Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biologia Cellulare e Molecolare

Ciclo XXXII

Settore Concorsuale: 05/E2

Settore Scientifico Disciplinare: BIO/11- Biologia molecolare

TITOLO TESI

Development of Mass Spectrometry-based analytical tools for characterization of bioconjugate vaccines in *E. coli*

Presentata da: Anna Galea

Coordinatore Dottorato

Prof. Giovanni Capranico

Supervisore

Prof. Vincenzo Scarlato

Dott.ssa Nathalie Norais

Esame finale anno 2020

Abstract

Antigen design is generally driven by the need to obtain enhanced stability, efficiency and safety in vaccines (Liljeroos *et al.*, 2015). Unfortunately, the antigen modification is rarely proceeded in parallel with analytical tools development for its characterization.

However, the analytical tools set up is required during several steps of vaccine manufacturing pipeline, as for vaccine production modifications and/or improvements or regulatory requirements (Lepenies, 2015b). Despite the relevance of bioconjugate vaccine development, robust and consistent analytical tools to evaluate the extent of carrier glycosylation are still missing (Sharma *et al.*, 2018).

Bioconjugation is a glycoengineering technology aimed to produce N-glycoprotein in vivo in *Escherichia coli* cells, based on the PglB-dependent system by *Campylobacter jejuni* (Kowarik et al., 2006a; Kowarik et al., 2006b; Wacker et al., 2002) (Cuccui *et al.*, 2013). (Wacker *et al.*, 2002) (Cuccui *et al.*, 2013; Kowarik *et al.*, 2006b). (Szymanski *et al.*, 1999).

It has recently found application in the delivery of a series of glycoconjugate vaccine candidates (Cuccui *et al.*, 2013; Cuccui & Wren, 2015; Garcia-Quintanilla *et al.*, 2014; Harding & Feldman, 2019; Harding *et al.*, 2019; Langdon *et al.*, 2009; Ravenscroft *et al.*, 2019; van den Dobbelsteen *et al.*, 2016; Wacker *et al.*, 2014; Wetter *et al.*, 2013). This wide applicability is due to glycocompetent *E. coli* cells ability to produce site-selective glycosylated protein which could be used, after few purification steps, as vaccines molecules in order to elicit both humoral and cell-mediate immune-response (Cuccui *et al.*, 2013; Langdon *et al.*, 2009).

In this work, *Staphylococcus aureus* Hla bioconjugated with type 5 capsular polysaccharide was used as proof of concept to perform rational analytical driven design of the glycosylation sites, suitable for the quantification of glycosylation extent by Mass Spectrometry.

The aim of the study was to develop a MS-based approach to quantify the extent of glycosylation for in-process monitoring of the bioconjugate production and for the final product characterization.

The three designed consensus sequences differ for a single amino acid residue and fulfilled the prerequisites to obtain engineered bioconjugate vaccines more appropriate from an analytical perspective. In details, we aimed to achieve an optimal MS detectability of the peptide carrying the consensus sequences, complying with the well-characterized minimal requirements for N-glycosylation by PglB (Kowarik *et al.*, 2006b; Picotti & Aebersold, 2012).

Hla carrier protein isoforms, bearing these consensus sequences, allowed a recovery of about 20 ng/µg of periplasmic proteins glycosylated at 40%.

The SRM MS here developed was successfully applied to evaluate the differential site occupancy when two consensus sequences are inserted in the carrier protein. In each glycosylation site, the extents of glycosylation were determined and ranging from 16 to 100%. The glycosylation extent of

carrier protein was influenced either by the overall source of protein produced and by the position of glycosite insertion.

The analytical driven design of the bioconjugated antigen and the development of accurate, precise and robust analytical method allowed to finely characterize the vaccine candidate.

Authors have declared the following interests: A. Galea is a PhD student at the University of Bologna and participate in a post graduate PhD studentship program sponsored and funded by GlaxoSmithKline Biological SA. N. Norais is employee of the GSK group of companies. V. Scarlato is the university supervisor at University of Bologna.

Index

Introduction

1. Glycoconjugate vaccines

1.1 Carbohydrate based vaccines

Bacterial and viral cell surfaces are decorated with diverse components among which carbohydrates are displayed under a huge variety of glyco-forms such as linked to protein, the glycoproteins, or linked to lipids, the glycolipids. Peptidoglycan, teichoic acid, capsular polysaccharides (CPS) and lipopolysaccharide (LPS) are some examples (La Placa, 2012).

LPS is a distinctive glycolipid of all Gram-negative bacteria covering the bacterial surface formed by a constant portion (antigen R) linked to a polysaccharide chain called O-antigen. The O-antigen differs along and within different bacterial species (Whitfield & Trent, 2014).

By contrast, Gram-positive bacteria are encapsulated with large molecular-weight surface polysaccharides called capsular polysaccharides. CPS consist of repeating units with extremely diverse structures, variable in the length of the repeating units, the composition in monosaccharides, the stereo-chemistry of glycosidic linkages, and the presence of branching points or glycan modifications (*i.e.* sulfation, acetylation, and phosphorylation) (Roberts, 1996).

CPS are covalently attached to the outer most layer of the bacterial cell, while LPS is anchored to outer membrane of the bacterial cells with the lipidic portion located within the outer membrane bilayer and the O-antigen exposed to the external (Roberts, 1996).

Therefore, O-antigen and the CPS are the first, and most abundant, microbial components encountered by the immune system during infection, because their location on the bacterial surface at the interface between the pathogen and the host. (Comstock & Kasper, 2006)

Moreover, since glycan epitopes are often highly pathogen-specific, they have been used as antigens for vaccine development in carbohydrate-based vaccines, playing fundamental roles in host-pathogen interactions and in the immune response (Dwek, 1996).

The history of carbohydrate-based vaccines dates back to 1923 when Heidelberger and Avery identified the carbohydrate nature of the pneumococcal capsule derived from *Streptococcus pneumoniae* and showed that *S. pneumoniae* CPS were immune reactive components (Heidelberger & Avery, 1923). Despite this key discovery, the introduction of antibiotics in the same period slow down the development of carbohydrate vaccines as they were preferably used for the treatment of infectious diseases rather than preventing infection by vaccination.

Between 1970s and the early 1980s, carbohydrate-vaccines were licensed against meningococcus, pneumococcus and *Haemophilus influenzae* type b. At the same time, Peltola *et al.* elucidated the reason why these vaccines failed to elicit adequate protection in high-risk groups, such as infants and children under 2 years of age, although they were efficacious in adults (Peltola *et al.*, 1977a; Peltola

et al., 1977b). Indeed, when polysaccharides are used alone in vaccine formulations, they usually act as T-cell independent antigens that activate B cells in the absence of T cell help. As a consequence, the primary immune response to carbohydrates does not stimulate immunoglobulin class switching from IgM to IgG isotypes and does not lead to the generation of T and B cell memory, although can stimulate existing memory B cell pools in pre-exposed adults (Comstock & Kasper, 2006)..

The poor immune response of polysaccharide vaccines has been solved by covalently attaching a polysaccharide to a protein carrier in a process known as conjugation (Avery & Goebel, 1929). Indeed, Avery and Goebel demonstrated that the "chemical union" of small saccharides to proteins results in enhanced immunogenicity of the saccharide and that saccharide specific antibody concentrations increase after reinjection of the conjugate (Avery & Goebel, 1929). Moreover, the first preclinical and clinical reports on glycoconjugate vaccines were about Hib PS–protein conjugates by the team of John Robbins at NIH (Schneerson *et al*., 1980) that demonstrated how glyco-conjugate vaccine triggers a T-cell-dependent immune response assuring efficacious vaccination of children and elderly (Lepenies, 2015b).

1.2 Glycoconjugate vaccines

Glycoconjugate vaccines have been used to control a variety of bacterial infections in recent years, and more vaccines are either under development at preclinical level or in clinical trials (Rappuoli *et al.*, 2019). The glycoconjugate vaccines licensed so far are obtained from CPS or derived fragments (Micoli *et al.*, 2018b). Traditionally, glycoconjugate vaccines are synthesized using a semi-synthetic approach where the polysaccharide is extracted from the target bacterium, purified, chemically modified and covalently linked to a carrier protein.

The classical approach used for glycoconjugate vaccines is based on polysaccharide extraction from bacterial fermentation, purification and chemical activation at random sites, or at reducing ends, prior the conjugation to the carrier protein. The carrier protein is also purified from bacteria by fermentation and subsequent purification and, depending on the chemistry used, it can be conjugated via its functional groups or, alternatively, it can be derivatized prior the linkage to the saccharide. Furthermore, in order to reduce the steric hindrance present between protein and saccharide, suitable linkers can facilitate the conjugation (Micoli *et al.*, 2018b). This approach has resulted in the commercial licensure of multiple glycoconjugate vaccines to prevent colonization and infection by *H. influenzae* type B, and multiple serotypes of *Streptococcus pneumoniae* and *Neisseria meningiditis* (Harding & Feldman, 2019)

Although chemically conjugated vaccines had successful application, their manufacturing processes are not without drawbacks:

i) batch to batch heterogeneous product formation due to the handling processes of the components;

ii) the need of large scale production of pathogenic organisms from which extract the bacterial components;

iii) the high manufacturing costs deriving from all these steps (Frasch, 2009) .

The key aspects to be evaluated in glycoconjugate vaccines development are the selection of the carrier protein and the antigen-processing operated by the immune system, as well as the choice of the glycan portion which is almost dictated by epidemiological studies on pathogens spread and their pathogenicity (Micoli *et al.*, 2018b).

In the last years, the synthetic approach was applied to produce the minimal rationally designed carbohydrate epitopes needed to elicit an efficient immune response (Verez-Bencomo *et al.*, 2004). Moreover, the production of glycoconjugate vaccines using peptides known to elicit a strong immune response (called MHC peptides) and site-selective conjugation appears to be appealing for the development of vaccines (Hu *et al.*, 2016). Recently, a novel technology based on the protein glycan coupling technology (PGCT or Bioconjugation technology), has been proposed as a promising approach for vaccine development (Broker *et al.*, 2017; Cuccui & Wren, 2015; Harding & Feldman, 2019; Kay *et al.*, 2019; Langdon *et al.*, 2009).

1.2.1 Criteria for selection of carrier proteins

Carrier proteins that are currently used for licensed conjugate vaccines include bacterial products such as tetanus toxoid (TT), diphtheria toxoid (DT), a non-toxic cross-reacting mutant of DT (CRM_{197}), protein D from non-typeable *H. influenzae*, the outer membrane protein complex of *N. meningitidis* type B (OMPC), and keyhole limpet hemocyanin (KLH) (Jones, 2005). These proteins can be isolated from the bacteria cultures and chemically detoxified or produced as nontoxic mutants by recombinant DNA techniques in *heterologous organisms* (*i.e. E. coli*).

The main aspects to evaluate for the selection of carrier protein are safety and large-scale manufacturability. In particular, toxic or enzymatic activity should be removed from the target protein, preferably by selective amino acids genetic detoxification. A typical example is CRM_{197}, a non-toxic mutant of the diphtheria toxin where a single amino acid substitution in the enzymatic site allowed to extensively employ this protein as carrier (Broker *et al.*, 2011). In addition, a carrier protein for glycoconjugate vaccine production might be suitable for large-manufacturability, allowing the availability of large amount of material at high level of purity, according to cGMP quality requirements, and reasonable production costs (Micoli *et al.*, 2018a) (Hu *et al.*, 2016).

Furthermore, extensive preclinical studies of the glycoconjugates are required to confirm the immunogenicity in comparison to a bench mark carrier (Micoli *et al.*, 2018a).

Thus, immunogenic proteins that are safe, stable and available in high amounts and purity level are considered suitable carrier (Micoli *et al.*, 2018a).

Recently, new carrier proteins have been investigated at preclinical level to be used as carrier for glycoconjugation or bioconjugation. Some examples are nontoxic recombinant α-toxin (Hla) and clumping factor (Clf) proteins from *S. aureus*, or recombinant nontoxic Exotoxin A (EPA) from *Pseudomonas aeruginosa*. In particular, these proteins have been investigated since has been demonstrated that TT, DT and CRM_{197} based glycoconjugates elicit also antibodies against the respective related carrier proteins (Broker *et al.*, 2017). In particular, it has been studied that pre-exposure or co-exposure to a given carrier can interfere with the anti-carbohydrate immune response, such as carrier priming or carrier-induced epitopic suppression (CIES), where the effect is a reduced immunogenicity reducing the against the conjugated polysaccharide (Pobre *et al.*, 2014).

Therefore, the selection of a carrier protein for glycoconjugation which could play the dual role of carrier and protective antigen might be a benefit for the formulation of vaccines targeting both carbohydrate and protein virulence factors of the same pathogen. For example, Hla bioconjugated with *S. aureus* type 5 CPS (CP5) has been shown by Wacker and colleagues to elicit in rabbits and mice immune response with antibody produced against both the glycan and the protein moiety with protective activity (Wacker *et al.*, 2014).

The Hla-CP5 bioconjugate described by Wacker *et. al* (Wacker *et al.*, 2014) was used as a Proof of Concept to achieve the purpose of the present study, as discussed in the Results and Discussion section of the present study.

1.2.2 Immunological perspective of glycoconjugate

The human immune system is a complex network of cells, organs and soluble mediators with the ability to protect the body against a wide range of pathogens, foreign molecules, or tumor cells. In vertebrates, the immune system is divided into the innate and the adaptive immune system (Medzhitov & Janeway, 1997). The innate immune system is evolutionarily ancient and represents a rapid and stereotyped response to a large but limited number of stimuli. In contrast, adaptive immunity requires more time for induction but it is highly antigen-specific. A key function of adaptive immunity is the development of an immunological memory that provides protection from reinfection with the same pathogen. Both arms of immunity strongly cooperate and enable the recognition and destruction of pathogens (Lepenies, 2015a).

Two models have been proposed to explain the mechanism by which glycoconjugates engage T cells (Fig. 1) (Comstock & Kasper, 2006).

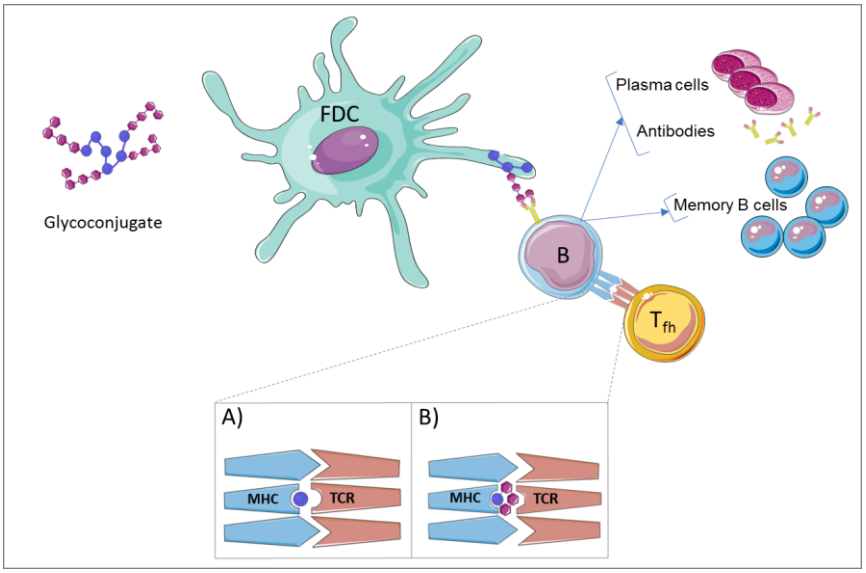


Figure 1: Models proposed to display the mechanisms of T cells engagement by glycoconjugate vaccines molecules. A) MHC molecule presents the peptide as epitope to engage the T-cells repertoire; B) glycopeptides are presented as epitopes where the glycans are recognized by the TCR. FDC: Follicular dendritic cells; B: B cells; Tfh: T-cells follicular helper; MHC: Major Histocompatibility Complex; TCR: T-cell Receptor. Adapted from ***(Rappuoli et al., 2019)***

In one model (A), the peptides derived from the carrier protein play as T cell epitopes for T-cells which recognize the protein, with the final result that T-cells repertoire has been recruited (**Fig.** 1A). In the second model (B), it has been proposed that the glycoconjugate may also be processed into glycopeptides functioning as epitopes where the peptidic part binds to the MHC molecules and its covalently linked sugar part could be recognized by T-cells (Fig**.** 1B) (Avci & Kasper, 2010).

It has been suggested the existence of specific T-cells recognizing carbohydrates, called T-carbs, since specific clones has been isolated by Avci and colleagues during immune response induced by glycoconjugate (Avci *et al.*, 2011). Although the mechanism of action of conjugate is not yet well-defined, it has been demonstrated that the structure of the polysaccharide is critical to the mechanisms involved in the antigen presentation to the T-cells (Sun *et al.*, 2019).

Accondingly, since the primary role of the carrier protein in glycoconjugate vaccines is to provide T-cell epitopes, it is fundamental to preserve the key T and B cell-epitopes of the carrier protein during the conjugation process. As a consequence, a rational design of the glycoconjugate vaccine molecules is required to address the impact of the saccharide chain on the protein folding, the size and the location of the saccharide chain and the protein to sugar ratio, in order to manage the influence of the glycoconjugation on the vaccine efficacy. It is well documented that glycosylation can profoundly

affect a protein intrinsic properties such as the conformations, protease stability, antigenicity, and immunogenicity (Petrescu *et al.*, 2006).

In this context, the need of analytical techinques suitable to map both protein and carbohydrate epitopes as well as techniques for glycoconjugate molecules characterization is of increasing interest in vaccine development (Micoli *et al.*, 2018a).

2. Bioconjugate vaccines

2.1 N-glycosylation system from *C. jejuni*

The post-translational modification of proteins with glycan structures was for many years thought to be a feature unique to eukaryotic organisms, but starting from 1990s there have been an increasing number of reports about N-linked and O-linked glycosylation pathways in bacteria.

While for O-glycosylation the glycans are linked to the Serine (S) or Threonine (T) residues on the carrier protein sequence, for N-glycosylation the modification take place on specific consensus sequences on the carrier protein, initially defined as N-X-S/T, where the N is the asparagine residue target of the glycan linkage and X could be any amino acid except Proline (P) (Kowarik *et al.*, 2006b).

N-glycosylation in bacteria has been described for the first time in 1999 by Szymanski and her group, evidenced from the genome sequencing of Gram-negative pathogen *C. jejuni* (Szymanski *et al.*, 1999). In the genetic locus 81-176, that resembled a coding region for polysaccharide biosynthesis, an open reading frame encoding for an oligosaccharyltransferase (OTase) has been found, that is able to modify asparagines residues within a protein by a β-glcyosidic linkage with a saccharide chain. The region was named *pgl* for protein glycosylation locus, and the corresponding product, PglB (Cuccui & Wren, 2015; Szymanski *et al.*, 1999; Szymanski & Wren, 2005).

Since then, a wide number of studies has been performed to better understand the *pgl* pathway and the potential applications of such N-glycosylation process in bacterial cells.

The evidence of the glycopeptide N-linkage nature has been obtained demonstrating that the glycan release did not occur when the glycopeptides are treated to obtain the β-elimination of the O-linked glycan, suggesting the linkage of the sugar via a glycosyl amide to an asparagine residue (Young *et al.*, 2002). Additionally, MS/MS collision-induced dissociation of the glycopeptide confirmed that the oligosaccharide was N-linked (Young *et al.*, 2002).

Extensive genetic and biochemical studies have been performed to identify the mechanism of action of PglB, revealing that the substrate saccharide of PglB is assembled on the lipid carrier undecaprenyl-pyrophosphate (UndPP), present in the inner membrane (Linton *et al.*, 2005; Wacker *et al.*, 2002). Such complex formed by UndPP-linked oligosaccharide (or lipid-linked oligosaccharide, LLO) is part of the conservative pathway of bacteria already known to take place for

the biosynthesis of all cell wall polysaccharides: the LPS-O-antigen (Wright *et al.*, 1965), the enterobacterial common antigen (ECA), the Peptidoglycan, Wall teichoic acid (WTA) and the CPS (Manat *et al.*, 2014).

Moreover, the most studied system for CPS biosynthesis is *E. coli*, where the main pathways described require the polysaccharides building on the UndPP in cytoplasmic side, starting from activated building blocks or antigen repeating units, and then the exporting across the membrane by a flippase (Wzx) or an ABC transporter (Whitfield, 2006). Once in the periplasm, the LLO are finally polymerized by a Wzy polymerase, which extends the growing chain one repeat unit at a time at the periplasmic face of the membrane (Whitfield, 2006). Here, PglB acts to transfer the glycan moiety to the asparagine residue of the consensus sequence on an acceptor protein (Glover *et al.*, 2005).

Fig.2 reports a schematic illustration of N-glycosylation system operated by *pgl* gene cluster.

The flipping mechanism for the translocation of LLO across the membrane bilayer is well characterized and the x-ray crystal structure of the flippase in distinct states has been solved (Perez *et al.*, 2015).

Moreover, the structure of PglB has been solved and its molecular mechanism of action has been elucidated (Lizak *et al.*, 2011; Napiorkowska *et al.*, 2017).

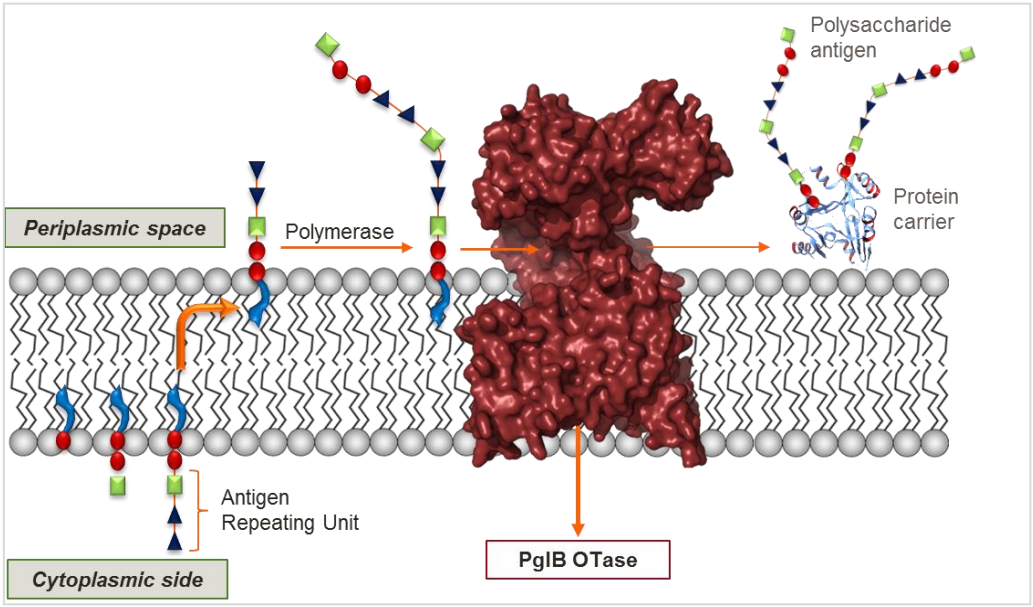


Figure 2: Schematic illustration of PglB N-glycosylation process on bacterial cells. PglB OTase in ruby red surface representation structure with PDB code 3RCE. Adapted from (Feldman et al., 2005)

2.2 N-glycosylation in *E. coli*

In 2002, Aebi and coworkers demonstrated that the *pgl* gene cluster can be functionally transferred into the model organism *E. coli*, where two *C. jejuni* periplasmic proteins, AcrA and PEB3, has been shown to be glycosylated. This work confirmed that the cluster contains all the genes needed for the

N-glycosylation machinery expression and functioning (Wacker *et al.*, 2002). Moreover, by glycoengineering *E. coli* it has been possible to explore the roles of the genes contained in the locus, to study PglB mechanism of action and to investigate the determinants of the acceptor consensus sequence as well as the features of the saccharide chain. Below, major achievements in the understanding of the enzyme mode of action are reported.

AcrA and PEB3 are predominantly annotated as periplasmic proteins, suggesting that the glycosylation machinery acts within the periplasmic space. Indeed, when the periplasmic signal sequence has been removed in an AcrA mutant, it has been shown as it results to be not glycosylated (Nita-Lazar *et al.*, 2005).

N-glycosylation is a site-selective modification of the proteins and it has been demonstrated that glycosylation operated by PglB requires stringent rules regarding the target consensus sequence (Kowarik *et al.*, 2006b). Indeed, considering as position 0 the N in the consensus sequence, a negatively charged amino acid residue (aspartate, D or glutamate, E) should be present in position -2 for an efficient N-glycosylation. Moreover, the presence of the Proline (P) in position -1 and/or in +1 inhibits the glycosylation, indicating the importance of peptide conformation in the glycosylation process; finally, in position +2, should be present an S or a T. The definition of these determinants for the consensus sequence has been obtained by mutational and bioinformatic analyses performed on native *C. jejuni* glycoproteins, as well as on non-campylobacter and heterologous proteins engineered to contain a glycosylation sites, expressed in "glycocompetent" *E. coli* (Kowarik *et al.*, 2006b). Glycocompetent *E. coli* strains have been obtained by the expression of plasmids bearing the *pgl* gene cluster.

The consensus sequence D/E-X-N-X-S/T, where X could be any amino acid except P, has been defined as the minimal amino acyl acceptor sequence required to achieve an efficient N-glycosylation by PglB (Kowarik *et al.*, 2006b; Nita-Lazar *et al.*, 2005).

The presence of such consensus sequence is required but not sufficient to achieve N-glycosylation of protein by PglB, since the structural conformation of the acceptor protein might be considered. The consensus sequence should be located within specific positions along the protein sequence, to be at once accessible to PglB enzyme and in domains sufficiently flexible to admit the steric hindrance of the saccharide chain and to allow the conformational changes of the enzyme (Kowarik *et al.*, 2006b; Nita-Lazar *et al.*, 2005).

The importance of protein folding state is supported by the analyses on the PglB glycosylation efficiency evaluated on a subset of native *C. jejuni* glycoproteins harboring the consensus sequence in structured regions (Silverman & Imperiali, 2016).

Significantly, when PEB3 protein is expressed with the signal sequence for Sec-pathway translocation (ssPelB), which is the classical pathway for translocation of unfolded protein in periplasm (De Geyter *et al.*, 2016), it is more efficiently glycosylated than when expressed with the signal peptide for TAT translocation (ssTorA) (De Geyter *et al.*, 2016), which drives the translocation of fully folded and oligomeric complex proteins in the periplasm. Moreover, these analyses suggested that a partial unfolding state is required for an efficient N-glycosylation, and therefore that the N-glycosylation is a co-translational process occurring during translocation into the periplasm, hence before the complete folding of the protein (Silverman & Imperiali, 2016), as in eukaryotic cells (Whitley *et al.*, 1996). In fig. 3 is reported a schematic representation.

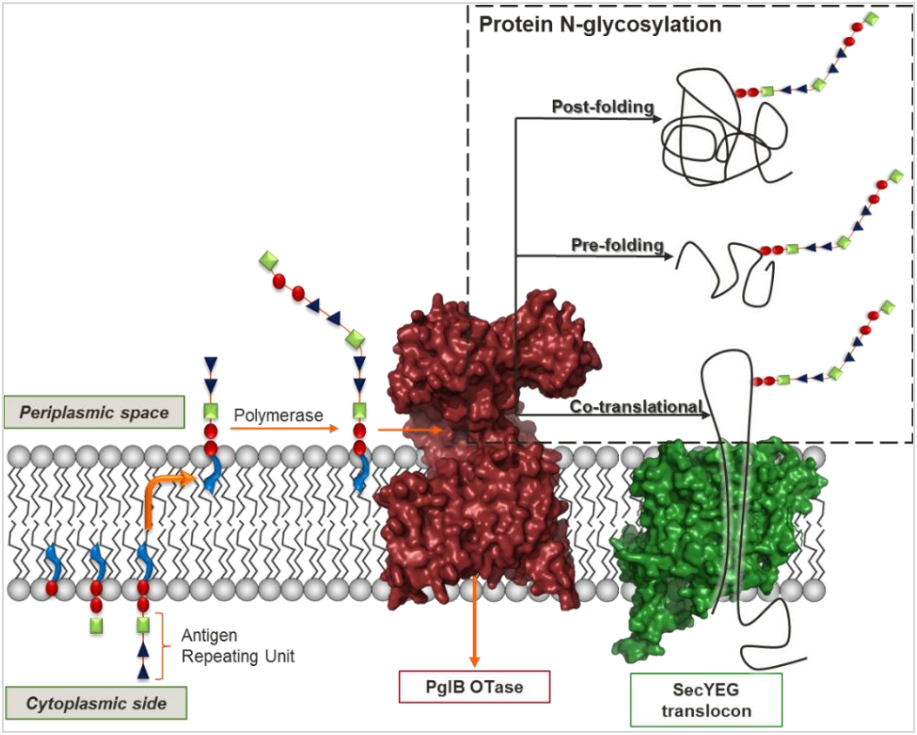


Figure 3: schematic illustration of the possible conformational states of substrate proteins during N-glycosylation process operated by PglB. In surface representation are shown in ruby red, PglB enzyme (PDB code 3RCE), in forest green the SecYEG translocon structure (PDB code: 5AWW). Adapted from (Silverman & Imperiali, 2016).

By contrast, Kowarik *et al*. and Fisher *et. al* have shown that PglB is able to glycosylate also folded protein, as long as the consensus sequence is sufficiently solvent-exposed and in a region of suitable flexibility to allow the sugar chain steric hindrance and the conformational changes of the enzyme (Fisher *et al.*, 2011; Kowarik *et al.*, 2006a).

The resolution of the PglB structure in functional state allowed a better understanding the molecular mechanism of action of the N-glycosylation process. The structure revealed a specific external loop, EL5, that has the dual role of binding peptide at C-terminal segment and LLO at N-terminal segment. The key $_{481}$DGGK motif in this external loop is near the active site of the enzyme, whose G_{482} residue

forms a hydrogen bond with the acceptor asparagine while G_{483} shows van der Waals interactions with the LLO moiety. Accordingly, this motif is involved in the alignment of the moieties for the PglB-catalyzed glycosylation thanks to the dynamic engagement and disengagement of the EL5 flexible loop in the catalytic site (Lizak *et al.*, 2011; Napiorkowska *et al.*, 2017). These data upon the structural features of PglB provided further evidence to support the requirements in terms of amino acidic residues of the consensus sequence suitable for N-glycosylation by PglB, already established by Kowarik and co-workers in 2006 (Kowarik *et al.*, 2006b).

Concerning the sugar moieties, PglB shows relatively relaxed substrate specificity on the LLO substrate since a wide variety of saccharides with different lengths and different saccharide subunits could be transferred from the UndPP-linked state to the proteins, with preference of hexose containing an acetamido group in position C2 (Wacker *et al.*, 2006). In details, when the ligase of the O-antigen (waaL) has been substituted by PglB in *E. coli*, polysaccharide chains have been transferred to the acceptor protein (Feldman *et al.*, 2005).

