# Comprehensive characterization of bacterial glycoconjugate vaccines by liquid chromatography: mass spectrometry

Marco, F. di; Ederveen, A.L.H.; Schaick, G. van; Moran, A.B.; Domínguez-Vega, E.; Nicolardi, S.; ... ; Wuhrer, M.

Carbohydrate Polymers 341 (2024) 122327

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

# Comprehensive characterization of bacterial glycoconjugate vaccines by liquid chromatography - mass spectrometry

Fiammetta Di Marco [a,1], Agnes L. Hipgrave Ederveen [a,1], Guusje van Schaick [a], Alan B. Moran [c], Elena Domínguez-Vega [a], Simone Nicolardi [a], Constantin Blöchl [a], Carolien A. Koeleman [a], Renzo Danuser [b], Ali Al Kaabi [b], Viktoria Dotz [c,d], Jan Grijpstra [c], Michel Beurret [c], Chakkumkal Anish [c], Manfred Wuhrer [a,*]

[a] Center for Proteomics and Metabolomics, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands
[b] Janssen Vaccines AG (Branch of Cilag GmbH International), Rehhagstrasse 79, CH-3018 Bern, Switzerland
[c] Bacterial Vaccines Discovery and Early Development, Janssen Vaccines and Prevention B.V., Archimedesweg 4-6, 2333 CN Leiden, the Netherlands
[d] BioTherapeutics Analytical Development, Janssen Biologics B.V., Einsteinweg 101, 2333 CB Leiden, the Netherlands

ARTICLE INFO

*Keywords:*
Bacterial vaccine
O-antigen
Glycoconjugate
Mass spectrometry
Glycopeptide analysis
Intact protein analysis

ABSTRACT

Bacterial pathogens can cause a broad range of infections with detrimental effects on health. Vaccine development is essential as multi-drug resistance in bacterial infections is a rising concern. Recombinantly produced proteins carrying O-antigen glycosylation are promising glycoconjugate vaccine candidates to prevent bacterial infections. However, methods for their comprehensive structural characterization are lacking. Here, we present a bottom-up approach for their site-specific characterization, detecting *N*-glycopeptides by nano reversed-phase liquid chromatography-mass spectrometry (RP-LC-MS). Glycopeptide analyses revealed information on partial site-occupancy and site-specific glycosylation heterogeneity and helped corroborate the polysaccharide structures and their modifications. Bottom-up analysis was complemented by intact glycoprotein analysis using nano RP-LC-MS allowing the fast visualization of the polysaccharide distribution in the intact glycoconjugate. At the glycopeptide level, the model glycoconjugates analyzed showed different repeat unit (RU) distributions that spanned from 1 to 21 RUs attached to each of the different glycosylation sites. Interestingly, the intact glycoprotein analysis displayed a RU distribution ranging from 1 to 28 RUs, showing the predominant species when the different glycopeptide distributions are combined in the intact glycoconjugate. The complete workflow based on LC-MS measurements allows detailed and comprehensive analysis of the glycosylation state of glycoconjugate vaccines.

## Hypothesis

The distribution of bacterial O-antigen polysaccharides in glycoconjugate vaccines can be unraveled and semi-quantified by glycopeptide and intact glycoform analysis using liquid chromatography - mass spectrometry approaches.

## 1. Introduction

The emergency of bacterial multi-drug resistance causing significant morbidity worldwide is demanding the development of alternative protection strategies, such as novel glycoengineered vaccines (Aslam et al., 2018; Kourtis et al., 2021; Micoli et al., 2019). Bacterial surface O-polysaccharide (O-PS) molecules play a pivotal role as virulence factors for many pathogens. The immunogenic O-antigen component of the

*Abbreviations:* DEN, Dopant-enriched nitrogen; EPA, *Pseudomonas aeruginosa* exoprotein A; Gal, Galactose; GalNAc, *N*-acetylgalactosamine; Glc, Glucose; GlcNAc, *N*-acetylglucosamine; Hex, Hexose; HexNAc, *N*-acetylhexosamine; isCID, In-source collisional induced dissociation; LC, Liquid chromatography; Man, Mannose; ManNAc, *N*-acetylmannosamine; MS, Mass spectrometry; MS/MS, Tandem mass spectrometry; O-PS, O-polysaccharide; Rha, Rhamnose; RP, Reversed phase; RUs, Repeat units.

* Corresponding author.
*E-mail address:* m.wuhrer@lumc.nl (M. Wuhrer).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.carbpol.2024.122327
Received 12 March 2024; Received in revised form 22 May 2024; Accepted 25 May 2024
Available online 30 May 2024

bacterial lipopolysaccharides, composed of repeat units (RUs) of two to seven monosaccharides, is exposed on the surface of the bacterial cell where it modulates the pathogenesis of the bacteria providing a promising candidate for protective vaccines (Lee et al., 2022). Among the diverse types of vaccines developed, glycoconjugate-based vaccines are attractive and effective protective systems against bacterial infections, consisting of an immunogenic, strain-specific O-PS covalently linked to a carrier protein (Huttner & Gambillara, 2018; Micoli et al., 2019; Stefanetti et al., 2019). The protein-conjugated O-PS is required to elicit an appropriate and lasting immune response (Micoli et al., 2018; Rappuoli, 2018).

Glycoconjugation can be achieved using linkers with specific chemical functional groups (chemical conjugation) or via the use of specific oligosaccharyltransferases (bioconjugation). For the bioconjugation approach, a carrier protein, such as the detoxified *Pseudomonas aeruginosa* exoprotein A (EPA) or the AcrA protein (Kay et al., 2019), is decorated with *N*-linked glycans by the oligosaccharyltransferase PglB in an engineered bacterial production strain (Feldman et al., 2005; Iwashkiw et al., 2012; Poolman & Wacker, 2016; van den Dobbelsteen et al., 2016). These glycoengineered proteins often contain several potential asparagine residues with the bacterial *N*-glycosylation consensus sequence D/E-Z-N-X-T/P (where Z, X $\neq$ P) (Kowarik et al., 2006), which upon glycoconjugation will carry heterogeneous mixtures of polysaccharides with a specific RU chain length distribution and site-occupancy for each glycosylation site, depending on the employed PglB variant (Terra et al., 2022). Usually, the O-polysaccharide of the versatile and diverse *E. coli* strains show different chemical structures and contains between 10 and 25 RUs, each consisting of two to seven (different) sugar residues (Liu et al., 2020; Stenutz et al., 2006). The bioconjugation process results in the production of a heterogeneous mixture of glycoconjugate. Product heterogeneity may arise from diversity of O-antigen RUs in cell, biases and preferences of the glycosyl transferases, differential expression of glycans due to fermentation conditions, heterogeneity due to different length regulators of the O-antigen synthesis and diverse decorations of monosaccharides such as capping sugars, branching sugars and acetylation. Moreover, the bioconjugation may results in potential differences of glycosylation efficiency on various glycosylation sites (site occupancy differences due to carrier protein structure) (Anish et al., 2021).

Thus, analytical strategies to monitor the efficiency and consistency of glycosylation are critical to ensure the efficacy of the glycoconjugate vaccines (Cuccui & Wren, 2015). Conventionally, information on the polysaccharide chain length distribution is obtained with SDS-PAGE-based methods and light scattering methods such as SEC-MALS (Goldman & Leive, 1980; Micoli et al., 2018). Orthogonal techniques employed for glycoconjugate characterization include NMR and monosaccharide linkage analysis using gas chromatography coupled to mass spectrometry (MS) (Jansson et al., 1987; Kim et al., 2005; Ravenscroft et al., 2016). Only a few studies in literature reported the analysis of glycopeptides by liquid chromatography – tandem mass spectrometry (LC-MS/MS) solely to determine the O-antigen attachment sites in *O*-glycosylated bioconjugates (Harding et al., 2019; Jiang et al., 2021, 2022; Scott et al., 2014). In addition, LC-MS of intact glycoconjugates carrying long polysaccharide chains has been reported (Harding et al., 2019; Jiang et al., 2021, 2022; Ravenscroft et al., 2019). A more direct and faster characterization of glycoconjugates was achieved using MALDI in-source decay FT-ICR MS. However, this approach only provided indirect and limited information on the glycosylation state of the glycoconjugate (Nicolardi et al., 2022).

While general methods are in place for the determination of polysaccharide attachment to the carrier protein, there is a lack of sensitive methodologies to monitor site-occupancy, the number of RUs and the glycan chain integrity in a site-specific manner, hampering polysaccharide conjugate development and production monitoring. Compared to the characterization of "conventional" biopharmaceuticals (Beck et al., 2013; Sandra et al., 2014; Struwe & Robinson, 2019; Yang, Franc, & Heck, 2017; Yang, Wang, et al., 2017), bacterial glycoconjugates present new analytical challenges. The mass of the polysaccharide can be as high as the carrier protein itself, with each glycosylation site carrying glycosylation chains of up to 15 kDa. In addition, the glycoconjugates may present a high level of heterogeneity due to the different number of attached RUs and the diverse modifications present on the bacterial glycans (Imperiali, 2019; Liu et al., 2020).

Approaches to characterize protein glycosylation using LC-MS-based methodologies remain challenging due to the poor ionization of glycoconjugates and the microheterogeneity of glycosylated species leading to the acquisition of complex mass spectra that display numerous ion signals carrying multiple charge states. Moreover, the presence of end-capped sugars can lead to a higher heterogeneity. The glycopeptides produced upon enzymatic protein digestion and the intact glycoconjugates are highly hydrophilic due to the attachment of the unusually long glycan chain composed of multiple repeat units, bringing an analytical challenge for their retention in a reversed-phase LC column, their ESI ionization and the possibility of in-source MS fragmentation of the long polysaccharide chain (Yu et al., 2018). Dopant-enriched nitrogen (DEN) gas has been found to minimize the negative ionization bias of glycosylated versus non-glycosylated species. Generally, non-glycosylated hydrophobic species show higher ionization efficiencies as compared to the hydrophilic glycosylated species (Mysling et al., 2010). The use of DEN gas during ESI-MS has been found to boost sensitivity for the analysis of released glycans, conventional glycopeptides and intact protein glycoforms (Alagesan & Kolarich, 2019; Kammeijer et al., 2016; Madunić et al., 2021; Van Schaick et al., 2023).

