N- and O-glycosylation analysis of human C1-inhibitor reveals extensive mucin-type O-glycosylation

Kathrin Stavenhagen1,2,*, H. Mehmet Kayili2,3,4,5+, Stephanie Holst1, Carolien A. M. Koeleman1, Ruchira Engel6,7, Diana Wouters6,7, Sacha Zeerleder6,7, Bekir Salih5, Manfred Wuhrer1,2

1 Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands
2 Division of BioAnalytical Chemistry, VU University Amsterdam, Amsterdam, The Netherlands
3 Department of Chemistry, Çankırı Karatekin University, Çankırı, Turkey
4 Department of Nutrition and Dietetics, Karabuk University, Karabuk, Turkey
5 Department of Chemistry, Hacettepe University, Ankara, Turkey
6 Department of Immunopathology, Sanquin Research and Landsteiner Laboratory of the AMC, Amsterdam, The Netherlands
7 Department of Hematology, Academic Medical Center, University of Amsterdam, The Netherlands
+ Authors equally contributed to this work
* Corresponding authors

Corresponding author contact information:
Phone: +31 71 5268701; Fax: +31 71 5266907; E-mail: kathrin.stavenhagen@gmail.com

Running title: Human C1-Inhibitor glycosylation
Abbreviations:

AmBiC: ammonium bicarbonate;
C1-Inh: C1-Inhibitor;
DHB: 2,5-dihydroxybenzoic acid;
ENeuAc: α2,6-linked N-acetylneuraminic acid;
EDC: 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride;
ETD: electron-transfer dissociation;
FA: formic acid;
Fuc: fucose;
GalNAc: N-acetylgalactosamine;
GlcNAc: N-acetylglucosamine;
Hereditary angioedema: HAE;
Hex: hexose;
HexNAc: N-acetylhexosamine;
HILIC-SPE: hydrophilic interaction liquid chromatography-solid phase extraction;
HOBt: hydroxybenzotriazol hydrate;
IAA: iodoacetamide;
IPQ score: isotopic pattern quality score
IT: ion trap;
LNeuAc: α2,3-linked N-acetylneuraminic acid;
Na₂HPO₄·2H₂O: sodium hydrogen phosphate dihydrate;
NaOH: sodium hydroxide;
NeuAc: N-acetylneuraminic acid;
NP-40: Nonidet P-40;
PCG: porous graphitized carbon
PNGase F: peptide-N-glycosidase F;
RP: reversed-phase;
sDHB: 2-hydroxy-5-methoxy-benzoic acid and 2,5-dihydroxybenzoic acid, 1:9;
Serpin: serine protease inhibitor;
SPE: solid phase extraction
Summary

Human C1-inhibitor (C1-Inh) is a serine protease inhibitor and the major regulator of the contact activation pathway as well as the classical and lectin complement pathways. It is known to be a highly glycosylated plasma glycoprotein. However, both the structural features and biological role of C1-Inh glycosylation are largely unknown. Here, we performed for the first time an in-depth site-specific N- and O-glycosylation analysis of C1-Inh combining various mass spectrometric approaches, including C18-porous graphitized carbon (PGC)-LC-ESI-QTOF-MS/MS applying stepping-energy collision-induced dissociation (CID) and electron-transfer dissociation (ETD). Various proteases were applied, partly in combination with PNGase F and exoglycosidase treatment, in order to analyze the (glyco)peptides. The analysis revealed an extensively O-glycosylated N-terminal region. Five novel and five known O-glycosylation sites were identified, carrying mainly core1-type O-glycans. In addition, we detected a heavily O-glycosylated portion spanning from Thr82-Ser121 with up to 16 O-glycans attached. Likewise, all known six N-glycosylation sites were covered and confirmed by this site-specific glycosylation analysis. The glycoforms were in accordance with results on released N-glycans by MALDI-TOF/TOF-MS/MS. The comprehensive characterization of C1-Inh glycosylation described in this study will form the basis for further functional studies on the role of these glycan modifications.
Introduction

Human C1-inhibitor (C1-Inh) is a serine protease inhibitor (serpin) and the major regulator of the contact activation pathway via inhibition of factor XIIa, kallikrein and factor XIa, as well as the classical and lectin complement pathways via C1s, C1r, and MASP (1). A C1-Inh deficiency, either due to decreased or dysfunctional expression, is associated with hereditary angioedema (HAE), which results in vascular permeability causing tissue swelling (1). For treatment and prophylaxis of HAE human plasma-derived C1-Inh replacement therapy is commonly applied (2). Also other diseases, such as inflammatory diseases, sepsis and endotoxic shock, may be targeted by C1-Inh therapy (1, 3). To meet this need, recombinant C1-Inh formats are currently being developed (4–7).

