

RESEARCH

Open Access



Defining albumin as a glycoprotein with multiple N-linked glycosylation sites

Kishore Garapati^{1,2,3†}, Anu Jain^{3†}, Benjamin J. Madden⁴, Dong-Gi Mun³, Jyoti Sharma^{1,2}, Rohit Budhraj³ and Akhilesh Pandey^{3,5*} 

Abstract

Background Glycosylation is an enzyme-catalyzed post-translational modification that is distinct from glycation and is present on a majority of plasma proteins. *N*-glycosylation occurs on asparagine residues predominantly within canonical *N*-glycosylation motifs (Asn-X-Ser/Thr) although non-canonical *N*-glycosylation motifs Asn-X-Cys/Val have also been reported. Albumin is the most abundant protein in plasma whose glycation is well-studied in diabetes mellitus. However, albumin has long been considered a non-glycosylated protein due to absence of canonical motifs. Albumin contains two non-canonical *N*-glycosylation motifs, of which one was recently reported to be glycosylated.

Methods We enriched abundant serum proteins to investigate their N-linked glycosylation followed by trypsin digestion and glycopeptide enrichment by size-exclusion or mixed-mode anion-exchange chromatography. Glycosylation at canonical as well as non-canonical sites was evaluated by liquid chromatography–tandem mass spectrometry (LC–MS/MS) of enriched glycopeptides. Deglycosylation analysis was performed to confirm N-linked glycosylation at non-canonical sites. Albumin-derived glycopeptides were fragmented by MS3 to confirm attached glycans. Parallel reaction monitoring was carried out on twenty additional samples to validate these findings. Bovine and rabbit albumin-derived glycopeptides were similarly analyzed by LC–MS/MS.

Results Human albumin is *N*-glycosylated at two non-canonical sites, Asn⁶⁸ and Asn¹²³. *N*-glycopeptides were detected at both sites bearing four complex sialylated glycans and validated by MS3-based fragmentation and deglycosylation studies. Targeted mass spectrometry confirmed glycosylation in twenty additional donor samples. Finally, the highly conserved Asn¹²³ in bovine and rabbit serum albumin was also found to be glycosylated.

Conclusions Albumin is a glycoprotein with conserved N-linked glycosylation sites that could have potential clinical applications.

Keywords Human serum albumin, HSA, BSA, Glycoproteomics, Microheterogeneity, Novel glycosylation site

[†]Kishore Garapati and Anu Jain contributed equally.

*Correspondence:

Akhilesh Pandey
pandey.akhilesh@mayo.edu

¹ Manipal Academy of Higher Education (MAHE), Manipal, Karnataka, India

² Institute of Bioinformatics, International Technology Park, Bangalore, Karnataka, India

³ Department of Laboratory Medicine and Pathology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

⁴ Proteomics Core, Mayo Clinic, Rochester, MN, USA

⁵ Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA

Background

Glycosylation is the commonest post-translational modification (PTM) of proteins [1]. It is distinct from glycation, a non-enzymatic process of protein modification by the addition of sugars on a background of hyperglycemia. Glycation affects a number of plasma proteins including albumin, haptoglobin and fibrinogen and is associated with microvascular damage and organ dysfunction in advanced diabetes [2]. By contrast, glycosylation is an enzyme-catalyzed physiological process which occurs on specific amino acids and is essential for protein stability,



folding and function [3]. N-linked glycosylation is the most complex form of protein glycosylation in humans, where oligosaccharide chains or glycans are covalently attached to proteins at asparagine (Asn) residues by an N-glycosidic bond [1]. Most secretory and plasma proteins are N-glycosylated at asparagines in a canonical motif in the primary amino acid sequence, Asn-X-Ser/Thr, where X is any amino acid except proline [4]. The hydroxyl group in the side chain of serine or threonine performs the hydrogen bond donor function that is necessary for the catalytic transfer of the N-glycan to asparagine [5]. However, the presence of this motif is not sufficient for, and does not always result in, glycosylation. It is estimated that only ~70% of such sites are glycosylated [4]. Further, N-glycosylation sites are occupied by glycans to different levels, defining glycosylation macroheterogeneity [6]. Besides the canonical motif, N-glycosylation occurs on asparagines within the non-canonical motif Asn-X-Cys of some proteins, with the sulfhydryl group of cysteine performing the hydrogen bond donor function. However, the sulfur on cysteine has less electronegativity than oxygen on the side chains of serine or threonine [7]. As a result, this motif is known to be glycosylated at low levels in several proteins including transferrin and von Willebrand Factor [8, 9]. Another non-canonical motif, Asn-X-Val, has been shown to be glycosylated to low levels in some proteins including alpha-1B-glycoprotein and apolipoprotein B-100 [10, 11].

Mass spectrometry (MS)-based analysis of deglycosylated peptides has historically played an important role in the identification of glycoproteins and their sites of N-glycosylation [12]. Advancements in MS technology over the past several years coupled with the development of appropriate database search tools have facilitated comprehensive glycopeptide profiling with identification of intact glycans and their sites of attachment [11]. We sought to deploy advanced MS methods to discover and characterize glycosylation events that might have been missed previously because of low abundance or because they occurred at non-canonical motifs. Among abundant plasma proteins, such motifs, i.e., Asn-X-Cys or Asn-X-Val are present in alpha-2-macroglobulin, alpha-1-acid glycoprotein 2, transferrin, immunoglobulin heavy chains, and albumin [13]. Albumin is the most abundant plasma protein and besides maintenance of colloidal osmotic pressure of plasma, it functions as a transporter, antioxidant and enzyme [14]. It has been considered a non-glycosylated protein because it does not contain a canonical motif in its amino acid sequence. However, asparagines at sites Asn⁶⁸ and Asn¹²³ are part of non-canonical N-glycosylation motifs Asn-X(Glu)-Val and Asn-X(Glu)-Cys, respectively [13]. We wondered if albumin is glycosylated at these sites at

levels that might not be detected by traditional methods of glycoprotein analysis [15]. Recently, one of these sites, i.e., Asn⁶⁸, was reported to be linked to two glycans (Hex₅HexNAc₄NeuAc₂ and Hex₅HexNAc₄NeuAc₁) based on MS/MS fragmentation data [10]. In our experience with the analysis of plasma and serum-derived glycopeptides enriched using alternate methods, we observe a greater degree of glycan microheterogeneity in glycopeptides derived from abundant plasma proteins [11]. We were intrigued if Asn⁶⁸ is occupied by a larger glycan repertoire and if Asn¹²³ is also glycosylated. Thus, we systematically investigated N-linked glycosylation of albumin in serum from volunteer donors using a multi-pronged approach.

Methods

Samples

Twenty-three serum samples used in this study were deidentified residual samples from volunteer donors (approved by Mayo Clinic IRB: 21-012890).

LC-MS/MS-based discovery analysis of serum-derived glycopeptides

Serum samples from volunteer donors were first enriched for 14 abundant serum proteins and digested with trypsin. Glycopeptides were enriched from the peptide mixture using either size exclusion chromatography or mixed-mode anion exchange cartridge (MAX), and analyzed by mass spectrometry (MS) in data dependent acquisition mode on an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific) [11, 16, 17]. Data was searched in pGlyco3 [18]. Commercial bovine (Thermo Scientific) and rabbit (Sigma) serum albumin were digested followed by glycopeptide enrichment using MAX. Details of sample preparation and MS analysis are provided in Additional file 1: Supplemental Methods.

Mapping N-glycosylation sites onto structure of albumin

The crystal structure of human albumin derived from pooled human plasma with the identifier 1A06 [19] was obtained from the PDB [20] and visualized using PyMOL (v2.5.7) [21]. N-linked glycosylation site Asn⁶⁸ was highlighted in red color. The structure was rotated by 90° to visualize the other glycosylation site, Asn¹²³, which was also highlighted in red.

