Multi-level structural and functional characterization of therapeutic glycoproteins by mass spectrometry

Lippold, S.

Citation

Lippold, S. (2021, November 30). *Multi-level structural and functional characterization of therapeutic glycoproteins by mass spectrometry*. Retrieved from https://hdl.handle.net/1887/3243984

1

INTRODUCTION

1.1. THERAPEUTIC PROTEINS

THERAPEUTIC proteins are successful and important biopharmaceuticals and many of them are on the World Health Organization list of essential medicines [1, 2]. Recombinant protein-based medicines have been introduced in recent decades, improving accessibility and enabling new treatment opportunities. Important milestones were the approval of recombinant human insulin (Humulin®, 1982), recombinant human erythropoietin (Epogen®, 1989) and the first monoclonal antibody (mAb) for targeted cancer therapy (Rituxan®, 1997) by the Food and Drug Administration (FDA) [3–5]. Nowadays, mAbs are the dominating entity of therapeutic proteins and have been established for therapeutic areas such as oncology, dermatology, hematology and neurology [1]. In addition, more complex formats, e.g. bispecific antibodies or antibody-drug conjugates, are currently emerging [6]. Numerous mAb-based drugs are on the market (100 FDA approvals, status April 2021) with a steadily increasing number in development [6, 7]. Recombinant therapeutic proteins are much larger compared to small molecule (synthetic) drugs and exhibit an inherently higher degree of structural complexity.

1.1.1. GLYCOSYLATION OF THERAPEUTIC PROTEINS

Glycosylation of (therapeutic) proteins is of high importance for their physico-chemical properties, such as solubility and stability, as well as for pharmacokinetics and pharmacodynamics [8, 9]. For example, glycosylation affects the half-life of erythropoietin (EPO) [10, 11]. Glycosylation also impacts the safety and efficacy of therapeutic proteins [12]. A protein may vary in the occupancy of a glycosylation site (macroheterogeneity) or in the glycan structure present at a single glycosylation site (microheterogeneity) [13]. Protein glycosylation is categorized into *N*- and *O*-glycosylation (**Figure 1.1**). Asn residues in a consensus sequence motive (Asn-x-Ser/Thr, x $\neq$ Pro) may be subjected to *N*-glycosylation [14]. In contrast, *O*-glycosylation (Ser/Thr) cannot be easily predicted by a specific sequence motif [15].

The most common building blocks of mammalian protein glycans are hexoses (galactose, mannose), *N*-acetylhexosamines (*N*-acetylglucosamine, *N*-acetylgalactosamine), deoxyhexose (fucose) and sialic acids (*N*-acetylneuraminic acid and *N*-glycolylneuraminic acid) (**Figure 1.1**). Furthermore, glycan building blocks may be decorated by other modifications such as acetylation or phosphorylation. *N*-glycans share a common core and are structurally divided into high mannose-, hybrid- and complex-type glycans. Mucin-type *O*-glycans show an *N*-acetylgalactosamine starting unit and are grouped according to eight different core structures [16]. On therapeutic proteins, the most often found *O*-glycans are of the core 1 type.

The glycosylation features of recombinant therapeutic proteins are highly dependent on the production host [17]. For therapeutic proteins relying on glycosylation for the mechanism of action, mammalian cell lines are usually chosen. The glycosylation profiles of mammalian cell derived proteins are closer to human glycosylation compared to bacterial or fungal production systems. Common cell lines for recombinant therapeutic protein production are Chinese hamster ovary (CHO), baby hamster kidney (BHK), murine myeloma cells (NS0, SP2/0) or human embryonic kidney 293 (HEK293) cells [18, 19]. Of note, hamster and murine cells additionally express non-human glycan features which negatively affect the safety of therapeutic proteins. For example, the presence of *N*-glycolylneuraminic acid and *α*1-3-linked digalactose motives may cause immunogenicity [20].

