

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

**Biologia Cellulare, Molecolare e Industriale:
Biologia Funzionale dei Sistemi Cellulari e
Molecolari**

Ciclo XXIV

**Settore Concorsuale di afferenza: 05/E2
Settore Scientifico disciplinare: BIO-11**

**Insights in the maturation of pathogenic bacteria
vaccine candidates using mass spectrometry
based approaches**

Presentata da: Danilo Donnarumma

Coordinatore Dottorato

Prof. Scarlato Vincenzo

Relatore

Dott.sa Norais Nathalie

Prof. Scarlato Vincenzo

Esame finale anno 2012

Index

1 Introduction	1
1.1 Brief history of vaccines	1
1.2 Maturation of proteins	4
<i>Non-covalent modifications</i>	5
<i>Covalent modifications</i>	7
1.3 Mass spectrometry	12
1.4 Aim of the thesis	23
2 Results and discussion	25
2.1 Investigation on the maturation of different vaccine candidates using native mass spectrometry	25
<i>Introduction</i>	25
<i>Assignment of the oligomerization states of proteic vaccine candidates</i>	26
<i>Characterization of the mass increase found on PSL1</i>	28
<i>Characterization of the mass increase found on GNA1030</i>	35
<i>Characterization of the mass decrease found on SAL1486</i>	42
<i>Isopeptide bonds in SdrC, an adhesin from S. aureus</i>	49
2.2 Development of MS-based approaches to identify unknown PTMs in pathogenic bacteria	57
<i>Introduction</i>	57
<i>Selection of the starting material for PTMs discovery</i>	58
<i>Characterization of the OMVs and PTMs discovery</i>	59
3 Conclusions	66
4 Materials and Methods	77
5 References	86

1 Introduction

1.1 Brief history of vaccines

Vaccination is one of the most effective tools for the prevention of infectious diseases. The basic paradigm of vaccine development established by Louis Pasteur at the end of the 19th century (i.e. isolation, inactivation and injection of the causative agent) constitutes the foundation of classical vaccinology and led vaccine development throughout the 20th century [1]. Conventional approaches on the basis of these empiric principles have provided vaccines from three major categories: inactivated microorganisms, live-modified agents and subunit vaccines (composed by purified portions of the infectious agent) [2]. Not all pathogens, however, can be grown in culture and some microorganisms may require specific, sometimes expensive, conditions for growth. Others limitations include the necessity of safety procedures for the pathogen manipulation, and the possibility of an insufficient killing or attenuation, that may result in the presence of virulent organisms in the final vaccine. Recombinant DNA technologies have been used for the design of second-generation vaccines, to obtain rationally attenuated strains or highly purified antigenic components. Examples include bacterial toxins detoxified by molecular engineering, such as the pertussis toxin [3]. This approach, however, even if more refined, could require years [4], and in some cases, the conventional empiric approach was just insufficient to find appropriate solutions for the development of universal vaccines (i.e. meningococcus B). Although the history of genomics research can be traced back to the 1970s, with the development of DNA sequencing technology, the late 1990s marked

the beginning of the so-called genomics era, with the first complete genome sequenced of the free-living organism *Haemophilus influenzae* in 1995 [5]. Since then, emerging technologies have allowed the sequencing of a genome to be completed very quickly and sequencing of entire genomes has become a commonly used practice in research [6]. The approach referred to as 'reverse vaccinology' uses the genome sequences of viral, bacterial or parasitic pathogens of interest rather than the cells as starting material for the identification of novel antigens, whose activity should be subsequently confirmed by experimental biology [7]. In general, the aim is the identification of genes potentially encoding virulence factors (on the basis of sequence similarities to known pathogenic proteins) and secreted or surface-associated proteins (using several computational tools identifying the presence of amino acid motifs responsible for targeting the mature protein to the bacterial surface (signal peptides, anchoring motifs), to the lipid bilayer (lipoproteins), to the integral membrane (transmembrane domains) or for recognition and interaction with host structures). The selected ORF were expressed in a heterologous system, *Escherichia coli*, purified and tested as vaccine in an animal model of infection. When an antigen provides promising protection in animal model, it is further characterized before entering in clinical [8]. The first example of a successful application of the reverse vaccinology approach was the identification of vaccine candidates against *Neisseria meningitidis* serogroup B (MenB) that are now close to enter in the market. The success of reverse vaccinology for meningococcus has led to the application of this approach to a variety of other human pathogens, such as *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Chlamydia pneumoniae*, *Bacillus*

anthracis, *Porphyromonas gingivalis*, *Mycobacterium tuberculosis*, *Helicobacter pylori* and others [9], [10], [11], [12], [13], [14].

In spite of the success of the reverse vaccinology, several aspects that could not be assessed by the approach are currently emerging. One of these aspects is the impossibility to obtain information about the post-translational modifications (PTMs) of the putative vaccine candidates. Moreover the necessity to use heterologous recombinant proteins may results in changes in the maturation, compared to the native proteins, which can affect their immunogenicity. Indeed, during the maturation the physico-chemical properties of a protein are rapidly and reversibly modified, influencing a number of cellular processes like:

- oligomerization state
- Interaction with other proteins or cofactors
- change of enzyme activity by covalent or non-covalent “ group addition”
- sub-cellular localization
- half-life.

Among these processes, the covalent modifications (PTMs) are reported to be of pivotal importance in eukaryotes. Indeed, cells of multicellular organisms are known for their extensive networks of PTMs in which different modification pathways converge in signal integration. By contrast, bacteria have often been erroneously considered as simple sacs of metabolites, optimized for fast growth and devoid of the regulatory network based on post-translational modifications. Over the past years, this appears not to be the case, and it become fundamental to assess how these post-translational

modifications occurs, either at the pathogen level or in the host system during recombinant protein expression, and how they influence antigen immunogenicity and host recognition.

1.2 Maturation of proteins

The process of gene expression is not over when the genetic code has been used to create the sequence of amino acids that constitutes a protein. To be useful to the cell, this new polypeptide chain must fold up into its unique three-dimensional conformation, bind small-molecule cofactors required for its activity or assemble correctly with other protein subunits. These modifications are mainly achieved by the formation of non-covalent bonds. Moreover the folded proteins can also be covalently modified by protein kinases or other protein-modifying enzymes (Figure 1).

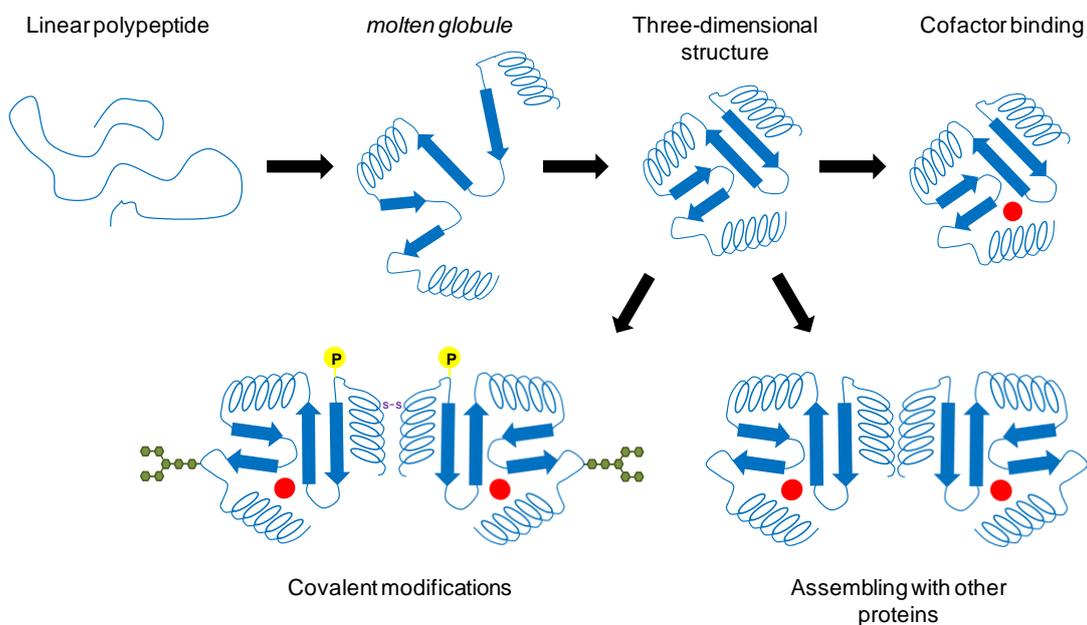


Figure 1: scheme of the maturation of a protein

During or after its synthesis, different regions of the polypeptide starts to acquire their secondary structures leading to the formation of the *molten globule*, this is the first step of the maturation. Subsequently the structure of the protein, alone or with the help of molecular chaperones, became more compact with the formation of appropriate hydrogen bonds and van der Waals, ionic, and hydrophobic interactions. In this way the linear, or one-dimensional, genetic message in the DNA is converted into the three-dimensional structure of the protein [15]. For some proteins the attainment of the tertiary structure it's the last step of the maturation but many others do not reach their final biologically active conformation until they have been modified by one or more processing reactions. These modifications can be divided in two main groups based on the type of chemical bonds involved, non-covalent or covalent.

Non-covalent modifications

Non-covalent modifications mainly include binding of cofactors and the assembling with other protein subunits and are very similar between prokaryotes and eukaryotes.

Cofactors binding: A cofactor is a non-protein chemical compound, organic or inorganic, that is bound to a protein and is required for the protein's biological activity. These proteins are commonly enzymes, and cofactors can be considered "helper molecules" that assist in biochemical transformations (e.g. alcohol dehydrogenase require the NAD cofactor). Cofactors can be divided into two broad groups: organic cofactors, such as flavin or coenzyme Q, and inorganic cofactors, such as the metal ions Mg^{2+} , Cu^{2+} , Mn^{2+} , or iron-sulfur

clusters. Some cofactors are also covalently bound and are called prosthetic group (e.g. heme) [16].

Assembling with other subunits (quaternary structure of proteins): Many proteins have multiple polypeptide subunits (from two to hundreds). The association of polypeptide chains can serve a variety of functions. Many multi-subunit proteins have regulatory roles; the binding of small molecules may affect the interaction between subunits, causing large changes in the protein's activity in response to small changes in the concentration of substrate or regulatory molecules (e.g. the lactose repressor involved in the regulation of the lac operon in *E. coli*). In other cases separate subunits take on separate but related functions, such as catalysis and regulation, or serve primarily structural roles. A multi-subunit protein is also referred to as a multimer. A multimer with just a few subunits is called an oligomer. If a multimer has non-identical subunits, the overall structure of the protein can be asymmetric and quite complicated. However, most multimers have identical subunits or repeating groups of non-identical subunits, usually in symmetric arrangement [17].

In the vaccine field it becomes fundamental to understand if a recombinant selected vaccine candidate will present epitopes similar to the native antigen that will be recognized by the host immune system at the moment of the infection. The growing evidences of the importance of conformational epitopes in the immune response [18] implicate that in order to develop efficacious vaccines, conformation and oligomerization state of selected antigens should be addressed. For this purpose the characterization of surface protein complexes and the evaluation of their relevance for the

immune response are essential. Epitopes arising from protein-protein interactions could not be induced by immunization with single or non correlated recombinant proteins. The works provided by the group of C.M. Ferreirós, on the characterization of neisserial PorA and PorB interactions show how the characterization of intact complexes is a crucial step in the search of potential vaccine candidates [19], [20] and [21]. Another example is the complex responsible for the endocytic entry of HCMV, composed by five proteins (gH, gL, UL128, UL130 and UL131). The work of McVoy and Gerna demonstrated that the majority of the neutralizing activity of convalescent human sera from HCMV-seropositive individuals targets the endocytic complex while sera from recipients of the gB vaccine (containing only the recombinant protein gB) showed an epithelial neutralizing titers that were, on average, 15-fold lower, indicating that the gH/gL/UL128/UL130/UL131 complex is important during the infections and can be considered as a potential vaccine candidate [22].

Covalent modifications

The covalent modifications or post-translational modifications (PTMs) are chemical modifications of a protein that occur after its translation. The post-translational modifications of amino acids extend the range of functions of the protein by attaching it to other biochemical functional groups (such as acetate, phosphate, various lipids and carbohydrates), changing the chemical nature of an amino acid, or making structural changes (e.g. formation of disulfide bridges or isopeptide bonds). The role of covalent PTMs in bacteria just start to become of interest; their functions include stability, protection

from proteases and signal transduction. Potential modifications on bacterial proteins include proteolytic processing, phosphorylations, glycosylations, methylations and lipidations. Other modifications such as sulfations, hydroxylations and myristoylations do occur on proteins, but at this stage appear specific to eukaryotes [23].

Proteolytic processing: Proteolytic processing is a major form of PTM which occurs when a protease cleaves one or more bonds in a target protein to modify its activity. This processing may lead to activation or inhibition of the protein's activity. Many cellular processes are triggered by proteolytic cleavage. The processing protease may remove a peptide segment from the C- or N-terminal extremity of the target protein (e.g. the activation of the trypsinogen by the enterokinase), but it may also cleave internal bonds in the protein that lead to major changes in the structure and function of the protein. An example of this last mechanism is the chemokine protease SpyCEP from *Streptococcus pyogenes* [24]. This protein is autocatalytically cleaved into 2 fragments and the N-terminal prodomain remains structurally associated to the C-terminal portion and participate in the formation of the active site by providing one of the 3 residues of the catalytic triad. Another example of proteolytic processing of proteins is secretory proteins or proteins targeted specific cellular compartments (like the periplasm), which have their signal peptide removed by specific signal peptidases before or just after release to the extracellular environment or specific compartments [25].

Phosphorylations: Protein phosphorylation is a post-translational modification in which a serine, a threonine or a tyrosine residue is phosphorylated by a protein kinase by the addition of a covalently bound phosphate group.

Regulation of proteins by phosphorylation is one of the most common modes of regulation of protein function, and is often termed "phosphoregulation". In almost all cases of phosphoregulation, the protein switches between a phosphorylated and an unphosphorylated form, and one of these two is an active form, while the other one is inactive. The first clearly established example of phosphorylation of a bacterial protein was on the isocitrate dehydrogenase [26], in 1979, 25 years after the discovery of protein phosphorylation in eukaryotes. This enzyme was reported to be phosphorylated on a serine residue. In subsequent years, numerous other bacterial proteins phosphorylated on Ser, Thr or Tyr were discovered and the corresponding protein kinases and P-protein phosphatases were identified [27]. These protein modifications regulate all kinds of physiological processes. Ser/Thr/Tyr phosphorylation in bacteria therefore seems to play a similar important role as in eukaryotes. In addition to phosphate group, others related groups can be found attached to bacterial proteins. As example the PilE protein subunit of the *Neisseria gonorrhoeae* type IV pilus (Tfp) colonization factor can be uniquely modified by the covalent addition of different phospho-forms (i.e. phosphoethanolamine, phosphocholine and phosphoglycerol) to specific serine residues; has been proposed that these phospho-form modifications may influence pilus structure, function, and interactions with both the adaptive and innate immune systems [28].

Glycosylations: Protein Glycosylation is the post-translational process by which saccharides are selectively added to specific protein residues utilizing two distinct mechanisms (N-linked or O-linked) in order to convey more structural stability or function to the native protein structure. Proteins are

glycosylated for several reasons; some glycoproteins are more stable once they have polysaccharides attached, others for cell recognition and communication, still some proteins are not able to fold properly without their accompanying glucidic chains. N-linked glycosylation is the most common form of glycosylation. It is widely employed by Eukaryotes and Archaea, but rarely in Prokaryotes. O-linked glycosylation is mainly found in Prokaryotes. The principal difference from between the N-linked and O-linked variants of protein glycosylation relies in the variety of the side-chain interactions, with type O- utilizing an oxygen based linker mechanism. Bacteria employ a wider variety of monosaccharides than Eukarya to glycosylate their proteins [29]. To date, protein glycosylation has not been related to bacterial signal transduction, but is recognized as an important pathogenicity determinant. In *Pseudomonas aeruginosa*, flagellin glycosylation plays a major role in virulence [30], and in *Streptococcus parasanguinis* and *Streptococcus gordonii* O-glycosylation of serine-rich adhesion proteins modulates the attachment to host cells [31].

Methylations: Methylation of proteins is the most common form of post-translational modifications. As observed with other post-translational modifications, protein methylation is involved in the regulation of protein-protein interactions resulting in a plethora of effects during key cellular events, including regulation of transcription, stress response, ageing and protein repair. Methylation on carboxylate side-chains covers up a negative charge and adds hydrophobicity. The carboxyl methylation mechanism is the memory component of the bacterial chemotaxis information processing system [32] that allows cells to respond to changes in attractant

concentrations rather than absolute levels (methylation and demethylation of the chemotaxis receptors allows the bacteria to measure the attractant concentrations). N-Methylation of lysines does not alter the cationic charge but increase hydrophobicity. In particular, dimethylation and trimethylation of lysine side-chains increase both hydrophobicity and steric bulk and can affect protein–protein interactions if they are in an interacting surface.

Lipidations: The covalent binding of a lipid group to a peptide chain, also known as lipidation, can affect the activity of the protein and/or its cellular localization. Most bacterial lipoproteins are anchored to the plasma membrane by either di- (in case of Gram-positive bacteria) or triacylglycerol (in case of Gram-negative bacteria) linked to the side chain of an N-terminal cysteine via the sulfur atom, this modification is required for the cleavage of the signal peptide. Finally the mature N-terminal chain is further acylated on the amino group of the modified cysteine to obtain its final form [33]. Bacterial lipoproteins have been shown to perform various roles, including nutrient uptake, signal transduction, adhesion, conjugation, and sporulation, and to participate in antibiotic resistance, transport (such as ABC transporter systems) and extracytoplasmic folding of proteins. In the case of pathogens, lipoproteins have been shown to play a direct role in virulence-associated functions, such as colonization, invasion, evasion of host defense, and immunomodulation [34].

Unfortunately, few reports evidence how covalent PTMs could influence antigen immunogenicity and host recognition. For example the Apa deglycosylated antigen is less active than native molecules in eliciting

protective immune response against BCG in animal [35]. More recently, evidence suggesting that *Pseudomonas* glycosylated pili provide O-antigen-specific protection via the mucosal and systemic routes of immunity have been reported [36]. Moreover, the work of the group of W. Zlotnick, on the characterization of the LP2086 protein from *Neisseria meningitidis* as vaccine candidate, shows that the lipidated version of the protein consistently elicited a greater immune response compared to the non-lipidated form [37].

The necessity of analytical methods, that allows studying the maturation steps of proteins in a fast and reliable way, becomes of pivotal importance in the vaccine research and the application of mass spectrometry is emerging as the method of choice in this field.

1.3 Mass spectrometry

Mass spectrometry is a highly reproducible methodology based on the determination of the molecular mass. This technique is based on the production of gas phase ions that are subjected to an electric field and can be resolved following their electro-dynamic attitude, which is dependent on their mass-to-charge ratios (m/z).

Mass spectrometers can be divided into three fundamental parts, namely the ion source, the analyzer and the detector. Samples under investigation have to be introduced into the ion source of the instrument where the sample molecules are ionized. These ions are extracted into the analyzer region of the mass spectrometer where they are separated according to their mass-to-charge ratios (m/z). The separated ions are detected and the generated signals sent to a data system where the m/z ratios are stored together with

their relative abundance for presentation in the format of an m/z spectrum. Mass spectrometers can be distinguished on the bases of their ionization system and on the type of analyzer, which is an essential component to define the accessible mass range, sensitivity and resolution. Between the different mass spectrometers available, the ones mainly used in this work are the MALDI/ToF and the ESI/Q-ToF. The ionization methods used are “soft” ionization process like Matrix Assisted Laser Desorption Ionization (MALDI) and Electrospray Ionization (ESI) in which the biological sample is not destroyed during the ionization and the transfer to the gas phase.

Matrix Assisted Laser Desorption Ionization (MALDI): MALDI is based on the bombardment of sample molecules with a laser light to induce sample ionization [38]. The sample is pre-mixed on a stainless steel plate with a highly absorbing matrix, i.e. a small aromatic molecule, that once excited is able to transfer energy and protons to the sample. Upon drying, the matrix molecules crystallize and solid sample/matrix co-crystals are finally formed. The MALDI-plate is then inserted into the ion source of the mass spectrometer under a high vacuum. A strong electrical field is applied between the plate and the extraction plate(s). A laser (generally a pulsed nitrogen laser at 337 nm) is fired onto the sample, resulting in a desorption event due to absorbance of the laser energy by the matrix molecules. Energy deposition into the matrix molecules leads to the conversion of the absorbed energy into heat. This rapid heating causes sublimation of the matrix crystals and subsequent expansion of matrix molecules and the co-crystallized analyte molecules into the gas phase. The ions are repelled from the target surface and accelerated into the mass analyzer (Figure 2).

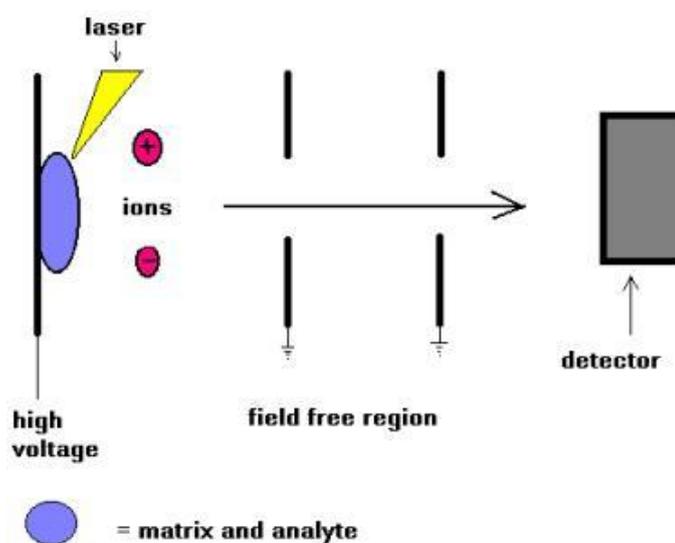


Figure 2: MALDI source

In positive ion mode the protonated molecular ions $(M+H)^+$ are usually the dominant species, although they can be accompanied by salt adducts, and a trace of the doubly charged molecular ion at approximately half the m/z value. In negative ionization mode the deprotonated molecular ions $(M-H)^-$ are usually the most abundant species, accompanied by some salt adducts.

