



Universiteit
Leiden
The Netherlands

Clinical glycomics of antibodies by mass spectrometry

Haan, N. de

Citation

Haan, N. de. (2019, February 14). *Clinical glycomics of antibodies by mass spectrometry*. Retrieved from <https://hdl.handle.net/1887/68328>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/68328>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The following handle holds various files of this Leiden University dissertation:

<http://hdl.handle.net/1887/68328>

Author: Haan, N. de

Title: Clinical glycomics of antibodies by mass spectrometry

Issue Date: 2019-02-14

A fascinating number of different components are continuously circulating in the human blood. While these components all have their own function, they largely play together in a complex arrangement to assure a - most of the time - healthy day-to-day life. The human circulation provides the infrastructure of the body, with the large arteries and veins serving as the highways, and the smaller, organ- and tissue-specific vessels facilitating the local transportation. It supplies the whole body with all necessities in the form of red blood cells, white blood cells, platelets, proteins, sugars, lipids, hormones, salts, vitamins and many other cells and substances. In addition to the nutrient-, defense-, and messenger-components that are constantly being distributed, the blood circulation ultimately picks up the waste created in the distant areas of the body, in order to remove it. Altogether, the specific composition of the human blood provides an extensive overview of a large part of the intriguing processes going on in the body. This can be sampled and studied with current medical and chemical technologies. In normal, healthy conditions, a blood sample represents the balanced processes of the body, homeostasis. Disruption of the homeostasis will result in the up- or down- regulation of one or more components. A minor, temporary imbalance might reflect the body restoring its own equilibrium. However, fierce or persisting disruptions of the normal situation might indicate the presence or emergence of a serious underlying pathology. Medical tests studying these processes are, for example, used for the diagnosis, prognosis or treatment monitoring of various diseases. Expanding and improving these tests will be beneficial for the early diagnosis of pathologies and personalized treatment of patients.

One of the most diverse groups of molecular species circulating in the blood is the group of the proteins, performing functions ranging from protection of the body and the coordination of biological processes, to aiding the synthesis, modification or transportation of other molecules. Fine-tuning the properties of the vast majority of these blood proteins happens by, so called, post-translational modifications of which phosphorylation, acetylation and *N*-linked glycosylation form the top three of most occurring modifications [1]. Of these three, *N*-linked glycosylation is by far the most diverse. This makes it highly challenging, and at the same time highly interesting, to investigate this protein modification for the identification of potential diagnostic and prognostic markers.

Antibodies are immune proteins that play an important role in the defense of the body against pathogenic substances, like viruses and bacteria. The working mechanism of antibodies is often highly dependent on their glycosylation, and the development of suitable analytical tools is crucial to map the glycans attached to these glycoproteins. The possibility to discriminate between differentially glycosylated antibodies has high potential in precision medicine by contributing to the sensitivity and specificity of medical

tests. Immunoglobulin G (IgG) is the major antibody in human plasma, plays an essential role in the humoral immunity and will be one of the key players in this thesis.

1.1 Protein glycosylation

Protein glycosylation describes the process of the co- or post-translational attachment of a combination of monosaccharides (also known as sugars or carbohydrates) to a protein [2]. *N*-glycosylation, comprising the addition of a glycan to the nitrogen atom of an asparagine residue, is one of the most occurring PTMs as well as the most occurring form of protein glycosylation. Another often occurring form of protein glycosylation involves the attachment of a glycan to the oxygen atom of serine or threonine residue (*O*-glycosylation) [1, 3]. As glycans often contribute to a large fraction of the total glycoprotein size and have distinct physicochemical properties, including a high hydrophilicity, one can imagine that the influence of a glycan on the structure and function of a protein is substantial [4]. The presence of a large hydrophilic structure indirectly alters the interaction possibilities of the glycoconjugate with other proteins by influencing its folding and/or stability [4]. In addition, specific monosaccharide residues can have a direct interaction with glycan recognizing proteins, like lectins [5-7].

1.1.1 *N*-glycan structure and biosynthesis

The biosynthetic coupling of an *N*-glycan to an asparagine almost exclusively occurs when the asparagine is present in the *N*-glycosylation consensus sequence, asparagine-x-serine/threonine, where “x” can be any amino acid except for proline [8]. Additionally, *N*-glycosylation is reported to take place, to a lower extent, in the asparagine-x-cysteine motive [9, 10]. During, or right after, the synthesis of the protein in the endoplasmic reticulum (ER), a precursor *N*-glycan structure, consisting of two *N*-acetylglucosamines (GlcNAc), nine mannoses (Man) and three glucoses (Glc; **Figure 1.1A**), is transferred in its entirety from a dolichol anchor to the asparagine in the consensus sequence [11]. During the folding of the protein, still in the ER, the terminal Glc residues are enzymatically removed, one-by-one, serving as a protein folding-quality control. Only after proper folding, the protein enters the Golgi apparatus, where the Man residues may be trimmed further, reaching a structure with two GlcNAcs and five Mans (Man₅, a high-mannose glycan; **Figure 1.1B**). Later in the Golgi, the glycan is extended with a GlcNAc, forming a hybrid-type glycan (**Figure 1.1B**). The ongoing interplay between glycosidases and glycosyltransferases results in the formation of complex-type *N*-glycans, which can be extended by different GlcNAc branches (antennae), a bisecting GlcNAc, core and/or antennary fucoses (Fuc) and antennary galactoses (Gal), *N*-acetylgalactosamines (GalNAc), Gal-GlcNAc (LacNAc) repeats, GalNAc-GlcNAc (LacdiNAc) repeats and *N*-acetylneuraminic acids (Neu5Ac; in humans) or *N*-glycolylneuraminic acids (Neu5Gc; in

other mammals; **Figure 1.1**) [11]. The latter two monosaccharides belong to the group of sialic acids (Sia), a group of large monosaccharides containing a carboxylic acid functional group [12].