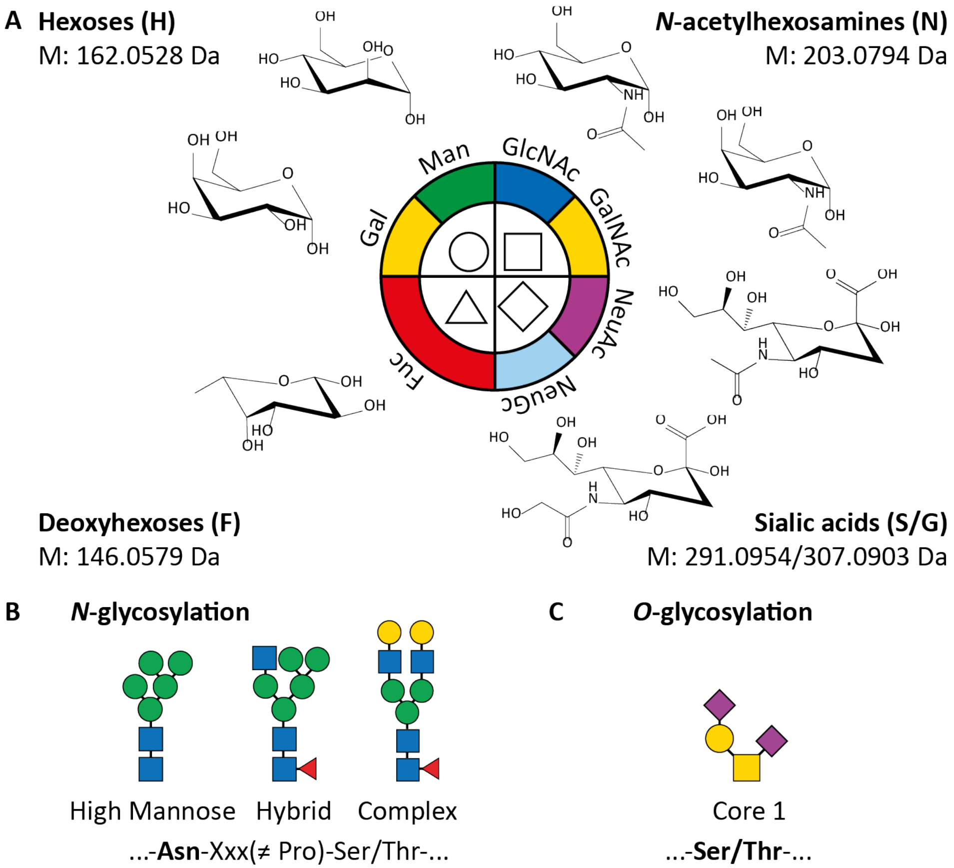

Figure 1.1: Common glycan building blocks of therapeutic protein glycosylation. **A** Monosaccharides are depicted with *α*-chair structure, symbol and color-coding. Monoisotopic masses (M) of monosaccharide residuals are indicated. **B** and **C** Representation of common *N*- and *O*-glycan structures observed for therapeutic proteins. Of note, all information is related to mammalian-based expression systems, which are usually used for therapeutic protein production.

1.1.2. STRUCTURE OF IMMUNOGLOBULIN G-BASED ANTIBODIES

The protein sequence of most therapeutic mAbs is based on human immunoglobulin G (IgG) [21]. IgG consist of two heavy and two light chains, arranged in a Y-form and connected via disulfide bonds (**Figure 1.2**). The antigen-binding fragment (Fab) is formed by the Fd (VH and CH1) and the light chain (VL and CL). Three complementarity-determining regions (CDR) in the Fab are responsible for antigen binding. The Fab is connected to the fragment crystallizable (Fc) domain (CH2 and CH3) via a hinge region. IgG subclasses 1, 2 and 4 are currently being used for therapeutic mAbs. The subclasses differ in specific characteristics, such as half-life, effector functions and stability, which affect the suitability for manufacturability and therapeutic purposes [22]. IgG1 is the most commonly used subclass due to characteristics such as high affinity to Fc receptors and long plasma half-life [21]. Of note, a recent review suggests to reconsider the therapeutic potential of engineered IgG3 due to enhanced effector functions and better antigen accessibility [23]. In addition, the different subclasses of IgG can be further subdivided into distinct allotypes [24, 25]. The Kabat numbering is commonly used to describe positions of amino acid residues [26]. It was the first standardized numbering system and facilitated the comparison of different antibodies.