Here, we present a workflow for the comprehensive characterization of the *N*-glycosylation state of five model glycoconjugates composed of five different protein carriers (four recombinantly produced EPA- and one recombinant AcrA protein) carrying three different O-PS (i.e., O21, O75 and O1A1) (Baumann et al., 1991; Erbing et al., 1978; Gupta et al., 1992; Jann et al., 1992; Li et al., 2011; Staaf et al., 1999). Our approach aimed at the characterization of the glycoconjugates at the two different structural levels of glycopeptides and intact bioconjugate glycoforms to obtain complementary information on the glycosylation state of the bacterial vaccines. A DEN gas-assisted bottom-up approach provided insights into the chain length distribution and characterization of the RU structural heterogeneity on a site-specific level. Using a complementary intact mass approach based on nano LC-MS, the heterogeneity of intact glycoconjugate proteoforms was determined, providing additional information on the total number of attached RU. The data obtained at the two structural levels were compared and integrated using the software MoFi as already reported in other studies (Lebede et al., 2021; Skala et al., 2018; Wohlschlager et al., 2018). To our knowledge, this is the first report on semi-quantitative site-specific glycosylation analysis of glycoengineered bacterial vaccines performed by LC-MS of glycopeptides carrying intact O-antigen polysaccharides. This sensitive workflow demonstrated its feasibility for key structural characterization, quality control and understanding of bioconjugate consistency and can be used as a guideline for glycoconjugate structural elucidation employing liquid chromatography coupled to mass spectrometry.

## 2. Material and methods

### 2.1. Materials

All chemicals were at least analytical grade quality. Acetonitrile (ACN) of LC–MS grade originated from Biosolve (Valkenswaard, The Netherlands). Rapigest SF Surfactant was obtained from Waters (Milford, MA). Acetic acid, ammonium bicarbonate, DL-dithiothreitol (DTT), formic acid, iodoacetamide (IAA), and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (Steinheim, Germany). Porcine trypsin, sequencing grade, was obtained from Promega (Madison, WI). Ultrapure deionized water was generated using a Purelab Chorus 1, maintained at $\geq$18.2 MΩ (Veolia Water Technologies

Netherlands B.V., Ede, the Netherlands).

The five glycoconjugates analyzed in this study were provided by Janssen Vaccines & Prevention B.V. (Leiden, The Netherlands) and encompassed: *i.* an aa-sequence-modified EPA protein embedding one glycosylation site at Asn 8 (EPAg1_O75) and *ii.* an aa-sequence-modified EPA protein singly-glycosylated at Asn 639 (EPAg4_O75) both expressed in engineered *E. coli* cells producing serotype O75 glycans (RU -3 [Manβ1-4] Galα1-4Rhaα1-3GlcNAcβ1-) (Erbing et al., 1978), *iii.* an aa-sequence-modified EPA 2 protein embedding two glycosylation sites at Asn 243 and 385 (EPAg2,3_O21), *iv.* an AcrA protein diglycosylated at Asn 105 and 255 (AcrAg1,2_O21) both carrying *E. coli* serotype O21 glycans -3[Galβ1-4] Galβ1-4[GlcNAcβ1-2] Glcβ1-3GalNAcβ1- (Li et al., 2011; Staaf et al., 1999) and *v.* a multi glycosylated EPA protein carrying *E. coli* serotype O1A1 glycans -3[ManNAcβ1-2] Rhaα1-3Rhaα1-3Rhaβ1-4GlcNAcβ1- (EPAg1,2,3,4_O1) (Baumann et al., 1991; Gupta et al., 1992; Jann et al., 1992) (Man: mannose, Rha: rhamnose, Glc: glucose, Gal: galactose, ManNAc: *N*-acetylmannosamine, GalNAc: *N*-acetylgalactosamine and GlcNAc: *N*-acetylglucosamine).

### 2.2. Glycopeptide analysis by nano RP-LC-MS

The glycoengineered EPA and AcrA carrying *E. coli* O21, O1A1 and O75 type O-PS (10 μg) were denatured in the presence of 0.5 % Rapigest SF surfactant and 2 mmol $L^{-1}$ DTT in 16 mmol $L^{-1}$ ammonium bicarbonate (pH 8) for 30 min at 60 °C. For alkylation, 6 mmol $L^{-1}$ IAA was added and the sample was kept in the dark for 30 min at room temperature. To avoid overalkylation DTT was added to a final concentration of 6 mmol $L^{-1}$, and the sample was kept in bright light for 20 min at room temperature. For the generation of glycopeptides, samples were proteolytically cleaved using trypsin (enzyme-substrate ratio; 1:50; *w/w*) with 16–18 h incubation at 37 °C in ammonium bicarbonate buffer (final concentration 35 mmol $L^{-1}$; pH 8). Digestion was stopped and the surfactant was hydrolyzed by adding TFA to a concentration of 1.5 % (*v/v*) followed by incubation at 37 °C for 45 min. Samples were centrifuged at 15,800 ×*g* for 10 min to remove the precipitated surfactant. The supernatant was dried by vacuum centrifugation at 45 °C. Prior to LC-MS analysis, the samples were dissolved in water to a concentration of 20 ng $\mu L^{-1}$. The tryptic digests of the bioconjugate samples were separated and analyzed by a C18 RP-LC-MS/MS. The Thermo Scientific™ Ultimate™ 3000 RSLCnano UHPLC system (Thermo Fisher Scientific, Waltham, MA) system was equipped with an Acclaim PepMap 100 trap column (100 μm × 20 mm, particle size 5 μm, Thermo Scientific) and an Acclaim PepMap RSLC C18 nano-column (75 μm × 150 mm, particle size 2 μm, Thermo Fisher Scientific). Five microliters of sample (100 ng) were injected and the glycopeptides were separated with a gradient from 99 % solvent A (0.1 % formic acid in water) and 1 % solvent B (0.1 % formic acid in 95 % ACN) to 30 % solvent B over 30 min, with a flow rate of 700 nL $min^{-1}$. The Ultimate™ 3000 RSLCnano UHPLC system (Thermo Fisher Scientific) was coupled to a maXis quadrupole time-of-flight-MS (q-TOF-MS; Bruker Daltonics, Bremen, Germany) equipped with a CaptiveSpray nanoESI source (Bruker Daltonics). Ionization was enhanced by applying DEN gas at 0.2 bar using acetonitrile as a dopant solvent. The nitrogen drying gas temperature was set at 150 °C with a flow of 3.0 L $min^{-1}$. Furthermore, ionization conditions were evaluated by a triplicate injection of 100 ng glycoconjugates. The nebulizer gas pressure was kept constant for the assessment of enriched $N_2$ (all conditions except for ambient air). Profile spectra were recorded in an *m/z* range from 50 to 4000 with a frequency of 1 Hz. The collision energy was 5 eV, the transfer time 115 μs and the pre-pulse storage 15 μs. Tandem mass spectra data were acquired at 1 Hz in the *m/z* range 50–2800 or 50–4000 with the precursor selection above *m/z* values 350 or 1000. In each MS cycle, one MS spectrum was followed by product-ion spectra of the four most intense precursors using an isolated width of 8–15 Da depending on the *m/z* values. Collision energies were increased linearly in an *m/z* dependent manner from 55 eV to 124 eV, for all charge states. Stepping mode was applied for the tandem MS collision energy (100 %–50 %) to 80 % of the TOF summations used for each spectrum; for the remaining 20 %, the collision energy was halved. In this mode of operation, ion counts from each subdivision are summed together. The energies were adjusted for the method with *m/z* range 50–4000 as follows, basic stepping mode was applied with collision energy (70 %–40 %) each 80 % and 20 % of the time.

### 2.3. Intact glycoform analysis by nano RP-LC-MS

The intact glycoconjugates were diluted in 150 mmol $L^{-1}$ aqueous solution of ammonium acetate to a final concentration of 100 ng $\mu L^{-1}$ and directly transferred into HPLC vials. Analysis of the intact glycoforms was performed by nano RP-LC-MS using an Ultimate™ 3000 RSLCnano UHPLC system (Thermo Fisher Scientific) equipped with a PepMap™ 300 C4 trapping column (300 Å C4 5 μm 0.3 × 5 mm; Thermo Fisher Scientific) and a Halo® Diphenyl nano column (1000 Å, 2.7 μm, 1.0 × 50 mm) from Advance Materials Technology. Mobile phase A consisted of $H_2O$ + 0.1 % TFA and mobile phase B of 95 % ACN + 0.1 % TFA. 100 ng of the sample (1 μL injection volume) was loaded on the trapping column and flushed for 5 min with 30 % B at a flow rate of 1 μL $min^{-1}$. Subsequentially, the flow was redirected to the nano LC column with a 6-port valve and a linear gradient from 30 % to 60 % B was carried out in 15 min. The column was then flushed for 5 min at 90 % B and equilibrated for 5 min at 30 % B for a total chromatographic analysis time of 30 min. The nano LC was coupled to a Bruker maXis impact™ high-resolution QTOF mass spectrometer equipped with a CaptiveSpray nanoESI source (Bruker Daltonics, Bremen, Germany). The ionization was enhanced by applying ACN-doped nebulizing gas. MS parameters were set as follows: positive ionization mode with a capillary voltage of 800 V, nebulizer gas of 0.10 bar, dry gas flow of 3.0 L $min^{-1}$, and dry temperature of 220 °C. The isCID was set to 70, 90 or 100 eV. The collision energy was 7 eV, the collision frequency was 2000 Vpp, the transfer time was 150 μs, and the pre-pulse storage was 20 μs. Profile spectra were acquired in the *m/z* range 300–4000 with a frequency of 1 Hz. The measurements were performed in triplicate.

### 2.4. Data analysis

Raw data were visually inspected in DataAnalysis v5.0 (Bruker Daltonics). The glycopeptide mass spectra were screened for exact mass, isotopic pattern, retention time and relative intensities. The initial manual glycopeptide assignments were used to generate a list of (possible) glycopeptides with their corresponding glycans. This list was imported for each glycopeptide in Skyline (MacCoss Lab, Department of Genome Sciences, University of Washington, v 22.2.0.351). Using this software, glycopeptide semi-quantitation was obtained based on the extracted ion current chromatograms (XICs) of MS1 ions. Glycopeptides were validated considering mass match below 20 ppm, theoretical and experimental isotopic pattern matching (idotp ≥ 0.8) and retention time. The following Skyline transition parameters were used: precursor charges from 1 to 11, ion charges 1, ion types "p", instrument *m/z* range 50–4.000, method match tolerance *m/z* to 0.055, full-scan $MS^1$ filtering isotope peaks included "percent", precursor mass analyzer "TOF", min % of base peak 10 %, and the resolving power was 30.000 at *m/z* 200. Information was obtained from the Skyline output file on the (relative) abundance of glycans for each glycosylation site. Each glycopeptide that contained modified sites was normalized individually so that the sum of all its glycovariant areas was set to 100 %.