Deglycosylation analysis of serum glycoproteins

Glycopeptides from serum proteins enriched by MAX were treated overnight with PNGase F (N-Zyme Scientifics) in either ¹⁶O or ¹⁸O water (97% ¹⁸O enriched, Sigma) at 37 °C. Deglycosylated peptides were analyzed by MS in parallel reaction monitoring mode as described in

the Additional file 1: Supplemental Methods. Spectral inspection and peak identification were done manually.

MS3 analysis of glycopeptides

Albumin was immunoprecipitated from pooled serum samples using anti-albumin antibody (Invitrogen) followed by trypsin digestion and MAX-enrichment of glycopeptides. Selected glycopeptides were analyzed in the MS3 mode on an Orbitrap Eclipse mass spectrometer. Precursor ions were detected in the Orbitrap at a resolution of 120,000 with a scan range of 800 to 1500 m/z. Precursor ions were selected and fragmented in the ion-trap using collision induced dissociation (CID). Fragment ions were detected in the ion-trap and selected fragment ions for each precursor were further fragmented using HCD. Data analysis and fragment annotation in MS2 and MS3 spectra was done manually. See Additional file 1: Supplemental Methods for details.

Targeted LC–MS/MS analysis

Glycopeptides derived from 20 volunteer donor serum samples were analyzed in targeted mode on an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) coupled with Ultimate 3000 liquid chromatography system. Inclusion list consisted of precursor ions for all the detected albumin glycopeptides. Data was analyzed using Skyline (v 22.2) [22]. Details are described in the Additional file 1: Supplemental Methods.

Results

We employed a rigorous multi-step LC–MS/MS approach to detect and confirm *N*-glycosylation at the two non-canonical sites of albumin along with attached glycans. First, we performed deep discovery analysis using donor serum samples to identify intact glycopeptides with sites Asn⁶⁸ and Asn¹²³. We then confirmed our findings using streamlined enrichment methods, targeted LC–MS/MS analysis of ¹⁸O-labeled deglycosylated peptides as well as MS3 analysis of intact glycopeptides. These findings were validated in serum samples from twenty additional donors by targeted glycopeptide detection. Further, we show that the highly conserved glycosylation motif at Asn¹²³ is also glycosylated in bovine and rabbit serum albumin.

A novel N-linked glycosylation site on albumin

For initial discovery, we analyzed serum from three volunteer donors using previously described glycoproteomic profiling methods [11]. First, we reduced the complexity of the serum glycoproteome by enriching the most abundant serum proteins using the Human 14 Multiple Affinity Removal (MARS 14) column prior to trypsin digestion. Second, we enriched glycopeptides from

peptide mixtures using size-exclusion chromatography (SEC). Eight fractions from SEC were analyzed using LC–MS/MS-based discovery pipeline [11] (Fig. 1A). The resulting data were searched using pGlyco3 for glycopeptide identification [18]. The search was performed against the UniProt human proteome database and the in-built human *N*-glycan database [13]. On average, 1933 glycopeptides were detected in the three samples. The most abundant glycopeptides were from abundant serum glycoproteins including haptoglobin, alpha-1-acid glycoprotein, immunoglobulin heavy chain and complement C3. These proteins accounted for >80% of the glycopeptide precursor peak areas. *N*-glycopeptides from albumin were detected with glycosylation at both sites Asn⁶⁸ (LVN⁶⁸EVTEFAK) and Asn¹²³ (QEPERN¹²³ECFLQHK, which contains a missed tryptic cleavage site N-terminal to the site of glycosylation). To our knowledge, this is the first report of *N*-glycosylation at Asn¹²³ of albumin. At both sites, complex sialylated *N*-glycans with the following compositions were identified: Hex₅HexNAc₄NeuAc₂, Hex₅HexNAc₄NeuAc₁, Hex₅HexNAc₄NeuAc₂Fuc₁ and Hex₄HexNAc₃NeuAc₁ (Fig. 1B). Two of these glycans, Hex₅HexNAc₄NeuAc₂Fuc₁ and Hex₄HexNAc₃NeuAc₁ have not been reported previously on Asn⁶⁸. To our surprise, albumin-derived glycopeptides accounted for <1% of the total intensity of glycopeptides derived from abundant serum proteins even though albumin is the most abundant serum protein. The relative contribution of individual glycoproteins enriched by MARS 14 to total glycopeptide intensity from these samples is shown in Fig. 1C. We were curious to observe the relationship between the abundance of these proteins and the abundance of corresponding glycopeptides. For comparison, we used protein-level data reported by Geyer et al., 2016, to plot the relative intensities of the same proteins from plasma samples [23]. As shown in Fig. 1C, though albumin accounted for 36% total peptide share among these proteins, it only contributed 1% of the glycopeptide signal. Because *N*-glycosylation occurs more commonly on exposed regions of proteins as compared to internal, more buried regions [24], we examined the location of both glycosylation sites in the three-dimensional structure of albumin. We visualized the crystal structure of albumin from Protein Data Bank and mapped the two *N*-glycosylation sites [20]. As shown in Fig. 1D, both Asn⁶⁸ and Asn¹²³ are located on the surface of the structure of albumin.

Next, we tested an alternate strategy for glycopeptide enrichment for analysis by single MS runs. Peptides from MARS 14-enriched proteins were subjected to glycopeptide enrichment using MAX [17]. LC–MS/MS analysis of enriched samples as a single fraction led to the identification of 409 glycopeptides in each sample on