Usually, MALDI is coupled to a time-of-flight (ToF) tube for mass analysis (Figure 3). The ToF tube is under a high vacuum (10^{-6} - 10^{-8} mbar) and is a field-free drift region. All ions enter the ToF tube at the same time and have a fixed kinetic energy, which is proportional to the applied voltage and the charge. This implicates the higher the mass of the ion, the lower its velocity and the longer it takes before the ion arrives to the detector. Based on their different velocities ions of different mass can be separated during their flight in the ToF-tube, measuring the time each ion takes to travel through the field free region. A detector amplifies and converts the signal triggered by the laser pulse and records the time-of-flight of the ions. Smaller ions fly faster

than larger ions, and their m/z ratio can be calculated from their flight time knowing the length of the tube and after calibration of the analyser using compounds with known masses. The length of the ToF is a crucial factor that affects both the resolution and the sensitivity of the mass spectrometer (the longer the tube, the higher the resolution but the lower the sensitivity).

Inherent to the MALDI ionization process is a spread of kinetic energy of ions resulting in different points in time and space of ion formation within the source. Thereby ions with the same mass obtain different kinetic energies and velocities during their extraction out of the ion source. This results in peak broadening of the ion signal at the detector. Thereby the mass resolution is limited. This peak broadening can be reduced by the use of an ion mirror (or reflectron) at the end of the linear flight tube and by delayed ion extraction out of the ion source.

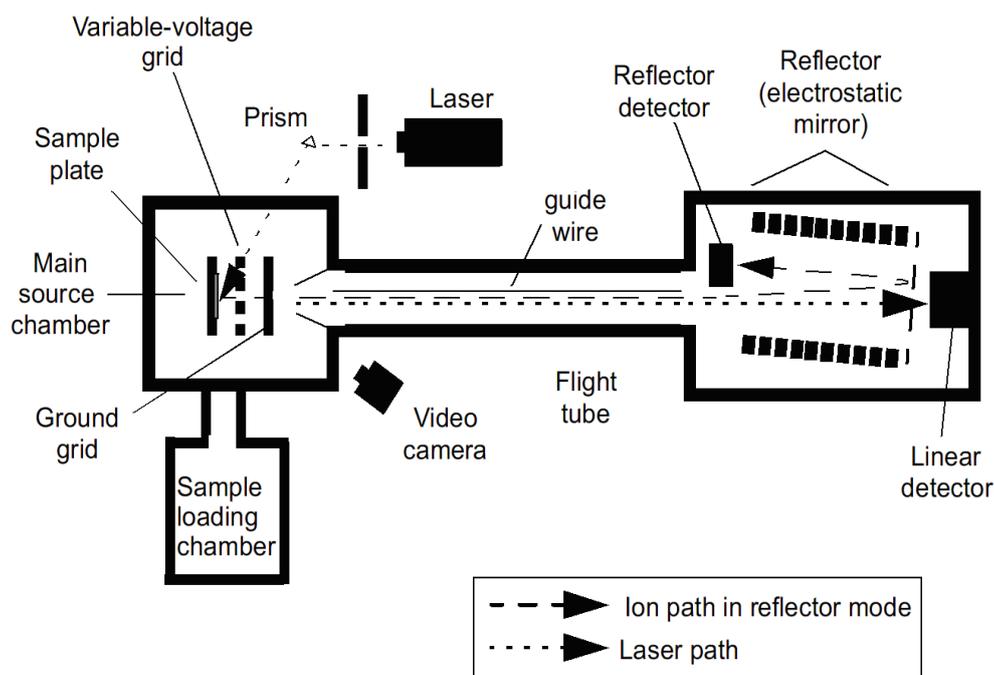


Figure 3: Scheme of a MALDI-TOF spectrometer

With delayed extraction (DE), an extraction voltage pulse is applied to the sample plate between 100 and 500 ns following the laser pulse. During this delay, ions are allowed to spread in the source and higher energy ions will move further away than lower energy ions with the same mass. The extraction voltage is now applied as a potential gradient over the ion source. This compensates for the distribution of initial kinetic energies, so that ions with identical m/z values will be grouped in space before leaving the source. The ions will be then accelerated before the entrance of the ToF to reach the same kinetic energy.

The reflectron has an applied voltage higher than that of the accelerating voltage in the ion source, resulting in ions slowing down and reversion of their flight path to the second detector. Ions with lower kinetic energies do not penetrate the reflectron as deep and thus turn around faster, catching up with ions of slightly greater kinetic energies that penetrate the reflectron deeper. Thereby, the flight times of ions with identical m/z values, but different kinetic energy values will be corrected when the ions arrive to the detector. Moreover the presence of the reflectron mirror allows to increase the travelling path of the ions in the ToF region without physically change the length of the tube; in this way is possible to achieve higher resolution with relatively small ToF devices.

Electrospray ionization: Electrospray Ionization (ESI) is one of the Atmospheric Pressure Ionization (API) techniques and is well suited for the analysis of polar molecules ranging from less than 100 Da to more than 1,000,000 Da in molecular weight [39].

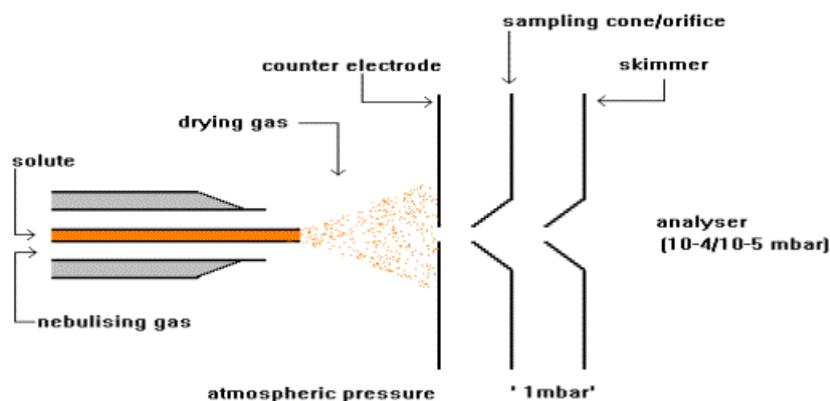


Figure 4: Electrospray source

During standard electrospray ionization samples are dissolved in a polar, volatile solvent and pumped through a narrow, stainless steel capillary. A high voltage of 2 up to 5 kV is applied to the tip extremity of the capillary situated within the ionization source of the mass spectrometer and the sample emerging from the tip is dispersed into a spray of highly charged droplets, a process that is aided by a co-axially introduced nebulising gas flowing around the outside of the capillary. This gas, usually nitrogen, helps to direct the spray emerging from the capillary tip towards the mass spectrometer (Figure 4). The charged droplets diminish in size by solvent evaporation, assisted by a warm flow of nitrogen which passes across the front of the ionization source or by the heating of the capillary, and undergo to a fission event giving birth to smaller droplets (coulombic explosions). After different fission events the electric field on the surface of the droplets become large enough to allow the desorption of sample ions (Iribarne and Thomson model or Ion Evaporation Model, IEM), some of which pass through a sampling cone or orifice into an intermediate vacuum region, and from there through a small orifice into the analyser of the mass spectrometer, which is

held under high vacuum. When the droplet contains very large molecules, like proteins for example, the molecules will not desorb, but are freed by evaporation of the solvent (Dole model or Charge Residue Model, CRM). This seems to occur when the molecular weight of the compounds exceeds 5000 to 10000 Da [40].

Electrospray is known as a “soft” ionization method as the sample is ionized by the addition or removal of a proton, with very little extra energy remaining to cause fragmentation of the sample ions. The peculiar aspect of this technique is the fact that it gives rise to multiply charged molecular-related ions such as $(M+nH)^{n+}$ in positive ionization mode and $(M-nH)^{n-}$ in negative ionization mode.

ESI source is generally coupled with quadrupole analyzers. A quadrupole mass analyser consists of four parallel rods that have fixed DC and alternating RF potentials applied to them. Ions produced in the source of the instrument are then focused and passed along the middle of the quadrupole. Their motion will depend on the electric fields so that only ions of a particular m/z will be in resonance and thus pass through to the detector. All other ions do not have a stable trajectory through the quadrupole mass analyzer and will collide with the quadrupole rods, never reaching the detector. The amplitude of the RF potential is modulated to bring ions of different m/z to be focused and thus transmitted to the detector to build up a mass spectrum (Figure 5).

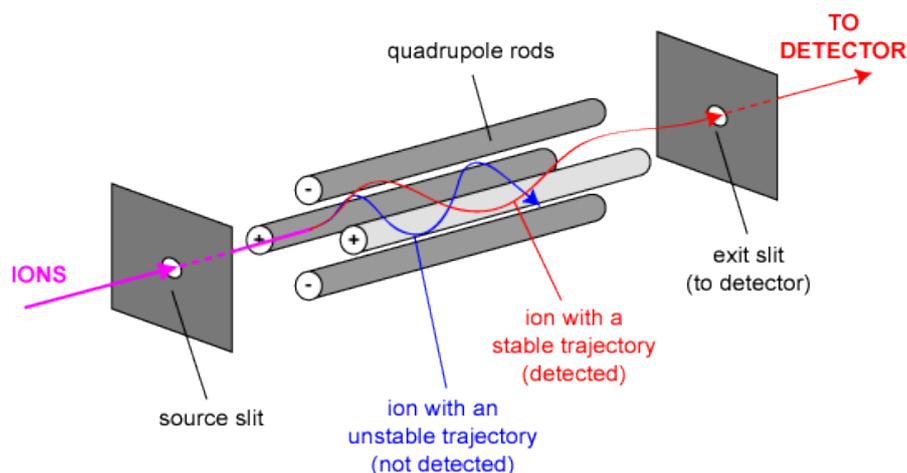


Figure 5: The quadrupole analyzer

Tandem mass spectrometry (MS/MS): Tandem mass spectrometry is used to obtain structural information about a compound by selecting and fragmenting specific ions inside the mass spectrometer and identifying the generated fragment ions. Tandem mass spectrometry also enables specific compounds to be detected in complex mixtures on account of their specific and characteristic fragmentation patterns. The fragmentation behaviour of a peptide is well understood and follows specific rules [41]; applying a low energy (i.e. below 100 eV) there are three different types of bonds that can fragment along the amino acid backbone: the NH-CH, CH-CO, and CO-NH bonds. Each bond breakage gives rise to two species, a charged one, that is monitored by the mass spectrometer and a neutral one; the charge can be retained on one of the two fragments depending on the chemistry and the relative proton affinity of the two species. Hence there are six possible fragment ions for each amino acid residue and these are labelled as in the Figure 6, with the a, b, and c ions having the charge retained on the N-terminal fragment, and the x, y and z ions having the charge on the C-

terminal fragment. The most common cleavage sites are at the CO-NH bonds which give rise to the b and/or the y ions.

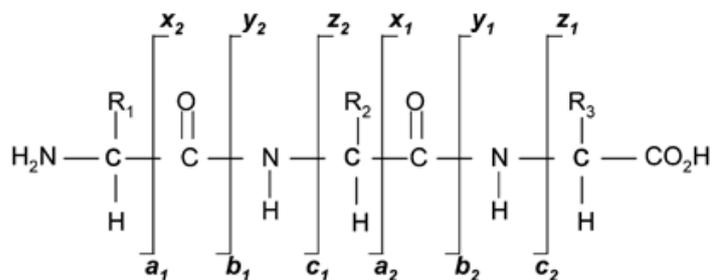


Figure 6: Scheme of peptide fragmentation

Generally in a tandem mass spectrometer the two analyzers are separated by a collision cell filled with an inert gas (e.g. argon, xenon). When performing a classical MS analysis the collisions of the ions with the gas decrease their kinetic energy and stabilize their trajectories (cooling effect). In MS/MS experiments the ions are accelerated before the entrance of the collision cell where they collide with the gas and bring out their fragmentation. An example of such an instrument is the Quadrupole Time-of-Flight (Q-ToF), where the two resolving mass analyzers are separated by a fragmentation cell (Figure 7) [42].

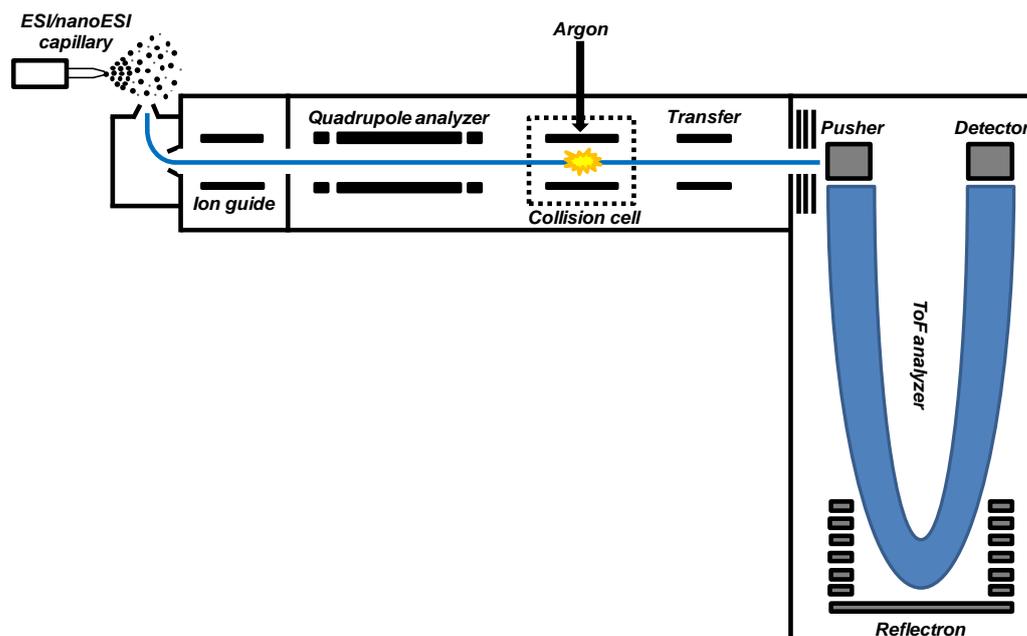


Figure 7: Scheme of an ESI/Q-ToF instrument

Native mass spectrometry: Cellular processes are not typically performed by a single protein but by higher order protein oligomers. Therefore rather than focusing on individual proteins, it became more and more important to develop structural biology approaches capable of investigating intact functional protein complexes [40]. Native MS is a powerful technology that allows the topology of intact protein complexes to be investigated with high dynamic ranges [43]. Compared to the common approaches used for structural research such as X-ray crystallography, nuclear magnetic resonance spectroscopy or cryo-electron microscopy, native MS only requires low sample concentrations which permit the analysis of protein complexes close to their physiological conditions. In addition, this technique can be applied to study the interactions between proteins and small ligands and to measure the binding affinity of the complexes in the gas phase. For native MS the sample has to be diluted in an aqueous solution containing a

volatile buffer compatible with MS (i.e., ammonium acetate). A nanoESI source is generally used. NanoESI is a development of ESI for spraying very low amounts of very low concentration samples (nmol/mL). The technique has an increased tolerance to high aqueous solvents and salt contamination. Spectra can be obtained from pg of material with very little clean up being required. This increased performance is the result of lowering the inner diameter of the spray needle and reducing potentials normally used in ESI. In this type of source the analyte is sprayed from a tip with an opening of less than 5 μm , smaller than the one used in standard ESI (130 μm), thus producing smaller droplets which require a low amount of sample and “softer” conditions, compared to ES, to generate ions in the gas phase (e.g. temperature and pressure) (Figure 8). Standard nanospray uses disposable tips, but has problems with signal reproducibility between tips.

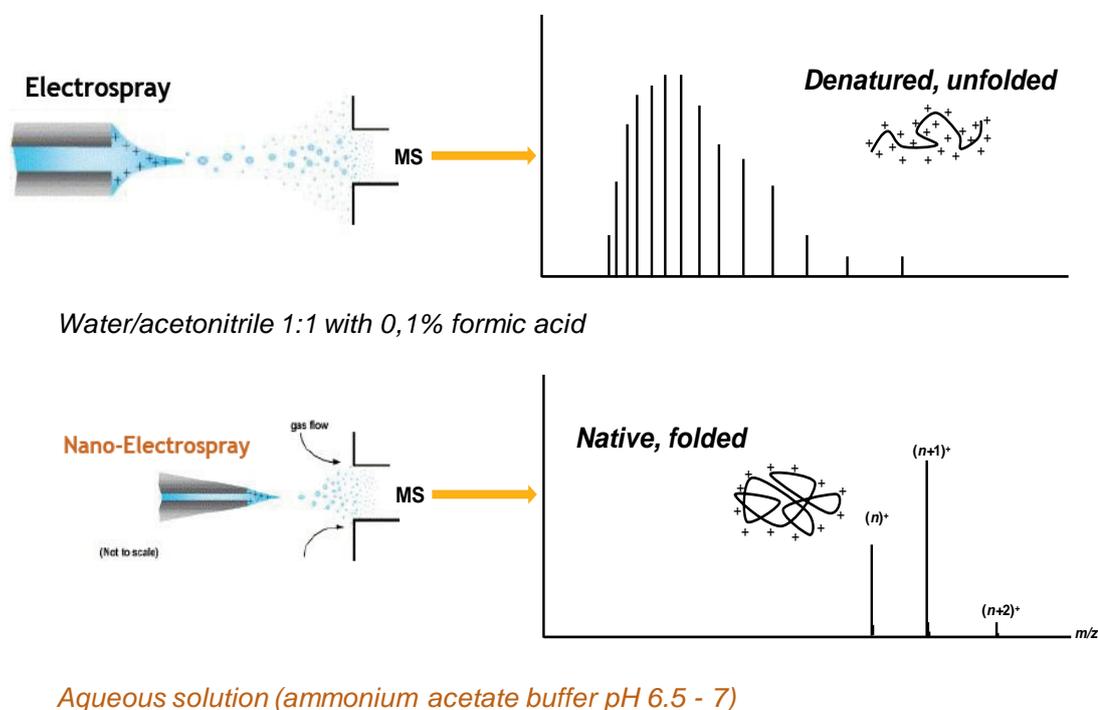


Figure 8: Comparison between denatured (on the top) e native (on the bottom) MS

Moreover the mass of intact proteins or protein complex is relatively high and the surface area exposed that can be charged is rather low, due to the retainment of the folded structure (Figure 8), hence the main type of analyzer used in native MS is the ToF, which theoretically does not have a limit in mass range. To obtain additional information from a native MS analysis, like the stoichiometry of hetero-oligomeric proteins or protein-ligand complexes, the spatial arrangement of the individual building blocks, and the stability of the complex, is necessary to use a hybrid mass spectrometer that allows the fragmentation of the analyzed sample. The Q-ToF mass spectrometer is well suited for this type of analysis coupling the capability to select ions of the quadrupole analyzer with the unrestricted mass range of the ToF. The main drawback of this type of instrument is the dynamic range of the quadrupole; up to now, the reduction of the RF frequency (around 300 kHz) allows to design quadrupole with a mass limit of 32 kDa, but it is not possible, for the moment, to use frequencies below 300 kHz for stability reasons [42].

1.4 Aim of the thesis

The study of the maturation process that occurs to a protein is of pivotal importance for the understanding of its function. This is true also in the vaccine field but in this case is also important to evaluate if inappropriate protein conformation and maturation play roles in the impairment of the functional immunogenicity of protein vaccines.

Mass spectrometry (MS) is the method of choice for the study of the maturation process since each modification that occurs during the maturation will lead to a change in the mass of the entire protein.

Therefore the aim of my thesis is the development of mass spectrometry-based approaches to study the maturation of proteins and the application of these methods to proteic vaccine candidates.

The thesis is divided in two main parts. In the first part, I focused my attention on the study of the maturation of different vaccine candidates using native mass spectrometry. The analyses in this case have been performed using recombinant proteins produced in *E. coli*. In the second part I applied different MS strategies for the identification of unknown PTMs on pathogenic bacteria surface proteins since modified surface proteins are now considered for vaccine candidate selection.

2 Results and discussion

2.1 Investigation on the maturation of different vaccine candidates using native mass spectrometry

Introduction

Many proteins acquire their biological active conformations only when they become part of higher order oligomers or interact with cofactors. These interactions are generally mediated by non-covalent bonds (i.e. hydrogen bonds, electrostatic and hydrophobic interactions) that are not retained using classical MS approaches (i.e. denaturing conditions). Native MS is a technique that allows the structural investigation of protein complexes without “destroying” non-covalent interactions, giving in the same time informations about covalent modifications [43]. Even though native MS is not as resolute as crystallography or NMR the sensitivity, speed, selectivity and theoretical unlimited mass range and accuracy of this approach allow us to access to structural information of huge protein complexes (up to 2.2 MDa) [44] using nearly physiological conditions and low quantities of material. Within this approach nano-electrospray ionization is the most popular technique to ionize protein/protein and protein/ligand complexes. It is a soft ionization method that preserves non-covalent bonds and thus allows the study of the oligomerization state and cofactor binding of the proteins of interest.

Assignment of the oligomerization states of proteic vaccine candidates

Oligomeric proteins abound in nature. They are composed of multiple subunits, which may be identical (homo-oligomeric proteins) or different (hetero-oligomeric proteins). It has been calculated that the average oligomeric state of cellular proteins is tetrameric, and a recent survey suggests that 35% or more of the proteins in a cell are oligomeric. Nevertheless, the proportion of oligomeric protein structures deposited in the Protein Data Bank (<http://www.pdb.org/pdb/home/home.do>) is significantly lower [17]. The study of protein oligomerization may be critical to understand the protein's physiological functions and is therefore necessary to fully characterize proteic vaccine candidates. Moreover the native MS analysis can give also important information about the presence of covalent modifications. We applied the native MS to seven recombinant proteic vaccine candidates or proteins of interest for vaccine development, belonging to three pathogenic microorganisms (Table 1), in order to study the oligomerization state and have preliminary informations about the presence of other modifications.

Protein	Function	Microorganism
GNA1030	Unknown	<i>Neisseria meningitidis</i>
GNA2091	Hemolysin (putative)	<i>Neisseria meningitidis</i>
NadR	Repressor of NadA	<i>Neisseria meningitidis</i>
NadA	Adhesin	<i>Neisseria meningitidis</i>
fHbp	Factor H binding	<i>Neisseria meningitidis</i>
PSL1	Unknown	<i>Staphylococcus aureus</i>
SAL1486	Backbone pilus subunit	<i>Streptococcus agalactiae</i>

Table 1: List of the proteins analyzed by native MS

All the proteins were expressed in *E. coli* and purified avoiding denaturing steps. The results of the native MS analysis are summarized in Table 2. Among the seven proteins tested, three were found monomeric (GNA2091, fHbp and SAL1486), two were dimeric (GNA1030 and NadR, as suggested in [45]) and one was trimeric (NadA, as suggested in [46]). Interestingly, two forms of PSL1 were identified: (i) a covalent dimer (through an inter-protein disulfide bridge), and (ii) a monomer with a mass increase of 765.6 Da compared to the theoretical mass of the monomer. These two species were also present in the mass spectrum acquired in denaturing conditions. In addition, as observed with PSL1, the measured molecular weight of GNA1030 (only in native MS) and SAL1486 (also in denatured MS) differs from the expected one (by + 1457 Da and – 51 Da, respectively). Finally some degradation products were also observed in the NadA sample (at the N-terminal part) but the degradation does not seem to affect its oligomerization state (trimeric).

