Proteomics
Proteomics and Systems Biology

RESEARCH ARTICLE

Similarities and differences in the structures and proteoform profiles of the complement proteins C6 and C7

Marie V. Lukassen[1,2] (ID) | **Vojtech Franc**[1,2] | **Johannes F. Hevler**[1,2] | **Albert J.R. Heck**[1,2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, The Netherlands

[2] Netherlands Proteomics Center, The Netherlands

Correspondence
Albert Heck, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
Email: a.j.r.heck@uu.nl

Marie V. Lukassen and Vojtech Franc contributed equally to this work

Abstract

The human complement system provides a first line of defence against pathogens. It requires a well-orchestrated sequential assembly of an array of terminal complement components (C5, C6, C7, C8, and C9), ultimately forming the membrane attack complex (MAC). Although much information about MAC assembly is available, the structure of the soluble C7 has remained elusive. The complement proteins C7 and C6 share very high sequence homology and exhibit several conserved domains, disulphide bridges, and C-mannosylation sites. Here, we used an integrative structural MS-based approach combining native MS, glycopeptide-centric MS, in-gel cross-linking MS (IGX-MS) and structural modelling to describe structural features, including glycosylation, of human serum soluble C7. We compare this data with structural and glycosylation data for human serum C6. The new structural model for C7 shows that it adopts a compact conformation in solution. Although C6 and C7 share many similarities, our data reveals distinct O-, and N-linked glycosylation patterns in terms of location and glycan composition. Cumulatively, our data provide valuable new insight into the structure and proteoforms of C7, solving an essential piece of the puzzle in our understanding of MAC assembly.

KEYWORDS
complement, cross-linking, glycosylation, mass spectrometry, proteoforms

1 | INTRODUCTION

The membrane attack complex (MAC) is part of the innate immune system's first defence line against pathogens. Upon infection and complement activation, the MAC pore assembly is initiated and proceeds through a well-orchestrated sequential assembly of soluble complement components [1]. Most of these components (C6, C7, C8, and

C9) share similar domains, such as thrombospondin type-1 (TSP1), low-density lipoprotein class A (LDL), and MAC/perforin (MACPF) domains. Despite these similarities, each complement component is playing a distinctive role in the MAC assembly. All these complement components are glycoproteins and are harbouring several N- and O-glycosylation sites and even several C-mannosylation sites [2–4]. Differences in the complement components' protein glycosylation and their potential role in complement activation and MAC assembly are still not very well-described.

Several structural models have been described, both for the soluble complement components and the MAC pore and even for some assembly intermediates [5–11]. From these studies, it is well known that all

---

the complement proteins undergo dramatic structural and conformational changes in their transition from soluble monomers to being part of the larger assemblies. A missing piece of information is still a fine-quality structural model for soluble C7.

Upon binding of C7 to the soluble C5b6 complex, the complex becomes lipophilic and tethered to the membrane [12]. This transition is believed to be caused by the unfurling of the cluster of helices (CH) regions of the MACPF domain to form a continuous beta-sheet. Unlike the other MAC components, the only known high-resolution structure of C7 is from the fully assembled MAC [9]. The much better-characterized C6, the closest homolog of C7, can adopt various conformations in its soluble state [5, 7, 9, 13]. Based on the sequence and domain similarities between C6 and C7, it is tempting to speculate that the soluble C7 could also possess this conformational flexibility. Current structural knowledge on soluble C7 is limited to a low-resolution structure derived from transmission electron microscopy and a NMR structure of the FIM domains [12, 14]. As C7 plays a unique membrane tethering role in the MAC assembly pathway, we deemed it important to characterize its structure and proteoform profile in more detail.

This study examines the structure and proteoform profile of C7 by an integrative structural MS-based approach combining native MS, glycopeptide-centric MS, in-gel cross-linking MS (IGX-MS) [13], and structural modelling. Furthermore, we present the proteome profile of C6 and extend on our earlier work, where we showed that C6 adapts a compact conformation in solution [13]. Comparisons to the C6 structure reveals, as expected, an overall high similarity in structure and domain organization. However, C7 harbours several distinctive features. The flexible linker from the globular main body to the C-terminal arm is shorter, thus limiting the conformational range of the compacted soluble structure. Furthermore, we found that O- and N-glycosylation of C7 was distinct in both location and composition compared to C6. Notably, an N-glycan, found only on C7, is buried inside our structural model, possibly explaining why it was found to be modified distinctively with only high-mannose glycans.

2 | MATERIALS AND METHODS

2.1 | Chemicals and materials

Complement component C6 and C7 proteins purified from pooled human blood plasma (more than three healthy donors) were acquired from Complement Technology, Inc. (Texas, USA). The samples were purified according to a standard protocol (the certificate of analysis is attached as Supporting Information 1 and 2) [15, 16]. Tris-HCl, sodium deoxycholate (SDC), Tris (2-carboxyethyl)phosphine hydrochloride (TCEP), chloroacetamide (CAA) and ammonium acetate (AMAC) were from Sigma-Aldrich (Steinheim, Germany). Formic acid (FA) and TFA were from Merck (Darmstadt, Germany). ACN was purchased from Biosolve (Valkenswaard, The Netherlands). Sequencing grade trypsin was obtained from Promega (Madison, WI). Glu-C was obtained from Roche (Indianapolis, USA) and neuraminidase from *Arthrobacter ureafaciens* from Merck KGaA (Darmstadt, Germany).

Significance Statement

Activation of the complement system by pathogens or antibodies results in the formation of the membrane attack complex (MAC). MAC assembly requires sequential association and conformational changes of complement components C5b, C6, C7, C8, and multiple C9. Though a lot is known about structures along the assembly pathway, C7 remains to be structurally characterized in its soluble form. Furthermore, the MAC is a highly-glycosylated complex and the function of these PTMs is still elusive. This study uses an integrative structural MS-based approach combining native MS, glycopeptide-centric MS, and recently introduced in-gel cross-linking MS (IGX-MS) to describe the structure and glycosylation of soluble C7. The data demonstrate that C7 in solution, like C6, adopts a compact conformation with the C-terminal domains binding to the cluster of helices (CH) regions of the MAC/perforin (MACPF). These CH regions are dynamic and undergoes conformational changes during MAC assembly. Proteoform profiling of C6 and C7 shows distinctive O- and N-glycosylation despite the high structural homology between the two proteins.

2.2 | Sequence alignment

The sequences of C6 and C7 (C6: P13671, C7: P10643) were aligned using EMBOSS Needle with the EBLOSUM62 matrix [17]. The open gap penalty was 10, and the extended gap penalty 0.5. One region (C6 836–863, C7 752–773) was realigned to align cysteine residues better. Here individual alignment of minor segments was performed. PTM information about C6 and C7 protein glycosylation sites previously reported were added [5, 18-23].