The aforementioned monosaccharides are the most common *N*-glycan building blocks in humans and other mammals. They usually form six-membered cyclic structures consisting of five carbons and one oxygen in the ring (**Figure 1.1A**) [13]. The five carbon atoms of the ring are all chiral (except from C-3 in sialic acids), resulting in the existence of multiple different monosaccharide stereoisomers. Monosaccharides that only differ from each other by the configuration of one of their chiral carbon atoms are known as epimers [14]. For example, Gal and Glc are epimers due to their different configurations at C-4. Additionally, exact mirror images, enantiomers, are denoted by the prefix L- or D-, based on the configuration at C-5. Most monosaccharides in humans are D-enantiomers, with the exception of L-fucose (**Figure 1.1A**) [14]. Finally, while the configurations at C-2 to C-5 are stable within a structure, the configuration of C-1 (or C-2 in sialic acids; the anomeric carbon) can be altered in solution, resulting in a mixture of α -anomers (when the configuration at C-1 and C-5 is the same) and β -anomers (when the configuration at C-1 and C-5 is different; **Figure 1.2A**) [14].

Two monosaccharides can be enzymatically linked by condensation between the hemiacetal group of one monosaccharide and one of the hydroxyl groups of the other monosaccharide [13]. Branched structures can be formed as multiple hydroxyl groups of one monosaccharide can form simultaneous glycosidic bonds [11]. This in contrast to, for example, the almost exclusively linear connection between amino acids to form a protein. The linkage between two monosaccharides can vary on the basis of the hydroxyl groups that are involved, resulting in different regioisomers. Additionally, the configuration around the involved anomeric center determines whether an α - or β -stereoisomer is formed (**Figure 1.2B**) [15]. The anomeric carbon not involved in a glycosidic bond is referred to as the reducing end of a glycan [13]. The formation of both regio- and stereoisomers is under specific enzymatic control and has a high influence on the 3D-structure of the resulting glycan, eventually influencing its function and activity [11].

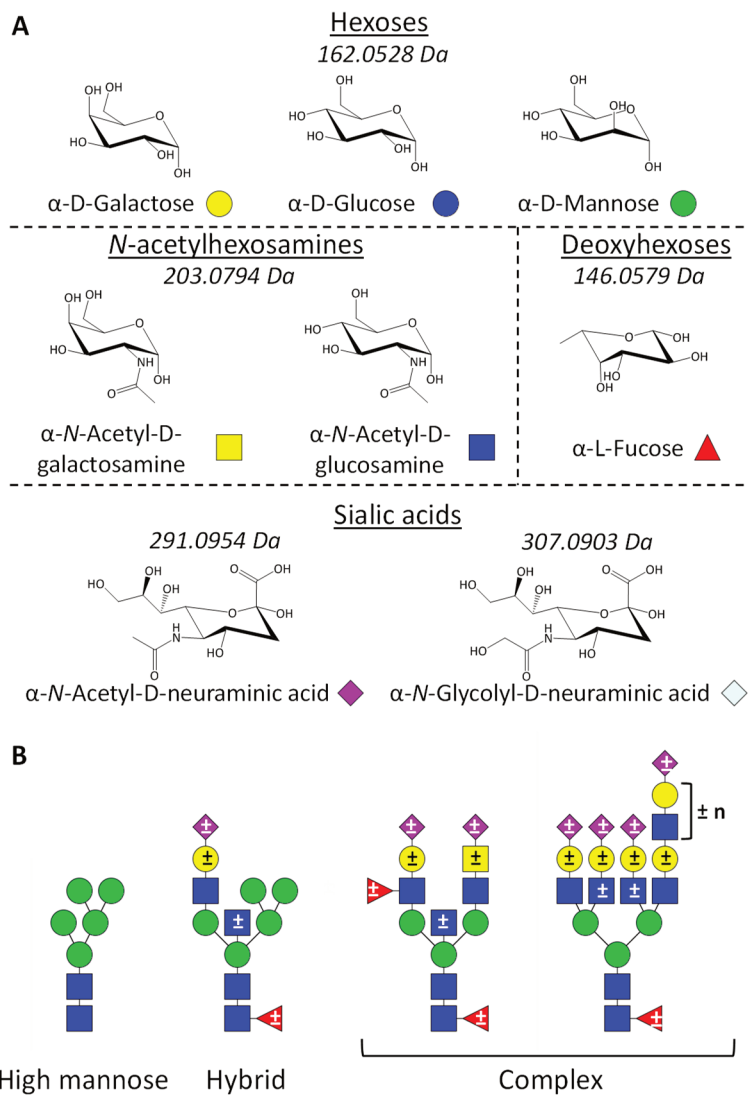


Figure 1.1 N- and O-glycan monosaccharides and the three common types of N-glycan structures. (A) The monosaccharides commonly found in mammalian N- and O-glycans and their representative symbols. While humans exclusively incorporate the sialic acid variant N-acetylneuraminic acid in their N- and O-glycans, other mammals additionally, or exclusively, feature N-glycolylneuraminic acids. The monosaccharide masses shown are the residue masses of the sugar units upon condensation with a glycan. (B) Symbolic representations of the three N-glycan types found on human proteins, including high mannose-, hybrid- and complex-type N-glycans. High mannose-type glycans carry five (as shown) to nine mannoses. Hybrid- and complex-type glycans carry optional core or antennary monosaccharide residues, indicated with a plus/minus sign (\pm).

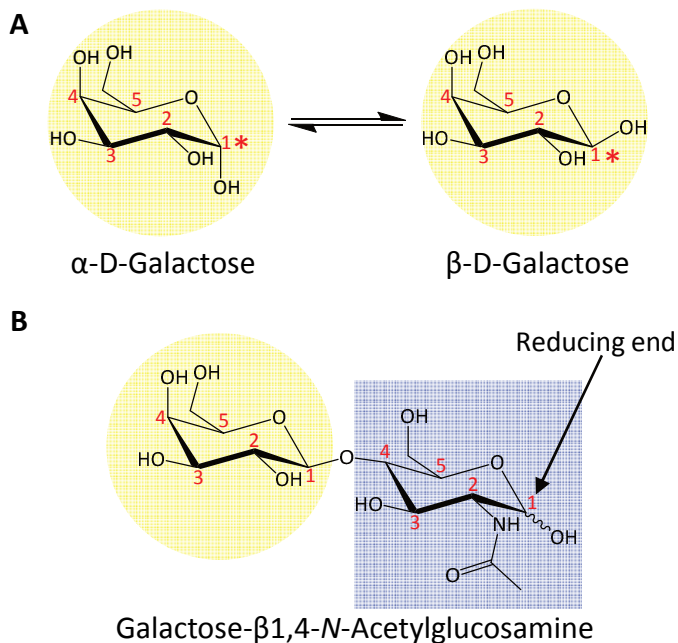


Figure 1.2 Monosaccharide anomers and linkages. (A) The α - and β -anomer of D-galactose. The anomeric center at C-1 is indicated by an asterisk (*). (B) A galactose linked to an *N*-acetylglucosamine via a β 1,4-linkage, the hydroxyl group at the anomeric center (reducing end) of the *N*-acetylglucosamine is connected with a wavy line, indicating the configuration is either α or β , alternately present in solution via an open ring intermediate.