C1-Inh is considered as one of the most heavily glycosylated proteins in human plasma (1). The protein consists of 478 amino acids and the calculated molecular mass of C1-Inh is approximately 55 kDa without glycans, while a much higher apparent molecular mass was observed on SDS-PAGE (> 80kDa) due to its heavy glycosylation (8, 9). It has been reported that C1-Inh possesses six occupied N- and up to 25 O-glycosylation sites (8, 10–13). Of the latter ones nine have been identified with their exact location (10, 11, 13). The protein consists of two domains: i) the C-terminal domain (serpin domain), which carries three of the six N-glycosylation sites, provides the inhibition activity of C1-Inh and is similar to other serpins; and ii) the N-terminal domain, which consists of approximately 135-142 amino acid residues (approximately 113-120 amino acids in the mature protein), featuring the remaining three N- and all O-glycosylation sites (11, 14). Despite the fact that protein glycosylation has a large impact on biological processes, protein stability, and protein functions (10, 15, 16), the structural features as well as biological role of C1-Inh glycosylation is still largely unknown.

To address this, we here present a detailed site-specific N- and O-glycosylation characterization of plasma derived C1-Inh using a panel of mass spectrometric approaches. The C1-Inh glycosylation as studied here will inform further functional studies in order to understand glycan involvement in C1-Inh function.
Furthermore, this plasma-derived human C1-Inh glycosylation will serve as a benchmark for evaluating the glycosylation profiles of recombinant C1-Inh.

**Methods**

**Materials**

Ammonium bicarbonate (AmBiC), hydroxybenzotriazole hydrate (HOBr), iodoacetamide (IAA), 2-mercaptoethanol, Nonidet P-40 (NP-40), PBS, formic acid (FA), sDHB (2-hydroxy-5-methoxy-benzoic acid and 2,5-dihydroxybenzoic acid, 1:9) as well as a 50% sodium hydroxide (NaOH) solution were obtained from Sigma-Aldrich (Steinheim, Germany). Sodium hydrogen phosphate dihydrate (Na₂HPO₄·2H₂O), sodium bicarbonate, DTT, ethanol, potassium dihydrogen phosphate, SDS, sodium chloride and TFA were purchased from Merck (Darmstadt, Germany). 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) originated from Fluorochem (Hadfield, UK) and HPLC SupraGradient ACN from Biosolve (Valkenswaard, The Netherlands).

Commercial C1-Inh (Ceto®), isolated from pooled cryo- and 4F-depleted plasma, using ion exchange chromatography and PEG precipitation, was obtained from Sanquin, Amsterdam, The Netherlands.

**In-solution PNGase F treatment for glycopeptide analysis**

For samples to be treated with exoglycosidases, de-N-glycosylated was achieved using an in-solution Peptide-N-glycosidase F (PNGase F; Roche Diagnostics, Mannheim, Germany) approach prior to in-gel digestion with proteolytic enzymes. Ten microliter C1-Inh (1 µg/µL) were first denatured and reduced with a mixture of 1 µL of 5% SDS and 0.4 M DTT by shaking thoroughly for 5 min at 95°C. Then, 20 µL to 30 µL PNGase F solution (1 U in 10% NP-40/5 x PBS) was added, followed by an overnight incubation at 37°C. De-N-glycosylated C1-Inh was then subjected to SDS-PAGE and in-gel proteolytic digestion as described below.
In-gel proteolytic digests for glycopeptide analysis

In-gel protease digestion (10 µg protein) of C1-Inh (supplemental Figure S1), was performed as previously described,(17) using either 0.15 µg of trypsin (sequencing grade; Promega, Madison, WI), 1 µg of Proteinase K (*Tritirachium album*, Sigma-Aldrich, Steinheim, Germany) or 1 µg Pronase (*Streptomyces griseus*, Sigma-Aldrich) in 30 µL of 25 mM AmBiC.

De-N-glycosylated glycopeptides were obtained by in-gel PNGase F treatment (details in released N-glycan section) followed by in-gel proteolytic digestion, except for glycopeptide samples that were further treated with exoglycosidases (see in-solution PNGase F treatment). After in-gel PNGase F treatment the gel pieces were washed twice for 5 min with 100 µL 25 mM AmBiC before in-gel proteolytic digestion.