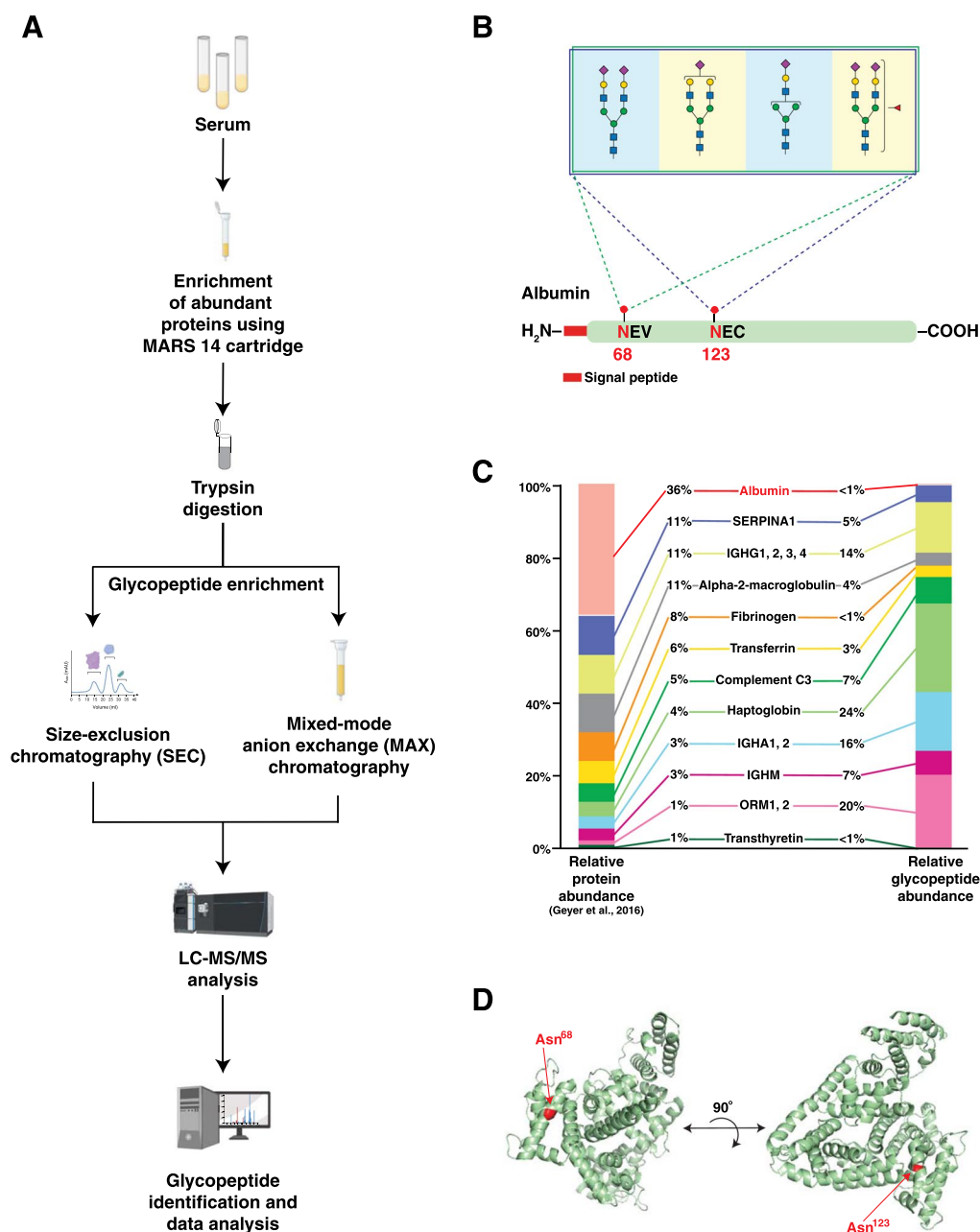


Fig. 1 N-linked glycosylation of albumin and other abundant serum proteins. **A** Experimental strategy for discovery-based analysis of site-specific glycosylation of abundant serum proteins. **B** Representation of glycopeptides identified at Asn⁶⁸ and Asn¹²³ in human albumin with glycans identified at each site (length not drawn to scale). **C** Stacked bar charts to show relative contributions from abundant serum proteins. Relative contribution to total glycopeptide intensity from proteins enriched by MARS 14 column is plotted on the right. The relative abundance levels among the same set of proteins in plasma, i.e., at the protein level, are plotted on the left (glycopeptide data from current study; protein-level data from plasma proteomics experiments, Geyer et al., 2016) [24]. **D** Schematic representation of the crystal structure of albumin highlighting the accessible positions of the two N-glycosylation sites (marked in red)

average. In this method also, the most abundant serum glycoproteins described above accounted for >80% of the glycopeptide precursor peak areas. Glycosylation at both non-canonical glycosylation sites of albumin, i.e.,

Asn⁶⁸ and Asn¹²³ was also detected in all three samples following MAX-enrichment. However, both sites were detected with only two glycans (Hex₅HexNac₄NeuAc₂, Hex₅HexNac₄NeuAc₁) using this method (Additional

file 2: Fig. S1) Glycopeptides detected from SEC- and MAX-enriched samples are listed in Additional file 3: Tables S1 and S2, respectively.

Relative abundance of *N*-glycans on Asn⁶⁸ and Asn¹²³

To determine the relative abundance of the glycopeptides identified from each site, we compared the peak intensity of precursor ions of the glycopeptides detected at each site in the SEC-based experiment. Glycopeptides with glycan compositions Hex₅HexNAC₄NeuAc₂ and Hex₅HexNAC₄NeuAc₁ were the most abundantly detected glycopeptides at both sites (Fig. 2A and B). MS/MS spectra were manually verified for evidence of oxonium ions including signature ions of sialic acid, peptide backbone ions with attached glycan fragments (Y ions) as well as fragments of the naked peptide (*b* and *y* ions) for all glycopeptides mapped to albumin. Annotated MS/MS spectra for glycopeptides from both sites are shown in Fig. 2C, D and (Additional file 2: Fig. S2A–F). These data confidently identify both Asn⁶⁸ and Asn¹²³ as *N*-glycosylation sites while also describing the microheterogeneity at each site.

Confirmation of *N*-linked glycosylation sites

Next, we sought to confirm *N*-glycosylation at sites Asn⁶⁸ and Asn¹²³ of albumin by analyzing enzymatically deglycosylated peptides. Serum proteins were digested using trypsin and glycopeptides were enriched using a MAX column. Glycopeptides were treated with PNGase F using either ¹⁶O or ¹⁸O-labeled water. Deglycosylated peptides were identified considering the mass shift expected after enzymatic removal of the *N*-glycan, which is accompanied by the conversion of asparagine (Asn) to aspartic acid (Asp) [25]. Deglycosylated Asn residues were identified with conversion to Asp showing a mass difference of 0.98 Da in case of ¹⁶O incorporation and 2.98 Da in case of ¹⁸O incorporation.

The non-glycosylated peptide with Asn⁶⁸ (LVN⁶⁸EVTEFAK) was identified with a charge state of +2 with *m/z* of 575.31. Upon treatment with PNGase F in ¹⁶O water, we detected the deglycosylated form of the formerly *N*-glycosylated peptide with a mass shift of 0.98 Da or 0.5 *m/z* (LVD⁶⁸EVTEFAK, *m/z* of 575.80, Fig. 3A). In samples treated with PNGase F in ¹⁸O-labeled water, we observed a mass shift of 2.98 Da or 1.5 *m/z*, corresponding to the deglycosylated peptide (LVD⁶⁸EVTEFAK, *m/z* of 576.81 *m/z*, Fig. 3B). The partial overlap of peaks from the ¹⁶O-labeled peptides with the ¹⁸O-labeled peptides is explained by the natural abundance of isotopes and purity of ¹⁸O-labeled water used [26] (Fig. 3B). This analysis demonstrates enzymatic deglycosylation of Asn⁶⁸, conclusively showing albumin glycosylation at this site.

Similarly, we detected the non-glycosylated peptide containing Asn¹²³ (QEPERN¹²³ECFLQHK) with a charge state of +3 and *m/z* of 572.27. Upon treatment with PNGase F in ¹⁶O water, we identified the deglycosylated form of the peptide (QEPERD¹²³ECFLQHK, *m/z* of 572.60 *m/z*) as depicted in Fig. 3C. With ¹⁸O incorporation, we observed the deamidated form QEPERD*¹²³ECFLQHK at the *m/z* of 573.25 *m/z* (Fig. 3D). This confirms glycosylation at Asn¹²³.

Confirmation of albumin glycopeptides by MS3 fragmentation

To further enhance the confidence in the identification of albumin-derived glycopeptides, we performed MS3 analysis using an Orbitrap Eclipse Tribrid mass spectrometer which incorporates a high-sensitivity ion-trap detector. Albumin was immunoprecipitated from pooled serum and glycopeptides were enriched by MAX. Precursor ions corresponding to four albumin-derived glycopeptides were isolated and fragmented using collision-induced dissociation (CID) followed by their detection in the ion-trap. At low collision energy, glycosidic bonds were expected to break forming ions consisting of the peptide backbone carrying glycan fragments (Y ions). Selected Y ions were fragmented at the MS3 level using higher-energy collisional dissociation (HCD) followed by detection in the ion-trap. MS3 fragmentation produced glycan oxonium ions confirming the presence of glycopeptides, as well as further fragments of the Y ions. The resulting spectra were manually inspected and annotated (Fig. 4).