IgGs carry a conserved *N*-glycosylation site in the heavy chain (CH2 domain) at Asn297 [27]. Since IgGs exhibit two glycosylated heavy chains, heterogeneous glycan pairings are commonly present (**Figure 1.2**). Due to the heterogeneity of glycan features and the different combination possibilities of mAb glycan pairings, the functional understanding of glycosylation is a major analytical challenge [28]. Further, non-canonical Fab *N*-glycosylation may be present in the variable domains [29]. Of note, *O*-glycans have also been reported for elongated therapeutic mAbs [30]. For the conserved Fc *N*-glycosylation, a dedicated nomenclature (GxF) is commonly used for describing complex-type diantennary glycans of IgGs. Gx indicates the number (x) of terminal galactoses and F the presence of core fucose. In addition, GxF-N (monoantennary), GxFN (bisection) and Mx (high mannose, x = number of mannoses) are differentiated [12].

1.1.3. FUNCTIONAL RELEVANCE OF IMMUNOGLOBULIN G GLYCOSYLATION

Glycosylation features of IgG Fc *N*-glycans are known to impact effector functions [12, 29]. For example, the interaction of IgG and the Fc gamma receptor III (FcγRIII) is known to be highly glycosylation dependent, which is attributed to unique glycan-glycan interactions [31]. Two different types of FcγRIII are expressed on natural killer cells or neutrophils. Both types initiate key immunological processes such as antibody-dependent cellular cytotoxicity or antibody-dependent cellular phagocytosis [32]. Hence, the effector functions of a therapeutic mAb are greatly affected by its glycosylation features. Fucosylation of IgG Fc glycans is known to tremendously (up to 50 fold) decrease the binding affinity towards FcγRIII [33]. Other glycosylation features

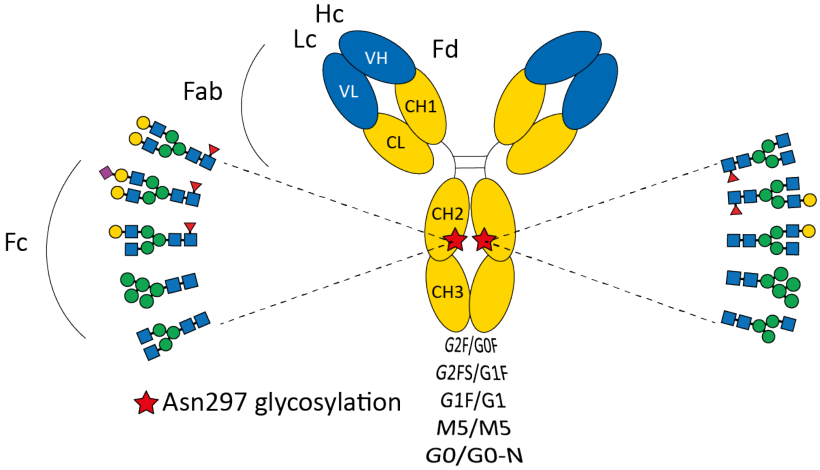

Figure 1.2: Scheme of IgG1-based mAb and commonly observed Asn297 glycans.

such as galactosylation or bisection also influence the binding affinity, although to a smaller extend [8, 34, 35]. Glycoengineering approaches to decrease fucosylation have been successfully proven to enhance the clinical efficacy of mAbs such as Gazyva® [29, 36, 37]. Besides effector functions, the pharmacokinetics are also affected by Fc glycosylation. For example, hybrid and high mannose glycoforms are cleared faster from circulation [38].

1.1.4. OTHER POST TRANSLATIONAL MODIFICATIONS OF PROTEINS

Besides glycosylation, additional post translational modifications (PTMs) occur naturally, or they are introduced accidentally during manufacturing, purification or storage of therapeutic proteins (**Figure 1.3**) [39]. Common PTMs are C-terminal Lys/Arg clipping-variants, deamidation (Asn, Gln) or oxidation (Met, Trp, Cys). Certain PTMs are known to impact the function of therapeutic proteins and hence are defined as critical quality attributes (CQAs) [40]. For example, Asn deamidation of mAbs has been reported to affect antigen binding or Fc receptor binding [41–43]. Fc oxidation decreases half-life, owing to decreased neonatal Fc receptor binding and scavenging [44, 45]. Co-occurring PTMs on therapeutic proteins cause a high degree of complexity. Glycosylation is the most complex PTM and hence a major contributor to proteoform heterogeneity. For complex glycoproteins such as EPO, several hundred proteoforms have been reported [46]. The FDA estimated the presence of 108 theoretical proteoforms for a therapeutic mAb [47]. However, it is critical to understand which of these proteoforms are abundant enough to cause a significant biological/clinical effect. Hence, the analytical characterization of PTMs is crucial and the elucidation of structure-function relationships is of high importance to assess the criticality of a PTM [48].

