Université de Montréal

Proteomic Approaches for the Detection of Unusual Post-Translational Modifications in Simple and Complex Bacterial Protein Mixtures

par
Michael Schirm

Département de chimie
Faculté des Arts et des Sciences

Thèse présentée a la Faculté des Études Supérieures
En vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
En Chimie

Octobre 2004

Université de Montréal
Faculté des Études Supérieures

Cette thèse intitulée:

Proteomic Approaches for the Detection of Unusual Post-Translational Modifications in Simple and Complex Bacterial Protein Mixtures

Présentée par:
Michael Schirm

A été évaluée par un jury composé des personnes suivantes:

Pr. Jeffrey W. Keillor....., président-rapporteur
Pr. Pierre Thibault....., directeur de recherche
Pr. Karen C. Waldron....., codirectrice de recherche
Pr. Marius D'Amboise....., membre du jury
Dr. John Kelly....., examinateur externe
Pr. France Daigle....., représentante du doyen de la FES

Thèse acceptée le:.....21/02/05.....

ABSTRACT

The research conducted in this dissertation is separated in three parts. In the first part the flagellin proteins from different human pathogens were analyzed for protein glycosylation. Studies on flagellin from *Helicobacter pylori, Listeria monocytogenes*, *Aeromonas caviae*, *Pseudomonas aeruginosa* and *Campylobacter jejuni* showed that prokaryotic protein glycosylation is a common theme and much more prevalent than previously thought. A novel approach using targeted capillary liquid chromatography coupled to tandem mass spectrometry (capLC/MS/MS) was developed and revealed that the flagellin from these bacteria were all post-translationally modified with O-linked glycans. Extensive structural analyses of these flagellin glycans showed the occurrence of different carbohydrate residues, indicating the use of different biosynthetic substrates and glycosyl transferases. It is noteworthy that *H. pylori*, *A. caviae* and *C. jejuni* all used substrates structurally related to pseudaminic acid, a novel nine-carbon sugar related to sialic acid, while in *L. monocytogenes* the flagellin was glycosylated with a single N-acetyl glucosamine. The nature of the glycosyl modifications varied significantly between different strains of *P. aeruginosa*, with a heterogeneous glycan comprising up to 11 monosaccharide units for strain PAK while in strains JJ692 and PAO, the glycans were identified as rhamnose and an unknown carbohydrate residue having a mass of 356 Da, respectively. Furthermore, genes involved in this flagellar glycosylation process were identified for *H. pylori*, *P. aeruginosa* and *L. monocytogenes* and their functional relationships were elucidated from mass spectrometry analyses.

The complete characterization of post-translational modifications and characterization of genes involved in this process is, a tedious and time consuming process even for simple protein mixtures that can take from several months to years. To simplify this approach, the second part of this thesis focused on developing a novel "top down" mass spectrometry approach to allow the rapid identification and characterization of post-translational modifications on intact proteins from simple protein mixtures. Furthermore, it enabled the rapid assignment of unknown gene functions by monitoring changes imparted to the sugar modifications in isogenic mutants derived from insertional inactivation of the individual genes involved in the flagellin glycosylation process. Compared to the traditional "bottom up" approach, the analysis time was significantly reduced.

To identify novel unknown glycoproteins in *H. pylori,* complex protein extracts were analysed in the third part of this dissertation. This analysis permitted the identification of approximately 875 unique proteins. More than 35000 MS/MS spectra were acquired and screened for potential glycopeptides. These analyses suggested that glycosylation is a relatively rare event primarily observed for FlaA, FlaB and the flagellar hook protein. These proteins were identified to be glycosylated with Pse5Ac7Ac. Offline two dimensional (2D)-capLC/MS/MS and capLC/MS/MS of gel-isolated proteins were used for the analysis of the bacterial extracts and the results indicated that the 2D-capLC/MS/MS approach was more efficient to detect these types of modified peptides, due to the enrichment of glycopeptides in specific salt fractions compared to unmodified peptides.

Keywords: post-translational modifications, mass spectrometry, capLC/MS prokaryotic protein glycosylation, flagellin

Sommaire

Le projet de recherche faisant l'objet de cette dissertation est séparé en trois parties. La première partie décrit l'étude de la glycosylation de flagellines extraites de différents pathogènes humains. Cette étude a été effectuée sur les flagellines obtenues des bacteries *Helicobacter pylori, Listeria monocytogenes, d'Aeromonas caviae*, *Pseudomonas aeruginosa et Campylobacter jejuni* et a demontré que la glycosylation des protéines procaryotes était beaucoup plus abondantes qu'initialement anticipée.

Une nouvelle approche utilisant la chromatographie liquide capillaire couplée à la spectrometrie de masse en tandem (capLC/MS/MS) a été développée et a permis de démontrer que les flagellines de ces bactéries ont subi des modifications post-traductionelles mettant en évidence la glycosylation ciblée d'acides amines sérine et thréonine. L'analyse structurale de ces modifications a montré la présence de différents hydrates de carbone dans plusieurs souches bactériennes, indiquant ainsi l'utilisation de différents substrats biosynthétiques et de glycosyl-transférases. Il est à noter que les bactéries *H. pylori*, *A. caviae* et *C. jejuni* utilisent toutes des substrats analogues à l'acide pseudaminique, un sucre de 9 carbones similaire à l'acide sialique, alors que les flagellines de *L. monocytogenes* sont glycosylées par un seul N-acétyl glucosamine. La nature de la glycosylation varie significativement entre les différentes souches de *P. aeruginosa.* En effet, la souche PAK comprend un glycan hétérogène comptant jusqu'à 11 monosaccharides, tandis qu'on retrouve dans les souches JJ692 et PAO le rhamnose et un hydrate de carbone inconnu ayant une masse

de 356 Da. De plus, les gènes impliqués dans le procédé de glycosylation des flagellines ont été identifiés dans *H. pylori*, *P. aeruginosa* et *L. monocytogenes* et leurs relations fonctionnelles ont été élucidées par spectrométrie de masse.

La caractérisation complète des modifications post-traductionelles et des gènes impliqués dans ce procédé est complexe et peut prendre des mois ou même des année, et ce pour un mélange simple de protéines. Afin de simplifier cette tâche, la seconde partie de cette thèse décrit une nouvelle approche spectrométrique permettant l'identification et la caractérisation rapide de modifications post-traductionelles directement sur les protéines intactes à partir de mélanges simples. De plus, elle permet d'assigner rapidement la fonction de gènes inconnus en observant les changements survenant chez les sucres issus de dérivés d'inactivation insertionnelle du gène individuel impliqué dans la glycosylation des flagellines. Lorsque comparé à l'approche traditionnelle, le temps d'analyse est significativement réduit avec l'utilisation de cette technique.

Afin d'identifier de nouvelles glycoprotéines dans *H. pylori*, des extraits protéiques complexes ont été analysés dans la troisième partie de cette dissertation. Cette analyse a permis l'identification d'approximativement 875 protéines uniques. Plus de 35000 spectres MS/MS ont été obtenus et examinés afin d'identifier des glycoprotéines potentielles. Ces analyses ont suggéré que la glycosylation est relativement rare et est principalement observée sur les proteines FlaA, FlaB ainsi que sur la protéine flagellaire crochet FlgE. Des analyses bi-dimensionelle (2D)-capLC/MS/MS et capLC/MS/MS de protéines séparées sur gel ont été utilisées pour analyser des extraits bactériens. Ces analyses ont révélées que l'approche 2D-

capLC/MS/MS était plus efficace pour détecter ce type de peptide modifié, à cause de l'enrichissement des glycopeptides dans des fractions de sel spécifiques permettant ainsi de cibler plus facilement l'identification de ces peptides modifiés.

Mots clés: modifications post-traductionelles, spectrométrie de masse, capLC/MS, flagellin, glycosylation des protéines procaryotes

Table of Content

List of Figures

List of Tables

ABBREVIATIONS AND SYMBOLS

2D	two dimensional
Ab	antibody
Asn	asparagine
capLC/MS/MS	capillary liquid chromatography mass spectrometry
CHO	carbohydrate
CID	collision induced dissociation
DHA	dehydroalanine
DHB	dehydrobutyric acid
dHex	deoxyhexose
dHexN	deoxyhexosamine
ECD	electron capture dissociation
ER	endoplasmatic reticulum
ES	electrospray
ESMS	electrospray mass spectrometry
FA	formic acid
Fuc	fucose
Gal	galactose
GalNAc	*N*-acetyl-D-galactosamine
GC	gas chromatography
GI	glycosylation island
Glc	Glucose

GlcNAc	*N*-acetyl-D-glucosamine
Hex	hexose
HexA	Hexuronic acid
HPLC	high-performance liquid chromatography
i.d.	Internal diameter
LC	liquid chromatography
LPS	lipopolysaccharide
[M+H]+	protonated molecule
m/z	mass-to-charge ratio
MALDI	matrix assisted laser desorption/ionization
MALT	mucosa-associated lymphoid tissue
Man	mannose
MCPs	methyl-accepting chemo taxis proteins
MS	mass spectrometry
MS/MS	tandem mass spectrometry
Neu	neuraminic acid
NMR	nuclear magnetic resonance
ORF	open reading frame
Pse5Ac7Ac	5,7-diamino-3,5,7,9-tetradeoxy-L-glycero-L-manno-nonulosonic acid; N-acetyl pseudaminic acid
PseAm	acetamidino form of pseudaminic acid
PTM	post-translational modification
PPE	periplasmic extract

Q	quadrupole
Q-TOF	quadrupole-time of flight
Sia	sialic acids
SCX	strong cation exchange chromatography
SDS-PAGE	sodium dodecyl sulphate - polyacrylamide gel electrophoresis
Ser	serine
Thr	threonine
TLR	toll-like receptor
Xyl	xylose
Wma	wrong mass assignment

Acknowledgments

I am most grateful to my two supervisors, Prof. Dr. Pierre Thibault and Prof. Dr. Karen Waldron, for their wise guidance, encouragement and their support throughout this work. Without their help this work would not have been possible.

I would like to thank Dr. Susan Logan for the great collaboration during the last three years, for the preparation of the samples, her advice and the valuable discussions.

I am very grateful to my colleagues, Dr. Alexandra Furtos-Matei, Dr. Evelyn Soo, Dr. Eric Bonneil, Karine Venne, Genevieve Mercier, Sylvain Tessier, Pascal Fex, Isabelle Migneault, Anik Forest, Sylvie Plante, Dmitri Sitnikov, Rachel Page-Belanger, Kevin Eng, Jason Yen, Carl Hemond, Jean-Louis Cabral, Jean-François St-Amant and Cathy Dartiguenave, for their support, advice and inspiration.

I want to thank the reviewers of this thesis, Prof. Dr. Jeffrey Keillor, Prof. Dr. Marius D'Amboise and Dr. John Kelly for their valuable comments, which led to considerable improvement of this text.

In addition, I would like to thank Caprion Pharmaceutical for financial support and for giving me access to mass spectrometry instrumentation.

Finally, but not least, I would like to thank my parents, Richard and Agnes Schirm, for their never-ending support.

1 Aims and scope

Knowledge about prokaryotic protein glycosylation is very limited and the full identification and characterization of the diverse glycans found on prokaryotic proteins is currently lacking. Due to glycan diversity amongst bacterial species and the often limiting quantity of individual glycoproteins, characterization is currently a very tedious and time consuming procedure. As a consequence of this and to develop more useful analytical tools the goal of this dissertation was to:

- Improve and develop new mass spectrometry methods for the rapid identification and characterization of protein glycosylation in simple protein mixtures from several important bacterial pathogens.
- Apply these new and improved methods to the identification and characterization of post-translational modifications (PTMs) on the flagellin protein of several important human pathogens.
- Develop simple and fast mass spectrometry methods to monitor changes in glycosylation profiles of isogenic mutants, enabling the functional characterization of selected genes implicated in the glycosylation process.
- Analyze complex protein mixtures by using techniques like SDS-PAGE and ion exchange chromatography followed by capLC/MS/MS to identify novel unknown glycoproteins in the bacterial pathogen *Helicobacter pylori*.

2 Review of the literature

2.1 Glycosylation

2.1.1 Eukaryotic glycosylation

Glycosylation is the addition of monosaccharides or polysaccharides to molecules such as proteins. The donor molecules for glycosylation are activated nucleotide sugars which are transferred onto the polypeptides via glycosyl transferases. Glycosylation is probably the most common post-translational modification in eukaryotic cells. It is estimated that around 50 % of all proteins are glycosylated.[1] In addition glycoconjugates represent the most structurally and functionally diverse molecules in nature and therefore the analysis of glycoproteins is a very challenging task.

In eukaryotes, four different kinds of glycoproteins exist, namely N-linked glycans, O-linked glycans, proteoglycans and GPI anchors. The focus in this dissertation will be the N-linked and O-linked glycans of bacteria. N-linked glycans are the most studied form of glycosylation. In eukaryotes, they occur in the endoplasmatic reticulum (ER) typically by addition of $Glc_3Man_9GlcNAc_2$ to the polypeptide via the enzyme oligosaccharyl transferase.[3] This process is site specific and only asparagines with the sequence Asn-X-Ser or Asn-X-Thr can be

glycosylated, where X can be any amino acid, but is generally not proline or aspartate. During processing the glycan is trimmed to $Man_5GlcNAc_2$ and from there additional sugars are added forming three major classes of N-glycans called High Mannose, Complex and Hybrid. All glycan classes have a common core structure of $Man_3GlcNAc_2$ to which additional sugars are added forming the three classes. An example of the three classes is shown in figure 2.1.

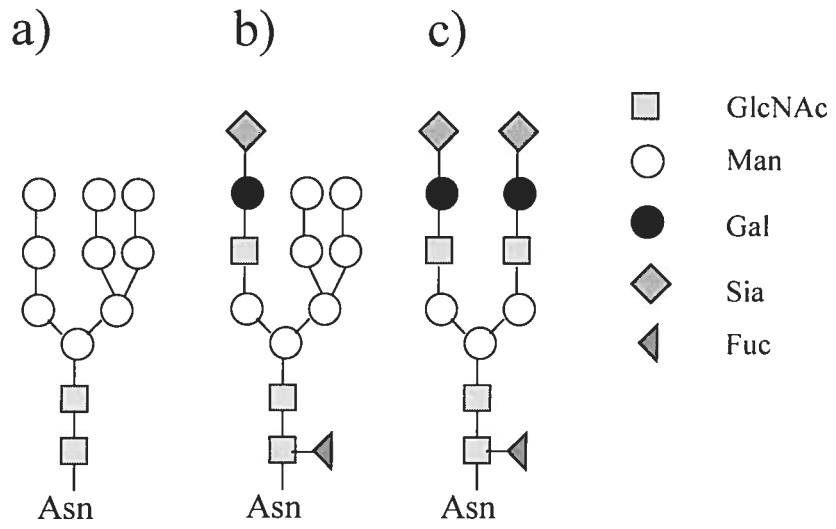

Figure 2.1: Major classes of N-glycans. a) "High-Mannose", b) "Hybrid" c) "Complex"

Sugars commonly found in N-linked glycans are D-Galactose (Gal), D-Glucose (Glc), D-Mannose (Man), *N*-Acetyl-D-Glucosamine (GlcNAc), ß-L-Fucose (Fuc), neuraminic Acid (Neu) and *N*-Acetyl-Neuraminic Acid (Neu5Ac). N-linked glycosylation plays an important role in protein folding, protein stability, turn over rate of proteins and in certain diseases, e.g. cancer.[4-9]

In contrast to N-linked glycosylation, in eukaryotic systems O-glycans comprise GalNac directly connected to the peptide backbone via the hydroxyl group of the amino acids serine or threonine. O-linked glycosylation occurs at a later stage

during protein processing and takes place in the Golgi apparatus. In contrast to N-linked glycosylation, O-linked sugars are added individually to Ser or Thr starting with GalNAc, transferred through the O-GlcNAc transferase and more sugars can be added afterwards.[10] O-linked glycosylation can also take place in the cytoplasm or nucleus but there, only a single sugar is added. A consensus sequence for O-linked glycosylation has not been identified so far, making the identification of the exact linkage site a challenging task. O-linked glycans play an important role in the regulation of transcription, translation, nuclear transport, cytoskeletal assembly, the cell cycle, diabetes, and in the regulation of protein turnover.[11-15]

2.1.2 Prokaryotic glycosylation

While it was originally thought that glycosylation was restricted to eukaryotes, protein glycosylation is now a well established fact in prokaryotic glycobiology with substantial evidence demonstrating that O and N glycosylation does occur.[16-20] A total list of identified glycoproteins is beyond the scope of this thesis and interested readers are referred to the three recent excellent reviews of Schmidt,[21] Benz[22] and Messner.[23]

The first prokaryotic glycoproteins identified and still today representing the most intensely studied, are the S-layer proteins in archae and bacteria. More recently, flagellar proteins, pili, fimbriae, extracelluar proteins and cell-envelope associated proteins from both archae and bacteria have been reported to be glycosylated. Among these bacteria are several important human pathogens such as *Campylobacter jejuni,*[24] *Chlamydia trachomatis,*[25] *Escherichia coli*[26] and *Mycobacterium tuberculosis*[27] (see Table 2.1 for more information). However, only limited information on the structural composition of bacterial glycans is available for many species. In comparison, protein glycosylation of the gastrointestinal pathogen *Campylobacter jejuni* has been extensively studied. Two unique glycosylation pathways have been characterised in this organism. Thibault and co-workers[24] showed in 2000 that the major structural protein, flagellin, is O-glycosylated with N-acetyl pseudaminic acid and derivatives. Additionally, in 2003 an N-linked glycosylation system, which uses the consensus sequence typical in eukaryotic systems (Asn-X-Thr/Ser) was reported by Young et. al.,[28] who identified more than 30 potential glycoproteins. These proteins were

glycosylated with a heptasaccharide of the structure GalNAc-α1,4-GalNAc-α1,4-[Glcβ1,3]GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-α1,*N*-Asn.[28] Interestingly, unlike eukaryotic N-linked glycosylation, the linkage sugar in this novel prokaryotic system is a deoxyhexose sugar unique to prokaryotes, bacillosamine.

While only limited information currently exists for eubacterial glycoproteins, it is already clear that the diversity in structure far exceeds the rather limited display in eukaryotes.[22] "Unusual" carbohydrates not common in eukaryotes have been identified in prokaryotic systems, ranging from single monosaccharides like GlcNAc, Glc, Gal, Xyl and Man residues to derivatives of pseudaminic acid, a novel 9 carbon sialic acid like sugar (Table 2.1). In many cases these novel carbohydrates are not common to eukaryotic systems, and so provide novel targets for therapeutic intervention. The biological role of glycosylation in prokaryotes is in most cases currently unknown but since many of the identified glycoproteins are surface associated in bacterial pathogens it seems likely that they will play a role in host/pathogen interactions.

Table 2.1 Examples of glycoproteins found in prokaryotic pathogens. Adapted from Benz et al.[22]

Organism of	Identified carbohydrates	Protein	Linkage	Function of carbohydrate modification
Borrelia burgdorferi	GlcNAc	OspA, OspB, minor proteins	Asn	Unknown
Campylobacter coli	Derivatives of pseudaminic acid	Flagellin	Ser/Thr	Unknown
Campylobacter jejuni	Derivatives of pseudaminic acid	Flagellin	Ser/Thr	Unknown
Campylobacter jejuni	GalNAc-α1,4-GalNAc-α1,4-[Glcβ1,3]GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-α1	> 30 proteins	Asn	Unknown
Chlamydia trachomatis	High mannose	40 kDa MOMP and minor proteins	Asn	Essential for attachment
Ehrlichia chaffeensis	Glc, Gal, Xyl	P120	Ser/Thr	Unknown
Ehrlichia canis	Glc, Gal, Xyl	P140	Ser/Thr	Unknown
Escherichia coli	Heptose	TibA, AIDA-I	Ser/Thr	Essential for stability and adherence.
Mycobacterium tubercolosis	α-(1-2)-Man1-3	45/47 kDa and 38/19 kDa	Thr	Prominent antigens for humoral and celluar immune responses
Pseudomonas aeruginosa	α-5NßOHC$_4$7NFmPse-(2→4)-ß-Xyl-(1→3)-ß-FucNAc	Pilin	Ser	Unknown

2.2 Flagella

The eubacterial flagellum consists of the helical filament (propeller), the hook, the rod, the L-P ring, the SMC ring complex and the motor. A scheme of the flagellum is shown in figure 2.2. The flagellum is the organelle for locomotion. It can be up to

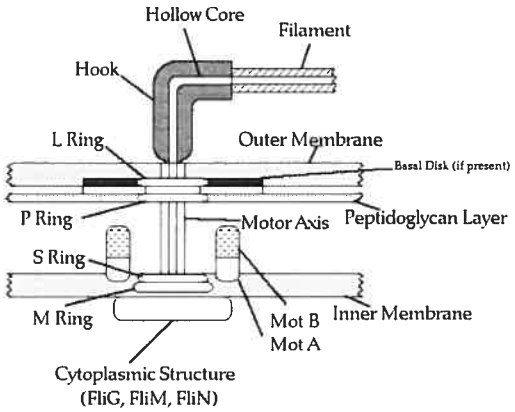

Figure 2.2: Scheme of the flagellum[30]

15 µm long with a diameter of 120-250 Å[29] and can rotate clockwise and counter clockwise.[31] Clockwise rotation results in a tumbling motion of the bacteria, causing a change in the direction of movement of the bacteria. Counter clockwise rotation leads to long, straight or curved runs without a change in direction. This motility helps the bacteria to stay in optimum medium via taxis. Taxis is the motile response to an environmental stimulus e.g. light, chemicals, temperature. It is regulated by special receptors, for example chemo taxis is regulated by chemo receptors located in the cytoplasmic membrane or periplasm of the bacterium. These receptors bind chemical attractants or repellents, which leads to either the methylation or demethylation of methyl-accepting chemo taxis proteins (MCPs), triggering either a counter clockwise or clockwise rotation of the flagellum.[32]

The filament of the flagella is a polymer consisting only of flagellin proteins and the crystal structure of the flagellar filament has recently been identified for Salmonella.[33] A scheme of the filament is shown in figure 2.3. From the crystal structure the flagellin protein can be separated into 4 domains (D0,D1,D2,D3).

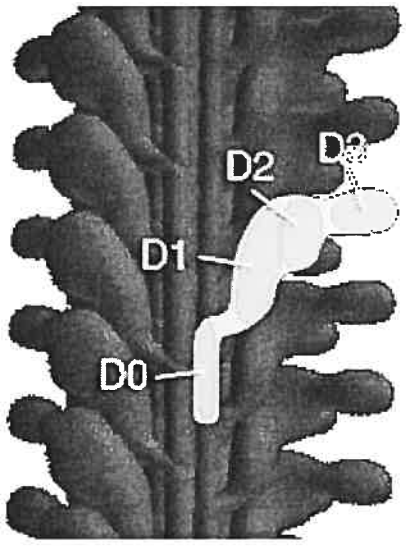

Figure 2.3: Structure of the filament[33]

Domain D1 forms the outer core of the filament and domains D2 and D3 form the projection on the filament surface (Fig. 2.3).[33]

Flagellin is an immunodominant protein that is recognized by toll-like receptor 5. Toll like receptors (TLR) bind to bacterial factors such as LPS, bacterial lipoproteins and flagella. Members of the TLR gene family convey signals stimulated by these factors, activating signal transduction pathways that result in transcriptional regulation and stimulate immune function.[34,35]

2.3 Bacterial pathogens

2.3.1 General description

Bacteria belong to the family of prokaryotes and are single-cell microorganisms. Their DNA is unbound in the cytoplasm and they possess no nuclear membrane, unlike eukaryotic cells. The size of prokaryotic cells is small (~1 μm) when compared to eukaryotic cells. (~10-100 μm). Prokaryotic cells consist of three different architectural regions. The appendages (proteins attached to the cell surface e.g. Flagella and pili), a cell envelope consisting of a cell wall, capsule and a plasma membrane, and a cytoplasmic region containing the cell genome, the ribosome and different sorts of inclusion bodies. Bacteria are furthermore separated into two groups, Gram-positive and Gram-negative, depending on their cell wall. Gram-positive bacteria have a thick cell wall consisting of murein. Murein is a polymer of disaccharides cross-linked by short amino acid chains. For Gram-negative bacteria the cell wall is relatively thin and consists only of a small layer of murein surrounded by the outer membrane. All of the bacteria analysed in this thesis were Gram-negative with the exception of *L. monocytogenes*, which is a Gram-positive bacterium. Some bacteria are pathogens and causes disease in humans, animals and plants. Important diseases caused by bacteria are for example tuberculosis, diphtheria and tetanus. Bacterial pathogens have special structural or biochemical properties that determine their virulence or pathogenicity. Important factors are the ability to colonize and

invade their host, the ability to resist or withstand the antibacterial defenses, and to produce various toxic substances that damage the host.[36]

2.3.2 Importance of motility in colonization of host

Motility as mentioned above is very important for pathogens and most pathogens are motile by means of flagella. Motility gives the bacteria the possibility to invade tissues and gives the ability to bypass or overcome the host defence mechanisms. Studies showed that less motile strains of the same bacteria were less able to colonize and survive in the host.[37-38] Motility is important for colonization of tissues. For example, the mucosal surfaces of the bladder or the intestines are constantly flushed, to avoid bacteria colonization. However, motile bacteria are able to swim away from this constant flow and reach the mucosal surfaces where they then make contact with the mucous membranes and are able to colonize the epithelial cell layer.

2.3.3 Selection of bacterial pathogens

Five bacterial species were analysed in this dissertation for post-translational modifications, namely, *Helicobacter pylori*, *Campylobacter jejuni*, *Pseudomonas aeruginosa*, *Listeria monocytogenes* and *Aeromonas caviae*. In this section, information about each bacteria and previous research done on glycosylation will be described.

2.3.3.1 *Helicobacter pylori*

Helicobacter pylori is a motile, Gram-negative, microaerophilic, spiral shaped organism that colonizes the stomachs of at least half the world's population. While most infected individuals are asymptomatic, a significant number develop a more serious pathology. *H. pylori* is the causative agent of chronic type B gastritis, a prerequisite for duodenal ulcers and more recently the organism has been associated with mucosa-associated lymphoid tissue (MALT) and with B-cell MALT lymphomas. Helicobacter species demonstrate highly efficient flagellar motility under conditions of elevated viscosity such as that found in the gastric lumen. Helicobacter cells possess a unipolar bundle of sheathed flagella. The complex filament is comprised of two flagellin subspecies, the more abundant FlaA protein which comprises the majority of the filament, and a second slightly larger protein, FlaB which appears to be exclusively located proximal to the hook within the assembled

filament.[39] *Helicobacter pylori* inhabits the mucus adjacent to the gastric mucosa and is able to survive in the acid environment of the stomach, because of urease, one of the highly expressed proteins in *H. pylori*. Urease converts urea to ammonium and bicarbonate, neutralizing gastric acid providing protection until it reaches the submucosal layer with its neutral pH value.

Josenhans provided evidence in 1999 that the flagellin of *H. felis* and other strains of *Helicobacter* are glycosylated by using periodic acid treatment and subsequent labeling of flagellin with dioxygenin-tagged hydrazine.[40] More recently, Josenhans et al. (2002) identified three genes (neuA, flmD and flmH) which were shown to be involved in the glycosylation of flagellin, although the functional basis of this proposal remained unclear. [41]

2.3.3.2 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is an opportunistic Gram-negative pathogen that causes life threatening infections in immunocompromised individuals and burn wound victims and chronic infections in patients with cystic fibrosis.[42] The organism is ubiquitous, being found in soil, marshes and costal marine habitats as well as on plant and animal tissues. Pseudomonas are motile by means of a simple flagella comprised of a single structural protein FliC. *P. aeruginosa* can be classified into two serogroups based on flagellin protein: type a or type b, depending on their molecular weight and reaction with specific antisera.[43] Type a flagellins have been shown to be hetererolgous by SDS-PAGE and to have molecular masses ranging from 42 kDa to

52 kDa, while type b flagellins appear to have a conserved sequence and a molecular mass of 53 kDa. The flagella of *P. aeruginosa* strains has been shown to be an important virulence factor, and non motile mutants are severely attenuated in the ability to cause disease in animal models.[44-45] The first evidence of protein glycosylation in Pseudomonas was presented in 1995 by Castric et al. who showed that the pilin of strain 1244 might be glycosylated due to an increased migration pattern in SDS-PAGE.[46] However, the pilin failed to react with sugar sensitive reagents. In 2001 the structure of the glycan was identified. Castric and co-workers[47] showed that the pilin in strain 1244 is glycosylated with 5-N-β-hydroxybutyryl-7-N-formyl-pseudaminic acid-(2,4)-β-Xyl-(1,3)-b-FucNAc trisaccharide linked to a Ser residue. This glycan is identical with the O-antigen of the LPS.

Preliminary evidence for post-translational modification of *Pseudomonas* flagellin has been provided previously. Variation in the observed molecular weight by SDS-PAGE of the FliC protein from the predicted sequence was the first evidence for post-translational modification.[48] Subsequently, Brimer and Montie verified that a-type flagellins were indeed glycosylated.[49] Most recently, a unique genomic island was identified in strains producing a-type flagellin which contained a cluster of 14 genes.[50] Coinheritance of a-type flagellin with this cluster of genes led to protein glycosylation. Based on limited homologies suggesting that the genes from the glycosylation island encoded enzymes involved in the synthesis, activation or polymerization of sugars necessary for flagellin glycosylation, it was proposed that the coinheritance of a-type flagellin with this cluster of genes led to protein glycosylation. Most recently, a comprehensive study on genetic variability within

glycosylation islands and flagellar genes amongst a-type strains has been completed.[51] These studies indicated that antigenic diversity of flagellin protein is likely influenced by glycosylation. However, the nature and type of post-translational modification remained unknown. In this thesis flagellin from strains PAK (a-type), jj692 (a-type) and PAO (a b-type flagellin that is thought not to be glycosylated) were analysed.

2.3.3.3 *Listeria monocytogenes*

L. monocytogenes is a Gram-positive bacterium responsible for listeriosis, a severe food borne infection associated with the consumption of contaminated foods. Clinically, symptoms of listeriosis include meningitis, septicemia, spontaneous abortion, perinatal infections and gastroenteritis.[52] Listeria species are found throughout the food processing environment[53-55] and can survive adverse conditions such as high salt and both pH[56] and temperature extremes.[57] However, the organism does not survive thermal processing; contamination of food generally results from contact of a processed food with a spoiled surface prior to packaging[58] or through the use of contaminated ingredients in minimally processed foods. Post processing contamination remains an important issue, since *L. monocytogenes* can proliferate within a contaminated food product at refrigeration temperatures.[58]

In *L. monocytogenes*, motility is a thermo-regulated phenotype.[59] At temperatures less than 30°C the organism is flagellated with 4-6 peritrichous flagella. However, motility is reduced as growth temperatures increase above 30°C. At 37°C

transcription of *flaA* appears to be shut down[59] although minor quantities of flagellin can be detected immunologically.

The serotyping of *L. monocytogenes* strains is based on agglutination reactions utilizing both "O" somatic and "H" flagellar antigens.[60] The flagellar "H" antigen correlates with the recognized pulse-field fingerprinting determined genomic divisions where 1/2a and 1/2c strains fall within Division I, and the 1/2b and 4b within Division II.[61] The identification of H-antigen type is based on differences in the agglutination profile using a series of four cross-adsorbed polyvalent antisera (A, AB, C and D), and allows the classification of the isolate into one of three groups having shared H-antigens (AB, ABC, and BD).[60] For example, strains falling within the 4b and 1/2b serotypes agglutinate with H-antisera A, AB, and C, whereas 1/2a serotypes agglutinate with only the A and AB antisera. However, the structural basis underlying the H-serotype specificity currently remains unknown.

Previous studies have indicated that Listeria flagellin undergoes some form of post-translational modification.[59,62] Peel et al.[59] found that the flagella could be separated into multiple bands following extended SDS-polyacrylamide gel electrophoresis and these bands display antigenic heterogeneity. In contrast, Dons *et al.*[62] determined that only a single flagellin gene was present within the genome, and noted a discrepancy between the predicted and actual MW of the *L. monocytogenes* flagellin. Strains Clip 23485, 2568, 568 and 394 were analysed in this thesis.

2.3.3.4 *Campylobacter jejuni*

Campylobacter jejuni is a Gram-negative slender, curved, and motile rod bacteria. It is motile by means of flagella. The bacteria is an important enteric pathogen and is the leading cause of gastrointestinal illness in the United States. *C. jejuni* causes diarrhea with fever and abdominal pain. Infections occurs most commonly through consumption of contaminated foods.[63]

Logan and co-workers presented in 1989 the first evidence that the flagellin protein is post-translationally modified[64] and Doig et al.[65] showed that the flagellin is sensitive to periodate oxidation and binds to a sialic specific lectin, indicating that the flagellin protein is glycosylated. The structural analysis of the flagellin protein was completed by Thibault et al.[24] in 2001 and showed that the flagellin of *C. jejuni* is O-linked glycosylated on up to 19 sites with pseudaminic acid and related derivatives. The major modifications are N-acetyl pseudaminic acid (Pse5Ac7Ac) and an acetamidino substituted pseudaminic acid (PseAm). Minor amounts of a dihydroxyproprionyl form (Pse5Pr7Pr) and an O-acetylated form (Pse5Ac7Ac8OAc) were reported too.

Very recently, Young and co-workers[28] identified an N-linked glycosylation system. More than 30 potential glycoproteins were identified. Analysis by mass spectrometry and NMR showed that the glycan was a heptasaccharide consisting of GalNAc-α1,4-[Glcβ1,3-]GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-β1,*N*-Asn. The glycan was linked to the proteins via asparagines with the

consensus sequence Asn-Xaa-Ser/Thr.

2.3.3.5 *Aeromonas caviae*

A. caviae is a Gram-negative rod shaped bacteria, facultative anaerobe. It is motile via a single polar flagellum. The bacteria is highly distributed in aquatic systems and is an important pathogen in fish. In addition, the organism is known to infect humans, and causes traumatic wound infection, gastroenteritis, septicemia, and meningitis.[66-67] Five genes (*flmA*, *flmB*, *flmD*, *neuA*, *neuB*) have been reported by Gryllos et al.[68] to be involved in motility and LPS O-antigen expression. Three of these five genes showed homology to flagellar glycosylation genes in *C. crescentus* and a knock-out mutant in flmB resulted in loss of motility, but did not result in altered LPS expression.[69] Mutant strains with inactive copies of *flmA* and *flmB* resulted in incomplete flagellar expression. However, the nature and kind of glycosylation remained unknown.