2.3 *In vivo* production of glycoconjugate: bioconjugate vaccines

The engineering of *E. coli* to produce recombinant glycoproteins has been a notable technological advance especially for glycoconjugate vaccines development (Cuccui *et al.*, 2013). Potentially, any non-campylobacter protein engineered to contain an accessible D/E-X-N-X-S/T sequon, expressed in *E. coli*-PglB expressing strains, can be converted to a glycoprotein by directing its secretion in the periplasm, where N-glycosylation occurs.

This pioneering technique has been employed to produce glycoconjugate molecules to be used as vaccines, called Protein Glycan Coupling Technology (PGCT) or Bioconjugation (Cuccui *et al.*, 2013).

Such technology has been developed by the Glycovaxyn swiss group (now Limmatech Biologics, part of GSK) to produce several glycoconjugate molecules, initially for vaccines against Gram-negative pathogens as *Salmonella enterica*, *Shigella spp* and *E. coli* by using as carrier protein exotoxin A (EPA) from *P. aeruginosa* (van den Dobbelsteen *et al.*, 2016; Wacker *et al.*, 2014; Wetter *et al.*, 2012; Wetter *et al.*, 2013). Successively, capsular polysaccharides from Gram-positive *S. aureus* serotype 5 and 8 conjugated to the EPA carrier were also shown to be immunogenic in animal models. The bioconjugation of *S. aureus* CP5 to the Hla antigen, also has been demonstrated to be protective against invasive staphylococcal diseases in mice (Wacker *et al.*, 2014).

Feldman and coworkers developed several glycoconjugate vaccine molecules by exploiting PglB conjugation of polysaccharides from *Burkholderia pseudomallei* to the *Campylobacter*-related

protein AcrA, whose has been able to induce protection in mice model (Garcia-Quintanilla *et al.*, 2014). Other applications of PGCT resulted in glycoconjugate molecules with enhanced immunogenicity in mice, as for *Francisella tularensis* O-antigen conjugated to EPA carrier protein described by Cuccui *et al.* in mice (Cuccui *et al.*, 2013) and for *S. pneumoniae* serotype 4 CPS linked to the NanA, PiuA, and Sp0148 proteins in a pneumococcal conjugate vaccine tested by Reglinski *et al.* (Reglinski *et al.*, 2018).The most recently work by Feldman and coworkers successfully obtained a bioconjugate vaccine against *S. pneumoniae* CPS containing glucose as the monosaccharide reducing end from multiple serotypes (Harding *et al.*, 2019).

The production of glycoconjugate by PGCT allowed to overcome several limitations of the classical chemical conjugation approach.

Firstly, by exploiting *E. coli* to produce polysaccharide chain as well as carrier/antigen protein, the handling of pathogen and the biosafety issues to produce the vaccines components has been extremely reduced. About vaccine components, a relatively restricted number of them need to be produced and purified since the whole enzymatic reaction to obtain the vaccine molecule takes place *in vivo* in *E. coli* cells. Finally, the downstream steps for vaccine preparation to achieve the final drug substance used in the formulation are reduced as well as batch-to-batch variations, resulting in an overall decrease of the number of analytical controls, times and costs of the production (Cuccui *et al.*, 2013; Frasch, 2009; Hu *et al.*, 2016). In particular, such site-directed conjugation allowed to obtain vaccines molecules with more defined physiochemical characteristics in comparison to the glycoconjugate produced by the chemical approach, resulting in the improvement of quality controls during the process development, and the reduced number of routine controls for product release (Frasch, 2009).

From the antigen design standpoint, an important advantage is the possibility to engineer heterologous carrier protein to contain a consensus sequence (glyco-sites) optimal for the glycosylation by PglB. In particular, such glyco-sites can be inserted in carrier protein domains chosen to avoid instability in the protein structure and potential interference in the antigenicity of the carrier protein, respecting the dual role of antigen and carrier of the target protein.

The plethora of studies performed on the structure, function and regulation of PglB provides strong knowledge helpful to optimize the efficiency of the enzyme by engineering carrier protein consensus sequences or to modify the catalytic site of the enzyme to improve the catalytic efficiency. Moreover, understanding the glycan transfer process might improve the substrate specificity of the enzyme expanding the related applications (Ihssen *et al.*, 2015; Napiorkowska *et al.*, 2017).

One of the limitations of bioconjugation technology has been for many years the substrate specificity of PglB, which although relatively relaxed, was restricted to transfer preferably hexoses containing an acetamido group in position C2 and unable in case of polysaccharides where the two sugars

proximal to the lipid carrier are connected via a β(1→4) linkage (Kay *et al.*, 2019; Wacker *et al.*, 2006). Recently, the employment of several Pgl-variants and the optimization of the enzyme allowed to overcome this issue (Harding *et al.*, 2019; Wacker *et al.*, 2006).

A balanced number of linkage sites per protein molecule and polysaccharides length is required in order to avoid the masking of the T-cell and B-cell epitopes and the following decrease in the immunogenicity. Moreover, the presence of multiple polysaccharide chains on the bioconjugate might confer an improved antigenicity to the pathogenic carbohydrates, as occurs for chemically conjugated vaccines which are decorated with multiple saccharide molecules (Micoli *et al.*, 2018a).

Indeed, engineering the carrier protein with multiple glycosylation sites and the definition of the minimal glycan epitopes would lead to the antigen design of the bioconjugate towards improved and enhanced vaccine molecules to increase the immune response.

In this context, the extent of glycosylation is considered a surrogate of the protein-to-carbohydrate ratio and, more specifically, as the location and the quantitative evaluation of the conjugated sites. The availability of analytical methods which are able to monitor the glycosylation process in terms of extent of glycosylation and location is an urgent need (Micoli *et al.*, 2018a).

3. The role of MS in the vaccine characterization

3.1 Regulatory requirements for glycoconjugate vaccines

Vaccination has as target population healthy individuals, such as infants, pregnant women and elderly, and it is aimed to prevent future infectious diseases. For this reason, the drug products must be safe and effective and therefore the pharmaceutical industry is among the most rigorously regulated industries worldwide (Sharma *et al.*, 2018).

Modern vaccines, in contrast to the older ones which contain attenuated or killed pathogens, derive from chemical synthesis (carbohydrate or DNA vaccines) or from biotechnological production (recombinant proteins or subunits), leading to more defined molecules in the final drug product. The most advanced are the glycoconjugate vaccines, where saccharide structures of a specific bacterial pathogens are conjugated to carrier proteins. However, their levels of impurities and degradation in drug substances (DS), and the presence of reaction intermediates in the drug product (DP), have to be evaluated and strictly controlled. These attributes have to meet the rigorous specifications required by National and International regulatory authorities in order to be licensed (FDA, EMA, ICH and WHO) (Lepenies, 2015b).

The regulatory requirements should be accomplished to obtain the approval along the whole vaccine development pipeline, either during the manufacturing process or in the preclinical and clinical phase. The main aspects involved are:

i) to ensure the delivery of a consistent and sterile product;

ii) to demonstrate the consistency of the manufacturing batches in comparison to those tested in clinical trials (batch release);

iii) to prove the safety, effectiveness and stability of the proposed vaccine.

Focusing on research and development levels, several tests are performed for the complete characterization of the vaccine molecules. In details, in these preclinical phases it is required to establish the stability of the vaccine, the storage and distribution conditions and to define the shelf life of the product. At this stage, the *in vivo* animal studies to define the immunization schedule supporting the further clinical phase and the antigen structure-to-immunogenicity correlation studies are performed (ICH harmonized tripartite guideline Q6B, http://www.ich.org) (Lepenies, 2015b).

For what concern glycoconjugate vaccines, WHO developed product-specific guidance documents for the different polysaccharide and glycoconjugate vaccines covering all aspects of the production and quality controls (QC) to be performed, to regulate the performance of clinical trials and analysis of the data (http://www.who.int/biologicals/technical_report_series/en/); European and US pharmacopeia also refer to glycoconjugate vaccines with specific chapters (Lepenies, 2015b).

Regulatory authorities defined the "characterization of starting materials, intermediates and final products, and of their stability" as one of the main expectation required. Accordingly, it results in an extensive physiochemical characterization to ensure the vaccine safety and effectiveness. In details, some of the required documentations are about:

- the quantification and identification of impurities,
- the molecular size of the saccharide conjugated to the protein,
- the quantification of the oligomeric forms of the protein,
- the degree of activation of saccharide and protein;
- the integrity of the final conjugate, in terms of "free" unconjugated protein or polysaccharides;
- the polydispersity of the protein isoforms in the final product;
- the polysaccharide-protein ratio (w/w);
- the residual reagents or uncovered activation sites of the final vaccine.

Furthermore, the regulatory agencies demand for validated analytical methods for batch release testing in manufacturing and quality controls processes, while at preclinical levels the validation is not usually required for characterization assays, although they should be demonstrated to be fit-for-purpose (Jenkins *et al.*, 2015; Lepenies, 2015b).

Hence, the glycoconjugate vaccine candidate should be tested to determine the polysaccharide-to-protein ratio, the extent of glycosylation and to quantify the unconjugated ("free") saccharides and

carrier proteins. Evaluation assays for residual conjugation reagents as activated protein and/or saccharides are required in case of chemically conjugated molecules (Lepenies, 2015b), while they could be avoided in case of bioconjugated molecules. Nevertheless, the extent of glycosylation and the quantification of the free carrier protein and saccharides should be evaluated for bioconjugate vaccines. Indeed, conjugates with excess of free carrier protein or with high polysaccharide-to-protein ratio would have immunological implications, as for example impairing the antigen-presenting process (Cui *et al.*, 2010). Pozsgay *et al.* observed that a high level of protein modification could be needed for optimal immune response of short synthetic *S. dysenteriae* type 1 oligomers. By contrast, a lower number of glycosites could be sufficient for polysaccharides, avoiding the lower immune response due to the masking of T cell epitopes of the protein (Pozsgay *et al.*, 1999). These findings suggest the importance of a right balance between number and length of saccharide chains attached to the carrier protein for the preservation of the protein T cell epitopes (Micoli *et al.*, 2018a).

3.2 MS role in vaccine development

The development of the current recombinant vaccines with more defined features, in comparison to the traditional ones containing live, attenuated, or killed pathogens, allowed the application of Mass Spectrometry (MS) techniques both in discovery and development field (Donnarumma *et al.*, 2016; Josefsberg & Buckland, 2012). Moreover, the MS role is growing also thanks to the technological advances in instrumentations and methods developments. MS detectors coupled to enhanced separation techniques, such as Ultra High-Performance Liquid Chromatography (UHPLC, or UPLC) and electrophoresis (*i. e.* Capillary Electrophoresis, CE) could be applied to gain a comprehensive physiochemical characterization of vaccine products. Some examples applied at research and development level are: *i)* the characterization of the major and minor isoforms in the final product; *ii)* the detection of uncommon post-translational modification of a protein constituent in a vaccine; *iii)* the detection and identification of a low abundance host cell protein in drug substance by a top-down proteomics approach; *iv)* the discovery of an unusual modifications in a protein-polysaccharide conjugate induced by conjugation chemistry (Sharma *et al.*, 2018).

A huge number of MS applicability examples for antigen characterization in vaccines demonstrates how MS-based methods feasible for QC and Good Manufacturing Practice (GMP) environments are emerging. The only limitation would be the method validation, since the classical validation rules as specified in the ICH guidelines are often unsuitable for MS-based methods (Sharma *et al.*, 2018). Nevertheless, the update of method validation requirements from the regulatory agencies, occurred for the implementation of HPLC or CE methods for batch-to-batch consistency in QC (D'Atri *et al.*, 2019) and for NMR spectroscopy for routine release test for structural characterization of

carbohydrate-based vaccines (Jones & Ravenscroft, 2008; Lepenies, 2015b), would be a benefit for the integration of MS as release method for vaccine products not only as tool for antigen characterization in research and development departments (Hickey *et al.*, 2016). Accordingly, MS characterization is widely applicable to understand the product's structure as well as to help establish clinical trial material release specifications and to ensure quality (Sharma *et al.*, 2018).
Some example of MS-based techniques applied for characterization of protein-based vaccines are reported in table 1.

MS methods for characterizaton of Protein-based Vaccines	
Attribute	Method
Intact Mass	LC-MS
Peptide mapping	LC-MS/MS
Glycan profiling	CE-LIF; HILIC-FLR; LC-MS/MS
Glycopeptide site-specific analysis	LC-MS/MS
Aggregation or higher order aggregates	HPLC SEC-MALS/MS
Epitope mapping and structure characterization	HDX-MS

Table 1:MS-based methods for characterization of protein vaccines. Adapted from (Sharma et al., 2018).

A successful case study resulted to be the measurement of hemagglutinin (HA) strain content in the inactivated influenza vaccine (Williams *et al.*, 2008). Quantitative MS-based method has been developed to be applied on the final drug product, thus without sample cleanup and in presence of detergents and other proteins of the vaccines. That achievement has been complied applying Stable Isotope Dilution method (SID) for accurate quantification of HA in combination with Selected Reaction Monitoring (SRM) technique (Williams *et al.*, 2008). Such analytical method resulted to be equivalent to the traditional approved one (Single-radial immunodiffusion, SRID), and since 2013 it has been employed as orthogonal supporting batch release test for the quantification of HA content in seasonal influenza vaccines (Ridenour *et al.*, 2015; Santana *et al.*, 2014; Sharma *et al.*, 2018).
In a comprehensive review performed by Sharma V. *et al.* in 2018, MS-based techniques have been reported to be extensively used to perform the structural characterization, glycosylation profiling and antigen quantitation during the development of different vaccines (Sharma *et al.*, 2018).
Focusing on glycoprotein and glycoconjugate vaccines, in all the works reported, the target vaccine antigen to be characterized need to be purified, enriched, deglycosylated, immunocaptured or, in other words, isolated from the complex biological matrix or from the final DP before to be analyzed by MS (Sharma *et al.*, 2018). Accordingly, some considerations could be arisen:

i) the absence of a targeted analytical method to characterize glycoproteins to accomplish an in-process monitoring of the conjugation and, in case of bioconjugated vaccines, of the bioconjugation levels during fermentation as well as the quantification of the "free" protein, without applying prior purification steps of the sample (which could lead to modification or alteration of the product);
ii) the lack of a consistent analytical method to be applied along all the vaccine antigen production steps as well as in the final product for evaluation of extent of glycosylation, in both DS and DP;
iii) the relevance of sample preparation in MS analyses. Without isolation, enrichment, or purification processing of the samples prior the analyses, the characterization of targeted glycoproteins has not been achieved.

Within this context, the targeted proteomics approaches as SRM or Parallel Reaction monitoring (PRM) techniques could have wide applicability due to highly specificity, sensitivity and reproducibility in the quantification of peptides in complex matrices. Such techniques are bottom-up approaches based on the precise knowledge of a peptide mass derived from the target protein, coupled to its unique fragmentation spectra.

3.3 Targeted proteomics technique: Selected reaction monitoring (SRM)

Accurate measurements of protein concentrations in biological or clinical samples traditionally rely on the use of immunoassays, such as the enzyme-linked immunosorbent assay (ELISA). However, a growing number of new proteins and antigens are currently discovered through proteomic methods (Grandi, 2001; Klade, 2002). Despite the high sensitivity of the immunoassays for the detection, antibody reagents could be not necessarily available for the precise quantification of novel proteins, antigens or biomarkers. Moreover, proteomics approaches hold great potential for quantifying protein isoforms or protein posttranslational modifications, for which effective antibody-based assays are unavailable (Shi *et al.*, 2012).

SRM, also referred to as multiple reaction monitoring (MRM), has emerged as a powerful alternative to immunoassays for targeted quantification of proteins in a number of biological samples (Shi *et al.*, 2012). It is a MS technique for the targeted detection and quantification of selected proteotypic peptides (PTPs) with known fragmentation properties in complex sample matrix. In vaccine and glycoprotein characterization fields, numerous studies could be cited to describe the enhanced specificity, sensitivity, and reproducibility of SRM MS technique. For example:

- Specificity: John R. Barr and his group developed an SRM method for simultaneous quantification of both HA and neuraminidase (NA) protein antigen of the three viral strain

subtypes found in seasonal vaccines. This method resulted rapid and accurate since the experiments has been carried out on several drug products licensed from different manufacturers. Moreover, it resulted easily adaptable to changes in annual vaccine strains. The enhanced specificity could be also highlighted considering that the analyzed drug products contained various surfactants and preservatives, and all were used without further purification prior MS analyses (Williams *et al.*, 2012).

- Sensitivity: Picotti *et al.* applied SRM to detect and quantify yeast proteins to evaluate the dynamic range of such MS targeted approach in a tryptic digest of a total cell lysate. SRM allowed the measurement of proteins down to a concentration in the low microgram or high nanogram per milliliter range, without any further enrichment or purification of the sample before the analyses (Picotti *et al.*, 2009).
- Reproducibility: a number of different studies reported the development of multiplexed SRM assays for the simultaneous quantification of plasma proteins and tested their accuracy and reproducibility, down to a concentration of 1 µg/mL and with coefficients of variation mostly <20% (Addona *et al.*, 2009; Anderson & Hunter, 2006; Kuzyk *et al.*, 2009).

SRM exploits the unique capability of triple quadrupole (QQQ or QqQ) mass spectrometers (Yost & Enke, 1979), to act as mass filters and to selectively monitor a specific analyte molecular ion and one or several fragment ions generated from the analyte by collisional dissociation. The number of such fragment ions that reach the detector is counted over time, resulting in a chromatographic trace with retention time and signal intensity as coordinates Figure 4. Precursor–fragment ion pairs, termed SRM transitions, are sequentially and repeatedly measured more quickly than the analyte chromatographic elution time, yielding chromatographic peaks for each transition that allow the parallel quantification of multiple analytes. When applied to proteomics, SRM measures peptides produced by the enzymatic digestion of a proteome as surrogates of the corresponding proteins. Molecular ions within a mass range (termed isolation window) centered around the mass of the targeted peptide are selected in the first mass analyzer (Q1), fragmented at the peptide bonds by collision-activated dissociation (in Q2 or q) and one or several fragment ions, uniquely derived from the targeted peptide, are measured by the second analyzer (Q3) (Lange *et al.*, 2008; Vidova & Spacil, 2017).

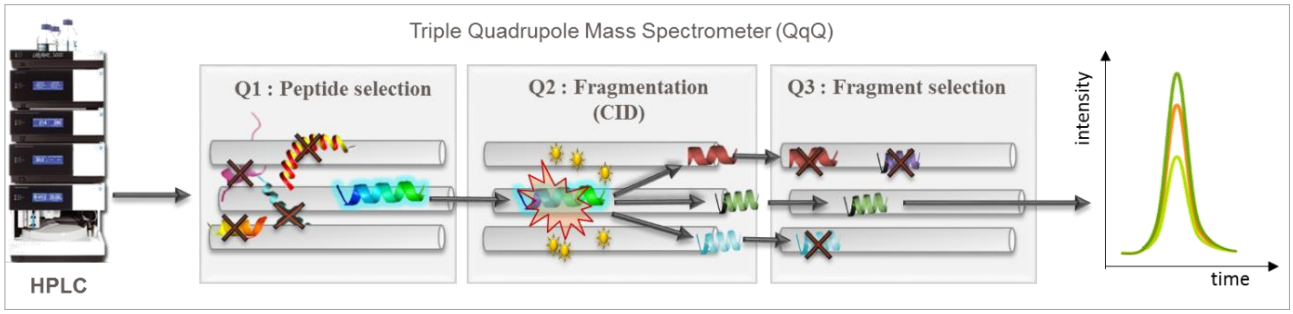


Figure 4: Selected Reaction Monitoring technique. Molecular ions of a specific analyte are selected in Q1 and fragmented in Q2. A specific fragment ion from the targeted analyte is selected in Q3 and guided to the detector.

The qualitative and quantitative protein assays set-up by SRM requires a more complex workflow in comparison to the conventional shotgun proteomic studies. Peptides serves as surrogate of the corresponding protein and should be selected and evaluated with respect to several attributes as proteotypicity, physiochemical properties, and quantitative performances, herein described:

1. Proteotypicity: refers to the characteristic properties of peptide sequence to uniquely represents the selected proteins or isoforms (biological specificity) in the background proteome of interest. This attribute should be accomplished due to the high sequence-based selectivity character of the SRM and may be firstly verified *in silico* using blasting tools available through protein databases as UniProtKB (www.uniprot.org) or NCBI (www.ncbi.nlm.nih.gov) (Lange *et al.*, 2008).

2. Physiochemical properties: the amino acid sequence and the length of the peptides determine the chemical polarity and hydrophobicity rate, that are implied in the peptide ionization efficiency and MS detectability. In general, short hydrophilic and long hydrophobic peptides are not preferred. For the same reason, peptide containing amino acids prone to post-translational modifications (PTM) or to chemically induced modifications, due to the sample processing, should be discarded in order to avoid misquantification. Another source of error would be the incomplete digestion of the protein and the occurrence of missed cleavages. Accordingly, peptides predestined for high rate of missed cleavage should be rejected (Brownridge & Beynon, 2011). As a consequence, at least two peptides should be monitored for each target protein to perform a reliable quantification and detection in SRM (Lange *et al.*, 2008).

3. Quantitative performances: to perform a reliable quantification in complex biological matrices, the selection of at least 3 SRM transitions per peptide would be recommended to preserve the selectivity of the assay. The optimal transitions include the most predominant fragment in terms of signal intensity, with a sufficient sequence specificity. The quantitative performances of proteotypic peptides are experimentally determined, performing replicated

analyses. The key parameters evaluated are the Limit of Detection (LOD), the Limit of Quantification (LOQ), the sensitivity (expressed by the slope of the calibration curve) and the technical reproducibility (Coefficient of Variation, CV) (Vidova & Spacil, 2017).

After the proteotypic peptides (PTP) selection, the SRM transitions need to be optimized and validated before moving forward to the quantification assay. In particular, the collision energy (CE) value associated for the ion fragmentation in Q2 is a fundamental parameter to be optimized individually for each peptide in order to provide the best signal intensity. This is obtained using a CE step variations analysis, monitoring the corresponding ion peak area intensities (Picotti & Aebersold, 2012). The validation of SRM transitions is performed by experimentally addressing that the chosen transitions and their intensities are associated with the targeted peptide, in complex matrices. The matrix might have different effects on the PTPs performances, such as shift of the retention times or reduction of the signal intensities but validating the transitions in complex biological matrix ensures that the signals of selected transitions are compatible with the quantification of the peptides under analysis (Lange *et al.*, 2008).

3.3.1 Stable Isotope Dilution (SID) Method and SRM MS

SRM MS technique is one of the approaches reported for absolute quantitative proteomics in recent years (Bantscheff *et al.*, 2007). The protein is not directly quantified, but it is quantitatively measured by signature peptides (PTP) that are generated during enzymatic digestion in stoichiometric amounts. The PTP is uniquely associated to the protein and its amount reflects the amount of protein present in sample.

By contrast to the label-free methodologies interesting thanks to their ease of implementation and low financial costs (Neilson *et al.*, 2001), SRM MS allows the generation of more analytically rigorous data using the established principle of stable isotope dilution-mass spectrometry (SID-MS) (Bantscheff *et al.*, 2007; Ciccimaro & Blair, 2010).

In SID, stable isotopically labeled peptide standards are synthesized for every peptide to be measured. To achieve absolute quantification, an accurately determined amount of such standard peptide is added to the sample. As the amount of standard added is known, the ratio of the mass spectrometric response between standard and analyte can be used to determine the quantity of analyte present. The integrated peak areas for transitions for the unlabeled peptide analyte are summed and then normalized to the summed peak areas for transitions from the corresponding labeled peptide standard. Therefore, the reported values obtained by SID method are “measured concentration” (Zhang *et al.*, 2011). The known amount of isotopically forms of PTPs are spiked in the biological sample before digestion, at the beginning of sample preparation:

i) to monitor the target PTP along the entire analysis since these standards are subjected to all of the sample preparation and cleanup steps as the peptides derived from the protein, leading to an accurate comparison of MS the signals. (Carr *et al.*, 2014);

ii) to increase the precision of the quantification (Holman *et al.*, 2012);

iii) to built-up a dose-response standard curve to assess the fitting of the linear regression model to the data obtained;

iv) to achieve the quantitation of unknown targeted analytes in the sample (Bantscheff *et al.*, 2007).

One of the methods employed is the use of chemically synthetized PTPs, containing amino acids enriched with heavy isotopes *e.g.* ^{13}C, ^{15}N, to allow mass resolution from the target unlabeled (light) analyte (Silva *et al.*, 2006). Being chemically equivalent to endogenous peptide to be quantified, the labelled peptide has identical yield through all sample purification steps, as well as the same ionization efficiency and the MS fragmentation behavior in the mass spectrometer. These peculiarities account for any variations in the analytical procedure. Thereby, it's possible also to evaluate variations in the sample preparation and to quantify assessed biases by the use internal standard (Carr *et al.*, 2014; Holman *et al.*, 2012).

3.3.2 SRM and Western blotting analysis

For its features, SRM is similar to Western blotting analysis; it is used for the identification and quantification of specific known set of proteins in a complex background. However, the methods differ substantially in their implementation, the reliability of the assay, and the quality of results they produce (Aebersold *et al.*, 2013).

A Western blot assay essentially depends on the specificity of the antibody used (Towbin *et al.*, 1979; Towbin *et al.*, 1992), in contrast to an SRM assay that depends on multiple parameters, such as the retention time, the mass-to-charge ratio of the precursor ions and fragment ions of the targeted peptide, and the relative signal intensities of the detected fragment (Picotti & Aebersold, 2012).

The quality of quantitative data obtained is very different: quantification by WB is based on the intensity of a band obtained by the reaction of specific antibody that may be poorly characterized, and a reference sample is not always available to test the performance of the assay. By contrast, SRM depend on isotopically labeled reference peptides, whose quality can be always verified in a fragment ion spectrum. Moreover, multiple independent peptides are monitored simultaneously to quantify a specific protein, which signals are integrated into a composite score indicating the protein quantity.

To date, SRM is considered the key technology to investigate deep into a proteome, to obtain precise and absolute quantitative information related to protein(s) expressed in wide dynamic range.

Moreover, it is applied to obtain sensitive data from different and very complex biological matrices, as cellular lysate and tissue in different filed as biology, clinical and in studies upon protein interaction and modification (Bantscheff *et al.*, 2007; Shi *et al.*, 2012; Silva *et al.*, 2006).

Aim

Glycoconjugate vaccines have been successfully employed to overcome limitations of the poor immunogenicity of long-lasting memory immune response elicited by the carbohydrate vaccines (Schneerson *et al.*, 1980). However, the manufacturability of chemical conjugated vaccines by the classical approach has some constrictions due to the complex processes applied to achieve the final product (Frasch, 2009). Recently, novel approaches to obtain glycosylated protein suitable as vaccines are progressing towards further technologies. In particular, Protein Glycan Coupling Technology (PGCT or bioconjugation) consists in exploiting a bacterial enzymatic pathway to obtain *in vivo* N-glycosylated recombinant carrier protein (Wacker *et al.*, 2002). In *E. coli* cells, the glycosylation machinery (PglB enzyme, pathogen saccharides and target carrier proteins) expressed to produce site-selective bioconjugated proteins leads to several advantages in terms of vaccine development streamlining (Langdon *et al.*, 2009; Ravenscroft *et al.*, 2019; Wacker *et al.*, 2014; Wetter *et al.*, 2013). Bioconjugation application, in comparison to the chemical glycoconjugate, reduces the steps of pathogen handling, the components formulation and the downstream vaccine preparations. Moreover, the site-selective character of the conjugation could imply an easier characterization of such molecules. Therefore, the use of bioconjugate vaccines with defined physico-chemical characteristics results in improved quality controls during the process development and in an easier characterization of such molecules (Frasch, 2009).

The regulatory agencies are harmonizing the requirements of glycoconjugate vaccines towards even more strictly controls. A precise characterization of such vaccines by robust analytical testing is one of the main expectations for licensing (Lepenies, 2015a). The use of site-selective approaches would increase the batch-to-batch consistency and the structure-biological activity correlation of the glycoconjugate vaccines, with advantages in the number of routine controls for product release and manufacturing times and costs (Hu *et al.*, 2016).

To date, most of the assays aimed to perform physio-chemical characterization of glycoconjugate vaccines have been carried on purified drug substances, to evaluate the polysaccharide-to-protein ratio, to confirm the conjugate integrity and to evaluate the unconjugated saccharide/carrier protein or the residual activated intermediates of conjugation reaction (Lepenies, 2015a).

However, no analytical tools aimed to evaluate the extent of glycosylation are available in the context of bioconjugate vaccines characterization for in-process production monitoring. Furthermore, until now prior purification or enrichment steps of the target glycoprotein are still needed for any characterization of the product.

Within this context, the enhanced specificity of LC-MS techniques would be an added value to the development of novel analytical methods for characterization of bioconjugate vaccines (Sharma *et al.*, 2018).

The aim of the study is to setup a MS-based approach for the quantification of the extent of glycosylation, suitable for in-process monitoring of the bioconjugate production and for the final product characterization. The assay applied was the Selected Reaction Monitoring (SRM) technique, that resulted the most suited to perform a targeted quantification of the extent of glycosylation since ensures to achieve the required specificity, sensitivity, accuracy, and precision in the intended matrix (Carr *et al.*, 2014; Picotti *et al.*, 2009; Sharma *et al.*, 2018; Vidova & Spacil, 2017)

A rational analytical driven design of the glycosylation sites combining the optimal MS detectability features and the minimal requirements for N-glycosylation by PglB was undertaken, in order to obtain engineered bioconjugate vaccines more appropriate from an analytical perspective.

As proof of concept, we focused our attention on Hla-CP5 bioconjugate molecule and a LC-SRM MS method was developed to quantify the extent of glycosylation in single and double glycosylated carrier protein isoforms.

Result and Discussion

Background: ***Staphylococcus aureus* Hla-CP5 bioconjugate molecule as Proof of Concept**

Bioconjugate vaccines are meant as molecules in which the covalent linkage of bacterial polysaccharides on a carrier protein selectively occurs on specific amino acids in the protein sequence, leading to a glyco-conjugated protein. Such bioconjugates are produced by exploiting a bacterial N-glycosylation system, firstly observed in bacterial domain in *C. jejuni* (Szymanski *et al.*, 1999). The discovery and characterization of the enzymatic pathway for protein N-glycosylation in bacteria has encouraged the production of recombinant glycoproteins in glycoengineered *Escherichia coli* cells (Langdon *et al.*, 2009).

Bioconjugate vaccine candidates have been recently proposed for the prevention of Gram-negative (*Salmonella enterica*, *Shigella spp*, pathogenic *E. coli*) and Gram-positive pathogen infections (*S. pneumoniae* and *S. aureus* serotype 5 or 8) (Ravenscroft *et al.*, 2019; van den Dobbelsteen *et al.*, 2016; Wacker *et al.*, 2014; Wetter *et al.*, 2013).

