PhD in co-tutoring between the Department of Chemical Sciences of University of Naples "Federico II" and GSK

Ph.D. School in Chemical Sciences - Cycle XXXVII

Development and analysis of glycoprotein based antiviral vaccines using Mass Spectrometry

Antonio Lembo

Supervisors:

Prof. Cristina De Castro

Prof. Antonio Molinaro

Dr. Massimiliano Biagini

Examiner:

Prof. Angela Amoresano

2021-2024

Coordinator: Prof. Angelina Lombardi

Abstract

Glycosylation is the most prominent post-translational modification occurring on protein antigens and its specific pattern depends both on the cells used for their expression (*e.g.* mammalian cells, insect cells, yeast) and the cells growth conditions. Oligosaccharide moieties on protein antigen surface play an important role in the modulation of its immunological properties; indeed, glycans can affect these features in different ways, like protein epitope shielding, formation of oligosaccharide epitopes and presence of non-self monosaccharides. Mass spectrometry (MS) represents a powerful tool for the in-depth analysis of proteins glycosylation pattern; indeed, it allows to describe the microheterogeneity of each glycosylation site in a reproducible manner, unlike X-ray crystallography, which fails to describe highly flexible structures like glycans.

Two recombinant glycoprotein antigens from Human Cytomegalovirus (HCMV) envelope have been considered as case study for the development of MS-based methods for glycans analysis: glycoprotein B (gB) and Pentamer (gH/gL/UL128/UL130/UL131); these two antigens are the most promising candidates for the development of a vaccine against HCMV. Furthermore, different expression conditions have been considered in order to compare antigens with different glycan patterns.

In this thesis the glycoprotein antigens have been characterized by using different hyphenated MS techniques (LC-MS/MS, LC-FLR-MS, LC-UV-MS, GC-MS), depending on the class of analytes originating from the antigen: glycopeptides, glycans and monosaccharides. The analysis of the first category of molecules relies on a bottom-up proteomic approach in which the *N*- and *O*-glycosites heterogeneity and occupancy have been investigated by using tandem mass-spectrometry; this study allowed to see that both gB and Pentamer *N*-glycosylation sequons possess glycans with different maturation degrees, depending on the accessibility of the processing enzymes to each site. In addition, it has been proved that, in cell culture conditions

that arrest glycans maturation to oligomannose structures, site-specific maturation is almost identical for each site.

Released *N*-glycans analysis has been performed by enzymatic glycans shaving followed by liquid chromatography with fluorescence detection and mass-spectrometry characterization; this analysis allowed to confirm the identities of the glycans found by site-specific analysis, also providing further information about the presence of highly processed tetrantennary complex glycans and Neu5Gc-containing oligosaccharides.

Antigens absolute monosaccharide composition has been assessed by gas chromatography-mass spectrometry, degrading the oligosaccharides in acetylated methyl glycosides and using the internal standard method for quantification; furthermore, a comparison with monosaccharide analysis by LC-UV-MS has been done, which showed to be a highly sensitive method.

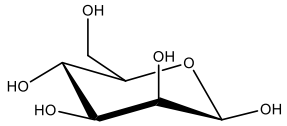
Immunogenicity of gB and Pentamer with different glycosylation pattern have been measured by neutralizing antibodies titer elicited in mice after vaccination; both gB and Pentamer showed higher immunogenicity when bringing only high-mannose structures.

To assess the shielding that glycans play on gB and Pentamer epitopes, tridimensional models of the oligosaccharides on these proteins have been built; therefore, it has been seen that antigens bringing only M5 glycans have more exposed epitopes than once bringing M9 glycans.

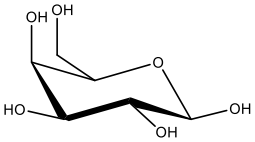
To conclude, glycoprotein oligosaccharide pattern can be broadly explored by MS techniques, providing a lot of information about the nature and the stoichiometry of the glycans on the antigens surface. The combination of the MS data with a glycoprotein modeling software is a strong tool to predict how glycans can affect the recognition of antigen epitopes.

Table of Contents

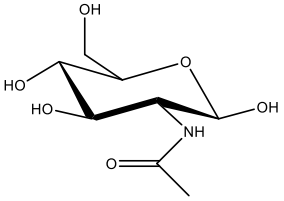
Monosaccharides structures and Symbol Nomenclature For Glycans (SNFG)



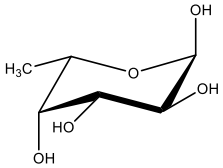
β-D-mannose



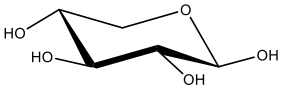
β-D-galactose



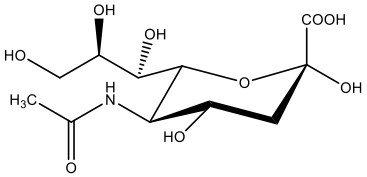
N-acetyl-β-D-glucosamine



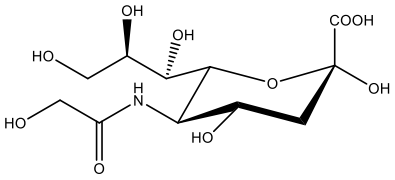
α-L-fucose



β-D-xylose



N-acetyl-α-D-neuraminic acid



N-glycolyl-α-D-neuraminic acid

SECTION I:

INTRODUCTION

Chapter 1 - Impact of glycosylation on viral vaccines

1.1 Protein glycosylation: an overview

Glycans are attached to the proteins and processed by a series of enzymes without the rigid direction of nucleic acid templates. Consequently, a single protein that is glycosylated normally emerges from the biosynthetic pathway as a mixture of glycosylation variants, known as glycoforms. Under the same experimental conditions, each glycoprotein has a reproducible and characteristic glycosylation profile because the glycan processing is directed by protein tridimensional structure (Rudd et al., 1997).

There are 3 main classes of glycosidic linkages to protein: *N*-glycosylation, *O*-glycosylation and glypiation (glycosylphosphatidylinositol mediated linkages); there also rare form of glycosylation like C-glycosylation (which involves a Trp residue), found on Ebola virus GP protein, and S-glycosylation, found on bacteriocin GccF from *L. plantarum* (Bloch et al., 2023; Falzarano et al., 2007; Stepper et al., 2011). Non-enzymatic covalent attachment of monosaccharides (called glycation) it is also known: one the most important case is the pathway by hemoglobin exposure to plasma glucose, which is irreversibly glycated at one or both N-terminal valines of the beta chains (Chauhan, 2017). However, nowadays glycation describes an unusually diverse set of PTMs, collectively known as advanced glycation end-products (AGEs), which are associated with aging and disease and are often regarded as random damage that occurs over time or in response to metabolic dysregulation (Martin et al., 2024).

N-glycosylation is a co-translational modification which starts in the endoplasmic reticulum (ER) while the protein is being synthetized and is still unfolded; this modification concerns about 70 % of the asparagines contained in the sequon Asn-X-Ser/Thr (with X different from Pro), which become functionalized with the precursor oligosaccharide Glc3Man9GlcNAc2 (G3M9) (Fig. 1.1).

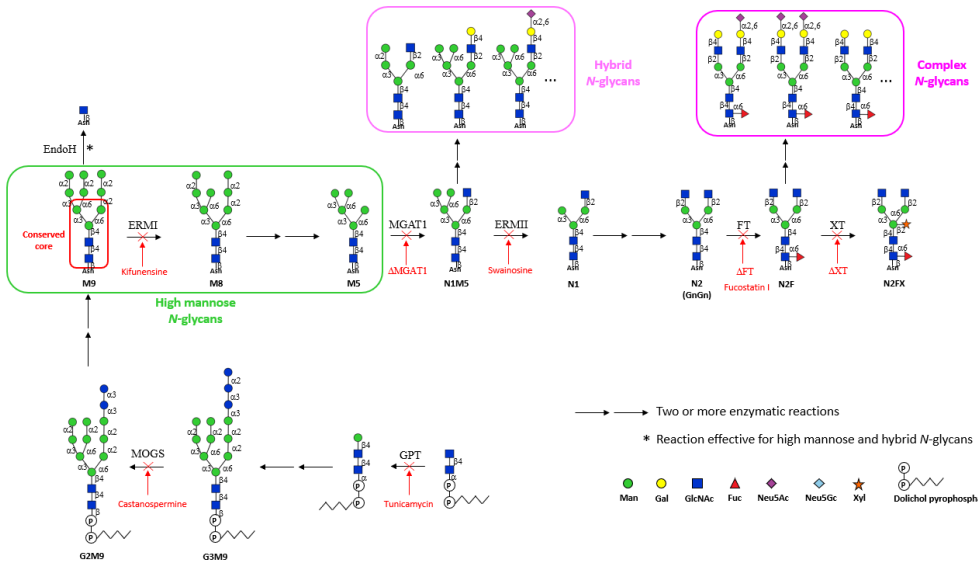


Figure 1.1. Enzymatic steps of *N*-glycan maturation affected by inhibitors or mutations. From *Impact of glycosylation on viral vaccines.* (Lembo et al., 2024).

On the contrary, *O*-glycosylation happens in the Golgi apparatus when the protein is fully formed and folded; it begins with the post-translational addition of a single monosaccharide on appropriate Ser or Thr side chains. In general, no consensus sequence has been clearly detected for *O*-glycosylation, however it has been noted that regions which contain high proportions of Ser, Thr and Pro residues are frequently *O*-glycosylated (Fig. 1.2).

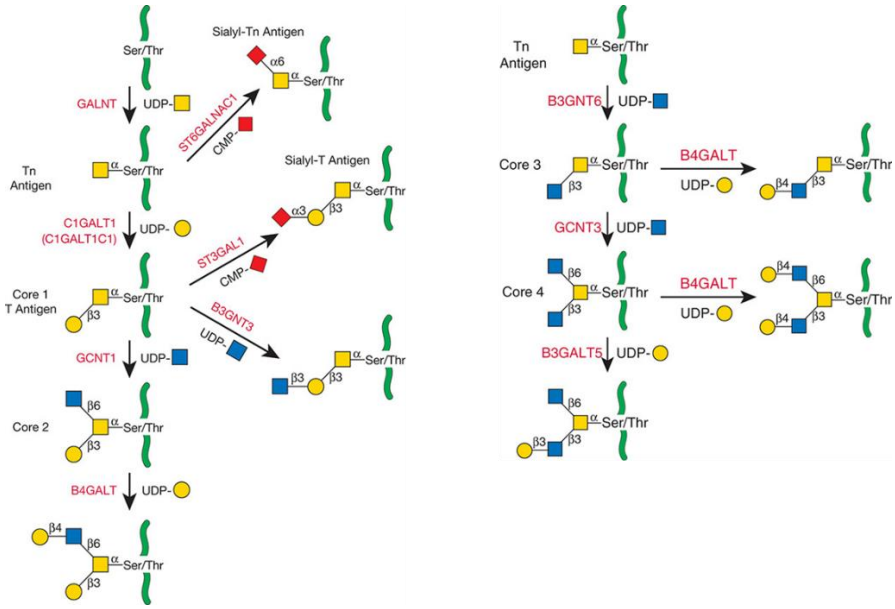


Figure 1.2. Biosynthesis of core 1, 2, 3 and 4 O-GalNAc glycans. From *Essential of Glycobiology*, Chapter 10, O-GalNAc glycans.

Usually, this is a non-stoichiometric modification difficult to be characterized (I. Bagdonaite et al., 2015; Tian et al., 2021). Furthermore, *O*-glycosylation takes place on exposed side chains, usually part of secondary structures. Statistical studies have shown that proline has a high frequency of occurrence at +3, +4 and -1, -4 positions relative to the *O*-glycosylated Ser/Thr (Veluraja, 2001). In addition, there are software that, given the protein primary structure, can predict the most probable *O*-glycosylation sequons, like NetOGlyc 4.0 (Steentoft et al., 2013). Finally, glypiation consists of the post-translational addition of a glycosylphosphatidylinositol (GPI) anchor to those proteins which bring a GPI signal sequence at the carboxy terminus, that is cleaved before modification (Rudd et al., 1997).

Typically, the same glycosylation site can present different glycans so that the same protein exists in distinct glycoforms: this feature is also called microheterogeneity and it depends on several cellular factors, like tissue/cell type-specific expression of

processing enzymes along with their concentrations and kinetic parameters, availability of the nucleotide sugar donors, competition of different modifying enzymes for the same oligosaccharide substrate and contact time in the reaction compartment (*e.g.*, Golgi apparatus). In addition, intrinsic protein structural properties can have a strong effect on site-specific *N*-glycan processing (Dicker et al., 2015). Examples of homogeneous glycoproteins are rare. Known examples are the soybean agglutinin from plants and the human serum transferrin: the first contains only the Man9GlcNAc2 glycan, while the second has exclusively the A2G2 glycan, which has a biantennary structure with composition Gal2Man3GlcNAc4 (Rudd et al., 1997).

About 70% of all proteins contain *N*-glycosylation sequons and ~70% of sequons carry a *N*-glycan (classified as high-mannose, hybrid and complex); this means that there are potential *N*-glycosylation sites that are not occupied. Thus, the Asn-X-Ser/Thr consensus is a necessary but not sufficient condition for *N*-glycosylation. Moreover, the same glycosylation site can bind or not a glycan in different protein molecules; this concept takes the name of site occupancy (or macroheterogeneity), and it refers to the stoichiometry of substitution of a certain glycosylation site. *N*-glycosylation efficiency depends on many factors, the most relevant are the nature of the central amino acid (X) in the glycosylation sequon and the competition between protein folding and glycan attachment reaction during protein biosynthesis (Rudd et al., 1997).

Glycans play a fundamental role in protein folding and stabilization: in fact, there is evidence for glycoprotein misfolding and aggregation after glycans removal, even if this is not true for all glycoproteins. It was suggested that glycans could have a ‘’chaperone-like’’ activity, affecting the height of the folding barrier, beyond their main function in the calnexin-calreticulin lectins quality control system (Shental-Bechor et al., 2008). Furthermore, thanks to their hydrophilicity, oligosaccharides contribute to increase protein solubility and/or its stability against proteolysis by shielding cleavage sites, extending protein half-life. Glycans also stabilize the protein

by making key interactions with nearby amino acids, so improving the protein thermal stability and slowing down the aggregation kinetic (Jayaprakash et al., 2017).

1.2 Implication of glycosylation in viral antigen design

Viruses exploit the host cell glycosylation machinery to decorate their envelope proteins with glycans typical of the infected cell; one exception was recently found in giant virus *Paramecium bursaria* Chlorella virus (PBCV-1) (De Castro et al., 2016), and other giant viruses (Notaro et al., 2021; Speciale, Notaro, Abergel, et al., 2022), which encode a set of glyco-related genes, as those involved in sugar-nucleotide metabolism, as well as glycosyltransferases, showing both uncommon N-glycans structures and sequons (Asn β-glucose linkage instead of Asn β-di-N-acetylchitobiose linkage).

Various viral vaccine proteins available are glycosylated and the composition of the attached glycans potentially influences the efficacy and immunogenicity of the antigen; in addition, protein glycosylation is cell-type dependent, making the choice of the expression system a crucial issue.

Independently from the type of vaccine considered (live-attenuated, inactivated, recombinant, nucleic acid), protein antigen glycosylation influences its immunological properties, for this reason the development of glycoengineering technology aims to design vaccines with a tailor-made glycosylation pattern. Enveloped viruses hijack host cell glycosylation machinery and adorn their own glycoproteins with host-derived glycans; these glycans play roles in structural stability, shielding the pathogen surface from immunodetection and are important for attachment to host receptors to gain entry into cells.

The first function that has been attributed to the glycans on the surface of the viral antigenic proteins consists in shielding the conserved protein epitopes, which hampers the recognition by immune system. This concept is exacerbated in HIV-1 Env, one of the most heavily glycosylated proteins known with glycans making up 50% of its total mass (Behrens et al., 2016). BG505 SOSIP.664 is a model HIV-1 antigen which

consists of a soluble recombinant trimer that mimics the structure of the native form of the virion-associated Env, which comprises a noncovalently linked trimer of gp120/gp41 heterodimers. Env is synthesized as a precursor glycoprotein, gp160 (Fig. 1.3), later processed into gp120/gp41 by the cells infected with HIV-1, this step is indispensable to confer infectivity to HIV-1 virions.

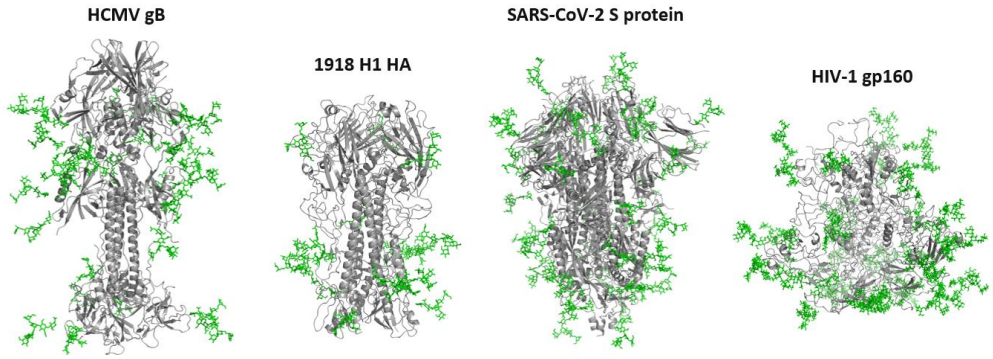


Figure 1.3. *N*-glycans shield of different peplomers (high-mannose M5 structure has been used for all the glycoproteins sequons). Generated using GLYCAM and the following PDB files: 1RUZ, 5CXF, 6PWU, 6VXX.

It is possible to identify four main categories of immunomodulating strategies, depending on the aim for which the glycan is used: immuno-shifting, immuno-focusing, immuno-broadening and immuno-altering. With immuno-shifting, glycan masking is used to hide selected epitopes so as to lead the immune response toward other epitopes that elicit high antibody titers endowed of low neutralization activity (‘’decoy epitopes’’, ‘’decotopes’’); multiple holes in the oligosaccharide shield may act as immunological decoys and thereby impede the development of neutralizing antibodies (Derking et al., 2020).

Immuno-focusing is a concept very similar to immuno-shifting. Indeed, the only difference refers to the insertion of a glycosylation site next to the epitope on which the immune response has to focus. So, while in immuno-shifting the goal is to elicit the immune response away from certain epitopes, in immuno-focusing the attention

is placed on a specific epitope. When hyper-variable regions are masked by glycans, the immuno-broadening concept applies, while immuno-altering is a term used when the titer of antibodies elicited is increased after antigen glycoengineering (Martina et al., 2023).

It is important to underline that protein architecture can have a substantial impact on the glycosylation pattern. Indeed, under-processed oligomannose-type and hybrid-type glycans are common on viral glycoproteins and arise because of either glycan- or protein-mediated steric clashes with ER and Golgi resident mannosidase enzymes, which are detrimental for the glycan processing (J. D. Allen et al., 2021). For example, on SARS-CoV-2 S protein there are at least seven *N*-glycosylation sites that present poorly processed glycans likely due to the limited access of mannosidases. In such cases, the oligomannose-type structures prevail over the other possible *N*-glycans, because seem protected by the nearby protein regions, as exemplified by N234 glycan (Fig. 1.4), which is partially sandwiched between the N-terminal and receptor binding domains (Watanabe et al., 2020).

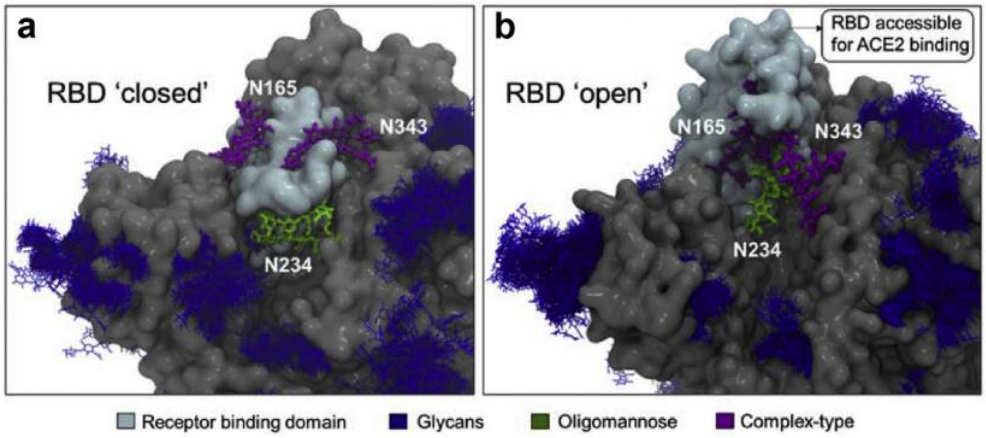


Figure 1.4. Effect of N165, N234 and N343 glycans on S protein RBD conformational equilibrium. From *Principles of SARS-CoV-2 glycosylation.* (Chawla et al., 2022).

Oligomannose-type glycan content is linked to the glycosylation density of a glycoprotein, defined as the number of glycans per surface area ($Å^2$). Compared to

other viral glycoproteins, SARS-CoV-2 S protein has an oligomannose-type glycan content of 28% therefore it is less densely glycosylated. Glycans can also affect glycoprotein conformation and consequentially receptor recognition; for example, SARS-Cov-2 S protein RBDs can exist in 2 distinct conformations (Fig. 1.4): 'up' for a receptor-accessible state (called also RDB_{open}) and 'down' (or RDB_{closed}) for a receptor-inaccessible state (Jackson et al., 2022; Watanabe et al., 2020). The N234 glycan is one the most ordered glycans observed in cryo-EM structures and this glycan has also been shown by MD simulations to stabilize the RBD open conformation and its changes (Chawla et al., 2022). Moreover, deletion of some key glycans (N165A, N234A, and N343A) through single point mutations led to a significantly reduced binding to the ACE2 receptor; further insight from MD and cryo-electron microscopy (cryo-EM) studies have shown that glycosylation at N343 is essential for gating the RBD opening and closing. Overall, the interaction of the clustered N165, N234, and N343 glycosylation sites with the surrounding protein helps to stabilize the orientation and dynamics of the open RBD conformation allowing thereby supporting the receptor recognition event (Chawla et al., 2022). Using equilibrium molecular dynamics simulations, Maity et al. (2023) showed that the absence of the N370 glycan on SARS-Cov-2 S protein enabled the interaction between the ACE2 glycans and the RBD domain of the viral protein (Maity et al., 2023). Based on these results and on the fact that the homologue proteins from bat RaTG13 and pangolin PCoV_GX coronavirus S trimers are instead glycosylated at this same position, the authors hypothesized that the lack of this glycan could have driven the spillover event (Maity et al., 2023). At N343 there is an amphipathic nature complex *N*-glycan, able to interact with the hydrophobic residues of a nearby beta sheet through its core GlcNAc-β(1- 4)GlcNAc and to engage hydrogen bonds with the surrounding water and with the aa 365–375 helical loop). Such interactions enhance the stability of the RBD architecture, bridging between the two partially helical loops that frame a highly hydrophobic beta sheet core (Ives et al., 2024). Removal of the glycan at N343 triggers a tightening of the loops, causing the detachment of the hydrophilic patch and the

misfolding from the ACE2-recognized conformation. These results are in agreement with the drastic reduction of viral infectivity observed upon deletion of both N331 and N343 glycosylation in the WHu-1 strain, where loss of structure may bring to the loss of function, indeed glycan N343 acts as a 'glycan gate' pushing the RBD from the 'down' to the 'up' conformation (Ives et al., 2024). Furthermore, the binding affinity of the RBD for the oligosaccharides of the monosialylated gangliosides GM1os and GM2os is modulated by N343 glycosylation. These were shown to function as co-receptors in WHu- 1 infection (Ives et al., 2024).

The nature of the glycans on the surface of a protein antigen can affect its immunogenicity; high-mannose-type and single-GlcNAc-type *N*-glycans are more effective than complex-type *N*-glycans in triggering activation and maturation in both mouse and human myeloid dendritic cells (hmDCs). It is likely that terminally sialylated complex-type *N*-glycans provide shielding against the activation of pattern recognition receptors (*e.g.*, mannose or another C-type lectin) on mDC surfaces; this phenomenon can be chemically explained considering that lectins recognize glycans mainly by CH−π stacking interactions, which involve the donation of the electrons from an aromatic−π system (phenylalanine, histidine, tyrosine, and tryptophan) to the C−H σ* orbitals of the monosaccharide unit (Spiwok, 2017). The aromatic π cloud is electron-rich, therefore, such interactions rarely involve anionic carbohydrate residues such as sialic acid; some monosaccharides stack on aromatic residues far more often than others, specifically, β-galactose, β-N-acetylgalactosamine, and β-mannose are strong CH−π donors (Kiessling et al., 2021). Furthermore, protein antigen immunogenicity can be increased by introducing non-self-motifs and/or monosaccharides (*e.g.* 1,3-fucose, xylose, or α-Gal epitope) in its glycans (Hodzic et al., 2020; Sandig et al., 2017).

In some cases, glycan shield can be exploited to elicit an immune response; indeed, although this "glycan shield" obscures conserved protein epitopes from immune recognition, the immune system can mount neutralizing responses against these

glycans. The first monoclonal antibody able to recognize viral glycans was the 2G12. It was isolated from an asymptomatic HIV-1 infected patient in 1990 and its neutralizing activity against HIV-1 strains and the binding to the glycoprotein 120 (gp120) was described for the first time in 1994 (Buchacher et al., 1994; Kosma et al., 2012; Trkola et al., 1996). 2G12 is a broadly neutralizing antibody (bNAb) which recognizes a conserved M9-glycans cluster (mannose patch) on the surface of HIV-1 gp120, located near the V3 and V4 protein loops (interacting glycans are N295, N332, N339, and N392) (Miller et al., 2022; Scanlan et al., 2002). Nowadays, several antibodies directed against the glycans of HIV-1 are known (Gillmann et al., 2023), for example, PGT antibodies are able to penetrate the HIV-1 glycan shield. These antibodies are classified in different families from different germline lineages such as PGT121–123/PGT133–134/10-1074, PGT125–128/PGT130–131, PGT135–137, PGT151–158, each able to recognize different glycan structures (Falkowska et al., 2014; Garces et al., 2014; Kusumoto et al., 2020; Sok et al., 2013). A recently discovered glycan-dependent antibody is S309, which is a neutralizing mAb isolated from an individual infected with SARS-CoV in 2003; it has a cross-neutralizing activity, in fact, it binds to both SARS-CoV-2 and SARS-CoV RBDs with nano-molar affinity. S309 recognizes a proteoglycan epitope on the domain 1 of SARS-CoV-2 S1, distinct from the receptor-binding motif; the glycan involved in this recognition is a complex glycan (Pinto et al., 2020).

1.3 Glycoengineering approaches

Glycoengineering is one of the approaches currently implemented to produce therapeutic proteins presenting a homogenous glycan composition. There are two different types of glycoengineering: *in vivo* and *in vitro*. *In vivo* glycoengineering uses molecular biology methods such as gene knockout, knockdown, knocking, overexpression, mutation, or the use of inhibitors to change the type and concentration of the glycosidases and glycosyltransferases available inside these cells with the effect of modifying the glycosylation patterns of interested proteins expressed by these cells

(Ma et al., 2020). In mammalian CHO cells, the improvement of antibody-dependent cell-mediated cytotoxicity (ADCC) effector function has been demonstrated successfully by the knockout of the fucosyltransferase 8 (*fut8*) gene to produce non-fucosylated recombinant monoclonal antibodies (Tejwani et al., 2018); this phenomenon is explained by the fact that core fucose causes a steric clash that weakens carbohydrate-carbohydrate interactions required for high affinity receptor recognition (Schneider et al., 2017; Tejwani et al., 2018). The crystal structure of the complex between afucosylated Fc (from human IgG) and FcγRIIIa receptor presents several hydrogen bonds between the first two core GlcNAc units of the receptor and GlcNAc1 of the Fc. Moreover, this interaction is further stabilized by the presence of a hydrogen bonding network mediated by five water molecules. The described carbohydrate-mediated interaction is responsible for up to 100-fold gain in binding affinity for afucosylated *versus* fucosylated IgGs (Ferrara et al., 2011).

CHO cells occasionally contain *N*-glycans capped with N-glycolylneuraminic acid (Neu5Gc). Neu5Gc in glycans can be a cause for concern as Neu5Gc-capped glycans act as "xenoautoantigens" in humans and cause the "xenosialitis," an inflammatory process initiated by the binding of naturally occurring antibodies against Neu5Gc in the human body. Various strategies have been attempted to eliminate the incorporation of Neu5Gc from the glycoprotein; an easy glycoengineering approach consists in the knockout of the gene encoding CMP-Neu5Ac hydroxylase which converts CMP-Neu5Ac to CMP-Neu5Gc. Similarly, the α2,3-sialyltransferase can be eliminated by genetic modification and the ectopic expression of the missing α2,6-sialyltransferase can lead to humanized *N*-glycans (Dicker et al., 2015). Besides modulating the glycan structure at specific glycosylation sites, glycoengineering can also be performed by changing the number of glycosylation sites. In this regard, the most representative example is the glycoengineering of human erythropoietin (hEPO) via site-directed mutagenesis, increasing the number of *N*-glycosylation sites from 3 to 5 (Ma et al., 2020). Moreover, N-acetylglucosaminyltransferase I deficient (GnT1−) human

embryonic kidney cells that are unable to synthesize complex type *N*-glycans (producing mainly M5 glycans) were used to produce secreted, transmembrane domain-truncated hemagglutinins; these high-mannose HAs were then further trimmed with the high-mannose-cleaving enzyme endoglycosidase H (EndoH), leaving a single GlcNAc residue, dramatically decreasing the size and shielding effect of *N*-glycans while still maintaining the native HA structure in its trimeric state; this technique has been shown to elicit cross-neutralizing antibodies against antigenically diverse strains of influenza viruses within a subtype (Chen et al., 2020).

It is possible to engineer *in vivo* the *N*-glycosylation pattern by introducing in the culture medium enzymatic inhibitors (chemical glycoengineering), namely small molecules which mimic the monosaccharides structures while inhibiting one or more enzymes involved in the *N*-glycan biosynthesis; this approach is operationally extremely simple and applicable across different cell culture systems, making it highly attractive from a practical viewpoint. A first inhibitor example is kifunensine, an alkaloid originally isolated from an actinobacterium (*Kitasatosporia kifunense*). This molecule is a strong inhibitor of α-mannosidase I in the ER (ERMI) and/or cis-Golgi apparatus (GMI), thus it prevents *N*-glycans maturation with the consequence that the glycoproteins essentially present only M9 (Fig. 1.1). A second inhibitor is castanospermine (Fig. 1.1), which blocks the removal of the first (and therefore all subsequent) D-glucose residue(s) during the initial trimming step in the ER (S. Li et al., 2021). Another relevant inhibitor is the indolizidine alkaloid swainosine (from the fungus *Rhizoctonia leguminicola*), which inhibits α-mannosidase II in the ER (ERMII) and/or medial-Golgi (GMII), bringing to the formation of hybrid N1M5 glycans (Choi et al., 2018) (Fig. 1.1). It is also possible to completely prevent the *N*-glycan binding to the protein by using inhibitors of the GlcNAc phosphotransferase (GPT), such as tunicamycin (a nucleoside from *Streptomyces chartreusis*) (Fig. 1.1). However, this often brings to misfolded proteins (Dawood et al., 2020). Finally, the synthetic molecule fucostatin I was found to be a potent inhibitors of antibody fucosylation (J. G. Allen et al., 2016) (Fig. 1.1).

On the contrary, *in vitro* glycoengineering uses chemoenzymatic glycan remodeling tools such as glycosidases and glycosyltransferases to trim or extend the sugar chains in an intact glycoprotein (L. X. Wang et al., 2019). The advantage of *in vitro* enzymatic glycoengineering is that, by improving the structural control of protein glycosylation, it allows for relatively easy access to homogeneous glycoforms (Ma et al., 2020). With such research examples, quantitative structure-function relationships can be derived and used to design new proteins to be afforded by glycoengineering research. However, the efficiency of this method is currently still limited due to the limited range of commercial substrates that are recognized can be tolerated by glycosynthases and to the difficulty to control glycosylation site selectivity; it is also challenging to use this method in large-scale.

1.4 Impact of glycosylation on different vaccine platforms

The parameters that a vaccine has to account are essentially: safety for the patient, strong protection against the pathogen and cost effectiveness for large-scale manufacturing. There are several viral vaccine platforms (Fig. 1.5), each one with a specific balance among the three listed criteria. An ideal vaccine should start by priming the innate immune response first to develop at a second stage the protective antigen-specific adaptive immune response, along with a long-lasting immunological memory (Schijns et al., 2021).

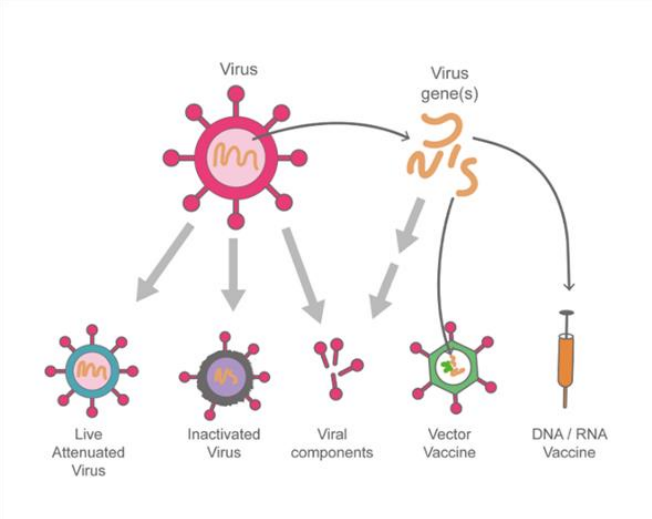


Figure 1.5. Viral vaccine platforms — The Open University, United Kingdom. From *Vaccines for Viruses and Covid-19*.

The effects of glycosylation depend on the vaccine platform. Considering whole virus vaccines, protein glycosylation affects the virus life cycle in several ways (promotion of expression, transport, fusion, binding to cell surface) (Feng et al., 2022); for example, the N-linked glycosylation at Asn-154 of the Zika virus (ZIKV) E protein was shown to affect virus assembly and life cycle (Gwon et al., 2020). Hepatitis C virus (HCV) encodes two envelope glycoproteins E1 and E2: mutations in glycosylation sites E1N1, E2N3, E2N7, E2N8, E2N10, and E2N11 have a negative impact on virus particle assembly and/or secretion as well as virus entry (Helle et al., 2010). The presence of the glycosidase inhibitor kifunensine, which generates only high-mannose glycans, abolishes Omicron SARS-CoV-2 pseudoviruses infectivity while it has no effects on earlier variants infectivity (D614G or B.1.617.2) (Lusvarghi et al., 2023). These modifications can have both negative and positive consequences, in fact, they can lower vaccine immune response, but they can be exploited to generate weakened or inactivated viral particles when the virion is sensitive to glycan pattern modifications. On the other side, recombinant vaccines allow a punctual control of the protein glycan pattern, due to the possibility to engineer the protein sequons,

hiding or exposing specific epitopes (epitope-focused vaccine design). The lower immunogenicity of this type of vaccine can be overcome by the advantage to expose, by glycan removal, conserved protein epitopes, which can elicit the production of bNAbs useful against viruses with high mutation rate like HIV-1. Furthermore, recombinant vaccines are useful when dealing with viruses capable of viral integration following exposure and establishment of life-long persistence, such as HIV-1 (Nkolola et al., 2024). Native-like Env trimers, such as BG505 SOSIP.664, are designed to present multiple bNAb epitopes, among which there are also glycan epitopes, such as the V3-glycan patch located at the base of the V3 loop region, and the V2-glycan epitope which includes an N-linked glycan at position 160 and a lysine-rich carboxy-terminal strand (Daniels et al., 2019). Therefore, the retaining of specific glycan structures on Env is fundamental to obtain a broad protection against HIV-1 variants.

A main drawback about the glycosylation impact on recombinant, inactivated and, to a lesser extent, nucleic acid vaccines concerns the degree of similarity between the vaccine antigen glycosylation pattern and the virion antigen; indeed, inflammation taking place during an infection can alter the glycosylation process yielding to glycoproteins with glycans different from those obtained by recombinant (or nucleic acid) technology, that can have a diverse elicitation of the immune response. It has been demonstrated that cell surface proteins during HIV transcription have different glycosylation profiles (Colomb et al., 2020). Furthermore, the expression system (even if uses human cell lines) will never give a glycosylation pattern identical to that of the immunized patient. The main advantage about nucleic acids vaccines concerns the fact that they encode proteins in the host with a glycosylation pattern almost identical to the protein produced by viral infection because the viral antigen is expressed by the patient cells (Donnelly et al., 1997). Like in recombinant protein vaccines, with nucleic acids vaccines it is possible to engineer the antigen glycosylation sites; for example, the immunogenicity of an HIV-1 Env expressing DNA immunogen has been boosted by elimination of N-linked glycosylation signals

known to be engaged in the shielding of gp120 neutralization epitopes (Bolmstedt et al., 2002).

Chapter 2 - Human Cytomegalovirus

2.1 Virology and pathogenicity of Human Cytomegalovirus

The human cytomegalovirus (HCMV) is a betaherpesvirus with a double stranded DNA of about 235kb, the mature virions have a size between 200-300 nm. There are about 30 to 40 viral proteins that compose the HCMV virion, distributed as follows: 50% tegument proteins, 30% capsid proteins, 13% envelope proteins, and 7% undefined proteins. The most abundant protein in the virion is the tegument protein pp65; the envelope most abundant proteins are gM, gB, gH, gL, gN, gO (Fig. 2.1) (Varnum et al., 2004). The human cytomegalovirus is perhaps the most ubiquitous human infection. Although better hygiene and lesser close contact between children and adults have decreased the prevalence of CMV in developed countries, virtually 100% of adults in low- and middle-income countries have been infected when young. The mode of HCMV transmission is presumed to occur through body fluids, including saliva, urine, blood, and tears; moreover, HCMV can be transmitted sexually and *via* breast milk, and it also propagates through transplanted organs.

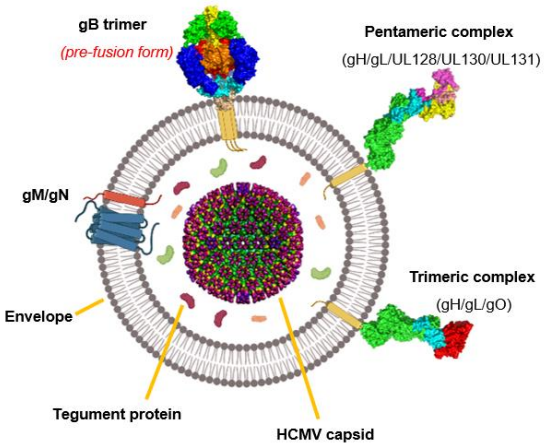


Figure 2.1. Schematic structure of HCMV virion.