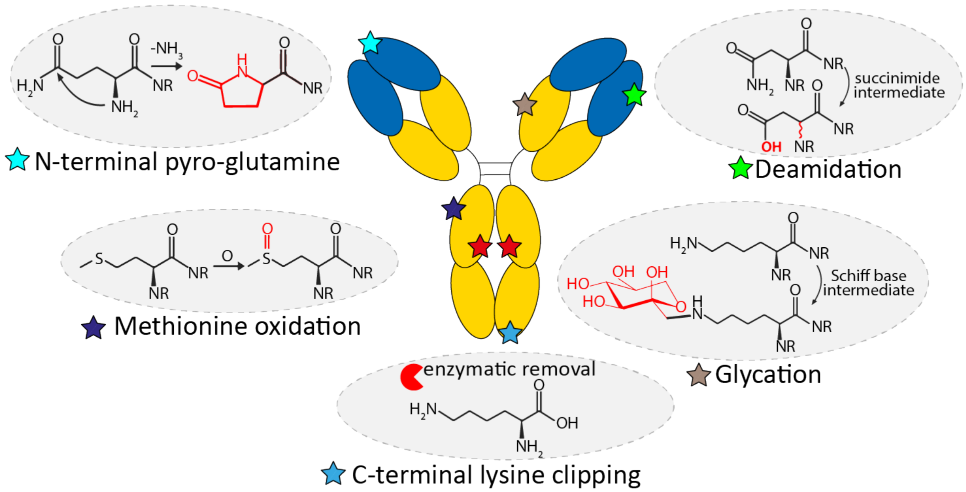

Figure 1.3: Visualization of different post-translational modifications commonly present in therapeutic proteins.

1.2. MASS SPECTROMETRIC ANALYSIS OF THERAPEUTIC GLYCOPROTEINS

Tremendous technological improvements in mass spectrometry (MS) instrumentation have been made in the last decades, which redefined the way (therapeutic) glycoproteins are analyzed [49–53]. The gold standard for glycosylation analysis was based on hydrophilic interaction liquid chromatography of released glycans with fluorescence detection [54]. However, technological advances allow for a protein-specific glycosylation analysis with MS detection. MS is a versatile technique, which is used extensively for the characterization of therapeutic glycoproteins in academia and industry [39, 55, 56]. In MS, the mass to charge ratio (*m/z*) of an ionized analyte is determined in the gas phase. The two most common ionization techniques for large biomolecules, such as proteins, are matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) [57, 58]. Both techniques allow for soft ionization of biomolecules, i.e. non-destructive ion formation. Several analytical MS-based approaches have been developed for (therapeutic) proteins and PTMs, namely bottom-up, middle-up and intact mass analysis (**Figure 1.4**). For a comprehensive analytical characterization of therapeutic proteins, the integration of multi-level analyses is pivotal [59–61].

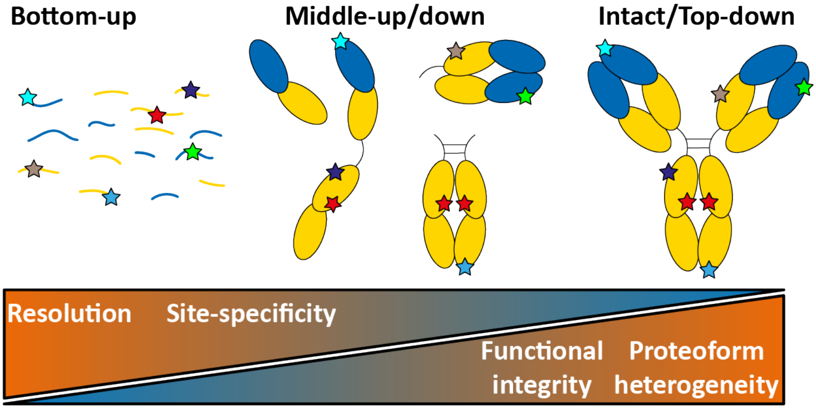

Figure 1.4: Schematic representation of different approaches for the MS analysis of mAb proteoforms.