Among them, *S. aureus* α toxin Hla bioconjugated with *S. aureus* type 5 CP (Hla-CP5) induces rabbit or mice protective antibody recognizing both the glycan and the protein moieties, exploiting the dual role of Hla protein as carrier and as protective antigen (Wacker *et al.*, 2014). This data is particularly relevant for the development of a vaccine preventing the diffusion of this pathogen, which is becoming challenging to fight due to the increase of the multi-drug resistant strains spreading around the world, including hospital and community-related infections strains (Ansari *et al.*, 2019). Despite the increasing relevance of bioconjugates in the vaccine development field, the robust analytical tools needed to evaluate efficacy of carrier glycosylation are still missing (Micoli *et al.*, 2018a). Quantification of vaccine antigens with a dual nature such as bioconjugates by immunoassays remains a challenging task, since antibody recognition depends both on the amount and antigenic variability of the two different moieties of the antigen. Nevertheless, this information is fundamental to respond to potential regulatory expectations and to monitor antigen production and characterization (Lepenies, 2015b).

For the quantification of the extent of glycosylation of a bioconjugate, our purpose was to apply a protocol developed by Zhu et al. (Zhu *et al.*, 2014) for absolute quantification of the glycosylation site occupancy of bovine fetuin protein expressed in *E. coli* cells. This approach has been developed without the usage of deglycosylase enzyme (*i. e.* PNGase-F) to remove the sugar before the MS analysis of the target protein, a step which is commonly used in protocols applied for the quantification and characterization of glycosylated protein by MS. It is based on the stable isotope dilution (SID) quantification method associated with a Selected Reaction Monitoring (SRM) technique, also referred as Multiple Reaction Monitoring (MRM) MS (Picotti & Aebersold, 2012),

in which isotopically labeled peptides identical to those deriving from the tryptic digestion of the target protein (proteotypic peptides, PTPs) were used either as surrogates for the total quantification of the carrier and for the quantification of unglycosylated site contain in a tryptic peptide. In this way, the abundance of the glycosylated portion of the protein is calculated by subtracting the non-glycosylated protein abundance from the overall protein concentration, and the site occupancy is then determined (Zhu *et al.*, 2014).

In the present study, the Hla-CP5 conjugate was used as a Proof of Concept to develop and test analytical methods for the characterization of the extent of glycosylation in the bioconjugate molecules (Wacker *et al.*, 2014).

With the purpose to apply the SRM-MS approach developed by Zhu *et al.* (Zhu *et al.*, 2014) to quantify the extent of N-glycosylation (see paragraph 2.2 of the Results and Discussion section), the selected peptides must fulfill precise prerequisites. In particular, the peptide carrying the consensus sequence for glycosylation (glycosite) must be unique for the carrier protein and show a strong signal in MS analysis.

Unfortunately, the necessary requirements needed for the application of the approach were not satisfied by the bioconjugate Hla-CP5 produced as reported in (Wacker *et al.*, 2014).

In fact, regarding the:

i. Uniqueness: considering that the bioconjugate molecules are produced in a glycocompetent *E. coli K-12* periplasm, the uniquely association of the tryptic peptide carrying the glycosite to the targeted protein Hla was investigated in such specific biological matrix.

   In details, from BLASTp analysis of the theoretical peptide carrying the glycosequence against the publicly available host strain proteome database (https://www.uniprot.org/proteomes/UP000182246) resulted in a significative match with a tryptic peptide belonging to the cytochrome o ubiquinol oxidase subunit I protein (NP_414965.1). Such protein is a component of the aerobic respiratory chain, located in the cell inner membrane with several topological domains exposed in the bacterial periplasm.

ii. Detectability by MS: as not all the peptides of a digested protein are detectable in an LC-MS/MS analyses due to the various physicochemical properties of the peptides that have an impact on ionization efficiency during electrospray ionization (ESI), the detection of such peptide was investigated (Schulz-Trieglaff *et al.*, 2008).

With this purpose, several tests were performed by digestion and LC-MS/MS analyses on unconjugated purified Hla protein of the Hla-CP5 bioconjugate molecule. Data are shown and discussed in Appendix 1.

Taken together, the generated data on the Hla-CP5 bioconjugate protein moiety carrying the glycosylation site, indicated that the DQNR (Hla-glycosite) was not MS detectable under our experimental conditions. As consequence, to obtain MS compatible and quantifiable peptides carrying glycosite a new analytic driven design strategy was developed.

Part I: Experimental strategy set-up for an analytic driven design of the bioconjugate

The figure 5 presents the workflow employed for the analytic driven design of a carrier proteins for bioconjugation.

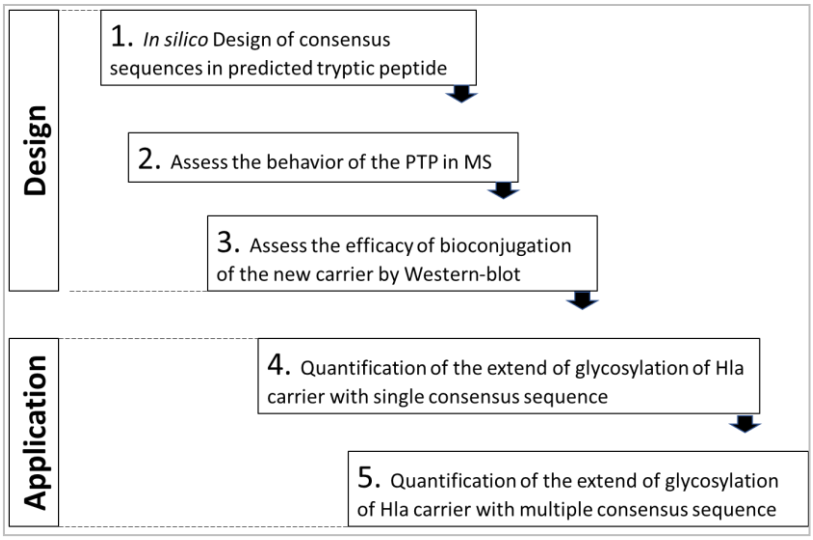


Figure 5: Workflow of the analytics-driven design approach developed.

The first step was the *in silico* design of the consensus sequences predicted to be substrates of the PglB enzyme (Kowarik *et al.*, 2006b) and able to generate tryptic peptides (referred as proteotypic peptides, PTPs) suitable for the quantification of the extent of glycosylation by LC-MS (Zhu *et al.*, 2014).

The designed PTPs were chemically synthetized in natural or heavy-labeled forms by incorporating ^{13}C-^{15}N in the arginine residue and quantified by the provider according to amino acid composition analysis. At this stage of the workflow, the synthetic peptides were used to investigate their behaviors in MS using a triple quadrupole mass spectrometer.

Once PTPs suitable for quantification by LC-MS were identified, the corresponding sequences were introduced in a carrier protein, and the efficacy of PglB enzyme to recognize and glycosylate the new carrier isoforms was evaluated by Western blot.

In this study, the *S. aureus* Hla was selected as model since the antigen has been demonstrated to be a carrier for bioconjugation able to induce a protective response in animal models (ref), and that the World Health Organization and Centers for Disease Control have highlighted that antimicrobial

resistance, with methicillin resistant *S. aureus* (MRSA) representing a major problem and that the development of a vaccine against *S. aureus* is of high importance.

The site occupancy for each consensus sequence was then determined through the absolute quantification of the non-glycosylated form of the glycopeptide, by using stable isotope dilution method (SID) with a Selected Reaction Monitoring (SRM) technique, also referred as Multiple Reaction Monitoring (MRM) MS (Picotti & Aebersold, 2012). In our approach, two sets of stable isotopically labeled peptide standards were spiked into the sample before proteolysis. One set of peptide standard was employed to determine the total carrier amount, while the other standard set monitors the non-glycosylated form of the carrier. The samples were digested with trypsin and the resulting peptides were analyzed by LC-SRM MS. The abundance of the glycosylated form of the protein was calculated by subtracting the unglycosylated protein abundance from the overall protein concentration, and the site occupancy was determined. The approach has been developed for the quantification of the extent of glycosylation of bovine fetuin, an eukaryotic (Zhu *et al.*, 2014). The enhanced precision of this approach derives from the avoidance of the deglycosylation step to remove the sugar before MS analyses, a step routinely and widely used in protocols for the quantification and characterization of glycosylated protein by MS (Gonzalez *et al.*, 1992; Kuster & Mann, 1999; Segu *et al.*, 2010; Zhang *et al.*, 2003). Moreover, sensitivity and specificity are ensured by the application of LC-SRM MS technique, proved to be suitable for quantification of vaccine antigen in complex matrices (Williams *et al.*, 2008)

Moreover, the same approach was applied to characterize carrier proteins with multiple sites of bioconjugation to evaluate the extent of glycosylation on each glycosite.

In the following paragraphs, each step of the workflow reported in fig. 5 is investigated, and a discussion of the obtained results is provided.

5.1 *In silico* design of consensus sequences

The consensus sequence substrate of *C. jejuni* PglB has been well characterized (Kowarik *et al.*, 2006b). Considering as position 0 the Asparagine residue (Asn, N) acceptor of the saccharide, the sequence is characterized by the presence of a negatively charged side chain amino acid residues in the position -2 (Asp or Glu, D or E) and a serine or threonine (Ser or Thr, S or T) in position +2, while amino acids in position -1 and +1, can be occupied by any amino acid residues except proline (Kowarik *et al.*, 2006b).

To design four consensus sequences predicted to be substrates of the PglB (Kowarik *et al.*, 2006b) and also able to generate peptides suitable for quantification by MS approach (Fig. 6B), the following aspects were taken into account:

(i) structure integrity: to minimize the possible impact on the carrier structure, the designed consensus sequences did not exceed nine amino acid residues and the insertion of hydrophobic and aromatic amino acid residues should be limited;

(ii) accurate quantification: to avoid underestimated quantification, amino acid residues prone to post-translational modifications (PTMs) such as oxidation (Met, Cys, Trp) and deamination (Asn and Gln) were limited. The consensus sequences were designed to be substrate of trypsin, selected for its high specificity, efficacy and ability to generate positively double charged peptides prompt to an efficient ionization in a MS analysis (Brownridge & Beynon, 2011; Glatter *et al.*, 2012);

(iii) site specificity of glycosylation: the frequencies of amino acid residues surrounding the Asn, acceptor of the saccharide, evidenced from the comparison of a data set containing 32 active *C. jejuni* N-glycosylation sites were taken into consideration (fig. 6A) (Kowarik *et al.*, 2006b);

(iv) proteotypicity: the designed consensus sequences were checked to be unique for the carrier protein isoforms when expressed in *E. coli K-12*. The uniqueness of the sequences was checked by blasting the designed sequences in the proteome database *of E. coli* K-12. Such property is crucial for the peptide detection and quantification specificity (Li *et al.*, 2010).

The four designed consensus sequences were (-3)KDSNXTSAR(+5) in which X is an Ile, Ser, Val or Ala amino acid residue, respectively.

After trypsin digestion, PTPs (-2)KDSNXTSAR(+5) were named PTP-i, PTP-s, PTP-v and PTP-a according to the amino acid residue present in the position +1 (Fig. 6B). The theoretical molecular monoisotopic masses for these peptides were 862.89, 836.81, 848.97 and 820.81 Da, respectively.

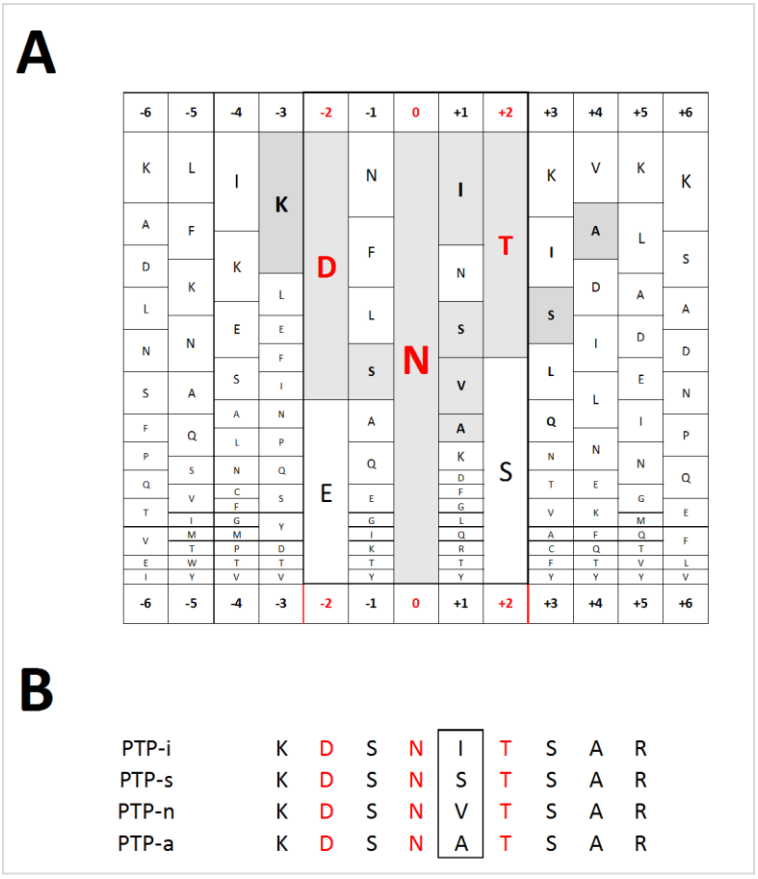


Figure 6: ***In silico design of consensus sequences.*** *(A) Statistical analysis of the occurrence of amino acids in the region from −6 to +6 of the glycosylated Asn residue found in 32 native C. jejuni glycoproteins as reported by Kowarik M, et al., 2006 (Kowarik et al., 2006b). The height in the box reflects the frequency of the amino acid residues in the consensus sequence. The Asn residue, site of glycosylation and the Asp and Thr residues in position -2 and +2 respectively, demonstrated to be crucial for an efficient glycosylation, are reported in bold red in a grey box. The amino acid residues in position -3, -1, +1, +3 and +4, respectively, represented in bold black in grey boxes, were the selected ones for the design of the four consensus sequences (B). They were selected as most frequently found and responding to the set-up criteria (limited selection of amino acid residues susceptible to post translational modification, hydrophobic and aromatic, see text "In-silico design of consensus sequences"). The amino acid Arg in position +5 (not reported in the statistical analysis), and the Lys residue in position -3 are the substrates of trypsin, required for the generation PTPs. The PTPs differ from each other only from the amino acid residue in position +1.*

In details, in position -1 amino acid residue Asn was excluded since susceptible to PTMs in MS, while Phe and Leu because of the hydrophobic character of such residue which could confer undesired destabilizing effect on protein stability. Accordingly, also Asn in position +1 and Ile and Val (in position +3 and +4, respectively) were excluded. Moreover, Lys in position +3 was excluded since the tryptic peptide that would be formed would have been too short for optimized MS detectability. It should be noticed many peptides fulfilling these characteristics could be designed.

The consensus sequence suitable for the quantification could be defined as:

$$(K/R+Y_n)\text{-}D/E\text{-}X_1\text{-}N\text{-}X_2\text{-}S/T\text{-}(Z_n+K/R)$$

where:

- X_1 and X_2 are not proline;
- Preferentially X_1 and X_2, Y_n and Z_n are not lysine or arginine residues;
- Preferentially X_1 and X_2, Y_n and Z_n are not cysteine, methionine, asparagine, glutamine residues and hydrophobic or aromatic;
- ($K/R+Y_n$) could be from 1 to 5-10 amino acid residues with the first amino acid residue being K or R;
- (Z_n+K/R) could be from 1 to 5-10 amino acid residues with the last amino acid residue K or R.

5.2 Targeted MS analysis to set-up SRM-MS method

5.2.1 Theoretical SRM transitions computation

In order to evaluate the MS fragmentation behavior of the designed peptides, PTP-i, PTP-s, PTP-v and PTP-a were chemically synthetized either in light and heavy-labeled forms by incorporating ^{13}C-^{15}N in the arginine residue and quantified by the provider according to amino acid composition analysis (JPT Peptide Technologies GmbH, Berlin).

The theoretical SRM precursor/fragment pairs (*i.e.* transition) were computed *in silico* using the Pinpoint software environment, selecting four transitions per precursor and assembling a list of total 16 transitions, summarized in table 2.

Moreover, the Collision Energies (CEs) were predicted for each PTP using the formula CE = 0.034 × (parent *m/z*) + 1.314 (Maclean *et al.*, 2010).

Name	Sequence	Charge state	Q1 Precursor ion (*m/z*)	Q3 Fragment ions (*m/z*)	Ion type	CE predicted
PTP-i	DSNITSAR	2	432.21	430.19	b4	16
				547.32	y5	
				661.36	y6	
				434.24	y4	
PTP-i*	DSNITSAR[HeavyR]	2	437.22	430.19	b4	16
				557.33	y5	
				671.37	y6	
				444.24	y4	
PTP-s	DSNSTSAR	2	419.19	404.14	b4	16
				635.31	y6	
				521.27	y5	
				434.24	y4	
PTP-s*	DSNSTSAR[HeavyR]	2	424.19	404.14	b4	16
				645.32	y6	
				531.28	y5	
				444.24	y4	
PTP-v	DSNVTSAR	2	425.21	416.18	b4	16
				533.30	y5	
				647.35	y6	
				434.24	y4	
PTP-v*	DSNVTSAR[HeavyR]	2	430.21	416.18	b4	16
				543.31	y5	
				657.35	y6	
				444.24	y4	
PTP-a	DSNATSAR	2	411.19	388.15	b4	15
				619.32	y6	
				505.27	y5	
				434.24	y4	
PTP-a*	DSNATSAR[HeavyR]	2	416.20	388.15	b4	15
				629.32	y6	
				515.28	y5	
				444.24	y4	

Table 2: ***Pinpoint computed list of transitions for PTP-i,, PTP-v, PTP-s and PTP-a.*** *In the table are reported the peptide names, the sequence, the charge state, the m/z values for precursor ions and fragment ions, the type of the fragment ion and the predicted CEs.*

5.2.2 Selection of transitions

The attributes considered for the selection of transitions were:

a) selectivity: transitions including the amino acid residues selective for each glycosylation site of the mutants (-i, -v, -a, -s versions) in the fragment ions;

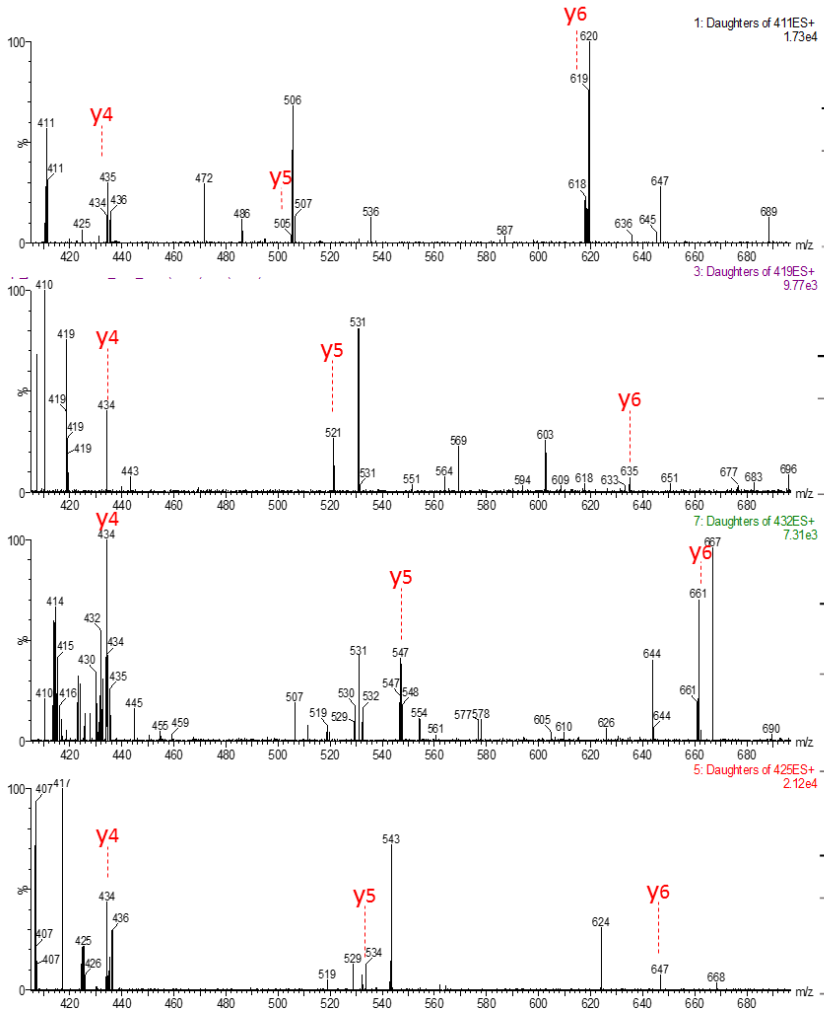
b) signal intensity: fragment -tsar (y4) which resulted to be the transition with the highest signal intensity was included in the final list although it is generated by fragmentation from all the precursors under analysis;

c) wide range: multiple transitions were selected for each peptide to reduce the likelihood to run into false positive during the analysis of complex mixture of peptides in which co-elute isobaric peptide exhibit similar primary sequence (Lange *et al.*, 2008).

To confirm the selected transitions for each target peptide and to investigate the quality of MS/MS spectra, an LC-MS/MS analysis was performed by using the isotopically labeled standard peptides mixed in 0.1% (v/v) formic acid (FA) solution both in Product Ion Scan and SRM modes.

As regards Product Ion Scan mode, intense signals and highest selectivity for each precursor in the spectral library were observed for three out of four predicted transitions. In details, the precursor/product pairs y5 and y6 containing the selective amino acid residue specific for each glycosylation site, and y4, common to all peptides were detected.

The MS/MS daughter/fragment ions obtained in product ion scan mode analysis are reported in fig. 7.



DSNATSAR (PTP-a)

z+	Q1	Q3	ion type
2	411.19	434.24	y4
		505.27	y5
		619.32	y6

DSNSTSAR (PTP-s)

z+	Q1	Q3	ion type
2	419.19	434.27	y4
		521.39	y5
		634.96	y6

DSNITSAR (PTP-i)

z+	Q1	Q3	ion type
2	432.21	434.21	y4
		547.15	y5
		661.47	y6

DSNVTSAR (PTP-v)

z+	Q1	Q3	ion type
2	425.21	434.23	y4
		533.30	y5
		647.34	y6

Figure 7: ***Product ion scan mode analysis of designed PTP.*** *MS/MS spectrum of the tree most intense fragment ions associated to each selected PTPs. In the right panel, the specific PTPs properties are summarized: m/z of parent ions (Q1) and fragment ions selected (Q3), charge state of the parental ion, (z+) and the ion type of the three selected fragments.*

5.2.3 Optimization and validation of transitions in SRM acquisition mode

Since bioconjugates are addressed, produced and purified in *E. coli* periplasm (Wacker *et al.*, 2002), the matrix effect on the detection of PTP-i, PTP-v, PTP-s and PTP-a was investigated. The matrix might have different effects on the PTPs performances, such as shift of the retention times or reduction of the signal intensities. Transitions validation allows to assess that the signals detected are

consistent with the quantification of the peptide under analysis. In particular, the CEs optimization and the chromatographic elution methods were investigated by spiking 0.1 pmol/µL of synthetic peptides in 50 µg of periplasmic fraction, prepared from an *E. coli* strain isogenic to the bioconjugate producer st9268 lacking for the Hla plasmid, and then digested with lysC/trypsin (table 11 in materials and methods section).

LysC/trypsin mix was chosen for digestion because of the high specificity and efficacy of the enzymes (Brownridge & Beynon, 2011). Specifically, the presence of lysC has been added to minimize miscleavages, since it has been demonstrated that when the cleavage site is located close to an acidic (D/E) amino acid residue a higher rate of miscleavages is registered using trypsin alone (Glatter *et al.*, 2012).

The generated peptides, in presence of spiked heavy PTPs, were separated on a reverse phase C18 UPLC coupled with a triple-quadrupole mass spectrometer for setting up the SRM-MS method. The predicted CEs were optimized in order to obtain the higher fragment ion signal intensities in matrix, applying ramped values panel of CEs (ranging from +6 eV to -6 eV of the predicted CE) to the precursor ions monitoring their fragmentations and the related peak area.

The chromatographic elution method was optimized to allow the separation of the PTPs. While well distinct retention times (RT) were obtained with strong MS signals in matrix for PTP-i, PTP-v, and PTP-s, ranging from 1.66 to 4.97 min (the minimal difference of 0,39 min was observed between peptide PTP-s and PTP-v), a deleterious effects of the matrix on the detection of PTP-a was observed, for which neither retention time and transitions were stable and comparable over repetitive analyses. For this reason, the bioconjugation consensus sequence DSNATSAR was not validated and not further investigated (PTP-a not validated) (Fig. 8 and table 3).

These conditions herein established were employed to detect and quantify the unglycosylated form of the Hla carrier protein in further analyses.

The designed PTPs compatible with MS detection can be used as universal target peptides for N-glycosylation of diverse carrier proteins selected to produce bioconjugate vaccine molecules. In fact, these sequences can be by bioengineering added to any carrier protein. Moreover, the availability of different peptides represents added value to engineer and characterize carrier protein with multiple glycosylation sites, since the extent of glycosylation can be differentially evaluated on individual site.

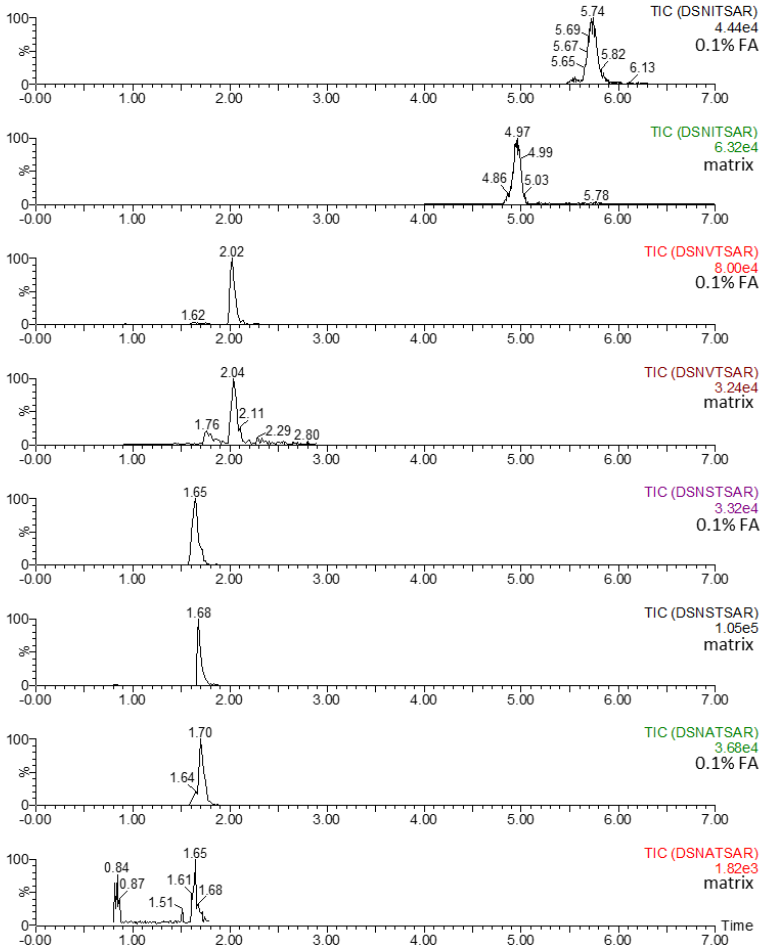


Figure 8: ***Total Ion Chromatogram (TIC) of PTPs designed in 0.1% FA and in matrix****. The detection in matrix for PTP-a (DSNATSAR) over repetitive analyses was not achieved.*

Name	Sequence	Charge state	Q1 Precursor ion (*m/z*)	Q3 Fragment ions (*m/z*)	Ion type	CE predicted	CE optimized	RT in neat (min)	RT in matrix (min)
PTP-i	DSNITSAR	2	432.2145	547.32	y5	16	14	5.74	4.97
				661.36	y6		14		
				434.24	y4		12		
PTP-i*	DSNITSAR[HeavyR]	2	437.2186	557.33	y5	16	14		
				671.37	y6		14		
				444.24	y4		12		
PTP-s	DSNSTSAR	2	419.1884	635.31	y6	16	14	1.65	1.66
				521.27	y5		14		
				434.24	y4		16		
PTP-s*	DSNSTSAR[HeavyR]	2	424.1926	645.32	y6	16	14		
				531.28	y5		14		
				444.24	y4		16		
PTP-v	DSNVTSAR	2	425.2066	533.30	y5	16	18	2.02	2.05
				647.35	y6		18		
				434.24	y4		12		
PTP-v*	DSNVTSAR[HeavyR]	2	430.2108	543.31	y5	16	18		
				657.35	y6		18		
				444.24	y4		12		
PTP-a	DSNATSAR	2	411.191	619.32	y6	15	17	1.70	N/A
				505.27	y5		17		
				434.24	y4		13		
PTP-a*	DSNATSAR[HeavyR]	2	416.1951	629.32	y6	15	17		
				515.28	y5		17		
				444.24	y4		13		

Table 3: ***Optimized SRM transitions, CEs and RT values for designed PTPs.*** *The names of the heavy labeled peptide are marked with an asterisk; the separation and detection of the PTP-a was not achieved in matrix (see details in the text 5.2.3 paragraph).*

5.3 Generation and characterization of Hla mutants constructs

The *S. aureus* Hla antigen was selected as proof of concept as reported in the Background paragraph of the present study. In particular, the sequences corresponding to the selected PTP-i, PTP-v and PTP-s were inserted into the position 130 of Hla, where the efficacy of bioconjugation has been demonstrated and the resulting bioconjugate has been proved to be protective in mice models (Wacker *et al.*, 2014).

The Polymerase Incomplete Primer Extension (PIPE) method was used for cloning procedure to perform insertional mutagenesis and to obtain plasmids carrying genes expressing the protein isoforms with the designed consensus sequences. This technique allows the insertion of a defined gene cluster in plasmid vector avoiding the classical use of restriction enzymes. Indeed, it is possible to clone a gene of interest into a target vector, both amplified by PCR using *ad hoc* designed primers (Klock & Lesley, 2009).

At this stage, PIPE was applied to perform site-directed mutagenesis on existing coding sequence of the Hla variant bearing the MS incompatible consensus sequence (Wacker *et al.*, 2014). In particular, *hla* gene was amplified by PCR using primers designed to anneal few base pairs (bp) before the encoding sequence of the glyco-site, but in which the corresponding bp annealing to the site mutation where designed to insert the nucleotides encoding for the desired bioconjugation consensus sequence. Moreover, the peculiarity of the PIPE results to be the achievement of PCR products with incomplete extended tails designed to anneal on the PCR products of the destination vector. Once the amplified DNA inserts were obtained for each of the clones needed, the PIPE-PCR products called I-PIPE (Insert-PIPE) were mixed to the PCR products of the V-PIPE, that is the PIPE-PCR performed on the destination vector (pEC415) to open the plasmid in which the insert need to be inserted.