HCMV establishes lifelong latent infections, and its reactivations are typically suppressed by competent immune systems. However, HCMV can cause disease in the immunocompromised individuals including symptoms such as gastrointestinal ulceration, hepatitis, pneumonitis or retinitis in solid organ transplant patients and retinitis in patients with AIDS, which can lead to blindness. HCMV is also a leading cause of congenital viral infections in newborns where it can cause permanent defects such as deafness, blindness, epilepsy, mental retardation and microcephaly. In the United States nearly 1 in 3 children is already infected with CMV by the age of 5 (P. Griffiths et al., 2015).

Hitherto, there is no available vaccine against HCMV and only antiviral drugs and immunoglobulins from seropositive individuals have been approved. The primary antiviral agents used to treat HCMV infections in patients with an impaired immune system are ganciclovir, foscarnet, and cidofovir, with the first one being the most effective and the standard of care (David R. Snydman, 1987). These drugs have improved the survival and quality of life of immunocompromised individuals suffering from HCMV, but they are far from ideal due to major hematologic, renal, and neutropenia toxicity; furthermore, ganciclovir has associated toxicity and cannot be administered to some patients such as pregnant women (Benoist et al., 2013).

When infecting a target cell, HCMV envelope glycoproteins positioned on the outside of infectious virions, interact with host receptors to mediate the fusion or endocytosis of the virion into the cell. Viral tegument proteins bound to the capsid are believed to interact with the host microtubule machinery to transport the viral capsids to the nuclear envelope and into the nucleus, where viral transcription, genome replication, and encapsidation occurs. At the same time, other tegument proteins are deposited in infected cells by incoming virions and targeted to different subcellular locations to inhibit the initial steps of immune response and to regulate viral gene expression. Capsids, assembled in the nucleus, egress through the nuclear double membrane by disruption of the nuclear lamina and formation of a nuclear egress complex. Once capsids reach the cytoplasm, the assembly and transport of virions occurs via the

integration of multiple cellular trafficking pathways. The cellular secretory machinery, including the ER, Golgi apparatus, and endosomal machinery, is hijacked for the formation of a cytoplasmic viral assembly complex (AC). At the AC, capsids acquire their tegument layer and viral envelope from intracellular vesicles. The generated infectious particles, as well as other noninfectious particles named dense bodies, are next released into the extracellular space (Fig. 2.2) (Jean Beltran et al., 2014).

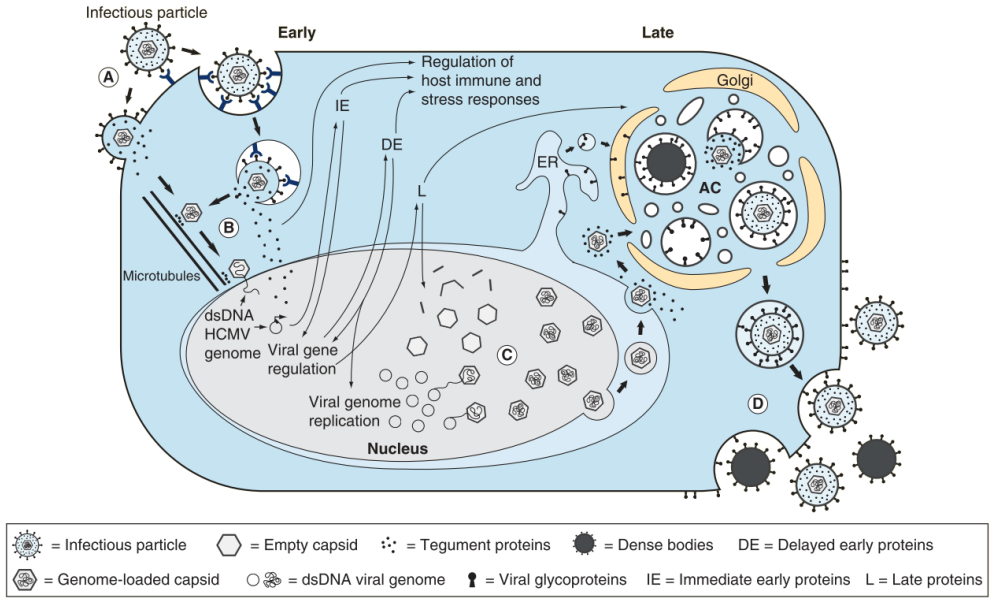


Figure 2.2. Life cycle of human cytomegalovirus infection. From *The life cycle and pathogenesis of human cytomegalovirus infection: Lessons from proteomics*. (Jean Beltran et al., 2014).

The development of HCMV vaccines began in the 1970s soon after the toll of the virus on infants in utero and transplant recipients became obvious. Two vaccine strains were attenuated starting with viruses that had been isolated for laboratory work: AD-169 and Towne (Elek et al., 1974; Plotkin et al., 1975). The Towne attenuated strain

showed interesting preliminary data about protection, however further studies demonstrated that immunization against infection was not statistically significant; moreover, the attenuated strain failed to prevent natural acquisition of HCMV by women exposed to children in day care.

A remarkable step in HCMV vaccine development was the successful purification of the envelope protein called glycoprotein B (gB), indeed, when combined with the oil-in-water adjuvant MF59, good levels of neutralizing antibodies were produced in humans after three injections over a six-month period.

Another important event was the discovery by researchers at Princeton University that a pentameric complex of proteins (pentamer) was present on the surface of HCMV and that this structure, consisting of glycoprotein H (gH), glycoprotein L (gL), and the products of genes UL128, 130 and 131, elicited far more neutralizing antibodies than gB; furthermore, the pentamer was associated with protection against transmission to the fetus (Itell et al., 2017).

The main candidate vaccines developed up to now belong to different classes. For example, Merck Sharp & Dohme Corp are focusing on a live-attenuated vaccine which uses a genetically engineered AD169 strain to express the pentamer; Sanofi Pasteur is instead developing a vaccine based on the recombinant subunit gB with MF59 adjuvant, while Moderna is prioritizing the mRNA platform.

2.2 Structure and function of HCMV antigens: gB and Pentamer

gB is an homotrimer protein whose function is to force the virus membrane to fuse with that of the target cell (fusogen) to start the infection event; each monomer has a molecular weight of ≈100 kDa and it has 17-18 *N*-glycosylation sequons (depending on the strain). The monomer has 9 different domains: DI, DII, DIII, DIV, DV, a membrane proximal region (MPR), a transmembrane domain (TM), a N-terminal domain and a cytodomain (Fig. 2.3a); however, to obtain the crystal, both the N-terminal domain and the cytodomain have been removed.

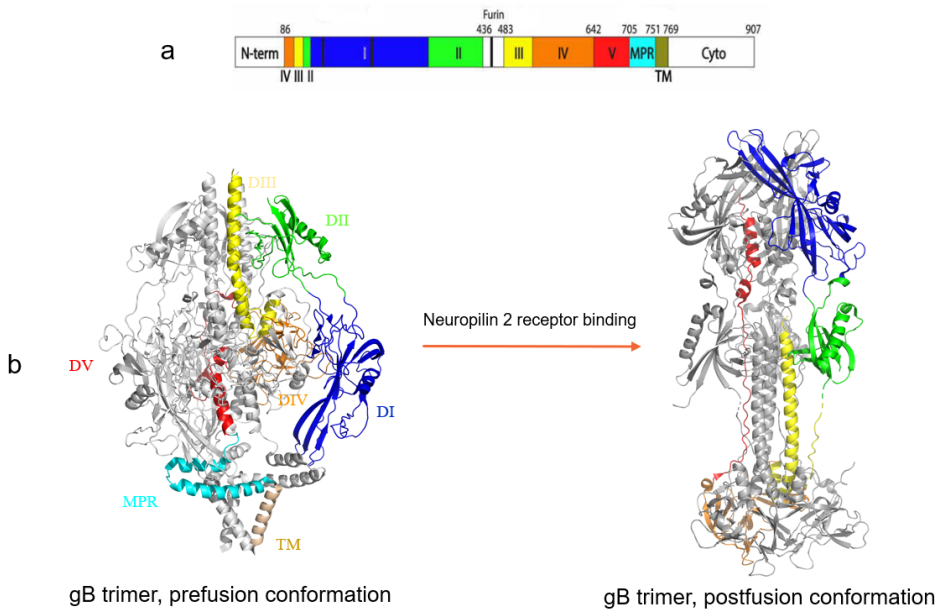


Figure 2.3. (a) gB schematic representation. (b) gB prefusion and postfusion conformations, with highlighted domains. Generated using the following PDB files: 7KDP and 5CXF.

gB trimer can exist in two different conformations: prefusion form and postfusion form (passing for an intermediate hemifusion form) (Fig. 2.3b). In the prefusion form the hydrophobic loops (fusion peptides) of the domain I are hidden inside the structure; upon binding of neuropilin 2 to pentamer, gB is activated and changes its conformation by rotating DI of 180° and displaying the fusion peptides towards the target cell membrane. After catching the target membrane (with hemifusion form) the protein drives the virus membrane towards the cell membrane to fuse with it (postfusion form) (Figs. 2.4, 2.5) (Liu et al., 2021).

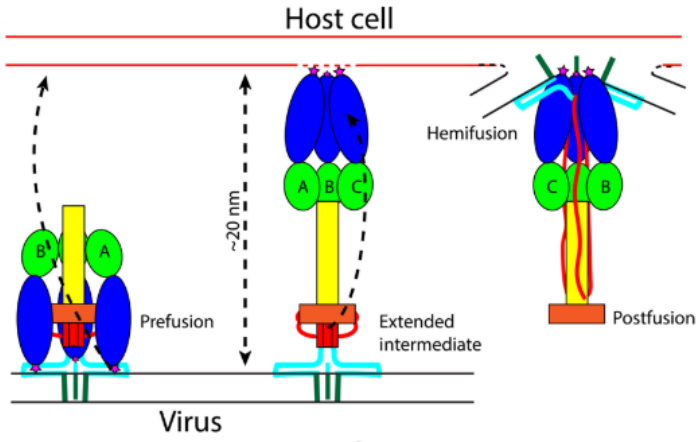


Figure 2.4. Illustration of the membrane fusion mechanism promoted by gB. From *Prefusion structure of human cytomegalovirus glycoprotein B and structural basis for membrane fusion.* (Liu et al., 2021).

The dimeric gH/gL complex is a covalent complex, which can interact either with gO or with the ULs proteins (UL128, UL130, UL131A) to form the pentamer gH/gL/gO and the pentamer gH/gL/UL128/UL130/UL131A, respectively (Fig. 2.1). The type of complex that will be assembled will define the cell tropism (kind of cell that virus can infect). In fact, HCMV entry into fibroblasts is mediated by gH/gL/gO (recent data suggest that it may contribute to entry into all cell types), while HCMV entry into dendritic, epithelial, endothelial cells and leukocytes requires the pentameric glycoprotein complex (Chandramouli et al., 2017). Mass spectrometry analysis demonstrated that one gH/gL dimer forms a disulfide bridge with a second gH/gL dimer, forming a new homodimer involving gL–Cys144 of a first monomer and gL–Cys144 of a second one.

Pentamer is a complex with an apparent molecular mass of ≈180 kDa with the function of defining cell tropism towards endothelial and epithelial cells (Ciferri et al., 2015). As for the 5 different proteins (gH/gL/UL128/UL130/UL131A), the gH/gL complex is covalently linked to UL128 through disulfide bond and non-covalently to UL130 and UL131A), and the entire complex adopts a helicoid shape 180 Å in length and 30

to 80 Å in cross-section. The gH/gL part of the complex has a domain organization similar to that of other herpesviruses, with gH domains (DI to DIV) stacking on top of each other. The polypeptide chains of the UL130 and UL131A form a gently curved subcomplex that binds to an extension at the N-terminus of gL.

When the pentameric complex recognizes neuropilin 2 (a receptor protein), it triggers a conformation switch which results in the dissociation from gB and in the activation of the fusion mechanism (function of priming gB) (Fig. 2.5). The pentameric complex has 11 *N*-glycosylation sequons: 6 in gH, 1 in gL, 3 in UL130 and 1 in UL131A (Chandramouli et al., 2017). Overall, the complex has two different faces, one decorated with 10 glycans while the other features a single glycan in UL130 and positively charged areas with clusters of exposed arginine and lysine residues.

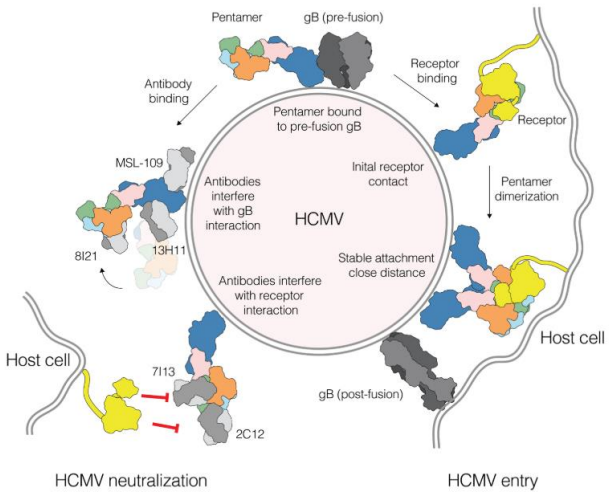


Figure 2.5. Model for HCMV Pentamer neutralization and Pentamer-mediated cell entry. From *Structural basis for HCMV Pentamer receptor recognition and antibody neutralization.* (Kschonsak et al., 2022).

2.3 Antigenic and immunological aspects of gB and Pentamer

gB displays 5 antigenic domains (AD-1, AD-2, AD-3, AD-4, AD-5) (Fig. 2.6), of which AD-4 and AD-5 are the most promising candidates for production of therapeutic immunoglobulins; in fact, they are both highly immunogenic and elicit the production of broadly neutralizing antibodies (nAb); two crystal structures of gB in complex with the two neutralizing antibodies, 1G2 Fab and SM5-1 Fab respectively, have been reported (Chandramouli et al., 2015; Chandramouli et al., 2017). The proteins have been deprived of the glycan shield (either by *E. coli* expression or Endo H treatment) to allow the formation of the crystals.

gB domain	Immunological features	NAb
AD-1	Strongest immunogenic domain	X
AD-2	Mildly immunogenic (unresolved)	V
AD-3	Immunogenic (cytodomain)	X
AD-4	Highly immunogenic	V
AD-5	Highly immunogenic	V

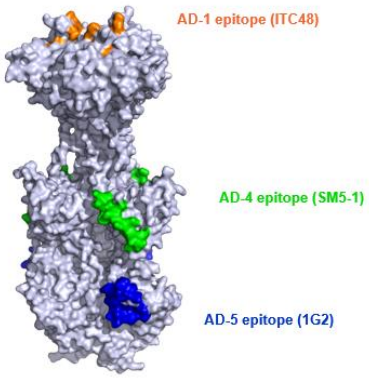


Figure 2.6. (a) Immunological details of gB antigenic domains. (b) Postfusion structure of gB with highlighted antigenic domains; AD-2 is not resolved in the crystal structure, while and AD-3 is in the cytodomain which has been removed from the shown structure. From *Structure of HCMV glycoprotein B in the postfusion conformation bound to a neutralizing human antibody.* (Chandramouli et al., 2015).

It has been demonstrated that two monoclonal antibodies, 1–223 and 12–874, isolated from a naturally-infected individual and a V160-vaccinated individual, respectively, had a 30-fold increase in binding to gB ectodomain treated with PNGaseF compared to untreated. Through site-directed mutagenesis it has been shown that glycosylation of the N73 asparagine shields antibody binding to the AD-2 region (Valencia et al., 2023).

Nowadays, only formulations containing the postfusion form of gB have been investigated; however, based on the analogy with the respiratory syncytial virus (RSV) vaccine candidate, it has been hypothesized that the prefusion gB may prove more effective than postfusion gB at eliciting neutralizing antibodies. The difficulties in expressing gB in its prefusion conformation lie in the stabilizing role that the pentamer and the envelope play towards the prefusion conformation (Ebel et al., 2022). Immunization with HCMV gB vaccine plus MF59 adjuvant resulted in a 40% to 50% efficacy in solid organ transplant patients and showed a correlation between the protection against the virus and the titer of antibodies against gB (Gomes et al., 2019).

Seven broadly neutralizing epitopes for the sites 1 to 7 of the pentamer have been identified previously (Chandramouli et al., 2017), with five of them (1, 2, 3/7, 4/6, and 5) recognizing distinct epitopes (Fig. 2.7).

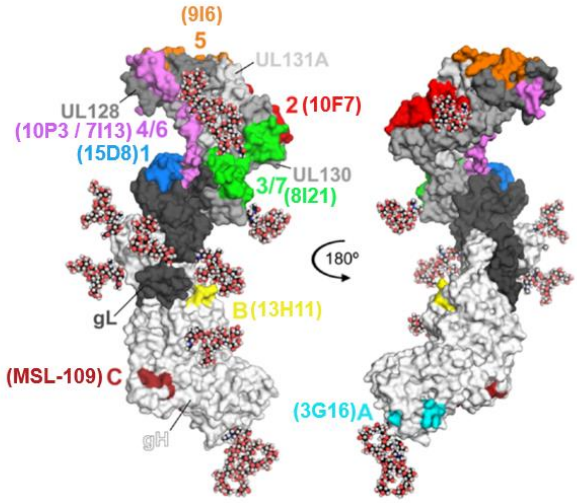


Figure 2.7. Two 180°-rotated surface views of Pentamer. The location of the epitopes for neutralizing antibodies is depicted with colored surface patches (blue for site 1, red for site 2, green for site 3/7, pink for site 4/6, and orange for site 5). Furthermore, additional neutralizing epitopes located on gH are mapped in cyan (site A), yellow (site B), and dark red (site C). The neutralizing antibody for each epitope is reported in the brackets. From *Structural basis for potent antibody-mediated neutralization of human cytomegalovirus.* (Chandramouli et al., 2017).

Antibodies binding to UL128 and UL131A residues located at the tip of Pentamer (9I6; site 5), UL128-a3 helix (15D8; site 1), and close to the linker connecting UL128-a2b4b5b6 to UL128-a3 (10P3 and 7I13; sites 4 and 6, respectively) prevent the pentamer binding to the cells. These antibodies may inhibit the interaction of Pentamer with a cell surface receptor, by either direct competition or steric hindrance, and their binding sites identify a surface of the Pentamer likely involved in the binding with a cell surface receptor (Chandramouli et al., 2017). Antibody binding to the elbow of the ULs arm (4N10 and 8I21; sites 3 and 7, respectively) and the solvent-exposed side of the UL130/UL131A b sheet (10F7; site 2) does not affect pentamer binding to cells, suggesting a different mechanism of neutralization; it is plausible that

site 2 and 3/7 antibodies may interfere with the interaction between the antigen and a co-receptor at the cell surface in a post-entry step (Chandramouli et al., 2017).

This pentameric complex is important for eliciting protective humoral immune response and it is the predominant target of neutralizing antibodies in human, accounting for >85% neutralizing activities in natural infection; furthermore, pentameric complex is associated with protection against transmission to the fetus. Pentamer-specific antibodies are 100 to 1000-fold more potent than those targeting gH/gL/gO or gB for the neutralization of epithelial and endothelial cell infection (A. Macagno et al., 2010).

2.4 The aim of the study

The viral proteins are glycosylated and their glycans have a huge impact on the immunological features of the antigen. Usually, oligosaccharides hamper by steric hindrance the recognition of protein epitopes by the antibodies, lowering the immune response elicited by the glycoprotein. In case of gB, the two most immunogenic domains (AD4 and AD5) contain 1 and 2 *N*-glycosylation sequons, respectively, which could potentially prevent the antibody binding (Fig 2.8).

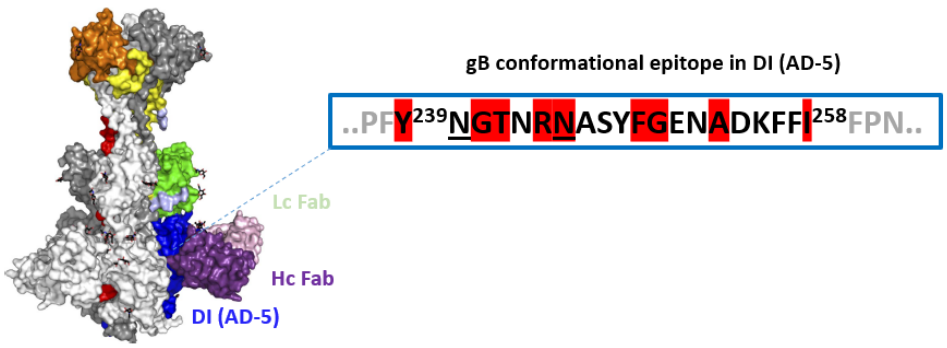


Figure 2.8. Crystal structure of the deglycosylated gB-1G2 Fab complex with the DI conformational epitope primary sequence. From *Structure of HCMV glycoprotein B in the postfusion conformation bound to a neutralizing human antibody.* (Chandramouli et al., 2015).

Then, the composition of oligosaccharides is relevant because if presenting either non-self monosaccharides or non-self configurations, it can trigger allergic reactions in patients. Equally important are the glycan profile of the protein and the site-specific information, both needed to undertake antigen design studies. In addition, the knowledge of the total glycan profile allows to explain the antigen pharmacokinetic In fact, the absence of sialic acid on complex-type *N*-glycans (Fig. 1.1) can negatively affect this parameter because it leaves exposed nonreducing mannose, fucose and N-acetylglucosamine residues, which promote the recognition of the glycoprotein by the mannose receptors (MRs) on liver Kupffer and endothelial cells, which are one of the avenues of serum clearance (Rocamora et al., 2023). Contrariwise, high-mannose-type and single-GlcNAc-type *N*-glycans are more effective than complex-type *N*-glycans in triggering activation and maturation in both mouse and human myeloid dendritic cells (mDCs); this because terminally sialylated complex-type *N*-glycans provide shielding against the activation of pattern recognition receptors (*e.g.,* mannose or another C-type lectin) on mDC surfaces.

Nowadays, mass spectrometry is the method of choice for the study of post-translational modifications; indeed, protein glycans are usually too flexible to be seen with other techniques (such as X-ray crystallography or cryo-EM). The aim of the study is to develop mass spectrometry based analytical methods which allow the glycans profile monitoring on glycoprotein antigens, using as starting antigens HCMV gB and Pentamer.

Chapter 3 – Mass-spectrometry as tool for protein glycans pattern analysis

3.1 Introduction to mass spectrometry

The birth of mass spectrometry (MS) is commonly attributed to the physicist J.J. Thomson with his discovery of the electron (1897), using an electric field inside a

cathode ray tube. This success led him to develop a crude ‘‘mass spectrograph’’ to measure the atomic weights of elements. Then two Thomson’s students, F.W. Aston and A.J. Dempster, developed more advanced versions of the initial instrument, giving birth to the first sector field mass spectrometer (1918), which brought Aston to the discovery of ^{20}Ne and ^{22}Ne isotopes, proving the existence of stable isotopes (John R Yates). In the same period electron ionization (EI), the only hard ion source known to date, was first described by Arthur J. Dempster in the article of "A new method of positive ray analysis’’. During World War II a preparative sector mass spectrometer (named Calutron) was used for separating the isotopes of uranium developed by Ernest O. Lawrence. In 1953 W. Paul developed the first quadrupolar analyzer (J. R. Yates, 2011).

In 1968, Malcolm Dole at Northwestern University found that large gas-phase ions (‘macroions’) could be transiently generated by “*electrospraying a dilute solution into an evaporation chamber containing nitrogen, and by allowing the volatile solvent to evaporate from the tiny drops so produced”*. The ‘’electrospray revolution’’ in the late 1980s dramatically increased the utility of mass spectrometry. Developed by John Fenn at Yale University, electrospray ionization (ESI) finally allowed the mass spectrometric analysis of biological macromolecules, including proteins, and was rapidly adopted in laboratories around the world. In 2002 J. Fenn became a Nobel laureate for giving “electrospray wings to molecular elephants’’ (J. Griffiths, 2008). Another soft ionization technique was discovered by K. Tanaka, which first reported in 1985 an ionization method, with higher sensitivity using a small organic compound as a matrix, that they named matrix-assisted laser desorption/ionization (MALDI); this innovative ionization technique replaced the fast atom bombardment (FAB) ionization, developed a few years earlier by Michael Barber. Tanaka shared the Nobel prize with Fenn in 2002. Recently a new member of the FTMS family (the other one is Fourier transform ion cyclotron resonance, FTICR) was developed by A. Makarov: the Orbitrap analyzer. It was presented at a conference of the American Society for Mass Spectrometry in 1999, it quickly made its debut in mainstream MS in 2005 as

an accurate and compact mass detector. In 2023, the proteomics community was electrified by the launch of the first conceptually novel mass analyzer in a commercial instrument since 2005 at the annual meeting of the American Society for Mass Spectrometry, the so-called Asymmetric Track Lossless (Astral).

At the end of the story, we can say that mass spectrometers are in continuous evolution, with always more outstanding performances.

Mass spectrometers are instrument characterized by three main components: ion source, analyzer and detector, the last two must necessarily be under strong vacuum (10^{-5} - 10^{-10} Torr) (Fig. 3.1); these kinds of instruments are able to study gas phase ions (produced by ion source) by using an analyzer which exploits an electric field, a magnetic field or a radiofrequency (at least one of them) to measure the ions mass/charge ratio (*m/z*, a dimensionless ratio), at the end of the ion path a detector will measure the intensity of each ion. The kind of information that can be obtained from MS analysis can be both quantitative (molecular mass) and qualitative (molecular structure).

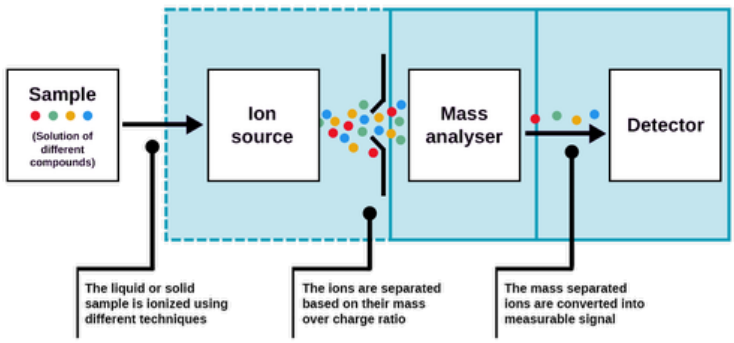


Figure 3.1. Schematic representation of a mass spectrometer.

According to F. W. McLafferty, several properties (with some exceptions) of mass spectrometers begin with the letter ‘’s”: specificity, selectivity, sensitivity, speed, stoichiometry.

Under the same experimental conditions, the same molecule will always produce the same mass spectrum, even when present in very small amounts (specificity and sensitivity). Mass spectrometry allows to correctly determine an analyte without interference from other compounds (selectivity); the time needed to record a mass spectrum is of the order of milliseconds (speed). It is possible to calculate the molecular formula of a chemical species (stoichiometry) and with the same instrument is possible to study different classes of molecules (versatility).

A mass spectrum is a two-dimensional plot in which on the x-axis there is the ion *m/z* value, while on the y-axis there is the ion count (intensity). An important performance factor of mass spectrometers is the resolving power, which is an intrinsic feature of the analyzer; this parameter is linked to the concept of resolution (Δm), which can be defined as the smallest difference in *m/z* that can be separated for a give signal. Resolving power can be defined both referring to two adjacent peaks and to a single peak (Fig. 3.2).

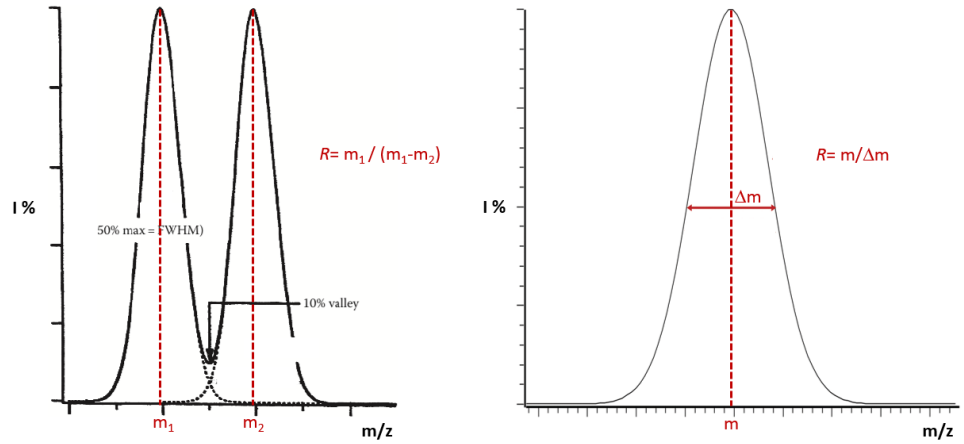


Figure 3.2. Definition of resolving power.

When using the single peak definition, the resolving power can be calculated using the peak full width at different intensity values, for this reason it must be always specified the height (in terms of percentage) at which the resolving power is obtained. A common full width used is the full width at half maximum (FWHM).

When the valley between two adjacent peaks is less than 10 % the resolution can be considered good.

The resolving power depends on the mass value at which it has been defined; moreover, different analyzers show a different relationship between resolving power and mass. In the case of FTICR there is R a $(m/z)^{-1}$, while for Orbitrap R a $(m/z)^{-1/2}$, for TOF R=constant (Fig. 3.3) (Lossl et al., 2014). Increasing resolution does not affect the relative intensity of the peaks, however, increased settings of resolving power are often obtained at the cost of transmission of the mass analyzer, thereby reducing the sensitivity (absolute signal intensity is reduced) (Lossl et al., 2014).

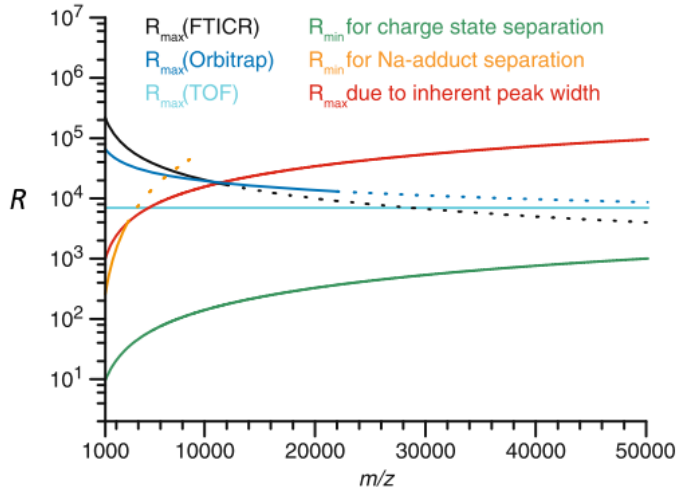


Figure 3.3. Theoretical (dashed) and experimental mass resolution plotted against *m/z*. R-versus-*m/z* plots for the three mass analyzers (FTICR, Orbitrap and TOF). From *Boundaries of mass resolution in native mass spectrometry*. (Lossl et al., 2014).

The nominal mass of an element is defined as the integer mass of its most abundant naturally occurring stable isotope. The nominal mass of a molecule is the sum of the nominal masses of its constituent elements.

The monoisotopic mass of an element is defined as the exact mass of the most abundant isotope of an element. The monoisotopic mass of a molecule is the sum of the monoisotopic masses of the elements in its empirical formula; the monoisotopic mass is not necessarily the naturally occurring isotope of lowest mass.

The relative atomic mass (or atomic weight) is calculated as the weighted average of the naturally occurring isotopes of an element, using the abundances of the isotopes as weights in the average. The relative molecular mass (M_r), or molecular weight, is calculated from the relative atomic masses of the elements contributing to the empirical formula.

The difference between the measured accurate mass (obtained experimentally) and calculated exact mass gives the absolute mass accuracy. Instead of stating the absolute mass accuracy in units of Da, it can also be given as relative mass accuracy, *i.e.*, absolute mass accuracy divided by the mass it is determined for. The relative mass accuracy is normally given in parts per million (ppm). Mass accuracy strongly depends on various parameters such as resolving power, scan rate, canning method, signal to noise ratio of the peaks.

Isotopes of the same element have equal atomic number but different mass number (because have different neutrons number). Even if the analyte is chemically perfectly pure, it represents a mixture of different isotopic compositions; therefore, a mass spectrum normally superimposes the mass spectra of all isotopic species involved. For example, ^{13}C has a natural abundance of 1.1 %, thus if we consider 1000 methane molecules there will be 11 molecules containing ^{13}C and 989 containing ^{12}C; the ratio of relative intensity of the peaks at *m/z* 16 and *m/z* 17 is defined by the ratio 989/11,

otherwise, 100/1.1. In a more general way, the ratio between ^{13}C and ^{12}C can be written as $r=c/(100-c)$, where c is the abundance of ^{13}C. Then, the probability to have only ^{12}C in a molecular ion M consisting of w carbons is given by:

$$P_M = \left(\frac{100-c}{100}\right)^w$$

While the probability of having exactly one ^{13}C atom in an ion with w carbon atoms is:

$$P_{M+1} = w\left(\frac{c}{100-c}\right)\left(\frac{100-c}{100}\right)^w$$

It is very helpful to read out the P_{M+1}/P_M ratio from a mass spectrum to calculate the approximate number of carbon atoms, provided no other element contributing to M+1 is present. Increasing the carbon number, the X+1 peak reaches the same intensity as the X peak and at higher carbon number, it becomes the base peak of the pattern (most intense peak) (Fig. 3.4).

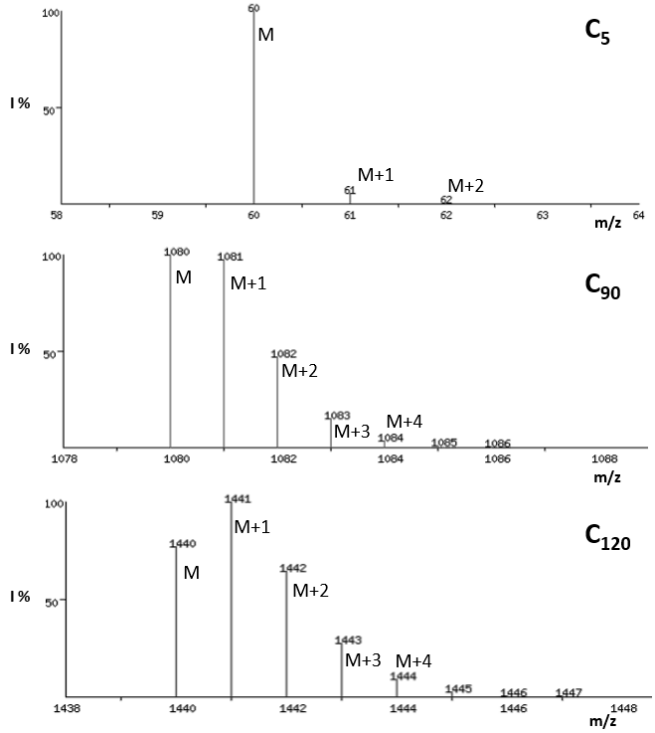


Figure 3.4. Calculated isotopic patterns for carbon.

All spectrometers require mass calibration before they enter in use. However, proper procedures and the number of required calibration points may largely differ between different types of mass analyzers. Typically, a robust mass calibration necessitates several peaks of well-known *m/z* values evenly distributed over the mass range of interest. These are supplied from a well-known mass calibration compound.

There are different mass reference compounds available; for example, perfluorokerosene (PFK) is a well-established mass calibration standard in electron ionization. It provides evenly spaced $C_xF_y^+$ fragment ions over a wide mass range; PFK mixtures are available from low-boiling to high-boiling grades which may be used up to 700-1100 *m/z*. Another common calibration standard in EI is

perfluorotributylamine (PFTBA, also termed FC-43), which yields peaks up to 614 *m/z*. Furthermore, cluster ions are frequently employed for mass calibration as they provide series of ions that are equidistant on the *m/z* axis; for example, CsI can be used for mass calibration in both MALDI and ESI-MS. This salt yields cluster ions of the general formula $[Cs(CsI)_n]^+$ in positive-ion and $[I(CsI)_n]^-$ in negative ion mode. Actually, in ESI-MS the CsI has been replaced by standards which consists of a mixture of different molecules (calmix), like caffeine, oligopeptide (such as MRFA, Met-Arg-Phe-Ala), Ultramark 1621 and n-butylamine.

The mass spectrometer calibration can be either external or internal. The external mass calibration is performed by recording a mass spectrum of the calibration compound and subsequent correlation of experimental *m/z* values to the mass reference list. Thus, the mass spectrum is recalibrated by interpolation of the *m/z* scale between the assigned calibration peaks to obtain the best match.

The internal mass calibration is done by introducing the calibration compound with the analyte prior to analysis; the highest mass accuracy is achieved through this kind of calibration because drift in instrumental parameters and space charge effects are effectively compensated; nevertheless, a separate inlet for introducing the calibration compound is preferred because the mixing of the calibration compounds with the analyte requires some operational skills (Romson et al., 2021).

Another operation that can be performed before mass spectrometric analysis is the tuning, which consists of the optimization of the signal of an analyte of interest by adjusting several mass spectrometer parameters that affect signal processing, as well as voltages and currents associated with ion source components, the mass analyzer, and detector; it is done after calibration by injecting only the analyte with known molecular mass.

3.1.2 Fragmentation techniques

The term tandem mass spectrometry, or briefly tandem MS (or MS/MS), encompasses the numerous techniques where mass-selected ions are subjected to a second mass

spectrometric analysis. Tandem mass spectrometry comprises the acquisition and study of the spectra of ionic products or precursors of m/z-selected ions, or of precursor ions of precursor ions of a selected neutral mass loss. A mass spectrometer designed for MS/MS requires to incorporate at least two stages of *m/z* analysis, often referred to MS^1 and MS^2, respectively.

There are two basic instrumental concepts for MS/MS. The first is tandem mass spectrometry in space, in which precursor and product ions spectra are recorded using spatially separated *m/z* analyzers (QqQ, QqTOF, QqOrbitrap, TOF/TOF). Specific *m/z* separation is performed so that in one section of the instrument ions are selected, then dissociated in an intermediate region, and the products thereof are finally transmitted to a second analyzer for mass analysis. The second approach, tandem mass spectrometry intime, employs a single *m/z* analyzer (QIT, LIT, FT-ICR) that may be operated by executing the discrete steps of ion selection, fragmentation and product ion analysis in the very same plate but sequentially in time.