2.4 Analytical techniques used for the characterization of prokaryotic glycosylation

To facilitate the complete characterization of a glycoprotein it is necessary to determine

- The extent of glycosylation
- The sugar composition
- The structure of the glycan
- The identification of the glycosylation site

While glycan structures are relatively well defined for eukaryotes, the identification of prokaryote glycoproteins is more problematic due to their unusual oligosaccharide structures and their low abundance in cell extracts. Identification of these unususal glycan structures can be tedious and time consuming as current methods are rather developed for eukaryote N-linked glycans.

Currently the complete structural characterisation of only a limited number of prokaryotic glycoproteins is available. For most proteins, post-translational modification is apparent from the fact that the molecular weight observed by SDS-PAGE analysis is different from the mass predicted from the cDNA sequence. Evidence for glycosylation is provided by altered migration by SDS-PAGE analysis following chemical treatment, by staining with specific dyes by formation of Schiff bases,[48,65,68,72] and the binding of specific lectins or metabolic labelling with radioactive sugars are techniques used to confirm the presence of glycans.[65,70-72]

While these methods confirm the presence of a post-translational modification on the protein they do not give information about the extent, the sugar composition, the structure of the glycan and the identification of the glycosylation site.

Due to the complexity and diversity in glycan composition there is no standardised method to characterize post-translational glycosylation in prokaryotic and eukaryotic systems. The extent of post-translational modification is normally determined by measuring the molecular mass of the protein by SDS-PAGE or mass spectrometry. The measured molecular mass is then compared to the mass predicted from the cDNA sequence. This gives a measure of the degree of glycosylation and the proportion of the total mass of the glycoprotein attributed to the PTM. Information about the sugar composition for prokaryotic proteins can be obtained by releasing the glycan chemical or enzymatically[22,73] and performing permethylation analysis using GC/MS, or proteolytic digestion and analysis of the glycan by NMR, MALDI-TOFMS and LC/ES-MS.[24,28,47,74] Chemical release of the glycan is a more generic sample preparation method for prokaryotic glycoproteins due to the narrow specifity of glycanase and lack of availability of enzyme applicable to prokaryotic glycans. The structure of the glycan can be identified by means of NMR supported by data of the GC/MS or MS/MS analysis.[24,74]

The identification of the linkage site is a very challenging task in the whole analysis. The linkage site can be determined by proteolytic digestion and N-terminal sequencing, site directed mutagenesis or by electrospray mass spectrometry. (ESMS).[74,75] The determination of site of attachment for O-linked glycosylation by ESMS is further complicated by the fact that no consensus sequence is known.

Furthermore, collision induced dissociation (CID) of glycopeptides results mainly in the loss of the glycan, leaving the site of attachment unmodified. Thus, in most cases it is not possible to determine the linkage site solely by mass spectrometry. For this purpose Rademaker et al.[76] developed a novel method in 1998 for the detection of the linkage site for O-linked glycosylation.

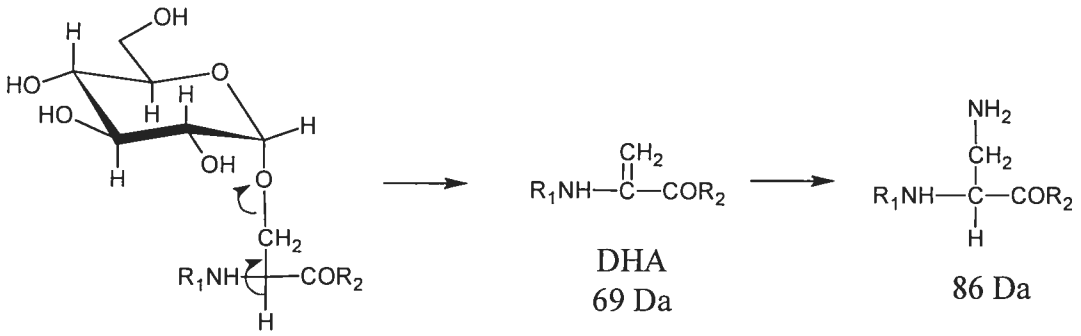

Figure 2.4: Reaction scheme of ß-elimination with NH_4OH

In the first step, the protein or tryptic digest is treated with ammonium hydroxide for ~18h at 50 °C. The glycan is released from the Ser/Thr via ß-elimination to yield dehydroalanine (DHA) for a serine residue or dehydrobutyric acid (DHB) for threonine. In the second step, NH_3 is added across the double bond via Michael addition forming a novel amino acid with a net –1 Da mass shift. This –1 Da mass shift can be detected via MS and identifies the linkage site of the former glycopeptide.

Two excellent examples of characterization of post-translational modifications on prokaryotic proteins were done on the pilin of *P. aerigunosa* and the flagellin of *C. jejuni*. Different approaches were used in each case, as explained here.

Castric identified in 1995 a gene responsible for the glycosylation of the pilin based on migration pattern on a SDS-PAGE gel.[46] In 2001 the structure of the glycan

was determined by proteolytic digestion of the pilin and analysis of the glycan by NMR.[47] Finally in 2002 Comer at al.[75] identified the linkage site of the glycan on the protein. They used a combination of different methods. They used N-terminal sequencing of an endopeptidase digest. The peptide containing the glycan was identified by a glycan specific monoclonal antibody. This analysis narrowed down the possible sites of glycan linkage to serine 148 and serine 118. Only site directed mutagenesis of serine 148 to alanine resulted in loss of glycosylation indicating that this serine was glycosylated.

The first fully characterized protein glycosylation on a flagellin protein was done for *C. jejuni* by Thibault et al. in 2001.[24] Different techniques were used for the characterization of the glycosylation. The extent of glycosylation was determined by a molecular mass determination of the intact protein by MS indicating that the protein is post-translationally modified. Identification of the glycan was achieved by tryptic digestion and analysis of the tryptic digest by using nanoLC/MS/MS. The acquired MS/MS spectra were manually searched for glycosylated peptides. The exact identification of the linkage site was achieved by fraction collection of the tryptic digest on a 2.1 mm x 15 cm RP 18 column and each fraction were analysed by nanospray MS/MS. Fractions containing glycopeptides were treated with NH_4OH to form a modified Ser/Thr residue with a –1 mass difference enabling the detection of the linkage site and were analysed after by nanospray MS/MS. To determine the anomeric structure of the glycan one tryptic glycopeptide was purified on a 2.1 mm x 15cm RP 18 column and after analysed by NMR.

2.5 Mass spectrometry

As mentioned in chapter 2.4, mass spectrometry techniques are often used for the characterization of proteins. Mass spectrometry is a technique for measuring the mass-to-charge ratio (m/z) of charged molecules in the gas phase and has become a powerful tool for the analysis of various compounds such as, small molecules, oligonucleotides, carbohydrates and biomolecules. The analysis of biomolecules by mass spectrometry has been revolutionized by the introduction of the soft ionization techniques matrix-assisted laser desorption/ionization (MALDI) in 1988 by Hillenkamp and co-workers[77] and electrospray (ES) by Fenn et al.[78] in 1989. These two techniques enabled the ionization and transfer of involatile or thermally labile molecules (i.e., most large biomolecules) into the gas phase.

2.5.1 Electrospray

Electrospray was developed as an interface for mass spectrometry by Fenn and co-workers in 1989.[78] The principal of ES is the transfer of charged molecules in the liquid phase into the gas phase and therefore is in most cases not an ionization technique. The experimental setup for ES is quite simple and consist of a capillary held at high voltage. The analyte solution is pumped through the capillary maintained at a high voltage, forming a mist of highly charged droplets. The size of droplets is reduced, through evaporation and by Coulomb explosion. The Coulomb explosion will occur when the magnitude of the electric field surrounding the charged analyte is

sufficient to overcome the surface tension (Rayleigh limit). From these small droplets desolvated ions are generated either by complete evaporation or by field desorption (Fig. 2.5).[79]

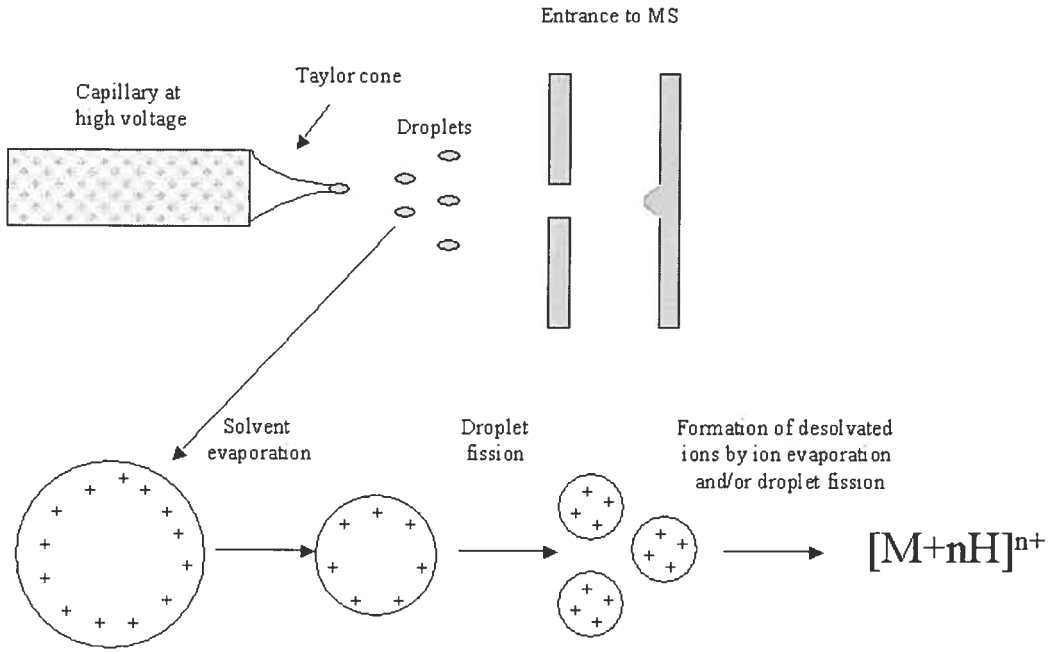

Figure 2.5: Production of desolvated ions by electrospray.

Sensitivity in electrospray ionization is dependent on sample flow rate. The lower the flow rate, the higher the sensitivity. Mann et al. developed a nanoelectrospray source (nanoES) using flow rates in the low nL/min range.[80] This device produces droplets in the nm range rather than μm range as for conventional ES. The efficiency of the transfer of the condensed-phase analyte into the gas phase was estimated to be approximately two orders of magnitudes better than conventional electrospray.[80] The experimental set-up involves a gold coated borosilicate capillary tapered to a fine tip of 1-2 μm. The capillaries are installed with a backing gas

pressure to improve flow stability but no pumping is involved. The flow rate depends on the applied potential, the solvent properties and the diameter of the tip.

2.5.2 Q-TOF mass analyzer

A Q-TOF is a hybrid instrument combining a quadrupole (Q) and a time of flight (TOF) analyzer. A schematic description of the instrument is shown in figure 2.7. The quadrupole mass analyzers are one of the most common analyzers in mass spectrometry. It is robust, inexpensive and has a high scan speed. The quadrupole separates ions into their corresponding m/z ratios by creating an electrical field, where only ions of specific m/z values have a stable trajectories and can pass the quadrupole analyzer. The quadrupole analyzer consists of four cylindrical rods. A direct current (dc) voltage and an oscillating voltage are applied to the rods, where adjacent rods are have the opposite polarity (Fig. 2.6). The ion trajectories are described by the Mathieu equation:

$$\frac{d^2u}{d\xi^2} + (a - 2q\cos 2\xi)u = 0 \qquad \text{(eq. 1)}$$

$$\xi = \frac{\omega t}{2} \qquad a = a_x = -a_y = \frac{8eU}{m\omega^2 r_o^2} \qquad q = q_x = -q_y = \frac{8eV}{m\omega^2 r_o^2}$$

u = ion trajectory in either x or y axis
ω = angular frequency
f = rf frequency
t = time
U = dc potential
m = mass
r_0 = half distance between the rods
V = rf voltage

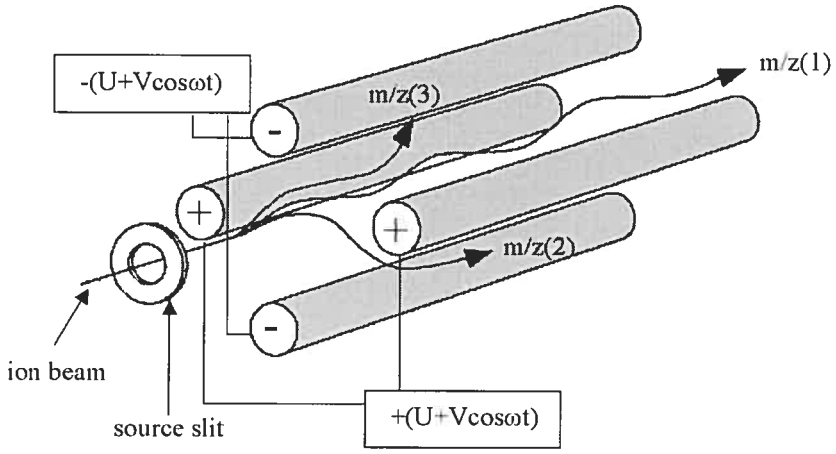

Figure 2.6 Schematic of a quadrupole mass analyser.

The quadrupole is either used as an ion guide (MS mode) simply transmitting all the ions to the TOF analyzer or as an ion filter (MS/MS mode), where only predefined m/z values are transmitted through the quadrupole analyzer.

The TOF analyzer separates ions based on their flight time through a fix drift tube. The time of arrival at the detector is directly proportional to the m/z value of the incoming ion (eq.2).

$$t = s * \left(\frac{m}{(2E_K)z} \right)^{1/2} \qquad \text{(eq. 2)}$$

t = drift time
s = drift distance
m = molecular weight
E_K = kinetic energy
z = charge of the ion

precursors from the quadrupole analyzer are fragmented through collision induced dissociation (CID) and the fragment ions are measured by the TOF analyzer. The advantages of this design are high resolution (up to 18000 in W-mode and 10000 in V-mode), high sensitivity (amol amounts can be detected) and high mass accuracy (<5 ppm by internal calibration) in MS and MS/MS mode. Today the Q-TOF instrument is one of the most widely used mass spectrometers for the characterization of proteins.

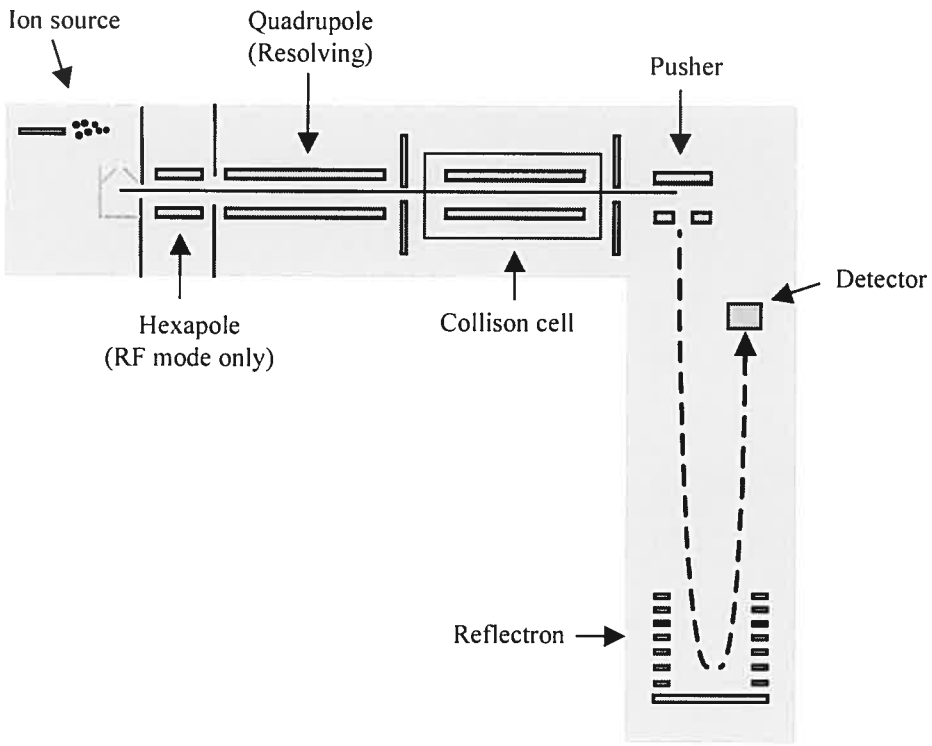

Figure 2.7: Schematic diagram of a Q-TOF mass spectrometer.

2.5.3 Protein identification using Q-TOF mass spectrometry

Intact proteins can be analyzed directly by mass spectrometry, though protein characterization typically requires their proteolytic digestion and LC/MS analysis of their corresponding peptides. The proteins are digested using typically the protease trypsin, an exceedingly specific protease cleaving proteins on the carboxy-terminal side of arginine and lysine residues.[81] The advantages of analyzing peptides are the convenience of a wide range of online separation techniques, the sensitivity of the MS is higher for peptides and the fact that sequence information can be more easily obtained for small peptides, rather than from intact proteins. After digestion the sample is analyzed by LC/MS and the mass spectrometer measures the mass-to-charge ratio of the eluting peptides from the LC column (Survey scan). Primary structure information of these peptides are obtained from the MS/MS spectra of the corresponding precursor ions. In the MS/MS experiment a specific precursor is isolated by the quadrupole, fragmented in the collision cell and the fragment ions are measured in the TOF analyzer. The most common fragmentation in the MS/MS experiment is a cleavage of the amide bond between amino acids. The resulting ions are called b-ions if the charge is retained by the N-terminal end of the peptide and y-ions when the charge is retained on the C-terminal end (Fig. 2.9).[82] From the observed fragment ions, peptide sequence information can be obtained. The mass difference of neighboring peaks of the same ion series corresponds to the residue mass of the amino acid. Additionally, post-translational modifications can be detected

from the fragmentation pattern of the peptide. In general, MS/MS spectra can be interpreted manually or with the help of computer algorithms. In both cases the success of this approach depends on the quality of the MS/MS spectra. It is easier to use a database-matching approach employed by software programs like Sequest (http://fields.scripps.edu/sequest/), ProteinProspector (http://prospector.ucsf.edu/) or Mascot (www.matrixscience.com). Automatic database searches compare MS/MS spectra to those predicted from existing protein databases and successful identification can be obtained even when only a partial amino acid sequence is available. [83] An example of a Mascot search parameter page is shown in figure 2.8. Basically, a data file containing a list of the charge and m/z value of the acquired precursor and the corresponding m/z values and intensities of the fragment ions is uploaded into the software. Different parameters can be adjusted such as the type of database, the enzyme used, fixed and variable modifications, the instrument type etc.. After choosing the parameters, the data file is searched against the selected database and a list of best matching proteins with their corresponding peptides is obtained.

During a LC/MS run the mass spectrometer can be used in data dependent mode such that in the survey scan the masses of the eluting peptides are measured and the most intense ions are automatically selected for the MS/MS experiment. During a 60 min LCMS run more than 1250 MS/MS spectra can be acquired.

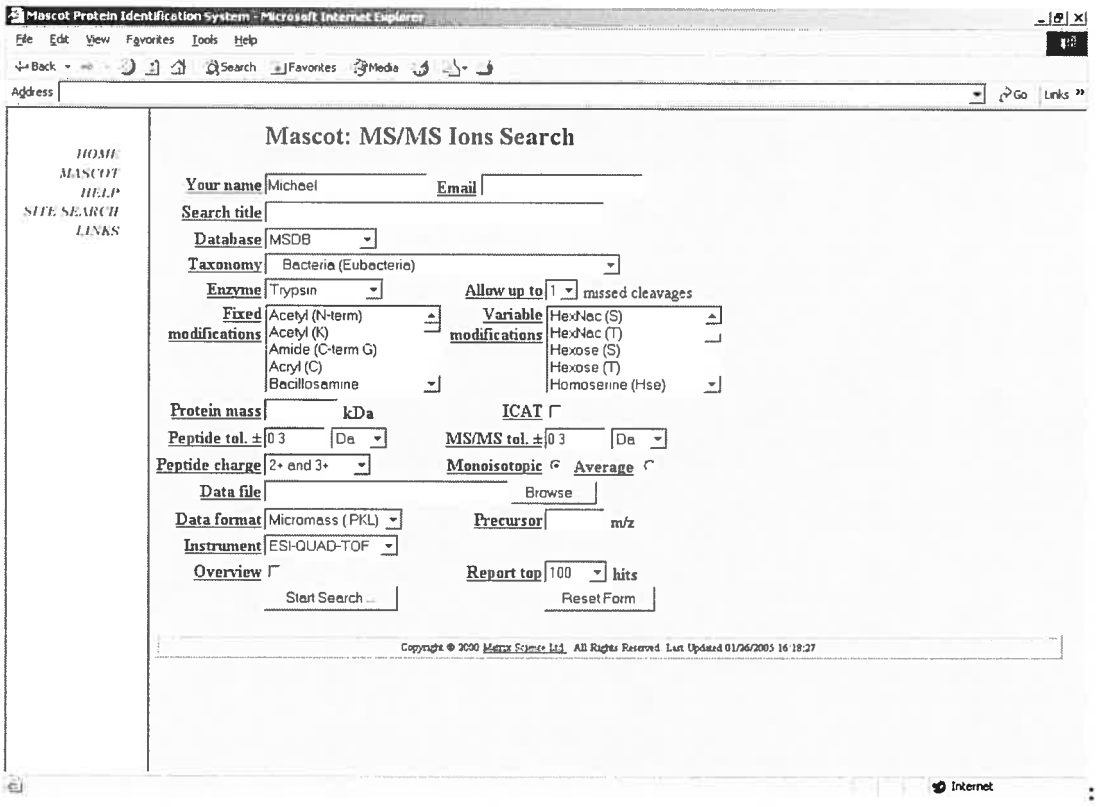

Figure 2.8: Screenshot of the MS/MS ion search parameters used with the Mascot program.

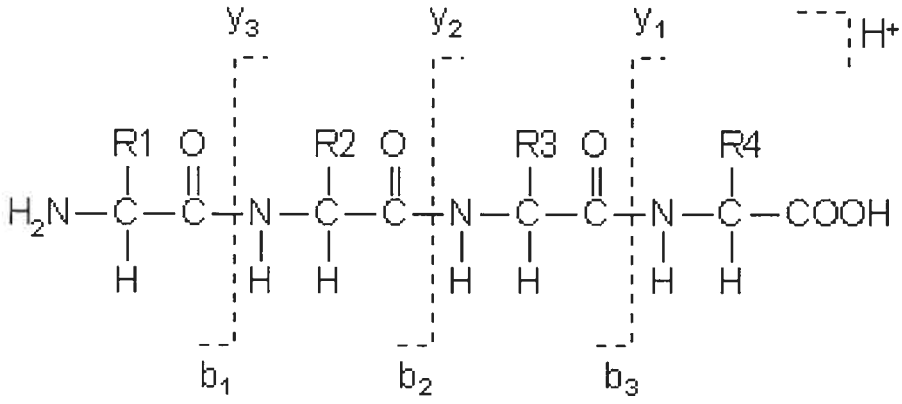

Figure 2.9: Fragmentation of peptides by CID.

The ions are labeled consecutively from the original amino terminus. The most common fragmentation is the cleavage of the amide bond between amino acids. Fragment ions with charge retention at the N or C-terminus are referred to as b and y-type ions. [82]

3 Experimental section

All chemicals were obtained from Sigma if not noted otherwise. Water (Optima grade), acetonitrile (HPLC grade) and formic acid (FA) were obtained from Fisher and the trypsin from Promega.

3.1 Instrumental methods

A photograph of the instrument used to analyze the samples is shown in figure 3.1. The system consists of a CapLC™ pump, a stream selector and a Q-TOF Ultima™ mass spectrometer (Waters). Peptide separation was achieved using a Waters Symmetry™ C_{18} precolumn cartridge and a home made Jupiter C_{18} analytical column 10 cm x 150 μm i.d., 5 μm particle size (Phenomenex). For each sample approximately 200 ng was loaded on the precolumn for 3 min at a flow rate of 30 μL/min from pump C (Fig. 3.2a). After that the stream selector valve was switched to position 2 (Fig. 3.2b) and the gradient was started. A linear gradient of 10-40 % acetonitrile (0.2% formic acid) in 60 minutes was used. The set-up of the instrument is shown in figure 3.2.

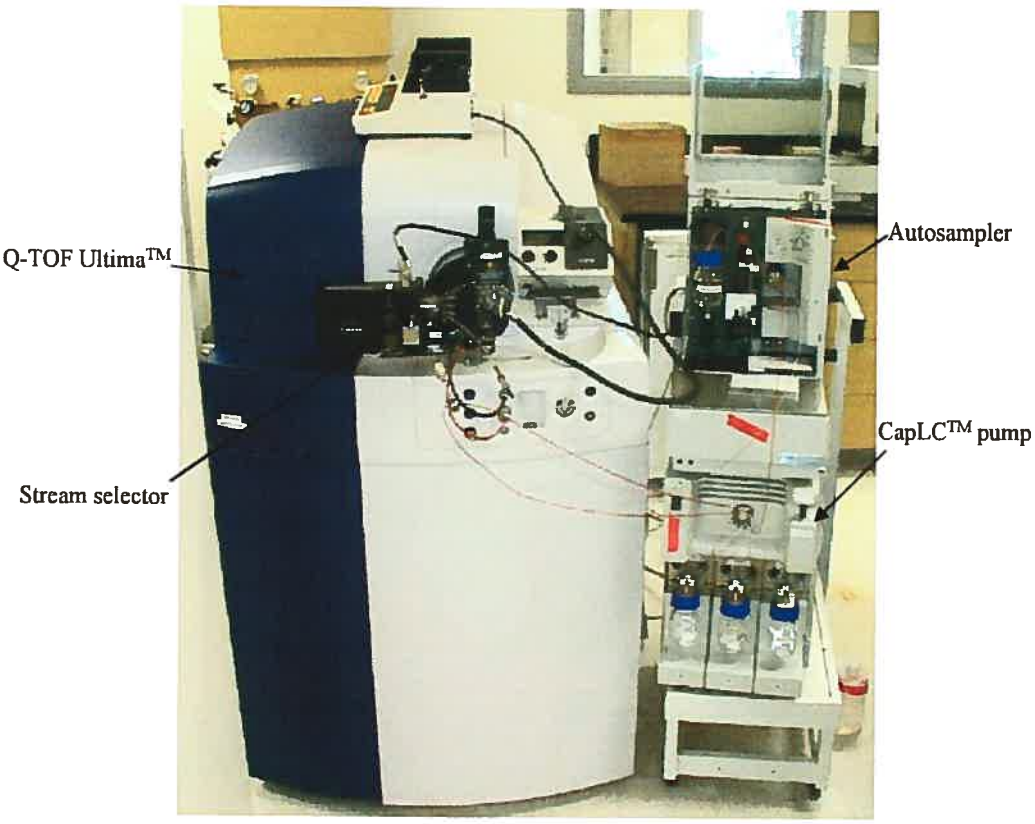

Figure 3.1: Photograph of a CapLC™ system coupled to a Q-Tof Ultima™

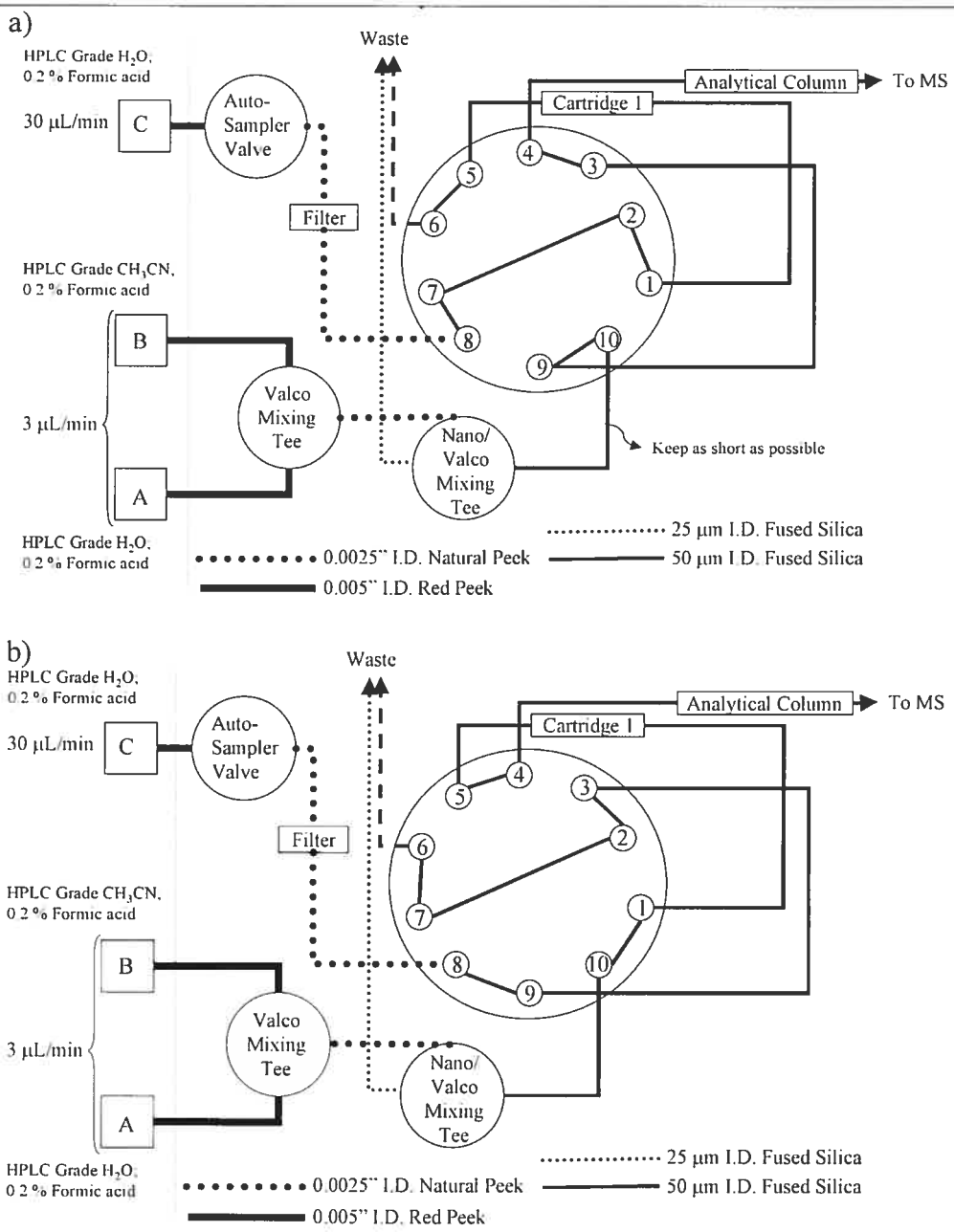

Figure 3.2: Instrument set-up of the capLC/MS system

a) Switching valve configuration for loading the sample. b) Valve configuration for eluting the sample

All MS/MS spectra were obtained in data-dependent mode selecting the three most intense ions per survey scan for the MS/MS experiment. The duration of the MS/MS aquisition per precursor was set to 2 s. For the analysis of the ß-eliminated sample an

inclusion list was generated with the predicted masses of the deglycosylated glycopeptides, therefore only former glycopeptides were selected for the MS/MS experiment. The duration of one MS/MS experiment was increased to 6 s.

Second generation product ion spectra were obtained by increasing the RF Lens1 from 50V to 125V to form fragment ions in the orifice/skimmer region, while the desired precursor ion was selected by the quadrupole.

3.2 Preparation of the flagellin samples

The flagellin samples for the bacteria and the isogenic mutants were prepared by Dr. Susan Logan (NRC, Ottawa, Canada). Briefly, flagellins were prepared by shearing the filaments from the cell surface using a waring blender. The flagellar filaments were collected by ultracentifugation (100,000 x g, 1 h) following removal of bacterial cells by low speed centrifugation.

Prior to infusion of the intact flagellin protein into the MS each flagellin sample was ultracentrifuged in aqueous 0.2 % (v:v) formic acid using Centricon YM-30 membrane filter (Millipore). For the flagellin sample of *L. monocytogenes* Centricon YM-10 membrane filters were used. These solutions were infused to the mass spectrometer at a flow rate of 0.5 µL/min. In the case of *C. jejuni* flagellin a solution of 50/50 H_2O/CH_3CN + 0.2 % FA was infused into the MS.

Sample digestions were performed overnight in 50 mM NH_4HCO_3 (5% CH_3CN) using trypsin (Promega) at a enzyme:substrate ratio of 1:50 at 37 °C. For

the flagellin of *P. aeruginosa*, PAO strain the tryptic digest was redigested for 4 h with chymotrypsin (Promega).

The precise identification of the glycosylation site was achieved using β-elimination with ammonium hydroxide to leave a modified Ser/Thr residue that could be located using tandem mass spectrometry. The tryptic digest was subjected to alkaline hydrolysis using 1 mL NH_4OH for 18 h at 50 °C, dried down, redissolved in water and analysed by capLC/MS/MS.

Fraction collection of the tryptic digest of flagellin of *P. aeruginosa* PAK was done on a Vydac C_{18} 2.1 mm x 20 cm, 5 μm particle size column. A linear gradient of 0-100% acetonitrile (0.2% formic acid) in 30 min was used and fractions were collected every minute. Fractions containing the two ion series were analysed by nanospray MS/MS on the Q-TOF Ultima™ machine before and after β-elimination.

3.3 Top down mass spectrometry

For the top down approach, flagellins were dialyzed in H_2O (0.2 % formic acid) using Centricon YM-30 membrane filter (Millipore) with a molecular weight cut off of 30 kDa to remove salts. These solutions were infused into a Waters Q-TOF™ Ultima mass spectrometer at a flow rate of 0.5 μL/min. MS/MS experiments were performed using Argon as collision gas with collision energies ranging from 10-25 V. Second generation fragment ion spectra were obtained by increasing the RF lens1 voltage from 50 to 125 V, thereby forming fragment ions in the high-pressure

region of the skimmer/cone region of the mass spectrometer and selecting the desired precursor m/z value for the MS/MS experiment.