By mixing such PCR products *E. coli* topTEN cells were transformed. The I-PIPE inserts anneal directionally across the complementary sequences encoded by the incomplete extended tails of the V-PIPE products and the bacterial DNA ligase repairs *in vivo* the nicks and the gaps on the plasmid sequence. The colonies were screened by plasmid purification and sequencing to select the clones of interest carrying Hla isoforms plasmids with the designed consensus sequences PTP-i, PTP-v, PTP-s. The carrier protein obtained were called hereinafter Hla-i, Hla-v and Hla-s, respectively.

A schematic illustration of the PIPE cloning method is reported in fig. 9.

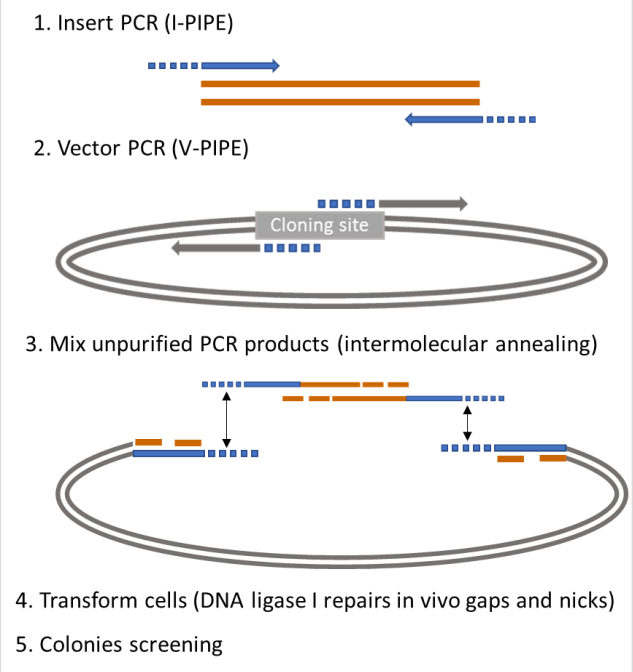


Figure 9: ***Schematic illustration of the PCR amplifications used in the PIPE-cloning method.*** *(1) I-PIPE (insert PCR). Primers are used to PCR amplify inserts from various templates in a way suitable for annealing to the vector amplified in (2) V-PIPE (vector PCR). PIPE cloning works by intermolecular annealing across the two annealing sites of the (1) and (2) PCRs (3). Mixing the unpurified PCR products and transforming bacterial cells, the colonies obtained would carry plasmids containing the targeted mutagenesis gene. Such M-PIPE (mutagenic PCR was used to create substitution mutants.*

5.4 Expression and glycosylation experiment on Hla isoforms to produce bioconjugate

Three sets of electrocompetent *E. coli* st9268(p1221) cells were transformed with plasmids Hla-i, Hla-v and Hla-s mutant constructs. The transformed strains were grown in 5 mL of Lysogenic Broth (LB) medium for culture, and the induction was performed overnight with 1 mM IPTG and0.2% arabinose (w/v) to induce the expression of plasmid p1221 for the PglB expression and of the p2872 for carrier protein expression, respectively.

To confirm the occurrence of bioconjugation in all the Hla isoforms, a parallel cell culture of *E. coli* (st9268) producing Hla-CP5 bioconjugate with the reference consensus sequence (Wacker *et al.*, 2014) incompatible with MS analysis, was used as positive control (lane 8, Fig. 10).

As negative control, cell cultures of the respective strains not carrying Hla plasmids, named strains st9268 (p1221, ΔHla), were performed.

Periplasmic fractions were prepared by harvesting 20 ODs of cell culture and 0.3 ODs of each culture and analyzed by Western blot analysis to assess the conjugation of CP5 to Hla (fig. 10). Bioconjugates were revealed using a rabbit polyclonal serum that specifically recognize Hla-CP5 bioconjugate. This antiserum was previously obtained by rabbit immunization using the purified Hla-CP5 reference bioconjugate.

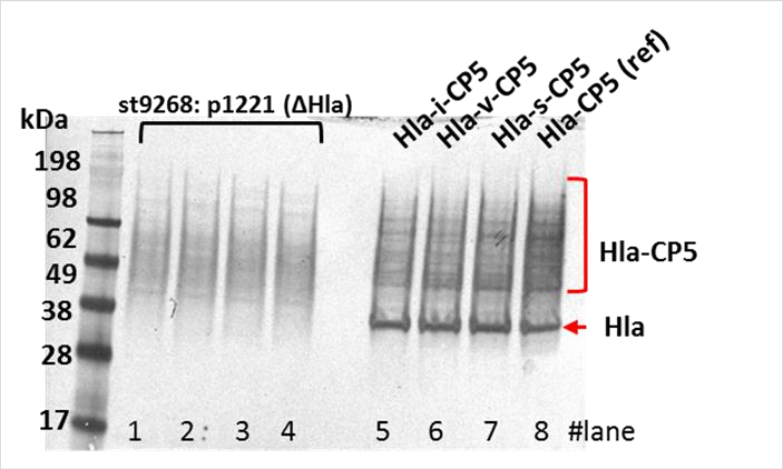


Figure 10: **Efficacy of bioconjugation of the newly designed carriers assessed by Western blot**. *The periplasmic fractions prepared from E. coli engineered for the expression of Hla-i-CP5, Hla-v-CP5 and Hla-s-CP5 (lanes 6-7-8, on the left) were analyzed by Western blot using a rabbit anti-Hla-CP5 serum. The levels of expression were compared to the Hla bearing the reference consensus sequence (KDQNRTK) which is not compatible with the MS analysis used as positive control to assess the bioconjugation occurrence (lane 9). As negative control the Western blot analysis of the periplasmic fractions prepared from the respective strains that do not express Hla are reported (lanes 1-4, on the left). The positive signal observed might be related to the reaction intermediate undecaprenyl-linked CP5 molecules, produced and assembled during the process.*

As showed in fig. 10, the carriers Hla-i-CP5, Hla-v-CP5 and Hla-s-CP5 (lanes 5-6-7) were characterized by a partially glycosylation pattern that was comparable among the three different constructs, although the levels of expression and bioconjugation obtained resulted lower than the signal showed for the positive control (in lane 9 in fig. 10). The glycosylation pattern appears as distinct bands that probably reflect the number repeating units of the sugar incorporated onto the carrier protein.

A weak smear from 38 to 198 kDa was revealed in the periplasmic fractions prepared from the negative controls. Such signal might be attributed to the intermediates of reaction *i.e.* Und-PP-linked CP5 (LLO) molecules, produced and assembled during the bioconjugation process. Several efforts were employed with the aim to purify, characterize and quantify such LLO intermediates of the biosynthesis of the saccharide chain. Nevertheless, this was a challenging task probably because of the low amount of material and its heterogeneity, in terms of both variable saccharide repeating units and dual chemical nature of LLO compounds (hydrophobicity of the lipid carrier and hydrophilic nature of the oligosaccharides). However, no conclusive data were obtained. A summary of the activities performed are provided and discussed in Appendix 2 of the present study.

These results indicated that the Western blot allowed a rough comparative analysis of the expression levels of Hla isoforms in cell periplasm and of the bioconjugation levels among the different carrier protein isoforms. By contrast, the quantification of the extent of glycosylation could not be evaluated

by this analysis. In fact, the scalar signals of conjugated proteins and the background interference make a densitometric analysis challenging. Moreover, the difference of specificity of the antibody for the carrier (un-modified or bioconjugated) and for the sugar moiety increases the complexity of the analysis.

Therefore, the application of a quantitative approach based on SRM-MS technique was needed.

Part II: Quantification of the extent of glycosylation by SRM approach

With the purpose to specifically quantify the glycosylation level of a target protein in a complex biological background, such as *E. coli* periplasmic extract, we employed SRM, a quantitative mass spectrometry technique for the detection and the quantification of specific, predetermined analytes with known fragmentation pattern (Picotti & Aebersold, 2012).

In particular, we applied the stable isotope dilution (SID) quantification method in which isotopically labeled peptides identical to those deriving from the tryptic digestion of the target protein (proteotypic peptides, PTPs) were used as surrogates for the final protein quantification. To be defined PTPs and to be suitable for the quantification of a specific protein, PTP should be unique for the target protein and owning well-defined MS detectability features (Lange *et al.*, 2008). In SID, known amount of isotopically forms of PTPs were spiked in the biological sample before digestion to increase the accuracy of the antigen quantification, to built-up a dose-response standard curve to assess the fitting of the linear regression model to the data obtained, and to achieve the quantitation of unknown targeted analytes in the samples.

The approach for the quantification of the extent of glycosylation applied in the present study was developed by Zhu et al. (Zhu *et al.*, 2014) for absolute quantification of the glycosylation site occupancy of commercial and purified eukaryotic bovine fetuin protein. This approach has been developed without the PNGase-F enzyme removal of the saccharide chain before the MS analyses In details, the step of deglycosylation of the protein by PNGase-F works causing deamidation of the asparagine formerly involved in the N-glycosylated peptide (Gonzalez *et al.*, 1992; Kuster & Mann, 1999; Segu *et al.*, 2010; Zhang *et al.*, 2003).This residue is converted in an Asp (D) residue and such peptide is measured by MS and assigned as the “deglycosylated” peptide. However, since chemical deamidation of asparagine could occur spontaneously during sample preparation, this could lead to misquantification of the peptides involved in N-glycosylation as long as non-glycosylated asparagine residues would be incorrectly assigned as product of deglycosylation by PNGase-F (Palmisano *et al.*, 2012).

The absolute quantification was performed by using two set of isotopically labelled peptides: one set was employed to quantify the total amount of the target protein in solution while the other set was

employed to evaluate the amount of unglycosylated protein by measuring the peptide carrying the consensus sequences which are not occupied by the sugar. The correlation between these two quantifications was described by the formula eq.1:

$$\text{Site Occupancy (\%)} = \frac{(Total - unmodified)\ carrier\ concentration}{Total\ carrier\ concentration} x100$$

and allow the quantification of the glycosylation site occupancy (Zhu *et al.*, 2014).

6.1 Selection of PTPs for total protein amount quantification of Hla Carrier protein

While the PTPs that characterize the consensus sequence for bioconjugation, as described in this study, could be universal and employed on diverse carriers (see paragraph 1.2.3), the PTPs necessary for the quantification of the total amount of carrier are specific for the protein of interest. The PTPs for the quantification of the protein total amount were experimentally selected from lysC/trypsin digestion of the purified unglycosylated Hla variant used for the reference bioconjugate Hla-CP5 and analyzed by LC-MS/MS in data dependent acquisition mode. The spectra obtained were processed using PEAKS X software against an *E. coli K-12* database containing the Hla variant protein sequence.

The results provided a list of thirty-one peptides covering 76% of the Hla amino acid sequence (table 4). From this list, three peptides were selected as PTPs (highlighted in red) representative for Hla. The PTPs were selected following the criteria described in literature for the selection of peptides for SRM analysis: intense MS signals, uniqueness for the target protein, low rate of missed cleavages sites and low rate of chemical or post-translational modifications (Picotti & Aebersold, 2012).

Peptide	Length	*m/z*	z	Area	PTM	% Hla Coverage
K.AADNFLDPNK.A	10	552.7707	2	1.92E+07		
K.QQTNIDVIYER.V	11	689.8526	2	1.82E+07		
K.YVQPDFK.T	7	448.73	2	1.62E+07		
A.SAADSDINIK.T	10	517.2632	2	1.58E+07		
K.TILESPTDK.K	9	502.2696	2	1.57E+07		
K.TGDLVTYDK.E	9	506.2547	2	1.55E+07		
K.VFYSFIDDK.N	9	567.2808	2	1.48E+07		
K.TGTTDIGSNTTVK.T	13	432.2219	3	1.30E+07		
K.SGLAWPSAFK.V	10	532.2838	2	1.22E+07		
G.GLIGANVSIGHTLK.Y	14	690.4052	2	1.08E+07		
K.VQLQLPDNEVAQISDYYPR.N	19	750.0482	3	6.64E+06		
R.VYSEEGANK.S	9	498.7363	2	4.18E+06		
K.TILESPTDKK.V	10	377.8808	3	3.50E+06		
K.ASSLLSSGFSPDFATVITMDR.K	21	734.6982	3	2.86E+06		
R.YKIDWEK.E	7	491.2572	2	2.29E+06		
A.ADSDINIK.T	8	438.2285	2	2.25E+06		82%
R.DSWNPVYGNQLFMK.T	14	566.9381	3	2.13E+06		
K.GTIAGQYR.V	8	433.2311	2	1.76E+06		
K.EEMTNGSHR.H	9	530.7288	2	1.69E+06		
K.ASSLLSSGFSPDFATVITM(+15.99)DR.K	21	740.0298	3	1.49E+06	Oxidation (M)	
R.DSWNPVYGNQLFM(+15.99)K.T	14	572.2698	3	1.21E+06	Oxidation (M)	
K.Q(-17.03)QTNIDVIYER.V	11	681.3395	2	6.71E+05	Pyro-glu from Q	
K.EEM(+15.99)TNGSHR.H	9	538.7271	2	4.31E+05	Oxidation (M)	
K.ASSLLSSGFSPDFATVITMDRK.A	22	777.3967	3	2.71E+05		
S.S(+210.99)LLSSGFSPDFATVITMDR.K	19	752.3355	3	1.63E+05	Tri nitro benzene	
R.VRDDYQLHWTSTNWK.G	15	650.3156	3	1.49E+05		
L.IGANVSIGHTLK.Y	12	605.3524	2	1.06E+05		
K.TGTTDIGSN(+.98)TTVK.T	13	648.3218	2	8.01E+04	Deamidation (NQ)	
I.LESPTDKK.V	8	459.2524	2	3.32E+04		
K.VQLQLPDNEVAQ(+.98)ISDYYPR.N	19	563.0394	4	2.25E+04	Deamidation (NQ)	
K.SGLAW(+31.99)PSAFK.V	10	548.2767	2	1.97E+04	Dihydroxy	

Table 4: ***List of 31 peptides obtained from the MS/MS spectra analysis by PEAKS X software.*** *In the table, peptide sequence, length,* m/z *value, charge, area and identified modified amino acids are reported.*

The selected peptides were 42T-50K (TGDLVTYDK, PTP-3), 225A-234K (AADNFLDPNK, PTP-2) and 58V-66K (VFYSDFIDDK, PTP-1). PTP-1 and PTP-3 are located at the N-terminal of the protein sequence, while PTP-2 is located proximally to the C-terminal.

Moreover, choosing peptides located in more than one region of the protein and quantifying them simultaneously and independently provides confidence in the accuracy of the quantification results (Williams *et al.*, 2012).

The synthetic peptides obtained for such PTPs were then investigated and the conditions of analysis were optimized as already described for PTP-i-s-v in paragraph 1.2.3. The best performing transitions selected are reported in table 5. Moreover, the PTP-1-2-3 resulted to have well distinct retention times in the chromatographic elution with respect to the designed PTP-i-s-v containing the consensus sequence.

Name	Sequence	Charge state	Q1 Precursor ion (m/z)	Q3 Fragment ions (m/z)	Fragment ions	CE predicted	CE optimized	RT (min)
PTP-1	VFYSFIDDK	2	567.28	724.35	y6	20	18	9.60
				887.41	y7	20	18	
				377.17	y3	20	24	
PTP-1*	VFYSFIDDK[HeavyK]	2	571.29	732.36	y6	20	18	9.60
				895.43	y7	20	18	
				385.18	y3	20	24	
PTP-2	AADNFLDPNK	2	552.78	473.23	y4	18	18	8.30
				586.32	y5	18	18	
				962.46	y8	18	18	
				733.39	y6	18	18	
				358.21	y3	18	18	
PTP-2*	AADNFLDPNK[HeavyK]	2	556.78	481.25	y4	18	18	8.30
				594.33	y5	18	18	
				970.47	y8	18	18	
				741.40	y6	18	18	
				366.22	y3	18	18	
PTP-3	TGDLVTYDK	2	506.25	526.25	y4	18	18	7.42
				625.32	y5	18	18	
				853.43	y7	18	18	
PTP-3*	TGDLVTYDK[HeavyK]	2	510.26	534.26	y4	18	18	7.42
				633.33	y5	18	18	
				861.44	y7	18	18	

Table 5: ***List of optimized SRM transitions for the PTPs selected for quantification of total amount of the protein.*** *For each PTP, PTP's name, peptide sequence, charge state, molecular mass, and the optimized transition and chromatographic condition (RT) were reported. For each selected fragment, the reproducibility of detection was assessed monitoring the TIC signal.*

6.2 Quantification of the extent of glycosylation in single glycosylated carrier protein

The developed SRM assay was used to measure the bioconjugation level of Hla carrier protein isoforms Hla-i-CP5, Hla-v-CP5 and Hla-s-CP5. Each quantification was deduced from three independent digestions, analyzed in triplicate by LC-MS/MS.

Light and heavy pairs of synthetic peptides were used to build-up a dose-response curve to calculate the amount of the unknown analytes. In particular, eight dilution points of the light form of peptides (final concentration ranging from 0.0125 to 1.6 pmol/µL) were prepared and mixed with a fixed amount of the heavy forms (final concentration 0.1 pmol/µL) in 50 µg of matrix prior digestion with lysC/trypsin.

To ensure precise quantification, scheduled SRM-MS analysis was employed to obtain high sensitivity and reproducibility, and to acquire enough data points to enhance the accuracy of the measurements (Picotti & Aebersold, 2012). In details, the settings of dwell time and cycle time to acquire enough data points were optimized based on the peak width given by the chromatographic elution profile.

A cycle time of 1.02 s with .03 s of dwell time for 34 transitions in total for the 6 peptides selected were set-up. By applying these parameters, it resulted an acquisition of 7 data points per peak across the elution profile.

The data generated by LC-SRM were analyzed by TargetLynX software (Waters®) and the integrated peak areas of transitions for the unlabeled peptide analyte (the light-labeled peptides in the dilution

points of the standard curve) were summed and then normalized to the summed peak areas of transitions from the corresponding heavy-labeled peptide standard. Such “peak area ratios” were the measured values to be interpolated into the dose-response standard curve to obtain the amount of the analyte in unknown samples. The dose-response standard curves of the PTP1-2-3 for the quantification of the total amount of Hla and the PTP-i-v-s for the quantification of the unglycosylated form of the protein are provided in figure 11, as well as slopes of the regression curves and the coefficient R^2.

The quantification values are expressed in pmol/µg, referring as pmol of PTPs detected for 1 µg of periplasmic proteins.

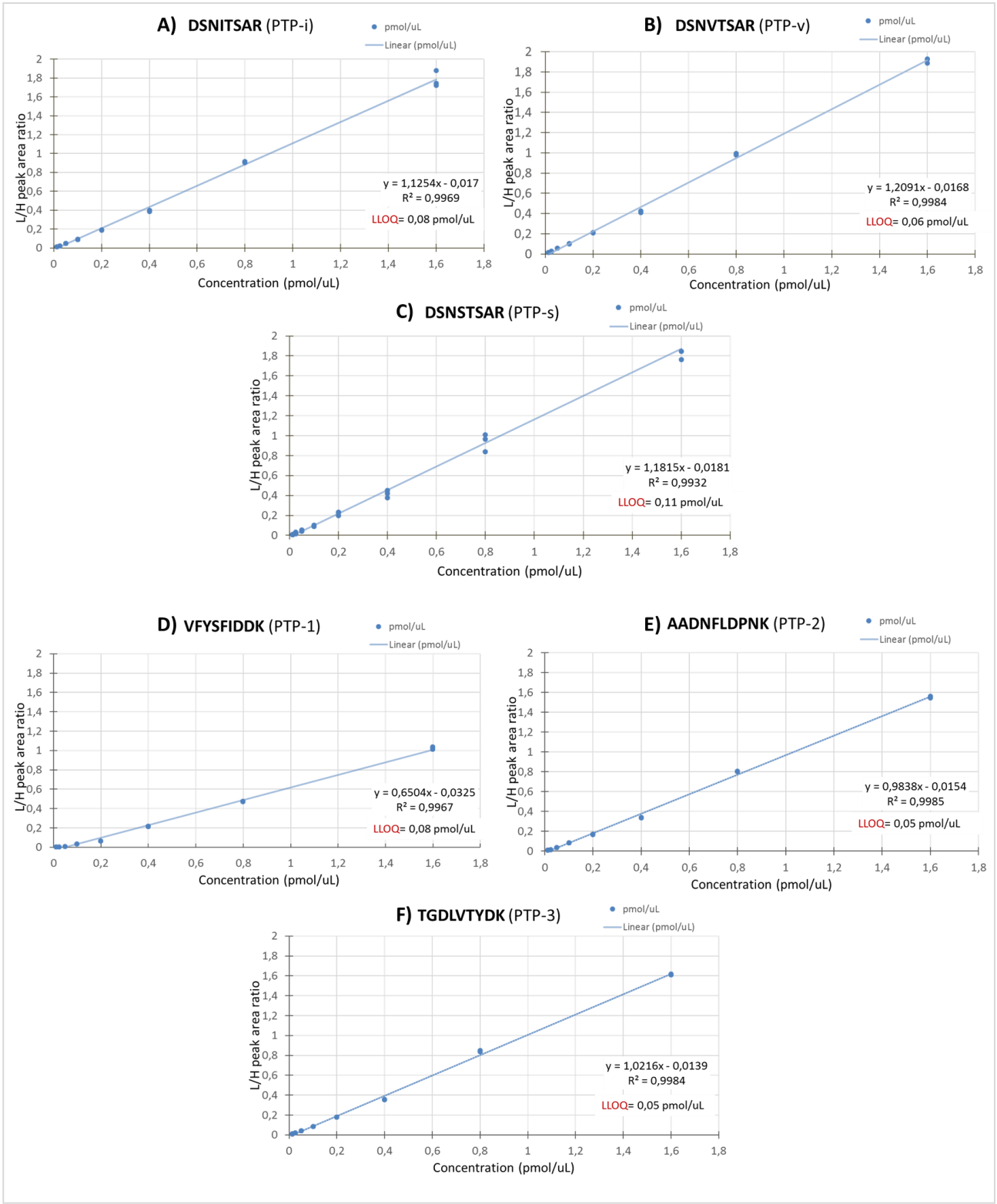


Figure 11: ***Dose-response curve of the PTPs under analysis.*** *The LLOQ considered was at 0,05 pmol/µL. Considering an average Hla isoforms molecular weight of 34,900 Da, the limit of quantitation is determined as 1.88 ng per 1 µg of periplasmic total protein concentration.*

It is worth of note in PTP-1 curve (fig 11D) that, although the linear regression model is maintained to describe the data analyzed (R^2 >0.99), the measured value responses and the concentration spiked in samples did not comply with the relation expected among them. Such observation could be

appreciated at the mid-concentration point of the curve (0.8 pmol/µL), where the equimolar amount of the light and heavy forms spiked in matrix should reflect the unitary relation in terms of signal acquired for the Ion peak area. Further analyses might be performed to assess the quantitative amount of PTP-1 standard (light and heavy form) provided by the supplier or to assess the stability of the peptide. For this reason, the PTP-1 was excluded for the quantification of the total amount of Hla.

The LLOQ for each peptide was set as the lowest concentration point on the fitted curve that can be quantitively detected and defined as 10 σ/S, where σ = the standard deviation of the response and S = the slope of the calibration curve by the International Conference on Harmonization (ICH) Guidelines (http://www.ich.org/products/guidelines/quality/article/quality-guidelines.html) Also the LOD or Limit of Detection (LOD = 3.3 σ/S) was defined based on ICH Guidelines.

For each isoform, the highest LLOQ calculated based on the molecular weight of the protein variant in the sample was considered and used as threshold for quantification.

The values of LOD, LLOQ, R^2, the slope and intercept of the curves are reported in table 6. All the R^2 of the curves were >0.99, confirming the fitting of the linear regression model to describe the data collected.

Protein isoform	MW (Da)	LLOQ ng/µg				
		PTP-2	PTP-3	PTP-i	PTP-v	PTP-s
Hla-i	34,969	1.88	1.95	**2.82**		
Hla-v					**2.23**	
Hla-s						**4.02**
Protein isoform	MW (Da)	LOD ng/µg				
Hla-i	34,969	0.62	0.64	**0.93**		
Hla-v					**0.73**	
Hla-s						**1.33**
		R2				
		0.9985	0.9984	0.9969	0.9984	0.9932
		Slope				
		0.9838	1.0216	1.1254	1.2091	1.1815
		Intercept				
		-0.0154	-0.0139	-0.0170	-0.0168	-0.0181

Table 6: ***LOD, LLOQ, R2, the slope and intercept of the dose-response curves of Fig. 11.*** *In bold black are reported values of the highest LOQ and LOD considered for the quantification of Hla isoforms by quantification of surrogate PTPs, both in pmol/µL and in ng of Hla per µg of total periplasmic protein amount.*

The quantification obtained are reported in pmol/µg in table 7.

	Total Hla							unmodified glyco-site		
	Dig1	Dig2	Dig3		Dig1	Dig2	Dig3	Dig1	Dig2	Dig3
	PTP-2 (AADNFLDPNK)				PTP-3 (TGDLVTYDK)			PTP-i (DSNITSAR)		
Hla-i	0,732	0,604	0,885		0,669	0,516	0,818	0,434	0,286	0,526
	0,719	0,603	0,879		0,665	0,512	0,817	0,418	0,297	0,545
	0,729	0,604	0,885		0,670	0,512	0,828	0,410	0,295	0,534
	Average				Average			Average		
	0,738				0,667			0,416		
	CV (%) run				CV (%) run			CV (%) run		
	0,90	0,15	0,36		0,35	0,44	0,77	2,98	1,96	1,73
	CV (%) digestion				CV (%) digestion			CV (%) digestion		
	16,43				20,00			25,29		
	% site occupancy = 40%									
	PTP-2 (AADNFLDPNK)				PTP-3 (TGDLVTYDK)			PTP-v (DSNVTSAR)		
Hla-v	0,561	0,475	0,845		0,522	0,434	0,805	0,288	0,231	0,497
	0,561	0,477	0,850		0,519	0,435	0,794	0,293	0,243	0,450
	0,540	0,467	0,865		0,523	0,434	0,796	0,291	0,249	0,475
	Average				Average			Average		
	0,627				0,585			0,335		
	CV (%) run				CV (%) run			CV (%) run		
	2,19	1,02	1,19		0,46	0,15	0,71	0,80	3,76	4,91
	CV (%) digestion				CV (%) digestion			CV (%) digestion		
	27,72				28,20			31,96		
	% site occupancy = 44%									
	PTP-2 (AADNFLDPNK)				PTP-3 (TGDLVTYDK)			PTP-s (DSNSTSAR)		
Hla-s	0,692	0,546	0,787		0,623	0,502	0,727	0,336	0,286	0,451
	0,692	0,578	0,789		0,627	0,499	0,741	0,372	0,294	0,429
	0,687	0,545	0,777		0,628	0,498	0,733	0,387	0,321	0,474
	Average				Average			Average		
	0,677				0,620			0,372		
	CV (%) run				CV (%) run			CV (%) run		
	0,47	3,40	0,78		0,39	0,47	0,94	7,20	6,18	5,04
	CV (%) digestion				CV (%) digestion			CV (%) digestion		
	14,76				16,39			18,39		
	% site occupancy = 42%									

Table 7: ***Hla isoforms PTPs quantification in pmol/ug.*** *Total amount of the protein was obtained by the quantification of PTP-2 and PTP-3; the amount of the unglycosylated form by PTP-i, PTP-v and PTP-s, one for each isoforms respectively. Values are expressed in pmol of peptides detected for 1* µg of protein. *Site occupancies were calculated by applying the eq. 1.*

The quantifications of the PTPs were reproducible, within-run coefficient of variation (CV) less than 4% in (in 22 out of 27 data) and almost all the CVs among the three different digestions inferior to 30% (8 out of 9). These values are in line with CVs observed in inter-laboratories studies reported in literature, demonstrating the precision and the reproducibility of the quantification achievable in complex digests (Anderson & Hunter, 2006; Huttenhain *et al.*, 2009; Kuzyk *et al.*, 2009).

Considering an average MW of the Hla isoforms of 34,960 Da, the total amounts of the three isoforms were quantified across 3 SRM experiments as 24.2, 20.9 and 22.4 ng/µg of total periplasmic proteins for for Hla-i, Hla-v and Hla-s, respectively. The overall amounts of Hla in periplasmic extracts were comparable between the three strains cultured simultaneously, highlighting comparable yields of carrier produced and reflecting the minor influence of the different consensus sequences on the bioconjugate production.

The amount of unglycosylated isoforms was deduced from the quantification of the designed PTP-i, PTP-v, PTP-s. Site occupancies were calculated by applying the eq. 1 and the values obtained were 40%, 44% and 42% for Hla-i-CP5, Hla-v-CP5 and Hla-s-CP5, respectively (fig. 12).

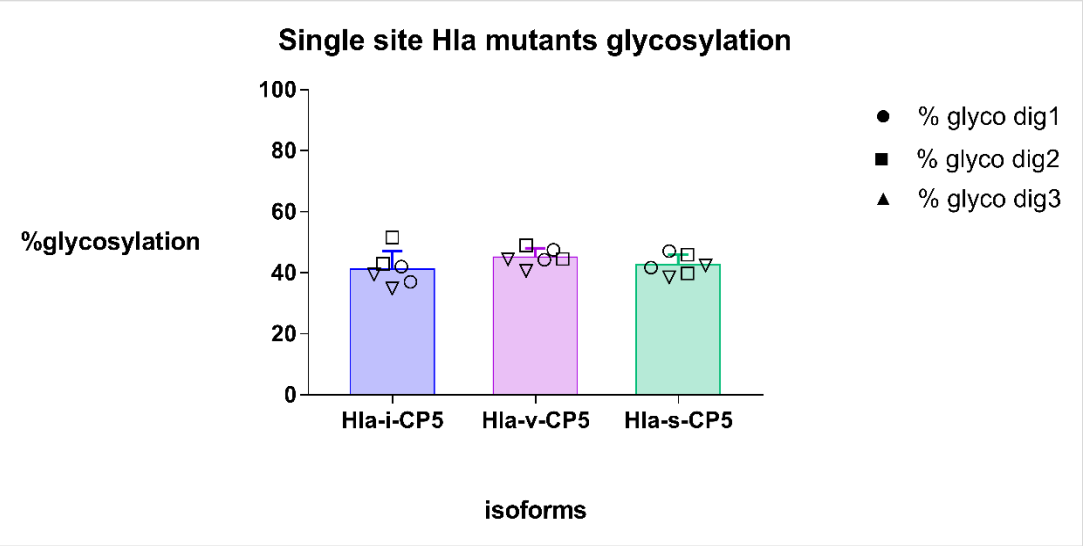


Figure 12: ***Percentage of Hla isoforms glycosylation measured by SRM approach reported in table 7.*** *The data were acquired in 3 SRM experiments; each bar represents the amount of the glycosylated carrier protein both in % of glycosylation and with respect to the ng of la isoform quantified by SRM approach reported in table 7. Each point on the graphs represents the values of the % of glycosylation between digestion (run 1, run 2 and run 3) calculated considering PTP-2 and PTP-3. The error bars are set at the mean and at the SD among the different digestions is shown.*

The extents of glycosylation assessed by SRM approach were in the same range for the three isoforms, consistent with the qualitative patterns observed in the Western blot analysis in fig. 10 (lanes 5, 6 and 7).