The precursor ions selected included the two most abundant glycopeptides at each glycosylation site, i.e., LVN⁶⁸EVTEFAK with Hex₅HexNAC₄NeuAc₁ (*m/z*=1021.7, charge state +3), LVN⁶⁸EVTEFAK with Hex₅HexNAC₄NeuAc₂ (*m/z*=1118.8, charge state +3), QEPERN¹²³ECFLQHK with Hex₅HexNAC₄NeuAc₁ (*m/z*=907.8, charge state +4), and QEPERN¹²³ECFLQHK with Hex₅HexNAC₄NeuAc₂ (*m/z*=980.6, charge state +4). As expected, prominent product ions generated from low energy CID fragmentation at MS2 level were glycopeptide Y ions (Fig. 4). Notably, we also detected singly charged oxonium ions (albeit with lower intensity) at *m/z* values of 274.0 (NeuAc with water loss), 292.1 (NeuAc), 366.1 (HexNAC and Hex), and 657.2 (HexNAC, Hex, and NeuAc), further confirming the presence of glycopeptides (as depicted in Fig. 4). Subsequently, fragment Y ions for each precursor ion generated at the MS/MS level underwent further fragmentation via HCD, yielding diagnostic MS3 fragment ions. The ion series with the serial loss of single monosaccharide residues validated the glycan composition of these glycopeptides. Further, the glycan oxonium ions at the MS3 level were detected

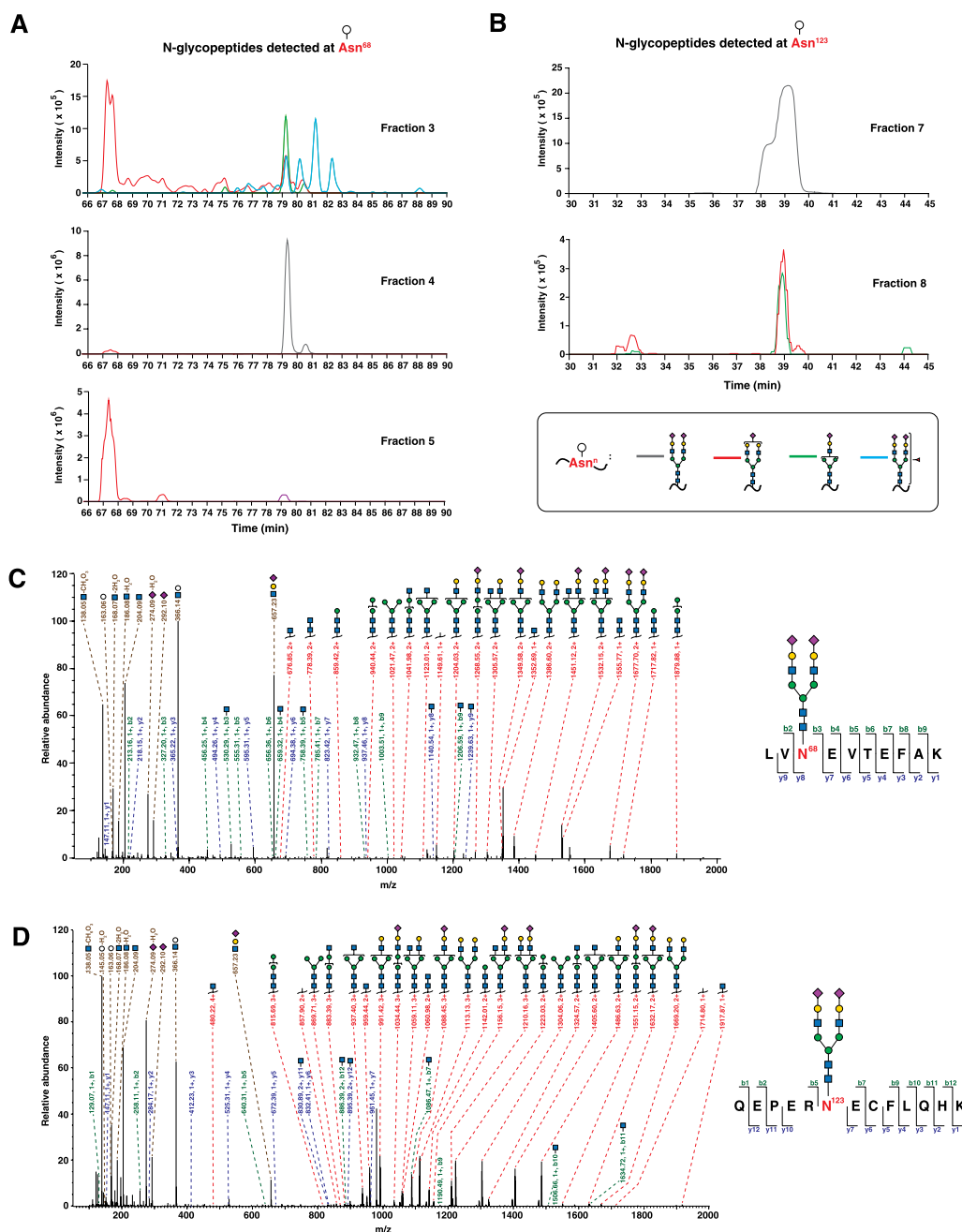


Fig. 2 Abundance and identification of albumin-derived glycopeptides. **A** Extracted ion chromatograms (XIC) showing relative abundance of glycopeptides corresponding to Asn⁶⁸ detected in different fractions of an individual sample from size-exclusion chromatography (SEC). In fraction 3, glycopeptides bearing the glycan Hex₅HexNAc₄NeuAc₂ at this site (represented by a grey line in other fractions) were identified with peak intensity of ~5 × 10⁷ at 79.3 min. To clearly depict the lower-abundance glycopeptides with other compositions which would otherwise be lost to scale, we omitted the XIC of the glycopeptide bearing the glycan Hex₅HexNAc₄NeuAc₂ at this site in fraction 3. **B** XICs showing the relative abundance of glycopeptides corresponding to Asn¹²³ detected in different fractions of an individual sample from SEC. In fraction 8, glycopeptides bearing the glycan Hex₅HexNAc₄NeuAc₂ at this site (represented by a grey line in other fractions) were identified with a peak intensity of ~3 × 10⁷ at 38.9 min. To clearly depict the lower-abundance glycopeptides with other compositions which would otherwise be lost to scale, we omitted the XIC of the glycopeptide bearing the glycan Hex₅HexNAc₄NeuAc₂ at this site in fraction 8. **C, D** Annotated MS/MS fragmentation spectra of representative glycopeptides derived from albumin with the glycan Hex₅HexNAc₄NeuAc₂ at sites Asn⁶³ and Asn¹²³, respectively