Protein	Oligomerization	Expected MW (Da)	Observed MW (Da)	Δ Mass (Da)
GNA1030	Dimeric	37186.4	38643.2	+1457
GNA2091	Monomeric	20983.6	20983.9	/
NadR	Dimeric	33728.6	33727.8	/
NadA	Trimeric	103676.4	103676.6	/
fHbp	Monomeric	27523.8	27523.8	/
PSL1	Dimeric (covalent)	54339.2	54340.2	/
	Monomeric	27170.6	27936.2	+765.6
SAL1486	Monomeric	67426.1	67375.2	-51

Table 2: Oligomerization state and mass differences observed for the analyzed proteins

Further analyses have been conducted to characterize the modifications found in the PSL1, GNA1030 and SAL1486 samples.

Characterization of the mass increase found on PSL1

PSL1 (Putative Staphylococcal Lipoprotein 1) is a *Staphylococcus aureus* lipoprotein with unknown function that is able to confer protection in mice immunization models [Bagnoli F., personal communication]. *Staphylococcus aureus* is a Gram-positive opportunistic pathogen that can cause a spectrum of infections in humans and animals that differ in severity. Some relatively minor skin infections, such as folliculitis, impetigo and cellulitis, can progress to life threatening diseases like sepsis, endocarditis, osteomyelitis and pneumonia [47]. Bacterial lipoproteins have been shown to perform various roles, including nutrient uptake, signal transduction, adhesion, conjugation, and sporulation, and to participate in antibiotic resistance, transport (such as ABC transporter systems) and extracytoplasmic folding of proteins. In the

case of pathogens, lipoproteins have been shown to play a direct role in virulence-associated functions, such as colonization, invasion, evasion of host defense, and immunomodulation, and have been described as promising vaccine candidates both in Gram negative [48] and gram positive bacteria [49]. Lipoproteins are initially translated as prelipoproteins, which possess a N-terminal signal peptide of around 20 amino acids with typical characteristic features of the signal peptides of secreted proteins. Lipoproteins are either di- (in case of Gram-positive bacteria) or triacylated (in case of Gram-negative bacteria) on a highly conserved cysteine present at the C region of the signal peptide, in a region referred to as lipobox (Leu/Val/Ile-Ala/Ser/Thr/Val/Ile-Gly/Ala/Ser-Cys). By the consecutive action of the three enzymes pre-lipoprotein diacyl glyceryl transferase (Lgt), lipoprotein signal peptidase (LspA) and apolipoprotein N-acyltransferase (Lnt), lipoproteins are post-translationally modified after translocation over the cytoplasmic membrane. Lgt attaches a diacylglyceryl residue to the universally conserved cysteine in the lipobox by thioether-linkage. Then LspA removes the lipoprotein signal peptide and leaves the cysteine of the lipobox as the new amino-terminal residue. Finally, Lnt attaches a third acyl chain to the amino group of the modified cysteine [29]. The recombinant PSL1 used for this study has been expressed with a truncated N-terminal lipobox composed by only two amino-acids (GC) in order to avoid the binding of the diacylglycerol moiety and thus the anchorage of PSL1 to the membrane. The entire mass analysis of the protein in denaturing condition (Figure 9, lower spectrum) revealed the presence of three species with a molecular weight of 27170.3 Da, 54340.2 Da and 27936.2 Da, respectively. The first two species

showed a mass in agreement with the monomeric and dimeric (covalent dimer through a disulfide bridge) form of the PSL1 protein respectively. The third specie showed a mass increase of 765.6 Da compared with the monomer, suggesting a covalent modification. In the entire mass spectrum performed in reducing conditions (Figure 9, upper spectrum) the only specie present is the monomer (27170.3 Da); these data confirms that the specie with a MW of 54340.2 Da in the unreduced spectrum is a covalent dimer through a disulfide bridge and reveal that the modification is also bound to the cysteine. Moreover, the presence of the modification only on the monomer further confirms this finding (in the dimer the cysteine residues are involved in the disulfide bridge formation, thus are not available for the modification).

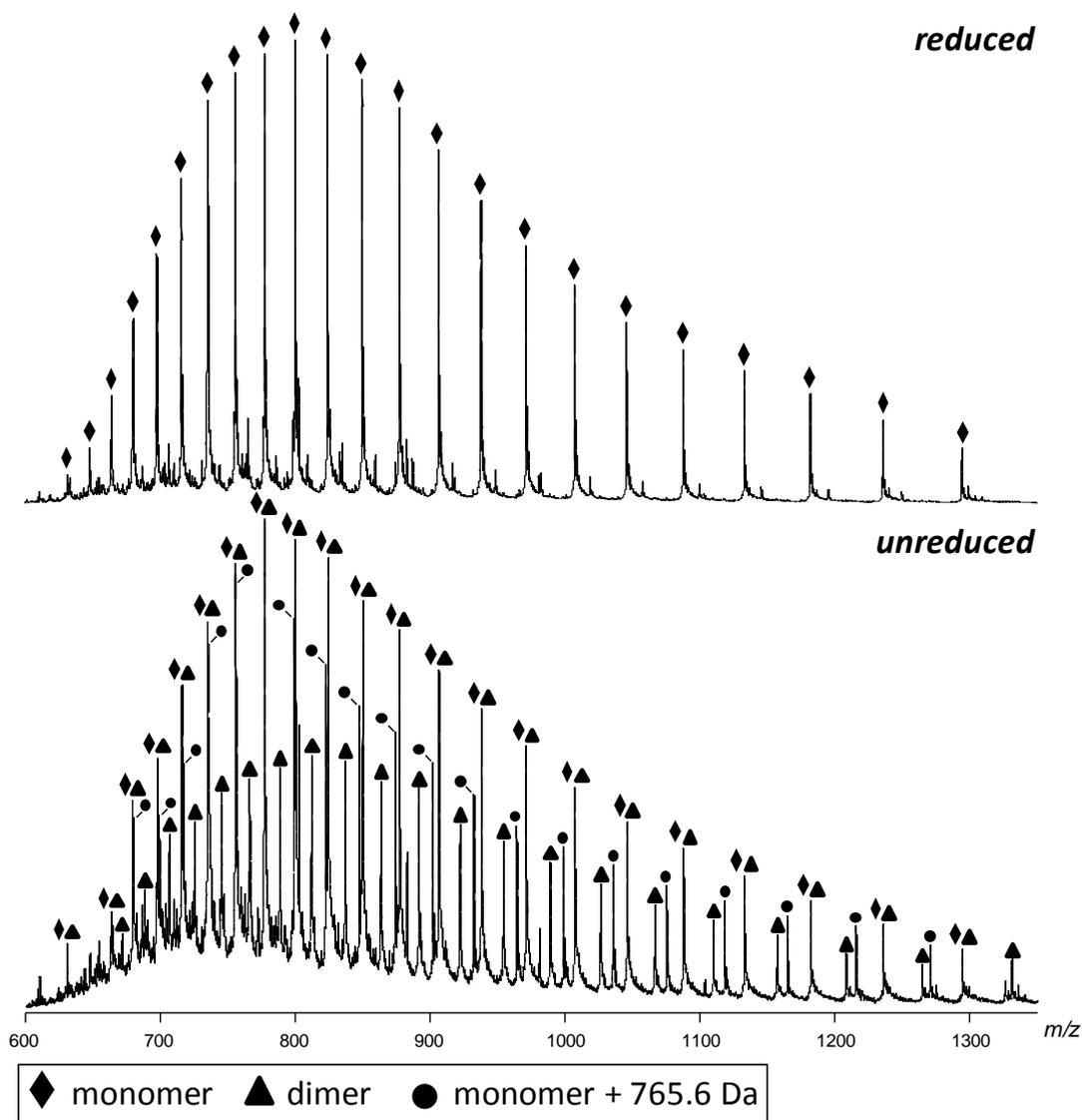


Figure 9: Entire mass spectrum of PSL1 in denaturing conditions with and without reducing agent (upper and lower panel, respectively)

In order to characterize the modification, the protein, with and without reducing agent, was analyzed by MALDI-ToF MS in negative ionization mode. As reported in Figure 10, a molecular ion at 766.6 m/z was only observed under reducing conditions. This peak can be assigned as the negative ion produced by the modification ($[M-H]^-$); therefore the mass of the molecule bound to the protein is 767.6 Da.

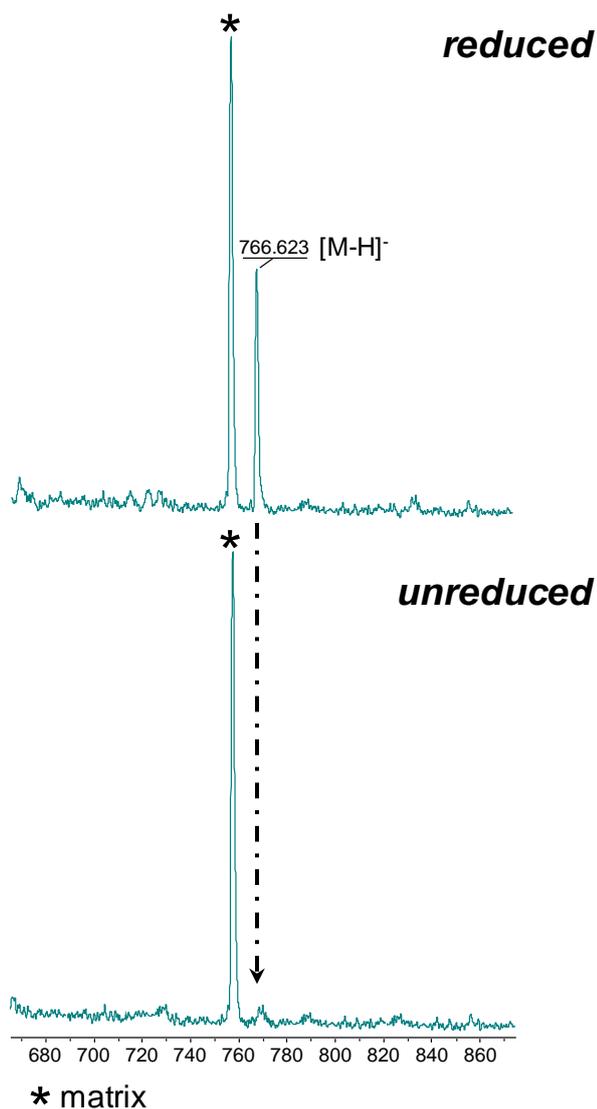


Figure 10: Zoom in of MALDI-ToF spectrum of the PSL1 protein with and without reducing agent (upper and lower panel respectively). The peak at m/z 766.7 can be assigned as the negative ion of the modification $[M-H]^-$

It has already been reported in the literature that molecules with free thiols are able to link cysteine residues present in proteins through disulfide bonds (S-thiolation); this modification is generally observed in recombinant proteins secreted from *E. coli* cells [50]. S-thiolation in this case can be a response to environmental stress experienced by the cells or to the (patho)-physiological burden brought on by the expressed proteins. Various thiol modifiers have

been identified on recombinant proteins secreted from *E. coli* by MS including glutathione (mass increase of 305 Da), gluconoylated glutathione (mass increase of 483 Da), 4-phosphopantetheine (mass increase of 356 Da), dephosphorylated coenzyme A (mass increase of 685 Da) and coenzyme A (mass increase of 765 Da). Between the known thiol modifiers the coenzyme A presents a MW of 767.5 Da, which is in perfect agreement with the mass of the modification found on the PSL1 protein (767.6 Da). Based on these observations, it is possible to identify the molecule bound through a disulfide bridge to the PSL1 protein as the coenzyme A (CoA).

The only cysteine present in the protein is the one of the lipobox that in nature is covalently attached to a diacylglycerol moiety, thus both the formation of the covalent dimer and the binding of the CoA are not physiological modifications.

The presence of the CoA covalently attached on the protein may be able to influence the immunogenicity of the protein in a positive mode since the structure of this molecule is similar to some Toll-like receptors (TLRs) agonists (e.g. Imiquimod and Resiquimod) (Figure 11) [51].

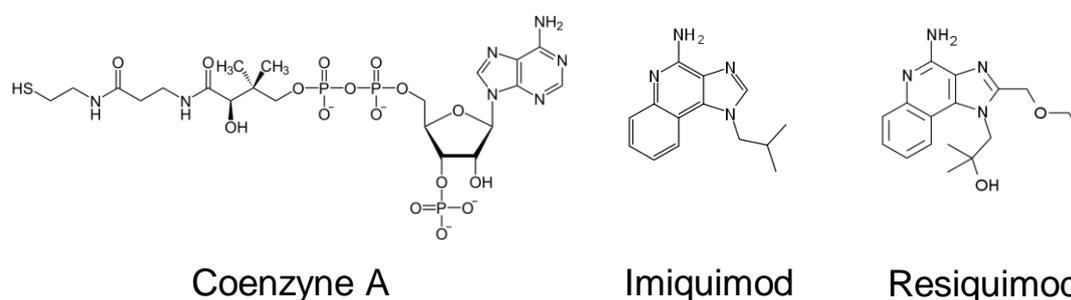


Figure 11: Structure of the coenzyme A and of two similar TLR agonists

TLRs are pattern recognition receptors that can recognize pathogens via pathogen-specific molecular patterns (PAMPs). TLRs play a crucial role in

both innate and adaptive immunity. Cells of the innate immune system can recognize invading pathogens as non-self through the TLRs leading to activation, maturation and induction of pro-inflammatory cytokines and other anti-microbial compounds. Immature dendritic cells (DCs) resident in peripheral tissues recognize these invading pathogens via numerous TLRs present on them. This leads to the activation, maturation and trafficking of the DCs to local lymph nodes and presentation of microbial antigens to naïve T cells leading to the induction of adaptive immunity against the invading pathogen. Furthermore, DCs can also regulate the T cell differentiation (Th1 versus Th2) based on the pro-inflammatory cytokines that are produced by them which in turn may depend on the TLR:PAMP interaction. Since these molecules are able to boost the cell mediated immune responses and induce the formation of memory B cell, they can be used as vaccine adjuvants [52]. The ability of the modified PSL1, attached to the CoA, to activate the TLRs has been tested measuring the cytokines production of adherent 293T cells, stably expressing the indicated TLR, exposed both to the modified and unmodified protein. No differences has been found between the two samples (data not shown), indicating that the CoA does not possess an adjuvant activity.

The effect of these modifications on the immunogenicity of the protein *in vivo* has not been tested, however to avoid any risk a mutated form of the protein, with the deletion of the cysteine residue, has been generated and named PSL1 Δ cys. The mutated protein is still able to confer protection in mice immunization models and, after native MS analysis, showed a monomeric

oligomerization state and an observed MW in agreement with the expected one (data not shown).

Characterization of the mass increase found on GNA1030

GNA1030 is a periplasmic protein with unknown function present as an accessory antigen fused to the main antigen NHBA (GNA2132) in the 4CMenB multicomponent vaccine against MenB developed by Novartis [8].

N. meningitidis is an encapsulated Gram-negative bacterium that colonizes the upper respiratory tract of about 5-10% of humans. With a frequency of one to three cases per 100,000 of the population, the bacterium enters the bloodstream where it multiplies to high density and causes a form of sepsis characterized by the dramatic disruption of the endothelium and microvasculature. From the bloodstream the bacterium can cross the blood–brain barrier and cause meningitis. The invasive infection is very dramatic, affecting mostly infants, children, and adolescents who do not have bactericidal antibodies raised against the infecting strain. Based on the chemical composition of the polysaccharide capsule, *N. meningitidis* strains can be classified into 13 different serogroups. Strains representative of five serogroups (A, B, C, Y, and W135) cause nearly all diseases in humans [52], [53]. Tetravalent vaccines composed of purified capsular polysaccharides of serogroups A, C, Y, and W135 have been available for three decades for use in adults, while conjugate vaccines, effective in all age groups, were developed a decade ago [54]. MenB differs from the A, C, Y, and W135 serogroups because it is decorated by a capsular polysaccharide identical to the polysialic acid [$\alpha(2-8)$ N-acetylneuraminic acid] present in many human

glycoproteins. This mimicry rules out the use of polysaccharide–protein conjugate vaccines to combat MenB disease; therefore, there is currently no universal vaccine available against this bacterium. An alternative approach to vaccine development is based on surface-exposed proteins contained in outer membrane vesicles (OMVs). These vaccines have been shown both to elicit serum bactericidal antibody responses and to protect against developing meningococcal disease in clinical trials. However, the limitation of OMV vaccines is that the major protein antigens (i.e. PorA and PorB) display sequence and antigenic variability among the different MenB strains and, although they induce protective antibodies against the homologous strains, they fail to induce protection against heterologous strains [55]. To develop a universal vaccine against MenB, the genome of the bacterium has been sequenced [56] in order to discover novel putative antigens using the reverse vaccinology [7]. The antigens selected were prioritized based on their ability to induce broad protection as inferred by bactericidal assay (BCA) or observed in passive protection in the infant rat or mouse protection assays. The top antigens that met the prioritization criteria were: *Neisseria* heparin binding antigen (NHBA), factor H binding protein (fHbp), *Neisseria* adhesin A (NadA), GNA1030 and GNA2091 [8]. Among the selected antigens, GNA1030 is the only one with an unknown function. It is a member of the Ycel-like family, a group of putative lipid binding proteins, showing 34% identity and 66% similarity with *E. coli* Ycel. Only few structures of proteins belonging to this family are present in the Protein Data Bank [57], [58], [59]. These proteins share a common domain with an eight stranded beta-barrel

fold containing a lipid molecule (with the exception of the Ycel-like protein from *C. jejuni*) (Figure 12).

Protein	Organism	PDB code	Ligand
Ycel	<i>H. pylori</i>	3HPE	Erucamide
Ycel-like	<i>E. coli</i>	1Y0G	8PP
PI-binding protein	<i>T. thermophilus</i>	1WUB	Polyisoprenoid
X158 (Ycel-like)	<i>S. degradans</i>	2X32	Octaprenylpyrophosphate
X158 (Ycel-like)	<i>S. degradans</i>	2X34	Ubiquinone-8
Ycel-like	<i>C. jejuni</i>	2FGS	No ligand observed

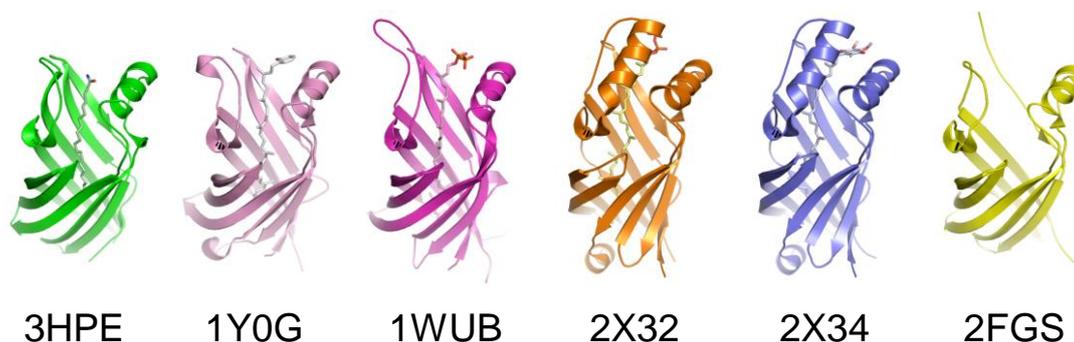


Figure 12: List of the proteins belonging to Ycel-like family present in the PDB with the relative structures

Therefore the mass increase found only during the native MS analysis of the GNA1030 protein (+ 1457 Da) could be due to the presence of a lipidic ligand non-covalently attached to the beta-barrel domain of GNA1030 (Figure 13, left inset). To demonstrate this hypothesis, an in source fragmentation experiment was performed during the native MS analysis of the protein. During electrospray ionization, ions entering through the sample cone are accelerated towards the extraction cone; the region between the sampling cone and the extraction cone is under a vacuum of 1-2 mbars (> 5 mbars during native MS experiments) (Figure 4). This pressure is relatively high owing to the presence of solvent vapor and nitrogen gas. In order to induce in-source fragmentation, the sampling cone voltage is increased, causing

ions entering the ion block to accelerate more quickly through the region (Figure 4); the resulting collisions can induce dissociation of the non-covalent complexes. After in source fragmentation, five different species are present with MW of 38642.6 Da, 37914.6 Da, 37186.3 Da, 19321.4 Da and 18593.3 Da respectively (Figure 12). These species can be assigned as the dimer plus 1456 Da, the dimer plus 728 Da, the dimer alone, the monomer plus 728 Da and the monomer alone. These data confirm the presence of a small ligand, with a MW of about 728 Da, non-covalently bound to the GNA1030 protein with a 1:1 stoichiometry (Figure 13).

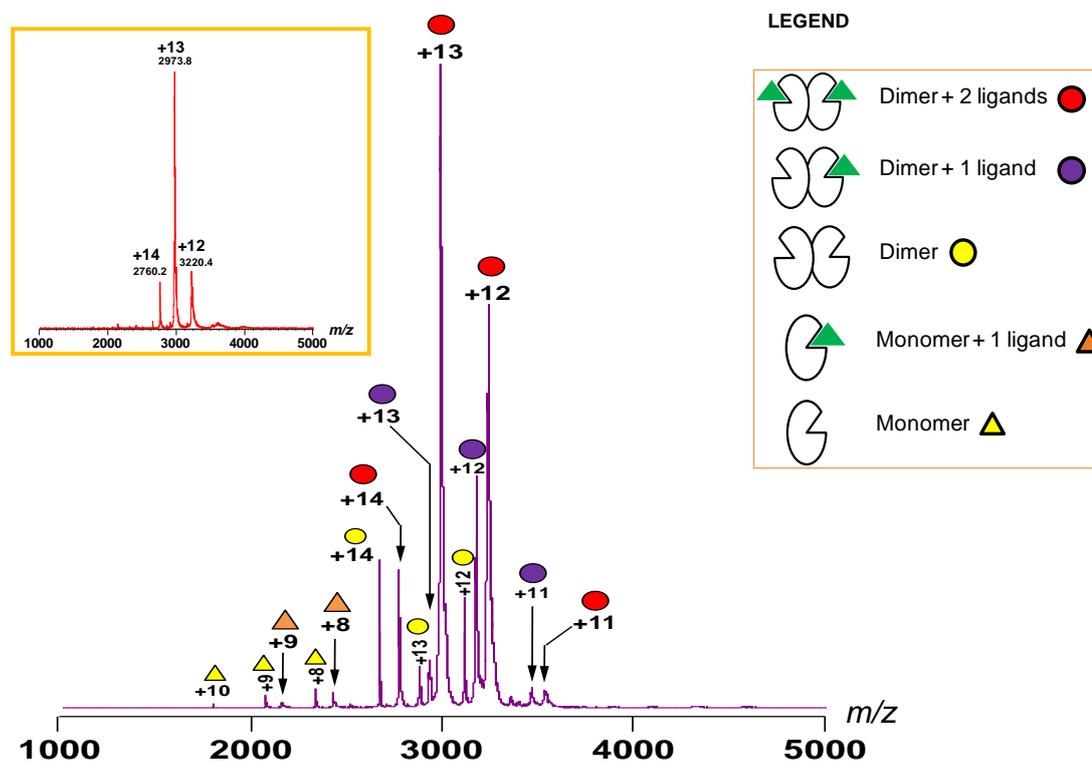


Figure 13: In source fragmentation of the GNA1030 protein (the native MS spectrum obtained before in source fragmentation is presented in the left inset)

In order to fully characterize the ligand, the molecular ion released after in source fragmentation (m/z of 728.6 in positive mode (Figure 14) and m/z 726.6 in negative mode), was selected and fragmented, in both positive and

negative mode, in the collision cell of the mass spectrometer (MS/MS experiment).

