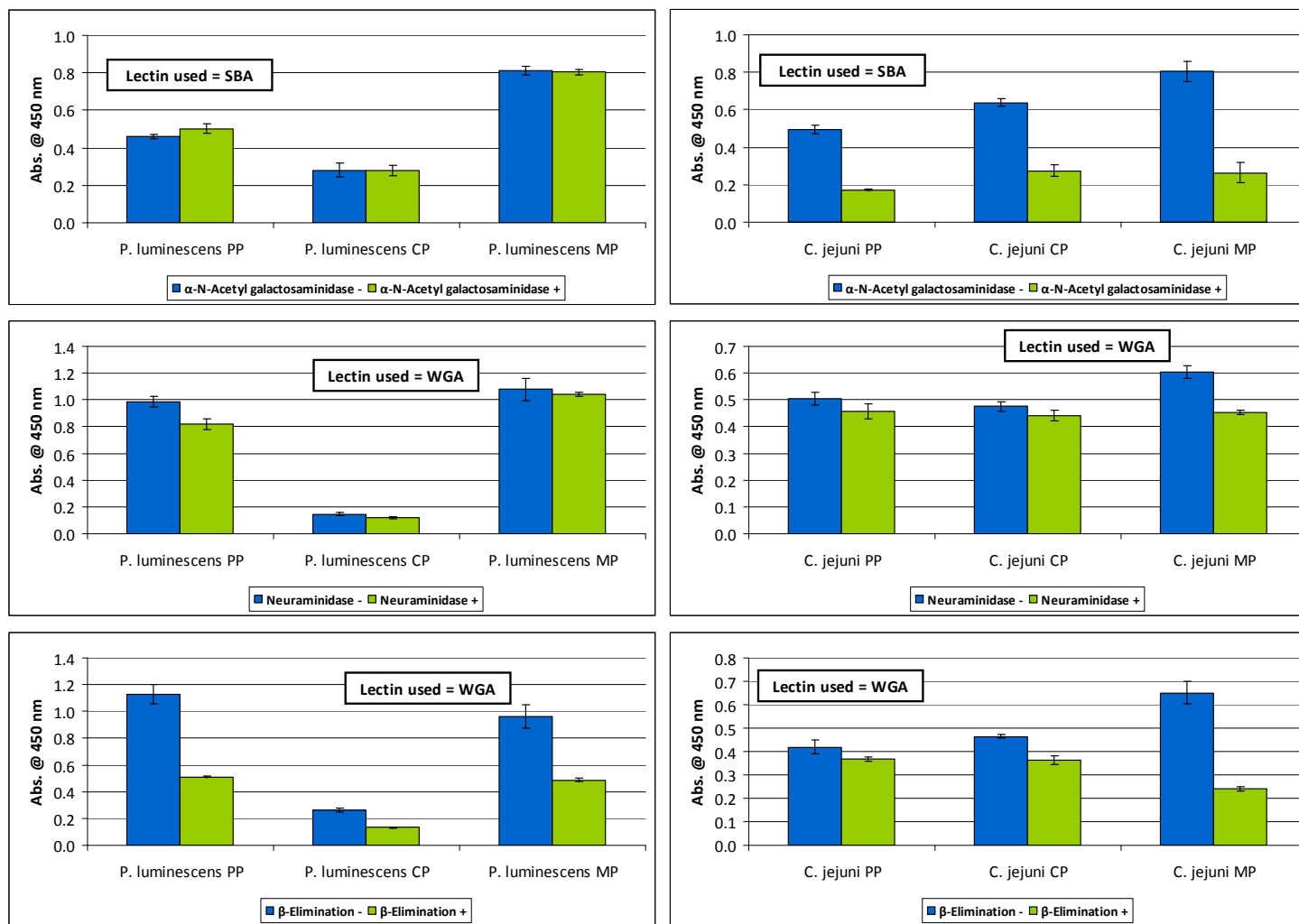
Figure 3.22 continued... ELLA analysis of glycosidase and β-elimination treated *P. luminescens* and *C. jejuni* proteins

Untreated samples are represented by the blue bars and treated samples are represented by the green bars.

126

In many cases the ELLA signal observed was reduced in the untreated samples compared to the treated samples. Endo H_f treatment reduced the binding of all protein samples to the lectin Con A. Upon researching published literature no incidences have been found where this enzyme has previously been used to cleave bacterial glycoproteins. In eukaryotic glycoproteins the enzyme cleaves the chitobiose core of an *N*-linked glycoprotein. Bacterial glycoproteins that have been discovered to date are not known to contain this chitobiose core so it is possible that this enzyme has a slightly different target in prokaryotic glycoproteins.

The galactosidase enzyme does not appear to be having any effect on *P. luminescens* proteins but treatment with this enzyme has significantly reduced the signal obtained for *C. jejuni* membrane proteins. β-*N*-acetyl-hexosaminidase$_f$ treatment does not seem to be affecting *P. luminescens* proteins or *C. jejuni* proteins to any great extent. Linton et al. (2002) also found this to be the case. However, α-*N*-acetyl-galactosaminidase treatment is having a big effect on the signal obtained for SBA binding to *C. jejuni* proteins. Again this result was seen by Linton et al. (2002). However, this enzyme is not causing a decrease in signal for the *P. luminescens* proteins. Neuraminidase treatment of *C. jejuni* membrane proteins produced a large reduction in signal. Neuraminidase targets the glycan moiety sialic acid and *C. jejuni* flagellin proteins are known to be modified with pseudaminic acid, a glycan that is very similar to sialic acid ((Thibault et al. 2001). Finally the β-elimination treatment is seen to have a significant effect on the binding of all *P. luminescens* proteins fractions and *C. jejuni* membrane proteins to WGA. Once again, it has been previously reported that *C. jejuni* flagellin proteins contain *O*-linked glycans (Szymanski et al. 2003) therefore, the fact that the signal for the membrane proteins is reduced is not surprising. These results indicate that many of the glycoproteins produced by *P. luminescens* are *O*-linked.

The fact that galactosidase and *N*-acetyl-galactosaminidase did not produce any effect on *P. luminescens* proteins is disappointing as the lectins GSL I and DBA gave the best responses in the initial western blots that were carried out (figure 3.6). This suggested that the glycoproteins that are produced by *P. luminescens* contain galactose and/or GalNAc. However the enzyme may not be able to cleave the glycans for a number of reasons including stearic hindrance and aberrant linkages between glycan moieties that the enzyme cannot recognise.

Chapter 4

Identification of a glycosylated outer membrane protein

from *P. luminescens*

4.1 Introduction

Peptide mass fingerprinting (PMF) is a technique widely used to identify unknown proteins. It involves digesting the protein with a protease enzyme such as trypsin, producing many peptides. The masses of the peptides are then accurately determined by mass spectrometry (MS) (Pappin et al. 1993). Trypsin cleaves proteins at the carboxyl side of lysine and arginine only so that these cleavages are specific. Various online tools exist that theoretically cleave all proteins from a database of interest into peptides using the protease of choice and calculates the masses of these peptides. These tools then compare the masses of the peptides generated from the digestion of the unknown protein to those of known proteins in the database. The protein with the best match is chosen. This method of protein identification is very advantageous because no information needs to be known about the protein except for the masses of the peptides. However, in order for the identification to be successful the protein must be present in the chosen database. This technique does not work well for complex protein mixtures so proteins must first be separated using SDS-PAGE or 2-D gel electrophoresis.

The gel is stained with silver stain and the protein band or spot of interest is excised from the gel. The protein in the band is then treated with dithiothreitol (DTT) to reduce any disulphide bonds. Next the protein is treated with iodoacetamide to modify cysteine residues in the reduced protein to produce carbamidomethyl cysteine which helps to increase the strength of the signal produced during mass spectrometric analysis. The protein is digested into peptides using trypsin and the resultant peptides are extracted from the gel with acetonitrile, dried under vacuum and then resuspended in the appropriate solvent before analysis by MS. The procedure used in this study for the in-gel trypsin digestion of proteins is described in section 2.25.5.

Matrix assisted laser desorption ionization MS (MALDI-MS) was used in this study to analyse digested peptides. The analyte (the peptides) must be ionised in order to be analysed by the mass spectrometer. Ionization is triggered by a laser beam. The solvent chosen to resuspend the analyte is important for facilitating ionization. For positive ionization an acidic solvent such as trifluoroacetic acid (TFA) or formic acid is used to resuspend the analyte. In MALDI-MS the solvent donates one proton (H^+) to each

peptide to give the peptides a positive charge. The observed mass of each peptide is therefore the molecular weight of the peptide plus the molecular weight of one proton $(M_r+H)^+$. In order to calculate the actual molecular weight of each peptide, 1.008 Da (the molecular weight of H^+) must be subtracted from the observed mass. Ionization by the laser is a very harsh process so the peptides are mixed with a matrix to protect them from being destroyed by being directly hit by the laser. The matrix absorbs the energy from the laser and passes it on to the analyte.

Once the sample is ionised, the ions are extracted to the analyser where they are separated based on their mass (m) to charge (z) ratio (*m/z*). The separated ions are detected and a spectrum is generated by the operating software indicating the *m/z* ratios of the components of the mixture and their relative abundance.

4.2 Identification of proteins isolated by lectin affinity chromatography using peptide mass fingerprinting

Chapter 3 described the isolation of many putative glycoproteins from the periplasm and cytoplasm of *P. luminescens* using LAC. Putative glycoproteins from the periplasm were chosen to attempt to identify them using PMF as described in section 4.1. Figure 4.1 shows the proteins isolated from the periplasm using LAC and those that were chosen to be identified using PMF. This gel was also shown previously in figure 3.9.
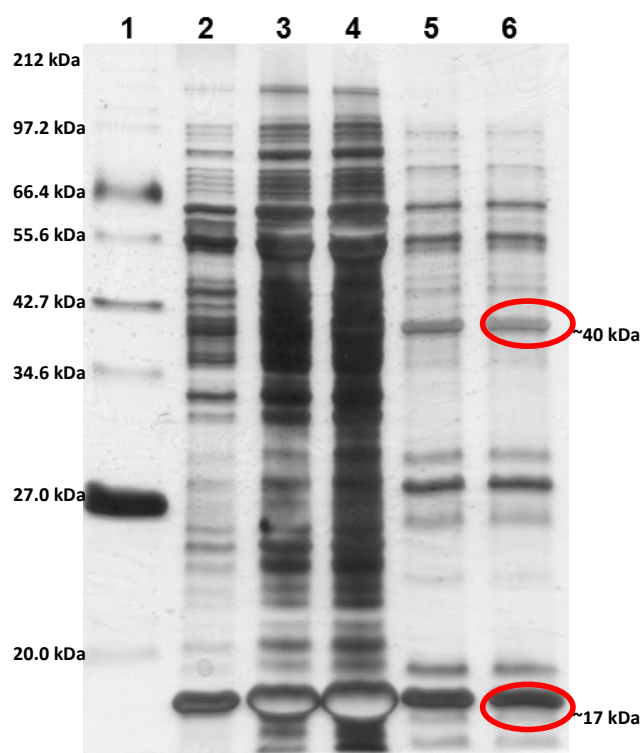
Figure 4.1 *P. luminescens* periplasmic proteins isolated by WGA lectin affinity chromatography

Proteins were separated on a 15% SDS-PAGE gel which was subsequently stained with silver stain. Lane 1 = NEB Broad Range Protein Marker, lane 2 = crude periplasmic fraction from water lysis method, lanes 3, 4 = unbound periplasmic proteins from LAC, lanes 5, 6 = eluted periplasmic proteins from LAC. The proteins circled in red at ~40 kDa and ~17 kDa were chosen for identification by PMF.

4.2.1 Identification of the isolated 40 kDa putative glycoprotein from the periplasm of *P. luminescens*

The 40 kDa band from figure 4.1 was excised from the gel and digested into peptides using the in-gel trypsin digestion method described in section 2.25.5. Generated peptides were analysed using MALDI-MS according to the procedure outlined in section 2.27. Two mass spectrometers were used to analyse the sample, the Q-TOF Ultima™ Global and the Q-TOF Premier. The Q-TOF Premier is an upgraded version of the Q-TOF Ultima™ Global with the main difference between the two instruments being the hit frequency of the laser. For the Q-TOF Ultima™ Global the maximum hit frequency obtainable is 10 Hz and for the Q-TOF Premier the maximum hit frequency is 200 Hz. Therefore the Q-TOF Premier can generate a much stronger signal compared

to the Q-TOF Ultima™ Global as the amount of times the laser is hitting the sample is increased twenty-fold. Some analytes do not ionise as well as others and so may create very low intensity peaks. This is true for glycopeptides. They do not tend to take on the positive charge as readily as a non-glycosylated peptide due to their low hydrophobicity (Zhang and Reilly 2009). Using an instrument with a higher hit frequency should increase the sensitivity of the instrument and allow more of the lower intensity peaks to be seen. This means that better protein coverage may be obtained and any glycopeptides that are present may be detected. Mass spectra were generated for the 40 kDa protein using both instruments and the results are shown in figure 4.2.
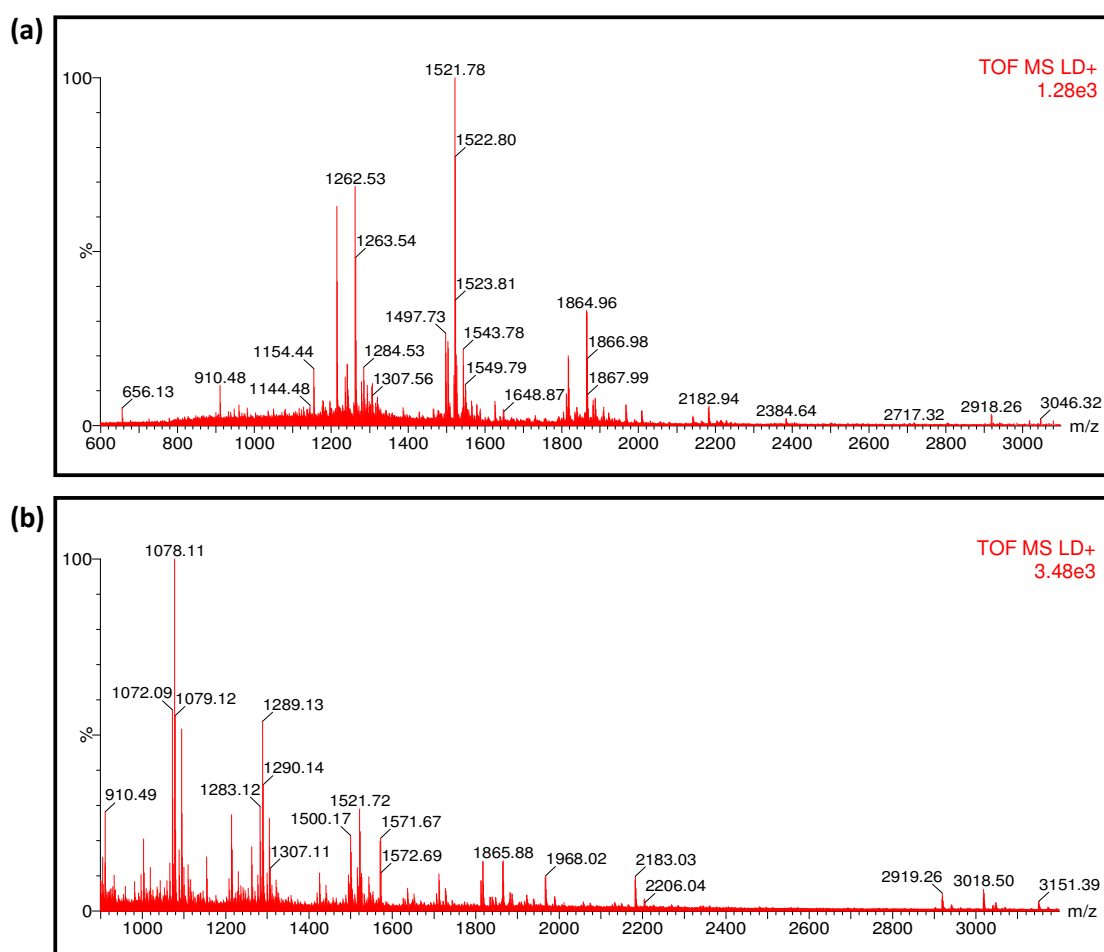
Figure 4.2 Mass spectra generated from the trypsin digestion of the 40 kDa protein isolated from the periplasm of *P. luminescens* by LAC

(a) Peptides were analysed using the Q-TOF Ultima™ Global mass spectrometer. (b) Peptides were analysed using the Q-TOF Premier mass spectrometer.

The total ion count (LD+) was higher for the Q-TOF Premier (3.48×10^3, figure 4.2 (b)) than for the Q-TOF Ultima™ Global (1.28×10^3, figure 4.2 (a)). Most of the peaks that are showing up are common to both spectra however there is a relatively large difference in intensity for some of the peaks between the two spectra.

As previously stated, the pattern of molecular weights of peptides generated from the proteolytic digestion of a protein may be used to identify that protein. MASCOT (section 2.12) is a search engine that uses mass data to identify proteins from primary sequence databases. MASCOT lets the user select different parameters to use in the search, for example, the enzyme used for digestion, any fixed or variable modifications of the peptides (e.g. carbamidomethylation of cysteine, oxidation of methionine, etc.), the primary sequence database which you would like to search (MSDB, NCBI or SwissProt), the number of missed cleavages allowable, the error allowed etc. A report is generated showing the closest protein match and the percentage sequence coverage. A MASCOT score is also calculated with any score over 78 being regarded as a significant match.

MASCOT was used to identify the protein from the masses of the peptides in figure 4.2 (b). The database searched was NCBI, one missed cleavage was allowed, and a tolerance of ±0.05 Da was used. The protein identified was an outer membrane protein N precursor (porin OmpN) from *P. luminescens* subsp. laumondii TTO1.

The MASCOT score obtained for the matched protein, OmpN, was 313 and the percentage sequence coverage was 44%. Table 4.1 shows the matched peptides, their position in the sequence, the difference between the expected molecular weight and the calculated molecular weight and the number of missed cleavages. Figure 4.3 shows the sequence of OmpN precursor.

Table 4.1 Peptides generated from the digestion of the 40 kDa protein (from figure 4.2 (b)) that matched to the sequence of OmpN

Residue	Observed mass (Da)	Expected Mr (Da)	Calculated Mr (Da)	Delta	Miss	Sequence
29 – 38	1122.56	1121.5527	1121.5717	-0.0189	1	K.DGNKLDLYGK.V
50 – 60	1154.48	1153.4727	1153.4636	0.0091	0	K.SGEDGDNSFAR.L
61 – 78	1967.01	1966.0027	1956.9909	0.0119	1	R.LGFKGETQVNDQLTGFGR.W
65 – 78	1521.72	1520.7127	1520.7220	-0.0092	0	K.GETQVNDQLTGFGR.W
85 – 96	1306.57	1305.5627	1305.5797	-0.0169	1	K.GNRAEGTDEETK.T
88 – 96	979.42	978.4127	978.4142	-0.0014	0	R.AEGTDEETK.T
88 – 98	1236.54	1235.5327	1235.5630	-0.0302	1	R.AEGTDEETKTR.L
150 – 157	910.49	909.4827	909.4620	-0.0092	0	R.STGLLTYR.N
190 – 219	3048.34	3047.3327	3047.3533	-0.0206	1	R.KDNGDGYGFSATYDVGYGI SVGGAYSNSAR.T
191 – 219	2920.27	2919.2627	2919.2584	0.0043	0	K.DNGDGYGFSATYDVGYGIS VGGAYSNSAR.T
249 – 257	975.44	974.4327	974.4818	-0.0494	0	R.AESWNIGAK.Y
258 – 273	1864.88	1863.8727	1863.8461	0.0266	0	K.YDANNIYLAAMYGETR.N
294 – 319	3017.51	3016.5027	3016.5294	-0.0267	0	K.TQNIELTAQYQFDFGLRPSL AYVQSK.G
339 – 350	1503.71	1502.7027	1502.7082	-0.0054	0	K.YISVGSYYYFNK.N
351 – 359	1116.54	1115.5327	1115.5499	-0.0172	0	K.NLTTYVDYK.I
360 – 370	1293.66	1292.6527	1292.6612	-0.0085	0	K.INLLDSNEFTK.A

```
  1 MMKRNVLAVV IPALLAAGAA NAAEVYNKDG NKLDLYGKVE VQHKFANNKS
 51 GEDGDNSFAR LGFKGETQVN DQLTGFGRWE YNIKGNRAEG TDEETKTRLA
101 FAGLKLSNYG SLNYGRNYGV VYDVNAWTDV LPVFGGDSMA QTDNFMTGRS
151 TGLLTYRNTD FFGLVDGLNF ALQYQGENSE NTANSRINRK DNGDGYGFSA
201 TYDVGYGISV GGAYSNSART SAQKTAFSVV EEKGKKDKPT GSTALGDRAE
251 SWNIGAKYDA NNIYLAAMYG ETRNMTRFGG NVDGVNWENI ANKTQNIELT
301 AQYQFDFGLR PSLAYVQSKG KNLLGDDGKE SSSRDLVKYI SVGSYYYFNK
351 NLTTYVDYKI NLLDSNEFTK ATGTSTDNVV GVGMIYQF
```

Figure 4.3 The sequence of the outer membrane porin protein N precursor, OmpN

The peptides that were generated from the digestion of the 40 kDa protein that matched to the sequence of OmpN are shown in bold red. 44% of the sequence was covered.

44% is a relatively high percentage sequence coverage. It is never possible to get 100% coverage as many of the peptides generated after digestion will be only one or two amino acids in length and so will be too small. Some of the peptides may not be capable of accepting a proton and so will not ionise. Sometimes the trypsin cannot completely

cleave a protein because some of the cleavage sites will be buried within the 3D structure of the protein and will not be accessible.

The data from figure 4.2 (a) was also inserted into MASCOT to confirm the identification of the protein. This data also matched to OmpN from *P. luminescens* (data not shown) however the sequence coverage was a lot lower at 22% and the MASCOT score was 80, just above the threshold level for a significant match. The reason for the lower coverage is that some of the lower intensity peptides were not being detected by the Q-TOF Ultima™ Global. For example peptides 29 – 38 (1121.5717 Da), 88 – 96 (974.4818 Da) and 294 – 319 (3016.5294 Da) were not detected by the Q-TOF Ultima™ Global. The potential error was also about ten times greater for the Q-TOF Ultima™ Global. The peptide tolerance used for the Q-TOF Ultima™ Global data was ±0.5 Da whereas for the Q-TOF Premier it was only ±0.05 Da. A much tighter calibration could be achieved using the Q-TOF Premier making the MASCOT score much higher.

To unambiguously confirm that the protein was indeed OmpN, a gel slice containing the protein of interest was sent to the company, Alphalyse. They carry out specialised protein identification using similar techniques as those used in this study. Alphalyse confirmed that the protein was indeed OmpN. They achieved a sequence coverage of 20% and a MASCOT score of 347. They also chose four of the matching peptides and carried out peptide sequencing using MS/MS (section 1.4.3). The sequences obtained for the four peptides using this technique matched exactly to the expected peptide sequences.

Porins are a family of proteins found in the outer membrane of Gram-negative bacteria. They are synthesised at the inner side of the cytoplasmic membrane as precursor proteins. The precursor protein contains a signal peptide at its N-terminus that functions in directing the transport of the protein across the cytoplasmic membrane into the periplasm. Once in the periplasm, the signal peptide is cleaved by a signal peptidase and the mature protein is integrated into the outer membrane (Lengeler et al. 1999). Porins are transport proteins that form aqueous channels in the outer membrane to facilitate passive diffusion of hydrophilic molecules in and out of the bacterial cell.

The OmpN porin precursor is a 388 amino acid protein with a molecular weight of 42,542 Da. OmpN and its orthologues contain a signal peptide at the beginning of the protein at the amino terminus (Inokuchi et al. 1982, Duchene et al. 1988). In OmpN from *P. luminescens* this signal peptide is 22 amino acids in length. This peptide is cleaved in the periplasm and the mature protein contains 366 amino acids and has a molecular weight of 40.4 kDa. This is the molecular weight of the peptide backbone only. It is expected that if the protein is glycosylated the molecular weight of the modified protein would be higher than this. On an SDS-PAGE gel it is difficult to determine an accurate molecular weight but it appears in figure 4.1 that OmpN is migrating at an apparent molecular weight of ~40 kDa. This means that the modifying glycan is not very large, perhaps only several glycan units in length.

Many of the peaks that were seen in figure 4.2 (b) did not match to the *P. luminescens* protein OmpN. There are several reasons for this. Some of the non-matching peaks could be contaminants. One of the most common contaminants in this type of analysis is keratin. Keratin is very abundant in the air and comes from the shedding of skin and hair. The unmatched peaks in figure 4.2 were compared against a list of the most common contaminant keratin peaks but none of them appeared as if they were coming from keratin contamination. However these were only the common contaminants and there are thousands of other possible peptides that could be contaminating the sample. Sometimes the matrix can form ions giving rise to contaminant peaks. Some of the unmatched peaks could be peptides that were cleaved at some point other than at lysine or arginine residues. Peptides may get fragmented during the preparation process or in the mass spectrometer itself if the conditions are harsh (Hillenkamp and Peter-Katalinić 2007).

Another possible reason that peptides may not match is that they could be glycosylated. It has already been discussed in section 1.3 that outer membrane proteins such as pili and flagella are glycosylated in some pathogenic organisms (*C. jejuni*, *Neisseria spp*. and *P. aeruginosa*). Apart from surface appendages, other structural outer membrane proteins have been discovered. The glycosylation of the outer membrane nitrate reductase, AniA in *Neisseria* spp. has already been discussed in section 1.3.4. Other glycosylated structural outer membrane proteins that have shown to be glycosylated include the major surface lipoproteins OspA and OspB from the Gram-negative

spirochete *Borrelia burgdorferi*, the causative agent of Lyme disease (Schwan and Piesman 2002). These proteins are differentially expressed during infection by this organism and have been shown to be directly implicated in pathogenesis. *Chlamydia trachomatis*, which causes sexually transmitted disease and the eye infection, trachoma, produces a 40 kDa major outer membrane protein that is the main structural protein of a complex called the infectious extracellular elementary body (Swanson and Kuo 1991). In the strain serovar L2/434/bu this 40 kDa protein has been shown to be glycosylated with an *N*-linked glycan containing 8 to 9 mannose residues. When this modification is absent the bacteria are no longer capable of attaching and infecting HeLa cells. Therefore, this glycosylation has a direct affect on the pathogenicity of this organism (Swanson and Kuo 1991, Kuo et al. 1996). The above examples are all instances of protein glycosylation of structural outer membrane proteins. Following extensive examination of the literature no instances could be found where outer membrane porins have been shown to be glycosylated.

The evidence provided so far in this study shows that there is a strong possibility that OmpN is glycosylated. Any of the unmatched peptides in figure 4.2 could be glycopeptides. Some of the larger peaks visible in figure 4.2 (b), for example 2183.03 Da, 2919.26 Da or 3018.50 Da, are likely to be glycopeptides because of their size and because their intensity is very low (as discussed earlier, glycopeptides do not ionise very well). In order to investigate if these peptides are glycopeptides the glycan would have to be removed by chemical or enzymatic means (see next section 4.2.2). The OmpN protein could also be purified and its exact molecular weight could be determined by MS analysis. By comparing the molecular weight of the peptide backbone to the measured mass of the purified protein, it could be determined if there are any modifications present (section 4.3.4).

4.2.2 PNGaseF treatment of OmpN to investigate *N*-linked glycosylation

Many previous studies have been done to identify sites of *N*-linked glycosylation using the enzyme PNGaseF in eukaryotic glycoproteins (Fan et al. 2004, Zhang et al. 2008). PNGaseF (also known as *N*-glycosidase F) is an endoglycosidase enzyme that cleaves the entire *N*-linked glycan from a protein. It cleaves between the innermost GlcNAc of

an *N*-linked glycan and the asparagine residue. In order to use PNGaseF to identify sites of *N*-linked glycosylation, a trypsin digested protein is treated with the enzyme to remove any *N*-linked glycans, then the spectrum obtained for the untreated protein digest and the PNGaseF treated protein digest are compared. If a glycopeptide is modified with an *N*-linked glycan, cleaving it will decrease the molecular weight of the peptide. The spectrum of the PNGaseF treated protein will contain a new peak that was not seen in the original spectrum. The sequence of the peptide may be determined by MS/MS peptide sequencing and the site of the modification found by analysing the peptide sequence for the consensus sequence Asn-X-Ser/Thr.