2.3 | Native MS of C6 and C7

For native MS analysis, unprocessed protein solution in a phosphate buffer at pH 7.2, containing ∼ 30–40 μg of C6 or C7, was buffer exchanged into 150 mM aqueous AMAC (pH 7.5) and concentrated to a volume of ∼ 50 μL by ultrafiltration (vivaspin500, Sartorius Stedim Biotech, Germany) using a 10 kDa cut-off filter. The resulting protein concentration was measured by UV absorbance at 280 nm and adjusted to 2–3 μM prior to native MS analysis. For desialylation, the enzyme sialidase (Neuraminidase from *Arthrobacter ureafaciens*) was used to remove sialic acids (Merck KGaA, Darmstadt, Germany). The desialylation was performed by incubating the sample with 0.02 U of sialidase 8 h at room temperature. All samples were buffer exchanged to 150 mM AMAC (pH 7.2) prior to native MS measurements. Native and desialylated proteins were subsequently analysed on a modified Exactive Plus Orbitrap instrument with extended mass range (EMR)

(Thermo Fisher Scientific, Bremen) as described previously [24]. The samples were introduced into the instrument via goal-coated borosilicate capillaries prepared in-house [25]. The measurements were conducted in positive mode and a standard m/z range of 500–15,000 was used [26]. The voltage offsets on the transport multi-poles and ion lenses were manually tuned to achieve optimal transmission of protein ions at elevated m/z. Nitrogen was used in the higher-energy-collision dissociation (HCD) cell at a gas pressure of 6–8×10^{-10} bar. The optimized MS parameters were used: spray voltage 1.2–1.3 V; source fragmentation 100 V; source temperature 250°C; collision energy 30 V; automatic gain control: fixed; noise level parameter 3.64; resolution (at m/z 200) 35,000; injection time 50 ms. The mass spectrometer was calibrated using CsI clusters, as described previously [24].

The masses of the observed proteoforms of C6 and C7 were obtained from the zero-charge deconvoluted native mass spectra using Intact Mass software (Protein Metrics ver. 3.8) [27]. Parameters for spectra deconvolution were used as follows: min difference between mass peaks was set to 15 Da, charge vectors spacing was set to 0.4, smoothing sigma was set to m/z 0.02, spacing was set to m/z 0.04 and peak sharpening was disabled. Other parameters were set to be default. The spectra were annotated using an in-house developed R script for semi-automated peak annotation, as described previously [28].

2.4 | Glycopeptide-centric MS analysis of C6 and C7

For LC-MS and MS/MS analysis, native human C6 and C7 proteins in PBS buffer (10 mM sodium phosphate, 145 mM NaCl, pH 7.3) at a concentration of 1 mg/mL were introduced into the digestion buffer containing 100 mM Tris-HCl (pH 8.5), 1% w/v SDC, 5 mM TCEP and 30 mM CAA. Proteins were digested overnight with trypsin at an enzyme-to-protein-ratio of 1:100 (w/w) at 37°C. Another C6 and C7 samples were digested for 4 h by using Glu-C 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. After, the SDC was precipitated by bringing the samples to 1% TFA. The supernatant was collected for subsequent desalting by an Oasis μElution HLB 96-well plate (Waters, Wexford, Ireland) positioned on a vacuum manifold. All desalted proteolytic digests containing modified glycopeptides were dried with a SpeedVac apparatus and stored at -20°C.

Prior LC-MS/MS analysis, the samples were reconstituted in 1% FA. Approximately 300 fmol of peptides was separated and analysed using the HPLC system (Agilent Technologies, Waldbronn, Germany) coupled on-line to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). The peptides were first trapped on a 100 μm $\times$ 20 mm trap column (in-house packed with ReproSil-Pur C18-AQ, 3 μm) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) and then separated on the in-tandem connected 50 μm $\times$ 500 mm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 μm) (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 66 min gradient was used as follows: 0–5 min, 100% solvent A; 13–44% solvent B within 35 min; 44–100% solvent B within 3 min; 100% solvent B for 5 min; 100% solvent A for 17 min.

The mass spectrometer was operated in positive ion mode, and the spectra were acquired in the data-dependent acquisition mode. 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. For the MS scans, the mass range was set from m/z 350 to 1800 at a resolution of 60,000 maximum injection time 50 ms, and the normalized AGC target set to 4 $\times 10^5$. For the MS/MS measurements, HCD and electron-transfer and higher-energy collision dissociation (EThcD) were used. MS/MS scans were performed with fixed first mass m/z 120. The resolution was set to 30,000; the AGC target was set to 4 $\times 10^5$ (with normalized AGC target 800%); the precursor isolation width was 1.6 Da and the maximum injection time was set to 250 ms. Stepped HCD was performed with a normalized collision energy of 15, 30, and 45%. For EThcD, the supplementary activation energy of 27% was used.

2.5 | Combining native MS and glycopeptide-centric proteomic data

Annotation of the resolved proteoforms in C6 or C7 was achieved by an integrative approach combining the native MS and glycoproteomic data. This hybrid MS approach has been described previously by Yang *et al.* [29] and further developed by implementing peak annotation in a site-specific manner by Lin *et al.* [28]. This procedure allows visualizing proteoforms in the native spectrum with their PTM compositions in a site-specific manner. The glycopeptide data were used for *in silico* data construction of the native-like protein spectra. The simulated spectra were compared to the experimental native spectrum of C6 or C7, respectively. The similarity between the two data sets was expressed by Pearson correlation. The scripts used for the spectra annotation are available at https://github.com/juer120/NativePTMannotation.

In the mass calculations for C6, we used the average mass of C6 sequence of two allotypes A and B (UniProt code: P13671, Allotype A (E119) and B (A119)) lacking the N-terminal signalling propeptide, corrected for the mass shift induced by the 32 disulphide bonds present in C6 (-64 $\times$ 1.0079 Da). Eventually, the mass of calcium ion was added (40.078 u). This calculation resulted in average backbone masses of 102445.1124 Da (Allotype A) and 102385.0762 Da (Allotype B).

In the mass calculations for C7, we used the average mass of the C7 sequence (UniProt code: P10643) lacking the N-terminal signalling propeptide, corrected for the mass shift induced by the 28 disulphide bonds present in C7 (-56 $\times$ 1.0079 Da). Eventually, the mass of calcium ion was added (40.078 u). This calculation resulted in an average backbone mass of 91098.3980 Da.

Glycan structures were deduced based on known biosynthetic pathways. Average masses were used for the PTM assignments, including 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 glycan structures in this work are based on compositional information retrieved from the mass spectrometry measurements. These structures do not contain information about the linkages between the monosaccharide units. Such detailed structural information is possible to obtain via advanced glycomics methods [30, 31]. All used symbols and text nomenclature are according to recommendations of the Consortium for Functional Glycomics.

2.6 | IGX-MS of C7

IGX of C7 was performed using methods previously described in detail [13]. Briefly, 5 μg of C7 was loaded onto a blue native PAGE (Invitrogen, California, USA). The band corresponding to C7 was cut and cross-linked in the gel using 1.5 mM DSS for 30 min at RT. Following cross-linking, the reaction was quenched and subjected to in-gel digestion by trypsin. The generated peptides were extracted by ACN and dried. The peptides were then resuspended in MS solvent (2 % FA in water) and analysed as described below. Experiments were done in triplicates at the cross-linking level.