1.1.2 *O*-glycan structure and biosynthesis

Different types of *O*-glycosylation exist, of which the *O*-GalNAc-type of glycosylation is most common (also known as mucin-type *O*-glycosylation). While the building blocks for the construction of a mucin-type *O*-glycan (further referred to as *O*-glycan) are matching the building blocks used for *N*-glycan synthesis (GalNAc, GlcNAc, Gal, Fuc and Sia; **Figure 1.1A**) and the glycosidic linkages formed are the same, the biosynthesis pathway is rather different between the two [16]. While *N*-glycosylation happens both co- and post-translationally and is initiated in the ER, *O*-glycosylation exclusively occurs post-translationally in the Golgi apparatus, starting with the enzymatic transfer of one GalNAc to a serine or threonine residue [16, 17]. Additionally, there is no known amino acid consensus sequence for the *O*-glycosylation to take place. A broad repertoire of enzymes can initiate the initial GalNAc transfer, which all have their own substrate specificity. Some of these enzymes prefer a serine as substrate and others a threonine. Additionally, it is known that a proline or another *O*-glycan in close proximity of the serine or threonine

facilitates *O*-glycosylation [17, 18]. After the attachment of the first GalNAc, the *O*-glycan can be extended by the interplay of a wide variety of glycosyltransferases, resulting in branched structures with one of the four common cores 1 to 4, consisting of GalGalNAc, GlcNAcGalGalNAc, GlcNAcGalNAc or GlcNAc₂GalNAc, respectively (**Figure 1.3**) [19].



Figure 1.3 Symbolic representations of the four common cores of mucin-type *O*-glycan structures. Yellow square: *N*-acetylgalactosamine (GalNAc); Yellow circle: galactose (Gal); Blue square: *N*-acetylglucosamine (GlcNAc). The linkages between monosaccharides are indicated at the connective lines.

1.2 IgG glycosylation

There are four human IgG subclasses, IgG1, 2, 3 and 4, all 150 kDa glycoproteins consisting of two heavy (H) and two light (L) chains, connected via disulfide bonds. The molecules consist of constant (C) domains (C_{H1} to C_{H3}, and C_L) and variable (V), antigen binding domains (V_H and V_L; **Figure 1.4**), which together form the Fab (C_L, C_{H1}, V_L, V_H), and Fc (C_{H2} and C_{H3}) region of the antibody. Fab and Fc are connected by the hinge region, which has a relatively flexible structure (**Figure 1.4**) [20]. All IgG subclasses can be subdivided into allotypes, based on the genetic variation between individuals in the heavy chain sequence of the antibody. With 19 different known allotypes, IgG3 is the subclass with the widest variety of constant domain sequence variants [21, 22].

Within the C_{H2} domains of the IgG Fc regions an *N*-glycosylation consensus sequence can be found, carrying an *N*-glycan in over 99% of the IgG in the human circulation [23, 24]. IgG Fc *N*-glycosylation is rather special in its composition as compared to the glycosylation of other glycoproteins, as it shows predominantly diantennary complex-type glycans, and a minor fraction of hybrid-type glycans [25-27]. The complex-type IgG Fc glycans are mainly fucosylated (~93% in healthy adults) and may carry a bisecting GlcNAc (~12%). Furthermore, their composition varies from having zero to two antennary galactoses (~40% per antenna) and zero to two terminating *N*-acetylneuraminic acids (~4% per antenna; further referred to as sialic acids when human glycosylation is concerned; **Figure 1.4**) [28, 29]. In addition to the aforementioned C_{H2} *N*-glycosylation, that is conserved for

all IgG subclasses, IgG3 carries an *N*-glycosylation site in the C_H3 domain of most allotypes [23, 30] and three *O*-glycosylation sites in the hinge [31]. The occupancy of these additional sites was determined to be about 12% for the C_H3 *N*-glycosylation and about 10% for the *O*-glycosylation sites [23, 31]. Furthermore, 15 to 25% of the IgG Fab regions carry one or more occupied *N*-glycosylation consensus sequences. When present in the antigen-binding domain of the antibody, the glycans enhance the variability of these regions [32, 33]. Compared to Fc glycans, Fab glycans show higher levels of galactosylation, sialylation and bisecting GlcNAcs, while their fucosylation is lower [29].

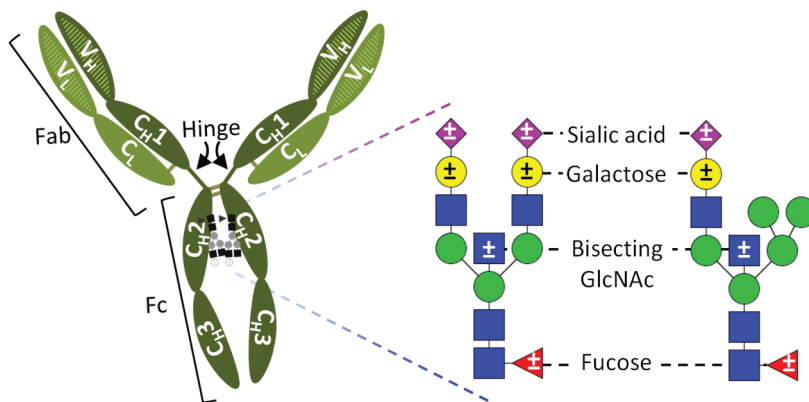


Figure 1.4 Schematic representation of IgG1 and its commonly found C_H2 glycan structures. IgG consists of two heavy chains (H; dark green) and two light chains (L; light green), connected via disulfide bridges (grey lines). Together, the heavy and light chain form the Fab and Fc portion, connected via the hinge region. The variable parts of the heavy and light chain (V) are both involved in the antigen binding sites (striped). On one of the constant domains (C), C_H2, a conserved *N*-glycosylation site is present. This site carries hybrid- and complex-type glycans with optional extra monosaccharide residues, like a core fucose, a bisecting *N*-acetylglucosamine (GlcNAc), galactoses and sialic acids, indicated with a plus/minus sign (±).