OmpN contains three possible *N*-linked glycosylation sites. These are at residue 274 (sequence = **N**MT), residue 292 (sequence = **N**KT) and residue 351 (sequence = **N**LT). The consensus sequence at residue 351 is likely not to be glycosylated because a peptide containing this residue (residues 351 – 359 (table 4.1)) was identified in the initial MS analysis. Therefore, this residue is unlikely to contain a modification. Peptides containing residues 274 and 292 were not identified and therefore they are possible sites of *N*-linked protein glycosylation. To investigate if OmpN contained any *N*-linked glycans the trypsin digested protein was treated with PNGaseF according to the procedure described in section 2.25.6. The treated peptides were analysed using MALDI-MS. The results are shown in figure 4.4.
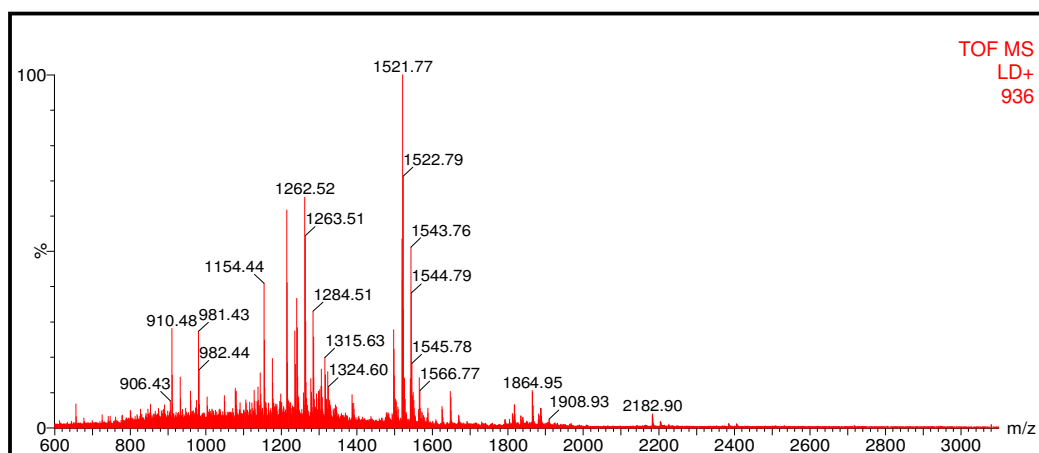
Figure 4.4 Mass spectrum generated from the analysis of the PNGaseF treated OmpN digest
Peptides were analysed using the Q-TOF Ultima™ Global mass spectrometer.

The peaks obtained for the PNGaseF treated OmpN digest were compared to the untreated digest in figure 4.2 (a). If *N*-linked glycans were present on either residue 274 or 292, removing them would yield peptides with molecular weights of 520.61 Da (NMTR) and 1733.83 Da (FGGNVDGVNWENIANK) or 2236.43 Da if there was one missed cleavage (NMTRFGGNVDGVNWENIANK). However, the spectra before and after PNGaseF treatment are almost identical. There are no new peaks showing up that were not seen before. This suggests that if this protein is glycosylated then the glycan is not *N*-linked but is perhaps *O*-linked.

Unfortunately there is no enzyme available commercially that removes an entire *O*-linked glycan like PNGaseF does for *N*-linked glycans. A sequential enzymatic deglycosylation may be performed using a range of exoglycosidases to remove monosaccharides one by one from the glycan but it is not possible to know what enzyme to use since it is not known which monosaccharide is at the terminus of the glycan. Glycosidases can be quite expensive and so this type of analysis may not be possible. An alternative is to use chemical deglycosylation to remove *O*-linked glycans. Alkaline β-elimination is the most common *O*-deglycosylation strategy. It uses a mixture of sodium hydroxide and sodium borohydride to release the *O*-glycans (Carlson 1968). However, alkaline β-elimination can sometimes damage the protein structure leading to difficult determination of the results.

4.2.3 Identification of the isolated 17 kDa putative glycoprotein from the periplasm of *P. luminescens*

A second protein isolated from the periplasm by LAC with the lectin WGA was also excised and identified using PMF. This protein came from the 17 kDa band circled in figure 4.1. The protein was digested with trypsin as before and the peptides were analysed using MALDI-MS. The spectrum can be seen in figure 4.5.
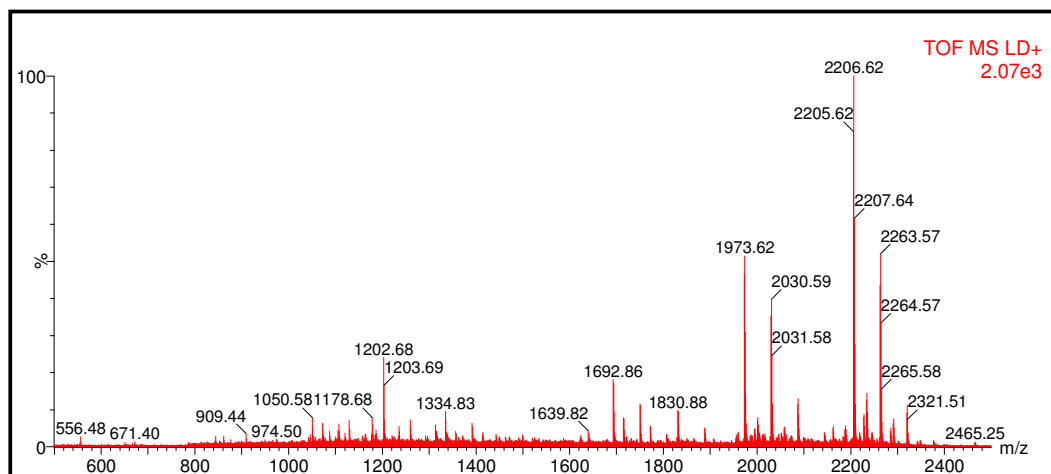
Figure 4.5 Mass spectrum generated from the trypsin digestion of the 17 kDa protein isolated from the periplasm of *P. luminescens* by LAC

Peptides were analysed using the Q-TOF Ultima™ Global mass spectrometer.

The signal obtained for this sample was quite strong with a total ion count of 2.07 x 10³. MASCOT was again used to identify the digested 17 kDa protein from the masses of the peptides in figure 4.5. The database searched was NCBI, one missed cleavage was allowed, and a tolerance of ±0.5 Da was used. The protein identified was hypothetical protein, plu3795 from *P. luminescens* subsp. laumondii TTO1. The MASCOT score was 180 and the percentage sequence coverage was 67%. Table 4.2 shows the matched peptides as before and figure 4.6 shows the sequence of plu3795.

Table 4.2 Peptides generated from the digestion of the 17 kDa protein (from figure 4.5) that matched to the sequence of plu3795

Residue	Observed mass (Da)	Expected Mr (Da)	Calculated Mr (Da)	Delta	Miss	Sequence
14 – 30	1832.87	1831.8627	1832.0077	-0.1450	1	K.LESMLDASVRTVITAVK.N
35 – 59	2206.6200	2205.6127	2205.1641	0.4486	0	K.VTVGPLPQTPELIPSASEQSR.Y
60 – 74	1692.86	1691.8527	1691.8770	-0.0243	0	R.YTIIPPVGPTTYQDK.V
60 – 77	2033.58	2032.5727	2032.1245	0.4482	1	R.YTIIPPVGPTTYQDKVLK.F
78 – 89	1334.84	1333.8327	1333.7606	0.0722	0	K.FTTVVQSQALLK.K
90 – 98	1178.6900	1177.6827	1177.6244	0.0583	1	K.KIWVFQDSR.D
91 – 98	1050.57	1049.5627	1049.5294	0.0333	0	K.IWVFQDSR.D
91 – 105	1830.87	1829.8627	1830.0040	-0.1412	1	K.IWVFQDSRDPLLTIK.Y
106 – 121	1972.62	1971.6127	1971.8751	-0.2624	0	K.YWVGEQFNYDDANPVR.V

```
  1 MSEMNEMKVK IEKLESMLDA SVRTVITAVK NDLSSSLKVT VGPLPQTPEL
 51 IPSASEQSRY TIIPPVGPTT YQDKVLKFTT VVQSQALLKK IWVFQDSRDP
101 LLTIKYWVGE QFNYDDANPV RVSAEGGGNK FMIINEDSDG IAGIAVTMY
```

Figure 4.6 The sequence of plu3795, a hypothetical protein from *P. luminescens*

The peptides that were generated from the digestion of the 17 kDa protein that matched to the sequence of plu3795 are shown in bold red. 67% of the protein sequence was matched.

plu3795 is a hypothetical protein. This means that the existence of the protein has been predicted usually through identification of open reading frames during analysis of the genome, but no experimental evidence exists that the protein is expressed *in vivo* (Zarembinski et al. 1998). BlastP searches using the sequence of plu3795 against all organisms revealed an orthologue in *Photorhabdus asymbiotica* which was also a hypothetical protein but no significant orthologues were found in any other organisms. A search of the sequence for conserved domains also yielded no results. Because so little is known about this protein no further work was carried out on it.

4.3 The cloning of *ompN*, and expression and purification of the encoded protein

PMF is a powerful tool for identifying protein samples (Henzel et al. 1993). Only µg quantities of sample are required and the protein does not have to be in pure form as the protein mixture is first separated using gel electrophoresis. However, in order to carry out other analyses, for example molecular weight determination of the intact protein, enzymatic/chemical removal of glycans and ELLA/western blot analysis of deglycosylated protein, a larger quantity of pure protein sample is required.

For this reason a recombinant OmpN protein was produced. The *ompN* gene was amplified using PCR then ligated to a broad host range cloning vector pBBR1MCS-5 (figure 2.1). The construct was transformed into *E. coli* then delivered to *P. luminescens* by conjugation in order to express OmpN in its natural environment so that the protein would be correctly glycosylated.

4.3.1 The cloning of *ompN*

P. luminescens genomic DNA was prepared as outlined in section 2.6. A 1176 kb region of DNA was amplified that included the *ompN* gene with its signal peptide and also its ribosome binding site, a five base sequence which lies upstream from the start codon, to ensure correct translation of the gene. Primers were designed to incorporate a HindIII restriction site at the beginning of the PCR product and a 6 x histidine tag followed by a BamHI restriction site at the end of the PCR product. The PCR reaction was carried out as outlined in sections 2.25.2 and 2.25.3. The region of DNA to be amplified and the primers used are shown in figures 4.7 and 4.8.

```
2089282 5' ATGAGGGTAATAATAATGATGAAACGCAATGTTCTTGCAGTAGTAATCCCAGCTCTGT
TAGCTGCTGGGGCAGCAAATGCAGCTGAAGTTTATAATAAAGACGGCAACAAACTGGACCTATACGGTAA
AGTTGAAGTCCAGCACAAATTCGCTAACAACAAAAGTGGCGAAGATGGCGATAACTCCTTCGCCCGTCTG
GGTTTCAAAGGTGAAACTCAAGTTAATGACCAACTGACCGGTTTTGGCCGTTGGGAATACAATATCAAAG
GTAACCGTGCAGAAGGTACTGACGAAGAAACCAAAACGCGTTTAGCTTTCGCTGGTTTGAAACTTTCTAA
CTACGGTTCACTGAACTACGGCCGTAACTACGGTGTGGTTTATGATGTCAACGCTTGGACCGATGTGTTG
CCAGTATTTGGTGGCGACTCAATGGCGCAGACTGACAACTTCATGACCGGCCGTTCTACTGGTCTGTTGA
CTTACCGCAATACCGATTTCTTCGGTTTAGTTGATGGCCTGAATTTTGCCCTGCAATATCAGGGTGAAAA
TAGTGAAAACACGGCTAATAGCCGTATCAACCGTAAAGACAACGGTGATGGTTACGGTTTCTCTGCAACT
TATGACGTAGGTTATGGTATTAGCGTCGGTGGTGCATACTCTAACTCTGCCCGTACTTCTGCGCAAAAAA
CTGCTTTTAGCGTTGTTGAGGAAAAAGGGAAAAAAGACAAACCTACTGGCTCTACTGCCTTGGGTGATCG
CGCTGAATCATGGAACATTGGCGCCAAATACGATGCCAACAACATCTACCTGGCTGCCATGTACGGTGAA
ACCCGTAATATGACTCGTTTCGGTGGAAATGTTGACGGTGTAAACTGGGAAAATATCGCTAACAAAACCC
AGAACATCGAATTGACCGCCCAATATCAGTTCGACTTCGGTTTACGTCCATCTCTGGCATACGTTCAGTC
CAAAGGTAAAAATCTGCTCGGTGACGATGGGAAAGAATCCAGCAGCCGCGATCTCGTAAAATATATTTCT
GTTGGTAGTTACTACTACTTCAACAAAAACTTGACTACCTATGTTGATTACAAAATCAACCTGCTGGATA
GCAATGAATTCACCAAAGCAACGGGTACCAGCACTGACAACGTTGTTGGCGTAGGTATGATTTACCAATT
CTAA 3' 2088101
```

Figure 4.7 The region of *P. luminescens* DNA to be amplified by PCR in order to clone the *ompN* gene

The top strand of DNA is written in the 5′ → 3′ direction. The coordinates of the DNA are shown. The ribosome binding site is shown in red. The signal peptide is shown in blue font. The start codon, ATG is underlined. The stop codon, TAA is represented by green font and is underlined.

ompN_F1:   5′ **GTCAGT****AAGCTT**ATGAGGGTAATAATAATGATGAAA 3′

ompN_F2:   5′ **GTCAGT****GGATCC****TTA****GTGATGGTGATGGTGATG**GAATT
GGTAAATCATACCTACG 3′

Figure 4.8 PCR primers used for the amplification of the region of *P. luminescens* DNA shown in figure 4.7, encoding OmpN

Non-binding regions of DNA are shown in bold type. Restriction sites are in red font. Six additional bases were added at the 5′ of the primers to enable efficient binding of the restriction enzymes to their restriction sites. The start codon, ATG is underlined in ompN_F1 and the stop codon is in green font in ompN_F2. A 6 x histidine tag has been incorporated into the reverse primer before the stop codon and is shown in blue.

The complete strategy for the cloning of *ompN* with a C-terminal histidine tag is shown in figure 4.9.
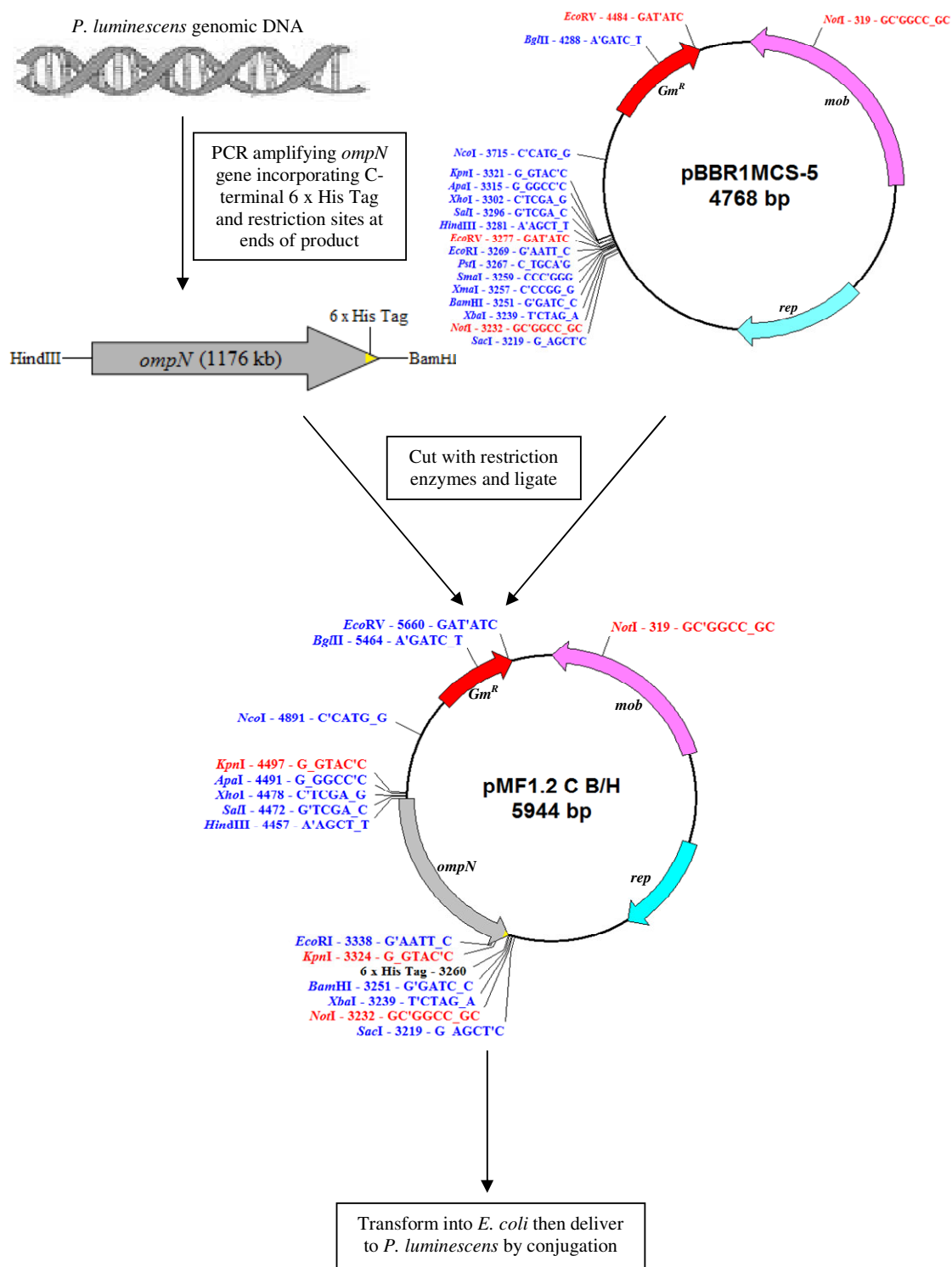
Figure 4.9 Strategy used for the cloning of *ompN* with a C-terminal histidine tag

pMF1.2 C B/H was constructed by restricting the *ompN* PCR product and the vector pBBR1MCS-5 with BamHI and HindIII restriction enzymes. The restricted pieces of DNA were then ligated together and transformed into *E. coli*. pMF1.2 C B/H was delivered from *E. coli* to *P. luminescens* by conjugation.

The *ompN* PCR product was purified as outlined in section 2.10. Plasmid DNA from pBBR1MCS-5 was prepared (section 2.8). Prepared pBBR1MCS-5 DNA and the purified *ompN* PCR product were restricted with BamHI and HindIII enzymes to generate cohesive ends and allow *ompN* to be ligated to pBBR1MCS-5. Prior to ligation the restriction digest products were purified by gel extraction (section 2.10) to remove the restriction reaction components. The ligation reaction was carried out as outlined in section 2.25.4 and the construct pMF1.2 C B/H was formed.

Chemically competent *E. coli* XL10-Gold cells were prepared (section 2.11.3) and transformed with the construct pMF1.2 C B/H (section 2.11.4) then grown on LB agar containing the selective antibiotic, gentamicin. Transformants were visible on the plates following overnight incubation. Plasmid DNA was prepared from five transformants using the 1, 2, 3 method (section 2.7) to confirm if the plasmids contained the *ompN* gene insert. The results are shown in figure 4.10.
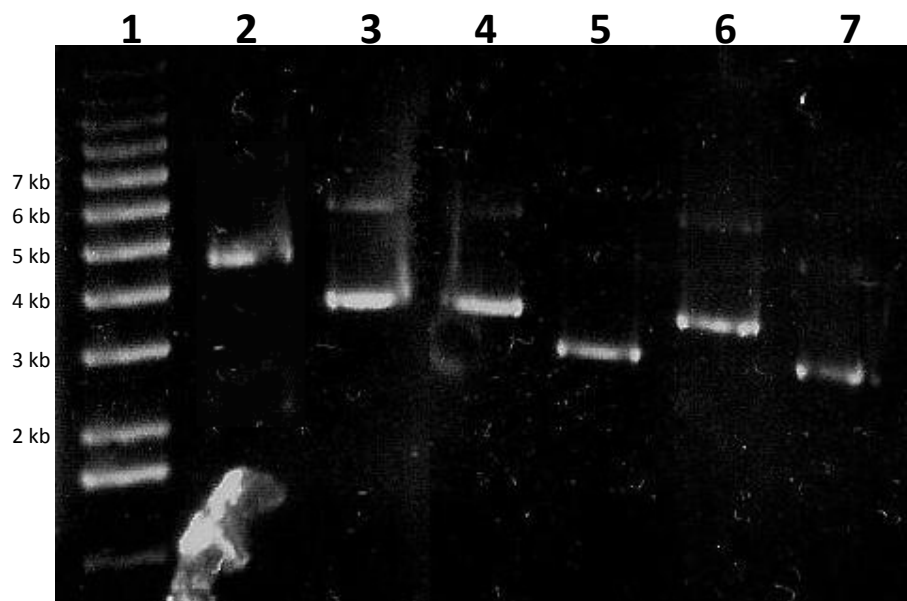
Figure 4.10 Plasmid DNA prepared from *E. coli* XL10-Gold cells transformed with pMF1.2C B/H
Plasmid DNA was prepared using the 1, 2, 3 method described in section 2.7. The DNA samples were separated on a 0.7 % agarose and stained with ethidium bromide. Lane 1 = Invitrogen 1 kb DNA ladder, lane 2 = pBBRMCS-5, lanes 3 – 7 = pMF1.2C B/H transformants 1-5.

In figure 4.10 two forms of plasmid DNA are visible in some lanes. The two forms are covalently closed circular (CCC) DNA and open circular (OC) DNA. The CCC form migrates further down the gel than the OC form as it is more compact. The CCC form is

also usually more concentrated that the OC form. Lane 2 contains pBBR1MCS-5 empty vector plasmid DNA. By comparing the sizes of the DNA bands obtained for the transformants to the empty vector it may be deduced that transformants 3 and 5 (lanes 5 and 7) do not contain the *ompN* gene, as their DNA is migrating to the same region of the gel as pBBR1MCS-5. Transformants 1, 2 and 4 do contain the insert (lanes 3, 4 and 6). Their plasmid DNA is larger than the empty vector.

Three of the five transformants chosen appeared to have the correct *ompN* insert. This was confirmed by restriction analysis. One of the transformants, number 1 (shown in lane 3), was chosen and digested with BamHI and HindIII in one reaction and with KpnI in a separate reaction. The results are shown in figure 4.11.
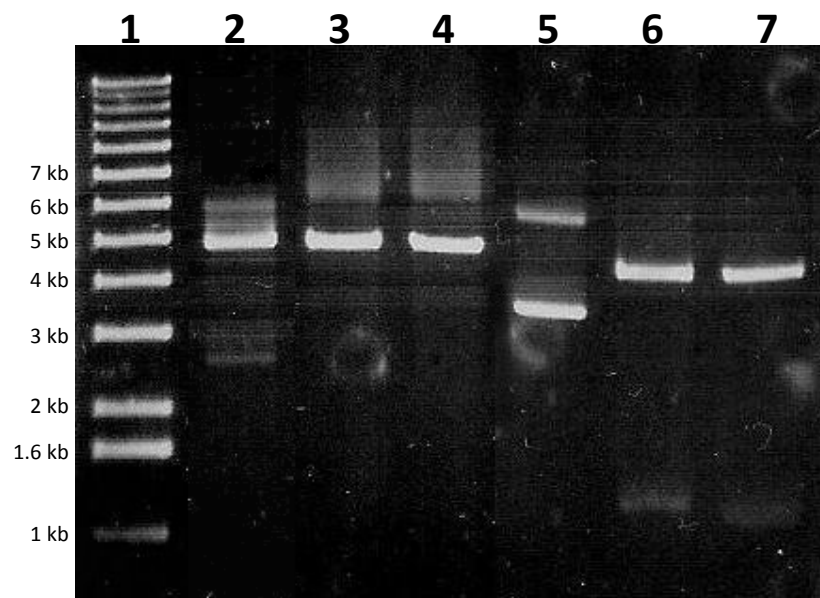
Figure 4.11 BamHI / HindIII and KpnI restriction digests of the plasmid pMF1.2C B/H to confirm that it contained the *ompN* insert

Restriction analysis was carried out and DNA fragments were separated on an agarose gel and stained with ethidium bromide. Lane 1 = Invitrogen 1 kb DNA ladder, lane 2 = pBBRMCS-5 uncut, lane 3 = pBBRMCS-5 cut with BamHI/Hind III, lane 4 = pBBRMCS-5 cut with KpnI, lane 5 = pMF1.2C B/H uncut, lane 6 = pMF1.2C B/H 5 cut with BamHI/Hind III, lane 7 = pMF1.2C B/H cut with KpnI.

Because the BamHI and HindIII restriction sites are only 30 bp apart in pBBR1MCS-5, cutting pBBR1MCS-5 with both of these enzymes yields only one visible linear fragment of 4.7 kb in size (lane 3). pBBR1MCS-5 contains one KpnI site so cutting with this enzyme linearises the DNA (lane 4). Lane 5 shows uncut pMF1.2 C B/H, with

the OC and CCC forms visible. Cutting with BamHI and HindIII removes the *ompN* insert yielding two linear fragments of ~4.7 kb and ~1.2 kb. pMF1.2 C B/H contains two KpnI sites, one of which lies within the *ompN* insert. Cutting with KpnI yields two fragments of ~4.7 and ~1.2 kb. This also shows that the insert is in the correct orientation. The restriction analysis shown in figure 4.11 confirms that the *ompN* insert is present in pMF1.2 C B/H.

In order to express OmpN in its natural environment, to ensure correct glycosylation, the construct was delivered to *P. luminescens* from *E. coli* by conjugation. Tri-parental mating was used for conjugation to occur. The method used for bacterial conjugation by tri-parental mating is described in section 2.13. The principles of how this method works is shown in figure 4.12.
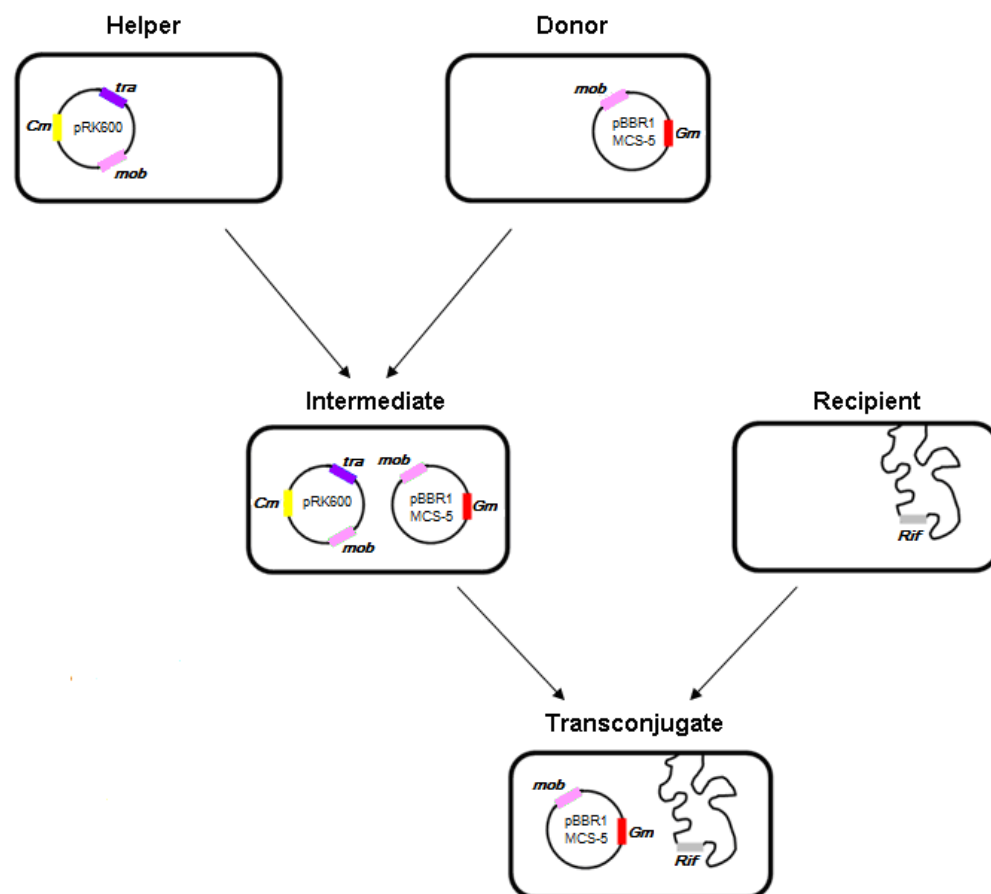
Figure 4.12 The principles of tri-parental mating

Transfer genes (*tra*) are highlighted in purple, mobilisation genes (*mob*) are highlighted in pink and antibiotic resistance genes are highlighted in yellow for chloramphenicol (*Cm*), red for gentamicin (*Gm*) and grey for rifampicin (*Rif*).

