

Proteoform Profile Mapping of the Human Serum Complement Component C9 Revealing Unexpected New Features of *N*-, *O*-, and *C*-Glycosylation

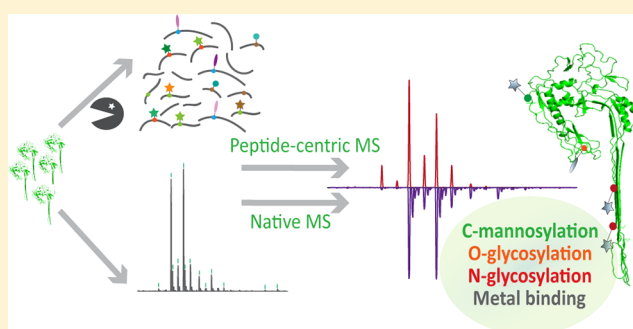
Vojtech Franc, Yang Yang, and Albert J. R. Heck*^{1b}

Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

S Supporting Information

ABSTRACT: The human complement C9 protein (~65 kDa) is a member of the complement pathway. It plays an essential role in the membrane attack complex (MAC), which forms a lethal pore on the cellular surface of pathogenic bacteria. Here, we charted in detail the structural microheterogeneity of C9 purified from human blood serum, using an integrative workflow combining high-resolution native mass spectrometry and (glyco)peptide-centric proteomics. The proteoform profile of C9 was acquired by high-resolution native mass spectrometry, which revealed the co-occurrence of ~50 distinct mass spectrometry (MS) signals. Subsequent peptide-centric analysis, through proteolytic digestion of C9 and liquid chromatography (LC)-tandem mass spectrometry (MS/MS) measurements of the resulting peptide mixtures, provided site-specific quantitative profiles of three different types of C9 glycosylation and validation of the native MS data. Our study provides a detailed specification, validation, and quantification of 15 co-occurring C9 proteoforms and the first direct experimental evidence of *O*-linked glycans in the *N*-terminal region. Additionally, next to the two known glycosylation sites, a third novel, albeit low abundant, *N*-glycosylation site on C9 is identified, which surprisingly does not possess the canonical *N*-glycosylation sequence N-X-S/T. Our data also reveal a binding of up to two Ca²⁺ ions to C9. Mapping all detected and validated sites of modifications on a structural model of C9, as present in the MAC, hints at their putative roles in pore formation or receptor interactions. The applied methods herein represent a powerful tool for the unbiased in-depth analysis of plasma proteins and may advance biomarker discovery, as aberrant glycosylation profiles may be indicative of the pathophysiological state of the patients.



Post-translational modifications (PTMs) of proteins regulate their activity, localization, turnover, interactions, and many other important physiological processes.^{1,2} Of all possible PTMs, protein glycosylation is one of the most abundant yet structurally diverse PTM, making it analytically and biochemically challenging to monitor.³ This is because the enzymes involved in the glycosylation machinery can produce diverse glycosylation patterns on proteins and heterogeneous populations of glycans at every occupied glycosylation site.^{4–6} In addition, these modifications are also present in nonstoichiometric amounts with multiple varieties of chemical moieties involved. Thus, new methods are needed for their detailed analysis. Recent progress in high-resolution native electrospray ionization (ESI) mass spectrometry (MS) can provide a novel means to facilitate the in-depth analysis of all coappearing modifications, at least when they are distinguishable in mass.^{7–9} In combination with peptide-centric proteomics, this approach is very useful for high detail analysis of PTMs on proteins from different biological sources and can provide direct assessment of the biosimilarity among similar therapeutic proteins.¹⁰ Through integrating native MS data and peptide-centric

proteomics, one can obtain information about composition, stoichiometry, site-specificity, and relative abundance of the modifications at each site. Moreover, a direct cross-evaluation of both data sets (native MS and peptide-centric MS) validates completeness of each approach and provides reliability for the quantitative profiling of protein proteoforms. So far, in the analysis of protein proteoforms, most MS methods typically use denaturing conditions prior to the ESI process.^{11–14} These conditions inevitably disrupt protein tertiary and quaternary structures. Although native MS is often regarded as somewhat less sensitive, it provides the advantage that the resulting mass spectra are less congested, as the ion signals are distributed over a substantially fewer number of charge states and over a wider *m/z* window. Collectively, such hybrid mass spectrometry strategies have the potential to become beneficial for the study of biologically important (glyco)proteins, whereby knowledge

Received: November 17, 2016

Accepted: February 21, 2017

Published: February 21, 2017

about their precise modifications is crucial in understanding their activity and function.

Most proteins in human blood plasma are decorated by a plethora of PTMs, particularly involving glycosylation, and the complement component protein C9 is not an exception.¹⁵ Human C9 is primarily produced in the liver and plays a key role in the formation of the membrane attack complex (MAC), together with the other complement proteins C5, C6, C7, and C8. While several cryoEM maps have recently become available for the MAC,^{16,17} no detailed structure is available for its C9 component. Still, amino acid alignments have identified several domains in C9 based on its homology to other proteins. These include the *N*-terminal type 1 thrombospondin (TSP) domain, a low-density lipoprotein receptor class A repeat (LDLRA), a number of potential transmembrane regions, and the *C*-terminal epidermal growth factor (EGF)-like domain (Figure 1).¹⁸ The majority of the detailed characterization studies of the

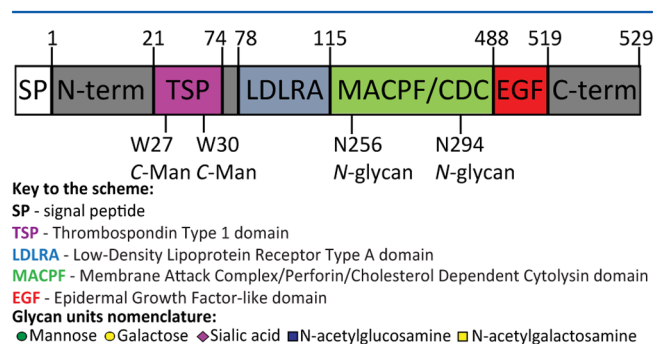


Figure 1. Schematic of domain composition and primary structure of C9. The scheme includes previously reported sites of C-mannosylation²² and N-glycosylation on C9.^{19–21} The glycan nomenclature used is indicated at the bottom.