3.4 Cloning and complementation studies in *H. pylori*

These analyses were performed by Dr. Susan Logan (NRC, Ottawa, Canada) and additional information can be found in Schirm et al..[84]

Briefly, PCR primers were designed to amplify each ORF using *H. pylori* 26695 sequence data. The product was cloned into pUC19 and plasmid DNA purified and sequenced. Briefly, each clone was disrupted by using reverse primers which were internal to each gene, in a PCR reaction and which resulted in deletion of 10-20bp within the ORF. This was followed by ligation of a Kan cassette to the gel purified product to make plasmids pAAHP0840kan, pAAHP0178kan, pAA0326A, pAA0326B and pAA0114. The mutated allele was returned to Helicobacter strains by natural transformation.

For complementation studies, wild type copies of HP0178, HP0114 and HP0326A and B were obtained by PCR of genomic 26695 DNA using the following primers (HP0178: 1F 5' CAAACACCCATTACTCTTAAATCATGCCAA3', 1R 5'CCTACAATGAGCGTTCTATATCAGCGCT3', HP0114: 1F 5' CGGGATCCAATTCAAAGGGGCGTTAGCCC 3', 1R 5' GGAATTCTTACCATTCTTTTAAAGCCATTTTGATCGCT3', HP0326A/B: 1F 5' CGGGATCCATGAGAGCGATCGCTATTGTTTTAGCCAGA3' , 1R 5' GGGGTACCTCAAATCTCTAAAAACTCCCTTAATGCACCCT3'). The cloned

genes were subsequently transferred to the pHEL shuttle vector and used to naturally transform the respective isogenic mutants in either M6 or 1061. Transformants were selected on chloramphenical and kanamycin and initially stabbed onto motility agar.

SDS-PAGE of whole cell lysates

PCR's were carried out by using a Perkin Elmer thermocycler and using PWO polymerase (Roche Molecular Biochemicals). For the DNA sequence analysis, PCR products and plasmid DNA's were sequenced using terminator chemistry and Taq cycle sequencing kits (Perkin Elmer Applied Biosystems) and analysed on an Applied Biosystems 373 DNA sequencer. Custom primers were synthesized on an Applied Biosystems 394 DNA/RNA synthesizer.

3.5 Cloning and complementation studies in *L. monocytogenes*

These analyses were performed by Dr. Susan Logan (NRC, Ottawa, Canada) and additional information can be found in Schirm et al..[85]

Briefly, flagellin genes were amplified from each respective strain by PCR using the following primers based on the N and C-terminal sequence of the full length cloned flagellin sequences. Primer F1 (5'-ATGAAAGTAAATACTAATATCATTAGCTTG-3') corresponding to the N-terminal amino acid sequence MKVNTNIISL and primer R1 (5'-GCTGTTAATTAATTGAGTTAACAT-3') corresponding to the C-terminal flagellin sequence MLTQLINS*. The resulting PCR amplified products were analysed by gel

electrophoresis, and the PCR products recovered using glass milk (Bio101, La Jola) and cloned into pCR2.1-TOPO (Invitrogen). Each PCR amplified flagellin gene was sequenced using forward and reverse M13 universal primers. An Applied Biosystems 373 DNA sequencer and Taq cycle sequencing kits with terminator chemistry was used to sequence the plasmid DNA. Custom primers were synthesized on an Applied Biosystems 394 DNA/RNA synthesizer.

3.6 Electron microscopy

The wild type and mutants of *H. pylori* were analyzed by electron microscopy by Dr. J. Austin (Health Canada, Ottawa, Canada). A grid covered with a carbon-coated parlodion film was floated onto a 20µl sample drop and left for approximately 2 min for adsorption of the sample to the grid. The grid was then removed from the drop and floated on a drop of 1% ammonium molybdate and left for approximately 2 minutes. Excess stain was removed by touching the edge of the grid to a piece of Whatman No.1 filter paper. All samples for electron microscopy were examined in a Zeiss EM902 transmission electron microscope (Carl Zeiss) operating at 80 kV with the energy loss spectrometer in place.

3.7 Monosaccharide analysis of flagellar glycan of *P. aeruginosa* PAK

The monosaccharide analysis by GC/MS was performed by Dr. E. Vinogradov (NRC, Ottawa, Canada). Flagellin (5 mg) was hydrolysed with 3 M CF_3COOH (2h, 120°), the sample dried under a stream of nitrogen, dissolved in 0.3 mL water, and treated with $NaBH_4$ (5 mg, 1 h, 25° C). The $NaBH_4$ reagent was destroyed with acetic acid (0.5 ml), dried under a stream of nitrogen, and methanol (1 mL) was added and the sample dried again under a stream of nitrogen. This was repeated twice, and the sample was then acetylated with acetic anhydride (0.5 mL, 1h, 85°), dried under a stream of nitrogen, water (1 mL) and chloroform (0.5 mL) were added, the sample shaken, and the chloroform layer passed through cotton wool. The derivatised carbohydrate extract was then injected into GC-MS instrument Varian Saturn 2000 equipped with a capillary column DB-1 (0.25mm x 50 m) with ion trap detector. The same procedure was used for the preparation of standards of perosamine (4-amino-4,6-dideoxymannose) and viosamine (4-amino-4,6-dideoxyglucose) from the O-specific polysaccharide of *E.coli* O157 and the LPS core sample from *P. mirabilis* O6 respectively.

3.8 Preparation of the periplasmic extract sample

The periplasmic extract was prepared by Dr. Susan Logan (NRC, Ottawa, Canada). Briefly, the periplasmic protein extract was obtained by harvesting cells

from 48 h culture (200 mL) grown in brucella broth. Cells were washed in 33 mM tris pH 7.4 and then osmotically shocked by resuspension on 4ml distilled H_2O with 1mM $MgCl_2$ and left on ice 10min. Cells were removed by centrifugation and the supernatant lyophilised.

For the analysis by SDS-PAGE 40 μg of the sample was loaded onto a 12% Bis-Tris NuPAGE mini gels (8 x 10 cm), 12 wells, 1mm thickness. The gel was run, fixed and excised into 24 bands. The excised bands were transferred into different wells of a 96 well plate. An aliquot of 100 μL of acetonitrile was added to each well and after 15 minutes the acetonitrile was removed. The gel pieces were dried at room temperature for 15 min. To break disulfide bonds between cysteine residues, 100 μL of H_2O_2 7% solution was added to the gel pieces. The plate was covered and incubated for 15 min in the dark. The H_2O_2 solution was removed and 100 μL of H_2O was added. After 15 min the water was removed and 100 μL CH_3CN was added for 15 minutes. The CH_3CN was removed and the gel pieces were dried at room temperature for 15 min. An aliquot of 55 μL of digestion buffer containing ~150 ng of trypsin was added to the well and the plate was covered and incubated at 37 °C for 4 hours. The solution was then transferred to another 96 well plate. An aliquot of 50 μL of the back extraction solution (0.5 M urea in 90% CH_3CN) was added to the gel pieces, incubated for 30 minutes and after the solutions were transferred to the corresponding wells in the other well plate. The well plate was dried down and stored at – 20 °C. For capLC/MS/MS analysis the samples were redissolved in 95/5 H_2O/CH_3CN (0.2% formic acid).

For the analysis of the periplasmic extract by SCX, around 250 µg of the sample was dissolved in H_2O + 0.2 % FA. This solution was ultracentifuged 3x in H_2O + 0.2% FA using Centricon YM-3 membrane filter (Millipore). The retentate was dried down and redissolved in 100 µL 8M urea. 100 µL of digestion buffer containing ~10 µg of Lys-C was added. This solution was incubated for 4 hours at 30 °C. After the incubation 800 µL of digestion buffer was added containing this time 10 µg of trypsin. The solution was incubated over night at 37 °C. The next day the solution was dried down, redissolved in 100 µL water, dried down again and stored at –20 °C in the freezer. For the SCX analysis a 2.1cm x 20 cm PolySULFOETHYL Aspartamide™ strong cation exchange column (Western Analytical) was used. The operating conditions were:

Buffer A: 10 mM ammonium formate + 15 % CH_3CN, pH 3.5

Buffer B: 1M ammonium formate +15 % CH_3CN, pH 3.5

The gradient was as follows:

T in min	A in %	B in %
0	100	0
3	100	0
13	90	10
23	70	30
28	40	60
32	0	100
41	0	100

Fractions were collected every minute from 0 to 30 minutes. The fractions were dried down, redissolved in 100 µL H_2O+ 0.2% FA and analysed by capLC/MS/MS

3.9 Preparation of the membrane extract sample

The membrane extract was prepared by Dr. S. Logan (NRC, Ottawa, Canada). Briefly, the membrane fraction of *H. pylori* was prepared by lysing cells with an emulsiflex high pressure system. Unbroken cells and debris was removed by low speed centrifuation. Total cell membranes were collected by centrifugation at 100000 xg and the pellet washed twice and resuspended in PBS prior to analysis.

The membrane extract was spun down at 100000 g for 1 h in an ultracentifuge. The supernatant was discarded and 1 mg of the membrane extract was redissolved in 50 µL of 8 M Urea, shaken and ultrasonicated for around 15 minutes. Then, 400 µL of digestion buffer containing 20 µg of trypsin was added to the sample. The sample was incubated at 37 °C for 15 minutes and sonicated for 15 minutes. These steps were repeated until the pellet was gone and completely dissolved. This solution was then incubated overnight at 37 °C. The sample was dried down and stored at –20 °C in the freezer prior analysis. For the SCX separation the same condition as for the analysis of the PPE were used.

3.10 Data analysis of the acquired MS/MS spectra from the periplasmic and membrane extract

The acquired mass spectra were analysed and screened for glycosylation by four different ways:

- Mascot search using the variable modifications: deamidation, wma+1, wma-1, M+16. Manual validation of the Mascot search data and elimination of redundancy by in-house developed software.
- Mascot search with variable modifications for Hex (+162Da), dHex (+146Da), HexNAc (+203Da), Bacillosamine (+228Da) and sialic acid (291Da) and manual interpretation of the search data.
- Extraction of MS/MS spectra with intense ions (>30%) of common carbohydrate oxonium ions, like 204, 229, 292, 298, 299, 316 and 317 by in-house developed software and manually interpretation of these spectra.
- Manual screening of all acquired MS/MS spectra for characteristic glycospectra.

4 Results and Discussion

This part of the thesis is divided into 3 sections. In part one the flagellin protein, of different important human pathogens are characterized for post-translational modifications by proteolytic digestion and analysis by capLC/MS/MS. In the second part a novel "top down" approach is used as a fast screening method for the identification and monitoring of post-translational modifications on intact proteins. In the third part, the analysis of a periplasmic and a crude membrane extract are explained.

4.1 Characterization of post-translational modifications (PTMs) on flagellin proteins from important human pathogens

4.1.1 Characterization of PTMs on the flagellin of *H. pylori*

Results of this study were published previously in Molecular Microbiology[84] and are further described in this section.

4.1.1.1 Structural analysis of *H. pylori* flagellin

The flaA and flaB flagellin structural genes from strain 1061 were amplified by PCR and subjected to DNA sequence analysis by Dr. Susan Logan (NRC, Ottawa, Canada) to determine the amino acid sequence of the translated proteins. The predicted aa sequence for flaA (1061) was identical to that of 26695 while the predicted aa sequence of the 1061 flaB gene had two amino acid substitutions at position 154 (T for an A) and 181 (E for an A) when compared to the predicted 26695 flaB sequence (GenbankAccession # AY155231 and AY155232). Electrospray mass spectrometry analyses of purified flagellin from *H. pylori* 1061 resulted in a multiply charged-state envelope (Fig. 4.1). The data was deconvoluted to obtain a profile of the molecular mass for the flagellin proteins and showed two well defined components at Mr:55049 and 55365 Da corresponding to the molecular mass of the

monomeric FlaA protein (Mr: 53153.4 Da) with additional modifications accounting for 1896 and 2212 Da respectively (Fig.4.2). These mass spectral analyses also indicated relatively little heterogeneity in isoform or glycoform distribution of the *H. pylori* flagellin (peak width 40 Da) in contrast to those obtained from *C. jejuni* flagellin where extensive heterogeneity in glycosylation profile resulted in a relatively large peak width extending over 700 Da.[24] It is noteworthy that two other components of significantly lower intensity (< 10% FlaA) were observed at 56683 Da and 56999 Da which were later shown to correspond to FlaB with modifications accounting for an additional 2794Da and 3160 Da respectively (see below).

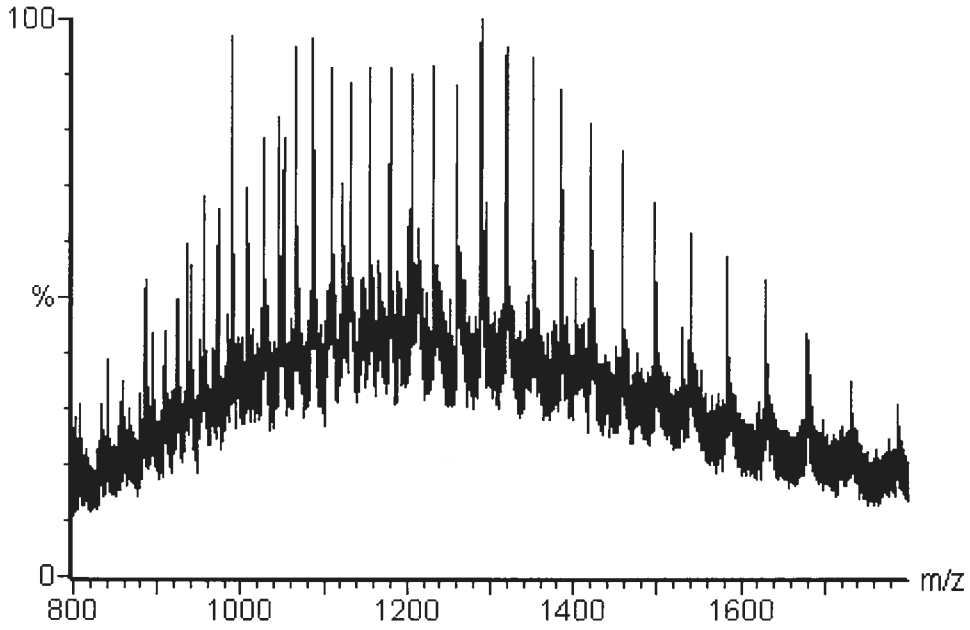

Figure 4.1: Molecular mass determination of intact bacterial flagellin of *H. pylori*.

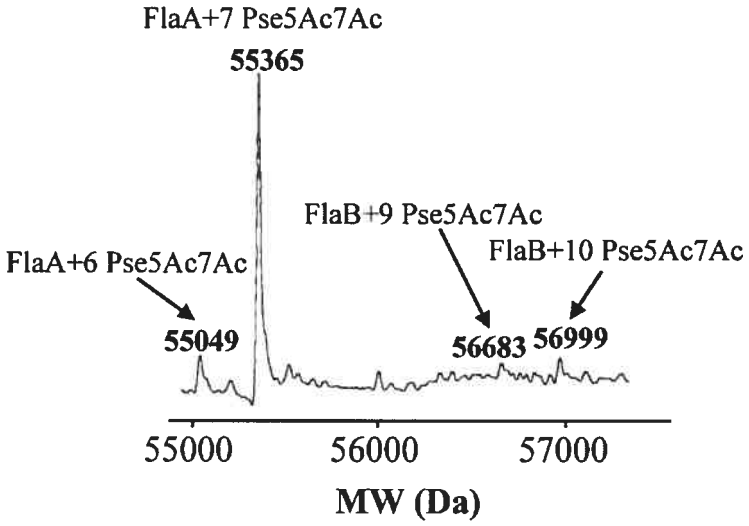

Figure 4.2: Deconvoluted electrospray mass spectrum of intact flagellin from *H. pylori* 1061. The reconstructed molecular mass profile is shown, and indicates two peaks at Mr: 55365 and 55049 Da corresponding to FlaA and two minor peaks at 56999 and 56683 Da from FlaB.

CapLC/MS/MS analysis of the tryptic digest and manual interpretation of the acquired MS/MS spectra identified several modified glycopeptides all containing a monosaccharide with a residue mass of 316 Da, the same mass as for Pse5Ac7Ac. (Fig. 4.3a). Second generation product ion spectrum of m/z 317 showed a fragmentation pattern identical to that of N-acetyl pseudaminic acid (Pse5Ac7Ac) (Fig. 4.4), an unusual O-linked monosaccharide previously identified in *C. jejuni* and *C. coli* flagellins.[24] No evidence for additional carbohydrate residues were observed from these tandem mass spectrometry experiments in contrast to Campylobacter flagellins which displayed acetamidino and hydroxyproprionyl substituents on Pse5Ac7Ac.[24] Results from these combined analyses suggested that *H. pylori* FlaA protein is modified with a total of 7 O-linked Pse5Ac7Ac residues (7 x 316.1 Da = 2212.7 Da) whereas FlaB contains up to 10 Pse5Ac7Ac residues (10x316.1Da = 3161Da).

In order to precisely assign the location of glycosylation sites on *H. pylori* FlaA protein, the flagellin tryptic digest was treated with NH_4OH to identify the site of O-linked attachment.[75] Upon ß-elimination and addition steps, O-linked Ser and Thr residues yield modified amino acids of neutral mass of 86 and 100 Da that can be identified by the corresponding mass shift in the product ion spectrum (Fig. 4.3). The sample was then analysed by capLC/MS/MS using an inclusion list, that contained the theoretical m/z values of the products of the ß-elimination. By using this approach it was possible to map all 7 sites of modifications on *H. pylori* flagellin (Fig. 4.5a). Figure 4.3 shows an example of a MS/MS spectrum corresponding to the tryptic glycopeptide IATGALITASGDISLTFK before and after ß-elimination. Before ß -

elimination an oxonium ion at m/z 317 and m/z 299 (loss of water) are observed from Pse5Ac7Ac and y-fragments from the peptide enabled the identification of the corresponding peptide. After ß-elimination the former glycopeptide shows a mass shift of –1 Da for the y_{11} fragment indicating that the Pse5Ac7Ac residue was attached to this threonine (Fig. 4.3b).

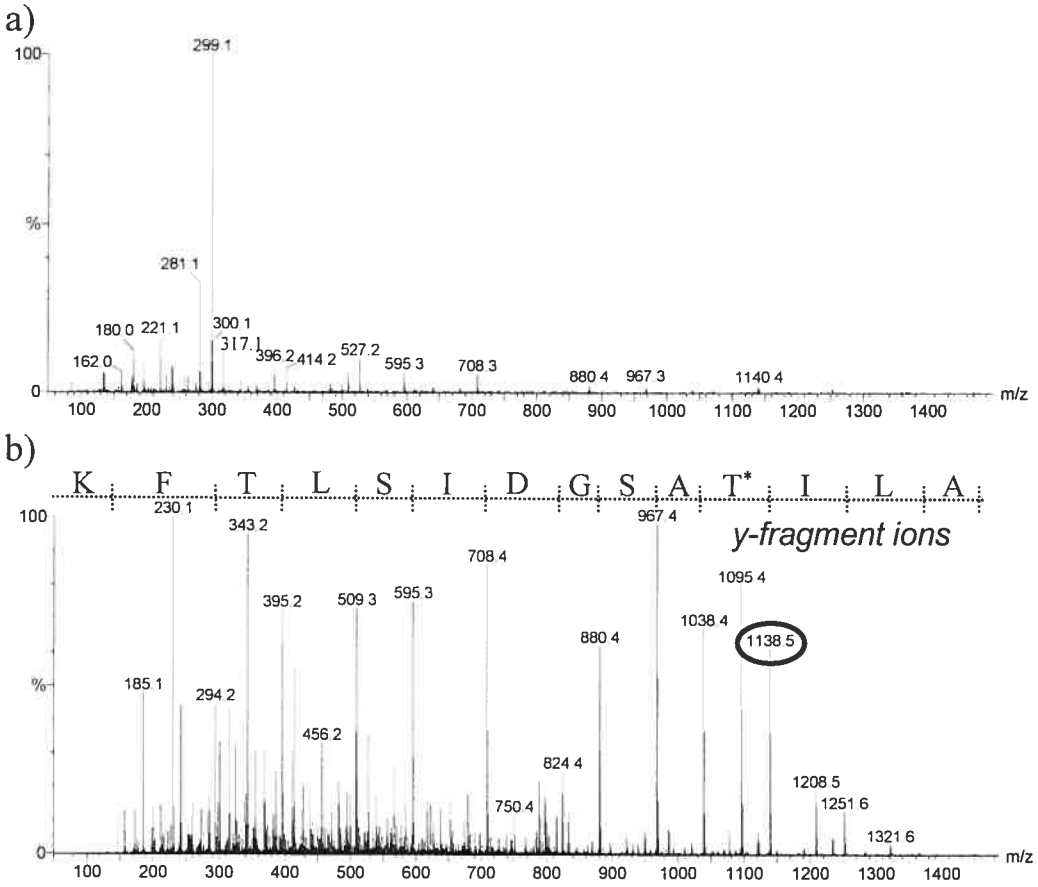

Figure 4.3: Tandem mass spectra of m/z 1047.85 (a) and its product from alkaline hydrolysis m/z 889.35 (b). The fragment ion at m/z 317 in (a) correspond to the oxonium ion of Pse5Ac7Ac. For MS/MS spectrum of m/z 317 see figure 4.4.

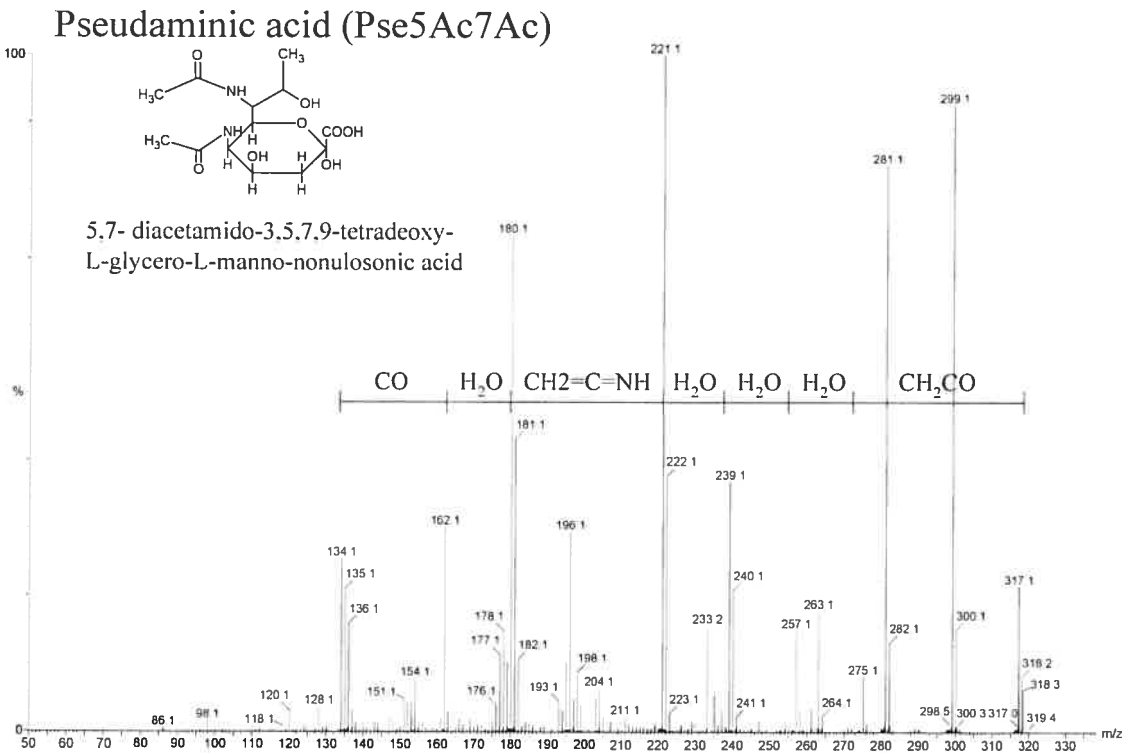

Figure 4.4: Second generation fragment ion spectrum of the unusual carbohydrate residue at m/z 317.

It is noteworthy that all glycosylation sites are located in the central core region consistent with that observed previously in Campylobacter flagellins. Similar analyses were also performed on FlaB present in this sample at a much lower concentration. While precise location of the glycosylation sites was very challenging in view of the overwhelming abundance of FlaA and the limited amount of flagellin sample available, it was possible to identify a total of 10 modification sites, all containing O-linked Pse5Ac7Ac in the central core region of the FlaB molecule (Fig. 4.5b).

FlaA

AFQVNTNINA MNAHVQSALT QNALKTSLER LSSGLRINKA ADDASGMTVA 50
DSLRSQASSL GQAIANTNDG MGIIQVADKA MDEQLKILDT VKVKATQAAQ 100
DGQTTESRKA IQSDIVRLIQ GLDNIGNTTT YNGQALLSGQ FTNKEFQVGA 150
YSNQSIKASI GSTTSDKIGQ VRIATGALIT ASGDISLTFK QVDGVNDVTL 200
ESVKVSSSAG TGIGVLAEVI NKNSNRTGVK AYASVITTSD VAVQSGSLSN 250
LTLNGIHLGN IADIKKNDSD GRLVAAINAV TSETGVEAYT DQKGRLNLRS 300
IDGRGIEIKT DSVSNGPSAL TMVNGGQDLT KGSTNYGRLS LTRLDAKSIN 350
VVSASDSQHL GFTAIGFGES QVAETTVNLR DVTGNFNANV KSASGANYNA 400
VIASGNQSLG SGVTTLRGAM VVIDIAESAM KMLDKVRSDL GSVQNQMIST 450
VNNISITQVN VKAAESQIRD VDFAEESANF NKNNILAQSG SYAMSQANTV 500
QQNILRLLT

Mr: 53153.4 Da

FlaB

SFRINTNIAA LTSHAVGVQN NRDLSSSLEK LSSGLRINKA ADDSSGMAIA 50
DSLRSQSANL GQAIRNANDA IGMVQTADKA MDEQIKILDT IKTKAVQAAQ 100
DGQTLESRRA LQSDIQRLLE ELDNIANTTS FNGQQMLSGS$_2$FSNKEFQIGA 150
YSNTTVKASI GSTSSDKIGH$_1$VRMETSSFSG EGMLASAAAQ NLTEVGLNFK 200
QVNGVNDYKI ETVRISTSAG TGIGALSEII NRFSNTLGVR ASYNVMATGG 250
TPVQSGTVRE LTINGVEIGT VNDVHKNDAD$_2$GRLTNAINSV KDRTGVEASL 300
DIQGRINLHS$_3$IDGRAISVHA ASASGQVFGG GNFAGISGTQ HAVIGRLTLT 350
R$_1$TDARDIIVS GVNFSHVGFH SAQGVAEYTV NLRAVRGIFD ANVASAAGAN 400
ANGAQAETNS QGIGAGVTSL KGAMIVMDMA DSARTQLDKI RSDMGSVQME 450
LVTTINNISV TQVNVKAAES QIRDVDFAEE SANFSKYNIL AQSGSFAMAQ 500
ANAVQQNVLR LLQ

Mr: 53839 Da

Figure 4.5: Sequence coverage map of *H. pylori* 1061 FlaA (A) and FlaB (B) proteins. The sites of O-linked Pse5Ac7Ac glycosylation are boxed. Numbers shown above boxes for FlaB correspond to Pse5Ac7Ac units found on the identified tryptic peptide.

Five genes involved in this glycosylation process were identified by Dr. Susan Logan (NRC, Ottawa, Canada). In all strains where these genes had been insertionally inactivated, cells were completely non-motile. The mutants were examined by transmission electron microscopy by Dr. J. Austin (Health Canada, Ottawa, Canada) and no flagella could be detected in these mutant strains while the characteristic

multiple polar, sheathed flagella were abundant on parent cells. These observations demonstrated that normal flagellar assembly was severely affected (Fig. 4.6).

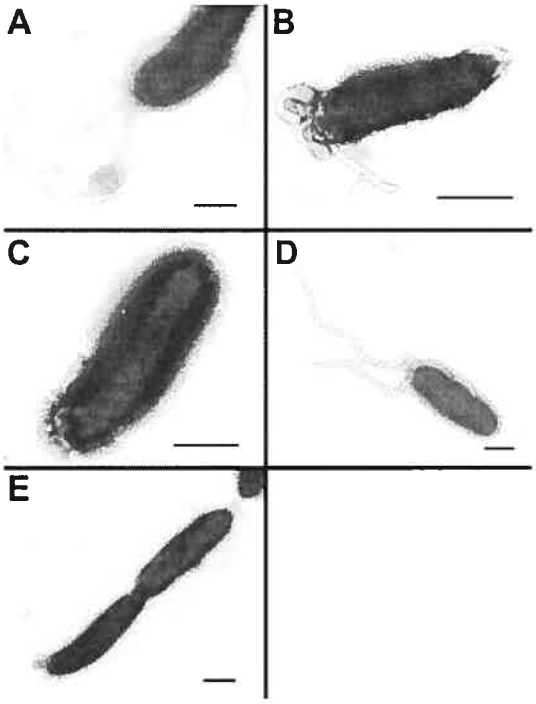

Figure 4.6: Transmission electron micrographs of negatively stained *H. pylori*.
A. *H. pylori* 1061; B. HP0178 isogenic mutant; C. HP0326A isogenic mutant; D. HP0326B isogenic mutant; E. HP0114 isogenic mutant. Scale bars represent 1μm. Note the appearance of empty flagellar sheaths in panels B and D, and the presence of truncated sheaths at the cell poles in panels C and E.

4.1.1.2 Discussion

Structural characterization of the flagellin proteins identified 7 sites of glycosylation on FlaA and 10 sites on FlaB where pseudaminic acid is O-linked at serine or threonine in the central region of the molecule. As for Campylobacter flagellin, glycosylation appears to be restricted to the central domain of the flagellin monomer a surface exposed region in the assembled filament. However, a major difference in the modification profile of *H. pylori* flagellin when compared to that found on Campylobacter flagellin is the lack of heterogeneity in the degree of modification as evidenced by the sharp intact mass profile indicating a total of seven sites. Moreover, only a single sugar species, Pse5Ac7Ac, was present on *H. pylori* flagellin in contrast to the numerous related derivatives found on Campylobacter flagellin.[24] This observation correlates well with the "glycosylation related" gene content of the genomes of Campylobacter and Helicobacter. While the Campylobacter genome contains 4 distinct carbohydrate biosynthetic loci with over 100 annotated carbohydrate biosynthesis genes, the two *H. pylori* genomes have only a very limited number of genes (<30) which show homology to either the larger pool of flagellar glycosylation genes found in the Campylobacter genome or to LPS/capsule biosynthesis genes. The relevance of this in terms of the pathogenesis of each organism remains to be established, and may reflect the particular mucosal environment in which each organism resides (stomach vs intestine) and/or in the presence of a sheath covering the flagellar filament of *H. pylori*. As for Campylobacter, the glycosylation of Helicobacter flagellin may occur as a

consequence of the local hydrophobicity surrounding selected serine/threonine residues. This report now expands the process of glycosylation of bacterial flagellins with the novel sugar Pse5Ac7Ac, to a second bacterial species which also has a polar complex flagella. Additional studies are required to establish if the presence of this charged acidic polysaccharide confers some novel property to the filament or if it is indeed an integral part of the assembly process.

Recent discoveries of derivatives of pseudaminic acid as components of cell surface glycopolymers or homologs of genes known to be involved in glycopolymer synthesis have been documented for a diverse number of bacterial species i.e. Clostridium[86], Legionella,[87] Pseudomonas,[47.50]Campylobacter,[24] Neisseria,[88] and Aeromonas.[68] This suggests that these new higher sugars are more common to bacteria than previously believed and may provide novel targets for intervention strategies. To date the biological significance of these unique carbohydrate moieties either as a component of LPS/capsules or as a constituent of a glycoprotein remains elusive but their existence in cell surface associated molecules suggests they are likely to contribute significantly in host interactions. In contrast to sialic acids, little is known regarding the biosynthesis of 5,7-diamino-3,5,7,9-tetradeoxynonulosonic acids, and *Helicobacter pylori* may provide a useful model system to determine the biosynthetic pathways and to develop functional assays for the enzymes involved in these glycosylation processes.

4.1.2 Characterization of PTMs on a-type flagellins of *P. aeruginosa*

Results of this study were published previously in Journal of Bacteriology[89] and are further described in this section.

4.1.2.1 Intact mass analysis of *P. aeruginosa* PAK flagellin

The capillary electrospray mass spectrometry analysis of purified flagellin from *P. aeruginosa* PAK revealed a broad and heterogeneous envelope of multiply-charged ions extending from m/z 1000 to 1800. The reconstructed molecular mass profile of the corresponding spectrum (Fig. 4.8) indicated that the observed flagellin mass was 4-7 % higher than that predicted from the *fliC* sequence, 39,905 Da. The molecular mass shift and the heterogeneous pattern of modifications observed here suggest that *P. aeruginosa* PAK flagellin is expressed with a broad distribution of glycoforms. In contrast, the reconstructed molecular mass profile of flagellin from a second strain producing a-type flagellin, JJ692, showed a sharp, well delineated peak at 39,576 Da consistent with the addition of two deoxyhexose residues on the protein backbone. The second minor peak of higher mass represents the sodium adduct of the modified flagellin protein. The assignment of rhamnose as the *O*-linked residue was confirmed by tandem mass spectrometry analyses on the tryptic digest of JJ692 flagellin and GC/MS monosaccharide analysis (see below).