Peptide hydrophobicity score was calculated using the freely available Peptide Synthesis and Proteotypic Peptide Analyzing Tool by Thermo Fisher Scientific (https://www.thermofisher.com/nl/en/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.html).

Intact MS data were analyzed by deconvolution of raw mass spectra of intact glycoforms using the 'Maximum Entropy' algorithm of the DataAnalysis software. Deconvolution parameters were set as follows:

**Table 1**
Overview of the proteins analyzed in this study. The *E. coli* serotypes are presented together with repeat unit (RU) structures, name of the glycoconjugate, protein type, tryptic peptides generated upon digestion and glycosylation sites. In addition, the peptide hydrophobicity is reported as a score with negative values indicating a more hydrophilic peptide. The symbol indicates monosaccharides: green circle mannose, yellow circle galactose, blue square *N*-acetylglucosamine, yellow square *N*-acetylgalactosamine, green square *N*-acetylmannosamine, green triangle rhamnose. EPA stands for *Pseudomonas aeruginosa* exoprotein A.

| *E. coli* serotype | RU | Glycoconjugate name | Carrier protein | Peptide sequence | Glycosylation site | Hydrophobicity |
|---|---|---|---|---|---|---|
| O75 |  | EPAg1_O75 | modified EPA | GSGGGDQNATGSGGGK | Asn 8 | −0.23 |
| | | EPAg4_O75 | modified EPA | LGSGGGDQNAT | Asn 639 | 4.66 |
| O21 |  | EPAg2,3_O21 | modified EPA | DNNNSTPTVISHR | Asn 243 | 11.54 |
| | | | | HDLDIKDNNNSTPTVISHR (1MC) | | 23.91 |
| | | | | DQNR | Asn 385 | 1.49 |
| | | AcrAg1,2_O21 | AcrA | DFNR | Asn 105 | 5.47 |
| | | | | AVFDNNNSTLLPGAFATITSEGFIQK | Asn 255 | 43.31 |
| O1A1 |  | EPAg1,2,3,4_O1 | EPA | GSGGGDQNATGSGGGK | Asn8 | −0.23 |

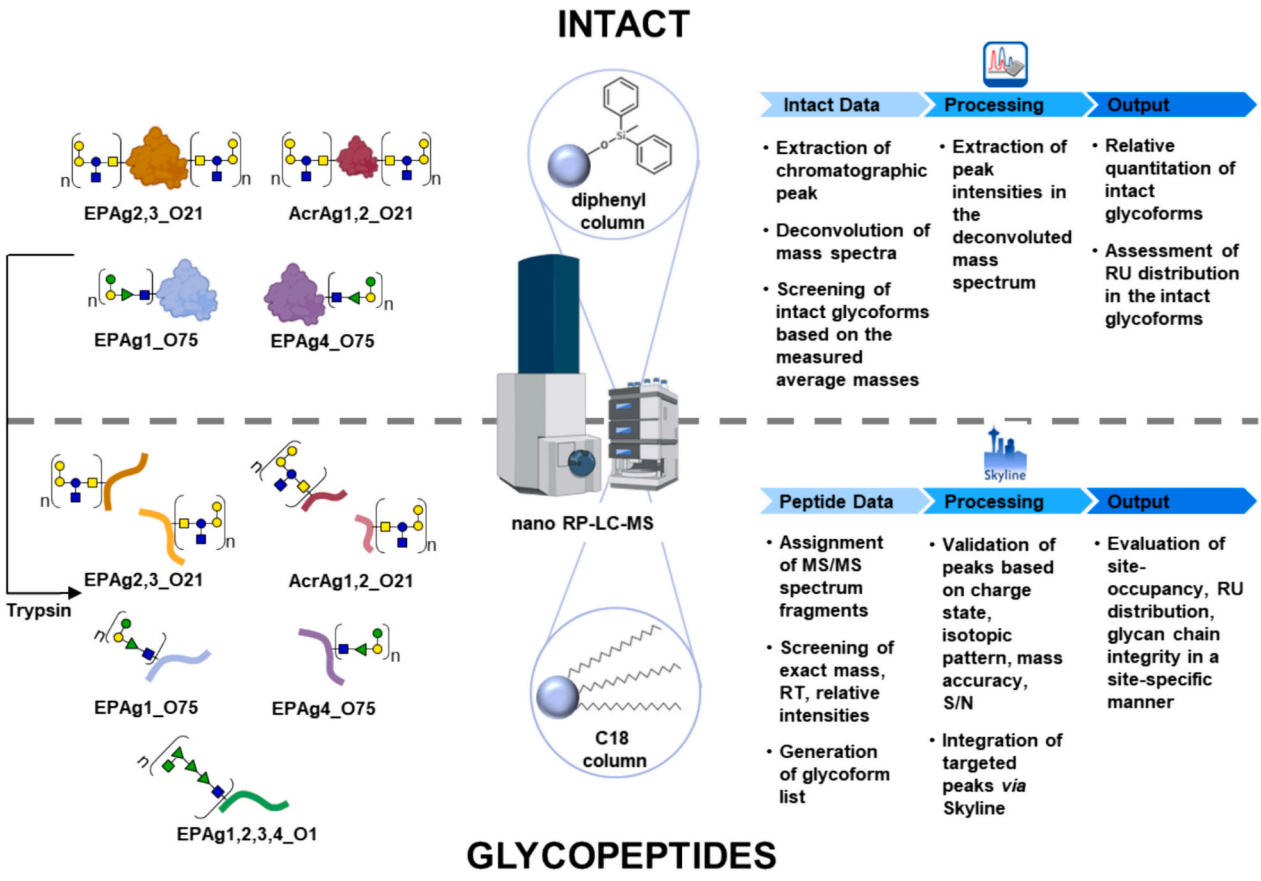


**Fig. 1.** Overview of the analytical workflows employed for the analysis of the five glycoconjugates EPAg2,3_O21, Acrag1,2_O21, EPAg1_O75, EPAg4_O75 and EPAg1,2,3,4_O1. RP-LC-MS: Reversed-phase liquid chromatography coupled to mass spectrometry.

adduct ion +H, low mass 60 kDa for EPA and 30 kDa for AcrA, high mass 100 kDa for EPA and 70 kDa for AcrA, data point spacing 1, instrument resolving power 5000, resolution normal. The entire *m/z* range in the raw mass spectra was deconvoluted. The data were manually screened for the glycoconjugate intact proteoforms based on their average masses. Relative quantitation of intact bioconjugate glycoforms was performed based on the signal intensities of the glycoforms in the deconvoluted mass spectrum. Moreover, deconvoluted mass spectra were annotated using the software tool MoFi (Skala et al., 2018) with the following parameters: disulfide bridges 0 for Acrag1,2_O21 and 4 for EPAg2,3_O21, mass set average (IUPAC), variable modification Hex and HexNAc and fixed modification +106.10 Da for Acrag1,2_O21 and +71.03 Da for EPAg2,3_O21, mass tolerance 50 ppm. The peak intensities of the deconvoluted mass spectrum were calculated as the average of three replicates and used in MoFi for the annotations. A site-specific semi-quantitative glycopeptide library was built from glycopeptide data excluding the RU carrying modifications and normalizing to 100 % the other variants. Peaks in the deconvoluted mass spectrum due to glycoforms carrying RU modification were also excluded from the annotations to avoid the explosion of the software combinatorial space.

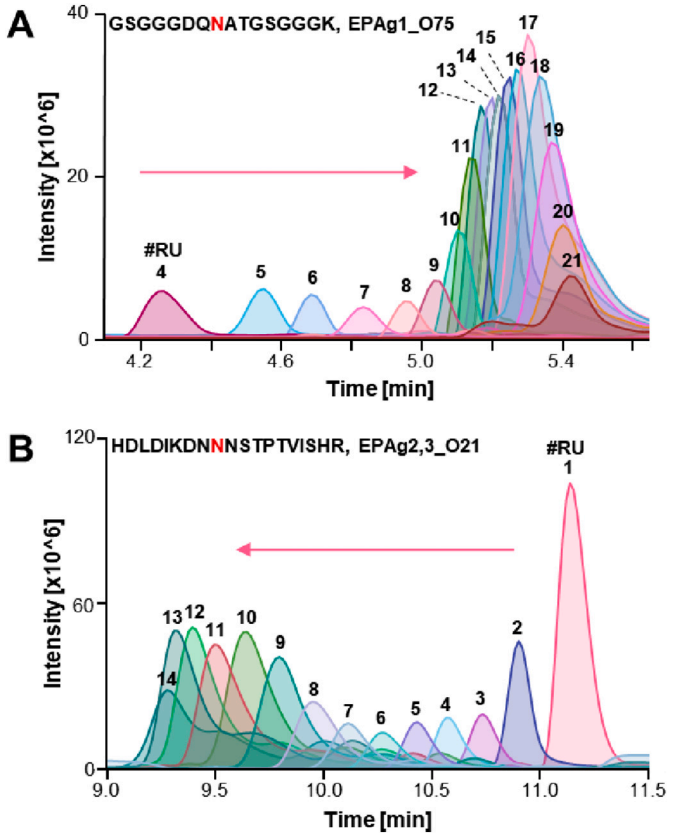


**Fig. 2.** Nano RP-LC-MS analysis chromatogram of tryptic glycopeptides. XICs of the main glycovariants of (A) the glycopeptide of EPAg1_O75 embedding the glycosylated Asn 8 and (B) the glycopeptide of EPAg2,3_O21 embedding the glycosylation site Asn 243. The numbers above the peaks indicate the number of RUs attached to the peptide moiety.

## 3. Results and discussion

### 3.1. Site-specific glycosylation characterization by nano RP-LC-MS analysis of glycopeptides

The five glycoconjugates analyzed in this study carried polydispersed O-PS from three different *E. coli* serotypes (Table 1), resulting in highly heterogeneous glycoproteins. Such complexity was tackled by an analytical strategy that combined bottom-up MS and intact protein analysis (Fig. 1).