Exoglycosidase treatment of glycopeptides samples

(Glyco)peptides were dried in a centrifugal vacuum concentrator and reconstituted by adding 16 µL of water, 2 µL sodium acetate (50 mM, pH 5.5), 1 µL sialidase (5 mU, Glyko sialidase A; Prozyme, Hayward, CA), and 1 µL of galactosidase (5 mU, Glyko beta-galactosidase, Prozyme). The digestions were carried out overnight at 37°C.

C18-PGC-LC-ESI-QTOF-MS/MS

C18-porous graphitized carbon (PGC)-LC-ESI-QTOF-MS/MS analysis was performed as described previously (18) using a maXis HD QTOF mass spectrometer equipped with a CaptiveSpray nanoBooster source (both Bruker Daltonics) coupled to an Ultimate 3000 x 2 dual analytical nanoUPLC system (Thermo Scientific). The LC-MS setup controlled by Hystar 3.2 (Bruker Daltonics) and data analysis was performed using DataAnalysis 4.2 (Bruker Daltonics).
A combined C18-PGC-LC approach was applied to separate Pronase- and Proteinase K- treated (glyco)peptides (18). The two valve nanoUPLC system was used with the following setup: valve 1 was equipped with a C18 precolumn (C18 PepMap 100, 300 µm x 5 mm, 5 µm, 100 Å, Thermo Scientific) and analytical column (Acclaim PepMap RSLC, 75 µm x 15 cm, 2 µm, 100 Å; Thermo Scientific) and valve 2 with a PGC precolumn (in-house made, 100 µm x 15 mm, 3 µm Hypercarb material; Thermo Scientific) and analytical column (in-house made, 50 µm x 150 mm, 3 µm Hypercarb material; Thermo Scientific).

While loading, both precolumns were switched in-line allowing the sample first to pass the C18 precolumn and then to directly load the flow-through, with all unbound compounds, onto the PGC precolumn. In a second step the valves were switched for sequential elution of the compounds from the two precolumns over their corresponding analytical columns. A post-column nano valve directed the flow of the C18 or PGC column system subsequently to the mass spectrometer.

Pronase and Proteinase K treated glycopeptides were diluted 10 times and 4 µL were loaded onto the precolumns with loading solvent (99% water/1% ACN/0.05% TFA) at a flow rate of 6 µL/min and column oven temperature of 36°C. The C18-PGC-LC setup was operated with solvent A (water containing 0.1% FA (v/v)) and solvent B (80% acetonitrile/20% water containing 0.1% FA (v/v)). First, (glyco)peptides from the C18 columns were eluted with a flow rate of 500 nL/min using a linear gradient (t= 5-35 min, c(B)=1-55%), followed by column washing and reconditioning. After 28 min a post-column nano-valve was switched and the flow from the PGC columns was sent to the MS. The elution of the PGC columns was performed with a linear gradient (t= 22-55 min, c(B)=1-40%) at a flow rate of 400 nL/min, followed by column washing and reconditioning.

Ionization was enhanced using a nanoBooster (Bruker Daltonics) with acetonitrile-enriched nitrogen at 0.2 bar. The source parameters were set to a dry gas flow of 3 L/min at 150°C and a capillary voltage of 1200 V. The mass spectrometer was calibrated using ESI-L-low concentration tuning mixture (Agilent Technologies, Santa Clara, CA). MS acquisition was performed within a mass range of m/z 50 to m/z 2800 at a spectra rate of 1 Hz. Basic stepping mode was applied for the MS/MS collision energy (80 and 140%)
each for 50% of the time. Collision energies were set as follows: For singly charged precursors 45 eV at 
m/z 500, 60 eV at m/z 800, 80 eV at m/z 1300; for doubly charged precursors 25 eV at m/z 500, 47 eV at 
m/z 800, 60 eV at m/z 1300, for precursors with three and more charges 20 eV at m/z 500, 45 eV at m/z 
800, 65 eV at m/z 1300. MS/MS was performed on the three most abundant precursor ions at a spectra rate 
of 0.5 Hz to 2 Hz depending on the precursor intensity (18).