The helper strain contains the plasmid pRK600 that provides transfer functions (*tra* genes). The plasmid in the donor strain contains the gene of interest. The helper plasmid is transferred to the donor strain to create an intermediate strain that contains all the elements to transfer the plasmid of interest to the recipient strain, *P. luminescens* (*tra* and *mob*). The pBBR1MCS-5 construct is transferred to the recipient by conjugation to create a transconjugate. Correct intermediates and transconjugates may be selected for by culturing on media containing antibiotics, chloramphenicol and gentamicin for intermediates and gentamicin and rifampicin for transconjugates.

Both empty vector, pBBR1MCS-5 and the construct pMF1.2 C B/H were transferred separately to *P. luminescens* by conjugation. A number of transconjugates from each mating were chosen and plasmid preparations were carried out. Plasmid DNA isolated from *P. luminescens* transconjugates was run on an agarose gel and compared to plasmid DNA from *E. coli*. The results are shown in figure 4.13.
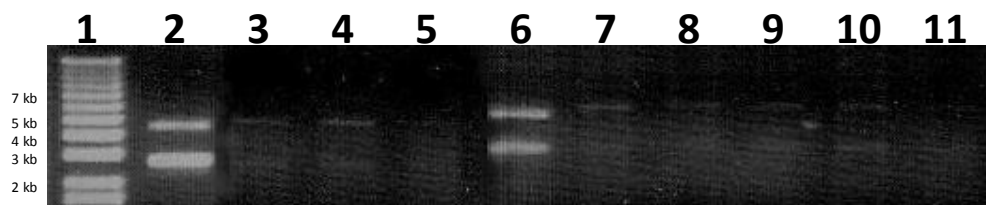
Figure 4.13 Plasmid DNA prepared from *P. luminescens* transconjugates
The pBBR1MCS-5 and pMF1.2C B/H plasmids were transferred from *E. coli* XL10-Gold to *P. luminescens* by tri-parental mating. Plasmid DNA was prepared from *P. luminescens* transconjugates and visualised on an ethidium bromide stained agarose gel. Lane 1 = Invitrogen 1 kb DNA ladder, lane 2 = pBBR1MCS-5 plasmid prep from *E. coli*, lanes 3 – 5 = pBBR1MCS-5 plasmid prep from *P. luminescens* (transconjugates), lane 6 = pMF1.2C B/H plasmid prep from *E. coli*, lanes 7 – 11 = pMF1.2C B/H plasmid prep from *P. luminescens* (transconjugates).

Plasmid DNA prepared from *P. luminescens* was much less concentrated than that from *E. coli*, however it can be seen that the plasmid is present in all transconjugates and is the correct size.

4.3.2 The expression of OmpN in *E. coli* and *P. luminescens*

Following the cloning of *ompN*, the gene was expressed in both *E. coli* XL10-Gold and *P. luminescens.* Because multiple copies of this plasmid exist in each bacterial cell, in theory a higher level of expression of *ompN* should be obtained in *P. luminescens* pMF1.2 C B/H than in the normal wildtype.

Protein expression was carried out as indicated in section 2.14. Following OmpN expression in *P. luminescens*, periplasmic proteins were isolated from the cells using the water lysis method (section 2.15). A wildtype *P. luminescens* culture was grown simultaneously and periplasmic proteins were also isolated from it as a comparison. The concentrations of the protein samples were measured using the BCA assay (section 2.17) and 20 µg of each protein sample was run on a 2-D gel as described in section 2.19. The gels were subsequently visualised by staining with silver stain (section 2.20). The images of both 2-D gels are shown in figure 4.14.
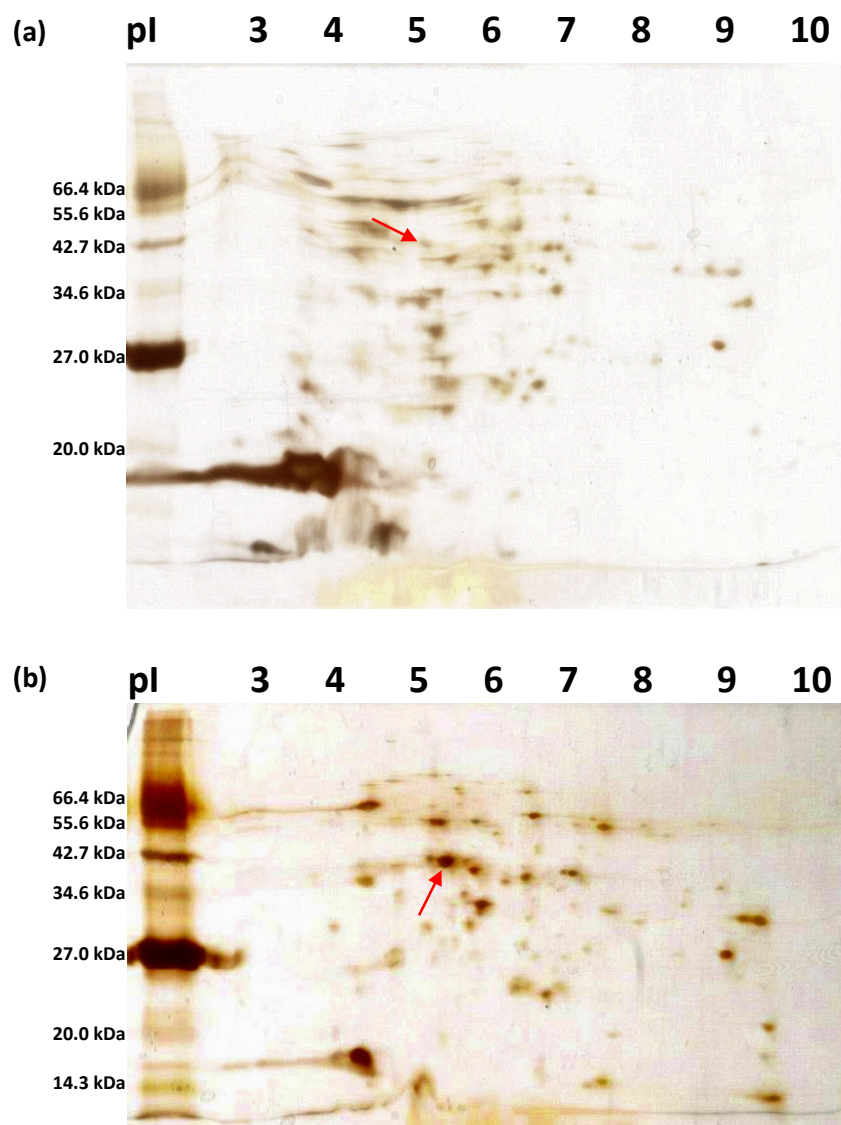
Figure 4.14 2-Dimensional gel electrophoresis comparing periplasmic proteins isolated from *P. luminescens* wildtype (a) and *P. luminescens* pMF1.2 C B/H (b)

Protein samples were separated using 2-D electrophoresis and stained with silver stain. The OmpN protein is indicated in both gels by a red arrow.

It can be seen from figure 4.14 that the level of expression of OmpN in *P. luminescens* pMF1.2 C B/H is much higher that the expression in the wildtype. OmpN is migrating at a molecular weight of ~40 kDa and a pI of ~5.3. Figure 4.14 indicates that recombinant OmpN is being produced by *P. luminescens* pMF1.2 C B/H. In order to carry out further analysis of the recombinant protein it was first purified from the mix of other proteins in the periplasm using its 6 x histidine affinity tag for purification using IMAC.

4.3.3 Purification of OmpN from *E. coli* and *P. luminescens* using IMAC

Recombinant OmpN was purified from periplasmic proteins prepared from *E. coli* XL10-Gold and *P. luminescens* pMF1.2 C B/H strains using the procedure described in section 2.16. A protein sample from before the purification (crude sample) and samples from throughout the purification were analysed on a 15% SDS-PAGE gel (section 2.18) and stained with silver stain (section 2.20). The results of the purification from *E. coli* are shown in figure 4.15 and from *P. luminescens* in figure 4.16.
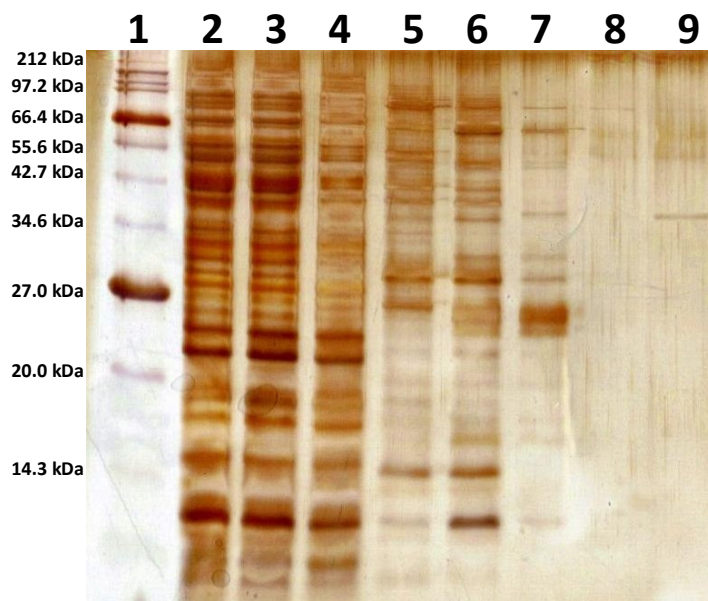
Figure 4.15 IMAC purification of OmpN from the periplasm of *E. coli* XL10-Gold pMF1.2C B/H
Periplasmic proteins were isolated from *E. coli* XL10-Gold pMF1.2C B/H using the water lysis method and were applied to an IMAC column. Proteins were eluted from the column with an increasing gradient of imidazole. Lane 1 = NEB broad range protein marker, lane 2 = crude *E. coli* XL10-Gold pMF1.2C B/H periplasmic proteins, lane 3 = flow through from IMAC resin, lane 4 = 10 mM imidazole wash, lane 5 = 20 mM wash, lane 6 = 40 mM wash, lane 7= 80 mM wash, lane 8 = 100 mM wash, lane 9 = 300 mM imidazole elution.

Figure 4.15 shows that a small amount of pure OmpN was isolated from the periplasm and can be seen in lane 9.
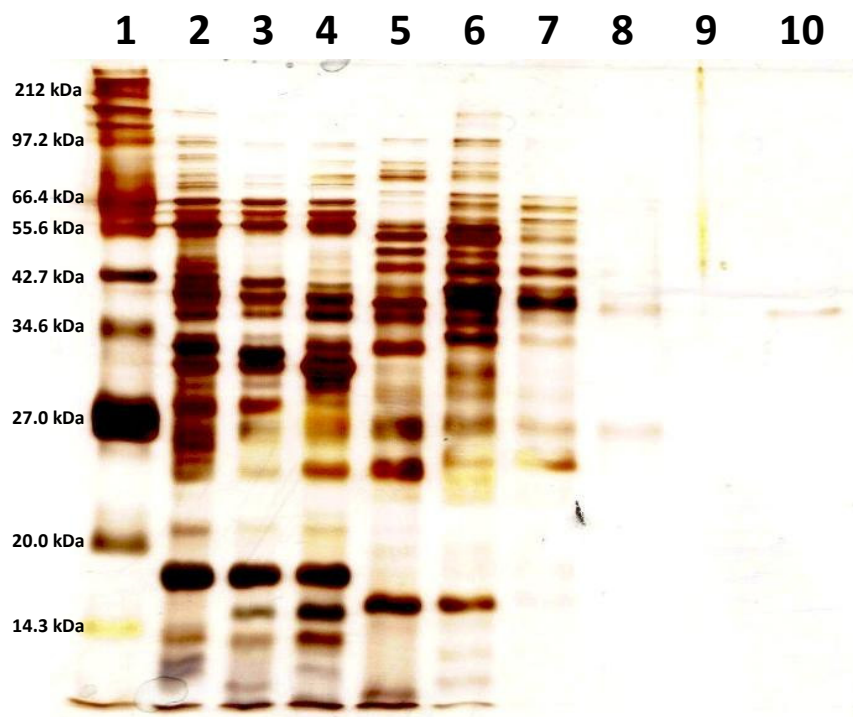
Figure 4.16 IMAC purification of OmpN from the periplasm of *P. luminescens* pMF1.2C B/H

Periplasmic proteins were isolated from *P. luminescens* pMF1.2C B/H using the water lysis method and were applied to an IMAC column. Proteins were eluted from the column with an increasing gradient of imidazole. Lane 1 = NEB broad range protein marker, lane 2 = crude *P. luminescens* pMF1.2C B/H periplasmic proteins, lane 3 = flow through from IMAC resin, lane 4 = 10 mM imidazole wash, lane 5 = 20 mM wash, lane 6 = 40 mM wash, lane 7= 60 mM wash, lane 8 = 80 mM wash, lane 9 = 100 mM wash, lane 10 = 300 mM imidazole elution.

The level of OmpN expression in both *E. coli* and *P. luminescens* is relatively low. This is because OmpN is a membrane protein and so forcing high levels of expression is usually not possible. However, more importantly, the samples are pure, which is vital for ELLA analysis where only small amounts of protein are needed for analysis.

4.3.4 Analysis of recombinant OmpN purified from *E. coli* and *P. luminescens*

After attaining a pure sample of OmpN from *P. luminescens* pMF1.2 C B/H, it was dried down in a vacuum drier to completely evaporate it and the sample was sent to the company Alphalyse for accurate relative molecular mass determination. It was hoped that the measured molecular mass would be higher than the calculated mass of the

protein backbone and so the size of the glycan(s) modifying the protein could be calculated. Unfortunately, the company were unable to determine the molecular weight because the spectrum obtained for the sample was too weak. This was probably due to insufficient sample amount. Due to limitations in funding another sample could not be sent for analysis and so lectins were employed to analyse the OmpN protein further.

OmpN purified from *P. luminescens* pMF1.2 C B/H was analysed using LAC. The lectin WGA was initially used to isolate OmpN from wildtype *P. luminescens* so this lectin was used again to see if the recombinant protein was still properly glycosylated and could still bind to WGA. The procedure used for LAC is described in section 2.24. The eluted protein sample was concentrated down in a vivaspin 15 centrifugal filter device. Purified OmpN, before application to the column, unbound protein from the WGA column and concentrated eluted protein were run on a 15% SDS-PAGE gel stained with silver stain. The results are shown in figure 4.17.
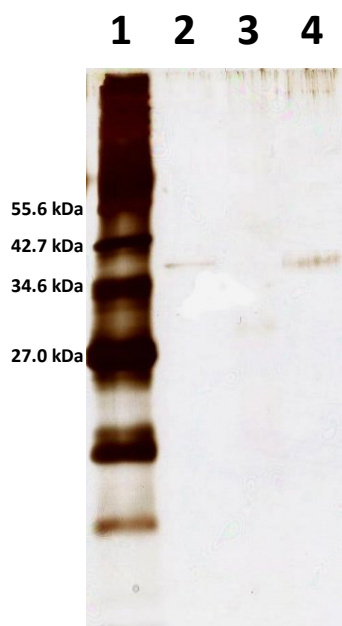
Figure 4.17 Investigation into whether OmpN purified from *P. luminescens* pMF1.2C B/H is capable of binding WGA in a lectin affinity chromatography column

OmpN purified from *P. luminescens* pMF1.2C B/H using IMAC was applied to a WGA LAC column. Unbound proteins were washed away and bound proteins were eluted using 0.5 M GlcNAc. Samples were run on a 15% SDS-PAGE gel and stained with silver stain. Lane 1 = Broad range protein marker, lane 2 = OmpN Purified from *P. luminescens* pMF1.2C B/H before addition to the column, lane 3 = unbound OmpN, lane 4 = OmpN eluted from WGA column.

Figure 4.17 shows that recombinant OmpN from *P. luminescens* still has the ability to bind to the WGA lectin and has been eluted from the column with the addition of 0.5 M GlcNAc (lane 4). The recombinant OmpN is binding quite strongly since little or no protein was visible in the unbound fraction (lane 3). In order to confirm that the interactions between WGA and OmpN were protein-glycan based and not protein-protein, OmpN purified from *E. coli* was used as a control. *E. coli* does not contain the same glycosylation machinery as *P. luminescens* and so OmpN from *E. coli* would not be glycosylated. An ELLA was carried out to investigate the interactions been OmpN (from both sources) and the lectins WGA and GSL I. Crude periplasmic proteins before IMAC purification and proteins from the first wash using 10 mM imidazole were also analysed. The results are shown in figure 4.18.
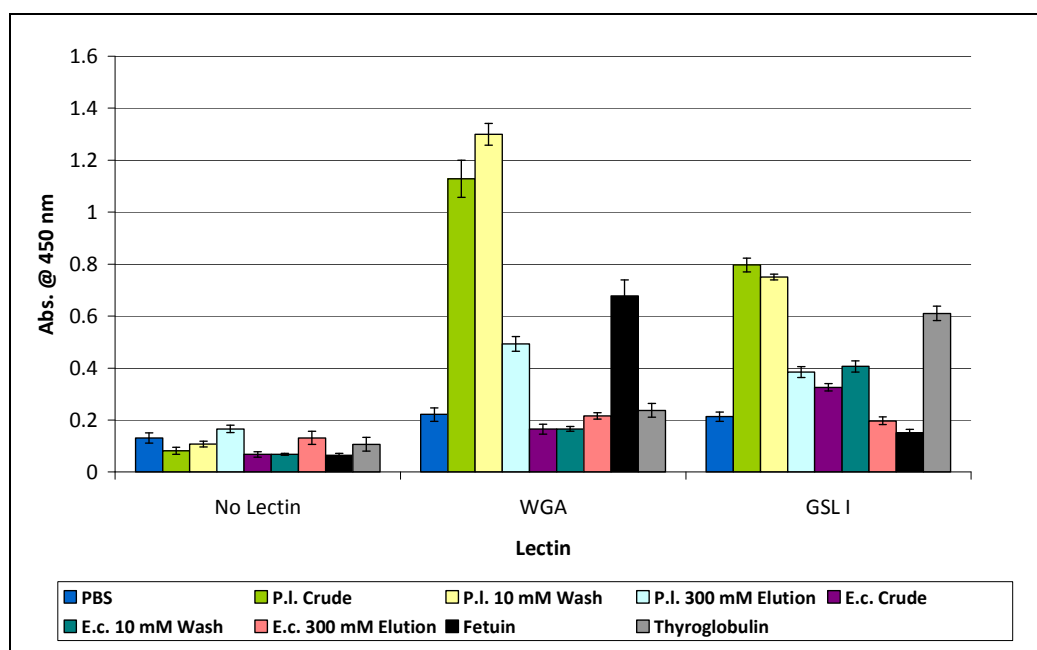
Figure 4.18 ELLA comparing interactions of OmpN purified from *P. luminescens* and *E. coli* with the lectins WGA and GSL I

P. l. = *P. luminescens* pMF1.2 C B/H, E. c. = *E. coli* pMF1.2 C B/H, Crude = Periplasmic proteins before IMAC purification, 300 mM elution = purified OmpN. Fetuin and Thyroglobulin are commercial glycoproteins that were used as assay control. Fetuin (black bars) is known to bind WGA and Thyroglobulin (grey bars) is known to bind GSL I.

It can be seen from figure 4.18 that proteins from the *P. luminescens* crude periplasmic sample (green) and 10 mM wash sample (yellow) bound very strongly to both lectins. No proteins from *E. coli* crude periplasmic sample (purple) and 10 mM wash sample

(dark green) bound to WGA but a signal was seen for both of these samples with the lectin GSL I. OmpN purified from *P. luminescens* (light blue) produced a significant signal with both lectins. However, in contrast the signal obtained for OmpN purified from *E. coli* (pink) was at baseline levels for both lectins. These results strongly indicate that OmpN from *P. luminescens* is glycosylated as it binds to the lectins WGA and GSL I, while the exact same protein produced in *E. coli* without glycosylation does not. The interaction between OmpN and the lectins is therefore concluded to be a glycan-protein binding and not a protein-protein interaction.

Chapter 5

Mutagenesis of *wbl* genes in *P. luminescens* and analysis
of altered glycosylation patterns

5.1 *In silico* analysis of the *P. luminescens* genome in search of glycosylation genes

The genome of *P. luminescens* subsp. laumondii TTO1 has been sequenced (Duchaud et al. 2003) and is accessible online. The *P. luminescens* genome was examined to identify encoded proteins with homology to the general protein glycosylation system in *C. jejuni* subsp. jejuni 81-176 using the BlastP program (NCBI). Figure 5.1 shows the genes encoding the general protein glycosylation system in *C. jejuni*. This cluster of genes is called the *pgl* gene cluster.
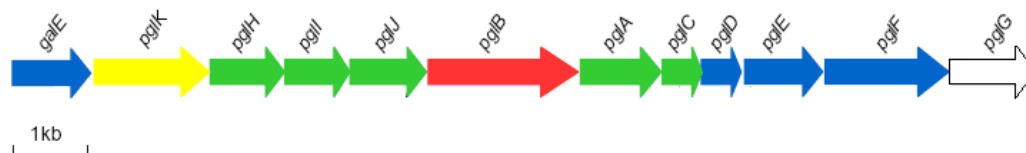
Figure 5.1 The *pgl* gene cluster in *C. jejuni*

Schematic representation of the *pgl* gene cluster that produces *N*-linked glycoproteins. ORFs are indicated by arrows. Glycosyltransferases are shown in green, enzymes involved in sugar biosynthesis are shown in blue, the oligosaccharyltransferase, *pglB* is shown in red and the predicted ABC transporter, *pglK* is shown in yellow. The function of *pglG* is unknown (Linton et al. 2005).

Almost all of the proteins encoded in this cluster were found to have orthologues in *P. luminescens*. Some of these orthologues were very similar to the *C. jejuni* proteins over the entire sequence, others had only part homology to the *C. jejuni* proteins. Table 5.1 summarises the results of the homology searches carried out between *C. jejuni* Pgl proteins and *P. luminescens* proteins.

Table 5.1 Proteins encoded by genes in the *pgl* gene cluster in *C. jejuni* and their orthologues in *P. luminescens*

C jejuni protein	Locus	Identification	JCVI role category[1]	Protein length	*P. luminescens* orthologue	Locus	Identification	JCVI role category	Protein length	Sequence similarity[2]
GalE	CJJ81176_1148	UDP-glucose 4-epimerase	1, 2	328 aa	GalE	plu4831	UDP-glucose 4-epimerase (galactowaldenase)	2	340 aa	38% (59%) (overall)
PglK	CJJ81176_1147	ABC transporter, permease/ATP-binding protein	1, 3	564 aa	plu3125	plu3125	unnamed protein product	4, 5	706 aa	43% (62%) (part)
PglH	CJJ81176_1146	General glycosylation pathway (GGP) protein	1	359 aa	plu3361	plu3361	unnamed protein product	1	404 aa	31% (51%) (part)
PglI	CJJ81176_1145	GGP protein	1	309 aa	plu4290	plu4290	unnamed protein product	1	329 aa	22% (45%) (overall)
PglJ	CJJ81176_1144	GGP protein	1	365 aa	WalW	plu4851	WalW protein	1	367 aa	27% (46%) (overall)
PglB	CJJ81176_1143	GGP protein	1	713 aa	no significant orthologue	N.A.	N.A.	N.A.	N.A.	N.A.
PglA	CJJ81176_1142	GGP protein	1	376 aa	WalR	plu4862	WalR protein	1	373	28% (51%) (part)
PglC	CJJ81176_1141	GGP protein	1	200 aa	WblF	plu4808	WblF protein	6, 7, 8	203 aa	34% (55%) (overall)

Table 5.1 Continued…

C jejuni protein	Locus	Identification	JCVI role category[1]	Protein length	_P. luminescens_ orthologue	Locus	Identification	JCVI role category	Protein length	Sequence similarity[2]
PglD	CJJ8117_1140	GGP protein	1	203 aa	LpxD	plu0682	UDP-3-_O_-[3-hydroxymyristoyl] glucosamine _N_-acyltransferase	1	342 aa	29% (49%) (part)
PglE	CJJ81176_1139	GGP protein	1	386 aa	WblK	plu4807	WblK protein	Not assigned	385 aa	27% (48%) (overall)
PglF	CJJ81176_1138	GGP protein	1	590 aa	WblM	plu4809	WblM protein	6, 7, 8	624 aa	36% (51%) (overall)
PglG	CJJ81176_1137	GGP protein	1	297 aa	no significant orthologue	N.A.	N.A.	N.A.	N.A.	N.A.

[1] **J. Craig Venter Institute (JCVI) Cellular Role Category:**

1 = Cell envelope: Biosynthesis and degradation of surface polysaccharides and lipopolysaccharides

2 = Energy metabolism: Sugars

3 = Transport and binding proteins: Other

4 = Transport and binding proteins: Unknown substrate

5 = Transport and binding proteins: Amino acids, peptides and amines

6 = Energy metabolism: Electron transport

7 = Regulatory functions: Other

8 = Protein synthesis: Ribosomal proteins: synthesis and modification

[2] **Sequence Similarity:** First value = % identities (exact match), second value = % positives (similar match)

It can be seen from the table that the orthologues found in *P. luminescens* are not all encoded adjacent to each other as the *pgl* genes in *C. jejuni* are. Apart from GalE, the three proteins in *P. luminescens* that showed the closest similarity overall to Pgl proteins in *C. jejuni* were WblK (PglE orthologue), WblF (PglC orthologue) and WblM (PglF orthologue). Section 1.3.2.3 described how PglC, PglE and PglF from *C. jejuni* are homologous to PglB, PglC and PglD from *N. meningitidis* respectively. Consequently PglB, PglC and PglD from *N. meningitidis* are homologous to WblF, WblK and WblM respectively from *P. luminescens*. PglB from *N. meningitidis* shows 54% similarity with WblF covering 42% of the protein sequence. PglC from *N. meningitidis* shows 58% similarity with WblK covering 69% of the protein sequence. Finally PglD from *N. meningitidis* shows 70% similarity with WblM covering almost all of the protein sequence (98%). The Pgl proteins from *N. meningitidis* are involved in the OT mediated *O*-linked glycosylation of pili and other proteins. In the case of the OT enzymes, PglB from *C. jejuni* and PglL from *N. meningitidis*, no significant orthologues have been found in *P. luminescens*.