PTMs occurring on C9 dates back to the previous century, when techniques used to perform such analysis were very cumbersome. In these early studies, C9 was reported to be N-glycosylated;^{19–21} however, no current evidence exists regarding the composition and heterogeneity of these N-linked glycans. Moreover, although suspected, no direct proof has been reported for the presence of O-linked glycans. C9 is additionally modified by a rarer type of glycosylation: C-mannosylation.²² With such a diverse repertoire of modifications, C9 not only presents a challenging analytical target but also imparts a potential variability in its physiological functioning. Exemplary findings to support this statement come from reports wherein the extracellular Ser phosphorylation of C9 by ecto-protein kinases in cancer cells K562 was proposed to serve as a protective mechanism against complement in tumor cells.²³ Moreover, C9 with fucosylated N-glycans has been suggested as a biomarker for squamous cell lung cancer, as patients tend to show overproduction of these proteoforms.²⁴ A detailed map of the full proteoform profile of serum-derived complement component C9 is, therefore, of high interest.

Here, we report an unbiased and in-depth analysis of the complement component C9 protein isolated from pooled human blood serum (of at least three donors) using modern, hybrid MS technologies. Our data provides a detailed view of the modifications co-occurring on C9. We validate all identified PTMs from high quality tandem mass spectrometry (MS/MS) spectra using a peptide-centric approach. In addition to the

earlier reported C9 modifications, our data revealed the attachment of mucin type of O-glycosylation in the *N*-terminal part of C9, providing the first experimental evidence of this modification on C9. Except information about PTMs, our native MS measurements suggest binding of up to two Ca²⁺ ions on C9. Since the *N*-terminal region of C9 seems to play a crucial role in the C9 polymerization process and thus also in the assembly of MAC,²⁵ the role of the here identified O-glycosylation site may act as a target for further functional investigations. Moreover, we identified a novel low abundant N-glycosylation site on N215 that is highly conserved throughout mammals and does not adhere to the typical N-glycosylation sequon N-X-S/T (X can be any amino acid except P).

MATERIALS AND METHODS

Chemicals and Materials. Complement component C9 (Uniprot Code: P02748) purified from pooled human blood plasma (more than three healthy donors) was acquired from Complement Technology, Inc. (Texas, USA). The sample was purified according to a standard protocol²⁶ (the certificate of analysis is attached in the [Supporting Information S5](#)). Dithiothreitol (DTT), iodoacetamide (IAA), and ammonium acetate (AMAC) were purchased from Sigma-Aldrich (Steinheim, Germany). Formic acid (FA) was from Merck (Darmstadt, Germany). Acetonitrile (ACN) was purchased from Biosolve (Valkenswaard, The Netherlands). POROS Oligo R3 50 μm particles were obtained from PerSeptive Biosystems (Framingham, MA, USA) and packed into GELoader pipet tips (Eppendorf, Hamburg, Germany). Sequencing grade trypsin was obtained from Promega (Madison, WI). Glu-C, Asp-N, PNGase F, and Sialidase were obtained from Roche (Indianapolis, USA).

Sample Preparation for Native MS. Unprocessed protein solution in a phosphate buffer at pH 7.2, containing ~30–40 μg of C9, was buffer exchanged into 150 mM aqueous AMAC (pH 7.5) by ultrafiltration (vivasp500, Sartorius Stedim Biotech, Germany) using a 10 kDa cutoff filter. The resulting protein concentration was measured by UV absorbance at 280 nm and adjusted to 2–3 μM prior to native MS analysis. The enzyme Sialidase was used to remove sialic acid residues from C9. PNGase F was used to cleave the N-glycans of C9.²⁶ All samples were buffer exchanged to 150 mM AMAC (pH 7.2) prior to native MS measurements.

Native MS Analysis. Samples were analyzed on a modified Exactive Plus Orbitrap instrument with extended mass range (EMR) (Thermo Fisher Scientific, Bremen) using a standard *m/z* range of 500–10 000, as described in detail previously.²⁷ The voltage offsets on the transport multipoles and ion lenses were manually tuned to achieve optimal transmission of protein ions at elevated *m/z*. Nitrogen was used in the higher-energy collisional dissociation (HCD) cell at a gas pressure of 6–8 × 10^{−10} bar. MS parameters used: spray voltage, 1.2–1.3 V; source fragmentation, 30 V; source temperature, 250 °C; collision energy, 30 V; resolution (at *m/z* 200), 30 000. The instrument was mass calibrated as described previously, using a solution of CsI.²⁷

Native MS Data Analysis. The accurate masses of the observed C9 proteoforms were calculated manually averaging over all detected charge states of C9. For PTM composition analysis, data were processed manually and glycan structures were deduced on the basis of known biosynthetic pathways. Average masses were used for the PTM assignments, including