C18-RP-LC-ESI-QTOF-MS/MS

C18-RP-LC-ESI-QTOF-MS/MS analysis was performed on the same LC-MS system as used for C18- 
PGC-LC-ESI-QTOF-MS/MS. All trypsin-treated samples were diluted 30 times with water and 5 µL were 
loaded onto a C18 precolumn (C18 PepMap 100, 100 µm x 2 cm, 5 µm, 100 Å, Dionex/Thermo 
Scientific, Breda, The Netherlands) with 15 µl/min of loading solvent (water containing 0.1% ACN/ 0.1% 
FA) for 2 min. The analytes were separated on a C18 analytical column (Acclaim PepMap RSLC, 75 µm 
× 15 cm, 2 µm, 100 Å, Dionex/Thermo Scientific) at 32°C column oven temperature. Elution was 
performed at a flow rate of 0.7 µl/min with solvent A (water containing 0.1% formic acid (FA) (v/v)) and 
solvent B (95% acetonitrile (ACN)/ 5% water containing 0.1% FA (v/v)). A linear gradient of 3-31.7% 
solvent B in 25 min was applied followed by column washing and reconditioning.

The MS was operated in stepping-energy CID mode as described previously (setup two in (19)). To 
acquire data for relative quantitation the MS was operated in MS only mode. For electron-transfer 
dissociation (ETD) experiments the MS parameters were set as described for stepping-energy CID, except 
of the collision RF, which was set to 500 and 800 Vpp in basic stepping mode (each 50% of the time). The 
ICC target was set to 3 Mio for an accumulation time of max. 600 ms. The reagent injection was 45 ms 
and the extended reaction time 5 ms. ETD precursors were selected using a target list obtained from CID 
runs.

Data evaluation of LC-MS/MS spectra
DataAnalysis 4.2 software (Bruker Daltonics) was used to analyze glycopeptides of C1-Inh (P05155) by manually scanning for glycan oxonium ions. The defined stepping-energy CID glycopeptide spectra were analyzed manually to identify the glycan structure and the mass of the peptide backbone as described previously (18), including carbamidomethylation as a fixed and oxidation as a variable modification. Additionally, also lower mass range oxonium ions of stepping-energy CID spectra were used to characterize the glycan portion. The ratio of HexNAc fragments in higher-energy CID can be a marker for the presence of GlcNAc in the glycopeptide (20). While a high ratio of \( m/z \) 138 ([HexNAc-CH\(_6\)O\(_3\)]\(^+\)) + \( m/z \) 168 ([HexNAc-2H\(_2\)O]\(^+\)) relative to \( m/z \) 126 ([HexNAc-C\(_2\)H\(_6\)O\(_3\)]\(^+\)) + \( m/z \) 144 ([HexNAc-C\(_2\)H\(_4\)O\(_2\)]\(^+\)) is diagnostic for a GlcNAc-containing glycopeptide, close to equal intensities of \( m/z \) 138 + \( m/z \) 168 compared to \( m/z \) 126 + \( m/z \) 144 are indicative for only GalNAc-containing glycopeptides.

For a selected list of identified glycopeptides with known compositions and retention times, based on CID spectra, also ETD spectra were manually analyzed, allowing a mass deviation of 10 ppm (QTOF) and 0.1 Da (IT).

For relative quantitation signal intensities of all tryptic glycopeptides and partially also of miss-cleaved ones (up to one miss cleavage) were extracted in an automated manner using LaCy tools (version 1.0.0) (21). LaCy tools settings were as follows: sum spectrum resolution = 100; mass window 0.07 Th; time window 18 s; minimum percentage of the total theoretical isotopic distribution = 95%, background window =10 Th. The analyte was included for relative quantitation based on the following criteria: signal-to-noise of at least 9; average mass error of ±10 ppm, average isotopic pattern quality (IPQ) score ≤0.25. The samples were analyzed in triplicates. The data was normalized based on the total intensity of all compounds and the standard deviation was calculated.

 Released \( N \)-glycan analysis

In-gel \( N \)-glycan release (10 \( \mu \)g protein) were performed as previously described(17) with minor modifications. Different from the protocol, the gel bands were washed with 25 mM sodium bicarbonate,
pH 8, instead of AmBiC. For the N-glycan release, 20 µL to 30 µL PNGase F solution (2 U (Roche Diagnostics) in 2% Nonidet P-40 (NP-40) and 2.5xPBS) were used.