Glycopeptide analysis was used to determine the RU distribution and site occupancy for each *N*-glycosylation site (Table 1). The use of C18 reversed-phase chromatography allowed to separate the different peptides based on their hydrophobicity. More intriguingly, glycopeptides with the same sequence were partially separated based on the number of RUs resulting in the elution of a cluster of different glycoforms of a peptide over a range of several minutes as exemplified for two glycopeptides of EPAg2,3_O21 and EPAg1_O75 (Fig. 2). Of note, the chromatographic elution order of the different glycoforms (different number of RUs attached) can be opposite depending on the length and the hydrophobicity of the peptides. Specifically, the hydrophilic glycopeptides of EPAg1_O75 (GSGGGDQNATGSGGGK) showed an increased retention time when a higher number of RUs was attached (Fig. 2A), while the opposite was observed for the more hydrophobic glycopeptides of EPAg2,3_O21 (HDLDIKDNNNSTPTVISHR) showing the chromatographic elution decreasing with a higher number of RU attached (Fig. 2B).

The steric hindrance of longer polysaccharides might shield the interaction of the peptide hydrophobic residues with the chromatographic column stationary phase leading, for some peptides, to a shorter retention time with longer polysaccharides. On the other side, carbohydrates may interact via hydrophobic interaction with the stationary phase with their alkyl portion, depending also on the mobile phase composition (Buttersack, 2017). Since the polysaccharide can be even heavier than the peptide, for shorter, rather hydrophilic peptide moieties these interactions may increase retention with increasing polysaccharide size.

Of note, the chromatographic method allowed the (partial) separation of structural isomers. The branching of the monosaccharide in the RU should be specific depending on the enzyme used for the synthesis, however for the strain O21, in both glycoconjugates EPAg2,3_O21 and AcrAg1,2_O21, we observed that multiple isomers of the same glycovariant are present because (partially) chromatographically separated peaks associated to the same glycopeptide can be observed (Fig. S1). For longer peptides and polysaccharides attached, we observed that the isomers are only partially separated resulting in multiple unresolved chromatographic peaks. From a drug quality control perspective, the capability of the chromatographic method to separate structural isomers may be interesting to monitor the polysaccharide linkage consistency in the different batches of glycoconjugates.

The hyphenation of LC to MS allowed the acquisition of glycopeptide mass spectra. The ionization parameters were carefully tuned to obtain good $m/z$ signals for the different glycovariants. For this, we enriched the ESI dry gas by organic solvent evaporation achieving a so-called dopant enrichment-nitrogen (DEN) gas. The introduction of DEN gas changes the atmospheric condition in the source and can increase ionization efficiency. We observed that, compared to the ionization in air or nitrogen, the use of DEN gas containing ACN positively affected the ionization efficiency and limited the spread of analyte signals over multiple $m/z$ values leading to less overlap of $m/z$ peaks and increasing the signal intensities of glycopeptides in up to 6-fold increase in signal intensity and a 4-fold increase in S/N ratios (Fig. S2).

The obtained glycopeptide mass spectra displayed a series of equidistant $m/z$ signals reflecting the different numbers of RU (Fig. 3). Depending on the length of the polysaccharide, the (glyco) peptides were observed at different charge states that spanned in a range of $z = 1$ for the non-glycosylated peptide up to $z = 8$ for the glycopeptides carrying long polysaccharide chains of between 14 and 21 RUs. Individual glycopeptides were observed at up to four different charge states. From manual inspection of the $MS^1$ data, information on RUs attached to the specific glycosylation sites was retrieved.

For the confirmation of the polysaccharide structures, a nano LC-MS/MS approach was employed with collision-induced dissociation (CID) of the glycopeptides. Using low collisional energy, glycosidic bond cleavages were predominantly observed resulting in B- and Y-ions. As examples, MS/MS spectra of three abundant glycovariants of peptide DFNR of AcrAg1,2_O21 and GSGGGDQNATGSGGGK of EPAg1_O75 and EPAg1,2,3,4_O1 are reported (Fig. 4). The three different glycan structures of the strain O75, O21 and O1A1 RU could be confirmed based on the annotated fragments of the tandem mass spectra. However, a limitation of this MS/MS approach is the impossibility to retrieve glycan linkage information, thus the RU structural corroboration is solely based on monosaccharide compositions. It is worth mentioning that most of the fragments present in the MS/MS spectra were always present for a specific strain and almost independent of the number of RU attached or the peptide sequence (Fig. S3). From a biopharmaceutical quality control perspective, the observation of MS/MS spectra may be useful to confirm the RU structure in terms of monosaccharide compositions. In fact, the presence of the signature diagnostic RU ions on MS/MS spectrum may be used as a quick screening method for RU structure confirmation and for batch to batch comparison.

Besides information on the number of RUs, the glycopeptide MS and MS/MS spectra revealed the presence of additional modifications of the glycan structures or truncation of the polysaccharide. In fact, EPAg1_O75 and EPAg4_O75 displayed the presence of one

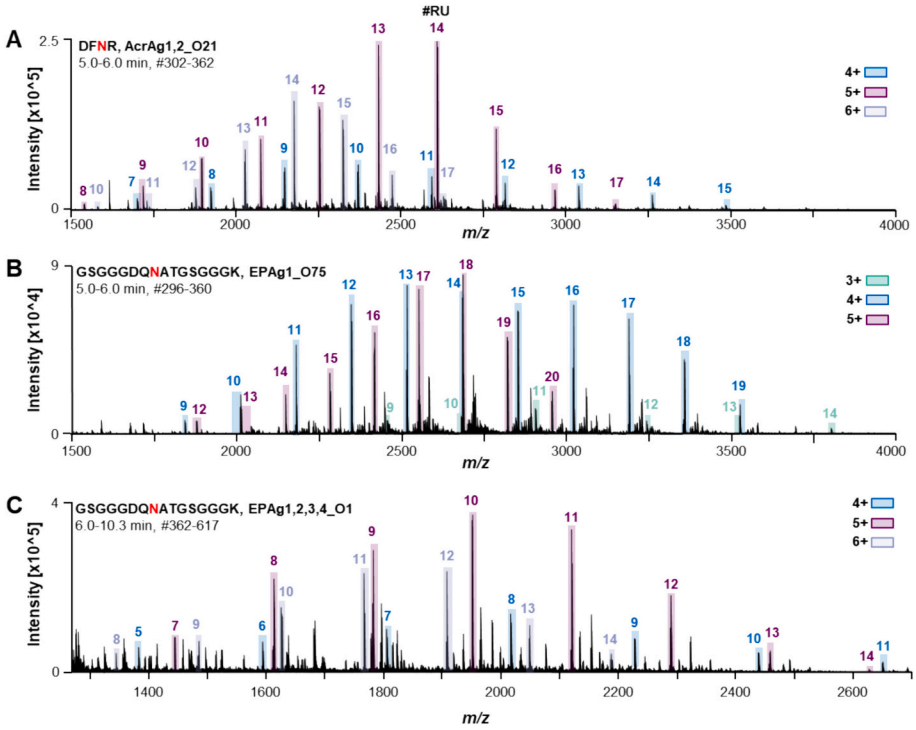


**Fig. 3.** Mass spectra of glycopeptides of (A) O21, (B) O75 and (C) O1A1 strains. In each mass spectrum, the peptide sequence, protein name, retention time window and MS scan range are reported. Charge states of the different glycovariants are indicated by different colors and the numbers above the peaks indicate the number of RUs attached.

phosphoglycerol residue attached to the RU chain (Fig. S4) and EPAg2,3_O21 and AcrAg1,2_O21 showed the truncation of the terminal *N*-acetylhexosamine or the addition of one hexose residue (Fig. S5). The investigation of modifications on the glycans is of crucial importance for bacterial polysaccharide production since the presence of polysaccharide modifications, such as acetylation, may be critical for the immunorecognition of the vaccines (Berti et al., 2018). The MS and MS/MS approach proposed is thus valuable to assess the presence and the chemical composition of these modifications and to prove the consistency of different vaccine batches.

After the manual inspection of MS and MS/MS spectra of the different glycopeptides, information about the number of the RUs attached per site and modification on the RU structure were retrieved and a library of the different glycovariants present was built. One of the advantages of acquiring LC-MS data of glycopeptides is the possibility to perform site-specific semi-quantitation of the different glycovariants based on XICs of the glycopeptide MS ions. In this study, for the first time, we applied this strategy for the analysis of glycoconjugates carrying O-PS. The semi-quantitation was assisted by the use of the freely available software Skyline which allowed the automatic extraction and peak area integration of glycopeptide MS ion XICs based on a list of glycovariants manually generated and integrated in the software. The obtained areas were normalized to 100 %, whereafter the glycovariants were relatively quantified per glycosylation site and plotted in bar charts (Fig. 5). For glycosylation site Asn 243 of EPAg2,3_O21 the glycopeptide with 1 missed cleavage was chosen for the semi-quantitation as signals were more abundant than for the fully cleaved one. For all the other glycopeptides a specific and complete cleavage was observed (Table 1). In this way, a semi-quantitative distribution of RU per glycosylation site was retrieved together with site occupancy. Of note, this quantitation can be slightly biased due to the different ionization efficiency of glycopeptides carrying a different length of polysaccharide chain. The results showed that the RU distributions varied depending on the different serotype O-PS but also on different glycosylation sites of the same glycoconjugate. Generally, glycoforms carrying 1 or 13 and 14 RUs were very prominent in AcrAg1,2_O21 and EPAg2,3_O21 embedding O21 strain RUs while for the glycoconjugates EPAg1_O75 and EPAg4_O75 glycoforms carrying 1 and 16, 17 and 18 RUs were very abundant. All the glycosylation sites were demonstrated to be almost fully occupied with the exception of EPAg4_O75 which showed a site occupancy of 71 % (Fig. 5).

### 3.2. *Intact glycoconjugate characterization by nano LC-MS analysis*

Orthogonal to glycopeptide analysis, intact glycoconjugate proteoforms were analyzed by nano RP-LC-MS employing a diphenyl stationary phase column. Compared to conventional micro LC approaches, a nano LC approach allowed the injection of a minute amount of sample (~100 ng) with an increase of analysis sensitivity and reduction of solvent consumption. Moreover, the hyphenation with mass spectrometry allowed the semi-automated acquisition of glycoconjugate intact mass spectra. The analysis at intact level provided complementary information on the glycosylation state of the glycoconjugates compared to the bottom-up approach as it allowed us to determine the total RU distribution of the intact glycoforms.