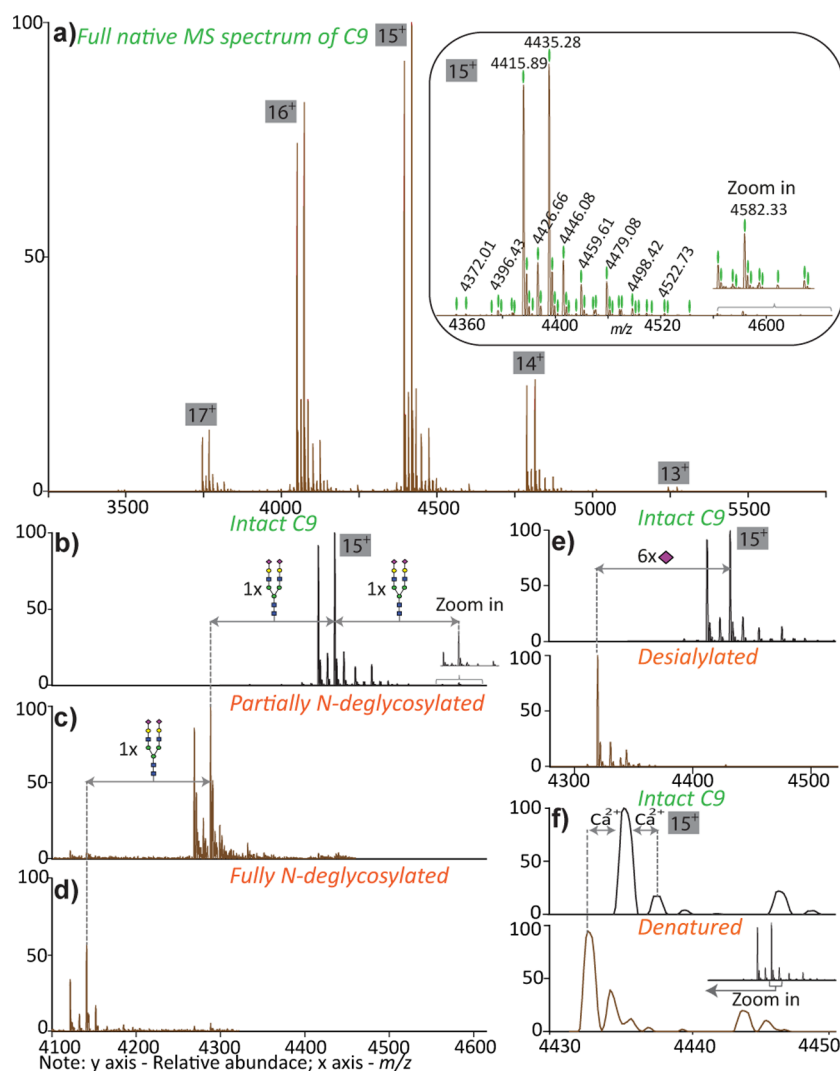


Figure 2. Full native ESI-MS spectrum of the intact C9 sprayed from aqueous ammonium acetate (a). The charge states are indicated. Zoom in on the 15^+ charged state in the inset reveals approximately 50 distinct ion signals. (b) Zoom in on the 15^+ charged state, centered around m/z 4500, in the native mass spectra of intact C9. In (c), alike spectrum of C9 treated with PNGase F for 4 h reveals partial N-deglycosylation. In (d), C9 was treated with PNGase F for 48 h. In (e), C9 was first enzymatically desialylated, and in (f), C9 was partially denatured, prior to native MS analysis. The differences in mass between C9 proteoforms in the unprocessed and treated samples allows the deduction of the PTM composition of these most abundant C9 proteoforms. The mass of the most abundant ion in the unprocessed sample, at m/z of 4435.28, is 66 516.20 Da, from which a calculated composition of PTMs can be derived, Hex₁₂HexNAc₉NeuAc₆ unambiguously. The 15 positive charges come from 13 H⁺ and 1 Ca²⁺ ion, due to the presence of a bound Ca²⁺ ion, as revealed in (f).

hexose/mannose/galactose (Hex/Man/Gal, 162.1424 Da), N-acetylhexosamine/N-acetylglucosamine (HexNAc/GlcNAc/GalNAc, 203.1950 Da), and N-acetylneuraminic acid (NeuAc, 291.2579 Da). All used symbols and text nomenclature are according to recommendations of the Consortium for Functional Glycomics.

In-Solution Digestion for Peptide-Centric Glycoproteomics. Intact human C9 protein in PBS buffer (10 mM sodium phosphate, 145 mM NaCl, pH 7.3) at a concentration of 1 mg/mL was reduced with 5 mM DTT at 56 °C for 30 min and alkylated with 15 mM IAA at room temperature for 30 min in the dark. The excess of IAA was quenched by using 5 mM DTT. C9 was digested overnight with trypsin at an enzyme-to-protein-ratio of 1:100 (w/w) at 37 °C. Another C9 sample was digested for 4 h by using Asp-N at an enzyme-to-protein ratio of 1:75 (w/w) at 37 °C, and the resulted peptide mixtures were further treated with trypsin (1:100; w/w) overnight at 37 °C. All proteolytic digests containing modified glycopeptides were

desalted by GELoader tips filled with POROS Oligo R3 50 μ m particles,²⁸ dried, and dissolved in 40 μ L of 0.1% FA prior to liquid chromatography (LC)-MS and MS/MS analysis.

LC-MS and MS/MS Analysis. All peptides (typically 300 fmol of C9 peptides) were separated and analyzed using an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) coupled online to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Reversed-phase separation was accomplished using a 100 μ m inner diameter 2 cm trap column (in-house packed with ReproSil-Pur C18-AQ, 3 μ m) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) coupled to a 50 μ m inner diameter 50 cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 μ m) (Agilent Technologies, Amstelveen, The Netherlands). Mobile-phase solvent A consisted of 0.1% FA in water, and mobile-phase solvent B consisted of 0.1% FA in ACN. The flow rate was set to 300 nL/min. A 45 min gradient was used as follows: 0–10 min, 100%

solvent A; 10.1–35 min, 10% solvent B; 35–38 min, 45% solvent B; 38–40 min, 100% solvent B; 40–45 min, 100% solvent A. Nanospray was achieved using a coated fused silica emitter (New Objective, Cambridge, MA) (outer diameter, 360 μm ; inner diameter, 20 μm ; tip inner diameter, 10 μm) biased to 2 kV. The mass spectrometer was operated in positive ion mode, and the spectra were acquired in the data dependent acquisition mode. For the MS scans, the mass range was set from 300 to 2000 m/z at a resolution of 60 000 and the AGC target was set to 4×10^5 . For the MS/MS measurements, HCD and electron-transfer and higher-energy collision dissociation (EThcD) were used. HCD was performed with normalized collision energy of 15% and 35%, respectively. A supplementary activation energy of 20% was used for EThcD. For the MS/MS scans, the mass range was set from 100 to 2000 m/z , and the resolution was set to 30 000; the AGC target was set to 5×10^5 . The precursor isolation width was 1.6 Da, and the maximum injection time was set to 300 ms.

LC/MS and MS/MS Data Analysis. Raw data were interpreted by using the Byonic software suite (Protein Metrics Inc.),²⁹ and further validation of the key MS/MS spectra was performed manually. The following parameters were used for data searches: precursor ion mass tolerance, 10 ppm; product ion mass tolerance, 20 ppm; fixed modification, Cys carbamidomethyl; variable modification: Met oxidation, Trp mannosylation, and both *N*- and *O*-glycosylation from mammalian glycan databases. A nonenzyme specificity search was chosen for all samples. The database used contained the C9 protein amino acid sequence (Uniprot Code: P02748). Profiling and relative quantification of PTM modified C9 peptides were achieved by use of the extracted ion chromatograms (XICs) from two independently processed C9 samples. The peptide mixtures were prepared with different combinations of proteolytic enzymes as described above ((1) Trypsin; (2) AspN + Trypsin). For peak area calculations, the first three isotopes were taken from each manually validated peptide proteoform. Integrated peak areas were normalized for all PTM sites individually, and the average peptide ratios from the two samples were taken as a final estimation of the abundance. The XICs were obtained using the software Skyline.³⁰ The glycan structures of each glycoform were manually annotated. Hereby reported glycan structures are depicted without the linkage type of glycan units, since the acquired MS/MS patterns do not provide such information.

Combining Native MS and Peptide-Centric Proteomic Data. Reliability and completeness of the obtained proteoform profiles of C9 were assessed by an integrative approach combining the native MS data with the glycopeptide centric proteomics data. Details of this approach have been described in detail previously.¹⁰ Briefly, *in silico* data construction of the “intact protein spectra” was performed on the basis of the masses and relative abundances of all site-specific PTMs derived from the glycopeptide centric analysis. Subsequently, the constructed spectrum was compared to the experimental native MS spectra of C9. The similarity between the two independent data sets (Native MS spectra and constructed spectra based on glycopeptide centric data) was expressed by a Pearson correlation factor. All R scripts used for the spectra simulation are available at github (<https://github.com/Yang0014/glycoNativeMS>). All C9 proteoforms predicted from the peptide-centric data were further filtered by taking 0.5% cutoff in relative intensity of the peaks in the experimental native spectrum, and mass deviations were manually checked.