Released N-glycans were derivatized by ethyl esterification (22) followed by glycan purification by hydrophilic interaction chromatography (HILIC)-solid phase extraction (SPE) using cotton thread modified from a protocol described previously (23). The N-glycans were eluted in 10 µL water. From this, 5 µL were used for mass spectrometric analysis by spotting them onto an anchor chip matrix-assisted laser desorption dissociation (MALDI) target plate (Bruker Daltonics, Bremen, Germany) and co-crystallized with 1 µL of 5 mg/mL sDHB (2-hydroxy-5-methoxy-benzoic acid and 2,5-dihydroxybenzoic acid, 1:9, Sigma-Aldrich) in 50% ACN/50% water containing 1 mM NaOH. MALDI-TOF-MS spectra were acquired using an UltrafleXtreme mass spectrometer (Bruker Daltonics) in positive ion reflector mode. Spectra were obtained over a mass window of m/z 1000 to m/z 5000 with suppression up to m/z 900 for a total of 20000 shots. Tandem mass spectrometry (MALDI-TOF/TOF-MS/MS) was performed for structural elucidation via fragmentation in gas-off TOF/TOF mode.

A compound list of C1-Inh N-glycans was manually curated and relative quantitation of the N-glycoforms was performed using an in-house developed software for automated data processing MassyTools 1.0 (24). Only glycan compositions which have been confirmed by MS/MS and their directly related compositions (± one monosaccharide) were taken into account for relative quantitation. Detailed information of the released N-glycan sample preparation and analysis is provided in the supplemental data.

**Experimental Design and Statistical Rational:**

A detailed N- and O-glycosylation analysis of C1-Inh was performed using various MS-based approaches. C1-Inh glycopeptides were generated by subjecting the purified glycoprotein to various protease treatments with and without the addition of PNGase F and exoglycosidases for N-glycan release and glycan trimming, respectively. This resulted in N-, O- and N-/O-glycopeptides of different complexity in order to achieve a high glycopeptide coverage. Additionally C1-Inhibitor N-glycans were released and analyzed. Sample preparation, followed by MS analysis was performed in triplicates (glycopeptide
analysis) or quadruplicates ($N$-glycan analysis) and the average relative distribution of glycopeptides, $N$-glycans and their respective standard deviations were calculated.

**Data availability**

All glycopeptide tandem MS spectra are available under the following link:

ftp://massive.ucsd.edu/MSV000081096

**Results**

**Site-specific $O$-glycosylation analysis of C1-Inh**

For $O$-glycosylation site identification and characterization C1-Inh was subjected to in-gel trypsin, Pronase and Proteinase K treatment. The latter two broad-specificity proteases commonly result in smaller peptide portions to reduce sample complexity for tandem MS analysis. To further decrease sample heterogeneity and enhance $O$-glycosylation site identification a portion of these digests were de-$N$-glycosylated by PNGase F and partially also processed with exoglycosidases such as sialidase and galactosidase. This approach aimed to trim down short mucin-type $O$-glycans, to obtain $O$-glycopeptides with a single $N$-acetylhexosamine (HexNAc) or HexNAc-hexose (Hex) moiety attached to the $O$-glycosylation site, which allowed a more reliable site-specific analysis. Samples were analyzed by nanoLC-MS/MS analysis applying different tandem MS modes to obtain more structural information of the glycopeptides. For this, we applied a C18-PGC-LC-ESI-QTOF-MS/MS approach that has recently been developed by us and facilitates higher glycopeptide coverage after Pronase and Proteinase K treatment.(18) Next to the targeted ETD mode, stepping-energy CID was applied for glycopeptide assignment, resulting in b- and y-type peptide fragments, in addition to B- and Y-type glycan-derived fragments (18, 19). A detailed description of the individual $O$-glycosylation sites is given in the following and a list with all identified $O$-glycopeptides is provided in supplemental Table S1.

The first $O$-glycosylation within the protein sequence of C1-Inh is either Thr27 or Ser28 identified within the peptide portion $^{23}$NPNATS$^{28}$ carrying a $N$-glycan (HexNAc$_4$Hex$_3$NeuAc$_1$) and a monosialylated core 1
$O$-glycan ($\text{HexNAc}_1\text{Hex}_1\text{NeuAc}_1$) (Figure 1A). The peptide and glycan portions were identified by the exact mass and several b- and y-ions, as well as B- and Y-ions. Diagnostic ions, such as $m/z$ 1462.5790 $[\text{M+H}]^+$ (peptide + $\text{HexNAc}_2\text{Hex}_1\text{NeuAc}_1$) revealed the $O$-glycan nature next to the $N$-glycan portion, however, the exact location could not be unambiguously identified.