Due to the complexity of the glycosylation profile of the glycoconjugates, the MS parameters needed to be carefully tuned to obtain high-quality mass spectra. The proper declustering of the intact proteoforms was pivotal to improve ionization efficiency and proved important in removing TFA adducts that are formed due to the presence of this chemical in the mobile phase. We noticed that the use of an appropriate DEN gas and the optimization of the in-source collisional induced dissociation (isCID) energy were two critical parameters for achieving high-quality mass spectra.

Therefore, different ionization conditions were tested for intact analysis of EPAg4_O75 glycoforms. In particular, the effect of isopropanol and ACN as DEN gas were compared to the ionization in absence of DEN solvent ($N_2$ only) (Fig. S6). ACN was selected as DEN gas because it allowed to reach the best compromise between the removal of TFA adducts and proper declustering of the glycoforms, (almost)

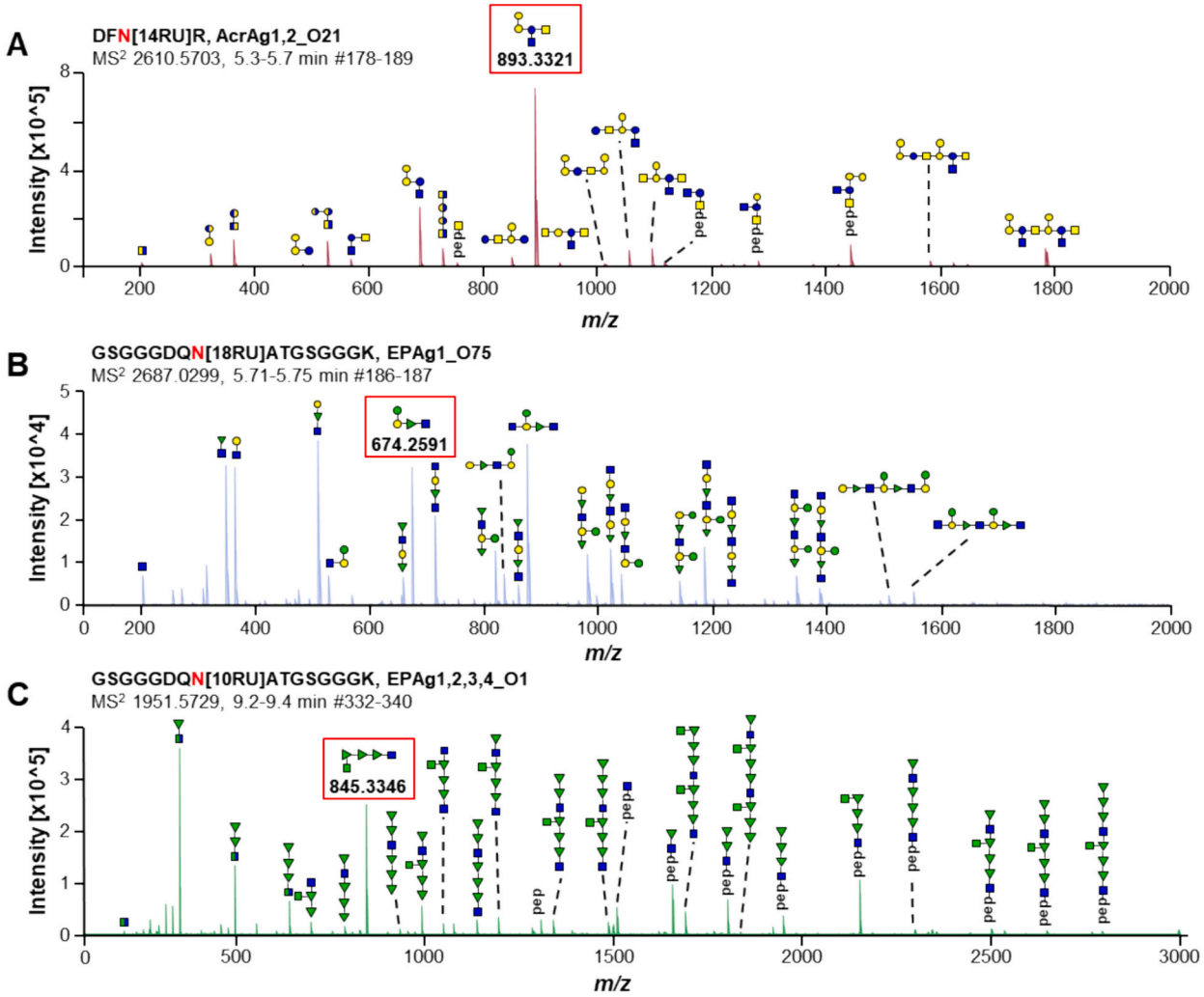


**Fig. 4.** MS/MS spectra of the most abundant glycovariant of the glycopeptide embedding (A) O21, (B) O75 and (C) O1A1 glycans. In each MS/MS spectrum, the peptide sequence, protein name, mass of the isolated ion, retention time window and MS scan range are reported. All the fragments indicated are singly protonated, and the assigned structures are tentative. A red box highlights the protonated single RU mass and its structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

avoiding the fragmentation of the glycan and/or protein chains. This balance was more challenging to find for the diglycosylated EPA2,3_O21 (Fig. S7), specifically for the glycoforms carrying a high number of RUs (from 18 to 28). Already at an isCID of 70 eV glycan chains started to fragment, but the distribution of glycoforms carrying between 18 and 28 RUs was only detected at isCID 100 eV. The latter energy was selected for the ionization of EPAg2,3_O21 but a potential bias of the glycoform distribution due to the fragmentation of the glycan chain should be taken into account. The removal of TFA from the mobile phase would reduce this issue since it is expected that the formation of adducts should occur at a lower extent, however, TFA is often required for proper elution of intact proteins in RP-LC.

In Fig. 6, the chromatogram, the raw and the deconvoluted mass spectrum of EPAg4_O75 intact glycoforms obtained using this RP-LC-MS approach are shown. As expected, glycoforms eluted slightly earlier than their non-glycosylated counterparts (Fig. 6A and B). The deconvoluted mass spectrum of a glycoconjugate (as example EPAg4_O75 reported in Fig. 6C) is characterized by a series of equidistant peaks corresponding to protein with a varying number of RU.

From the deconvoluted mass spectra of the different glycoconjugates, the intensities of different mass peaks were retrieved and semi-quantitation of RU distribution in the intact glycoforms of EPAg1_O75, EPAg4_O75, EPAg2,3_O21, and AcrAg1,2_O21 was carried out (Fig. 7). The results showed two series of glycoform distributions for the diglycosylated AcrAg1,2_O21 and EPAg2,3_O21 spanning from 9 to 19 and 20 to 29 RUs for AcrAg1,2_O21 and 7 to 17 and 18 to 28 RUs for EPAg2,3_O21. For the monoglycosylated EPAg1_O75, a prominent glycoform carrying 1 RU and a distribution of glycoforms carrying from 10 to 20 RUs was observed. EPAg4_O75 showed to be mainly non-glycosylated at 60 % and displayed a distribution of glycoforms with 10 to 20 RUs attached. Noteworthy, EPAg2,3_O21 glycoforms showed a mass shift of ~+71 Da compared to the theoretical masses, and the glycoforms of AcrAg1,2_O21 a mass shift of ~+106 Da. Since the glycan structural integrity and the modifications present were confirmed at glycopeptide level for both glycoconjugates, we attributed these two mass shifts to an amino acid sequence variant. Overall, the intact mass approach provided a fast and straightforward way to compare of average number of RU of manufactured glycoconjugates and permitted the monitoring of RU structural integrity as well as differences in glycosylation levels.

### 3.3. Comparison and integration of glycopeptide and intact data

Finally, the data at the two structural levels of glycopeptide and intact glycoprotein analysis were compared. A prompt comparison can be made between the glycopeptide and the intact glycoform patterns of