In principle, both instrumental concepts can be expanded to allow for multiple-stage mass spectrometry (MS^n); however, with tandem mass spectrometry in space the presence of additional analyzers and fragmentation cells is needed if we want to increase the number of MS/MS experiments, for this reason the most common configuration for tandem MS in space is the MS^2. On the contrary, tandem MS in time allows to make MS^n experiments without adding further analyzers but using always the same single device (Fig. 3.42).

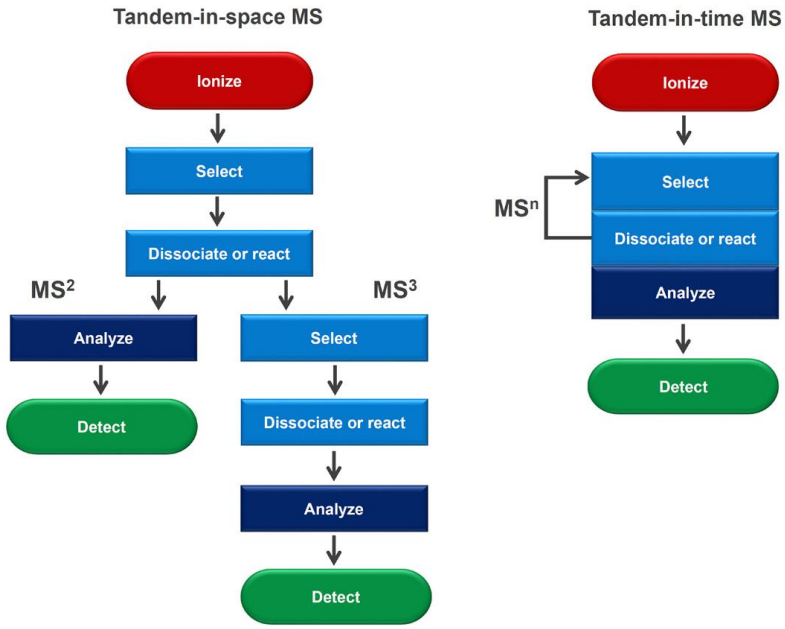


Figure 3.42. Comparison of tandem-in-space and tandem-in-time MS.

The success of any tandem mass spectrometric experiment depends on the occurrence of some kind of reaction between the consecutive steps of precursor ion selection and product ion analysis. This demands that ions entering the zone and/or period for reaction either possess or receive sufficient internal energy for doing so; alternatively, a partner for reaction may be presented. There are various ways to fragment the parent ion and depending on the method chosen different types of MS/MS spectra are collected. The parent ion dissociation can be induced by collision with an inert gas, by interaction with electrons or by interaction with a radiation.

The most prominent fragmentation technique is collision-induced dissociation (CID), called also collisionally activated dissociation (CAD). CID is realized by passing an ion beam through a collision cell where the collision gas (He, N_2, Ar) is set to a pressure considerably above that of the surrounding high vacuum; the cell consists of a RF-only multipole, which acts as a field free region. The ion is accelerated towards the gas with a kinetic energy of some kiloelectronvolts; after the collision with the

gas, the energy is redistributed in the vibrational modes of the ion before its fragmentation, which occurs involving the weakest bonds. The CID process can be regarded as a two-step process: first, the activated species AB^{+*} is formed, second, after randomization of the internal energy has occurred, AB^{+*} will dissociate along any fragmentation pathway available at this specific level of internal energy.

I) $AB^+ + N \rightarrow AB^{+*} + N'$ (fast step, collision)

II) $AB^{+*} \rightarrow A^+ + B$ (energy dissipation and fragmentation)

The amount of internal energy transferred to the ion during the inelastic collision with a gas molecule is given by the equation:

$$E_{CM} = (m_N / (m_{AB} + m_N)) E_{LAB}$$

E_{LAB} represents the kinetic energy of the ion as received from passing through an acceleration stage; this is the parameter meant to be the collision energy. Not all the kinetic energy is converted as internal energy, it hinges on the masses of the ion and inert gas. The bigger the ion is the less CID will be effective, while if we increase the gas mass, more internal energy will be transferred during collision. This fragmentation technique works well for not too big molecules; CID provides an ergodic fragmentation mechanism, which means that the dissociation rate is low with respect to the rate of redistribution of the excitation energy among all of the internal modes. The increase of the collision energy from 100 eV to 1 keV (mechanism termed high-energy collision dissociation, HCD), leads to the cleavage of bonds that would never break in the classic CID conditions (*e.g.* peptide bonds on glycopeptides), yielding to higher level of information; indeed, when a glycopeptide undergoes CID, only more labile glycosidic bonds will break instead of the peptide bonds (Fig. 3.43).

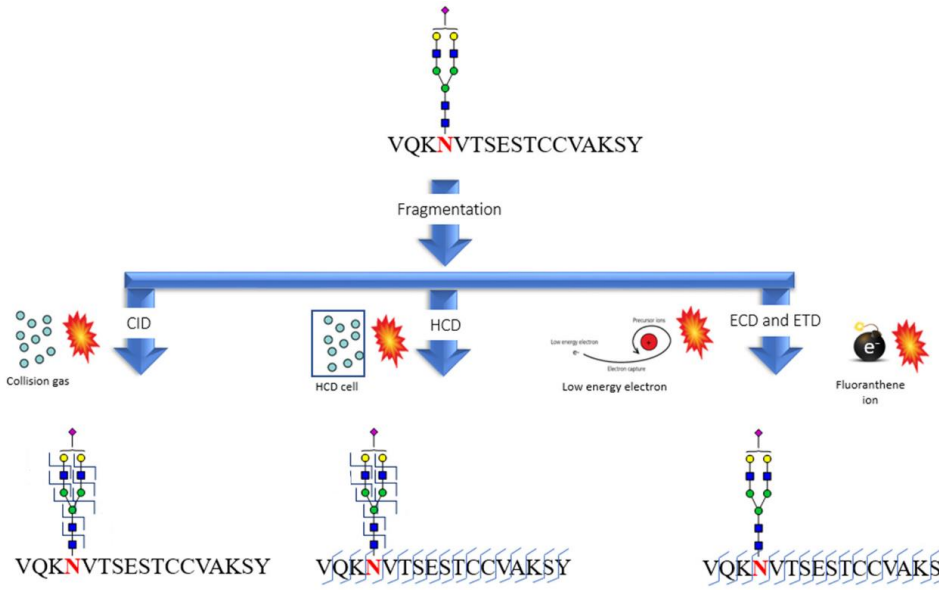


Figure 3.43. Mechanism of collision induced dissociation (CID), high-energy collision dissociation (HCD), electron-capture dissociation (ECD), and electron transfer dissociation (ETD) fragmentation applied to a glycopeptide. From *Protein Glycosylation Investigated by Mass Spectrometry: An Overview.* (Illiano et al., 2020).

The peptide fragment ions are appointed according a common nomenclature, introduced by Roepstorff and Fohlmann in 1984 (Roepstorff et al., 1984) and modified by Biemann in 1988 (Biemann, 1988). According to this convention the three possible cleavage points of the peptide backbone are called a_n, b_n and c_n when the charge is retained at the N-terminal fragment of the peptide and x_n, y_n and z_n when the charge is retained by the C-terminal fragment; the numbering indicates which peptide bond is cleaved counting from the N- and C-terminus respectively, and thus also the number of amino acid residues in the fragment ion. Protonated peptide precursor ions activated under CID conditions fragment primarily at amide bonds along the peptide backbone; this leads to a series of b and y ions which contain, respectively, the N-terminus and C-terminus amino acid residues (Fig. 3.44).

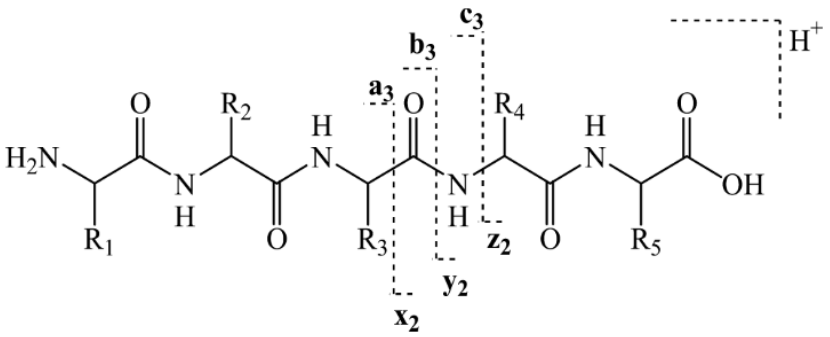


Figure 3.44. Roepstorff-Fohlmann-Biemann nomenclature for peptide fragment ions. From *Mass spectrometry of peptides and proteins.* (Wysocki et al., 2005).

In 1988 Domon and Costello introduced systematic nomenclature for carbohydrate fragmentations in MS/MS spectra of glycoconjugates (Domon et al., 1988). According to this nomenclature, when the charge is retained on the carbohydrate portion, fragments are designated as A_i, B_i and C_i, where i represents the number of the glycosidic bond cleaved, counted from the non-reducing end (the case of branched oligosaccharides will be discussed below). On the other hand, ions containing the aglycone (or the reducing sugar unit in the case of oligosaccharides) are labeled as X_j, Y_j and Z_j, where j is the number of the interglycosidic bond counted from the aglycone (Fig. 3.45).

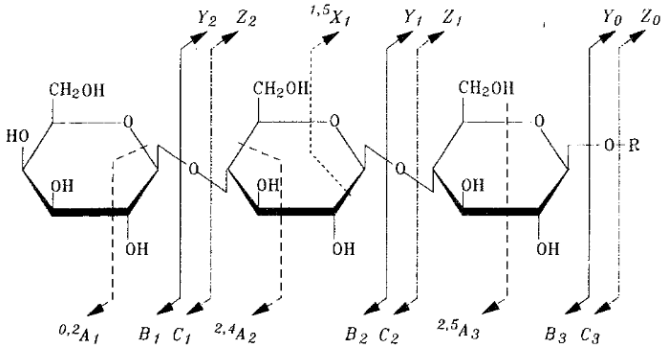


Figure 3.45. Domon-Costello nomenclature for oligosaccharide fragment ions. From *A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates.* (Domon et al., 1988).

Akin to CID is surface-induced dissociation (SID), where ions undergo a collision with a surface target often composed of a fluorinated alkanethiol self-assembled monolayer on gold; this surface provides a large effective mass for collision of projectile ions. The fluorocarbon chains are relatively rigid so that they do not severely dampen the energy of the colliding projectile, and the fluorocarbon resists facile electron transfer from the metal surface to the incoming ions. SID is a fast, high-energy deposition process, and thus allows for better characterization of the oligomer assembly, because faster structurally informative dissociation pathways can outcompete the slower monomer unfolding (rearrangement) dissociation pathway observed in CID (Fig. 3.46) (Stiving et al., 2019).

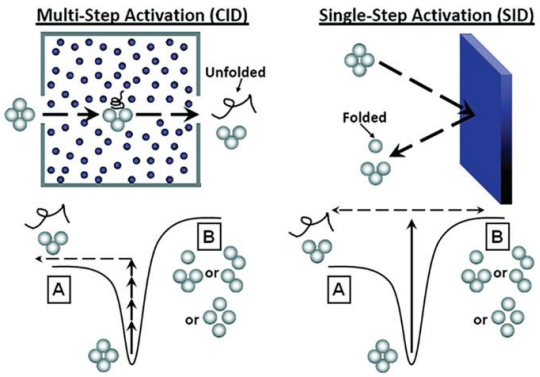


Figure 3.46. Simplified representation of the pathways a protein complex may take when undergoing CID (left) or SID (right). CID involves multiple low-energy collisions (top left) that generally result in dissociation via the lowest energy pathway (bottom left). SID involves a single-step, high energy deposition via collision with a surface (top right) and typically results in dissociation via faster, alternative dissociation pathways and dissociation products form that are often compact and reflective of the native structure (bottom right). From *Surface-Induced Dissociation: An Effective Method for Characterization of Protein Quaternary Structure.* (Stiving et al., 2019).

SID dissociation of a streptavidin homotetramer initially results in the cleavage of the smaller dimer-dimer interfaces yielding C2 dimers, a structurally informative pathway that is not seen by the common CID activation method. SID dissociation at higher energies results in primary cleavage to dimers and secondary cleavage of the larger monomer-monomer interface within the C2 dimer to produce monomers (Fig. 3.47) (Quintyn et al., 2015).

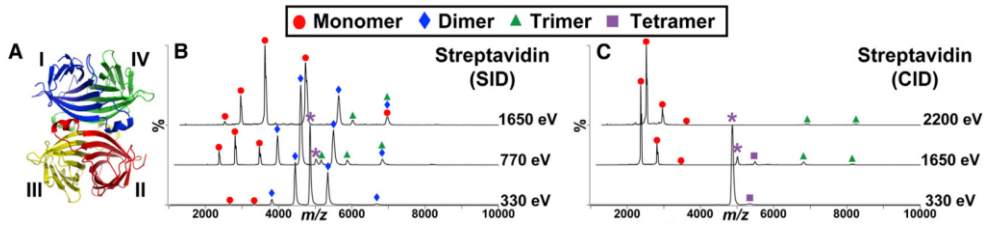


Figure 3.47. Nano-electrospray SID (left panel) and CID (right panel) mass spectra of the charge-reduced +11 precursor of streptavidin. From *Surface-induced dissociation of homotetramers with D2 symmetry yields their assembly pathways and characterizes the effect of ligand binding.* (Quintyn et al., 2015).

Alternative fragmentation techniques are electron capture dissociation (ECD), developed in 1998, and electron transfer dissociation (ETD), developed in 2004 (McLuckey et al., 1998), (Zubarev, 2004). It is important to underline that these ionization methods only work on multiply charged cations. ECD involves the direct introduction of low energy electrons (<0.2 eV), generated by a heated tungsten filament, for the activation of the precursor ion (Zubarev, 2004). On the other side, ETD uses anions with sufficiently low electron affinities (like fluoranthene) as vehicles for delivering one electron to the parent ion (Syka et al., 2004). Both ECD and ETD have been described as non-ergodic processes, where the energy is not redistributed over the whole molecule and the weakest bonds are not preferentially broken following ion activation. In fact, the dissociation of the ion is restricted to specific protonation sites where the electron is captured. When a peptide undergoes ETD, it generates radical cation fragment ions which are of c-type and z-type.

The advantages of ECD and ETD are about the possibility to identify post-translational modifications which remain difficult for labile bonds such as *O*-glycosylation, phosphorylation, sulfation, γ-carboxylation. If the primary electrons are more energetic (about 20 eV) than thermal (<0.2 eV), there is a substantial chance to promote detachment of one of them from an anion; this process can cause backbone

cleavage due to the formation of opposite charge sites and is called electron detachment dissociation (EDD). EDD is useful to study negatively charged analytes not amenable for ECD (or ETD), like peptides bringing sulphates, phosphates or acid groups.

There are fragmentation techniques which exploit the matter-radiation interaction too, which are generally called photodissociations. In infrared multiphoton dissociation (IRMPD) the dissociation event is promoted by infrared radiation due to a multiple photon absorption process (Bianco et al., 2015); ions are excited to high vibrational states and dissociation occurs when the accumulated internal energy exceeds the dissociation energy threshold. IRMPD is usually promoted by an infrared laser, the most common laser source is a continuous wave CO_2 laser, operating at $\lambda = 10.6$ μm; medium to large molecules have vibrational frequencies in this range. The fragmentation pattern obtained with IRMPD (mainly b and y ions) depends on the power of the laser (from 1 to 50 W), the time of irradiation (from ms to s) and the molecule absorption coefficient. Compared to CID, IRMPD allows better detection of low *m/z* values, by breaking both labile and strong bonds. In UVPD instrument, the ultraviolet light (usually with $\lambda = 198$ nm) is used for photodissociation, so that the absorption of one single photon is sufficient to fragment the precursor ion. The increase in the ion size causes a decrease of dissociation rates; for this reason, efficient dissociation of large ions in a mass spectrometer requires deposition of large amounts of internal energy, like achieved by using multiphoton excitation (IRMPD).

3.2 Glycans site specific analysis

Site-specific *N*-glycans analysis (or microheterogeneity analysis) of highly glycosylated proteins is generally performed by using a proteomic bottom-up approach; according to this method, proteins of interest are digested with an enzyme such as trypsin, and the resulting “tryptic peptides” are analyzed by electrospray ionization (ESI). The bottom-up approach is especially useful for identifying proteins,

because tryptic peptides are readily solubilized and separated, tasks that are considerably more difficult for the parent proteins. In addition, many tryptic peptides can be readily analyzed by mass spectrometry analysis, providing useful fragmentation ladders that often yield sufficient information to identify the parent protein (Chait, 2006).

Glycoprotein digestion can be performed on a gel ('in-gel' digestion), in solution ('in-solution' digestion) or on a filter (Filter Aided Sample Preparation, FASP). The first digestion involves solubilization of proteins with detergents, separation of proteins by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis and digestion of the gel-trapped proteins. The second method is detergent-free, comprising protein extraction with strong chaotropic reagents such as urea and thiourea, protein precipitation and digestion of proteins under denaturing conditions. Advantages of in-gel digestion include its robustness against impurities, which interfere with digestion, but the gel may prevent peptide recovery and the method cannot easily be automated. In-solution digestion is more readily automatable and minimizes sample handling, but the proteome may be incompletely solubilized, and digestion may be impeded by interfering (Wisniewski et al., 2009). The FASP is an innovative sample preparation method, it was introduced by J. R. Wisniewski in 2009; according to this technique, a common ultrafiltration device acts as a 'proteomic reactor' for detergent removal, buffer exchange, chemical modification and protein digestion. The key feature of the method is the ability of the filter membrane to retain high-molecular-weight substances (proteins and DNA) and to allow through low-molecular weight substances (impurities and digested peptides); the main drawback of this method concerns the retention of the certain analytes (*e.g.* glycopeptides) by non-specific interaction with the membrane.

Another method to analyze glycoproteins consists of the top-down approach, where intact protein ions are introduced into the gas phase by ESI and are subsequently fragmented in the mass spectrometer; the advantages of this approach are almost null sample preparation and information loss limitation (almost 100 % coverage)

(Catherman et al., 2014). However, this approach has been applied only to low glycosylated proteins and it is less suitable for the quantitation of all the glycoforms on each sequon (Going; et al., 2016).

Liquid Chromatography-Mass Spectrometry (LC-MS) can be used to analyze glycoproteins oligosaccharide pattern by site specific analysis. In this case the glycan remains linked to a peptide, and the mixture of glycopeptides and peptides obtained by protease digestion is purified by LC (usually using a reverse phase C18 column) so that only a single glycopeptide per time undergoes ionization and reaches the detector, thus enabling the elucidation of its structure by tandem mass spectrometry (MS/MS) experiments (Ieva Bagdonaite et al., 2022). More in detail, the glycopeptide parent ion is fragmented by a certain mechanism that depends on the type of instrument used, and the fragment ions produced are used both to confirm the presence of a glycopeptide and to rebuild peptide and glycan primary structures (Dalpathado et al., 2008). Moreover, the use of a proteomic software implemented with a glycans mass database (*e.g.*, PEAKS Studio) enables to make an educate guess about the glycan structure. Of note, as important requirement it is necessary to know which type of cells were used to express the glycoproteins to minimize structures misinterpretation (Zhang et al., 2012). Importantly, the ionization efficiency of glycopeptides is lower compared to that of non-glycosylated peptides due to the presence of the bulky oligosaccharide portions, while charged oligosaccharides like those containing sialic acid cause further decrease in the glycopeptide ionization efficiency especially if the analysis is in positive ion mode (Kolarich et al., 2012; Stavenhagen et al., 2013). Nevertheless, glycopeptide protonation during the ionization phase occurs on the peptide portion, so the signal intensity is affected mainly by amino acid composition; this is an important observation which establishes the starting point for a correct site-specific analysis assessment of viral antigens (Tajiri et al., 2005; Wada et al., 2007). Glycans site-specific analyses of several viral envelope proteins have been reported over the last few years thanks to technological improvements in both mass spectrometry and bioinformatic fields; indeed, structural

knowledge of the immunologically important glycans displayed on the surface of glycoproteins has remained elusive until a few years ago, due to difficulties in seeing them via X-ray crystallography (highly flexible structures).

For example, site-specific *N*-glycosylation analysis of HIV-1 BG505 SOSIP.664 showed that this antigen has 28 *N*-glycosylation sites per monomer, all with an almost quantitative occupancy (the site with the lowest occupancy, N625, had about 60 % of glycosylation) (Fig. 3.46). (Cao et al., 2017).

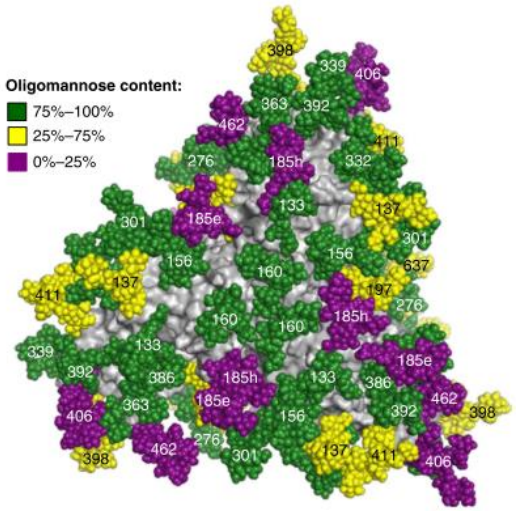


Figure 3.46. Mapping of site-specific glycosylation onto the structure of BG505 SOSIP.664 trimer. From *Global site-specific N-glycosylation analysis of HIV envelope glycoprotein.* (Cao et al., 2017).

Recently this approach was applied for the first time also for the semi-quantitative site-specific glycosylation analysis of glycoengineered bacterial vaccines carrying intact O-antigen polysaccharides. More in detail, glycoconjugates carrying polydisperse *E. coli* O-antigen polysaccharides have been characterized by LC-MS/MS (Fig. 3.47) (Di Marco et al., 2024).

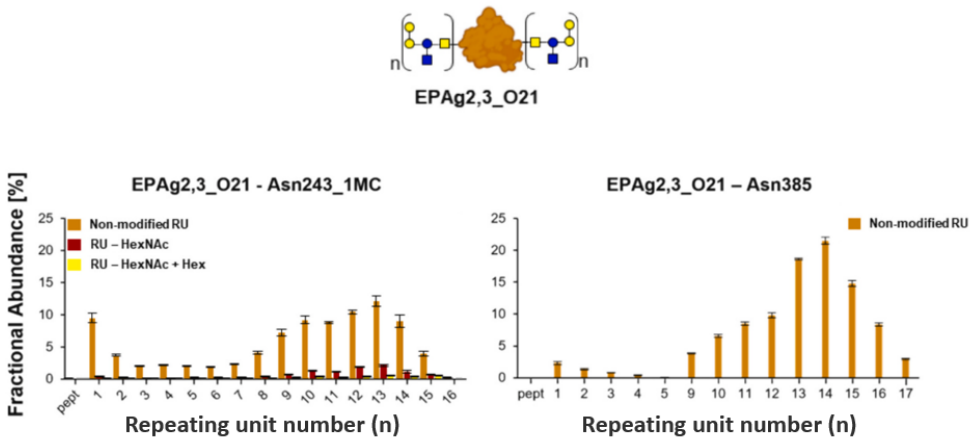


Figure 3.47. Bar charts reporting the site-specific semi-quantitation of the glycopeptide repeating unit (RU) distribution and their related modifications of the glycoconjugate EPAG2,3_O21. 1 MC stands for a peptide with one missed cleavage. From *Comprehensive characterization of bacterial glycoconjugate vaccines by liquid chromatography - mass spectrometry.* (Di Marco et al., 2024).

3.3 Released glycans analysis

The notion that glycans should be studied as a totality (Glycomics), as well as simply one glycan or glycoprotein at a time, arose when it became apparent that glycans form patterns on cells that change during development, cancer progression, infection and many other diseases. Thus, the term “Glycomics” was coined to describe the many aspects of glycobiology that can be understood only with a systems-level analysis. No systems-level analysis of a biological process is complete without interrogating the glycome and the glycoproteome in addition to the genome, transcriptome, and proteome.

The starting material in a Glycomics analysis can be glycoproteins embedded in SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) gels, whole cell lysates, homogenized tissue, enriched membranes, or serum and other body fluids. For high-throughput analyses of the glycome, intact *N*-glycans are most often released from glycoproteins using an amidase (peptide N-glycosidase F, PNGase F). PNGase F cleaves the linkage between the core GlcNAc and the asparagine residue of all classes of *N*-glycans, with the exception of specific *N*-glycans found in plant and insect glycoproteins that contain fucose α(1,3) linked to the GlcNAc residue attached to the protein. For the analysis of the latter, PNGase A, an enzyme extracted from almond emulsion, may be used to release all fucosylated *N*-glycans. Other enzymes cleave between the two GlcNAc residues within the chitobiose core, the most famous are: endoglycosidase D (Endo D), which releases some high-mannose type oligosaccharides (those for which the Man-a-l,3 residue is not substituted at C-2), Endo H, which selectively cleaves oligomannose- and hybrid-type structures and various types of Endo F (Endo F1, Endo F2, Endo F3), where Endo F3 cleaves within the chitobiose core of N-linked fucosylated-biantennary and triantennary complex oligosaccharide (O'Neill, 1996). Before treatment, denaturation with or without trypsin digestion can be used to relax the three-dimensional (3D) structure of the protein and improve enzyme accessibility. For O-linked glycans the release method of choice is the reductive β-elimination of O-linked glycans from serine and threonine; however, nowadays methods which allow *in situ* labeling of the oligosaccharide reducing end are available (*e.g.* 1-phenyl-3-methyl-5-pyrazolone labeling). Very recently, an oxidative strategy making use of NaClO to release free reducing N-and O-glycans from glycoproteins and glycosphingolipids was shown to be good alternative for glycan analysis and a promising methodology to generate a library for glycan array preparation (Fig. 3.48) (Song et al., 2016).

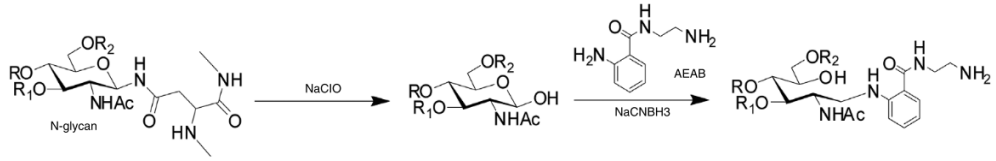


Figure 3.48. Chemical scheme of NaClO treatment. From *Oxidative release of natural glycans for functional glycomics*. (Song et al., 2016).

Fluorescent tags increase the sensitivity and limit of quantitation and detection. Permethylation and peracetylation, when all the mobile protons (present on hydroxides, carboxyls, and sometimes amides) in a glycan are substituted by alkyls (*e.g.*, -methyl) or esterified (*e.g.*, -acetyl) converts the glycans from being hydrophilic to hydrophobic, greatly improving sensitivity and linkage determination by MS-based analyses.

3.4 Monosaccharides analysis

After total hydrolysis of a glycan into its monosaccharide constituents, colorimetric reactions (like phenol-sulfuric acid method) can be used to determine the total amount of hexose, hexuronic acid, or hexosamine in a sample. These approaches only require common reagents and a spectrophotometer, but determination of total glycan content may not always be accurate because of variations in the sensitivities of different linkages to hydrolysis, variations in the degradation of individual saccharides, or a lack of specificity and/or sensitivity in the assays. Since the early 1960s, a variety of gas chromatography (GC) methods have been developed to quantify monosaccharides.

The depolymerization procedure consists in acid catalyzed solvolysis, using either water or other solvents. This is a very critical first step as ideal conditions should guarantee 100% depolymerization and 0% monosaccharide degradation. These conditions are hardly met because of the reluctance of the glycosidic linkage to be

cleaved (Q. C. Wang et al., 2016). The level of reluctancy depends on the monosaccharide considered and on the type of sugar ring isomer; furanic forms are more prone to lysis than the pyranose isomers due to higher carbocation stabilization. Among sugars in the pyranose form, aminosugars and uronic acids are more resistant than hexoses, whereas deoxy-hexoses and ketoses are more labile. This behavior finds its explanation by following the hydrolytic mechanism according to which, the protonation of the exocyclic oxygen leads to the generation of the cyclic oxonium ion intermediate (Fig. 3.49, structure on the right).

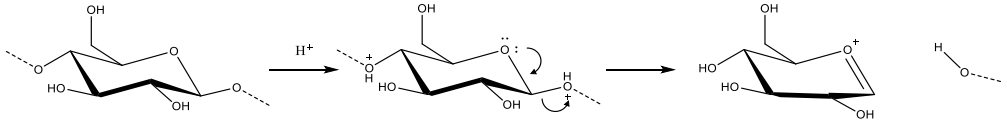


Figure 3.49. Partial reaction scheme of glycan hydrolysis.

As for aminosugars, the amino group loses the acetyl (or acyl) function usually present to be converted into an ammonium group; similarly, protonation for uronic acid (Fig. 3.50) takes place throughout a different mechanism. Nevertheless, both types of sugars have a net positive charge that hinders the protonation of the exocyclic oxygen by electrostatic repulsion, so that cleavage of the glycosidic linkage is inhibited and occurs at a lower rate.

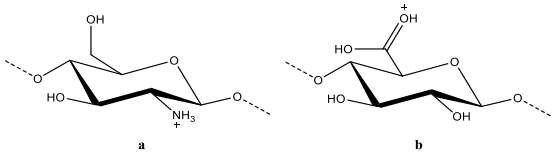


Figure 3.50. a) Protonated amino sugar. b) Protonated uronic acid.

This mechanism does not operate for the other monosaccharides. The lability of deoxysugars depends on the minor tensions experienced by the ring during conversion into the half-chair conformation, that is necessary for the formation of the reaction intermediate, the oxonium ion (Fig. 3.49). As for ketoses, the stabilization of the cyclic oxonium ion by the alkyl group that replaces the hydrogen at the anomeric carbon, facilitates cleavage of the glycosidic linkage. Consequently, whenever hydrolysis proceeds, some glycosidic linkages are cleaved before others (Uhliarikova et al., 2021).

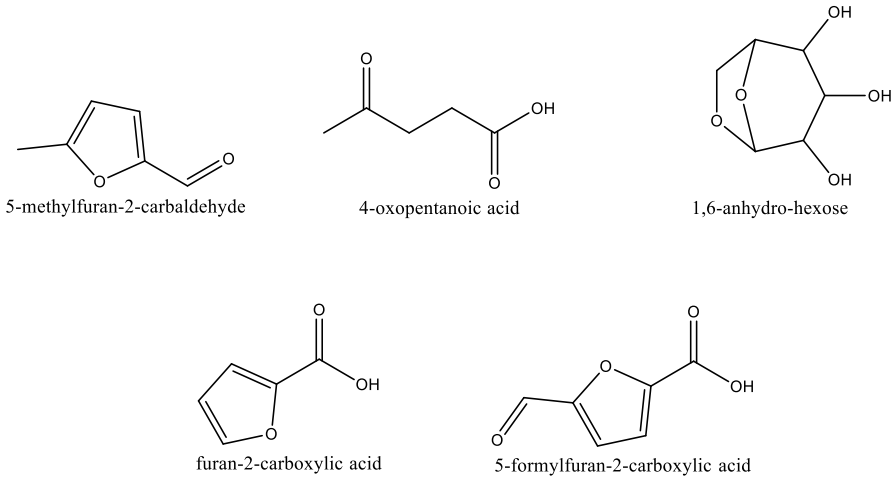


Figure 3.51. Monosaccharides main degradation products.

The stability of the released monosaccharides at given hydrolysis conditions must also be considered as they can be degraded to various furan derivatives through reactions involving the free aldehyde function (furan-2-carbaldehyde for pentoses; 5-methylfuran-2-carbaldehyde and 4-oxopentanoic acid and 1,6-anhydro-derivatives for hexoses; and furan-2-carboxylic acid and 5-formylfuran-2-carboxylic acid for uronic acids) (Fig. 3.51). The monosaccharides that are cleaved first expose their free aldehyde function for a longer time than those more resistant, for this reason amino sugars are more stable to degradation compared to uronic acids, which are even less

stable than neutral monosaccharides. These side reactions make the quantification of the labile component not trustful and can mask completely its occurrence. Loss of a residue can also occur when cleavage of the glycosidic linkage is not complete. In this case the monosaccharide remains attached to another residue so that the resulting disaccharide in overlooked in the GC-MS analysis.

The choice of an alternative solvent, usually methanol, greatly reduces parasite reactions that involve the aldehyde function. The free aldehyde group of the monosaccharide is never exposed. Accordingly, solvolysis can be prolonged in order to ensure the complete transformation of all the monosaccharide constituents of the glycan into the corresponding methyl glycosides. If hydrolysis is performed on a glycan, the obtained monosaccharides need to be reduced and acetylated to be analyzed by GC-MS (acetylated alditols, AA, analysis); otherwise, when methanolysis is performed on a glycan, the obtained monosaccharides only need to be reduced to be analyzed by GC-MS (acetylated methyl glycosides, AMG, analysis) (Fig. 3.52).

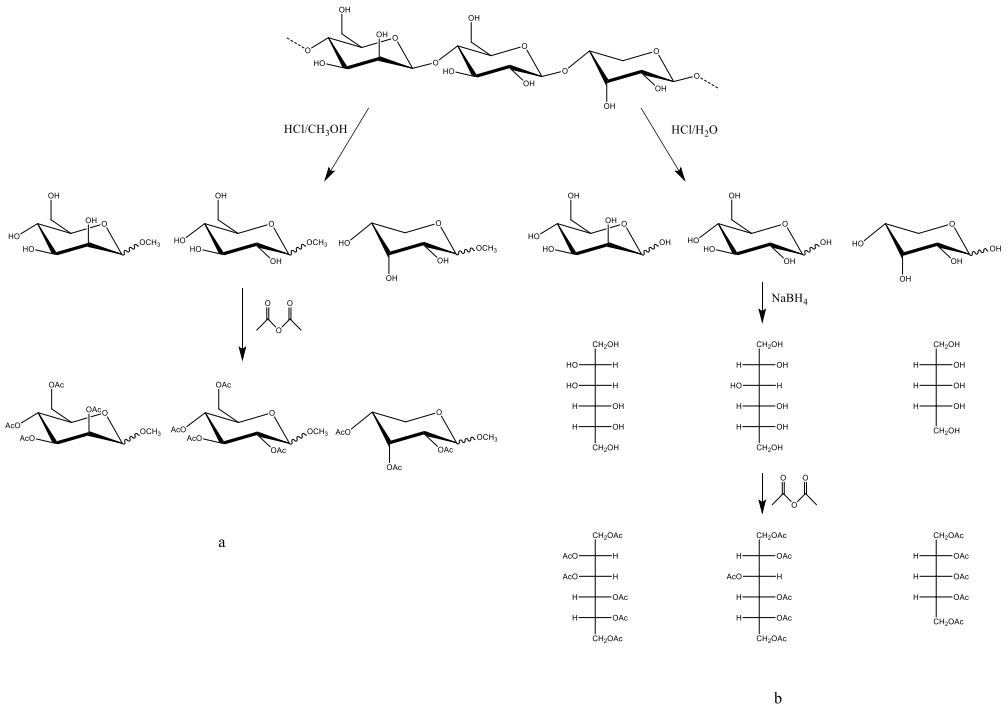


Figure 3.52. Sequence of the reactions to produce AMG (a) and MGA (b).

GC-MS analysis can be exploited also to have deeper structural information about a glycan; in fact, linkage analysis can be performed by combination of methylation reaction (Williamson reaction) on the intact polysaccharide and reduction with acetylation reaction on the hydrolyzed glycan; the final products will be partially methylated and acetylated alditols (PMAA) (Fig. 3.53).

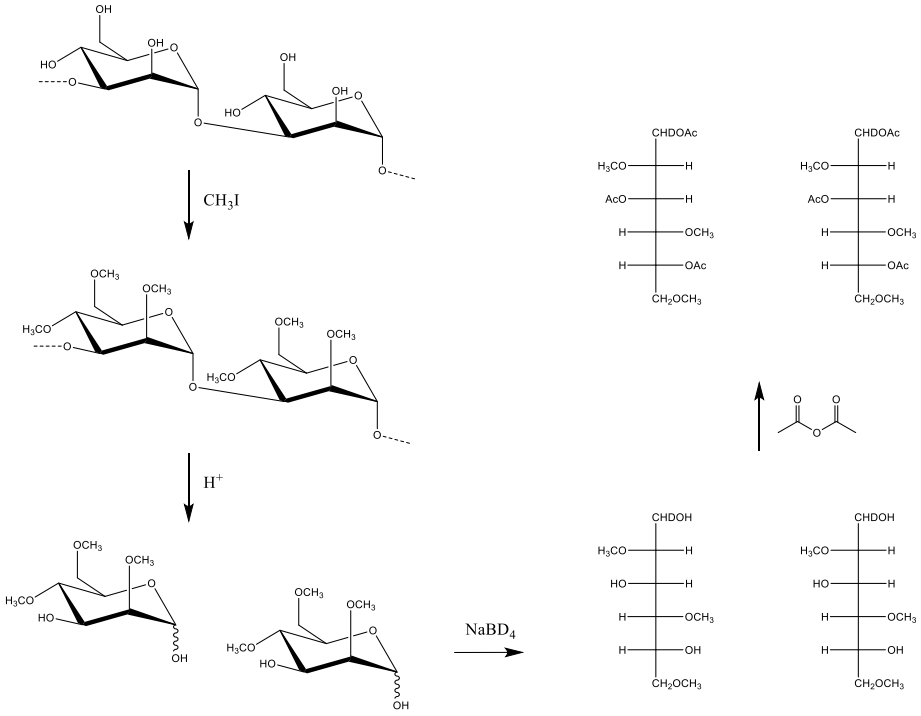


Figure 3.53. Sequence of the reactions to produce PMAA.

The interpretation of the EI-MS spectrum of PMAA according to rules of fragmentation of the alditols backbone brings to the correct attribution of the polysaccharide regiochemistry (Fig. 3.54).

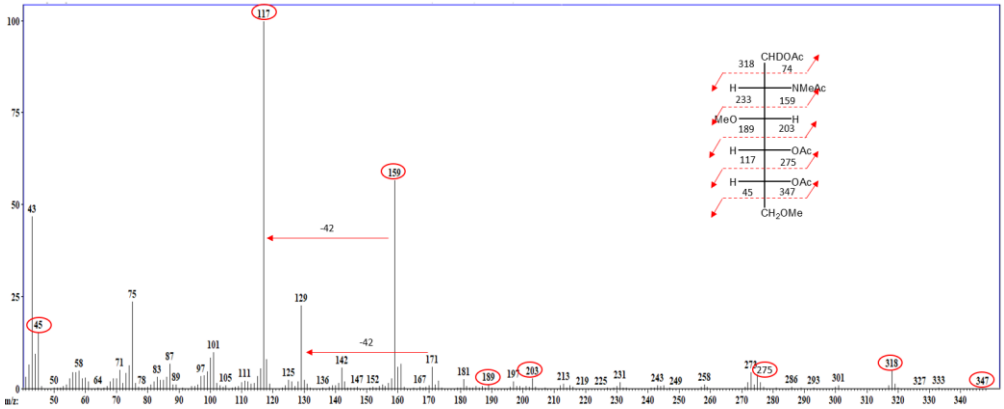


Figure 3.54. Example of PMAA EI-MS: spectrum of a 4-linked Hexp2NAc.