RESULTS

Native MS Analysis Provides Hints about Novel Unexpected PTMs and Ca^{2+} Binding to C9. We started our investigation by first acquiring high-resolution native ESI-MS spectra of the human complement component C9 (Figure 2a). The recorded native MS spectrum of C9 shows at least five different charge states, ranging from $[\text{M} + 13\text{H}]^{13+}$ to $[\text{M} + 17\text{H}]^{17+}$. Each charge state contains various ion series that correspond to different masses and thus different proteoforms of C9. On the basis of their distinguishable masses, taking a 1% cutoff in relative intensity of the peaks, we can distinguish at least ~ 50 co-occurring MS signals. Since we suspected these could correspond to different proteoforms of C9, we set out to further examine and validate our findings.

To simplify the visualization of the C9 proteoform profile, we focused on the most intense charge state (15^+). The average mass of the protein backbone of C9 is 60 954.02 Da. In this mass calculation, we used the mass of the C9 backbone sequence lacking the *N*-terminal signal peptide, corrected by the mass shift induced by the 12 disulfide bonds present in C9 (-24×1.0079 Da). Compared to, for instance, chicken ovalbumin³¹ and CHO derived erythropoietin,¹⁰ which were previously analyzed by high-resolution native mass spectrometry, the native mass spectra of C9 are remarkably less heterogeneous, especially since, according to previously published data, C9 has been shown to be *C*-mannosylated²² at the TPS domain and *N*-glycosylated^{19–21} at two sites in the MACPF domain.

Looking at the inset in Figure 2a, the mass difference of 656 Da between the abundant peaks with m/z of 4435.28 and 4479.02 corresponds to the glycan composition HexNAc₁-Hex₁NeuAc₁. The same mass difference can be observed between the abundant peaks with m/z of 4415.89 and 4459.61. This may correspond to either variability in the number of antennae on the *N*-glycans or the additional attachment of mucin type *O*-glycans. To address this, we next treated the protein with various enzymes that cleave off parts of the glycan moieties. The removal of *N*-glycans or sialic acid residues resulted in specific mass shifts, allowing us to calculate and partly predict the PTM composition of C9. For cleavage of *N*-glycosylations, we used PNGaseF (Figure 2b–d) and sialidase for the specific removal of sialic acids (Figure 2e). We subsequently subjected these treated C9 samples to native MS analysis. The incubation of C9 with PNGase F for 4 h at room temperature resulted in a removal of one of the two *N*-glycan chains (Figure 2c). The mass difference of 2206 Da between the most abundant intact C9 proteoform and the *N*-deglycosylated C9 indicated the attachment of a *N*-glycan with the composition of HexNAc₄Hex₅NeuAc₂. A prolonged treatment (48 h; 37 °C) of C9 with PNGase F resulted in a second major mass shift of 2206 Da corresponding to a loss of the second *N*-glycan (Figure 2d). A closer look at the native MS spectrum of C9 exposed another lower abundant ion series in the higher mass range region suggesting the existence of an unexpected third *N*-glycosylation site. The observed mass difference of 2206 Da in the intact C9 between the peaks with m/z of 4435.28 and 4582.33 strongly indicates the presence of a third *N*-glycan with a similar composition as the other two known sites (Figure 2b–d).

The heterogeneity in the proteoform profile of the fully *N*-deglycosylated C9 remained very similar (Figure 2b,d). Therefore, the observed mass differences of 656 Da most

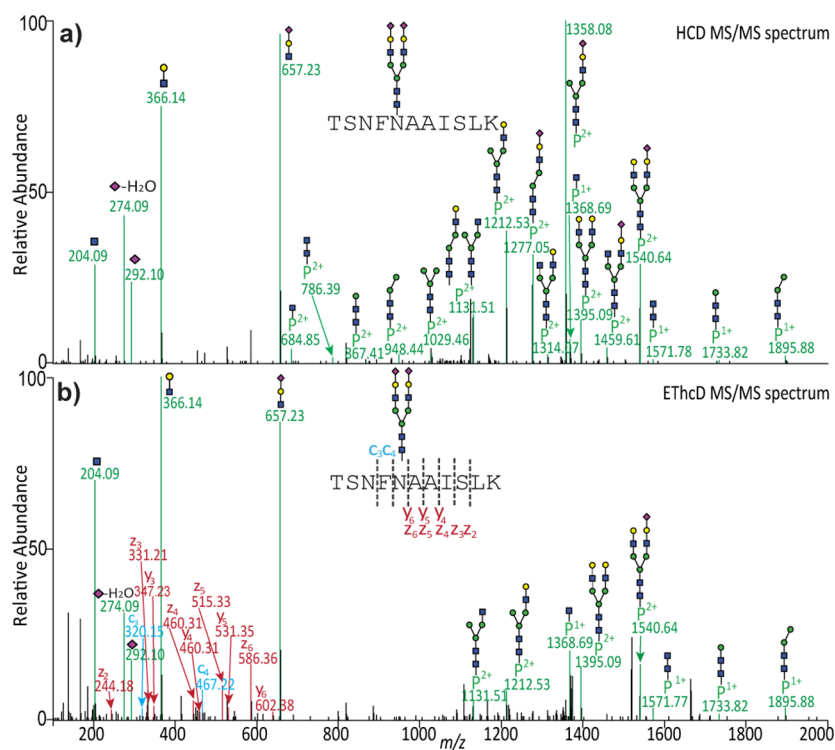


Figure 3. Low energy HCD MS/MS (a) and EThcD MS/MS (b) spectra of the peptide harboring the novel, noncanonical *N*-glycosylation site at N215, derived by tryptic digestion of C9. Both spectra were acquired for the same precursor with m/z of 1124.13 Da. Sequential fragmentation of the *N*-glycan moiety in the spectrum (a) allowed deduction of its glycan composition while the EThcD spectrum (b) provided confirmation of the peptide sequence and position of the *N*-glycan. “P” = peptide backbone of the glycopeptide.

likely correspond to *O*-linked glycan chains with the overall composition of HexNAc₁Hex₁NeuAc₁. Additionally, a mass difference of 947 Da between the peaks with m/z of 4435.28 and 4498.41 and between 4415.89 and 4459.61 suggests the occurrence of disialylated mucin *O*-glycans with the overall composition of HexNAc₁Hex₁NeuAc₂. For the further assessment of the PTM composition of C9, we treated C9 with sialidase. The removal of sialic acid residues resulted in a pronounced simplification of the structural heterogeneity in C9 (Figure 2e). The proteoform occurring at 4435.284 m/z , which is known to contain only two occupied *N*-glycosylation sites, shifted to 4318.80 m/z , corresponding to an overall loss of six sialic acid moieties. Four out of six of these sialic acids were cleaved off from the two *N*-glycans. The remaining two sialic acids are thus most likely attached to *O*-glycans. This assumption is based on the mass difference of 1147.72 Da between the overall mass of C9 PTMs (66 513.76–60 954.02 Da = 5559.74 Da) and the total mass of *N*-glycans in the most abundant C9 proteoform represented by the peak with m/z of 4435.28 (2×2206.01 Da = 4412.02 Da). Since the proteoform occurring at 4435.28 m/z contains one *O*-glycan, further sequential mass differences (656 and 947 Da respectively) among the peaks with m/z of 4435.28, 4498.41, and 4542.22 suggest at least three *O*-glycans are attached to C9.