1.2.1 IgG Fc glycosylation effector mechanisms

IgG plays an important role in humoral immune responses, e.g. the immune events taking place in the extracellular space, by destroying microorganisms and foreign substances. IgG molecules bind antigens with their Fab regions and clear them via neutralization, opsonization by Fc receptor binding or activation of the complement system [34]. The latter two events involve the Fc region of the antibody and are highly influenced by both the presence [35, 36] and the structure [37] of the C_H2 *N*-glycan. Total absence of the Fc glycan minimizes the affinity between the IgG-Fc and most of its interaction partners by significantly changing the 3D-conformation of the Fc [38-40]. The presence or absence of

individual monosaccharides, on the other hand, modulates the binding affinity of the Fc for its interaction partners in a more subtle and specific way [41]. A large part of the studies into Fc glycan effector functions were performed in murine models, however murine glycosylation comprises some distinct glycosylation features as compared to human glycosylation. Most noticeable are the presence of different sialic acids, *N*-glycolylneuraminic acids, and the possibility to carry terminal α 1,3-linked galactoses attached to the β 1,4-linked antennary galactoses [42-44]. As it has been shown that functional results obtained from mouse experiments are not always straightforwardly translatable to the human situation [45, 46], in the following, only findings are presented obtained from human based *in vitro* studies, unless reported otherwise.

Fucosylation

The absence of a core fucose on complex-type IgG1 Fc glycans is known to enhance Fc γ receptor (Fc γ R)IIIa and b binding by 20- to 100-fold [37, 47], thereby strongly increasing the subsequent antibody-dependent cellular cytotoxicity (ADCC) [37, 48]. Both protein-glycan and glycan-glycan interactions between the uniquely glycosylated Fc γ RIIIa/b and the non-fucosylated Fc showed to play a key role in the increase in binding affinity [49].

Galactosylation

In contrast to the aforementioned core fucoses, terminal galactose residues are unlikely to be in close enough proximity to the Fc binding site to provide a direct interaction between the glycan and a receptor(-glycan) [50]. Still it has been shown that galactosylation slightly improves the binding of IgG1 to Fc γ RI, Fc γ RIIa/b/c and Fc γ RIIIa/b [51, 52], which might be due to an increase in Fc stability caused by glycan-protein and glycan-glycan interactions within the antibody [40, 50]. Although previous research indicated this galactosylation effect to be independent from the fucosylation effect [51], a more recent study showed the increased IgG1 affinity for Fc γ RIIIa/b upon galactosylation only to be present for afucosylated glycoforms [37]. In addition to the effect on Fc γ R binding, an increase in Fc galactosylation has been shown to improve complement activation via C1q binding and in consequence enhance complement-dependent cytotoxicity (CDC) [37, 53].

Sialylation

A lot of debate is going on about the effect of IgG Fc sialylation. Changes in sialylation often follow the effect of the galactosylation of the glycan, as the terminal galactose is a substrate for sialyltransferases [54, 55]. Therefore, the effects of sialylation and galactosylation in a biological system are often hard to distinguish. Whereas some report no effect of IgG1 sialylation on Fc γ R binding [52], others report a slight decrease in Fc γ RIIIa/b affinity for sialylated, afucosylated, bisected IgG1 as compared to its non-

sialylated, afucosylated, bisected form [37]. Furthermore, sialylation was both reported to increase [37] and to decrease [56] C1q binding. The latter discrepancy was suggested to be caused by the difference in monoclonal antibody studied, as the spatial distribution of the monoclonal antibodies on the cell surface might have a large influence on the activation of the complement [37]. In addition to FcγR and C1q binding, sialylation was reported to exert its effect via binding to the cell surface receptor DC-SIGN and its murine orthologue SIGN-R1, which results in the upregulation of anti-inflammatory components [57, 58]. However, also this is under debate as others reported the affinity of IgG to DC-SIGN to be independent of Fc glycosylation, but rather influenced by Fab glycosylation [59]. The potential cause of this confusion is the large difference between humans and mice in DC-SIGN and SIGN-R1 cellular and tissue distribution and the different sialic acid variants found in both species [45, 60].

Bisection

Although the presence or absence of a bisecting GlcNAc has a large effect on the overall Fc *N*-glycan structure [61], it did not seem to directly influence C1q or FcγR binding in any way [37]. Little is known about the biological relevance of Fc bisection, although presence of a bisecting GlcNAc was reported to inhibit the subsequent fucosylation of *N*-glycans [62]. Additionally, when bisection happens early in the biosynthesis pathway, that is, prior to full mannose trimming in the Golgi apparatus, the conversion from hybrid- to complex-type glycan is blocked, resulting in the propagation of bisected hybrid-type glycoforms [62, 63]. Together, this suggests that the presence of a bisecting GlcNAc might rather play a role as regulatory tool early in the glycan biosynthesis pathway, instead of having a direct effect on Fc interactions.

Hybrid-type glycans

Hybrid-type glycans occur on IgG in minor amounts [26, 27] and are present when the protein leaves the biosynthesis process before the formation of complex-type glycans was initiated. Although very little is known about the biological relevance of hybrid-type glycans on IgG, one study showed the negative effect of afucosylated, bisected, hybrid-type glycans on complement activation, as compared to afucosylated, bisected complex-type glycans [62].