The common way to assess the absolute configuration (D or L) of a monosaccharide is to transform it in a mixture of diastereomer by using an optically pure alcohol (as 2-octanol or 2-butanol) that can be resolved by the GC-MS chromatographic conditions used.

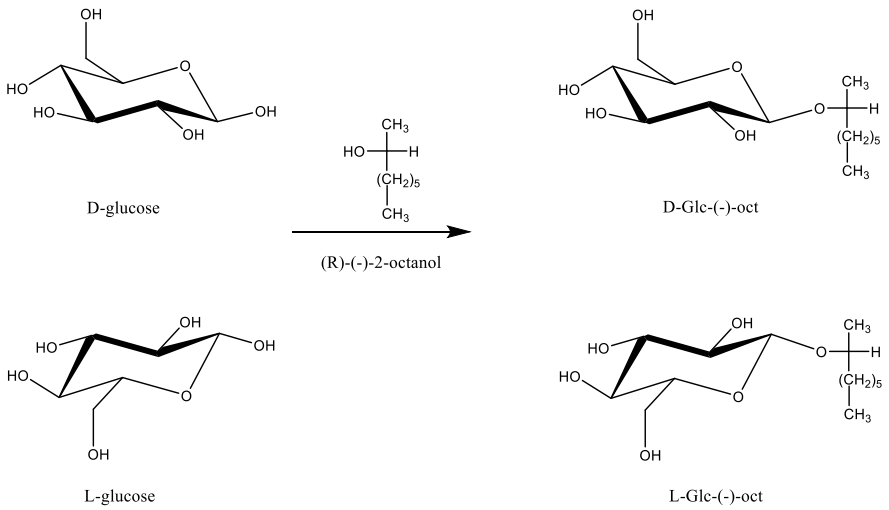


Figure 3.54. Scheme of solvolysis of an enantiomeric mixture of glucose using chiral (R)-(-)-2-octanol.

3.5 Glycoproteomic software

Peptide identification from tandem mass spectrometry data is a central task in proteomics. Many software tools have been developed for correct peptide primary sequence attribution; these can be broadly divided into two categories: *de novo* sequencing and database search. In the first case the peptide sequence is derived directly from the MS/MS spectrum by its interpretation, whereas in the second case the software queries a sequence database for the best peptide to explain the peaks in the MS/MS spectrum. Representative de novo sequencing software packages include PEAKS, PepNovo, NovoHMM, Lutefisk, and representative database search software packages include Mascot, SEQUEST, MaxQuant (Zhang et al., 2012).

Database search software are generally believed to be a simpler than *de novo* ones because they provide a limited space for the software to search the protein sequence. Therefore, when a protein sequence database is available, a database search is the most common method for peptide identification; however, this kind of software suffer from low identification rates (low sensitivity) and high false discovery rates (low

accuracy), for these reasons, the improvement of database search performance has always been a central aim of bioinformatic research.

False discovery rate (FDR) is defined as the percentage of the false identifications in all identifications above the score threshold; to define FDR, decoy peptides are needed, which are produced by randomly shuffling the amino acids in each protein.

$$\text{FDR} = \frac{\text{number of peptide spectrum matches with decoy peptides}}{\text{number of peptide spectrum matches with target peptides}}$$

An FDR= 1 % means that on 100 identified peptides there is 1 with wrong attribution. Accuracy can be accomplished by increasing the score threshold. However, this will at the same time reduce the sensitivity. To improve both sensitivity and accuracy in comparison to other database search software, a combination of both database search and *de novo* tool was introduced with PEAKS DB, which exploits a database search approach for peptide identification, but it improves the filtration and the scoring function using a *de novo* sequencing too.

The software PEAKS studio 11 provides a module (Glycan module) to deeply characterize glycopeptides MS/MS and quantify glycoforms on each glycosite; according to this tool, the glycopeptide spectrum is quickly searched against the protein database to identify candidate peptides, candidate glycans are then selected from the glycan database based on the precursor mass and the respective candidate peptides. If the spectrum remains unidentified after the peptide-based search, it proceeds to the glycan-based search which first searches for candidate glycans and then infers candidate peptides (Fig. 3.55). To pass a 1% FDR filter both peptide and glycan FDRs have to be less than or equal 1%; in specific, glycans FDR is generated by applying random mass shifts of fragment ions to create decoy spectra (Sun et al., 2023). If the glycan structure is not found in the database, a new glycan will be defined by *de novo* sequencing; this procedure exploits a deep learning module which

constructs the oligosaccharide structure from root to leaves, starting from the peptide (root) and iteratively adding monosaccharides (leaves) to the tree.

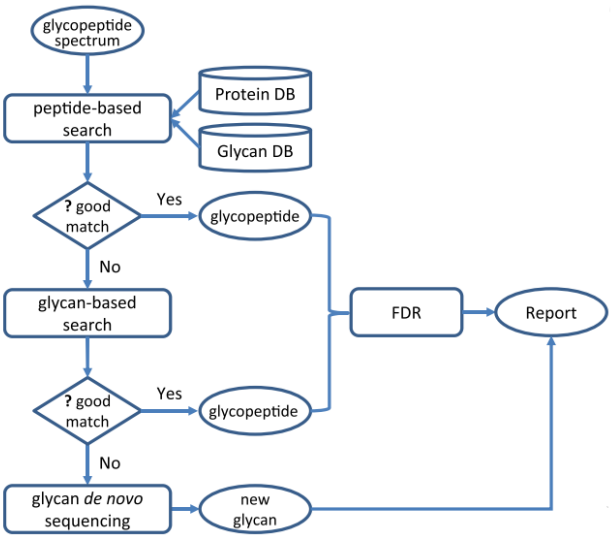


Figure 3.55. Workflow of PEAKS 11 for glycopeptide MS/MS analysis. From *Glycopeptide database search and de novo sequencing with PEAKS GlycanFinder enable highly sensitive glycoproteomics.* (Sun et al., 2023).

SECTION II:

RESULTS AND DISCUSSION

Chapter 4 – *N*-glycans site specific analysis

4.1 Introduction

N-glycans analysis of a glycoprotein is a critical quality attribute, indeed, it can affect half-life, efficacy and safety of the product; in addition, defining the detailed composition of the glycan shield increases the understanding of bNAb epitopes and helps the rational design of glycoprotein-based vaccines (Gupta et al., 2018). Protein architecture can have a substantial impact on glycosylation pattern. Indeed, under-processed oligomannose-type and hybrid-type glycans are common on viral glycoproteins and arise from the glycan- or protein-mediated steric clashes with ER and Golgi resident mannosidase enzymes, which might be detrimental for the glycan processing (J. D. Allen et al., 2021). The first function that has been attributed to the glycans on the surface of viral antigenic proteins consists in the shielding of the conserved protein epitopes, which hampers the recognition by the immune system. This concept is exacerbated in HIV-1 Env, one of the most heavily glycosylated proteins known with glycans composing up 50% of its total mass (Behrens et al., 2016). However, in some cases protein glycans can form distinct epitopes. The first monoclonal antibody able to recognize viral glycans has been the 2G12, isolated from an asymptomatic HIV-1 infected patient in 1990.

Furthermore, glycans can also affect glycoprotein conformation and consequentially receptor recognition, like in the case of the SARS-CoV-2 S protein, where three N-glycans (N165A, N234A, and N343A) play a key role in the stabilization of the receptor binding domain (RBD) up conformation (Chawla et al., 2022).

In this study we used different mass spectrometry techniques to characterize in a site-specific way the *N*- glycosylation pattern of recombinant gB and Pentamer (Towne strain) secreted from mammalian CHO cells. Moreover, the *N*-glycans pattern of CHO-produced gB in the presence of kifunensine (a mannosidase I inhibitor) in the culture medium has been compared with the CHO-produced gB without inhibitor. The

use of kifunensine offers the advantage to produce *N*-glycans from CHO cells without Neu5Gc.

4.2 *N*-glycans microheterogeneity analysis by LC-MS/MS

After denaturation, gB was digested with an appropriate protease (bottom-up approach) and then injected into a liquid chromatography-tandem mass spectrometry (LC-MS/MS) system. More in detail, to cover all the peptides sequences containing *N*-glycosylation sequons, different proteases (or a combination of them) have been used; the used proteases consist of trypsin, chymotrypsin, Glu-C, alpha-lytic protease. Their selection was done in order to obtain glycopeptides with specific physico-chemical requirements: minimum number of consensus sequences (ideally only one), specific cleavage, glycoforms of homogeneous peptide chain length and clear MS/MS characterization. In addition, to correctly describe a glycoform the following criteria have been used: the glycopeptide has to be characterized by at least two MS/MS spectra and each spectrum has to contain at least one oxonium ion from the glycan fragmentation (Fig. 4.1); furthermore, each MS/MS spectrum should give a good peptide sequence coverage, this aim is reached using the function "glycan search" of the software PEAKS Studio version 11, which allows to obtain a full characterization of the glycopeptides MS/MS spectra, accounting also for b/y ion series which bring variable glycan fragments to the peptide backbone (Sun et al., 2023).

Oxonium ion (B ion)	Exact mass (Da)
$[HexNAc_1\text{-}C_2H_6O_3]^+$	126.0550
$[HexNAc_1\text{-}CH_6O_3]^+$	138.0550
$[Hex_1]^+$	163.0601
$[HexNAc_1\text{-}H_2O]^+$	186.0761
$[HexNAc_1]^+$	204.0867
$[HexNAc_1Hex_1]^+$	366.1395
$[HexNAc_1NeuAc_1]^+$	495.1921
$[HexNAc_1Hex_2]^+$	528.1923
$[Fuc_1]^+$	147.0652
$[Fuc_1\text{-}H_2O]^+$	129.0546
$[NeuAc_1]^+$	292.1027
$[NeuAc_1\text{-}H_2O]^+$	274.0921
$[HexNAc_1Hex_1NeuAc_1]^+$	657.2349
$[HexNAc_1Hex_2NeuAc_1Fuc_1]^+$	803.2928
$[HexNAc_1Hex_2NeuAc_1]^+$	819.2877
$[NeuGc_1]^+$	308.09761
$[NeuGc_1\text{-}H_2O]^+$	290.08704

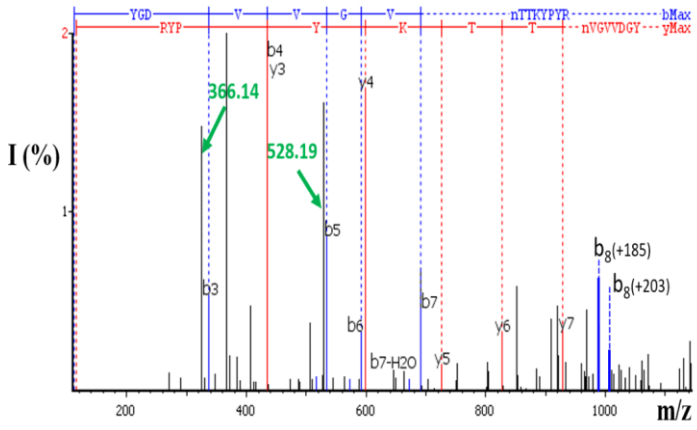


Figure 4.1. Oxonium ions list and glycopeptide MS/MS spectrum.

Site-specific glycans relative quantitation was done by summing the ion intensity of the corresponding glycopeptide over all the identified charge states, using the corresponding XICs, and dividing it by the total intensity of all the glycopeptide peaks present for a specific site (label free approach) (Fig. 4.2); this can be done because protonation of glycopeptides occurs mainly on the peptide backbone, which drives ionization efficiency (Behrens et al., 2016). Moreover, ionization efficiencies of peptides are driven by the nature of the constituent amino acids, and they increase as the number of amino acid residues increases; however, from a length of five amino acid residues the ionization efficiency level off and further increase in the backbone length has only a minor effect. To account for this feature, we excluded from the relative abundance measurements the glycopeptides quite different coverages (Liigand et al., 2019).

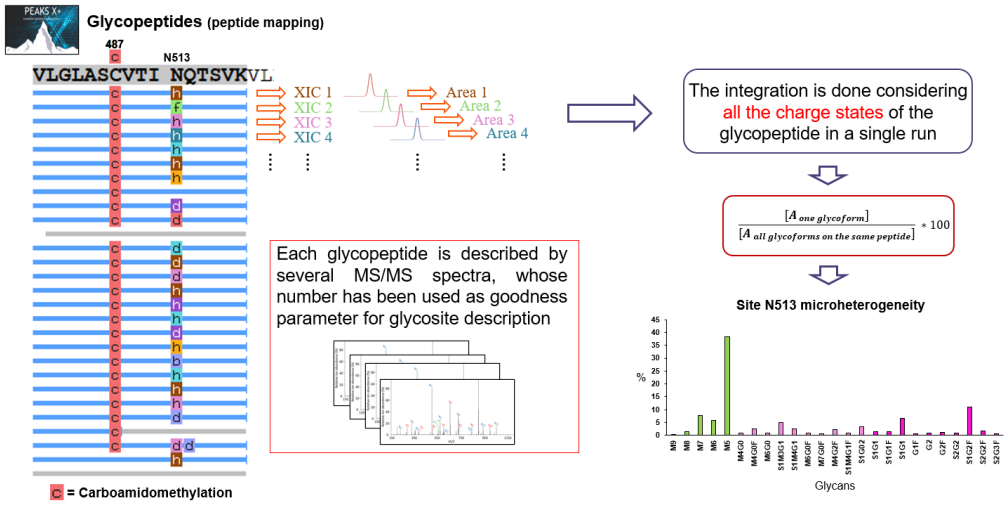


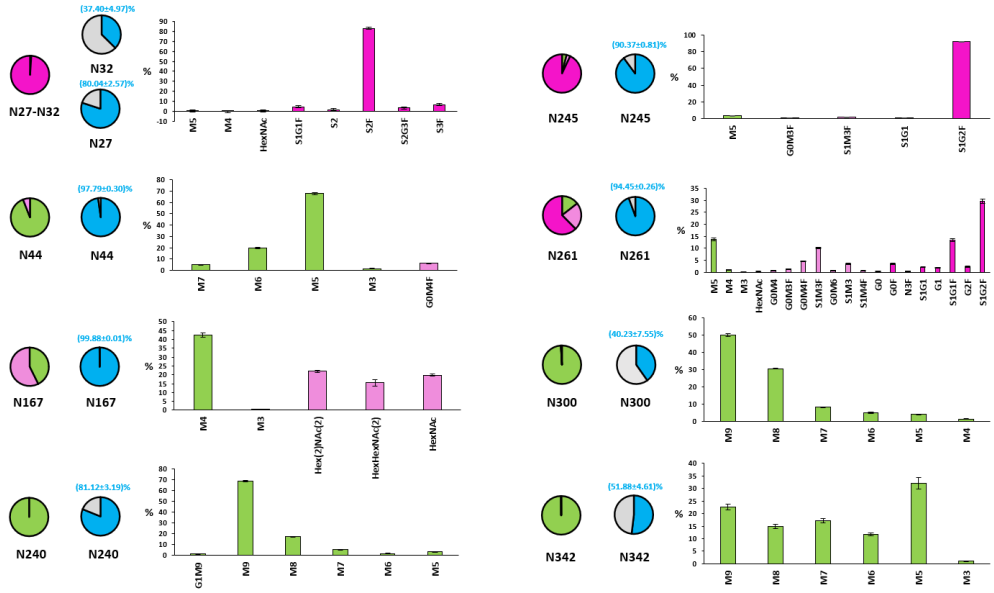
Figure 4.2. Workflow for glycoforms relative abundancies calculation.

4.3.1 Site-specific *N*-Glycosylation profile of the recombinant protein gB expressed in CHO cells.

The expressed gB contains 18 predicted *N*-glycosylation sites (NxS/T, with x≠P): N27, N32, N44, N167, N240, N245, N261, N300, N342, N364, N368, N376, N406, N411, N423, N424, N513, N544. Among these sites, N406, N423 and N424 were not found glycosylated; on the other hand, the glycosylated sites have shown to have several different glycoforms, with a certain *N*-glycans maturation degree for each sequon. *N*-glycans have been classified into three categories (oligomannose, hybrid and complex), depending on branching and fucosylation.

It was found out that five sites are principally occupied by oligomannose-type glycans: N44, N167, N240, N300 and N342. The predominant oligomannose-type glycans observed across the protein are M5 (Man5GlcNAc2) and M9 (Man9GlcNAc2). The site N240 was found to be the most underprocessed since it presented also, showing only a glucose containing G1M9 (GlcMan9GlcNAc2)

structure. The sites bringing mainly M5 structures suggested that they were largely accessible to α-1,2-mannosidases but poor substrates for GlcNAcT-I (Fig. 4.3). Moreover, sites N261, N376, N513 and N544 consisted of a mixture of oligomannose- and complex-type glycans. The remaining sites N27-32, N245, N261, N364-368, had mainly complex glycans; in addition, on site N411 only one glycoform has been found (Fig. 4.4). Sometimes isobaric species may coelute, when it happens glycoforms identified on peptides bringing more than one sequon cannot be assigned to any individual *N*-glycosite, for this reason we treated the glycosylation sites couples N27-32 and N364-368 as one, giving an average description of the glycan microheterogeneity of each pair.



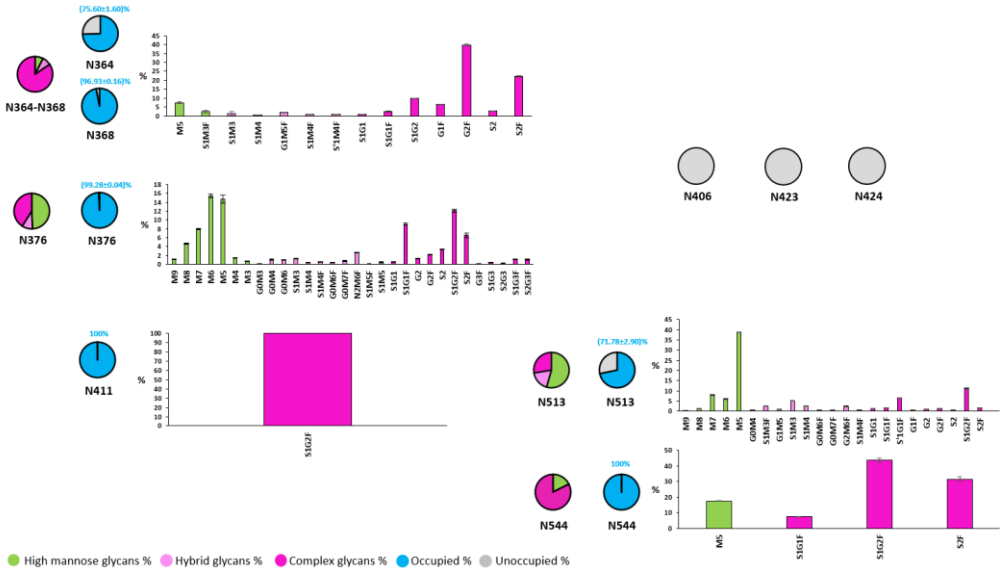


Figure 4.4. Site specific *N*-glycosylation profile of gB expressed in CHO cells (glycans relative abundancies and occupancies). The final output is given by the total average of the three average values. The detailed percentage values with their corresponding standard deviations are given in Tab. 4.1. The glycan structures with their corresponding names are given in Fig. 4.5.

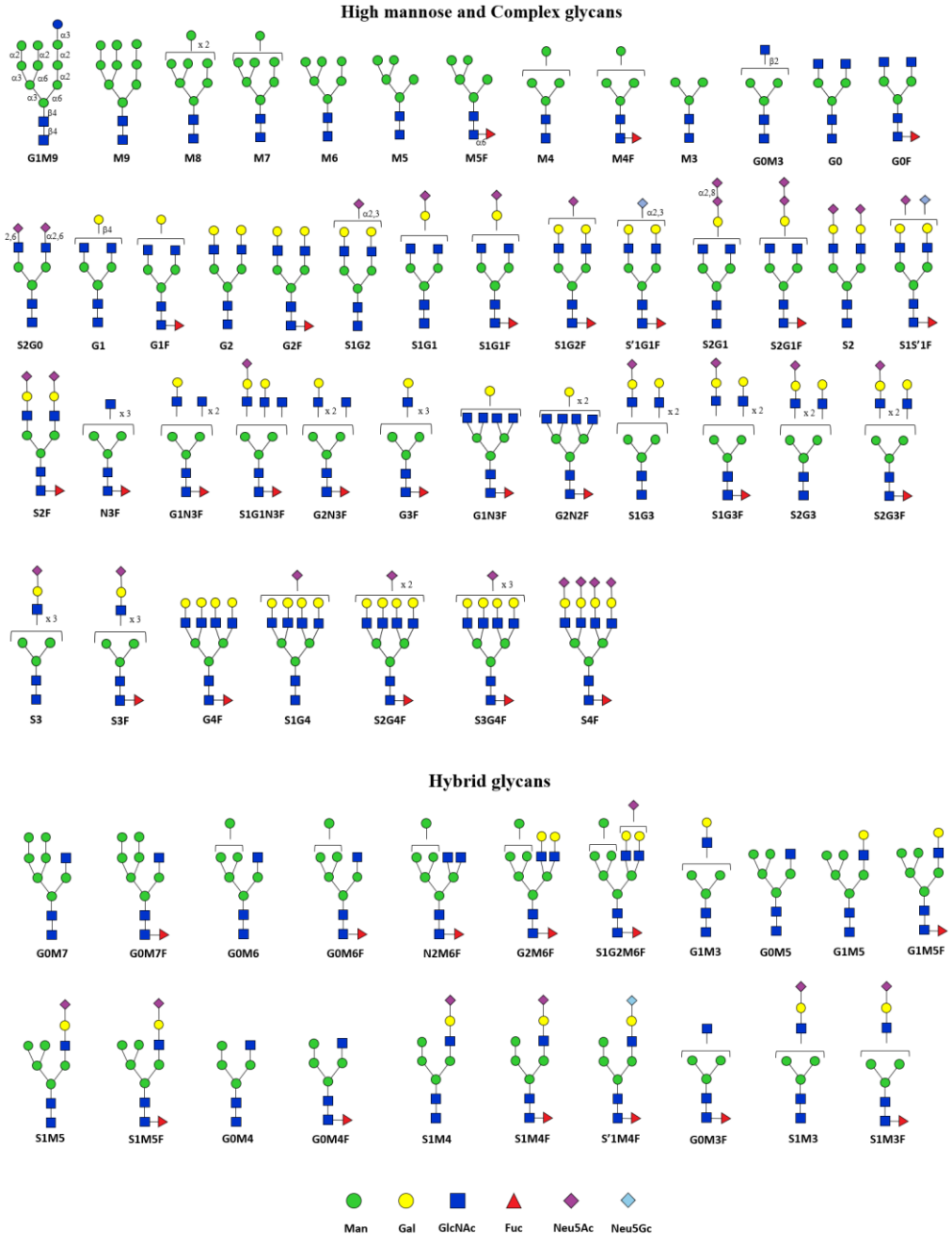


Figure 4.5. List of all the glycan structures found on the analyzed glycoproteins.

Table 4.1. Site specific *N*-glycans analysis from CHO expressed gB analysis data. The proteolytic enzyme used for the glycopeptide quantitation is shown in brackets (T=Trypsin, G=Glu-C); the most predominant glycoform for each site is highlighted in red. The glycoforms relative abundance values are expressed as percentages. The glycan structures with their corresponding names are given in Fig. 4.5.

	***N*-glycosylation site**														
Glycan	27(T)	32(T)	44(T)	167(G)	240(T)	245(T)	261(G)	300(T)	342(G)	364(T)	368(T)	376(T)	411(G)	513(T)	544(T)
G1M9					1.37±0.03										
M9					68.63±0.62			49.99±0.87	22.51±1.18			1.28±0.11		0.24±0.03	
M8					17.45±0.05			30.69±0.03	14.77±0.84			5.01±0.11		1.44±0.09	
M7			4.93±0.29		5.51±0.18			8.34±0.21	17.18±0.85			8.52±0.11		7.74±0.26	
M6			19.49±0.61		1.94±0.06			5.19±0.26	11.77±0.52			16.52±0.42		5.91±0.24	
M5	0.27±0.01		67.84±0.80		3.50±0.29	3.85±0.23	13.67±0.59	4.25±0.23	32.03±2.35	7.45±0.42		15.77±0.92		38.25±0.30	17.57±0.39
M4	0.01±0.001			42.53±1.32	1.59±0.14		1.04±0.05	1.52±0.16	1.06±0.08			1.53±0.07			
M3			1.52±0.11	0.15±0.03			0.24±0.01					0.76±0.06			
GlcNAc	0.22±0.01			19.87±0.65			0.37±0.07								
ManGlcNAc(2)				15.50±1.88											
Man(2)GlcNAc(2)				21.96±0.46											
G0M3												0.15±0.02			
G0							0.48±0.01								
G0F							3.50±0.21								
G1							1.91±0.13								
G1F										6.49±0.11				0.69±0.03	
G2												1.43±0.05		0.98±0.03	
G2F							2.28±0.24			39.91±0.46		2.34±0.10		1.24±0.06	
S1G2										9.82±0.15					
S1G1						0.83±0.03	2.10±0.12			0.91±0.03		0.65±0.43		1.43±0.09	
S1G1F	4.56±0.16						13.41±0.55			2.44±0.24		9.70±0.28	100	1.49±0.10.	7.52±1.33
S1G2F						91.97±0.26	29.58±0.82					12.97±0.04		10.91±0.32	43.46±0.23
S'1G1F														6.49±0.13	
S2	1.80±0.01									2.65±0.21		3.62±0.17		0.74±0.05	
S2F	82.93±0.90									22.19±0.48		7.02±0.61		1.66±0.09	31.43±1.82
N3F							0.64±0.04								
G3F												0.18±0.01			
S1G3												0.46±0.01			
S1G3F												1.22±0.05		0.69±0.04	
S2G3												0.25±0.03			
S2G3F	3.35±0.19											1.10±0.15			
S3F	6.86±0.57														
G0M7F												0.77±0.16		0.71±0.06	
G0M6							0.80±0.06					1.13±0.04			
G0M6F												0.46±0.01		0.84±0.05	
G2M6F												2.82±0.09		2.37±0.21	
S1G2M6F												0.07±0.01			
G1M5														0.94±0.02	
G1M5F										2.14±0.02					
S1M5										0.84±0.05					
G0M4							0.80±0.08					1.18±0.16		0.82±0.01	
G0M4F			6.22±0.28			2.19±0.05	4.62±0.18								
S1M4										0.61±0.02		0.52±0.01		2.54±0.13	
S1M4F							0.66±0.04			0.84±0.10		0.63±0.02		0.80±0.03	
S'1M4F												0.50±0.11			
G0M3F						0.94±0.06	1.37±0.10								
S1M3							3.42±0.27			1.22±0.10		1.39±0.07		5.04±0.21	
S1M3F							10.10±0.34			2.47±0.69				2.49±0.07	

4.3.2 Site-specific *N*-Glycosylation profile of the recombinant protein gB expressed in CHO cells treated with kifunensine.

To assess if the effect of kifunensine on glycan maturation is the same for each sequon, eleven out eighteen *N*-glycosylation sites of gB from CHO cells treated with kifunensine have been analyzed. As expected, it was found out that all the sites are principally oligomannose-type, with M9 glycan as most abundant structure. Some sites (*e.g.* N27, N32) have shown a significative percentage of also M8 glycan, probably due to higher accessibility of mannosidase than other sites. Low percentages of other glycans (*e.g.* M7, M6, M5) have been found on similar amount on all the analyzed glycosylation sites. These data are aligned with the predicted effect that kifunensine has on the *N*-glycans maturation (Figs. 1.1, 4.6).

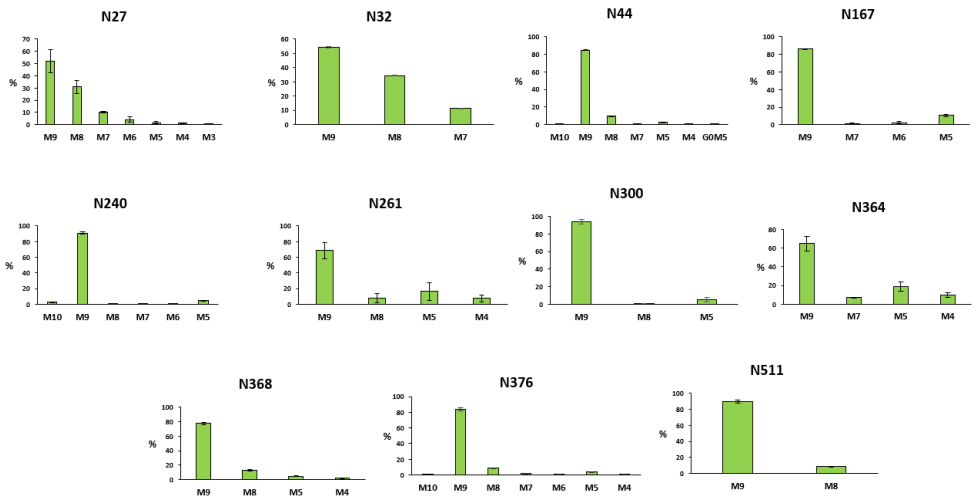


Figure 4.6. Site specific *N*-glycosylation profile of gB expressed in CHO cells treated with kifunensine (glycans relative abundancies). The final output is given by the total average of the three average values. The detailed percentage values with their corresponding standard deviations are given in Tab. 4.2. The glycan structures with their corresponding names are given in Fig. 4.5.

Table 4.2. Site specific *N*-glycans analysis from CHO treated with kifunensine expressed gB analysis data. The proteolytic enzyme used for the glycopeptide quantitation is shown in brackets (T=Trypsin, G=Glu-C); the most predominant glycoform for each site is highlighted in red. The glycoforms relative abundance values are expressed as percentages. The glycan structures with their corresponding names are given in Fig. 4.5.

	***N*-glycosylation sites**										
Glycan	27(T)	32(T)	44(T)	167(G)	240(T)	261(G)	300(T)	364(T)	368(T)	376(T)	511(T)
G1M9			0.949±0.008		3.22±0.01					0.20±0.11	
M9	51.86±9.51	54.25±0.45	84.77±0.09	85.96±0.58	90.89±1.55	68.56±10.55	94.14±2.63	65.00±7.91	79.27±0.12	84.08±2.19	91.44±0.96
M8	30.93±5.39	34.45±0.21	9.28±0.06		0.73±0.44	8.05±6.06	0.66±0.02		13.30±0.69	8.57±0.29	8.56±0.96
M7	10.23±0.76	11.30±0.23	1.08±0.03	1.28±0.50	0.47±0.51			7.19±0.33		1.79±0.5	
M6	4.08±2.30			2.16±1.24	0.42±0.28					0.91±0.57	
M5	1.57±0.80		3.00±0.10	10.60±1.17	4.28±0.33	16.11±11.79	5.21±2.65	18.50±5.08	4.68±0.45	3.60±0.40	
M4	1.25±0.24		0.65±0.02			7.27±4.81		9.31±2.51	2.31±0.05	0.84±0.32	
M3	0.070±0.005		0.28±0.14								

4.3.3 Site-specific *N*-Glycosylation profile of the recombinant protein Pentamer from CHO cells.

Pentamer (or pentameric complex) is a heteropentameric glycoprotein, formed by the monomers gH, gL, UL128, UL130 and UL131A; among these, subunit UL128 is the only one without any *N*-glycosylation sequon. Recombinant pentameric complex contains eleven predicted N-glycosylation sites: N32, N39, N44, N169, N618, N677 (gH), N44 (gL), N60, N93, N176 (UL130) and N63 (UL131A) (Chandramouli et al., 2017). This analysis revealed that each site presents a differential level of oligomannose, hybrid, and complex-type glycan populations (Fig. 4.7). Among these sites, only N44 (gL) and N63 (UL131A) have shown a high percentage of oligomannose structures (56 % and 91 % respectively), probably due to the higher steric hindrance experienced by gL and UL131A; the remaining *N*-glycosylation sites showed only complex glycan structures, where the most frequent were sialylated mono- and biantennary glycans (Fig. 4.7).

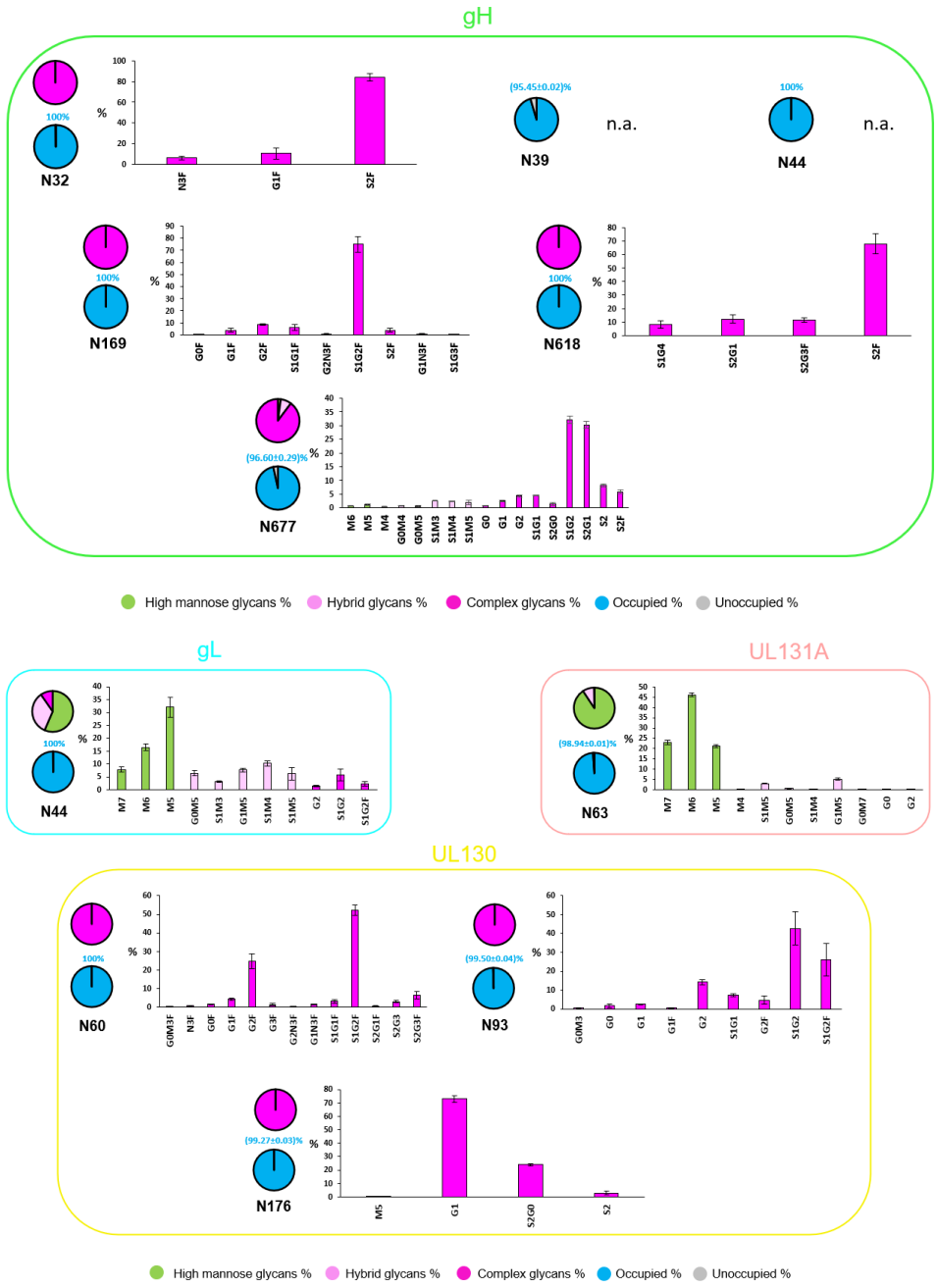


Figure 4.7. Site specific cells *N*-glycans profile in Pentamer expressed in CHO (glycans relative abundancies and occupancies). The final output is given by the total average of the

three average values. The detailed percentage values with their corresponding standard deviations are given in Tab. 4.3. The glycan structures with their corresponding names are given in Fig. 4.5.

Table 4.3. Site specific *N*-glycans analysis from CHO expressed Pentamer analysis data. The proteolytic enzyme used for the glycopeptide quantitation is shown in brackets (T=Trypsin, C=Chymotrypsin, A= alpha-lytic protease); the most predominant glycoform for each site is highlighted in red. The glycoforms relative abundance values are expressed as percentages. The glycan structures with their corresponding names are given in Fig. 4.5.

	***N*-glycosylation site**								
Glycan	32 (gH, T)	169 (gH, A)	618 (gH, C)	677 (gH, C)	44 (gL, C)	63 (UL131A, T)	60 (UL130, T)	93 (UL130, T)	176 (UL130, T)
M7					7.84±0.96	23.10±1.15			
M6				0.65±0.14	16.50±1.20	46.21±0.77			
M5				1.14±0.19	32.11±3.84	21.11±0.71			0.37±0.20
M4				0.36±0.11		0.30±0.06			
G0M3								0.49±0.18	
G0				0.73±0.09		0.10±0.01		1.65±0.84	
G0F		0.69±0.15					1.50±0.10		
S2G0				1.32±0.60					23.94±1.08
G1				2.36±0.26				2.47±0.15	72.92±2.57
G1F	10.29±5.36	3.90±1.65					4.25±0.55	0.45±0.19	
G2				4.24±0.34	1.57±0.27.	0.05±0.03		14.19±1.40	
G2F		8.69±0.78					24.81±3.92	4.71±2.15	
S1G2				32.16±1.39	5.71±2.33			42.67±8.94	
S1G1				4.56±0.20				7.29±0.74	
S1G1F		6.01±2.60					3.08±0.89	26.06±8.47	
S1G2F		74.86±6.71			2.36±0.97		52.09±2.78		
S2G1			11.54±1.72	30.33±1.20					
S2G1F							0.64±0.04		
S2				8.18±0.61					2.77±1.41
S2F	83.77±3.66		12.13±3.01	5.85±0.75					
N3F	5.94±1.71						0.73±0.02		
G2N3F		0.81±0.18					0.41±0.11		
G3F							1.37±0.67		
G1N3F							1.59±0.26		
S1G3F		0.46±0.15							
S2G3							2.95±0.59		
S2G3F			8.34±2.82				6.46±2.16		
S1G4			67.99±7.54						
G0M7						0.07±0.01			
G0M5				0.50±0.19	6.49±1.07	0.65±0.10			
G1M5					7.67±0.67	5.13±0.47			
G1M5F									
S1M5				1.92±0.80	6.26±2.48	2.89±0.19			
G0M4				0.75±0.15					
S1M4				2.36±0.07	10.33±1.04	0.39±0.03			
G0M3F							0.12±0.05		
S1M3				2.58±0.06	3.17±0.29				

4.4 *N*-glycans occupancy analysis by LC-MS/MS

Glycopeptides ionize weakly, compared to nonglycosylated peptides, for this reason small changes in the presence of non-glycosylated species could impart large variability in a semi-quantitative assay (Rebecchi et al., 2009). As consequence, *N*-glycosylation sites occupancies have been evaluated by removing glycans from the corresponding asparagines with the enzyme PNGase F, which transforms Asn into Asp.