LC-MS data for cross-linked samples was acquired using an Ultimate 3000 system coupled on-line to an Orbitrap Exploris (Thermo Scientific, San Jose, CA). Firstly, peptides were trapped using a 0.3 × 5 mm PepMap-100 C18 pre-column (Thermo Scientific, San Jose, CA) of 5 μm particle size and 100 Å pore size for 1 min in solvent A (0.1 % FA) at a flow rate of 0.03 mL/min. The peptides were subsequently separated on an analytical column (50 cm of length, 75 μm inner diameter; packed in-house with Poroshell 120 EC-C18, 2.7 μm) with a 60 min gradient as follows: 9–13 % solvent B (0.1 % FA in 80 % v/v ACN) in 1 min, 13–41 % in 55.5 min, 41–99 % in 1.5 min and finally 99 % for 4 min. The flow was 300 nL/min and the mass spectrometer was operated in a data-dependent mode. A full scan MS spectra from *m/z* 375–2200 were acquired in the Orbitrap at a resolution of 60,000 with the AGC target set to 2×10^6 and a maximum injection time of 25 ms. Cycle time for MS^2 fragmentation scans was set to 2 s. Only peptides with charge states 3–8 were fragmented, and dynamic exclusion properties were set to n = 1 for a duration of 15 s. Fragmentation was performed using a stepped HCD collision energy mode (28, 31, 34 %) in the ion trap and acquired in the Orbitrap at a resolution of 30,000. The AGC target was set to standard with an isolation window of *m/z* 1.4 and a maximum injection time of 55 ms.

The resulting raw files from the MS-analysis were searched in MaxQuant (version 1.6.10.0) to generate libraries for the C7 bands. The data was searched against the reviewed *Homo Sapiens* Uniprot database (2019_08, downloaded from UniProt). Trypsin was set as a digestion enzyme with two allowed missed cleavages. Carbamidomethyl was set as fixed modification and oxidation (M) and acetylation (protein N-term) as variable modifications. For cross-link analysis the raw files were analysed with the Proteome Discoverer software suite version 2.3 (Thermo Fisher Scientific) with the incorporated XLinkX node to analyse the cross-linked peptides. The non-cleavable cross-link search option with the MaxQuant generated

FASTA file was used for the XlinkX search. The precursor mass tolerance was set to 10 ppm, the maximum FDR rate set to 1 % and ΔXlinkX score ≥ 40. Carbamidomethyl was set as fixed modification and oxidation (M) and acetylation (protein N-term) as variable modifications. Only cross-links observed in two out of the three replicates were included for further analysis. The cross-links were plotted onto the respective structures using PyMol to obtain Cα-Cα distances. Cross-link sequence overviews were generated using xiNET [32].

2.7 | Comparative modelling of C7

Consurf [33] and HHPRED [34] services were used to identify three suitable template structures (Table S1, Supporting Information 3) for comparative modelling using Modeller 9.24 [35]. Modelling was performed with disulphide bond restraints (Table S2, Supporting Information 3) as well as with cross-link restraints (mean = 25 Å, stdev = 1), resulting in 20 soluble C7 models. The generated models were optimized using the variable target function method following 100 iterations, with the speed of optimization schedule set to slow. Afterwards, the models were refined following molecular dynamics optimizations [36]. All optimizations were repeated two times and run until obj.func. > 1E^6. The model with the best DOPE score (and GA341 score) [37–39] and the model satisfying the distance restraints the best was further refined using Robetta [40]. The final model was validated using Mol-Probity, [23] ProSA-web [41] and MNXL [42]. For presented structures (soluble C6 [13], soluble C7), the most abundant glycans identified on each site (Supporting Information 4) were added to the structures using the GLYCAM-Web service (http://glycam.org). The size of added glycan structures corresponds to identified composition based on mass spectrometry. The linkages in these glycan structures were not experimentally confirmed. Therefore, our models do not show the exact glycan conformation but represent a visualization of approximate spacing of the glycans in the protein structures.

3 | RESULTS

3.1 | Sequence alignment of C6 and C7 suggests differences in PTMs

Domain assignment of C6 and C7 shows a highly similar domain structure with a TSP1 domain, followed by an LDL domain, the central MACPF domain, and a second TSP1 domain (Figure 1A). Additionally, C6 has an extra N-terminal TSP1 domain. The C-terminal is composed of a pair of complement control protein (CCP) domains and a pair of factor I-MAC (FIM) domains (Figure 1A). Pairwise sequence alignment shows a near-complete alignment of cysteine residues and provides a complete prediction of disulphide bonds covering all cysteine residues in C7 (Figure 1B and Table S2, Supporting Information 3). This observation suggests a similar domain fold and organization. The only cysteine residue that does not directly align (C6:Cys623, C7:Cys560) is located in the linker between the TSP1 and CCP1 domain. This cysteine
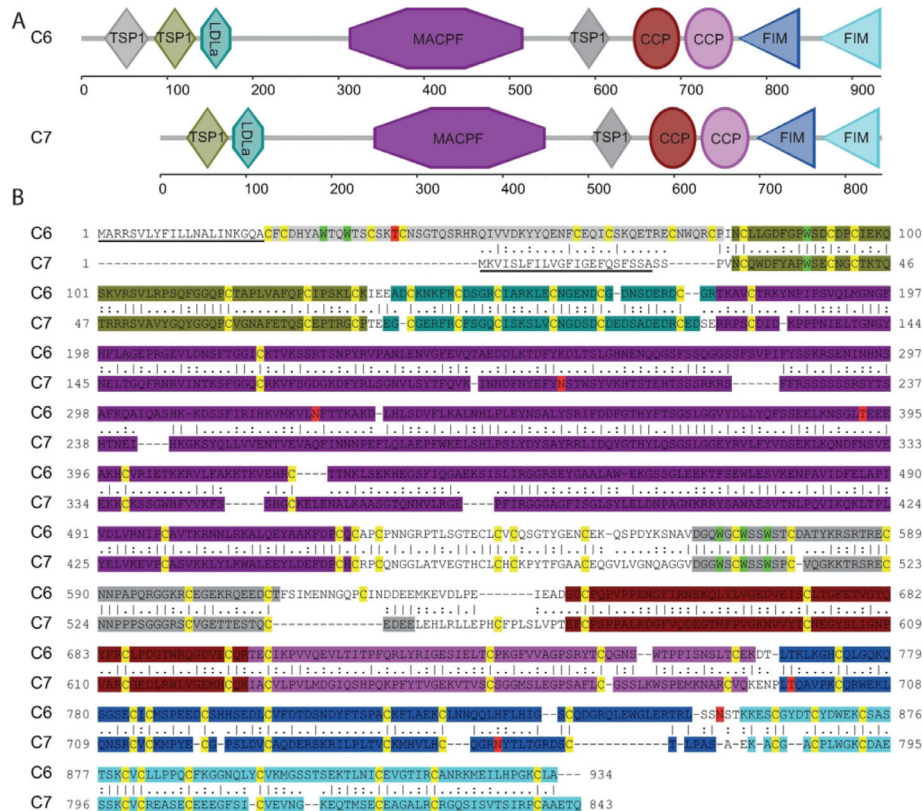
FIGURE 1 Complement proteins C6 and C7 share many characteristics. (A) Domain architecture of C6 (top) and C7 (bottom) as assigned by SMART [45]. (B) Pairwise sequence alignment of C6 and C7 using EMBOSS Needle complemented with manual annotation [17]. The domains are highlighted using the same color code as in (A), and the signal sequence is underlined. The many conserved cysteine residues are highlighted in yellow. Reported sites of C-mannosylation, notably in the TSP1 domains, are highlighted in green and O- or N-glycosylation sites in red