Alternative mechanisms influenced by Fc glycosylation

The effect of Fc glycosylation on immunological responses is best studied for FcγR binding and complement activation via C1q protein. However, IgG has additional interaction partners that might influence an immunological response. The dependence on specific Fc glycosylation of these interaction partners is less well studied. One example is the DC-SIGN

receptor, briefly mentioned above [57, 59]. Others include the family of Fc receptor-like proteins (FcRL1 to 6), the neonatal Fc receptor (FcRn) and various lectins [64-66]. FcRL4 and 5 are expressed on B cells and have been shown to bind IgG. Furthermore, FcRL5 is reported to bind IgG in a glycan-specific manner, although it is unknown whether this is based on Fc or Fab glycosylation [65]. The FcRn binds IgG-Fc in the placenta thereby mediating the transport of IgG molecules from the mother to the fetus. Additionally, FcRn present on epithelial cells helps to recycle IgG, resulting in its relatively long half-life of about three weeks [67]. Differences – albeit not striking – in binding affinity for FcRn were found between differently glycosylated Fc regions [64]. Additionally, the IgG Fc glycosylation between mothers and their newborns was reported to be slightly different [68]. Although there is no consensus about the biological relevance of Fc glycosylation for FcRn binding, large effects can rather safely be ruled out [69, 70]. Furthermore, while for other glycoproteins, and also for Fab glycosylation, the asialoglycoprotein receptor expressed in the liver binds asialylated glycoforms and in that way reduces the half-life of the proteins, for IgG-Fc this receptor plays no role, likely due to restricted accessibility of the glycans [71, 72]. For IgG based biopharmaceuticals it is known that the presences of high mannose-type glycans of the Fc region does play a role in reducing its half-life, possibly caused by the involvement of specific lectins such as the mannose receptor [73]. This might be a reason why this glycoform is hardly found on endogenous IgG in the circulation.

1.2.2 IgG Fc glycosylation in health and disease

Fc glycosylation plays an important role in modulating IgG effector functions, as described above. IgG glycosylation is partly genetically regulated [74, 75], but also highly influenced by environmental factors [76-78]. The biological relevance of Fc glycosylation is supported by its associations found with altered physiological states, both in health and disease. First, Fc glycosylation associates with basic descriptors of the human population, like age and sex [28, 79-81]. Most pronounced is the lower galactosylation observed with older age [28, 79-81]. Notably, while women have higher galactosylation in early adulthood than men, this changes around female menopause, and results in similar galactosylation for women and men in late adulthood [28]. Fc sialylation follows this trend in the adult population, while bisection increases in early adulthood and fucosylation stays relatively constant [28, 79, 82]. For children, especially of the youngest age category, less is known about their Fc glycosylation differences connected to age and sex. The studies done until now only describe IgG glycosylation on the level of released glycans [83, 84], which ignore the effects of potential Fab glycosylation or differences in IgG subclass distributions. This issue will be assessed later in this thesis (Chapter 3). However, what is known until now is that in children agalactosylated species on IgG are lower with higher age and monogalactosylated species stay relatively constant [83]. Additionally, for children

between 6 and 18 years of age, it was found that bisection increases for boys and fucosylation decreases for girls [84]. Furthermore, it was proposed that most of the sex-related Fc glycosylation differences are introduced during puberty, under the influence of sex-related hormones [79, 84].

Moreover, Fc glycosylation is well characterized in healthy pregnant women [29, 85] and their newborns [86, 87], showing a substantial increase in galactosylation with pregnancy, which normalizes after delivery [29, 85]. Newborns exhibit similarly high galactosylation levels, as their IgG is completely mother-derived [86, 87]. Furthermore, pregnant women have high levels of IgG sialylation, while the levels of bisection and fucosylation are lower during pregnancy than after delivery [29, 85].

While age, sex and pregnancy are natural descriptors of the population, there are indications that also lifestyle influences IgG glycosylation. For example, smoking results in a slightly higher level of bisection [77, 88], and the level of Fc galactosylation is correlated to living area (e.g. rural or urban) [76]. Additionally, the food substance all-trans retinoic acid, when directly applied on cultured B cells, has shown to influence the galactosylation profiles of the IgGs produced in these cells [78].

All described genetic, environmental and physiological influencers of the Fc glycosylation suggest large inter-individual differences in this IgG feature, which were indeed observed in numerous IgG Fc glycosylation studies [28, 74, 89]. Although this might complicate the discovery of disease-related markers, intra-individual glycosylation changes are in general quite minor over restricted periods of time [90, 91] and common disease-related changes in IgG Fc glycosylation have previously been identified in the adult population. For example, patients suffering from various autoimmune diseases like rheumatoid arthritis [92-94], inflammatory bowel diseases [95] and lupus [96] have shown to carry a low Fc galactosylation on their total pool of IgG as compared to healthy controls. For ANCA-associated vasculitis, IgG Fc galactosylation and sialylation levels have shown to be able to predict a relapse of the disease [97]. Additionally, other inflammatory conditions, like HIV [98], alloimmune cytopenias [99, 100], and active tuberculosis infections [101], and also cancers such as ovarian cancer [102] and colorectal cancer [103] did show to associate with specific features of the total Fc glycosylation. Interestingly, antigen-specific subsets of IgG showed distinct glycosylation profiles from the total IgG. For example, HIV gp120-, HPA-1a- and RhD-specific IgGs all carry relatively high levels of afucosylated glycan structures, as compared to the total pool of IgG [98-100]. For the RhD-specific IgGs in hemolytic disease of the fetus and newborn it was shown that this low fucosylation was correlated to their ability to induce FcγRIIIa mediated ADCC of the red blood cells, and in that way likely influences the severity of the disease [99].

Both the innate and adaptive immune system of the human child are relatively immature at birth and their maturation process continues until early adulthood [104]. Therefore, one can imagine that the IgG glycosylation effects found in children might be different from the ones observed in adults, both in healthy and diseased conditions. In general, information on IgG Fc glycosylation in children is scarce, especially when compared to the numerous studies performed in human adults. A recent study showed juvenile idiopathic arthritis to be associated with a lowered galactosylation and sialylation on the total IgG pool [66]. Furthermore, studies into pediatric patients with allergic diseases and asthma showed no disease-specific IgG glycosylation effects [76, 105]. Evaluating the similarities and differences between IgG glycosylation responses in children and adults might contribute to the better understanding of the immune system as a whole and the humoral immune responses in children, specifically.

