Pieter Langerhorst, Melissa Baerenfaenger, Purva Kulkarni, Simon Nadal, Charissa Wijnands, Merel A. Post, Somayya Noori, Martijn M. vanDuijn, Irma Joosten, Thomas Dejoie, Alain J. van Gool, Jolein Gloerich, Dirk J. Lefeber, Hans J.C.T. Wessels and Joannes F.M. Jacobs*

N-linked glycosylation of the M-protein variable region: glycoproteogenomics reveals a new layer of personalized complexity in multiple myeloma

https://doi.org/10.1515/cclm-2023-1189
Received October 24, 2023; accepted January 28, 2024;
published online February 9, 2024

Abstract

Objectives: Multiple myeloma (MM) is a plasma cell malignancy characterized by a monoclonal expansion of plasma cells that secrete a characteristic M-protein. This M-protein is crucial for diagnosis and monitoring of MM in the blood of patients. Recent evidence has emerged suggesting that N-glycosylation of the M-protein variable (Fab) region contributes to M-protein pathogenicity, and that it is a risk factor for disease progression of plasma cell disorders. Current methodologies lack the specificity to provide a site-specific glycoprofile of the Fab regions of M-proteins. Here, we introduce a novel glycoproteogenomics method that allows detailed M-protein glycoprofiling by integrating patient specific Fab region sequences (genomics) with glycoprofiling by glycoproteomics.

Methods: Glycoproteogenomics was used for the detailed analysis of *de novo* N-glycosylation sites of M-proteins. First, Genomic analysis of the M-protein variable region was used to identify *de novo* N-glycosylation sites. Subsequently glycopeptide analysis with LC-MS/MS was used for detailed analysis of the M-protein glycan sites.

Results: Genomic analysis uncovered a more than two-fold increase in the Fab Light Chain N-glycosylation of M-proteins of patients with Multiple Myeloma compared to Fab Light Chain N-glycosylation of polyclonal antibodies from healthy individuals. Subsequent glycoproteogenomics analysis of 41 patients enrolled in the IFM 2009 clinical trial revealed that the majority of the Fab N-glycosylation sites were fully occupied with complex type glycans, distinguishable from Fc region glycans due to high levels of sialylation, fucosylation and bisecting structures.

Conclusions: Together, glycoproteogenomics is a powerful tool to study *de novo* Fab N-glycosylation in plasma cell dyscrasias.

Keywords: glycoproteogenomics; multiple myeloma; M-protein; N-glycosylation; variable region

Introduction

Multiple myeloma (MM) is a plasma cell malignancy characterized by a monoclonal expansion of plasma cells in the bone marrow. These malignant plasma cells produce and secrete a characteristic monoclonal immunoglobulin (M-protein) [1]. M-protein diagnostics is imperative for diagnosis and monitoring of patients with MM [2]. Similar to other antibodies, both the heavy and the light chain of M-proteins consists each of a constant (Fc) and a variable (Fab) region. The Fc region is

Hans J.C.T. Wessels and Joannes F.M. Jacobs contributed equally to this work.

***Corresponding author: Joannes F.M. Jacobs**, Department of Laboratory Medicine, Radboud University Medical Center, Laboratory Medical Immunology (Route 469), Geert Grooteplein 10, 6525 GA Nijmegen, The Netherlands, Phone: +0031 (0)24 3614625,
E-mail: h.jacobs@radboudumc.nl
Pieter Langerhorst, Charissa Wijnands, Irma Joosten, Alain J. van Gool, Jolein Gloerich and Hans J.C.T. Wessels, Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands. https://orcid.org/0000-0001-9057-8751 (C. Wijnands)
Melissa Baerenfaenger, Department of Neurology, Donders Institute for Brain, Cognition, and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands; and Division of BioAnalytical Chemistry, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Purva Kulkarni, Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands; Department of Medical BioSciences, Radboud University Medical Center, Nijmegen, The Netherlands; and Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands
Simon Nadal, CY Cergy Paris Université, CNRS, BioCIS, Cergy-Pontoise, France. https://orcid.org/0000-0002-0748-0284
Merel A. Post and Dirk J. Lefeber, Department of Neurology, Donders Institute for Brain, Cognition, and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands
Somayya Noori and Martijn M. vanDuijn, Department of Neurology, Erasmus University Medical Center, Rotterdam, The Netherlands. https://orcid.org/0000-0002-6654-994X (M.M. vanDuijn)
Thomas Dejoie, Biochemistry Laboratory, Centre Hospitalier Universitaire (CHU), Nantes, France

immunoglobulin (Ig) isotype specific and harbors the effector functions of the antibody, and the Fab region defines the binding specificity of the antibody and is both patient- and myeloma specific [3]. The variable Fab domain arises through unique rearrangements of germline *V(D)J* genes and subsequent somatic hypermutations [4]. These unique regions and their derived clonotypic peptides can be used for blood-based monitoring of MM minimal residual disease activity [5–8].

Protein N-glycosylation is an important post-translational modification (PTM), where oligosaccharides are enzymatically attached to asparagine residues in proteins [9–11]. Immunoglobulins are the most abundant glycoproteins in the blood. All immunoglobulins (Ig) contain glycans that are linked to conserved motifs in the Fc-region. This Ig glycosylation is known to affect Ig function in Fc Receptor and complement component 1q binding [12, 13]. In addition, the Fab region may also acquire N-linked glycans. This, however, requires the introduction of amino acid consensus motifs for N-glycosylation during somatic hypermutation [12]. The estimated incidence rate of Fab N-glycosylation is ~15 %, yet the exact role of Fab N-glycans is not yet established [13]. Recent studies suggest that Fab glycans are involved in immune regulation, thereby playing a role in autoimmunity and also in the clinical effect of intravenous Ig products [12–15].

Recently, evidence has emerged that the presence of M-protein light chain (LC) Fab N-glycosylation contributes to the M-protein pathogenicity in amyloid light chain (AL)-amyloidosis [16]. Additionally, M-protein Fab-glycosylation of the LC is reported as a risk factor for Monoclonal Gammopathy of Undetermined Significance (MGUS) disease progression to AL-amyloidosis, MM and other plasma cell dyscrasias [17].