The MultAlin and Genedoc programs (section 2.12) were used to perform global sequence alignments of the three Pgl proteins from *C. jejuni* that showed high similarity to Wbl proteins from *P. luminescens* and these alignments are shown in figures 5.2 to 5.4.

```
                  *        20         *        40         *        60         *
C.j._PglE : ---MRFFLSPHHGGNEIKYIEEVFKSNYIAPLGEFVNRFEQSVKDYSKSE-NALALNSATAALHLAIRVAGVK :  70
P.l._WblK : MNSSFLPFALPEISQEEIDEVIDSLKIGWITT-GPKTRRFEQDFAQYLGSDVEAIAVNSATAGLHLALEAIGIQ :  73
                    P 6   E6 6     K3 5I  G    RFEQ   Y  S    A6A6NSATA LHLAL   G6

             80         *        100        *        120        *        140
C.j._PglE : QDLIVIASSFTEIASVAPICYIKAKPVFIDC-DETYNIIVDLIKIAIKECEKKEKALILTHLYGNAAKMIEIVE : 143
P.l._WblK : PGLEVIVPTYTETATAEVVRYICAHPIFVISLPDSLNIIPIALEKAITA---KTKAIMPVHFAGLSCDMLAIIA : 144
                 D V6  35TF A3   6 YL A P6F6D   3 NID   6  AI    K KA66 H  G    MD I6

             *        160        *        180        *        200        *        220
C.j._PglE : ICKENEIVLIEDAAEALGSFYKNALGTIFG-EFGAYSYNGNKIITTSGGGMLIGKNKEKIEKAR---------- : 206
P.l._WblK : IAKKFNLRIIEDAAHSFPILYKGKKIGLLDTDVTVFSEYANKTMTIAEGGMLVSRHPDIIKRAKIMRLHGISKD : 218
               I K   6 6IEDAA   3 YK K 6GT     5S5  NK 6TT  GGML6 4    I 4A4

             *        240        *        260        *        280        *
C.j._PglE : -FYSTCARENCLHYEHIDYGYNYRLSNVLGAIGVAQMEVIEQRVLKKREIYEWYKEFIGEYFSFIDELENSRSN : 279
P.l._WblK : ALDRYQSKIPAWYYEVIEPGYKYNAPDICASIGIHQLQKIDTFHKKREANAKLYDDEIKDLPIILPT-KPQESN : 291
                F   Q 4     YE 6   GY Y 6 16   IG6 Q62 6   K4 6    Y  L     L     SN

             300        *        320        *        340        *        360        *
C.j._PglE : RWLSTAIIDFDKNELNACQKDVNISQKNIALHPKISKIIEDIKNEQIETRPLWKAMHIQEVFKGAKAYILNGNSE : 353
P.l._WblK : SQHAWHLYTIRILND------QVKISRDLFILKMAEKNIGCSVHFIPIHKQPIWKNTYSLNENNFPVAEEN---- : 355
                  L    N         V IS  1  6    6  6    6  P6WK   3       A N

             380        *        400
C.j._PglE : LFFCKGICLPSGTAMSKDIVCEISKLILKSIKA : 386
P.l._WblK : --FKRIVSIELYIKMTKEIQIYVIKTIKEILA- : 385
                 F 4 6 6P  T M3K D   6 K I    6
```

Figure 5.2 Sequence alignment of *C. jejuni* PglE with *P. luminescens* WblK

Black indicates conserved amino acids

```
                   *        20         *         40         *         60         *
C.j._PglC  : --MYEKVFKRIFDFILALVILVLFSPVILITAILLKI-TQGSVIFTQNRFGLDEKIFKIYKFKTMSDERDEKG :  70
P.l._WblF  : MMTYYLAIKRLFDFTSSLLGIIILLPPILITIAIWIKIDSKGPIFFRCKRVGQYNQDFYIHKFRSMTTGSEQKG :  73
              Y     KR6FDF   L6 66L   P666   A6 6KI 3 G 6 F   Q R G     F I KF43M3    2KG

                   80        *        100         *        120         *        140
C.j._PglC  : EL-ISDELRIKAFGKIVRSLSLDEILQLFNVLKGDMSFVGPRPLIVEYISLYNEEQKLRH-KVRPGIIGWAQV : 141
P.l._WblF  : QLTIGHDLRITHSGKFIRKYKIDEIAQLIDVVRGKMSLVGPRPEVPQFMDKYPDDVRNKILSVKPGIIDIASI : 146
              2L 6   LR6   GK 6R    6DEL QL 1V64G MS VGPRP 6 256  Y     4 4   V4PGIT  A 6

                   *        160         *        180         *        200
C.j._PglC  : NGRNAISWQKKFELDVYYVKNISFIIDLKIMFLTAIKVLKRSGVSKEGHVLTEKFNGKN : 200
P.l._WblF  : EMILENQILSQYCDPHQAYIDIIMFIKAKY-YIEYIDKKSFFYDLKIIFKIIYKVVFR- : 203
              1     52      1I   6 K  5L 6          K   T  K   4
```

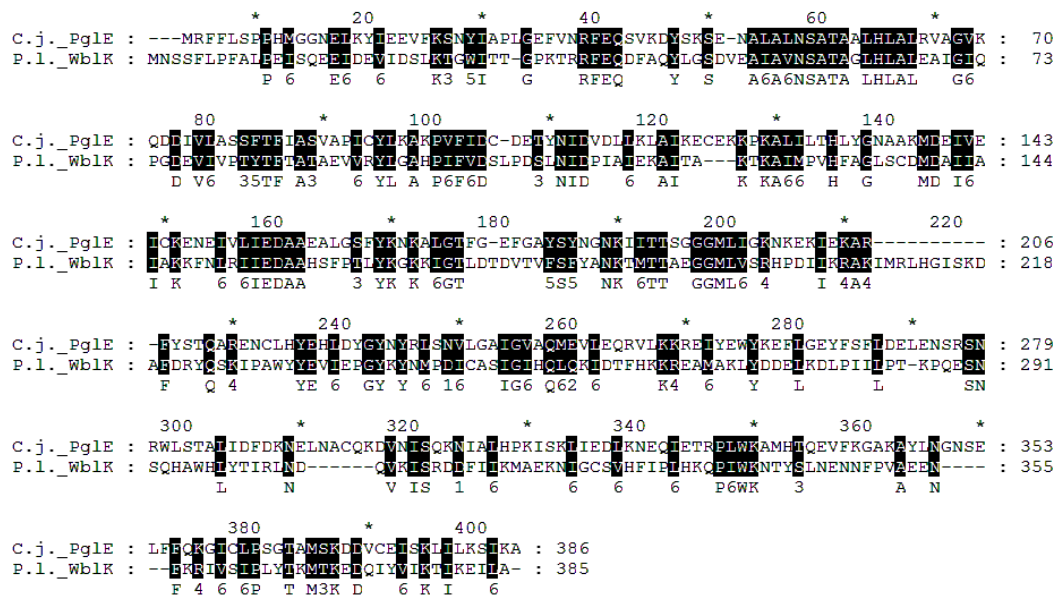**Figure 5.3 Sequence alignment of *C. jejuni* PglC with *P. luminescens* WblF**

Black indicates conserved amino acids

```
                   *        20         *         40         *         60         *
C.j._PglF  : ---MIFYKSKRIA---FFLISDIVLILLSVYIAFSLRFSGDIPSIFYHGMMVSAIILIVLKISFLFVFRIY :  65
P.l._WblM  : MNKRIFFNQDRLIKTIVLLIVDIFIIISYWISMWLRLDRBVP-IYSLQHWLIILFTIPFTLFIFLRIGFY :  70
              IF5   RL     LI DI 6I66S 5L   LR    6P I5    63 I  6   L      Y

                   80        *        100         *        120         *        140
C.j._PglF  : KVAWRFFSINEARKIFIALLIAEFCFFLIFYFFSDFFNPFPRSAIVILFVLSYMFIGTIFISKFMIVLFKP : 136
P.l._WblM  : RSILKYVNMSILKWAILGSFISTLLLTALSLYQQAFL---PRTVPIIYESFLVILSLCGTRFIYRTLRNL-- : 136
              4  45 6   4   I  L      I 5   F   PR3 6I F   6 6    R   R L 1

                   *        160         *        180         *        200         *
C.j._PglF  : SRMKEEETFCIVVGATSKALHLLKGAKEGSLGLFPVGVVLARKELIGTYCDKFIVEEKEKIKSYAEQ-GVK : 206
P.l._WblM  : --LKYKGAFVIIYGAGESGRQLIPILREHR-EFNPIAFVIDNAKIHNLSIHGVYVFSSEKITILVFKYNIK : 204
              6K    P I6 GA     LL   4E     P6 VD   L       V   EKI   E    6K

              220        *        240         *        260         *        280
C.j._PglF  : TAIIAIRLEQEELKKLFBELVAYGICDV-KIFSETR--------NEARDISIFDLIARKPKCLDDSAVAAF : 268
P.l._WblM  : KILLAIPSASISNRKKILERIQKEHCFILTIPSENDLVEGKAQINSLKRVSINDLIGREQVLPLQSILSKN : 275
              66A6       4K    E 6   C 6  I SF       N   4 6SI DLL R   D   S 6

              *        300         *        320         *        340
C.j._PglF  : LKDKVVLVSGAGGTIGSELCKCCIKFGAKHLIMVDHSEYNLYKINDDINLYKEK------ITPIILSILLK : 333
P.l._WblM  : IKNKNVLITGAGGSIGSELCRQIITQSPSQIVIFELTEYCLYSIEKBIQKINSDRDLNITITAIIGDIKIS : 346
              6K1K VL63GAGG3IGSELC4Q I          666   3EY LY I   L         IT IL  I D

              360        *        380         *        400         *        420
C.j._PglF  : CSIDBVLKTYKPELILHAAAYKHVPICFCNPHSAVINNILGTKILCDSAKENKVAKFVMISTDKAVRPTNI : 404
P.l._WblM  : EKIARLINKENINTIYHAAAYKHVPLVEMNTIEGINNNIFGTLSIAQAAINQKVEKFVLISTDKAVRPTNT : 417
              2 6  66   5     I HAAAYKHVPL E N   6 NNI GT  L   A    KV KFV6ISTDKAVRPTN

              *        440         *        460         *        480         *
C.j._PglF  : MGCTKRVCEIYTLSMSDENF--EVACVRFGNVLGSSGSVIEKEKAQIANNEPLTLTHPDIVRYFNIVAEAV : 473
P.l._WblM  : MGATKRFAEIILQAFAKFHSYTKFSMVRFGNVLGSSGSVVFIFEKQIKSGGPITLTHKDITRYFMTIPEAA : 488
              MG TKR   EL          E       VRFGNVLGSSGSV6P F   QI    P6TLTH DI RYFM 6 EA

              500         *        520         *        540         *        560
C.j._PglF  : QLVLQAGAIAKGGELFVLDMGKFVKIIDLAKKMIILSN---------RNDLEIKITGLRKGEKLYEELLID : 535
P.l._WblM  : QLVIQAGAIGNNGDVFVLDMGESVKIYDLARKMIHLSGLTVKDETNPHGDIEIKITGLRBGEKLYEELLIG : 559
              QLV6QAGA6   G 6FVLDMG  VKI  DLA4KM6 LS          D6EIKITGLR GEKLYEELLI

              *        580         *        600         *        620
C.j._PglF  : ENDAKIQYESIFVAKNEKVDINWLNKEIENIQI-CEDIS----EALIKIVFEFKHNKEGV----- : 590
P.l._WblM  : DNVSGTNHPRIMTANEVFITWIKINPLINDIKTSCENYDFENIRQIIINAFIDFFPKDYIYDLLK : 624
              N    T    I A      6 1 LN  6 1L   CE1       6L   P   H K  6
```

Figure 5.4 Sequence alignment of *C. jejuni* PglF with *P. luminescens* WblM

Black indicates conserved amino acids

The three proteins are very similar to their orthologues in sequence and size. However, the three Wbl proteins have not been assigned the same cellular role by the JCVI as their orthologues. According to the JCVI, PglC, E and F are involved in biosynthesis

and degradation of surface polysaccharides and LPS, while WblK has not been assigned to a role category and WblF and WblM are described as being involved in electron transport, the synthesis of ribosomal proteins and to have regulatory functions.

In the NCBI database WblF is listed as being similar to putative glycosyltransferases. It contains a similar conserved domain as proteins found in the Bac_transf (bacterial sugar transferase) superfamily. WblK is listed as being highly similar to putative hydro-lyases (enzymes that catalyze the removal of water from a substrate by breakage of a carbon-oxygen bond). WblK contains a conserved domain found in WecE proteins. These proteins are predicted to be pyridoxal phosphate-dependent enzymes involved in the regulation of cell wall biogenesis. WblK also contains a conserved domain from the AHBA_syn (3-amino-5-hydrozybenzoic acid synthase) family. The members of this family are involved in various biosynthetic pathways for secondary metabolites. Some members of this family are involved in the synthesis of an amine containing sugar called perosamine. Perosamine is an important element in the glycosylation of several cell products such as LPS. Another well studied protein in this family is PglA, a galactosyltransferase involved in pilin glycosylation in *N. meningitidis*. In the NCBI database WblM is listed as being similar to proteins involved in galactose modification. It contains conserved domains from proteins in two superfamilies, LbetaH and NADB-Rossmann. The LbetaH (left-handed parallel beta-helix) superfamily contains acetyltransferases and proteins involved in ion transport or translation initiation. The NADB-Rossmann (Rossmann-fold NAD(P)(+))-binding proteins are involved in many metabolic pathways. WblM also contains a conserved domain from nucleoside-diphosphate sugar epimerase proteins involved in cell envelope biogenesis and carbohydrate transport and metabolism. All three proteins contain conserved domains that suggest that they are involved in glycosylation in some way. They could be involved in LPS production or protein glycosylation or these enzymes could be common to both pathways.

Finally, the amino acid sequences of WblK, WblF and WblM were compared against the NCBI database of protein sequences using the BlastP program to find other orthologues to these proteins to gain a better insight into the possible functions of these proteins. The results of these BlastP searches are shown in figures 5.5 to 5.7. The three

closest orthologues to WblK, WblF and WblM were aligned using the MultAlin and Genedoc programs.
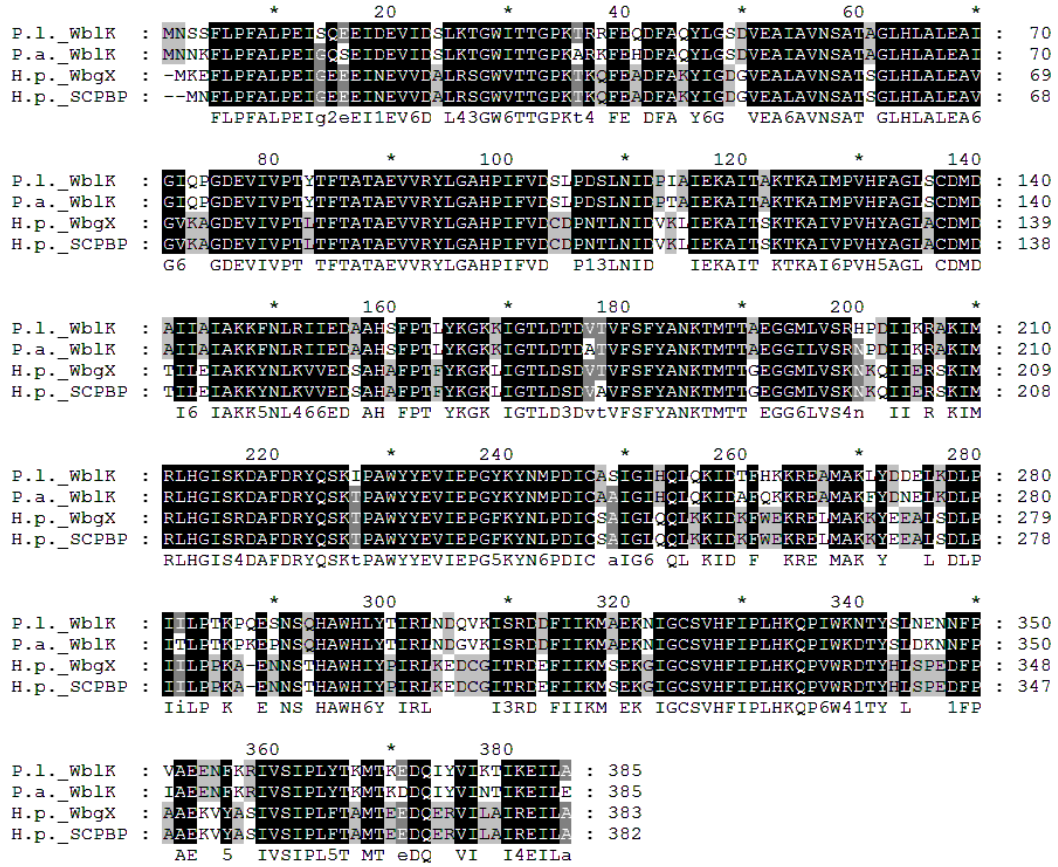

Figure 5.5 Multiple sequence alignments of WblK from *P. luminescens* with its three closest orthologues

P.l._WblK = WblK from *P. luminescens*, P.a._WblK= WblK from *P. asymbiotica*, H.p._WbgX = WbgX from *Haemophilus parasuis* SH0165 and H.p._SCPBP = spore coat polysaccharide biosynthesis protein from *H. parasuis* 29755. Sequence highlighted in black = 100% conservation, dark grey = 75% conservation and light grey = 50% conservation.

Figure 5.5 shows that WblK from *P. luminescens* has 92% (96%) homology with WblK from *P. asymbiotica*, 71% (87%) homology with WbgX and 71% (87%) homology with SCPBP. WblK from *P. asymbiotica* is reported as being an aminotransferase, WbgX is also an aminotransferase involved in LPS and spore coat polysaccharide biosynthesis and SCPBP is involved in spore coat polysaccharide biosynthesis.

```
                  *         20          *         40          *         60          *
P.l._WblF  : MMTYYLAIKRLFDFTSSLIGITLLFPILLTIAIWIKIDSKGPIFFRQKRVGCYNQDEYIHKFRSMTTGSEQK :  72
P.a._WblF  : MKTTYSIIKRLFDLISSSIGLITLSPLLIVIATWIKIDSKGPIFFRQKRVGFYNQNEYIHKFRSMTTGSEQK :  72
S.p._1414  : ------MIKRTFDFILSGVGLLGLSPLGLIVAIWIKVDSEGPIFFRQSRVGIFGRDEIIHKFRTMRVDSESS :  66
Sh.s._WbgY : -------MKRIFDVIVAGLGLLFLFPVFLIVSMLIVADSKGGVFFRCYRVGFEGKDERIHKFRTMFIDSEKK :  65
             6KR FD    s 6G66 L P6 6 6a wIk DSkGp6FFRQ RVG 5   1F IHKFR3M   SE k

                  *         80          *        100          *        120          *        140
P.l._WblF  : GGLTIGHDLRITHSGKFIRKYKIDELAQLIDVVRGKMSLVGPRPEVEQFMDKYPDDVRNKILSVKPGITDLA : 144
P.a._WblF  : GGLTIGHDLRITNSGKFIRKYKIDELAQLIDVIQGKMSLVGPRPEVEQFMDKYPDNVRNKILSVKPGITDLA : 144
S.p._1414  : GRLTVGKLSRITRSGHFLRKSKIDELPQLIDVILGTMSLVGPRPEVREFIDCYPDDVKEKVLSVRPGITDRA : 138
Sh.s._WbgY : GRITVGQIARVTRVGWYLRKYKIDELPQLIDVLSGTMSLVGPRPEVREFILEYPDDIREKVLSVRPGITDIA : 137
             G 6T6G D R6T sG 56RKyKIDEL QLIDV6 G MSLVGPRPEV 2F6D YPD164 K6LSV4PGITDlA

                  *        160          *        180          *        200
P.l._WblF  : SIEMIDENQILSCYQDPHQAYIDIIMPIKAKYYLEYIDKKSFFYDLKIIFKTIYKVVFR- : 203
P.a._WblF  : SIEMIDENQILSCYQDPHQAYIDIIMPIKAKYYLEYIDKKSFFYDLKIIFKTIYKVVFR- : 203
S.p._1414  : SIEMVDENEILSKYEDPRQAYIEHILPIKQNYYIEYVKNNLLLDVTLIFQTLILKIIGKR : 198
Sh.s._WbgY : SIEMVDENEILSSYDDPRRAYIDIILPIKQRYYLDYANNSVKYLCVIIWKTIIKIISR-  : 196
             SIEM6DEN2ILS Y DP qAYIdiI6PIK  YY6eY6   3  yD  6I5kT6 K66 4
```

Figure 5.6 Multiple sequence alignments of WblF from *P. luminescens* with its three closest orthologues

P.l._WblF = WblF from *P. luminescens*, P.a._WblF = WblF from *P. asymbiotica*, S.p._1414 = Spea_1444 from *Shewanella pealeana* ATCC 700345 and Sh.s._WbgY = WbgY from *Shigella sonnei* Ss046.

WblF from *P. luminescens* has 90% (95%) homology with WblF from *P. asymbiotica*, 60% (75%) homology with Spea_1444 and 59% (79%) homology with WbgY. All three orthologues are reported as being glycosyltransferases.

Figure 5.7 Multiple sequence alignments of WblM from *P. luminescens* with its three closest orthologues

P.l._WblM = WblM from *P. luminescens*, P.a._WblM= WblM from *P. asymbiotica*, Y.r._TrsG = TrsG like protein from *Yersinia rohdei* ATCC 43380 and Pl.s._WbgZ = WbgZ from *Plesiomonas shigelloides*.

WblM from *P. luminescens* has 93% (97%) homology with WblM from *P. asymbiotica*, 58% (77%) homology with TrsG and 57% (74%) homology with WbgZ. WblM from *P. asymbiotica* is reported as being a polysaccharide biosynthesis protein, TrsG and WbgZ are both reported as being nucleoside-diphosphate epimerases.

The results of the BlastP searches on WblK, WblF and WblM suggest that WblK is an aminotransferase, WblF is a glycosyltransferase and WblM is an epimerase. In order to confirm the function of the three proteins they can be mutated by insertional inactivation using a kanamycin resistance (Kan^R) cassette (as described in section 5.2). The mutated gene may be introduced into the *P. luminescens* genome by triparental mating and recombination mediated reciprocal exchange (as described in section 5.2). It is anticipated that if these genes are involved in protein glycosylation in *P. luminescens*, mutating them will lead to changes in the glycan patterns of the glycoproteins isolated from the organism. Figure 5.8 shows the location of the genes *wblK*, *wblF* and *wblM* in the genome of *P. luminescens*.
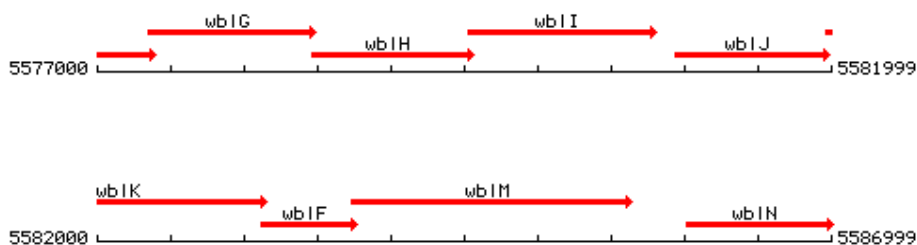
Figure 5.8 The location of *wblK*, *wblF* and *wblM* in the genome of *P. luminescens*
The coordinates of *wblK* are 5581963→5583120, the coordinates of *wblF* are 5583121→5583732 and the coordinates of *wblM* are 5583742→5585616. *wblK* is 1158 bp, *wblF* is 612 bp and *wblM* is 1875 bp in length. All three genes are found in the positive orientation in the genome.

The three genes *wblK*, *wblF* and *wblM* lie adjacent to each other in the genome. There is no gap between *wblK* and *wblF* and only ten base pairs separate *wblF* and *wblM*. This suggests that all three genes are controlled by the same promoter so mutation of *wblK* will likely lead to mutation of *wblF* and *wblM* also.

5.2 The principles of mutagenesis by insertional inactivation, triparental mating and recombination mediated reciprocal exchange

With the availability of so many sequenced microbial genomes and rapid advances in PCR it has become possible to target a specific gene in a microorganism for mutagenesis. Antibiotic resistance cassettes are widely used to mutate targeted genes by

insertional inactivation. Suicide vectors such as pJQ200sk+ (figure 2.2) (Quandt and Hynes 1993) have been developed to introduce the mutated gene into the organism.

In order to mutate the gene of interest using insertional inactivation, a 2 – 4 kb region of *P. luminescens* DNA carrying the target gene of interest is ligated to the pJQ200sk+ vector. An antibiotic resistance cassette is inserted into the recombinant target gene to cause the mutation. The orientation of the resistance cassette should be opposite to that of the target gene to ensure that transcriptional read-through cannot occur. For the mutagenesis of a *P. luminescens* gene, pJQ200sk+ vector constructs are prepared in *E. coli* (the donor strain) and delivered by conjugation into *P. luminescens* (the recipient strain) using tri-parental mating with the help of a plasmid such as pRK600 (Finan et al. 1986) in a helper strain. The principles of tri-parental mating are illustrated and discussed in figure 4.12.

The vector pJQ200sk+ is a narrow host range plasmid. Once the plasmid has been delivered to the recipient, *P. luminescens*, it cannot replicate in this host and so it secures and maintains itself by integrating into the chromosome by homologous recombination (figure 5.9).
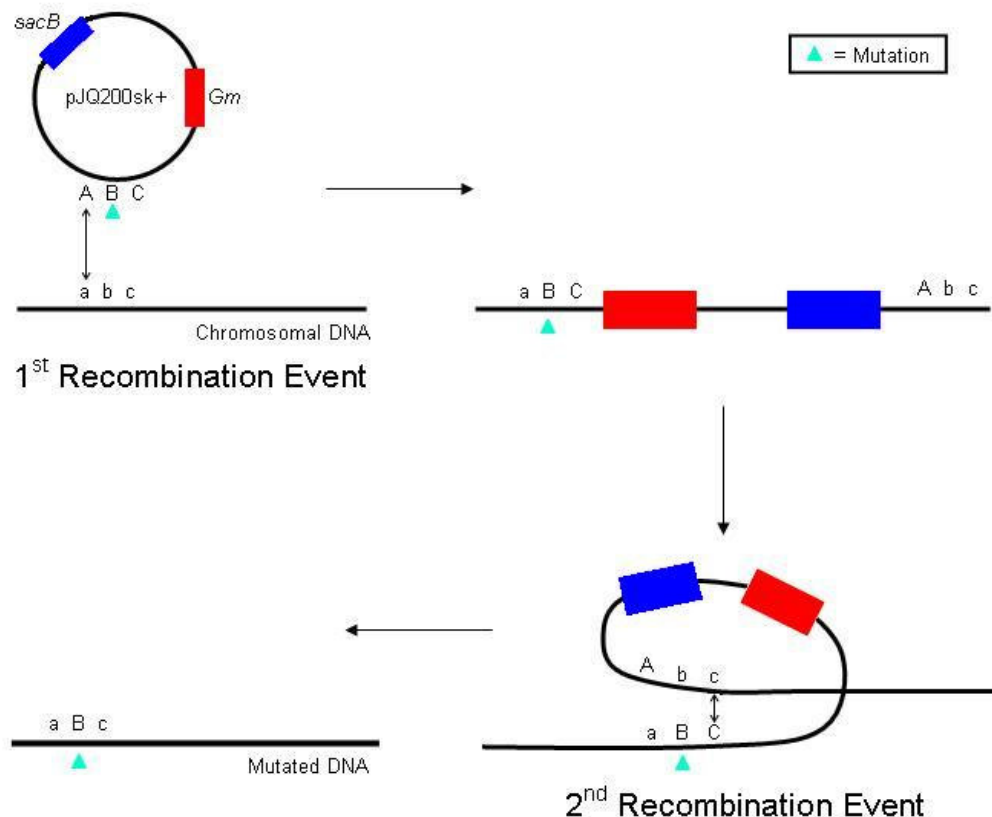
Figure 5.9 Homologous recombination

The circular suicide vector, pJQ200sk+ is shown, containing the gentamicin (*Gm*) resistance gene (red) and *sacB* gene (blue) and the recombinant DNA from *P. luminescens* (upper case letters) that has been mutated (light blue triangle). The straight black lines indicate a section of chromosomal DNA. The wild type gene is indicated by lower case letters. Double ended arrows indicate homologous regions of DNA. After the 2nd recombination event the mutation is incorporated into the chromosomal DNA.

About 1 – 2 kb of homologous sequence is required on either side of the insertion to facilitate successful recombination. The first recombination event integrates the entire pJQ200sk+ construct into the chromosomal DNA. The pJQ200sk+ vector encodes gentamicin resistance which allows for selection in *P. luminescens*. In order to excise the vector from the chromosomal DNA, leaving behind only the mutated gene a second recombination event must occur. Second recombination events occur randomly and at low frequency. pJQ200sk+ also encodes the *sacB* gene that results in a suicide effect in the presence of 5% sucrose. This facilitates the selection of second recombinant events at a detectable level. Any cells that have undergone the second recombination event will not contain the *sacB* gene and so will survive in the presence of 5% sucrose.