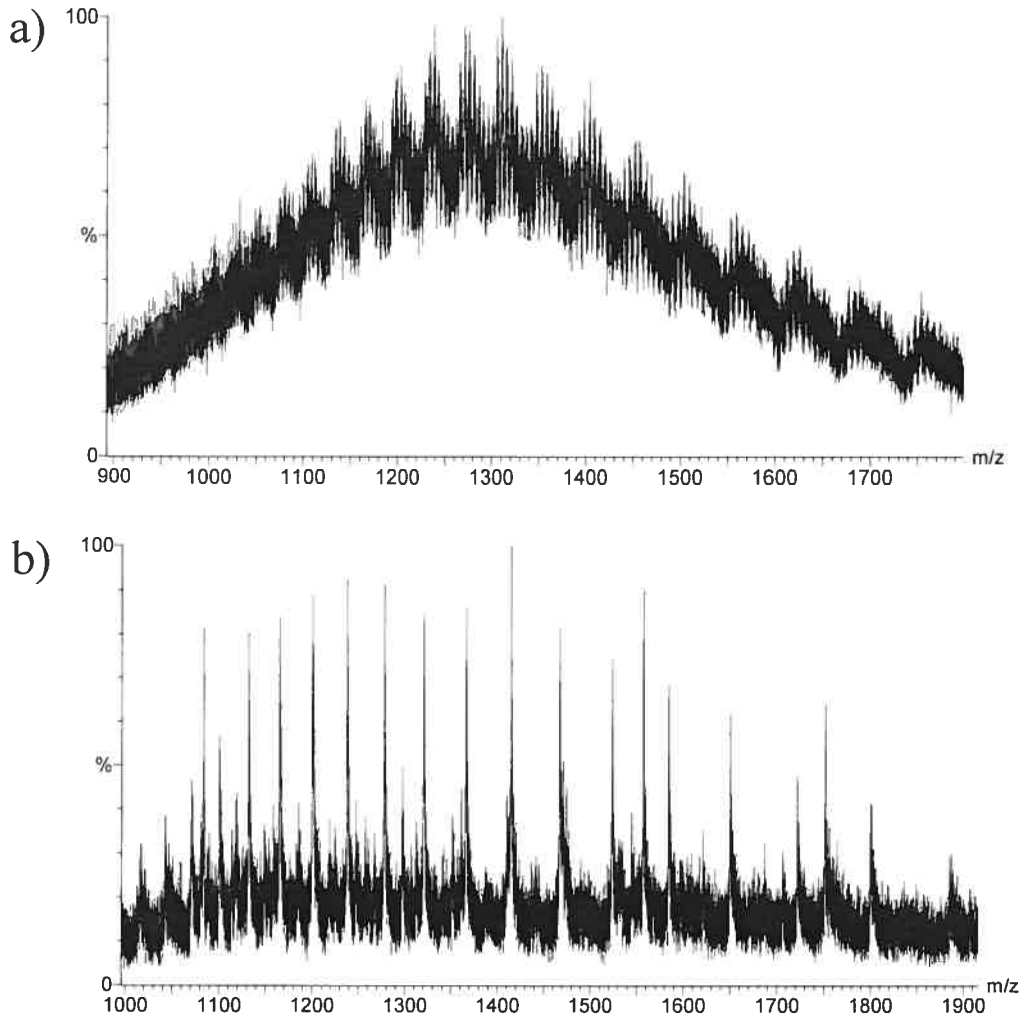

Figure 4.7: Molecular mass determination of intact bacterial flagellin.

a) *P. aeruginosa* PAK b) *P. aeruginosa* JJ692. The reconstructed molecular mass profiles of a) and b) are shown in figure 4.8.

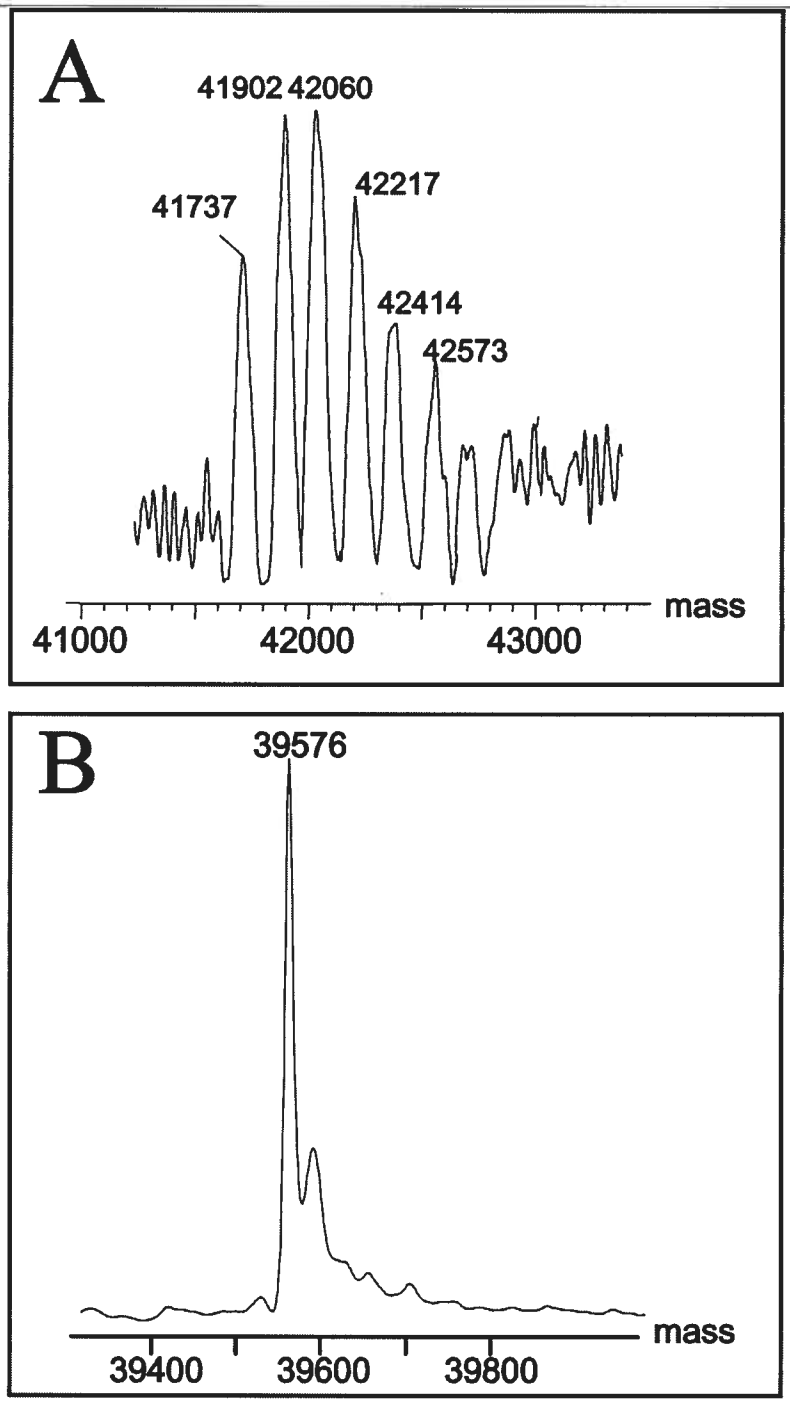

Figure 4.8: Intact mass analysis of *Pseudomonas* flagellins.

The reconstructed molecular mass profile of flagellin from each strain is shown. A). PAK flagellin (M_{theo}: 39905 Da). Broad molecular weight distribution for PAK flagellin indicates extensive heterogeneity in glycoform distribution. B) JJ692 flagellin (M_{theo}: 39288 Da). Flagellin from JJ692 shows a peak at 39576 Da, corresponding to the attachment of two deoxyhexose molecules to the protein.

4.1.2.2 Analysis of the tryptic digest from PAK flagellin

To precisely assign the type and location of the post-translational modification, PAK flagellin protein was digested with trypsin and then analysed by capillary liquid chromatography-nanoelectrospray quadrupole time of flight mass spectrometry. The Q-TOF mass spectrometer was used in data dependent mode, selecting the three most intense multiply charged ions from each survey scan for MS/MS experiments. An initial Mascot search from the acquired MS/MS spectra was performed and provided a sequence coverage of at least 66%, missing only a few small peptides (<800 Da) and two larger peptides ($T_{175-205}$: 2700.4 Da and $T_{250-287}$: 3320.7 Da) in the central region of the flagellin protein.

Comparison of the capLC/MS data before and after ß-elimination showed two ion series, which were not present after ß-elimination anymore (Fig. 4.9), suggesting O-linked glycosylation. MS/MS spectra from the two ion series revealed that the two ion series corresponding to the two peptides $T_{175-205}$ and $T_{250-287}$. The mass spectrometry data indicated that the two unaccounted tryptic peptides were modified by an heterogeneous oligosaccharide chain as reflected by a broad series of triply-protonated peptide ions as exemplified for $T_{250-287}$ in figure 4.9a. Extensive glycosylation on both $T_{175-205}$ and $T_{250-287}$ resulted in a broad series of more than 40 different oligosaccharides appended to each peptide. After HPLC fractionation, glycopeptide ions were subjected to nanospray MS/MS analyses to identify the nature of glycosylation for each intense signal from the two ion series.

These analyses confirmed that both tryptic peptides were modified with oligosaccharides comprising up to 11 monosaccharide residues with a complex micro

heterogeneity. Structural information was limited due to the poor ionisation of the oligosaccharide although partial information on the glycan composition was obtained. A mass difference of 14 Da was typically accounted for by a substitution of a pentose (132 Da) to a deoxyhexose (146 Da) or from a hexose (162 Da) to a hexuronic acid (176 Da) whereas a shift of 16 Da represented a substitution of a deoxyhexose (146 Da) for a hexose (162 Da). Rationalization of the glycan structures observed here is presented in Table 4.1. Several structural features characterized the oligosaccharide chains observed on *P. aeruginosa* PAK flagellin. For example, in all cases the proximal carbohydrate residue through which each glycan moiety is *O*-linked to the peptide backbone has a mass of 146 Da corresponding to the mass of a deoxyhexose sugar, and which was later shown to be rhamnose (see below). Additionally, in the majority of cases the glycan moiety is capped by a trisaccharide consisting of an amino-deoxyhexose, deoxyhexose and an unknown modification of mass 174 Da. The middle region of the glycan displays considerable variability in its structural composition represented by monosaccharides of mass corresponding to pentose, hexose, deoxyhexose and hexuronic acid. No fragments from the middle section of the oligosaccharides were observed. The assignment in Table 4.1 is based on the best fit of the unaccounted mass differences to common sugar combinations and considering the GC/MS monosaccharide analysis of the glycan (see below) A structural model accounting for the glycan chains depicted in Table 4.1 is presented in Figure 4.9.

The monosaccharide (alditol acetate) GC/MS analysis of PAK flagellin protein was performed by Dr. E. Vinogradov (NRC, Ottawa, Canada), who identified

rhamnose, mannose, glucose and 4-amino-4,6-dideoxyglucose (viosamine) as major components. Trace levels of ribose and arabinose were also present. Uronic acids were not determined. Monosaccharide analysis of flagellin from JJ692 confirmed that the deoxyhexose sugar identified by nanospray MS/MS on this flagellin was indeed rhamnose (data not shown).

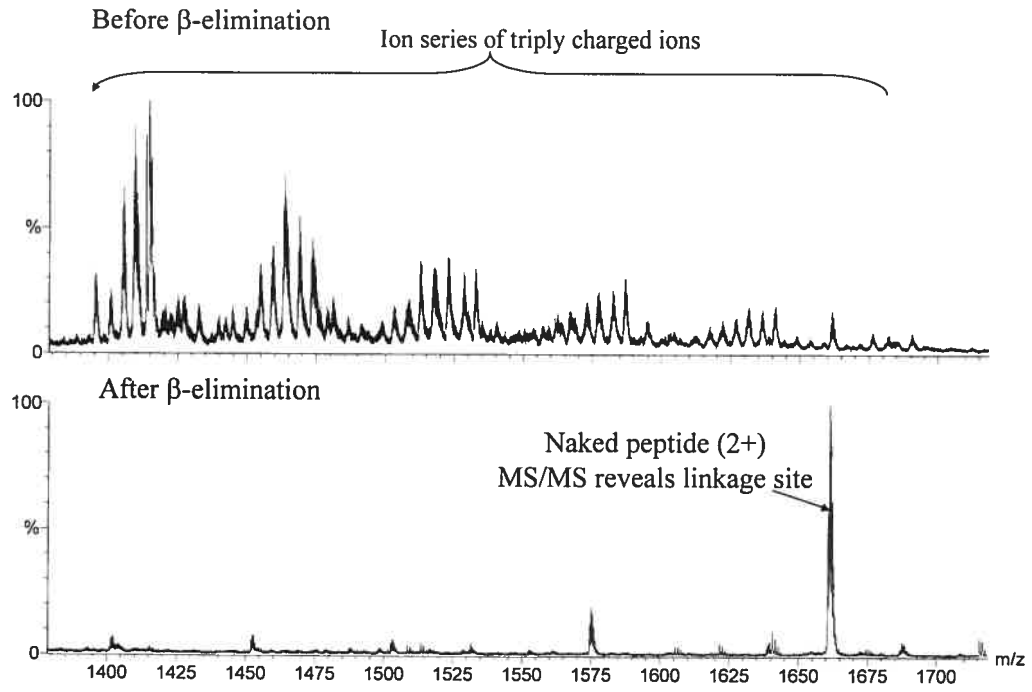

Figure 4.9: Ion spectra of glycosylated tryptic peptide T_{250}-T_{287}. before and after ß-elimination. a) Observed triply charged ion series for the peptide T_{250}-T_{287}. Each signal corresponds to a different oligosaccharide attached to the peptide. For information on the oligosaccharides see Table 4.1. b) After ß-elimination the ion series was not present anymore and instead the naked peptide was observed. MS/MS of this peptide revealed the linkage site (Fig. 4.11)

Table 4.1: Oligosaccharide chains found in *P. aeruginosa* PAK flagellin.

m/z Glycopeptide (3x charged)	Mass Oligosaccharide	Proximal carbohydrate	Middle section Mass of remaining oligosaccharide with a possible CHO structure	Terminal end (assignment based on fragments from MS/MS data)			
1350,9	729,1	dHex	292 dHex,dHex	dHexN	dHex		
1394,9	861,1	dHex	424 dHex,dHex,Pen	dHexN?	dHex		
1404,9	891,1	dHex	454 dHex,dHex,Hex	dHexN?	dHex		
1408,9	903,1	dHex	292 dHex,dHex	dHexN	dHex	174	
1414,2	919,1	dHex	308 dHex,Hex	dHexN	dHex	174	
1418,9	933,1	dHex	322 dHex,HexA	dHexN	dHex	174	1. Cluster
1424,2	949,1	dHex	338 Hex,HexA	dHexN	dHex	174	
1428,9	963,2	dHex	352 HexA,HexA	dHexN	dHex	174	
1452,9	1035,2	dHex	424 Pen,dHex,dHex	dHexN	dHex	174	
1458,3	1051,3	dHex	440 Pen,dHex,Hex	dHexN	dHex	174	
1463,0	1065,3	dHex	454 dHex,dHex,Hex	dHexN	dHex	174	
1468,3	1081,2	dHex	470 dHex,Hex,Hex	dHexN	dHex	174	2. Cluster
1472,9	1095,1	dHex	484 dHex,Hex,HexA	dHexN	dHex	174	
1478,3	1111,3	dHex	500 Hex,Hex,HexA	dHexN	dHex	174	
1502,3	1183,3	dHex	572 Pen,Pen,dHex,Hex	dHexN?	dHex	174	
1507,0	1197,4	dHex	586 Pen,dHex,dHex,Hex	dHexN?	dHex	174	
1512,3	1213,4	dHex	602 Pen,dHex,Hex,Hex	dHexN	dHex	174	
1517,0	1227,3	dHex	616 dHex,dHex,Hex,Hex	dHexN	dHex	174	3. Cluster
1522,3	1243,3	dHex	632 dHex,Hex,Hex,Hex	dHexN	dHex	174	
1527,7	1259,4	dHex	646 Hex,Hex,Hex,Hex	dHexN	dHex	174	
1532,3	1273,3	dHex	662 Hex,Hex,Hex,HexA	dHexN	dHex	174	
1556,3	1345,3	dHex	734 Pen,Pen,dHex,Hex,Hex	dHexN?	dHex	174	
1561,0	1359,3	dHex	748 Pen,dHex,dHex,Hex,Hex	dHexN?	dHex	174	
1566,4	1375,5	dHex	764 Pen,dHex,Hex,Hex,Hex	dHexN	dHex	174	
1571,7	1391,5	dHex	780 Pen,Hex,Hex,Hex,Hex	dHexN	dHex	174	4. Cluster
1576,4	1405,5	dHex	794 dHex,Hex,Hex,Hex,Hex	dHexN?	dHex	174	
1581,7	1421,5	dHex	810 Hex,Hex,Hex,Hex,Hex	dHexN?	dHex	174	
1586,4	1435,6	dHex	824 Hex,Hex,Hex,Hex,HexA	dHexN?	dHex	174	
1610,4	1507,6	dHex	896 Pen,Pen,dHex,Hex,Hex,Hex	dHexN?	dHex	174	
1620,4	1537,5	dHex	926 Pen,dHex,Hex,Hex,Hex,Hex	dHexN?	dHex	174	
1625,8	1553,7	dHex	942 Pen,Hex,Hex,Hex,Hex,Hex	dHexN?	dHex?	174	
1630,4	1567,7	dHex	956 dHexHex,Hex,Hex,Hex,Hex	dHexN?	dHex	174	5. Cluster
1635,7	1583,6	dHex	972 Hex,Hex,Hex,Hex,Hex,Hex	dHexN?	dHex	174	
1640,4	1597,6	dHex	986 Hex,Hex,Hex,Hex,Hex,HexA	dHexN?	dHex	174	

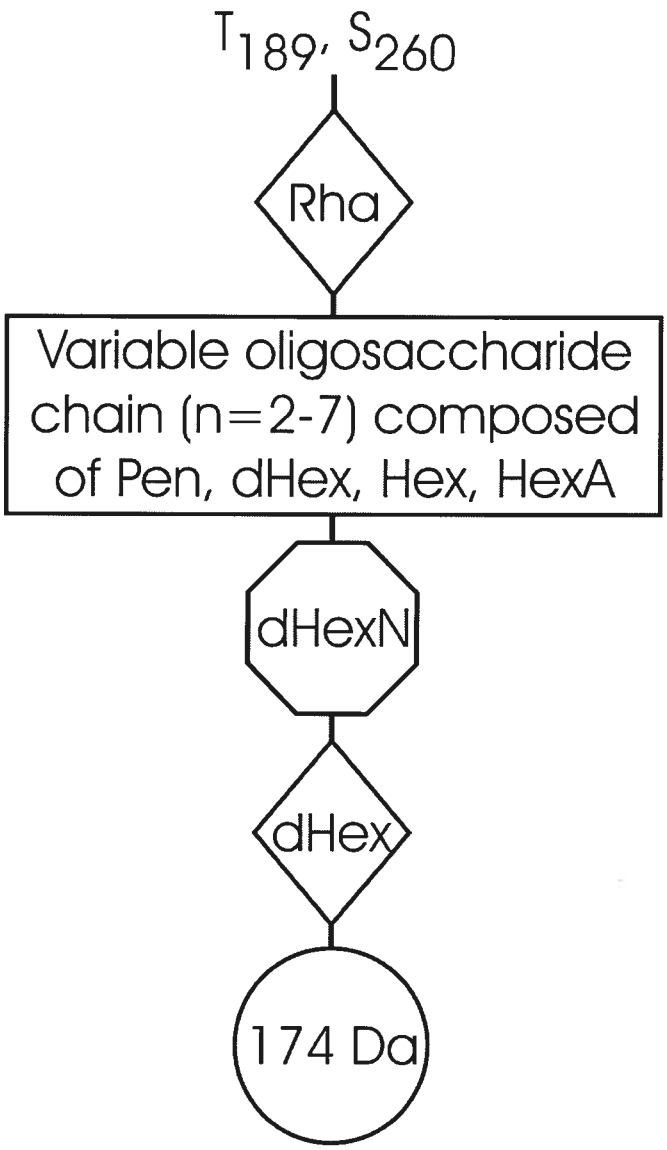

Figure 4.10: Structural assignment of oligosaccharide chains found on PAK flagellin. The residues Rha, Pen, Hex, dhexN correspond to rhamnose, pentose, hexose, deoxyhexosamine.

4.1.2.3 Determination of glycan attachment site

To determine the precise site of attachment of the glycan moiety to the modified peptides, purified glycopeptide fractions were subjected to base catalysed hydrolysis in the presence of NH_4OH whereby the β-elimination product incorporated a newly formed amino group of distinct mass (-1 Da) which can be detected by mass spectrometry. Accordingly, the amino acids Thr_{189} and Ser_{260} were modified with an O-linked glycan (see Fig. 4.12). The tandem mass spectra of the tryptic peptide precursor ion $T_{250\text{-}287}$ before and after β-elimination is shown in figure 4.11. It shows the tryptic glycopeptide $T_{250\text{-}287}$ modified with an O-linked glycan identified by the singly-charged fragment ion at m/z 892.5. The naked doubly-charged peptide fragment ion is observed at m/z 1661.4 together with a doubly charged fragment ion at m/z 1734.5 corresponding to the addition of a dHex residue. The singly-charged y-type fragment ions at m/z 792.5, 863.5, 962.5 and 1019.6 are consistent with the sequence assignment for $T_{250\text{-}287}$ whereas oxonium ions extending from m/z 601.4, 746.5 and 892.5 (identified by asterisk) correspond to backbone cleavage products of the hexasaccharide 4 x dHex dHexN Hex. In Fig. 4.11b the spectrum of the peptide after ß-elimination is shown. The sequence assignment of $T_{250\text{-}287}$ is validated by the b-type fragment ion series shown in the expanded region of the mass spectrum (inset). The site of the O-linked attachment to Ser_{260} is confirmed by the identification of b ions shifted by -1 resulting from the substitution of the hydroxyl Ser side chain for an amine group (see below).

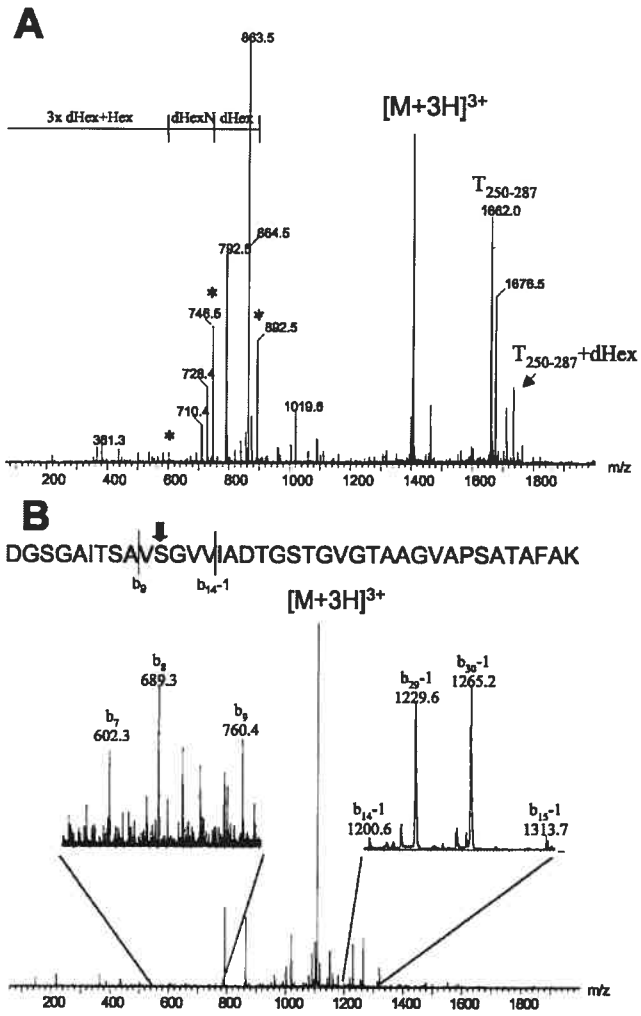

Figure 4.11: Determination of the carbohydrate attachment site on peptide $T_{250-287}$.

A) The spectrum shows the tryptic glycopeptide $T_{250-287}$ modified with an O-linked glycan of mass 891.5 Da. The naked doubly-charged peptide fragment ion is observed at m/z 1661.4 together with a doubly charged fragment ion at m/z 1734.5 corresponding to the addition of a dHex residue. Oxonium ions extending from m/z 601.4, 746.5 and 892.5 (identified by asterisk) correspond to backbone cleavage products of the hexasaccharide dHex4, dHexN, Hex.

B) Tandem mass spectrum of $[M+3H]^{3+}$ at m/z 1107.6 corresponding to the ß-elimination product of the glycopeptide shown in Fig. 4.11a. The sequence assignment of T250-287 is validated by the b-type fragment ion series shown in the expanded region of the mass spectrum (inset). The site of the O-linked attachment to Ser260 is confirmed by the identification of b ions shifted by -1 resulting from the substitution of the hydroxyl Ser side chain for an amine group.

ALTVNTNIAS	LNTQRNLNNS	SASLNTSLQR	LSTGSRINSA	KDDAAGLQIA	50
NRLTSQVNGL	NVATKNANDG	ISLAQTAEGA	LQQSTNILQR	MRDLSLQSAN	100
GSNSDSERTA	LNGEVKQLQK	ELDRISNTTT	FGGRKLLDGS	FGVASFQVGS	150
AANEIISVGI	DEMSAESLNG	TYFKADGGGA	VTAATASGTV	DIAIGITGGS	200
AVNVKVDMKG	NETAEQAAAK	IAAAVNDANV	GIGAFSDGDT	ISYVSKAGKD	250
GSGAITSAVS	GVVIADTGST	GVGTAAGVAP	SATAFAKTND	TVAKIDISTA	300
KGAQSAVLVI	DEAIKQIDAQ	RADLGAVQNR	FDNTINNLKN	IGENVSAARG	350
RIEDTDFAAE	TANLTKNQVL	QQAGTAILAQ	ANQLPQSVLS	LLR	400

Mr: 39905 Da

Figure 4.12: Assignment map of PAK flagellin. The sites of O-linked glycosylation are boxed. Peptides identified by MS/MS analysis are presented in bold whereas missing peptides are indicated in normal font and are underlined.

4.1.2.4 Functional characterization of genes involved in flagellar glycosylation

Mass spectral analysis of intact flagellin from PAK and JJ692 revealed significant differences in the extent of glycosylation found on the respective flagellins (Fig. 4.8). These two primary amino acid sequences are different only by a few amino acid substitutions and two small deletions of codons for 3 and 4 amino acids at positions 254 and 279, respectively (GI accession # AAP33174 , P21184) none of which involves the sites of glycan attachment. However, genomic comparisons of the respective glycosylation islands have revealed a substantial deletion in strain JJ692.[51] Most notably, the JJ692 genomic island is approximately 5.4 kilobase pairs shorter and five *orfs* that are present in the PAK island, *orfs I,J,K,L,M* are absent in JJ692 and a frameshift mutation in *orfE* results in premature termination and which results in a polypeptide of 30aa as compared to PAK where the predicted *orfE* protein is 211aa long. Arora *et al.*[19] suggested that the proteins encoded within the glycosylation island shared significant homology to enzymes involved in the synthesis, activation and polymerization of sugars. In particular, the *orfA* gene shows 58% identity with *vioA* of *E. coli*, which encodes a nucleotide sugar aminotransferase involved in the biosynthesis of viosamine. The protein encoded by *orfN*, which displays homology to a number of glycosyltransferases[19] is likely involved in the attachment of a glycan moiety to the flagellin protein. To ascertain the role of these two genes in flagellin glycosylation, chromosomal mutations were made in *orfA* and *orfN* of the PAK glycosylation island . Flagellin purified from these PAK isogenic mutants were then

examined by mass spectrometry. Intact mass analysis revealed that glycosylation of PAK flagellin had been considerably reduced and the extensive heterogeneity observed in wild type flagellin was no longer apparent in either mutant (Fig. 4.13). Instead, in the case of *orfA* isogenic mutant, the major flagellin protein had an intact mass of 40,194 Da corresponding to the protein with two rhamnose residues. MS/MS analysis of the tryptic digest confirmed the attachment of a single rhamnose at the same two glycosylation sites found on PAK flagellin. Additionally, a second minor species (20%) of modified flagellin was observed of mass 40,341 Da which corresponds to a third deoxyhexose attachment. capLC/MS/MS analyses of the tryptic digest confirmed the addition of this residue to either of the two sites identified by β-elimination (data not shown). It appears therefore that insertional inactivation of *orfA* results in an inability to synthesize the heterogeneous glycan observed in the parent flagellin although rhamnose attachment to serine and threonine is unaffected.

In the case of the isogenic mutant of *orfN* which had significant homology to a putative glycosyltransferase, *rfbC*,[90] three distinct intact masses were apparent for this flagellin sample (Fig. 4.13b). The major peak corresponded to flagellin of the predicted mass based on amino acid sequence with no glycosylation. The two other minor peaks corresponded to modification with either a single rhamnose or two rhamnose molecules. In this case, the ability to glycosylate flagellin is severely impaired with only limited addition of the rhamnose monosaccharide.

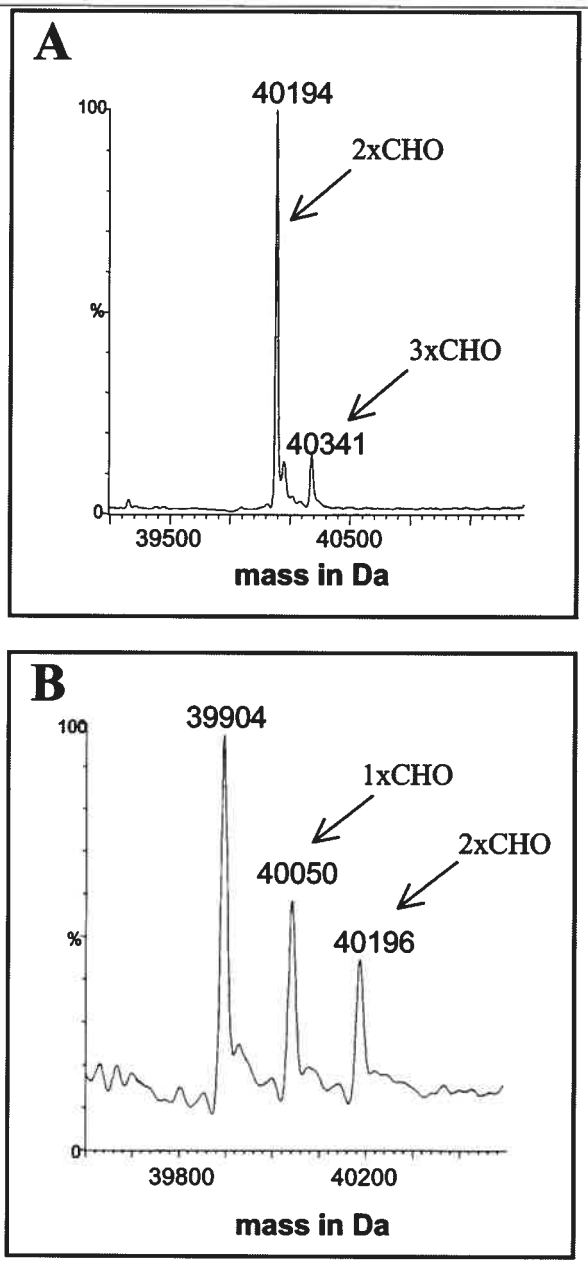

Figure 4.13: Intact mass analysis of flagellin from PAK glycosylation island isogenic mutants. The reconstructed molecular mass profile of flagellin from PAK isogenic mutants *orfA* (A) and *orfN* (B)

4.1.2.5 Discussion

In this report we have determined the structural basis of flagellin glycosylation in *P. aeruginosa* strains producing a-type flagellin. Flagellin from two strains, PAK and JJ692, which contain a long and a short glycosylation island (GI) respectively, were examined and each strain was shown to produce a unique pattern of glycosylation on their respective flagellins. In both cases the glycosyl moiety is *O*-linked through a rhamnose residue. The long glycosylation island found in PAK contains the additional genetic information responsible for the synthesis of a heterogeneous glycan of up to 11 monosaccharide residues while flagellin from strain JJ692 is glycosylated with a single rhamnose residue at each site. This short glycosylation island has been found on a number of strains but the *orfs* that are found in JJ692 do not necessarily represent the glycosylation potential of all strains carrying short islands due to the mutations found in this strain. Other strains with this short island show more extensive glycosylation as judged by SDS-PAGE of flagellin.[51] However, their glycosylation islands have not been sequenced.

The level of glycosylation found on the flagellin of *Pseudomonas* is notably different from that found on flagellin from two other Gram negative, polarly flagellated organisms. *Campylobacter* is the most heavily glycosylated prokaryotic protein identified to date with at least 16 sites/monomer while *Helicobacter* flagellin has been shown to have 7 sites of glycosylation on the FlaA protein monomer and 10 sites on the FlaB protein (see section 4.1.1). In contrast, the flagellin from *P. aeruginosa* has only two sites of modification, although the modified residues are

located in the central, surface exposed, variable region rather than the highly conserved N and C terminal regions of the protein, as was the case for *Campylobacter* and *Helicobacter* flagellin. Again, the site of attachment appears not to be based on sequon specificity but may reflect local structural restraints or surface accessibility of individual serine and threonine residues in the folded protein.

The complex heterogeneity displayed on the glycan of PAK flagellin is indicative of significant differences in the biosynthetic pathway as compared to that described for a second major *Pseudomonas* cell surface glycoprotein; pilin.[47] In the pilin glycosylation process no heterogeneity in glycan structure was observed and only an intact O-chain repeat unit was present on the pilin monomer. It has been shown that the undecaprenyl bound O-antigen repeat unit is utilized as the substrate for pilin glycosylation and so LPS assembly and pili production are therefore closely connected. In contrast, in the case of *Pseudomonas* flagellin glycosylation, no serotype O:6 LPS O-antigen repeat unit was found in the flagellar glycan of PAK and it appears that a novel biosynthetic pathway is utilized in the production of this glycan. Glycosylation of the flagellin protein appears to occur by an alternative mechanism, one of which at least partially involves the sequential attachment of individual sugars. Such a process would lead to the complex heterogeneity in glycan composition observed at each site. Chemical analysis of the flagellin glycan revealed significant quantities of the deoxyhexose monosaccharide, rhamnose a component also found in the LPS of O:6 serotypes.[91] Significant levels of mannose, glucose and the novel 4-amino-4,6 dideoxyhexose, viosamine sugar were also identified by GC/MS analysis. However, both GC/MS analysis of the flagellin glycan and mass

measurements of glycan fragment ions confirmed, that other O-chain specific monosaccharides of LPS appear not to be components of the heterogeneous flagellin glycan.

Polymorphism in the glycosylation islands from numerous strains of *Pseudomonas* has been extensively analysed by microarray analysis although the precise function of individual genes from this locus was unknown.[51] In this study we have demonstrated a functional role of GI genes in flagellar glycan production. Structural analysis of flagellin from strains where *orfA* and *orfN* genes were mutated in the glycosylation island of PAK clearly demonstrate the role of this locus in the biosynthesis of this glycan moiety. While addition of the rhamnose residue is unaffected in an *orfA* mutant, no heterogeneous glycan is added. *OrfA* shows homology to an LPS gene from *E. coli* which is responsible for the biosynthesis of 4-amino-4,6-dideoxyhexose, viosamine[92] through nucleotide activation and the presence of this monosaccharide in the flagellar glycan capping trisaccharide has been confirmed in this study. It appears that the *orfA* protein from PAK is indeed responsible for the activation of this unique sugar.