Moreover, a relationship between percentage of glycosylation and total amount of Hla in the periplasm could be graphed (Fig 13). The higher the amount of Hla, the lower the extent of conjugation was observed. This observation might reflect the lack of efficacy of PglB enzyme to glycosylate increasing amount of substrate, although at this stage of the analysis it could not be excluded that the difference of the observed glycosylation percentage was due to the different sequence amino acid residues in position +1 of the consensus sequences.

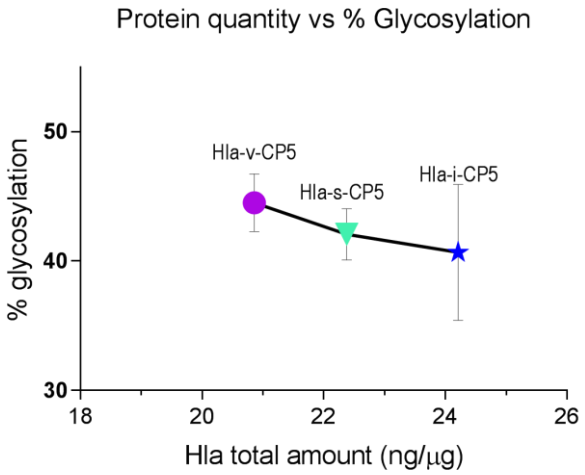


Figure 13: ***Relationship between percentage of glycosylation and total amount of Hla.*** *Quantification values are reported in ng of Hla protein in 1 µg of periplasmic extract.*

6.3 Quantification of the extent of glycosylation in double glycosylated carrier protein

As already mentioned in the paragraph 2.3 of the introduction section, one of the strategies to improve the immunogenicity of the glycoconjugate produced by bioconjugation technology is the engineering of the carrier protein with multiple consensus sequences to increase the number of oligosaccharides chains on the bioconjugate molecules. With the purpose to increase the valency of glycosylation in the carrier protein, the analytical driven approach described in part I was applied to generate Hla carrier protein isoforms with two glycosites of bioconjugation.

To insert the second consensus sequence, several aspects were taken into consideration:

1. the preservation of the protective epitopes of Hla as antigen;
2. the dynamics and solvent-exposed regions of Hla protein.

6.3.1 Preservation of the known protective epitopes

The Staphylococcal Hemolysin A, also known as α-toxin or Hla, is a toxin and thus needs to be detoxified in order to be used as a vaccine antigen. Wild-type Hla monomers assemble to form a heptamer in presence of lipid-bilayer of cell membranes, creating a penetrating pore in the membrane of human erythrocytes which cause cell lysis (Gouaux *et al.*, 1994). Such hemolytic activity of Hla has been inhibited by mutation of amino acid residues involved in pore formation, which are located in the N-terminal domain. As described in literature (Menzies & Kernodle, 1994; Walker & Bayley, 1995), one single amino acid substitution mutant, called HlaH35L, showed greatly reduced heptamer formation, no hemolytic activity and no toxicity when administered to mice. HlaH35L has been used

in experimental vaccines against *S. aureus* infection, including to produce the bioconjugate molecule described by Wacker *et al.* (Wacker *et al.*, 2014).

From a structural standpoint, apart for the major differences in the stem and in the latch region of Hla, the structure of the monomer was demonstrated to be nearly identical to the monomer derived from the heptameric structure (Foletti *et al.*, 2013).

For the insertion of the additional consensus sequences in Hla regions, we considered location far away from the epitopes recognized by two functional and well characterized mAbs LMT14 and MEDI4893 (Foletti *et al.*, 2013; Oganesyan *et al.*, 2014). LMT14 has been shown to inhibit the Hla binding to the cell membranes both directly and through ADAM10 receptor, binding a non-linear epitope between the cap and the rim domain (fig. 14, in dark and light magenta) (Foletti *et al.*, 2013). MEDI4893 is a therapeutic mAb in phase 2 clinical trial that has been demonstrated to sterically block the binding of Hla to the membrane, by preventing the interaction with ADAM10 of the rim domain in the opposite side if compared to LTM14 and inhibiting the formation of the heptameric pore responsible of the lytic activity (fig. 14, in dark and light cyan) (Oganesyan *et al.*, 2014).

When mapped on the x-ray crystal structure of the Hla H35L monomer (PDB code: 4IJD), the two mAbs recognize epitopes located on the same half of the protein (fig. 14). In details, amino acid residues Arg66, Glu70, Arg200, Asp255, and Glu276 are involved in the binding with mAb LMT14. The two segments spanning residues Asn177–Arg200 and Thr261–Lys271 are recognized by mAb MEDI4893 (Foletti *et al.*, 2013; Oganesyan *et al.*, 2014).

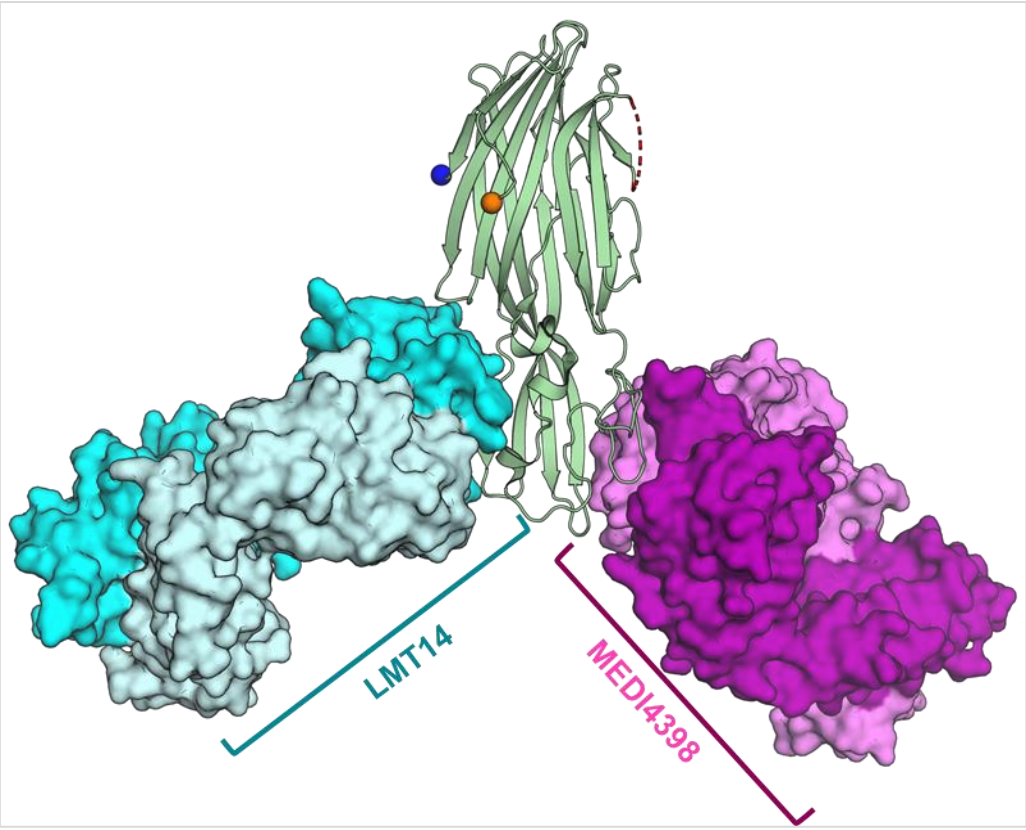


Figure 14: ***X-ray crystal structure reconstruction of Hla monomer in complex with MEDI4893 (PDB code: 4U6V) and LMT14 (PDB code: 4IJD)****. The Hla monomer structure is reported in ribbon representation: in red is colored the position of the reference glycosylation site pos130 (Wacker et al., 2014) while the spheres represent the N-terminal and in blue the C-terminal; mAbs are shown in surface representation with the heavy chain and the light chain in dark and light color, respectively.*

6.3.2 The dynamics and solvent-exposed regions of Hla protein

The regions of Hla protein presenting different dynamics, as folded or unfolded domains, flexible and solvent-exposed or quite structured regions, were studied for the design of Hla constructs bearing two glycosites.

The combination of amide hydrogen/deuterium exchange (HDX) with mass spectrometry (MS) is becoming a favored tool for characterizing protein dynamics (Konermann *et al.*, 2011). HDX is a chemical reaction in which a covalently bonded hydrogen atom is replaced by a deuterium atom, or vice versa when a protein is incubated in a deuterated buffer. The HDX provides structural information based on solvent accessibility of peptide bonds in tertiary and quaternary structure of each protein, because exchange rates between hydrogen and deuterium in proteins depend on the degree of a proton's exposure to the protein surface and flexibility of the surrounding tertiary structure in permitting access of protons to solvent. By HDX-MS is possible to analyze proteins in solution in native condition, in complex mixture and using small amount of material (Donnarumma *et al.*, 2018). Hydrogen/deuterium exchange measurements can be used to sense changes in protein structure on a specific timescale. In structured species, the exchange rate of amide hydrogens depends on the degree

of exposure to the solvent. For instance, amide hydrogens located at the surface of proteins exchange very rapidly, while those buried in the hydrophobic core of the protein may not exchange for hours, days, or months. Therefore, the movements of proteins, and the rate of such movements, disclose the flexibility of the protein. The determination of the deuterium incorporation levels on each peptide allows to achieve a precise definition of the dynamics of restricted portion of the protein under investigation. For unstructured species, as proteins in solution containing chaotropic agents or unstructured peptides, the extent of deuteration solely depends on the deuterium content in the reaction mixture and on their primary amino acid sequence. In details, considering every backbone amides completely exposed to the solvent, hydrogens fully exchange with the deuterium atoms in a few seconds or milliseconds.

In order to evaluate the flexibility and solvent exposure of the different regions of Hla, we performed a HDX-MS experiment on the available purified wild-type variant of the protein, because of its monomeric character in solution and of the high identity with the monomer employed by Wacker *et al.* (Wacker *et al.*, 2014).

Hla was incubated in deuterated PBS buffer (89% deuterium content in the reaction mixture) for 0.25 min and 100 min on ice. The deuterium incorporation was stopped by quenching the reaction through dropping the pH to 2.5 and the temperature to 0°C, then the samples were quickly frozen. The labelled protein was on-line digested with pepsin, and the resulting peptides were analyzed by LCMS. In parallel, Hla was digested on-line and peptides eluting from the pepsin column were collected. The aqueous solvent was evaporated and the dry peptides resuspended in deuterated PBS buffer (89% deuterium content identical to the intact protein) for 1h on ice, before quenching the exchange reaction. In this way, the backbone amide hydrogens of the unstructured peptides generated by Hla digestion were fully exchanged with deuterium, obtaining the so-called “maximally labelled control”. In details, comparing the conformation and dynamics of the intact protein and its unstructured species, it was possible to discriminate regions fully solvent exposed and highly dynamic with respect to the structured ones.

Sixty-five peptides covering 98.3% of the protein sequence were identified and analyzed. The extent of deuterium incorporation of the various sub-localized protein regions was calculated as percentage of the deuterium uptake of the maximally labelled control. These fractional values are depicted on the available crystallographic structure of Hla monomer (PDB code: 4IJD) in fig 15.

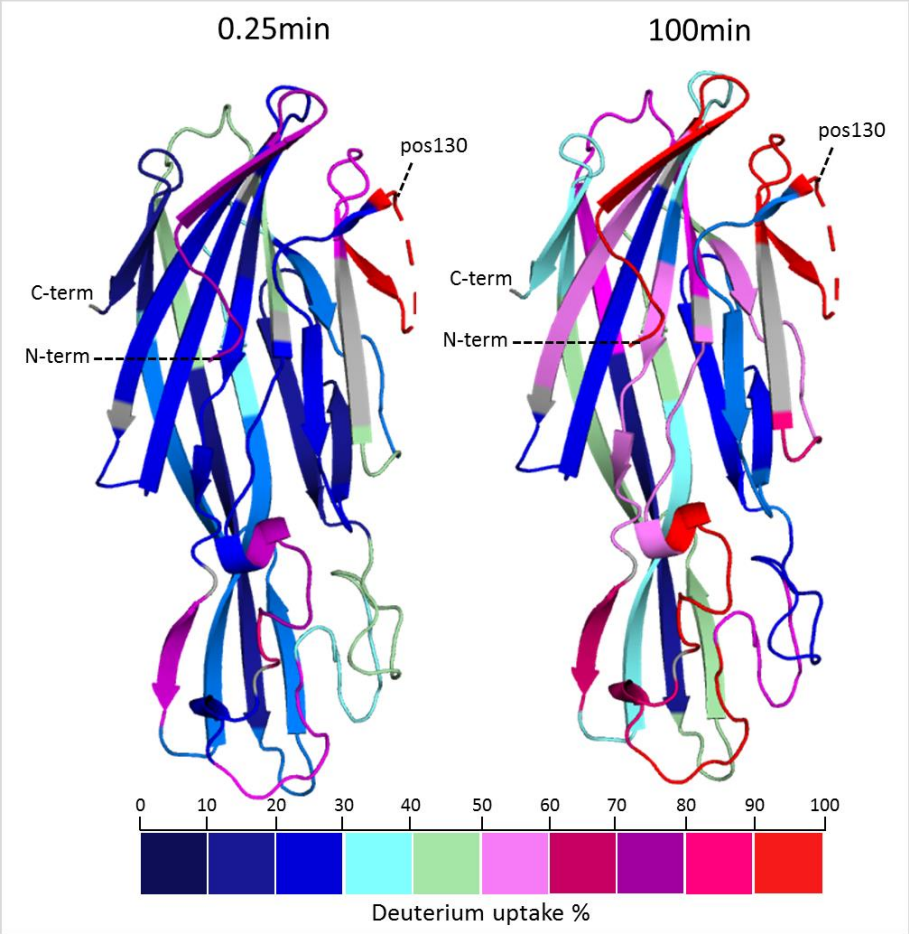


Figure 15: ***Extent of deuterium incorporation*** *of the sub-localized protein regions calculated as percentage of deuterium uptake of the maximally labelled control. The pos130 indicates the location of the single glycosite inserted in Hla-I, Hla-v and Hla-s isoforms. The ribbon diagram of the protein is colored following the deuterium uptake range from 0 to 100% showed below obtained in the two time points performed (0.25 min and 100 min of deuteration).*

The data agreed with the available crystal structure (PDB code: 4IJD). As represented in fig 15, the N-terminal protein region spanning residues 1N-13T (peptide 1NIKTGTTDIGSNT13) and the region encompassing residues 127V-142N (peptide 127VTGDDTGKIGGLIGAN142) showed almost similar (1N-13T) and identical (127V-142N) deuterium incorporation after 0.25 min and 100 min of labelling. Moreover, the deuterium uptake of these peptides resulted comparable to the maximally labelled peptides. These data suggested that those regions are fully solvent exposed and highly dynamic. Remarkably, the regions corresponding to the glycosite in pos130 and five amino acid residues at N-terminal end of the protein lack of crystal structure determination, indicating extremely flexible and disordered domains (Foletti *et al.*, 2013). Conversely, the C-terminal region reached only 10% and 40% of deuterium uptake at 0.25 min and 100 min respectively, revealing that this domain is highly structured and folded.

6.3.3 Design of Hla construct harboring two glycosylation sites.

In addition to the consensus sequence in position 130 of Hla protein (Wacker *et al.*, 2014), for the design of Hla constructs harboring two glycosites we decided to insert new consensus sequences adding eight amino acid residues at the N-terminal or at C-terminal regions of the protein.

The N-terminal resulted fully solvent-exposed and highly dynamic and therefore optimal for the insertion of the glycosite, since it has been demonstrated that the structural context of the target sequons could influence the PglB efficiency (Silverman & Imperiali, 2016).Moreover, the insertion of consensus sequence on the N-terminal region is consistent with several studies reporting the co-translocational mechanism of PglB enzymatic process (Silverman & Imperiali, 2016). By contrast, although the C-terminal region resulted structured and folded, the choice of this region to insert the consensus sequence was driven by evidences that PglB is able to perform N-glycosylation also on folded protein (Fisher *et al.*, 2011).

Furthermore, the new glycosites resulted to be located on the opposite Hla side of epitope recognized by mAbs LMT14 and MEDI4893 (Fig. 14), in order to preserve the key epitopes of such well-characterized mAbs.

Three constructs were designed to carry PTP-i and PTP-v peptides that resulted well separated during the LC-MS/MS chromatographic setup. The consensus sequences were inserted in alternative combination at N-terminal domain, at C-terminal domain and in the position at amino acid 130 (pos130), as represented in a schematic illustration in Fig. 16. The constructs were named Hla-ds1 (double site construct 1), Hla-ds2 (double site construct 2) and Hla-ds3 (double site construct 3).

Moreover, the design strategy took into account to insert preferably the same consensus sequence in the same position in more than one construct, in order to evaluate the possible relations among different constructs. In details, PTP-i was located in pos130 in two constructs (Hla-ds1 and Hla-ds2) and PTP-v at N-terminal in Hla-ds1 and Hla-ds3.

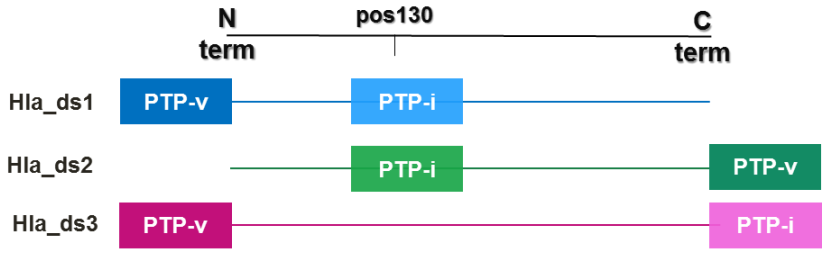


Figure 16: ***Schematic illustration of the constructs designed for Hla carrying multiple glycosylation sequences****. PTP-i and PTP-v are alternatively located at N-term, C-term and in the reference position 130 on the carrier protein. In black, a schematic illustration of the Hla protein sequence without any glycosite. In blue is reported Hla-ds1, in green the Hla-ds2 and in pink the Hla-ds3 construct. The glycosites are colored in light and dark color according to the construct.*

6.3.4 Generation and characterization of Hla mutants carrying two glycosylation sites.

The DNA plasmid backbones containing the encoding sequences for the designed double sites constructs were obtained from GeneArt® supplier and the DNA inserts of Hla constructs were cloned by PIPE method, as reported above. The forward and reverse primers used for amplification of the DNA inserts were designed to anneal on the target nucleotide sequence for the initial 20 bp and to anneal on the destination vector (pEC415) for the following 20-30 bp. The I-PIPE PCR products were mixed to the PCR products of the V-PIPE performed on the destination vector to linearize the destination plasmid. The transformation of chemically competent *E. coli* topTEN (Thermo®) by using such mutagenesis PCR reaction products allowed to obtain colonies carrying the specific plasmids constructs (Klock & Lesley, 2009).

Plasmids from selected clones were purified and sequenced to confirm and select the Hla constructs of interest. Glycocompetent *E. coli* (st9268, p1221) bearing the machinery for glycosylation by PglB were transformed with plasmids bearing Hla-ds1, Hla-ds2 and Hla-ds3 constructs, to produce Hla-CP5 bioconjugate isoforms with two glycosites. In table 11 of the Materials and Methods section are reported the strains, the primer sequences used and the resulting plasmids; in paragraph 7.1.2 and 7.2 are reported the growth conditions and the periplasmic extraction protocol, respectively

6.3.5 Quantification of the extent of glycosylation and differential site occupancy evaluation by LC-SRM

The LC-SRM MS analyses were performed in the same experimental conditions already described for the single glycosylated constructs (part I of this chapter). Also in this case, the three combined isoforms were analyzed with three independent LysC-trypsin digestion and for each, the CVs inter-run and inter-digestion were investigated. All the obtained results were summarized in table 9.

For each isoform, the highest LLOQ was considered and used as threshold for quantification, calculated on the basis of the molecular weight of the protein variant in the sample. The values of LOD, LLOQ, R^2, the slope and intercept of the curves are reported in table 8. All the R^2 of the curves were >0.99, confirming the fitting of the linear regression model to describe the data collected.

Protein isoform	MW (Da)	LLOQ ng/µg			
		PTP-2	PTP-3	PTP-i	PTP-v
Hla-ds1	37,049	1.98	2.05	**2.19**	1.96
Hla-ds2	36,606	1.96	2.02	**2.17**	1.94
Hla-ds3	37,478	2.01	2.07	**2.22**	1.99
Protein isoform	MW (Da)	LOD ng/µg			
Hla-ds1	37,049	0.65	0.67	**0.72**	0.65
Hla-ds2	36,606	0.65	0.67	**0.72**	0.64
Hla-ds3	37,478	0.66	0.68	**0.73**	0.66
		R2			
		0.999	0.999	0.998	0.999
		Slope			
		0.964	0.911	0.842	1.042
		Intercept			
		-0.017	-0.010	-0.011	-0.007

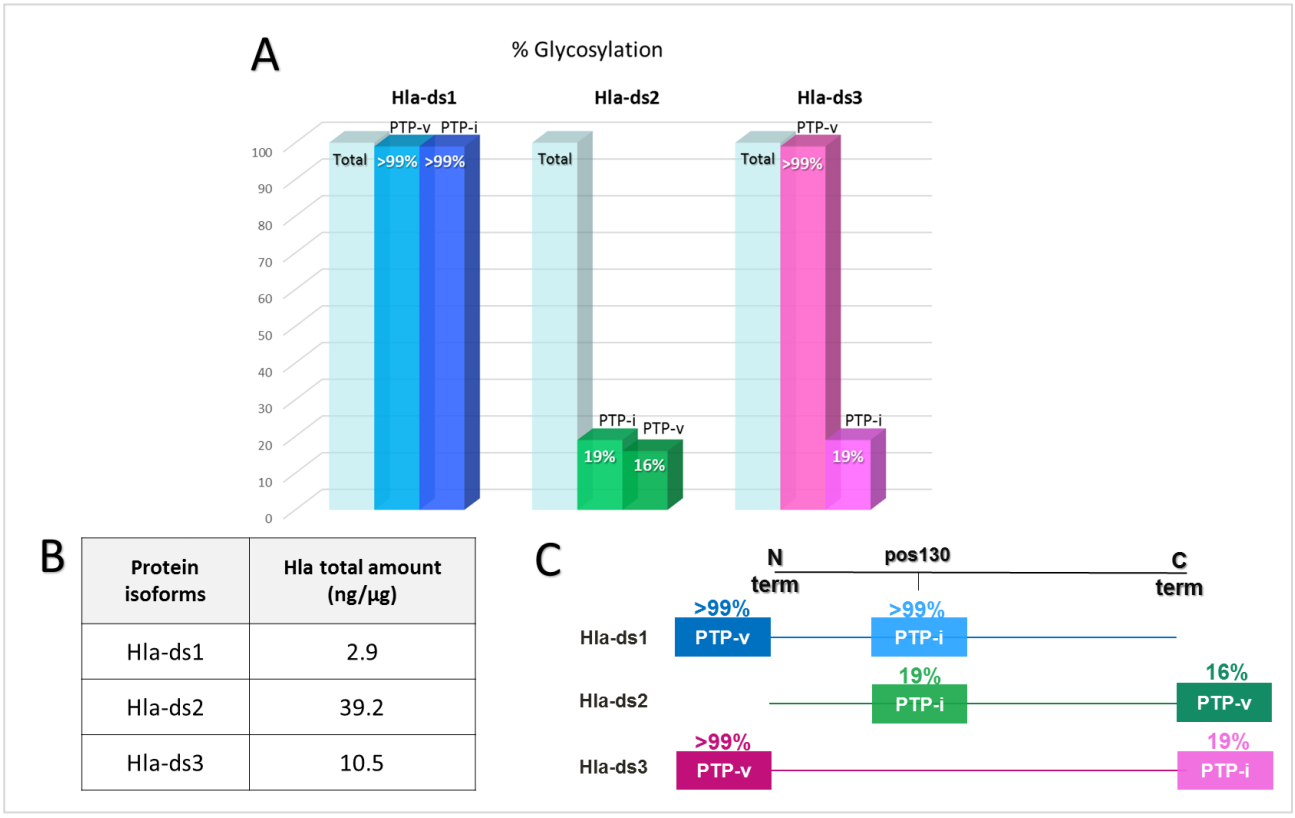
Table 8: ***LOD, LLOQ, R2, the slope and intercept of the dose-response curves.*** *In bold black are reported values of the highest LOQ and LOD considered for the quantification of Hla-ds isoforms by quantification of surrogate PTPs, expressed in ng of Hla per µg of total periplasmic protein amount.*

	Total Hla						unmodified glyco-site			unmodified glyco-site		
	Dig1	Dig2	Dig3	Dig1	Dig2	Dig3	Dig1	Dig2	Dig3	Dig1	Dig2	Dig3
	PTP-2 (AADNFLDPNK)			PTP-3 (TGDLVTYKD)			PTP-i (DSNITSAR)			PTP-v (DSNVTSAR)		
Hla_ds1	0,094	0,075	0,079	0,086	0,065	0,072	<LLOQ			<LLOQ		
	0,095	0,074	0,079	0,086	0,066	0,071						
	0,095	0,073	0,080	0,087	0,065	0,071						
	Average			Average			Average			Average		
	0,08			0,07								
	CV (%) run			CV (%) run			CV (%) run			CV (%) run		
	0,86	1,71	0,39	0,66	0,96	1,29						
	CV (%) digestion			CV (%) digestion			CV (%) digestion			CV (%) digestion		
	11,38			12,39								
	Siete occupancies						*>99%*			*>99%*		
	PTP-2 (AADNFLDPNK)			PTP-3 (TGDLVTYKD)			PTP-i (DSNITSAR)			PTP-v (DSNVTSAR)		
Hla_ds2	1,119	0,956	1,152	1,079	0,922	1,123	0,888	0,748	0,955	0,955	0,759	0,940
	1,125	0,938	1,167	1,068	0,927	1,114	0,885	0,742	0,962	0,949	0,771	0,979
	1,129	0,947	1,163	1,075	0,896	1,121	0,904	0,737	0,930	0,952	0,732	0,987
	Average			Average			Average			Average		
	1,08			1,04			0,86			0,89		
	CV (%) run			CV (%) run			CV (%) run			CV (%) run		
	0,4124	0,9443	0,6438	0,54	1,82	0,42	1,15	0,77	1,80	0,31	2,68	2,59
	CV (%) digestion			CV (%) digestion			CV (%) digestion			CV (%) digestion		
	9,22			10,36			10,80			11,73		
	Site occupancies						*19%*			*16%*		
	PTP-2 (AADNFLDPNK)			PTP-3 (TGDLVTYKD)			PTP-i (DSNITSAR)			PTP-v (DSNVTSAR)		
Hla_ds3	0,260	0,325	0,308	0,240	0,299	0,284	0,205	0,242	0,238	<LLOQ		
	0,257	0,325	0,315	0,241	0,294	0,281	0,199	0,246	0,241			
	0,258	0,319	0,312	0,240	0,296	0,279	0,201	0,252	0,251			
	Average			Average			Average			Average		
	0,30			0,27			0,23					
	CV (%) run			CV (%) run			CV (%) run			CV (%) run		
	0,5400	1,0600	1,0782	0,29	0,88	0,92	1,52	2,05	2,69			
	CV (%) digestion			CV (%) digestion			CV (%) digestion			CV (%) digestion		
	10,10			10,55			9,66					
	Site occupancies						*19%*			*>99%*		

Table 9: ***Hla isoforms PTPs quantification in pmol/µg.*** *Total amount of the protein was obtained by the quantification of PTP-2 and PTP-3; the amount of the unglycosylated form by PTP-i and PTP-v for each respective glycosylation site. Values are expressed in pmol of peptides detected for 1 µg of protein. Site occupancies were calculated by applying the eq. 1.*

Analyzing the results, the produced acceptable quantitative data demonstrated within-run CV <2% in the 82% of analyses (n=27) and CVs inferior to the 12% associated to the inter-digestion data. These values are in line with CVs observed in inter-laboratories studies reported in literature, demonstrating the precision and the reproducibility of the quantification achievable in complex digests (Anderson & Hunter, 2006; Huttenhain *et al.*, 2009; Kuzyk *et al.*, 2009).

The total amount of the three Hla isoforms in the periplasmic fraction resulted to be 2.9 ng/µg of periplasmic extract for Hla-ds1, 39.2 ng/µg for Hla-ds2, that resulted the most expressed isoform, and 10.5 ng/µg for Hla-ds3 that is (fig. 18B). The data collected for each peptide were used for the quantification of extent of glycosylation. Considering the amount of unglycosylated PTP-i and PTP-v, the site occupancy was deduced in each glycosylation site by applying eq. 1 (table 9 and in fig.18). In our experimental conditions, the constructs Hla-ds1 and Hla-ds3 showed the lower level of expression of Hla protein in comparison to the Hla-ds2. These constructs harbor the PTP-v glycosite at N-terminal end of the protein sequences suggesting that the rate of the protein expression and the PglB glycosylation process could be mutually influenced.



Protein isoforms	Hla total amount (ng/µg)
Hla-ds1	2.9
Hla-ds2	39.2
Hla-ds3	10.5

Figure 17: (A) Site occupancies (%) for each of the different glyco-sites in the Hla isoforms carrying double consensus sequences for glycosylation, as reported in panel C. Light colors represent the PTP-i glycosite occupancies, while darker colors the PTP-v occupancies. In panel B the total amount quantification of the three Hla isoforms is reported.

The following aspects could be evinced:

i) For Hla-ds1, both peptides containing the glyco-sequence (PTP-i and PTP-v) resulted not quantifiable with values detected under the LLOQ of the analysis. For this reason, both sites were considered fully glycosylated;

ii) Hla-ds2 resulted to be glycosylated ~20% in both PTP-i (19%) and PTP-v glyco-sites (16%);

iii) Hla-ds3 isoform resulted to be completely glycosylated in PTP-v since no quantitative values were measured for such peptide, while it resulted to be ~19% glycosylated in PTP-i.

By the comparison of the quantification data obtained, the overall read-out is that when the glycosite is located at the N-terminal domain of the carrier protein, it results to be fully glycosylated rather than when it is located at C-terminal end (as observed in Hla-ds1 and Hla-ds3 for PTP-v). This is consistent with the studies supporting the hypothesis that N-glycosylation operated by PglB in bacterial cells occurs simultaneously to the translocation of the protein in periplasm (Silverman & Imperiali, 2016). By contrast, when the first glycosite translocated into the periplasm is in a central (pos130) and the second at C-terminal position (as in Hla-ds2, and in Hla-ds3 for the PTP-v at C-term), a partial folding of the protein could occur probably affecting the PglB ability to directly interact with the glycosite. This hypothesis could be a confirmation that a more efficient N-glycosylation requires a partial unfolding of the carrier protein as well as the presence of the consensus sequence in a sufficiently solvent-exposed, likely in a flexible and accessible loop.