5.3 Antibiotic resistance cassette mutagenesis of *wblK*

wblK was selected for mutagenesis based on the results obtained from the *in silico* analysis in section 5.2. However, because *wblK*, *wblF* and *wblM* are adjacent to each other and appear to be part of the same operon, mutating *wblK* will likely prevent transcription of *wblF* and *wblM* also. Mutagenesis of *wblK* was achieved by insertional inactivation of the gene using a Kan^R cassette. Genomic DNA from *P. luminescens* was prepared (section 2.6) and used as a template to amplify two adjacent 2040 bp regions of the genome containing the *wblK* gene using PCR. Primers were designed to incorporate a BamHI site at the beginning and a PstI site at the end of the first region (wblK1) and the primers for the second region (wblK2) incorporated a PstI site at the beginning and an XhoI site at the end of the region. This is indicated by figure 5.10. The PstI site would be used later to insert the Kan^R cassette into the gene.
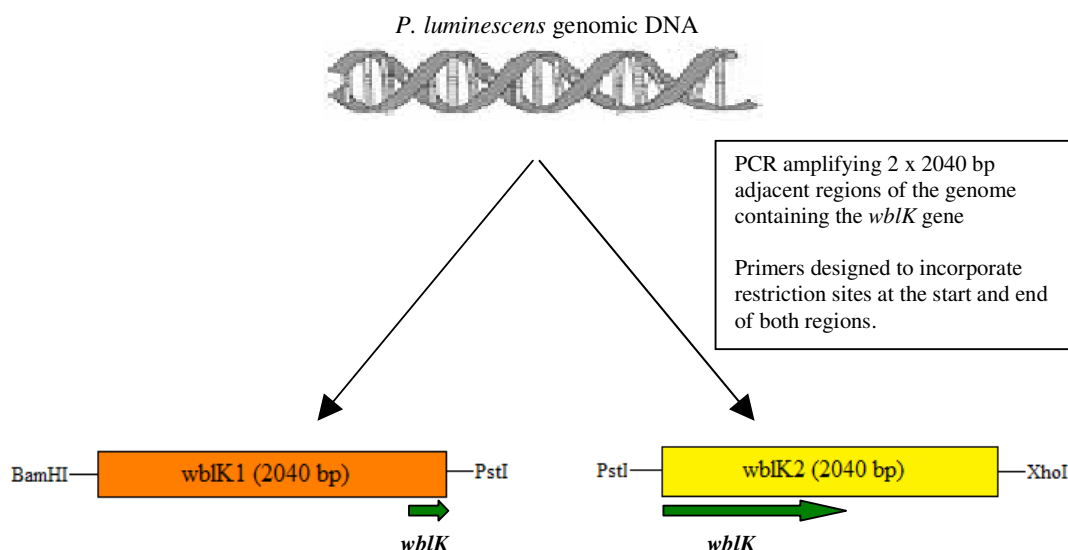
Figure 5.10 PCR amplifying two regions of *P. luminescens* DNA incorporating the *wblK* gene
The *wblK* gene is split over the two regions. The Kan^R cassette may be inserted at the PstI site in the opposite orientation to interrupt the gene.

The primers also contained nine additional bases at the beginning to enable efficient recognition of the restriction site by the restriction enzyme. The primers used for *wblK* mutagenesis are shown in figure 5.11. The PCR reaction components used are shown in section 2.25.2 and the PCR program cycle used is shown in section 2.25.3.

wblK_F1:   5′ **ACGTGACATGGATCC**TTCCATGCTTCCTAAAGTAC 3′

wblK_R1   5′ **ACGTGACATCTGCAG**ATATTGCGCAAAATCTTGTTC 3′

wblK_F2   5′ **ACGTGACATCTGCAG**TTAGGTTCCGATGTTGAAGC 3′

wblK_R2   5′ **ACGTGACATCTCGAG**TAGCGTTCTATAGATAAATCTTG 3′

Figure 5.11 PCR primers used for the amplification of the region of *P. luminescens* DNA encoding *wblK*

Non-binding regions of DNA are shown in bold type. Restriction sites are shown in red. wblK_F1 and wblK_R1 were used to amplify wblK1 and wblK_F2 and wblK_R2 were used to amplify wblK2.

The suicide vector pJQ200sk+ (figure 2.2) was employed to introduce the *wblK* fragment containing the Kan^R cassette into the genome of *P. luminescens*. Plasmid DNA from pJQ200sk+ was prepared (section 2.8). Prepared pJQ200sk+ DNA and the PCR product, wblK1, which had been purified (section 2.10), were restricted with BamHI and PstI to generate cohesive ends to allow wblK1 to be ligated to pJQ200sk+. pJQ200sk+ and wblK1 digests were purified by gel extraction (section 2.10) to remove the restriction reaction components prior to ligation. The ligation reaction is described in section 2.25.4. This created the first construct in the mutagenesis strategy called pMF2.0 B/P which is shown in figure 5.12.
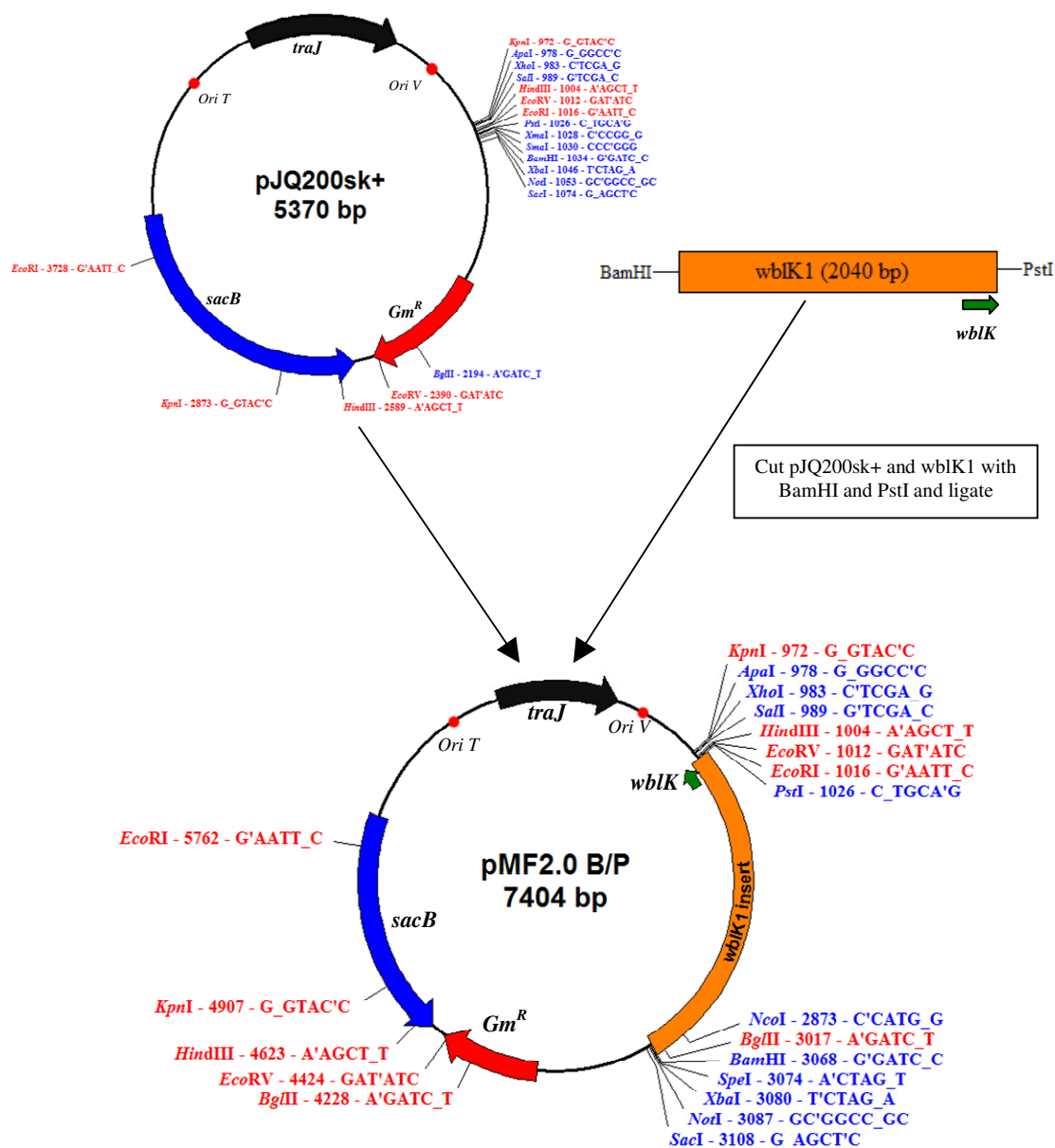
Figure 5.12 The construction of pMF2.0 B/P

pMF2.0 B/P was constructed by ligation of BamHI/PstI restricted wblK1 and pJQ200sk+. wblK1 contains the 5′ region of the *wblK* gene.

Electrocompetent *E. coli* XL1-Blue cells were prepared (section 2.11.1) and transformed using electroporation (section 2.11.2) with the clone pMF2.0 B/P then grown on agar containing gentamicin to select for transformants. Plasmid DNA from twelve transformants was prepared using the 1, 2, 3 method (section 2.7) to confirm that they contained the wblK1 insert. The results are shown in figure 5.13.

Figure 5.13 Plasmid DNA prepared from *E. coli* XL1-Blue cells transformed with pMF2.0 B/P
Plasmid DNA was prepared using the 1, 2, 3 method described in section 2.7. The DNA samples were separated on a 0.7% agarose gel and stained with ethidium bromide. Lane 1 = Invitrogen 1 kb ladder, lane 2 = pJQ200sk+ empty vector and lanes 3 – 14 = transformants 1 – 12.

In figure 5.13 CCC and OC plasmid DNA is visible in lanes 3-14. Lane 2 contains pJQ200sk+ plasmid DNA and only the CCC form is visible. By comparing the sizes of the DNA bands obtained for the transformants to the empty vector it may be deduced that transformants 1 (lane 3), 6 (lane 8) and 9 (lane 11) do not contain the wblK1 insert as their CCC DNA form is migrating to the same region of the gel as the empty vector in lane 1. Transformants 2-5, 7, 8 and 10-12 do contain the wblK1 insert. Their plasmid DNA is larger than the empty vector. Nine of the twelve transformants appeared to have the correct wblK1 insert. This was confirmed by restriction analysis. Transformant four was digested with BamHI and PstI. The results are shown in figure 5.14.
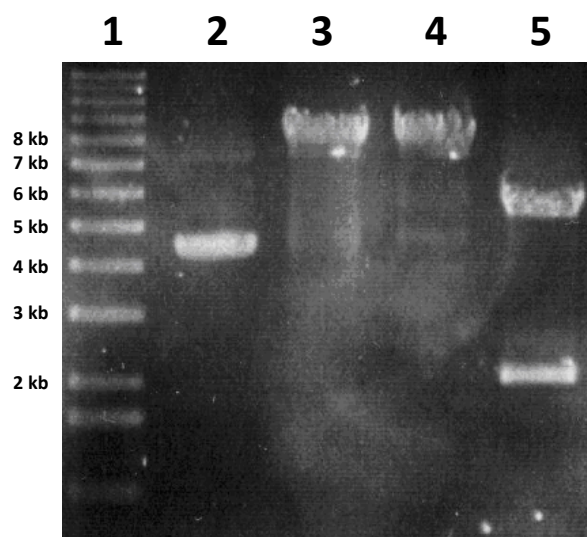
Figure 5.14 BamHI / PstI restriction digest of the plasmid pMF2.0 B/P to confirm that it contained the wblK1 insert

Restriction analysis was carried out and DNA fragments were separated on an agarose gel and stained with ethidium bromide. Lane 1 = Invitrogen 1 kb ladder, lane 2 = uncut pMF2.0 B/P, lane 3 = pMF2.0 B/P cut with BamHI only, lane 4 = pMF2.0 B/P cut with PstI only and lane 5 = pMF2.0 B/P cut with BamHI and PstI.

The CCC form of uncut pMF2.0 B/P is showing up strongly in lane 2. When cut with BamHI only (lane 3) and PstI only (lane 4) pMF2.0 B/P is linearised as there is only one restriction site in the plasmid for each of these enzymes. The linearised pMF2.0 B/P appears at the expected size of ~7.5 kb. When pMF2.0 B/P is cut with BamHI and PstI, two fragments are seen, one at 2 kb (the wblK1 insert) and one at ~5.5 kb. The restriction analysis shown in figure 5.14 confirms that the pMF2.0 B/P clone does contain wblK1. This was further confirmed by carrying out a PCR on pMF2.0 B/P using the primers wblK_F1 and wblK_R1. The PCR yielded a 2 kb fragment, the wblK1 insert (results not shown).

Because directional cloning was used it was expected that the insert wblK1 should have ligated to pJQ200sk+ in the desired orientation. In order to confirm this, restriction analysis using the enzyme BglII was carried out. It can be seen from figure 5.12 that pJQ200sk+ contains one BglII site and wblK1 also contains one BglII site. Restriction of pMF2.0 B/P with this enzyme should therefore cut the plasmid twice yielding two fragments of 1211 bp and 6193 bp if the insert is in the correct orientation. Figure 5.15 shows the result of the restriction analysis.

173
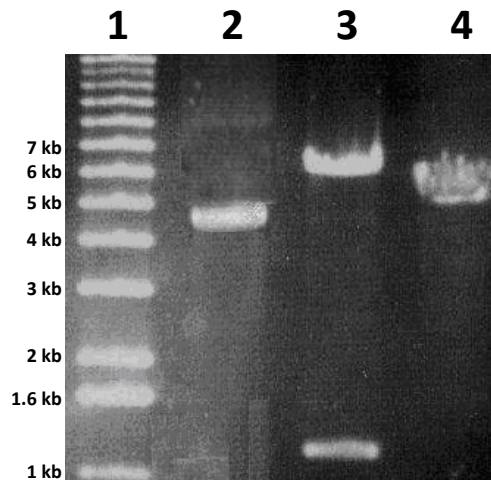
Figure 5.15 BglII restriction digest of pMF2.0 B/P to confirm orientation of wblK1

DNA fragments generated from the restriction analysis using BglII were visualised on an agarose gel that had been stained with ethidium bromide. Lane 1 = Invitrogen 1 kb DNA ladder, lane 2 = uncut pMF2.0 B/P, lane 3 = pMF2.0 B/P cut with BglII and lane 4 = pJQ200sk+ cut with BglII.

pMF2.0 B/P cut with BglII yielded two fragments of ~1.2 kb and 6.2 kb in size (lane 3). These were the sizes expected if the fragment was in the correct orientation. BglII cuts pJQ200sk+ only once as expected giving a linear fragment of 5.5 kb in size (lane 4).

The next step in the mutagenesis strategy was the construction of the second construct, pMF2.1 P/X. Plasmid DNA from pMF2.0 B/P and the purified PCR product wblK2, containing the 3′ region of the *wblK* gene were restricted with the enzymes PstI and XhoI. The digests were gel extracted as before and ligated together to create pMF2.1 P/X as shown in figure 5.16.
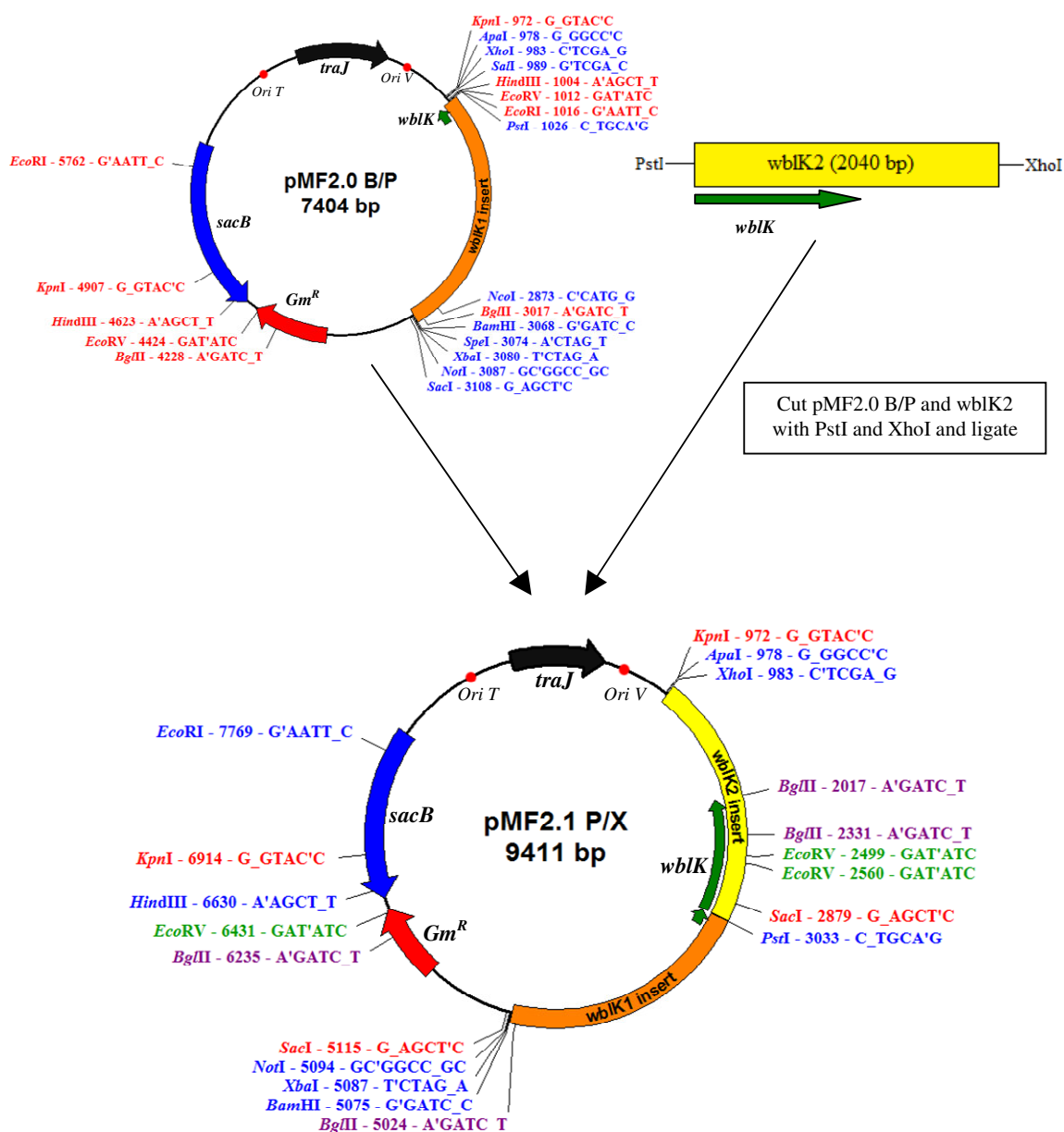
Figure 5.16 The construction of pMF2.1 P/X

pMF2.1 P/X was constructed by ligation of PstI/XhoI restricted wblK2 and pMF2.0 B/P. wblK2 contains the 3′ region of the *wblK* gene.

Electrocompetent *E. coli* XL1-Blue cells were transformed with the clone pMF2.1 P/X using electroporation and transformants were visible following overnight incubation on LB agar containing gentamicin. Plasmid DNA from eight transformants was prepared using the 1, 2, 3 method to confirm if they contained the wblK2 insert. The results are shown in figure 5.17.
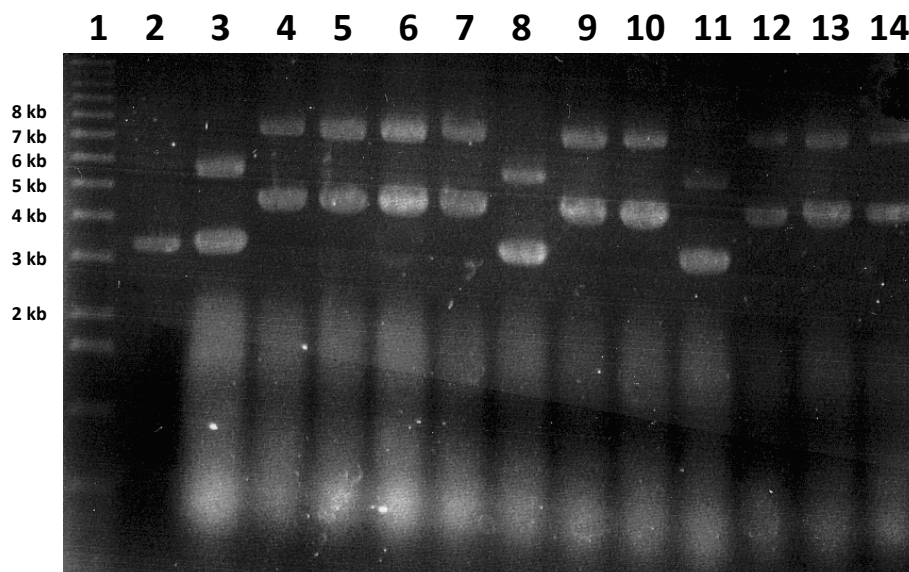
Figure 5.17 Plasmid DNA prepared from E. coli XL1-Blue cells transformed with pMF2.1 P/X
Plasmid DNA was prepared using the 1, 2, 3 method and separated on an agarose gel. DNA was visualised on a UV transilluminator following staining with ethidium bromide. Lane 1 = Invitrogen 1 kb DNA ladder, lanes 2 – 9 = Transformants 1 – 8, lane 10 = pMF2.0 B/P.

From figure 5.17 it can be deduced that transformants 5 (lane 6) and 8 (lane 10) do not contain the wblK2 insert as their CCC DNA form is migrating to the same region of the gel as pMF2.0 B/P in lane 10. Transformants 1 – 4, 6 and 7 do contain the wblK2 insert. Their DNA is larger than that obtained for pMF2.0 B/P.

One of the transformants was chosen for restriction analysis to confirm that wblK2 was present and that it was in the correct orientation. Again, directional cloning was used so it is expected that wblK2 should have ligated to pMF2.0 B/P in the desired direction. SacI and BglII restriction digests were carried out. pMF2.0 B/P was also digested as a control. The results of the restriction analysis are shown in figure 5.18.
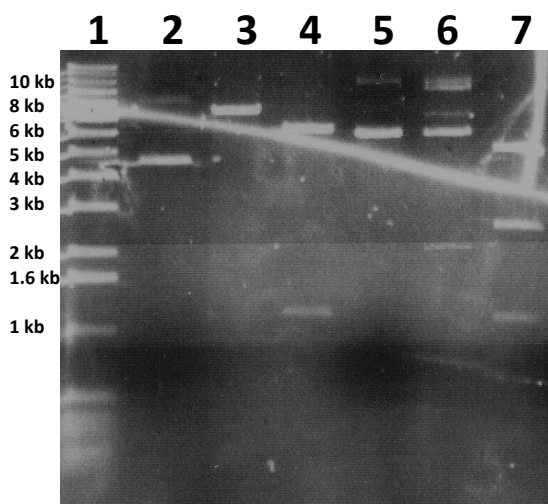
Figure 5.18 SacI and BglII restriction digests of the plasmid pMF2.1 P/X to confirm the presence and orientation of wblK2 in the plasmid

DNA fragments generated from the restriction analysis using SacI and BglII were separated on an agarose gel and visualised after staining with ethidium bromide. Lane 1 = Invitrogen 1 kb DNA ladder, lane 2 = uncut pMF2.0 B/P, lane 3 = pMF2.0 B/P cut with SacI, lane 4 = pMF2.0 B/P cut with BglII, lane 5 = uncut pMF2.1 P/X, lane 6 = pMF2.1 P/X cut with SacI (partial digestion), lane 7 = pMF2.1 P/X cut with BglII.

SacI cuts pMF2.0 B/P only once to yield a linear fragment of DNA with a size of 7.5 kb (lane 2). pMF2.0 B/P contains two BglII restriction sites and is cut twice to yield fragments of the expected sizes of 6.5 kb and 1.2 kb. The insertion of wblK2 introduces one extra SacI site and two extra BglII sites. In lane 6, four bands are visible. The bands migrating at 7.5 kb and 2 kb are the SacI digestion fragments. However incomplete digestion occurred and bands containing unrestricted OC and CCC pMF2.1 P/X DNA are also still visible. In lane 7 three fragments are visible. BglII cuts pMF2.1 P/X four times but one of these fragments is only 314 bp in size and so is not visible on this gel. The other three fragments are at 1.2 kb, 2.7 kb and 5.2 kb and these are the expected size.

This restriction analysis confirms that the wblK2 insert is present in pMF2.1 P/X and is in the correct orientation because restriction sites for SacI and BglII lie inside the insert and following digestion the correct sized fragments were obtained.

A final confirmation was carried out on pMF2.1 P/X. The primers that were used to initially amplify wblK2 were used in a PCR with pMF2.1 P/X DNA as the template (figure 5.19).
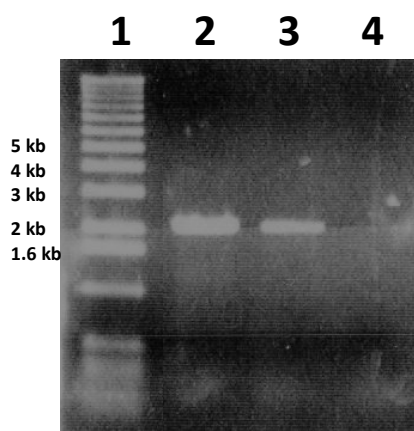
Figure 5.19 PCR products generated from the amplification of the *wblK* gene

A PCR reaction was carried out to confirm that the complete *wblK* gene was present in pMF2.1 P/X. PCR DNA was run on an agarose gel, stained with ethidium bromide and visualised. Lane 1 = Invitrogen 1 kb DNA ladder, lane 2 = PCR on pMF2.1 P/X, lane 3 = PCR on *P. luminescens* genomic DNA (positive control), lane 4 = PCR on pMF2.0 B/P (negative control).

The PCR produced a 2 kb DNA fragment for pMF2.1 P/X DNA (lane 2) and *P. luminescens* genomic DNA (lane 3), which was used as a positive control. No PCR product was formed using pMF2.0 B/P as template DNA (lane 4) because the wblK2 insert is not present.