More in detail, after digestion and work-up, gB glycopeptides were dried and treated with the enzyme PNGase F in $H_2{}^{18}O$ before injection into the LC-MS/MS system; in this way we have a more accurate occupancy measurement. By using isotopic labelling, we can distinguish between enzymatic deamidated peptides from non-enzymatic deamidated peptides; indeed non-enzymatic deamidation of Asn and Gln occurs spontaneously on proteins and peptides both *in vivo* and *in vitro*; the deamidation rate of peptides increases significantly under alkaline conditions and at relatively high temperature (Hao et al., 2011), (Pace et al., 2013); therefore, slightly acid pH and reduced heating times have been used during protein denaturation to minimize this effect. Deamidation of Asn leads to aspartic and isoaspartic acid forms, thereby introducing a negative charge at the deamidation site; the reaction passes through a succinimide intermediate (Fig. 4.8). Usually, when asparagine is followed by small amino acids such as glycine or serine, the rate of deamidation is higher (Spanov et al., 2022).

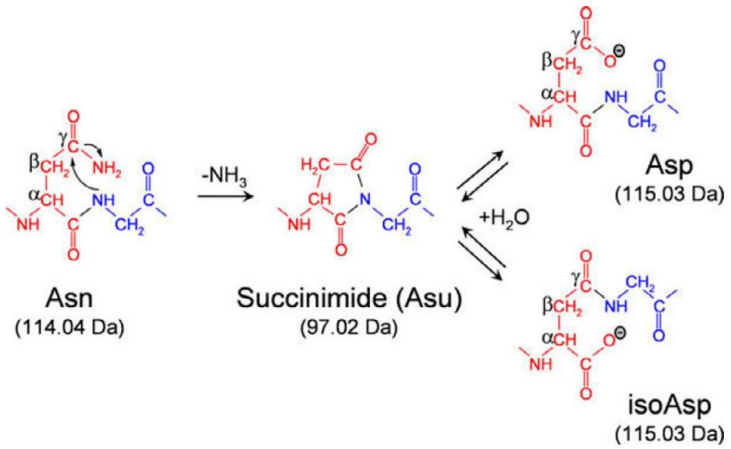


Figure 4.8. Mechanism of asparagine deamidation through a succinimide intermediate. From *Analytical tools for the characterization of deamidation in monoclonal antibodies.* (Spanov et al., 2022).

For occupied sites, the enzymatic reaction will bring to peptides with a + 2.9840 Da mass shift, while for peptides with unoccupied glycosites these treatments produce no mass shift or a mass shift of 0.9840 Da if deamidation has occurred (Cao et al., 2017) (Fig. 4.9). Furthermore, tryptic incorporation of one or two ^{18}O into the C-terminus of peptides during N-linked glycosylation site mapping has been considered, indeed this modification can lead to incorrect and ambiguous identification of sites of N-linked glycosylation; this issue has been overcome by allowing ^{18}O incorporation into the C-terminus as a variable modification during the database search (+2.0042 Da or + 4.0082 Da) (Angel et al., 2007).

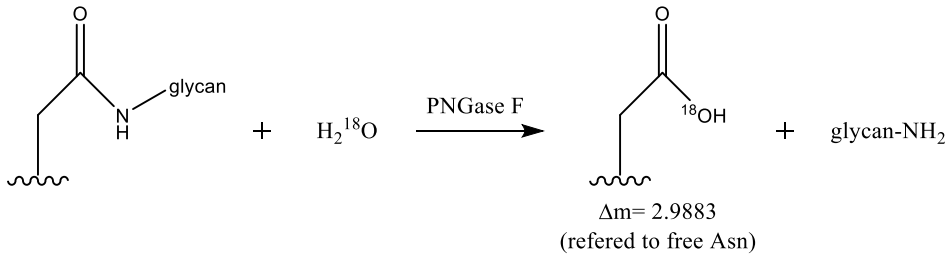


Figure 4.9. Enzymatic deamidation labeling reaction.

From the data analysis, it results that recombinant gB has three unoccupied sites (N406, N423 and N424), three low-occupancy sites (N32, N300 and N342) and the remaining sites have occupancies of at least 75 % (Fig. 4.4).

The same analysis was performed on the Pentameric complex, which showed all the *N*-glycosylation sites completely occupied by glycans (Fig. 4.7)

Standard deviation of each percentage value was computed considering three average values obtained repeating the analysis three times for each of the three technical replicates. The final output is given by the total average of the three average values. The detailed percentage values with their corresponding standard deviations are given in Fig. 4.4 and Fig. 4.7.

4.5 Discussion

N-glycans site specific analysis has been performed on gB, $gB_{kifunensine}$ and Pentamer using high resolution LC-MS/MS combined with a bottom-up proteomic strategy. This approach allowed to characterize both *N*-glycans microheterogeneity and *N*-glycans occupancy through relative quantification. gB *N*-glycosylation sequons showed a multifaceted glycan maturation degree, with some sites bringing only complex glycans, others only oligomannose glycans and some both; furthermore, gB

showed some unoccupied or low occupancy *N*-glycosylation sites. All these data suggest that there are protein regions less accessible to the processing enzymes than others, probably due to glycan structures that are sandwiched between other two. Furthermore, it has been proved that, in presence of kifunensine, the site-specific *N*-glycan profile is the same for eleven sites.

The Pentameric complex showed a more homogeneous glycosylation profile than gB, in fact, it has all the *N*-glycosylation sites occupied, with only two of them bearing mainly high-mannose structures.

Chapter 5 – *N*-glycans relative quantitative profiling on gB and Pentamer

5.1 Introduction

Glycan mixtures can be analyzed in several ways. They can be labelled with a tag which allows their detection by fluorescence after chromatographic separation. Another powerful technique to analyze glycans mixtures is the ion mobility spectrometry (IMS), a gas-phase separation technique in which ions are guided through a gas-filled drift cell by an electric field. On their way, they undergo collisions with a buffer gas, which separates them based on their charge, size, and shape; this technique is able to separate glycans even with minimal structural differences: not only structural isomers but also the α and β reducing end anomers of each isomer as well as different conformers; furthermore, differently from LC-FLR, it does not require sample derivatization (Faleh et al., 2022). Glycan microarray is a different approach and a powerful tool for antigenicity studies. Indeed, known glycans can be used as probes to investigate if they are recognized by certain mAbs. To reach this target, glycans are immobilized glycans on a solid support and the analytical strategy for the detection consists of tagging the glycan binding protein (GBP) by using either a fluorescent antibody or fluorescent streptavidin with biotinylated GBP (Gillmann et

al., 2023). Finally, nuclear magnetic resonance (NMR) spectroscopy is one of the most powerful tools to elucidate the molecular structure of mono, oligo and polysaccharides, indeed the application of this technique to carbohydrates radically changed and simplified the way to deduce the structure of these molecules (Marchetti et al., 2021; Speciale, Notaro, Garcia-Vello, et al., 2022). This spectroscopic tool provides information on the glycan primary structure, including the type of glycosidic linkages, the stereochemistry of the residues and their sequence. However, its application requires a certain amount of pure sample.

5.2 *N*-glycans relative quantitative profiling by LC-FLR-MS/MS: principle overview.

Oligosaccharides can be analyzed after *N*-glycans shaving with a specific enzyme (*e.g.*, PNGase F), by using Liquid Chromatography with Fluorescence detection (LC-FLR); in fact, released glycans can be labelled with a tag which allows their detection by fluorescence with high sensitivity and the advantage to enable the quantitative estimation of each species. One of the most common fluorescent probes is 2-aminobenzamide (2AB), while other reagents are becoming very popular like RapiFluor-MS and InstantPC, which consist of both fluorescent and ionizable tags suitable for mass spectrometry (Fig. 5.1). These tags present further advantages: reduced preparation time and increased detection sensitivity, and good reproducibility (Lim et al., 2019; Xie et al., 2021). The only drawback for 2-AB labeling is the need of acidic conditions that promote the loss of sialic acids from glycans (Furuki et al., 2017).

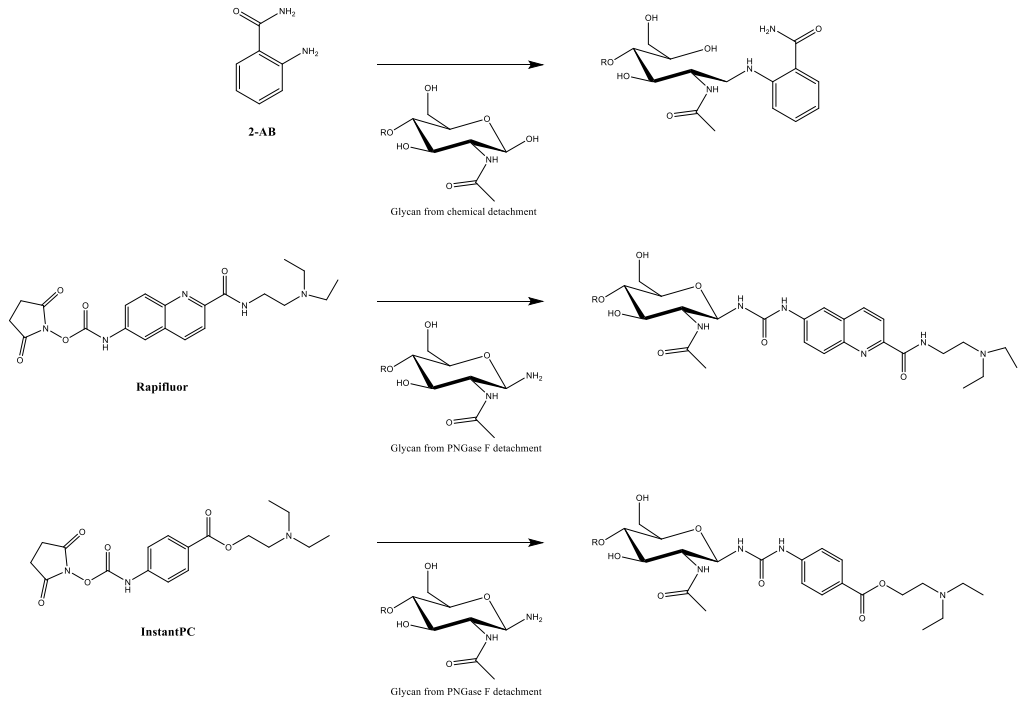


Figure 5.1. Most common labelling reactions of glycans.

Hyphenation of LC-FLR with MS/MS, not only allows the immediate identification of the *N*-glycans structures (from the molecular mass value) without needing any standard, but can also give deeper information about the glycan structure and the presence of non-self monosaccharides; indeed, it is hard to obtain this kind of information from the precursor ion mass value, due to the presence of isobaric species. The presence of non-self monosaccharides (such as Neu5Gc) and motifs (such as α-Gal epitope) in a glycan structure can be confirmed by the detection specific oxonium ions, such as *m/z* 308.10, *m/z* 290.09 for Neu5Gc and *m/z* 528.19 for α-Gal epitope (Fig. 5.2).

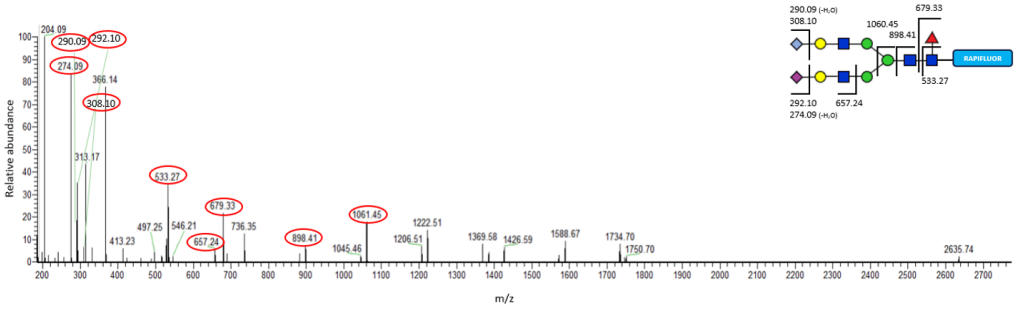


Figure 5.2. MS/MS spectrum of a complex *N*-glycan containing both Neu5Gc and Neu5Ac.

5.2.1 *N*-glycans relative quantitative profiling of the recombinant gB.

To confirm the identities and abundances of the glycans found by site-specific analysis and have a broader view of the glycan composition, a total *N*-glycans analysis of gB expressed in different culture conditions (from CHO cells without and with kifunensine, a mannosidase I inhibitor, Fig. 1.1) has been performed by liquid chromatography-fluorescence-tandem mass spectrometry (LC-FLR-MS/MS) system. More in detail, *N*-glycans shaving on gB has been done by denaturing the protein and using PNGaseF, then the enzymatically released *N*-glycans have been labeled with Rapifluor-MS. In addition, Rapifluor-MS normalizes the signal of each oligosaccharide, because the fluorescence is owed to the tag quinoline portion; the ionizable Rapifluor-MS amino group allows the detection by mass spectrometry, which is useful for glycan identification (Deris et al., 2021).

The *N*-glycans pattern of gB expressed with kifunensine (gB_{kif}) in the culture medium has shown only a predominant mannose glycan M9 structure as expected because the glycan maturation is arrested to the mannosidase I level (Fig. 1.1). Lower levels of shorter high-mannose structures like M8, M7, M5 and M4 were observed; no complex structures have been found (Fig. 5.3, Tab. 5.1).

The *N*-glycans pattern of gB expressed without the inhibitor has shown two predominant structures, which are the sialylated complex glycans S1G1F and S2F, meaning that glycans have high maturation degree. However, a lower but significant

percentage of high-mannose glycans M5 and M9 has been detected (Fig. 5.4, Tab. 5.2). Glycan structures containing Neu5Gc have been detected with a very low relative abundance, indeed, their intensities were comparable with the background noise; the confirmation of the presence of Neu5Gc has been done by detecting the *m/z* 308.09761 (Neu5Gc) and *m/z* 290.08759 (Neu5Gc-H_2O) oxonium ions in the glycans MS/MS spectra.

Released *N*-glycans analysis by LC-FLR-MS allowed to detect a lower number of glycan structures than site-specific analysis (36 against 48 structures); nevertheless, tetrantennary complex oligosaccharides (G4F, S2G4F, S3G4F, S4F) have been detected only by LC-FLR-MS, probably due to their strong retention on SPE cartridges used for site specific analysis (HLB Oasis). Moreover, non-human biantennary structure S1S'1F has been detected only by LC-FLR-MS.

However, the other structures found by LC-FLR-MS are the same found by site-specific analysis and the relative abundancies data of both methods are in agreement. On the other side, site-specific analysis allowed to detect very short glycans (GlcNAc, ManGlcNAc(2), Man(2)GlcNAc(2))), which cannot be analyzed by fluorescence due to their weak retention; furthermore, tryptic peptide portion in site-specific analysis allows a better ionization of the analyte, increasing the number of glycoforms detected.

Despite the lower number of glycoforms detected, LC-FLR-MS analysis is a powerful analytical tool to relatively quantify a great number of protein antigen glycans from different production lots; this method has a greater robustness than site-specific analysis and a very quick sample preparation (half day).

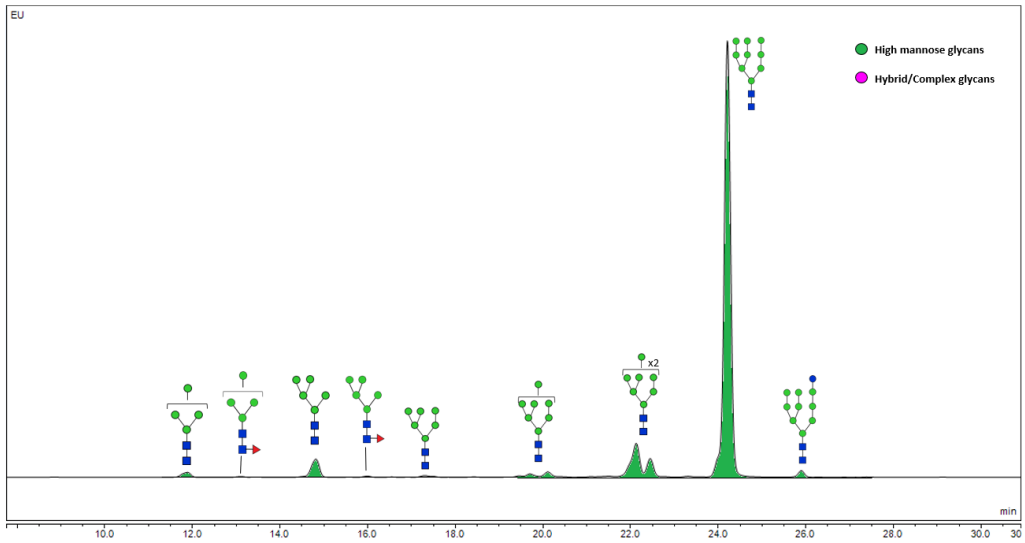


Figure 5.3. *N*-glycans profiling of $gB_{kifunensine}$.

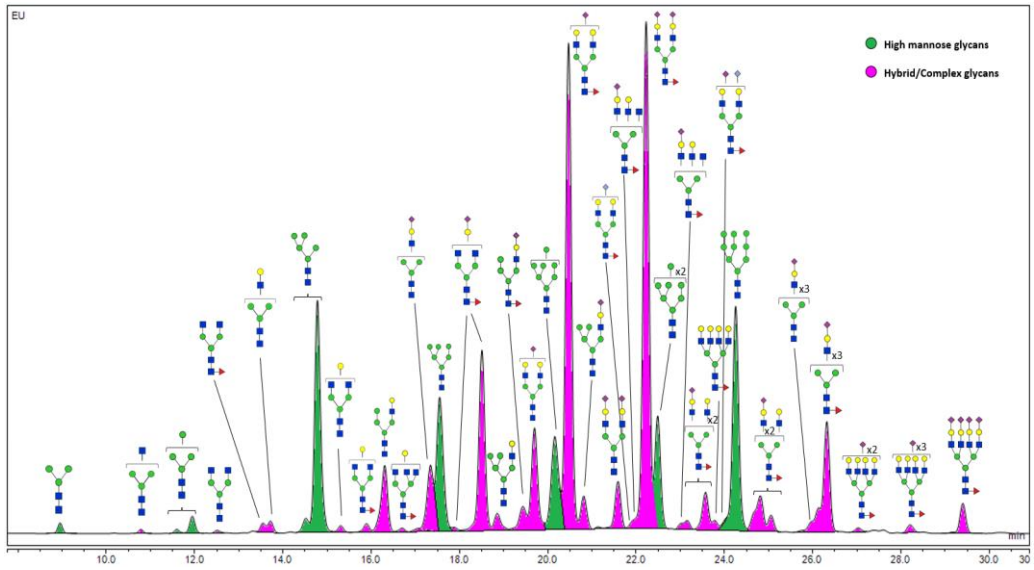


Figure 5.4. *N*-glycans profiling of gB.

Table 5.1. Released *N*-glycans from $gB_{kifunensine}$ analysis data.

***N*-glycan**	**Relative abundance (%)**
M4	1,28±0.01
M4F	0,15±0.01
M5	4,25±0.03
M5F	0,22±0.01
M6	0,44±0.02
M7	1,80±0.01
M8	10,95±0.18
M9	79,70±0.18
G1M9	1,19±0.08

Table 5.2. Released *N*-glycans from gB analysis data.

***N*-glycan**	**Relative abundance (%)**
M3	0,31±0.01
N1	0,11±0.01
M4	0,66±0.02
N2	0,08±0.01
N2F	0,27±0.01
G1M3	0,37±0.01
M5	8,79±0.03
G1	0,23±0.01
G1F	0,32±0.01
G1M4	2,65±0.01
G1N3F	0,14±0.01
S1M3F	2,62±0.01
M6	5,38±0.03
S1G1F	6,63±0.03
G1M5	0,70±0.01
S1M4F	1,03±0.01
S1G2	3,80±0.01
M7	4,11±0.01
S1G2F	17,11±0.07
S1M5G1	1,18±0.01
S2G2	1,61±0.01
S'1G2F+S1G1N3F2	0,20±0.02
S2G2F	18,16±0.02
M8	3,83±0.02
S1G2N1F	0,16±0.01
S1G3F	1,50±0.20
S1S'1F+ G4F	0,50±0.20
M9	7,17±1.37
S2G3F	3,43±1.02
S3	0,46±0.40
S3G3F	4,10±0.78
S2G4F	0,95±0.80
S3G4F	0,27±0.02
S4G4F	0,94±0.11

5.2.2 *N*-glycans relative quantitative profiling of the recombinant Pentamer.

Total *N*-glycans analysis has been performed also on Pentamer expressed in different cell types (CHO cells and Expi293^{GnTI-} cells), has been performed by liquid chromatography-fluorescence-tandem mass spectrometry (LC-FLR-MS/MS) system. Like in the previous case, *N*-glycans shaving on Pentamer have been done using PNGaseF, then the released *N*-glycans were labeled with Rapifluor-MS.

Expi293^{GnTI-} cells are human embryonal kidney cells that do not have N-acetylglucosaminyltransferase I (GnTI) activity, encoded by the MGAT1 gene; the deletion of this gene arrests the glycans maturation at the high-mannose structure M5 (Fig. 1.1). This glycoengineering strategy offers the advantage to produce shorter oligosaccharide chains, exposing eventually shielded protein epitopes.

The *N*-glycans pattern of Pentamer expressed in Expi293^{GnTI-} cells (Pentamer$_{GnTI-}$) has shown only a predominant structure, which is the high-mannose glycan M5; in this case the glycan maturation is arrested at the MGAT1 level. Lower levels of longer high-mannose structures like M5F, M6, M7 have been detected; no complex structures have been found (Fig. 5.5, Tab. 5.3).

The *N*-glycans pattern of the Pentamer expressed in CHO cells (Pentamer) has shown S1G2F as most abundant structure, meaning that glycans reach the highest maturation degree. Significant percentages of other complex glycans, like S2G2F and S2G2 have been also found. Only two high-mannose glycans have been detected with relevant abundances: M6 and M5 (Fig. 5.6, Tab. 5.4).

Even in the case of Pentamer, released N-glycans analysis by LC-FLR-MS allowed to detect a lower number of glycan structures than site-specific analysis (28 against 37 structures); however, as already seen with gB, tetrantennary complex oligosaccharides (S2G4F, S3G4F, S4F) have been detected only by LC-FLR-MS, probably due to their strong retention on SPE cartridges used for site specific analysis (HLB Oasis).

However, the other structures found by LC-FLR-MS are the same found by site-specific analysis and the relative abundancies data of both methods are in agreement.

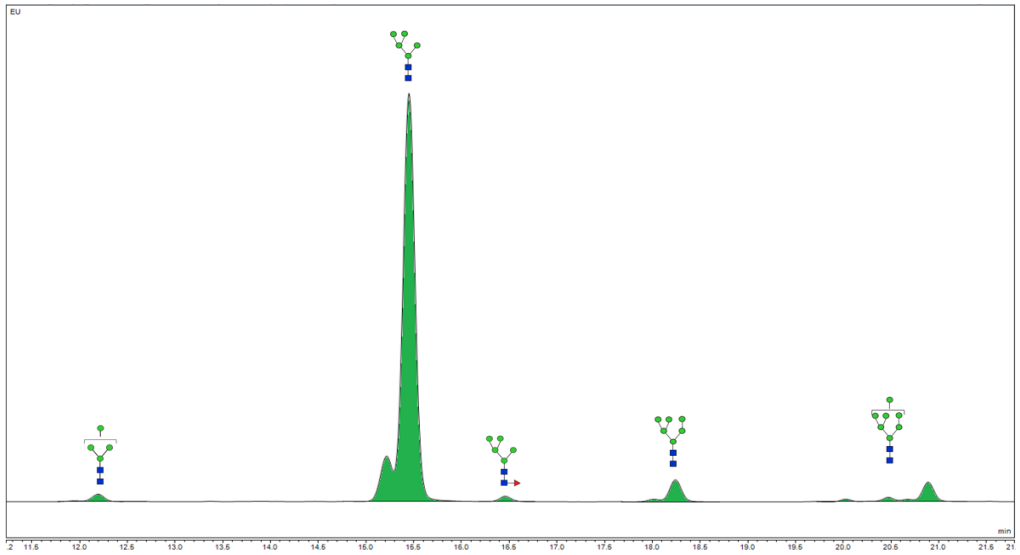


Figure 5.5. *N*-glycans profiling of Pentamer$_{GnTI-}$.

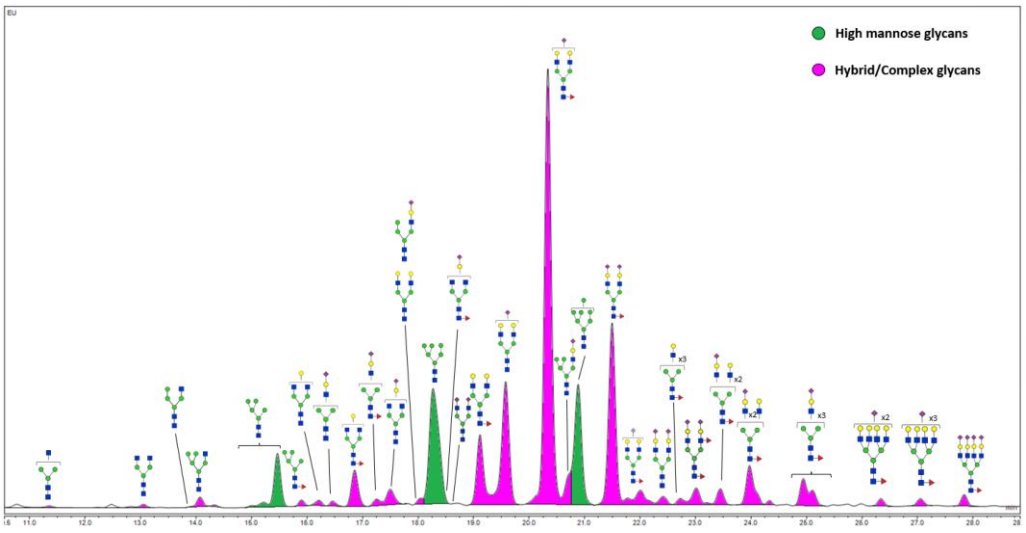


Figure 5.6. *N*-glycans profiling of Pentamer.

Table 5.3. Released *N*-glycans from Pentamer$_{GnTI-}$ analysis data.

***N*-glycan**	**Relative abundance (%)**
M4	1.73±0.04
M5	90.18±0.03
M5F	2.35±0.24
M6	4.78±0.05
M7	3.28±0.08

Table 5.4. Released *N*-glycans from Pentamer analysis data.

***N*-glycan**	**Relative abundance (%)**
G0M3	0,10±0.01
G0M4	0,47±0.31
G0M5	0,15±0.05
M5	3,83±0.02
M5F	0,33±0.01
G1	0,37±0.01
S1M3	0,31±0.01
G1F	2,41±0.01
S1M3F	0,49±0.01
S1G1	1,22±0.01
G2+S1M4	0,36±0.01
M6	10,90±0.01
G2F	4,93±0.04
S1G2	9,07±0.10
S1G2F	29,71±0.10
S1M5	1,70±0.10
M7	8,62±0.01
S2F	12,17±0.07
S1'G2F	0,52±0.02
S2	1,42±0.01
G3F	0,56±0.02
S2F2	1,26±0.04
S1G3F	1,22±0.02
S2G3F	3,39±0.04
S3F	2,78±0.01
S2G4F	0,47±0.02

S3G4F	0,52±0.02
S4G4F	0,71±0.01

5.2.3 Orthogonal confirmation of gB *N*-glycans profiling by MALDI-MS.

Total *N*-glycans analysis of gB expressed in CHO cells has been performed by MALDI-MS. More in detail, *N*-glycans shaving on gB have been done using PNGaseF, then the enzymatically released *N*-glycans have been permethylated with methyl iodide and extracted with chloroform for three times (Fig. 5.7). The samples have been loaded on a MALDI plate using DHB as matrix.

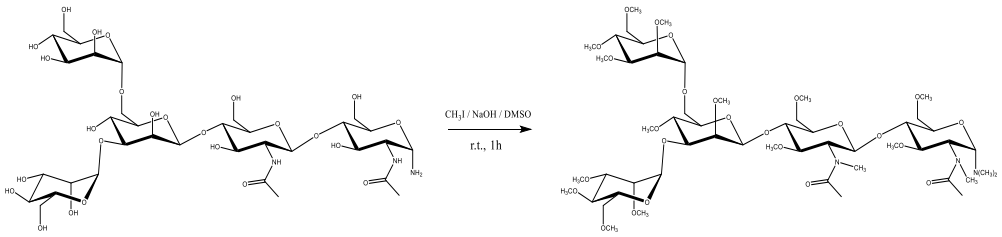


Figure 5.7. *N*-glycan permethylation reaction.

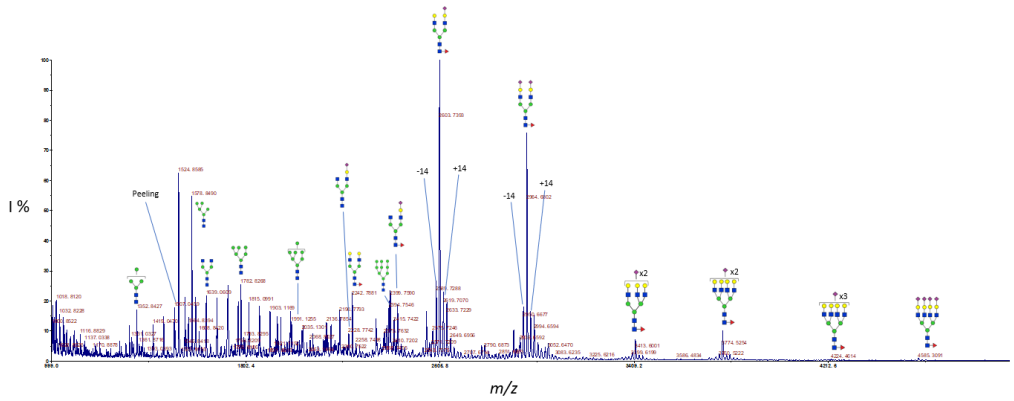


Figure 5.8. *N*-glycans profiling of gB by MALDI-MS.

The glycan profiling obtained by MALDI-MS showed S1G1F and S2G2F as the most abundant structures, in agreement with the LC-FLR profile, however the abundancies of these two species is inverted in MALDI-MS due to different ionization efficiencies (Fig. 5.8). The number of *N*-glycans found by MALDI-MS is lower than the number of *N*-glycans identified by LC-FLR.

5.2.4 Building of glycans tridimensional models on the structure of HCMV recombinant proteins gB and Pentamer.

To visualize the possible effect that the glycans found by mass spectrometry play on glycoprotein epitopes, a tridimensional model presenting the most abundant glycoform on each *N*-glycosylation site of gB was built (Fig. 5.9). It was found that two highly immunogenic epitopes AD-4 and AD-5, which produce only neutralizing antibodies, are hindered by the glycans at N240, N245 (AD-5) and N342 (AD-4), the latter with lower occupancy. This fact suggests that the glycans have an active role in the lowering of the immune response (Chandramouli et al., 2015; Spindler et al., 2014). The model has been built using GLYCAM Glycoprotein Builder function.

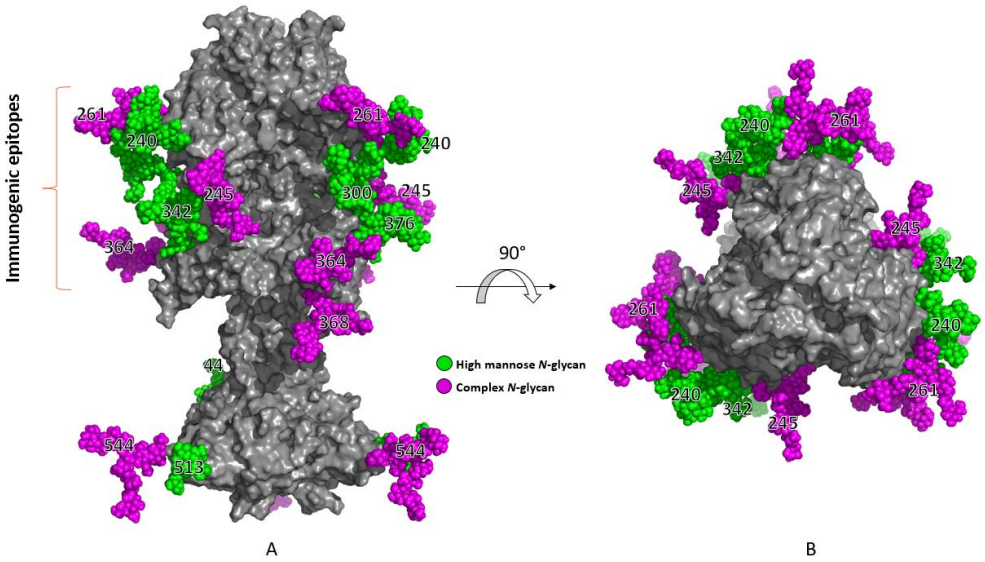


Figure 5.9. The trimer is shown as a grey surface and the glycans are shown as spheres of different colors depending on the type of glycoform most abundant at that site (highlighted in red in Tab. 4.1). (A) Side view of the model. (B) Top view of the model. N368 is resolved only in chain B and C. N44 is resolved only in chain C. X-ray crystallography could not resolve the glycans at all the site: N27, N32, N406, N411, N423, N424 are not resolved. 11 out of 18 *N*-glycosylation sites are visualized in the crystal and the glycan structures detected are uncomplete because only the first 3-6 residues are visible by X-ray crystallography. The glycoprotein region containing the most immunogenic epitopes is emphasized. The model has been built using GLYCAM Glycoprotein Builder function (Burke et al., 2015) and using the publicly available 3D structure of gB (PDB: 5CXF).

To assess the shielding that glycans play on Pentamer surface, a tridimensional model of the oligosaccharides on the protein has been built using GlycoShape; indeed, glycans are much more dynamic than folded polypeptides, exploring extensive arrays of conformations over tens of nanoseconds, hence they create molecular shields that mask large swaths of protein surface. Sequon N44 on gL is located on a flexible loop at the center of the groove formed by gH, gL and UL130 (Fig. 5.10, A); the

oligosaccharide on this site is surrounded by two glycans (N32 from gH and N60 from UL130). The steric hindrance played by the two lateral glycan structures hampers processing enzymes accessibility to glycan N44, giving rise to the oligomannose structures M7, M6 and M5 (Fig. 4.7). Glycosylation site N63 on UL131A is the most underprocessed one, with a high-mannose structures percentage of 90 % (Fig. 4.7); this consensus is sited nearby the contact region between UL128 and UL131A, which partially shields the glycan from the action of glycosidases, the glycan is wrapped up by the helicoidal protein backbone and the most external N93 glycan (Fig. 5.10, C). The most abundant complex glycan is structure S1G2F, which has been found on half of the processed glycosylation sites (UL130, N60, N93 and gH N169); otherwise, biantennary complex glycans (S2F) and non-sialylated glycans (G1) have been found. Despite occupied at nearly 100 %, sites N39 and N44 on gH did not show glycoforms according to the analytical method used; nevertheless, the presence of glycan structures on these sites has been further confirmed by glycopeptide mapping experiments after *N*-glycans trimming with the enzymes EndoH and EndoF3, which selectively cleave different types of *N*-glycan structures within the chitobiose core, leaving a single GlcNAc residue attached to the asparagine.

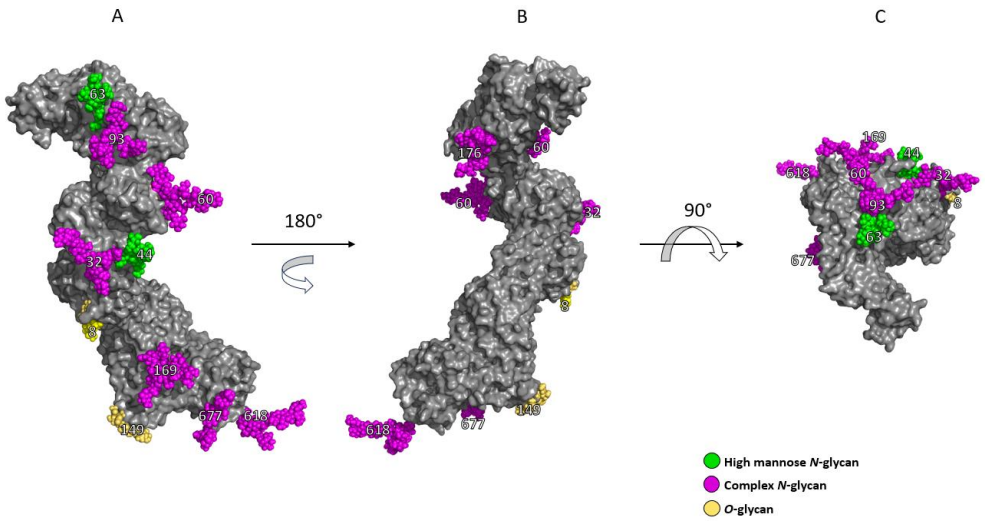


Figure 5.10. The Pentamer protein backbone is shown as a gray surface and the glycans are shown as spheres of different colors depending on the most abundant glycoform at that site. A) Side view of the model. B) Side view of the model. C) Top view of the model. Sites N39 (gH), and N44 (gH) have not been reported, S4 (gL) is not present in the crystal (9 out 11 *N*-glycosylation sites and 2 out 3 *O*-glycosylation sites [see Chapter 7 for detailed results]) are visualized in the crystal). The glycosylated model was constructed using the publicly available 3D structure (PDB: 5VOD) and GlycoShape (Ives, 2024).