Methods to study Fab N-glycosylation have classically been based on glycan binding lectins or released glycan analysis by mass spectrometry (MS) of enriched Fab regions of IgGs [12–14, 18–20]. Recently, direct analysis of intact M-protein light chains by MALDI-TOF MS has been proposed to detect mass shifts corresponding with glycan mass distances [16]. However, all of these methods cannot resolve the M-protein glycosylation status at the level of individual N-glycosylation sites required to gain novel insights into M-protein glycobiology. Since N-glycosylation is a highly heterogenous PTM with variable glycan structures and occupancy rates, we believe that site-specific glycosylation data is key to discern the biological or clinical consequences of Fab N-glycosylation [21]. We hypothesize that the unique M-protein Fab N-glycosylation status can be characterized by personalized genomic analysis of the myeloma plasma cells to derive reference M-protein sequences to enable subsequent glycoproteomics for detailed site-specific glycosylation analysis [22]. This may expand our knowledge on M-protein pathogenicity.

Here, we introduce a novel glycoproteogenomics method, through integration of genomics and glycoproteomics, to predict and confirm *de novo* Fab N-glycosylation profiles of M-proteins from patients with MM in a site-specific manner. Based on genomic data, we showed that M-proteins from patients with MM more often have a Fab N-glycosylation site compared to Ig from healthy individuals. Furthermore, we revealed that the M-protein Fab glycoprofiles are mainly complex type glycans that structurally differ significantly from Fc-glycans. Together, glycoproteogenomics opens new opportunities to study the clinical relevance of *de novo* M-protein Fab N-glycosylation.

Materials and methods

Sequence analysis from public databased to determine M-protein Fab N-glycosylation site

The construction of M-protein Fab region sequences has been described elsewhere [3, 6]. In short, RNA-seq datasets from myeloma plasma cells were processed by MiXCR to construct the CDR3 sequences [23]. In case of incomplete sequences, Targeted Assembly of Short Sequence Reads was used [24]. Nucleotide sequences were translated into (poly)peptide sequences and germline amino acid alterations were determined by HIGH-VQUEST on the IMGT platform.

The Multiple Myeloma Research Foundation (https://themmrf.org/finding-a-cure) provided access to the data from the CoMMpass study on the NIH dbGAP platform (Accession: phs000748.v6.p4). At the time of analysis, 107 full length Fab region sequences were available for analysis.

The cAb-Rep is a database of curated immunoglobulin repertoires from both healthy and diseased donors [25]. The cAb-Rep database was accessed through the web browser interface, where FASTA files containing nucleotide sequences of the Fab region of immune repertoires of all healthy donors were retrieved. Sequences from either IgG heavy chain (HC) or kappa or lambda LCs were used for sequence analysis.

The Fab region sequences were analyzed by an in-house made R script (Version 3.5.3) that determines whether an N-glycosylation motif is present in the sequence. The assignment of an N-glycosylation motif was solely based on the sequences and defined as a asparagine residue followed by any amino acid besides proline followed by a serine, threonine or cysteine (non-canonical) residue (Asn-X-Ser/Thr/Cys) [26].

Samples from patients with multiple myeloma

Genomic data and serum samples were collected from the IFM 2009 clinical study (ClinicalTrials.gov identifier NCT01191060 trial) after written informed consent, and clinical and genomic data were de-identified in accordance with the Declaration of Helsinki and approval for this study was provided by our Institutional Review Board (2018-4140) [27]. The construction of Fab M-protein sequences has been described extensively elsewhere [6]. In short, Fab sequences were determined by analysis of a 100,000 read sample extracted from RNA sequencing data. Full length sequences were obtained by targeted assembly of reads from highly expressed mRNA based on MiXCR obtained seeding sequences. All serum samples were stored at −80 °C upon arrival.

M-protein purification, digestion and LC-MS/MS analysis

The purification and digestion of M-proteins is described in the Supplementary Materials & Methods. In short, M-proteins are purified by light chain specific affinity matrix. Subsequently, the M-proteins were reduced and alkylated followed by enzymatic digestion with Trypsin (Promega, Madison, Wisconsin, USA) and LysC (FUJiFILM Wako, Akasaka, Minato, Tokyo, Japan). The (glyco)peptide mixtures were measured by LC-MS/MS on the timsToF PRO 2 Mass Spectrometer (Bruker Daltonics, Billerica, Massachusetts, USA).

Data analysis

The data analysis of glycoproteomic LC-MS/MS datasets and their integration with genomic data is described in the supplementary information (Supplementary Materials & Methods). In short, LC-MS/MS datasets were analyzed to identify glycopeptide fragmentation spectra, where the glycan moiety and peptide were identified based on database searches. For the peptide database search, a custom database was created containing the M-protein Fab region sequences. The Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [28] partner repository with the dataset identifier PXD040718.

3D modeling of the Fab regions and their N-glycosylation sites

The 3D structures of the M-protein Fab regions were modeled by close homology crystal structures in AbyMod [29]. To acquire the model both a HC and LC Fab sequence has to be known. Since the M-protein of patient 7 consists of a LC only (Table 1) this patient was excluded from analysis. Sequences were entered and the known crystal structures with the highest sequences homology and resolution were selected. 3D structures of the Fab regions were exported and analyzed in pymol (version 2.4.0). The solvent-accessible surface area (SASA) was calculated with the function “get.area”.

Results

Increased *de novo* M-protein N-glycosylation sites in patients with multiple myeloma

The presence of Fab N-glycosylation on the M-protein LC has proven prognostic value for plasma cell dyscrasia, mainly in AL amyloidosis [16, 17]. Here, we investigated whether Fab N-glycosylation sites are more commonly observed in the M-proteins of patients with Multiple Myeloma compared to immunoglobulins in healthy individuals. To determine the occurrence of an N-glycosylation sites in the M-proteins of patients with MM, we reconstructed the M-protein Fab sequences based on RNA-seq data present in the CoMMpass database. We could reconstruct the full-length sequence of 104 HCs and 93 LCs. An N-glycosylation site was observed in 11.5 % of the HCs and in 11.8 % of the LCs (Figure 1A). This implies that Fab-glycosylation sites were observed in 22 % of the M-proteins. The NxS motif was most dominantly observed in HCs (9 out of 11) and the NxT motif (7 out of 11) was most often seen in LCs. The non-canonical N-glycosylation motif NxC was observed in one LC sequence (Figure 1B). The location of the glycosylation sites was predominantly confined to the region between Framework 1 (FR1) and FR2 in the LC sequences (Figure 1C, red). N-glycosylation sites of the HC were observed over the entire Fab region, with a hot spot at complementary determining region (CDR) 2 and FR3 (Figure 1C, blue).