1.2.1. IONIZATION TECHNIQUES

MALDI was introduced by Karas and Hillenkamp [62, 63]. Several advantages, such as short analysis time, low spectrum complexity and straightforward data interpretation, make MALDI highly suitable for high-throughput applications. In MALDI, the analyte is first co-crystallized with a MALDI matrix on a target plate. Small aromatic compounds, which show absorption in the ultraviolet or infrared wavelengths, are usually used as MALDI matrices. A laser pulse irradiation under vacuum causes a rapid heating of the crystals. Of note, the exact ionization mechanism of MALDI is not fully grasped. Ion formation is suggested either via a proton transfer mechanism in the solid phase prior to desorption or in the gas phase upon ablation [64, 65]. Predominantly singly charged ions are produced by MALDI, which simplifies the data interpretation. Most crucial for successful MALDI experiments is the selection of an appropriate matrix, which depends on the analyte of interest [66].

ESI of proteins was first described by Fenn et al. [58]. Separation techniques may be hyphenated to MS via ESI, which is highly advantageous when analyzing complex mixtures. In ESI, the liquid flow is dispersed into a spray under high voltage and highly charged droplets are produced. Solvent molecules are evaporated, and ions are formed via desorption of charged analytes from the surface of the droplets or after their complete desolvation. In contrast to MALDI, ions produced by ESI are multiply charged. This increases the potential mass range of analytes and MS resolution. Furthermore, ESI allows for analyzing proteins or protein complexes retaining a native-like structure in the gas phase [67].

1.2.2. BOTTOM-UP ANALYSIS

In bottom-up approaches, proteins are analyzed at the peptide level. The advantages are the high resolution and the site-specific information obtained on modifications. In bottom-up workflows, proteins are usually reduced and alkylated followed by protease cleavage. Trypsin is the gold standard protease. However, other specific proteases exist and are useful alternatives to obtain complementary information [68]. Peptides are then separated, mostly using reversed-phase (RP) liquid chromatography (LC) prior to ESI-MS. For identification purposes, peptides are fragmented e.g. via gas-phase collisions or radical transfer reactions to assign sequences and potential PTMs. In the biopharmaceutical industry, bottom-up methods are popular for multiple-attribute monitoring, meaning the simultaneous assessment of different PTMs. This is related to the higher assignment confidence and specificity achieved by bottom-up approaches compared to middle-up or intact analysis. However, the actual information of proteoforms is lost and artificial modifications may be introduced during the long sample preparation [69]. Bottom-up methods are very data-rich and efficient data processing remains a bottleneck, in particular for glycopeptides.

1.2.3. INTACT ANALYSIS

Intact analysis of proteins is performed without prior chemical or enzymatic sample preparation. In contrast to bottom-up approaches, information about the combination of PTMs on one molecule is revealed, e.g. pairing of glycans on a mAb. Further, better insights on the actual proteoform heterogeneity are achieved since the distribution of PTMs, i.e. glycans, does not follow a random statistical distribution. However, intact protein analysis lacks sensitivity and has a high degree of proteoform assignment ambiguities, because intact mass profiles may be very complex. Particularly, glycoproteins with multiple glycosylation sites, such as EPO, suffer from unresolved isobaric and isomeric PTMs [59].

Intact protein analysis by MS is greatly facilitated by online hyphenation to separation techniques such as size-exclusion chromatography (SEC) or ion exchange chromatography (IEC) [70, 71]. Alternatively, Fc receptor affinity chromatography (AC) is a recently introduced strategy for functional separation of mAb proteoforms [44, 72]. The commercialization of Fc receptor affinity columns as analytical tool is currently emerging. AC allows to decomplexify intact glycoprotein mixtures. However, hyphenation of functional AC and MS is yet unexplored in literature for mAb analysis.