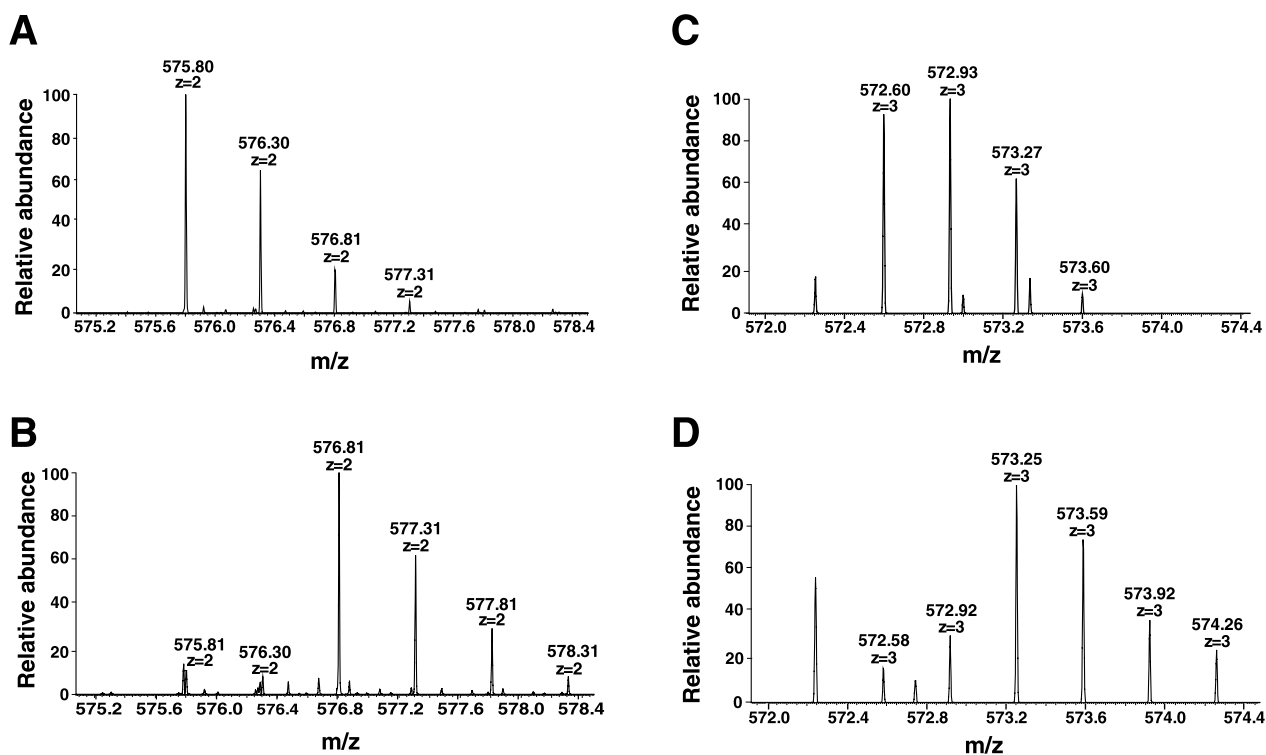


Fig. 3 Mass spectra showing the detection of deglycosylated peptides of albumin after treatment with PNGase F in stable isotope-labeled water. **A, B** Precursor mass spectra of the albumin-derived glycopeptide with glycosylation at Asn⁶⁸, detected in a charge state of +2 after deglycosylation by PNGase F treatment in the presence of H₂¹⁶O with a mass shift corresponding to 0.98 Da (**A**) or in the presence of H₂¹⁸O with a mass shift corresponding to 2.98 Da (**B**). **C, D** Precursor mass spectra of the albumin-derived glycopeptide with glycosylation at Asn¹²³, detected in a charge state of +3 after deglycosylation by PNGase F treatments in H₂¹⁶O with a mass shift corresponding to 0.98 Da (**C**) and in H₂¹⁸O with a mass shift corresponding to 2.98 Da (**D**)

with higher intensities, confirming the presence of glycopeptides. Spectra resulting upon fragmentation of precursor ions with m/z of 1118.8 (charge state +3) and 980.6 (charge state +4) with peptide sequence and glycan composition mentioned above are shown in Fig. 4A and 4B respectively.

Albumin glycosylation in a larger cohort of volunteer donors

To assess if glycosylation of albumin is a general phenomenon and validate our findings, we analyzed serum samples from twenty volunteer donors by targeted MS. Eight albumin-derived glycopeptides identified in the discovery experiment were targeted, i.e., glycopeptides with sequences LVN⁶⁸EVTEFAK and QEPERN¹²³ECFLQHK, each bearing one of four glycans, Hex₅HexNAc₄NeuAc₂, Hex₅HexNAc₄NeuAc₁, Hex₅HexNAc₄NeuAc₂Fuc₁ and Hex₄HexNAc₃NeuAc₁. MAX-enriched *N*-glycopeptides from serum proteins were analyzed by parallel reaction monitoring-mass spectrometry (PRM-MS). In all the twenty individuals that were tested, we detected glycosylation at both Asn⁶⁸ and Asn¹²³ of albumin. The

heterogeneity in the overall glycopeptide complement detected among the individuals is shown in Table S3 (Additional file 3).

Albumin glycosylation in other species

Because albumin is a highly conserved protein, we were curious if its orthologs in other mammalian species are also glycosylated. Examining the amino acid sequences of albumin orthologs from cow, rabbit, dog and mouse revealed that only albumin from mouse has canonical Asn-X-Ser/Thr motifs, but without annotation for *N*-linked glycosylation on UniProt [13]. However, these orthologs have multiple non-canonical *N*-glycosylation motifs. Multiple sequence alignment showed that the non-canonical motif Asn¹²³-Glu-Cys, is highly conserved, whereas site Asn⁶⁸ is not an evolutionarily conserved glycosylation site or amino acid (Fig. 5A). Therefore, to test if this site is glycosylated in other species, we analyzed bovine serum albumin (BSA) and rabbit serum albumin, which are commonly used in molecular biology and MS applications. Commercially available BSA was digested using trypsin followed by MAX-based enrichment of