Next, we determined the occurrence of Fab regions containing a *de novo* N-glycosylation site in the polyclonal immunoglobulin population of healthy donors available in the cAb-Rep database [25]. This database provided access to sequences of polyclonal HC of 19 healthy individuals (average sequences per donor=4,645) and sequences of

Table 1: Overview of Fab N-glycosylation motifs.

Patient ID	M-protein type	Glycosylated Fab	Glycosylation site	Glycosylation location	Glycopeptide
1	IgG-kappa	Kappa	N26	FR1	A**NQT**IFSTYLAWYQQK
		Kappa	N93	CDR3	LEPEDFAVYFCQHY**NTS**PLFTFGPGTK
2	IgG-kappa	IgG	N59	FR3	DSDP**NYS**PSFQGLVTMSSDK
		IgG	N76	FR3	S**NNT**AFLEWSSLK
3	IgG-kappa	Kappa	N26	FR1/CDR1	S**NQS**LLHDDGK
4	IgG-lambda	Lambda	N25	FR1/CDR1	VAISCSG**NNS**NIGFSSVSWYQQLPGTAPK
5	IgA-kappa	IgA	N28	CDR1	LTCAGSGF**NFS**TYWMHWVR
6	IgA-kappa	Kappa	N30	CDR1	ASQSI**NSS**FLVWYQQK
		IgA	N23	FR1	PSETLSLTC**NVS**GGSLISNSYYWGWIR
7	FLC kappa	Kappa	N28	CDR1	ASQ**NIS**FYLTWYQQNPGK

Amino acid alterations compared to germline are underlined and N-glycosylation motifs are indicated in bold. FR, framework; CDR, complementary determining region.

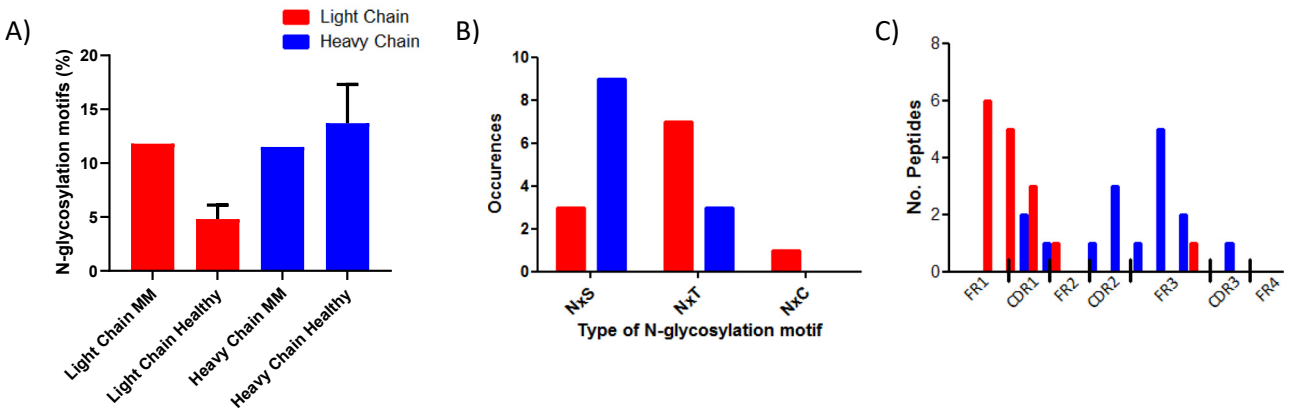

Figure 1: An increase in the occurrence of light chain N-glycosylation sites of the M-protein Fab region is associated with multiple myeloma. (A) A comparison of the occurrence of Fab N-glycosylation sites (%) on the heavy and light chain of M-protein in patients with multiple myeloma and healthy individuals (n=19 for the heavy chains and n=31 for the light chains), reveals that light chain N-glycosylation is more than 2 fold more common in M-proteins than expected from the polyclonal background. Error bars represent the standard deviation. (B) NxT N-glycosylation motifs are more common in light chains (red bars) and NxS motifs are more common in heavy chains (blue bars) of M-proteins in patients with MM. (C) Glycosylation sites in light chains (red bars) are confined to the Fr1/CDR1 in the *V(D)J* gene, while in heavy chains (blue bars) the glycosylation sites can be identified across the entire gene. The y-axis represents the frequence of occurrence of the site in a glycopeptide. CDR, complementary determining region; FR, framework; MM, multiple myeloma.

polyclonal kappa + lambda LCs of 31 healthy individuals (average sequences per donor=4,809). On average, 13.7 % of the HC Fab regions harbored a Fab N-glycosylation site (SD=3.6 %, range = 4.9–21.10 %). On average, 4.8 % (SD=1.3 %, range = 2.2–7.4 %) of the LCs harbored a Fab N-glycosylation site (Figure 1A), which is significantly lower compared to the HC ($p<0.0001$).

In conclusion, the occurrence of Fab N-glycosylation sites in the HC s of M-proteins in patients with MM was similar in healthy donors (11.5 vs. 13.7 %, Figure 1A). For the LC a more than 2-fold increase in the occurrence of Fab glycosylation sites in M-proteins compared to healthy donors was observed (11.8 vs. 4.8 %).

Studying M-protein *de novo* glycosylation using glycoproteogenomics

Protein N-glycosylation is a non-template driven process capable of generating thousands of unique N-glycan structures for several hundred unique sugar compositions. As such, the collection of N-glycan structures and their stoichiometry that decorate a single N-glycosylation site cannot be derived *in silico* from sequence information. Moreover, the mere presence of an N-glycosylation sequon does not guarantee any glycan occupancy. Therefore, we established a glycoproteogenomics workflow to characterize the glycosylation status of *de novo* N-glycosylations sites of patient-specific genomic alterations.

Firstly, based on RNA-seq datasets from malignant plasma cells in the bone marrow of patients with MM, the sequence of the Fab region of the M-proteins was predicted (Figure 2, genomics). These Fab protein sequences were used to identify the presence of N-glycosylation sites and *in silico* digestions were performed to predict the theoretical glycopeptides. Secondly, glycoproteomic analysis of the M-proteins was performed (Figure 2, glycoproteomics). Thirdly, glycoproteomics data were mapped on the RNA-seq derived Fab sequences of the M-proteins to annotate the clonotypic glycopeptides and determine their glycoprofiles (Figure 2, integration of genomics and glycoproteomics). Furthermore, the Fc region N-glycosylation was also monitored. An example of the generated data is shown for patient 1 (Figure 2B), in whom we identified two N-glycosylation sites in the Fab sequence (N15 and N128). We were able to identify site specific N-glycan profiles and differentiate them from Fc region N-glycosylation. Next, we applied this workflow to a cohort of 41 patients with MM.