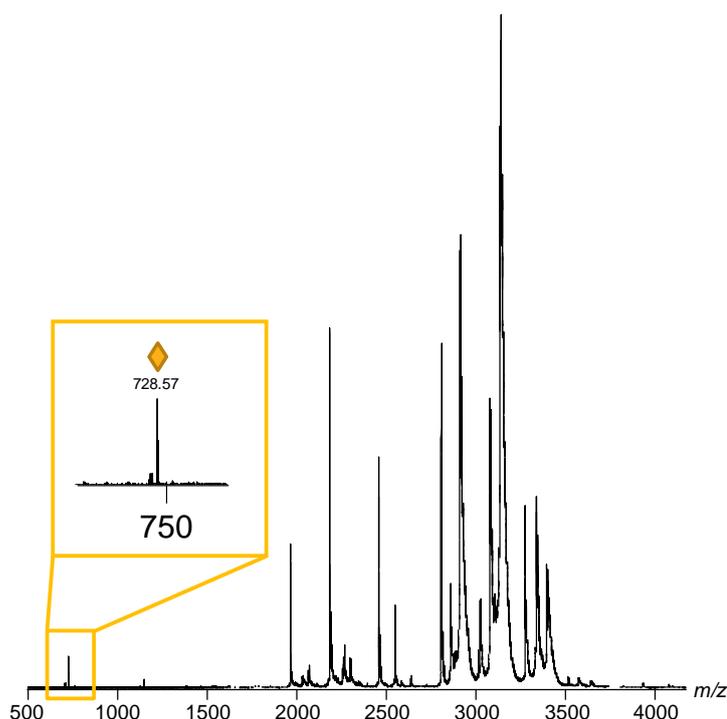


Figure 14: Selection of the peak corresponding to the ligand after in source fragmentation of the protein in positive mode (a zoom in of the spectra showing the ion released is reported in the inset).

In the negative MS/MS spectrum, the presence of two ions with -15 Da and -30 Da compared to the parental ion, implies the neutral loss of one and two methyl group $\{([M-H]^-)-CH_3$ and $([M-H]^-)-2CH_3\}$ (Figure 14, A). Moreover, the presence of a repetition of seven ions with a delta mass of 68.1 Da indicates the presence of at least six isoprenyl groups in the ligand (Figure 15, A). The positive MS/MS spectrum is characterized by the presence of two ions with m/z of 197.1 and 235.2 typical of the fragmentation of ubiquinones [60], known as tropylium and pyrylium ions respectively (Figure 15, B). Based on these data, it was possible to identify the ligand of the GNA1030 protein as the ubiquinone 8, also known as coenzyme Q8 (Figure 15, C).

The ubiquinone 8 has also been identified as the ligand of GNA1030 purified from the cytoplasm of *E. coli* (recombinant protein expressed without the signal sequence) as well as from the periplasm of MenB. The presence of the same molecule associated to the GNA1030 protein purified from different organisms and different cellular compartments suggests that the protein binds the ligand specifically.

E. coli YceI, one of the closest homologs of GNA1030, is a periplasmic protein that is induced by high pH [61] and high salt concentration [62]. The upstream gene adjacent to *yceI* encodes for the putative cytochrome b561, which is a member of PF01292 or the cytochrome b561 family in the Pfam database. The genes of this family and the GNA1030 homologs often exist as neighbors, such as in *Bacillus subtilis*, *Vibrio cholerae*, *Pseudomonas aeruginosa*, *Yersinia pestis*, *Mesorhizobium loti*, and *Xylella fastidiosa*. One of the open reading frames of *Caulobacter crescentus* encodes a fusion of the two proteins. Thus, it is likely that members of the YceI like family of proteins play a role in the electron transport system by binding polyisoprenoid molecules [58]. In the case of GNA1030, the upstream gene does not encode for a cytochrome but for an enzyme involved in leucine biosynthesis. The GNA1030 protein can also act as a carrier of the ubiquinone 8 from the cytoplasm to the periplasm, or can be involved in the stress response of the bacteria (the coenzyme Q8 possesses strong antioxidant capabilities) [60], [63]. In conclusion data from native mass spectrometry allowed us to assign a putative function to an uncharacterized protein vaccine candidate.

Characterization of the mass decrease found on SAL1486

SAL1486 is the backbone pilus subunit of *Streptococcus agalactiae* Pilus Island 2a. *S. agalactiae*, also known as Group B Streptococcus (GBS), is a Gram-positive pathogen that causes life-threatening pneumonia, sepsis, and meningitis in newborn and young infants [64]. This microorganism is classified into 10 capsular polysaccharide serotypes, each antigenically and structurally unique. Although major efforts have been made in the development of multivalent capsular conjugate vaccines, there is currently no vaccine against GBS [65]. To overcome serotype-specific immunity and the increasing number of nontypeable isolates, vaccines based on conserved protective proteins are highly desirable [66]. Recently, three pilus types have been discovered in GBS as important virulence factors as well as potential vaccine candidates [67]. The genes involved in pilus assembly are clustered in characteristic genomic loci, named Pilus Island (PI), specifically PI-1, PI-2a and PI-2b, each encoding three proteins containing a LPXTG motif, representing the structural components of the pilus, and two sortase enzymes, catalyzing protein polymerization. Each of the three pilus types carries two protective antigens [67] and among these the backbone protein of pilus type 2a (SAL1486 or BP-2a), is able to significantly mediate opsonophagocytic activity and to confer protection in mice only against strains expressing the homologous allele. The crystal structure of this protein (Figure 16, ~190 aa from the N terminus (D1 domain) were absent in the crystal, D1 is likely to be cleaved off during crystallization [68]), reveals three IgG-like fold domains (namely D2 (residues 190–332), D3 (residues 333–455), and D4

(residues 456–641), each one characterized by a putative stabilizing isopeptide bond.

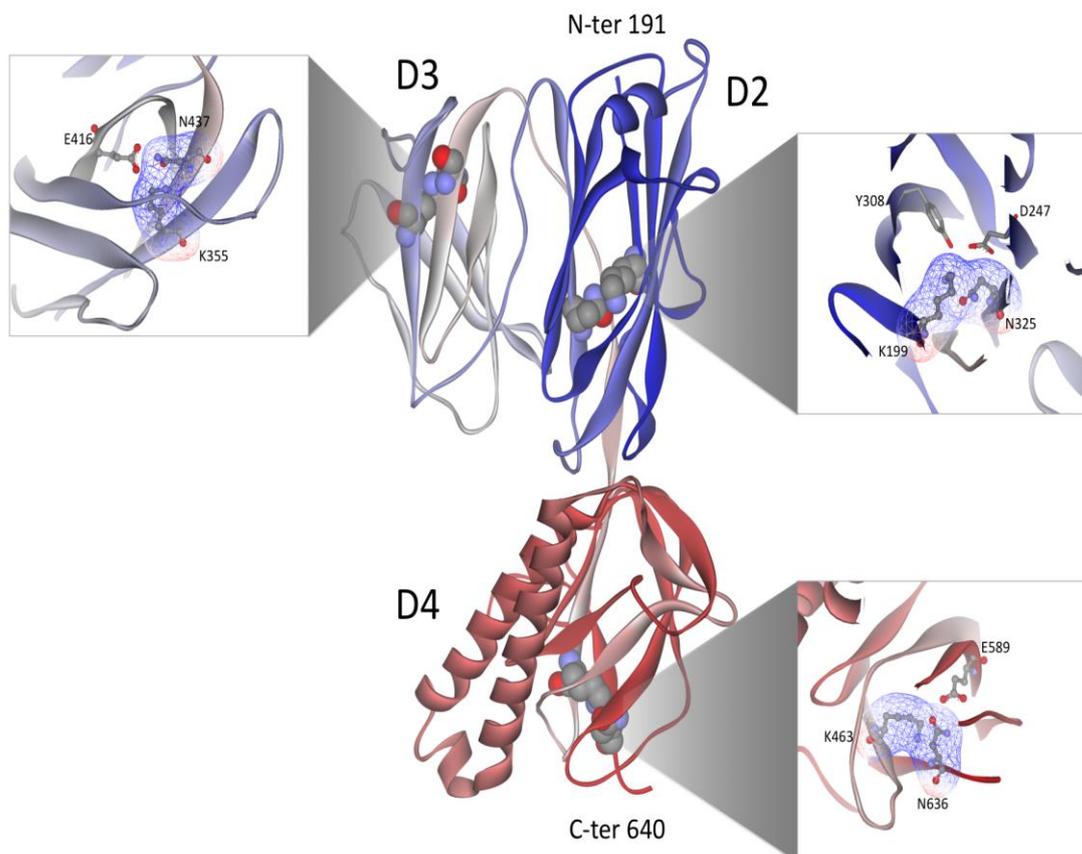


Figure 16: Ribbon representation of BP-2a at 1.75 Å showing the position of the three putative isopeptide bonds (ball and stick representation).

Isopeptide bonds are amide bonds with the same structure as peptide bonds (from the Greek *isos*, meaning equal), but formed between the sidechains of two aminoacids, at least one of which is not an α -amino or α -carboxy group. The best known examples involve the ϵ -amino group of Lys in reaction with a main chain α -carboxy group [69]; however, variations involving side chain carboxy or carboxyamide groups also occur. Isopeptide bonds formed through lysine residues mediate several crucial biological processes triggered

by the intermolecular cross-linking of proteins. Examples include ubiquitylation [70], sumoylation [71], sortase-mediated cell surface protein anchoring [72] and pilus formation [73], [74]. Although they differ in details, these processes share several features:

- (i) all involve the reaction of a Lys ϵ -amino group on one protein with a main chain α -carboxy group on the other.
- (ii) all are enzyme-mediated, and involve a transient thioester intermediate formed by the catalytic residue Cys present in the active site of the participating enzyme. This intermediate is then resolved through nucleophilic attack by the lysine ϵ -amino group to complete an isopeptide bond.

In contrast to these enzyme-dependent processes, one example is known where inter-molecular isopeptide bonds form autocatalytically. This occurs during capsid assembly of the *Escherichia coli* bacteriophage HK97, in which a precursor form of the viral capsid undergoes expansion and reorganization. In this maturation process, isopeptide bonds are formed between Lys and Asn residues of different subunits, resulting in covalent rings of protein subunits that are topologically linked into protein chain mail [75].

It was in 2007 that intramolecular isopeptide bonds were first discovered in the crystal structure of Spy0128, the protein that forms the polymeric shaft of the pili present in *Streptococcus pyogenes* serotype M1 [69]. Continuous electron density joining the side chains of Lys³⁶ and Asn¹⁶⁸ in the N-terminal domain and Lys¹⁷⁹ and Asn³⁰³ in the C-terminal domain suggested the presence of isopeptide bonds joining each Lys–Asn pair. These bonds were

confirmed by mass spectrometric analysis of both the recombinant protein and the native pili. Both bonds were buried in the hydrophobic cores of their respective domains and were associated with a neighboring Glu residue (Glu¹¹⁷ and Glu²⁵⁸, respectively) whose carboxyl group was hydrogen bonded to the isopeptide moiety. In each case, the substitution of the acidic residue by Ala resulted in the loss of the isopeptide bond, indicating that the bonds were generated by an autocatalytic, single-turnover intramolecular reaction catalysed by the Glu residue (Figure 17).

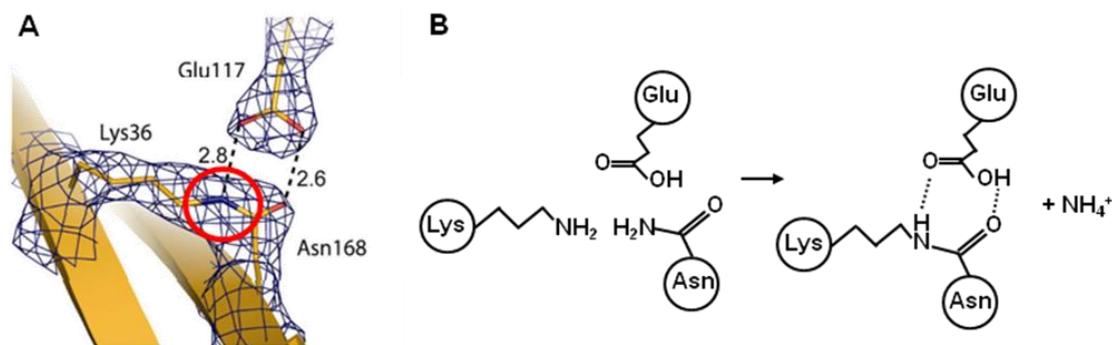


Figure 17: Intramolecular isopeptide bond between Lys³⁶ and Asn¹⁶⁸ on Spy0128 (A, red circle) and scheme of its formation (B)

Recent crystallographic and mass spectral analyses of protein subunits from Gram-positive pili indicate that intra-molecular isopeptide bonds like those observed in Spy0128 are present in all major pilin subunits so far analyzed, despite wide variations in sequence, size and domain organisation. Thus, the shaft-forming major pilins from *Corynebacterium diphtheriae* (SpaA) [76], *Bacillus cereus* (BcpA) [77] and *Streptococcus pneumoniae* (RrgB) [68] follow the pattern seen in Spy0128; therefore internal isopeptide bonds are a consistent feature of the pilus shafts.

The mass difference found on the SAL1486 protein is compatible with the presence of three isopeptide bonds ($- 17 \text{ Da} \times 3 = - 51 \text{ Da}$), as hypothesized by crystallographic study. Thus, a specific MS-based approach was developed to confirm the presence of these three post-translational modifications. For this purpose, the full length SAL1486 was expressed and purified from *E. coli*. The strategy consists in the complete digestion of the recombinant BP-2a using the endoprotease Lys-C, followed by the mass spectrometry analysis of the proteolysis products. In order to easily sort-out and identify the cross-linked peptides (containing the isopeptide bonds), a C-terminal derivatization strategy of the digestion products was developed. The digestion products were treated with O-methylisourea that modifies the C-terminal lysine in homoarginine, leading to a mass increase of 42 Da for each modified C-terminal extremity. Since the cross-linked peptides contain two C-terminal extremities, their derivatization lead to a double mass shift (i.e. +84 Da) (Figure 18).

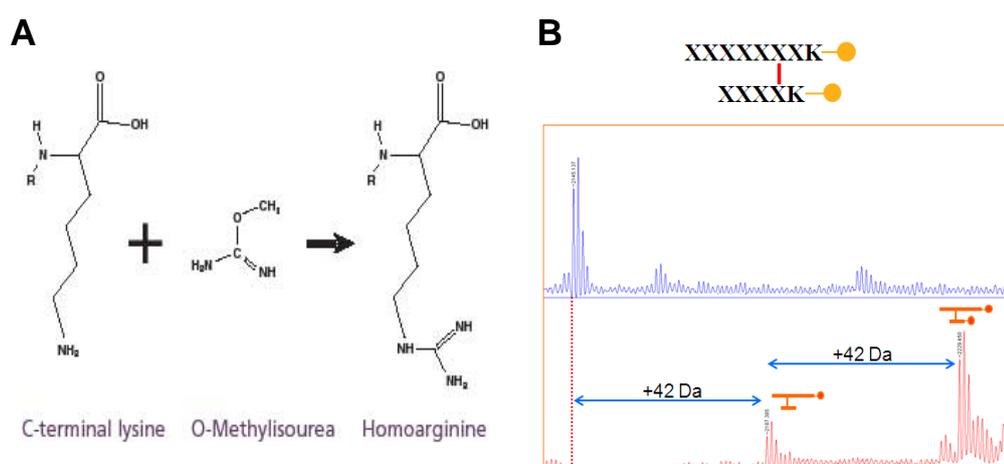


Figure 18: (A) scheme of the derivatization reaction of a C-terminal lysin with O-methylisourea. (B) Example of the mass spectra of a peptide containing an isopeptide bond before and after the derivatization step (in blue and in red, respectively).

When “in solution” digestion was performed, the SAL1486 protein was found to be particularly resistant (with the exception of the D1 domain). The best proteolysis and so the larger sequence coverage was obtained by “in gel” digestion with Lys-C after SDS-PAGE of entire recombinant forms, probably for the strong denaturing effect of the SDS. The peptides produced by the digestion were either directly analyzed by MALDI-ToF mass spectrometry (Figure 18, upper panels) or were modified with O-methylisourea prior the analysis (Figure 18, lower panels). Comparing the two spectra obtained it was possible to identify three isopeptide bonds, one for each domain except for D1. The involved aminoacids are Lys¹⁹⁹ and Asn³²⁵ in the D2 domain (Figure 19, A), Lys³⁵⁵ and Asn⁴³⁷ in the D3 domain (Figure 19, B), and Lys⁴⁶³ and Asn⁶³⁶ in the D4 domain (Figure 19, C). The surrounding area around these bonds is largely hydrophobic, comprising several aromatic residues, in agreement with observations made for the isopeptide bonds in several pilus proteins. Each of the four domains appears to fold independently, as demonstrated by expressing and purifying each domain, selecting the N and C termini based on the domain boundaries defined in the crystal structure of SAL1486 (Figure 16). All four domains were expressed in soluble form in *E. coli*, and MS analysis of tryptic digests of D2, D3, and D4 revealed that the domains carried the same isopeptide bonds found in the full-length protein. This finding suggested that the overall structural organization of the independently expressed domains was sufficiently preserved to bring the lysine and asparagine residues at a suitable distance and allow the autocatalytic reaction [78].

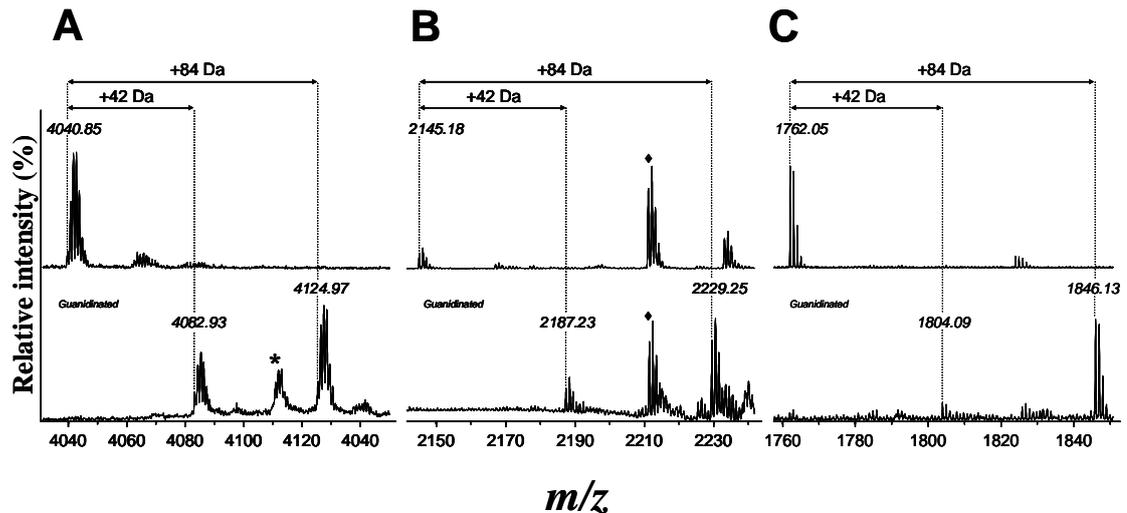


Figure 19: Identification of internal isopeptide bonds by MALDI TOF mass spectrometry.

The recombinant protein BP-2a was run on a 4-12% acrylamide SDS-PAGE. The protein was “in gel” digested with Lys-C. The peptides produced by the digestion were either directly analyzed by MALDI TOF mass spectrometry (upper panel) or were modified with O-methylisourea prior to the analysis (lower panel). Isopeptide in D2 domain was evidenced by the molecular ion of m/z 4040.85 (A, upper panel) that was in agreement with the molecular mass of peptide $^{190}\text{NTETKQVDKNFADK}^{204}$ linked by an isopeptide bond to the peptide $^{307}\text{ITYSATLNGSAVVEVLETNDVK}^{328}$ (expected molecular mass 4040.07 Da). The reaction of guanidination induced a signal mass shift of 42 and 84 Da corresponding to a single and the double C-terminal peptide derivatization, respectively and confirmed the linkage of the two peptides (A, lower panel). By the same way, isopeptide bonds in domains D3 and D4 were assigned from the ion of m/z 2145.18 in agreement with the molecular mass of peptide $^{351}\text{ITVNKTWAVDGNVKNK}^{366}$ linked to peptide $^{437}\text{NNK}^{439}$ (expected molecular mass 2145.13 Da) and from the ion of m/z 1762.05 in agreement with the molecular mass of the peptide $^{461}\text{FVKTNK}^{466}$ linked by an isopeptide bond to the peptide $^{630}\text{DAQQVINKK}^{638}$ (expected molecular mass 1761.90 Da), B and C upper panels, respectively. The reaction of guanidination confirmed the covalent linkage of the peptides by inducing the double mass shift of 42 and 84 Da (lower panels).
(?) Trypsin autodigestion product. (*) Peak not identified.

Finally, to evaluate the function of these isopeptide bonds in the antigenicity of BP2-a, the wild type protein and a mutated form of BP2-a (in which the three lysine residues involved in the isopeptide bonds were substituted into alanine residues) were both tested in a mouse maternal immunization model. Two groups of adult female CD1 mice were immunized with the purified recombinant proteins (with and without isopeptide bonds). After three immunizations, mice were mated and the resulting offspring were challenged with a dose of GBS calculated to kill about 90% of the pups. The high levels of protection observed with the mutated form of the protein revealed that the loss

of isopeptide bonds does not interfere with the capacity of the protein to confer protection in mice and to elicit opsonic antibodies (Figure 20) [78].