As mentioned above, the expected total mass of PTMs for the proteoform present at 4435.28 m/z is 5559.74 Da. This estimation is based on the assumption that all charges on the C9 in the native MS spectrum come from H⁺. However, the total mass of experimentally proven PTM composition (Hex₁₂HexNAc₉NeuAc₆) is 5522.01 Da. The remaining mass of 37.73 Da (5559.74–5522.01 Da) can be explained by the presence of a Ca²⁺ ion (40.076 Da) in the structure of C9. This

means that the 15 charges present on the C9 at 4435.28 m/z come from 13 H⁺ and one from Ca²⁺. In that case, the deprotonated mass of the most abundant proteoform with all PTMs and a Ca²⁺ ion is 66 516.20, which corresponds to a small standard deviation of ± 0.05 Da with respect to the calculated C9 mass: 66 516.11 Da. This hypothesis was supported by the recording native MS profile of the partly denatured C9 (Figure 2f). Denaturation of C9 through the addition of 1% formic acid resulted in a release of Ca²⁺ and corresponding mass shifts of all ions in the spectrum. This indicates that C9 can bind two Ca²⁺ ions, whereby the most abundant proteoforms contain just one bound Ca²⁺ ion.

Analysis of *N*-Glycosylation Revealed a Noncanonical *N*-Glycosylation Site on Asn215. C9 is a glycoprotein containing two known canonical *N*-glycosylation sites (N256 and N394). Our native MS measurements suggest the presence of a novel third low abundant *N*-glycosylation site. To validate this finding, C9 was digested using different proteolytic enzymes and the resulting peptide mixtures were analyzed by peptide centric LC-MS/MS. Data interpretation provided information about the site location, glycan type, composition, and abundance of all three *N*-glycans. Low energy HCD MS/MS spectra of the tryptic peptides with amino acid sequences FSYSKNETYQLFLSYSSK and AVNITSENLI DDVVSILR clearly revealed the composition of the two known *N*-glycans on N256 and N394 (Figure S1a,b). Improved sequence coverage obtained by EThcD MS/MS further confirmed the amino acid sequence of these *N*-glycopeptides. In addition, we also identified the third *N*-glycosylation site at N215, using the tryptic peptide TSNFNAAISLK (N215). Low energy HCD (Figure 3a) and EThcD MS/MS (Figure 3b) spectra of this glycopeptide unambiguously confirmed its amino acid

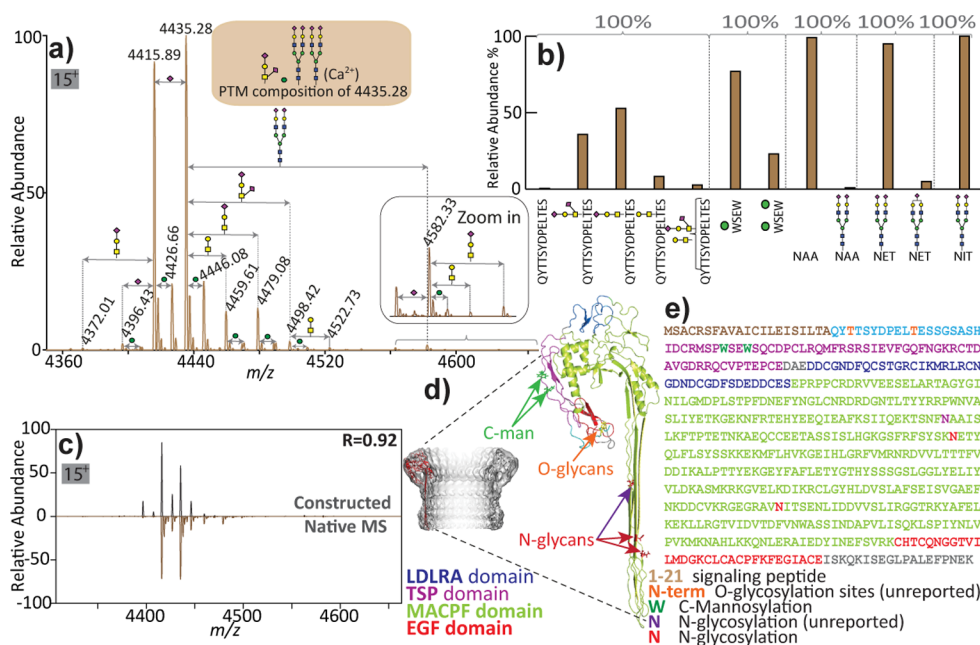


Figure 4. Summary of assignments of PTMs on C9. (a) Native MS spectrum of C9, zoomed in on charge state $[M + 13H + Ca^{2+}]^{15+}$. The overall PTM composition of the most abundant proteoform with m/z of 4435.28 was deduced as described in Figure 2. Mass differences among the peaks correspond to various glycan units or glycans. (b) Relative abundances of peptide proteoforms were estimated from their corresponding peptide ion currents (XICs). Each PTM modified peptide was normalized individually so that the sum of all proteoforms was set to 100%. For clarity, only parts of the peptide sequence carrying PTMs are shown below the graph. (c) A comparison of the intact C9 native MS spectrum with the *in silico* constructed spectrum based on the peptide-centric proteomics data. The correlation is very high ($R = 0.92$). (d) Structural model of poly-C9¹⁶ and mono-C9, whereby the sites of the modifications are indicated. The poly-C9 model was chosen as template for the mono-C9 using I-Tasser,⁴⁴ and the model was processed by means of the PyMOL Molecular Graphic System, Version 1.8 Schrödinger, LLC. (e) Overview of the C9 sequence with all identified PTM sites, wherein the newly discovered O-glycosylation sites at the N-terminus and the novel N-glycosylation site at residue 236 are highlighted in orange and purple, respectively.

sequence, localization, and the composition of the N-glycan. Markedly, this newly determined N-glycosylation site does not adhere to the canonical N-X-S/T motif.