Profiling *de novo* N-glycosylation sites in the Fab regions of M-proteins

Based on the RNA-seq derived Fab region sequences, we identified 10 N-glycosylation sites in 7 out of 41 patients with MM (Table 1). 9 out of 10 sites were the result of amino acid alterations and were thus *de novo* introduced into the Fab region. Only the IgG N59 site in patient 2 was already present

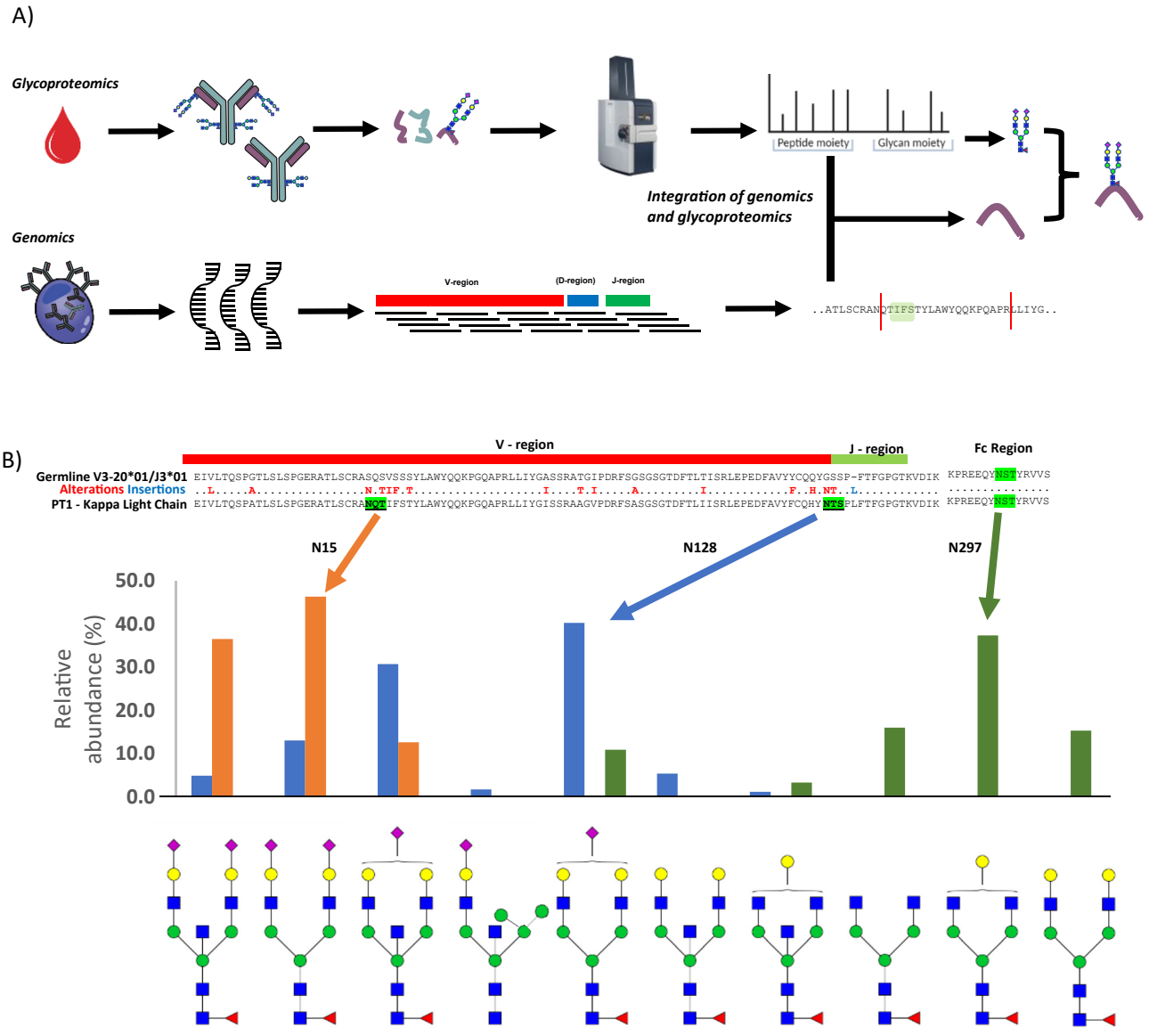

Figure 2: Glycoproteogenomics, the integration of genomics and glycoproteomics to study *de novo* Fab N-glycosylation of M-protein. (A) Schematic overview of the glycoproteogenomics workflow where genomics is integrated with LC-MS/MS based glycoproteomics to study *de novo* Fab N-glycosylation. (B) An example of generated data for a patient, whereby, based on genomic analysis, two *de novo* N-glycosylation sites were predicted (highlighted in green). Subsequent glycoproteomics analysis revealed two distinct glycoprofiles (orange and blue) on the two glycosylation sites. In parallel, the glycoprofile of the Fc was also detected (darkgreen), showing a different glycoprofile compared to the Fab region.

in the germline sequence (*IGHV5-10-1*04*). Fab glycosylation sites were present on both the LC (n=6) and HC (n=4) of the M-protein. For patient 1 and 2, two Fab glycosylation sites were identified in a single chain, while for patient 6 a glycosylation site in both the LC and HC was present. In Table 1 each patient- and myeloma specific glycopeptide (clonotypic glycopeptide) is shown.

For the 7 out of 10 of the Fab glycosylation sites we obtained glycan signals for all matching clonotypic glycopeptides, indicating that these sites were 100 % occupied by glycans or near 100 % due to peptide specific sensitivities (Figure 3A). For two sites, we observed partial glycan occupancy (51 % for the Fab site Kappa N128 from patient 1 and 94 % for Kappa N30 from patient 6), while the IgG N59 Fab site from patient 2, present in the germline sequence, was completely unoccupied.