The protein encoded by *orfN* appears to be responsible for the addition of the deoxyhexose sugar, rhamnose to the protein backbone. The identification of rhamnose as the linkage sugar on the flagellin protein expands the known list of sugar-amino acid linkages as described by Spiro.[93] Flagellin from an *orfN* mutant was predominantly unglycosylated. We hypothesize that the residual rhamnosyl transferase activity seen is likely due to the activity of a second rhamnosyl transferase elsewhere in the genome which may indeed utilize the same or related substrate but

has only limited specificity for the protein acceptor. The rhamnosyl transferases utilized in LPS biosynthesis or alternatively RhlB and C proteins from the rhamnolipid biosynthetic pathway are potential candidates for this limited activity .[94-95]

In this study we were unable to provide a structural assignment for the unique fragment ion corresponding to 174 Da. Based on the observed mass a possible assignment could be a diamino hexuronic acid or a trideoxynonulose although the poor ionization indicates such an assignment is unlikely. The annotations of the remaining genes in the PAK glycosylation island are based on relatively ill defined functions which include fatty acid biosynthesis, naphthalene catabolic pathway and nodulation factor biosynthesis and as such provide few clues as to the likely structure of this unique flagellar glycan component. However, *P. aeruginosa* produces extracellular glycolipids composed of L-rhamnose and 3-hydroxyalkanoic acid known as rhamnolipids which are recognized as a major virulence factors in cystic fibrosis patients.[96] The biosynthetic pathway of this glycolipid has recently been investigated and precursors of the pathway may indeed be utilized in the biosynthesis of the 174 Da modification found on the flagellar glycan linked to rhamnose.[97] It is important to note that as the distal residue of the capping structure, it will undoubtedly be significant in any type of interactions based on this flagellar glycan.

We have also shown that the composition of the GI determines the type of glycosylation found on the respective flagellins. JJ692 which possesses a truncated version of the island and lacks *orfsI, J, K, L, M* and displays polymorphisms in *orfs D,E, H* and *N* produces flagellin glycosylated with only rhamnose. In contrast, the

large glycosylation island found in strain PAK is capable of the addition of the heterogeneous glycan to either the A1 type flagellin of PAK or the A2 type flagellin of JJ692.

Interestingly, in *Pseudomonas* the absence of glycosylation of the flagellin protein does not lead to a loss of flagellar filament assembly and coincidental loss of motility. Such a situation is the case for two other well studied polarly flagellated, motile organisms. Both *Campylobacter* and *Helicobacter* become non motile and are unable to synthesise a flagellar filament when the glycosylation process is prevented (see section 4.1.1).[24] In contrast to *P. aeruginosa*, the flagellar filament of these two organisms is complex and comprised of two flagellin monomeric proteins FlaA and FlaB and the level of glycosylation per protein monomer is substantially higher. It remains to be established if either of these features are the reason for a more dramatic effect on flagellar assembly when glycosylation is inhibited.

The biological significance of each unique flagellar glycan structure can now be explored. As with *Campylobacter* strains which also display considerable heterogeneity in glycan structure[99] it is currently unknown if the diversity in glycan structure found on individual *Pseudomonas* strains confers a unique advantage in particular environments or indeed plays a role in modification of host immune responses to the flagellin protein. Specifically, in terms of the innate immune response, it has been shown that TLR5 recognition sequences are localized within the conserved D1 domain of bacterial flagellins.[33,100,101] It remains to be established if indeed steric hindrance due to glycan size may prevent this type of receptor interaction in flagellins which are glycosylated. While *O*-linked glycosylation is now

recognized as a common prokaryotic process, the diverse nature of glycan structures identified is suggestive of unique biological roles. Flagellin glycosylation in the plant pathogen *P. syringea* has recently been shown to be involved in determining plant host specificity.[102] In the case of flagellar glycosylation in *P. aeruginosa*, it is now important to investigate the diversity in flagellar glycan structure amongst isolates and the role of these structures in clinical outcomes of infection.

4.1.3 Characterization of PTMs on b-type flagellin of *P. aeruginosa*

4.1.3.1 Intact mass analysis of P. aeruginosa PAO flagellin

The mass spectrometry analysis of purified intact flagellin from strain PAO resulted in the typical multiply charged envelope for intact proteins (Fig. 4.14). Deconvolution revealed three sharp, well delineated peaks at 49402 Da, 49611 Da and 49820 Da, corresponding to excess masses of 291 Da, 500 Da and 709 Da from the predicted cDNA sequence (Fig. 4.15), indicating heterogeneity in the post-translational modifications. The peak at 49402 Da is consistent with the addition of two deoxyhexose residues to the flagellin protein, as it was observed for the flagellin of *P. aeruginosa* strain jj692. The other two peaks correspond to the addition of one 209 Da residue to each of the two deoxyhexose residues. (see below)

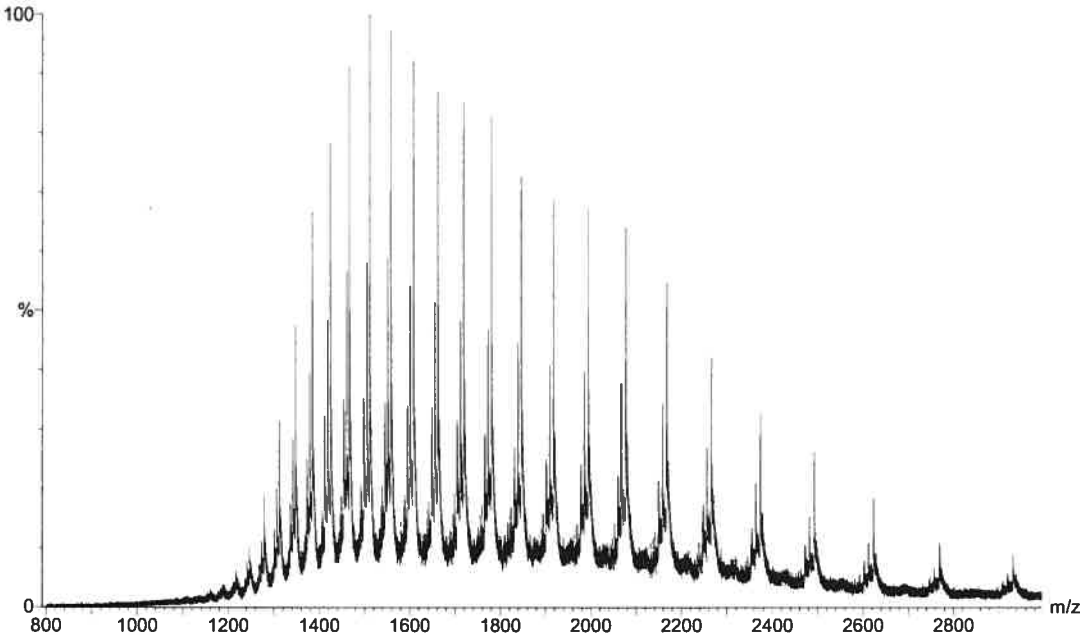

Figure 4.14: Molecular mass determination of intact bacterial flagellin of *P. aeruginosa* PAO.

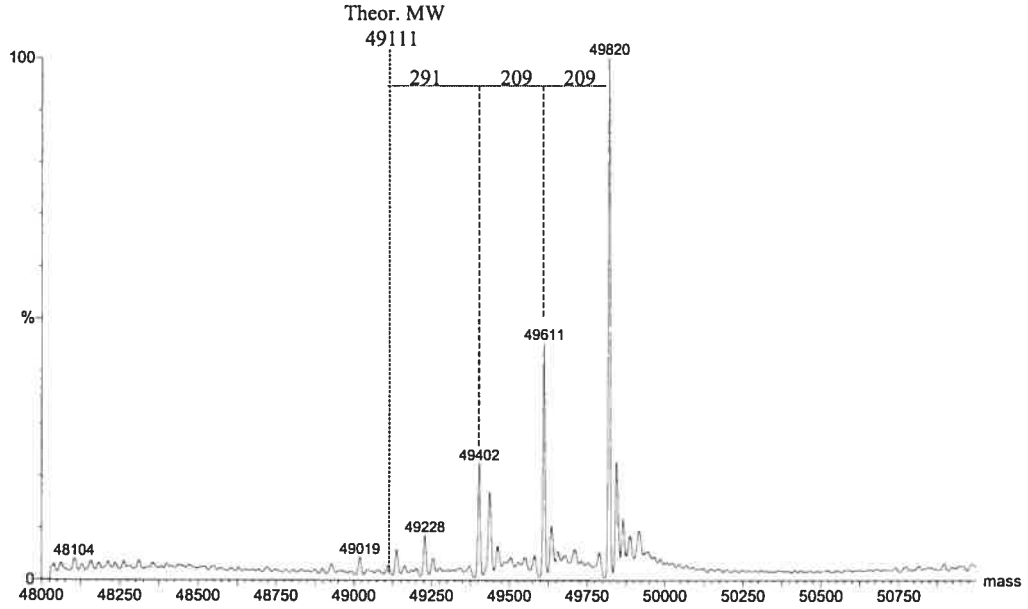

Figure 4.15: Deconvoluted mass spectrum of intact flagellin from *P. aeruginosa* b-type PAO.

The reconstructed molecular mass profile of flagellin for strain PAO is shown.

4.1.3.2 capLC/MS/MS analysis of chymotryptic/tryptic digest of P. aeruginosa PAO flagellin

To identify and characterize the post-translational modification on the flagellin protein, the protein was tryptic digested and analysed by capLC/MS/MS on a Q-TOF Ultima^{TM}. The acquired MS/MS spectra were inspected manually for post-translational modifications, but no typical glycosylation spectra were observed. However, in the central region of the flagellin protein the expected tryptic peptide $T_{136-207}$ of mass 6809.3 Da was not identified. In general peptides of this size are difficult to detect and require more extensive digestion with chymotrypsin. Chymotrypsin cleaves at the carboxyl side of phenylalanine, tryptophan, tyrosine, isoleucine, valine and leucine. Analysis of the chymotryptic/tryptic digest identified peptides modified with either a deoxyhexose as found in strain jj692 or with an abundant oxonium ion at m/z 356 Da (Fig. 4.16). This figure shows the peptide GTATASGIASGTVNLVGGQVK modified with two 355 Da residues. The mass spectrum is dominated by the oxonium ion at m/z 356 and by signals corresponding to the loss of 129 Da and 226 Da from the precursor mass. The loss of 226 Da corresponds probably to a deoxyhexose plus a phosphate group, but the nature of the 129 Da modification is still unknown. The second generation fragment ion spectrum of the oxonium ion at m/z 356 is shown in figure 4.17. It is noteworthy that no loss of H_3PO_4 is observed suggesting that the phosphate group could be linked to the carbohydrate and the unknown 129 Da residue via a phosphoester bond. An exact mass measurement was performed using the "top down" approach (see section 4.2).

The exact mass of the 356 residue was determined to be 356.112 Da +/- 10 ppm (Fig. 4.18). Fragment ions from the N-terminal end of the protein backbone were used for internal calibration enabling the very precise determination of the molecular mass. Based on the assumption that the post-translational modification consist of a deoxyhexose and a phosphate group the only possible elemental formula for the 129 Da residue is $C_6H_{12}NO_2$. Possible structures therefore would be for example leucine/isoleucine or an aminohexanoic acid attached to the deoxyhexose and phosphate group. However, the true nature of the modification cannot be determined solely by mass spectrometry.

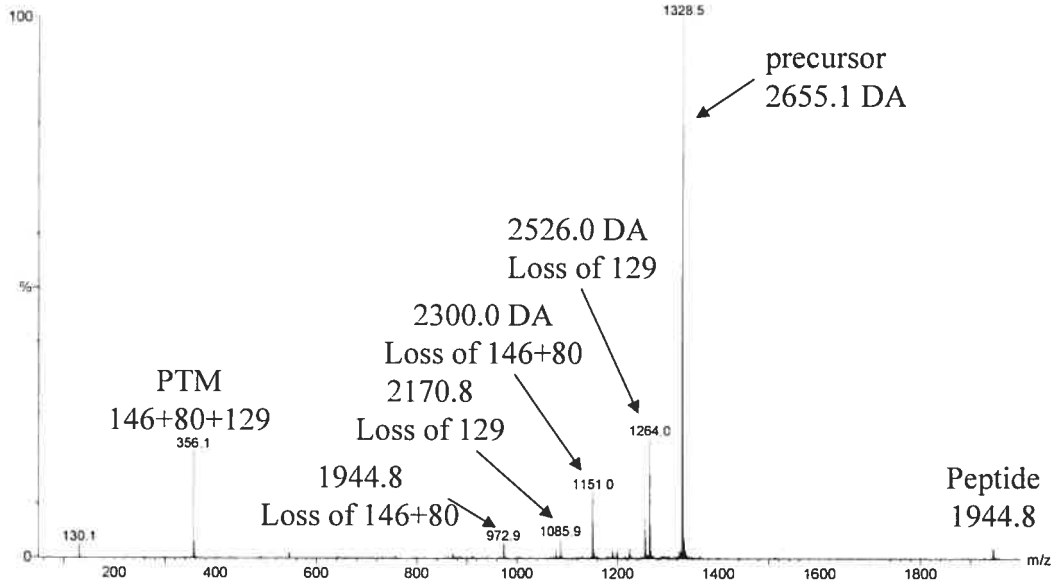

Figure 4.16: Tandem MS/MS spectra of m/z 1327.0.

The MS/MS spectrum corresponds to the peptide GTATASGIASGTVNLVGGQVK modified with two 355 residues. In the low mass region an oxonium ion at m/z 356 from the PTM is observed.

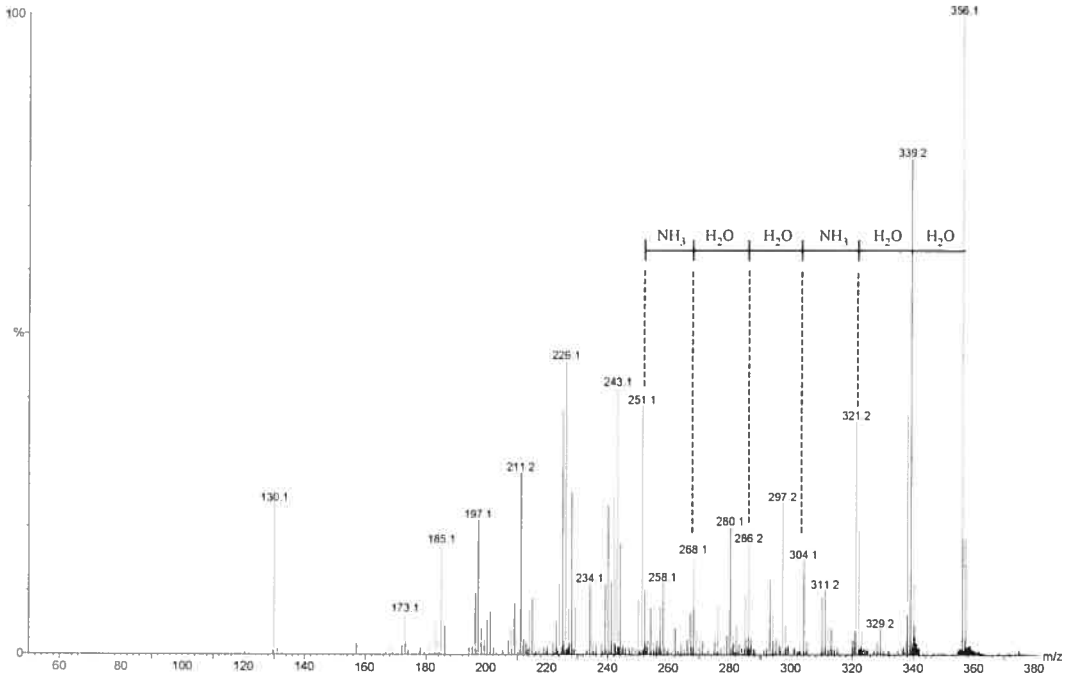

Figure 4.17: Second generation fragment ion spectrum of m/z 356.

The MS/MS spectrum shows several times the loss of H_2O which is common for carbohydrates. In the spectrum no loss of 98 is observed, suggesting that the phosphate group is not terminally attached to the carbohydrate.

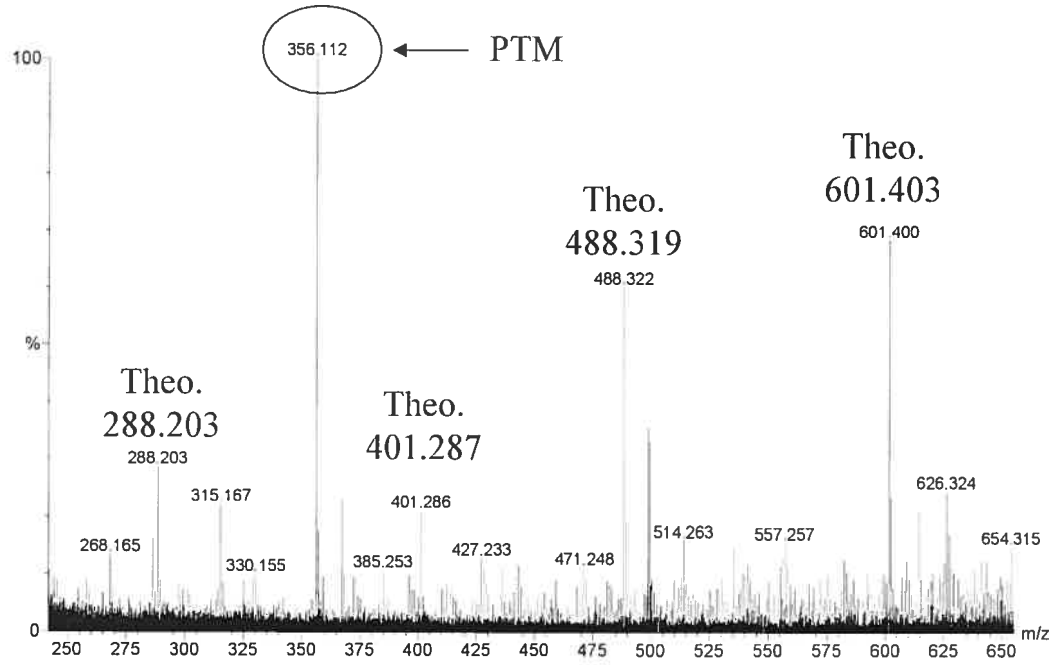

Figure 4.18: Exact mass measurement of the oxonium ion at m/z 356.
N-terminal y-fragment ions from the protein backbone were used for internal calibration. The theoretical values for the fragments are shown above the peaks.

4.1.3.3 Determination of glycan attachment site

To precisely determine the site of attachment the tryptic/chymotryptic digests were subjected to ß-elimination by NH_4OH. This analysis identified serine (T^{191}) and (T^{198}) to be modified with an O-linked glycan. The tandem MS/MS spectra of the ß-eliminated chymotryptic/tryptic peptide $T_{172\text{-}207}$ is shown in figure 4.19. The y-fragments showed after y_9 an increase of +1 Da resulting from deamidation of asparagine to aspartic acid. The y_{14} and y_{17} ions showed respectively a mass shift of –1 Da from the predicted mass indicating that serine (T^{191}) and (T^{198}) are glycosylated with the glycan.

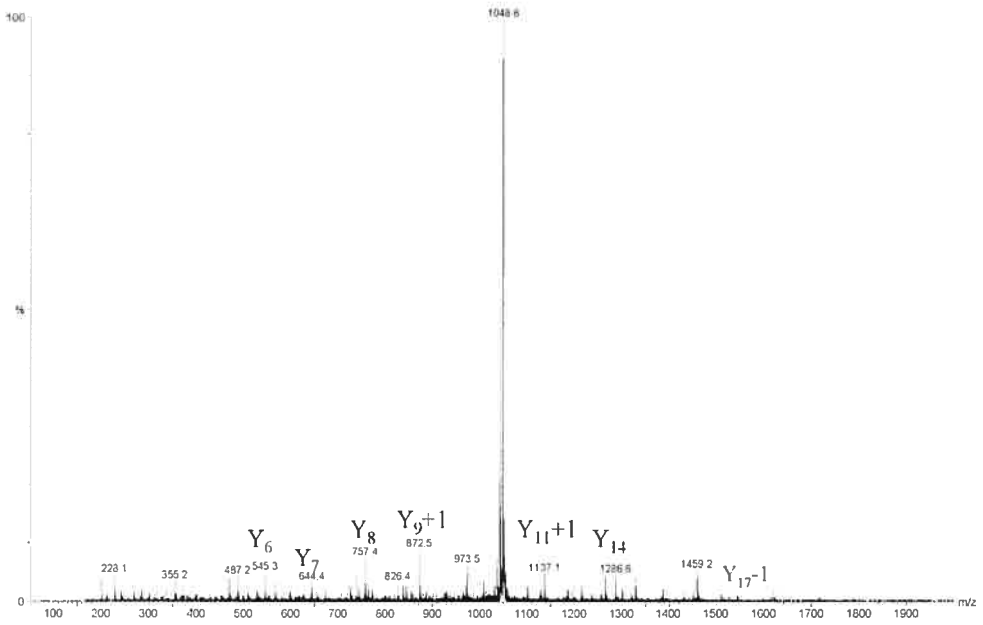

Figure 4.19: MS/MS spectrum of m/z 1048.3 after ß-elimination. The MS/MS spectrum of the peptide QVGSNGAGTVASVAGTATASGIASGTVNLVGGGQVK after ß-elimination is shown. Fragment ions showing a mass shift of either 1 or 2 m/z units enabled the identification of the linkage site of the glycan.

4.1.3.4 Functional characterization of genes involved in flagellar glycosylation

In contrast to strain PAK which has an extended glycosylation island containing 14 genes, the equivalent region in PAO has been shown to comprise only 4 genes, PA1088, PA1089, PA1090 and PA1091. Mutations were made in all four genes and their flagellins were analysed by infusion mass spectrometry. Intact mass analysis of PA1088 and PA1089 flagellins revealed that the glycosylation process was partially disrupted. Flagellin from PA1088 displayed a mass of 49402 Da corresponding to the flagellin protein modified with two rhamnose residues. In

contrast, flagellin from PA1089 gave a heterogeneous profile with two major peaks corresponding to flagellin with intact glycosylation (49819 Da) and flagellin modified with only two deoxyhexose residues (49400 Da). In contrast, we were unsuccessful in obtaining results for the flagellin isolated from PA1090 mutant. While we are uncertain of the reason for this, it should be noted that yields of flagellin from this particular mutant were much reduced and the sample may contain significant amounts of membrane material as a consequence of the mutation having a more drastic effect on membrane stability. Flagellin protein from the fourth isogenic mutant (PA1091, rfbC) revealed a single peak (49409 Da) corresponding to flagellin which is unglycosylated (Mr predicted 49111 Da).

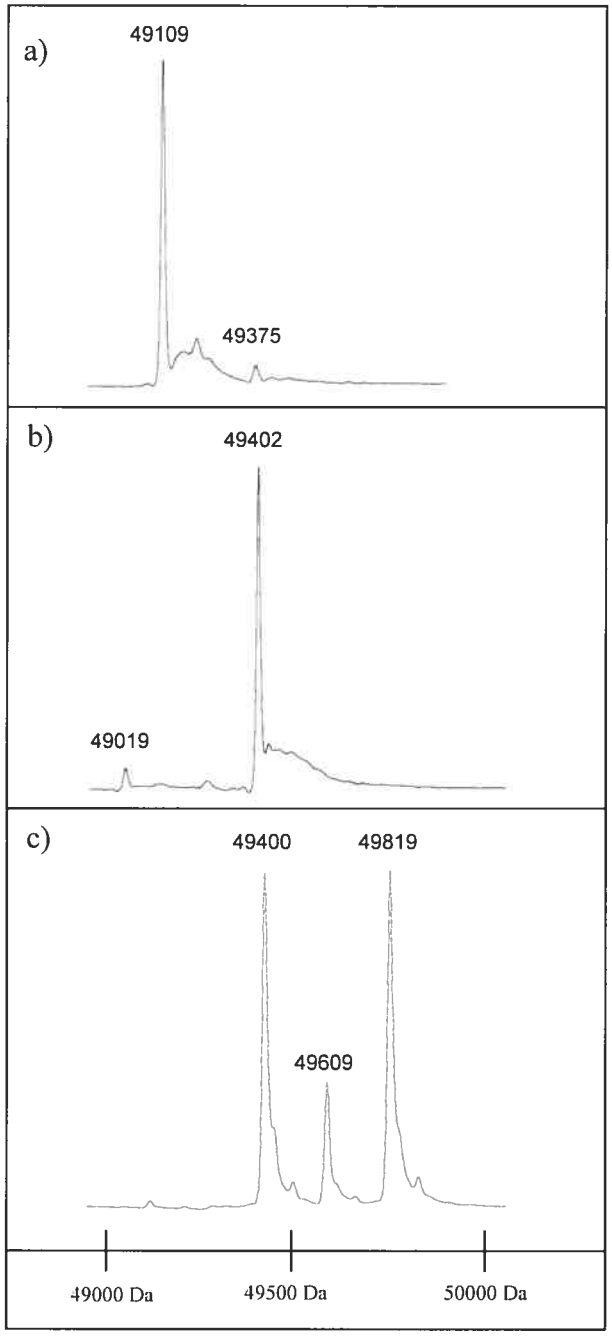

Figure 4.20: Intact mass analysis of flagellin from isogenic mutants a) PA1091 b) PA1088 and c) PA1089.

4.1.3.5 Discussion

To the best of our knowledge, this is the first report showing that b-type flagellin of *P. aeruginosa* PAO is glycosylated. It was shown by mass spectrometry analyses that the flagellin of the b-type strain PAO of *P. aeruginosa* is O-linked glycosylated at two sites either with a deoxyhexose residue or an unknown glycan of mass 356 Da. The structure of the glycan could not be revealed unambiguously but from the fragmentation pattern and the exact mass measurement, it may consists of a deoxyhexose a phosphate group and an unknown modification with the elemental formula $C_6H_{12}NO_2$. The presence of the phosphate group is supported by a report from Kelly-Winterberg et al..[103] They reported that the flagellin of PAO was phosphorylated on tyrosine residues. As methods they used ^{32}P labelling and an antibody against phosphotyrosine which showed a positive reaction. However, all tyrosines in the flagellin protein were identified as non-modified. The used antibody probably cross reacted with the phospho group in the glycan.

Isogenic mutants in the four genes in the glycosylation island of strain PAO were analyzed by mass spectrometry and for 3 of the 4 mutants results were obtained. For genes PA1088 and PA1089 the glycosylation process was partially disrupted and for gene PA1091 no glycan was attached anymore to the flagellin protein. As we had shown with PAK, inactivation of this ORF (rfbC) results in an inability to attach rhamnose at the sites of glycosylation and the flagellin protein remains unmodified. This gene is annotated as belonging to glycosyltransferase family 2 and so may be responsible for the transfer of rhamnose to threonine on the protein backbone.

4.1.4 Characterization of PTMs on the flagellin of *Listeria monocytogenes*

Results of this study were published previously in Journal of Bacteriology[85] and are further described in this section.

The flaA structural gene from strains, 568, 394, 2568 and CLIP23485, belonging to serotypes 1/2a, 1/2b, 1/2c and 4b respectively, were analysed by Dr. Susan Logan (NRC, Ottawa, Canada). A very high degree of conservation was found amongst the predicted flagellin DNA sequences from the four serotypes examined when compared to *L. monocytogenes* EGD[104] and only a single amino acid substitution was found for each flagellin protein (Fig. 4.21). The predicted amino acid sequences of flagellin from strain 394 (serotype 1/2b) and CLIP23485 (serotype 4b) were identical, as would be expected as both strains fall within the Division II sharing a common lineage.[105] The published flagellin protein sequences from serotype 1/2a strains, EGD and strain F6854,[104,106] differed in a single residue (position number 263, Leu to Pro) from strain 568 which is also a serotype1/2a strain. Strain 2568 (serotype 1/2c) differed from all of these strains by a single residue (position number 141, Thr to Ile) within the central core region of the flagellin protein. Serotype 1/2a and 1/2c strains fall within the Division I lineage.[105]

```
EGD (1/2a)          MKVNTNIISLKTQEYLRKNNEGMTQAQERLASGKRINSSLDDAAGLAVVTRMNVKSTGLD
394 (1/2b)          MKVNTNIISLKTQEYLRKNNEGMTQAQERLASGKRINSSLDDAAGLAVVTRMNVKSTGLD
CLIP 23485(4b)      MKVNTNIISLKTQEYLRKNNEGMTQAQERLASGKRINSSLDDAAGLAVVTRMNVKSTGLD
568 (1/2a)          MKVNTNIISLKTQEYLRKNNEGMTQAQERLASGKRINSSLDDAAGLAVVTRMNVKSTGLD
2568(1/2c)          MKVNTNIISLKTQEYLRKNNEGMTQAQERLASGKRINSSLDDAAGLAVVTRMNVKSTGLD

EGD (1/2a)          AASKNSSMGIDLLQTADSALSSMSSILQRMRQLAVQSSNGSFSDEDRKQYTAEFGSLIKE
394 (1/2b)          AASKNSSMGIDLLQTADSALSSMSSILQRMRQLAVQSSNGSFSDEDRKQYTAEFGSLIKE
CLIP 23485(4b)      AASKNSSMGIDLLQTADSALSSMSSILQRMRQLAVQSSNGSFSDEDRKQYTAEFGSLIKE
568 (1/2a)          AASKNSSMGIDLLQTADSALSSMSSILQRMRQLAVQSSNGSFSDEDRKQYTAEFGSLIKE
2568(1/2c)          AASKNSSMGIDLLQTADSALSSMSSILQRMRQLAVQSSNGSFSDEDRKQYTAEFGSLIKE

EGD (1/2a)          LDHVADTTNYNNIKLLDQTATGAATQVSIQASDKANDLINIDLFNAKGLSAGTITLGSGS
394 (1/2b)          LDHVADTTNYNNIKLLDQTATGAATQVSIQASDKANDLINIDLFNAKGLSAGTITLGSGS
CLIP 23485(4b)      LDHVADTTNYNNIKLLDQTATGAATQVSIQASDKANDLINIDLFNAKGLSAGTITLGSGS
568 (1/2a)          LDHVADTTNYNNIKLLDQTATGAATQVSIQASDKANDLINIDLFNAKGLSAGTITLGSGS
2568(1/2c)          LDHVADTTNYNNIKLLDQTAIGAATQVSIQASDKANDLINIDLFNAKGLSAGTITLGSGS
                                        *

EGD (1/2a)          TVAGYSALSVADADSSQQATEAIDELINNISNGRALLGAGMSRLSYNVSNVNNQSIATKA
394 (1/2b)          TVAGYSALSVADADSSQEATEAIDELINNISNGRALLGAGMSRLSYNVSNVNNQSIATKA
CLIP 23485(4b)      TVAGYSALSVADADSSQEATEAIDELINNISNGRALLGAGMSRLSYNVSNVNNQSIATKA
568 (1/2a)          TVAGYSALSVADADSSQQATEAIDELINNISNGRALLGAGMSRLSYNVSNVNNQSIATKA
2568(1/2c)          TVAGYSALSVADADSSQQATEAIDELINNISNGRALLGAGMSRLSYNVSNVNNQSIATKA
                                    *

EGD (1/2a)          SASSIEDADMAAEMSEMTKYKILTQTSISMLSQANQTPQMLTQLINS
394 (1/2b)          SASSIEDADMAAEMSEMTKYKILTQTSISMLSQANQTPQMLTQLINS
CLIP 23485(4b)      SASSIEDADMAAEMSEMTKYKILTQTSISMLSQANQTPQMLTQLINS
568 (1/2a)          SASSIEDADMAAEMSEMTKYKIPTQTSISMLSQANQTPQMLTQLINS
2568(1/2c)          SASSIEDADMAAEMSEMTKYKILTQTSISMLSQANQTPQMLTQLINS
                                        *
```

Figure 4.21: Comparison of the amino acid sequences of *Listeria monocytogenes* flagellins. Amino acid residues are written in the single letter code and variable amino acids are underlined in bold and with asterisks.

4.1.4.1 Structural analysis of *L. monocytogenes* flagellin.

To determine the degree of glycosylation and to identify the sites of modification on the flagellin monomer, detailed structural analyses were conducted using mass spectrometry. Electrospray mass spectrometry analysis of purified flagellin from *L. monocytogenes* CLIP23485 showed four well-defined components at Mr 31050, 31,254, 31,457 and 31660 Da corresponding to the molecular mass of the monomeric FlaA protein (Mr 30444Da) with additional modifications accounting

for 606, 810, 1013 and 1215 Da. The predominant flagellin species was of molecular mass 31,254 Da. (Fig. 4.22)

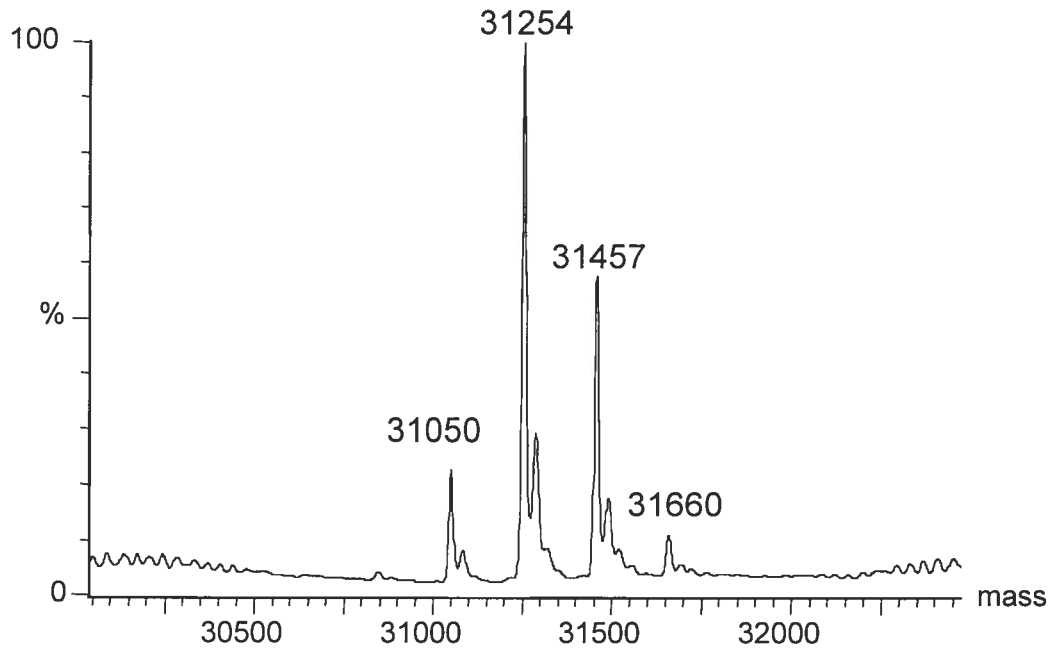

Figure 4.22: Intact mass analysis of CLIP 23485 flagellin.