residue is linked to Cys499 of the MACPF domain in C6, which aligns with Cys433 in C7. Thus, Cys433 and Cys560 are most likely forming a disulphide bond in C7. The linker's length from this disulphide bond to the CCP1 domain is thereby ten residues longer in C6 (18 aa) than in C7 (8 aa). The location of predicted C-mannosylated tryptophan residues perfectly aligns between the two sequences [20]. The major difference between C6 and C7 are the location and numbers of reported O- and N-glycans (Figure 1B). C6 has two reported O-linked glycosites, one in the N-terminal TSP1 domain (T38) and one in the MACPF domain (T392) [5]. These glycosites are not present in C7. C7 has only one reported O-glycan located in the FIM1 domain (T696) [19]. C6 and C7 both have two N-glycosites. One is localized in the MACPF domain (C6:N324, C7:N202) and does not align in the sequence, likely resulting in different spatial positioning in the structure [5, 18, 22]. C7 has a second reported N-glycan located in the FIM1 domain (N754), while C6 has it between the two FIM domains (N855) [21, 22].

3.2 | Proteome profiling and PTM analysis of C6 and C7

Applying an integrative MS approach combining glycopeptide-centric analysis with native MS proteoform profiling revealed the composi-

tion of the PTMs of all significant C6 and C7 proteoforms (Figure 2). This approach allows for semi-automated site-specific annotation of high-resolution native MS spectra of glycoproteins [28, 29]. We first recorded high-resolution native MS spectra of C6 and C7, revealing C6 and C7 charge state distribution ranging from $[M + 19H]^{19+}$ to $[M + 23H]^{23+}$ for C6 and $[M + 17H]^{17+}$ to $[M + 23H]^{23+}$ for C7, respectively (Figure S1A and S2A, Supporting Information 3). Each charge state contains various ion series, which correspond to different proteoforms of C6 and C7. Based on their different masses, we can distinguish at least ~30 co-occurring proteoforms in C6 and ~40 in C7.

To aid in the annotation, we next treated the samples with sialidase. The removal of sialic acid residues resulted in apparent mass shifts in both proteins (Figure S1B and S2B, Supporting Information 3). Zooming into the most abundant charge states (+21) reveals that the C6 proteoform occurring at m/z 5135.14 shifted to m/z 5079.26, corresponding to an overall loss of four sialic acid moieties (Figure S1C-D, Supporting Information 3), which were cleaved off from the two known N-glycans on C6. The same experiments performed on C7 showed the same mass shift occurring on the C7 proteoform at m/z 4798.49, which shifted to m/z 4740.29 (Figure S2C-D, Supporting Information 3). Since C7 is known to contain two N-glycosylation sites, we assumed that sialic acids were released from N-glycans. Surprisingly, the subsequent glycopeptide-centric analysis revealed that C7 contains on the N202
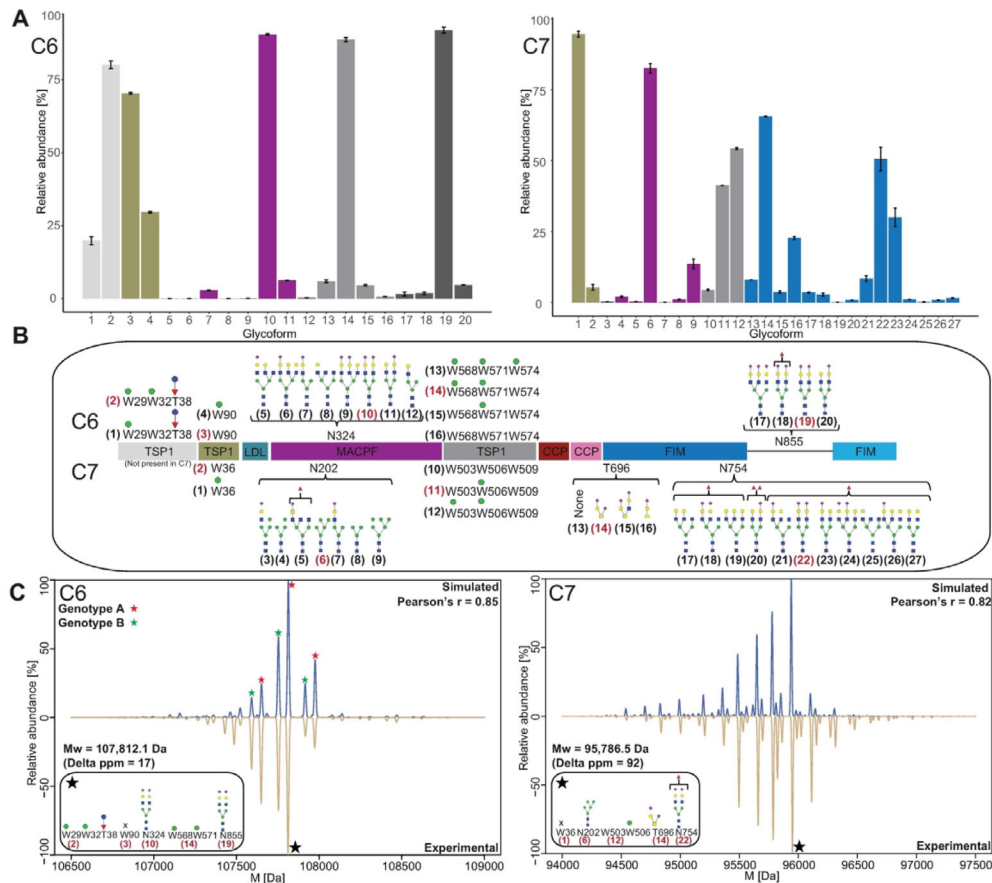
FIGURE 2 Proteoform profiling of C6 and C7. (A) Relative abundances of C6/C7 peptide glycoforms, as estimated from their corresponding extracted ion chromatograms (XICs). The glycoforms are assigned with numbers that correspond to the PTM compositions described in panels B and C. (B) Schematic of the domain composition of C6 and C7 with all found PTM sites, their occupancy, and microheterogeneity. The glycoforms highlighted by red numbers are present in the most abundant proteoform of C6/C7. (C) A comparison of the zero-charge deconvoluted native C6/C7 native MS spectra with the in silico reconstructed spectra based on the quantitative glycopeptide-centric proteomics data. The correlation between these spectra is for both C6 and C7 more than 0.8. Notably, the C6 proteoform profile is the consequence of a mixture of two genotypes A (red stars) and B (green stars), due to the frequently occurring E119A mutation, while C7 originates from a single gene variant. In summary, the complete proteoform profiling of C6 and C7 reveals huge similarities, especially in C-mannosylation, but also distinctive features in O- and N-glycosylation

site in the vast majority (≥ 95%) high mannose N-glycans (Figure S3, Supporting Information 3), suggesting that two out of four sialic acids were released from other glycan types. C7 is known to contain one O-glycosylation located in the FIM1 domain (T696) [19]. Detailed inspection of the MS/MS spectra revealed O-glycans with core 1 and 2 on T696 with variable sialylation level (Figure S4 and S5, Supporting Information 3).