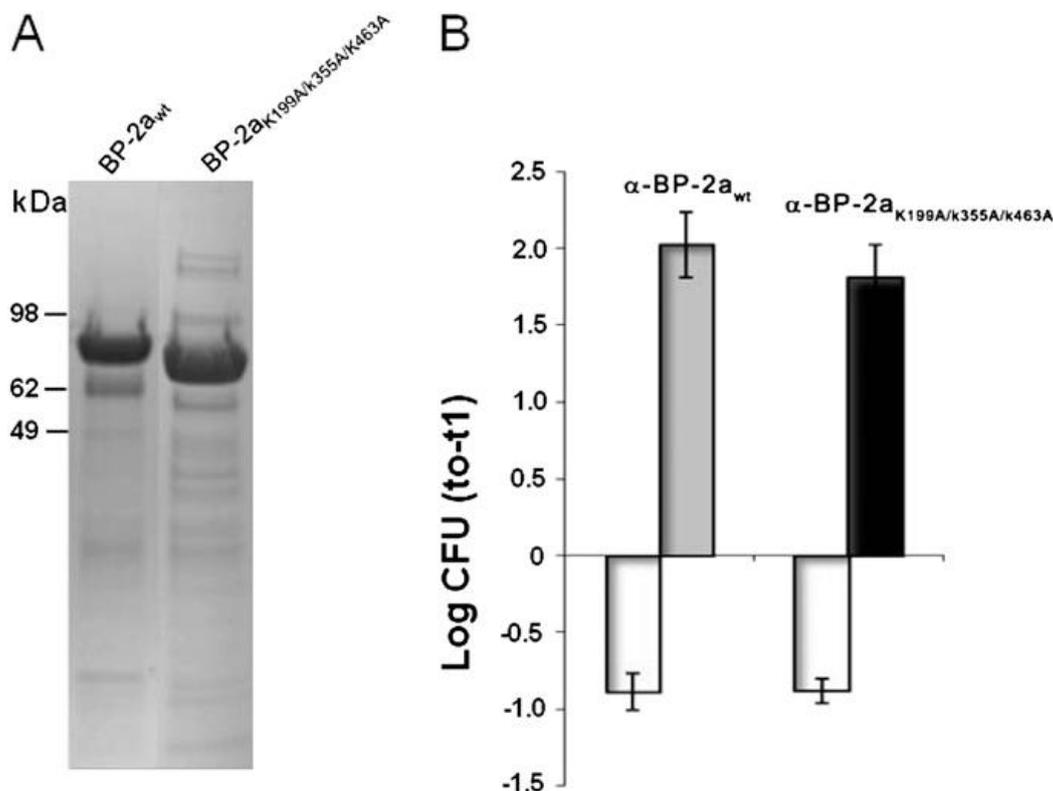


Figure 20: Comparison between the wild type and the mutated form of SAL_1486

(A) SDS/PAGE of purified recombinant wild type BP-2a-515 containing intramolecular isopeptide bonds (BP-2a_{wt}) and BP-2a-515K199A/K355A/K463A, the mutant form lacking the ability to form IPs. The mutant protein showed an electrophoretic mobility slower respect to the naive form. (B) Opsonophagocytosis activity of mice antisera raised against BP-2a_{wt} and BP-2a-515K199A/K355A/K463A in the presence of phagocytic HL60 cells and baby rabbit complement by using Group B Streptococcus strain 515. A 1:30 dilution serum was used in this assay. The log₁₀ difference between Group B Streptococcus CFUs at time 0 and time 1 h are shown. The antigens used are recorded above each bar. White bars represent preimmune sera from the same animals; error bars indicate standard deviation of three independent experiments.

Isopeptide bonds in SdrC, an adhesin from S. aureus

In addition to pilus-associated proteins, other cell-surface adhesins are now known to contain intra-molecular isopeptide bonds. One example is the *Staphylococcus aureus* adhesin Cna that possesses isopeptide bonds in both its collagen-binding A region (CnaA domain) and its repetitive B domains (CnaB domains) [69]. Other examples include the adhesin Ace from

Enterococcus faecalis [79], the *Streptococcus gordonii* antigen I/II adhesin SspB, which has Lys–Asn isopeptide bonds in its two C-terminal CnaA-type domains [80], and the fibronectin-binding protein FbaB from *Streptococcus pyogenes*, which contains a Lys–Asp bond in one of its CnaB domains [81], [82]. The presence of isopeptide bond in bacterial pili and in multidomain adhesins suggests a crucial role not only in resisting mechanical stress, given that these long, thin adhesive structures experience strong tensile forces along the long axis during host cell attachment, but also in the protection against proteolysis. Since the attention given to this type of PTMs is continuously growing, especially in the vaccine field where the characterization of the virulence factors is of pivotal importance, a MS-based approach to rapidly assess the presence of isopeptide bonds has been developed and applied to the *Staphylococcus aureus* adhesin SdrC.

A critical factor for the pathogenic success of *Staphylococcus aureus* depends on its ability to adhere effectively to multiple host tissues [83], [84]; the adhesins mediating staphylococcal adherence and colonization often target the extracellular matrix of the host and hence belong to the MSCRAMM (Microbial Surface Components Recognizing Adhesive Matrix Molecules) family [85]. This class of proteins is typified by the *S. aureus* adhesin Cna and shares a similar structural organization. These proteins contain an amino terminal signal sequence followed by an A-region that often harbors the ligand-binding sites (ligands are matrix molecules such as fibrinogen and collagen). The A-region is comprised of sub-domains (CnaA domains) adopting an immunoglobulin G-like (IgG-like) fold. Sometimes the A-region is followed by a B-region containing repeated β -sandwich modules

of unknown function (CnaB domains) [62]. In the case of the Sdr subfamily of staphylococcal MSCRAMMs, the B-region is accompanied by a repeat domain composed of multiple Ser-Asp dipeptide repeats (SD-repeat or Sdr); SdrC is one of the proteins of this subfamily [86]. Since the presence of isopeptide bonds has already been reported in the Cna protein (both in the CnaA and CnaB domains), as well as in others members of the MSCRAMM family, it has been hypothesized their presence by sequence similarity into the CnaB domains of the SdrC protein. For this study the region containing the CnaB domains was expressed in *E. coli* and purified, and the recombinant protein (named SdrC CnaB) was analyzed by denatured MS. The purpose of this analysis was to measure the intact molecular weight since for each isopeptide bond present, the mass of the protein decrease of 17 Da.

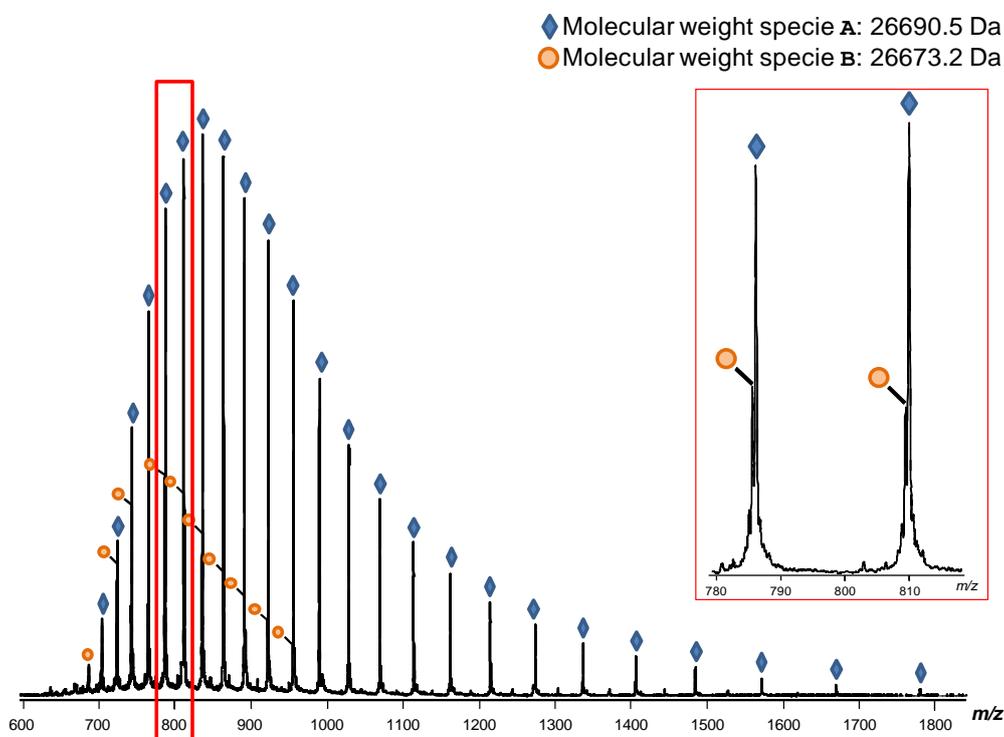


Figure 21: ESI MS spectra of the SdrC CnaB protein in denatured conditions

The spectrum in figure 21 evidences the presence of two species with a molecular weight (MW) of 26690.5 Da and 26673.2 Da, respectively (error: 3.7 ppm). The first one is the SdrC CnaB protein (expected MW 26690.4 Da) while the second one is the same protein with a mass reduction of about 17 Da (no other proteins were identified in the sample after *in solution* digestion with trypsin and LC-MS/MS analysis). This mass reduction is compatible with a loss of a NH_3 group somewhere in the protein and it can be due to a deamidation (succinimide conversion of an asparagine or a glutamine residue) or to the formation of an isopeptide bond. To discriminate between these two cases the protein dynamic in solution was investigated using hydrogen-deuterium exchange mass spectrometry (HDX-MS). This technique takes advantage of a natural process occurring when a protein is in solution; hydrogen located on polar side chains or at the N/C termini and bonded to heteroatoms such as $-\text{N}$, $-\text{O}$, or $-\text{S}$ exchange quite easily with hydrogens in the surrounding solvent. This $\text{H} \rightarrow \text{H}$ conversion cannot be detected by mass analysis. In contrast, exposing a protein to a D_2O -containing environment leads to $\text{H} \rightarrow \text{D}$ replacements that increase the mass of the protein by one unit per exchange event [87]. The reaction is then quenched and analyzed by MS, these steps are performed in an aqueous solution, which means that exchange will continue at a slow rate leading to a partial reversion of deuterated positions after the quench step; this process is referred to as back-exchange. Although HDX takes place also at side chains, the back exchange of these hydrogens is so fast that is not compatible with the timescale of the experiment, thus only the backbone amide hydrogens, that require a longer time both for the exchange and the back exchange, are

taken in account for the analysis. Every residue (with the exception of prolines and the N-terminal amino acid) possesses an amide N–H group, and therefore HDX can probe features affecting the entire protein. Isotope exchange is fastest for completely solvent-exposed amides that are not involved in hydrogen bonding and located on the surface of the protein. The rate constant of N–H→N–D conversion measured is referred to as k_{ex} (“exchange” rate constant). For each individual amide N–H, the value of k_{ex} is determined by the flanking side chains, as well as the pD ($\text{pH}=\text{pD}+0.4$), the temperature and the ionic strength of the solvent. HDX can proceed with acid or base-catalysis, and k_{ex} has its minimum around pH 2.5 (for the amide hydrogens). Catalysis by OD^- is most prevalent under typical conditions. Above pD 3, k_{ex} increases by one order of magnitude with each pD unit, reaching values on the order of 10^3 s^{-1} at pD 9 [88]. The possibility to tune k_{ex} by controlling the pH of the solvent is crucial for many HDX/MS strategies. Key to the application of HDX for structural studies is the fact that the N–H→N–D conversion rate is modulated by the conformational properties of the protein. Ordered regions possess a multitude of intramolecular hydrogen bonds that reduce the solvent accessibility of many amide sites; thus the overall rate constant k_{HDX} is much smaller than k_{ex} . The corresponding protection factors $P=k_{\text{ex}}/k_{\text{HDX}}$ sometimes exceed 10^6 for natively folded proteins. Nonetheless, even protected amides can undergo HDX at measurable rates. These exchange events are mediated by conformational fluctuations of the protein. Specifically, isotope labeling can only occur during short-lived transitions to an “open” conformation. The ongoing occurrence of opening/closing events is a manifestation of the protein thermal motions.

Some of these structural fluctuations may represent events that are quite localized, whereas others are more global. Rate constants of the opening and closing transitions are designated as k_{open} and k_{closed} , respectively. The overall exchange mechanism can thus be described as in Figure 22.

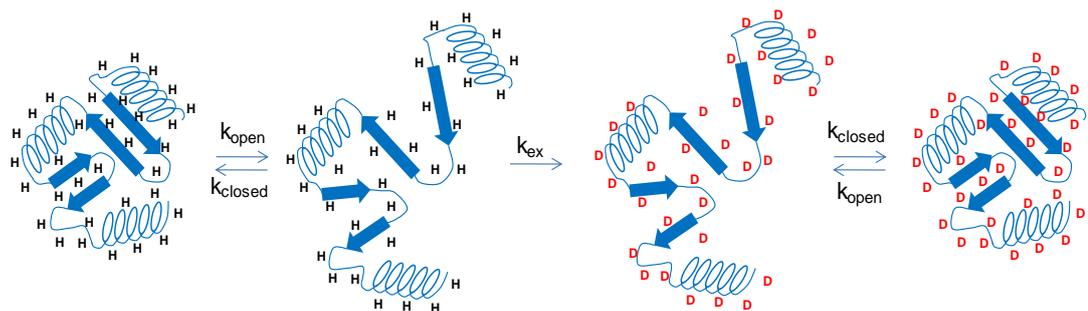


Figure 22: Schematic mechanism of HDX

This methodology allows us to investigate the mass difference found in the SdrC CnaB protein by comparing the dynamic of the two species present in the sample. The rationale is that the presence of a deamidation should not affect the overall dynamic of the protein (Figure 23, A). On the other hand an internal isopeptide bond can block a region of the protein reducing its capability to switch to an “open” conformation resulting in a protection effect of the hydrogens present in that region and therefore in a difference in the overall deuteration pattern and a significative decrease in the number of incorporated deuterium in comparison with the form without the isopeptide bond (Figure 23, B).

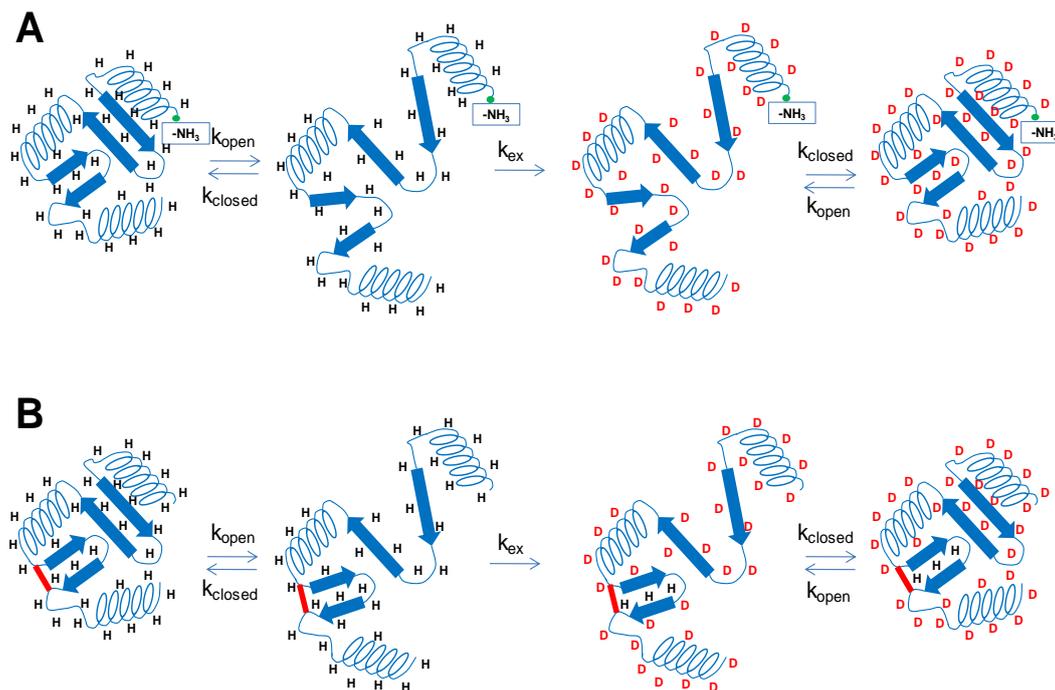


Figure 23: Schematic mechanism of HDX of a protein with a deamidation (A) or with an isoepitope bond (B)

To test this hypothesis, the averaged deuterium exchange behavior of the SdrC CnaB protein was measured at seven time points (from 30 seconds to 8.5 hours). The mass spectra of the +32 charge state of the two forms of the protein in some of the time points analyzed is reported in Figure 24 A; the deuteration was performed at 0 °C in order to decrease the global exchange rate and increase the resolution at lower time-points. After the deuteration, the mass difference between the two forms of the protein is constant for all the time points and it is about 50 Da (Figure 24, B and C).

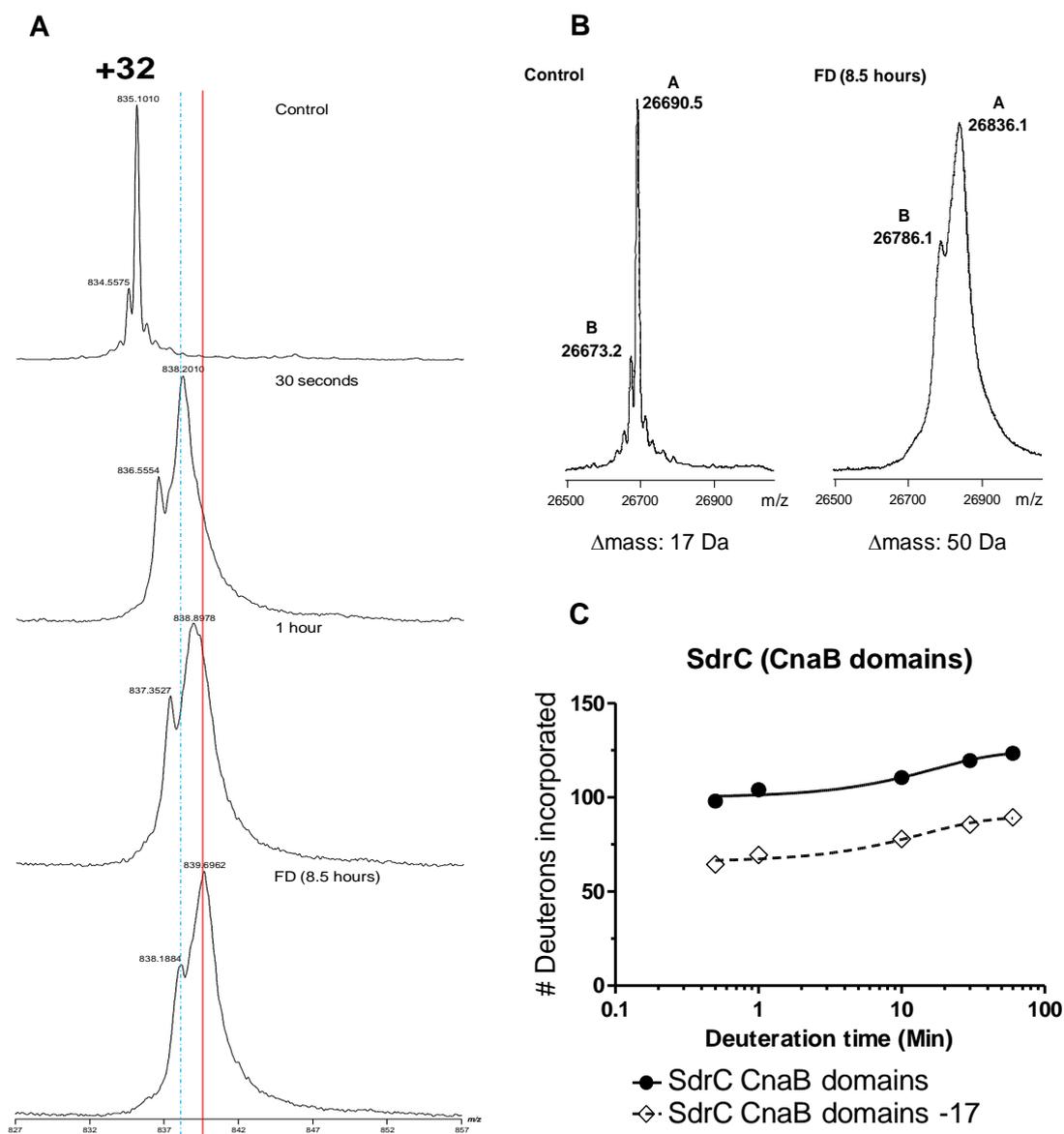


Figure 24: Global HDX-MS analysis of the SdrC CnaB protein

(A) Mass spectra of the +32 charge state of SdrC at different time points (FD, Fully deuterated). The labeling was performed at 0° to increase the resolution at lower time-points. The lines are provided for visual guidance. (B) Deconvoluted spectrum of SdrC prior (on the left) and after 8.5 hours of deuteration (right panel). The mass difference between the two species change from 17 Da to 50 Da. (C) Estimation of the number of deuterons incorporated from the two species. The excel-based program HDX-Express was used to extract the centroid mass of each charge state as a function of the labeling time.

The mass difference between the two species after the deuteration is higher than the one expected for a deamidation suggesting that the initial difference of 17 Da is due to an intramolecular isopeptide bond.

2.2 Development of MS-based approaches to identify unknown PTMs in pathogenic bacteria

Introduction

Recently, heightened attention has been drawn towards post-translational modified proteins in pathogenic bacteria. While the full significance of protein modifications has yet to be precisely defined in prokaryotic systems, post-translational modifications (PTMs) provide additional sources for protein structural and functional diversity. Thus, in a number of human pathogens such as *Streptococcus agalactiae* and *Campylobacter jejuni* [89], PTMs localized on surface proteins have been shown to be directly involved in adhesion, colonization, pathogenicity and virulence. Therefore, modified surface proteins are now considered for vaccine candidate selection.

MS represents a powerful tool for detecting and mapping PTMs since this processing step leads to a mass modification relative to the theoretical molecular weight of the protein. PTMs identification by MS is generally achieved using a two steps analytical strategy. First, the presence (and in some cases the number) of PTMs is revealed by direct mass measurement of the entire protein. Following this step, the modified regions of the protein as well as the nature of the PTMs are further characterized using proteolytic digestions in combination with tandem mass spectrometry experiments [90]. While this approach sounds very "simple", the identification and characterization of PTMs by MS represent a non-trivial task mainly due to the diversity of these modifications and the complexity of the samples to be analyzed. The main objective of this part of the work was to set-up mass spectrometry-based approaches for the identification and characterization of

unknown PTMs on the surface proteins of pathogenic bacteria. The pathogen used for this analysis was *Neisseria meningitidis* serogroup B.