1.3 Mass spectrometric glycosylation analysis

In order to unravel the exact glycosidic profile of any glycoprotein in a certain sample taken from a biological system, analytical methods are required that can assess protein glycosylation in a precise, accurate, and preferably protein- and glycosylation site-specific manner. Not one method exists that can easily perform the complete characterization of the glycoproteome of a complex sample, however various complementary separation and detection methods are around that can together give an in-depth overview. One important branch of glycosylation analysis methods makes use of mass spectrometry (MS) to separate and detect glycans or glycoconjugates based on their mass-to-charge ratio (m/z). Using MS, one can have roughly three different approaches studying the glycoproteome: 1) By the chemical or enzymatic release of glycans from proteins, resulting in a mixture of all glycan structures present in the sample, irrespective from which protein they were derived. This provides the opportunity to perform an in-depth characterization of the exact chemical structure of all glycans, however this approach fails to provide protein- or site-specific information. 2) By the enzymatic digestion of the glycoprotein backbone into smaller protein fragments or peptides, resulting in a mixture of glycosylated and glycan-free protein fragments. In this way, protein-specificity is maintained, as well as site-specificity, if only one glycosylation site is present per created fragment. As a downside, this approach will lead to a rapid increase of sample complexity when more than one kind of protein is present, making detailed structural glycan analysis laborious and the generation of protein- and site-specific glycopeptides challenging. 3) By the analysis of intact glycoproteins, often preceded by their specific isolation. Besides protein-specific data, this approach also provides the opportunity to assess the combination of different glycan structures and other post-translational modifications

present at one protein at the same time, however losing knowledge on site-specificity. For this approach high-resolution instruments are required.

The current work will focus on approach number 2, making use of a protease to create (glyco)peptides, which subsequently can be characterized by MS. Using MS, the glycan composition of a glycopeptide can be determined based on the theoretical mass of the peptide and the individual monosaccharides. However, stereoisomers cannot be distinguished, neither those arising from isomeric monosaccharides nor those arising from linkage isomerism, as they all result in identical m/z values (**Figure 1**). Additional information on the structure of specific glycan compositions and on the sequence of the peptide moiety can be obtained by applying tandem mass spectrometry (MS/MS) on selected ions, creating fragmentation spectra of the analytes of interest. Depending on the fragmentation mode used, either the glycan or the peptide portion of the glycopeptide is fragmented. Lower energy collision-induced dissociation (CID) results in glycan fragments, while higher energy CID, electron-transfer dissociation or electron-capture dissociation provide information about the peptide sequence [106]. A combination of high and low energy CID or post-source decay fragmentation results in the fragmentation of both portions of the glycopeptide [106, 107].

1.3.1 Mass spectrometers

Two important parts of a mass spectrometer are its ion source and mass analyzer, which both exist in different flavors and can be used in various combinations, all providing their own advantages in particular situations [108]. One frequently, yet not exclusively, used mass analyzer for glycopeptides is the reflectron time-of-flight (TOF) analyzer, which is able to separate analytes in a broad m/z range in a relatively short period of time [109]. This analyzer initially exposes all ions to an electrical field causing their acceleration while obtaining the same kinetic energy per charge. After their acceleration, the ions enter the flight tube of the TOF analyzer, where they travel with a constant velocity. Under these conditions, their velocity (and thus the flight time) is proportional to the m/z of the ions, with low mass or highly charged analytes traveling faster than high mass or lowly charged analytes [108]. Prior to entering the acceleration field of the TOF, during their ionization, the ions did acquire an initial velocity, which is not necessarily the same for all of them. This results in a dispersion of the final velocity (the sum of the initial velocity and the velocity acquired during acceleration) of identical species, which hampers the resolving power of TOF instrumentation. However, this is largely overcome by the implementation of delayed extraction of the ions during their acceleration [110] and the reflectron in the flight tube [111], both correcting for the spread in kinetic energy between ions with the same m/z .

As (glyco)peptides are not volatile and are preferably analyzed in their intact form, soft ionization technologies are required that allow ionization of the analytes from a liquid or solid sample. Next to obtaining a charge during the ionization, the molecules need to be brought into the gas phase, either via their own sublimation or desorption, or via the evaporation of their solvent or of a matrix. While some ionization techniques cause the molecules to fragment during ionization, soft ionization techniques will leave them largely intact. Two suitable approaches, often used in combination with a reflectron TOF analyzer, are matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) [108], both discussed in detail in the following sections.

1.3.2 MALDI-TOF-MS of glycopeptides

Ionization with MALDI takes place by the irradiation with a laser of a solid sample in vacuum. For this procedure the analyte molecules need to be embedded in a matrix, consisting of small organic compounds that absorb the light of the laser, which has a wavelength in either the UV or IR range, and are easy to sublimate [108, 112]. The exact mechanism behind MALDI is not completely understood as of yet, but ionization likely happens either via reactions in the gas phase after matrix ablation, or via the desorption and desolvation of ions formed in the solid sample [108, 113]. For the ionization of glycopeptides using MALDI with an UV-laser, examples of suitable matrices are α -cyano-4-hydroxycinnamic acid (CHCA), 4-chloro- α -cyanocinnamic acid (Cl-CCA) and 2,5-dihydroxybenzoic acid (DHB) [108, 114]. These matrices are soluble in diluted organic solvents that are compatible with the glycopeptides and carry an acidic group that stimulates the protonation of the analytes.

MALDI-TOF-MS is a widely recognized approach for the analysis of glycopeptides as the technique allows rapid acquisition of data, which are, in addition, relatively straightforward to interpret as the ions are usually singly charged [115, 116]. A typical MALDI glycopeptide workflow includes the enzymatic digestion of the (isolated) glycoproteins, enrichment of the glycopeptides, and co-application of the sample and matrix on the MALDI target [117]. Glycosylation profiles obtained by MALDI-TOF-MS of glycopeptides have shown to correlate well with the profiles obtained by other techniques, based on the high-performance liquid chromatography (HPLC)-fluorescence detection of released glycans [117]. However, this is only valid for the non-sialylated subset of the glycoproteome, as sialic acids are highly labile during MALDI-TOF-MS [117, 118]. In addition to the loss of sialic acids due to in-source or metastable decay, the negative charge of these sugar residues results in ionization biases for sialylated species, both in negative and positive ion mode, and in salt formation. The labile nature of the sialylated glycopeptides is especially problematic when they are analyzed in reflectron mode TOF-MS, as metastable fragments travel together in the field free region and are only separated after passing through the reflectron. Therefore, sialylated glycoconjugates

are often analyzed in linear mode TOF-MS, highly compromising the resolving power of the measurements. Other means for improving the detection of sialylated glycopeptides are the use of cold matrices like DHB or Cl-CCA and/or performing the analyses in negative ion mode [114]. However, none of these approaches fully prevent the loss of sialylated species, their ionization biases and salt formation.