Analysis of O-Glycosylation Confirmed the Presence up to Three O-Glycans. Peptide-centric LC MS/MS analysis also provided direct evidence of the existence of three O-glycosylation sites on C9. Figure S2 displays annotated EThcD MS/MS spectra of the C9 N-terminal peptide with amino acid sequence QYTTSYDPELIESSGSASHIDCR derived from proteolytic digestion of the C9 with AspN and trypsin. This peptide contains an N-terminal Q and exhibits a significant level of cyclization to pyroglutamic acid. Such N-terminal modifications are frequently reported in LC-MS/MS analysis.³² Manual inspection of the MS/MS spectra exposed the structural composition of the attached O-glycans. These were mucin type O-glycans with 0, 1, or 2 sialic acids connected to the core structure HexNAc₁Hex₁. The peptide was found to be modified with up to two O-glycans; however, their precise location could not be determined, as the fragmentation spectra lack sufficient signature ions. A peptide bearing the third O-glycosylation site, which is present in the native MS spectrum, was not detected in LC MS/MS, presumably due to its low abundance.

C9 TSP Domain Harbors Two C-Mannosylation Sites. C9 contains a thrombospondin (TSP) domain, which may be C-mannosylated.²² Looking back on the native MS profile of the intact C9, the most abundant peaks with m/z of 4415.89 and 4435.28 are in a vicinity with their lower abundant forms (4426.65 m/z and 4446.08 m/z) differing in mass by 162 Da. The same mass difference can also be observed in the N-

deglycosylated and the desialylated C9, which corresponds to the presence of a single Hex. Peptide-centric LC MS/MS detected peptides originating from the MSPWSEWSQC-DPCLR sequence. This peptide contains the well-known sequence motif WXXW, which is frequently mannosylated in proteins with TSP repeats.³³ Our EThcD MS/MS spectra unequivocally confirmed that in C9 W27 is fully occupied by C-mannosylation, whereas W30 is only partially occupied (Figure S3a,b).

Defining the Overall Structural Heterogeneity, Validation, and Quantification of C9 Proteoforms. Next, we also assessed the relative abundance of the different proteoforms of C9. Figure 4a shows the overall view on all detected C9 proteoforms in the annotated native spectrum. The site-specific characterization of C9 is based on peptide-centric proteomics data (Figure 4b, Table S1). Two N-glycosylation sites (N256 and N394) are fully occupied by complex biantennary N-glycans with sialylated antennas, each with a low degree of structural heterogeneity (only ~4% of the N-glycans at N256 have one of the antennas nonsialylated). The third novel noncanonical N-glycosylation site at N215 is occupied only in ~1% of C9, carrying the same type of glycan as the other two N-glycosylation sites. In contrast to the N-glycosylation, the O-glycosylation exhibits higher variability in its degree of sialylation. EThcD MS/MS spectra conclusively confirmed the composition of the presented O-glycan chains. The N-terminal peptides that contain an amino acid sequence of QYTTSYDPELIESSGSASHIDCR possess mucin type O-glycans with a different level of sialylation. In 36% of C9, the peptide is occupied by an O-glycan with a sialic acid attached to

both glycan units, HexNAc and Hex. The most abundant form (52%) represents *O*-glycan chains with one sialic acid attached to Hex. The lowest abundant form (8%) contains a nonsialylated *O*-glycan core. Unfortunately, our EThcD MS/MS spectra did not reveal the exact location of the *O*-glycosylated sites; however, the observed dual series of *c*, *z*, *y*, and *b* ions indirectly suggest that T11 is likely occupied (Figure S2b–d). The *O*-glycopeptide with two attached *O*-glycans is presented in less than 3% of relative abundance. The *N*-terminal peptide was also found to be unmodified (Figure S2a); however, its abundance is almost negligible (0.5%). Moreover, *C*-mannosylation contributes to the C9 heterogeneity with a variable occupancy of *W* residues by Man in the sequence motif WXXW. Tryptophan 27 is 100% occupied by Man, while W30 is only partially modified (23%).

A comparison of the intact C9 native MS spectrum with *in silico* constructed MS spectrum (Figure 4c) reveals a high degree of consistency between the native MS and peptide-centric MS approach ($R = 0.92$). Using this method, 15 C9 proteoforms could be validated (Table S2). The unmatched peaks mostly correspond to C9 proteoforms containing two Ca^{2+} ions and species that were not detected during the peptide centric analysis. For example, the peak at m/z 4479.08 most likely corresponds to the C9 proteoform containing two *O*-glycans with one and two sialic acids on the core HexNAc₁Hex₁. Since a peptide with this PTMs composition was not detected by LC-MS/MS, the reconstructed native-like spectrum does not contain this proteoform. Several non-annotated low abundant ions signals correspond to Na^+ and/or K^+ adducts, which are very frequent artifacts in native MS. The constructed spectrum also contains systematic artifacts, caused by the fact that labile PTMs are easily lost during the peptide-centric LC MS/MS analysis.³⁴ In our case, the most prevalent artifacts are glycopeptides that lose partly their sialic acids and *O*-linked glycan chains. Other MS signals with relative abundance below 0.5% were eventually filtered out during the validation process.

This cross-correlation of the data pinpoints the advantages of uniting native MS and peptide-centric LC-MS/MS, as well as the weaknesses associated with the use of only a peptide centric approach. Without the information obtained from the native MS measurements, the undesired peptide artifacts and modifications that are not detected or overlooked at the peptide-centric level would not have been discovered. Still, the consistency between both data sets is very high ($R = 0.92$), verifying that we did cover and validate most of detected signal of the C9 proteoforms and their PTMs in a comprehensive manner.

DISCUSSION

All previous reports on C9 PTMs were largely focused on only one type of PTM. From these studies, it has been known that C9 is *N*-glycosylated^{19–21} and *C*-mannosylated²² with 8% of the total mass of C9 originating from these attached glycans.³⁵ The majority of these glycan chains are distributed between the two previously known *N*-glycosylation sites (N256 and N394); however, their exact monosaccharide composition had so far remained elusive. A study suggested that the *N*-linked glycans of C9 are of the tri- or tetra-antennary complex type since neither endo F nor endo H released the glycans.²¹ Here, we unambiguously determined the composition of these *N*-glycans using low energy HCD. In agreement with previous observations, the *N*-deglycosylation reaction proceeded sequen-

tially and in a time dependent manner.²¹ Incubation of C9 with PNGase F for 4 h at 37 °C resulted in the cleavage of a single *N*-linked glycan. The second *N*-glycan (N256) was released only after prolonged incubation (48 h). Surprisingly, the native MS profile of *N*-deglycosylated C9 was shifted in mass but did not substantially change in comparison with the native MS profile of the intact C9, revealing the homogeneous *N*-glycosylation patterns on C9.