Next, we made a correlative comparison between the native MS spectra of C6/C7 using *in silico* constructed MS spectra based on all the quantitative information gathered from the glycopeptide-centric data (Figure 2A-B and Supporting Information 4–6). This procedure resulted in a list of annotated C6 and C7 proteoforms in a site-specific manner (Supporting Information 4). This comparative analysis revealed a fair degree of consistency between our native MS and glycopeptide-centric MS data (R ∼ 0.85), explaining the signal of (nearly) all significant proteoforms detected in the native MS spectra (Figure 2C).

3.3 | IGX-MS of soluble C7 reveals a compact conformation

We next performed IGX-MS to obtain distance restraints for modelling the soluble C7 structure (Supporting Information 7). Plotting the cross-links on the C7 sequence revealed several cross-links between the C-terminal domains (CCP2, FIM1, and FIM2) and the MACPF domain (Figure 3A). These cross-links indicate a compacted conformation of C7 in solution where the C-terminal folds down and interacts with the MACPF domain. The soluble C7 model was generated based on three distinct template structures (Table S1, Supporting Information 3), using both the cross-links and disulfide bonds as additional restraints. Furthermore, the final model had an improved MNXL score compared to the initial model before refinement and thereby satisfies the soluble accessible surface distances (SASD) of the cross-links better [42]. In the top-ranked structural model, the soluble C7 structure consists of
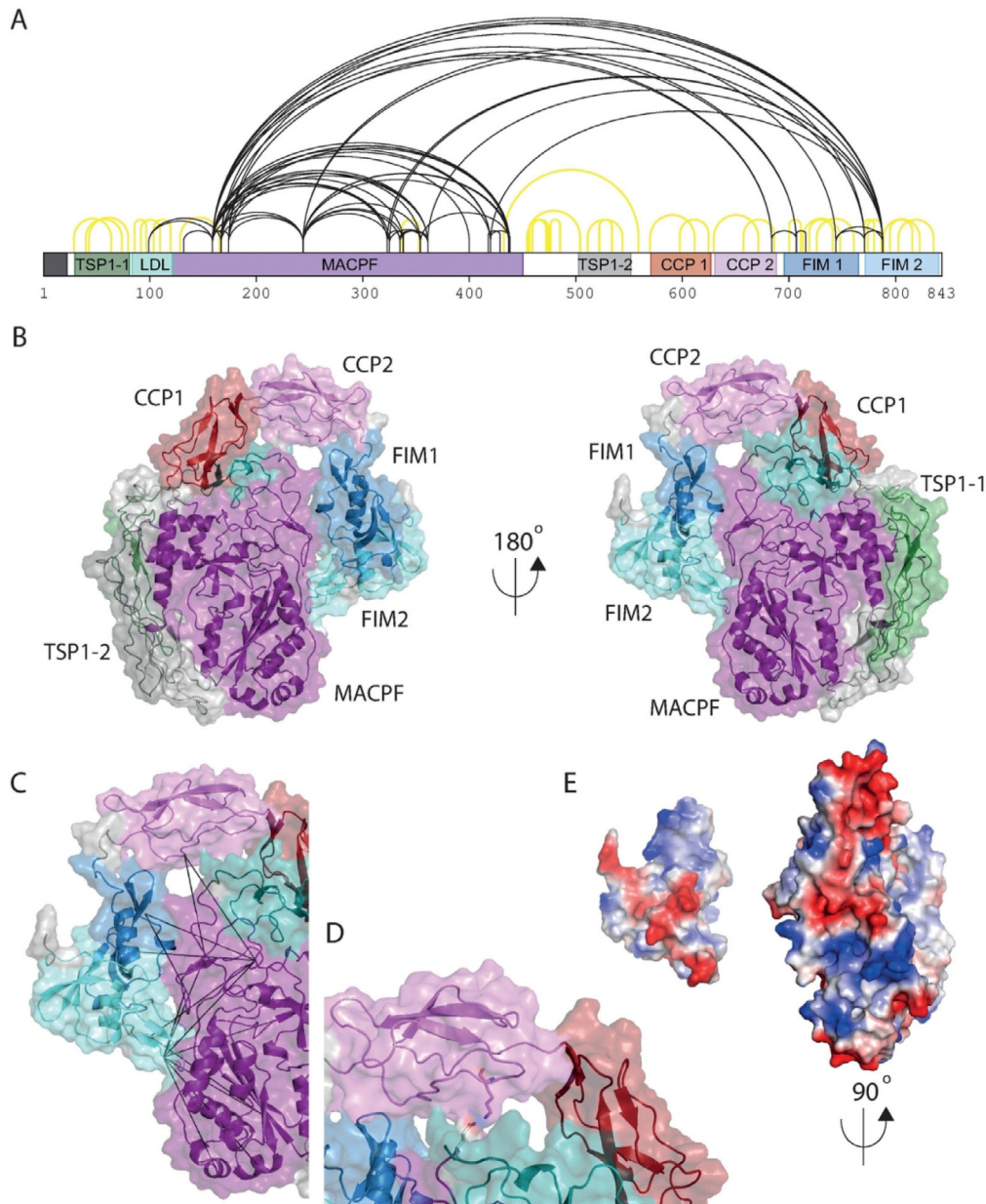
FIGURE 3 Structural model of soluble C7. (A) By IGX-MS detected cross-links (black lines) mapped on the C7 sequence with all domains color-coded. Yellow lines indicate the disulphide bonds, which are also used for the structural modelling (Table S2, Supplemental Information 3). (B) Final structural model of soluble C7 based on the cross-linking and disulphide-link restraints. The structure is shown in two orientations rotated by 180°. (C) Cross-links between the C-terminal domains (CCP2, FIM1, and FIM2) and the MACPF domain mapped onto the soluble C7 structure. (D) Zoom-in on the interaction surface between the CCP2 and the MACPF domain. Residues forming a salt bridge are shown as sticks (K678 and E123). (E) Surface electrostatics of the interaction interface between the FIM domains (left) and the MACPF (right) domain

a globular core composed of the TSP1, LDL, and MACPF domains with the C-terminal domains (CCP and FIM) wrapped around (Figure 3B). The FIM domains are interacting with the exposed side of the MACPF domain in line with the observed cross-links, all with Cα-Cα distances < 36 Å (Figure 3C). The CCP domains mainly interact with the LDL domain and CCP2 forms a salt bridge (K678-E123) to the linker between the LDL and MACPF domain (Figure 3D). Electrostatic interactions also mediate the interaction between the FIM2 and MACPF domains. An exposed negatively charged loop (783–788) of FIM2 is inserted into a positively charged pocket formed by the CH2 and CH3 regions of the MACPF (Figure 3E).