5.3 Discussion

N-glycans profiling experiments have been performed on gB, $gB_{kifunensine}$, Pentamer and $Pentamer_{GnTI-}$. The different expression conditions showed a totally different *N*-glycans profile both for gB and Pentamer, with $gB_{kifunensine}$ and $Pentamer_{GnTI-}$ having simpler profiles with only few high-mannose structures. On the other hand, gB and Pentamer from CHO cells without inhibitor showed crowded profiles due to the presence of several complex glycans structures, moreover, Pentamer bears a lower percentage of high-mannose oligosaccharides than gB, probably because of its less compact structure.

A comparison between the *N*-glycans profile by LC-FLR with that by MALDI-MS has been also performed: the latter detected a lower number of glycan structures, nevertheless, the most abundant structures are S1G2F and S2G2F in both, but their intensities are reversed.

Chapter 6 – Monosaccharide identification and quantitative analysis on recombinant gB

6.1 Introduction

Despite being employed for nearly 60 years for sugar analysis, Gas Chromatography-Mass Spectrometry (GC-MS) still remains a powerful method for the determination of the monosaccharide composition of glycoproteins. In fact, GC-MS allows to obtain information about the nature of the glycans building blocks with excellent sensitivity and resolution. This kind of analysis is based on the quantitative derivatization of monosaccharides into volatile molecules by acylation reaction (De Castro et al., 2010; Niedermeier, 1978; Ruiz-Matute et al., 2011).

Oligosaccharides can be cleaved to give the monomers by different strategies and depending on the kind of needed information. Indeed, lysis can be done either by acid hydrolysis or by acid methanolysis protocols. The hydrolysis protocol requires the transformation of monosaccharides in the corresponding alditols; the advantage of this approach consists of the fact that aldoses give only one peak, a feature that increases both the sensitivity of the detection and the accuracy of the quantitation. However, this procedure is limited to neutral or basic monosaccharides. On the other hand, methanolysis procedure causes less degradation of the monosaccharides, it is very effective in cleaving glycosidic linkages, and it allows the simultaneous analysis of neutral monosaccharides, aminosugars, uronic acids, ulosonic acids like Kdo and sialic acids (which are very sensitive to hydrolysis conditions). The main drawback is that the chromatogram appears crowded because each residue gives more than one peak due to presence of different anomeric configurations (Gerwig, 2007).

Monosaccharides can be analyzed by using liquid chromatography techniques, as well. Over the last 20 years, the monosaccharides analysis by LC-MS has undergone significant advances; nevertheless, derivatization of monosaccharides is indispensable to improve the sensitivity of detection, due to the low ionization efficiency of monosaccharides. Monosaccharides can be analyzed by using liquid chromatography techniques, as well. Over the last 20 years, the monosaccharides analysis by LC-MS has undergone significant advances; nevertheless, derivatization of monosaccharides is indispensable to improve the sensitivity of detection, due to the low ionization efficiency of monosaccharides.

The 1-phenyl-3-methyl-5-pyrazolone (PMP) is one of the most popular fluorescent labels that reacts with reducing monosaccharides under mild conditions, where two PMP molecules form a bis-adduct with the monosaccharide by Michael addition (Bailly et al., 2020); monosaccharides derivatization with PMP increases their hydrophobicity, allowing the usage of reverse phase C18 columns for their separation (L. Zhang et al., 2003). In this way mass spectrometry qualitative outputs support the quantitative spectrophotometric detection (Fan et al., 2019).

Another powerful LC technique for monosaccharides analysis is high-performance anion-exchange chromatography (HPAE) hyphenated with pulsed amperometric detection (PAD). HPAE-PAD is a fast monosaccharide analysis method, which directly quantifies nonderivatized carbohydrates with high sensitivity and selectivity. This chromatographic technique exploits the carbohydrates weak acidity, which at high pH are partially ionized and thus can be separated by anion-exchange mechanisms. The PAD system consists of a specific electrode surface which measures the current resulting from the oxidation or reduction of a small aliquot of the analyte (Rohrer et al., 2013). Typically, monosaccharides are separated on a Thermo Scientific Dionex CarboPac PA10 or PA20 anion-exchange column, preceded by a Thermo Scientific Dionex AminoTrap guard column, using a hydroxide eluent (10-100 mM NaOH or KOH). In glycoprotein monosaccharides analysis the trifluoroacetic acid

hydrolysis is optimized for neutral sugar (Fuc, Gal, Glc, Man), whereas the HCl hydrolysis is optimized for amino sugar (GalN, GlcN) (Basumallick et al., 2016).

6.2 Monosaccharide profile by LC-UV-MS/MS

Monosaccharide profiling has been performed also using LC-UV-MS/MS; this analysis has been done performing monosaccharides labeling by 1-phenyl-3-methyl-5-pyrazolone (PMP), which is a chromophore with UV absorption at 245 nm; each reducing monosaccharide reacts with two PMP molecules, giving a bis-PMP adduct (Fig. 6.1). Unlike GC-MS analysis, the PMP labeling does not allow to detect Neu5Ac because of its α-keto-acidic structure.

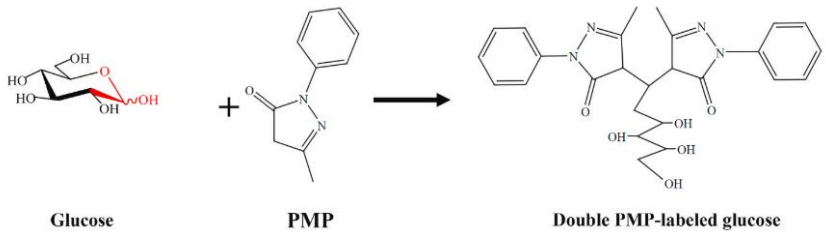


Figure 6.1. Reaction between reducing sugar and PMP.

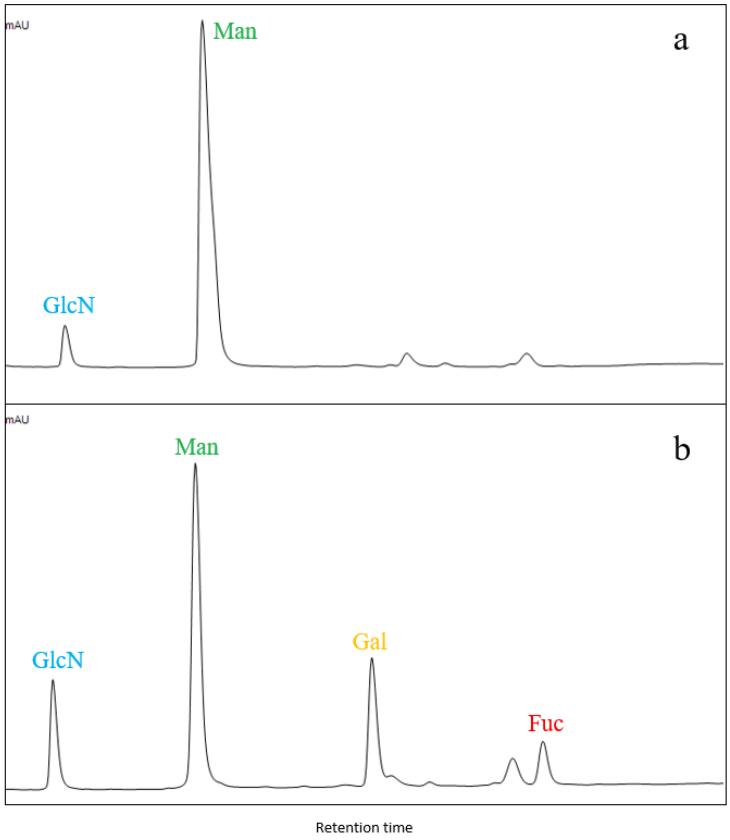


Figure 6.2. LC-UV chromatograms of monosaccharides from gB (a) and $gB_{kifunensine}$ (b). *Contaminant originating from sample preparation.

The monosaccharides found on gB, according to the composition of the eukaryotic complex *N*-glycans are: fucose, mannose, galactose and *N*-acetylglucosamine (Fig. 6.2, a).

Two different monosaccharides have been found on $gB_{kifunensine}$, reflecting the kifunensine effect on glycan maturation: mannose and *N*-acetylglucosamine (Fig. 6.2, b).

Monosaccharides relative abundancies have been measured with good reproducibility (Tab. 6.1). $gB_{kifunensine}$ showed very high mannose percentage (about 92 %), which is slightly overestimated compared to the M9 structure (on which mannose has

contribute of 82 %); this deviation can be explained by the fact that the cleavage of *N*-glycosidic bond between the first GlcNAc residue and the Asn from protein does not occur, furthermore, internal GlcNAc residue hydrolysis proceeds slower than Man, therefore, uncleaved GlcNAc can be present.

The most abundant monosaccharide on gB is mannose (Tab. 6.1, about 56 %) in agreement with the fact that it is always present in the core motif of each glycan structure; however, in this case it is not possible to check the correct monosaccharides stoichiometry using a single glycan structure, due to the presence of a big number of different oligosaccharides with significative abundances.

Table 6.1. $gB_{kifunensine}$ and gB monosaccharides relative quantification by LC-UV.

$gB_{kifunensine}$	
GlcN (%)	**Man (%)**
7.64±0.54	92.36±0.54

gB			
GlcN (%)	**Gal (%)**	**Man (%)**	**Fuc (%)**
16.86±1.31	20.09±1.65	55.89±1.04	7.16±0.16

6.3 Monosaccharide absolute quantification by GC-MS

The monosaccharide composition of the glycans from of gB produced in CHO cells, with ($gB_{kifunensine}$) and without (gB) kifunensine was determined using the Internal Standard (ISTD) method, by gas chromatography-mass spectrometry (GC-MS). To analyze monosaccharides by GC-MS it is necessary both to degrade glycans in their monosaccharide building blocks and derivatize the latter to make them volatile; this has been carried out by methanolysis followed by acetylation (Fig. 6.3). Acetylated methyl glycosides (AMG) procedure is suitable for the detection of almost all type of monosaccharides, (neutral or basic aldoses, uronic and ulosonic acids), while it fails

in the detection of ketoses, as fructose. Differently from hydrolysis, methanolysis offers milder reaction conditions, preventing the monosaccharides transformation in dehydration products, allowing in this way the detection of labile monosaccharides too, such as sialic acids (Gerwig, 2007).

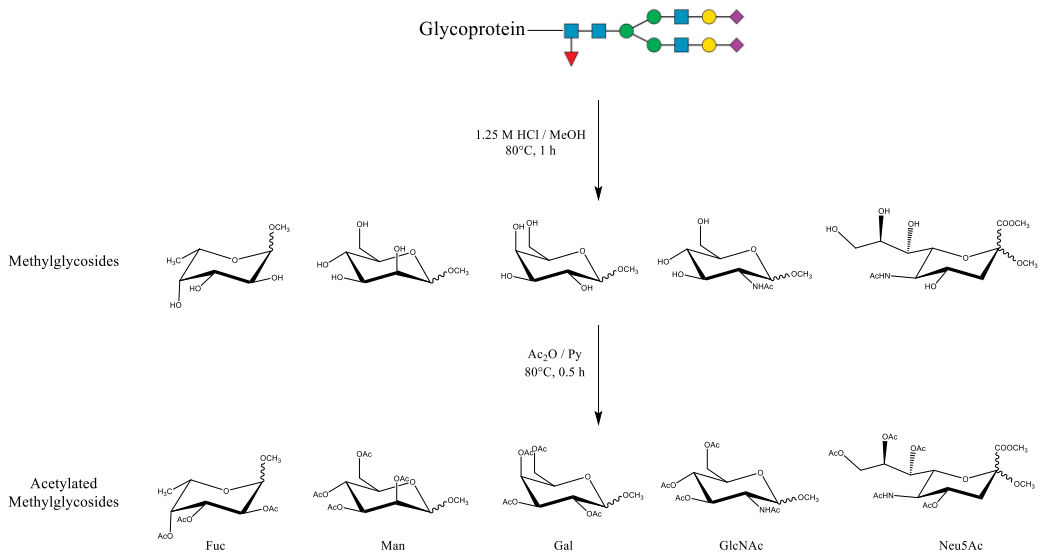


Figure 6.3. Glycans methanolysis and monosaccharides derivatization reactions.

Glycoprotein monosaccharides identities were attributed via both EI-MS interpretation (Fig. 6.4) and retention times comparison with monosaccharide standards; indeed, epimeric monosaccharides (*e.g.,* mannose and galactose) show the same MS fragmentation spectra, hampering their distinction with MS, on the other side they have different retention times.

This kind of analysis is very useful to assess the presence of non-self monosaccharides (such as Neu5Gc) and to identify the nature of all the monosaccharide building blocks in a univocal way; in fact, oligosaccharides with different monomers can give isobaric species which make difficult monosaccharide composition identification if MS-tandem experiments are not available.

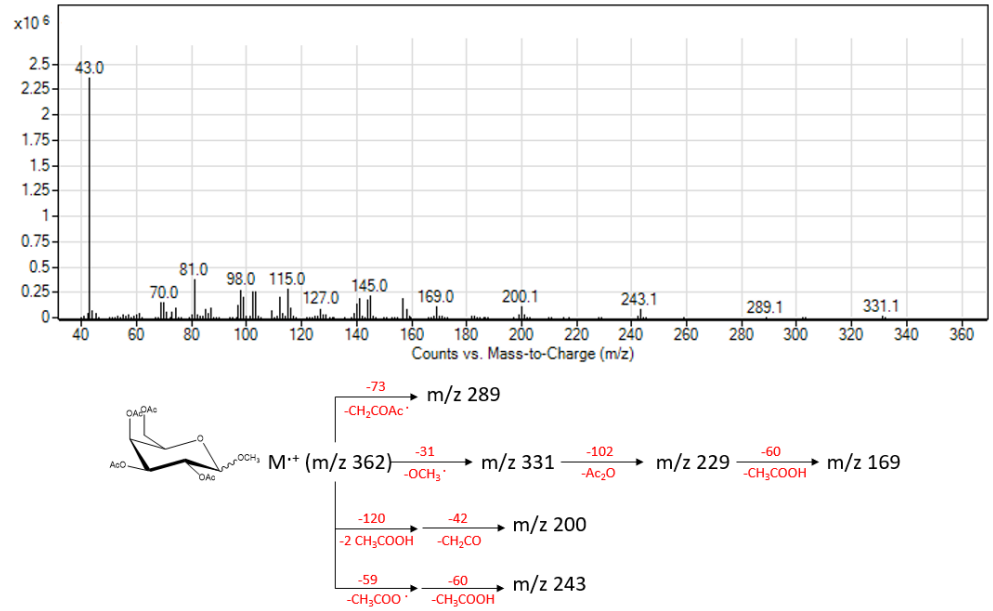


Figure 6.4. Galactose AMG electron ionization spectrum.

Two different monosaccharides have been found on $gB_{kif.}$, according to the composition of the eukaryotic high-mannose *N*-glycans: mannose and *N*-acetylglucosamine (Fig. 6.5).

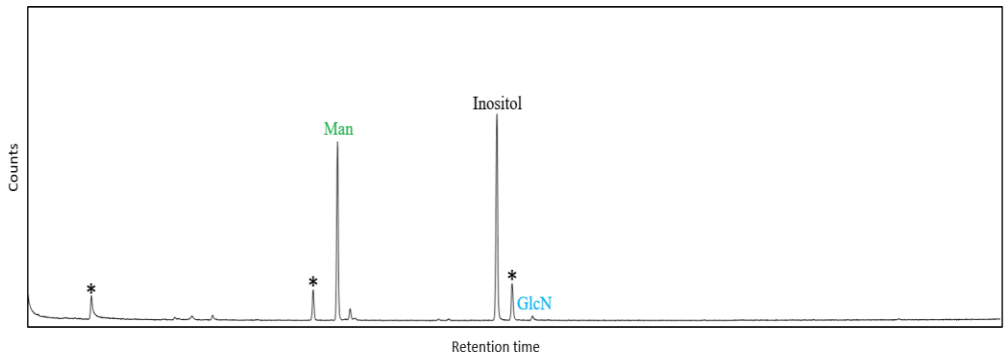


Figure 6.5. Gas chromatogram of acetylated methyl glycosides from $gB_{kif.}$ (with inositol). *Contaminants originating from used solvents.

Five different monosaccharides have been found in gB, according to the composition of the eukaryotic complex *N*-glycans: fucose, mannose, galactose, N-acetylglucosamine and N-acetylneuraminic acid. It is interesting to underline that fucose signals are split in 2 adjacent peaks, which correspond to the α and β anomers.

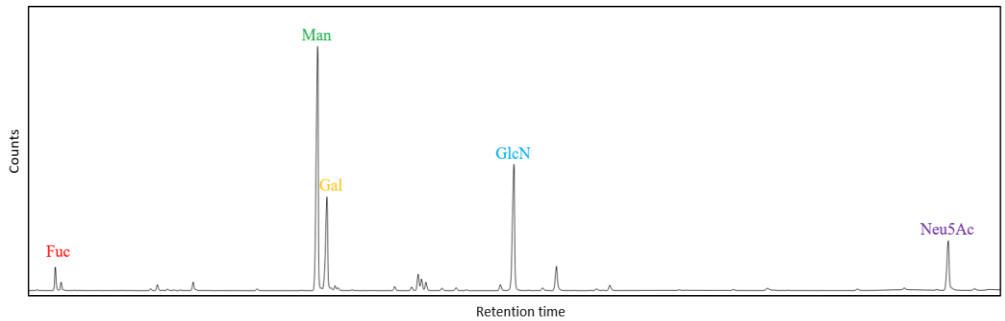


Figure. 6.6. Gas chromatogram of acetylated methyl glycosides from gB (without inositol).

Monosaccharides absolute quantification has been done by internal standard method (ISTD method), using *myo*-inositol as internal standard; this molecule shares a molecular structure similar to the monosaccharides and its retention time does not overlap with any of them. A calibration curve has been built for each monosaccharide, using five calibration levels: 0.069, 0.172, 0.345, 0.448, 0.552 mg/mL (Fig. 6.7).
For each calibration curve the internal standard concentration is of 0.270 mg/mL.

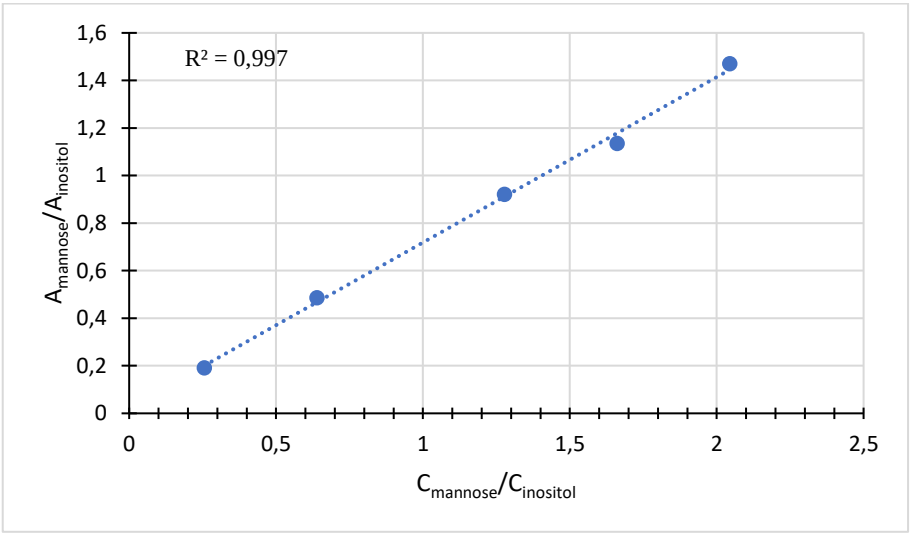


Figure 6.7. Example of calibration curve: calibration curve of mannose.

In the case of $gB_{kifunensine}$, the most abundant monosaccharide was mannose, while GlcNAc peak had very low area. In the case of gB it was found that the most abundant monosaccharides were mannose and *N*-acetylglucosamine. This was expected since these two monosaccharides are shared by all glycans because they constitute the conserved core structure.

Furthermore, a considerable amount of *N*-acetylneuraminic acid was found, confirming the presence of complex glycan structures.

Table 6.2. $gB_{kifunensine}$ and gB monosaccharides absolute quantification by GC-MS.

$gB_{kifunensine}$				
Fuc (µg/mg)	**Man (µg/mg)**	**Gal (µg/mg)**	**GlcNAc (µg/mg)**	**Neu5Ac (µg/mg)**
-	362.85±5.3	-	-	-

gB				
Fuc (μg/mg)	**Man (μg/mg)**	**Gal (μg/mg)**	**GlcNAc (μg/mg)**	**Neu5Ac (μg/mg)**
8.85±1.70	86.11±1.00	28.83±0.60	45.18±2.77	27.04±4.69

Importantly, Neu5Gc has not been detected in full MS scan mode, suggesting that the quantity of this monosaccharide is very low. Furthermore, the GC-MS analysis has been repeated in SIM mode to increase sensitivity, monitoring the ion with *m/z* 504 from Neu5Gc; even in this case the corresponding peak was among the background noise.

Neu5Gc is a labile monosaccharide, which converts into NeuN when exposed to acid pH and high temperature too for long (Fig. 6.8).

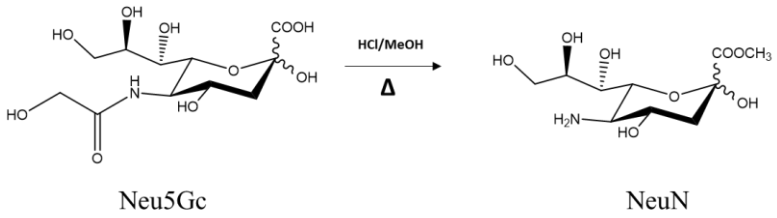


Figure 6.8. Reaction of degradation of Neu5Gc.

To optimize the methanolysis conditions in order to minimize Neu5Gc degradation, the effect of temperature, pH and time on the methanolysis reaction of Neu5Gc have been studied.

It was found that the minimum temperature to have quantitative monosaccharides derivatization was of 80°C with an HCl concentration of 1.25 M. The reaction time that allows for a compromise to be achieved between Neu5Gc degradation and oligosaccharides methanolysis is of at least 1h; by measuring three technical

replicates, it has been seen that in these conditions Neu5Gc degradation is of (54.6 ± 3.4) %. After 4 hours of methanolysis, the Neu5Gc is almost totally converted in Neu5Ac (Fig. 6.9).

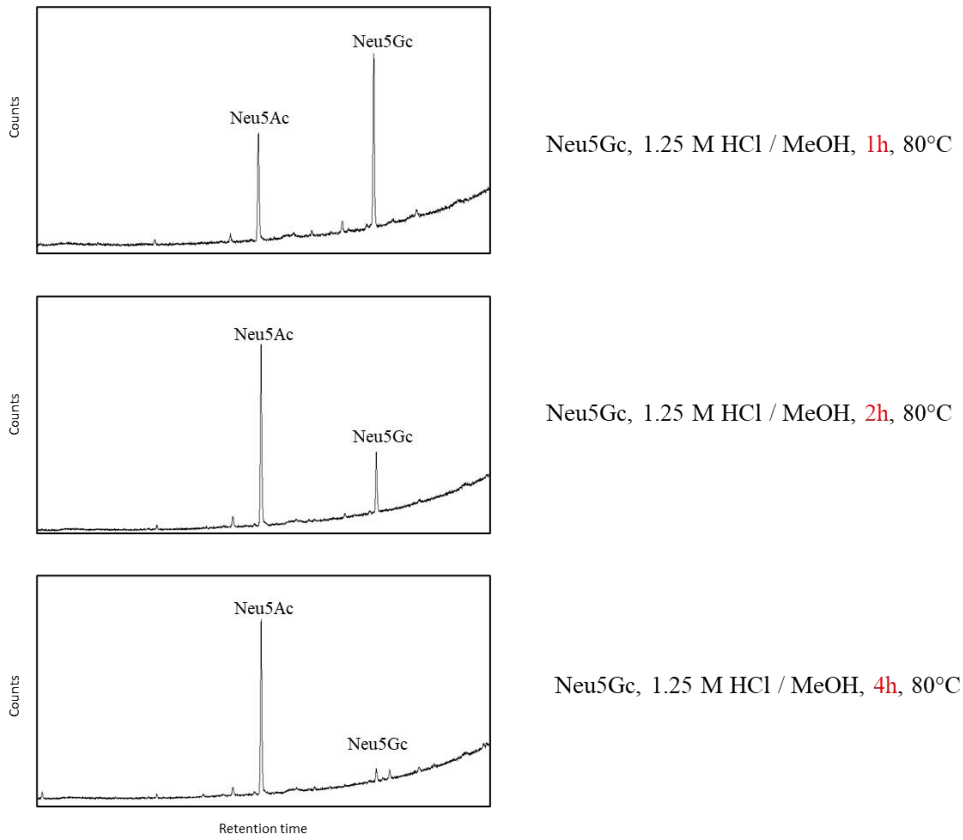


Figure 6.9. Gas chromatograms of acetylated methyl glycosides of Neu5Ac and Neu5Gc, obtained starting from a Neu5Gc standard and varying the methanolysis time.

The contribution of glycans to gB molecular weight measured by GC-MS analysis is of 20 %, against the 23 % found with denaturing gel analysis after PNGase F treatment (Tab. 6.2, Fig. 6.10). The percentage is lower in the first case because GlcNAc is underestimated due to no quantitative cleavage of GlcNAc-Asn linkage.

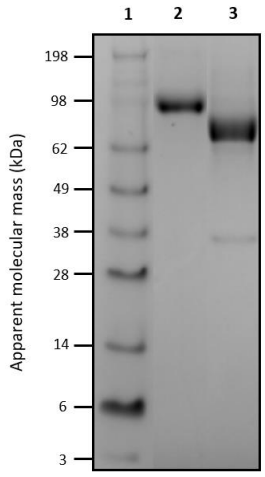


Figure 6.10. Denaturing gel electrophoresis of gB from CHO cells. Lane 1: SeeBlue Plus2 Pre-stained Protein Standard. Lane 2: gB without PNGase F treatment. Lane 3: gB with PNGase F treatment. Proteins were separated on a NuPAGE 4–12% Bis-Tris gel and stained by Coomassie-blue.

6.4 Discussion

Two methods have been developed to determine the quantitative monosaccharide composition of gB and $gB_{kifunensine}$ as model compounds by LC-UV-MS and GC-MS. Unlike LC-UV-MS, GC-MS allows to detect sialic acid too, which is not detectable in the first case because it does not react with PMP. Monosaccharide profiles of gB from different expression conditions have been compared.

One of the main advantages of LC-UV-MS method is its higher sensitivity, indeed it showed a higher limit of detection (LOD) than GC-MS method.

$$LOD_{LC\text{-}UV\text{-}MS} = 1.40\ \mu g/mL$$

$$LOD_{GC\text{-}MS} = 47.54\ \mu g/mL$$

The limit of detection (LOD) of an analytical method can be defined as the minimum sample concentration that can be analyzed with the method in exam. The $LOD_{5\%}$ value refers to a concentration value with a 5 % probability to give a false negative result;

LOD $_{5\%}$ can be obtained from two calibration curve parameters: the error on the intercept (σ_a) and the slope value (b):

$$\text{LOD}_{5\%} = \frac{3.3\sigma_a}{b}$$

Furthermore, LC-UV-MS method showed better repeatability than GC-MS when doing relative quantification.

Chapter 7 – *O*-glycosylation sites mapping on Pentamer

7.1 Introduction

The most common *O*-glycosylation pathway is through GalNAc-*O*-glycosylation occurring in the Golgi apparatus, after protein folding; *O*-GalNAc glycans are involved in almost every aspect of biology, including cell–cell communication, cell adhesion, signal transduction, immune surveillance, epithelial cell protection, and host–pathogen interactions (Magalhaes et al., 2021). The *O*-GalNAc glycans of mucins have four major core structures (cores 1–4) (Fig. 1.2). Each core can be extended by a variety of sugar residues to give linear or branched chains that resemble those on *N*-glycans and glycolipids. Core 1 and 2 O-GalNAc glycans are found in glycoproteins and mucins produced in many different cell types. However, core 3 and 4 O-GalNAc glycans are more restricted to mucins and glycoproteins in gastrointestinal and bronchial tissues.

O-glycans of viral glycoproteins appear in two types of patterns (Fig. 7.1): single *O*-glycans scattered along the amino acid sequence of certain viral glycoproteins and multiple *O*-glycans in densely packed clusters, called mucin-like domains (MLDs) (Olofsson et al., 2023).

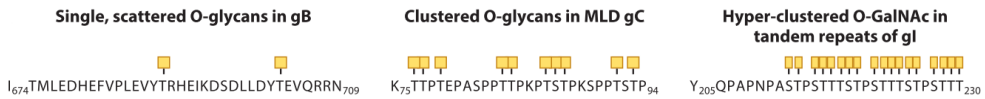


Figure 7.1. Examples of scattered and clustered *O*-glycans from the HSV-1 glycoproteins gB, gC, and gI. From *Structure and Role of O-Linked Glycans in Viral Envelope Proteins.* (Olofsson et al., 2023).

Two types of mature *O*-glycans are of special interest in the viral context: *O*-glycans terminated by N-acetylneuraminic acid (Neu5Ac), whose negatively charged carboxyl group can be relevant for early viral interactions, and large *O*-glycans terminated with rarely expressed carbohydrate epitopes (*e.g.,* sialyl Lewis X).

O-glycans of the viral MLDs ease the diffusion of GAG-binding viruses toward the cell surface through the thick and sticky glycosaminoglycan (GAG) layers. The model is based on the interactions of three key actors: the GAG-binding site and the MLD of HSV-1 gC on one hand and the surface GAGs (here CS) of the target cell on the other. Thus, charged O-glycans in the gC MLD determine the MLD ability to electrostatically counteract gC binding to CS (Fig. 7.2).

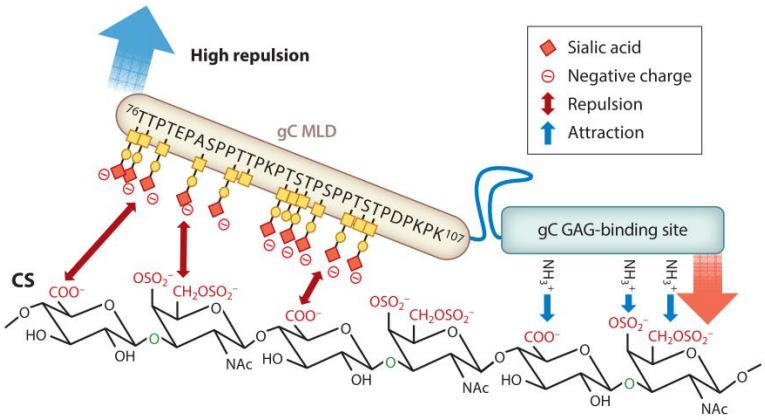


Figure 7.2. Model explaining the high regulatory potential of a viral glycoprotein MLD. From *Structure and Role of O-Linked Glycans in Viral Envelope Proteins.* (Olofsson et al., 2023).

7.2 *O*-glycans microheterogeneity analysis by LC-MS/MS (EThcD)

The need to detail site-specific glycosylation patterns on glycoproteins is increasing due to the need to understand the role glycans play within biological systems. This has driven the interest in "bottom-up" glycomic approaches that eliminate the need to remove (and potentially degrade) the glycan from the glycoprotein. Bottom up or "shotgun glycoproteomic" utilizes a protease to cleave the glycoprotein into shorter glycopeptides. ECD has also been shown using low energy electrons (<0.2 eV) to preferentially cleave the protein backbone of glycopeptides, helping to outline the site-specific binding of glycans. Comparably, ETD has been suggested as a technique for similar fragmentation of the peptide back- bone. Electron-Transfer Dissociation (ETD) has the additional benefit of maintaining O-GlcNAc linkages to the protein compared to fragmentation techniques such as CID. In this instance, because O-GlcNAc is a labile PTM group similar to phosphorylation, it is particularly susceptible to glycosidic cleavage by CID, making site determination difficult (Wilkinson et al., 2020).

To describe the Pentamer *O*-glycosylation pattern we used nano-LC-MS/MS in combination with different fragmentation methods; indeed, HCD allows to obtain information about both peptide primary sequence and glycan structure, however site-specific information is lost due to O-GlcNAc linkage lability. On the other hand, ETD promotes only the fragmentation of the peptide portion, maintaining O- GlcNAc linkages to the protein (Wilkinson et al., 2020). In this study, a combination of both HCD and ETD, called EThcD, has been exploited, in order to retain all the information concerning peptide structure, glycan structure and glycosylation site (Yu et al., 2017) (Fig. 7.3); in fact, complementary b/y and c/z ion series are obtained in the same MS/MS spectrum.

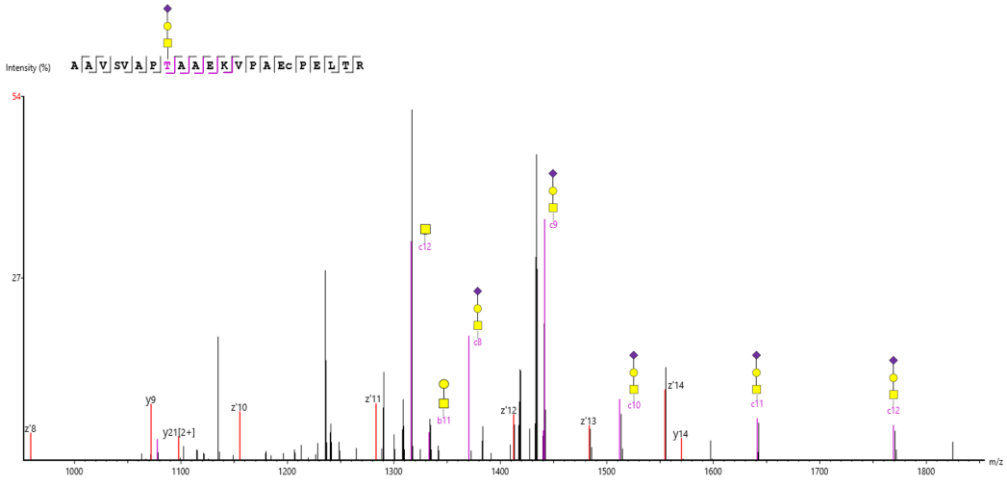


Figure 7.3. EThcD MS/MS spectrum of a tryptic *O*-glycopeptide from Pentamer.

A bottom-up glycoproteomic approach was used to study Pentamer *O*-glycosylations; after denaturation, Pentamer glycopeptides were generated by digestion with either trypsin or chymotrypsin and subjected to in-line LC-MS. As well as N-glycopeptides, *O*-glycopeptides were selected according to specific analytical requirements: specific cleavage, presence peptide ions bringing the intact glycan portion and good primary sequence coverage (both c and z ion series must be detected). The *O*-glycosylation mapping has been performed using the function ‘’glycan search’’ of PEAKS Studio 11, which allows to obtain a full characterization of the glycopeptides MS/MS spectra. With this strategy three Pentamer O-glycosylation sites have been identified: Ser^{4} and Thr^{8} (on gL) and T^{149} (on gH). Ser4 brings only 1 type of glycan structure (sialyl T antigen), Thr^{8} brings three different types of glycan structures (sialyl T antigen, T antigen and Tn antigen), as well as Thr^{149} (sialyl T antigen, Core 2 and GlcNAc sT antigen). Through the corresponding extracted ion chromatogram and its integration both microheterogeneity and occupancy have been assessed for each *O*-glycosylation site; site S4 has an occupancy of 0.95 % and only one glycan structure has been identified on it, while site Thr^{8} showed an occupancy of 22.6 % and three different

glycoforms on it, with sT antigen as most abundant one, site Thr[149] showed an occupancy of 52.5 % and three different glycoforms, with Core 2 oligosaccharide as most abundant (Fig. 7.4).

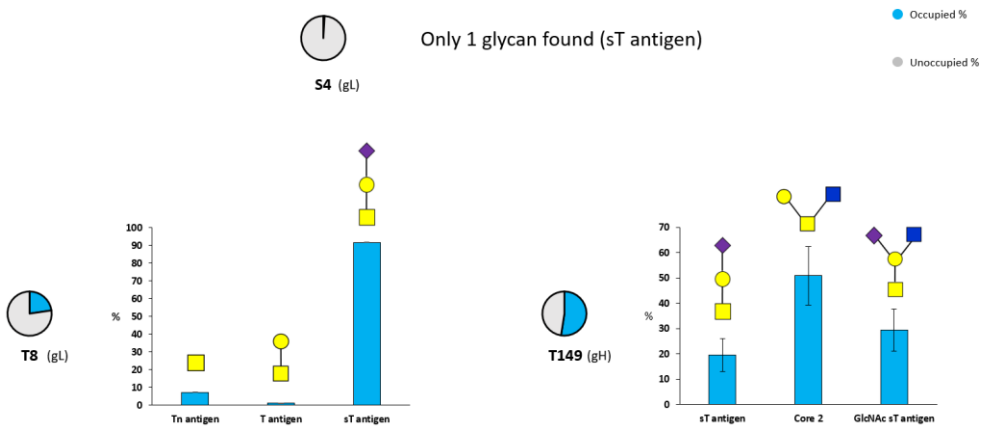


Figure 7.4. Occupancy (pie charts) and microheterogeneity (histogram) of gL *O*-glycosylation sites S4 and T8.

7.3 Discussion

Pentamer *O*-glycosylation pattern has been studied using tandem mass spectrometry with different fragmentation methods: HCD, ETD and EThcD. A boost in sensitivity for this kind of analysis has been obtained by using a nano-LC system. The analysis brought to the identification of three *O*-glycosylation sites (Ser[4], Thr[8] and Thr[149]) and five different glycoforms (sialyl T antigen, T antigen and Tn antigen, Core 2 and GlcNAc sT antigen).

Chapter 8 – Immunogenicity evaluation in animal model and impact of glycans

8.1 Introduction

High titers of serum neutralizing antibodies (nAbs) after antigen inoculation usually correlate with protection against the infection. There are several definitions of neutralization, two of which are widely accepted. One is “the loss of infectivity which ensues when antibody molecule(s) bind to a virus particle, and usually occurs without the involvement of any other agency’’; a second definition of neutralization is “the reduction in viral infectivity by the binding of antibodies to the surface of viral particles (virions), thereby blocking a step in the viral replication cycle that precedes virally encoded transcription or synthesis’’. For enveloped viruses, this block occurs before virus entry into a host cell but for non-enveloped viruses, it can occur after entry. The ability of nAbs to block viral entry by enveloped viruses *in vitro* requires that the antibodies bind to functional entry molecules on the surface of infectious virions, typically envelope (Env) protein spikes (Burton, 2023); after binding, the neutralization of enveloped virions can occur according different mechanisms (disassembly/conformational modification, virion aggregation, steric obstruction, blocking conformational changes) (Fig. 8.1).