Selection of the starting material for PTMs discovery

Since bacterial membrane proteins are virulence factors that play important roles during infections and are well exposed on the surface of the pathogens, they are considered as potential vaccine candidates. However, their hydrophobic nature makes them difficult to study and requires specific enrichment methods.

To select the best starting material for PTMs discovery, a classical preparation of membrane proteins extracted with sodium carbonate was compared with a preparation of outer membrane vesicles (OMVs) obtained with the *N. meningitidis* MC58 $\Delta gna33$ mutated strain [91]. This strain is deleted for the *gna33* gene, involved in membrane assembly/septation, and is able to release spontaneously relevant quantities of OMVs into the growth medium without requiring any chemical/physical treatment. Both samples were separated by SDS-PAGE and the main bands were identified by MALDI peptide mass fingerprint after in gel digestion.

Figure 25 shows the comparison between the OMVs preparation (lane 2) and the preparation obtained after sodium carbonate extraction (lane 1). The OMVs were selected for PTMs discovery as they contain more outer-membrane proteins and appear less contaminated compared to the classical preparation of extracted membrane proteins.

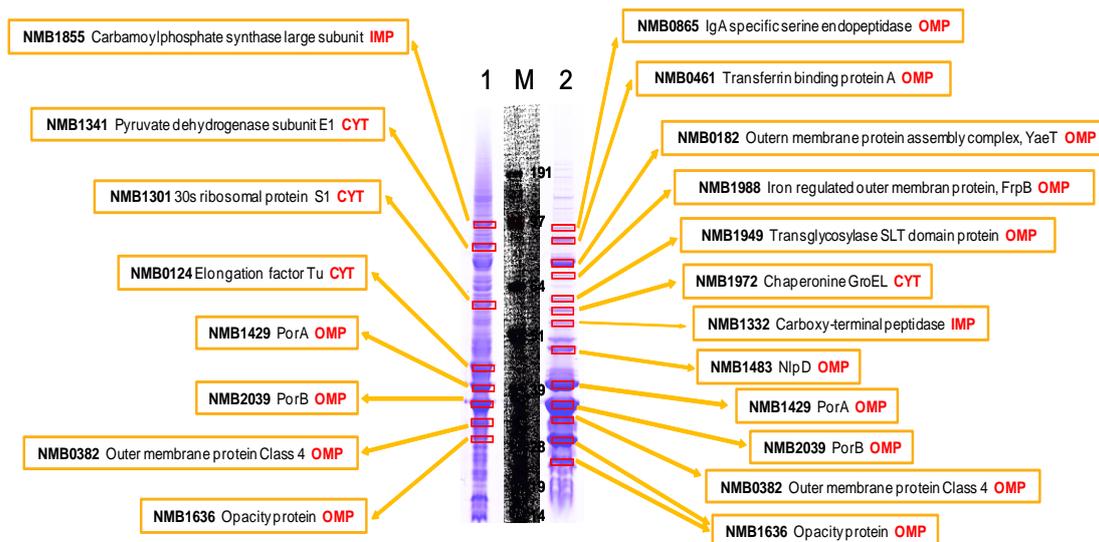


Figure 25: SDS-PAGE analysis comparing the main composition of a membrane preparation extracted with sodium carbonate extraction (lane 1) and a OMVs preparation from the *N. meningitidis* MC58 Δ gna33 mutated strain (lane 2).

CYT = cytoplasmic; IMP = inner membran protein; OMP = outer membrane protein

Characterization of the OMVs and PTMs discovery

In order to characterize proteins associated to the vesicles and to identify PTMs, a combined proteomic approach was set-up. A part of the OMVs preparation was first separated by SDS-PAGE and proteins were identified by MALDI peptide mass fingerprints after in gel digestion. In parallel, OMVs were directly subjected to trypsin digestion and the generated peptides identified by nano-LC/MS/MS. Mass spectra were processed either manually or with a local version of the Mascot search engine (using a database containing protein sequences deduced from the sequenced MenB genomes, downloaded from NCBI/nr) in order to identify specific neutral losses and/or reporter ions corresponding to unknown PTMs (Figure 26).

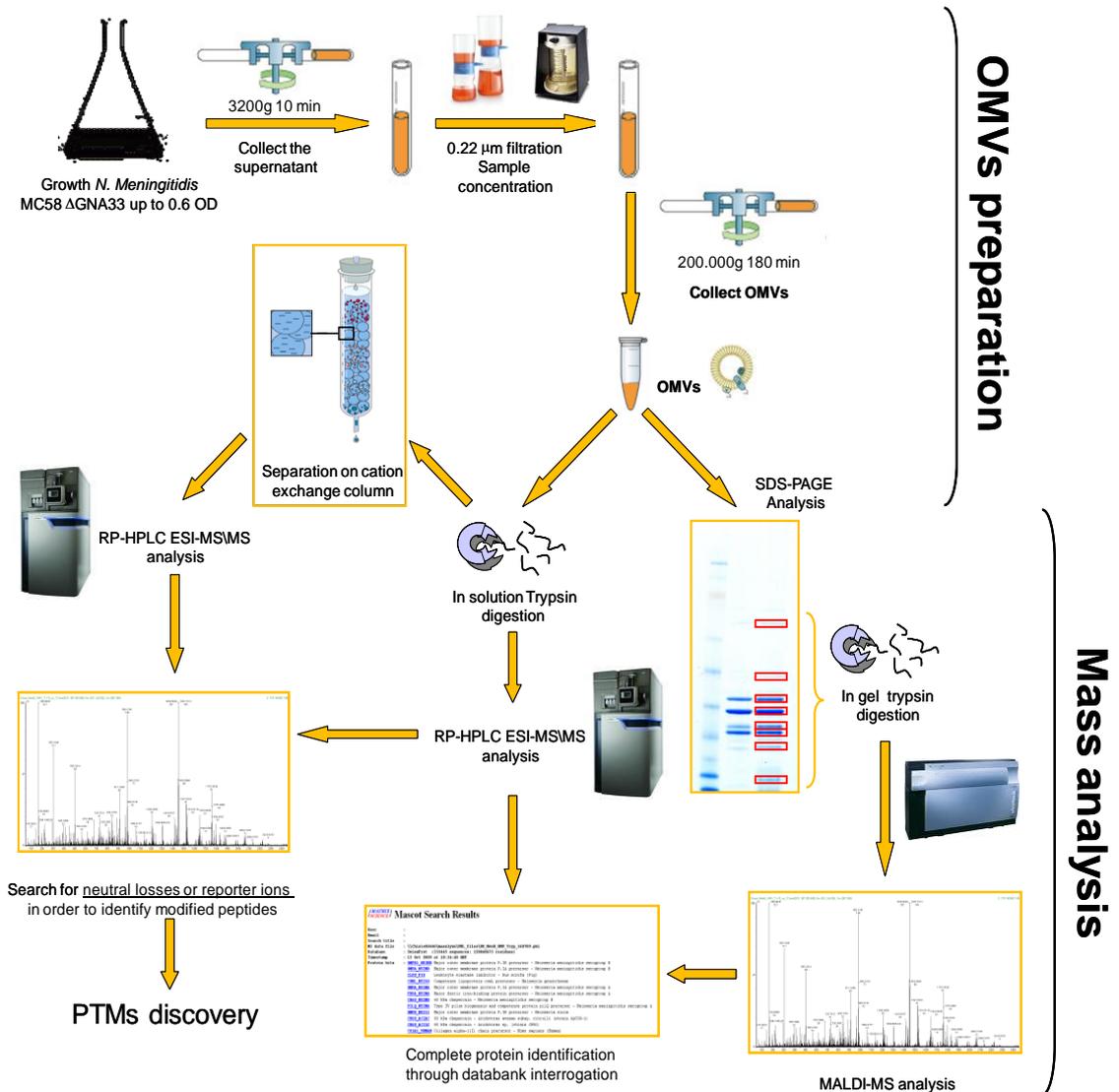


Figure 26: Schematic overview of the approach used for the identification of PTMs on OMVs proteins.

For the total characterization of the proteins present on the OMVs, an in solution digestion with trypsin was performed and the peptides were analyzed by nanoLC-MS/MS. A total of 60 proteins were identified. Most of the proteins (88%) were classified as outer-membrane proteins according to PSORT prediction, 4 proteins (7%) were classified as periplasmic and 3 proteins (5%) as cytoplasmic (Table 3).

ID	Name
Outer Membrane Proteins	
NMB2039	major outer membrane protein IB (porB)
NMB1429	major outer membrane protein IA (porA)
NMB0018	pilin PilE (pilE)
NMB0382	ompA family protein
gij120866875	putative lipoprotein (orf 731)
NMB0703	competence lipoprotein comL (comL)
NMB1057	gamma-glutamyltransferase (ggt)
NMB1053	outer membrane protein OpcA (opcA)
NMB0345	conserved hypothetical protein
NMB1483	LysM domain-M23 peptidase domain protein
NMB2091	phospholipid-binding domain protein
NMB0088	outer membrane protein, OMPP1-FadL-TodX family
NMB0550	putative thiol:disulfide interchange protein DsbC
NMB0707	rare lipoprotein B family
NMB0928	putative lipoprotein
NMB0182	outer membrane protein assembly complex, YaeT protein
NMB0204	lipoprotein, SmpA-OmlA family
NMB0281	surA-PPIASE domain protein
NMB0294	DSBA thioredoxin domain protein
NMB0663	outer membrane protein NsgA
gij2150054	opacity protein
NMB1124	putative lipoprotein
NMB1030	Ycel family protein
NMB1812	type IV pilus secretin PilQ (pilQ)
NMB0109	LysM domain protein
NMB1126	CsgG family protein
NMB1870	lipoprotein NMB1870
gij2315235	Opa1800 outer membrane protein
gij1841506	opacity outer membrane protein
NMB0700	IgA-specific serine endopeptidase
NMB1309	type IV pilus biogenesis-stability protein (pilF)
NMB1497	TonB-dependent receptor
NMB1567	macrophage infectivity potentiator
NMB0460	transferrin-binding protein
NMB2132	transferrin-binding protein-related protein
NMB0992	adhesin
NMB1519	thiol:disulfide interchange protein DsbD
NMB0181	outer membrane protein OmpH, putative
NMB1961	VacJ-related protein
NMB1398	Cu-Zn-superoxide dismutase
NMB1985	adhesin
NMB0783	conserved hypothetical protein
NMB0346	conserved hypothetical protein
NMB2095	conserved hypothetical protein
NMB0035	conserved hypothetical protein
NMB1557	conserved hypothetical protein
NMB1125-1163	hypothetical protein NULL
NMB2139	conserved hypothetical protein
NMB0039	hypothetical protein
NMB1963	conserved hypothetical protein
NMB1620	conserved hypothetical protein
NMB2147	hypothetical protein
NMB1468	hypothetical protein
Periplasmic Proteins	
NMB1946	D-methionine ABC transporter, periplasmic D-methionine-binding protein (metQ)
NMB0634	iron(III) ABC transporter, periplasmic iron(III)-binding protein (fbpA)
NMB0355	lipopolysaccharide ABC transporter, periplasmic lipopolysaccharide-binding protein (lptA)
NMB1332	C-terminal processing peptidase
Cytoplasmic Proteins	
NMB1285	phosphopyruvate hydratase (eno)
NMB0124	translation elongation factor Tu (tuf)
NMB1972	chaperonine GroEL

Table 3: Proteins identified on MenB OMVs.

Tryptic peptides were separated off-line using a strong cationic exchange resin prior to nano-LC-MS/MS analysis. Mass spectra were processed with a local version of the Mascot search engine using a database containing protein sequences deduced from the sequenced MenB genomes, downloaded from NCBIInr.

Due to the high number of membrane proteins identified, these proteins should be carefully considered as components of the membrane compartment. After automatic analysis of the MS/MS data with MASCOT, the unidentified spectra were all manually interpreted in order to select MS/MS spectra of peptides containing a neutral loss or a reporter ion with a mass corresponding to the mass difference observed between the modified and unmodified peptide (Figure 27).

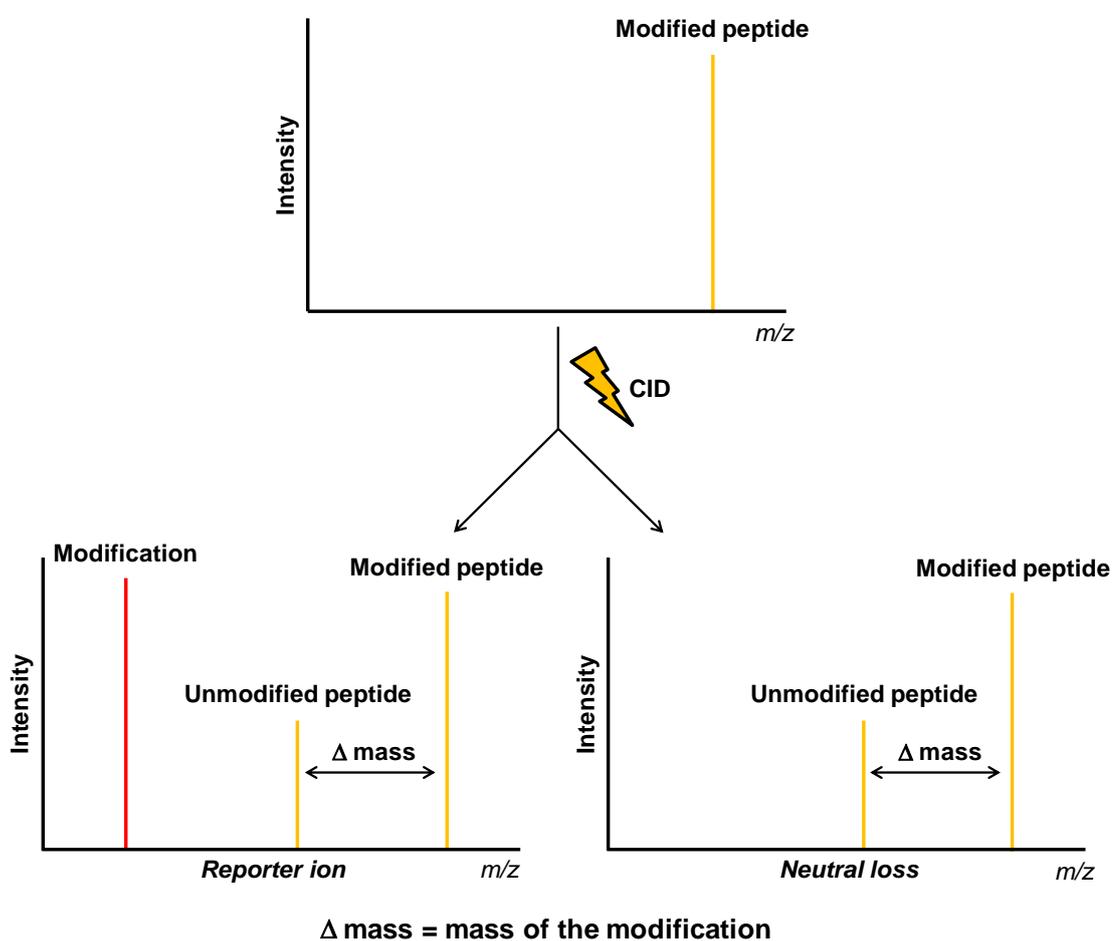


Figure 27: Rationale of the mass spectrometric approach used to identify new PTMs (CID, collision induced dissociation)

Using this strategy two modified peptides, belonging to the protein encoded by the *orf 731* (Figure 28, A) and the NMB 0382 (Figure 28, B) displaying a delta

mass of 166 Da, have been identified. In both cases, the fragmentation pattern contains a reporter ion with an m/z value of 167. The peptides were fully sequenced and the modified residue was identified as a cysteine. This is the first time that such a modification is reported thus suggesting the presence of a new type of PTM.

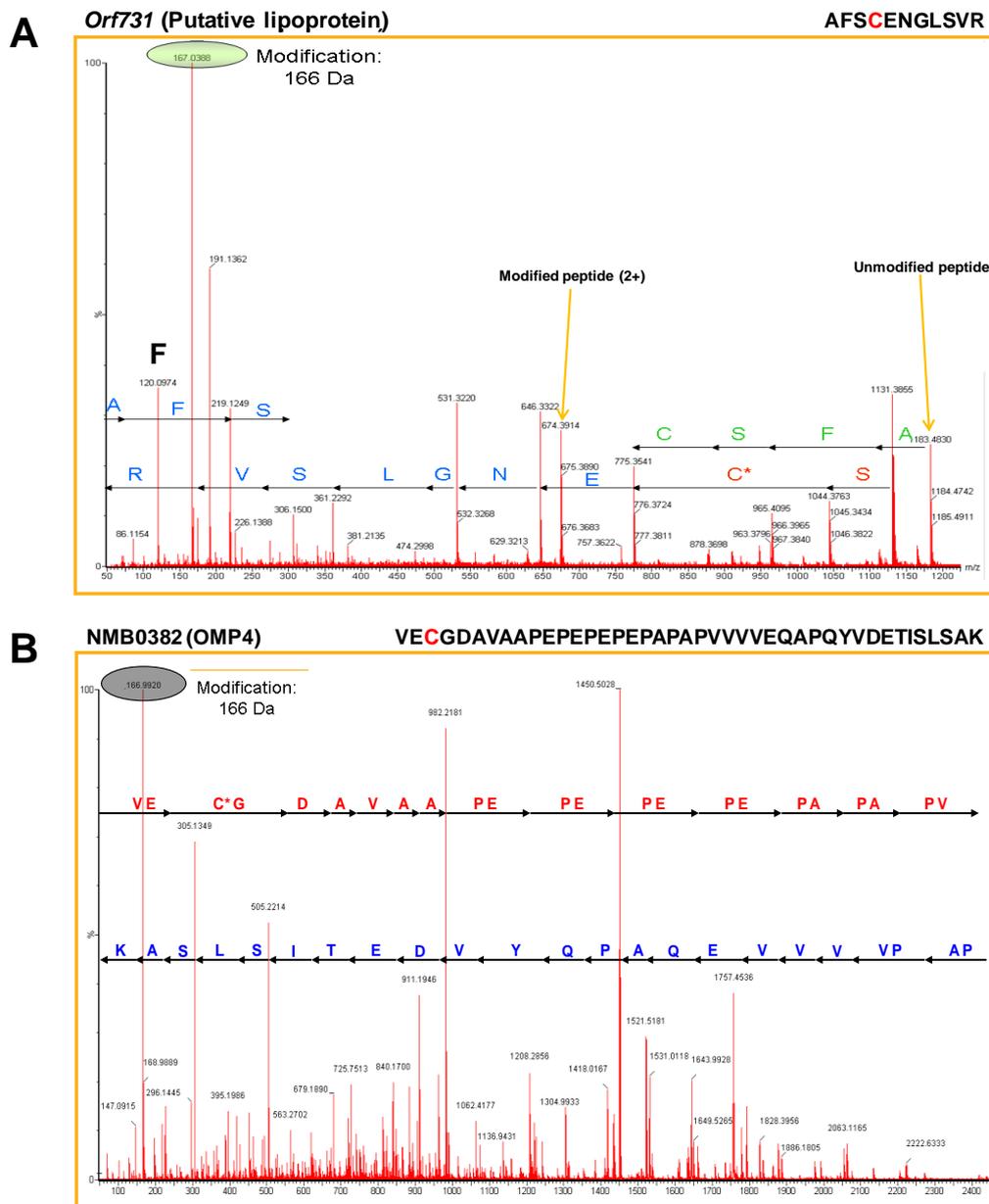


Figure 28: MS/MS spectra of the peptides carrying the putative PTM (*orf 731* panel A, NMB 0382 panel B). The reporter ion is highlighted (green ellipses).

Because of their surface localization, these two proteins could be considered as potential vaccine candidates.

The *orf731* codify for a putative lipoprotein well conserved among different neisserial strains. In literature there are no available data about this protein. In the Pfam database this protein, belong to the MliC (membrane bound lysozyme inhibitor of c-type lysozyme) superfamily, this family of proteins possesses lysozyme inhibitory activity and confers increased lysozyme tolerance [92]. Lysozyme is part of the innate immune system, it is an enzyme that hydrolyze the peptidoglycan by cleaving the glycosidic bond that connects N-acetylmuramic acid with the fourth carbon atom of N-acetylglucosamine; it is abundant in a number of secretions, such as tears, mucus, human milk, and especially saliva. Bacteria have evolved various mechanisms to evade this bactericidal enzyme, one being the production of lysozyme inhibitors. Since the ecological niche of *Neisseria meningitidis* is the human nasopharynx where it is continuously exposed to lysozyme, it is possible to hypothesize a crucial role of the protein coded by the *orf731* in the protection against this enzyme.

NMB0384 is a class 4 outer membrane protein known also as RmpM [93]. NMB0384 is highly conserved in all serogroups of *N. meningitidis* (around 99% sequence identity) and shares 94% sequence identity with its gonococcal orthologue, protein III. The NMB0384 sequence can be divided into four parts: a 22-residue signal sequence which is cleaved by a signal peptidase during translocation of the protein to the periplasm, an N-terminal domain of approximately 40 amino acids, followed by a 20-residue hinge region rich in proline. The C-terminal domain of RmpM consists of

approximately 150 amino acids sharing 35% sequence identity with the C-terminus of *E. coli* OmpA, and is therefore called an OmpA-like domain. C-terminal, OmpA-like domains, found in many Gram-negative bacterial proteins, have been suggested to associate non-covalently with peptidoglycan [94], [95]. Although NMB0384 has been identified as an outer membrane protein, it is not clear how it associates with the outer membrane. NMB0384 has no modifiable N-terminal cysteine residue which could accept a lipidic moiety, and the N-terminal part of the protein encompasses only 40 amino acids, which is too short to form a monomeric transmembrane β -barrel structure. However, this protein fractionates with outer membranes [96] and has been shown to interact with integral outer membrane proteins. NMB0384 forms heterooligomeric complexes with the two meningococcal major porins, PorA and PorB [97], and with the TonB-dependent transporters, TbpA (transferrin binding protein A) and LbpA (lactoferrin binding protein A) [98]. Because NMB0384 contains an OmpA-like domain and is able to interact with outer membrane proteins, it can work as a structural protein, linking the outer membrane to the peptidoglycan layer [95] and [98]. This link is essential for the integrity of the cell. For example, a $\Delta ompA-lpp$ *E. coli* strain, lacking both OmpA and the major outer membrane lipoprotein which interacts covalently with peptidoglycan, shows defects such as hypersensitivity to toxic compounds, the release of periplasmic proteins and the formation of outer membrane vesicles [99]. A $\Delta NMB0382$ *N. meningitidis* strain does not show such severe defects: the mutant has the same morphology and growth characteristics as the parental strain [90]. This suggests that other proteins can fulfill the role of NMB0384 in *N. meningitidis*.