The restriction analysis and PCR results confirmed conclusively that pMF2.1 P/X contains wblK2. This construct now contained the entire *wblK* gene. The next step in the mutagenesis process was to mutate this gene by insertional inactivation with a Kan[R] cassette. The Kan[R] cassette was isolated from the pUK4K vector (figure 2.3, Amersham Pharmacia) by restriction with PstI as shown in figure 5.20.
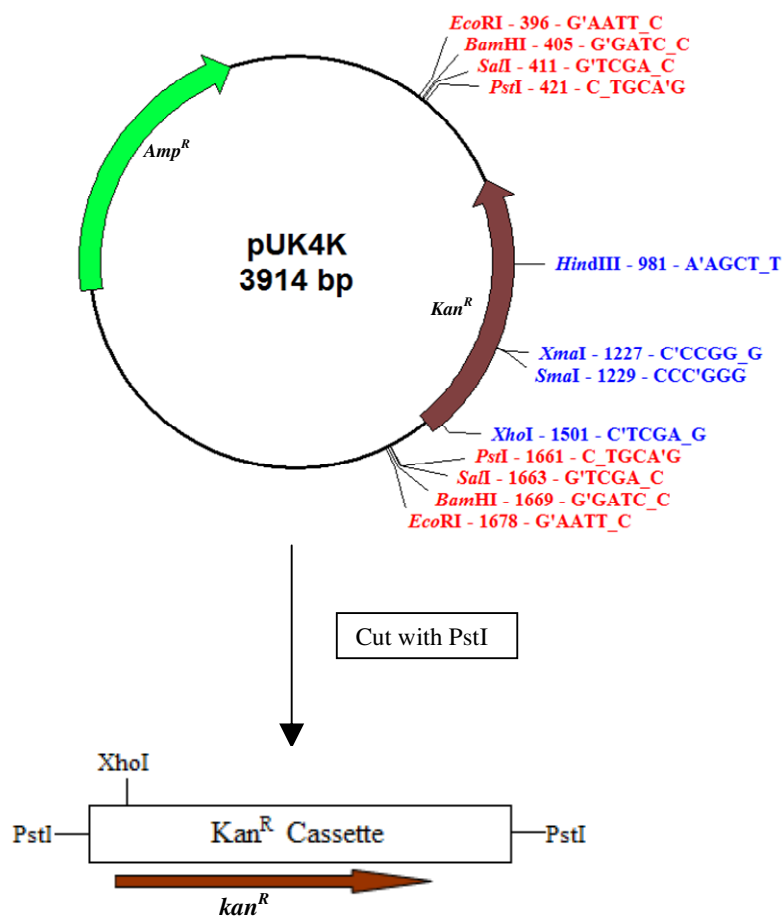
Figure 5.20 Schematic representation of how the kanamycin resistance cassette was isolated from the pUK4K plasmid

PstI was used to isolate the Kan^R cassette (1240 bp) which contained the *kan^R* gene (brown). The *kan^R* gene contains an XhoI site near its 5′ end.

pMF2.1 P/X was restricted with PstI. The pMF2.1 P/X digestion and the isolated Kan^R cassette were purified by gel extraction and ligated together to yield the final construct for the mutagenesis of *wblK*, pMF2.2 P, as shown in figure 5.21.
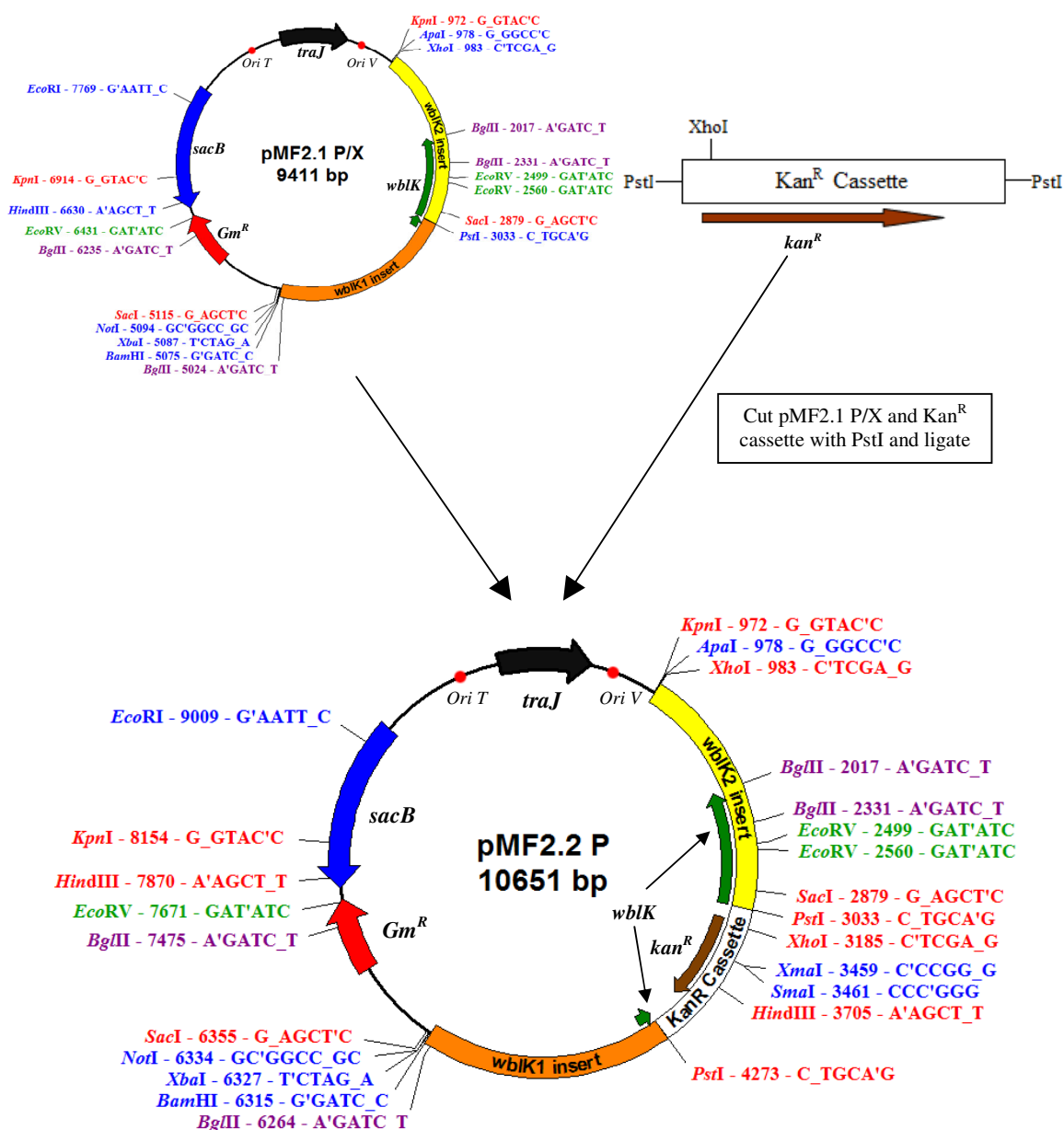
Figure 5.21 The construction of pMF2.2 P

pMF2.1 P/X was cut with PstI to generate cohesive ends so that the Kan[R] cassette could be ligated to it. The Kan[R] cassette interrupts the *wblK* gene (green). The *kan[R]* gene (brown) is in the opposite orientation to *wblK* to ensure that transcriptional read-through does not occur.

pMF2.2 P was transformed into chemically competent *E. coli* XL10-Gold cells (section 2.11.4) and transformants were selected as before and screened for the correct insert. Transformants are shown in figure 5.22.
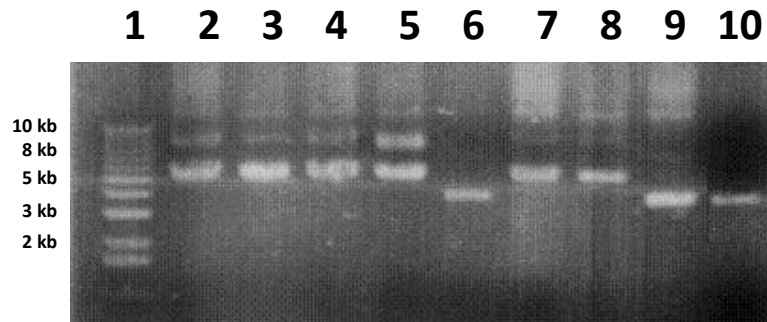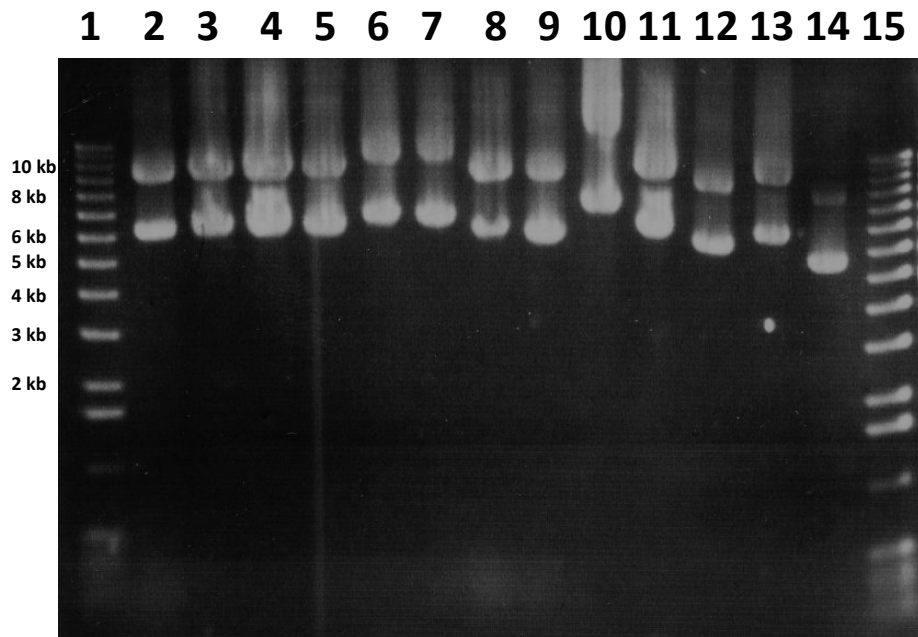
Figure 5.22 Plasmid DNA prepared from *E. coli* XL10-Gold cells transformed with pMF2.2 P
Plasmid DNA was prepared using the 1, 2, 3 method described in section 2.7. The DNA samples were separated on an agarose gel and DNA was stained using ethidium bromide. Lanes 1, 15 = Invitrogen 1 kb DNA ladder, lanes 2 – 13 = transformants 1-12, lane 14 = pMF2.1 X/P.

By comparing the sizes of the DNA fragments obtained for the transformants to the size of pMF2.1 P/X, it appears that all transformants contain an insert. Transformants 1-4, 7, 8 and 11 have the correct insert. Transformants 5, 6, 9, 10 and 12 have more than one insert.

In the case of pMF2.2 P it is very important to confirm the orientation of the *kan*^{R} gene as the Kan^R cassette may ligate to pMF2.1 P/X in either orientation. The XhoI site in the *kan*^{R} gene was used to confirm the orientation as shown in figure 5.23. Restriction analysis was also carried out on pJQ200sk+ DNA as a control.
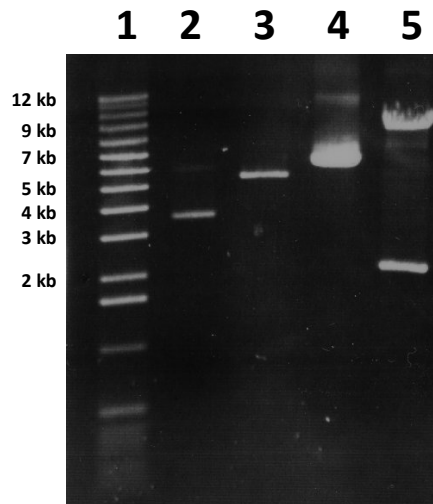
Figure 5.23 XhoI restriction digest of the plasmid pMF2.2 P to confirm that it contained the Kan^R

Figure 5.23 XhoI restriction digest of the plasmid pMF2.2 P to confirm that it contained the Kan^R cassette in the correct orientation

Following restriction analysis, DNA fragments were visualised by separating them on a 0.7% agarose gel and staining them with ethidium bromide. Lane 1 = Invitrogen 1 kb DNA ladder, lane 2 = uncut pJQ200sk+, lane 3 = pJQ200sk+ digested with XhoI, lane 4 = uncut pMF2.2 P, lane 5 = pMF2.2 P digested with XhoI.

XhoI cuts pJQ200sk+ once yielding a linear DNA fragment of 5.5 kb in size (lane 3). XhoI cuts pMF2.2 P twice, once inside the *kan^R* gene, yielding two fragments of 8.5 kb and 2.2 kb in size (lane 5). These are the expected sizes and thus confirm that the orientation of the Kan^R cassette is correct.

The final construct pMF2.2 P contained the *wblK* gene interrupted by *kan^R* in the opposite orientation. This plasmid was used to introduce the mutation into *P. luminescens*. The plasmid was introduced into *P. luminescens* by tri-parental mating. Transconjugates were selected by growing on LB agar containing rifampicin and kanamycin (section 2.4).

Homologous recombination occurred as described in section 5.2. One transconjugate colony was chosen and used to test for 2nd recombinants. The colony was grown in LB broth containing kanamycin. The following day, serial dilutions were performed on the culture and the dilutions were plated on media containing 5% sucrose and kanamycin. Any colonies that grew on these plates did not contain the *sacB* gene and so were able

to survive in the presence of sucrose. This meant that they were 2nd recombinants because the *sacB* gene had been excised from the chromosomal DNA.

As a final test, any colonies that grew on sucrose media were transferred to replica plates. One plate contained kanamycin and the other contained gentamicin. The gentamicin resistance gene should also be excised in 2nd recombinants and so the colonies should grow on the plates containing kanamycin and not on the ones containing gentamicin. Ten colonies were chosen for replica plating and all ten grew in the presence of kanamycin and not gentamicin.

Genomic DNA was prepared from the *P. luminescens wblK* mutant strain (section 2.6). Two separate PCRs were carried out using *P. luminescens* wildtype DNA and *P. luminescens wblK* mutant DNA as templates. The primers wblK_F1 and wblK_R2 were used in the reactions. wblK_F1 was the forward primer initially used to amplify the wblK1 fragment and wblK_R2 was the reverse primer used to amplify the wblK2 fragment. Using these primers together should amplify the entire 4080 base region of DNA in the wildtype strain containing the complete *wblK* gene. In the mutant strain the *wblK* gene is interrupted with the Kan^R cassette and so the size of the PCR product should be 5320 bases. The results of the PCRs are shown in figure 5.24.
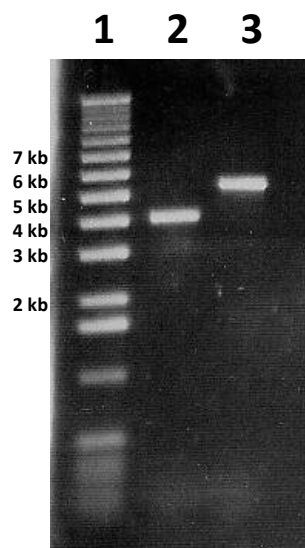
Figure 5.24 PCR products generated from the amplification of the entire wblK fragment from *P. luminescens* wildtype genomic DNA and *P. luminescens wblK* mutant genomic DNA

PCR products were separated on an agarose gel and visualised after ethidium bromide staining. Lane 1 = Invitrogen 1 kb DNA Ladder, lane 2 = wildtype PCR, lane 3 = *wblK* mutant PCR.

The size of the PCR fragment obtained for the wildtype (lane 2) is just over 4 kb and for the *wblK* mutant (lane 3) it is ~5.5 kb. The size of the KanR cassette is 1240 bp. This PCR analysis confirms unequivocally that in the strain *P. luminescens wblK* mutant, the *wblK* gene has been interrupted with a kanamycin resistance cassette.

5.4 Analysis of glycoproteins from *P. luminescens wblK* mutant

Once the *P. luminescens wblK* mutant was prepared, the next step was to analyse the mutant to examine how this mutation affects the glycan patterns of the glycoproteins produced by the organism.

A 50 ml culture of *P. luminescens wblK* mutant was prepared as described in section 2.5.1. Periplasmic, cytoplasmic and membrane proteins were isolated as before using the water lysis method (section 2.15). In order to examine the glycoproteins being produced by the *wblK* mutant similar methods were employed that had previously been used to analyse glycoproteins from the wildtype strain i.e. lectins were utilised in ELLAs, western blots and bound to agarose resins to detect and isolate glycoproteins. The anti-P.l. polyclonal antibody that had been prepared against whole cells of the *P. luminescens* wildtype strain (section 2.21) was also used to detect differences between the two strains.

Initially a number of ELLAs were carried out to profile the interactions between the periplasmic, cytoplasmic and membrane proteins from both strains with a range of lectins (table 3.2). Proteins prepared from both the wildtype strain and the mutant strain were adjusted to the same concentration and used in the same experiment so that the results could be accurately compared. The results from this experiment were very similar to those in figure 3.5 and there was no significant difference seen between the two strains with regard to the intensity of the signal generated with each lectin (results not shown).

Next western blots were carried out to compare the band pattern generated by proteins from the two strains when probed with the lectin SBA or anti-P.l. antibody. SBA was chosen as it gave a very strong response previously with *P. luminescens* proteins.

(WGA was not chosen as it tends to cause very high background staining and so does not work well in western blots). The anti-P.l. antibody was prepared against wildtype whole cells. The generated antibody would be specific for antigens from the surface of the cell including glycans on glycoproteins. If the *wblK* mutation affected the glycosylation of any surface proteins then this could affect the binding of the antibody to them. Figure 5.25 shows the result of the western blot analysis.
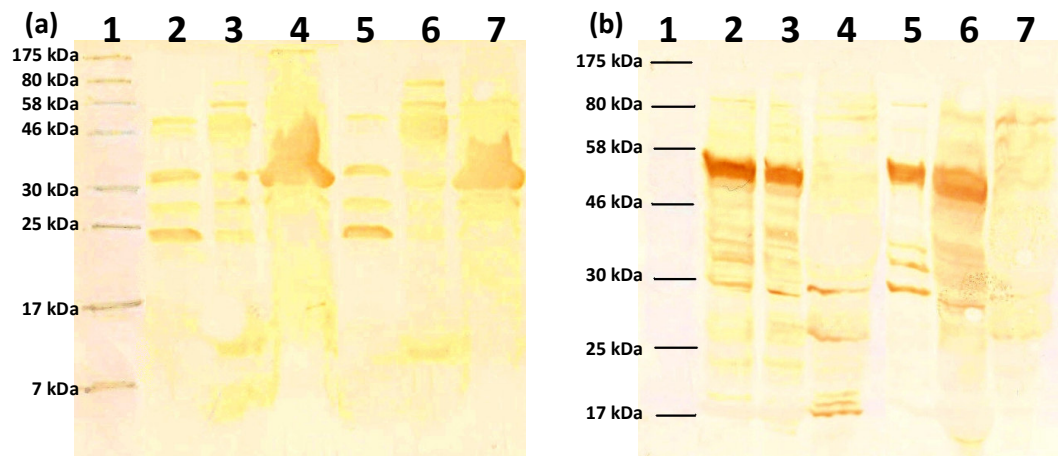
Figure 5.25 Western blot analysis showing interactions between *P. luminescens* wildtype and *wblK* mutant protein fractions with anti-P.l. antibody (a) and SBA lectin (b)

Lane 1 = NEB prestained protein marker, lane 2 = wildtype periplasmic proteins, lane 3 = wildtype cytoplasmic proteins, lane 4 = wildtype membrane proteins, lane 5 = *wblK* mutant periplasmic proteins, lane 6 = *wblK* mutant cytoplasmic proteins, lane 7 = *wblK* mutant membrane proteins.

In figure 5.25 there were no significant differences detected in the band pattern between the wildtype and mutant strains on the anti-P.l. blot. However, on the SBA blot when the membrane proteins are compared for both strains, bands are visible at ~17 kDa in the wildtype (figure 5.25 (b), lane 4) that are not as clear in the mutant (figure 5.25 (b), lane 7). On the original blot there are some small proteins showing up in lane 7 (b) at ~17 kDa that did not show up well when the blot was scanned. Therefore, from this blot it is not possible to conclude if protein glycosylation may have been affected by the mutation in *wblK*. For the initial ELLA and western blot analysis described above, crude periplasmic, cytoplasmic and membrane proteins were used. If the *wblK* mutation is having any effect on the glycoproteins produced by the organism it was not immediately obvious when looking at the large mix of total proteins produced by *P. luminescens*. Because so many proteins are produced by the organism, the bands seen in

figure 5.25 could contain not just one but several proteins of the same molecular weight and so any differences in the band pattern may be hidden. In order to make the analysis simpler and get a more detailed look at glycoproteins specifically, LAC was used to isolate glycoproteins from the other proteins in the cell.

5.4.1 Isolation of glycoproteins from *P. luminescens wblK* mutant periplasmic proteins using LAC

LAC was carried out using WGA resin as before and periplasmic proteins were again used because they produced the protein profiles that were observed as being least complex. Samples from the crude proteins (periplasmic proteins before separation on the column), unbound proteins and eluted proteins were run on an SDS-PAGE gel and stained with silver stain (figure 5.26 (a)). A western blot was also carried out with these samples which was probed with anti-P.l. antibody (figure 5.26 (b)).
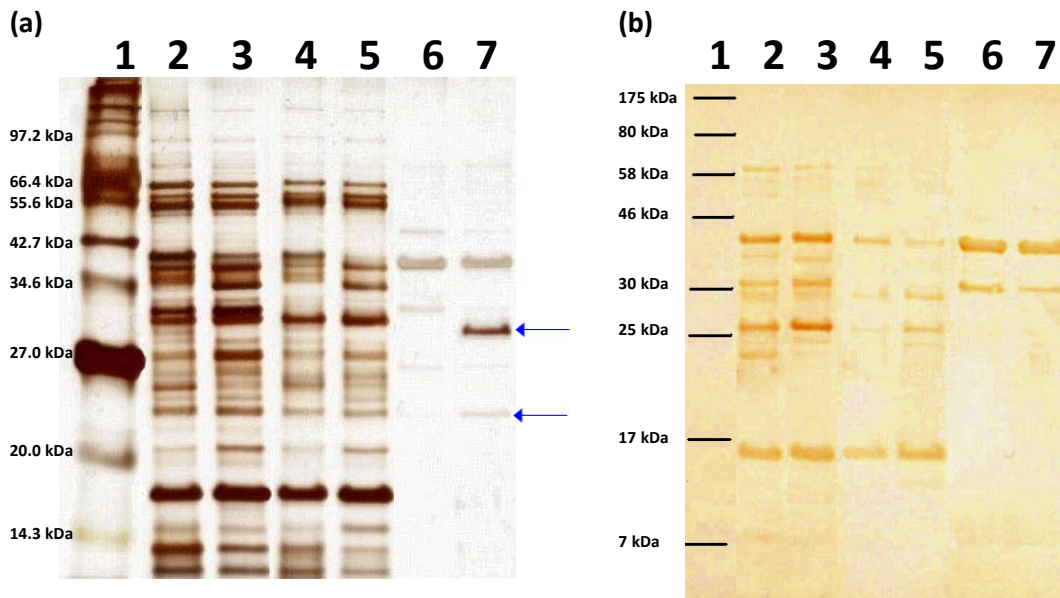
Figure 5.26 *P. luminescens* wildtype and *wblK* mutant WGA LAC periplasmic protein fractions separated on a 15% SDS-PAGE gel (a) and on a western blot probed with anti-P.l. antibody (b)
Lane 1 = NEB protein marker, lane 2 = crude wildtype periplasmic proteins, lane 3 = crude *wblK* mutant periplasmic proteins, lane 4 = unbound wildtype periplasmic proteins, lane 5 = unbound *wblK* mutant periplasmic proteins, lane 6 = eluted wildtype periplasmic proteins, lane 7 = eluted *wblK* mutant periplasmic proteins. Blue arrows indicate proteins that were isolated from the *wblK* mutant that were not present in the wildtype strain.

Many interesting observations may be made about the images in figure 5.27. Firstly, in the silver stained gel (a) there is a new band showing up in lane 3 (crude *wblK* mutant periplasmic proteins) at 34 kDa that is not visible in the wildtype (lane 2). There is also a band in lane 2 at ~40 kDa that appears to be absent in lane 3, although it is difficult to tell because of overlapping bands. It is possible that these bands are the same proteins and that mutating *wblK* could have altered the molecular weight of the protein from 40 kDa to 34 kDa if its glycosylation was altered. Neither of these proteins bound to WGA and are both clearly visible in the unbound fractions (lanes 4 and 5).

There is another band in lane 2 (a) at ~25 kDa that is not showing up in the mutant strain. Again this protein does not bind to WGA and is visible in the unbound fraction in lane 4. In the anti-P.l. blot there is a band visible just below the 25 kDa marker in lane 2 that is absent in lane 3. This could be the same protein that is seen in the silver stained gel. The fact that this protein is absent in the anti-P.l. blot in the mutant strain suggests that it may have been altered in some way by the mutation.

The most interesting difference between the two strains is seen when the eluted protein samples are compared (lanes 6 and 7). In the silver stained gel two new proteins are showing up at 30 kDa and 23 kDa (blue arrows) in the *wblK* mutant compared to the wildtype. The protein at 30 kDa is particularly strong. These new bands do not appear in the anti-P.l. blot. A number of possible scenarios could exist to explain the appearance of these new bands and the most likely scenario will now be explained. WGA binds terminal GlcNAc. WblK could be responsible for adding a further sugar moiety onto the glycan chain so that in the normal wildtype GlcNAc is not the terminating sugar and so it does not bind to WGA. However, when WblK is absent in the mutant, the glycan is not elongated any further and terminates with GlcNAc.

One final observation to make about figure 5.27 is that the glycosylation of the protein identified as OmpN (chapter 4) has not been affected by the mutation of *wblK*. The band is showing up in both the wildtype and mutant strain at 40 kDa as before and is still able to bind to WGA in the mutant. This means that either WblK is not needed for the glycosylation of OmpN or perhaps OmpN is glycosylated by a completely different pathway altogether. Many instances of multiple glycosylation pathways in the same

bacterium have been discovered already (chapter 1) so there is no reason to believe that *P. luminescens* does not have more than one protein glycosylation system.

In order to get a more detailed insight into the proteins eluted from the WGA resin in both the wildtype and mutant strain the proteins were run on a 2-D gel for better separation as shown in figure 5.27.
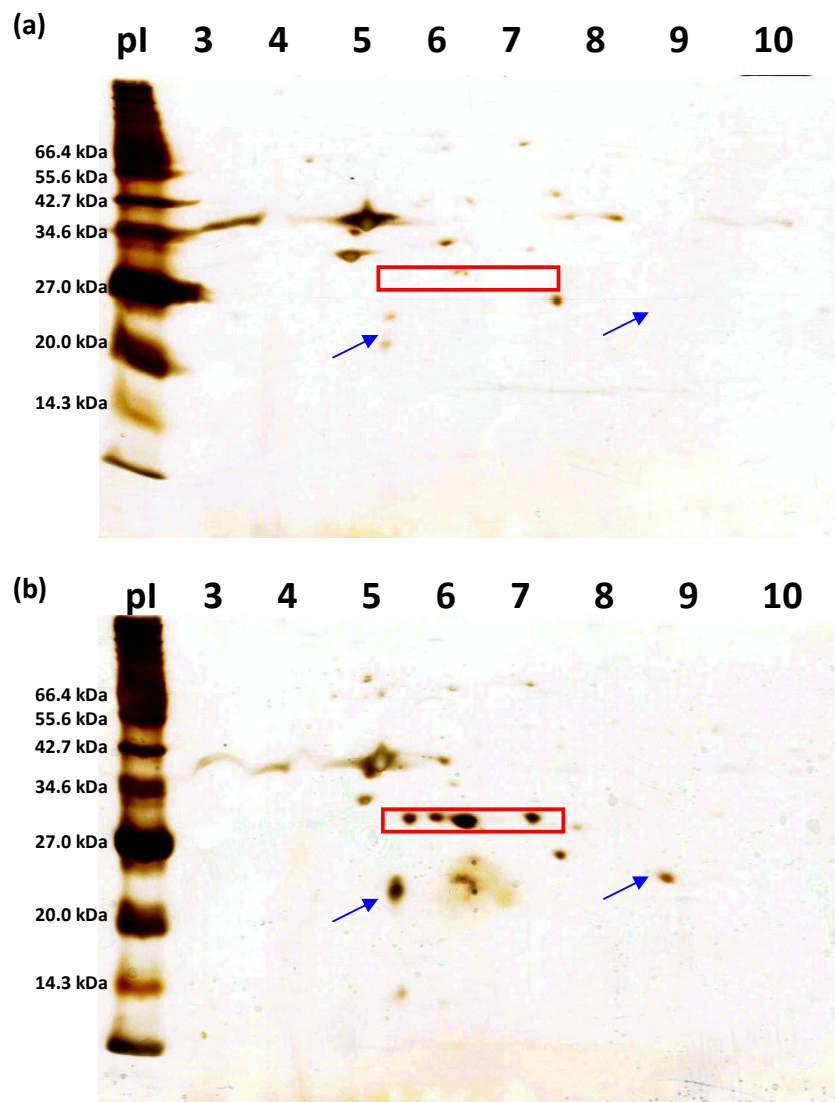
Figure 5.27 *P. luminescens* **wildtype (a) and** *wblK* **mutant (b) periplasmic proteins eluted from WGA agarose and separated on a 2-dimensional gel**

Periplasmic proteins from *P. luminescens* wildtype and *wblK* mutant strains that bound to WGA in a LAC column, were eluted from the column and separated on 2-dimensional gels, which were subsequently stained with silver stain. 30 kDa proteins that are visible in the mutant and not the wildtype are highlighted by a red box. Proteins with a molecular mass of ~23 kDa that are visible in the mutant and not the wildtype are highlighted by blue arrows.

It can be seen from figure 5.27 that the 30 kDa band that appeared in the *wblK* mutant eluted proteins (figure 5.26, lane 7) that was not present in the wildtype (figure 5.26, lane 6) actually consists of four proteins with almost identical molecular weights but different pI values. These spots are highlighted by a red box in figure 5.27 (b). These spots are not visible in figure 5.27 (a). The 23 kDa band highlighted in figure 5.26, lane 7, also seems to be made up of two proteins with very different pI values. These spots are highlighted by blue arrows in figure 5.27 (b) and again, these proteins are not seen in figure 5.27 (a).

The proteins highlighted in the red box in figure 5.27 (b) could be the same protein with slightly different glycosylation patterns which would give them different pI values. In other words they could be glycoforms. In order to investigate this two of the spots, the first spot at pI 5.5 and the third spot at pI 6.1 were chosen for identification.