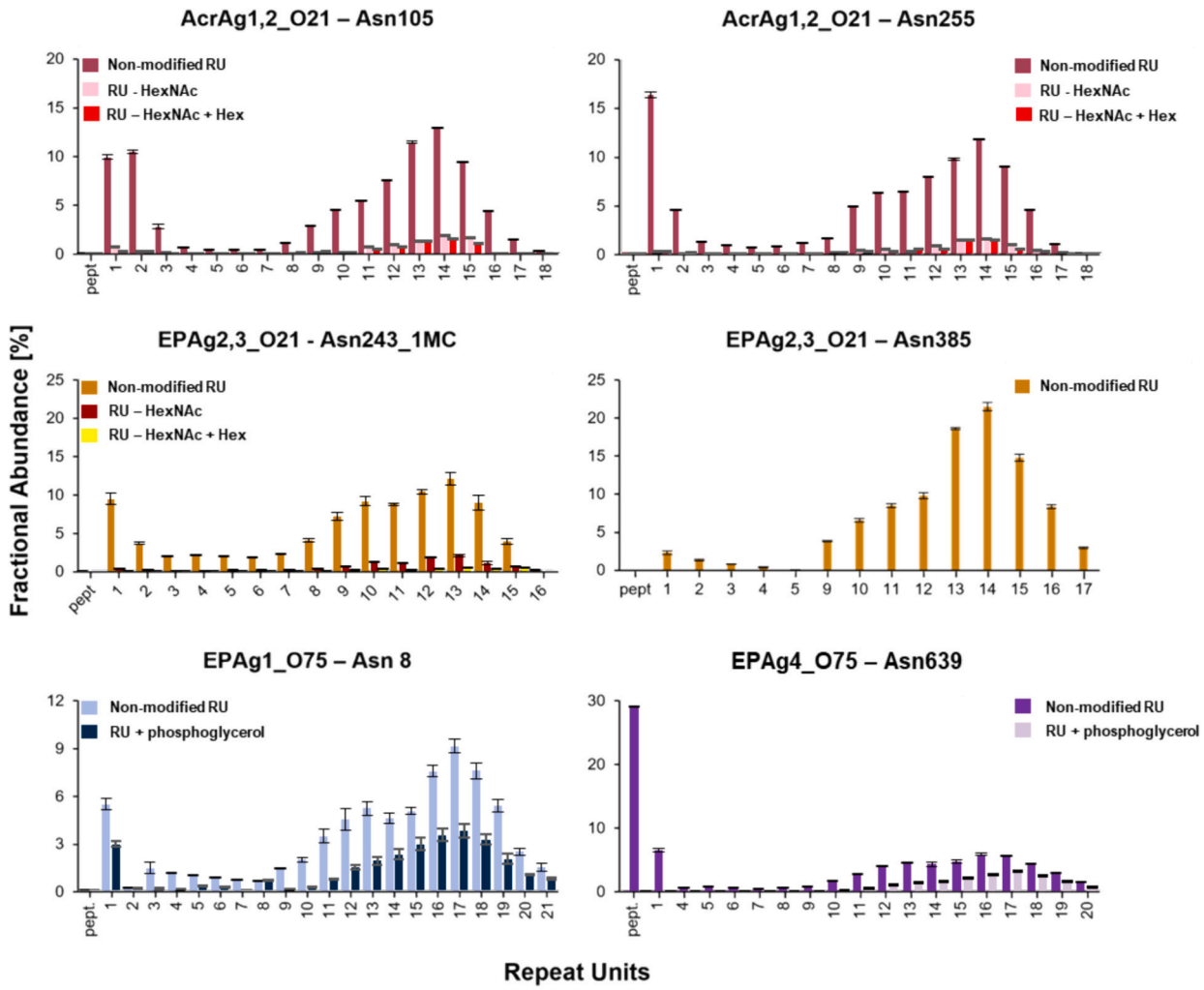


**Fig. 5.** Bar charts reporting the site-specific semi-quantitation of the glycopeptide RU distribution and their related modifications of the different samples. The bars indicate the average of three replicates and the error bars the standard deviation in percentage. 1 MC stands for a peptide with one missed cleavage. Numerical values are reported in supplementary information 2. The quantitation is relative and based on the areas of glycoform chromatographic peaks in the XICs.

the monoglycosylated conjugates. In fact, only one glycosylation site is involved and no preferential combination of RUs attached to the different glycosylation sites may influence the intact glycoform assembly. Thus, for the monoglycosylated EPAg1_O75 and EPAg4_O75, the glycopeptide (Fig. 5) and intact glycoforms (Fig. 7) RU distribution were expected to match. However, some differences were observed. First, at the glycopeptide level, low abundant glycovariants carrying from 2 to 9 RU were detected while this was not the case for the intact analysis. As a matter of fact, it is well known that bottom-up approaches are generally more sensitive than intact glycoform analysis. Moreover, the glycoform abundances did not match between the two structural levels. Only the distribution of the highly glycosylated variants (from 11 to 20 RU) correlated well for glycopeptide and intact data (Fig. S8). On the contrary, non-glycosylated proteoforms or proteoforms carrying 1 RU were more abundant at intact level than at glycopeptide level. This can be due to the fact that, at intact level, the declustering of these species in the ionization process is more efficient compared to the highly glycosylated species leading to a better ionization, thus resulting in a bias towards low glycosylated variants. On the other side, glycopeptide ionization is mainly dependent on the glycan chains attached since the mass of this portion can be even heavier than the peptide itself. Thus, at peptide level, it is expected that the ionization efficiency (even using DEN gas) would differ between highly glycosylated peptides versus peptides with low glycosylation levels. These bias issues could be overcome in the future using isotope-labeled internal standards for both glycopeptide and intact glycoforms to gather numerical factors that could potentially correct for this ionization bias.

Lastly, the data at the two structural levels were integrated for the diglycosylated AcrAg1,2_O21 and EPAg2,3_O21 and intact mass spectra were annotated. For this purpose, the software tool MoFi was employed as previously reported (Lebede et al., 2021; Skala et al., 2018; Wohlschlager et al., 2018). In this way, information on how the glycans attached to the two different glycosylation sites combine in the intact glycoforms could be retrieved. The annotations were performed only on the mass peaks corresponding to glycoforms not carrying modification on the RU to avoid the explosion of the combinatorial space in MoFi. The results revealed 263 co-existing glycoforms for AcrAg1,2_O21 (see Annotations_AcrAg1,2_O21.csv) and 213 glycoforms for EPAg2,3_O21 (see Annotations_EPAg2,3_O21.csv), excluding the proteoforms carrying the modified-glycan RU. The glycoforms were hierarchically listed based on their contribution to the peak intensity of a specific mass. This was calculated by MoFi based on the fractional abundances of the different glycopeptide variants and displayed with a parameter called permutation score. For instance, the most intense peak (mass 81,496.95 Da) in the deconvoluted mass spectrum of EPAg2,3_O21, corresponding to 14 RU composition, was annotated with 12 different isobaric glycoforms. These comprised different numbers of RUs attached in the two glycosylation sites with the most abundant variant carrying 1 RU at Asn 243

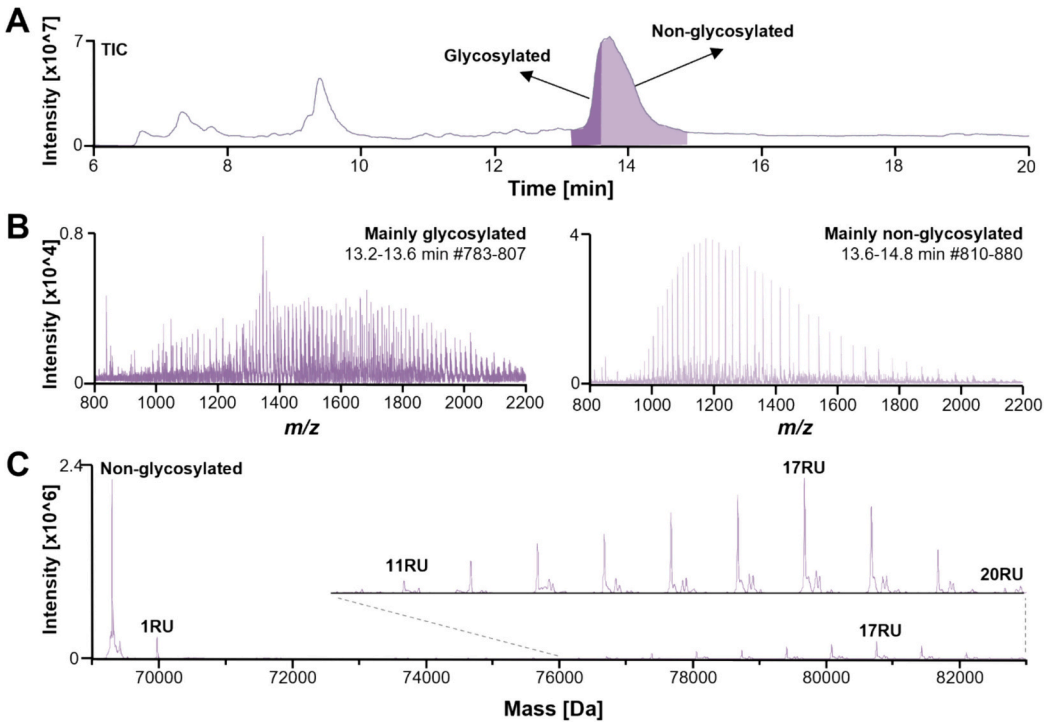


**Fig. 6.** Results of nano RP-LC-MS analysis of EPAg4_O75 intact glycoforms. A) Total ion current chromatogram (TIC) of EPAg4_O75 intact glycoforms. In shades of violet the region of the chromatographic peak where glycosylated and non-glycosylated conjugate proteoforms eluted are indicated. B) Raw mass spectra associated to the colored chromatographic peak in the TIC. Retention time window and number of MS scans are reported. In the raw mass spectra showed, (partially) overlapped *m/z* peaks related to the differently charged glycoform ions can be observed. C) Deconvoluted mass spectrum of EPAg4_O75 intact glycoforms. The number of RU attached to the proteoforms are reported. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and 13 RU at Asn385 contributing to 58.1 % (permutation score) to the peak intensity (see Annotations_EPAg2,3_O21.csv). The integration of the glycopeptide data into MoFi to annotate the deconvoluted mass spectrum of the intact glycoconjugates proved to be key in unravelling the glycoform complexity of the vaccine. The annotation results showed that numerous isobaric glycoforms contributed to a single peak in the deconvoluted mass spectrum. In fact, ~20 mass peaks in the deconvoluted mass spectra of AcrAg1,2_O21 and EPAg2,3_O21 resulted in hundreds of glycoforms annotated. Thus, the data integration of the two structural levels added new insights on the glycosylation state of the intact glycoforms, allowing the exploration of the intact glycoforms profile obtained by intact mass analysis. This proposed data integration can be even more insightful as the number of glycosylation sites of the carrier protein increases. In fact, more glycosylation sites bring more possible combinations of the different RU distributions related to the different glycosylation sites in the intact glycoconjugate. Thus, the exploration of the chemical space of the different isobaric co-existing glycoforms can be performed using the proposed bioinformatic data integration workflow.

## 4. Conclusion

In this study, we demonstrated the feasibility of performing in-depth characterization of the RU distribution of glycoconjugates carrying polydisperse *E. coli* O-antigen polysaccharides by RP-LC-MS (MS). To our knowledge, this is the first report in literature that describes such an approach where the RU distribution and RU modifications of glycoconjugate O-PS at the different glycosylation sites were obtained by bottom-up MS analysis and the total RU distribution was determined by intact glycoconjugate mass analysis. The use of dopant enrichment nitrogen gas for the ESI ionization proved to be key for increasing (glyco) peptides detection while for intact mass analysis it provided a suitable environment for removal of TFA adducts. Further improvement of such a characterization strategy would require the implementation of quantitative analyses (e.g., using isotopically labeled analytes) to assess the ionization bias between glycosylated and non-glycosylated variants thus, further research is warranted. This workflow provides insights into critical factors driving the glycosylation heterogeneity of lead molecules. It can thus be integrated into current biopharmaceutical characterization strategies to facilitate the development of complex glycoconjugate bacterial vaccines. In fact, the intact mass analysis can provide a fast profiling of glycoconjugate glycosylation pattern for vaccine batch comparison while glycopeptide analysis can be applied for the corroboration of the O-PS structures and their modifications and the site-specific quantitative comparison of RU distribution.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.carbpol.2024.122327.

## Funding

This work was supported by Janssen Vaccines and Prevention B.V.