The + 166 Da modification found on these proteins still need to be characterized but, since in both the proteins the modified residue is a cysteine not included in any functional domain, the putative PTM seems to be not directly involved in their functions. Nevertheless, further analyses are required in order to confirm the presence of this putative modification and to assign a possible chemical structure and a biological and immunological function.

3 Conclusions

In the reverse vaccinology process, protein vaccine candidates are selected following 4 main steps: (i) antigen selection; (ii) cloning/expression of the selected genes and purification of the recombinant forms of the antigens; (iii) *in vitro* and *in vivo* assays to define protection and toxicity; and (iv) structural, functional, epidemiological and immunological characterizations of the recombinant antigens that demonstrates protection in animal model and no toxicity. In spite of the success of the reverse vaccinology, several aspects that could not be assessed by the approach are currently emerging. One of these aspects is the impossibility to obtain information about the post-translational modifications (PTMs) of the putative vaccine candidates. Moreover the necessity to use heterologous recombinant proteins may results in changes in the maturation, compared to the native proteins, which can affect their immunogenicity.

Overexpression of a protein in a foreign host, such as *Escherichia coli*, is frequently the first step toward biochemical, enzymatic, and structural studies and is instrumental when purification from the natural source(s) is hardly

achievable. High-level production of functional heterologous proteins in *E. coli* often remains difficult in spite of the improvements achieved in the past decade. Indeed, heterologous protein overexpression in *E. coli* continues to be a challenging task for proteins possessing numerous disulfide bridges and/or being the target of post-translational modifications or when genes enriched in rare codons (i.e., codons that are used with very low frequency in this host) have to be expressed. Despite these limitations, bacterial expression often yields reasonable amounts of proteins that can then be extensively studied to get biological and structural insights. The key issue in these studies is to obtain large amounts of the purified recombinant protein with a homogeneity as high as possible prior to proceeding to its biochemical, functional and structural characterization. This requirement is deeply interconnected with the necessity of precisely determining the identity of the recombinant protein and of fully unraveling its primary structure, as well as with the need of unveiling any possible chemical modifications leading to undesirable microheterogeneities.

Traditional approaches used for quality control of recombinant proteins are based on bottom-up proteomics methodologies. Although a wealth of literature reports pointed out the successful use of this approach, the latter suffers from some limitations when it comes to determining the full complexity of a protein sample. For this purpose the top-down MS/MS approach has been developed. This combines the measurement of the intact experimental mass with the recording of MS/MS data on the full-length protein. Such a technique is becoming more and more popular since it allows an extensive description of protein properties. In addition to this “classical” mass

spectrometry approaches, together with equally spectacular advances in mass spectrometric instrumentation, a new field has emerged, termed native protein mass spectrometry, which focuses on the structural and functional analysis of the dynamics and interactions occurring in protein complexes. Native MS gives information about the composition, topological arrangements, dynamics, and structural properties of protein complexes. The mass range is theoretically unlimited and highly dynamic, allowing the detection of small subunits and large complexes within the same measurement and the amount of protein needed for an analysis is, compared to most other structural biology methods, very low. In the past years, the use of this methodology led to exciting applications ranging from the detailed study of equilibria between different quaternary structures as influenced by environmental changes or binding of substrates or cofactors, to the analysis of intact nano-machineries.

The first part of the work herein presented is related to the development of mass spectrometry-based approaches to study the maturation of recombinant proteins and the application of these methods to proteic vaccine candidates. I analyzed seven recombinant proteic vaccine candidates, belonging to three pathogenic microorganisms (Table 1). All the proteins were expressed in *E. coli*, purified avoiding denaturing steps and their oligomerization state was assigned using native MS (Table 2). Among the proteins tested, three were found monomeric (GNA2091, fHbp and SAL1486), two were dimeric (GNA1030 and NadR, as suggested in [45]) and one was trimeric (NadA, as suggested in [46]). Three proteins (PSL1,

GNA1030 and SAL1486) presented a mass difference between the expected and the observed MW and required further investigations.

PSL1 was present in two forms: a covalent dimer, through a disulfide bridge, and a monomer with a mass increase of 765.6 Da, also linked through an S-S bond. Both these modifications are not physiological since the only cysteine present in the protein is covalently attached, in nature, to a diacylglycerol moiety. In order to characterize the modification, the protein, with and without reducing agent, was analyzed by MALDI-ToF MS in negative ionization mode and a signal at 766.6 m/z (MW of 767.6 Da) was present only in the reduced sample (Figure 10) and was identified as the coenzyme A (MW of 767.5 Da). In literature is already reported that molecules with free thiols are able to link to protein cysteine through disulfide bonds (S-thiolation) [50]; this modification is commonly observed in recombinant proteins secreted from *E. coli* cells. Various thiol modifiers have been identified by MS including glutathione, gluconoylated glutathione, 4-phosphopantetheine, dephosphorylated coenzyme A and coenzyme A. S-thiolation in this case can be a response to environmental stress experienced by the cells during the high cell density growth, or to the (patho)-physiological burden brought on by the expressed proteins. Moreover, the attachment of the CoA could affect the immunogenicity of the protein, since the structure of this molecule is similar to some TLR agonists (Figure 11) [51]. Thus the ability of the modified and unmodified PSL1 to activate the TLRs has been tested but no differences has been found between the two samples (data not shown), indicating that the CoA does not possess an adjuvant activity. To avoid any risk a mutated form of the protein, with the deletion of the cysteine

residue, has been generated. The mutated protein is still able to confer protection in mice immunization models and, after native MS analysis, showed a monomeric oligomerization state and an observed MW in agreement with the expected one.

The mass increase found on the GNA1030 instead (+ 1457 Da) is present only in the native MS analysis thus indicating a non covalent modification. Since in literature is reported that many homologs of this protein are able to bind a lipid molecule (Figure 12) [57], [58], [59], has been hypothesized that also GNA1030 is bound to a small molecule that is responsible for the increase of MW in native conditions. This hypothesis has been demonstrated by in source fragmentation during the native MS analysis of the protein, showing the presence of a small ligand, with a MW of about 728 Da, non-covalently bound to the GNA1030 protein with a 1:1 stoichiometry (Figure 13). In order to fully characterize the ligand, the correspondent ion has been selected and fragmented, in both positive and negative mode (Figure 15, A and B respectively). Based on these data, it was possible to identify the ligand as the ubiquinone 8 (Figure 15, C). The ubiquinone 8 has been also identified as the ligand of GNA1030 purified from both the cytoplasm of *E. coli* and the periplasm of MenB, suggesting that the protein binds the ligand specifically. In literature is reported that the genes of the cytochrome b561 family and the GNA1030 homologs often exist as neighbors, such as in *Bacillus subtilis*, *Vibrio cholerae*, *Pseudomonas aeruginosa*, *Yersinia pestis*, *Mesorhizobium loti*, *Xylella fastidiosa*, and *Caulobacter crescentus* [58]. Thus, it is likely that these members of the Ycel like family of proteins play a role in the electron transport system by binding polyisoprenoid molecules. In

the case of GNA1030, the upstream gene does not encode for a cytochrome but for an enzyme involved in leucine biosynthesis. The GNA1030 protein can also act as a carrier of the ubiquinone 8 from the cytoplasm to the periplasm, or can be involved in the stress response of the bacteria. This represents, in our best knowledge, a first report where native MS was a key strategy to define a putative biological role to a protein with unknown function.

Finally the mass difference found on the SAL1486 protein is compatible with the presence of three isopeptide bonds ($- 17 \text{ Da} \times 3 = - 51 \text{ Da}$), as hypothesized by crystallographic study. Thus, a specific MS-based approach was developed to confirm the presence of these modifications. The strategy developed consists in the derivatization of the C-term extremities of cross-linked peptides followed by MS analysis. This strategy allowed us to identify the three isopeptide bonds hypothesized in SAL1486 protein (Figure 16). Moreover, to evaluate the function of these PTMs in the antigenicity of SAL1486, the wild type protein and a mutated form of SAL1486 were both tested in a mouse maternal immunization model. The results revealed that the loss of isopeptide bonds does not interfere with the capacity of the protein to confer protection in mice and to elicit opsonic antibodies (Figure 20). Thus the isopeptide bonds in this protein are probably required for the resistance to mechanical stress and for the protection against proteolysis [78].

In addition to pilus-associated proteins, other cell-surface adhesins are now known to contain intra-molecular isopeptide bonds. One example is the *Staphylococcus aureus* adhesin Cna that possesses isopeptide bonds in both its CnaA and CnaB domains [62]. The high presence of isopeptide bond in

bacterial pili and in multidomain adhesins suggests a crucial role of these PTMs in the stabilization of such virulence factors. Since the attention given to this type of PTMs is continuously growing, especially in the vaccine field, a MS-based approach to rapidly assess the presence of isopeptide bonds using hydrogen-deuterium exchange has been developed and applied to the *Staphylococcus aureus* adhesin SdrC. The region of the sdrC protein containing the CnaB domains was expressed in *E. coli* and purified, and the recombinant protein (named SdrC CnaB) was analyzed by denatured MS. In the entire mass spectra two species are present. The first one is the SdrC CnaB protein while the second one is the same protein with a mass reduction of about 17 Da. This mass reduction is compatible with a loss of a NH_3 group somewhere in the protein and it can be due to a deamidation (succinimide conversion of an asparagine or a glutamine residue) or to the formation of an isopeptide bond. To discriminate between these two cases the protein dynamic in solution was investigated using HDX-MS. The data obtained revealed that the mass difference between the two species after the deuteration is bigger than the one expected for a deamidation suggesting that the initial difference of 17 Da is due to an intramolecular isopeptide bond.

Obviously these two strategies can be applied for the assessment and the identification of isopeptide bonds in every protein of interest providing useful tools in the study of this type of post-translational modifications.

It is clear that native MS may play a vital, and rather unique, role in structural biology and particularly in the study of the protein maturation, providing not only detailed information about protein complex stoichiometry, but also about

the effects on structures and changes in the environment, the binding of cofactors, and the interactions and dynamics with other proteins or protein complexes, giving in the same time informations about covalent modifications.

In the second part of the thesis I applied different MS approaches to identify unknown PTMs in *Nisseria meningitidis* serogroup B surface proteins. As a matter of fact, for most bacterial pathogens, the proteins that are likely to induce a protective immuno-response are those well expressed and well exposed on the cell surface, These proteins have in fact the highest chances to come into contact with the host immune system. Therefore a complete characterization of the protein composition of the bacterial surface, along with their post-translational modifications, is of great relevance in vaccine research. However, their hydrophobic nature makes them difficult to study and requires specific enrichment methods. To select the best starting material for PTMs discovery, a classical preparation of membrane proteins extracted with sodium carbonate was compared with a preparation of OMVs obtained with the *N. meningitis* MC58 $\Delta gna33$ mutated strain [91]. The OMVs were selected for PTMs discovery, as they contain more outer-membrane proteins and appear less contaminated compared to the sodium carbonate extraction (Figure 25), and fully characterized using a combined proteomic approach including both MALDI-ToF/MS and nano LC/MS/MS (Table 3).

The main issue in the identification of bacterial PTMs is the lack of dedicated bioinformatics tools able to identify low represented or unknown modifications. Indeed, many approaches were developed to identify peptides

and proteins using amino acid sequence information from MS/MS and have been extended to identify modified peptides and proteins. However, many of them took into account only a few types of PTMs during the analysis, ignoring all the others and the investigators had to guess in advance which PTMs exist in a sample. Moreover, some search tools compared an MS/MS spectrum with all possible combinations of PTMs for each peptide from a database, thus, requiring extremely expensive computation.

In this work, the MS/MS spectra obtained from the characterization of the MenB vesicles were then manually processed in order to identify specific neutral losses and/or reporter ions corresponding to unknown PTMs (Figure 27). Using this strategy two modified peptides, belonging to the protein encoded by the *orf 731* (Figure 28, A) and the NMB0382 (Figure 28, B) displaying a mass increase of 166 Da, have been identified. In both cases, the fragmentation pattern contains a reporter ion with an m/z value of 167 and the modified residue was identified as a cysteine. Both the identified proteins are outer membrane proteins. The *orf731* encodes for a putative lipoprotein well conserved among different neisserial strains that belong to the MliC (membrane bound lysozyme inhibitor of c-type lysozyme) superfamily. This family of proteins possesses lysozyme inhibitory activity and confers increased lysozyme tolerance. Since the ecological niche of *Neisseria meningitidis* is the human nasopharynx where it is continuously exposed to lysozyme, it is possible to hypothesize a crucial role of the protein coded by the *orf731* in the protection against this enzyme. NMB0384 is a class 4 outer membrane protein known also as RmpM [93]. This protein is able to interact with other outer membrane proteins and possess a C-terminal OmpA-like

domain, needed to non-covalently associate with peptidoglycan [94], [95]. Up to now, it is not clear how the NMB0384 associates with the outer membrane since has no modifiable N-terminal cysteine residue which could accept a lipidic moiety, and the N-terminal part of the protein is too short to form a monomeric transmembrane β -barrel structure. Because NMB0384 contains an OmpA-like domain and is able to interact with outer membrane proteins, it can work as a structural protein, linking the outer membrane to the peptidoglycan layer [95], [98]. This link is essential for the integrity of the cell; for example, a $\Delta ompA-lpp$ *E. coli* strain, lacking both OmpA and the major outer membrane lipoprotein which interacts covalently with peptidoglycan, shows defects such as hypersensitivity to toxic compounds, the release of periplasmic proteins and the formation of outer membrane vesicles [99]. A $\Delta NMB0382$ *N. meningitidis* strain does not show such severe defects [90] suggesting that other proteins can fulfill the role of NMB0384. This is the first time that such a modification is reported thus suggesting the presence of a new type of PTM that requires further investigations in order to be fully characterized.

In literature there are only few works dedicated to the study of the protein maturation in bacteria, especially about PTMs. Therefore, further investigations are required to understand not only the type and the number of the modifications, but also their biological function. Moreover in the vaccine field it is of pivotal importance to study the maturation of the proteic vaccine candidates in order to understand their role in parasite-host interaction and

their capacity to generate diversity and to influence antigenicity. In this thesis I developed different MS approaches to study specific protein modifications revealed by the native MS analysis of recombinant proteins. Such molecular accidents illustrate the chemically dynamic nature of the cellular milieu and emphasize the need to check that any new recombinant protein has the expected structure, especially for proteic vaccine candidates. The MS methods developed can be then applied to other proteins of interest. In addition, I applied the mass spectrometric approach to identify unknown post-translational modifications on pathogenic bacteria surface proteins reporting a putative new PTM on *Neisseria meningitidis* serogroup B. Due to its sensitivity and specificity, mass spectrometry is the methodology of choice for the identification and characterization of both covalent and non covalent protein modification. Nevertheless, the MS data analysis softwares, to support such work, still need to be developed.

4 Materials and Methods

Protein expression and purification

All the recombinant proteins used in this thesis were produced and purified by the Protein Biochemistry Unit of Novartis Vaccines and Diagnostics srl (Siena, Italy).

Active immunization

All the immunizations were performed by the Animal Care Platform of Novartis Vaccines and Diagnostics srl (Siena, Italy).

Entire mass analysis in denaturing conditions

The samples were diluted with formic acid (Sigma-Aldrich, St. Louis, MO, USA) 0.1%, with or without 10 mM TCEP (Sigma), and injected into a LC-20ADXR Prominence HPLC system controlled by a CMB-20A module (Shimadzu Corporation, Kyoto, Japan). The protein samples were trapped and desalted for 2 min at a flow rate of 220 μ L/min using a Protein Micro Trap column (Michrom BioResources, Inc., Auburn, CA, USA) equilibrated with 100% buffer A (0.1% formic acid in water). Proteins were directly eluted into the mass spectrometer at a flow rate of 60 μ L/min with 55% solvent B (acetonitrile (J.T. Baker, Philipsburg, USA)/water (9/1), 0.1% formic acid). Mass spectra were acquired on a Waters Synapt G2 mass spectrometer (Waters, Milford, MA, USA) equipped with a standard ESI source using the following instrument settings: capillary voltage, 3.0 kV; sampling cone, 35V; extraction cone, 4V; source temperature, 80°C; desolvation gas flow and

temperature, 600 L/h and 180°C, respectively; cone gas flow, 20 L/h; trap collision energy, 4V. Mass spectra were acquired in resolution mode (m/z 100-2000) and the calibration was performed in positive mode using a 2 mg/mL cesium iodide (Sigma) solution prepared in 50% isopropanol. The spectra were processed with MassLynx 4.1 software (Waters).

Native mass analysis

For native mass spectrometry, protein samples were buffer exchanged against 250 mM ammonium acetate (pH 8.0) using Zeba spin desalting columns with a 7-kDa molecular weight cut-off (Thermo Fisher Scientific Inc., Waltham, MA, USA). The concentration of each desalted protein was measured using the Bradford reagent. Samples were analyzed on a SynaptG2 HDMS mass spectrometer (Waters) equipped with a nanoelectrospray source. The instrument was calibrated in resolution mode (m/z 1000-7000) using a 100 mg/mL cesium iodide solution prepared in water and the quadrupole profile was adjusted to ensure the best transmission in the selected mass range. To preserve the integrity of noncovalent complexes in the gas phase, the instrument settings were carefully adjusted to the following values: capillary voltage, 1.5 kV; sampling cone, 80 V; extraction cone, 0 V; backing pressure, 5.2 millibars; cone gas, 20 L/h; source temperature, 35°C. Spectra were acquired in positive mode for 5 to 10 minutes to obtain a good signal-to-noise ratio and processed with MassLynx 4.1 software (Waters) with minimal smoothing. During in source fragmentation experiments the sampling cone voltage was increased up to 130 V. For MS/MS analysis of the ligand, the complex was first dissociated in the source

and the ligand selected for MS/MS analysis in the quadrupole. MS/MS was performed in the Trap region of the instrument using Argon as collision gas and a collision energy between 4V and 45V.

SDS poly-acrilamide gel electrophoresis (SDS PAGE)

Proteins were denatured for 5 min at 99 °C in SDS-PAGE sample buffer containing 2% (w/v) SDS and 50 mM dithiothreitol (DTT) (Sigma). Proteins were loaded onto 4-12% (w/v) acrylamide gels (BioRad, Hercules, USA). Gels were run in 3-(N-morpholino)propanesulfonic acid (MOPS) buffer (BioRad) and stained with colloidal Coomassie Blue G-250 (BioRad).

In-gel protein digestion and MALDI-ToF analysis

Stained bands were excised from the gels, washed with 50 mM ammonium bicarbonate (Fluka) / acetonitrile (J.T. Baker) 50/50 (vol/vol), washed once with pure acetonitrile and air dried. Dried spots were digested for 8 hours at 37°C in 12 µl of 0.012 µg/µl sequencing grade modified trypsin (Promega, Madison, USA), in 5 mM ammonium bicarbonate. After digestion, 0.6 µl were loaded on a matrix PAC target (Prespotted Anchorchip 96, set for Proteomics, Bruker Daltonics, Bremen, Germany) and air-dried. Spots were washed with 0.6 µl of a solution of 70% ethanol (J.T. Baker), 0.1% trifluoroacetic acid (Sigma). Mass spectra were acquired on an Ultraflex MALDI ToF-ToF mass spectrometer (Bruker Daltonics) in reflectron, positive and negative mode, in the mass range of 500 to 3500 *m/z*. Ions generated by laser desorption at 337 nm (N₂ laser) were recorded with an acceleration voltage of 25 kV in the

reflector mode. In general, about 200 single spectra were accumulated for improving the signal/noise ration and analyzed by FlexAnalysis (version 2.4, Bruker Daltonics). Monoisotopic peaks were annotated with FlexAnalysis default parameters and manually revised. Protein identification was carried from the generated peaklist using the Mascot program (Mascot server version 2.2.01, Matrix Science). Mascot was run on a MenB database containing protein sequences deduced from the sequenced *Neisseria meningitidis* serogroup B genomes, downloaded from NCBI. Search parameters were: variable modifications= Oxidation of methionine, cleavage by Trypsin (cleaves the C-term side of KR unless next residue is P), mass tolerance= 150 ppm, missed cleavage= 1, mass values= M+H⁺ monoisotopic. Known contaminant masses (trypsin, m/z = 842.5094, 1045.5637, 1165.5853, 1179.6010, 1300.5302, 1713.8084, 1716.8517, 1774.8975, 1993.9767, 2083.0096, 2211.1040, 2283.1802, 2825.4056) were excluded. Identifications were validated when the Mowse score was significant according to Mascot. If peptides matched to multiple of a protein family here is reported the protein identified as first hit (top rank) by Mascot.

For the SAL1486 protein and single domains D1, D2, D3 and D4, the stained bands were excised from gel, destained and *in-gel* digested in 5 mM ammonium bicarbonate with 12 µg/ml of modified Lys-C protease (Roche, Basel, Switzerland) overnight at 37 °C. Guanidination of the C-term lysine residues was performed using the ProteoMass Guanidination Kit (Sigma) following manufacturer's instructions. Modified and unmodified peptide solutions were directly spotted to a Prespotted AnchorChip MALDI target (Bruker Daltonics) and treated as previously described.

Opsonophagocytosis assay

The opsonophagocytosis assay was performed using GBS strains as target cells and HL-60 cell line (ATCC; CCL-240), differentiated into granulocyte-like cells, by adding 100 mM N, N dimethylformamide (Sigma) to the growth medium for 4 d. Midexponential bacterial cells were incubated at 37°C for 1 h in the presence of phagocytic cells, 10% baby rabbit complement (Cedarlane Labs, Burlington, Ontario, Canada), and heat-inactivated mouse antisera. Negative controls consisted of reactions either with preimmune sera, or without HL-60, or with heat-inactivated complement. The amount of opsonophagocytic killing was determined by subtracting the log of the number of colonies surviving the 1-h assay from the log of the number of CFU at the zero time point.