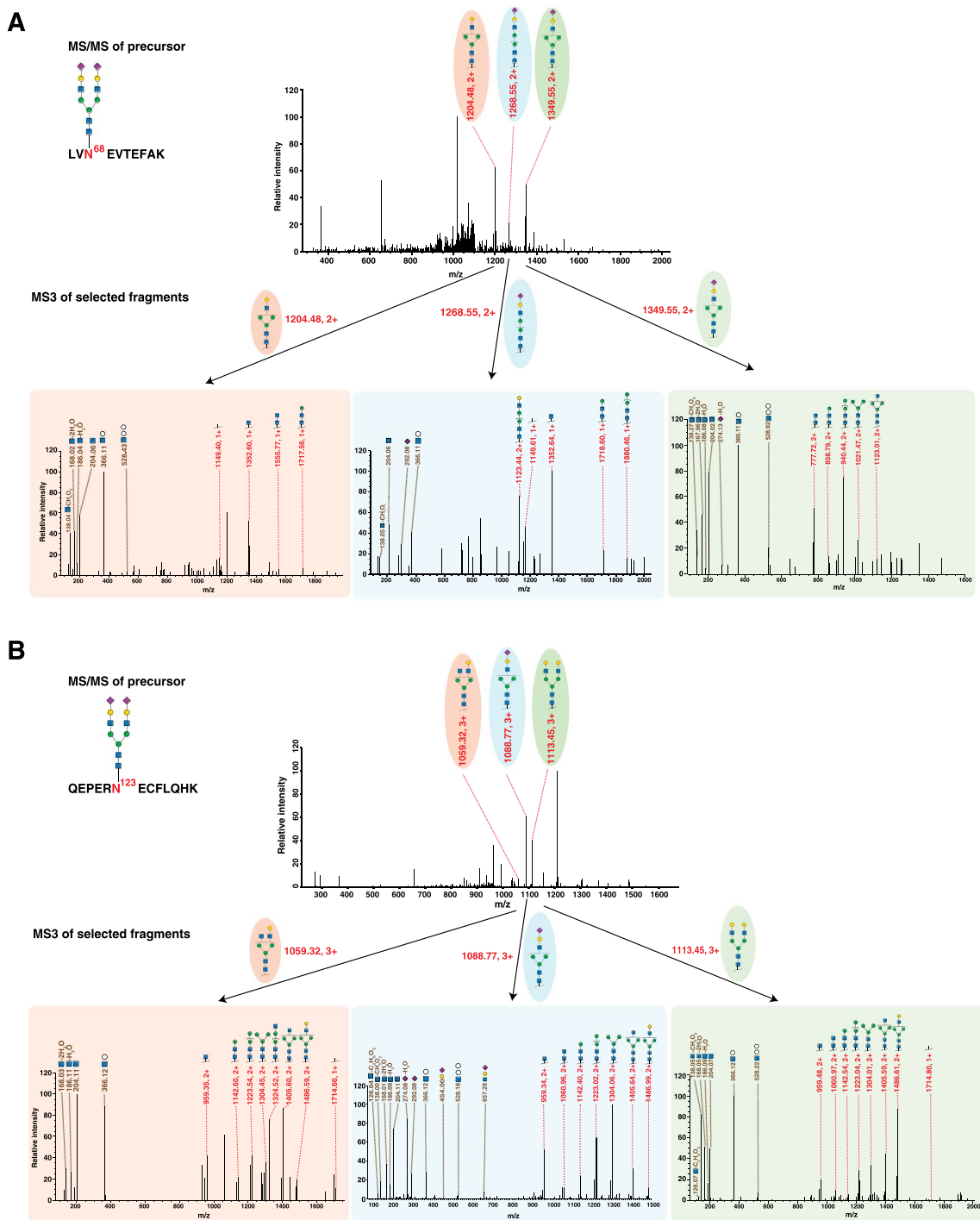


Fig. 4 MS3 analysis of albumin-derived glycopeptides. Precursor fragmentation is shown along with fragmentation of MS/MS-derived fragments using MS3. **A** Glycopeptide LNVN⁶⁸EVTEFAK with Hex₅HexNAC₄NeuAc₂ (m/z = 1118.8, charge state +3). The MS/MS fragmentation of the precursor glycopeptide is shown in the top half of the panel. The three annotated fragment Y ions (peptide backbone with glycan fragments still attached to it) from this scan were selected and further fragmented by MS3. Labeled arrows under the peak of each selected fragment ion in the MS/MS scan indicate MS3 fragmentation spectra for the corresponding Y ions. **B** Glycopeptide QEPERN¹²³ECFLQHK with Hex₅HexNAC₄NeuAc₂ (m/z = 980.6, charge state +4). The MS/MS fragmentation of the precursor glycopeptide is shown in the top half of the panel. The representations of the MS/MS spectrum, selected fragments and their MS3 spectra are as described in A

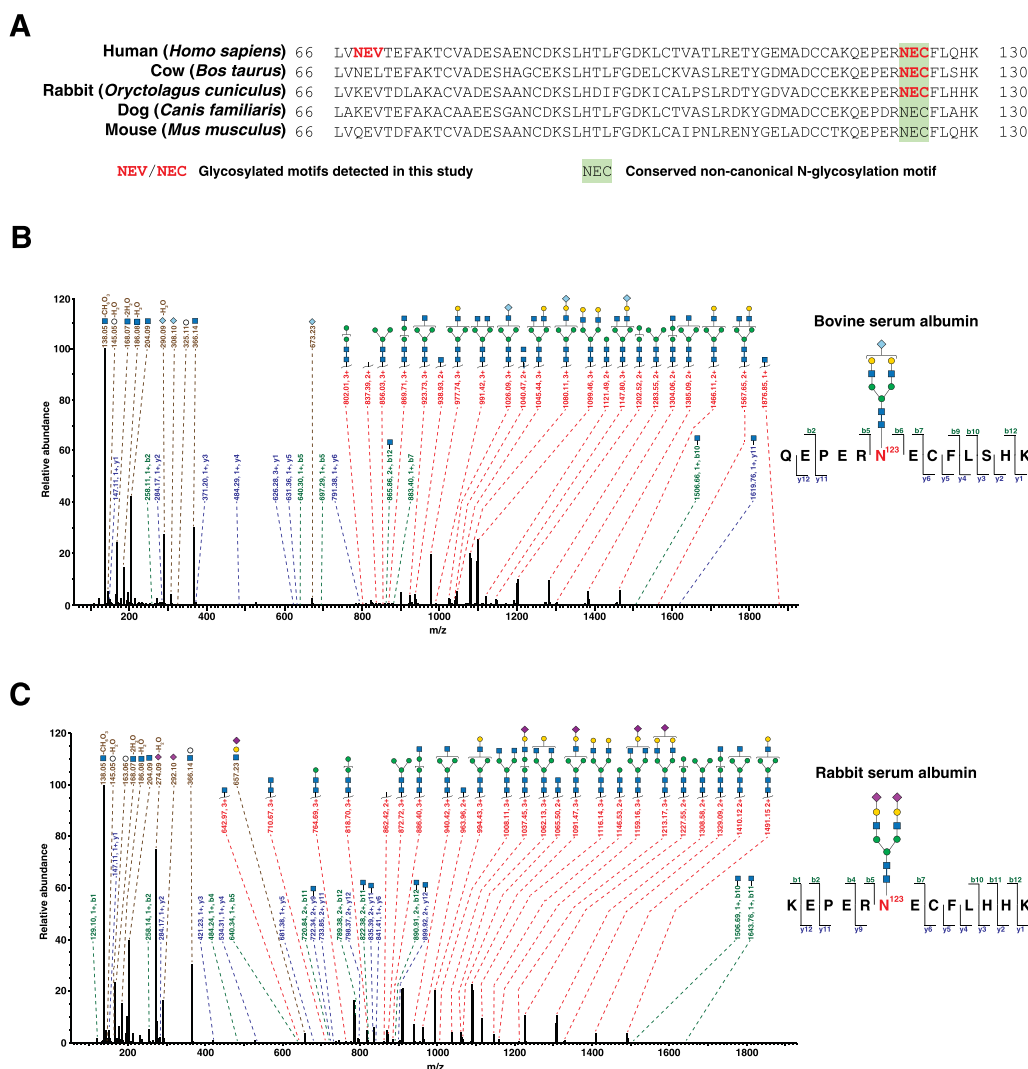


Fig. 5 Glycosylation of albumin orthologs in other mammalian species. **A** Multiple sequence alignment of the region of human albumin containing the two N-glycosylation sites with orthologs from selected mammalian species. The conserved non-canonical N-glycosylation motif with Asn¹²³ is shown highlighted in green. Non-canonical N-glycosylation motifs that were detected with glycosylation in this study are shown in red font. **B** Annotated MS/MS fragmentation spectrum of glycopeptide derived from bovine serum albumin (BSA) at site Asn¹²³ with the glycan Hex₅HexNAc₄NeuGc₁. **C** Annotated MS/MS fragmentation spectrum of glycopeptide derived from rabbit serum albumin at site Asn¹²³ with the glycan Hex₅HexNAc₄NeuAc₂, with annotations as described in **B**

glycopeptides followed by LC–MS/MS analysis for glycopeptide discovery. Database searching for glycopeptides was done using pGlyco3 with the UniProt bovine proteome database for peptide sequences. As bovine N-glycans are similar in composition to human N-glycans except for the presence of an additional sialic acid (N-glycolylneuraminic acid or NeuGc) which is also present in mouse, we used the in-built mouse N-glycan database for this search [27]. We detected BSA-derived glycopeptides with Asn¹²³ glycosylated by three complex sialylated glycans, i.e., Hex₅HexNAc₄NeuGc₁, Hex₅HexNAc₄NeuAc₁NeuGc₁ and Hex₅HexNAc₄NeuAc₁NeuGc₁ (Fig. 5, Additional

file 2: Fig. S3A and S3B respectively). Interestingly, besides glycosylation at the conserved site Asn¹²³, we also detected glycopeptides from BSA with glycosylation at Asn¹⁸⁵ with two glycans, Hex₅HexNAc₄NeuAc₂ and Hex₅HexNAc₄NeuAc₁NeuGc₁ (Additional file 2: Fig. S3C and S3D respectively). However, this non-canonical glycosylation site, which is in the motif Asn¹⁸⁵-Gly-Val, is not conserved across the species listed above. Additionally, in a separate experiment performed identically but with commercially available rabbit serum albumin and searched against the rabbit proteome and mouse N-glycan database, the conserved non-canonical

N-glycosylation site Asn¹²³ was also detected with two complex sialylated *N*-glycans, i.e., Hex₅HexNAc₄NeuAc₂ and Hex₅HexNAc₄NeuAc₁ (Fig. 5C; Additional file 2: Fig. S3E respectively). Overall, these data provide evidence for the glycosylation of albumin at the conserved non-canonical *N*-glycosylation site orthologous to Asn¹²³ of human albumin in two additional mammalian species. Glycopeptides detected in bovine and rabbit serum albumin are listed in Additional file 3: Tables S4 and S5 respectively.