## CRediT authorship contribution statement

**Fiammetta Di Marco:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Agnes L. Hipgrave Ederveen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Guusje van Schaick:** Writing – review & editing, Supervision, Project administration, Data curation. **Alan B. Moran:** Writing – review & editing, Project administration. **Elena Domínguez-Vega:** Writing – review & editing, Supervision. **Simone Nicolardi:** Writing – review & editing, Supervision, Conceptualization. **Constantin Blöchl:** Writing – review & editing, Methodology. **Carolien A. Koeleman:** Methodology, Formal analysis. **Renzo Danuser:** Writing – review & editing. **Ali Al Kaabi:** Writing –

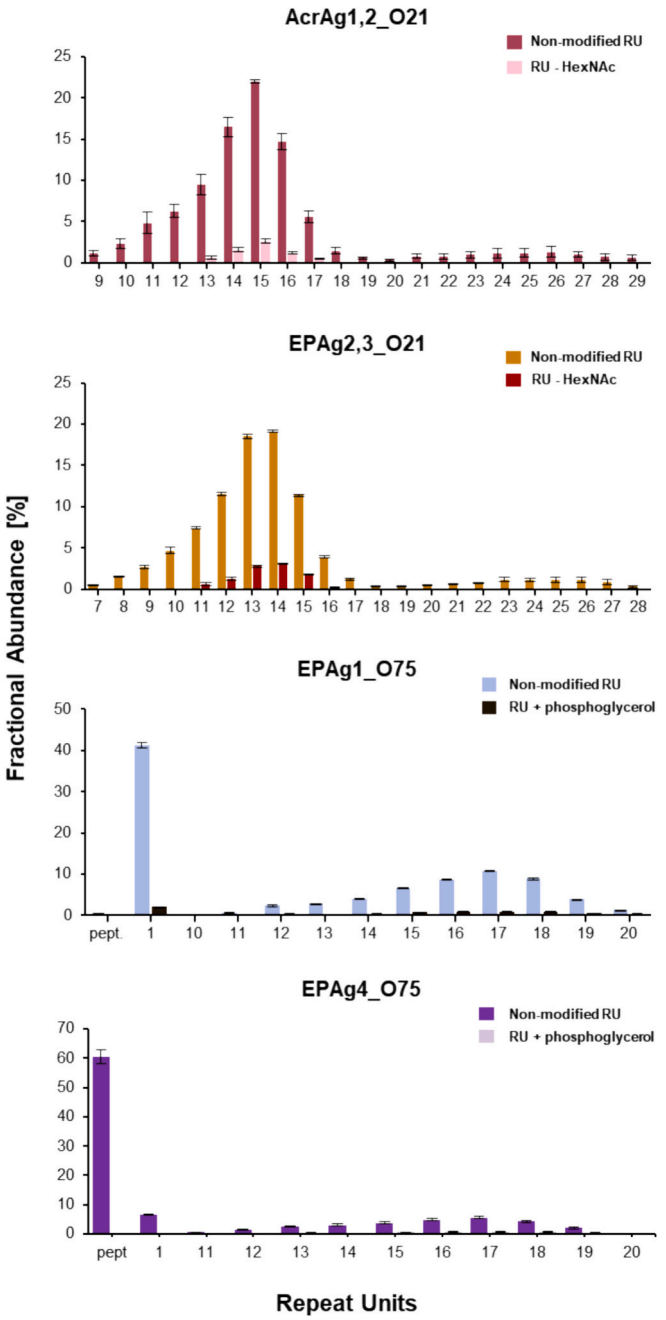


**Fig. 7.** Bar charts reporting the semi-quantitation of intact glycoform RU distribution and their related modifications. Bars indicate the average of three replicates and error bars the standard deviation in percentage. Numerical values are reported in supplementary information 2. The quantitation is relative and based on intensity of deconvoluted mass spectra.

review & editing. **Viktoria Dotz:** Writing – review & editing. **Jan Grijpstra:** Writing – review & editing. **Michel Beurret:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Chakkumkal Anish:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Manfred Wuhrer:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: A.B.M., R.D., V.D., A.A.K., J.G., M.B., and C.A. are employees of Janssen Vaccines. All other authors declare no conflict of interest.

## Data availability

The data that has been used is confidential.

## Acknowledgments

We thank Dr. Ieva Palubeckaitė for her valuable comments on the manuscript.

## References

Alagesan, K., & Kolarich, D. (2019). To enrich or not to enrich: Enhancing (glyco) peptide ionization using the Captive Spray nanoBoosterTM. *BioRxiv*. https://doi.org/10.1101/597922

Anish, C., Beurret, M., & Poolman, J. (2021). Combined effects of glycan chain length and linkage type on the immunogenicity of glycoconjugate vaccines. *npj Vaccines, 150*, 1–13. https://doi.org/10.1038/s41541-021-00409-1

Aslam, B., Wang, W., Arshad, M. I., Khurshid, M., Muzammil, S., Rasool, M. H., … Baloch, Z. (2018). Antibiotic resistance: A rundown of a global crisis. *Infection and Drug Resistance, 11*, 1645–1658. https://doi.org/10.2147/IDR.S173867

Baumann, H., Jansson, P. E., Kenne, L., & Widmalm, G. (1991). Structural studies of the Escherichia coli O1A O-polysaccharide, using the computer program CASPER. *Carbohydrate Research, 211*(1), 183–190. https://doi.org/10.1016/0008-6215(91)84159-C

Beck, A., Wagner-Rousset, E., Ayoub, D., Van Dorsselaer, A., & Sanglier-Cianférani, S. (2013). Characterization of therapeutic antibodies and related products. *Analytical Chemistry, 85*(2), 715–736. https://doi.org/10.1021/AC3032355/ASSET/IMAGES/LARGE/AC-2012-032355_0010.JPEG

Berti, F., De Ricco, R., Rappuoli, R., Kovac, P., Xu, P., & Pfister, H. (2018). Role of O-acetylation in the immunogenicity of bacterial polysaccharide vaccines. *Molecules, 23*, 1340–1350. https://doi.org/10.3390/molecules23061340

Buttersack, C. (2017). *Hydrophobicity of carbohydrates and related hydroxy compounds*. https://doi.org/10.1016/j.carres.2017.04.019

Cuccui, J., & Wren, B. (2015). Hijacking bacterial glycosylation for the production of glycoconjugates, from vaccines to humanised glycoproteins. *The Journal of Pharmacy and Pharmacology, 67*(3), 338–350. https://doi.org/10.1111/JPHP.12321

Erbing, C., Kenne, L., Lindberg, B., & Hammarstr6m, S. (1978). Structure of the O-specific side-chains of the Escherichia coil 0 75 lipopolysaccharide: A revision. *Carbohydrate Research, 60*, 400–403.

Feldman, M. F., Wacker, M., Hernandez, M., Hitchen, P. G., Marolda, C. L., Kowarik, M., … Aebi, M. (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*(8), 3016–3021. https://doi.org/10.1073/PNAS.0500044102/ASSET/9F255D80-EB19-4FC8-A679-5EA86EA31C5D/ASSETS/GRAPHIC/ZPQ0070573890005.JPEG

Goldman, R. C., & Leive, L. (1980). Heterogeneity of antigenic-side-chain length in lipopolysaccharide from Escherichia coli 0111 and Salmonella typhimurium LT2. *European Journal of Biochemistry, 107*(1), 145–153. https://doi.org/10.1111/J.1432-1033.1980.TB04635.X

Gupta, D. S., Shashkov, A. S., Jann, B., & Jann, K. (1992). Structures of the O1B and O1C lipopolysaccharide antigens of Escherichia coli. *Journal of Bacteriology, 174*(24), 7963–7970. https://doi.org/10.1128/JB.174.24.7963-7970.1992

Harding, C. M., Nasr, M. A., Scott, N. E., Goyette-Desjardins, G., Nothaft, H., Mayer, A. E., … Feldman, M. F. (2019). A platform for glycoengineering a polyvalent pneumococcal bioconjugate vaccine using E. coli as a host. *Nature Communications, 10*, 891–902. https://doi.org/10.1038/s41467-019-08869-9

Huttner, A., & Gambillara, V. (2018). The development and early clinical testing of the ExPEC4V conjugate vaccine against uropathogenic Escherichia coli. *Clinical Microbiology and Infection, 24*, 1046. https://doi.org/10.1016/j.cmi.2018.05.009

Imperiali, B. (2019). Bacterial carbohydrate diversity-a Brave New World. *Current Opinion in Chemical Biology, 53*, 1–8. https://doi.org/10.1016/j.cbpa.2019.04.026

Iwashkiw, J. A., Fentabil, M. A., Faridmoayer, A., Mills, D. C., Peppler, M., Czibener, C., … Feldman, M. F. (2012). Exploiting the Campylobacter jejuni protein glycosylation system for glycoengineering vaccines and diagnostic tools directed against brucellosis. *Microbial Cell Factories, 11*(13), 1–11. https://doi.org/10.1186/1475-2859-11-13

Jann, B., Shashkov, A. S., Gupta, D. S., Panasenko, S. M., & Jann, K. (1992). The O1 antigen of Escherichia coli: Structural characterization of the O1A1-specific polysaccharide. *Carbohydrate Polymers, 18*(1), 51–57. https://doi.org/10.1016/0144-8617(92)90187-U

Jansson, P. E., Lennholm, H., Lindberg, B., Lindquist, U., & Svenson, S. B. (1987). Structural studies of the O-specific side-chains of the Escherichia coli O2 lipopolysaccharide. *Carbohydrate Research, 161*(2), 273–279. https://doi.org/10.1016/S0008-6215(00)90084-3

Jiang, X., Bai, J., Yuan, J., Zhang, H., Lu, G., Wang, Y., Jiang, L., Liu, B., Wang, L., Huang, D., & Feng, L. (2021). High efficiency biosynthesis of O-polysaccharide-based vaccines against extraintestinal pathogenic Escherichia coli. *Carbohydrate Polymers, 255*, Article 117475. https://doi.org/10.1016/j.carbpol.2020.117475

Jiang, X., Bai, J., Zhang, H., Yuan, J., Lu, G., Wang, Y., Jiang, L., Liu, B., Huang, D., & Feng, L. (2022). Development of an O-polysaccharide based recombinant glycoconjugate vaccine in engineered E. coli against ExPEC O1. *Carbohydrate Polymers, 277*. https://doi.org/10.1016/J.CARBPOL.2021.118796