HDX analysis

The labeling was initiated by dilution of the proteins with 10-fold PBS 1x (pD 7.0) in 99.9% D₂O (Sigma). All exchange reactions were performed on ice. Over the time course of the experiment (spanning from 30 sec to 8.5 h), 15 µL of deuterated samples (55 pmoles) were removed and quenched with 35 µL of an ice-cold 200 mM Sodium Phosphate solution (pH 2.4) to lower the pH. The quenched samples were immediately frozen in dry ice and stored at -80°C for less than 24 h. Labeled samples were thawed rapidly to 0°C and injected into a Shimadzu LC-20ADXR Prominence HPLC system controlled by a CMB-20A module. The injector, switching valve, columns, solvents and all associated tubings were placed on ice to limit back-exchange. Protein samples were trapped and desalted for 2 min at a flow rate of 220 µL/min

using a Protein Micro Trap column (Michrom BioResources, Inc.) equilibrated with 100% buffer A (0.1% formic acid in water). The 2 min desalting step allows deuteriums located at fast exchanging sites (i.e. side chains and amino/carboxy terminus) to be replaced with hydrogens. Proteins were directly eluted into the mass spectrometer at a flow rate of 60 $\mu\text{L}/\text{min}$ with 55% solvent B (acetonitrile/water (9/1), 0.1% formic acid). Mass spectra were acquired on a Waters SynaptG2 mass spectrometer equipped with a standard ESI source using the instrument settings previously reported. Mass spectra were acquired in resolution mode (m/z 100-2000) and the calibration was performed in positive mode using a 2 mg/mL cesium iodide solution prepared in 50% isopropanol.

Bacterial strains and growth conditions

N. meningitidis MC58 Δgna33 mutant strain was grown in 200 mL GC culture medium (BD Biosciences, San Jose, CA, USA) in a humidified atmosphere containing 5% CO_2 until OD_{600} 0.6. Bacteria were collected by 10 min centrifugation at 3500 x g.

Carbonate Extraction

Bacteria were washed twice with PBS 1x and collected by 10 min centrifugation at 3500 x g. Cells were resuspended in 10 ml of PBS 1x and sonicated on ice. Any unbroken cells were discarded by centrifugation at 3500 x g for 10 min and the supernatant was retained for carbonate extraction. The supernatant was diluted 10 times with ice-cold 100 mM Sodium Carbonate solution (Sigma), pH 11 and slowly stirred for 1 h in an ice bath. The cell

membranes are collected by ultracentrifugation of the carbonate extraction solution (Beckman Coulter Inc., Brea, CA, USA). The supernatant was discarded and the membrane pellet was resuspended in 1 mL PBS 1x. The membrane fraction was finally collected by ultracentrifugation at 115000 x g for 20 min at 4°C, resuspended with PBS 1x and stored at -20°C.

OMVs preparation

The culture media recovered after the growth was filtered through a 0.22 mm pore size filter (Millipore, Bedford, MA, USA). The filtrates were subjected to ultracentrifugation (200000 x g, 180 min). The pellets constituting the m-OMVs were then resuspended with PBS and stored at -20°C.

In-solution digestion and Nano-LC/MS/MS analysis

The samples were denatured and reduced with Rapigest® (Waters) and 5 mM DTT at 100 °C, respectively, for 10 min. The pH was then adjusted to 8.0 using Ammonium Bicarbonate, and digested overnight with 2 µg of trypsin (Promega) at 37°C. The digestion reaction was stopped with formic acid at 0.1% final concentration. The peptide mixtures were then desalted using OASIS cartridges (Waters) following the manufacturer's protocol. Part of the digested sample was separated on 500 µl of SP Sepharose Fast Flow SCX resin (GE Healthcare, UK Limited) following manufacturer's instructions, and desalted using OASIS cartridges (Waters). Desalted peptides were concentrated with a Centrivap Concentrator (Labconco, Kansas City, KS) and kept at -20 °C until further analysis. Peptides were separated by nano-LC on a NanoAcquity UPLC system (Waters) connected to a Q-ToF Premier ESI

mass spectrometer equipped with a nanospray source (Waters). Samples were loaded onto a NanoAcquity 1.7- μm BEH130 C18 column (75 μm \times 25 mm; Waters) through a NanoAcquity 5- μm Symmetry® C18 trap column (180 μm \times 20 mm; Waters). Peptides were eluted with a 120-min gradient of 2–40% of solvent B (98% acetonitrile, 0.1% formic acid) at a flow rate of 250 nl/min. The eluted peptides were subjected to an automated data-dependent acquisition using the MassLynx software, version 4.1 (Waters) where an MS survey scan was used to automatically select multicharged peptides over the m/z ratio range of 300–2,000 for further MS/MS fragmentation. Up to eight different peptides were individually subjected to MS/MS fragmentation following each MS survey scan. After data acquisition, individual MS/MS spectra were combined, smoothed, and centroided using ProteinLynx, version 3.5 (Waters) to obtain the peak list file. Protein identification was carried from the generated peaklist using the Mascot program (Mascot server version 2.2.01, Matrix Science). Mascot was run on a MenB database containing protein sequences deduced from the sequenced *Neisseria meningitidis* serogroup B genomes, downloaded from NCBI. Search parameters were: variable modifications= methionine oxidation and glutamine and asparagine deamidation, cleavage by Trypsin (cleaves the C-term side of KR unless next residue is P), peptide mass tolerance= 0.3 Da, peptide MS/MS tolerance= 0.3 Da, missed cleavage= 2, mass values= $M+H^+$, $M+2H^+$ and $M+3H^+$. Only significant hits were considered as defined by the Mascot scoring and probability system. A computational analysis of each identified protein sequence was performed with the PSORTb version 2.0 to predict the subcellular localization. For PTMs discovery the unidentified spectra, after

MASCOT analysis, were manually interpreted searching for neutral losses or reporter ions.

5 References

- 1 Serruto, D. & Rappuoli, R. Post-genomic vaccine development. *FEBS Lett.* **580**: 2985–92 (2006).
- 2 Moylett, E.H. & Hanson, I.C. Immunization. *J. Allergy Clin. Immunol.* **111** (2): S754–65 (2003).
- 3 Pizza, M. *et al.* Subunit S1 of pertussis toxin: mapping of the regions essential for ADP-ribosyltransferase activity. *Proc. Natl. Acad. Sci. U. S. A.* **85**: 7521–25 (1988).
- 4 Fraser, C.M. & Rappuoli, R. Application of microbial genomic science to advanced therapeutics. *Annu. Rev. Med.* **56**: 459–74 (2005).
- 5 Fleischmann, R.D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512 (1995).
- 6 Kaushik, D.K. & Sehgal, D. Developing antibacterial vaccines in genomics and proteomics era. *Scand. J. Immunol.* **67**: 544–52 (2008).
- 7 Rappuoli, R. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* **19**: 2688–91 (2001).
- 8 Giuliani, M.M. *et al.* A universal vaccine for serogroup B meningococcus. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 10834–39 (2006).
- 9 Wizemann, T.M. *et al.* Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. *Infect. Immun.* **69**: 1593–8 (2001).
- 10 Montigiani, S. *et al.* Genomic approach for analysis of surface proteins in *Chlamydia pneumoniae*. *Infect. Immun.* **70**: 368–79 (2002).
- 11 Ariel, N. *et al.* Search for potential vaccine candidate open reading frames in the *Bacillus anthracis* virulence plasmid pXO1: in silico and in vitro screening. *Infect. Immun.* **70**: 6817–27 (2002).
- 12 Ross, B.C. *et al.* Identification of vaccine candidate antigens from a genomic analysis of *Porphyromonas gingivalis*. *Vaccine* **19**: 4135–42 (2001).
- 13 Chakravarti, D.N. *et al.* Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. *Vaccine* **19**: 601–12 (2000).
- 14 Betts, J.C. Transcriptomics and proteomics: tools for the identification of novel drug targets and vaccine candidates for tuberculosis. *IUBMB Life* **53**: 239–42 (2002).
- 15 Anfinsen, C. The formation and stabilization of protein structure. *Biochem. J.* **128** (4): 737–49 (1972).

- 16 Rucker, R.B. & Wold, F. Cofactors in and as posttranslational protein modifications. *FASEB J.* **2** (7): 2252-61 (1988)
- 17 Ali, M.H. & Imperiali B. Protein oligomerization: How and why? *Bioorganic & Medical Chemistry* **13**: 5013-20 (2005)
- 18 Ito, I. *et al.* Conformational difference in HMGB1 proteins of human neutrophils and lymphocytes revealed by epitope mapping of a monoclonal antibody. *J. Biochem.* **136** (2):155-62 (2004).
- 19 Sánchez, S. *et al.* Analysis of outer membrane protein complexes and heat-modifiable proteins in *Neisseria* strains using two-dimensional diagonal electrophoresis. *J. Proteome Res.* **4** (1):91-5 (2005).
- 20 Sánchez, S. *et al.* Cross-linking analysis of antigenic outer membrane protein complexes of *Neisseria meningitidis*. *Res. Microbiol.* **157** (2):136-42 (2006).
- 21 Sánchez, S. *et al.* Characterisation and immune responses to meningococcal recombinant porin complexes incorporated into liposomes. *Vaccine* **27** (39): 5338-43 (2009).
- 22 Sung, H. & Schleiss, M.R. Update on the current status of cytomegalovirus vaccines. *Expert Rev. Vaccines* **9** (11): 1303-14 (2010).
- 23 Cordwell, S.J. Exploring and exploiting bacterial proteomes. *Methods Mol Biol*, **266**: 115–135 (2004).
- 24 Zingaretti, C. *et al.* *Streptococcus pyogenes* SpyCEP: a chemokine-inactivating protease with unique structural and biochemical features. *FASEB J.* **24** (8): 2839-48 (2010).
- 25 Hedge, R.S. & Bernstein, H.D. The surprising complexity of signal sequences. *Trends Biochem. Sci.* **31** (10): 563-71 (2006).
- 26 Garnak, M. & Reeves, H.C. Purification and properties of phosphorylated isocitrate dehydrogenase of *Escherichia coli*. *J. Biol. Chem.* **254** (16):7915-20 (1979).
- 27 Grangeasse, C. *et al.* Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology. *Trends Biochem. Sci.* **32** (2): 86-94 (2007).
- 28 Hegge, F. T. *et al.* Unique modifications with phosphocholine and phosphoethanolamine define alternate antigenic forms of *Neisseria gonorrhoeae* type IV pili. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 10798–10803 (2004).
- 29 Abu-Qarn, M. *et al.* Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr. Opin. Struct. Biol.* **18** (5): 544-50 (2008).

- 30 Arora, S.K. *et al.* Role of motility and flagellin glycosylation in the pathogenesis of *Pseudomonas aeruginosa* burn wound infections. *Infect. Immun.* **73** (7): 4395-8 (2005).
- 31 Zhou, M. & Wu, H. Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. *Microbiology* **155** (Pt 2): 317-27 (2009).
- 32 Springer, M.S. *et al.* Protein methylation in behavioural control mechanisms and in signal transduction. *Nature* **280** (5720): 279-84 (1979).
- 33 Rezwan, M. *et al.* Lipoprotein synthesis in mycobacteria. *Microbiology* **153** (Pt 3): 652-8 (2007).
- 34 Kovacs-Simon, A. *et al.* Lipoproteins of bacterial pathogens. *Infect. Immun.* **79** (2): 548-61 (2011).
- 35 Romain, F. *et al.* Deglycosylation of the 45/47-kilodalton antigen complex of *Mycobacterium tuberculosis* decreases its capacity to elicit in vivo or in vitro cellular immune responses. *Infect. Immun.* **67** (11):5567-72 (1999).
- 36 Horzempa, J. *et al.* Immunization with a *Pseudomonas aeruginosa* 1244 pilin provides O-antigen-specific protection. *Clin. Vaccine Immunol.* **15**(4):590-7 (2008).
- 37 Fletcher, L.D. *et al.* Vaccine potential of the *Neisseria meningitidis* 2086 lipoprotein. *Infect. Immun.* **72** (4): 2088-100 (2004).
- 38 Hillenkamp, F. *et al.* Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chemistry* **63**: 1193-203 (1991).
- 39 Fenn, J.B. *et al.* Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246** (4926): 64-71 (1989).
- 40 Kebarle, P. & Verkerk, U.H. Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spec. Reviews* **28**:898-917 (2009).
- 41 Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom.* **11** (11): 601 (1984)
- 42 Sobott, F. *et al.* A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chemistry* **74** (6): 1402-7 (2002).
- 43 Lorenzen, K. & Van Dujin, E. Native mass spectrometry as a tool in structural biology. *Curr. Protoc. Protein Sci.* **62**: 17.12.1-17 (2010).
- 44 Sanglier, S. *et al.* Comparative ESI-MS study of approximately 2.2 MDa native hemocyanins from deep-sea and shore crabs: from protein oligomeric state to biotope. *J. Am. Soc. Mass Spectrom.* **14** (5):419-29 (2003).

- 45 Metruccio, M.M. *et al.* A novel phase variation mechanism in the meningococcus driven by a ligand-responsive repressor and differential spacing of distal promoter elements. *PLoS Pathog.* **5** (12): e1000710 (2009).
- 46 Cecchini, P. *et al.* The Soluble recombinant *Neisseria meningitidis* adhesin NadA_{Δ351-405} stimulates human monocytes by binding to extracellular Hsp90. *PLoS One* **6** (9): e25089 (2011).
- 47 Lowy, F.D. *Staphylococcus aureus* infections. *N. Eng.l J. Med.* **339**: 520-32 (1998).
- 48 Pizza, M. *et al.* Factor H-binding protein, a unique meningococcal vaccine antigen. *Vaccine* **26** Suppl 8:I46-8 (2008).
- 49 Lei, B. *et al.* Identification of new candidate vaccine antigens made by *Streptococcus pyogenes*: purification and characterization of 16 putative extracellular lipoproteins. *J. Infect. Dis.* **189** (1):79-89 (2004).
- 50 Liu, P. *et al.* Characterization of S-thiolation on secreted proteins from *E. coli* by mass spectrometry. *Rap. Com. Mass Spec.* **23**: 3343-9 (2009).
- 51 Vasilakos, J.P. *et al.* Adjuvant activities of immune response modifier R-848: comparison with CpG ODN. *Cell. Immunol.* **204** (1): 64-74 (2000).
- 52 Gotschlich, E.C. *et Al.* Human immunity to the meningococcus. IV. Immunogenicity of group A and group C meningococcal polysaccharides in human volunteers. *J. Exp. Med.* **129** (6):1367-84 (1969).
- 53 Gotschlich, E.C. *et Al.* Human immunity to the meningococcus. 3. Preparation and immunochemical properties of the group A, group B, and group C meningococcal polysaccharides. *J. Exp. Med.* **129** (6):1349-65 (1969).
- 54 Costantino, P. *et Al.* Development and phase 1 clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* **10** (10):691-8 (1992).
- 55 Tappero, J.W. *et Al.* Immunogenicity of 2 serogroup B outer-membrane protein meningococcal vaccines: a randomized controlled trial in Chile. *J. Am. Med. Assoc.* **281**:1520-7 (1999).
- 56 Tettelin, H. *et Al.* Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287** (5459): 1809-15 (2000).
- 57 Sisinni, L. *et Al.* *Helicobacter pylori* acidic stress response factor HP1286 is a YceI homolog with new binding specificity. *FEBS J.* **277**: 1896-905 (2010).
- 58 Handa, N. *et Al.* Crystal structure of a novel polyisoprenoid-binding protein from *Thermus thermophilus* HB8. *Protein Sci.* **14**: 1004-10 (2005).
- 59 Vincent, F. *et Al.* Structure of a polyisoprenoid binding domain from *Saccharophagus degradans* implicated in plant cell wall breakdown. *FEBS Lett.* **584**: 1577 (2010).

- 60 Padilla, S. *et Al.* Demethoxy-Q, An Intermediate of Coenzyme Q Biosynthesis, Fails to Support Respiration in *Saccharomyces cerevisiae* and Lacks Antioxidant Activity. *J. Biol. Chem.* **279** (25): 25995-6004 (2004).
- 61 Stancik, L.M. *et Al.* pH-Dependent expression of periplasmic proteins and amino acid catabolism in *Escherichia coli*. *Journal of Bacter.* **184** (15): 4246-58 (2002).
- 62 Weber, A. *et Al.* Time-dependent proteome alterations under osmotic stress during aerobic and anaerobic growth in *Escherichia coli*. *Journal of Bacter.* **188** (20): 7165-75 (2006).
- 63 Song, H.S. *et al.* Antioxidant effect of CoQ(10) on N-nitrosodiethylamine-induced oxidative stress in mice. *Korean J. Physiol. Pharmacol.* **13** (4): 321-6 (2009).
- 64 Johri, A.K. *et al.* Group B Streptococcus: Global incidence and vaccine development. *Nat. Rev. Microbiol.* **4**:932-42 (2006).
- 65 Baker, C.J. & Edwards, M.S. Group B streptococcal conjugate vaccines. *Arch. Dis. Child.* **88**:375-8 (2003).
- 66 Edwards, M.S. *et Al.* Group B streptococcal colonization and serotype-specific immunity in healthy elderly persons. *Clin. Infect. Dis.* **40**:352-7 (2005).
- 67 Margarit, I. *et al.* Preventing bacterial infections with pilus-based vaccines: The group B streptococcus paradigm. *J. Infect. Dis.* **199**:108-15 (2009).
- 68 Spraggon, G. *et al.* Supramolecular organization of the repetitive backbone unit of the *Streptococcus pneumoniae* pilus. *PLoS One* **5**: e10919 (2010).
- 69 Kang, H.J. *et al.* The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 16967-71 (2009).
- 70 Pickart, C.M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**: 503-33 (2001).
- 71 Hay, R.T. SUMO: a history of modification. *Mol. Cell* **18**: 1-12 (2005).
- 72 Marraffini, L.A. *et al.* Sortases and the art of anchoring proteins to the envelopes of Gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **70**: 192-221 (2006).
- 73 Mandlik, A. *et al.* Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol.* **16**: 33-40 (2008).
- 74 Ton-That, H. & Schneewind, O. Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol. Microbiol.* **50**: 1429-1438 (2003).
- 75 Duda, R.L. Protein chainmail: catenated protein in viral capsids. *Cell*: **94**: 55-60 (1998).

- 76 Kang, H.J. *et al.* Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* **318**: 1625–28 (2007).
- 77 Budzik, J.M. *et al.* Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 19992–97 (2009).
- 78 Nuccitelli, A. *et al.* Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B Streptococcus infections. *Proc. Natl. Acad. Sci. U. S. A.* **108** (25): 10278–83 (2011).
- 79 Liu, Q. *et al.* The *Enterococcus faecalis* MSCRAMM ACE binds its ligand by the collagen hug model. *J. Biol. Chem.* **282**: 19629–37 (2007).
- 80 Forsgren, N. *et al.* Two intramolecular isopeptide bonds are identified in the crystal structure of the *Streptococcus gordonii* SspB C-terminal domain. *J. Mol. Biol.* **397**: 740–51 (2010).
- 81 Oke, M. *et al.* The Scottish Structural Proteomics Facility: targets, methods and outputs. *J. Struct. Funct. Genomics* **11**: 167–80 (2010).
- 82 Hagan, R.M. *et al.* NMR spectroscopic and theoretical analysis of a spontaneously formed Lys-Asp isopeptide bond. *Angew. Chem. Int. Ed. Engl.* **49**: 8421–25 (2010).
- 83 Clarke, S.R. & Foster, S.J. Surface adhesins of *Staphylococcus aureus*. *Adv. Microb. Physiol.* **51**: 187–224 (2006).
- 84 Patti, J.M. *et al.* MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu. Rev. Microbiol.* **48**: 585–617 (1994).
- 85 Patti, J.M. & Hook, M. Microbial adhesins recognizing extracellular matrix macromolecules. *Curr. Opin. Cell. Biol.* **6**: 752–8 (1994).
- 86 Foster, T.J. & Hook, M. Surface protein adhesins of *Staphylococcus aureus*. *Trends Microbiol.* **6**: 484–8 (1998).
- 87 Konermann, L. *et al.* Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**: 1224–34 (2011)
- 88 Xiao, H. *et al.* Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I. A generalized model for a two-state protein and comparison with experiment. *Protein Sci.* **14** (2): 543–57 (2005).
- 89 Guerry, P. *et al.* Changes in flagellin glycosylation affect *Campylobacter* autoagglutination and virulence. *Mol. Microbiol.* **60** (2): 299–311 (2006).
- 90 Johnson, H. & Eyers, C.E. Analysis of post-translational modifications by LC-MS/MS. *Methods. Mol. Biol.* **658**: 93–108 (2010).
- 91 Adu-Bobie, J. *et al.* GNA33 of *Neisseria meningitidis* is a lipoprotein required for cell separation, membrane architecture, and virulence. *Infect. Immun.* **72** (4): 1914–9 (2004).

- 92 Leysen, S. et al. Molecular basis of bacterial defense against host lysozymes: X-ray structures of periplasmic lysozyme inhibitors PliI and PliC. *J. Mol. Biol.* **405** (5): 1233-45 (2011).
- 93 Klugman, K.P. et al. Sequence of the structural gene (rmpM) for the class 4 outer membrane protein of *Neisseria meningitidis*, homology of the protein to gonococcal protein III and *Escherichia coli* OmpA, and construction of meningococcal strains that lack class 4 protein. *Infect. Immun.* **57**: 2066-71 (1989).
- 94 De Mot, R. & Vanderleyden, J. The C-terminal sequence conservation between OmpA-related outer membrane proteins and MotB suggests a common function in both Gram-positive and Gram-negative bacteria, possibly in the interaction of these proteins with peptidoglycan. *Mol. Microbiol.* **12**: 333-4 (1994).
- 95 Koebnik, R. Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins. *Mol. Microbiol.* **16**: 1269-70 (1995).
- 96 McDade, R.L. & Johnston, K.H. Characterization of serologically dominant outer membrane proteins of *Neisseria gonorrhoeae*. *J. Bacteriol.* **141**: 1183-91 (1980).
- 97 Jansen, C. et al. Biochemical and biophysical characterization of in vitro folded outer membrane porin PorA of *Neisseria meningitidis*. *Biochim. Biophys. Acta.* **1464**: 284-98 (2000).
- 98 Prinz, T. & Tommassen, J. Association of iron-regulated outer membrane proteins of *Neisseria meningitidis* with RmpM (class 4) protein. *FEMS Microbiol. Lett.* **183**: 49-53 (2000).
- 99 Sonntag, I. et al. Cell envelope and shape of *Escherichia coli*: multiple mutants missing the outer membrane lipoprotein and other major outer membrane proteins. *J. Bacteriol.* **136**: 280-5 (1978).

Vorrei ringraziare la Dott.ssa Nathalie Norais per avermi dato la possibilità di svolgere l'attività di ricerca nei laboratori dell'unità di Spettrometria di Massa Strutturale e Proteomica, ed il Prof. Vincenzo Scarlato per il sostegno datomi durante i tre anni di dottorato.

Ringrazio Chiara Tani, Massimiliano Biagini, Agnese Faleri e Manuela Garibaldi per avermi supportato e sopportato durante tutto il periodo di dottorato, dentro e fuori il laboratorio.

Un ringraziamento speciale va a Sebastièn Brier, che mi ha permesso di crescere professionalmente condividendo con me la sua esperienza, ed alla mia ragazza Antonina Torre, che è stata sempre al mio fianco.