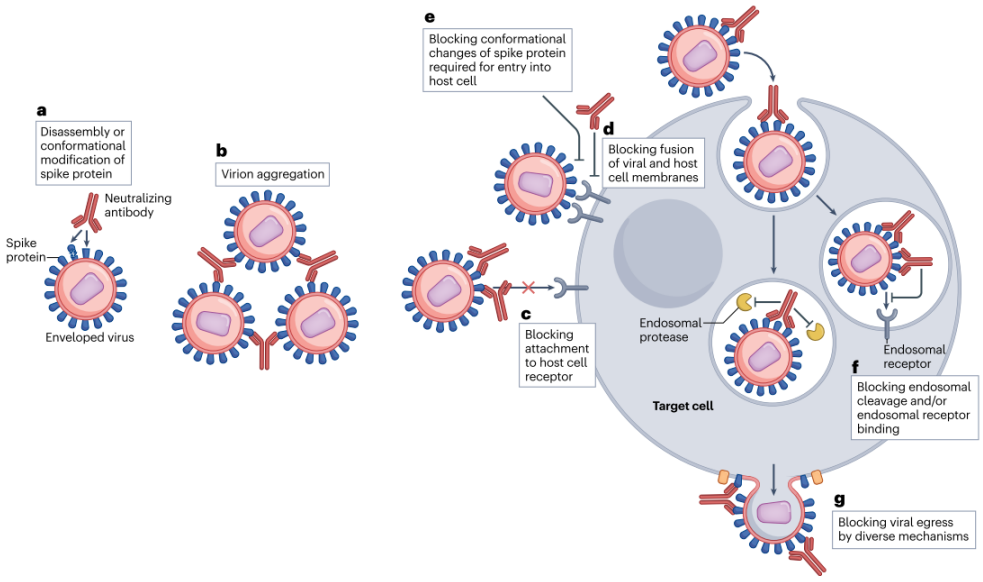


Figure 8.5. Mechanisms of neutralization of an enveloped virus by neutralizing antibodies *in vitro*. From *Antiviral neutralizing antibodies: from in vitro to in vivo activity.* (Burton, 2023).

8.2 nAbs titers of stressed gB immune mouse sera and Pentamer immune mouse sera

To determine whether $gB_{kifunensine}$ and $Pentamer_{GnTI-}$ elicit higher nAbs titers than gB and Pentamer from CHO cells without modifications, we investigated the titers induced after gB, Pentamer, $gB_{kifunensine}$ and $Pentamer_{GnTI-}$ vaccination in mice.

The nAbs titer is expressed as geometric mean (GMT) of all the measured values (over 10 biological replicates), where each measurement corresponds to the dilution according to which 95% of cells are not infected (IC_{95}) which is based on a specific starting serum dilution (the lowest dilution tested). The data have been fitted to a nonlinear regression model, which allowed us to determine the neutralizing titer elicited by each antigen at 95% confidence interval (CI).

$gB_{kifunensine}$ elicited a higher titer of nAbs (about five times) than gB; in the same way, $Pentamer_{GnTI-}$ showed a higher titer of nAbs (about six times) than Pentamer (Fig. 8.2).

Furthermore, both Pentamer and Pentamer$_{GnTI-}$ showed to be more immunogenic than gB and gB$_{kifunensine}$ correspondingly.

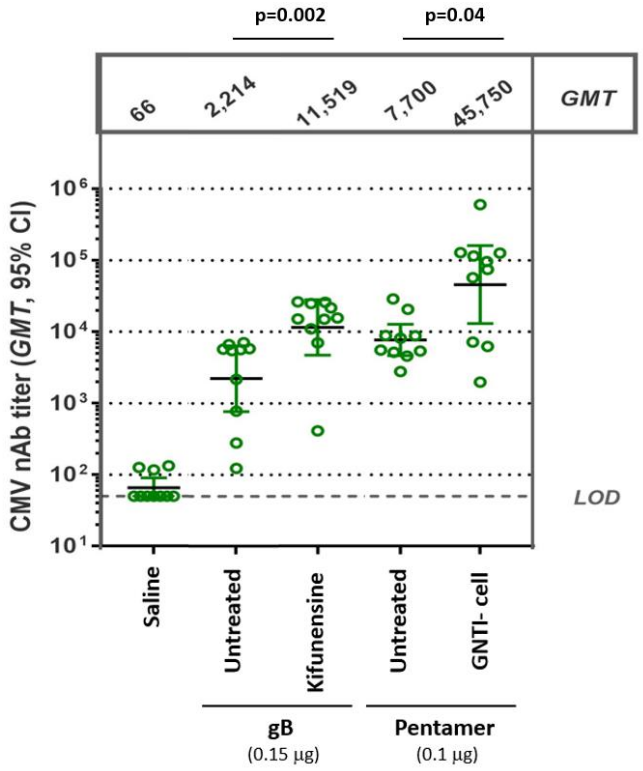


Figure 8.2. Immunogenicity of gB (from CHO cells and kifunensine-treated CHO cells) and Pentamer (from CHO cells and Expi293^{GnTI-} cells). GMT=Geometric Mean. LOD= Limit of Detection. Study in collaboration with Asma Ashraf from GSK US molecular biology and virology group.

8.3 Epitope shielding assessment

To visualize the effect that the different glycan patterns could have on immunogenicity of the studied glycoprotein antigens, tridimensional models of gB and Pentamer with different oligosaccharide shields have been compared. More in detail, the shielding of the most prominent protein epitopes has been assessed.

In the case of gB from CHO cells and gB from CHO cells treated with kifunensine, the shielding played by the glycans is almost identical, with a slightly higher masking in the second case due to the presence of only M9 structures on all the glycosylation sites, which are bulkier than M5 structure on site N342 (on the immunogenic head, epitope 1G2, Fig. 2.6) of gB from CHO cells.

For this reason, the fact that the higher immune response has been elicited by gB from CHO cells treated with kifunensine (Fig. 8.2), could be attributed to the presence of glycan epitopes involving M9 structures and not to unshielded epitopes.

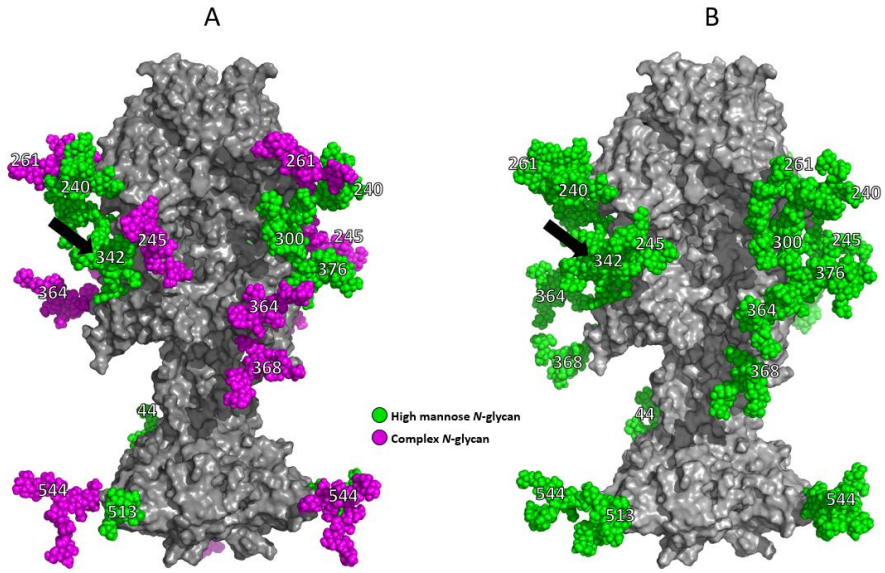


Figure 8.3. The gB backbone is shown as a gray surface and the glycans are shown as spheres of different colors depending on the most abundant glycoform at that site. A) gB from CHO cells. B) gB from CHO cells treated with kifunensine. Glycans with diverse bulkiness between the two structures are indicated by the arrows. N44 is resolved only in chain C. X-ray crystallography could not resolve the glycans at all the site: N27, N32, N406, N411, N423, N424 are not resolved. 11 out of 18 N-glycosylation sites are visualized in the crystal and the glycan structures detected are uncomplete because only the first 3-6 residues are visible by X-ray crystallography. The model has been built using GLYCAM Glycoprotein Builder function (Burke et al., 2015) and using the publicly available 3D structure of gB (PDB: 5CXF).

Otherwise, Pentamer from Expi293^{GnTI-} cells, is decorated with only M5 structures glycans, which play a lower steric hindrance than Pentamer from CHO cells; indeed, it is possible to visualize from the tridimensional model of Pentamer from Expi293^{GnTI-} cells how there is a higher portion of protein backbone exposed (*e.g.*, glycans N32, N93, N169 and N677) (Fig. 8.4).

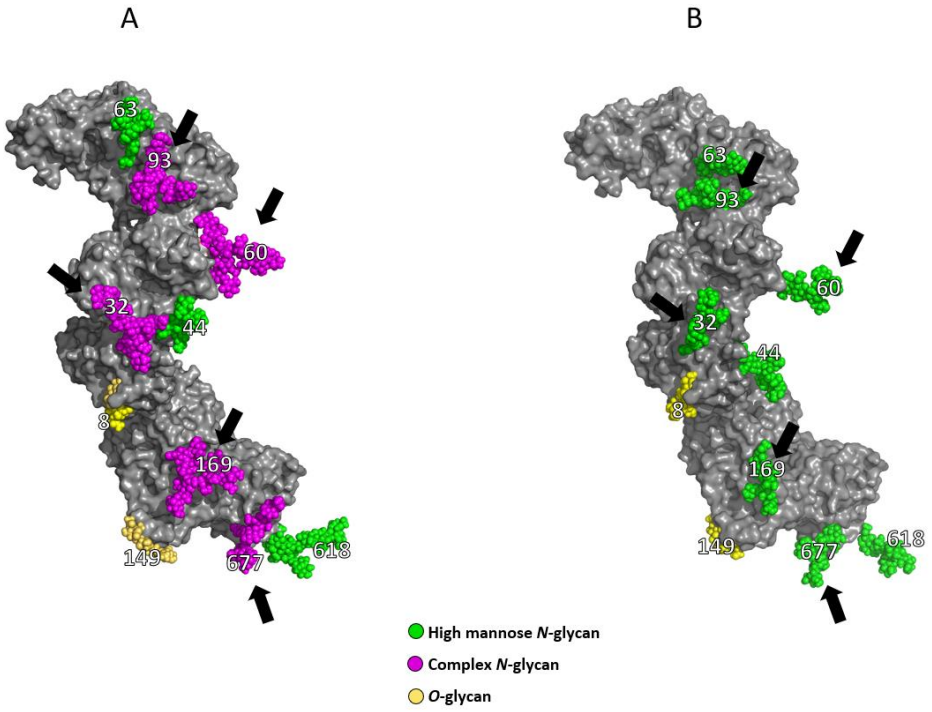


Figure 8.4. The Pentamer protein backbone is shown as a gray surface and the glycans are shown as spheres of different colors depending on the most abundant glycoform at that site. A) Pentamer from CHO cells. B) Pentamer from Expi293^{GnTI-} cells. Glycans with diverse bulkiness between the two structures are indicated by the arrows. Sites N39 (gH), and N44 (gH) have not been reported, S4 (gL) is not present in the crystal (9 out 11 *N*-glycosylation sites and 2 out 3 *O*-glycosylation sites [see Chapter 7 for detailed results]) are visualized in the crystal). The glycosylated model was constructed using the publicly available 3D structure (PDB: 5VOD) and GlycoShape (Ives, 2024).

The different shielding effect of the two glycan patterns on Pentamer has been quantified by GlycoShape, using a 3D heatmaps of accessibility (Fig. 8.5); from these maps we can see how regions next to epitopes for neutralizing antibodies (8I21, MSL-109,13H11, Fig. 2.7) are less shielded in the Pentamer from Expi293^{GnTI-} cells.

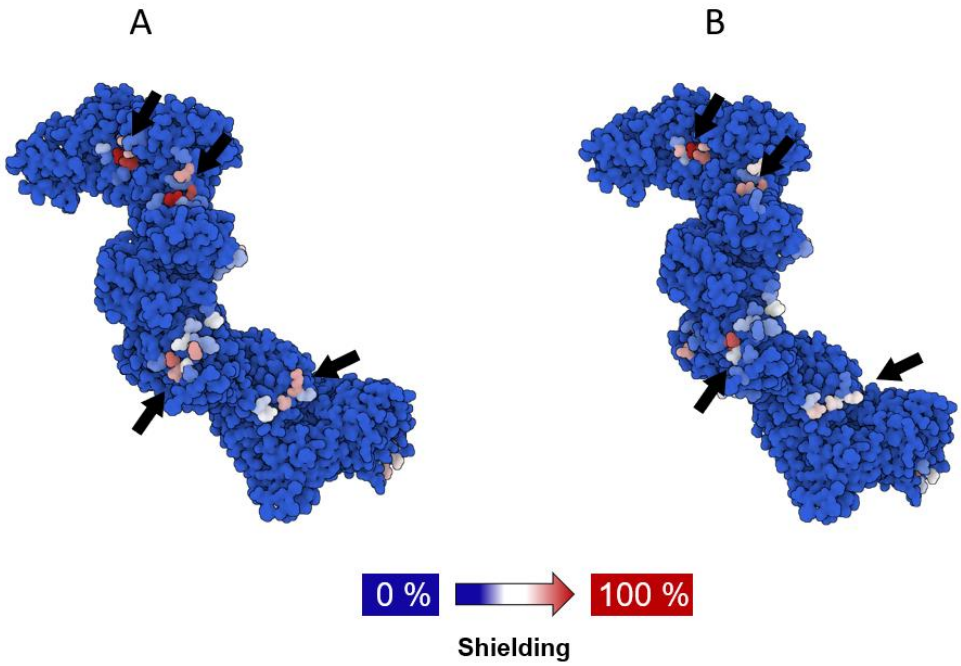


Figure 8.5. 3D heatmaps of accessibility scores of Pentamer from CHO cells (A) and Pentamer from Expi293^{GnTI-} cells (B) by GlycoShape (Ives, 2024). Higher color intensities indicate higher shielding. Protein regions with diverse shielding between the two structures are indicated by the arrows.

Differently from gB, the increase of immunogenicity of Pentamer with only M5 glycans can be attributed to two different causes: the possible presence of glycan epitopes involving M5 oligosaccharides and the less shielding of protein epitopes.

8.4 Discussion

The ability to elicit the production of neutralizing antibodies in mice has been studied both for gB and Pentamer with different glycosylation patterns; in specific the immunogenicity of gB from CHO cells, gB from CHO cells treated with kifunensine, Pentamer from CHO cells and Pentamer from Expi293^{GnTI-} cells have been compared. Both the Pentamer forms have shown to have higher immunogenicity than both gB forms; moreover, gB from CHO cells treated with kifunensine showed higher immune response than gB from CHO cells and Pentamer from Expi293^{GnTI-} cells showed higher immune response than Pentamer from CHO cells. The effect of the glycan shield on the protein epitopes has been assessed by building 3D models of the antigens. Both the forms of gB showed to have comparable shielding of the immunogenic epitopes, while Pentamer from Expi293^{GnTI-} cells showed a lower shielding of the immunogenic epitopes.

General conclusions

Mass spectrometry is a powerful technique to fully understand how *N*- and *O*-glycans decorate the surface of glycoproteins. This class of oligosaccharides plays a crucial role in the recognition phenomena between antigen and immune system and can significantly alter the elicitation of neutralizing antibodies induced by a glycoprotein. In this thesis, new MS methods for glycoprotein oligosaccharides analysis have been developed, providing useful analytical tools for the analysis of glycopeptides, released glycans and monosaccharides.

Furthermore, new insights into the understanding of the glycosylation pattern of two HCMV antigens (gB and Pentamer) have been provided; indeed, site-specific glycan heterogeneity, glycosylation stoichiometry and protein epitope shielding are crucial information to rationally design and make vaccines with tailor-made glycosylations. It is worth noting that different glycoengineering methods (enzymatic inhibitors, gene knockout) have been applied to the mentioned antigens and that the diverse glycan patterns have been monitored by MS, showing how the used glycoengineering techniques radically change the glycoforms distribution on the sequons, shifting type of glycans from complex to oligomannose.

Site-specific glycan analysis of both gB and Pentamer has given important insights on how *N*-glycans maturation is affected by the processing enzymes accessibility on protein surface; as matter of the fact, the presence of glycan clusters (*e.g.* N240 and N342 on gB) and glycans wrapped by the protein backbone (*e.g.* N63 on Pentamer) correlates with the high abundance of oligomannose structures on the involved glycosylation sites. Moreover, site-specific analysis exploiting different fragmentation techniques (HCD, ETD, EThcD) also allowed *O*-glycosylation sites characterization which showed to bring different glycoforms (*e.g.* sialyl T antigen, T antigen and Tn antigen).

The development of MS methods for the quick analysis of released *N*-glycans (by LC-FLR-MS) and their monosaccharide composition (by GC-MS) has shown to be a

reliable analytical tool; in fact, both the identities and abundances of the *N*-glycans found by site specific analysis are in agreement with global glycans profile: the two methods, therefore, can be defined as orthogonal analytical methods. Monosaccharide composition analysis is very useful to assess the presence of non-self monosaccharides (such as Neu5Gc) and to identify the nature of all the monosaccharide building blocks in a univocal way; in fact, oligosaccharides with different monomers can give isobaric species which make difficult monosaccharide composition identification if MS-tandem experiments are not available. The analysis through this technique allowed to measure the monosaccharide absolute composition of the glycoprotein glycans, which can be a useful parameter to monitor during product manufacturing.

Finally, the ability to elicit the production of neutralizing antibodies in mice has been studied both for gB and Pentamer with different glycosylation patterns and the analysis outputs have been explained by building 3D models of the antigens and evaluating the shielding of the immunogenic epitopes, which was lower for the Pentamer bringing only M5 high-mannose glycans.

To conclude, in this project it has been showed how the combination of different MS methods can guide the development of glycoprotein antigens through the characterization of their glycan patterns and the monitoring of their variations after glycoengineering experiments, providing a powerful tool for the interpretation of antigens immunogenicity.

SECTION III:

EXPERIMENTAL SECTION

Chapter 9 – Experimental section

9.1 Materials and methods *N*-glycans site specific analysis

Protein denaturation and reduction was done by diluting the sample (100 μg of gB or Pentamer from CHO cells) with 82 μL 8M urea in 100 mM ammonium acetate (pH=6), adding 1.5 μL of 1 M DTT (with final concentration of 15 mM) and incubating for 30 minutes at 50°C; then the free cysteines were alkylated by adding 7.5 μL of iodoacetamide to obtain a 30 mM solution, the sample was incubated for 30 minutes at room temperature in absence of light. The remaining iodoacetamide was quenched by adding 3 μL of DTT 1M to obtain a 30 mM solution. The sample was buffer exchanged with 10 mM potassium phosphate buffer pH=7 (in the case of digestion with Glu-C) or 50 mM ammonium bicarbonate buffer pH=7.8.

The glycoprotein was digested overnight at 37°C with a proteolytic enzyme (trypsin or Glu-C) using an enzyme: substrate ratio of 1:20. To describe the very close *N*-glycosylation sequons N423 and N424 a combined digestion protocol has been used: after an overnight chymotrypsin digestion at 37°C a 4 hours Glu-C digestion at 37°C was performed (both in phosphate buffer at pH=7).

Glycopeptides were purified before LC-MS/MS analysis by SPE with OASIS HLB Extraction Cartridges 1cc. The elution volume was then split in two halves. A first half was dried, dissolved in 100 μL of 0.1% (*v/v*) HCOOH, centrifuged and injected (5 μL) at LC-MS/MS system form site-specific glycans microheterogeneity analysis.

The second half of the volume was dried and dissolved in 50 mM ammonium bicarbonate in $H_2{}^{18}O$, 2 μL of PNGase F (500 units/mL, in $H_2{}^{18}O$) were added and the sample was incubated at 37°C for 1 hour.

The sample then was acidified with 10 % (*v/v*) HCOOH and injected at the instrument for site-specific glycans occupancy analysis; each sample was prepared in triplicate and each injection was repeated three times.

The analyses were performed on an ACQUITY UPLC I-Class (Waters) to a Q-Exactive Plus Mass Spectrometer (Thermo Scientific). Peptides and glycopeptides

were separated using an Acquity UPLC Peptide CSH C18, (130 Å, 1.7 um, 1 x 150 mm, Waters) the mobile phases were 3 % (*v/v*) acetonitrile, 0.1 % (*v/v*) formic acid in water (phase A) and 0.1 % (*v/v*) formic acid in acetonitrile (phase B), the column temperature was kept at 60°C. The Q-Exactive Plus Mass Spectrometer was operated in Full MS/data dependent MS^2 mode (Full MS/dd-MS^2), 3×10^6 ions were accumulated in the Orbitrap (Automatic Gain Control, AGC), the mass range used was *m/z* 300 – *m/z* 3500; dynamic exclusion was enabled for 12 s, scans were acquired at 70,000 mass resolving power (full MS), normalized Collision Energy (NCE) of 26 was used.

The electrospray ion source temperature was set at 320°C. Xcalibur was used for chromatogram integration and quantitative analysis. PEAKS Studio 11 was used to perform the peptide mapping analysis and to attribute the glycoforms at each *N*-glycosylation site. A false discovery rate (FDR) of 1% was used to detect glycopeptides; the precursor ion mass error tolerance was of 15 ppm, the fragment mass error tolerance was of 0.02 Da. The glycoforms abundances were calculated by integration of the MS1 chromatogram, using the software Xcalibur.

9.2 Materials and methods N-glycans profiling

15 μg of gB (or Pentamer) from CHO cells and 15 μg of gB from kifunensine treated CHO cells were dispensed in two different 1.5 mL Eppendorf. Protein denaturation was done by adding 6 mL of RapiGest [5 % (*w/v*)] and 23 μL of water, then the sample it has been heat denatured for 5 min at 90°C. To remove *N*-glycans from protein, 1.2 μL of Rapid PNGase F have been added and the sample has been incubated at 50°C for 10 minutes. For *N*-glycans labelling, 12 μL of 7 % (*w/v*) RapiFluor solution (in dimethylformamide) have been added to the test tube and it has been incubated 5 minutes at room temperature. Protein was precipitated by adding 50 mL of a mix dimethylformamide/acetonitrile [32 % / 68 % (*v/v*)]. This procedure was repeated twice to obtain 2 technical replicates.

The analyses were performed on an ACQUITY UPLC I-Class (Waters) coupled to an ACQUITY Premier fluorescence (FLR) detector and to a Q-Exactive Plus Mass Spectrometer (Thermo Scientific). Labelled *N*-glycans were separated using an Acquity UPLC Glycan BEH Amide, (130 Å, 1.7 um, 2.1 x 150 mm, Waters) the mobile phases were 50 mM ammonium formate, pH 4.4 (phase A) and acetonitrile (phase B), the column temperature was kept at 60°C.

The FLR detector operated with excitation wavelength of 265 nm and emission wavelength of 425 nm, the FLR data rate was 2 points/sec. The Q-Exactive Plus Mass Spectrometer was operated in Full MS/data dependent MS^2 mode (Full MS/dd-MS^2), 3×10^6 ions were accumulated in the Orbitrap (Automatic Gain Control, AGC), the mass range used was *m/z* 300 – *m/z* 3500; dynamic exclusion was enabled for 12 s, scans were acquired at 70,000 mass resolving power (full MS), normalized Collision Energy (NCE) of 26 was used.

The electrospray ion source temperature was set at 320°C. Chromeleon was used for chromatogram integration and quantitative analysis.

9.3 Materials and methods MALDI-TOF MS *N*-glycans profiling

Fifty (50) µg of gB from CHO cells were dispensed in a 1.5 mL Eppendorf. Protein denaturation was done by heating the sample for 5 min at 90°C. To remove *N*-glycans from protein, 1.2 µL of Rapid PNGaseF have been added and the sample has been incubated at 50°C for 10 minutes. *N*-glycans permethylation was performed at room temperature by adding 50 µL of DMSO, 0.1 mg of NaOH and 50 µL of CH_3I, the sample was incubated for 1 hour. Liquid-liquid extraction with CH_3Cl/H_2O was performed three times and the organic phase was dried under air flow.

For MALDI-TOF MS analysis, the samples have been solubilized in H_2O/CH_3OH (50:50, *v/v*), then 4 µL of sample were mixed with 4 µL of 2,5-dihydroxybenzoic acid (DHB) solution (10 mg/mL DHB in H_2O/CH_3OH, 50:50 *v/v*) and 2 µL of mixture were spotted on MALDI plate.

MALDI-TOF MS and the MS/MS experiments were all performed in reflectron mode, positive polarity, on an ABSCIEX TOF/TOF 5800 Applied Biosystems (Foster City, CA, USA) mass spectrometer equipped with an Nd: YAG laser (λ=349 nm), with a 3 ns pulse width and a repetition rate of up to 1000 Hz, and also equipped with delayed extraction technology. For MS experiments each spectrum was derived by the accumulation of 2000 laser shots. All experiments were executed in technical triplicates.

9.4 Materials and methods monosaccharide analysis by LC-UV-MS/MS

Fucose (Fuc), mannose (Man), galactose (Gal), glucosamine (GlcN) stock solutions were prepared from dried standards. The same volumes of each monosaccharide standard solution were mixed and dried (this has been done for different calibration levels); then, 200 μL of labeling solution (35 mM PMP in 1.3 % (*v/v*) NH_3 MeOH solution) were added to each calibration level and they were incubated at 70°C for 90 minutes. Each calibration level was dried under air flow, 100 μL of 30 mM HCl were added and the mixture was extracted three times with chloroform to remove the excess of unreacted PMP. Standards were dried under air flow, solubilized in 50 mL of 50 % (*v/v*) H_2O/MeOH and injected at the LC-UV-MS/MS system.

For glycoprotein analysis 50 μg of gB from CHO cells), 3% (*w/v*) sucrose was buffer exchanged 4 times with Amicon Ultra-0.5 Centrifugal Filter 3 kDa to minimize the sucrose content, then the sample was solubilized in 30 μL of 4 M trifluoroacetic acid (TFA) and kept at 100°C for 4 h to release all the monosaccharides. After acid hydrolysis, TFA was removed by air flow, then, 200 mL of labeling solution (35 mM PMP in 1.3 % (*v/v*) NH_3 in H_2O/MeOH solution) was added to the sample and it was incubated at 70°C for 90 minutes. The sample was dried under air flow, 100 mL of 30 mM HCl were added and the mixture was extracted three times with chloroform to remove the excess of unreacted PMP. Sample was dried under air flow, solubilized in 50 mL of 50 % (*v/v*) H_2O/MeOH and injected at the LC-UV-MS/MS system.

The analyses were performed on an ACQUITY UPLC I-Class (Waters) coupled to an ACQUITY Premier photodiode array (PDA) detector and to a Q-Exactive Plus Mass Spectrometer (Thermo Scientific). Labelled monosaccharides were separated using an ACQUITY UPLC BEH C18, (130 Å, 1.7 um, 1 x 150 mm, Waters) the mobile phases were 1 % (*v/v*) HCOOH in water (phase A) and 1 % (*v/v*) HCOOH in acetonitrile (phase B), the column temperature was kept at 50°C.

The PDA detector was operated with excitation wavelength of 245 nm, the PDA data rate was 20 points/sec. The Q-Exactive Plus Mass Spectrometer was operated in Full MS/data dependent MS^2 mode (Full MS/dd-MS^2), 3×10^6 ions were accumulated in the Orbitrap (Automatic Gain Control, AGC), the mass range used was *m/z* 300 – *m/z* 3500; dynamic exclusion was enabled for 12 s, scans were acquired at 70,000 mass resolving power (full MS), normalized Collision Energy (NCE) of 26 was used.

The electrospray ion source temperature was set at 320°C. Chromeleon was used for chromatogram integration and quantitative analysis.

9.5 Materials and methods monosaccharide analysis by GC-MS

4.6 µg/µL Fucose (Fuc), 6.9 µg/µL mannose (Man), 4.8 µg/µL galactose (Gal), 4.2 µg/µL *N*-acetylglucosamine (GlcNAc), 4.3 µg/µL *N*-acetylneuraminic acid (Neu5Ac) and 5.4 µg/µL inositol (Inos) stock solutions were prepared from dried standards. The same volumes of each monosaccharide standard solution were mixed, added with a constant volume of inositol solution (as internal standard) and dried (this has been done for different calibration levels); then, to obtain the monosaccharides methyl glycosides, 1.5 mL of 1.25 M HCl-methanol were added to the tube and the mixture was incubated at 80°C for 1 h. After that, the standards mixture was dried and peracetylated by adding 200 µL of dry pyridine and 100 µL of dry acetic anhydride; the reaction tube was incubated at 80°C for 0.5 h. The sample was dried under a stream of air and dissolved in 500 µL of acetone, then it was transferred in a 500 µL vial and

injected at GC-MS. 5 different calibration levels were prepared using different volumes of stock solutions.

100 µL of gB 5.522 µg/µL (from CHO cells), 3% (*w/v*) sucrose were buffer exchanged 3 times with Amicon Ultra-0.5 Centrifugal Filter 3 kDa to minimize the sucrose content; then 20 µL of 5.4 µg/µL inositol were added to the sample and it was dried in a PYREX® Culture Tube by Reacti-Therm. To obtain the monosaccharides methyl glycosides from protein glycans, 1.5 µL of 1.25 M HCl-methanol were added to the tube and the sample was incubated at 80°C for 4 h. After that, the sample was dried and peracetylated by adding 200 µL of dry pyridine and 100 µL of dry acetic anhydride; the reaction tube was incubated at 80°C for 0.5 h. The sample was dried and it was dissolved in 500 µL of acetone, then it was transferred in a 500 µL vial and injected at GC-MS. This procedure was repeated 3 times, in order to obtain 3 technical replicates.

The analyses were performed on an Agilent 7820A gas chromatograph hyphenated with an Agilent 5977B MSD (single quadrupole mass spectrometer), using an automatic injector 7693A; monosaccharides acetylated methyl glycosides were separated using a 5% phenyl polysiloxane ZB-5 capillary column (30 m x 0.25 mm i.d. x 0.25 mm f.t., Phenomenex), the carrier gas was helium with a flow of 1.2 mL/min; the column oven temperature program was as follows: initial temperature 150°C; hold time 3 min; to 280 °C at 3°C/min; post run temperature 300°C; run time 51.3 min; post run time 5 min. The autosampler was set up with a 10 µL syringe (Agilent) to draw up 5.0 µL of sample. The injector was operated in the split mode at 280 °C with a split ratio of 50 :1.

The Agilent 5977B MSD was operated with EI at 70 eV, the acquisition mode was scan (frequency of 2.9 scans/s) with a solvent delay of 3 min, the mass range used was *m/z* 40 – *m/z* 550. MS source temperature 230°C; MS quadrupole temperature 150°C. Agilent MassHunter was used for chromatogram integration and quantitative analysis.

9.6 Materials and methods SDS-PAGE

2 µL of gB protein solution (5.5 µg/µL) were mixed with 5 µL of sample buffer (LDS NuPAGE), 1 µL of 50 mM DTT and 12 µL of water. The solution was incubated at 99°C for 10 minutes. Meanwhile, 11 µL of digest were mixed with 5 µL of sample buffer (LDS NuPAGE), 1 µL of 50 mM DTT and 3 µL water; the digest was incubated at 99°C for 10 minutes. For the enzyme control, 11 µL of enzyme solution were mixed with 5 µL of sample buffer, 1 µL of 50 mM DTT and 3 µL of water; the enzyme was incubated at 99°C for 10 minutes. 5 µL of Marker (SeeBlue NuPAGE) have been used. A 4-12% NuPAGE polyacrylamide gel was used for the run, with MES (1x) as running buffer and a voltage of 150 V. A 0.1% (*w/w*) Coomassie G-250 solution was used for bands detection. A 40% ethanol, 10% (*v/v*) acetic acid solution was used for gel destaining.

9.7 Materials and methods *O*-glycosylation sites mapping on Pentamer

Protein denaturation and reduction was done by diluting 28 µg of Pentamer from CHO cells with 82 µL 8M guanidinium chloride, adding 1.5 µL of 1 M DTT (with final concentration of 15 mM) and incubating for 30 minutes at 50°C; then the free cysteines were alkylated by adding 7.5 µL of iodoacetamide to obtain a 30 mM solution, the sample was incubated for 30 minutes at room temperature in absence of light. The remaining iodoacetamide was quenched by adding 3 µL of DTT 1M to obtain a 30 mM solution. The sample was buffer exchanged with 50 mM ammonium bicarbonate buffer pH=7.8.

The glycoprotein was digested overnight at 37°C with a proteolytic enzyme (trypsin or chymotrypsin) using an enzyme: substrate ratio of 1:20.

Glycopeptides were purified before nLC-MS/MS analysis by SPE with OASIS HLB Extraction Cartridges 1cc. Then the sample was dried, dissolved in 50 µL of 0,1% (*v/v*) HCOOH, centrifuged and injected (5 µL) at nLC-MS/MS system form site-specific *O*-glycans analysis.

The analyses were performed on an Vanquish Neo UHPLC (Thermo Scientific) coupled to a Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Scientific). Peptides and glycopeptides were separated using an µPAC Neo HPLC Column, (20 µm x 50 mm, Thermo Scientific) the mobile phases were 3 % (*v/v*) DMSO, 0.1 % (*v/v*) formic acid in water (phase A) and 3 % (*v/v*) DMSO, 0.1 % (*v/v*) formic acid, 80 % acetonitrile (phase B).

The Orbitrap Eclipse Tribrid Mass Spectrometer was operated in Full MS/data dependent MS^2 mode (Full MS/dd-MS^2), a 400000 AGC target has been used, the mass range used was *m/z* 350 – *m/z* 1200; dynamic exclusion was enabled for 30 s, scans were acquired at 120,000 mass resolving power (full MS), EThcD was the activation type used, a supplemental activation (25%) of the charge-reduced species was used in the ETD analysis to improve fragmentation.

The electrospray ion source temperature was set at 275°C and the capillary voltage was set to 1.9 kV. Xcalibur was used for chromatogram integration and quantitative analysis. PEAKS Studio 11 was used to perform the peptide mapping experiment and to attribute the glycoforms at each *N*-glycosylation site. A false discovery rate (FDR) of 1% was used to detect glycopeptides; the precursor ion mass error tolerance was of 15 ppm, the fragment mass error tolerance was of 0.05 Da. The glycoforms abundances were calculated by integration of the MS1 chromatogram, using the software PEAKS Studio 11.

9.8 Materials and methods immunological assays

Animal studies were ethically reviewed and carried out in compliance with EU DIR 63/2010.

For each antigen, immunization has been performed on ten female laboratory mice, strain C57BL/6, at 7–10 weeks of age; 0.10 µg of gB and 0.15 µg of Pentamer were administered by intramuscular vaccinations with 50 µL/animal; all 50 µL in 1 hind leg (quadriceps). The mice have been vaccinated every three weeks, for a total of three immunizations; blood samples were collected at day 21, 42 and 64 after first

immunization. Serum was collected after blood centrifugation (10,000×*g*) for 10 min at room temperature.

To measure the nAbs titer, virus solution was preincubated with serial dilutions (1:1000, 1:2000, 1:4000, 1:8000, 1:16000, 1:32000, 1:64000, 1:128000) of mouse serum in 37°C, 5% CO_2 incubator for 2 hours [DMEM + 2% (*w/v*) FBS+ 1% (*w/v*) PSG was used as dilution medium]; then, 50 µL of the virus/serum mixture were added to the wells of the assay plate containing a layer of ARPE-19 cells and they incubated overnight at 37 °C, 5% (*w/v*) CO_2 incubator. Therefore, the cells have been washed with PBS and fixed with 50 µL of 4% (*w/v*) paraformaldehyde at room temperature for 20 minutes. Cells have been washed once again with PBS and 50 µL of 0.1% (*w/v*) Triton X-100 have been added to the wells to permeabilize cells. After washing with PBS, 50 µL of primary antibody have been added to the plate (mouse anti-CMV IE1 monoclonal antibody); the system has been incubated in 37 ºC incubator for 1 hour. Then, the cells have been washed with PBS and 50µL of secondary antibody (AlexFluor mouse anti-IgG) and 4',6-diamidino-2-phenylindole (DAPI), a fluorescent substrate, have been added; the system has been incubated in 37 ºC incubator for 1 hour. After PBS washing with PBS, the plates have been read using the High Content Imaging-CX7.

9.9 Materials and methods CHO cells expression

Recombinant HCMV gB and Pentamer were expressed in CHO cells (CHO-K1 cell line), grown in CD-CHO medium (Invitrogen, Carlsbad, CA, USA), in the presence of 30% of O_2; pH was controlled at the set point of 7.00 with a dead band of 0.05 and controlled either via addition of 1.0 N NaOH or by sparging CO_2 as needed, with an agitation set at 100 rpm. The cells were fed with two different feed supplements over the course of the culture. Starting day 2, a defined amount of ADCF Antifoam, irradiated (3% Hyclone (Foamaway)) was regularly added. Temperature was set at 37.0°C during the 2 first days and then shifted to 33°C to initiate the production phase, which lasted 12 days.

After the production phase, the harvest was clarified using a series of depth filters (D0SP, X0SP, 0.5/0.2 μm filters), then the antigen was purified with a Nuvia C-prime resin (Bio-RAD) in bind/elute mode; adventitious virus inactivation was performed by adjusting the pool to a target concentration of 2.0% (*w/v*) PS80 and stored overnight at room temperature after a 0.22 μm filtration. The virus-inactivated pool is diluted and loaded onto an anion exchange chromatography column with HP-Q resin (Bio-RAD) in bind/elute mode. The HP-Q pool is adjusted in conductivity and loaded onto on a multimodal chromatography column with CHT type I resin (Bio-Rad) in flowthrough mode. At the end of the chromatographic purification the pool undergoes a series of filtrations (nanofiltration on a Planova 20N, ultrafiltration using Biomax PES 50 kDa membrane and final filtration on Sartopore 2 filter), which at the end generate the drug substance.

To express the gB bringing only high-mannose glycans, 1 μL of 0.5 mM kifunensine per 1 mL transfection culture have been added 3 hours post transfection.

Nuvia cPrime is mixed-mode chromatography (MMC) resin that allows both hydrophobic and electrostatic interactions to be utilized for the purification of a variety of biomolecules. The stationary phase consists of a macroporous highly crosslinked polymer, with particle size of 70 μm (Fig. 9.1).