3.4 | Location of glycosylation on the structures of soluble C6 and C7

To gain insights into the spatial localization of the various PTMs on C6 and C7, we added the identified O- and N-linked glycosylations of
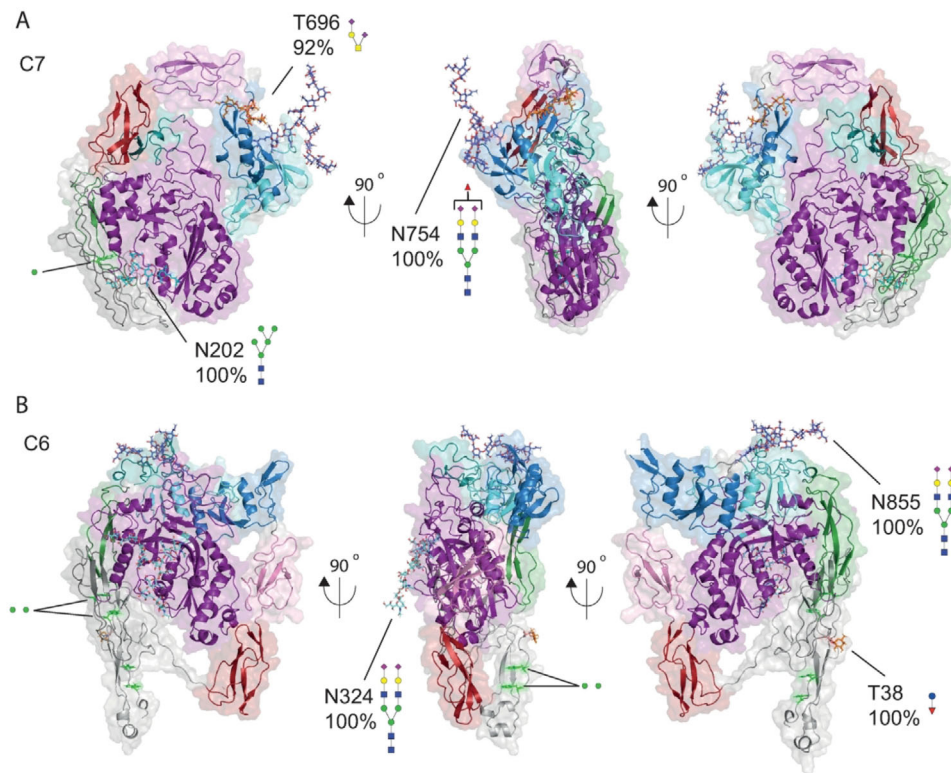
FIGURE 4 Certain glycosylation features distinct soluble C7 from C6. The most common O- and N-linked glycoforms were added onto the (A) soluble C7 and (B) soluble C6 structure [13] using the GLYCAM-Web service (http://glycam.org). The protein structures are shown in three different orientations rotated 90° respective to each other. The shown glycan structures indicate the experimentally observed most dominant O- and N-glycans. The C-mannosylated tryptophan residues are shown as green sticks. O- and N-glycans' occupancy at each site is indicated below the residue number with a schematic of the most abundant glycoform next to it. Besides the in between C7 and C6 strongly conserved C-mannosylations sites, the glycosylation features of C7 and C6 are distinct in O- and N-glycosylation. The glycan on N202, unique for C7, is somewhat buried inside the structural model, possibly explaining why it is mainly harbouring high-mannose glycans

the most abundant proteoform on the soluble C6 and C7 structures (Figure 4). We showed in a previous study that soluble C6 is in a compact conformation, similar to what we observe for C7, and not in an extended structure as seen in the X-ray structure [5, 13]. The interaction of the C-terminal domains with the MACPF domain is different for C6, probably due to the longer linker region. The O- and N-glycosylation of the C7 FIM2 domain (T696 and N855) are both located on the surface of the soluble structure. Distinctively for C7, the high-mannose N-glycan of the C7 MACPF domain (N202) is buried inside the structure between the MACPF and TSP1-2 domains (Figure 4A and Figure S6, Supporting Information 3). The C-mannosylated tryptophan residues are all located on one side of the structure of C6 and C7. The most common C7 proteoform harbours only one C-mannosylation located in the TSP1-2 domain (Figure 4A). The corresponding site is also occupied in C6. The C6 TSP1-1 domain, which is not present in C7, contains two C-mannosylation sites and one O-linked disaccharide (GlcFuc). These unusual glycosylation patterns have been identified on other proteins such as Thrombospondin-1 protein and properdin (Figure 4B) [29]. The three fully occupied O- and N-linked glycosites of C6 are all located on the structure's surface. N-glycan of the C6 MACPF domain (N324) is located towards the LDL domain, which is different from the C7 MACPF (N202) glycosylation, that is buried between the MACPF and TSP1-2 domains.

4 | DISCUSSION

Assembly of the MAC involves dramatic conformational changes to each of the individual complement components to facilitate proper pore formation. Several studies have presented protein structures shedding light on this assembly process, but there are still important gaps to be filled [5-9]. One of these is the structure of soluble C7, for which only low-resolution structures have been reported [12]. All the components of the MAC are also known to be modified with several types of glycosylation. However, the function and composition of these are not well-characterized. Here, we used an integrative structural MS-based approach complemented with structural modelling to generate a structure of soluble C7 and characterize the proteoforms of C6 and C7. Combining these results allows us to access the glycans' spatial localization and speculate on their potential role during the MAC assembly.

C6 and C7 share a similar domain organization highlighted by an identical disulphide bond pattern. Cross-linking of the soluble C7

by the recently developed IGX-MS approach provided us with distance restraints for the modelling [13]. The resulting structural model revealed a compacted conformation with the FIM2 domain binding to the CH2 and CH3 regions of the MACPF domain (Figure 4). We suggest that this interaction stabilizes the CH regions and prevents conformational changes. Upon binding to the C5b6 complex, the C-terminal is released and extends along the backside of C5b. This conformational change allows the CH3 region to shift and the CH1 and CH2 regions unfurl into beta-hairpins. Thereby, the C5b7 complex becomes more lipophilic, as previously described [12]. A previous study showed that C6, like C7, also adopts a similar compact conformation in solution [13]. The elongated X-ray structure of C6 is believed to be stabilized by crystal contacts to symmetry mates [5, 13]. The IGX-MS of soluble C6 showed a similar pattern to C7, with the FIM domains interacting with the MACPF domain. However, for C6, also cross-links between the CCP1 and TSP1-3 domains to MACPF were observed [13]. These contacts were not identified for C7 and might be caused by a smaller conformational range of the C7 linker between the core and the C-terminal arm. Thus, the C-terminal arm of C6 wraps around the MACPF in a different conformation than C7 (Figure 4).