Therefore, the position of the glycosite and the amount of protein in periplasm resulted the key aspects influencing the extent of glycosylation of the final bioconjugated protein.

To investigate the correlation between the amount of protein produced and the percentage of glycosylation observed, the comparison among the single (Hla-i-CP5) and double glycosylated Hla isoforms (Hla-ds1 and Hla-ds2) containing the PTP-i glycosite in the same position (pos130) could be done (fig. 19).

The inverse relation between protein produced and glycosylation efficacy of PglB, already observed and reported in fig.12 for the single glycosylated isoforms, is maintained. The increasing amount of protein produced resulted in a decreased efficiency of PglB to fully glycosylate Hla, although the same consensus sequence (PTP-i) is located in the same position on the protein sequence (pos130).

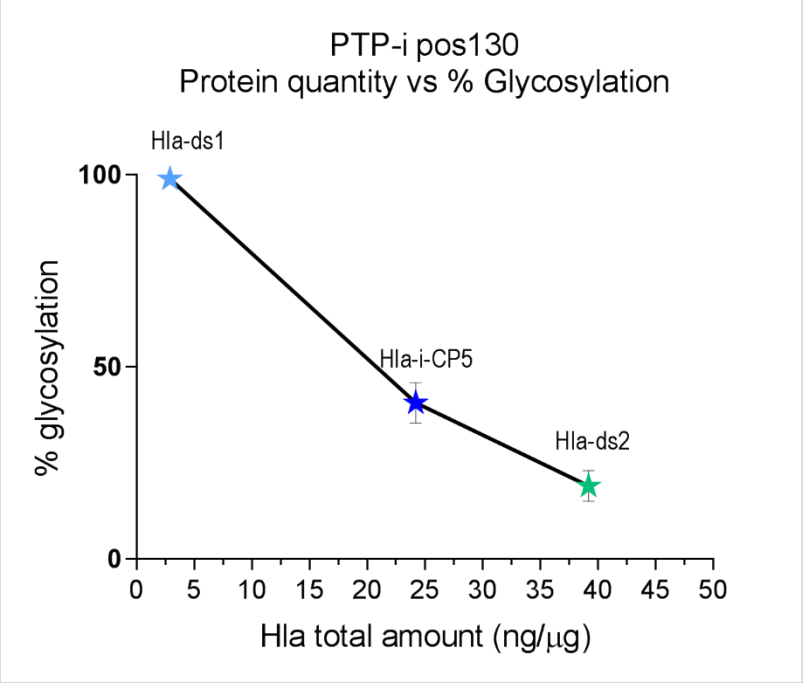


Figure 18: Correlation between the position of PTP-i consensus sequence and the amount of protein produced comparing different constructs (single site Hla-i-CP5, double sites Hla-ds1 and Hla-ds2).

From the quantification data, only the Hla-ds1 isoform resulted to be fully glycosylated in both glycosites and it is characterized by the lowest Hla yield.

All these results confirmed that the overall source of protein expressed and translocated in the cell periplasm plays a pivotal role on PglB efficiency, as much as the availability of sugar moieties and the position of the glycosite along the protein sequence.

Moreover, in our experimental conditions, the identical induction time applied for all the constructs growth conditions for the bioconjugate production (both single and double) allowed to hypothesize that the glycosylation occurring could delay the protein expression and translocation in the periplasm, although both the enzyme and the protein induction are concomitant. This is suggested by the observation that the obtained fully glycosylated constructs resulted to be the ones with glycosites located at N-terminal and with the lower Hla yield (Hla-ds1 and Hla-ds3).

Nevertheless, the yields of bioconjugate produced could be increased by optimizing the PglB enzyme to increase its efficiency and improving the cell cultures growth conditions in the fermentation process for large-scale production. However, although the processes employed in the present study show several differences compared to the ones of the large-scale fermentation, the overall protocol of cell growths and of the periplasmic fraction isolation by osmotic shock were maintained. For this reason, in our conditions, an optimization of the PglB enzyme would be suggested to increase the bioconjugate yield, especially in case of high amount of carrier protein produced (Ihssen *et al.*, 2015).

Additionally, also the saccharide production could be improved by boosting the system with additional plasmid expressing the saccharide gene cluster.

Furthermore, the conjugation of the Hla-ds isoforms was assessed by Western blot analyses. Three analyses were performed, comparing the glycosylation patterns of the isoforms among samples normalized to OD_{600} (0.3) of bacterial culture (fig. 17A) or normalized to ng of total Hla quantified by SRM (fig 17, panel B and C). The separation was performed on 14% polyacrylamide gels and revealed using a rabbit anti Hla-CP5 (fig. 17A-C) or anti Hla sera (fig. 17B).

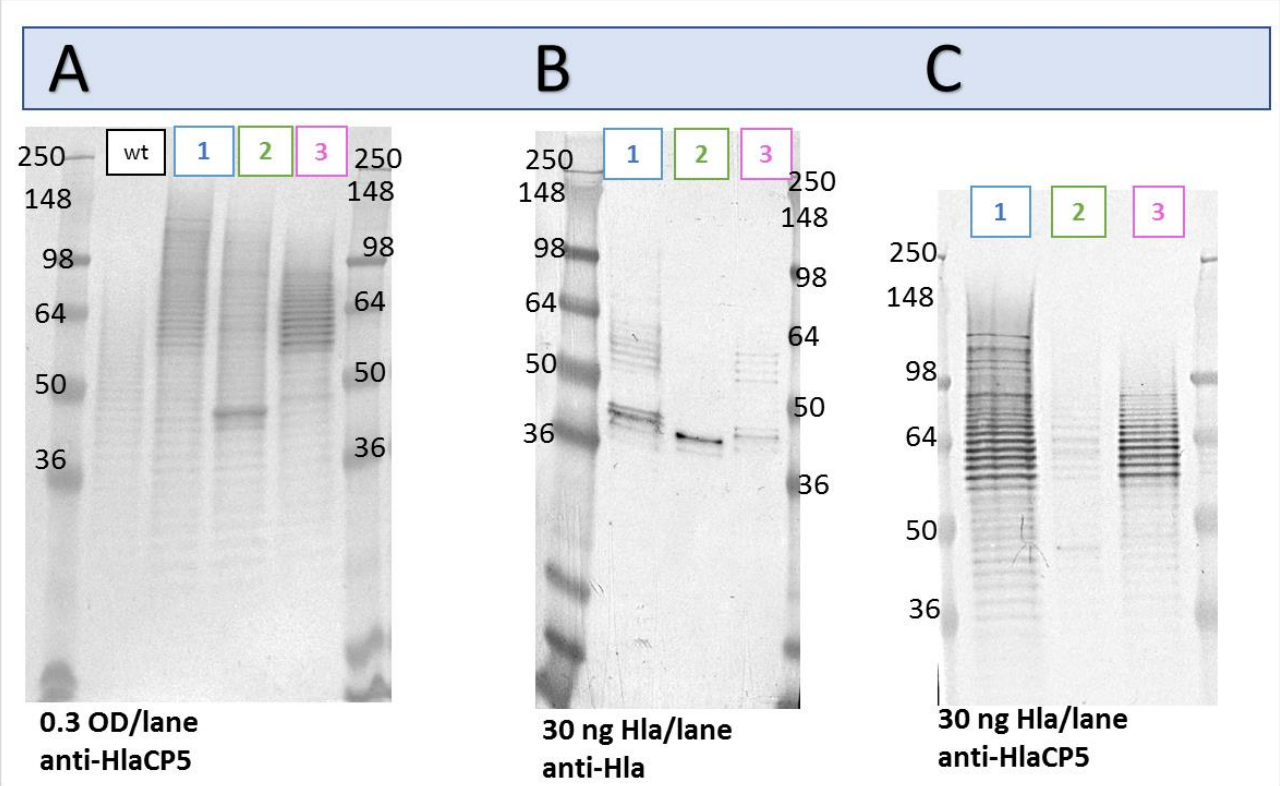


Figure 19: ***Efficacy of bioconjugation of the designed carriers carrying two glycosylation sites assessed by Western blot. A)*** *Western blot analysis using anti-HlaCP5 rabbit serum, samples normalized to OD600 (0.3 OD/lane for each construct) of bacterial culture;* ***B)*** *Western blot analysis using anti-Hla rabbit serum on samples normalized to ng of total Hla protein quantified by SRM (30 ng of Hla/lane);* ***C)*** *anti-HlaCP5 serum on samples normalized to ng of total Hla protein quantified by SRM (30 ng of Hla/lane). The periplasmic fractions were prepared from E. coli engineered for the expression of Hla_ds1 (1, in blue), Hla_ds2 (2, in green) and Hla_ds3 (3, in pink). As negative control, the analysis of the periplasmic fraction prepared from the strain that does not carry Hla-ds plasmid is reported (lanes 1, reported as WT, on the panel A).*

In panel A of fig. 17 is reported the Western blot analysis using anti Hla-CP5 serum, performed on samples normalized to OD_{600} (0.3 OD/lane for each construct) of bacterial culture. In lane 1, is reported the periplasmic fraction of an *E. coli* (st9268) strain as negative control.
Different glycosylation patterns could be appreciated for the strains expressing Hla-ds constructs:

i) Hla-ds1 showed fully glycosylated pattern, reaching high apparent molecular weight (MW), ranging from 55-60 to 148 kDa without any detectable unglycosylated form of Hla protein;

ii) Hla-ds2 showed a less intense glycosylation pattern and a strong signal agreed with the apparent molecular weight related to the unglycosylated form of Hla-ds2 carrier protein;

iii) Hla-ds3 presented a glycosylation pattern with high intensity ranging from 55-60 to 80 kDa of apparent MW, although with minor ladder pattern when compared to Hla-ds1, revealing that the carrier protein might be less glycosylated. Moreover, in this case, a signal at apparent MW of 40 kDa could be appreciated with a slightly higher electrophoretic mobility compared to the most intense signal observed for Hla-ds2.

In panel B and C of fig. 17 are reported the Western blot analyses performed on samples normalized to ng of total Hla protein quantified by SRM approach (30 ng of Hla/lane). The signals were revealed using a rabbit anti Hla and an anti Hla-CP5 sera, reported in panel B and in panel C respectively.

In the case of anti Hla Western blot (17B), the antibody does not efficiently reveal the bioconjugation patterns. It could be appreciated that the most intense signal is revealed at apparent molecular weight corresponding to Hla-ds2 (36,606 Da), that resulted the less glycosylated constructs and therefore the sample with a high amount of unglycosylated carrier.

By contrast, considering that Hla-ds1 and Hla-ds3 are fully glycosylated, the ladder signals observed in panel B at the apparent molecular weight corresponding to their molecular masses (37,049 and 37,478 Da, respectively) reveal the protein forms bioconjugated with short-length of saccharide repeating units. Indeed, the peculiar pattern of conjugated proteins in Western blot analyses appears as a ladder pattern with distinct bands, probably reflecting the repeating units incorporated onto the carrier protein (as showed in panel C).

In panel C, the anti Hla-CP5 serum was used and the pattern observed suggests that the antibody clearly reveal with different intensities the more glycosylated conjugate (Hla-ds1 and Hla-ds3) than the lesser one (Hla-ds2), suggesting different affinities of the antibody for the carrier protein and CP5 recognition. In lane 1 and 3, the glycosylation patterns observed for Hla-ds1 and Hla-ds3 in panel A were confirmed. Moreover, in lane 2 the signal corresponding to the apparent molecular weight of the Hla-ds2 isoform could be slightly appreciated (lane 2, panel B).

However, these considerations reflect that the presence of Hla-CP5 bioconjugate products could be evaluated only by a qualitative perspective by Western blot analyses. The glycosylation extent and the differential site occupancy evaluation of bioconjugated molecule would need robust and reproducible analytical tools to be addressed.

6.3.6 Insights on the evaluation on the double glycosylated samples heterogenicity

Several speculations from these data related to the diverse Hla protein forms present in the sample could be done.

In case of multiple glycosylated carrier when fully glycosylation of the sites is not achieved, multiple forms of the protein are present in the sample as reported in the schematic illustration in fig. 20 (on the left). In details, on the left, the possible forms of the protein present in the sample are shown: A is the totally unglycosylated form, B is the one where both of sites are occupied by the sugar, and C and D are the forms where only one of the site is occupied. In such configurations, the yellow box represents the total amount of the protein (obtained by quantification of PTP-2 and PTP-3) while the blue and the green boxes represent the glycosite (PTP-i and PTP-v).

In presence of construct Hla-ds1, only B form of the protein is present, that is the fully glycosylated one with both sites occupied by the sugar (fig. 20, Hla-ds1).

In sample of construct Hla-ds3, one or both of the sites are occupied, thus the amount of the unglycosylated form (C) is directly obtained by the difference between the total amount of Hla and the unglycosylated site value of the C form.

On the other hand, in samples of construct Hla-ds2 where the fully unglycosylated form is present, the quantification of each form could not be achieved, since the calculation of the C and D forms would need to consider the presence of the A form. However, after the purification of the bioconjugate sample to exclude the totally unglycosylated form, from our quantification data could be possible to evaluate the amount of the different forms of the glycosylated protein in the sample.

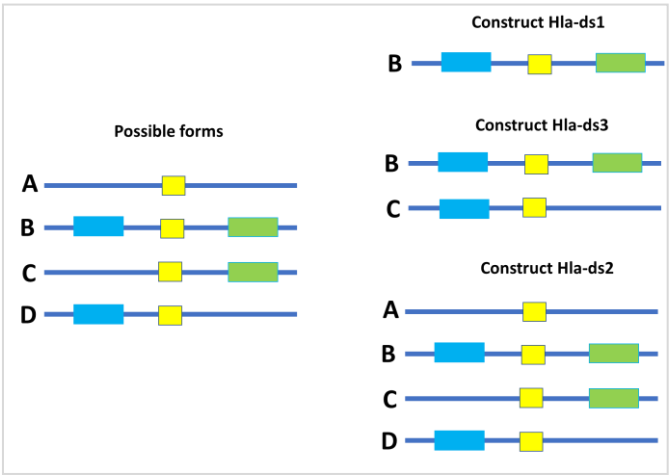


Figure 20: On the left the possible forms of Hla present in sample of multiple glycosylated constructs. On the right, the possible forms present in sample of the different construct Hla-ds1, Hla-ds3 and Hla-ds2.

Conclusion

Glycans coating the bacterial surface are key antigens for the development of vaccines against bacterial diseases. Over the last decades, glycoconjugate vaccines licensed to combat *N. meningitidis*, *H. influenzae* and *S. pneumoniae* have proven to be safe, efficacious and cost effective (Delea *et al.*, 2017; Delgleize *et al.*, 2016; Sharma *et al.*, 2012; Zarei *et al.*, 2016).

Nevertheless, glycoconjugates produced by chemical or synthetic conjugation approaches are considered time demanding and expensive for vaccine manufacturing due to the complex multistep processes and quality control flows (Frasch, 2009). Nowadays different tools are available for the glycoconjugates manufacturing, that could accelerate the development of novel carbohydrate-based vaccines (Harding & Feldman, 2019; Hu *et al.*, 2016; Kay *et al.*, 2019; Langdon *et al.*, 2009).

In particular, bioconjugation technology allows to produce N-glycosylated proteins directly in *E. coli* cells, exploiting an enzymatic pathway firstly discovered in *C. jejuni* (Szymanski *et al.*, 1999), and employed to produce several glycoconjugate vaccine candidates (Langdon *et al.*, 2009; van den Dobbelsteen *et al.*, 2016; Wacker *et al.*, 2002; Wacker *et al.*, 2014) (Cuccui *et al.*, 2013).

The use of site-selective approach can aid to achieve the optimal carbohydrate density and to select the attachment site on the carrier protein, supporting the rational design of vaccines with improved immunogenicity and preserved key protective epitopes (Broker *et al.*, 2017; Hu *et al.*, 2016).

Moreover, it would confer to conjugated vaccines higher batch-to-batch consistency and a more robust structure-biological activity correlation in comparison to the ones produced by the classic methods (Hu *et al.*, 2016). The better defined physico-chemical characteristics of the vaccines would result in improved quality controls during the process development and reduced number of routine controls for product release, giving indisputable advantages in terms of quality standards and manufacturing costs (Hu *et al.*, 2016).

All these requirements might be fulfilled to respond to regulatory agencies for licensing new vaccines (Jenkins *et al.*, 2015; Lepenies, 2015b). The characterization of the vaccine antigen is required to be achieved by validated analytical methods, although at the same time the regulatory agencies always look for new analytical methods to support the old-fashioned validated ones (Sharma *et al.*, 2018).

In the last years several MS methods have been applied to evaluate the protein-to-polysaccharide ratio and the extent of protein glycosylation, some of the key attributes required for glycoconjugated vaccine release (Sharma *et al.*, 2018). However, these techniques have shown to be applied only after previous purification or elaborated preparation of the samples, either during vaccine development or antigen discovery in preclinical phases and for the final product characterization (Sharma *et al.*, 2018).

To date, no analytical methods are available to determine the degree of glycosylation. A consistent method is still missing to monitor the polysaccharide-protein ratio and the extent of glycosylation in bioconjugate vaccines either during fermentation process or at drug substance (DS) and drug product (DP) levels.

The aim of the present study was to develop an analytical method for bioconjugate vaccine characterization, by the setup of a MS-based approach for the quantification of the extent of glycosylation. The assay applied was the Selected Reaction Monitoring (SRM) technique, that results the most suitable to perform a targeted quantification of the glycosylation extent since ensures to achieve the required specificity, sensitivity, accuracy, and precision in the complex matrices (Vidova & Spacil, 2017). In details, the approach here applied was firstly developed by Zhu *et al*. for absolute quantification of the glycosylation site occupancy of the commercially available and purified eukaryotic bovine fetuin protein (Zhu *et al.*, 2014). As proof of concept, we focused our attention on Hla-CP5 bioconjugate molecule, demonstrated to elicit in mice antibodies protective against *S. aureus* infections (Wacker *et al.*, 2014).

The mandatory requirements needed for the application of SRM-MS (Lange *et al.*, 2008) were not satisfied by the bioconjugate Hla-CP5 molecule produced as reported by Wacker *et al*. In details, the peptide carrying the consensus sequence for N-glycosylation did not comply with the uniqueness and MS detectability properties needed for the SRM-MS applicability. Consequently, we decided to apply an analytic driven design approach on the biconjugate antigen to obtain MS compatible and quantifiable peptides carrying the glycosite.

The design of MS compatible Proteotypic Peptides (PTPs, PTP-i-v-s) carrying N-glycosylation consensus sequence was assessed by the detection of the isotopically labelled PTPs forms in *E. coli* periplasmic extract matrix. This result confirmed the suitability of SRM-MS for the selective characterization of bioconjugated molecule in complex biological matrices. Additionally, it provided the basis to employ and optimize a universal analytical method for the purpose, since the designed PTPs are suitable to be inserted and to engineer any carrier protein, with the aim to act as universal tag for N-glycosylation by PglB.

The Hla carrier protein engineering was performed to insert single or double glycosites and allowed to obtain an optimized and well-defined antigen, suitable for an improved characterization. Moreover, the bioconjugated antigen would gain an increased valency of glycosylation, in the case of carrier protein doubly glycosylated.

Considering the results obtained, several considerations concerning the efficiency and efficacy of PglB enzyme could be done.

In terms of efficacy, the designed PTPs fulfilled the crucial requirements, defined in literature, for a favorable N-glycosylation in bacteria (Kowarik *et al.*, 2006b; Silverman & Imperiali, 2016; Wacker *et al.*, 2006). The Western blot analyses performed on both, single and double glycosylated forms, confirmed that the bioconjugation of the protein occurred. This suggested that probably the introduction of the variable amino acid residue (Ile, Val or Ser, *i.e.* PTP-i-v-s) in the consensus sequence did not significantly affect the efficacy of PglB enzyme.

By contrast, the efficiency of bioconjugation resulted to be strictly related both to the overall source of protein produced and to the position of the glycosites on the protein sequence. Such considerations raised from the comparison of diverse constructs carrying the same PTPs. The correlation between the amount of carrier protein and the percentage of glycosylation resulted significant for the constructs Hla-i-CP5, Hla-ds1 and Hla-ds2, which are related since all three constructs carry PTP-i in position 130. For these constructs, the extent of glycosylation is inversely proportional to the protein amount produced, suggesting that an optimization of the PglB enzyme would benefit to boost the bioconjugate yields as much as the antigen design of the vaccine candidate itself (Ihssen *et al.*, 2015).

The comparison of the constructs carrying PTP-v in N-terminal or C-terminal position allowed to assess that also the position of the glycosite along the protein sequence influences the extent of glycosylation. In details, constructs with PTP-v located at N-terminal position (Hla-ds1 and Hla-ds3) resulted constantly fully occupied by the glycans, independently of the protein amount produced (2.9 ng/µg and 10.5 ng/µg, respectively). By contrast, in the construct Hl-ds2 PTP-v is located to the C-terminal position and it resulted poorly glycosylated with an extent of 16%.

Moreover, from HDX-MS experiment on wild type Hla protein was determined that N-terminal region of the protein is highly flexible and solvent-exposed, while the C-terminal region showed well-structured features.

These results are in agreement with the comprehensive studies on PglB function and mechanism of action (Silverman & Imperiali, 2016). Although PglB is able to glycosylate folded protein, as occurred in Hla-ds2 and Hla-ds3, our results strongly suggest that the co-translocational nature of PglB N-glycosylation process is the most significant in terms of efficiency.

The developed SRM-MS approach proved to be applicable at various levels of antigen characterization thanks to its remarkably quantitative performances (Hickey *et al.*, 2016).

The enhanced selectivity allows to perform an in-process monitoring of the bioconjugation process directly from the *E. coli* periplasmic extract where the bioconjugation occurs, without previous purification or enrichment step of the sample. SRM-MS could be applied to evaluate progress of

bioconjugate production and to highlight if the manufacturing process is moving outside of the conditions for which it was validated (Sharma *et al.*, 2018).

Moreover, it showed high sensitivity, quantifying down to a concentration in the range of low ng per µg of total periplasmic proteins.

The quantification resulted reproducible in our experimental conditions, with coefficients of variation mostly <30% inter independent digestions (SRM experiments) for all the protein isoforms analyzed.

The Western blot analyses allowed to evaluate that this assay could be employed exclusively for a qualitative identification of specific known proteins in complex background and not for quantitative purposes in presence of glycoprotein peculiar patterns.

The application of the SRM-MS technique to quantify the glycosylation extent provides a number of information on the bioconjugated vaccine antigen, about:

i) the trend of bioconjugation process in terms of product yields and the feasibility to implement corrective actions to evaluate the glycosylation;
ii) the amount of the "free" carrier protein and therefore the conjugate integrity of the product;
iii) the consistency of the process, since it allows to assess if the applied experimental conditions are providing reproducible products, with consistent conjugated/unconjugated amount of the carrier protein;
iv) the heterogeneity of the bioconjugated sample, during production and in the purified final product. In fact, varied forms of the protein could be present for doubly bioconjugated isoforms in addition to the conjugated and unconjugated forms.

Therefore, MS results a valuable tool for the comprehensive analytical characterization of glycoconjugate vaccines. Although with rare exceptions, it is routinely used during the early research and development stages of protein-based molecules.

The implementation of MS methods in chemistry, manufacturing and quality control (QC) environments for vaccines has been of longstanding interest but has faced so far regulatory challenges that require validate methods (Jenkins *et al.*, 2015; Sharma *et al.*, 2018). Nevertheless, the validation could demonstrate that during the performance of the assay, no variability between different instruments and operators arises. The main rational for the MS-based methods implementation in the vaccine manufacturing could be to integrate and to support the existing validated traditional analytical methods, without replacing them.

The insights reported in this study raise the possibility of novel glycoconjugate vaccines with chemical and physical properties specifically designed in a view of antigen characterization. Moreover, the analytical antigen design developed considered also the preservation of the antigenicity of the key epitopes.

The antigen design might consider several aspects as the further analyte characterization needed to respond regulatory agencies and the optimization and increase of the bioconjugate yields in manufacturing processes. Moreover, it may allow to perform all the measurement aimed to test critical quality requirements of the antigen, as the stability of the final product, the safety, the batch-to-batch variation and the lot release testing.

Until the development of bioconjugate and site-selective conjugated vaccines, glyco-conjugation has been carried out as a random process to link carbohydrate and protein, without deep considerations on the optimal design driven by the accomplishment of regulatory requirements. The application of an analytical driven design approach on the antigen and the development of accurate, precise and robust analytical methods for vaccine antigen characterization would render the vaccine discovery and development more straightforward and focused processes.

Appendix

Appendix 1: Detection and characterization of Hla-CP5 glycosite

Diverse trials were performed to test the MS detection and quantification of the peptide carrying the consensus sequence for the reference glycosylation (glycosite), the key aspects of the peptide to perform an SRM-MS analysis.

Firstly, 1 µg of the digested protein was injected in a C18 UPLC column in-line with a Q-exactive MS instrument. The data generated were analyzed by PEAKSX software to identify the protein by matching the peak list against the *E. coli K-12* proteome database previously used, that include also the carrier protein sequence, in order to obtain the coverage map of the Hla protein. In fig. 22 is reported the coverage map achieved.

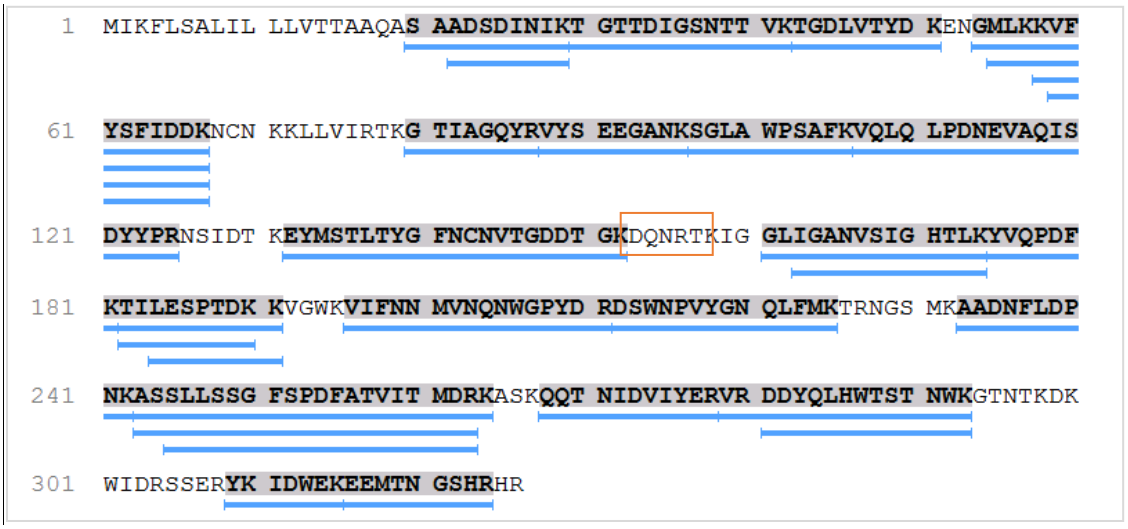


Figure 21: ***Coverage map of Hla carrier protein digested by lysC/trypsin.*** *In the red box is reported the peptide carrying the glyco-sequence.*

As marked in fig. 22 (red box), no coverage for the peptide carrying the glycosite was obtained.

Moreover, to improve the MS sensitivity, a targeted MS/MS analysis was performed using 1 µg of digested Hla in a C18 nanoUPLC column in-line with a triple quadrupole instrument (TSQ Vantage). The SRM MS acquisition mode was set-up for an unscheduled analysis in order to detect during the entire chromatographic elution method the transitions selected to target the peptides DQNR and DQNRTK, obtained by trypsin and lysC digestion, respectively.

In fig. 23 are reported the selected transitions. In fig. 24A are reported the different acquisition channels of the SRM analysis (each corresponding to a specific transition monitored), while in fig. 24B the fragment ions selected to be acquired are shown. In details, only the b4 and y3 ion type of the DQNRTK peptide were detected, although with and not reproducible low intensity signals (5.36E2).

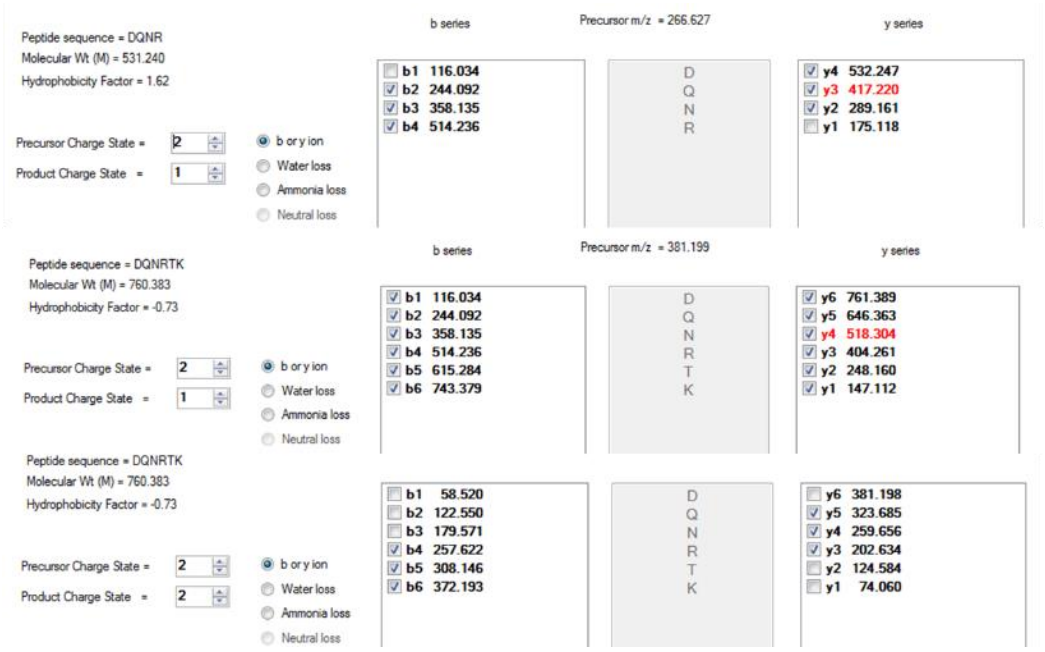


Figure 22: Selected transition for detection of DQNR and DQNRTK.