In recent years, the neutralization and stabilization of sialylated glycans and glycoconjugates has appeared as an attractive approach for their MALDI-TOF-MS analysis. One method for this is the permethylation of released glycans or glycopeptides, resulting in the methylation of all hydroxyl, amine and carboxyl moieties in the molecules [119, 120]. Other approaches are sialic acid-specific and include the esterification or amidation of the sialic acid carboxyl groups [115]. This involves, on the one hand, the uniform modification of all sialic acids, for example by their amidation with the nucleophiles acetohydrazide [121] or methylamine [122]. On the other hand, linkage-specific sialic acid modifications have been reported, allowing the differentiation between α 2,3- and α 2,6-linked sialic acids based on their analysis by MALDI-TOF-MS. For example, α 2,6-linked sialic acids can be methylesterified [123], ethylesterified [124, 125] or isopropylamidated [126], while α 2,3-linked sialic acids are simultaneously lactonized [123, 124] or methylamidated [125, 126]. Specifically for glycopeptides, the application of a sialic acid linkage-specific derivatization method is more complicated than for released glycans, as the peptide moieties also carry carboxylic acids (at least one at their C-terminus and additionally one at every aspartic acid and glutamic acid residue). Based on the local environment of these peptide carboxylic acids, they might either react with the added nucleophile or create by-products by forming lactones or lactams with neighboring hydroxyl or amine groups, respectively. This was shown for glycopeptides derived from IgG that were subjected to the ethylesterification reaction proven to be suitable for released glycans [124], resulting in three modification variants of the peptide moiety [127]. On the other hand, the uniform derivatization of glycopeptide sialic acids by methylamidation resulted in consistent peptide modifications for glycopeptides derived from both IgG and fetuin [128].

1.3.3 ESI-TOF-MS of glycopeptides

In contrast to MALDI, ESI requires a liquid sample that is passed through a capillary and which creates a spray at the end of the capillary upon the application of a strong electric field. The solvent droplets in the spray evaporate and charged analytes either desorb from the surface of the droplets (small molecules) or stay in the droplet until complete evaporation took place (larger (glyco)peptides or proteins) [108]. While MALDI results in singly charged analytes, ESI causes multiple charged ions to arise [129]. This has the advantage that the mass analyzer used is required to cover a narrower m/z range,

however it complicates data analysis as one analyte might be spread over multiple signals (charge states) in the mass spectrum.

A typical ESI-TOF-MS workflow for glycopeptides includes the digestion of (isolated) glycoproteins, the online or offline enrichment or fractionation of the glycopeptides and their subsequent analysis by ESI-TOF-MS [130]. Sialylated glycans and glycoconjugates ionized by ESI suffer less from instability as compared to these analytes ionized by MALDI, and usually no derivatization of glycopeptides is incorporated in the ESI-TOF-MS workflow [131]. However, one should take into account that also with ESI, sialylated species can possibly suffer from ionization biases and salt formation. Additionally, like for MALDI-TOF-MS, linkage-specific sialic acid derivatization in ESI-TOF-MS has the advantage of differentiating between sialic acid linkages without any further separation or fragmentation needed [132, 133].

The online separation of glycopeptides is often achieved by the hyphenation of ESI-TOF-MS to liquid chromatography (LC). A separation step prior to MS enables the analysis of more complex samples and might provide additional structural information on both the peptide and the glycan moiety of the glycopeptides [130]. Reversed-phase (RP)-C18-LC is an often used separation method, which results in clustering of the glycopeptides that share the same peptide moiety and carry equally sialylated glycans. The elution of the sialylated glycopeptides relative to the neutral glycopeptides depends on the ion-pairing agent added to the mobile phase [134, 135]. Simultaneous elution of all glycoforms on a single peptide moiety results in their straightforward identification and relative quantification [25, 130, 131, 136]. LC separation techniques complementary to RP-LC include hydrophilic interaction liquid chromatography (HILIC) and porous graphitized carbon (PGC)-LC. HILIC is especially suitable for the separation of glycopeptides with short peptide sequences, which might not be retained by RP-LC [137] and it allows the efficient separation between glycosylated and non-glycosylated peptides in complex mixtures [134, 138]. Additionally, for sialylated glycopeptides, HILIC enables the differentiation between α 2,3- and α 2,6-linked sialic acids without any prior modification of the molecules [139]. Similar to HILIC, PGC-LC provides a platform for the separation of glycopeptides with short peptide moieties, with the advantage that samples do not have to be loaded in high concentrations of organic solvent, as is the case for HILIC and which might cause solubility issues [23, 140]. However it should be taken into account that both highly sialylated glycopeptides and glycopeptides with longer peptide sequences might be irreversibly retained on the PGC stationary phase [141, 142]. As the three mentioned separation approaches, RP-LC, HILIC and PGC-LC, all have complementary properties, also combinations are used for the characterization of glycopeptides [23, 143-145].

Capillary electrophoresis (CE) provides an additional separation method that is regularly coupled to ESI-TOF-MS. The separation mechanism of CE is complementary to RP-LC as glycopeptides are mainly separated based on their glycan moiety (if the peptide portions are short enough), resulting in the co-elution of similar glycoforms attached to different peptides [146]. Additionally CE-ESI-MS is extremely sensitive, enabling the profiling of glycans on low abundant proteins [146, 147]. While CE-ESI-MS of glycopeptides was reported to be used in combination with sialic acid linkage-specific derivatization [132], it has also been shown that CE is able to distinguish between species with differently linked sialic acids without extra sample preparation after the proteolytic digestion [148].