5.4.2 Identification of the isolated 30 kDa glycoprotein from the periplasm of *P. luminescens wblK* mutant using peptide mass fingerprinting

In order to deduce if the 30 kDa proteins highlighted in figure 5.28 (b) were different proteins or different glycoforms of the same protein, two of the protein spots were chosen for identification using PMF. The proteins were sent to the company Alphalyse for identification. The method used for the generation of peptides was similar to that outlined in section 2.25.5. MALDI-MS using a Bruker Autoflex III MALDI TOF/TOF instrument in positive ion mode was employed to analyse the digested protein. The first sample was called WM1 and the second sample was called WM2. The MALDI spectra generated for WM1 and WM2 are shown in figure 5.28.
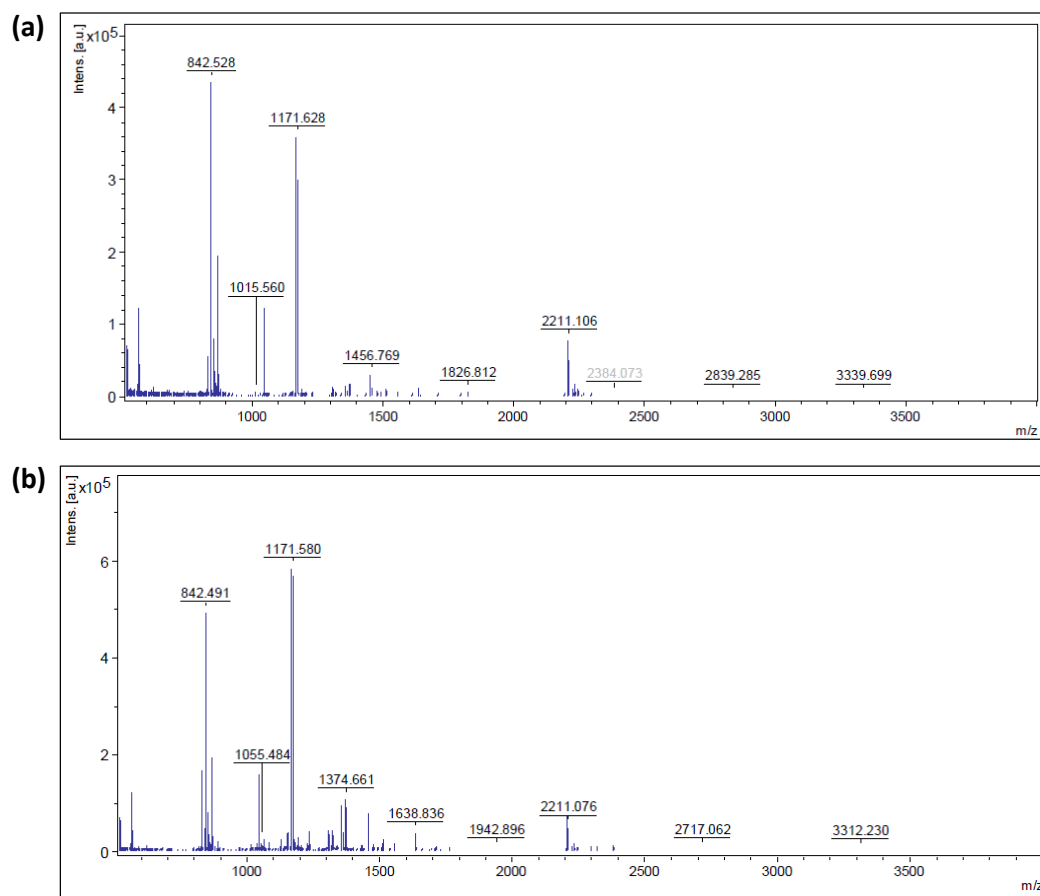
Figure 5.28 Mass spectra generated from the trypsin digestion of WM1 (a) and WM2 (b), 30 kDa proteins isolated from the periplasm of *P. luminescens wblK* mutant by WGA LAC

Peptides were analysed using a Bruker Autoflex III MALDI TOF/TOF mass spectrometer.

It can be seen from analysing the spectra from both protein samples that they are almost identical suggesting that the samples contain the same protein. The company also chose four peptides from each sample for fragmentation analysis i.e. partial sequencing using MALDI MS/MS. The data from this analysis as well as the data from the MS spectra (figure 5.28) were combined and used for database searching using the MASCOT software. The data was searched against in-house protein databases containing over 12 million known non-redundant protein sequences. The results from this analysis identified both samples to be the same protein. The protein was identified as plu3611, a hypothetical protein. The MASCOT score obtained for WM1 was 211 and for WM2 was 303, values that are far above the threshold of 78 for a significant match. Tables 5.2 and 5.3 show the peptides generated from the digestion of the WM1 and WM2 samples that matched to the sequence of plu3611. Those in bold were sequenced using MS/MS.

Figures 5.29 and 5.30 show the sequence of plu3611 and the peptides from the digested WM1 and WM2 samples that matched to it.

Table 5.2 Peptides generated from the digestion of WM1 (from figure 5.28 (a)) that matched to the sequence of plu3611

Residue	Observed mass (Da)	Expected Mr (Da)	Calculated Mr (Da)	Delta	Miss	Sequence
73 - 78	832.43	831.42	831.42	0.00	0	K.YFDFLK.N
73 - 84	1510.77	1509.76	1509.76	0.00	1	K.YFDFLKNHGIEK.K
85 - 93	1015.56	1014.55	1014.55	0.00	1	K.KDIDAANLR.T
86 - 93	887.49	886.48	886.45	+0.03	0	K.DIDAANLR.T
94 - 106	1515.66	1514.66	1514.65	+0.01	0	R.TQPDYEYDNVSGK.S
124 - 137	**1555.87**	**1554.86**	**1554.86**	**0.00**	**1**	**R.KLDQLNDLLDGALK.A**
183 - 192	**1171.63**	**1170.62**	**1170.60**	**+0.02**	**0**	**K.VGSIYSINYR.A**
193 - 201	1055.52	1054.52	1054.51	+0.01	0	R.APEVIDHMK.Y Oxidation (M)
193 - 203	**1358.70**	**1357.69**	**1357.68**	**+0.01**	**1**	**R.APEVIDHMKYR.N**
193 - 203	**1374.69**	**1373.69**	**1373.68**	**+0.01**	**1**	**R.APEVIDHMKYR.N Oxidation (M)**

```
  1 MKLKSLGLAA MLSAGVPLAA QAADLPGVPH VITSGNAVVK AEPDIATLLI
 51 NVNIAAKDAS GAKKQVDELV AKYFDFLKNH GIEKKDIDAA NLRTQPDYEY
101 DNVSGKSVLK GYRAIRSVEV KVRKLDQLND LLDGALKAGL NEIVSVQFGV
151 DNPQKYRDEA RQKAIENAIE QAGALAKGFN SKVGSIYSIN YRAPEVIDHM
201 KYRNTDVMMS GGAAAGVGET YQQDSINFSD QVDVVFELKP
```

Figure 5.29 The sequence of plu3611, a hypothetical protein from *P. luminescens*

The peptides that were generated from the digestion of WM1 that matched to the sequence of plu3611 are shown in bold red. The percentage sequence coverage was 28%.

Table 5.3 Peptides generated from the digestion of WM2 (from figure 5.28 (b)) that matched to the sequence of plu3611

Residue	Observed mass (Da)	Expected Mr (Da)	Calculated Mr (Da)	Delta	Miss	Sequence
64 - 72	1029.56	1028.55	1028.59	-0.04	1	K.KQVDELVAK.Y
73 - 78	832.39	831.39	831.42	-0.03	0	K.YFDFLK.N
73 - 84	1510.74	1509.73	1509.76	-0.03	1	K.YFDFLKNHGIEK.K
85 - 93	1015.51	1014.51	1014.55	-0.04	1	K.KDIDAANLR.T
86 - 93	887.44	886.43	886.45	-0.02	0	K.DIDAANLR.T
94 - 106	1515.63	1514.62	1514.65	-0.03	0	R.TQPDYEYDNVSGK.S
94 - 110	1942.90	1941.89	1941.93	-0.04	1	R.TQPDYEYDNVSGKSVLK.G
107 - 113	822.47	821.46	821.48	-0.02	1	K.SVLKGYR.A
124 - 137	**1555.84**	**1554.83**	**1554.86**	**-0.03**	**1**	**R.KLDQLNDLLDGALK.A**
183 - 192	**1171.58**	**1170.57**	**1170.60**	**-0.03**	**0**	**K.VGSIYSINYR.A**
193 - 201	1039.47	1038.46	1038.52	-0.06	0	R.APEVIDHMK.Y
193 - 201	1055.48	1054.48	1054.51	-0.03	0	R.APEVIDHMK.Y Oxidation (M)
193 - 203	**1358.67**	**1357.66**	**1357.68**	**-0.02**	**1**	**R.APEVIDHMKYR.N**
193 - 203	**1374.66**	**1373.65**	**1373.68**	**-0.03**	**1**	**R.APEVIDHMKYR.N Oxidation (M)**

```
  1 MKLKSLGLAA MLSAGVPLAA QAADLPGVPH VITSGNAVVK AEPDIATLLI
 51 NVNIAAKDAS GAKKQVDELV AKYFDFLKNH GIEKKDIDAA NLRTQPDYEY
101 DNVSGKSVLK GYRAIRSVEV KVRKLDQLND LLDGALKAGL NEIVSVQFGV
151 DNPQKYRDEA RQKAIENAIE QAGALAKGFN SKVGSIYSIN YRAPEVIDHM
201 KYRNTDVMMS GGAAAGVGET YQQDSINFSD QVDVVFELKP
```

Figure 5.30 The sequence of plu3611, a hypothetical protein from *P. luminescens*

The peptides that were generated from the digestion of WM2 that matched to the sequence of plu3611 are shown in bold red. The percentage sequence coverage was 35%.

It can be seen from the above data that both digested samples matched very well to plu3611. The percentage sequence coverage obtained for the WM2 sample was a little higher than for WM1 probably because the protein spot that contained WM2 was much bigger than WM1 and so would have had a higher concentration of sample.

plu3611 is a hypothetical protein. Cellular roles have been assigned to this protein by the JCVI. It is stated as being involved in electron transport, ribosomal protein synthesis and to have other regulatory functions.

The BlastP program (NCBI, section 2.12) was used to search for other proteins homologous to plu3611. The search returned many matches that were very similar to plu3611. The closest matches were seen in *P. asymbiotica*, *Xenorhabdus* spp. and *Yersinia* spp. However all of the matches to plu3611 came back as hypothetical proteins or proteins of unknown function or uncharacterised proteins so orthologues to plu3611 have not been well studied in any organism.

Searches were also carried out to identify conserved domains in this protein. plu3611 was found to have a conserved domain from oxidative stress defence proteins. It also contains two other conserved domains that have not been characterised. All three conserved domains are found in proteins from the SIMPL superfamily. Members from this superfamily have so far only been well studied in mice.

Although nothing is known about the function of the protein plu3611, this study shows that the protein is most likely glycosylated. Aberrant glycosylation of plu3611 was caused by mutating the *wblK* gene suggesting that the WblK protein (or WblF/WblM or all three proteins) is involved in normal glycosylation of this protein. Also, at least two different glycoforms exist in the *wblK* mutant as the two spots from figure 5.28 that were identified had different pI values. The calculated pI of the protein backbone of plu3611 is 5.97. On the 2-D gel WM1 and WM2 were shown to have pIs of 5.5 and 6.1. Also the molecular weight of the protein backbone is 25893 Da. The protein appears to have a molecular weight of ~30 kDa on a protein gel. The significant difference between the calculated and observed molecular weights may be accounted for by glycosylation. No other post translational modification could cause a difference of ~4 kDa in size. The work presented here also shows that it is likely that more than one glycosylation system could exist in *P. luminescens* since the mutation of *wblK* did not seem to alter the glycosylation of OmpN.

5.4.3 *O*-deglycosylation of glycoproteins from the periplasm of *P. luminescens* wildtype and *wblK* mutant strains

Proteins isolated from the periplasm of the two *P. luminescens* strains using LAC were treated with β-elimination reagent in order to remove any *O*-linked glycans. The

procedure for *O*-deglycosylation is found in section 2.26. Untreated and treated proteins were analysed by ELLA analysis to see if they still bound to the lectin WGA, which was initially used to isolate them. The results of the ELLA analysis are shown in figure 5.31.
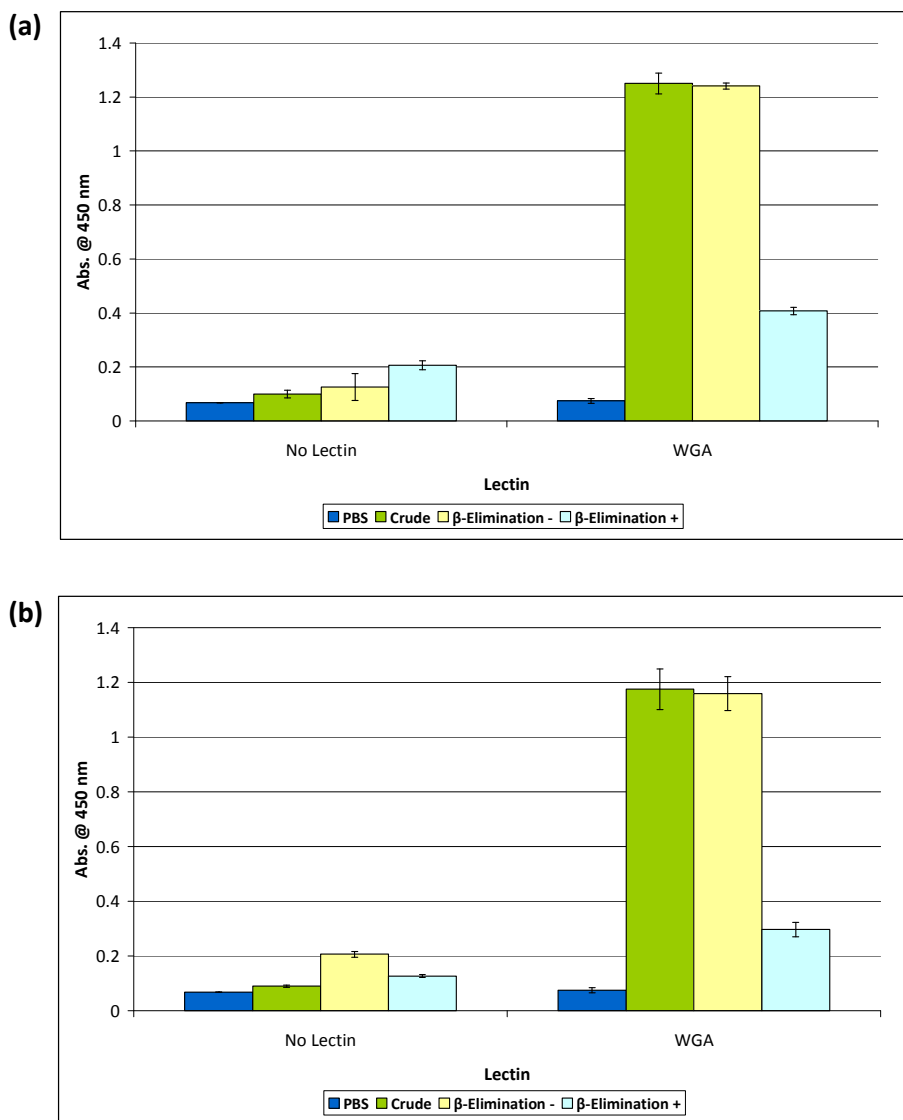
Figure 5.31 ELLA profiling interactions between *O*-deglycosylated *P. luminescens* wildtype (a) and *wblK* mutant (b) periplasmic proteins eluted from a WGA LAC column with the lectin WGA
Green bars represent proteins eluted from the WGA LAC column that were not treated with β-elimination reagent (crude sample). The light blue bars represent proteins that were *O*-deglycosylated (β-Elimination +). A control reaction was done where all of the steps of the β-elimination reaction were carried out except that dH$_2$O was used instead of β-elimination reagent (yellow bars, β-Elimination -).

It can be seen from the results shown in figure 5.31 that in both wildtype and *wblK* mutant samples, treatment with β-elimination reagent dramatically reduced the signal produced with the lectin WGA. These data provide strong evidence that most of the proteins isolated from the periplasm of *P. luminescens* using WGA LAC are *O*-glycosylated. Treatment with β-elimination reduced the ELLA signal generated with WGA almost down to baseline levels. In chapter 3, β-elimination was carried out on crude periplasmic, cytoplasmic and membrane proteins isolated from *P. luminescens* wildtype. It was also shown to significantly reduce the binding of the lectin WGA to proteins in each of the fractions (figure 3.22). Although some of the proteins produced by *P. luminescens* may be *N*-glycosylated it appears so far that *O*-linked glycosylation prevails, however, this requires further investigation.

Chapter 6

Concluding remarks

Interest in the area of bacterial glycoproteomics has increased significantly in the past twenty years especially since the breakthrough discovery of the *N*-linked general protein glycosylation pathway in *C. jejuni* in 2002 (Young et al. 2002). The types of bacterial glycans that have been discovered are very varied and complex in terms of the monosaccharide units that they consist of and the linkages between them. Many unusual monosaccharides have been found in bacterial systems that are not present in eukaryotic glycoproteins. It also appears that different bacterial organisms have evolved diverse glycosylation pathways, for example the stepwise *N*-linked glycosylation system seen in *H. influenzae* (Gross et al. 2008) and OT mediated *O*-linked glycosylation in *Neisseria* spp. (Faridmoayer et al. 2007) and *P. aeruginosa* (Smedley III et al. 2005). These types of glycosylation pathways have not been seen in any eukaryotic system. There is no doubt that what is known so far in the area of bacterial glycoproteomics is only the tip of the iceberg and there is much more that remains to be discovered.

The focus of this research was to investigate protein glycosylation in the Gram-negative, Gammaproteobacterium, *P. luminescens*. Analysis of the *P. luminescens* genome revealed genes encoding proteins homologous to Pgl proteins from *C. jejuni* involved in the general OT mediated *N*-linked protein glycosylation pathway (Linton et al. 2005) and Pgl proteins from *N. meningitidis* involved in the OT mediated *O*-linked protein glycosylation pathway (Power et al. 2003a). This information combined with knowledge about how adhesion and invasion play a crucial role in the life cycle of *P. luminescens* provided strong evidence to suggest that the organism does produce glycosylated proteins.

Lectins (carbohydrate binding proteins) were employed to analyse the entire pool of proteins produced by *P. luminescens* for the presence of glycoproteins. Initial screening of extracellular proteins and proteins from the cytoplasm, periplasm and membrane of the bacterium with various lectins in ELLAs and western blots revealed that many proteins from each fraction were interacting with a range of lectins especially mannose, galactose, GalNAc and GlcNAc binding lectins. This was consistent with the types of glycans that had already been discovered in other bacterial glycoproteins. In order to validate the results obtained for *P. luminescens* similar experiments were carried out using *C. jejuni* proteins. The results obtained for the *C. jejuni* proteins were consistent with what was already presented previously in the literature (Linton et al. 2002).

Lectin affinity chromatography, using a range of agarose bound lectins, Con A, GSL I and WGA was used to isolate putative glycoproteins from the cytoplasm and periplasm of *P. luminescens*. The addition of free sugar, known to inhibit glycan-binding of the particular lectin, was used to elute putative glycoproteins from the resin. This strongly suggests that that the binding of putative glycoproteins from *P. luminescens* to lectins is occurring through glycan-protein interaction rather than protein-protein or other non-specific interactions. Putative glycoproteins that were isolated in this way produced strong signals in western blots and ELLAs with the lectin that was used to isolate them initially, as well as with other lectins, suggesting that they are modified with more that one monosaccharide.

Two putative glycoproteins that had been isolated from the periplasm of *P. luminescens* by LAC using WGA agarose were identified by mass spectrometry. Peptide mass fingerprinting was the technique employed to identify the proteins. Bands containing the two proteins were excised from an SDS-PAGE gel and digested into peptides using the enzyme trypsin and the masses of the generated peptides were found using MALDI-MS. One putative glycoprotein was identified as plu3795, a hypothetical protein. No information was available about that protein and so it was not pursued any further. The second putative glycoprotein to be identified was an outer membrane porin, OmpN. The OmpN protein sample that had been digested with trypsin was also treated with the enzyme PNGaseF to remove any *N*-linked glycans and the sample was re-analysed by MALDI-MS. The PNGaseF treated sample contained no new peptides in its spectrum compared to the non-treated sample and so it was assumed that the glycosylation was not *N*-linked but possibly *O*-linked. The glycosylation of many structural outer membrane proteins has been discussed in the literature however no instances were found where an outer membrane porin was shown to be glycosylated.

In order to carry out further analysis of the OmpN protein, the *ompN* gene was cloned with a histidine tag, ligated to a broad host range plasmid and expressed in its natural background, *P. luminescens*. This enabled purification of OmpN using immobilised metal affinity chromatography. The plasmid containing the recombinant *ompN* gene was also transformed into *E. coli* and expressed in order to have a non-glycosylated form of the protein for comparison. Expression of OmpN in both *E. coli* and *P. luminescens* was found to be very low, however, a pure sample of OmpN was isolated

from both strains for analysis. Purified OmpN from *P. luminescens* retained its ability to bind WGA in LAC indicating that glycosylation of the recombinant protein was still occurring. OmpN purified from *P. luminescens* was also shown to bind WGA in an ELLA. It was also shown to bind the lectin GSL I. In contrast, OmpN purified from *E. coli* did not bind either of these lectins in an ELLA. The *E. coli* stain used does not contain the same glycosylation machinery as *P. luminescens*, therefore the only difference between the proteins produced in the two stains should be in how they are glycosylated. This fact and the results from the ELLA analysis of OmpN from both species strongly indicates that OmpN is glycosylated in *P. luminescens*.

The glycosylation of the porin OmpN could possibly be significant in enabling *P. luminescens* carry out its life cycle. Following the sequencing of the genome of *P. luminescens* many genes encoding proteins that are likely to be involved in host colonisation and invasion of the insect were identified. These proteins include a number of adhesins, lectins and fimbrial genes (Duchaud et al. 2003). The mechanisms for how *P. luminescens* colonises the host and infects the insect are not entirely understood. Because OmpN is an outer membrane, surface exposed protein, it is possible that it could also play a role in colonisation and adhesion and that the glycosylation of this protein is aiding in these processes. The best way to discover if this is true would be to mutate the genes that are involved in glycosylating OmpN and carry out *in vivo* studies using the mutant organism to investigate if it is still capable of colonising the nematode and infecting the insect.

As previously stated a number of genes encoding proteins homologous to genes involved in protein glycosylation in *C. jejuni* and *N. meningitidis* were identified in *P. luminescens*. One such gene, *wblK*, was mutated by insertional inactivation. From analysis of the genome of *P. luminescens* it appeared that *wblK* was part of an operon, with the genes *wblF* and *wblM* lying adjacent to it, with no gaps between the genes. Therefore it was anticipated that if *wblK* was knocked out, *wblF* and *wblM* would also be mutated. The *wblK* mutant was constructed and verified. Initial analysis of proteins isolated from the cytoplasm, periplasm and membrane of the *wblK* mutant showed no significant variations in the binding to a range of lectins in ELLAs compared to the wildtype. Proteins from the various cell fractions were also examined on a western blot and probed with anti-P.l. antibody, a polyclonal antibody raised against *P. luminescens*

wildtype whole cells. Again no changes in the protein band pattern were observed. However, when proteins from the *wblK* mutant periplasm were passed over a WGA LAC column and bound proteins were eluted with GlcNAc, two new proteins were visible on an SDS-PAGE gel in the eluted protein fraction compared to the wildtype.

The eluted proteins from the *wblK* mutant were further separated using 2-D gel electrophoresis. It was found that the 30 kDa band that appeared on the 1-D gel actually comprised four different spots on the 2-D gel. It was proven that at least two of the spots consisted of the same protein as they were both identified to be the protein plu3611 using peptide mass fingerprinting. Plu3611 is a hypothetical protein with very little information available about it. The molecular weight of the protein backbone is 26 kDa and the observed molecular weight on a protein gel is ~30 kDa. This large variation of molecular weight can be accounted for by glycosylation. The fact that at least two forms of the same protein with slight variations in their pI values were isolated through their binding to the lectin WGA suggests that these proteins are glycoforms of each other.

The mutation of *wblK* caused a change in the glycosylation of plu3611 and at least one other protein with a molecular weight of ~23 kDa. These proteins were not isolated by WGA LAC in the wildtype strain. However, mutating *wblK* did not seem to alter the glycosylation of OmpN as this protein was still isolated from the periplasm of the *wblK* mutant. This suggests that OmpN is being glycosylated through a different glycosylation system. This is not unreasonable to imagine as many other bacteria have been shown to contain more than one protein glycosylation system.

At present it appears more likely that proteins that are glycosylated in *P. luminescens* are modified with *O*-linked glycans rather than *N*-linked glycans. *P. luminescens* putative glycoproteins isolated by WGA LAC were treated with β-elimination reagent (an *O*-deglycosylation agent). The signal observed with the binding of WGA in an ELLA was dramatically reduced in treated samples compared to untreated samples. It is also more likely that the glycosylation is not OT mediated. No orthologues to known OT enzymes (PglB from *C. jejuni*, PglL from *N. meningitidis*, PilO from *P. aeruginosa* and HMW1C from *H. influenzae*) have been found in *P. luminescens*.

A number of interesting experiments could be carried out in order to continue forward from the work outlined in this thesis. For example, the protein plu3611, which was isolated from *P. luminescens wblK* mutant could be analysed further with respect to its glycan content and its function. Glycan analysis could be carried out using NMR spectroscopy and mass spectrometry. The function of plu3611 could be elucidated by knocking out the gene encoding plu3611 and analysing its phenotype. The gene could also be cloned and expressed and the protein purified. The purified protein could be analysed using two-hybrid screening or immunoprecipitation to discover protein-protein interactions.

The *wblK* mutation could also be complemented to see if the phenotype is restored to that which is seen in the wildtype. This may be done by cloning the *wblK* gene and ligating it to a plasmid such as pBBR1MSC-5 and expressing it in *P. luminescens wblK* mutant. It could also be complemented with *pglE* from *C. jejuni* in order to ascertain if PglE could carry out the same function in *P. luminescens* as it does in *C. jejuni*.

In vivo studies could be carried out to examine the effect of the *wblK* mutation on the life cycle and/or pathogenesis of *P. luminescens*. The organism can be injected directly into insect larva. Parameters such as numbers of bacteria and time taken for death of the insect to occur could be compared to the wildtype. Random mutagenesis could be used to find other glycosylation genes, for example those that may be involved in OmpN glycosylation. Again these mutants could be analysed using *in vivo* studies.

Research in the area of bacterial glycoproteomics is now proving to be very important for the biopharmaceutical industry. As discussed in section 1.5, scientists are now looking to prokaryotes as new expression systems for glycoprotein therapeutics, because of the ease with which they can be grown, maintained and manipulated compared to eukaryotic organisms (Pandhal and Wright 2010). Any knowledge that can be gained about the components of a bacterial glycosylation system would be very useful and could open up the potential to undertake the production of biotherapeutics in bacteria, where effective glycosylation could be engineered (Langdon et al. 2009).

Another reason for the interest in bacterial glycoproteomics is that it has been realised that glycoproteins can play a huge role in the pathogenicity of an organism. Knowledge

about the glycome of a pathogenic organism can be used to help prevent and treat infection from medically relevant pathogens. *P. luminescens* is not considered a human pathogen but its close relative *P. asymbiotica* is an opportunistic pathogen of humans (Gerrard et al. 2004). *P. luminescens* is also a promising natural insecticide (Blackburn et al. 2005), therefore knowledge of the mechanisms this organism uses to invade the insect could be valuable. Glycosylation may play a role in invasion by *P. luminescens*.

In eukaryotes protein glycosylation is essential for the correct function of many proteins. Glycosylation can be essential for protein trafficking in eukaryotes and changes in the glycosylation pattern of some eukaryotic glycoproteins have been directly implicated in certain diseases (Taylor and Drickamer 2006, Varki 1999). While it is not likely that glycosylation is of equal relevance in prokaryotes, the discovery of glycosylation systems in bacteria is likely to open up new insights into how bacterial cells function.

References

1. Abu-Qarn, M., Eichler, J. and Sharon, N. 2008. Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr. Opin. Struct. Biol.* 18 (5), pp544-550.

2. Aebi, M. and Hennet, T. 2001. Congenital disorders of glycosylation: genetic model systems lead the way. *Trends Cell. Biol.* 11 (3), pp136-141.

3. Ambrosi, M., Cameron, N. R. and Davis, B. G. 2005. Lectins: tools for the molecular understanding of the glycocode. *Org. Biomol Chem.* 3 (9), pp1593-1608.

4. Anttinen, H. and Hulkko, A. 1980. Regulation of the glycosylations of collagen hydroxylysine in chick embryo tendon and cartilage cells. *Biochimica et Biophysica Acta (BBA) - General Subjects* 632 (3), pp417-427.

5. Apweiler, R., Hermjakob, H. and Sharon, N. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1473 (1), pp4-8.

6. Arora, S. K., Bangera, M., Lory, S. and Ramphal, R. 2001. A genomic island in *Pseudomonas aeruginosa* carries the determinants of flagellin glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* 98 (16), pp9342-9347.