Discussion

Although most abundant serum proteins are glycoproteins, albumin itself has been considered a notable exception until recently [10]. Through discovery analysis and rigorous testing using different enrichment strategies [11, 15] and high-resolution LC-MS/MS methods, we report a novel *N*-glycosylation site on albumin (Asn¹²³) and expand the glycan heterogeneity on another site (Asn⁶⁸). Effective enrichment strategies are key to MS-based identification of glycopeptides owing to glycan heterogeneity [15, 28]. In the discovery experiments, SEC, which is based on physical properties and used here as a method for simultaneous enrichment and fractionation, resulted in identification of three times more glycopeptides in comparison to the single MS runs after MAX-based enrichment. Albumin glycopeptides at sites Asn⁶⁸ and Asn¹²³ were identified by both methods. Interestingly, though albumin is the most abundant plasma protein, glycopeptides from albumin accounted for <1% of identified glycopeptide precursor peak areas, indicating low site occupancy (Fig. 1C). This follows our expectation based on previous reports on other proteins that non-canonical *N*-glycosylation motifs have lower stoichiometry of glycosylation [8, 9]. We also show that Asn¹²³, which occurs within a highly conserved Asn-Glu-Cys motif is also glycosylated in bovine and rabbit serum albumin. In the case of BSA, we detected two glycopeptides containing NeuGc, a sialic acid that is not present in humans because the gene encoding an essential synthetic enzyme, cytidine monophosphate-*N*-acetylneuraminic acid hydroxylase (*CMAH*), is inactive in humans [29]. Though BSA is routinely used as a tool for quality control for MS, we believe that its glycosylation has generally been missed previously because of the absence of suspicion owing to lack of a consensus *N*-glycosylation motif.

Physiologically, albumin is involved in several functions including binding and transportation of molecules such as fatty acids, hormones, drugs, vitamins and metal ions [30, 31]. These ligand-binding and antioxidant functions of albumin are influenced by its various post-translational modifications (PTMs) [30] including cysteinylolation, oxidation and nitrosylation [31]. Additionally, glycation is

present at 20–30% in circulating albumin in hyperglycemic individuals, and this modification alters its binding properties [2, 32]. Traditional methods of protein analysis, e.g., isoelectric focusing (IEF) and two-dimensional gel electrophoreses (2DE) did not raise any suspicions of glycosylation of albumin on record, even though some such studies report separation of albumin into fractions based on isoelectric point [33]. In light of the current report, we wonder if the smears and unexplained spots annotated for albumin on IEF and 2DE experiments may be explained, at least in part, by albumin *N*-glycoforms [33, 34]. Additional studies may determine functional effects of glycosylation on the ligand-binding and antioxidant properties of albumin, along with its susceptibility to undergo other PTMs [35]. For example, Cys¹²⁵, which is the C-terminal amino acid in the motif that Asn¹²³ is part of (Asn¹²³-Glu¹²⁴-Cys¹²⁵), participates in the formation of a disulfide bridge in the secondary structure of albumin [36]. It has been previously shown that degree of glycosylation at sites in Asn-X-Cys motifs is likely related to the rate of translation as well as the rate of disulfide bond formation [7]. Hence, the rate of glycosylation at Asn¹²³ may be altered in states such as liver disease and metabolic syndrome where liver function is affected [37].

Conclusions

To conclude, we report that albumin is a glycoprotein with multiple N-linked glycoforms at two non-canonical sites. As these findings are discordant with the long-held notion that albumin is a non-glycosylated protein, we confirmed them by multiple additional lines of investigation. Serum albumin level is used as a marker for several diseases including renal, hepatic and cardiovascular disorders [38]. Pathological modifications of albumin including glycation and cysteinylolation are also associated with diabetes and liver disease [39]. In fact, glycated albumin has been shown to complement glycated hemoglobin as a marker of prediabetes [40]. Given this importance of albumin in clinical practice, glycosylated albumin could also have clinical significance. Indeed, we have recently found reduced levels of the glycopeptide bearing Hex₅HexNAc₄NeuAc₁ at Asn¹²³ in patients with a congenital disorder of glycosylation (CDG) [41]. This indicates that glycosylation events on albumin could potentially be of diagnostic or other clinical uses. Future studies may determine the exact role of glycosylation of albumin and how it is altered in other diseases associated with altered protein glycosylation. Our findings alter the prevailing paradigm by showing that albumin is not a non-glycosylated protein and may expand our understanding of its structure and function, and its clinical and biochemical applications.

Abbreviations

| | |
|------------------|--|
| MARS 14 | Multiple affinity removal spin cartridge human-14 |
| SEC | Size-exclusion chromatography |
| MAX | Mixed-mode anion exchange |
| LC-MS/MS | Liquid chromatography-tandem mass spectrometry |
| PRM | Parallel reaction mode |
| Asn ^x | Asparagine at amino acid site x in a polypeptide sequence |
| PTM | Post-translational modification |
| MS | Mass spectrometry |
| TEABC | Triethylammonium bicarbonate |
| DTT | Dithiothreitol |
| IAA | Iodoacetamide |
| TFA | Trifluoroacetic acid |
| FA | Formic acid |
| ACN | Acetonitrile |
| DDA | Data-dependent analysis |
| AGC | Automatic gain control |
| HCD | Higher-energy collisional dissociation |
| FDR | False-discovery rate |
| MS/MS | Tandem mass spectrometry |
| BSA | Bovine serum albumin |
| PDB | Protein Data Bank |
| PBS | Phosphate-buffered saline |
| CID | Collision-induced dissociation |
| Hex | Hexose |
| HexNAc | N-Acetylhexosamine |
| NeuAc | N-Acetylneuraminic acid |
| Fuc | Fucose |
| NeuGc | N-Glycolylneuraminic acid |
| SNFG | Graphical representations of glycans are made using Symbol Nomenclature for Glycans [42] |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-024-05000-5>.

Additional file 1: Supplemental Methods. Additional experimental details, materials and methods.

Additional file 2: Additional figures. Additional supporting figures providing additional information on the glycopeptides identified by discovery analysis of human serum, bovine serum albumin and rabbit serum albumin samples.

Additional file 3: Additional Tables. Additional information on glycopeptides identified by SEC- and MAX-based enrichment; albumin-derived glycopeptides identified in additional volunteer donor samples; glycopeptides identified from bovine and rabbit serum albumin samples.

Acknowledgements

We thank Kiran B. Gaikwad for help with structural analysis and Richard K. Kandasamy for helpful discussions.

Author contributions

KG, AJ and AP conceived and designed the study. KG and AJ performed experiments. KG, AJ, BJM, DGM, and RB analyzed the data. KG, AJ, JS and AP wrote the manuscript. KG, AJ and JS made the figures. All authors read and reviewed the manuscript.

Funding

We thank Mayo Clinic DERIVE Office and Mayo Clinic Center for Biomedical Discovery for financial support and a grant from DBT/Wellcome Trust India Alliance entitled "Center for Rare Disease Diagnosis, Research, and Training" (IA/CRC/20/1/600002) to AP.

Availability of data and materials

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [43] partner repository with the dataset identifier PXD047863.

Declarations

Ethics approval and consent to participate

Serum samples used in this study were deidentified residual samples from volunteer donors collected with consent (approved by Mayo Clinic IRB: 21-012890).

Consent for publication

Not applicable.

Competing interests

All authors declare no conflicts of interest.