As C9 contains only two canonical *N*-glycosylation sequons, the observation of a third *N*-glycosylation, located at N215 (NAA), was not expected. The *N*-glycosylation sequon N-X-S/T is well established. However, a few reports have shown *N*-glycosylation in the sequence N-X-C^{36–38} or N-X-V.^{39,40} In this regard, a high-resolution structural model of the bacterial oligosaccharyltransferase (OTase) provided evidence that an amino acid in the +2 position increases the binding affinity of *N* to the active site of OTase, but it is not the key requirement for a glycosylation. Rather, key requirements are that the position of the *N* residue be at the surface of a protein, neutrality of the amino acid in the +2 position, and compatibility with correct folding.⁴¹ Nevertheless, our findings seem to represent a very unique case as an *N*-glycan attached to the *N* in the NAA sequence has, as far as we know, no analogy in the available literature. All three found *N*-glycans are biantennary complex type and only slightly vary in their degree of terminal sialylation. Despite the fact that experimental evidence about *N*-glycosylation is available mostly for human C9, sequence analysis of C9 from various species and their alignment provided an interesting view on the C9 glycosylation in general (see Figure S4).

Although the presence of *O*-glycans on C9 was proposed more than 20 years ago,²¹ no direct evidence has been previously reported. Mucin type *O*-linked glycan chains, which were found in this study, are mostly located at the *N*-terminal part of the C9. The amino acid sequence QYTTSYDPELT-ESSGSASHIDCR of the *N*-terminal peptide does not seem to contain any distinct sequence motif for *O*-glycosylation; however, our integrated hybrid MS approach unambiguously demonstrated that *O*-glycans are attached to C9. The highly labile nature of *O*-glycans and their typically weak ionization response may have hampered their detection in earlier MS analysis. EThcD MS/MS spectra of the C9 *O*-glycosylated peptides revealed the composition of the *O*-linked glycan chains that are consistent with a core 1 type (GalNAc, also known as T-antigen). This structure is further terminally modified by a sialylation whereby the disialyl core 1 type was found to be the most abundant. Here, we report *O*-glycosylation of C9 for the first time, so the possible functional significance of these modifications remains speculative. Interestingly, it is known that the domain within the first 16 amino acids at the *N*-terminus of C9 is crucial in regulating the self-polymerization of C9.²⁵ Since the *N*-termini and also the putative *O*-glycosylation sites are not highly conserved among mammals (Figure S4), it seems unlikely that the *O*-glycans would play a significant role in the polymerization process, unless this is species specific as well. *O*-Linked glycans can also bind bacterial carbohydrate binding receptors.⁴² It is therefore tempting to speculate that the *O*-glycans on C9 protein in the MAC complex may interact with receptors or glycans at the surface of encountering pathogens.

The third type of glycosylation harbored on C9 is *C*-mannosylation. This less common glycosylation has been previously found to occur on C9 using MS, Edman

degradation, and nuclear magnetic resonance (NMR).²² It was already known that the N-terminal type-1 TSP domain of C9 can be modified with two mannoses. Here, we report data supporting these findings. While in human C9 W27 is 100% occupied, W30 is C-mannosylated in only ~23% of the C9.

All these PTMs are covalently attached to the backbone sequence of C9. Our native MS data provided additional information regarding the noncovalent binding of Ca²⁺ to C9. The only direct experimental proof about the binding of Ca²⁺ by intact C9 was reported two decades ago through the use of plasma emission spectroscopy. It was determined that C9 binds one Ca²⁺ ion per molecule with a dissociation constant of ~3 μM at physiological pH and ionic strength.⁴³ Here, we confirm the findings of that earlier report. Notably, standard bottom-up/middle-down proteomics approaches are not capable of identifying metal ion binding, due to the fact that such analyses happen under denaturing conditions. Our data indicate that C9 can possibly bind up to two Ca²⁺ ions. The exact site of Ca²⁺ binding in C9 has not been identified yet; nevertheless, one of the possible candidate sites is located at cysteine-rich LDLRA domain of C9.⁴³ Another possible Ca²⁺ binding site is at the C9 N-terminal part, which also harbors O-glycosylation sites. The 12 N-terminal amino acids of C9 exhibit a strong negative charge, and the 16 N-terminal amino acids contain a consensus sequence for Ca²⁺ binding proteins.²⁵ As mentioned above, the N-terminal region of C9, located at the outside of the MAC (Figure 4d,e), has been shown to play a significant role in C9 polymerization, hinting at a possible regulating role of the here identified O-glycosylation and Ca²⁺ binding in the assembly of the MAC or interactions with some other molecules or receptors.

In conclusion, here, we provide an unbiased detailed specification of three different types of glycosylation on C9 isolated from human blood serum and quantification and validation of 15 C9 proteoforms. In total, we achieved more than 90% correlation between the native MS data and peptide centric data. Mapping all sites of modifications on a structural model of C9, as present in the MAC, hints at their putative roles in pore formation or receptor interactions. In general, the applied combination of MS methods represents a powerful tool for the in-depth analysis of plasma proteins and thus may advance biomarker discovery.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b04527.

MS/MS spectra of N-, O-, and C-glycosylated peptides; multiple amino acid sequence alignment of the C9 protein; the certificate of analysis of the purified C9 sample; peptide-centric proteomic data; native MS data (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: a.j.r.heck@uu.nl

ORCID

Albert J. R. Heck: 0000-0002-2405-4404

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the support from The Netherlands Organization for Scientific Research (NWO) for funding the large-scale proteomics facility *Proteins@Work* (project 184.032.201) embedded in The Netherlands Proteomics Centre. Y.Y. and A.J.R.H. are supported by the EU funded ITN project ManiFold, grant 317371. This project has received additional funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 668036 (RELENT) and 686547 (MSMed). V.F. and A.J.R.H. acknowledge further support by the NWO TOP-Punt Grant 718.015.003.