7. Barondes, S. H. 1988. Bifunctional properties of lectins: lectins redefined. *Trends Biochem. Sci.* 13 (12), pp480.

8. Bartels, K., Funken, H., Knapp, A. 2011. Glycosylation Is Required for Outer Membrane Localization of the Lectin LecB in *Pseudomonas aeruginosa*. *J. Bacteriol.* 193 (5), pp1107-1113.

9. Baumeister, W. and Lembcke, G. 1992. Structural features of archaebacterial cell envelopes. *J. Bioenerg. Biomembr.* 24 (6), pp567-575.

10. Beeley, J. G. 1977. Peptide chain conformation and the glycosylation of glycoproteins. *Biochem. Biophys. Res. Commun.* 76 (4), pp1051.

11. Benz, I. and Schmidt, M. A. 2002. Never say never again: protein glycosylation in pathogenic bacteria. *Mol. Microbiol.* 45 (2), pp267-276.

12. Benz, I. and Schmidt, M. A. 2001. Glycosylation with heptose residues mediated by the *aah* gene product is essential for adherence of the AIDA-I adhesin. *Mol. Microbiol.* 40 (6), pp1403-1413.

13. Bernatchez, S., Szymanski, C. M., Ishiyama, N. 2005. A Single Bifunctional UDP-GlcNAc/Glc 4-Epimerase Supports the Synthesis of Three Cell Surface Glycoconjugates in *Campylobacter jejuni*. *J. Biol. Chem.* 280 (6), pp4792-4802.

14. Birnboim, H. C. and Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* 7 (6), pp1513-1523.

15. Blackburn, M. B., Domek, J. M., Gelman, D. B. and Hu, J. S. 2005. The broadly insecticidal *Photorhabdus luminescens* toxin complex a (Tca): activity against the Colorado potato beetle, *Leptinotarsa decemlineata*, and sweet potato whitefly, *Bemisia tabaci*. *J.Insect Sci.* 5, pp32-42.

16. Böhme, U. and Cross, G. A. M. 2002. Mutational analysis of the variant surface glycoprotein GPI-anchor signal sequence in *Trypanosoma brucei*. *J. Cell. Sci.* 115 (4), pp805-816.

17. Bolivar, F., Rodriguez, R. L., Greene, P. J. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2 (2), pp95-113.

18. Bugg, T. D. H. and Brandish, P. E. 1994. From peptidoglycan to glycoproteins: Common features of lipid-linked oligosaccharide biosynthesis. *FEMS Microbiol. Lett.* 119 (3), pp255-262.

19. Campbell, C. and Yarema, K. 2005. Large-scale approaches for glycobiology. *Genome Biol.* 6 (11), pp236.

20. Carlson, D. M. 1968. Structures and Immunochemical Properties of Oligosaccharides Isolated from Pig Submaxillary Mucins. *J. Biol. Chem.* 243 (3), pp616-626.

21. Caroff, M. and Karibian, D. 2003. Structure of bacterial lipopolysaccharides. *Carbohydr. Res.* 338 (23), pp2431.

22. Castric, P. 1995. *pilO*, a gene required for glycosylation of *Pseudomonas aeruginosa* 1244 pilin. *Microbiology* 141 (5), pp1247-1254.

23. Castric, P., Cassels, F. J. and Carlson, R. W. 2001. Structural Characterization of the *Pseudomonas aeruginosa* 1244 Pilin Glycan. *J. Biol. Chem.* 276 (28), pp26479-26485.

24. Comer, J. E., Marshall, M. A., Blanch, V. J. 2002. Identification of the *Pseudomonas aeruginosa* 1244 Pilin Glycosylation Site. *Infect. Immun.* 70 (6), pp2837-2845.

25. Dell, A., Galadari, A., Sastre, F. and Hitchen, P. 2010. Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int. J. Microbiol.* 2010pp148178.

26. DiGiandomenico, A., Matewish, M. J., Bisaillon, A. 2002. Glycosylation of *Pseudomonas aeruginosa* 1244 pilin: glycan substrate specificity. *Mol. Microbiol.* 46 (2), pp519-530.

27. Dobos, K. M., Khoo, K. H., Swiderek, K. M. 1996. Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. *J. Bacteriol.* 178 (9), pp2498-2506.

28. Doig, P., Kinsella, N., Guerry, P. and Trust, T. J. 1996. Characterization of a post-translational modification of *Campylobacter* flagellin: identification of a sero-specific glycosyl moiety. *Mol. Microbiol.* 19 (2), pp379-387.

29. Duchaud, E., Rusniok, C., Frangeul, L. 2003. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotech.* 21 (11), pp1307.

30. Duchene, M., Schweizer, A., Lottspeich, F. 1988. Sequence and transcriptional start site of the *Pseudomonas aeruginosa* outer membrane porin protein F gene. *J. Bacteriol.* 170 (1), pp155-162.

31. Endo, T. 1999. *O*-Mannosyl glycans in mammals. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1473 (1), pp237-246.

32. Evans, J. N. S. 1995. *Biomolecular NMR spectroscopy.* Oxford ; New York: Oxford University Press.

33. Fan, X., She, Y., Bagshaw, R. D. 2004. A method for proteomic identification of membrane-bound proteins containing Asn-linked oligosaccharides. *Anal. Biochem.* 332 (1), pp178-186.

34. Faridmoayer, A., Fentabil, M. A., Mills, D. C. 2007. Functional Characterization of Bacterial Oligosaccharyltransferases Involved in *O*-Linked Protein Glycosylation. *J. Bacteriol.* 189 (22), pp8088-8098.

35. Feldman, M. F., Wacker, M., Hernandez, M. 2005. Engineering *N*-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli. Proc. Natl. Acad. Sci. U.S.A.* 102 (8), pp3016-3021.

36. Finan, T. M., Kunkel, B., De Vos, G. F. and Signer, E. R. 1986. Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. *J. Bacteriol.* 167 (1), pp66-72.

37. Fletcher, C. M., Coyne, M. J., Villa, O. F. 2009. A General *O*-Glycosylation System Important to the Physiology of a Major Human Intestinal Symbiont. *Cell,* 137 (2), pp321-331.

38. Forst, S. and Nealson, K. 1996. Molecular biology of the symbiotic-pathogenic bacteria *Xenorhabdu*s spp. and *Photorhabdus* spp. *Microbiol. Rev.* 60 (1), pp21-43.

39. Forst, S., Dowds, B., Boemare, N. and Stackebrandt, E. 1997. *Xenorhabdus* and *Photorhabdus* spp. :Bugs That Kill Bugs. *Ann.l Rev. Microbiol.* 51 (1), pp47-72.

40. Fry, B. N., Feng, S., Chen, Y. 2000. The galE Gene of *Campylobacter jejuni* Is Involved in Lipopolysaccharide Synthesis and Virulence. *Infect. Immun.* 68 (5), pp2594-2601.

41. Gaastra, W. and Svennerholm, A. 1996. Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends Microbiol.* 4 (11), pp444-452.

42. Gabius, H., André, S., Jiménez-Barbero, J. 2011. From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* 36 (6), pp298-313.

43. Gerrard, J., Waterfield, N., Vohra, R. and ffrench-Constant, R. 2004. Human infection with Photorhabdus asymbiotica: an emerging bacterial pathogen. *Microbes Infect.* 6 (2), pp229-237.

44. Geyer, H. and Geyer, R. 2006. Strategies for analysis of glycoprotein glycosylation. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* 1764 (12), pp1853.

45. Glover, K. J., Weerapana, E. and Imperiali, B. 2005. *In vitro* assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic *N*-linked glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* 102 (40), pp14255-14259.

46. Goon, S., Kelly, J. F., Logan, S. M. 2003. Pseudaminic acid, the major modification on *Campylobacter* flagellin, is synthesized via the Cj1293 gene. *Mol. Microbiol.* 50 (2), pp659-671.

47. Grass, S., Buscher, A. Z., Swords, W. E. 2003. The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Mol. Microbiol.* 48 (3), pp737-751.

48. Gross, J., Grass, S., Davis, A. E. 2008. The *Haemophilus influenzae* HMW1 Adhesin Is a Glycoprotein with an Unusual *N*-Linked Carbohydrate Modification. *J. Biol. Chem.* 283 (38), pp26010-26015.

49. Grünewald, S., Matthijs, G. and Jaeken, J. 2002. Congenital Disorders of Glycosylation: A Review. *Pediatr. Res.* 52 (5), pp618-624.

50. Haltiwanger, R. S. and Stanley, P. 2002. Modulation of receptor signaling by glycosylation: fringe is an *O*-fucose-[beta]1,3-*N*-acetylglucosaminyltransferase. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1573 (3), pp328-335.

51. Hamilton, S. R. and Gerngross, T. U. 2007. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr. Opin. Biotechnol.* 18 (5), pp387-392.

52. Henzel, W. J., Billeci, T. M., Stults, J. T. 1993. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. U.S.A.* 90 (11), pp5011-5015.

53. Hillenkamp, F. andPeter-Katalinić, J. 2007. *MALDI MS: A practical guide to instrumentation, methods and applications.* Wiley-Vch Verlagsgesellschaft Mbh : Weinheim.

54. Hitchen, P. G. and Dell, A. 2006. Bacterial glycoproteomics. *Microbiology* 152 (6), pp1575-1580.

55. Horzempa, J., Dean, C. R., Goldberg, J. B. and Castric, P. 2006. *Pseudomonas aeruginosa* 1244 Pilin Glycosylation: Glycan Substrate Recognition. *J. Bacteriol.* 188 (12), pp4244-4252.

56. Inokuchi, K., Mutoh, N., Matsuyama, S. and Mizushima, S. 1982. Primary structure of the *ompF* gene that codes for a major outer membrane protein of *Escherichia coli* K-12. *Nucl. Acids Res.* 10 (21), pp6957-6968.

57. Inoue, H., Nojima, H. and Okayama, H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene,* 96 (1), pp23-28.

58. Jennings, M. P., Virji, M., Evans, D. 1998. Identification of a novel gene involved in pilin glycosylation in *Neisseria meningitidis. Mol. Microbiol.* 29 (4), pp975-984.

59. Kahler, C. M., Martin, L. E., Tzeng, Y. 2001. Polymorphisms in Pilin Glycosylation Locus of *Neisseria meningitidis* Expressing Class II Pili. *Infect. Immun.* 69 (6), pp3597-3604.

60. Karlyshev, A. V., Champion, O. L., Churcher, C. 2005. Analysis of *Campylobacter jejuni* capsular loci reveals multiple mechanisms for the generation of structural diversity and the ability to form complex heptoses. *Mol. Microbiol.* 55 (1), pp90-103.

61. Karlyshev, A. V., Ketley, J. M. and Wren, B. W. 2005a. The *Campylobacter jejuni* glycome. *FEMS Microbiol. Rev.* 29 (2), pp377-390.

62. Kawamura, T. and Shockman, G. D. 1983. Purification and some properties of the endogenous, autolytic *N*-acetylmuramoylhydrolase of *Streptococcus faecium*, a bacterial glycoenzyme. *J. Biol. Chem.* 258 (15), pp9514-9521.

63. Kaya, H. K. and Gaugler, R. 1993. Entomopathogenic Nematodes. *Annu. Rev. Entomol.* 38 (1), pp181-206.

64. Keeler, J. 2005. *Understanding NMR spectroscopy.* Chichester, England ; Hoboken, NJ: Wiley.

65. Ketley, J. M. 1997. Pathogenesis of enteric infection by *Campylobacter*. *Microbiology* 143 (1), pp5-21.

66. Kilpatrick, D. C. 2002. Animal lectins: a historical introduction and overview. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1572 (2-3), pp187.

67. Konami, Y., Yamamoto, K., Osawa, T. and Irimura, T. 1995. A putative carbohydrate-binding domain of the lactose binding *Cytisus sessilifolius* anti-H(O) lectin has a similar amino acid sequence to that of the L-fucose-binding *Ulex europaeus* anti-H(O) lectin. *Glycoconj. J.* 12 (2), pp128.

68. Kovach, M. E., Elzer, P. H., Hill, D. S. 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166, pp175–176.

69. Kowarik, M., Numao, S., Feldman, M. F. 2006a. *N*-Linked Glycosylation of Folded Proteins by the Bacterial Oligosaccharyltransferase. *Science* 314 (5802), pp1148-1150.

70. Kowarik, M., Young, N. M., Numao, S. 2006b. Definition of the bacterial *N*-glycosylation site consensus sequence. *EMBO J.* 25 (9), pp1957-1966.

71. Krinos, C. M., Coyne, M. J., Weinacht, K. G. 2001. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* 414 (6863), pp555-558.

72. Ku, S. C., Schulz, B. L., Power, P. M. and Jennings, M. P. 2009. The pilin *O*-glycosylation pathway of pathogenic *Neisseria* is a general system that glycosylates AniA, an outer membrane nitrite reductase. *Biochem. Biophys. Res. Commun.* 378 (1), pp84-89.

73. Kuno, A., Uchiyama, N., Koseki-Kuno, S. 2005. Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nat. Meth.* 2 (11), pp851.

74. Kuo, C., Takahashi, N., Swanson, A. F. 1996. An *N*-linked high-mannose type oligosaccharide, expressed at the major outer membrane protein of *Chlamydia trachomatis*, mediates attachment and infectivity of the microorganism to HeLa cells. *J. Clin. Invest.* 98 (12), pp2813-2818.

75. Laemmli, U. K. 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 227 (5259), pp680-685.

76. Laine, R. A. 1997. The Information Storing Potential of the Sugar Code. *Glycosciences* By: Gabius H. J, Gabius, S. (eds), London : Weinheim : Chapman and Hall, pp1-14.

77. Lamblin, G., Degroote, S., Perini, J. 2001. Human airway mucin glycosylation: A combinatory of carbohydrate determinants which vary in cystic fibrosis. *Glycoconj. J.* 18 (9), pp661-684.

78. Langdon, R. H., Cuccui, J. and Wren, B. W. 2009. *N*-linked glycosylation in bacteria: an unexpected application. *Future Microbiology* 4 (4), pp401-412.

79. Lengeler, J.W., Drews, G. and Schlegel, H.G. 1999. *Biology of the prokaryotes.* Stuttgart : Blackwell Science.

80. Lindenthal, C. and Elsinghorst, E. A. 1999. Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infect. Immun.* 67 (8), pp4084-4091.

81. Linton, D., Allan, E., Karlyshev, A. V. 2002. Identification of *N*-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. *Mol. Microbiol.* 43 (2), pp497-508.

82. Linton, D., Dorrell, N., Hitchen, P. G. 2005. Functional analysis of the *Campylobacter jejuni N*-linked protein glycosylation pathway. *Mol. Microbiol.* 55 (6), pp1695-1703.

83. Lis, H. and Sharon, N. 1998. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. *Chem. Rev.* 98 (2), pp637-674.

84. Logan, S. M., Kelly, J. F., Thibault, P. 2002. Structural heterogeneity of carbohydrate modifications affects serospecificity of *Campylobacter* flagellins. *Mol. Microbiol.* 46 (2), pp587-597.

85. Lommel, M. and Strahl, S. 2009. Protein *O*-mannosylation: Conserved from bacteria to humans. *Glycobiology* 19 (8), pp816-828.

86. Madera, M., Mechref, Y., Klouckova, I. and Novotny, M. V. 2006. Semiautomated High-Sensitivity Profiling of Human Blood Serum Glycoproteins through Lectin

Preconcentration and Multidimensional Chromatography/Tandem Mass Spectrometry. *J. Proteome Res.* 5 (9), pp2348-2363.

87. Marceau, M., Forest, K., Beretti, J. 1998. Consequences of the loss of *O*-linked glycosylation of meningococcal type IV pilin on piliation and pilus-mediated adhesion. *Mol. Microbiol.* 27 (4), pp705-715.

88. Marth, J. D. and Grewal, P. K. 2008. Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* 8 (11), pp874-887.

89. Mattick, J. S. 2002. Type IV Pili and Twitching Motility. *Annu. Rev. of Microbiol.* 56 (1), pp289-314.

90. McGuckin, W. F. and McKenzie, B. F. 1958. An Improved Periodic Acid Fuchsin Sulfite Staining Method for Evaluation of Glycoproteins. *Clin. Chem.* 4 (6), pp476-483.

91. Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A. 2000. The Characteristics of Peptide Collision-Induced Dissociation Using a High-Performance MALDI-TOF/TOF Tandem Mass Spectrometer. *Anal. Chem.* 72 (3), pp552-558.

92. Mescher, M. F. and Strominger, J. L. 1976. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*. *J. Biol. Chem.* 251 (7), pp2005-2014.

93. Mescher, M. F., Strominger, J. L. and Watson, S. W. 1974. Protein and Carbohydrate Composition of the Cell Envelope of *Halobacterium salinarium*. *J. Bacteriol.* 120 (2), pp945-954.

94. Michell, S. L., Whelan, A. O., Wheeler, P. R. 2003. The MPB83 Antigen from *Mycobacterium bovis* Contains *O*-Linked Mannose and (1→3)-Mannobiose Moieties. *J. Biol. Chem.* 278 (18), pp16423-16432.

95. Nakamura, N., Lyalin, D. and Panin, V. M. 2010. Protein *O*-mannosylation in animal development and physiology: From human disorders to *Drosophila* phenotypes. *Semin. Cell. Dev. Biol.* 21 (6), pp622-630.

96. Nita-Lazar, M., Wacker, M., Schegg, B. 2005. The N-X-S/T consensus sequence is required but not sufficient for bacterial *N*-linked protein glycosylation. *Glycobiology* 15 (4), pp361-367.

97. North, S. J., Hitchen, P. G., Haslam, S. M. and Dell, A. 2009. Mass spectrometry in the analysis of *N*-linked and *O*-linked glycans. *Curr. Opin. Struct. Biol.* 19 (5), pp498-506.

98. Nothaft, H. and Szymanski, C. M. 2010. Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Micro.* 8 (11), pp765-778.

99. Pandhal, J. and Wright, P. 2010. *N*-Linked glycoengineering for human therapeutic proteins in bacteria. *Biotechnol. Lett.* 32 (9), pp1189-1198.

100. Pappin, D. J. C., Hojrup, P. and Bleasby, A. J. 1993. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* 3 (6), pp327-332.

101. Parge, H. E., Forest, K. T., Hickey, M. J. 1995. Structure of the fibre-forming protein pilin at 2.6 A resolution. *Nature* 378 (6552), pp32.

102. Paulson, J. C. 1989. Glycoproteins: what are the sugar chains for? *Trends Biochem. Sci.* 14 (7), pp272-276.

103. Poretz, R. D. and Goldstein, I. J. 1971. Protein-carbohydrate interaction : On the mode of binding of aromatic moieties to concanavalin A, the phytohemagglutinin of the jack bean. *Biochem. Pharmacol.* 20 (10), pp2727-2739.

104. Power, P. M. and Jennings, M. P. 2003. The genetics of glycosylation in Gram-negative bacteria. *FEMS Microbiol. Lett.* 218 (2), pp211-222.

105. Power, P. M., Roddam, L. F., Rutter, K. 2003a. Genetic characterization of pilin glycosylation and phase variation in *Neisseria meningitidis*. *Mol. Microbiol.* 49 (3), pp833-847.

106. Power, P. M., Roddam, L. F., Rutter, K. 2000. Genetic characterization of pilin glycosylation in *Neisseria meningitidis*. *Mol. Microbiol.* 49 (3), pp833-847.

107. Power, P. M., Seib, K. L. and Jennings, M. P. 2006. Pilin glycosylation in *Neisseria meningitidis* occurs by a similar pathway to *wzy*-dependent O-antigen biosynthesis in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 347 (4), pp904-908.

108. Quandt, J. and Hynes, M. F. 1993. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene* 127 (1), pp15-21.

109. Roberts, I. S. 1996. The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu. Rev. Microbiol.* 50pp285-315.

110. Samuel, G. and Reeves, P. 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr. Res.* 338 (23), pp2503.

111. Schaffer, C. and Messner, P. 2004. Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. *Glycobiology* 14 (8), pp31R-42.

112. Schirm, M., Arora, S. K., Verma, A. 2004. Structural and genetic characterization of glycosylation of type a flagellin in *Pseudomonas aeruginosa*. *J. Bacteriol.* 186 (9), pp2523.

113. Schirm, M., Soo, E. C., Aubry, A. J. 2003. Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* 48 (6), pp1579-1592.

114. Schmidt, M. A., Riley, L. W. and Benz, I. 2003. Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol.* 11 (12), pp554-561.

115. Schwan, T. G. and Piesman, J. 2002. Vector interactions and molecular adaptations of Lyme disease and relapsing fever spirochetes associated with transmission by ticks. *Emerg. Infect. Dis.* 8 (2), pp115-121.

116. Sharon, N. and Lis, H. 1990. Legume lectins-a large family of homologous proteins. *FASEB J.* 4 (14), pp3198-3208.

117. Sherlock, O., Dobrindt, U., Jensen, J. B. 2006. Glycosylation of the Self-Recognizing *Escherichia coli* Ag43 Autotransporter Protein. *J. Bacteriol.* 188 (5), pp1798-1807.

118. Shevchenko, A., Tomas, H., Havlis, J. 2007. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protocols* 1 (6), pp2856-2860.

119. Siuti, N. and Kelleher, N. L. 2007. Decoding protein modifications using top-down mass spectrometry. *Nat. Meth.* 4 (10), pp817-821.

120. Sleytr, U. B. and Thorne, K. J. 1976. Chemical characterization of the regularly arranged surface layers of *Clostridium thermosaccharolyticum* and *Clostridium thermohydrosulfuricum*. *J. Bacteriol.* 126 (1), pp377-383.

121. Smedley III, J. G., Jewell, E., Roguskie, J. 2005. Influence of Pilin Glycosylation on *Pseudomonas aeruginosa* 1244 Pilus Function. *Infect. Immun.* 73 (12), pp7922-7931.

122. Sojar, H.T., Bahl, O.P. and Victor, G.1987. [27] Chemical deglycosylation of glycoproteins. *Methods in Enzymology.* Academic Press. pp341-350.

123. Stimson, E., Virji, M., Makepeace, K. 1995. Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol. Microbiol.* 17 (6), pp1201-1214.

124. Sumner, J. B. 1919. The Globulins of the Jack Bean, *Canavalia ensiformis*. Preliminary Paper. *J. Biol. Chem.* 37 (1), pp137-142.

125. Sumner, J. B. and Howell, S. F. 1936. Identification of Hemagglutinin of Jack Bean with Concanavalin A. *J. Bacteriol.* 32 (2), pp227-237.

126. Swanson, A. F. and Kuo, C. C. 1991. Evidence that the major outer membrane protein of *Chlamydia trachomatis* is glycosylated. *Infect. Immun.* 59 (6), pp2120-2125.

127. Szymanski, C. M., Burr, D. H. and Guerry, P. 2002. *Campylobacter* Protein Glycosylation Affects Host Cell Interactions. *Infect. Immun.* 70 (4), pp2242-2244.

128. Szymanski, C. M., Logan, S. M., Linton, D. and Wren, B. W. 2003. *Campylobacter* - a tale of two protein glycosylation systems. *Trends Microbiol.* 11 (5), pp233-238.

129. Szymanski, C. M. and Wren, B. W. 2005. Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Micro.* 3 (3), pp225-237.

130. Szymanski, C. M., Yao, R., Ewing, C. P. 1999. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* 32 (5), pp1022-1030.

131. Taylor, M.E. and Drickamer, K. 2006. *Introduction to glycobiology.* 2nd. Oxford ; New York: Oxford University Press.

132. Thibault, P., Logan, S. M., Kelly, J. F. 2001. Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *J. Biol. Chem.* 276 (37), pp34862-34870.

133. Twine, S. M., Paul, C. J., Vinogradov, E. 2008. Flagellar glycosylation in *Clostridium botulinum*. *FEBS J.* 275 (17), pp4428-4444.

134. Twine, S. M., Reid, C. W., Aubry, A. 2009. Motility and Flagellar Glycosylation in *Clostridium difficile*. *J. Bacteriol.* 191 (22), pp7050-7062.

135. University of Bath 2004. (News). [Online]. Available from : <http://www.bath.ac.uk/pr/releases/humaninfectionsfrominsects.htm> [Accessed 21 January 2008].

136. Upreti, R. K., Kumar, M. and Shankar, V. 2003. Bacterial glycoproteins: Functions, biosynthesis and applications. *Proteomics* 3 (4), pp363-379.

137. USDA Agricultural Research Service. [Online]. Available from : <http://www.ars.usda.gov/is/kids/insects/story12/nematodes.htm> [Accessed 21 January 2008].

138. Varki, A. 1999. *Essentials of glycobiology.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

139. Verma, A., Schirm, M., Arora, S. K. 2006. Glycosylation of b-Type Flagellin of *Pseudomonas aeruginosa*: Structural and Genetic Basis. *J. Bacteriol.* 188 (12), pp4395-4403.

140. Vik, A., Aas, F. E., Anonsen, J. H. 2009. Broad spectrum *O*-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae. Proc. Natl. Acad. Sci. U.S.A.* 106 (11), pp4447-4452.

141. Virji, M., Kayhty, H., Ferguson, D. J. P. 1991. The role of pili in the interactions of pathogenic *Neisseria* with cultured human endothelial cells. *Mol. Microbiol.* 5 (8), pp1831-1841.

142. Wacker, M., Feldman, M. F., Callewaert, N. 2006. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci. U.S.A.* 103 (18), pp7088-7093.

143. Wacker, M., Linton, D., Hitchen, P. G. 2002. *N*-Linked Glycosylation in *Campylobacter jejuni* and its Functional Transfer into *E. coli. Science* 298 (5599), pp1790-1793.

144. Walsh, G. and Jefferis, R. 2006. Post-translational modifications in the context of therapeutic proteins. *Nat. Biotech.* 24 (10), pp1241-1252.

145. Wang, Y., Wu, S. and Hancock, W. S. 2006. Approaches to the study of *N*-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap-Fourier transform mass spectrometry. *Glycobiology* 16 (6), pp514-523.

146. Ward, A., Sanderson, N. M., O'Reilly, J. 2000. The amplified expression, identification, purification, assay, and properties of hexahistidine-tagged bacterial membrane transport proteins. *Membrane transport—a practical approach.* Oxford, United Kingdom : Oxford University Press, pp141-166.

147. Waterfield, N. R., Ciche, T. and Clarke, D. 2009. *Photorhabdus* and a Host of Hosts. *Annu. Rev. Microbiol.* 63 (1), pp557-574.

148. Weerapana, E., Glover, K. J., Chen, M. M. and Imperiali, B. 2005. Investigating Bacterial *N*-Linked Glycosylation: Synthesis and Glycosyl Acceptor Activity of the Undecaprenyl Pyrophosphate-Linked Bacillosamine. *J. Am. Chem. Soc.* 127 (40), pp13766-13767.

149. Weerapana, E. and Imperiali, B. 2006. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* 16 (6), pp91R-101.

150. Wehmeier, S., Varghese, A. S., Gurcha, S. S. 2009. Glycosylation of the phosphate binding protein, PstS, in *Streptomyces coelicolor* by a pathway that resembles protein *O*-mannosylation in eukaryotes. *Mol. Microbiol.* 71 (2), pp421-433.

151. Werner, R. G., Kopp, K. and Schlueter, M. 2007. Glycosylation of therapeutic proteins in different production systems. *Acta Paediatr.* 96, pp17-22.

152. Whitfield, C. 1995. Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol.* 3 (5), pp178-185.

153. Young, N. M., Brisson, J., Kelly, J. 2002. Structure of the *N*-Linked Glycan Present on Multiple Glycoproteins in the Gram-negative Bacterium, *Campylobacter jejuni. J. Biol. Chem.* 277 (45), pp42530-42539.

154. Zarembinski, T. I., Hung, L., Mueller-Dieckmann, H. 1998. Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics. *Proc. Natl. Acad. Sci. U.S.A.* 95 (26), pp15189-15193.

155. Zeituni, A. E., McCaig, W., Scisci, E. 2010. The Native 67-Kilodalton Minor Fimbria of *Porphyromonas gingivalis* Is a Novel Glycoprotein with DC-SIGN-Targeting Motifs. *J. Bacteriol.* 192 (16), pp4103-4110.

156. Zhang, L. and Reilly, J. P. 2009. Extracting both peptide sequence and glycan structural information by 157 nm photodissociation of *N*-linked glycopeptides. *J. Proteome Res.* 8 (2), pp734-742.

157. Zhang, Y., Go, E. P. and Desaire, H. 2008. Maximizing Coverage of Glycosylation Heterogeneity in MALDI-MS Analysis of Glycoproteins with Up to 27 Glycosylation Sites. *Anal. Chem.* 80 (9), pp3144-3158.