Reconstructed molecular mass profile of CLIP 23485 flagellin. Flagellin (M_{theo} 30444Da) shows peaks at 31050Da, 31253Da, 31456Da and 31495Da corresponding to the attachment of 3, 4, 5 and 6 HexNAc residues respectively. No signal corresponding to unmodified flagellin was observed.

Analysis of the tryptic digest identified the tryptic peptides T_{135}-T_{154} as having been modified with one to two HexNAc residues and T_{168}-T_{214} with two to three HexNAc residues. Sequence coverage of 90% was obtained (Fig. 4.24). Second generation fragment ion spectrum of fragment ion m/z 204 resulted in a fragmentation pattern consistent with that of a HexNAc residue. Taken together these results suggest that *L. monocytogenes* flagellin is modified with a variable number of HexNAc residues ranging from 3 to 6 (3x203=609Da, 4x203=812Da, 5x203=1015Da, 6x203=1218Da). To assign precisely the location of the glycosylation sites on flagellin, the flagellin tryptic digest was then subjected to base catalysed hydrolysis in the presence of NH_4OH, followed by targeted capLC/MS/MS analysis. By using this approach it was possible to map 5 sites of attachment on *L. monocytogenes* CLIP 23485 flagellin (Fig. 4.24). Threonine residues T_{141},T_{145}, T_{173}, T_{175} and serine residue S_{181} were identified as the sites of O-linked glycosylkation attachment. An example for the tryptic glycopeptide T_{135}- T_{154} before and after ß-elimination is shown in figure 4.23.

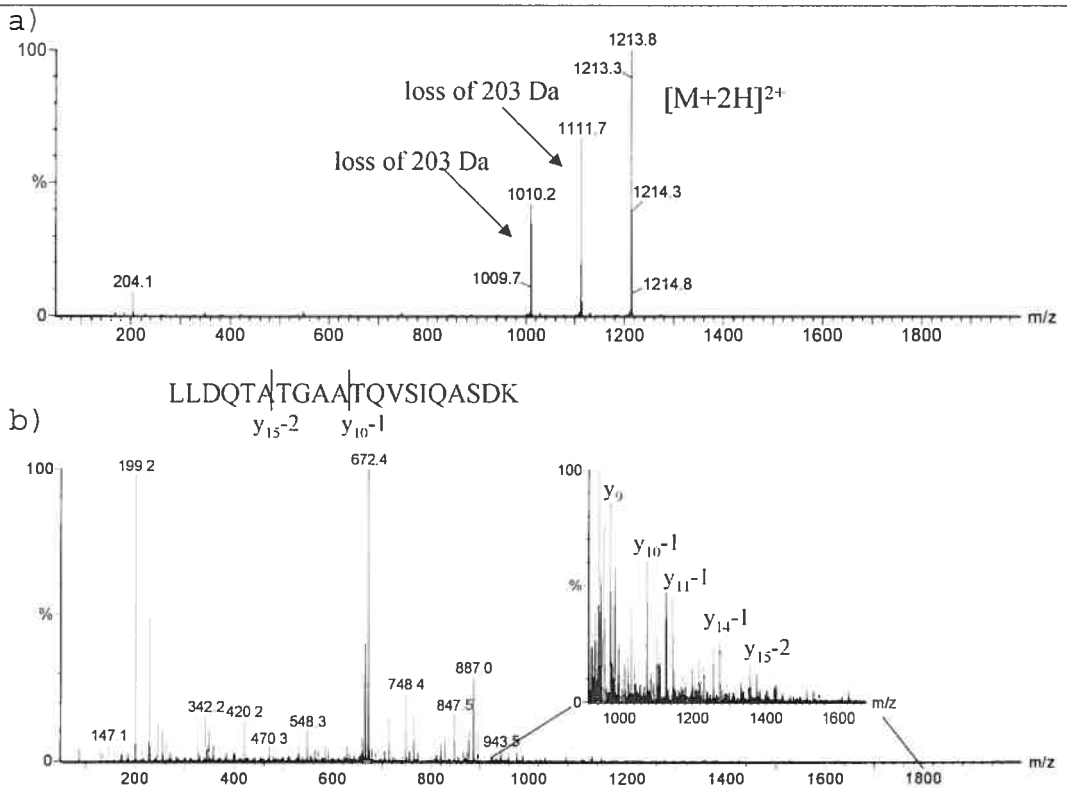

Figure 4.23: MS/MS spectra of a) m/z 1213.3 from the original tryptic glycopeptide and b) m/z 672.7 following subsequent ß-elimination. An expanded view enables the identification of fragment ions showing a mass shift of either 1 or 2 m/z units corresponding to the amino acid segments baring the modified residues.

MKVNTNIISLKTQEYLR**KNNEGMTQAQER**LASGK**RINSSLDDAAGLAVVTR**M
NVKSTGLDAASK**NSSMGIDLLQTADSALSSMSSILQR**MR**QLAVQSSNGSFSDE
DRKQYTAEFGSLIKELDHVADTTNYNNIKLLDQTATGAATQVSIQASDKAND
LINIDLFNAKGLSAGTITLGSGSTVAGYSALSVADADSSQEATEAIDELINNISN
GALLGAGMSRLSYNVSNVNNQSIATKASASSIEDADMAAEMSEMTKYKILTQ
TSISM LSQANQTPQMLTQLINS**

Figure 4.24: Assignment map of CLIP 23485 flagellin. The primary amino acid sequence of FlaA from CLIP 23485 is shown. Bold letters indicate peptides identified by mass spectrometry and underlined peptides were shown to be glycosylated. Boxed residues show carbohydrate attachment site. Peptides not identified are indicated in normal font. Sequence coverage was 90%.

4.1.4.2 Conservation of glycosylation sites amongst Listeria serotypes

Intact mass analyses of flagellin from strains 386, 568 and 2568 confirmed that these proteins were also glycosylated with HexNAc. Interestingly, in strain 2568, T_{141} is replaced with an isoleucine, one of the glycosylation sites identified by alkaline hydrolysis in CLIP23485 flagellin. As a result of this amino acid substitution, the intact mass analysis of this strain showed a different molecular mass profile. The intact glycoprotein comprising 6 HexNAc residues was no longer observed. Rather the mass spectrum showed peaks corresponding to the glycoprotein with only 3, 4 or 5 HexNAc residues attached. (Data not shown)

The monosaccharide (alditol acetate) GC-MS analysis done by Dr. E. Vinogradov (NRC, Ottawa, Canada) of hydrolysed flagellin protein from strain CLIP23485 identified GlcNAc as the N-acetylated amino sugar giving the singly charged m/z 204 fragment ion in the tryptic peptide MS analysis.

A monoclonal antibody which recognizes O-GlcNAc in a β -O glycosidic linkage to both serine and threonine was utilized by Dr. Susan Logan (NRC, Ottawa, Canada) in immunoblotting experiments to determine the nature of the linkage found on Listeria flagellin and indicated that the GlcNAc residue is attached to the protein backbone through a β-O glycosidic linkage. Flagellin protein from strains 568, 394, 2568 were also positive by immunoblotting with this antibody (data not shown).

4.1.4.3 Discussion

Our results demonstrate that *L. monocytogenes* flagellin is uniquely modified with O-linked β-N-acetylglucosamine at up to six sites on each protein monomer. The precise site of linkage for 5 of these residues was determined. These findings confirm previous studies suggesting that the flagellins of Listeria are post-translationally modified.[59,62] Furthermore, results from ESMS analysis of flagellin from three additional flagellar H-groups demonstrated that the modification is conserved regardless of the serotype. Although indirect evidence for glycosylation of flagellin from *Azospirillum brasileinse,*[107] *Aeromonas caviae,*[68] *Spirochaeta aurantia,*[108] *Clostridium tyrobutyricum,*[109] and *Caulobacter crescentus*[84] has been provided in the literature; to date, the only detailed structural characterization of the glycan moiety found on flagellins is that reported for *Pseudomonas aeruginosa* (see section 4.12) *Campylobacter jejuni*[24] and *Helicobacter pylori* (see section 4.1.1). This work represents the first structural characterization of flagellar glycosylation from a Gram-positive organism. Unlike the novel glycans found on flagella of these Gram negative organisms, glycosylation of Listeria flagella is restricted to up to 6 sites, each containing a single O-linked GlcNAc residue.

Posttranslational modification of proteins by β-N-acetylglucosamine (O-GlcNAc) is a well characterized monosaccharide modification of a number of eukaryotic nuclear and cytoplasmic proteins.[110-111] In comparison, no equivalent modification has been identified to date in prokaryotic systems. To our knowledge this represent the first description of a β-O-GlcNAc posttranslational modification

on a prokaryotic protein.[93] Unlike the glycosylation of *Campylobacter* and *Helicobacter flagellin* with pseudaminic acid, the addition of GlcNAc to Listeria flagellin would not require a novel glycan biosynthetic pathway as GlcNAc is a common biosynthetic precursor of numerous biochemical pathways. However, in eukaryotic systems an important component of the O–GlcNAc modification is the O-GlcNAc transferase enzyme.

As with other glycosylated bacterial flagellins, the sites of modification are located in the central surface exposed region of the protein.[33] This region is responsible for the antigenic variability observed for the flagella from a variety of bacteria.[115] While diversity in glycan structure has been shown to contribute to serospecificity of Campylobacter flagella, this is unlikely to be the case for Listeria. The flagellins from three different serotypes (1/2a, 1/2B and 4b) displayed an identical pattern of glycosylation with β-N-acetylglucosamine clearly demonstrating that this post translational modification is not likely to be a contributing factor in the serological differentiation of these *L. monocytogenes* strains, but is likely to represents a shared antigen common among all of the H- serotypes.[60]

Inhibition of the flagellar glycosylation process through site directed mutagenesis in Campylobacter and Helicobacter led to loss of motility, while in Pseudomonas flagellar assembly and motility was unaffected by lack of flagellin glycosylation. The future identification and characterization of Listeria glycosylation genes will permit the determination of the role of the glycosylation process in flagellar assembly and motility of this organism.

Flagella have been shown to be important in the initial attachment of the organism to various surfaces at temperatures less than 30°C.[113-114] Flagellar glycosylation may have an important role in the environmental interactions particularly within the food processing environment, where temperatures are generally maintained at much lower levels than ambient and at which flagella are optimally produced. It remains to be established if the presence of O-linked GlcNAc on the flagella filament acts to facilitate such interactions through lectin like association. Whether glycosylation is essential for flagellar assembly or in some way facilitates interactions of Listeria either with other cells or with environmental substrates is currently unknown. While flagellin expression has been shown to be downregulated at 37°C[115] and may not be essential for pathogenesis in human infection, a recent publication indicated that flagella of *L. monocytogenes* grown at 20°C played a role in Caco-2 cell invasion.[116] The molecular basis of this process was not determined but the potential of O-linked GlcNAc residues on the flagellin protein facilitating such interactions cannot now be overlooked.

4.2 Identification of unusual bacterial glycosylation from intact proteins using a top down mass spectrometry approach

4.2.1 Molecular weight determination

Infusion of the dialyzed flagellin solution into the mass spectrometer resulted in a characteristic multiply-charged state envelope for all four samples (Fig. 4.25). The data were deconvoluted to obtain a molecular mass profile that could be compared to the expected masses from their corresponding cDNA sequences. (Table 4.2) In all four cases the measured molecular masses were higher than that predicted from the cDNA sequence, indicating post-translational modifications. Distinct mass excess of 2212 Da was observed on *H. pylori* and of around 6501 Da on *C. jejuni* flagella, in good agreement with measurements reported previously by Thibault et al.[24] The *L. monocytogenes* flagellin revealed 4 glycan structures corresponding to masses 605, 808, 1011 and 1214 Da. The molecular mass profile of *A. caviae* flagellin showed mass excess of 2241 Da, 2464 Da and 2766 Da from the predicted mass of the FlaA structural protein (31244 Da).

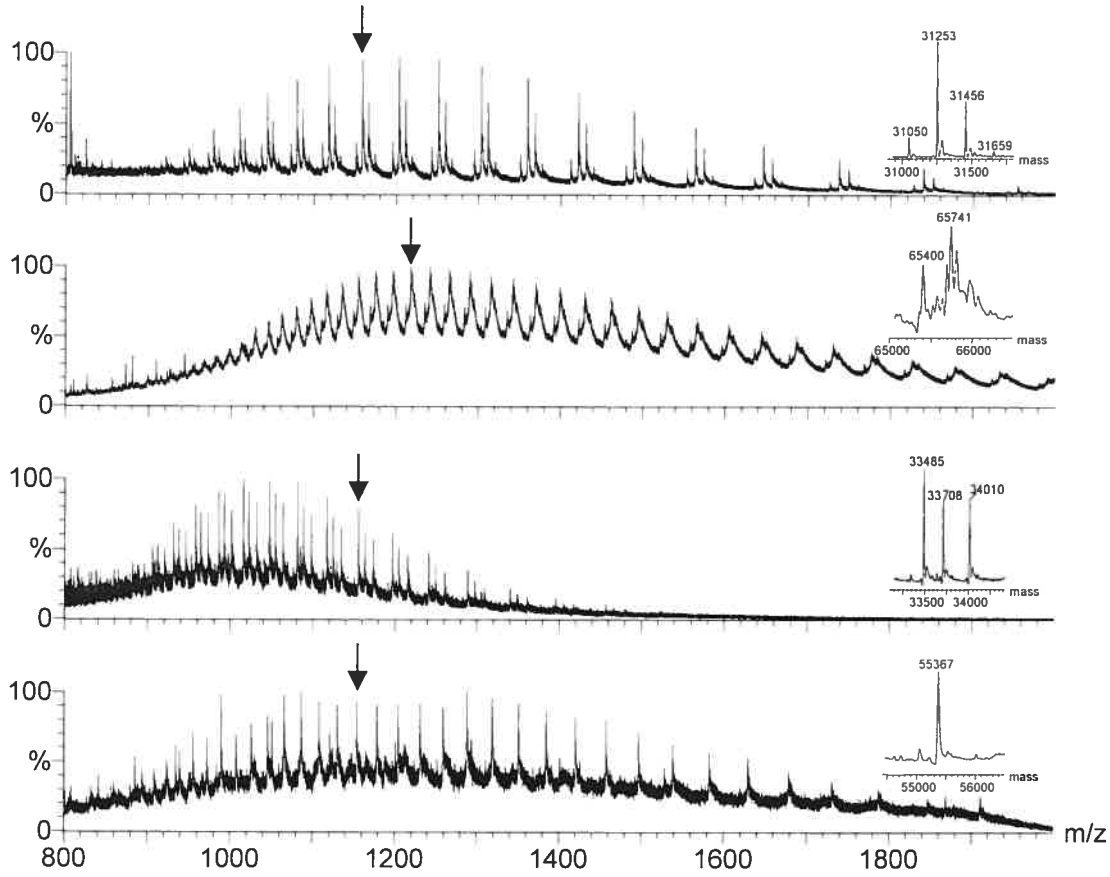

Figure 4.25: Molecular mass determination of intact bacterial flagellin.

a) *Listeria monocytogens*, b) *Campylobacter jejuni*, c) *Aeromonas caviae* and d) *Helicobacter pylori*. Ions marked with an arrow were subjected to MS/MS analysis (Fig. 4.26). The molecular mass profile obtained from each corresponding MS spectrum is presented on the right panels. In all cases the observed molecular mass was higher than that predicted from the cDNA sequence suggesting post-translational modifications.

Table 4.2: Molecular masses of different prokaryotic bacteria.

Bacterium	Strain	Determined MW	Predicted MW	Amount of PTM
Helicobacter pylori	1061	55367 Da	53153 Da	~4%
Aeromonas caviae	UU51	33485 Da, 33708 Da and 34010 Da	31244 Da	~7 %
Listeria monocytogenes	1174	31050 Da, 31253 Da, 31456 Da and 31659 Da	30445 Da	~2%-4%
Campylobacter jejuni	81-176	65400 Da, 65741 Da	59240 Da	~11%

After the determination of the molecular mass profile, one of the multiply-protonated ions from each bacterial flagellin was subjected to a tandem mass spectrometry experiment (Fig. 4.26). The corresponding MS/MS spectra showed abundant carbohydrate oxonium ions at the low m/z range of the spectra and attest to the labile nature of the corresponding glycosidic bonds.

4.2.2 Results for *Campylobacter jejuni*

The multiply charged protein signal at m/z 1196 was subjected to a CID experiment. The results are shown in figure 4.26b. Intense fragment ions were observed in the low mass region at m/z 486.2, 317.1, 316.1, 299.1 and 281.1. Each signal was subjected to second generation tandem mass spectrometry experiment by increasing the RF Lens 1 from 50V to 125V forming fragments in the orifice/skimmer region. The fragment ions at m/z 317.1 and 316.1 corresponds to Pse5Ac7Ac and PseAm.[24] The fragment ions at m/z 299.1 and m/z 281 correspond to loss of water molecules from Pse5Ac7Ac. In addition, an oxonium ion at m/z 486 was also observed in the spectra of this CID experiment which had not been observed in the earlier "bottom up" analysis of *C. jejuni* flagellin and was indicative of a third novel glycan structure The fragmentation pattern revealed that this novel glycan consisted of a disaccharide, consisting of PseAm and a second monosaccharide of mass 170 Da. The most probable structural assignment of the second monosaccharide is acetamidino-2,4,6-trideoxy-hexose, a molecule which is structurally related to bacillosamine, where the 2 amino group is converted to acetamidino and the 4 amino

group is removed. Not only would this structure comply with the mass of the remaining residue, its structural similarity to bacillosamine previously reported for the N-linked glycosylation system in *C. jejuni*[28] would make it a likely candidate. However, an unambiguous identification of the structure was not possible due to the limited amount of sample available for NMR structure determination.

4.2.3 Results for *Helicobacter pylori*

Tandem mass spectrometry of the multiply protonated ion at m/z 1154 showed intense signals at m/z 317, 299 and 281 (Fig. 4.26d) corresponding to Pse5Ac7Ac as confirmed by the MS/MS spectrum. (Fig. 4.27d) The MS/MS spectrum was identical to that reported previously by Thibault for Pse5Ac7Ac[24] and that obtained from the "bottom up" approach (see section 4.1.1).

As for *C. jejuni* hardly any fragments from the protein backbone were observed enabling the easy identification of the post-translational modification. The amount of attached Pse5Ac7Ac to the protein backbone can be derived from the excess mass determined by the molecular weight determination. The excess mass of 2212 corresponds to 7 sites modified with Pse5Ac7Ac. (7*316.1 Da=2212.7 Da)

4.2.4 Results for *Listeria monocytogenes*

The MS/MS spectra on the four different multiply charged envelope ions for *L. monocytogenes* flagellin resulted in all cases in a prominent oxonium ion at m/z 204 indicating the presence of HexNAc residues (Fig. 4.26a). This was further

confirmed by the fragmentation pattern observed for the oxonium ion at m/z 204 from *L. monocytogenes* flagellin (Fig. 4.27a), consistent with data obtained from the "bottom up" approach. (see section 4.1.3)

4.2.5 Results for *Aeromonas caviae*

MS/MS experiment of the multiply charged ion series from *A. caviae* showed in all cases an abundant oxonium ion at m/z 374. As an example the MS/MS spectrum of m/z 1155 is shown in figure 4.26. The oxonium ion at m/z 374 was subjected to a MS/MS experiment. (Fig. 4.27c) The fragmentation pathway was similar to Pse5Ac7Ac indicating structural similarity. Based on the mass of this residue and its characteristic fragmentation pattern, the monosaccharide was tentatively assigned to Pse3Ac5Ac7Ac. Alternatively, Pse5Ac7Ac could be modified by an extra glycine via an ester bond which could possibly explain the loss of 31 Da (CH_3-NH_2). However, unambiguous identification of this saccharide would require NMR analyses.

For the sample three different ion series were observed and all showed a characteristic oxonium ion of m/z 374. Molecular mass comparison with that predicted from the cDNA sequence, FlaA (31244 Da) and FlaB (31094 Da), suggested that the peak with a molecular mass of 33485 Da (Fig. 4.25c) corresponds to FlaA with 6 Pse3Ac5Ac7Ac (MW_{theo} 33483 Da) and the peak at 33708 Da to FlaB with 7 Pse3Ac5Ac7Ac attached (MW_{theo} 33706 Da). The origin of the peak corresponding to a molecular mass of 34010 Da is unknown and may corresponds to a third flagellin protein.

In separate experiments the flagellin was digested with trypsin and analysed by capLC/MS/MS. The only glycan oxonium ion observed was at m/z 374 and peptides bearing this modification were identified. However, no assignment could be made since the protein sequence was still unknown.

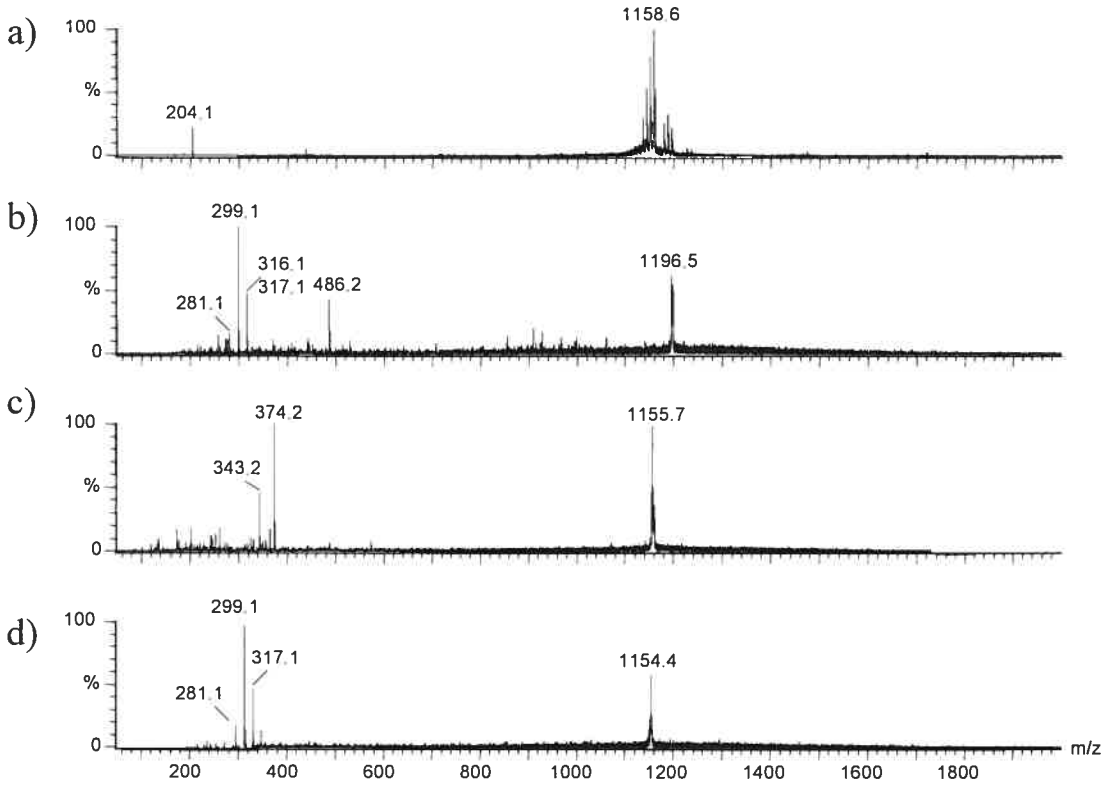

Figure 4.26: MS/MS spectra of multiply charged ions from intact flagellin.

a) MS/MS spectrum of m/z 1158 from *Listeria monocytogens*. b) MS/MS spectrum of m/z 1196 from *Campylobacter jejuni*. c) MS/MS spectrum of m/z 1155 from *Aermonas caviae*. d) MS/MS spectrum of m/z 1154 from *Helicobacter pylori*.

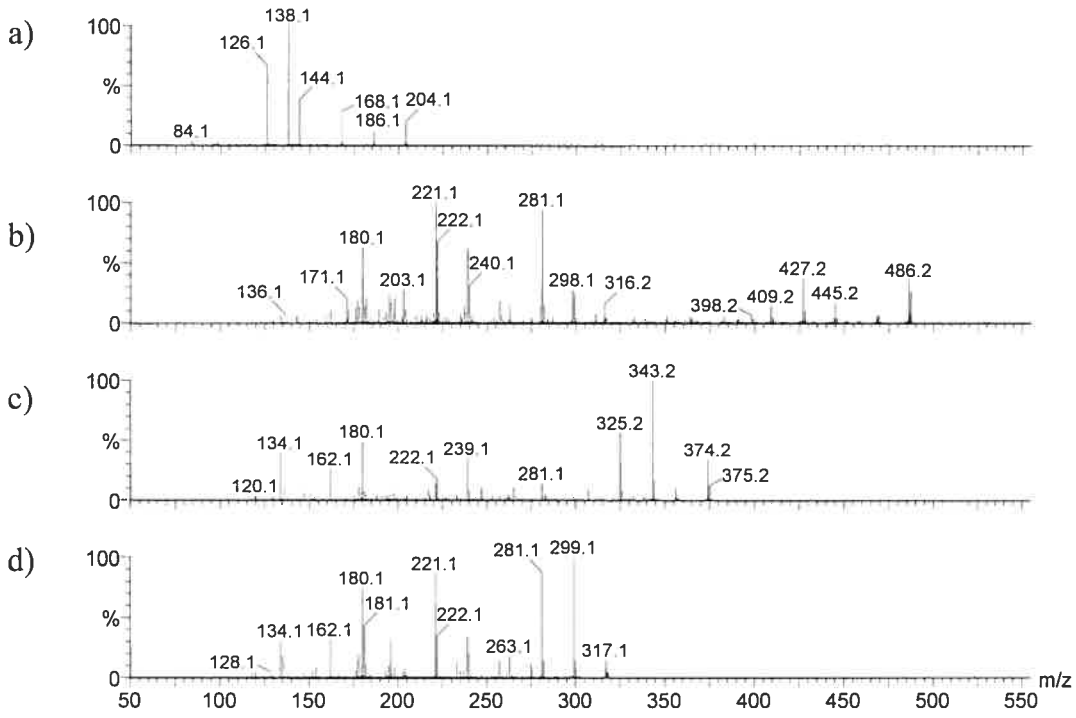

Figure 4.27: Second generation fragment ion spectra of the unusual PTMs.

Second generation fragment ion spectra of the unusual carbohydrate residues found in bacterial flagella. a) MS/MS of m/z 204 from *Listeria monocytogenes*. b) MS/MS spectrum of m/z 486 from *Campylobacter jejuni*. c) MS/MS spectrum of m/z 374 from *Aeromonas caviae*. d) MS/MS spectrum of m/z 317 from *Helicobacter pylori*. Note that the MS/MS spectra were acquired by increasing the RF Lens 1 from 50 to 125 forming fragments in the orifice/skimmer region to promote the formation of the oxonium ions from the native flagellin.

4.2.6 Differentiation of glycosylated and not glycosylated proteins

The application of this "top-down" approach to identify glycoproteins from complex protein extracts was evaluated for a *H. pylori* flagellin protein extract consisting predominantly of flagellin A, neutrophil activating protein and urease, as determined by the "bottom up" approach. Two different charge envelopes were observed originating from neutrophil activating protein and flagellin. Other signals were observed too, probably from urease and other protein contaminants in the sample. Three different m/z values were subjected to MS/MS analysis indicated by the arrow in figure 4.28. However, only the MS/MS spectrum of flagellin (m/z 1154, charge state +48) showed characteristic carbohydrate oxonium ions in the low mass region (Fig. 4.29a) whereas these characteristic features were not observed in the MS/MS spectra of neutrophil activating protein (m/z 842, charge state +20) (Fig. 4.29c) and m/z 1017 (Fig. 4.29b).

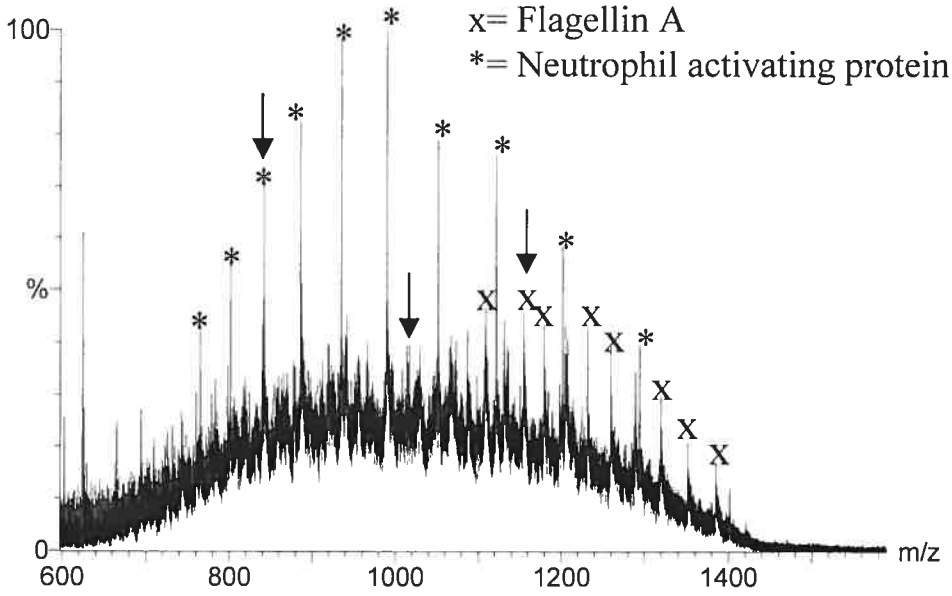

Figure 4.28: Infusion of a protein mix from *Helicobacter pylori*.

A protein mix consisting mainly of, Neutrophil activating protein, Urease and Flagellin A was infused into the mass spectrometer at a flow rate of 0.5 µL/min. Three multiply charged ions, indicated by the arrow, were subjected to a tandem MS/MS experiment. (m/z 842, m/z 1017 and m/z 1154). Each subjected precursor belongs to a different protein.

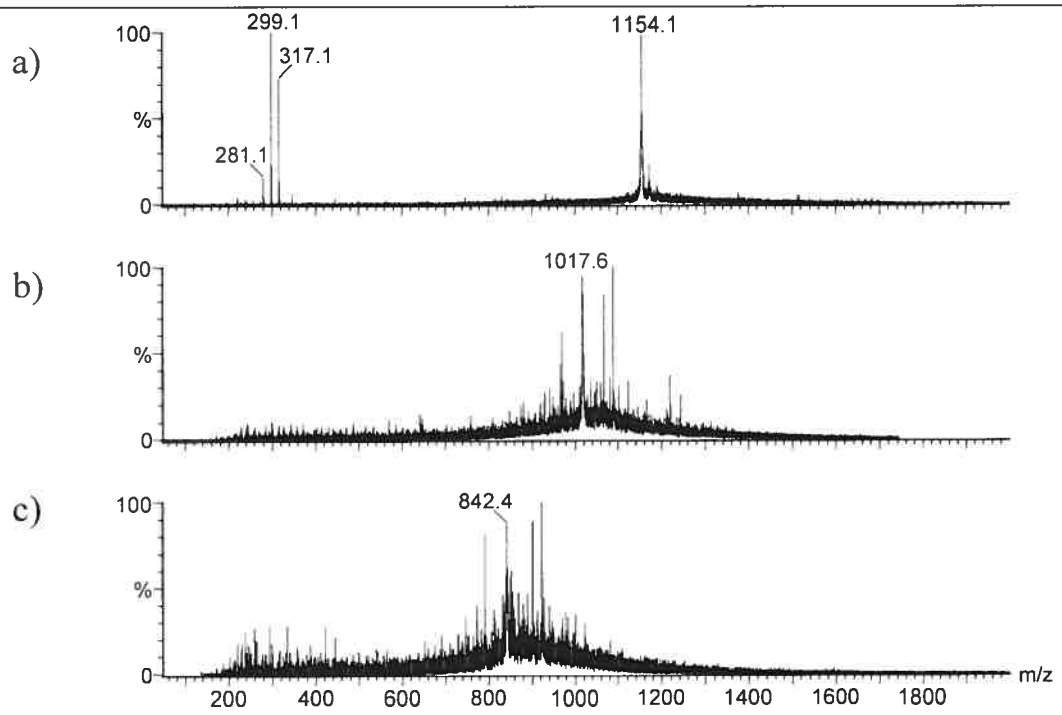

Figure 4.29: Tandem MS/MS spectra of glycosylated and not-glycosylated proteins. Spectrum a) shows the MS/MS spectrum of m/z 1154.1 (glycosylated flagellin A). Intense oxonium ions from the carbohydrate are observed enabling the easy identification of the PTM. Spectra b) MS/MS spectra of m/z 1017.6 and c) MS/MS spectra of 842.4 (neutrophil activating protein) do not show intense oxonium ions from carbohydrates indicating that these two proteins are not glycosylated.