1.3.4 Mass spectrometric analysis of IgG glycopeptides

Specifically for the analysis of IgG glycopeptides, often trypsin is used for proteolytic cleavage [131, 133]. For the glycosylation site on the C_H2 domain, tryptic digestion results in glycopeptides with the sequences EEQYNSTYR for IgG1, EEQFNSTFR for IgG2, EEQFNSTFR, EEQYNSTFR or TKPWEEQYNSTFR for IgG3 and EEQFNSTYR for IgG4 [22]. The different variants of the IgG3 peptide moieties are a result of the numerous different allotypes known for IgG3, of which the occurrence is dependent on the ethnical background of the population studied. For the Caucasian population it is often assumed that the IgG3 sequence is identical to the IgG2 sequence, resulting in indistinguishable glycosylation profiles on these subclasses. However, it cannot be ruled out that the IgG3 sequence corresponds in mass to the IgG4 sequence (EEQYNSTFR vs. EEQFNSTYR) or has an unique longer sequence (TKPWEEQYNSTFR) as long as the allotypes occurring in the samples are not determined [21].

The MALDI- and ESI-TOF-MS approaches for glycopeptide analysis described above are reported to be used for the glycosylation profiling of IgG glycopeptides. For example, numerous studies report the use of RP-LC-ESI-TOF-MS for the IgG glycosylation profiling of glycopeptides derived from the C_H2 domain [25, 131, 136]. Additionally, CE-ESI-MS showed to be suitable for the highly sensitive analysis of IgG glycopeptides, at the same time being able to differentiate between α 2,3- and α 2,6-linked sialylated species [147, 148]. Specifically for IgG3, PGC- and RP-LC were combined to characterize the *N*-glycosylation of the C_H2 and C_H3 domains and the *O*-glycosylation of the hinge region in the same run, after its non-specific proteolytic treatment with pronase [23]. Finally, for the high-throughput analysis of the tryptic C_H2 domain glycopeptides of IgG, the application of MALDI-TOF-MS methods with or without sialic acid derivatization were reported [114, 120, 127, 128, 131].

1.4 Scope

The reliable characterization of antibody glycosylation is of utmost importance, both in unraveling (patho)physiological processes and in identifying potential clinical markers for disease diagnosis, progression and treatment. The aim of the work described in this thesis is, on the one hand, to develop new methodologies for the analysis of antibody glycosylation, and, on the other hand, the application of glycopeptide profiling methods on clinical samples to gain a better understanding of the behavior and role of antibody glycosylation in health and disease.

In **Chapter 1**, the general concept of protein glycosylation is introduced. Specifically for the antibody IgG, a more in-depth description of its glycosylation features and functions is given. Additionally, the currently used methodologies for the characterization of protein glycosylation are described, with a special focus on the MALDI- and ESI-TOF-MS analysis of glycopeptides, as these technologies are used throughout the presented work.

While MALDI-TOF-MS is a relatively straightforward and rapid manner to study IgG glycopeptides, obtaining an unbiased overview of both the neutral and the sialylated glycopeptides is still a major challenge. In **Chapter 2**, the sialic acid linkage-specific stabilization of sialylated glycoforms, which before was reported for released glycans, is extended to sialic acid linkage-specific IgG Fc glycopeptide derivatization followed by MALDI-TOF-MS analysis.

IgG Fc glycosylation was already extensively studied in various clinical settings, however both the function and regulation of IgG Fc glycosylation is not completely understood as of yet. Additionally, most studies were performed in human adults, while knowledge on pediatric IgG glycosylation is lagging behind. **Chapter 3** uses the MALDI-TOF-MS method developed in **Chapter 2** to characterize the IgG Fc glycosylation of healthy newborns and children. To get an insight into the regulation of IgG Fc glycosylation, in **Chapter 4** this glycosylation is assessed in children in need of a hematopoietic stem cell transplantation, using a previously developed RP-LC-ESI-TOF-MS method. Samples of the patients before and after transplantation are characterized and compared to the profiles of the donors.

One of the aspects of clinical marker development is being able to distinguish between individuals with and without a specific disease. Additionally, it might be very valuable if these markers can also tell something about the subtype and/or severity of the disease, or provide information on disease progression or treatment success, as this offers the opportunity to tailor treatment in an individualized manner. **Chapter 5** deals with the characterization of IgG Fc glycosylation in pediatric meningococcal sepsis. Next to differentiating between diseased and healthy children, glycosylation signatures specific for

disease severity are assessed and glycosylation features are reported that have the potential to predict disease outcome shortly after patient submission to the hospital.

In **Chapter 6**, IgG Fc glycosylation is evaluated in the case of inflammatory bowel diseases in adults. In this study, specific glycosylation signatures are correlated to disease subtypes (Crohn's disease or ulcerative colitis) and progression. Additionally, the influence of treatment, either by medication or surgery, on the Fc glycosylation profile is described.

The majority of functional studies into IgG Fc glycosylation have been done using mice as model organisms. However, murine IgG glycosylation possesses different characteristics as compared to human IgG glycosylation and hitherto no systematic and subclass-specific profiling of mouse IgG glycosylation has been performed. In **Chapter 7**, the IgG Fc glycosylation of four commonly used mouse strains (both inbred and outbred) is profiled in a subclass- and site-specific manner using RP-LC-ESI-TOF-MS.

While IgG in human plasma is an extensively studied antibody with regard to its Fc glycosylation, information on other antibody subtypes and on antibodies from other sources than plasma is scarce. **Chapter 8** provides a method to isolate both IgG and IgA from human saliva and uses RP-LC-ESI-TOF-MS for their site-specific glycosylation analysis. Finally, a comparison is made between the glycosylation of IgG and IgA derived from saliva and plasma of the same individuals.

The final chapter of this thesis, **Chapter 9**, offers a general discussion of the work described, with a specific focus on mass spectrometric method development for the analysis of antibody glycosylation in a clinical setting.



- ¹ *Leiden University Medical Center, Center for Proteomics and Metabolomics,
Leiden, The Netherlands*
- ² *Roche Diagnostics GmbH, Pharma Biotech Development Penzberg,
Penzberg, Germany*
- ³ *VU University Amsterdam, Division of BioAnalytical Chemistry,
Amsterdam, The Netherlands*