1.2.4. MIDDLE-UP ANALYSIS

As alternative to intact protein analysis, middle-up strategies are often applied for the analysis of antibodies, since they reduce the proteoform heterogeneity and increase the MS resolution and sensitivity. Workflows consist of hinge-specific proteolysis and/or simple reduction of disulfide bonds. Due to vastly different ionization efficiencies of the subunits, middle-up experiments are commonly performed in combination with a separation dimension. In recent years, hinge-specific proteases have been commercialized and have facilitated the subunit analysis of therapeutic IgGs [73–75]. IgG hinge-proteases are very specific and require only a short sample preparation. In addition, some hinge protease can cleave under native conditions, which allows to analyze the distinct function of Fc and Fab fragments. Middle-up experiments complement the information obtained by bottom-up or intact protein analysis and are a good compromise of structural integrity and site-specificity. Hence, this approach is very helpful for assigning PTMs to a specific subunit.

1.3. SCOPE

The analytical characterization of therapeutic glycoproteins is crucial to assess their safety and efficacy. However, the complexity of glycosylation is a major challenge for analytical methods. The aim of the work in this thesis was to develop new mass spectrometry-based techniques for the characterization of therapeutic glycoproteins. The developed methods address current issues in glycoproteomic data-processing and sample preparation and show the feasibility of MALDI for intact glycoform profiling. They also demonstrate the power of intact- and middle-up mAb Fc receptor affinity chromatography - mass spectrometry for proteoform-resolved structure-function studies.

In **Chapter 1** the importance of therapeutic glycoproteins is emphasized. PTMs, with focus on glycosylation, are described with respect to their impact on proteoform heterogeneity and structure-function relationships. Further, the most common ionization techniques for protein analysis, MALDI and ESI are briefly introduced. Finally, different MS-based analytical strategies for the characterization of mAbs are elaborated.

Bottom-up approaches using RP LC-MS/MS are commonly used for glycoproteomic experiments. These approaches are very powerful for comprehensive glycosylation characterization. However, the vast complexity of data remains a major bottleneck in glycoproteomic workflows. In **Chapter 2,** bioinformatic tools are combined into a semi-automated workflow for glycoproteomics data analysis. Glycopeptide coverage and the reliability of relative quantification could be increased.

Protease selection is a crucial step for successful glycoproteomic experiments. One important selection criterion for a protease is that it generates singly glycosylated glycopeptides. In **Chapter 3** an alternative cysteine alkylation strategy, aminoethylation, is applied for the first time for the bottom-up analysis of EPO. Aminoethylation allows tryptic cleavage at newly introduced cleavage positions between two glycosylation sites, which could be previously only achieved by less specific proteases.

Intact protein analysis has the unique ability to reveal the combination of glycoforms of multiply glycosylated proteins. MALDI MS is usually not performed for comprehensive intact glycoform profiling, owing to the low glycoform resolution and low stability of glycosylation features, i.e. sialic acids. **Chapter 4** demonstrates the feasibility of glycoform-resolved intact analysis of EPO by MALDI Fourier-transform ion cyclotron resonance MS. The method shows a high stability of sialic acid moieties and facilitates the data analysis as no charge deconvolution is required.

Intact mass analysis of mAbs is commonly done by direct infusion MS or separation-based techniques such as RP-HPLC MS. However, glycoforms are usually not separated in this manner and no functional information is derived. In **Chapter 5**, a method for functional FcγRIIIa AC with online MS hyphenation is presented. New insights into the FcγRIIIa affinity of mAb glycoform pairings are derived from this analytical setup.

The glycoform assignments of intact mAbs are highly ambiguous, if additional glycosylation is present in the Fab. To expand the applicability of the developed method from Chapter 5, different IgG hinge proteases are evaluated in **Chapter 6** for an FcγRIIIa AC-MS method of Fc-moieties. The advantages with respect to proteoform resolution and assignment ambiguities, are demonstrated for Fab-glycosylated cetuximab.

In **Chapter 7**, an AC-MS method is described for FcγRIIIb mAb interactions. This is the first report on an FcγRIIIb AC, demonstrating the suitability of AC to study low-affinity interactions.

A discussion about the challenges and outlooks of MS data analysis for glycoproteins and the potential of functional AC-MS for biopharmaceutical research and development is provided in **Chapter 8**.