The significant difference between C6 and C7 lies within their proteoform profiles. The C-mannosylations of the TSP1 domains align between C6 and C7, but the O- and N-glycosylations do not (Figure 1B). Proteoform analysis of C6 and C7 revealed distinctive patterns. Surprisingly, we found that the C7 N202 N-glycosite carries a high-mannose N-glycan, unlike the complex ones found at the other N-glycosites. The presence of immature glycan forms could indicate that the glycosylation is buried in the structure and thus not accessible for processing in Golgi. It has been shown previously that the accessibility of the glycosylated asparagine residue and the attached glycan is critical for the glycan maturation process on glycoproteins [43, 44]. Indeed, placing the glycan onto the soluble C7 structure indicates its location buried within the MACPF domain (Figure 4A). Alignment of C6 and C7 to other species shows that the C-mannosylation sites are all conserved (Figure S6, Supporting Information 3). C-mannosylation is a specific feature of TSP1 domains, which is also found in the other MAC components C8 and C9. C-mannosylation's function is still not clear, but these modifications have been suggested to be involved in the regulation of complement proteins [4]. The C-mannosylations are located in the solvent-exposed outer edge of the MAC, where they align at the interface to the neighbouring complement component [9]. We found that the C-mannosylation sites' occupancy varied in both C6 and C7 (Figure 3 and 4). The glycosylation sites of the MACPF domain are also conserved across species suggesting that they likely play an important role in the MAC function. Upon unfolding the CH regions of C7 into beta-hairpins, the high-mannose glycan (N202) faces the MAC pore's cavity. C8 and C9 also have glycans facing the centre of the pore, and the deletion of these glycosylation sites was shown to cause irregular pores with distorted curvature [9]. In the MAC, the C6 MACPF domain's glycosylation is facing the terminal C9 of the pore and could thus provide proper spacing and termination of the pore. The last conserved site is the N-glycosylation site of the C7 FIM1 domain.

The FIM domains are not resolved in the current MAC structure, but they are predicted to be located near the C345c domain of C5b [9]. It could thus play an important role in the interaction with C5b or solubility of C7. The other glycosites of the FIM domains are not conserved, and their role remains elusive. For the sake of completeness, the previously reported O-glycosylation of the C6 MACPF domain (T392) was not identified in our study [5]. However, this modification was also not observed in other structural studies [7, 9].

In conclusion, we present a well-supported structural model of C7 in its soluble form. The model fits the predicted disulphide assignment, obtained distance restraints, and the solvent accessibility of glycosites. Furthermore, we compared structures of C6 and C7 at the proteoform level and revealed substantial distinctive features. Our models of C6 and C7 show the spatial distribution of the most abundant PTMs, bringing a more general understanding of the structural heterogeneity of these terminal complement proteins.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

The MS data from this publication have been deposited to the ProteomeXchange partner MassIVE database and assigned the identifier MSV000087019. The deposited data includes raw files from native MS, bottom-up glycopeptide analysis MS and IGX-MS together with the cross-linking PD search output.

ORCID

Marie V. Lukassen https://orcid.org/0000-0003-3237-2696

REFERENCES

1. Bayly-Jones, C., Bubeck, D., & Dunstone, M A., (2017). The mystery behind membrane insertion: a review of the complement membrane attack complex. *Philosophical Transactions of the Royal Society B*, *20160221*(1726). https://doi.org/10.1098/rstb.2016.0221.
2. Franc, V., Yang, Y., & Heck, A. J. R., (2017). Proteoform profile mapping of the human serum complement component C9 revealing unexpected new features of N-, O-, and C-glycosylation. *Analytical Chemistry*, *89*(6), 3483–3491. https://doi.org/10.1021/acs.analchem.6b04527.
3. Franc, V., Zhu, J., & Heck, A. J. R., (2018). Comprehensive proteoform characterization of plasma complement component c8alphabetagamma by hybrid mass spectrometry approaches. *Journal of the American Society for Mass Spectrometry*, *29*(6), 1099–1110. https://doi.org/10.1007/s13361-018-1901-6.
4. Ritchie, G. E., Moffatt, B. E., Sim, R. B., Morgan, B. P, Dwek, R A., & Rudd, P M., (2002). Glycosylation and the complement system. *Chemical Reviews*, *102*(2), 305-320-319. https://doi.org/10.1021/cr990294a.

5. Aleshin, A. E., Schraufstatter, I. U., Stec, B., Bankston, L A., Lidding-ton, R C., & Discipio, R. G., (2012). Structure of complement C6 suggests a mechanism for initiation and unidirectional, sequential assembly of membrane attack complex (MAC). *Journal of Biological Chemistry*, *287*(13), 10210–10222. https://doi.org/10.1074/jbc.M111.327809.

6. Fredslund, F., Laursen, N. S., Roversi, P., Jenner, L., Oliveira, C. L. P., Pedersen, J. S., Nunn, M. A., Lea, S. M., Discipio, R., Sottrup-Jensen, L., & Andersen, G. R., (2008). Structure of and influence of a tick complement inhibitor on human complement component 5. *Nature Immunology*, *9*(7), 753–760. https://doi.org/10.1038/ni.1625.

7. Hadders, M. A., Bubeck, D., Roversi, P., Hakobyan, S., Forneris, F., Morgan, B. P, Pangburn, M. K., Llorca, O., Lea, S. M., & Gros, P., (2012). Assembly and regulation of the membrane attack complex based on structures of C5b6 and sC5b9. *Cell Reports*, *1*(3), 200–207. https://doi.org/10.1016/j.celrep.2012.02.003.

8. Lovelace, L. L., Cooper, C. L., Sodetz, J. M., & Lebioda, L., (2011). Structure of human C8 protein provides mechanistic insight into membrane pore formation by complement. *Journal of Biological Chemistry*, *286*(20), 17585–17592. https://doi.org/10.1074/jbc.M111.219766.

9. Menny, A., Serna, M., Boyd, C. M., Gardner, S., Joseph, A. P., Morgan, B. P, Topf, M., Brooks, N. J., & Bubeck, D., (2018). CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers. *Nature Communications*, *9*(1), 5316. https://doi.org/10.1038/s41467-018-07653-5.

10. Dudkina, N. V., Spicer, B. A., Reboul, C. F., Conroy, P J., Lukoyanova, 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., (2016). Structure of the poly-C9 component of the complement membrane attack complex. *Nature Communications*, *7*, 10588. https://doi.org/10.1038/ncomms10588.

11. Bubeck, D., Roversi, P., Donev, R., Morgan, B. P, Llorca, O., & Lea, S M., (2011). Structure of human complement C8, a precursor to membrane attack. *Journal of Molecular Biology*, *405*(2), 325–330. https://doi.org/10.1016/j.jmb.2010.10.031.

12. Discipio, R. G., Chakravarti, D. N., Muller-Eberhard, H. J., & Fey, G. H., (1988). The structure of human complement component C7 and the C5b-7 complex. *Journal of Biological Chemistry*, *263*(1), 549–560.

13. Hevler, J. F., Lukassen, M. V., Cabrera-Orefice, A., Arnold, S., Pronker, M. F., Franc, V., & Heck, A. J. R., (2021). Selective cross-linking of coinciding protein assemblies by in-gel cross-linking mass spectrometrys. *EMBO Journal*, e106174. https://doi.org10.15252/embj.2020106174.

14. Phelan, M. M., Thai, C. - T., Soares, D. C., Ogata, R. T., Barlow, P. N., & Bramham, J., (2009). Solution structure of factor I-like modules from complement C7 reveals a pair of follistatin domains in compact pseudosymmetric arrangement. *Journal of Biological Chemistry*, *284*(29), 19637–19649. https://doi.org/10.1074/jbc.M901993200.

15. Morgan, B. P., (2000). Immunoaffinity methods for purification of complement components and regulators. *Methods in Molecular Biology*, *150*, 53–60. https://doi.org/10.1385/1-59259-056-X:53.