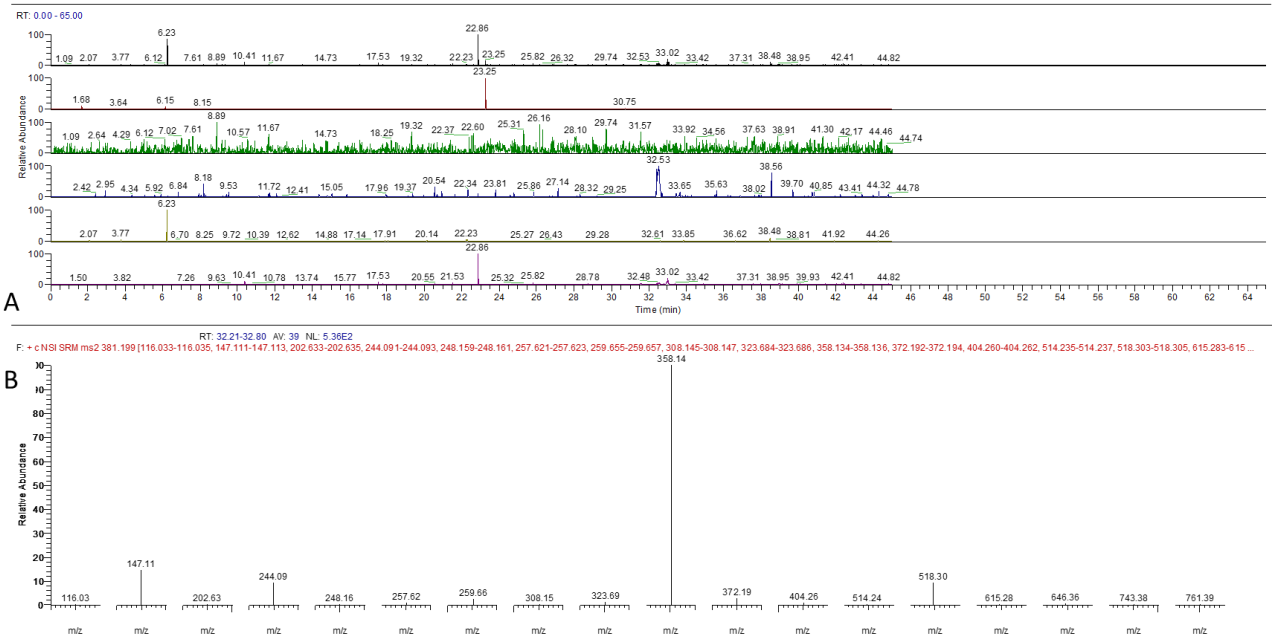


*Figure 23: **(A)** acquisition channels of the selected precursor/fragment pairs selected for the detection of DQNRTK; **(B)** fragment ions selected to be monitored for the detection of DQNRTK. As could be noticed by the fig. 2, the detection of the peptide DQNR could be monitored by the detection of fragment ions deriving from DQNRTK.*

Under these experimental conditions, the generated data on the peptide carrying the glycosequence DQNR (Hla-glycosite) indicated that such peptide was not MS detectable. Since to apply LC-SRM MS approach the selection of peptides that can be uniquely mapped to a protein and that are detectable in a mass spectrometer is crucial, we decided to focus on an analytic driven design approach to obtain peptides MS compatible and quantifiable peptides carrying glycosite.

Appendix 2: LLO intermediates isolation and detection

LLO intermediates are involved in the O antigen biosynthesis pathway, which is the conserved process of bacterial cells to produce saccharide transferred by PglB on the carrier protein exploited by the bioconjugation technology. LLO intermediates formation consists in individual building blocks of the sugar that are transferred sequentially from the sugar-nucleotide precursors onto the Und-PP to form the repeating units of the O-antigen (RU), on the cytoplasmic side of the plasma membrane (Wright *et al.*, 1967). Several efforts were employed in order to isolate, detect and quantify such LLO, in order to investigate the efficiency of *E. coli* K-12, to produce the saccharide chain and the influence of such process on the final bioconjugate production yields.

Western blot analysis was performed using a rabbit anti-CP5 polyclonal serum on total lysate of bacterial cell cultures obtained for bioconjugate expression of single Hla-CP5 isoforms (paragraph 5.4). In fig. 25 is possible to appreciate the different lengths of the sugar chains in the LLO intermediates.

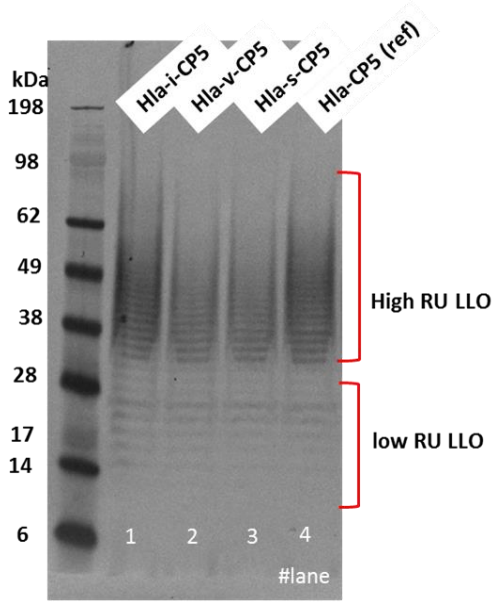


Figure 24: Western blot analysis using anti-CP5 rabbit serum of total lysate of E. coli strains producers (0.3 OD) of the bioconjugates Hla-i-CP5, Hla-v-CP5, Hla-s-CP5 and Hla-CP5 reference strain (as positive control, lane 4).

Several attempts were performed by applying different extraction protocols as the Bligh & Dyer protocol (chloroform/methanol and chloroform/methanol/water) and the hot phenol extraction, widely used in literature to isolate the LLO with low RU from bacterial cells, although no acceptable results were achieved (Arsang *et al.*, 2014; Bligh & Dyer, 1959; Davis & Goldberg, 2012). In details, such protocols were optimized and adapted to obtain LLO isolation in polar phase to be suitable to chromatographic separation for LC-MS analysis. Particularly, by applying the hot phenol extraction on lyophilized bacterial pellets, no LLO were detected by Western blot analysis. On the other hand,

by applying the Bligh & Dyer protocol after the substitution of the water phase with a solution of RapiGestSF™ 1% (a detergent/surfactant suitable for MS analysis) to try to recruit molecules with amphipathic features, LLO intermediates were detected by WB analysis with very low signals (data not shown), indicating a low level of LLO intermediates due to the poor availability of the Und-PP carrier in the bacterial cells which has been estimated to be the 0.5% of the total phospholipids in the bacterial cells (Kramer *et al.*, 2004). Indeed, it has been well studied the recycling of such lipid carrier which is present in a fixed pool in the bacterial cells, common to several pathways which compete for the same substrate (Jorgenson & Young, 2016). However, it is difficult to determine its amount, and when it has been achieved, the measurements were performed by employing radioactive labelling (Barreteau *et al.*, 2009).

In parallel, the chromatographic separation and the MS detection of an undecaprenol standard lipid (Und-MPDA, Undecaprenol monophosphoryl di-ammonium salt) were developed and optimized. The chromatographic separation was carried out on an HILIC column by testing methanol/chloroform elution method with 50 mM ammonium acetate, with the aim to optimize the chromatographic conditions in view of a final separation of the LLO intermediates of bacterial cells, where the conserved lipid (non-polar) is linked to oligosaccharide chains of different lengths in repeating units (polar intermediates). Consequently, the chromatographic method was set up with an increasing methanol gradient, to allow the separation of the different intermediates with increasing saccharide units repetition and, thus, increasing polar behaviors. For this reason, the separation was performed on a HILIC column, which is a normal phase UPLC column whose matrix type has been already employed in literature to enrich glycan-linked compounds (i.e. glycopeptides) (Hagglund *et al.*, 2004) and polar lipids (Rampler *et al.*, 2018).

In details, 100 pmol of the Und-MPDA standard were injected in a Shimadzu HPLC connected in-line to a Synapt G2Si Q-TOF analyzer, and it was detected in negative ion mode with a RT of 14.5 min, eluting at 11% of chloroform (Fig.26A).

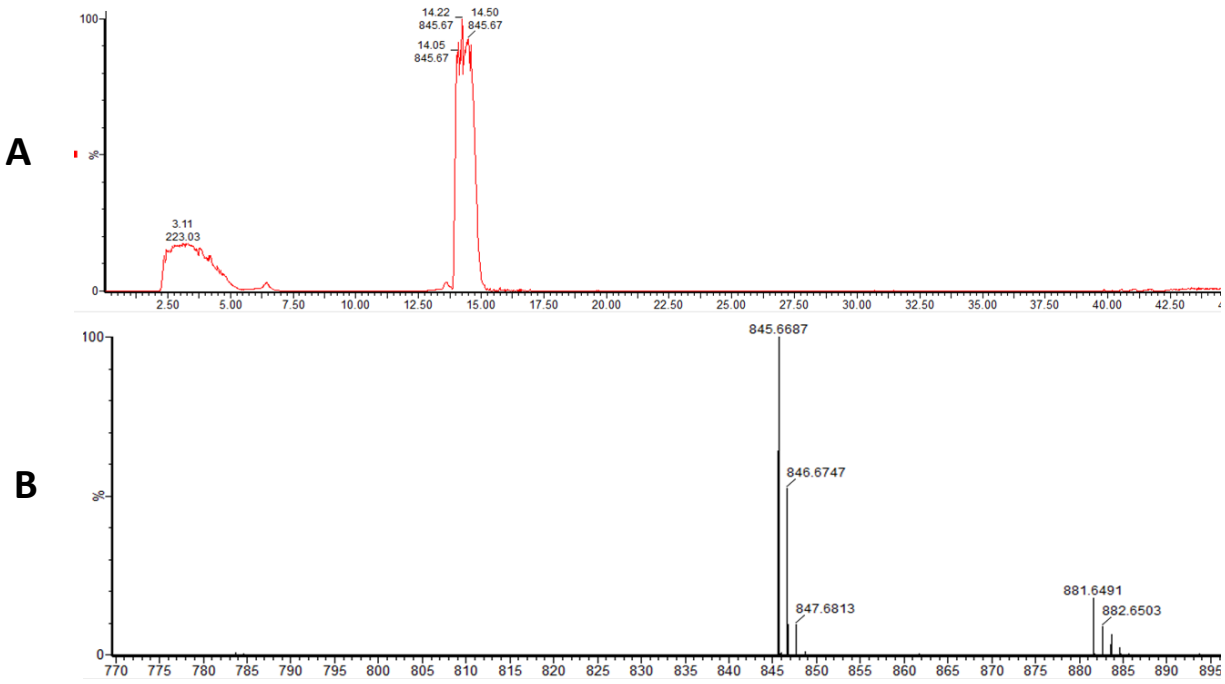


*Figure 25: **(A)** Chromatographic elution using HILIC column and **(B)** MS detection on Synapt G2 Q-TOF of Und-MPDA in negative ion mode.*

Once the detection was achieved (fig.20B), the MS/MS spectra were acquired applying increasing CEs to characterize the Und-MPDA (ranging from 15 to 35 eV). The fragmentation pattern shows how the lipid (845,65 *m/z* for Und-P and 881,65 *m/z* when adducts with ammonium are formed) is progressively fragmentated in the isoprene units (68 *m/z*), and the loss of the phosphate ion (78,9 *m/z*) can be detected. The main adduct form in negative mode was the diammonium $2(NH_4)^-$ adduct ions, and the loss of phosphate was observed at 783.7 *m/z* from the adducted with $(NH_4)^-$.

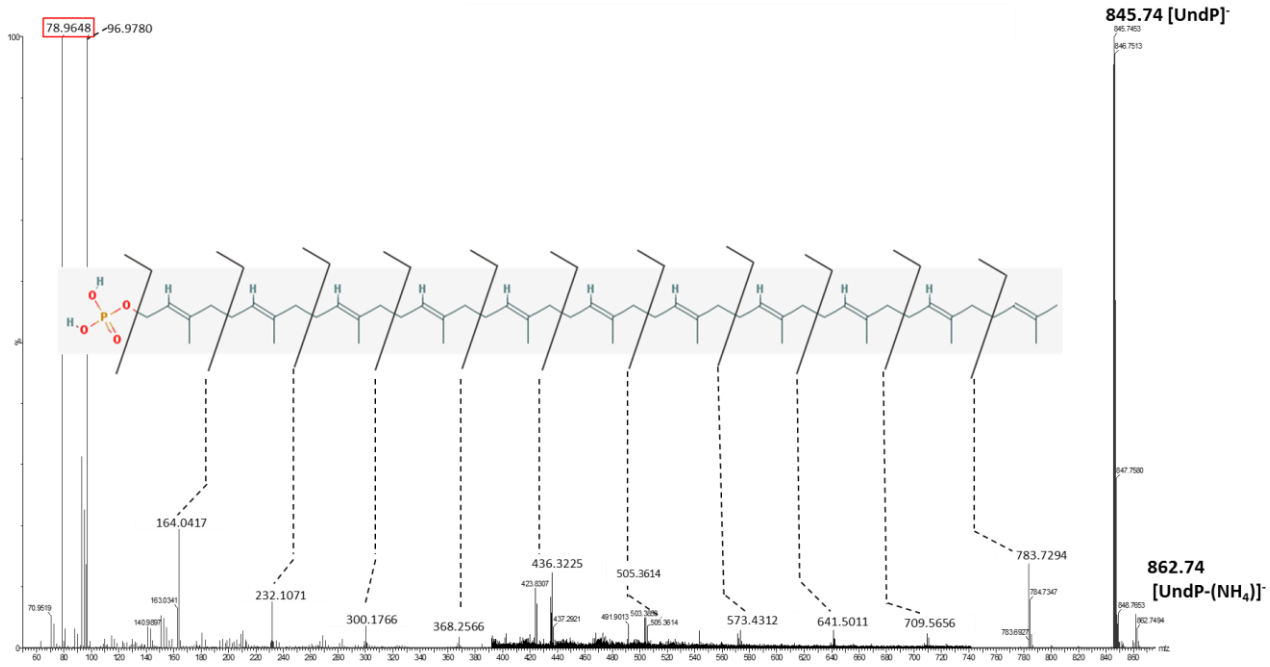


Figure 26: ***Und-MPDA MS/MS fragmentation pattern****. In the box the chemical structure of C55 isoprenoid Undecaprenyl-phosphate. In red the phosphate ion.*

Materials and Methods

7.1 Bioconjugate molecules engineering, production and extraction

7.1.1 Bacterial strains, cloning and purification of plasmids.

Bacterial strains and plasmids employed in this study are summarized in table 10.

Strain	Name	Description	Reference or source
topTEN	topTEN	**Cloning: F**- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG	Thermo
W3110	st9268	**Glycocompetent *E. coli*** (StGVXN9268): waaL::pglB4xmut; ΔrmlB -wecG; O16(rfb)::O11wbjB-wbpM-J23100-cap5HIJK	GVXN9268
W3110	st9268 (p1221, ΔHla)	**Glycocompetent *E. coli*** (st9268): bearing vector p1221	This study
		Plasmids	
pEC415: p2557-FlgI: p2533-Hla	p2872	**Expression vector** of *S. aureus* Hla H35L-G122C-H48C, FlgI signal sequence, cross-liked with glycosite KDQNRTK 130 and HRHR tag, via NheI/ XmaI arabinose inducible promoter, Kan	GVXN2872
pEXT21: PglBcuo N311V-K482R-D483H-A669V	p1221	**Expression vector of *C. jejuni* PglB** codon usage optimized containing N311V-K482R-D483H-A669V mutations, Lac promoter, IPTG inducible, Sp	GVXN1221
pEC415:Bcp218	pHla-i	Expression vector of *S. aureus* **Hla** (p2872) with glycosite PTP-i in 130	This study
pEC415:Bcp234	pHla-v	Expression vector of *S. aureus* **Hla** (p2872) with glycosite PTP-v in 130	This study
pEC415:Bcp235	pHla-a	Expression vector of *S. aureus* **Hla** (p2872) with glycosite PTP-a in 130	This study
pEC415:Bcp233	pHla-s	Expression vector of *S. aureus* **Hla** (p2872) with glycosite PTP-s in 130	This study
pEC415:Hla-ds1	pHla-ds1	Expression vector of *S. aureus* **Hla** (p2872) with double glycosite N-term (PTP-v) and pos130 (PTP-i)	This study, GeneArt supplied
pEC415:Hla-ds2	pHla-ds2	Expression vector of *S. aureus* **Hla** (p2872) with double glycosite pos130 (PTP-i) and C-term (PTP-v)	This study, GeneArt supplied
pEC415:Hla-ds3	pHla-ds3	Expression vector of *S. aureus* **Hla** (p2872) with double glycosite N-term (PTP-v) and C-term (PTP-i)	This study, GeneArt supplied

Table 10: Bacterial strains and plasmids description.

The PIPE method (Klock & Lesley, 2009)was applied for mutagenesis and cloning experiments to obtain all the plasmids carrying mutated Hla protein sequence (single and double glycosites).

Bcp218, Bcp233, Bcp234 and Bcp235 (table 11) carry the Hla-i Hla-s, Hla-v and Hla-a carriers, respectively, where the bioconjugation consensus sequence in position 130 of the reference form of Hla (plasmid p2872) (Wacker *et al.*, 2014) was mutated by site-directed mutagenesis (M-PIPE) employing ad-hoc designed primers, as described in the results and discussion chapter. The sequences of primers employed are reported in table 11.

Plasmids pHlads1, pHla-ds2 and pHla-ds3 were obtained by PIPE cloning of DNA strings carrying the desired sequence, supplied by GeneArt, into pEC415 destination vector (p2872).

All the obtained PCR products were subjected to DpnI (cod. R0176S, BioLabs New England) digestion at 37° for 5h to cleave methylated genomic DNA, before the mix with the V-PIPE products.

The obtained mutated I-PIPE DNA amplified insert were mixed with the V-PIPE product and transformed overnight in *E. coli* topTEN (cod. C404010 One Shot™ top10 chemically competent E. coli, ThermoScientific®) to obtain the target plasmid following the manufacturer instructions. The day after colony screening was performed on 10 colonies of each construct by PCR using primers selective for the mutated region in order to select the transformed colonies harboring the desired sequence. Colonies were picked and added to the PCR mastermix as reported in table 12.

Each PCR mix was prepared and cycled as shown in Table 12. The sequences of primers employed are reported in table 11.

Plasmid DNA were purified using the MiniPrep Kit (cod. A1330, Promega®). The concentration of nucleic acids was determined using a NanoDrop spectrophotometer (Thermo®) by measuring light absorption at 280 nm. After this purification steps a DNA sequencing was performed by the internal Sequencing facility. The mix for the sequencing was prepared adding in a total volume of 14 µl 200 ng of the plasmids and 2.17 pmol of the primers used for V-PIPE PCR (forward and reverse).

Plasmid name	Forward primer	Reverse primer
p2872	V-PIPE: GATCCGTACCAAAGGCACCA	V-PIPE: CTGCTCGCTTTGTTCGGATC
pHla-i	M-PIPE: AAAGATAGCAATATTACTAGCGCGCGCGGACTGAT TGGCGCGAACGTG	M-PIPE: GCGCGCGCTAGTAATATTGCTATCTTTGCCGGTAT CATCGCCGGTCAC
pHla-v	M-PIPE: GCGCGCGCTAGTCACATTGCTATCTTTGCCGGTATC ATCGCCGGTCAC	M-PIPE: AAAGATAGCAATGTGACTAGCGCGCGCGGACTGA TTGGCGCGAACGTG
pHla-a	M-PIPE: GCGCGCGCTAGTCGCATTGCTATCTTTGCCGGTATC ATCGCCGGTCAC	M-PIPE: AAAGATAGCAATGCGACTAGCGCGCGCGGACTGA TTGGCGCGAACGTG
pHla-s	M-PIPE: AAAGATAGCAATAGCACTAGCGCGCGCGGACTGAT TGGCGCGAACGTG	M-PIPE: GCGCGCGCTAGTGCTATTGCTATCTTTGCCGGTAT CATCGCCGGTCAC
pHla-ds1	V-PIPE: ATGTGTTATTCCTCCTTATTTAAAA	V-PIPE: GTCGACTCTAGAGGATCCCCGGGTA
pHla-ds2		
pHla-ds3	I-PIPE: TACCCGGGGATCCTCTAGAGTCGACTTAATGATGA TGATGATGATGGCTGC	I-PIPE: TTTTAAATAAGGAGGAATAACACATATGATCAAAT TCCTGTCTGCGCT

Table 11: Sequences of primer designed and employed in the present study. V-PIPE (vector); M-PIPE (mutagenesis); I-PIPE (insert).

V-PIPE PCR component reaction mix	Volume
Template DNA vector	1 µl
Pfu 2x Taq mix	25 µl
Primer a+b (stock 10um each, final 0.5um each)	2.5 µl
ddH_2O	21.5 µl
Total volume	50 µl

V-PIPE PCR cycling protocol			
Reaction step	Temperature	Time	Number of cycle
Initial activation step	98°	2min	1
Denaturation	98°	30sec	
Annealing	55°	30sec	30
Extension	72°	6.5min	
Final extension	72°	0min	
End of PCR	4°	Hold	

I-PIPE PCR component reaction mix	Volume
Template DNA	1 µl
Pfu 2x Taq mix	25 µl
Primer a+b (stock 10um each, final 0.5um each)	2.5 µl
ddH2O	21.5 µl
Total volume	50 µl

I-PIPE PCR cycling protocol			
Reaction step	Temperature	Time	Number of cycle
Initial activation step	94°	3min	1
Denaturation	94°	30sec	
Annealing	55°	30sec	30
Extension	72°	1.5min	
Final extension	72°	7min	
End of PCR	4°	Hold	

Colony screening PCR component reaction mix	Volume
Template DNA vector	0µl (colony picked)
GoTaq mix	10 µl
Primer a+b (stock 10um each, final 0.5um each)	0.5 µl
ddH_2O	9.5 µl
Total volume	20 µl

Colony PCR cycling protocol			
Reaction step	Temperature	Time	Number of cycle
Initial activation step	95°	3min	1
Denaturation	95°	30sec	
Annealing	55°	30sec	30
Extension	72°	30sec	
Final extension	72°	10min	
End of PCR	4°	Hold	

Table 12: PCR conditions. On the right the PCR reaction component mixes, on the left the cycling protocols

7.1.2. Growth condition for bioconjugate production

E. coli W3110 st9268 was co-transformed by electroporation with the plasmids encoding the *S. aureus* single and double isoforms and the *C. jejuni* OTase p1221.

Transformed bacteria were grown overnight on Lysogeny broth (LB) selective agar plates supplemented with the two antibiotics [50 µg/ml] kanamycin and [80 µg/ml] spectinomycin for the maintenance of the plasmid encoding for Hla and PglB, respectively. Bacteria were inoculated in 50 ml LB containing antibiotics and shaken in Erlenmeyer flask overnight at 37°C, 180 rpm. A culture of 50 ml HTMC medium supplemented with [50 µg/ml] kanamycin (Kan) and [80 µg/ml] spectinomycin (Sp) was inoculated to a dilution of 0.1 optical density at 600nm (OD_{600nm}), incubated at 37°C in a shaker at 180 rpm, until an average OD_{600nm} of 0.8 – 1.0 and then induced overnight, by using 0.2% arabinose for the induction of Hla expression and 1 mM IPTG for the induction of PglB

expression. The cultures were shaken overnight at 37°C, 180rpm, and 20 OD_{600nm} of bacteria were harvested after 20 hours of induction. The supernatants were discarded and the pellets were immediately used for periplasmic extraction.

7.2 Periplasmic extraction

Periplasmic extraction (PPE) was performed on 20 OD_{600nm} of bacterial pellets recovered after 20 min centrifugation (4000 rcf at 4°C), resuspended in 600 µl of Lysis buffer (30 mM Tris HCl pH 8.5, 1mM EDTA, 20% (w/v) Sucrose) and then treated for 20 min in a rotating shaker at 4°C with 1 mg/mL final Lysozyme. After 20 min of centrifugation at 16000 rcf 4°C, supernatants were immediately collected and stored at -20°C until their use. Total protein content was assessed by Bicinchoninic acid assay (BCA Kit Reducing Agent Compatible, Thermo Fischer Scientific).

7.3 Characterization of bioconjugate by Western Blot analysis

Glycosylation status of the periplasmic Hla was analyzed by Western Blot analysis. The separation was performed on 4-12% SDS–PAGE Bis Tris gel in 1X 2-(N-morpholino)-ethanesulfonic acid (MES) at 150V for 1h, while the immunoblotting with Hla-CP5 polyclonal antiserum (1:1000) was performed. Secondary antibody HRP-conjugated was a commercial Goat Anti-Rabbit Ab and was used in a 1:5000 dilution (DAKO® Ab). WB analysis in fig. 20B was performed separating the sample using 14% Tris-Glycine SDS-page polyacrylamide gel in 1X Tris-Glycine at 150V for 1h. The molecular mass marker used was the SeeBlue™ Plus2 Pre-stained Protein Standard (cod. LC5925, Thermo®).

7.4 HDX MS experiment

7.4.1. Sample preparation for HDX-MS analyses

Hla protein (15 pmol) was incubated in deuterated PBS buffer (89% deuterium content in the reaction mixture) for 15 sec and 1h 40 min on ice. The deuterium incorporation was stopped by quenching the reaction through dropping of the pH to 2.5 and the temperature to 0°C, then the samples were quickly frozen. The labelled protein was on-line digested with pepsin, and the resulting peptides were analyzed by LCMS. In parallel, in order to perform the maximally labelled control, Hla was digested on-line and peptides eluting from the pepsin column collected. The aqueous solvent was evaporated and the dry peptides resuspended in deuterated PBS buffer (89% deuterium content identical to the intact protein) for 1h on ice, before quenching the exchange reaction.

7.4.2. HDX-MS analyses

Labeled samples were thawed rapidly to 0 °C and injected into a Waters nanoACQUITY UPLC with HDX Technology. The injector, switching valve, columns, solvents and all associated tubings were at 0 °C to limit back-exchange. For local HDX-MS, protein samples were on-line digested for 2.5 min at 20 °C with a flow rate of 200 µL/min using a Poroszyme Immobilized Pepsin Cartridge (2.1 mm x 20 mm, Applied Biosystems) equilibrated with 100 % buffer A (2% ACN, 0.1% FA in water). The generated peptides were immediately trapped, concentrated and desalted using a VanGuard BEH Pre-column (1.7 µm, 2.1x5 mm, Waters). The 2.5 min digestion and desalting step allows deuterons located at fast exchanging sites (i.e. side chains and amino/carboxy terminus) to be replaced with hydrogens. Peptides were then separated on an ACQUITY UPLC BEH C18 reverse phase column (1.7 µm, 1.0x100mm, Waters) with a linear gradient from 10 to 40% solvent B (acetonitrile/water (9/1), 0.1 % formic acid) over 6.8 min at 40 µL/min.

The maximally labelled samples were acquired without Pepsin Cartridge.

7.4.3. Mass spectra acquisition and interpretation

Mass spectra were acquired in resolution mode (*m/z* 100-2000) on a Waters SynaptG2Si Q-TOF mass spectrometer equipped with a standard micro ESI source. Mass accuracy was ensured by continuously infusing a [Glu1]-Fibrinopeptide B solution (100 fmol/µL in 50% acetonitrile, 0.1% formic acid) through the reference probe of the ESI source. The identity of each peptide was confirmed by MS^E analyses. MS^E was directly performed by a succession of low (6V) and high collision (25V) energies in the transfer region of the mass spectrometer. All fragmentations were performed using argon as collision gas. Data were processed using ProteinLynx™ Global Server 2.5 (Waters) and each fragmentation spectrum was manually inspected to confirm the assignment. The DynamX™ software (Waters) was used to select the peptides considered for the analysis and to extract the centroid mass of each of them, and for each charge state, as a function of the labeling time. Only the peptic peptides present in repeated digestions of the unlabeled proteins were considered for the analysis.

7.5 LC-MS/MS sample preparation

7.5.1 In solution digestion

Purified unglycosylated Hla variant (20 µg) used for the reference bioconjugate Hla-CP5 (Wacker *et al.*, 2014) were denatured and reduced with 0.1% (w/v) RapiGestSF™ (Waters, USA) and 5 mm DTT at 100 °C, respectively, for 10 min. After a cool down at room temperature, 50 mM ammonium bicarbonate and lysC/trypsin protease mix [1/25 (w/w), enzyme/substrate ratio] (Promega®) were added. Digestion was carried out at 37 °C overnight. The digestion was stopped adding 0.1% (v/v)

formic acid (FA) until reaching acid pH (pH=2). The peptide mixtures were then desalted using OASIS™ cartridges (Waters®) following the manufacturer protocol. Desalted peptides were concentrated with a Centrivap™ Concentrator (Labconco, Kansas City, KS) and suspended in 50 µl of 0.1% (v/v) FA.

The generated peptides were analyzed by LC-MS/MS in data dependent acquisition mode, performed on a Acquity UPLC system (Waters®) coupled with a ThermoScientific Q-Exactive™ plus mass spectrometer equipped with a micro electron spray ESI source (Thermo®). Samples were loaded using a full loop injection at a flow rate of 40 µl/min in a mobile phase A (0.1% Formic Acid, FA). Peptide were then separated on a Acquity UPLC Peptide BEH™ C18 Column (75 µm x 100 mm) (Waters®) using a 60min gradient 3-98% mobile phase B containing 0.1% (v/v) FA in 98% (v/v) Acetonitrile (ACN) with at a flow rate of 40 µl/min.

7.5.2 Sample preparation for LC-SRM analysis of bacterial periplasmic extract.

50 µg of PPE fractions were digested by using an in-stage-tips (iST) sample preparation method supplied by PreOmics® (Martinsried, Germany). It is a three-step protocol performed on a cartridge: 1) lysis denaturation reduction and Alkylation; 2) proteolytic digestion by lysC and trypsin; 3) peptide elution operated as recommended by the provider. Recovered peptides were dried under vacuum at 45°C and resuspended in 0.1% FA to a final concentration of 1µg/µl and stored at – 20°C until the MS analysis.

7.6 SRM-MS method setup and quantitative analyses

7.6.1 Selection of PTPs for Hla total protein amount quantification.

The PTPs for the quantification of the protein total amount were experimentally selected from lysC/trypsin digestion of the purified unglycosylated Hla variant used for the reference bioconjugate Hla-CP5 (Wacker *et al.*, 2014). The spectra obtained after the analysis of in solution digested protein were processed using PEAKSX software against an *E. coli* K-12 database (https://www.uniprot.org/proteomes/UP000182246) integrated with the target Hla variant protein sequence. Search parameters as variable modifications were: methionine oxidation, glutamine and asparagine deamidation, trypsin cleavage (cleaves the C-term side of KR unless next residue is P), peptide mass tolerance as 0.15 Da, peptide MS/MS tolerance as 0.15 Da, missed cleavage= 2, ion charge states: +2, +3, +4).

Suitable PTPs were selected based on the following criteria: *i)* peptides specific for Hla, *ii)* peptides showing strong MS signal intensities either for the parental or fragment ions, *iii)* peptides that do not

contain methionine and tryptophan residues, which are susceptible to oxidation, and N-terminal glutamine, to avoid cyclization.

7.6.2 PTP Dose-Range Linearity Responses and Hla Quantification.

The dose-range linearity response of the selected PTPs was assessed in a lysed periplasmic bacterial sample prepared from *E. coli* glycocompetent cells (st9268) used as reference matrix.