The amino acid sequence context of the N-glycosylation site can influence the glycan occupancy [26]. Classically, a proline between the asparagine and serine/threonine residue prevents N-glycosylation. Similarly, a proline residue

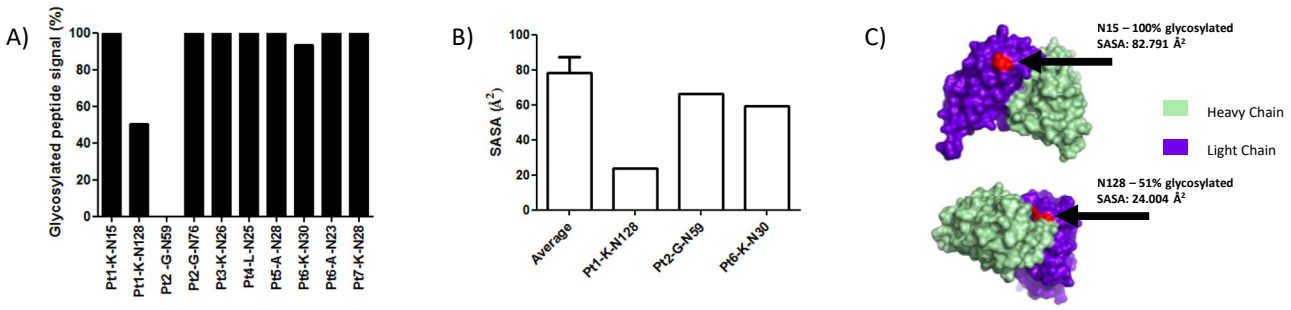

Figure 3: Occupancy of *de novo* Fab N-glycosylation sites on the Fab region of M-proteins. (A) The relative signal of the measured glycosylated peptide signal per Fab N-glycosylation reveals three sites with a less than 100 % glycan occupancy. Note that Pt2_G-N59 is an N-glycosylation site encoded in the germline. (B) The surface accessibility (SASA) was calculated based on 3D modeling of the Fab regions. For the three sites with a less than 100 % occupancy the SASA is depicted, as well as the average of the SASA of the fully occupied sites (n=7). Only for site N128 from patient 1 a decrease in SASA was observed. (C) 3D structures of the glycosylation sites on the M-protein of patient 1 unravels the buried position of N128. In green the heavy chain, and in purple the light chain is depicted.

succeeding the serine or threonine residue (NxS/TP) can also hamper N-glycosylation [26]. The succeeding proline in the glycosites of patient 1 (NTSP) and the pre- and succeeding proline in patient 2 (PNYSP) may thus explain the (partial) N-glycosylation observed in these M-proteins (Table 1). The loss of glycosylation observed at the Fab glycosylation site in patient 6 cannot be explained by the presence of a proline in the sequence. Surface accessibility of the asparagine residue for N-glycosylation enzymes might be of importance. To study this, we determined the surface accessibility (SASA) for the corresponding asparagine residues within the M-proteins. 3D modeling of the Fab region of each M-protein showed that the SASA of 24.004 Å^2 in Pt1-K-N128 (51 % glycosylated clonotypic glycopeptide signal) was significantly lower compared to the average SASAs present on the other N-glycosylation sites (Figure 3B). In the 3D model it is apparent that the asparagine is hidden in between the LC and HC (Figure 3C). The low SASA indicates a decreased accessibility for protein glycosylation enzymes and could explain the low site occupancy observed on Pt1-K-N128.

Together, these data highlight that the genomics and glycoproteomics ought to be combined in order to decipher *de novo* N-glycosylation status on M-proteins.

Comparison of M-protein Fab and Fc region N-glycosylation

We further characterized the glycoprofiles of M-proteins from patients with MM. The Heavy Chain Fab region and Light Chain Fab region are decorated with fairly similar glycans (Figure 4A, yellow and pink shade). These are mainly complex type, highly sialylated and fully fucosylated glycans with various degrees of bisecting structures, such as H5N5S1F1 and H5N4S2F1. Next, we compared the glycans observed on the Fab region with the ones on the Fc regions. We confirmed that N-glycosylation occurred on the Fc region of both IgA (N459 and N263, Figure 4A, green shades) and IgG (N297, Figure 4A blue shades) M-proteins. Similar to the Fab region, the Fc regions are also mainly decorated with complex type glycans. High mannose glycans were exclusively observed on the Fc region of IgA. Even though the major structural class is similar between Fab and Fc region glycans, within this class a clear distinction between the glycans observed on the two regions can be made. The Fc region glycans are for the majority simpler glycans containing either only fucose (H3N4F1 IgG Fc region) or sialic acid (H5N4S1 on IgA) compared to the fully decorated glycans observed on the Fab of M-proteins.

Next, we grouped and quantified the observed glycans based on structural elements; sialylation, fucosylation and bisecting GlcNaC containing structures. The level of sialylated glycans was comparable between the Fab and IgA Fc region (Figure 4B). The sialylation of glycans in the IgG constant region was significantly lower ($p<0.0001$). The fucosylation levels were comparable between Fab and IgG Fc regions, but were significantly lower for Fc IgA, with an almost complete absence of fucosylation on site N263 (Figure 4C). Finally, a striking difference was observed regarding the abundancy of bisecting structures in the Fab glycans compared to an almost complete absence of bisecting Fc glycans (Figure 4D). These data indicate site-specific glycoform structures with significant differences between M-protein Fc and Fab N-glycosylation.