Kammeijer, G. S. M., Kohler, I., Jansen, B. C., Hensbergen, P. J., Mayboroda, O. A., Falck, D., & Wuhrer, M. (2016). Dopant enriched nitrogen gas combined with sheathless capillary electrophoresis-electrospray ionization-mass spectrometry for improved sensitivity and repeatability in glycopeptide analysis. *Analytical Chemistry, 88*(11), 5849–5856. https://doi.org/10.1021/ACS.ANALCHEM.6B00479/ASSET/IMAGES/LARGE/AC-2016-00479A_0006.JPEG

Kay, E., Cuccui, J., & Wren, B. W. (2019). Recent advances in the production of recombinant glycoconjugate vaccines. *Vaccines, 4*(1), 1–8. https://doi.org/10.1038/s41541-019-0110-z

Kim, J. S., Laskowich, E. R., Arumugham, R. G., Kaiser, R. E., & Macmichael, G. J. (2005). Determination of saccharide content in pneumococcal polysaccharides and conjugate vaccines by GC-MSD. *Analytical Biochemistry, 347*, 262–274. https://doi.org/10.1016/j.ab.2005.09.022

Kourtis, A. P., Sheriff, E. A., Weiner-Lastinger, L. M., Elmore, K., Ellyn Preston, L., Dudeck, M., & Clifford McDonald, L. (2021). Clinical infectious diseases antibiotic multidrug resistance of Escherichia coli causing device-and procedure-related infections in the United States reported to the National Healthcare Safety Network, 2013–2017. *Clinical Infectious Diseases, 73*(11), 4552–4559. https://doi.org/10.1093/cid/ciaa1031

Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., … Aebi, M. (2006). Definition of the bacterial N-glycosylation site consensus sequence. *The EMBO Journal, 25*(9), 1957–1966. https://doi.org/10.1038/SJ.EMBOJ.7601087

Lebede, M., Di Marco, F., Esser-Skala, W., Hennig, R., Wohlschlager, T., & Huber, C. G. (2021). Exploring the chemical space of protein glycosylation in noncovalent protein complexes: An expedition along different structural levels of human chorionic gonadotropin by employing mass spectrometry. *Analytical Chemistry, 93*(30), 10424–10434. https://doi.org/10.1021/acs.analchem.1c02199

Lee, S., Inzerillo, S., Lee, G. Y., Bosire, E. M., Mahato, S. K., & Song, J. (2022). Glycan-mediated molecular interactions in bacterial pathogenesis. *Trends in Microbiology, 30*(3), 254–267. https://doi.org/10.1016/j.tim.2021.06.011

Li, Y., Perepelov, A. V., Guo, D., Shevelev, S. D., Senchenkova, y. N., Shahskov, A. S., … Zelinsky, N. (2011). Structural and genetic relationships of two pairs of closely related O-antigens of Escherichia coli and Salmonella enterica: E. coli O11/S. enterica O16 and E. coli O21/S. enterica O38. *FEMS Immunol Med Microbio, 61*, 258–268. https://doi.org/10.1111/j.1574-695X.2010.00771.x

Liu, B., Furevi, A., Perepelov, A. V., Guo, X., Cao, H., Wang, Q., … Widmalm, G. (2020). Structure and genetics of Escherichia coli O antigens. *FEMS Microbiology Reviews, 44*(6), 655–683. https://doi.org/10.1093/FEMSRE/FUZ028

Madunić, K., Wagt, M. S., Zhang, T., Wuhrer, M., & Lageveen-Kammeijer, G. S. M. (2021). Dopant-enriched nitrogen gas for enhanced electrospray ionization of released glycans in negative ion mode. *Analytical Chemistry, 93*, 6919–6923. https://doi.org/10.1021/acs.analchem.1c00023

Micoli, F., Adamo, R., & Costantino, P. (2018). Protein carriers for glycoconjugate vaccines: History, selection criteria, characterization and new trends. *Molecules, 23*(6). https://doi.org/10.3390/MOLECULES23061451

Micoli, F., Del Bino, L., Alfini, R., Carboni, F., Romano, M. R., & Adamo, R. (2019). Glycoconjugate vaccines: Current approaches towards faster vaccine design. *Expert Review of Vaccines, 18*(9), 881–895. https://doi.org/10.1080/14760584.2019.1657012

Mysling, S., Palmisano, G., Hojrup, P., & Thaysen-Andersen, M. (2010). Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Analytical Chemistry, 82*(13), 5598–5609. https://doi.org/10.1021/AC100530W/SUPPL_FILE/AC100530W_SI_001.PDF

Nicolardi, S., Danuser, R., Dotz, V., Domínguez-Vega, E., Al Kaabi, A., Beurret, M., … Wuhrer, M. (2022). Glycan and protein analysis of glycoengineered bacterial E. coli vaccines by MALDI-in-source decay FT-ICR mass spectrometry. *Analytical Chemistry, 94*, 4979–4987. https://doi.org/10.1021/acs.analchem.1c04690

Poolman, J. T., & Wacker, M. (2016). Extraintestinal pathogenic Escherichia coli, a common human pathogen: Challenges for vaccine development and progress in the field. *The Journal of Infectious Diseases, 213*(1), 6–13. https://doi.org/10.1093/INFDIS/JIV429

Rappuoli, R. (2018). Glycoconjugate vaccines: Principles and mechanisms. *Science Translational Medicine, 10*(456), 4615. https://doi.org/10.1126/SCITRANSLMED.AAT4615/ASSET/9B2AAFA9-E687-43E5-A8B5-A029ED1ED9BB/ASSETS/GRAPHIC/AAT4615-F4.JPEG

Ravenscroft, N., Braun, M., Schneider, J., Dreyer, A. M., Wetter, M., Haeuptle, M. A., … Kowarik, M. (2019). Characterization and immunogenicity of a Shigella flexneri 2a O-antigen bioconjugate vaccine candidate. *Glycobiology, 29*(9), 669–680. https://doi.org/10.1093/GLYCOB/CWZ044

Ravenscroft, N., Haeuptle, M. A., Kowarik, M., Fernandez, F. S., Carranza, P., Brunner, A., … Wacker, M. (2016). Purification and characterization of a Shigella conjugate vaccine, produced by glycoengineering Escherichia coli. *Glycobiology, 26*(1), 51–62. https://doi.org/10.1093/GLYCOB/CWV077

Sandra, K., Vandenheede, I., & Sandra, P. (2014). Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization. *Journal of Chromatography A, 1335*, 81–103. https://doi.org/10.1016/J.CHROMA.2013.11.057

Scott, N. E., Kinsella, R. L., G Edwards, A. V., Larsen, M. R., Dutta, S., Saba, J., … Feldman, M. F. (2014). Diversity within the O-linked protein glycosylation systems of acinetobacter species. *Molecular & Cellular Proteomics, 13*, 2354–2370. https://doi.org/10.1074/mcp.M114.038315

Skala, W., Wohlschlager, T., Senn, S., Huber, G. E., & Huber, C. G. (2018). MoFi: A software tool for annotating glycoprotein mass spectra by integrating hybrid data from the intact protein and glycopeptide level. *Analytical Chemistry, 90*, 5728–5736. https://doi.org/10.1021/acs.analchem.8b00019

Staaf, M., Urbina, F., Weintraub, A., & Widmalm, G. (1999). Structural elucidation of the O-antigenic polysaccharides from Escherichia coli O21 and the enteroaggregative Escherichia coli strain 105. *European Journal of Biochemistry, 266*(1), 241–245. https://doi.org/10.1046/J.1432-1327.1999.00850.X

Stefanetti, G., Okan, N., Fink, A., Gardner, E., & Kasper, D. L. (2019). Glycoconjugate vaccine using a genetically modified O antigen induces protective antibodies to Francisella tularensis. *PNAS, 116*(14), 7062–7070. https://doi.org/10.1073/pnas.1900144116

Stenutz, R., Weintraub, A., Oran Widmalm, G., &̈. (2006). The structures of Escherichia coli O-polysaccharide antigens. *FEMS Microbiology Reviews, 30*, 382–403. https://doi.org/10.1111/j.1574-6976.2006.00016.x

Struwe, W. B., & Robinson, C. V. (2019). Relating glycoprotein structural heterogeneity to function – Insights from native mass spectrometry. *Current Opinion in Structural Biology, 58*, 241–248. https://doi.org/10.1016/J.SBI.2019.05.019

Terra, V. S., Mauri, M., Sannasiddappa, T. H., Smith, A. A., Stevens, M. P., Grant, A. J., … Cuccui, J. (2022). PglB function and glycosylation efficiency is temperature dependent when the pgl locus is integrated in the Escherichia coli chromosome. *Microbial Cell Factories, 21*(1). https://doi.org/10.1186/S12934-021-01728-7

van den Dobbelsteen, G. P. J. M., Faé, K. C., Serroyen, J., van den Nieuwenhof, I. M., Braun, M., Haeuptle, M. A., … Poolman, J. T. (2016). Immunogenicity and safety of a tetravalent E. coli O-antigen bioconjugate vaccine in animal models. *Vaccine, 34*(35), 4152–4160. https://doi.org/10.1016/J.VACCINE.2016.06.067

Van Schaick, G., Wuhrer, M., & Domínguez-Vega, E. (2023). Dopant-enriched nitrogen gas to boost ionization of glycoproteins analyzed with native liquid chromatography coupled to nano-electrospray ionization. *Analytica Chimica Acta, 1265*. https://doi.org/10.1016/j.aca.2023.341271

Wohlschlager, T., Scheffler, K., Forstenlehner, I. C., Skala, W., Senn, S., Damoc, E., … Huber, C. G. (2018). Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals. *Nature Communications, 9*, 1713–1722. https://doi.org/10.1038/s41467-018-04061-7

Yang, Y., Franc, V., & Heck, A. J. R. (2017). Glycoproteomics: A balance between high-throughput and in-depth analysis. *Trends in Biotechnology, 35*(7), 598–609. https://doi.org/10.1016/j.tibtech.2017.04.010

Yang, Y., Wang, G., Song, T., Lebrilla, C. B., & Heck, A. J. R. (2017). Resolving the micro-heterogeneity and structural integrity of monoclonal antibodies by hybrid mass spectrometric approaches. *MAbs, 9*(4), 638. https://doi.org/10.1080/19420862.2017.1290033

Yu, A., Zhao, J., Peng, W., Banazadeh, A., Williamson, S. D., Goli, M., … Mechref, Y (2018). Advances in mass spectrometry-based glycoproteomics. *Electrophoresis, 39*, 3104–3122. https://doi.org/10.1002/elps.201800272