4.2.7 Identification of gene functions involved in the glycosylation process

This approach can also be used for monitoring the changes in glycosylation for bacterial mutants, where certain genes were knocked out. Top down mass spectrometry of the intact protein signals of the Cj1333 mutant showed a different glycosylation profile than for the parent strain. (Fig. 4.30) Pse5Ac7Ac (m/z 317) was still present but PseAm (m/z 316) and the new unusual modification at m/z 486 were not observed anymore. Instead of m/z 486 a new signal at m/z 487 appeared. Tandem mass spectrometry of this novel 487 modification showed clearly that the

acetamidino-2,4,6-trideoxy-hexose was now attached to Pse5Ac7Ac instead of PseAm. (Fig. 4.31)

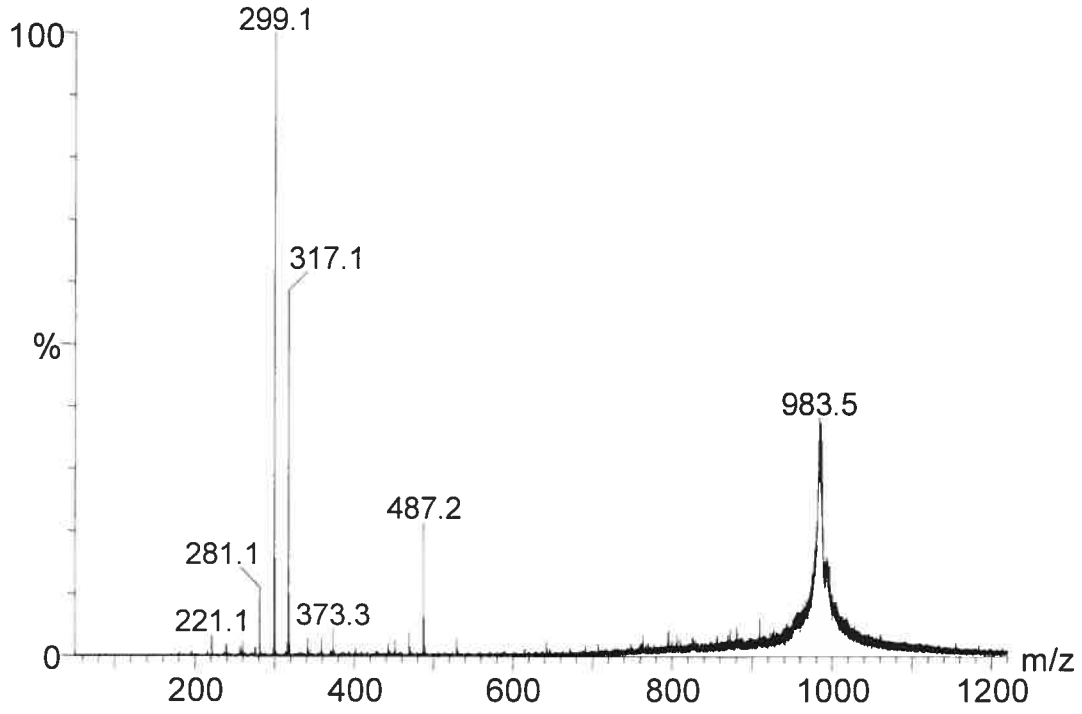

Figure 4.30: CID of multiply charged ion from *C. jejuni* mutant

Tandem MS/MS spectrum of m/z 983 from a *C. jejuni* mutant of gene Cj1333. This mutant still utilizes Pse5Ac7Ac as substrate for flagellin glycosylation, but however does not produce anymore the carbohydrates at m/z 316 and m/z 486 compared to the parent strain (Fig. 4.26b). Additionally, a novel sugar (m/z 487) is now attached to the flagellin protein.

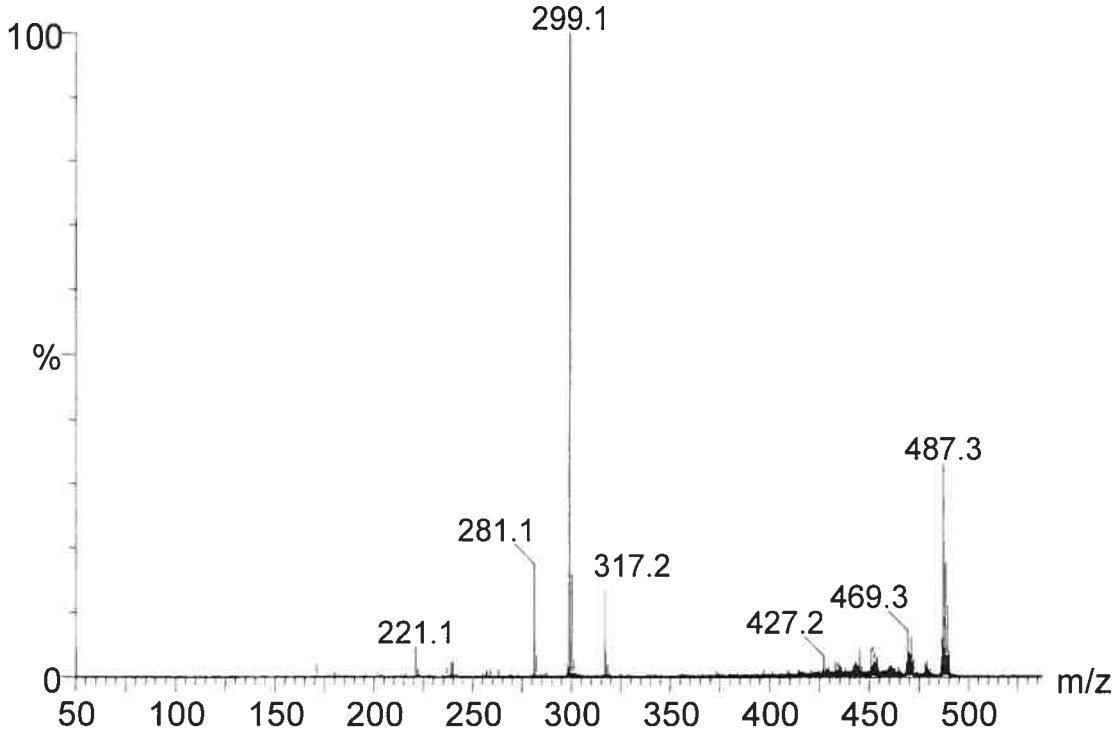

Figure 4.31: Second generation fragment ion spectra of m/z 487.

The MS/MS spectrum shows intense signals at m/z 317 and m/z 299 indicating that the monosaccharide of mass 170 Da is now attached to Pse5Ac7Ac instead of PseAm as observed in the parent strain.

4.2.8 Discussion

We present here a new approach to analyse samples which are believed to contain novel glycoproteins. Top down mass spectrometry is a straightforward and easy procedure for the fast screening of unusual post-translational glycosylation of proteins.

The first step, measuring the molecular weight of the intact protein, gives an immediate result in terms of the extent of glycosylation present, if the primary amino acid sequence is available through sequencing efforts. The second step, CID of one of the intact multiply charged protein ions, resulted in extensive gas-phase deglycosylation and intense oxonium ions from the PTM's in the low mass region. The only significant signals in the low mass region were from the carbohydrates and not from the protein backbone. The third step, second generation product ion scan of the possible modifications, gives structural information about the post-translational modification, enabling the identification of the carbohydrate. In addition, by increasing the collision offset it is possible to generate peptide fragment ions, that can be used for internal calibration to determine the exact mass (see section 4.1.3.2).

However, precision in mass measurement is inversely proportional to the square root of the intensity of the signal. Q-TOF mass spectrometer are able to determine the mass accuracy with a precision of around 5 ppm for relative intense signals (~300 counts/sec). However, most glycoproteins ionize poorly (e.g, *C. jejuni*) resulting in low intensity signals of precursor ions from which oxonium fragment ions of the carbohydrates would be generated (*C. jejuni* ~5 counts/sec). This increases the

theoretical achievable mass accuracy to around 50 ppm on the Q-TOF machine and is not sufficient to determine the elemental formula of the glycan.

This new analysis method provides a timely and relatively straightforward method to routinely screen protein samples from genetically manipulated strains and consequently assign putative functions to selected genes in terms of the glycosylation defects/changes observed. In contrast to the laborious and costly bottom up approach where particular modifications may indeed be missed, the approach described here will facilitate a much higher throughput in terms of numbers of samples examined and may indeed lead to the identification of novel modifications.

The flagellin of *C. jejuni* was also shown to be glycosylated with a novel glycan of mass 486 Da. This novel PTM had not been identified by the traditional bottom up approach possibly as a consequence of its stability. The flagellin of *A. caviae* has shown to be glycosylated with a monosaccharide, structurally related to Pse5Ac7Ac, at six sites on FlaA and seven sites on FlaB. The structure could not be determined unambiguously solely by mass spectrometry and further NMR analyses would be required.

Top-down mass spectrometry has been described for the detection of post-translational modifications directly from the intact proteins, using either Fourier-Transform Mass Spectrometry with electron capture dissociation (ECD)[117-118] or ion/ion proton transfer on a ion trap instrument.[119] ECD cleaves peptide backbones primarily at the C_α-N bond leaving labile side-chain modifications such as phosphorylation or glycosylation intact enabling the identification of the linkage site.[120] Identification of these modifications is based on the observation of common

mass discrepancy of e.g. 80 for phosphorylation or 162 for the addition of a hexose residue and can only be used for known PTMs. In the present approach extensive gas-phase deglycosylation and formation of an intense oxonium ion in the low mass region is observed and thus enables the identification of novel glycans. However, neutral monosaccharide residues such as hexose or deoxyhexose (*P. aeruginosa* JJ692) cannot be analysed by this approach as these do not yield abundant oxonium ions. In addition, the exact linkage site of the post-translational modifications cannot be determined and assignment still requires using the traditional bottom up approach.

4.3 Identification of novel unknown glycoproteins in complex protein extracts of *H. pylori*

As part of our continued interest in the identification of novel glycoproteins from prokaryotes, a periplasmic and a crude membrane extract from *H. pylori* were analysed. These extracts are relative complex and contain hundreds of different proteins yielding more than 25000 peptides following tryptic digestion. However, the maximum number of peptides detectable in one LC/MS/MS run is approximately 1500. To improve the number of detectable peptides the samples were separated into different fractions by SDS-PAGE and SCX. Both methods were compared for their ability to identify novel glycoproteins from complex mixtures.

4.3.1 Analysis of a periplasmic extract of *H. pylori*

4.3.1.1 SDS-PAGE

The periplasmic extract was separated by SDS-PAGE (Fig.4.32). A total of 24 bands were excised, digested with trypsin and each band was analysed by two iterative capLC/MS/MS exclusion runs. A collection of 10733 MS/MS spectra with charge state +2 and +3 were acquired and a Mascot search was conducted against a non-redundant protein database. The Mascot search data results were manually validated and 3769 out of the 10733 MS/MS spectra matched 609 protein

assignments, corresponding to a total of 296 unique proteins after accounting for redundancy of different proteins sharing the same peptide sequence. A total of 6964 MS/MS spectra were not matched to an entry in the database which corresponds to approximately 65% of all spectra. The high proportion of unassigned spectra is accounted for spectra of poor quality, unspecific cleavages or post-translational modifications. The analysis of the extracted MS/MS spectra with intense signals for common carbohydrate oxonium ions, identified six glycopeptides showing an ion at m/z 317 in band 12 and one in band 10. Glycopeptides from band 12 were identified to FlaA and FlaB, previously known to be glycosylated (see section 4.1.1). The glycopeptide identified in band 10 belongs to the tryptic peptide ISFTNDSAVSR from the flagellar hook protein (Fig. 4.33). The signal was not very intense (~50 counts) and although not many y-fragment ions from the peptide backbone were observed a unique sequence tag identified the flagellar hook protein. The assignment was further confirmed by the identification of unmodified tryptic peptides from the flagellar hook protein in the same band. The bands comprising glycopeptides were very complex as illustrated in the LC/MS analysis of tryptic peptides from band 10 (Fig. 4.33). 1322 doubly, 262 triply and 3 quadruply charged peptides were detected for band 10.

Glycopeptide spectra can be typically identified from unmodified tryptic peptides as they show intense fragment ions in the low mass region, with relative few peptide sequence fragment ions and an intense signal for the naked peptide fragment ion. All acquired MS/MS spectra were manually inspected for these characteristics, but only peptides modified with Pse5Ac7Ac were identified.

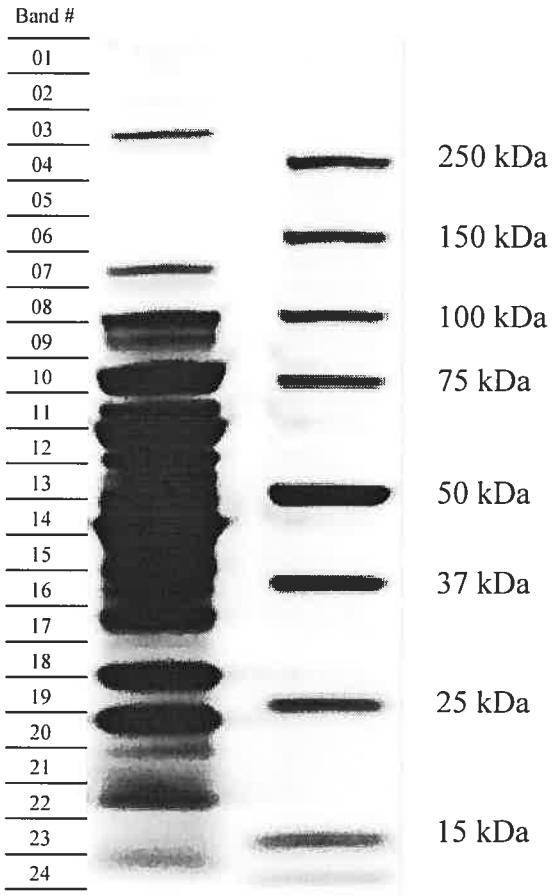

Figure 4.32: SDS-PAGE gel of a periplasmic extract of *H. pylori*.

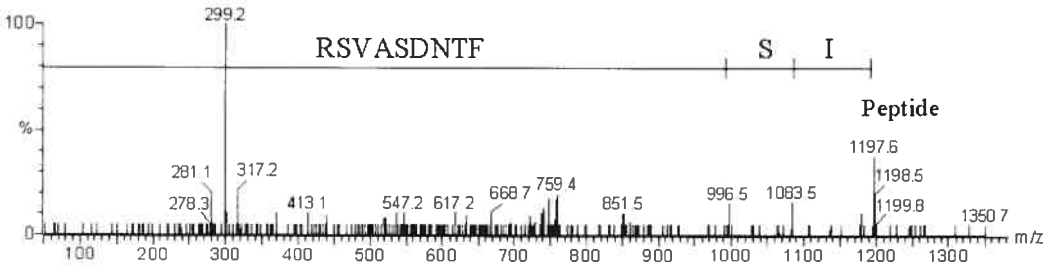

Figure 4.33: MS/MS spectrum of m/z 756.7.

The spectrum shows an intense oxonium ion at m/z 317 corresponding to Pse5Ac7Ac. The mass of the peptide and a y-fragment ion series observed matches that of the tryptic peptide ISFTNDSAVSR of the flagellar hook protein.

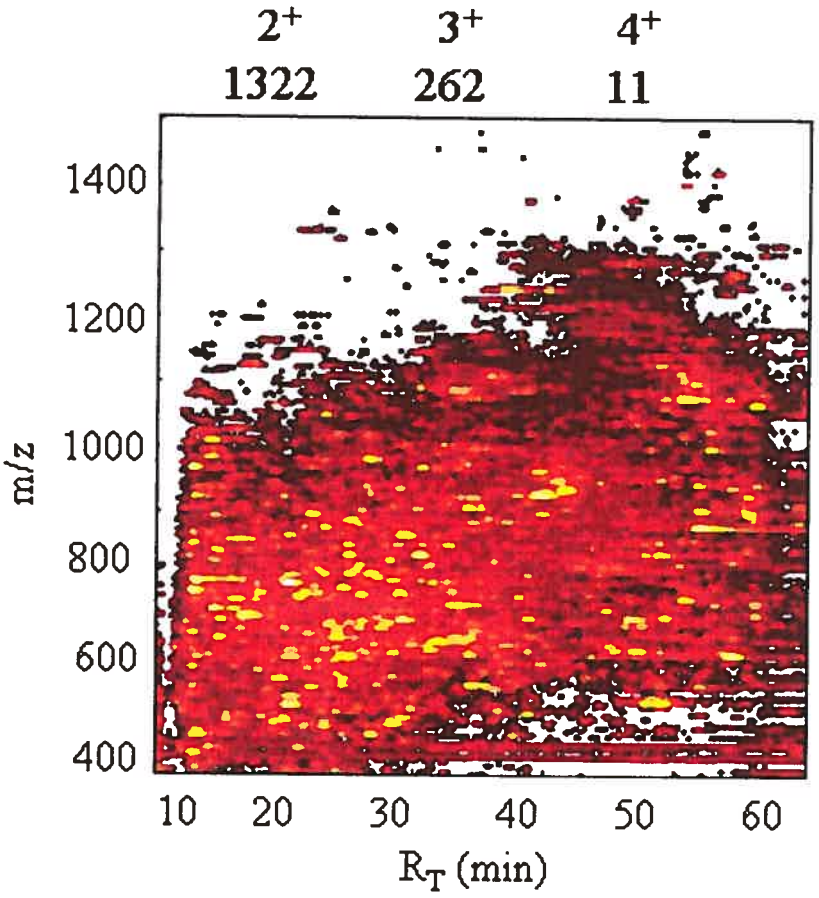

Figure 4.34: LC/MS isotope map of Band 10 from the SDS-PAGE separation of a *H. pylori* periplasmic extract sample.

4.3.1.2 2D-capLC/MS/MS

The tryptic digest of the periplasmic extract was separated into 30 fractions by SCX chromatography (Fig. 4.35) and each fraction was analysed by capLC/MS/MS. A total of 20234 MS/MS spectra with a precursor charge of +2 and +3 were acquired out of which 5610 were assigned to 362 unique protein clusters. The MS/MS spectra showing intense signals at masses of common carbohydrate oxonium ions were extracted by in-house software and manually interpreted. More than 50 spectra showed intense oxonium ions at m/z 317. These spectra were manually interpreted and identified. All of these spectra were assigned to FlaA, FlaB and one spectrum to the flagellar hook protein. A significant extent of unspecific cleavages of the peptide backbone was observed, explaining the unusually high number of detected glycopeptides. Interestingly, all of the detected glycopeptides eluted in the first 10 minutes of the gradient and a huge proportion of them eluted in the dead volume. The isotope map of the fraction containing the dead volume and of fraction 8 is shown in figure 4.36. The isotope map of the dead volume showed a peptide density similar to the fractions from the bands of the SDS-PAGE approach, but the signals from the glycopeptides were much more intense with signal intensities normally above 1000 counts. A Mascot search on this fraction only identified one regular unmodified tryptic peptide. The other signals were either non peptidic, glycopeptides or tryptic peptides with carbamylated N-termini because urea was used as denaturing agent for the tryptic digestion. The other early fractions showed less dense isotope maps as illustrated in figure 4.36b for fraction 8. In fraction 8 the glycopeptide

ISFTNDSAVSR from the flagellar hook protein was identified and the MS/MS spectrum is shown in figure 4.37, showing several y-fragment ions from the peptide backbone.

All acquired MS/MS spectra were manually inspected for characteristic glycospectra, but no other glycan was identified.

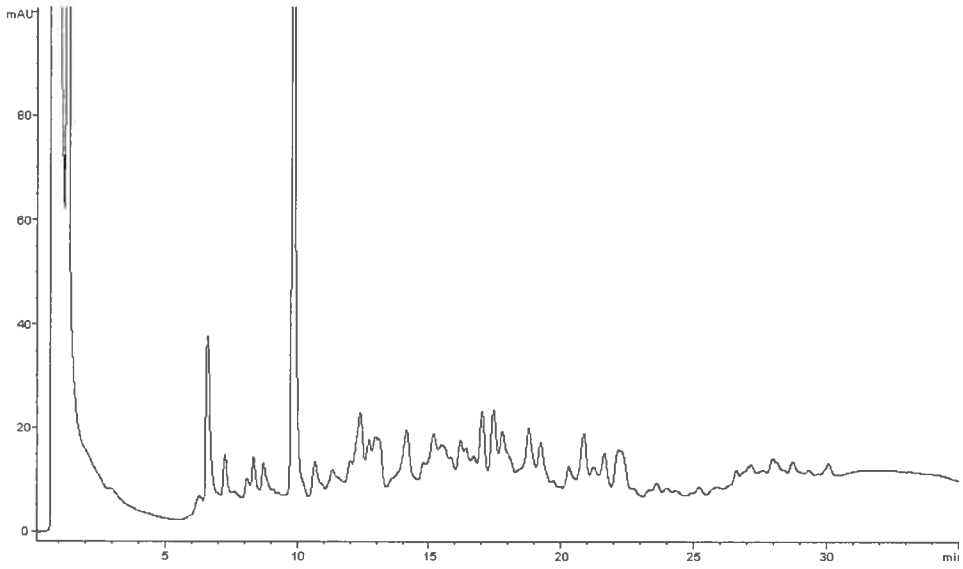

Figure 4.35: SCX fraction collection of the tryptic digest of the periplasmic extract from *H. pylori*. Fractions were collected every minute. All glycopeptides eluted in the first 10 min.

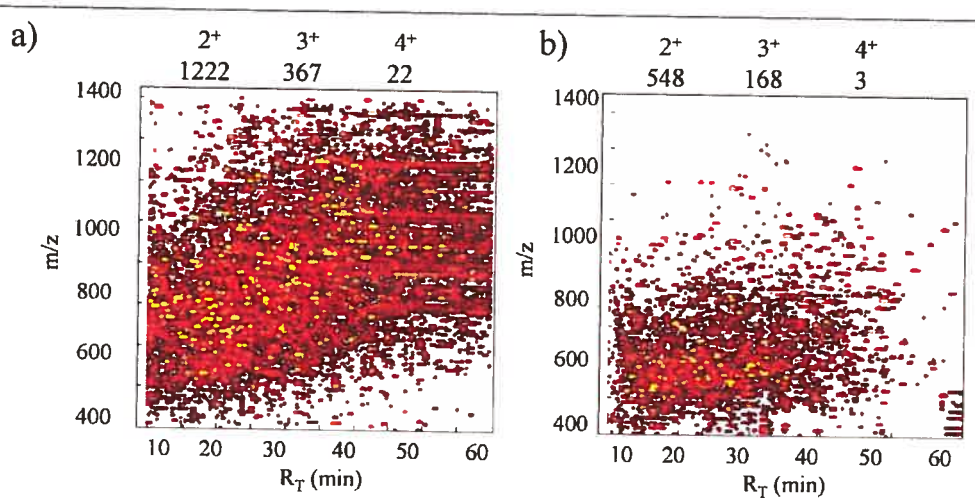

Figure 4.36: Isotope maps of a) fraction 2 and b) fraction 8 of the SCX separation of the PPE. These fractions contained several glycopeptides. Fraction 2 showed a complexity similar to the bands of the SDS-PAGE approach. Fraction 8 was less complex and contained the glycopeptide ISFTNDSAVSR from the flagellar hook protein. The amount of observed multiply-charged ions observed in each capLC/MS analysis is shown above the corresponding isotope maps.

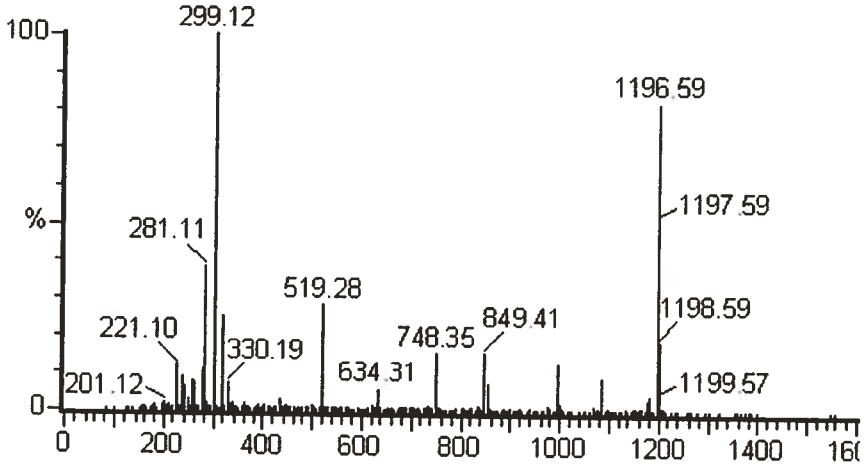

Figure 4.37: MS/MS spectrum of the glycopeptide ISFTNDSAVSR from the flagellar hook protein. Due to the higher intensity of the precursor ion compared to the SDS-PAGE approach the quality of the MS/MS spectrum was much better. Intense y-fragment ions enabled the easy identification of the peptide.

4.3.2 Analysis of a membrane extract

The membrane extract was only separated by SCX because of the better results obtained by this technique for the analysis of the periplasmic extract. The UV spectrum of the SCX run is shown in figure 4.38. Each fraction was analyzed by capLC/MS/MS and a total of more than 35000 MS/MS spectra were acquired, comprising 27702 MS/MS spectra with charge state +2 and +3. Out of these spectra, 12965 MS/MS were matched to 2684 unique proteins. After manual validation of the spectra and removal of redundancy a total of 875 unique proteins were identified and table 4.3 lists the top 30 hits based on the corresponding protein score. The interpretation of the extracted spectra with intense oxonium ions at common carbohydrates masses revealed more than 60 spectra showing intense oxonium ions at m/z 317 for Pse5Ac7Ac, all belonging to FlaA, FlaB and one to the flagellar hook protein. The signal intensities for the glycopeptides were very high abundant, in general above 1000 counts for the FlaA and FlaB protein and around 800 counts for the glycopeptide of the flagellar hook protein. The digestion of the membrane extract also yielded extensive unspecific cleavages as reflected by the high number of peptides modified with Pse5Ac7Ac.

In a different experiment performed by Dr. Susan Logan (NRC, Ottawa, Canada) a polyclonal antibody against the flagellin protein was employed on a whole cell extract. The result is presented in figure 4.39 and shows three bands giving a positive reaction. We believe that one of the antibodies was directed towards Pse5Ac7Ac on the flagellin protein and that this antibody reacted with the flagellar

hook protein which was shown here to be glycosylated with Pse5Ac7Ac. The MW of these bands were correlated to FlaA, FlaB and the flagellar hook protein.

A modified Mascot search using variable modifications for masses of common neutral sugars, which do not show oxonium ions from the carbohydrates was performed, but did not identify any novel glycoproteins. After this analysis, the ~35000 spectra were manually inspected for characteristic glycopeptide spectra but no new glycoproteins were identified. It is noteworthy that Fisher et al.[121] reported the RecA protein, a protein involved in DNA repair, to be post-translational modified on the first asparagine residue of the peptide MNTTLIFINQIR. The nature of the post-translational modification is still unknown and will probably not be Pse5Ac7Ac because we would expect then a reaction of the polyclonal antibody with the RecA protein (Fig. 4.39). The glycosylation of this peptide could however not be confirmed. The peptide was identified only unmodified in the membrane extract (Fig. 4.40). A possibility is that this peptide might be only partially glycosylated and only the non-glycosylated form was identified in this experiment.

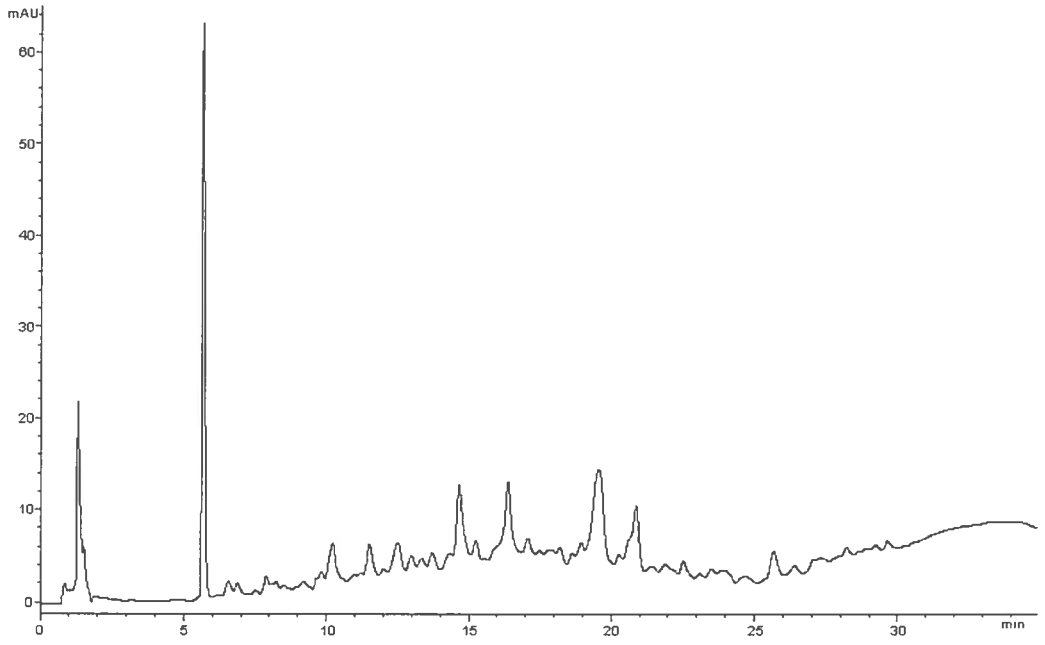

Figure 4.38: UV chromatogram of the SCX separation of the membrane extract of *H. pylori*.

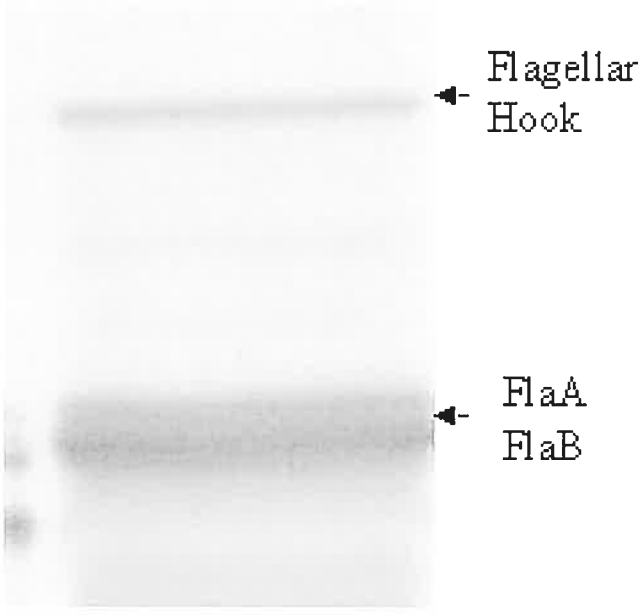

Figure 4.39: Whole cell extract of *H. pylori* probed with an antibody against Pse5Ac7Ac.
Three bands showed a positive reaction with the antibody. The MW of these bands were correlated to FlaA, FlaB and the flagellar hook protein.

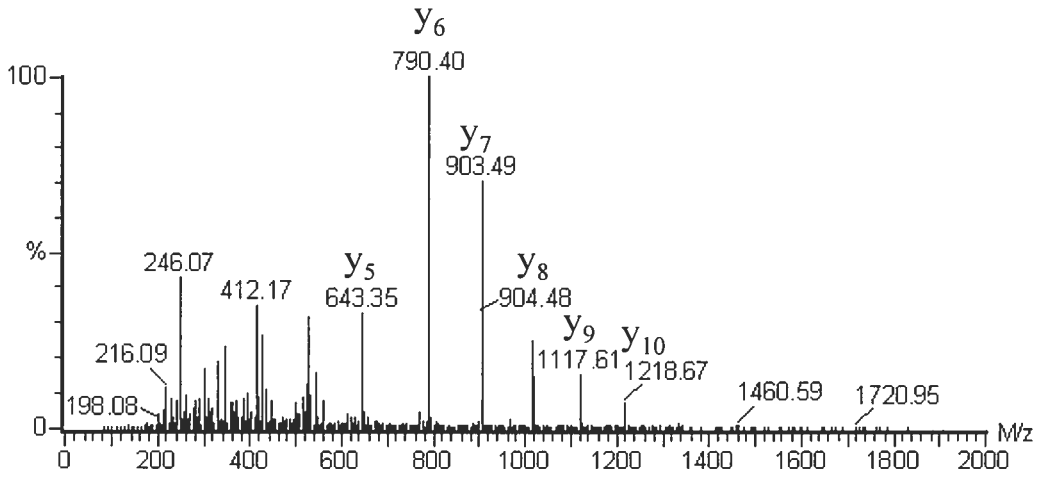

Figure 4.40: MS/MS spectrum of m/z 732.4.

The MS/MS spectrum shows the peptide MNTTLIFINQIR from the RecA protein. No post-translational modification was found to be attached to this peptide.

Table 4.3: Top 30 protein hits identified in the membrane extract. (Based on the protein score)

Accession	Protein Score	Mass kDa	Coverage(%)	Description
gi|15645812	4790	323,57	39,86	DNA-directed RNA polymerase beta subunit (rpoB)
gi|19338968	2510	58,24	66,67	heat shock protein B subunit
gi|15645398	2179	92,79	47,83	aconitase B (acnB)
gi|15645819	1933	43,65	74,44	translation elongation factor EF-Tu (tufB)
gi|15645318	1823	78,53	53,86	hydantoin utilization protein A (hyuA)
gi|15611615	1738	53,28	61,96	FLAGELLIN A [Helicobacter pylori J99]
gi|2197129	1712	51,45	61,41	F1F0-ATPase beta subunit
gi|15646010	1710	94,83	47,98	iron(III) dicitrate transport protein (fecA)
gi|15645027	1677	62,83	60,79	ribosomal protein S1 (rps1)
gi|1743248	1674	58,64	54,26	KatA catalase
gi|15644729	1606	74,47	41,78	methyl-accepting chemotaxis protein (tlpA)
gi|15645405	1601	99,08	42,2	preprotein translocase subunit (secA)
gi|15612491	1566	94,76	44,95	putative IRON(III) DICITRATE TRANSPORT PROTEIN
gi|15611138	1543	61,68	52,72	UREASE BETA SUBUNIT
gi|15645256	1535	64,4	59,69	quinone-reactive Ni/Fe hydrogenase large subunit (hydB)
gi|15645319	1483	86,51	37,52	N-methylhydantoinase
gi|15611171	1463	67,12	48,71	70kDa chaperone
gi|15611248	1437	80,12	46,22	Fumarate reductase
gi|2058519	1437	80,16	41,4	FrdA
gi|15611119	1401	135,22	27,51	Proline/pyrroline-5-carboxylate dehydrogenase
gi|15645414	1392	51,99	63,86	trigger factor (tig)
gi|15611642	1367	64,39	51,21	Hydrogenase large subunit
gi|15611688	1356	90,23	38,32	RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE 1 ALPHA CHAIN
gi|15611790	1318	98,89	34,91	PREPROTEIN TRANSLOCASE SUBUNIT
gi|15644716	1307	50,71	50,44	conserved hypothetical protein
gi|15645880	1278	94,23	28,55	NADH-ubiquinone oxidoreductase NQO3 subunit (NQO3)
gi|15645989	1274	94,42	39,52	ATP-dependent protease (lon)
gi|15644687	1226	135,14	22,7	delta-1-pyrroline-5-carboxylate dehydrogenase
gi|18075728	1198	55,27	44,14	ATP synthase F1 subunit alpha
gi|15646039	1155	77,39	47,17	conserved hypothetical ATP-binding protein

4.3.2.1 Discussion

When complex mixtures are analyzed it is important to extract and enrich the glycoproteins or glycopeptides to simplify the analysis. Probably, the best way to do this is to extract the glycoproteins by lectin extraction. This can be done for eukaryotic glycoproteins and was used to detect the N-linked glycans in *C. jejuni*.[28] However, this is not possible in most cases for prokaryotic glycoproteins, because of the unusual structure of oligosaccharides for which lectins may not be available. (see section 4.1). Therefore, the periplasmic extract was simplified by SDS-PAGE and SCX and analysed by capLC/MS/MS. The amount of identified proteins was similar in both techniques, and 296 unique protein clusters were obtained by SDS-PAGE and 362 by SCX, respectively. However, the number of identified glycopeptides was significantly different. Seven glycopeptides were identified by the SDS-PAGE approach, whereas more than 50 were obtained by the SCX approach resulting in a better coverage of the glycosylation on FlaA and FlaB. Due to a high amount of non-specific cleavages in the SCX approach the total number of detected glycopeptides was higher than the theoretical number of glycopeptides. However, the SCX approach proved to be more efficient at separating glycopeptides, as these were enriched in specific salt fractions. Most of the glycopeptides were not retained by the SCX column and eluted in the first few fractions. These fractions, except the fraction of the dead volume, were less complex, showing lower suppression effects, and higher amount of samples could be injected thus favoring their detection by the MS/MS experiment. Peptides found in the unretained fraction were either non peptidic,

glycopeptides or peptides carbamylated at the N-terminus. Carbamylation is a side reaction when using urea as denaturing agent during the tryptic digestion. The reaction scheme is shown in figure 4.41. Urea is decomposed to ammonium and cyanate. Isocyanic acid then reacts with the N-terminal end of the peptide and forms a carbamylated peptide. Interestingly, this small change shifts the isoelectric point of the peptide making it more "neutral" under the used pH condition used. (pH 3.5) leaving the C-terminal Lys/Arg residues as the only site for protonation. Similarly glycopeptides bearing the negatively charged Pse5Ac7Ac residue are typically more acidic than their corresponding non glycosylated tryptic peptides.