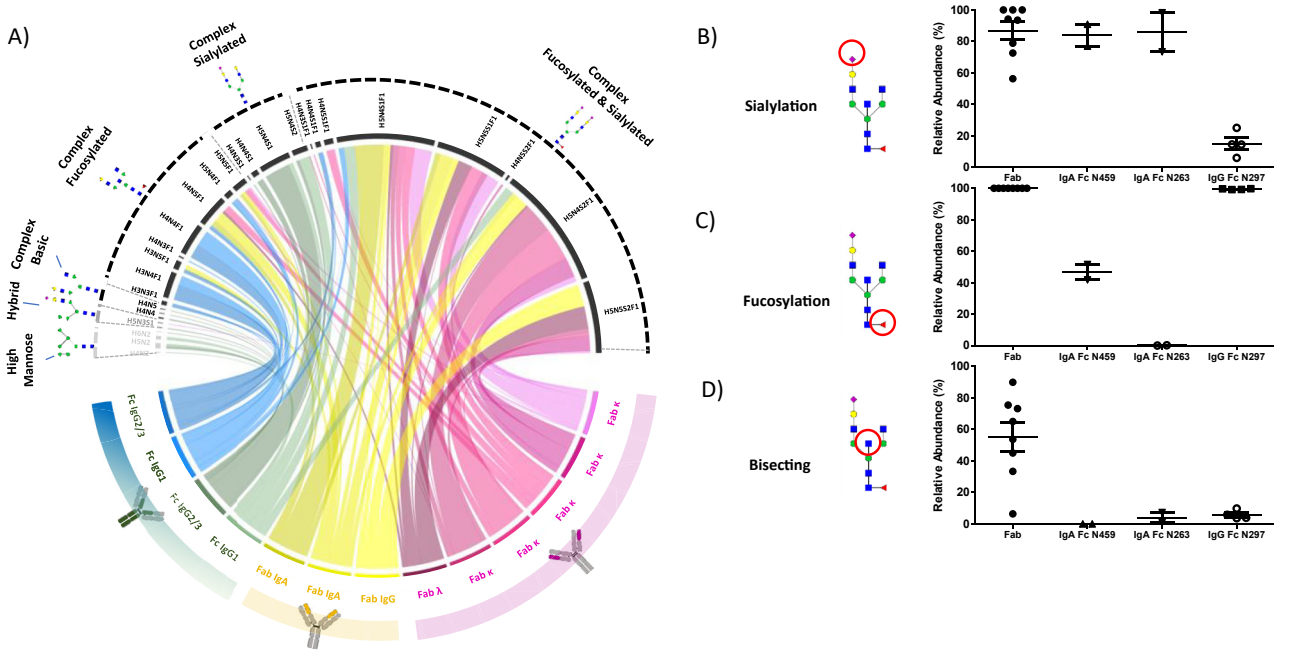

Figure 4: Glycan types and structures differ between Fab and Fc N-glycosylation. (A) Cord plot connecting the N-glycosylation sites on the bottom with observed N-glycans on the top (>1 % abundance) reveals unique profiles that distinguish Fc glycans and Fab glycans. The cord size represents relative abundances and the colors are grouped based on the chain of the N-glycosylation site. (B) The sialylation levels of Fab glycans were higher compared to IgG Fc glycans, while (C) fucosylation levels were lower in IgA Fc region glycans. (D) Bisecting structures were more abundant in Fab region glycans compared to Fc regions.

Discussion

Malignant plasma cells in MM produce and secrete a characteristic monoclonal immunoglobulin (M-protein) that contains a unique Fab region. The uniqueness of the Fab region arises due to V(D)J rearrangements and somatic hypermutations. This can lead to the introduction of myeloma- and patient-specific *de novo* N-glycosylation sites. Recently, evidence for the prognostic value of Fab N-glycosylation of the M-protein in certain monoclonal gammopathies was provided [16, 17]. However, methods to characterize these N-glycosylation sites thus far lacked specificity to determine the occupancy and glycoforms of both the HC and LC in a site-specific manner [13, 14, 16]. Here, we present a glycoproteogenomics approach that integrates genomics with glycoproteomics, which we applied to characterize *de novo* Fab N-glycosylation on the M-proteins of patients with MM in unprecedented detail. Acquiring in-depth information on the exact M-protein glycoprofiles might provide a biomarker to predict disease progression or contribute to our understanding of M-protein pathogenicity.

The strength of the combination of –omics technologies became apparent when we applied the glycoproteogenomics approach to a cohort of 41 patients with MM. Ten Fab N-glycosylation sites on the M-protein of seven patients were predicted based on genomic data. Subsequent glycoproteomic analysis confirmed glycan occupancy in nine of these sites, for the first time we were able to characterize multiple N-glycosylation sites on the same Fab and study their glycoform profiles separately. Two Fab glycosylation sites were partially occupied (51 and 94 % for patient 1 N128 and patient 6 N30, respectively). One site, already present in germline, appeared to be fully unoccupied. Based on the amino acid sequence information of the Fab regions, we could not predict the degree of glycan-occupancy of a specific site. The fully unoccupied site was the only Fab N-glycosylation site present in the germline sequence (*IGHV5-10-1*). Relatively few immunoglobulin Fab alleles contain such a conserved N-glycosylation site (*IGHV1-8, IGHV4-34, IGHV5-10-1, IGLV3-12,* and *IGLV5-37*) [30]; due to somatic hypermutations a proline residue is present in the N-glycosylation sequon preceding the asparagine residue. We hypothesize that during antigen binding affinity selection there was an advantage for this non-glycosylated version of the B cell receptor.

The glycans observed on the Fabs of the M-proteins of patients with MM were almost all of the complex type with a high degree of sialylation, fucosylation and bisecting GlcNAc containing structures. Heterogeneity of glycans structures were observed between patients, chains and between the

same Fab chain of one M-protein. The M-protein Fab glycan structures were distinguishable from the glycans on the Fc region of the M-protein in terms of sialylation (low degree of sialylation in Fc IgG), fucosylation (low degree of fucosylation in Fc IgA), and both the Fc IgG and Fc IgA had a lower degree of bisecting glycans. It is hypothesized that the limited Fc sialylation in IgG antibodies is due to the partially shielded inner facing position of the glycans [12].

There was a more than two-fold increase in the occurrence of Fab N-glycosylation sites in the M-protein LCs of patients with MM (11.8 %) compared to polyclonal antibodies in healthy individuals (4.8 %), indicating an association between MM and an increase in M-protein LC Fab N-glycosylation. Polyclonal antibodies in healthy individuals were determined in IgG^+ B-cells and excludes sequences coming from B-cells that did not undergo the process of affinity maturation. By selecting these mature IgG^+ B-cells we ensured that the M-protein sequences derived from plasma cell RNA-seq datasets from the CoMMpass database are not compared to sequences of naive B-cells that are still in their germline configuration.