16. Peterson, B., (1998). Complement: A practical approach; Edited by: H. W. Dodds and R. B. Sim. pp 274. IRL Press at Oxford University Press, Oxford. 1997. £29.95 ISBN 0-19-963539-0. *Biochemical Education*, *26*(2), 195. https://doi.org/10.1016/S0307-4412(98)00060-0.

17. Madeira, F., Park, Y Mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R., (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636-W641. https://doi.org/10.1093/nar/gkz268.

18. Bunkenborg, J., Pilch, B. J., Podtelejnikov, A. V., & Wiśniewski, J. R., (2004). Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics*, *4*(2), 454–465. https://doi.org/10.1002/pmic.200300556.

19. Halim, A., Nilsson, J., Rüetschi, U., Hesse, C., & Larson, G., (2012). Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Molecular and Cellular Proteomics*, *11*(4), M111 013649. https://doi.org/10.1074/mcp.M111.013649.

20. Hofsteenge, J., Blommers, M., Hess, D., Furmanek, A., & Miroshnichenko, O., (1999). The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues. *Journal of Biological Chemistry*, *274*(46), 32786–32794. https://doi.org/10.1074/jbc.274.46.32786.

21. Jia, W., Lu, Z., Fu, Y., Wang, H. - P., Wang, Le-H, Chi, H., Yuan, Z. - F., Zheng, Z. - B., Song, Li-Na, Han, H. - H., Liang, Yi-M, Wang, J. - L., Cai, Y., Zhang, Yu-K, Deng, Yu-L, Ying, W. - T., He, Si-M, & Qian, X. - H., (2009). A strategy for precise and large scale identification of core fucosylated glycoproteins. *Molecular and Cellular Proteomics*, *8*(5), 913–923. https://doi.org/10.1074/mcp.M800504-MCP200.

22. Liu, T., Qian, W. - J., Gritsenko, M. A., Camp, D. G., Monroe, M. E., Moore, R. J., & Smith, R. D., (2005). Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *Journal of Proteome Research*, *4*(6), 2070–2080. https://doi.org/10.1021/pr0502065.

23. Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S M., Arendall, W. B, Snoeyink, J., Adams, P D., Lovell, S. C., Richardson, J. S., & Richardson, D. C., (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*, *27*(1), 293–315. https://doi.org/10.1002/pro.3330.

24. Rosati, S., Yang, Y., Barendregt, A., & Heck, A. J. R., (2014). Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. *Nature Protocols*, *9*(4), 967–976. https://doi.org/10.1038/nprot.2014.057.

25. Van De Waterbeemd, M., Fort, K. L., Boll, D., Reinhardt-Szyba, M., Routh, A., Makarov, A., & Heck, A. J. R., (2017). High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. *Nature Methods*, *14*(3), 283–286. https://doi.org/10.1038/nmeth.4147.

26. Rosati, S., Rose, R. J., Thompson, N. J., Van Duijn, E., Damoc, E., Denisov, E., Makarov, A., & Heck, A. J. R., (2012). Exploring an orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angewandte Chemie (International Ed. in English)*, *51*(52), 12992–12996. https://doi.org/10.1002/anie.201206745.

27. Bern, M., Caval, T., Kil, Y J., Tang, W., Becker, C., Carlson, E., Kletter, D., Sen, K. I, Galy, N., Hagemans, D., Franc, V., & Heck, A J. R., (2018). Parsimonious charge deconvolution for native mass spectrometry. *Journal of Proteome Research*, *17*(3), 1216–1226. https://doi.org/10.1021/acs.jproteome.7b00839.

28. Lin, Yu-H, Zhu, J., Meijer, S., Franc, V., & Heck, A. J. R., (2019). Glycoproteogenomics: A frequent gene polymorphism affects the glycosylation pattern of the human serum fetuin/alpha-2-HS-glycoprotein. *Molecular and Cellular Proteomics*, *18*(8), 1479–1490. https://doi.org/10.1074/mcp.RA119.001411.

29. Yang, Y., Liu, F., Franc, V., Halim, L. A., Schellekens, H., & Heck, A J. R., (2016). Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nature Communications*, *7*, 13397. https://doi.org/10.1038/ncomms13397.

30. Jensen, P. H., Karlsson, N. G., Kolarich, D., & Packer, N. H., (2012). Structural analysis of N- and O-glycans released from glycoproteins. *Nature Protocols*, *7*(7), 1299–1310. https://doi.org/10.1038/nprot.2012.063.

31. Reinhold, V., Zhang, H., Hanneman, A., & Ashline, D., (2013). Toward a platform for comprehensive glycan sequencing. *Molecular and Cellular Proteomics*, *12*(4), 866–873. https://doi.org/10.1074/mcp.R112.026823.

32. Combe, C. W., Fischer, L., & Rappsilber, J., (2015). xiNET: Crosslink Network Maps With Residue Resolution. *Molecular and Cellular Proteomics*, *14*(4), 1137–1147. https://doi.org/10.1074/mcp.O114.042259

33. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N., (2016). ConSurf 2016: an improved methodology to

estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, *44*(W1), W344–W350. https://doi.org/10.1093/nar/gkw408.

34. Soding, J., Biegert, A., & Lupas, A. N., (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(Web Server issue), W244–W248. https://doi.org/10.1093/nar/gki408.

35. Webb, B., & Sali, A., (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Protein Science*, *86*, 2 9 1–2 9 37. https://doi.org/10.1002/cpps.20.

36. Šali, A., & Blundell, T. L., (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, *234*(3), 779–815. https://doi.org/10.1006/jmbi.1993.1626.

37. John, B., & Sali, A., (2003). Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Research*, *31*(14), 3982–3992. https://doi.org/10.1093/nar/gkg460.

38. Melo, F., Sánchez, R., & Sali, A., (2002). Statistical potentials for fold assessment. *Protein Science*, *11*(2), 430–448. https://doi.org/10.1002/pro.110430.

39. Shen, M. - Y., & Sali, A., (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science*, *15*(11), 2507–2524. https://doi.org/10.1110/ps.062416606.

40. Kim, D. E., Chivian, D., & Baker, D., (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*, *32*(Web Server issue), W526–W531. https://doi.org/10.1093/nar/gkh468.

41. Wiederstein, M., & Sippl, M. J., (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(Web Server issue), W407–W410. https://doi.org/10.1093/nar/gkm290.

42. Bullock, J. M. A., Schwab, J., Thalassinos, K., & Topf, M., (2016). The importance of non-accessible crosslinks and solvent accessible surface distance in modeling proteins with restraints from crosslinking

mass spectrometry. *Molecular and Cellular Proteomics*, *15*(7), 2491–2500. https://doi.org/10.1074/mcp.M116.058560.

43. Suga, A., Nagae, M., & Yamaguchi, Y., (2018). Analysis of protein landscapes around N-glycosylation sites from the PDB repository for understanding the structural basis of N-glycoprotein processing and maturation. *Glycobiology*, *28*(10), 774–785. https://doi.org/10.1093/glycob/cwy059.

44. Thaysen-Andersen, M., & Packer, N. H., (2012). Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology*, *22*(11), 1440–1452. https://doi.org/10.1093/glycob/cws110.

45. Schultz, J., Milpetz, F., Bork, P., & Ponting, C. P., (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(11), 5857–5864. https://doi.org/10.1073/pnas.95.11.5857.