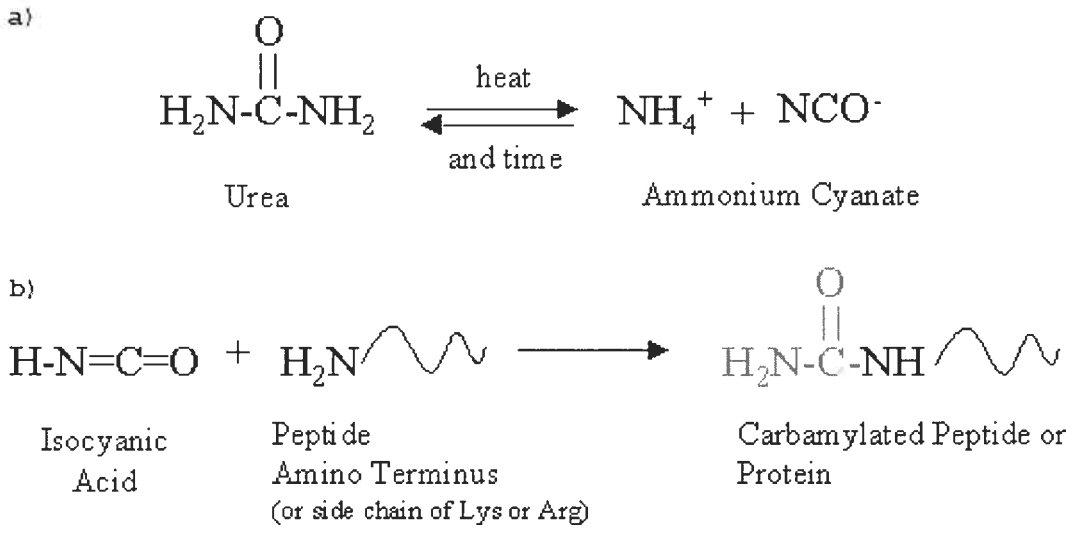

Figure 4.41: a) Reaction scheme of the decomposition of urea. b) Carbamylation reaction of the peptide/protein.

The higher proportion of glycopeptides identified using the two dimensional separation approach prompted us to examine the *H. pylori* membrane extract using this analytical strategy. A total of 875 unique protein clusters were identified. The

whole genome of this organism consist of 1590 predicted ORFs, and our sequencing efforts led to the identification of 55 % of these anticipated protein population.[122] Our results correspond to the highest number of identified proteins from *H. pylori*. Extensive data analysis of all fractions and more particularly of the first few fractions identified three proteins to be glycosylated with Pse5Ac7Ac, namely FlaA, FlaB and the flagellar hook protein. This is the first report showing that the flagellar hook protein is glycosylated with Pse5Ac7Ac. We also observed a significant number of unspecific cleavages of the peptide backbone which complicated the data interpretation. In general, MS/MS spectra of glycopeptides lack y-fragment ions from the peptide backbone, or are of low intensity. Only the mass of the peptide can be identified for most glycopeptide MS/MS, which represent a sizeable problem when analysing complex mixtures. The mass of specific peptides can be correlated to several theoretical tryptic peptides, and unambiguous identification without an adequate sequence tag is not possible. A situation which is made worse by the occurrence of unspecific cleavages.

Only three proteins were identified to be glycosylated with Pse5Ac7Ac. This is consistent with the results obtained from the immunoblotting experiment with the antibody against Pse5Ac7Ac which only showed positive reactions for three bands later identified as FlaA, FlaB and the flagellar hook protein. From these combined results we propose that glycosylation involving Pse5Ac7Ac is a very specific process, occurring primarily on proteins involved in the flagellum assembly.

To identify different glycans as Pse5Ac7Ac all glycospectra were manually inspected and a Mascot search with variable modifications for common neutral sugars

was performed to detect glycans not showing intense oxonium ions in the low mass region. However, no additional glycopeptides were identified using these approaches. Fischer et al.[120] described the RecA protein of *H. pylori* to be post-translational modified, but this could not be confirmed by our analysis. The peptide mentioned to be post-translational modified was identified as unmodified, but it may be only partially glycosylated and only the non-glycosylated form was identified by our experiments.

Based on this analysis it seems that glycosylation in *H. pylori* is a very rare process. Further, experiments involving better enrichment techniques would be required to confirm if other forms of glycosylation exists in *H. pylori*.

5 Conclusion

Studies on the flagellin from *H. pylori*, *Listeria monocytogenes*, *Aeromonas caviae* and *Pseudomonas aeruginosa* showed that protein glycosylation of the flagellin protein is much more prevalent than previously thought. The flagellin of each bacteria was shown to be post-translationally modified with O-linked glycans. The structures of the glycans were analyzed and showed the presence of different carbohydrate residues as substrates in the glycosylation process as well as variations in the nature of the carbohydrate moieties in different strains of the same bacteria. For example, *H. pylori* was glycosylated on 7 sites on FlaA and 10 sites on FlaB; *Listeria monocytogenes* was glycosylated with a single N-acetyl glucosamine on 4 to 6 sites; *P. aeruginosa* strain PAK was glycosylated at 2 sites with an hetererolgous glycan comprising up to 11 monosaccharide residues. The proximal carbohydrate residue where the glycan is linked to the protein backbone is a rhamnose residue and the structure is capped by a unique trisaccharide unrelated to the O chain repeat unit of the LPS. Strain jj692 was glycosylated at 2 sites with a single rhamnose residue and strain PAO at two sites with either a rhamnose or a glycan residue of mass 355 Da. The exact structure of the 355 Da modification could not be revealed unambiguously solely by mass spectrometry. However, it was possible to propose a potential structure consisting of a rhamnose, a phosphate group and an unknown residue of mass 129 Da with the elemental formula $C_6H_{12}NO_2$, based on the observation of characteristic fragment ions and the exact mass measurement. The genetic

glycosylation island found in strains jj692 and PAO was less complex than in strain PAK, as reflected by the glycan attached to the flagellin protein. *Aeromonas caviae* was glycosylated with a carbohydrate residue of 374 Da, which is related to Pse5Ac7Ac and was tentatively assigned as Pse5Ac7Ac9Ac. For the flagellin of *C. jejuni*, a novel carbohydrate with a residue mass of 485 Da was identified, in addition to the residues Pse5Ac7Ac and PseAm, which were already reported by Thibault et al.[24] Based on structural analysis of this 485 residue, it is proposed to be a disaccharide comprising an acetamido dideoxyhexose linked to PseAm. In general, the identification of the exact glycan structure is very challenging and cannot be determined solely by mass spectrometry. Further experiments by NMR are required if the exact structure of the glycan is to be determined.

The linkage site of these modifications was determined for *H. pylori*, *L. monocytogenes*, *P. aeruginosa* strain PAK, strain jj692 and strain PAO. In all cases the glycans were O-linked to serine or threonine residues with no specific consensus sequence. Rather glycosylation appears to be related to surface accessibility of Ser/Thr residues in the folded protein. Interestingly, in every case the glycosylation was located in the highly variable, central, surface exposed region of the flagellin protein. (Fig. 5.1). A summary of the results obtained for the different bacteria is shown in table 5.1.

Table 5.1: Summary of the identified O-linked glycans from bacterial flagellin.

Bacterium	Strain	Extent of PTM	Carbohydrate structure
Helicobacter pylori	1061	~4%	Pse5Ac7Ac
Aeromonas caviae	UU51	~7 %	Pse5Ac7Ac9Ac
Listeria monocytogenes	568, 394, 2568, 23485	~2%-4%	HexNAc
Pseudomonas aeruginosa	PAK	~4-7%	Heterologous glycan compromising up to 11 carbohydrate residues
	jj692	~1%	Rhamnose
	PAO	~1 %	Rhamnose and an unknown residue of mass 355 Da

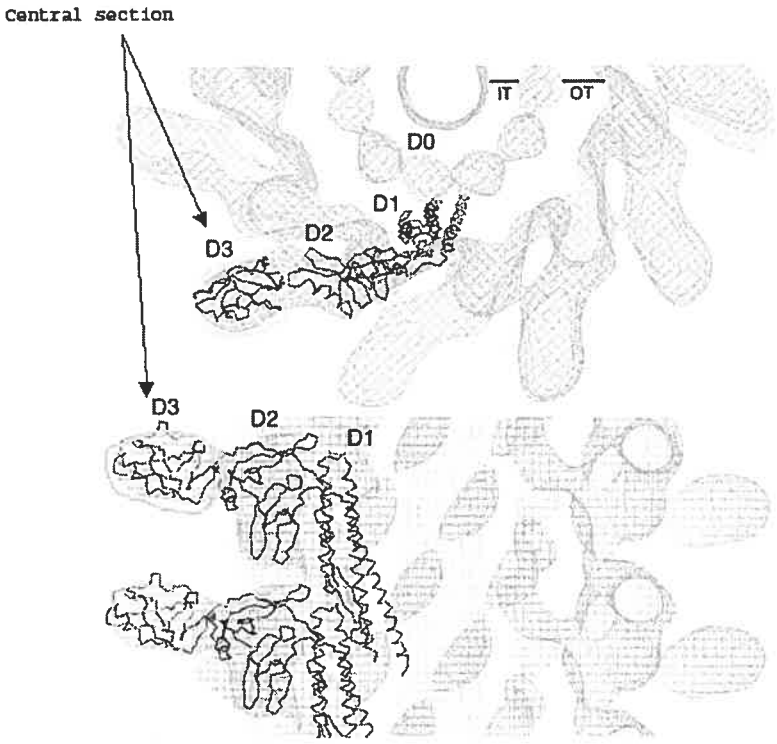

Figure 5.1: Structure of the flagellin filament of *Salmonella* typhimurium.[33]

Top panel shows the end-on view from the top of the filament and the bottom panel shows the side view. The continuous lines represent single flagellin proteins. D3 corresponds to the central region of the flagellin protein that is surface exposed. This domain was in all cases shown to be glycosylated.

As mentioned in the review from Benz et al.[22] the role of glycosylation might be divided into two groups: First the carbohydrate residue could directly affect the function of the protein, such as the assembly of the flagellin as observed for *C. jejuni*[24] and *H. pylori* (see section 4.1.1), where glycan-deficient mutants were not able to assemble the flagellin protein properly. Secondly, the glycosylation could influences the immune response of the host.[22] However, glycosylation of bacterial flagella utilizes novel carbohydrate moieties and we believe that the biosynthesis of these structures and assembly in the flagella filament offers a range of new targets for therapeutic intervention.

Genes involved in the glycosylation process were identified for *H. pylori*, *C. jejuni*, *Pseudomonas* PAK, strain PAO and *L. monocytogenes*. Isogenic mutants were generated and analysed by mass spectrometry. These analyses confirmed that these genes were indeed involved in the glycosylation process of the flagellin protein and their functions could be established from these analyses.

The complete characterization of protein glycosylation and the identification of genes involved in this glycosylation process is a very tedious and time consuming procedure, which can take from several months to years. To simplify this analysis, a novel "top down" approach was developed, and reduced the analysis time significantly. It is based on conducting MS/MS experiments on the multiply charged ions from the intact proteins. Due to the labile nature of the post-translational modifications, the glycan is cleaved from the protein backbone and forms oxonium ions, which give intense signals in the low mass region. This approach directly gives information about, the amount, the mass and structural information concerning the

post-translational modifications. Additionally, changes in the glycosylation pattern of isogenic mutants can now be quickly monitored, without requiring the time consuming "bottom up" approach (see section 4.2). However, the linkage site cannot be determined by this approach and neutral sugars like deoxyhexose or hexose cannot be detected, as these sugars do not form intense oxonium fragment ions. In cases where the linkage site needs to be determined or when the top down approach is not successful, the traditional "bottom up" approach needs to be used. The protocol for the bottom up approach used by Thibault et al.[24] was optimised in this thesis. This was achieved by using targeted capLC/MS/MS. An inclusion list was generated from the theoretical masses of the glycopeptide after ß-elimination. This approach enabled the maximization of useful MS/MS time while simultaneously increasing the quality of the collected data, without requiring fraction collection. A combination of the "top down" with the targeted "bottom up" approach gives the scientists a powerful tool for rapid characterization of post-translational modifications on proteins. This combined analytical approach is summarized in figure 5.2.

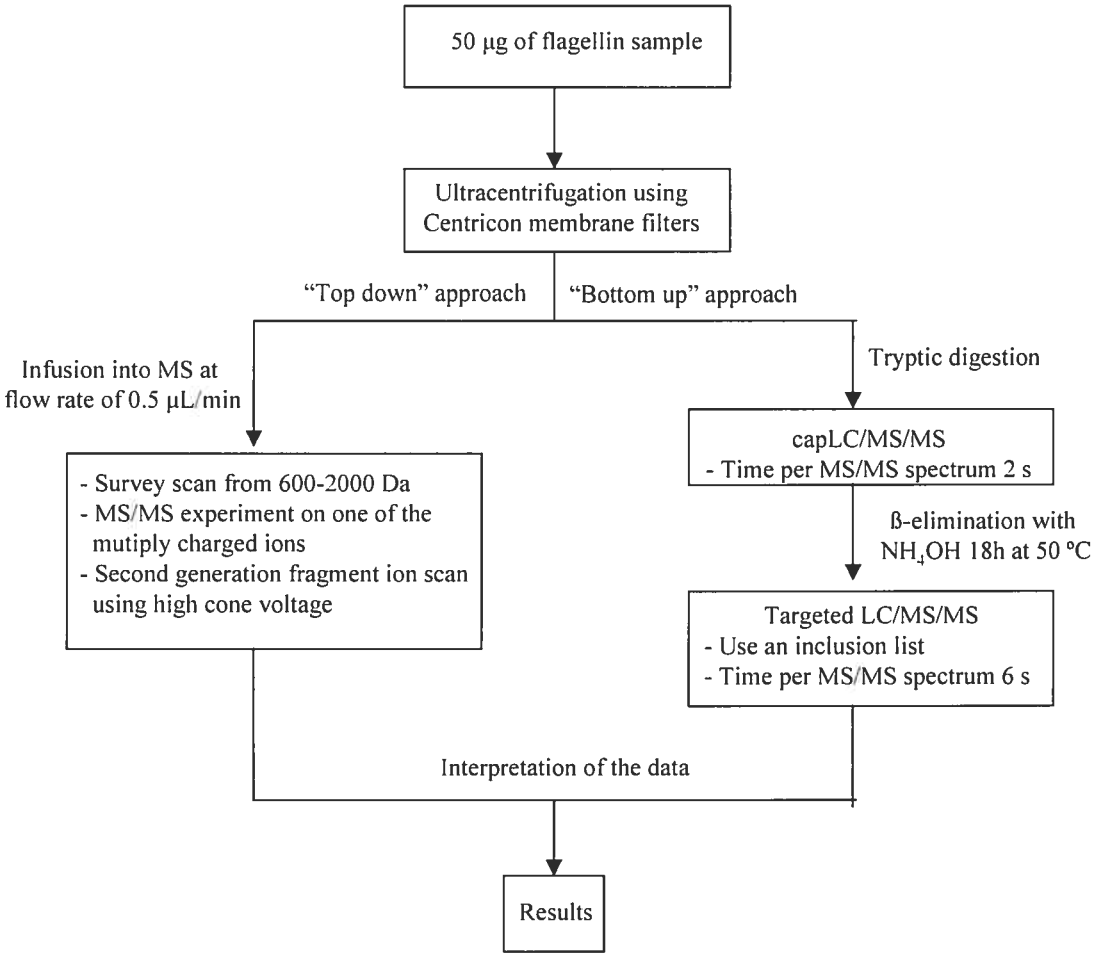

Figure 5.2: Analytical approach for the rapid characterization of post-translational modifications in simple protein mixtures.

Analysis of complex protein extracts of *H. pylori* showed that more efficient detection of glycopeptides was obtained using ion exchange chromatography compared to in-gel digest of excised bands of interest. This is due in part to the enrichment of glycopeptides in specific salt fractions. Interestingly, small modifications of the peptide backbone like carbamylation of the N-terminal end or glycosylation with Pse5Ac7Ac resulted in elution of the modified peptides in the unretained fraction or in the first few fractions of the salt gradient. These first fractions were extensively analysed by mass spectrometry using exclusion runs. Based on these analyses, glycosylation in *H. pylori* appears to be a rare event, primarily observed on FlaA, FlaB and the flagellar hook protein. (see section 4.3). This is consistent with results obtained with a polyclonal antibody sensitive to the flagellin protein and possibly displaying cross reactivity to the glycosylation on the flagellin protein. The antibody showed on a whole cell extract of *H. pylori* three positive reactions at masses corresponding to FlaA, FlaB and the flagellar hook proteins. From the capLC/MS/MS analyses of the membrane extract, more than 35000 spectra were acquired, corresponding to 875 unique protein clusters. Extensive data mining performed both manually and by in-house developed software of all acquired MS/MS spectra could not identify other types of glycans beside Pse5Ac7Ac. While we cannot rule out the presence of other glycoproteins, these results suggest that if present they would be expressed at levels significantly below our detection limits.

The main problem in the identification of glycopeptides in complex mixtures is their rather general fragmentation pattern (see section 4.3) making identification of

the corresponding protein in complex mixtures a challenging task, due to the lack of intense y-fragment ions from the protein backbone.

The results presented in this thesis have raised many interesting questions, which should be addressed in future work, such as:

- How is protein glycosylation in prokaryotic systems regulated?
- What motif on the flagellin molecule is recognized by the glycosyltransferase?
- What is the role of glycosylation in prokaryotes? Its assembly? Evasion of immune response?

Methods developed as part of this thesis will provide useful tools to further understand glycosylation pathways in general. Furthermore, new methods need to be developed to optimise the analysis of complex protein mixtures for PTMs, such as improving the extraction of glycopeptides, and their specific detection using dissociation techniques like electron capture dissociation (ECD) to address the problem with the fragmentation pattern of glycopeptides by collision induced dissociation.

6 References

1 Apweiler R, Hermjakob H, Sharon N, Biochim Biophys Acta. 1999;1473:4-8

2 Hart GW in "Capillary electrophoresis of Carbohydrates" (Thibault P, Honda S, Eds) Humana Press. 2003;3-26

3 Silberstein S, Gilmore R, FASEB J. 1996;10:849-58

4 Hurtley SM, Helenius A, Annu Rev Cell Biol. 1989;5:277-307

5 Imperiali B, O'Connor SE, Curr Opin Chem Biol. 1999;3:643-9

6 Wormald MR, Dwek RA, Structure Fold Des. 1999;7:155-60

7 Kundra R, Kornfeld S, Biol Chem. 1999;274:31039-46

8 Hakomori S, Adv Exp Med Biol. 2001;491:369-402

9 Couldrey C, Green JE, Breast Cancer Res. 2000;2:321-3

10 Van den Steen P, Rudd PM, Dwek RA, Opdenakker G, Crit Rev Biochem Mol Biol. 1998;33:151-208

11 Comer FI, Hart GW, Biochim Biophys Acta. 1999;1473:161-71

12 Slawson C, Hart GW, Curr Opin Struct Biol. 2003;13:631-6

13 Whelan SA, Hart GW, Circ Res. 2003;93:1047-58

14 Zachara NE, Hart GW, Biochim Biophys Acta. 2004;1673:13-28

15 Hanisch FG, Schwientek T, Von Bergwelt-Baildon MS, Schultze JL, Finn O, Eur J Immunol. 2003;33:3242-54

16 Okuda S; Weinbaum G, Biochemistry, 1975;30:23-56

17 Virji M, Gene. 1997;192:141-7

18 Castric P, Cassels FJ, Carlson RW, J Biol Chem. 2001;276:26479-85

19 Arora SK, Bangera M, Lory S, Ramphal R, Proc Natl Acad Sci U S A. 2001; 98:9342-7

20 Espitia C, Mancilla R, Clin Exp Immunol. 1989;77:378-83

21 Schmidt MA, Riley LW, Benz I, Trends Microbiol. 2003;11:554-61

22 Benz I, Schmidt MA, Mol Microbiol. 2002;45:267-76

23 Messner P, Schaffer, C. Fortschr Chem Org Naturst. 2003;85:51-124

24 Thibault P, Logan SM, Kelly JF, Brisson JR, Ewing CP, Trust TJ, Guerry P, J Biol Chem. 2001;276:34862-70

25 Kuo C, Takahashi N, Swanson AF, Ozeki Y, Hakomori S, J Clin Invest. 1996; 98:2813-8

26 Lindenthal C, Elsinghorst EA, Infect Immun. 1999;67:4084-91

27 Dobos KM, Khoo KH, Swiderek KM, Brennan PJ, Belisle JT, J Bacteriol. 1996; 178:2498-506

28 Young NM, Brisson JR, Kelly J, Watson DC, Tessier L, Lanthier PH, Jarrell HC, Cadotte N, St Michael F, Aberg E, Szymanski CM, J Biol Chem. 2002; 277:42530-9

29 Yonekura K, Maki-Yonekura S, Namba K, Nature. 2003;424:643-50

30 http://www.bio.cmu.edu/Courses/03441/TermPapers/old_tp/FlagellaMotor/ ExamplePaper1.gif

31 Larsen SH, Reader RW, Kort EN, Tso WW, Adler J, Nature. 1974; 249:74-7

32 Falke J, Bass RB, Butler SL, Chervitz SA, Danielson MA,. Annu Rev Cell Dev Biol. 1997;13:457-512

33 Samatey FA, Imada K, Nagashima S, Vonderviszt F, Kumasaka T, Yamamoto M, Namba K, Nature. 2001;410:331-7

34 Medzhitov R, Nat Rev Immunol. 2001;1:135-45

35 Aderem A, Ulevitch RJ, Nature. 2000;406:782-7

36 Monack DM, Mueller A, Falkow S, Nat Rev Microbiol. 2004;2:747-65

37 Eaton KA, Morgan DR, Krakowka S, Infect Immun. 1989;57:1119-25

38 Eaton KA, Morgan DR, Krakowka S, J Med Microbiol. 1992;37:123-7

39 Kostrzynska M, Betts JD, Austin JW, Trust TJ, J Bacteriol. 1991; 173:937-46

40 Josenhans C, Ferrero RL, Labigne A, Suerbaum S, Mol Microbiol. 1999; 33:350-62

41 Josenhans C, Vossebein L, Friedrich S, Suerbaum S, FEMS Microbiol Lett. 2002;210:165-72

42 Bodey GP, Bolivar R, Fainstein V, Jadeja L, Rev Infect Dis. 1983;5:279-313

43 Allison JS, Dawson M, Drake D, Montie TC, Infect Immun. 1985;49:770-4

44 Feldman M, Bryan R, Rajan S, Scheffler L, Brunnert S, Tang H, Prince A, Infect Immun. 1998;66:43-51

45 Montie TC, Doyle-Huntzinger D, Craven RC, Holder IA, Infect Immun. 1982; 38:1296-8

46 Castric P, Microbiology. 1995;141:1247-54

47 Castric P, Cassels FJ, Carlson RW, J Biol Chem. 2001;276:26479-85

48 Totten PA, Lory SJ, Bacteriol. 1990;172:7188-99

49 Brimer CD, Montie TC J, Bacteriol. 1998;180:3209-17

50 Arora SK, Bangera M, Lory S, Ramphal R, Proc Natl Acad Sci U S A. 2001; 98:9342-7

51 Arora SK, Wolfgang MC, Lory S, Ramphal RJ, Bacteriol. 2004;186:2115-22

52 Destro MT, Leitao MF, Farber JM, Appl Environ Microbiol. 1996;62:705-11

53 Johansson T, Rantala L, Palmu L, Honkanen-Buzalski T, Int J Food Microbiol. 1999;47:111-9

54 Miettinen MK, Bjorkroth KJ, Korkeala HJ, Int J Food Microbiol. 1999;46:187-92

55 Rorvik LM, Caugant DA, Yndestad M, Int J Food Microbiol. 1995;25:19-27

56 O'Driscoll B, Gahan CG, Hill C, Appl Environ Microbiol. 1996;62:1693-8

57 Walker SJ, Archer P, Banks JG, J Appl Bacteriol. 1990;68:157-62

58 Midelet G, Carpentier B, Appl Environ Microbiol. 2002;68:4015-24

59 Peel M, Donachie W, Shaw A, Gen Microbiol. 1988;134:2171-8

60 Seelinger, HPR, Hohne K, Methods in microbiology 1997;13:31-49

61 Brosch R, Chen J, Luchansky JB, Appl Environ Microbiol. 1994;60: 2584-92

62 Dons L, Rasmussen OF, Olsen JE, Mol Microbiol. 1992;6:2919-29

63 Butzler JP, Oosterom J, Int J Food Microbiol. 1991;12:1-8

64 Logan SM, Trust TJ, Guerry P, J Bacteriol. 1989;171:3031-8

65 Doig P, Kinsella N, Guerry P, Trust TJ, Mol Microbiol. 1996;19:379-87

66 Janda JM, Clin Microbiol Rev. 1994;4:397-410

67 Kirov SM, Barnett TC, Pepe CM, Strom MS, Albert MJ Infect Immun. 2000; 68:4040-8

68 Gryllos I, Shaw JG, Gavin R, Merino S, Tomas JM, Infect Immun. 2001; 69:65-74

69 Power MP, Jennings PM, FEMS Microbiology Letters. 2003;218:211-222

70 Sambri V, Stefanelli C, Cevenini R, Arch Microbiol. 1992;157:205-8

71 Garcia-Patrone M, Tandecarz JS, Mol Cell Biochem. 1995;145:29-37

72 Moens S, Vanderleyden J, Arch Microbiol. 1997;168:169-75

73 Linton D, Allan E, Karlyshev AV, Cronshaw AD, Wren BW, Mol Microbiol. 2002;43:497-508

74 Lechner J, Wieland F, Annu Rev Biochem. 1989;58:173-94

75 Comer JE, Marshall MA, Blanch VJ, Deal CD, Castric P, Infect Immun. 2002; 70:2837-45

76 Rademaker GJ, Pergantis SA, Blok-Tip L, Langridge JI, Kleen A, Thomas-Oates JE, Anal Biochem. 1998;257:149-60

77 Karas M, Hillenkamp F, Anal Chem. 1988;60:2299-301

78 Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM, Science. 1989; 246:64-71

79 Gaskell JS, J Mass Spectrom. 1997;32:677-88

80 Wilm M, Mann M, Anal Chem. 1996;68:1-8

81 Olsen JV, Ong SE, Mann, M Mol Cell Proteomics. 2004;3:608-14

82 Roepstorff P, Fohlman, J Biomed Mass Spectrom. 1984;11:601

83 Steen H, Mann M, Nat Rev Mol Cell Biol. 2004;5:699-711

84 Schirm M, Soo EC, Aubry AJ, Austin J, Thibault P, Logan SM, Mol Microbiol. 2003;48:1579-92

85 Schirm M, Kalmokoff M, Aubry A, Thibault P, Sandoz M, Logan SM, J Bacteriol. 2004;186:6721-7

86 Nolling J et al., J Bacteriol. 2001;183:4823-38

87 Luneberg E, Zetzmann N, Alber D, Knirel YA, Kooistra O, Zahringer U, Frosch M, Int J Med Microbiol. 2000;290:37-49

88 Stimson E, Virji M, Makepeace K, Dell A, Morris HR, Payne G, Saunders JR, Jennings MP, Barker S, Panico M, Mol Microbiol. 1995;17:1201-14

89 Schirm M, Arora SK, Verma A, Vinogradov E, Thibault P, Ramphal R, Logan SM, J Bacteriol. 2004;186:2523-31

90 Guo D, Bowden MG, Pershad R, Kaplan HB, J Bacteriol. 1996;178:1631-9

91 Knirel YA, Vinogradov EV, Shashkov AS, Dmitriev BA, Kochetkov NK, Stanislavsky ES, Mashilova GM, Eur J Biochem. 1985;150:541-50

92 Marolda CL, Valvano MA, J Bacteriol. 1993;175:148-58

93 Spiro RG, Glycobiology. 2002;12:43-56

94 Rahim R, Burrows LL, Monteiro MA, Perry MB, Lam JS, Microbiology. 2000; 146:2803-14

95 Rahim R, Ochsner UA, Olvera C, Graninger M, Messner P, Lam JS, Soberon-Chavez G, Mol Microbiol. 2001;40:708-18

96 Kownatzki R, Tummler B, Doring G, Lancet. 1987;1:1026-7

97 Deziel E, Lepine F, Milot S, Villemur R, Microbiology. 2003;149:2005-13

98 Goon S, Kelly JF, Logan SM, Ewing CP, Guerry P, Mol Microbiol. 2003; 50:659-71

99 Szymanski CM, Logan SM, Linton D, Wren BW, Trends Microbiol. 2003; 11:233-8

100 Donnelly MA, Steiner TS, J Biol Chem. 2002;277:40456-61

101 Jacchieri SG, Torquato R, Brentani RR, J Bacteriol. 2003;185:4243-7

102 Takeuchi K, Taguchi F, Inagaki Y, Toyoda K, Shiraishi T, Ichinose Y, J Bacteriol. 2003;185:6658-65

103 Kelly-Wintenberg K, Anderson T, Montie TC, J Bacteriol. 1990;172:5135-9

104 Glaser P et al., Science. 2001;294:849-52

105 Brosch R, Chen J, Luchansky JB, Appl Environ Microbiol. 1994;60:2584-92

106 Nelson KE et al., Nucleic Acids Res. 2004;32:2386-95

107 Moens S, Michiels K, Keijers V, Van Leuven F, Vanderleyden J, J Bacteriol. 1995;177:5419-26

108 Brahamsha B, Greenberg EP, J Bacteriol. 1989;171:1692-7

109 Bedouet L, Arnold F, Robreau G, Batina P, Talbot F, Binet A, Microbios. 1998;94:183-92

110 Torres CR, Hart GW, J Biol Chem. 1984;259:3308-17

111 Wells L, Hart GW, FEBS Lett. 2003;546:154-8

112 Joys TM, SAAS Bull Biochem Biotechnol. 1991;4:56-9

113 Gorski L, Palumbo JD, Mandrell RE, Appl Environ Microbiol. 2003;69:258-66

114 Vatanyoopaisarn S, Nazli A, Dodd CE, Rees CE, Waites WM, Appl Environ Microbiol. 2000;66:860-3

115 Way SS, Thompson LJ, Lopes JE, Hajjar AM, Kollmann TR, Freitag NE, Wilson CB, Cell Microbiol. 2004;6:235-42

116 Dons L, Eriksson E, Jin Y, Rottenberg ME, Kristensson K, Larsen CN, Bresciani J, Olsen JE, Infect Immun. 2004;72:3237-44

117 Fridriksson EK, Beavil A, Holowka D, Gould HJ, Baird B, McLafferty FW, Biochemistry. 2000;39:3369-76

118 Sze SK, Ge Y, Oh H, McLafferty FW, Proc Natl Acad Sci U S A. 2002; 99:1774-9

119 Reid GE, Stephenson JL Jr, McLuckey SA, Anal Chem. 2002;74:577-83

120 Kelleher NL, Zubarev RA, Bush K, Furie B, Furie BC, McLafferty FW, Walsh CT, Anal Chem. 1999;71:4250-3

121 Fischer W, Haas R, Bacteriol. 2004;186:777-84

122 Tomb JF, White O et al., Nature. 1997;388:539-47

List of Publications and Presentations

A. Publications in refereed journals:

1. Schirm M, Soo EC, Aubry AJ, Austin J, Thibault P, Logan SM, Structural, genetic and functional characterization of the flagellin glycosylation process in Helicobacter pylori. Mol Microbiol. 2003 Jun;48(6):1579-92

2. Schirm M, Arora SK, Verma A, Vinogradov E, Thibault P, Ramphal R, Logan SM, Structural and genetic characterization of glycosylation of type a flagellin in Pseudomonas aeruginosa. J Bacteriol. 2004 May;186(9):2523-31

3. Schirm M, Kalmokoff M, Aubry A, Thibault P, Sandoz M, Logan SM, Flagellin from Listeria monocytogenes is glycosylated with beta-O-linked N-acetylglucosamine. J Bacteriol. 2004 Oct;186(20):6721-7

B. Oral presentations at conferences:

1. The 13th Annual Frederick conference on Capillary Electrophoresis; October 21-22, 2002 (Frederick, Usa). Identification and characterization of protein glycosylation on bacterial flagellin by nanoLC/MS and CE/MS.

2. 51th ASMS conference 2003 June 8-12, 2003 (Montreal, Canada). Application of proteomics for mapping post-translational modifications in *Pseudomonas aeruginosa.*

3. The 16th Annual Tandem Mass Spectrometry Workshop, December 4-6, 2003 (Lake Louise, Canada). A novel approach in mapping O-linked glycosylation in bacteria through comparison of isotope maps before and after ß-elimination followed by targeted MS/MS.