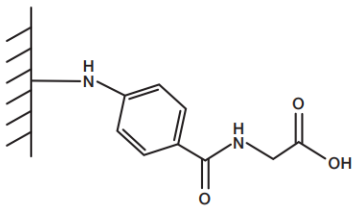


Figure 9.1. Ligand for Nuvia cPrime hydrophobic cation exchange resin.

In the MMC two interactions should both contribute to the retention of the solutes, by employing stationary phases with different functional groups (Fig. 9.2).

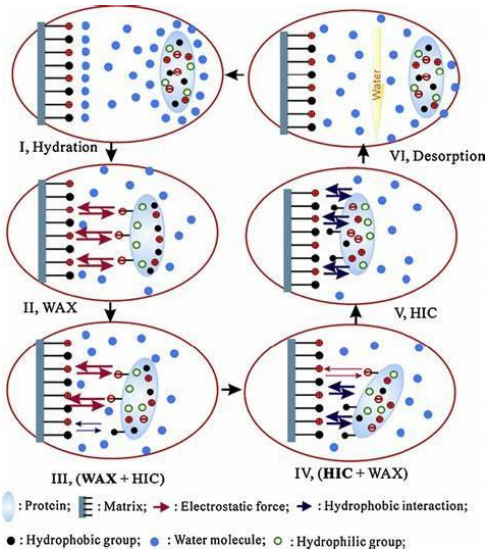


Figure 9.2. Interactions involved in MMC. The bold letter denotes the predominant mode by which protein retention occurs (depending on salts concentration).

Single column exhibited comparable or better resolution than commercial weak cation exchange (WCX) or hydrophobic interaction chromatography (HIC) column when used in their corresponding modes, respectively; indeed, in comparison with conventional single-mode chromatography, MMC offers several advantages which are summarized as follows: high selectivity, high loading capacity, reduced waste and/or overconsumption of materials (Yang et al., 2011).

Two more chromatographic purification steps are needed: first an anion exchange chromatography with HP-Q resin (Bio-Rad) in bind/elute mode and then a multimodal chromatography column with Ceramic Hydroxyapatite (CHT) type I resin (Bio-Rad). CHT $(Ca_5(PO_4)_3OH)_2$ support is a leading purification medium for demanding downstream process industry. This mixed-mode support offers unique selectivity and often separates biomolecules that appear homogeneous using other chromatographic methods; CHT ceramic hydroxyapatite is available in two distinct material types: Type I (lower pore size) and Type II (higher pore size). Hydroxyapatite contains two types of binding sites, positively charged calcium and negatively charged phosphate groups.

These sites are distributed regularly through the crystal structure of the matrix. Solute species dominantly interact through cation exchange *via* the phosphate groups and/or metal affinity via the calcium atoms (Fig. 9.3).

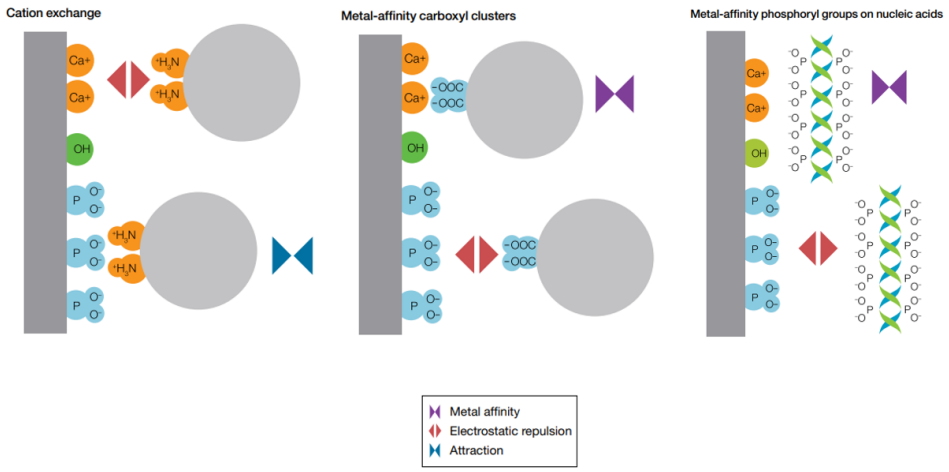


Figure 9.3. Schematic representation of CHT binding mechanisms.

After the production phase, the harvest is clarified using a series of depth filters (D0SP, X0SP, 0.5/0.2 µm filters), then the antigen is purified with a Nuvia C-prime resin (Bio-RADRad) in bind/elute mode; adventitious virus inactivation is then performed by adjusting the pool to a target concentration of 2.0% (*v/v*) PS80 and stored overnight at room temperature after a 0.22 µm filtration. The virus-inactivated pool is diluted and loaded onto an anion exchange chromatography column with HP-Q resin (Bio-RADRad) in bind/elute mode. The HP-Q pool is adjusted in conductivity and loaded onto on a multimodal chromatography column with CHT type I resin (Bio-Rad) in flowthrough mode. At the end of the chromatographic purification the pool undergoes a series of filtrations (nanofiltration on a Planova 20N, ultrafiltration using Biomax PES 50 kDa membrane and final filtration on Sartopore 2 filter), which at the end generate the purified bulk of the recombinant protein (drug substance.).

9.10 Material and methods Expi293^{GNTI-} cells expression.

HCMV Pentamer (Merlin strain), was stably expressed in human embryonic kidney (HEK) 293S cells bringing GnTI- mutation. Expression medium was loaded directly onto a StrepTrap HP column (GE Healthcare), and the proteins were eluted with 2.5 mM desthiobiotin (IBA Lifesciences). The Strep tag was cleaved off by incubating with TEV protease (AcTEV, Thermo Fisher Scientific) overnight at 4°C. The sample was diluted with 20 mM Hepes buffer (pH 7.0) to lower total salt concentration to 50 mM and loaded onto Mono S 10/30 column (GE Healthcare). Pentamer was eluted on a salt gradient from 50 mM to 1.0 M NaCl.

BIBLIOGRAPHY

A. Macagno, N. L Bernasconi, F. Vanzetta, E. Dander, A. Sarasini, M. G Revello, . . . Lanzavecchia, A. (2010). Isolation of human monoclonal antibodies that potently neutralize human cytomegalovirus infection by targeting different epitopes on the gH/gL/UL128-131A complex. *Journal of Virology,* 82(2).

Allen, J. D., Chawla, H., Samsudin, F., Zuzic, L., Shivgan, A. T., Watanabe, Y., . . . Crispin, M. (2021). Site-Specific Steric Control of SARS-CoV-2 Spike Glycosylation. *Biochemistry,* 60(27), 2153-2169.

Allen, J. G., Mujacic, M., Frohn, M. J., Pickrell, A. J., Kodama, P., Bagal, D., . . . McCarter, J. D. (2016). Facile Modulation of Antibody Fucosylation with Small Molecule Fucostatin Inhibitors and Cocrystal Structure with GDP-Mannose 4,6-Dehydratase. *ACS Chemical Biology,* 11(10), 2734-2743.

Angel, P. M., Lim, J. M., Wells, L., Bergmann, C., & Orlando, R. (2007). A potential pitfall in 18O-based N-linked glycosylation site mapping. *Rapid Commun Mass Spectrom,* 21(5), 674-682.

Bagdonaite, I., Malaker, S. A., Polasky, D. A., Riley, N. M., Schjoldager, K., Vakhrushev, S. Y., . . . Scott, N. E. (2022). Glycoproteomics. *Nature Reviews Methods Primers,* 2(1).

Bagdonaite, I., Norden, R., Joshi, H. J., Dabelsteen, S., Nystrom, K., Vakhrushev, S. Y., . . . Wandall, H. H. (2015). A strategy for O-glycoproteomics of enveloped viruses--the O-glycoproteome of herpes simplex virus type 1. *PLoS Pathogens,* 11(4), e1004784.

Bailly, C., Hecquet, P. E., Kouach, M., Thuru, X., & Goossens, J. F. (2020). Chemical reactivity and uses of 1-phenyl-3-methyl-5-pyrazolone (PMP), also known as edaravone. *Bioorganic & Medicinal Chemistry,* 28(10), 115463.

Basumallick, L., & Rohrer, J. (2016). Glycoprotein Monosaccharide Analysis Using HPAE-PAD with Eluent Generation. In T. F. Scientific (Ed.), (pp. 1-10).

Behrens, A. J., Vasiljevic, S., Pritchard, L. K., Harvey, D. J., Andev, R. S., Krumm, S. A., . . . Crispin, M. (2016). Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. *Cell Reports,* 14(11), 2695-2706.

Benoist, G., Leruez-Ville, M., Magny, J. F., Jacquemard, F., Salomon, L. J., & Ville, Y. (2013). Management of pregnancies with confirmed cytomegalovirus fetal infection. *Fetal Diagn Ther,* 33(4), 203-214.

Bianco, G., Buchicchio, A., & Cataldi, T. R. (2015). Structural characterization of major soyasaponins in traditional cultivars of Fagioli di Sarconi beans investigated by high-resolution tandem mass spectrometry. *Analytical and Bioanalytical Chemistry,* 407(21), 6381-6389.

Biemann, K. (1988). Contributions of mass spectrometry to peptide and protein structure. *BIOMEDICAL AND ENVIRONMENTAL MASS SPECTROMETRY,* 16(1-12), 99-111.

Bloch, J. S., John, A., Mao, R., Mukherjee, S., Boilevin, J., Irobalieva, R. N., . . . Locher, K. P. (2023). Structure, sequon recognition and mechanism of tryptophan C-mannosyltransferase. *Nature Chemical Biology*.

Bolmstedt, A., Hinkula, J., Rowcliffe, E., Biller, M., Wahren, B., & Olofsson, S. (2002). Enhanced immunogenicity ofa human immunodeficiency virus type 1 env DNA vaccine by manipulating N-glycosylation signals Effects of elimination of the V3 N306 glycan. *Vaccine,* 20, 397–405.

Buchacher, A., Predl, R., Strutzenberger, K., Steinfellner, W., Trkola, A., Purtscher, M., . . . Katinger, H. (1994). Generation of Human Monoclonal Antibodies against HIV-1 Proteins; Electrofusion and Epstein-Barr Virus Transformation for Peripheral Blood Lymphocyte Immortalization

AIDS RESEARCH AND HUMAN RETROVIRUSES, 10(4), 359-369.

Burton, D. R. (2023). Antiviral neutralizing antibodies: from in vitro to in vivo activity. *Nat Rev Immunol,* 23(11), 720-734.

Cao, L., Diedrich, J. K., Kulp, D. W., Pauthner, M., He, L., Park, S. R., . . . Paulson, J. C. (2017). Global site-specific N-glycosylation analysis of HIV envelope glycoprotein. *Nature Communications,* 8, 14954.

Catherman, A. D., Skinner, O. S., & Kelleher, N. L. (2014). Top Down proteomics: facts and perspectives. *Biochem Biophys Res Commun,* 445(4), 683-693.

Chandramouli, S., Ciferri, C., Nikitin, P. A., Calo, S., Gerrein, R., Balabanis, K., . . . Carfi, A. (2015). Structure of HCMV glycoprotein B in the postfusion

conformation bound to a neutralizing human antibody. *Nat Commun,* 6, 8176.
Chandramouli, S., Malito, E., Nguyen, T., Luisi, K., Donnarumma, D., Xing, Y., . . . Carfi, A. (2017). Structural basis for potent antibody-mediated neutralization of human cytomegalovirus. *SCIENCE IMMUNOLOGY* 2.
Chauhan, N. (2017). Laboratory Diagnosis of HbA1c: A Review. *Journal of Nanomedicine Research,* 5(4).
Chawla, H., Fadda, E., & Crispin, M. (2022). Principles of SARS-CoV-2 glycosylation. *Current Opinion in Structural Biology,* 75, 102402.
Chen, J. R., Liu, Y. M., Tseng, Y. C., & Ma, C. (2020). Better influenza vaccines: an industry perspective. *Journal of Biomedical Science,* 27(1), 33.
Choi, H. Y., Park, H., Hong, J. K., Kim, S. D., Kwon, J. Y., You, S., . . . Kim, D. I. (2018). N-glycan Remodeling Using Mannosidase Inhibitors to Increase High-mannose Glycans on Acid alpha-Glucosidase in Transgenic Rice Cell Cultures. *Scientific Reports,* 8(1), 16130.
Ciferri, C., Chandramouli, S., Donnarumma, D., Nikitin, P. A., Cianfrocco, M. A., Gerrein, R., . . . Carfi, A. (2015). Structural and biochemical studies of HCMV gH/gL/gO and Pentamer reveal mutually exclusive cell entry complexes. *Proceedings of the National Academy of Sciences of the United States of America,* 112(6), 1767-1772.
Colomb, F., Giron, L. B., Kuri-Cervantes, L., Adeniji, O. S., Ma, T., Dweep, H., . . . Abdel-Mohsen, M. (2020). Sialyl-Lewis(X) Glycoantigen Is Enriched on Cells with Persistent HIV Transcription during Therapy. *Cell Reports,* 32(5), 1-14.
Dalpathado, D. S., & Desaire, H. (2008). Glycopeptide analysis by mass spectrometry. *Analyst,* 133(6), 731-738.
Daniels, C. N., & Saunders, K. O. (2019). Antibody responses to the HIV-1 envelope high mannose patch. *Advances in Immunology,* 143, 11-73.
David R. Snydman, M. D., Barbara G. Werner, Ph.D., Beverly Heinze-Lacey, R.N., M.P.H., Victor P. Berardi, Nicholas L. Tilney, M.D., Robert L. Kirkman, M.D., Edgar L. Milford, M.D., Sang I. Cho, M.D., Harry L. Bush, Jr., M.D., Andrew S. Levey, M.D., Terry B. Strom, M.D., Charles B. Carpenter, M.D., Raphael H. Levey, M.D., William E. Harmon, M.D., Clarence E. Zimmerman, II, M.D., Michael E. Shapiro, M.D., Theodore Steinman, M.D., Frank LoGerfo, M.D., Beldon Idelson, M.D., Gerhard P.J. Schröter, M.D., Myron J. Levin, M.D., James McIver, Ph.D., Jeanne Leszczynski, DR.P.H., and George F. Grady, M.D. (1987). Use of Cytomegalovirus Immune Globulin to Prevent Cytomegalovirus Disease in Renal-Transplant Recipients. *The New England Journal of Medicine,* 317(17).

Dawood, A. A., & Altobje, M. A. (2020). Inhibition of N-linked Glycosylation by Tunicamycin May Contribute to The Treatment of SARS-CoV-2. *Microbial Pathogenesis,* 149, 104586.

De Castro, C., Parrilli, M., Holst, O., & Molinaro, A. (2010). Microbe-associated molecular patterns in innate immunity: Extraction and chemical analysis of gram-negative bacterial lipopolysaccharides. *Methods in Enzymology,* 480, 89-115.

De Castro, C., Speciale, I., Duncan, G., Dunigan, D. D., Agarkova, I., Lanzetta, R., . . . Van Etten, J. L. (2016). N-Linked Glycans of Chloroviruses Sharing a Core Architecture without Precedent. *Angewandte Chemie International Edition,* 55(2), 654-658.

Deris, H., Cindric, A., Lauber, M., Petrovic, T., Bielik, A., Taron, C. H., . . . Trbojevic-Akmacic, I. (2021). Robustness and repeatability of GlycoWorks RapiFluor-MS IgG N-glycan profiling in a long-term high-throughput glycomic study. *Glycobiology,* 31(9), 1062-1067.

Derking, R., Allen, J. D., Cottrell, C. A., Sliepen, K., Seabright, G. E., Lee, W.-H., . . . Sanders, R. W. (2020). Enhancing glycan occupancy of soluble HIV-1 envelope trimers to mimic the native viral spike. *Cell Reports,* 35(1).

Di Marco, F., Hipgrave Ederveen, A. L., van Schaick, G., Moran, A. B., Dominguez-Vega, E., Nicolardi, S., . . . Wuhrer, M. (2024). Comprehensive characterization of bacterial glycoconjugate vaccines by liquid chromatography - mass spectrometry. *Carbohydr Polym,* 341, 122327.

Dicker, M., & Strasser, R. (2015). Using glyco-engineering to produce therapeutic proteins. *Expert Opinion on Biological Therapy,* 15(10), 1-16.

Domon, B., & Costello, C. E. (1988). A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates. *Glycoconjugate,* 5, 397-409.

Donnelly, J. J., Ulmer, J. B., Shiver, J. W., & Liu, M. A. (1997). DNA VACCINES. *Annual Review of Immunology,* 15, 617–648.

Ebel, H., Benecke, T., & Vollmer, B. (2022). Stabilisation of Viral Membrane Fusion Proteins in Prefusion Conformation by Structure-Based Design for Structure Determination and Vaccine Development. *Viruses,* 14(8).

Elek, S. D., & Stern, H. (1974). DEVELOPMENT OF A VACCINE AGAINST MENTAL RETARDATION CAUSED BY CYTOMEGALOVIRUS INFECTION IN UTERO. *The Lancet* 303(7845).

Faleh, A., Warnke, S., Bansal, P., Pellegrinelli, R. P., Dyukova, I., & Rizzo, T. R. (2022). Identification of Mobility-Resolved N-Glycan Isomers. *Analytical Chemistry,* 94(28), 10101-10108.

Falkowska, E., Le, K. M., Ramos, A., Doores, K. J., Lee, J. H., Blattner, C., . . . Burton, D. R. (2014). Broadly neutralizing HIV antibodies define a glycan-dependent

epitope on the prefusion conformation of gp41 on cleaved envelope trimers. *Immunity,* 40(5), 657-668.
Falzarano, D., Krokhin, O., Van Domselaar, G., Wolf, K., Seebach, J., Schnittler, H. J., & Feldmann, H. (2007). Ebola sGP--the first viral glycoprotein shown to be C-mannosylated. *Virology,* 368(1), 83-90.
Fan, B., Li, T., Song, X., Wu, C., & Qian, C. (2019). A rapid, accurate and sensitive method for determination of monosaccharides in different varieties of Osmanthus fragrans Lour by pre-column derivatization with HPLC-MS/MS. *International Journal of Biological Macromolecules,* 125, 221-231.
Feng, T., Zhang, J., Chen, Z., Pan, W., Chen, Z., Yan, Y., & Dai, J. (2022). Glycosylation of viral proteins: Implication in virus-host interaction and virulence. *Virulence,* 13(1), 670-683.
Ferrara, C., Grau, S., Jager, C., Sondermann, P., Brunker, P., Waldhauer, I., . . . Benz, J. (2011). Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcgammaRIII and antibodies lacking core fucose. *Proc Natl Acad Sci U S A,* 108(31), 12669-12674.
Furuki, K., Toyo'oka, T., & Ban, K. (2017). Highly sensitive glycosylamine labelling of O-glycans using non-reductive beta-elimination. *Anal Bioanal Chem,* 409(9), 2269-2283.
Garces, F., Sok, D., Kong, L., McBride, R., Kim, H. J., Saye-Francisco, K. F., . . . Wilson, I. A. (2014). Structural evolution of glycan recognition by a family of potent HIV antibodies. *Cell,* 159(1), 69-79.
Gerwig, J. P. K. a. G. J. (2007). Strategies for the structural analysis of carbohydrates. 1-68.
Gillmann, K. M., Temme, J. S., Marglous, S., Brown, C. E., & Gildersleeve, J. C. (2023). Anti-glycan monoclonal antibodies: Basic research and clinical applications. *Current Opinion in Chemical Biology,* 74.
Going;, C., Huguet;, R., Ferrer;, D. L., Jenny Ho, V. Z., Huhmer;, A. F., & Pitteri, S. (2016). Characterization of Glycoproteins by Top Down UVPD Analysis. In T. F. Scientific (Ed.).
Gomes, A. C., Griffiths, P. D., & Reeves, M. B. (2019). The Humoral Immune Response Against the gB Vaccine: Lessons Learnt from Protection in Solid Organ Transplantation. *Vaccines (Basel),* 7(3).
Griffiths, J. (2008). A Brief History of Mass Spectrometry. *Analytical Chemistry,* 80(15), 5678–5683.
Griffiths, P., Baraniak, I., & Reeves, M. (2015). The pathogenesis of human cytomegalovirus. *J Pathol,* 235(2), 288-297.
Gupta, S. K., & Shukla, P. (2018). Glycosylation control technologies for recombinant therapeutic proteins. *Appl Microbiol Biotechnol,* 102(24), 10457-10468.

Gwon, Y. D., Zusinaite, E., Merits, A., Overby, A. K., & Evander, M. (2020). N-glycosylation in the Pre-Membrane Protein Is Essential for the Zika Virus Life Cycle. *Viruses,* 12(9), 1-18.

Hao, P., Ren, Y., Alpert, A. J., & Sze, S. K. (2011). Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Mol Cell Proteomics,* 10(10), O111 009381.

Helle, F., Vieyres, G., Elkrief, L., Popescu, C. I., Wychowski, C., Descamps, V., . . . Dubuisson, J. (2010). Role of N-linked glycans in the functions of hepatitis C virus envelope proteins incorporated into infectious virions. *Journal of Virology,* 84(22), 11905–11915.

Hodzic, A., Mateos-Hernandez, L., de la Fuente, J., & Cabezas-Cruz, A. (2020). alpha-Gal-Based Vaccines: Advances, Opportunities, and Perspectives. *Trends in Parasitology,* 36(12), 992-1001.

Illiano, A., Pinto, G., Melchiorre, C., Carpentieri, A., Faraco, V., & Amoresano, A. (2020). Protein Glycosylation Investigated by Mass Spectrometry: An Overview. *Cells,* 9(9), 1-23.

Itell, H. L., Nelson, C. S., Martinez, D. R., & Permar, S. R. (2017). Maternal immune correlates of protection against placental transmission of cytomegalovirus. *Placenta,* 60 Suppl 1(Suppl 1), S73-S79.

Ives, C. M. (2024). Restoring protein glycosylation with GlycoShape. *Nat Methods*.

Ives, C. M., Nguyen, L., Fogarty, C. A., Harbison, A. M., Durocher, Y., Klassen, J., & Fadda, E. (2024). Role of N343 glycosylation on the SARS-CoV-2 S RBD structure and co-receptor binding across variants of concern. *eLife,* 13.

Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology,* 23(1), 3-20.

Jayaprakash, N. G., & Surolia, A. (2017). Role of glycosylation in nucleating protein folding and stability. *Biochemical Journal* 474(14), 2333-2347.

Jean Beltran, P. M., & Cristea, I. M. (2014). The life cycle and pathogenesis of human cytomegalovirus infection: lessons from proteomics. *Expert Rev Proteomics,* 11(6), 697-711.

Kiessling, L. L., & Diehl, R. C. (2021). CH-pi Interactions in Glycan Recognition. *American Chemical Society Chemical Biology,* 16(10), 1884-1893.

Kolarich, D., Jensen, P. H., Altmann, F., & Packer, N. H. (2012). Determination of site-specific glycan heterogeneity on glycoproteins. *Nature Protocols,* 7(7), 1285-1298.

Kosma, P., & Muller-Loennies, S. (2012). *Anticarbohydrate Antibodies*.

Kschonsak, M., Johnson, M. C., Schelling, R., Green, E. M., Rougé, L., Ho, H., . . . Ciferri, C. (2022). Structural basis for HCMV Pentamer receptor recognition and antibody neutralization. *Science Advances*.

Kusumoto, M., Ueno-Noto, K., & Takano, K. (2020). Systematic Interaction Analysis of Anti-Human Immunodeficiency Virus Type-1 Neutralizing Antibodies with High Mannose Glycans Using Fragment Molecular Orbital and Molecular Dynamics Methods. *Journal of Computational Chemistry,* 41(1), 31-42.

Lembo, A., Molinaro, A., De Castro, C., Berti, F., & Biagini, M. (2024). Impact of glycosylation on viral vaccines. *Carbohydr Polym,* 342, 122402.

Liigand, P., Kaupmees, K., & Kruve, A. (2019). Influence of the amino acid composition on the ionization efficiencies of small peptides. *J Mass Spectrom,* 54(6), 481-487.

Lim, M. S., So, M. K., Lim, C. S., Song, D. H., Kim, J. W., Woo, J., & Ko, B. J. (2019). Validation of Rapi-Fluor method for glycan profiling and application to commercial antibody drugs. *Talanta,* 198, 105-110.

Liu, Heim, K. P., & Che, Y. (2021). Prefusion structure of human cytomegalovirus glycoprotein B and structural basis for membrane fusion. *Science Advances,* 7 1-9.

Lossl, P., Snijder, J., & Heck, A. J. (2014). Boundaries of mass resolution in native mass spectrometry. *J Am Soc Mass Spectrom,* 25(6), 906-917.

Lusvarghi, S., Stauft, C. B., Vassell, R., Williams, B., Baha, H., Wang, W., . . . Weiss, C. D. (2023). Effects of N-glycan modifications on spike expression, virus infectivity, and neutralization sensitivity in ancestral compared to Omicron SARS-CoV-2 variants. *PLoS Pathog,* 19(11), e1011788.

Ma, B., Guan, X., Li, Y., Shang, S., Li, J., & Tan, Z. (2020). Protein Glycoengineering: An Approach for Improving Protein Properties. *Frontiers in Chemistry* 8, 622.

Magalhaes, A., Duarte, H. O., & Reis, C. A. (2021). The role of O-glycosylation in human disease. *Molecular Aspects of Medicine,* 79, 100964.

Maity, S., & Acharya, A. (2023). Many Roles of Carbohydrates: A Computational Spotlight on the Coronavirus S Protein Binding. *ACS Applied Bio Materials*.

Marchetti, R., Forgione, R. E., Fabregat, F. N., Di Carluccio, C., Molinaro, A., & Silipo, A. (2021). Solving the structural puzzle of bacterial glycome. *Current Opinion in Structural Biology,* 68, 74-83.

Martin, M. S., Jacob-Dolan, J. W., Pham, V. T. T., Sjoblom, N. M., & Scheck, R. A. (2024). The chemical language of protein glycation. *Nat Chem Biol*.

Martina, C. E., Crowe, J. E. J., & Meiler, J. (2023). Glycan masking in vaccine design: Targets, immunogens and applications. *Frontiers in Immunology,* 14, 1126034.

McLuckey, S. A., & Stephenson, J. L. (1998). Ion/ion chemistry of high-mass multiply charged ions. *Mass Spectrometry Reviews,* 17(6), 369-407.

Miller, N. L., Subramanian, V., Clark, T., Raman, R., & Sasisekharan, R. (2022). Conserved topology of virus glycoepitopes presents novel targets for repurposing HIV antibody 2G12. *Scientific Reports*

12(1), 2594.

Niedermeier, D. g. P. a. W. (1978). Sensitive Gas Chromatographic determination of the glycoproteins *Journal of Chromatography,* 152, 487-494.

Nkolola, J. P., & Barouch, D. H. (2024). Prophylactic HIV-1 vaccine trials: past, present, and future. *Lancet HIV,* 11(2), e117-e124.

Notaro, A., Coute, Y., Belmudes, L., Laugieri, M. E., Salis, A., Damonte, G., . . . De Castro, C. (2021). Expanding the Occurrence of Polysaccharides to the Viral World: The Case of Mimivirus. *Angewandte Chemie International Edition,* 60(36), 19897-19904.

O'Neill, R. A. (1996). Enzymatic release of oligosaccharides from glycoproteins for chromatographic and electrophoretic analysis. *Journal of Chromatography A*(720), 201-215.

Olofsson, S., Bally, M., Trybala, E., & Bergstrom, T. (2023). Structure and Role of O-Linked Glycans in Viral Envelope Proteins. *Annu Rev Virol,* 10(1), 283-304.

Pace, A. L., Wong, R. L., Zhang, Y. T., Kao, Y. H., & Wang, Y. J. (2013). Asparagine deamidation dependence on buffer type, pH, and temperature. *J Pharm Sci,* 102(6), 1712-1723.

Pinto, D., Park, Y. J., Beltramello, M., Walls, A. C., Tortorici, M. A., Bianchi, S., . . . Corti, D. (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature,* 583(7815), 290-295.

Plotkin, S. A., Furukawa, T., Zygraich, N., & Huygelen, C. (1975). Candidate Cytomegalovirus Strain for Human Vaccination. *INFECTION AND IMMUNITY,,* 12(3).

Quintyn, R. S., Yan, J., & Wysocki, V. H. (2015). Surface-Induced Dissociation of Homotetramers with D2 Symmetry Yields their Assembly Pathways and Characterizes the Effect of Ligand Binding. *Chem Biol,* 22(5), 583-592.

Rebecchi, K. R., Wenke, J. L., Go, E. P., & Desaire, H. (2009). Label-free quantitation: a new glycoproteomics approach. *J Am Soc Mass Spectrom,* 20(6), 1048-1059.

Rocamora, F., Peralta, A. G., Shin, S., Sorrentino, J., Wu, M. Y. M., Toth, E. A., . . . Lewis, N. E. (2023). Glycosylation shapes the efficacy and safety of diverse protein, gene and cell therapies. *Biotechnol Adv,* 67, 108206.

Roepstorff, P., & Fohlman, J. (1984). Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomedical Mass Spectrometry* 11(11).

Rohrer, J. S., Basumallick, L., & Hurum, D. (2013). High-performance anion-exchange chromatography with pulsed amperometric detection for carbohydrate analysis of glycoproteins. *Biochemistry (Moscow),* 78(7), 697-709.

Romson, J., & Emmer, Å. (2021). Mass calibration options for accurate electrospray ionization mass spectrometry. *International Journal of Mass Spectrometry,* 467.

Rudd, P. M., & Dwek, R. A. (1997). Glycosylation: heterogeneity and the 3D structure of proteins. *Critical Reviews in Biochemistry and Molecular Biology,* 32(1), 1-100.

Ruiz-Matute, A. I., Hernandez-Hernandez, O., Rodriguez-Sanchez, S., Sanz, M. L., & Martinez-Castro, I. (2011). Derivatization of carbohydrates for GC and GC-MS analyses. *J Chromatogr B Analyt Technol Biomed Life Sci,* 879(17-18), 1226-1240.

Sandig, G., von Horsten, H. H., Radke, L., Blanchard, V., Frohme, M., Giese, C., . . . Hinderlich, S. (2017). Engineering of CHO Cells for the Production of Recombinant Glycoprotein Vaccines with Xylosylated N-glycans. *Bioengineering* 4(2), 1-12.

Scanlan, C. N., Pantophlet, R., Wormald, M. R., Ollmann Saphire, E., Stanfield, R., Wilson, I. A., . . . Burton, D. R. (2002). The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of alpha1-->2 mannose residues on the outer face of gp120. *Journal of Virology,* 76(14), 7306-7321.

Schijns, V., Majhen, D., van der Ley, P., Thakur, A., Summerfield, A., Berisio, R., . . . McClean, S. (2021). Rational Vaccine Design in Times of Emerging Diseases: The Critical Choices of Immunological Correlates of Protection, Vaccine Antigen and Immunomodulation. *Pharmaceutics,* 13(4), 1-29.

Schneider, M., Al-Shareffi, E., & Haltiwanger, R. S. (2017). Biological functions of fucose in mammals. *Glycobiology,* 27(7), 601-618.

Shental-Bechor, D., & Levy, Y. (2008). Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences of the United States of America,* 105(24), 8256–8261.

Sok, D., Laserson, U., Laserson, J., Liu, Y., Vigneault, F., Julien, J. P., . . . Poignard, P. (2013). The effects of somatic hypermutation on neutralization and binding in the PGT121 family of broadly neutralizing HIV antibodies. *PLoS Pathog,* 9(11), e1003754.

Song, X., Ju, H., Lasanajak, Y., Kudelka, M. R., Smith, D. F., & Cummings, R. D. (2016). Oxidative release of natural glycans for functional glycomics. *Nat Methods,* 13(6), 528-534.

Spanov, B., Govorukhina, N., Merbel, N. C. v. d., & Bischoff, R. (2022). Analytical tools for the characterization of deamidation in monoclonal antibodies. *Journal of Chromatography Open,* 2.

Speciale, I., Notaro, A., Abergel, C., Lanzetta, R., Lowary, T. L., Molinaro, A., . . . De Castro, C. (2022). The Astounding World of Glycans from Giant Viruses. *Chemical Reviews,* 122(20), 15717-15766.

Speciale, I., Notaro, A., Garcia-Vello, P., Di Lorenzo, F., Armiento, S., Molinaro, A., . . . De Castro, C. (2022). Liquid-state NMR spectroscopy for complex carbohydrate structural analysis: A hitchhiker's guide. *Carbohydrate Polymers,* 277, 118885.

Spindler, N., Diestel, U., Stump, J. D., Wiegers, A. K., Winkler, T. H., Sticht, H., . . . Muller, Y. A. (2014). Structural basis for the recognition of human cytomegalovirus glycoprotein B by a neutralizing human antibody. *PLoS Pathog,* 10(10), e1004377.

Spiwok, V. (2017). CH/pi Interactions in Carbohydrate Recognition. *Molecules,* 22(7).

Stavenhagen, K., Hinneburg, H., Thaysen-Andersen, M., Hartmann, L., Varon Silva, D., Fuchser, J., . . . Kolarich, D. (2013). Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *Journal of Mass Spectrometry,* 48(6), 627-639.

Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., . . . Clausen, H. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO Journal,* 32(10), 1478-1488.

Stepper, J., Shastri, S., Loo, T. S., Preston, J. C., Novak, P., Man, P., . . . Norris, G. E. (2011). Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Letters,* 585(4), 645-650.

Stiving, A. Q., VanAernum, Z. L., Busch, F., Harvey, S. R., Sarni, S. H., & Wysocki, V. H. (2019). Surface-Induced Dissociation: An Effective Method for Characterization of Protein Quaternary Structure. *Anal Chem,* 91(1), 190-209.

Sun, W., Zhang, Q., Zhang, X., Tran, N. H., Ziaur Rahman, M., Chen, Z., . . . Shan, B. (2023). Glycopeptide database search and de novo sequencing with PEAKS GlycanFinder enable highly sensitive glycoproteomics. *Nature Communications*

14(1), 4046.

Syka, J. E. P., Coon, J. J., & Schroeder, M. J. S., Jeffrey Hunt, Donald F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *PNAS,* 101(26), 9528–9533.

Tajiri, M., Yoshida, S., & Wada, Y. (2005). Differential analysis of site-specific glycans on plasma and cellular fibronectins: application of a hydrophilic affinity method for glycopeptide enrichment. *Glycobiology,* 15(12), 1332-1340.

Tejwani, V., Andersen, M. R., Nam, J. H., & Sharfstein, S. T. (2018). Glycoengineering in CHO Cells: Advances in Systems Biology. *Biotechnol Journal,* 13(3), e1700234.

Tian, W., Li, D., Zhang, N., Bai, G., Yuan, K., Xiao, H., . . . Gao, G. F. (2021). O-glycosylation pattern of the SARS-CoV-2 spike protein reveals an "O-Follow-N" rule. *Cell Research,* 31(10), 1123-1125.

Trkola, Martin Purtscher, Thomas Muster, Claudia Ballaun, Andrea Buchacher, Nancy Sullivan, . . . Katinger, a. H. (1996). Human Monoclonal Antibody 2G12 Defines a Distinctive Neutralization Epitope on the gp120 Glycoprotein of Human Immunodeficiency Virus Type 1. *Journal of Virology,* 70(2), 1100-1108.

Uhliarikova, I., Matulova, M., & Capek, P. (2021). Optimizing acid hydrolysis for monosaccharide compositional analysis of Nostoc cf. linckia acidic exopolysaccharide. *Carbohydr Res,* 508, 108400.

Valencia, S. M., Rochat, E., Harnois, M. J., Dennis, M., Webster, H. S., Hora, B., . . . Permar, S. R. (2023). Vaccination with a replication-defective cytomegalovirus vaccine elicits a glycoprotein B-specific monoclonal antibody repertoire distinct from natural infection. *NPJ Vaccines,* 8(1), 154.

Varnum, S. M., Streblow, D. N., Monroe, M. E., Smith, P., Auberry, K. J., Pasa-Tolic, L., . . . Nelson, J. A. (2004). Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. *J Virol,* 78(20), 10960-10966.

Veluraja, T. H. T. C. a. K. (2001). Database Analysis of O-Glycosylation Sites in Proteins. *Biophysical Journal,* 80 952–960.

Wada, Y., Azadi, P., Costello, C. E., Dell, A., Dwek, R. A., Geyer, H., . . . Taniguchi, N. (2007). Comparison of the methods for profiling glycoprotein glycans--HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology,* 17(4), 411-422.

Wang, L. X., Tong, X., Li, C., Giddens, J. P., & Li, T. (2019). Glycoengineering of Antibodies for Modulating Functions. *Annual Review of Biochemistry,* 88, 433-459.

Wang, Q. C., Zhao, X., Pu, J. H., & Luan, X. H. (2016). Influences of acidic reaction and hydrolytic conditions on monosaccharide composition analysis of acidic, neutral and basic polysaccharides. *Carbohydr Polym,* 143, 296-300.

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., & Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science,* 369, 330–333.

Wilkinson, H., & Saldova, R. (2020). Current Methods for the Characterization of O-Glycans. *J Proteome Res,* 19(10), 3890-3905.

Wisniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat Methods,* 6(5), 359-362.

Wysocki, V. H., Resing, K. A., Zhang, Q., & Cheng, G. (2005). Mass spectrometry of peptides and proteins. *Methods,* 35(3), 211-222.

Xie, Y., Mota, L. M., Bergin, A., O'Flaherty, R., Jones, A., Morgan, B., & Butler, M. (2021). High-throughput and high-sensitivity N-Glycan profiling: A platform for biopharmaceutical development and disease biomarker discovery. *Analytical Biochemistry,* 623, 114205.

Yang, Y., & Geng, X. (2011). Mixed-mode chromatography and its applications to biopolymers. *J Chromatogr A,* 1218(49), 8813-8825.

Yates, J. R. A century of mass spectrometry: from atoms to proteomes. *nature methods,* 8(8).

Yates, J. R. (2011). A century of mass spectrometry: from atoms to proteomes. *nature methods,* 8(8), 633-637.

Yu, Q., Wang, B., Chen, Z., Urabe, G., Glover, M. S., Shi, X., . . . Li, L. (2017). Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. *J Am Soc Mass Spectrom,* 28(9), 1751-1764.

Zhang, Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., . . . Ma, B. (2012). PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics,* 11(4), M111 010587.

Zhang, L., Xu, J., Zhang, L., Zhang, W., & Zhang, Y. (2003). Determination of 1-phenyl-3-methyl-5-pyrazolone-labeled carbohydrates by liquid chromatography and micellar electrokinetic chromatography. *Journal of Chromatography B* 793(1), 159-165.

Zubarev, R. A. (2004). Electron-capture dissociation tandem mass spectrometry. *Current Opinion in Biotechnology,* 15(1), 12-16.

This work was sponsored and funded by GlaxoSmithKline Biologicals SA