For Hla quantification, labeled PTPs (final concentration 0.1 pmol/µL) and non-labeled PTPs (final concentration from 1.6 pmol/ µl to 0,0125 pmol/µl) were spiked in 50 µg of periplasmic fraction prior to digestion, and SRM experiments.

For each PTP, concentrations were plotted as ratio of peak area response of light (variable)/peak area heavy (constant) measured and the fitted curve was used to obtain the unknown concentration of selected PTPs in the analyses. The LLOQ for each PTP was set as the lowest concentration point on the fitted curve and defined as 10 σ/S, where σ = the standard deviation of the response and S = the slope of the calibration curve by the International Conference on Harmonization (ICH) Guidelines (http://www.ich.org/products/guidelines/quality/article/quality-guidelines.html) Also the LOD or Limit of Detection (LOD = 3.3 σ/S) was defined based on ICH Guidelines.

The Hla concentrations were reported in nanograms per microgram of total protein extract considering an average MW of the single glycosylated Hla isoforms of 34,960 Da, the total amounts of protein were quantified across 3 SRM experiments. In double glycosylated Hla isoforms, the individual MW of the protein was considered (37,049 Da for Hla-ds1; 36,606 Da Hla-ds2; 37,478 Da Hla-ds3).

The quantification was obtained by the interpolation of each peptide-response value in the related dose-response linearity curve.

7.6.3 SRM-MS method set-up

Detection and chromatographic elution optimizations of the peptides were performed with 1 pmol synthetic PTPs form of the PTP-i-s-v peptides in a mix solution of light and heavy forms in 0.1% of FA using a reverse phase column (ACQUITY UPLC HSS T3™ Column, 100Å, 1.8 µm, 2.1 mm X 30 mm, Waters, USA) coupled to a Xevo TQ™ triple quadrupole mass spectrometer associated to an UPLC (Waters, USA). The elution gradient was optimized developed with mobile phase A 0.1% FA in water and B 0.1% FA in ACN. The synthetic PTPs were used to optimize collision energy (CE) values starting from the predicted ones, computed using the formula CE = 0.034 × (parent *m/z*) + 1.314 (Carr *et al.*, 2014), applying ramping panels of different CE to the precursor ions during

transitions analyses. The transitions were then validated the transitions in 50 µg of *E. coli* periplasmic extract as matrix.

The optimization of the chromatographic separation was performed in an unscheduled SRM acquisition mode by using the optimized CE and the selected transitions, both in neat and in matrix background. The retention times (RTs) for each peptide were registered. The optimized gradient setup in unscheduled SRM and applied in scheduled SRM analyses for quantification is provided in table 13.

Time	mL/min	%A	%B
0.00	0.080	97.0	3.0
1.00	0.080	97.0	3.0
3.00	0.080	95.0	5.0
10.00	0.080	65.0	35.0
13.00	0.100	10.0	90.0
15.00	0.100	10.0	90.0
15.01	0.100	10.0	90.0
17.00	0.100	93.0	7.0
17.01	0.100	93.0	7.0

Table 13: Gradient elution method for chromatographic separation of peptides in LC-SRM MS analyses.

7.6.4 SRM-MS analysis.

SRM was performed by injecting 10 ug of a periplasmic fraction in column per run, and each sample were analyzed in triplicate. The following parameters were used: Q1 isolation window 1.0 m/z, Q3 Isolation window 0.7 m/z, 0.03 s of dwell time and 0.02 m/z of scan width; collision cell exit and entrance potentials were set at 30 V. A spray voltage of 1700 V was used with a heated ion transfer setting of 270 °C for desolvation temperature. Data were acquired applying a scheduled SRM mode accordingly to the RTs optimized along the chromatographic elution, and the data were acquired using MassLynx™ software (version 2.1.0; Waters®).

In scheduled SRM experiments, the transitions of a specific peptide were only acquired by the MS during a time window around its elution time; briefly, the instrument repetitively cycles through the list of transitions spending on each a defined time (dwell time), resulting in a quantification with higher sensitivity and reproducibility compared to unscheduled SRM experiments. In our case, by targeting 6 peptides with 34 transitions in total and 0.03 s of dwell time, an intensity value was recorded for each transitions every 1.02 seconds. By applying these parameters, it resulted an acquisition of 7 data points per peak across the elution profile. To ensure precise LC-MS/MS

quantification, enough data points should be acquired in order to enhance the accuracy of the measurements since the peaks cannot be correctly reconstructed with too few data points (Picotti & Aebersold, 2012). The settings of dwell time and cycle time are optimized on the basis of the peak width given by from chromatographic elution profile.

The peak area quantification was determined with TargetLynx™ software (version 1.0.0.1; Waters®) after confirming the coelution of all transitions for each peptide and following the best practices reported in Carr *et al.* (Carr *et al.*, 2014).

Bibliography

Addona, T. A., Abbatiello, S. E., Schilling, B. & other authors (2009). Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* **27**, 633-641.

Aebersold, R., Burlingame, A. L. & Bradshaw, R. A. (2013). Western blots versus selected reaction monitoring assays: time to turn the tables? *Mol Cell Proteomics* **12**, 2381-2382.

Anderson, L. & Hunter, C. L. (2006). Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Molecular & cellular proteomics : MCP* **5**, 573-588.

Ansari, S., Jha, R. K., Mishra, S. K., Tiwari, B. R. & Asaad, A. M. (2019). Recent advances in Staphylococcus aureus infection: focus on vaccine development. *Infect Drug Resist* **12**, 1243-1255.

Arsang, A., Yari, S., Masoumi, M., Nour Neamatollahi, A., Vaziri, F., Nejati, M., Bahremand, A. R. & Siadat, S. D. (2014). Extraction and Purification of Haemophilus influenzae Type b Lipooligosaccharide by Modified Phenol Method. *Vaccine Research* **1**, 28-30.

Avci, F. Y. & Kasper, D. L. (2010). How bacterial carbohydrates influence the adaptive immune system. *Annual review of immunology* **28**, 107-130.

Avci, F. Y., Li, X., Tsuji, M. & Kasper, D. L. (2011). A mechanism for glycoconjugate vaccine activation of the adaptive immune system and its implications for vaccine design. *Nature medicine* **17**, 1602-1609.

Avery, O. T. & Goebel, W. F. (1929). CHEMO-IMMUNOLOGICAL STUDIES ON CONJUGATED CARBOHYDRATE-PROTEINS : II. IMMUNOLOGICAL SPECIFICITY OF SYNTHETIC SUGAR-PROTEIN ANTIGENS. *The Journal of experimental medicine* **50**, 533-550.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry* **389**, 1017-1031.

Barreteau, H., Magnet, S., El Ghachi, M., Touze, T., Arthur, M., Mengin-Lecreulx, D. & Blanot, D. (2009). Quantitative high-performance liquid chromatography analysis of the pool levels of undecaprenyl phosphate and its derivatives in bacterial membranes. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences* **877**, 213-220.

Bligh, E. G. & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology* **37**, 911-917.

Broker, M., Costantino, P., DeTora, L., McIntosh, E. D. & Rappuoli, R. (2011). Biochemical and biological characteristics of cross-reacting material 197 CRM197, a non-toxic mutant of diphtheria toxin: use as a conjugation protein in vaccines and other potential clinical applications. *Biologicals : journal of the International Association of Biological Standardization* **39**, 195-204.

Broker, M., Berti, F., Schneider, J. & Vojtek, I. (2017). Polysaccharide conjugate vaccine protein carriers as a "neglected valency" - Potential and limitations. *Vaccine* **35**, 3286-3294.

Brownridge, P. & Beynon, R. J. (2011). The importance of the digest: proteolysis and absolute quantification in proteomics. *Methods* **54**, 351-360.

Carr, S. A., Abbatiello, S. E., Ackermann, B. L. & other authors (2014). Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Molecular & cellular proteomics : MCP* **13**, 907-917.

Ciccimaro, E. & Blair, I. A. (2010). Stable-isotope dilution LC-MS for quantitative biomarker analysis. *Bioanalysis* **2**, 311-341.

Comstock, L. E. & Kasper, D. L. (2006). Bacterial glycans: key mediators of diverse host immune responses. *Cell* **126**, 847-850.

Cuccui, J., Thomas, R. M., Moule, M. G. & other authors (2013). Exploitation of bacterial N-linked glycosylation to develop a novel recombinant glycoconjugate vaccine against Francisella tularensis. *Open biology* **3**, 130002.

Cuccui, J. & Wren, B. (2015). Hijacking bacterial glycosylation for the production of glycoconjugates, from vaccines to humanised glycoproteins. *J Pharm Pharmacol* **67**, 338-350.

Cui, C., Carbis, R., An, S. J., Jang, H., Czerkinsky, C., Szu, S. C. & Clemens, J. D. (2010). Physical and chemical characterization and immunologic properties of Salmonella enterica serovar typhi capsular polysaccharide-diphtheria toxoid conjugates. *Clinical and vaccine immunology : CVI* **17**, 73-79.

D'Atri, V., Fekete, S., Clarke, A., Veuthey, J. L. & Guillarme, D. (2019). Recent Advances in Chromatography for Pharmaceutical Analysis. *Anal Chem* **91**, 210-239.

Davis, M. R., Jr. & Goldberg, J. B. (2012). Purification and visualization of lipopolysaccharide from Gram-negative bacteria by hot aqueous-phenol extraction. *J Vis Exp*.

De Geyter, J., Tsirigotaki, A., Orfanoudaki, G., Zorzini, V., Economou, A. & Karamanou, S. (2016). Protein folding in the cell envelope of Escherichia coli. *Nature Microbiology* **1**.

Delea, T. E., Weycker, D., Atwood, M., Neame, D., Alvarez, F. P., Forget, E., Langley, J. M. & Chit, A. (2017). Cost-effectiveness of alternate strategies for childhood immunization against meningococcal disease with monovalent and quadrivalent conjugate vaccines in Canada. *PloS one* **12**, e0175721.

Delgleize, E., Leeuwenkamp, O., Theodorou, E. & Van de Velde, N. (2016). Cost-effectiveness analysis of routine pneumococcal vaccination in the UK: a comparison of the PHiD-CV vaccine and the PCV-13 vaccine using a Markov model. *BMJ open* **6**, e010776.

Donnarumma, D., Faleri, A., Costantino, P., Rappuoli, R. & Norais, N. (2016). The role of structural proteomics in vaccine development: recent advances and future prospects. *Expert review of proteomics* **13**, 55-68.

Donnarumma, D., Maestri, C., Giammarinaro, P. I., Capriotti, L., Bartolini, E., Veggi, D., Petracca, R., Scarselli, M. & Norais, N. (2018). Native State Organization of Outer Membrane Porins Unraveled by HDx-MS. *J Proteome Res* **17**, 1794-1800.

Dwek, R. A. (1996). Glycobiology: Toward Understanding the Function of Sugars. *Chemical reviews* **96**, 683-720.

Feldman, M. F., Wacker, M., Hernandez, M. & other authors (2005). Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 3016-3021.

Fisher, A. C., Haitjema, C. H., Guarino, C. & other authors (2011). Production of Secretory and Extracellular N-Linked Glycoproteins inEscherichia coli. *Applied and Environmental Microbiology* **77**, 871-881.

Foletti, D., Strop, P., Shaughnessy, L. & other authors (2013). Mechanism of action and in vivo efficacy of a human-derived antibody against Staphylococcus aureus alpha-hemolysin. *Journal of molecular biology* **425**, 1641-1654.

Frasch, C. E. (2009). Preparation of bacterial polysaccharide-protein conjugates: analytical and manufacturing challenges. *Vaccine* **27**, 6468-6470.

Garcia-Quintanilla, F., Iwashkiw, J. A., Price, N. L., Stratilo, C. & Feldman, M. F. (2014). Production of a recombinant vaccine candidate against Burkholderia pseudomallei exploiting the bacterial N-glycosylation machinery. *Frontiers in microbiology* **5**, 381.

Glatter, T., Ludwig, C., Ahrné, E., Aebersold, R., Heck, A. J. R. & Schmidt, A. (2012). Large-Scale Quantitative Assessment of Different In-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion. *Journal of Proteome Research* **11**, 5145-5156.

Glover, K. J., Weerapana, E., Numao, S. & Imperiali, B. (2005). Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. *Chemistry & biology* **12**, 1311-1315.

Gonzalez, J., Takao, T., Hori, H., Besada, V., Rodriguez, R., Padron, G. & Shimonishi, Y. (1992). A method for determination of N-glycosylation sites in glycoproteins by collision-induced dissociation analysis in fast atom bombardment mass spectrometry: identification of the positions of carbohydrate-linked asparagine in recombinant alpha-amylase by treatment with peptide-N-glycosidase F in 18O-labeled water. *Analytical biochemistry* **205**, 151-158.

Gouaux, J. E., Braha, O., Hobaugh, M. R., Song, L., Cheley, S., Shustak, C. & Bayley, H. (1994). Subunit stoichiometry of staphylococcal alpha-hemolysin in crystals and on membranes: a heptameric transmembrane pore. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 12828-12831.

Grandi, G. (2001). Antibacterial vaccine design using genomics and proteomics. *Trends Biotechnol* **19**, 181-188.

Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O. N. & Roepstorff, P. (2004). A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res* **3**, 556-566.

Harding, C. M. & Feldman, M. F. (2019). Glycoengineering bioconjugate vaccines, therapeutics, and diagnostics in E. coli. *Glycobiology* **29**, 519-529.

Harding, C. M., Nasr, M. A., Scott, N. E. & other authors (2019). A platform for glycoengineering a polyvalent pneumococcal bioconjugate vaccine using E. coli as a host. *Nat Commun* **10**, 891.

Heidelberger, M. & Avery, O. T. (1923). THE SOLUBLE SPECIFIC SUBSTANCE OF PNEUMOCOCCUS. *The Journal of experimental medicine* **38**, 73-79.

Hickey, J. M., Sahni, N., Toth, R. T. t., Kumru, O. S., Joshi, S. B., Middaugh, C. R. & Volkin, D. B. (2016). Challenges and opportunities of using liquid chromatography and mass spectrometry methods to develop complex vaccine antigens as pharmaceutical dosage forms. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences* **1032**, 23-38.

Holman, S. W., Sims, P. F. & Eyers, C. E. (2012). The use of selected reaction monitoring in quantitative proteomics. *Bioanalysis* **4**, 1763-1786.

Hu, Q. Y., Berti, F. & Adamo, R. (2016). Towards the next generation of biomedicines by site-selective conjugation. *Chemical Society reviews* **45**, 1691-1719.

Huttenhain, R., Malmstrom, J., Picotti, P. & Aebersold, R. (2009). Perspectives of targeted mass spectrometry for protein biomarker verification. *Current opinion in chemical biology* **13**, 518-525.

Ihssen, J., Haas, J., Kowarik, M., Wiesli, L., Wacker, M., Schwede, T. & Thony-Meyer, L. (2015). Increased efficiency of Campylobacter jejuni N-oligosaccharyltransferase PglB by structure-guided engineering. *Open biology* **5**, 140227.

Jenkins, R., Duggan, J. X., Aubry, A. F. & other authors (2015). Recommendations for validation of LC-MS/MS bioanalytical methods for protein biotherapeutics. *The AAPS journal* **17**, 1-16.

Jones, C. (2005). Vaccines based on the cell surface carbohydrates of pathogenic bacteria. *Anais da Academia Brasileira de Ciencias* **77**, 293-324.

Jones, C. & Ravenscroft, N. (2008). Chapter 2 - NMR Assays for Carbohydrate-Based Vaccines. In NMR Spectroscopy in Pharmaceutical Analysis, pp. 341-368. Edited by U. Holzgrabe, I. Wawer & B. Diehl. Amsterdam: Elsevier.

Jorgenson, M. A. & Young, K. D. (2016). Interrupting Biosynthesis of O Antigen or the Lipopolysaccharide Core Produces Morphological Defects in Escherichia coli by Sequestering Undecaprenyl Phosphate. *J Bacteriol* **198**, 3070-3079.

Josefsberg, J. O. & Buckland, B. (2012). Vaccine process technology. *Biotechnology and bioengineering* **109**, 1443-1460.

Kay, E., Cuccui, J. & Wren, B. W. (2019). Recent advances in the production of recombinant glycoconjugate vaccines. *npj Vaccines* **4**.

Klade, C. S. (2002). Proteomics approaches towards antigen discovery and vaccine development. *Current opinion in molecular therapeutics* **4**, 216-223.

Klock, H. E. & Lesley, S. A. (2009). The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods in molecular biology (Clifton, NJ)* **498**, 91-103.

Konermann, L., Pan, J. & Liu, Y. H. (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society reviews* **40**, 1224-1234.

Kowarik, M., Numao, S., Feldman, M. F., Schulz, B. L., Callewaert, N., Kiermaier, E., Catrein, I. & Aebi, M. (2006a). N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science (New York, NY)* **314**, 1148-1150.

Kowarik, M., Young, N. M., Numao, S. & other authors (2006b). Definition of the bacterial N-glycosylation site consensus sequence. *The EMBO journal* **25**, 1957-1966.

Kramer, N. E., Smid, E. J., Kok, J., de Kruijff, B., Kuipers, O. P. & Breukink, E. (2004). Resistance of Gram-positive bacteria to nisin is not determined by lipid II levels. *FEMS microbiology letters* **239**, 157-161.

Kuster, B. & Mann, M. (1999). 18O-labeling of N-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. *Anal Chem* **71**, 1431-1440.

Kuzyk, M. A., Smith, D., Yang, J., Cross, T. J., Jackson, A. M., Hardie, D. B., Anderson, N. L. & Borchers, C. H. (2009). Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Molecular & cellular proteomics : MCP* **8**, 1860-1877.

La Placa, M. (2012). Principi di microbiologia medica Ed. Società editrice Esculapio.

Langdon, R. H., Cuccui, J. & Wren, B. W. (2009). N-linked glycosylation in bacteria: an unexpected application. *Future microbiology* **4**, 401-412.

Lange, V., Picotti, P., Domon, B. & Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology* **4**, 222.

Lepenies, B. (2015a). Carbohydrate-based Vaccines: Methods and Protocols: Humana Press.

Lepenies, B. (2015b). Carbohydrate-Based Vaccines. Preface. *Methods in molecular biology (Clifton, NJ)* **1331**, v-vi.

Li, Y. F., Arnold, R. J., Tang, H. & Radivojac, P. (2010). The importance of peptide detectability for protein identification, quantification, and experiment design in MS/MS proteomics. *Journal of proteome research* **9**, 6288-6297.

Liljeroos, L., Malito, E., Ferlenghi, I. & Bottomley, M. J. (2015). Structural and Computational Biology in the Design of Immunogenic Vaccine Antigens. *Journal of immunology research* **2015**, 156241.

Linton, D., Dorrell, N., Hitchen, P. G. & other authors (2005). Functional analysis of theCampylobacter jejuniN-linked protein glycosylation pathway. *Molecular microbiology* **55**, 1695-1703.

Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. (2011). X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350-355.

Maclean, B., Tomazela, D. M., Abbatiello, S. E., Zhang, S., Whiteaker, J. R., Paulovich, A. G., Carr, S. A. & Maccoss, M. J. (2010). Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal Chem* **82**, 10116-10124.

Manat, G., Roure, S., Auger, R., Bouhss, A., Barreteau, H., Mengin-Lecreulx, D. & Touze, T. (2014). Deciphering the metabolism of undecaprenyl-phosphate: the bacterial cell-wall unit carrier at the membrane frontier. *Microb Drug Resist* **20**, 199-214.

Medzhitov, R. & Janeway, C. A., Jr. (1997). Innate immunity: impact on the adaptive immune response. *Current opinion in immunology* **9**, 4-9.

Menzies, B. E. & Kernodle, D. S. (1994). Site-directed mutagenesis of the alpha-toxin gene of Staphylococcus aureus: role of histidines in toxin activity in vitro and in a murine model. *Infection and immunity* **62**, 1843-1847.

Micoli, F., Adamo, R. & Costantino, P. (2018a). Protein Carriers for Glycoconjugate Vaccines: History, Selection Criteria, Characterization and New Trends. *Molecules (Basel, Switzerland)* **23**.

Micoli, F., Costantino, P. & Adamo, R. (2018b). Potential targets for next generation antimicrobial glycoconjugate vaccines. *FEMS microbiology reviews* **42**, 388-423.

Napiorkowska, M., Boilevin, J., Sovdat, T., Darbre, T., Reymond, J. L., Aebi, M. & Locher, K. P. (2017). Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat Struct Mol Biol* **24**, 1100-1106.

Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S. C. & Haynes, P. A. (2001). Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* **11**, 535-553.

Nita-Lazar, M., Wacker, M., Schegg, B., Amber, S. & Aebi, M. (2005). The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology* **15**, 361-367.

Oganesyan, V., Peng, L., Damschroder, M. M., Cheng, L., Sadowska, A., Tkaczyk, C., Sellman, B. R., Wu, H. & Dall'Acqua, W. F. (2014). Mechanisms of neutralization of a human anti-alpha-toxin antibody. *The Journal of biological chemistry* **289**, 29874-29880.

Palmisano, G., Melo-Braga, M. N., Engholm-Keller, K., Parker, B. L. & Larsen, M. R. (2012). Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J Proteome Res* **11**, 1949-1957.

Peltola, H., Kayhty, H., Sivonen, A. & Makela, H. (1977a). Haemophilus influenzae type b capsular polysaccharide vaccine in children: a double-blind field study of 100,000 vaccinees 3 months to 5 years of age in Finland. *Pediatrics* **60**, 730-737.

Peltola, H., Makela, H., Kayhty, H. & other authors (1977b). Clinical efficacy of meningococcus group A capsular polysaccharide vaccine in children three months to five years of age. *The New England journal of medicine* **297**, 686-691.

Perez, C., Gerber, S., Boilevin, J., Bucher, M., Darbre, T., Aebi, M., Reymond, J. L. & Locher, K. P. (2015). Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature* **524**, 433-438.

Petrescu, A. J., Wormald, M. R. & Dwek, R. A. (2006). Structural aspects of glycomes with a focus on N-glycosylation and glycoprotein folding. *Current opinion in structural biology* **16**, 600-607.

Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. (2009). Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **138**, 795-806.

Picotti, P. & Aebersold, R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature methods* **9**, 555-566.

Pobre, K., Tashani, M., Ridda, I., Rashid, H., Wong, M. & Booy, R. (2014). Carrier priming or suppression: understanding carrier priming enhancement of anti-polysaccharide antibody response to conjugate vaccines. *Vaccine* **32**, 1423-1430.

Pozsgay, V., Chu, C., Pannell, L., Wolfe, J., Robbins, J. B. & Schneerson, R. (1999). Protein conjugates of synthetic saccharides elicit higher levels of serum IgG lipopolysaccharide antibodies in mice than do those of the O-specific polysaccharide from Shigella dysenteriae type 1. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5194-5197.

Rampler, E., Schoeny, H., Mitic, B. M., El Abiead, Y., Schwaiger, M. & Koellensperger, G. (2018). Simultaneous non-polar and polar lipid analysis by on-line combination of HILIC, RP and high resolution MS. *The Analyst* **143**, 1250-1258.

Rappuoli, R., De Gregorio, E. & Costantino, P. (2019). On the mechanisms of conjugate vaccines. *Proceedings of the National Academy of Sciences* **116**, 14.

Ravenscroft, N., Braun, M., Schneider, J. & other authors (2019). Characterization and immunogenicity of a Shigella flexneri 2a O-antigen bioconjugate vaccine candidate. *Glycobiology* **29**, 669-680.

Reglinski, M., Ercoli, G., Plumptre, C., Kay, E., Petersen, F. C., Paton, J. C., Wren, B. W. & Brown, J. S. (2018). A recombinant conjugated pneumococcal vaccine that protects against murine infections with a similar efficacy to Prevnar-13. *npj Vaccines* **3**, 53.

Ridenour, C., Johnson, A., Winne, E. & other authors (2015). Development of influenza A(H7N9) candidate vaccine viruses with improved hemagglutinin antigen yield in eggs. *Influenza and other respiratory viruses* **9**, 263-270.

Roberts, I. S. (1996). The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annual review of microbiology* **50**, 285-315.

Santana, W. I., Williams, T. L., Winne, E. K., Pirkle, J. L. & Barr, J. R. (2014). Quantification of viral proteins of the avian H7 subtype of influenza virus: an isotope dilution mass spectrometry method applicable for producing more rapid vaccines in the case of an influenza pandemic. *Anal Chem* **86**, 4088-4095.

Schneerson, R., Barrera, O., Sutton, A. & Robbins, J. B. (1980). Preparation, characterization, and immunogenicity of Haemophilus influenzae type b polysaccharide-protein conjugates. *The Journal of experimental medicine* **152**, 361-376.

Schulz-Trieglaff, O., Pfeifer, N., Gröpl, C., Kohlbacher, O. & Reinert, K. (2008). LC-MSsim – a simulation software for liquid chromatography mass spectrometry data. *BMC Bioinformatics* **9**, 423.

Segu, Z. M., Hussein, A., Novotny, M. V. & Mechref, Y. (2010). Assigning N-glycosylation sites of glycoproteins using LC/MSMS in conjunction with endo-M/exoglycosidase mixture. *J Proteome Res* **9**, 3598-3607.

Sharma, H. J., Patil, V. D., Lalwani, S. K. & other authors (2012). Assessment of safety and immunogenicity of two different lots of diphtheria, tetanus, pertussis, hepatitis B and Haemophilus influenzae type b vaccine manufactured using small and large scale manufacturing process. *Vaccine* **30**, 510-516.

Sharma, V. K., Sharma, I. & Glick, J. (2018). The expanding role of mass spectrometry in the field of vaccine development. *Mass Spectrom Rev*.

Shi, T., Su, D., Liu, T., Tang, K., Camp, D. G., 2nd, Qian, W. J. & Smith, R. D. (2012). Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. *Proteomics* **12**, 1074-1092.

Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. & Geromanos, S. J. (2006). Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Molecular & cellular proteomics : MCP* **5**, 144-156.

Silverman, J. M. & Imperiali, B. (2016). Bacterial N-Glycosylation Efficiency Is Dependent on the Structural Context of Target Sequons. *The Journal of biological chemistry* **291**, 22001-22010.

Sun, X., Stefanetti, G., Berti, F. & Kasper, D. L. (2019). Polysaccharide structure dictates mechanism of adaptive immune response to glycoconjugate vaccines. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 193-198.

Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. & Guerry, P. (1999). Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Molecular microbiology* **32**, 1022-1030.

Szymanski, C. M. & Wren, B. W. (2005). Protein glycosylation in bacterial mucosal pathogens. *Nature Reviews Microbiology* **3**, 225-237.

Towbin, H., Staehelin, T. & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A* **76**, 4350-4354.

Towbin, H., Staehelin, T. & Gordon, J. (1992). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. 1979. *Biotechnology* **24**, 145-149.

van den Dobbelsteen, G., Fae, K. C., Serroyen, J. & other authors (2016). Immunogenicity and safety of a tetravalent E. coli O-antigen bioconjugate vaccine in animal models. *Vaccine* **34**, 4152-4160.

Verez-Bencomo, V., Fernandez-Santana, V., Hardy, E. & other authors (2004). A synthetic conjugate polysaccharide vaccine against Haemophilus influenzae type b. *Science (New York, NY)* **305**, 522-525.

Vidova, V. & Spacil, Z. (2017). A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Analytica chimica acta* **964**, 7-23.

Wacker, M., Linton, D., Hitchen, P. G. & other authors (2002). N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science (New York, NY)* **298**, 1790-1793.

Wacker, M., Feldman, M. F., Callewaert, N. & other authors (2006). Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7088-7093.

Wacker, M., Wang, L., Kowarik, M. & other authors (2014). Prevention of Staphylococcus aureus infections by glycoprotein vaccines synthesized in Escherichia coli. *J Infect Dis* **209**, 1551-1561.

Walker, B. & Bayley, H. (1995). Key residues for membrane binding, oligomerization, and pore forming activity of staphylococcal alpha-hemolysin identified by cysteine scanning mutagenesis and targeted chemical modification. *The Journal of biological chemistry* **270**, 23065-23071.

Wetter, M., Goulding, D., Pickard, D., Kowarik, M., Waechter, C. J., Dougan, G. & Wacker, M. (2012). Molecular characterization of the viaB locus encoding the biosynthetic machinery for Vi capsule formation in Salmonella Typhi. *PloS one* **7**, e45609.

Wetter, M., Kowarik, M., Steffen, M., Carranza, P., Corradin, G. & Wacker, M. (2013). Engineering, conjugation, and immunogenicity assessment of Escherichia coli O121 O antigen for its potential use as a typhoid vaccine component. *Glycoconjugate journal* **30**, 511-522.

Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annu Rev Biochem* **75**, 39-68.

Whitfield, C. & Trent, M. S. (2014). Biosynthesis and export of bacterial lipopolysaccharides. *Annu Rev Biochem* **83**, 99-128.

Whitley, P., Nilsson, I. M. & von Heijne, G. (1996). A nascent secretory protein may traverse the ribosome/endoplasmic reticulum translocase complex as an extended chain. *The Journal of biological chemistry* **271**, 6241-6244.

Williams, T. L., Luna, L., Guo, Z., Cox, N. J., Pirkle, J. L., Donis, R. O. & Barr, J. R. (2008). Quantification of influenza virus hemagglutinins in complex mixtures using isotope dilution tandem mass spectrometry. *Vaccine* **26**, 2510-2520.

Williams, T. L., Pirkle, J. L. & Barr, J. R. (2012). Simultaneous quantification of hemagglutinin and neuraminidase of influenza virus using isotope dilution mass spectrometry. *Vaccine* **30**, 2475-2482.

Wright, A., Dankert, M. & Robbins, P. W. (1965). Evidence for an intermediate stage in the biosynthesis of the Salmonella O-antigen. *Proceedings of the National Academy of Sciences of the United States of America* **54**, 235-241.

Wright, A., Dankert, M., Fennessey, P. & Robbins, P. W. (1967). Characterization of a polyisoprenoid compound functional in O-antigen biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **57**, 1798-1803.

Yost, R. A. & Enke, C. G. (1979). Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal Chem* **51**, 1251-1264.

Young, N. M., Brisson, J. R., Kelly, J. & other authors (2002). Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. *The Journal of biological chemistry* **277**, 42530-42539.

Zarei, A. E., Almehdar, H. A. & Redwan, E. M. (2016). Hib Vaccines: Past, Present, and Future Perspectives. *Journal of immunology research* **2016**, 7203587.

Zhang, H., Li, X. J., Martin, D. B. & Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **21**, 660-666.

Zhang, H., Liu, Q., Zimmerman, L. J. & other authors (2011). Methods for peptide and protein quantitation by liquid chromatography-multiple reaction monitoring mass spectrometry. *Molecular & cellular proteomics : MCP* **10**, M110.006593.

Zhu, Z., Go, E. P. & Desaire, H. (2014). Absolute quantitation of glycosylation site occupancy using isotopically labeled standards and LC-MS. *J Am Soc Mass Spectrom* **25**, 1012-1017.