In this study we observed a similar level of Fab N-glycosylation sites in the M-protein HCs of patients with MM (11.5 %) compared to polyclonal antibodies in healthy individuals (13.7 %), as in agreement with earlier studies [31]. To further increase the number full-length V(D)J sequences from RNA-seq datasets, future studies could employ a recently improved version of the MiXCR bioinformatics tool [32]. For M-proteins from patients with monoclonal gammopathy of undetermined significance, a benign monoclonal gammopathy, it was shown that the LC Fab N-glycosylation occupancy is 6 % [17]. This is similar to the occupancy we observed in healthy controls. In AL-amyloidosis, a LC plasma cell malignancy, the occurrence of monoclonal LC N-glycosylation was also increased (16–36 %) [16]. In AL amyloidosis the Fab LC glycans contribute to the pathogenicity of the M-proteins [33]. Our glycoproteogenomics workflow described here can precisely locate the N-glycosylation site within the Fab of the M-protein. This might help identifying light chain sequences more likely to be pathogenic. M-protein pathogenicity in MM is poorly understood [34, 35]. Based on both mice-models and human data, Westhrin et al. showed that abnormal M-protein Fc glycosylation may promote bone loss in MM and altering IgG glycosylation may be a therapeutic strategy to reduce bone loss [36]. Thus far, no causal relationship between Fab N-glycosylation and M-protein pathogenicity has been described in MM, however it is speculated that Fab N-glycosylation might contribute to M-protein pathogenicity [35]. Recent data strongly suggests that the presence of M-protein Fab N-glycosylation is a risk factor for progression from MGUS to AL-amyloidosis and other plasma cell disorders [17]. Methods to study this have relied on the discrimination between glycosylated and non-glycosylated M-protein LCs. Even though there are ongoing efforts to optimize these techniques to also elucidate glycan classes and profiles, the level of insight glycoproteogenomics offers is unprecedented, with unparalleled site specificity [37]. We hypothesize that the additional information about glycosylation of the Fab region of the HC, the specific glycan profiles and the exact site on the M-protein Fab, obtained by the unprecedented detail of glycoproteogenomics, will aid in unraveling the clinical role of M-protein Fab glycosylation. Further studies in larger, independent cohorts are warranted to study this, not only in MM but in other plasma cell dyscrasias as well.

Proteogenomics, the integration of genomics and proteomics, is an emerging technological development in the multi-omics field and has the potential to improve cancer characterization, diagnosis, and even advance cancer therapy [38–40]. Furthermore, such integration efforts are also being used to sequence antibody repertoires by MS [41]. With glycoproteogenomics, we can now relate genomic alterations to detailed glycosylation phenotypes in a clinical setting. In the current study, glycoproteogenomics is applied to study *de novo* Fab N-glycosylation, but we believe that this can also contribute to other fields. Acquisition of holistic genomic data of patients is becoming more custom in clinical practice. In such genomic data the deletion or acquisition of glycosylation sites through mutations/truncations can be distilled, followed by subsequent glycoproteomics analysis to dissect the glycosylation phenotype and whether this has clinical consequences for the patients. The glycoproteogenomics workflow introduces a novel level of personalized diagnostics and can aid in improving patient care.

Acknowledgments: The authors thank the following organizations for funding this work: the Dutch Cancer Society (Grants #10817 and #14465), Sebia (Lisses, France), ZonMw Medium Investment Grant (40-00506-98-9001), The Netherlands X-omics Initiative (Project 184.034.019), and Health~Holland (Project LSHM21032).

Research ethics: Written informed consent and clinical and genomic data were de-identified in accordance with the Declaration of Helsinki, and approval for this study was provided by our Institutional Review Board (2018-4140).

Informed consent: Informed consent was obtained from all individuals included in this study, or their legal guardians or wards.

Author contributions: PL performed experiments, data analysis and writing of the manuscript. MB/PK performed data analysis and contributed to writing of the manuscript. SNa, CW, MP, SNo, MD, IJ, AG, JG contributed to writing the manuscript. TD provided samples from the global biobank and contributed to writing. DL helped conceptualize the study and contributed to

writing. HW/JJ conceptualized the study, contributed to data analysis and writing of the manuscript. The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: JFMJ and TML hold a patent related to the field of myeloma diagnostics. The funding organizations played no role in the design, preparation, or approval of the manuscript. All other authors state no conflict of interest.

Research funding: This work was financially supported by two grants from the Dutch Cancer Society for JFMJ (#10817 and #14465). This research was supported by ZonMw Medium Investment Grant (40-00506-98-9001) and was part of the Netherlands X-omics Initiative supported by collaboration project "EnFORCE" (LSHM21032) which is co-funded by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships. This work was supported by funding from Sebia (Lisses, France). The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Data availability: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD040718.