■ REFERENCES

- (1) Seo, J.; Lee, K. *J. Biochem. Mol. Biol.* **2004**, *37*, 35–44.
- (2) Spiro, R. G. *Glycobiology* **2002**, *12*, 43R–56R.
- (3) Hart, G. W. *Curr. Opin. Cell Biol.* **1992**, *4*, 1017–1023.
- (4) Rudd, P. M.; Dwek, R. A. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32*, 1–100.
- (5) Chalabi, S.; Panico, M.; Sutton-Smith, M.; Haslam, S. M.; Patankar, M. S.; Lattanzio, F. A.; Morris, H. R.; Clark, G. F.; Dell, A. *Biochemistry* **2006**, *45*, 637–647.
- (6) Arnold, J. N.; Wormald, M. R.; Sim, R. B.; Rudd, P. M.; Dwek, R. A. *Annu. Rev. Immunol.* **2007**, *25*, 21–50.
- (7) van de Waterbeemd, M.; Lossel, P.; Gautier, V.; Marino, F.; Yamashita, M.; Conti, E.; Scholten, A. M.; et al. *Angew. Chem., Int. Ed.* **2014**, *53*, 9660–9664.
- (8) Rosati, S.; Rose, R. J.; Thompson, N. J.; van Duijn, E.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R. *Angew. Chem., Int. Ed.* **2012**, *51*, 12992–12996.
- (9) Lössel, P.; van de Waterbeemd, M.; Heck, A. J. *EMBO J.* **2016**, *35*, 2634–2657.
- (10) Yang, Y.; Liu, F.; Franc, V.; Halim, L.; Schellekens, H.; Heck, A. J. R. *Nat. Commun.* **2016**, *7*, 13397.
- (11) Staub, A.; Guillaume, D.; Schappler, J.; Veuthey, J. L.; Rudaz, S. *J. Pharm. Biomed. Anal.* **2011**, *55*, 810–822.
- (12) Bush, D. R.; Zang, L.; Belov, A. M.; Ivanov, A. R.; Karger, B. L. *Anal. Chem.* **2016**, *88*, 1138–1146.
- (13) Haselberg, R.; de Jong, G. J.; Somsen, G. W. *Anal. Chem.* **2013**, *85*, 2289–2296.
- (14) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, M. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. *Nature* **2011**, *480*, 254–258.
- (15) Ritchie, G. E.; Moffatt, B. E.; Sim, R. B.; Morgan, B. P.; Dwek, R. A.; Rudd, P. M. *Chem. Rev.* **2002**, *102*, 305–320.
- (16) Dudkina, N. V.; Spicer, B. A.; Reboul, C. F.; Conroy, P. J.; Lukyanova, N.; Elmlund, H.; Law, R. H. P.; Ekkel, S. M.; Kondos, S. C.; Goode, R. J. A.; Ramm, G.; Whisstock, J. C.; Saibil, H. R.; Dunstone, M. A. *Nat. Commun.* **2016**, *7*, 10588.
- (17) Serna, M.; Giles, J. L.; Morgan, B. P.; Bubeck, D. *Nat. Commun.* **2016**, *7*, 10587.
- (18) Stanley, K. K.; Kocher, H. P.; Luzio, J. P.; Jackson, P.; Tschopp, J. *EMBO J.* **1985**, *4*, 375–382.
- (19) Biesecker, G.; Müller-Eberhard, H. J. *J. Immunol.* **1980**, *124*, 1291–1296.
- (20) Biesecker, G.; Gerard, C.; Hugli, T. E. *J. Biol. Chem.* **1982**, *257*, 2584–2590.
- (21) Kontermann, R.; Rauterberg, E. W. *Mol. Immunol.* **1989**, *26*, 1125–1132.
- (22) Hofsteenge, J.; Blommers, M.; Hess, D.; Furmanek, A.; Miroshnichenko, O. *J. Biol. Chem.* **1999**, *274*, 32786–32794.
- (23) Paas, Y.; Bohana-Kashtan, O.; Fishelson, Z. *Immunopharmacology* **1999**, *42*, 175–185.
- (24) Narayanasamy, A.; Ahn, J. M.; Sung, H. J.; Kong, D. H.; Ha, K. S.; Lee, S. Y.; Cho, J. Y. *J. Proteomics* **2011**, *74*, 2948–2958.

- (25) Taylor, K. M.; Trimby, A. R.; Campbell, A. K. *Immunology* **1997**, *91*, 20–27.
- (26) Rosati, S.; Yang, Y.; Barendregt, A.; Heck, A. J. R. *Nat. Protoc.* **2014**, *9*, 967–976.
- (27) Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R. *Nat. Methods* **2012**, *9*, 1084–1086.
- (28) Kussmann, M.; Nordhoff, E.; Rahbek-Nielsen, H.; Haebel, S.; Rossel-Larsen, M.; Jakobsen, L.; Gobom, J.; Mirgorodskaya, E.; Kroll-Kristensen, A.; Palm, L.; Roepstorff, P. *J. Mass Spectrom.* **1997**, *32*, 593–601.
- (29) Bern, M.; Kil, Y. J.; Becker, C. *Current protocols in bioinformatics*; Wiley: New York, 2012; Chapter 13: Unit 13.20; DOI: [10.1002/0471250953.bi1320s40](https://doi.org/10.1002/0471250953.bi1320s40).
- (30) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. *Bioinformatics* **2010**, *26*, 966–968.
- (31) Yang, Y.; Barendregt, A.; Kamerling, J. P.; Heck, A. J. *Anal. Chem.* **2013**, *85*, 12037–12045.
- (32) Purwaha, P.; Silva, L. P.; Hawke, D. H.; Weinstein, J. N.; Lorenzi, P. L. *Anal. Chem.* **2014**, *86*, 5633–5637.
- (33) Hofsteenge, J.; Huwiler, K. G.; Macek, B.; Hess, D.; Lawler, J.; Mosher, D. F.; Peter-Katalinic, J. *J. Biol. Chem.* **2001**, *276*, 6485–6498.
- (34) Budnik, B. A.; Lee, R. S.; Steen, J. A. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764*, 1870–1880.
- (35) DiScipio, R. G.; Hugli, T. E. *J. Biol. Chem.* **1985**, *260*, 14802–14809.
- (36) Vance, A.; Wu, W.; Ribaldo, R. K.; Segal, D. M.; Kearse, K. P. *J. Biol. Chem.* **1997**, *272*, 23117–23122.
- (37) Miletich, J. P.; Broze, G. J. *J. Biol. Chem.* **1990**, *265*, 11397–11404.
- (38) Yasuda, D.; Imura, Y.; Ishii, S.; Shimizu, T.; Nakamura, M. *FASEB J.* **2015**, *29*, 2412–2422.
- (39) Zielinska, F.; Gnad, J.; Wisniewski, M.; Mann, M. *Cell* **2010**, *141*, 897–907.
- (40) Sun, S.; Zhang, H. *Anal. Chem.* **2015**, *87*, 11948–11951.
- (41) Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P. *Nature* **2011**, *474*, 350–355.
- (42) Hounsell, E. F.; Davies, M. J.; Renouf, D. V. *Glycoconjugate J.* **1996**, *13*, 19–26.
- (43) Thielens, N. M.; Lohner, K.; Esser, A. F. *J. Biol. Chem.* **1988**, *15*, 6665–6670.
- (44) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. *Nat. Methods* **2015**, *12*, 7–8.