Received: 12 January 2024 Accepted: 14 February 2024

Published online: 13 May 2024

References

- Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 2002;12(4):43R-56R.
- Rondeau P, Bourdon E. The glycation of albumin: structural and functional impacts. *Biochimie*. 2011;93(4):645-58.
- Schjoldager KT, Narimatsu Y, Joshi HJ, Clausen H. Global view of human protein glycosylation pathways and functions. *Nat Rev Mol Cell Biol*. 2020;21(12):729-49.
- Stanley P, Moremen KW, Lewis NE, Taniguchi N, Aebi M. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor (NY) 2022, 103-16.
- Bause E, Legler G. The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the N-glycosylation step during glycoprotein biosynthesis. *Biochem J*. 1981;195(3):639-44.
- Hulsmeier AJ, Tobler M, Burda P, Hennen T. Glycosylation site occupancy in health, congenital disorder of glycosylation and fatty liver disease. *Sci Rep*. 2016;6:33927.
- Lowenthal MS, Davis KS, Formolo T, Kilpatrick LE, Phinney KW. Identification of novel N-glycosylation sites at noncanonical protein consensus motifs. *J Proteome Res*. 2016;15(7):2087-101.
- Canis K, McKinnon TA, Nowak A, Haslam SM, Panico M, Morris HR, et al. Mapping the N-glycome of human von Willebrand factor. *Biochem J*. 2012;447(2):217-28.
- Satomi Y, Shimonishi Y, Takao T. N-glycosylation at Asn(491) in the Asn-Xaa-Cys motif of human transferrin. *FEBS Lett*. 2004;576(1-2):51-6.
- Sun S, Hu Y, Jia L, Eshghi ST, Liu Y, Shah P, et al. Site-specific profiling of serum glycoproteins using N-linked glycan and glycosite analysis revealing atypical N-glycosylation sites on albumin and alpha-1B-glycoprotein. *Anal Chem*. 2018;90(10):6292-9.
- Saraswat M, Garapati K, Mun DG, Pandey A. Extensive heterogeneity of glycopeptides in plasma revealed by deep glycoproteomic analysis using size-exclusion chromatography. *Mol Omics*. 2021;17(6):939-47.
- Zielinska DF, Gnad F, Wisniewski JR, Mann M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell*. 2010;141(5):897-907.
- UniProt C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49(D1):D480-9.
- Quinlan GJ, Martin GS, Evans TW. Albumin: biochemical properties and therapeutic potential. *Hepatology*. 2005;41(6):1211-9.
- Riley NM, Bertozzi CR, Pitteri SJ. A pragmatic guide to enrichment strategies for mass spectrometry-based glycoproteomics. *Mol Cell Proteomics*. 2021;20: 100029.
- Budhraja R, Saraswat M, De Graef D, Ranatunga W, Ramarajan MG, Mousa J, et al. N-glycoproteomics reveals distinct glycosylation alterations in NGLY1-deficient patient-derived dermal fibroblasts. *J Inher Metab Dis*. 2023;46(1):76-91.
- Yang W, Shah P, Hu Y, Toghi Eshghi S, Sun S, Liu Y, et al. Comparison of enrichment methods for intact N- and O-linked glycopeptides using strong anion exchange and hydrophilic interaction liquid chromatography. *Anal Chem*. 2017;89(21):11193-7.

18. Zeng WF, Cao WQ, Liu MQ, He SM, Yang PY. Precise, fast and comprehensive analysis of intact glycopeptides and modified glycans with pGlyco3. *Nat Methods*. 2021;18(12):1515–23.
19. Sugio S, Kashima A, Mochizuki S, Noda M, Kobayashi K. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng*. 1999;12(6):439–46.
20. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The protein data bank. *Acta Crystallogr D Biol Crystallogr*. 2002;58(Pt 6 No 1):899–907.
21. Schrodinger, LLC. The PyMOL molecular graphics system, Version 1.8. 2015.
22. Pino LK, Searle BC, Bollinger JG, Nunn B, MacLean B, MacCoss MJ. The Skyline ecosystem: informatics for quantitative mass spectrometry proteomics. *Mass Spectrom Rev*. 2020;39(3):229–44.
23. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma proteome profiling to assess human health and disease. *Cell Syst*. 2016;2(3):185–95.
24. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology*. 2004;14(2):103–14.
25. Kuster B, Mann M. 18O-labeling of N-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. *Anal Chem*. 1999;71(7):1431–40.
26. Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol*. 2003;21(6):667–72.
27. Nwosu CC, Aldredge DL, Lee H, Lerno LA, Zivkovic AM, German JB, et al. Comparison of the human and bovine milk N-glycome via high-performance microfluidic chip liquid chromatography and tandem mass spectrometry. *J Proteome Res*. 2012;11(5):2912–24.
28. Bagdonaitė I, Malaker SA, Polasky DA, Riley NM, Schjoldager K, Vakhru-shev SY, et al. Glycoproteomics. *Nat Rev Methods Primers*. 2022;2(1):48.
29. Chou HH, Takematsu H, Diaz S, Iber J, Nickerson E, Wright KL, et al. A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proc Natl Acad Sci USA*. 1998;95(20):11751–6.
30. Fasano M, Curry S, Terreno E, Galliano M, Fanali G, Narciso P, et al. The extraordinary ligand binding properties of human serum albumin. *IUBMB Life*. 2005;57(12):787–96.
31. Rahali MA, Lakis R, Sauvage FL, Pinault E, Marquet P, Saint-Marcoux F, et al. Posttranslational-modifications of human-serum-albumin analysis by a top-down approach validated by a comprehensive bottom-up analysis. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2023;1224: 123740.
32. Fanali G, di Masi A, Trezza V, Marino M, Fasano M, Ascenzi P. Human serum albumin: from bench to bedside. *Mol Aspects Med*. 2012;33(3):209–90.
33. Chromy BA, Gonzales AD, Perkins J, Choi MW, Corzett MH, Chang BC, et al. Proteomic analysis of human serum by two-dimensional differential gel electrophoresis after depletion of high-abundant proteins. *J Proteome Res*. 2004;3(6):1120–7.
34. Ong SE, Pandey A. An evaluation of the use of two-dimensional gel electrophoresis in proteomics. *Biomol Eng*. 2001;18(5):195–205.
35. Zacchi LF, Schulz BL. N-glycoprotein macroheterogeneity: biological implications and proteomic characterization. *Glycoconj J*. 2016;33(3):359–76.
36. Bocedi A, Cattani G, Stella L, Massoud R, Ricci G. Thiol disulfide exchange reactions in human serum albumin: the apparent paradox of the redox transitions of Cys(34). *FEBS J*. 2018;285(17):3225–37.
37. Levitt DG, Levitt MD. Human serum albumin homeostasis: a new look at the roles of synthesis, catabolism, renal and gastrointestinal excretion, and the clinical value of serum albumin measurements. *Int J Gen Med*. 2016;9:229–55.
38. Ballmer PE. Causes and mechanisms of hypoalbuminaemia. *Clin Nutr*. 2001;20(3):271–3.
39. Domenicali M, Baldassarre M, Giannone FA, Naldi M, Mastroberto M, Biselli M, et al. Posttranscriptional changes of serum albumin: clinical and prognostic significance in hospitalized patients with cirrhosis. *Hepatology*. 2014;60(6):1851–60.
40. Sumner AE, Duong MT, Bingham BA, Aldana PC, Ricks M, Mabundo LS, et al. Glycated albumin identifies prediabetes not detected by hemoglobin A1c: the Africans in America Study. *Clin Chem*. 2016;62(11):1524–32.
41. Garapati K, Budhraj R, Saraswat M, Kim J, Joshi N, Sachdeva GS, et al. A complement C4-derived glycopeptide as a biomarker for PMM2-CDG. *JCI Insight*. 2024;In Press.
42. Neelamegham S, Aoki-Kinoshita K, Bolton E, Frank M, Lisacek F, Lutteke T, et al. Updates to the symbol nomenclature for Glycans guidelines. *Glycobiology*. 2019;29(9):620–4.
43. Perez-Riverol Y, Bai J, Bandla C, Garcia-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res*. 2022;50(D1):D543–52.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.