References

1. Röllig C, Knop S, Bornhäuser M. Multiple myeloma. Lancet 2015;385: 2197–208.
2. Willrich MAV, Murray DL, Kyle RA. Laboratory testing for monoclonal gammopathies: focus on monoclonal gammopathy of undetermined significance and smoldering multiple myeloma. Clin Biochem 2018;51: 38–47.
3. Langerhorst P, Brinkman AB, VanDuijn MM, Wessels HJCT, Groenen PJTA, Joosten I, et al. Clonotypic features of rearranged immunoglobulin genes yield personalized biomarkers for minimal residual disease monitoring in multiple myeloma. Clin Chem 2021;67: 867–75.
4. Di Noia JM, Neuberger MS. Molecular mechanisms of antibody somatic hypermutation. Ann Rev Biochem 2007;76:1–22.
5. Barnidge DR, Tschumper RC, Theis JD, Snyder MR, Jelinek DF, Katzmann JA, et al. Monitoring M-proteins in patients with multiple myeloma using heavy-chain variable region clonotypic peptides and LC-MS/MS. J Proteome Res 2014;13:1905–10.
6. Langerhorst P, Noori S, Zajec M, De Rijke YB, Gloerich J, van Gool AJ, et al. Multiple myeloma minimal residual disease detection: targeted mass spectrometry in blood vs next-generation sequencing in bone marrow. Clin Chem 2021;67:1689–98.
7. Zajec M, Langerhorst P, VanDuijn MM, Gloerich J, Russcher H, van Gool AJ, et al. Mass spectrometry for identification, monitoring, and minimal residual disease detection of M-proteins. Clin Chem 2020;66: 421–33.
8. Martins CO, Huet S, Yi SS, Ritorto MS, Landgren O, Dogan A, et al. Mass spectrometry–based method targeting Ig variable regions for assessment of minimal residual disease in multiple myeloma. J Mol Diagn 2020;22:901–11.
9. Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. Nat Rev Nephrol 2019;15:346–66.
10. Gudelj I, Lauc G, Pezer M. Immunoglobulin G glycosylation in aging and diseases. Cell Immunol 2018;333:65–79.
11. Stanley P, Taniguchi N, Aebi M. N-glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, editors, et al. Essential of glycobiology. NY: Cold Spring Harbor; 2015:99–111 pp.
12. van de Bovenkamp FS, Hafkenscheid L, Rispens T, Rombouts Y. The emerging importance of IgG fab glycosylation in immunity. J Immunol 2016;196:1435–41.
13. Van De Bovenkamp FS, Derksen NIL, Ooijevaar-de Heer P, Van Schie KA, Kruithof S, Berkowska MA, et al. Adaptive antibody diversification through N-linked glycosylation of the immunoglobulin variable region. Proc Natl Acad Sci U S A 2018;115:1901–6.
14. Bondt A, Rombouts Y, Selman MHJ, Hensbergen PJ, Reiding KR, Hazes JMW, et al. Immunoglobulin G (IgG) fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. Mol Cell Proteomics 2014;13: 3029–39.
15. Koers J, Sciarrillo R, Derksen NIL, Vletter EM, Fillié-Grijpma YE, Raveling-Eelsing E, et al. Differences in IgG autoantibody Fab glycosylation across autoimmune diseases. J Allergy Clin Immunol 2023;151:1646–54.
16. Kourelis T, Murray DL, Dasari S, Kumar S, Barnidge D, Madden B, et al. MASS-FIX may allow identification of patients at risk for light chain amyloidosis before the onset of symptoms. Am J Hematol 2018;93: E368–70.
17. Dispenzieri A, Larson DR, Rajkumar SV, Kyle RA, Kumar SK, Kourelis T, et al. N-glycosylation of monoclonal light chains on routine MASS-FIX testing is a risk factor for MGUS progression. Leukemia 2020;34: 2749–53.
18. Kinoshita N, Ohno M, Nishiura T, Fujii S, Nishikawa A, Kawakami Y, et al. Glycosylation at the Fab portion of myeloma immunoglobulin G and increased fucosylated biantennary sugar chains: structural analysis by high-performance liquid chromatography and antibody-lectin enzyme immunoassay using lens culinaris agglutinin. Cancer Res 1991;51: 5888–92.
19. Shrivastava A, Joshi S, Guttman A, Rathore AS. N-Glycosylation of monoclonal antibody therapeutics: a comprehensive review on significance and characterization. Anal Chim Acta 2022;1209:339828.
20. Peng W, den Boer MA, Tamara S, Mokiem NJ, van der Lans SPA, Bondt A, et al. Direct mass spectrometry-based detection and antibody sequencing of monoclonal gammopathy of undeterminder significance from patient serum: a case study. J Proteome Res 2023;9: 3022–8.
21. Schulte D, Peng W, Snijder J. Template-based assembly of proteomic short reads for de novo antibody sequencing and repertoire profiling. Anal Chem 2022;94:10391–9.
22. Wessels HJCT, Kulkarni P, van Dael M, Suppers A, Willems E, Zijlstra F, et al. Plasma glycoproteomics delivers high-specificity disease biomarkers by detecting site-specific glycosylation abnormalities. bioRxiv 2022;2022.05.31.494121.
23. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods 2015;12:380–1.
24. Warren RL, Holt RA. Targeted assembly of short sequence reads. PLoS One 2011;6:e19816.

25. Guo Y, Chen K, Kwong PD, Shapiro L, Sheng Z. cAb-Rep: a database of curated antibody repertoires for exploring antibody diversity and predicting antibody prevalence. Front Immunol 2019;10:2365.
26. Gavel Y, Heijne GV. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. Protein Eng 1990;3:433.
27. Attal M, Lauwers-Cances V, Hulin C, Leleu X, Caillot D, Escoffre M, et al. Lenalidomide, bortezomib, and dexamethasone with transplantation for myeloma. N Engl J Med 2017;376:1311–20.
28. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res 2022;50:D543–52.
29. Swindells MB, Porter CT, Couch M, Hurst J, Abhinandan KR, Nielsen JH, et al. abYsis: integrated antibody sequence and structure-management, analysis, and prediction. J Mol Biol 2017;429:356–64.
30. Lefranc M-P, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, et al. IMGT(R), the international ImMunoGeneTics information system(R). Nucleic Acids Res 2009;37:D1006–12.
31. Zhu D, McCarthy H, Ottensmeier CH, Johnson P, Hamblin TJ, Stevenson FK. Acquisition of potential N-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. Blood 2002;99: 2562–8.
32. Nau A, Shen Y, Sanchorawala V, Prokaeva T, Morgan GJ. Complete variable domain sequences of monoclonal antibody light chains identified from untargeted RNA sequencing data. Front Immunol 2023; 14:11627235.
33. Nevone A, Girelli M, Mangiacavalli S, Paiva B, Milani P, Cascino P, et al. An N-glycosylation hotspot in immunoglobulin κ light chains is associated with AL amyloidosis. Leukemia 2022;36:2076–85.
34. Fermand J-P, Bridoux F, Dispenzieri A, Jaccard A, Kyle RA, Leung N, et al. Monoclonal gammopathy of clinical significance: a novel concept with therapeutic implications. Blood 2018;132:1478–85.
35. Sirac C, Herrera GA, Sanders PW, Batuman V, Bender S, Ayala MV, et al. Animal models of monoclonal immunoglobulin-related renal diseases. Nat Rev Nephrol 2018;14:246–64.
36. Westhrin M, Kovcic V, Zhang Z, Moen SH, Vikene Nedal TM, Bondt A, et al. Monoclonal immunoglobulins promote bone loss in multiple myeloma. Blood 2020;136:2656–66.
37. Miller ID, Kohlhagen MC, Ladwig PM, Dasari S, Kumar S, Dispenzieri A, et al. Characterizing M-protein light chain glycosylation via mass spectrometry. Clin Biochem 2022. https://doi.org/10.1016/j.clinbiochem.2022.09.004.
38. Ang MY, Low TY, Lee PY, Wan M, Nazarie WF, Guryev V, et al. Proteogenomics: from next-generation sequencing (NGS) and mass spectrometry-based proteomics to precision medicine. Clin Chim Acta 2019;498:38–46.
39. Hyung D, Baek MJ, Lee J, Cho J, Kim HS, Park C, et al. Protein-gene expression nexus: comprehensive characterization of human cancer cell lines with proteogenomic analysis. Comput Struct Biotechnol J 2021;19:4759–69.
40. Rodriguez H, Zenklusen JC, Staudt LM, Doroshow JH, Lowy DR. The next horizon in precision oncology: proteogenomics to inform cancer diagnosis and treatment. Cell 2021;184:1661–70.
41. de Graaf SC, Hoek M, Tamara S, Heck AJR. A perspective toward mass spectrometry-based de novo sequencing of endogenous antibodies. mAbs 2022;14:1–17.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/cclm-2023-1189).