The $O$-glycosylation site Ser31 was identified by applying ETD to the glycopeptide $\text{SSQDPESL}_{37}$ with a monosialylated core 1 $O$-glycan attached (Figure 1B). The fragment ions $z_8$ and $z_7$ with the $O$-glycan attached were observed next to $c_2$ and $c_3$ peptide + glycan fragments identifying Ser31 as glycosylation site.

The analysis of tryptic de-$N$-glycosylated $O$-glycopeptides revealed the presence of non-, mono- and disialylated core 1 $O$-glycans attached to the peptide portion $\text{NPDATSSSSQDPESLQDR}_{40}$ (D indicating deamidation afterPNGase F treatment) covering Thr27/Ser28 and Ser31 (supplemental Table S1B). In addition, lower stepping-energy CID was applied in order to obtain more intense oxonium ions for glycan structural assignment (supplemental Figure 2B/C/D), while the stepping-energy CID spectrum was used to confirm the peptide portion (supplemental Figure 2A). Relative quantitation of the de-$N$-glycosylated $O$-glycopeptides containing the sites Thr27/Ser28 and Ser31 shows that around 82% of the glycopeptides contain one monosialylated core 1 $O$-glycan attached to the peptide at a time (Figure 3B, inset, supplemental Table S4).

The glycosylation sites Thr47 and Thr48 were identified based on the Pronase-treated glycopeptide $\text{ATTV}_{49}$ with two monosialylated core 1 structures attached (Figure 1C). The peptide and glycan portions were identified by the exact mass and several b- and y-ions, as well as B- and Y-ions. During this study also lower mass range oxonium ions of stepping-energy CID spectra were taken into account to identify the glycan portion. As reported by Halim et al. the ratio of HexNAc fragments in higher-energy CID can be a marker for the presence of GlcNAc in the glycopeptide (20). While a high ratio of $m/z$ 138 ($[\text{HexNAc-CH}_6\text{O}_3]^+$) + $m/z$ 168 ($[\text{HexNAc-2H}_2\text{O}]^+$) relative to $m/z$ 126 ($[\text{HexNAc-C}_2\text{H}_6\text{O}_3]^+$) + $m/z$ 144
([HexNAc-C$_2$H$_4$O$_2$]) is diagnostic for a GlcNAc-containing glycopeptide, close to equal intensities of $m/z$ 138 + $m/z$ 168 compared to $m/z$ 126 + $m/z$ 144 are indicative for only GalNAc-containing glycopeptides. A representative example of an $O$-glycopeptide (GalNAc only) and the same glycopeptide with an additional $N$-glycan attached is given in supplemental Figure S3. The presence of four GlcNAc residues in the $N$-glycan portion clearly results in elevated relative signals for $m/z$ 138 + $m/z$ 168 compared to $m/z$ 126 + $m/z$ 144. Based on close to equal intensities of these signals for the glycopeptide 46ATTV$_{49}$ (Figure 1C) it can be concluded that GalNAc is prevailing and consequently two monosialylated core 1 structures are attached to the peptide portion. In order to further study the glycan microheterogeneity of these glycosylation sites the tryptic $O$-glycopeptides with the peptide portion 45VATTVISK$_{52}$ were analyzed by lower stepping-energy CID (supplemental Figure S4/D/E and S5E/F/G). Compared to the stepping-energy CID spectrum (supplemental Figure S4C) the lower stepping-energy CID spectrum (supplemental Figure S4D) contains more intense oxonium ions for glycan structural elucidation. Using this approach it was possible to identify non-, mono- and disialylated core 1, as well as a monosialylated core 2 structure attached to the glycosylation sites Thr47/Thr48 (supplemental Figure S4D/E and S5E/F/G). Based on the lower mass range oxonium ion pattern it can be further verified that the glycosylation sites Thr47/Thr48 contain mainly core 1 $O$-glycans (supplemental Figure S4B and S5A/B/C) and partially also core 2 $O$-glycans (supplemental Figure S4A). Relative quantitation of tryptic $O$-glycopeptides with the peptide portion 46VATTVISK$_{52}$ revealed approx. 73% of the glycopeptides to be occupied by core 1 $O$-glycans (Figure 3C, supplemental Table S4). However, the missed cleaved peptide species 41GEGKVATTVISK$_{52}$ contains almost exclusively glycopeptides with the composition HexNAc$_2$Hex$_2$NeuAc$_2$ and HexNAc$_3$Hex$_3$NeuAc$_3$, indicating a bias in trypsin cleavage based on the $O$-glycan portion. With both peptide variants the glycan compositions containing HexNAc$_3$Hex$_3$NeuAc$_{2-4}$ indicative of a core 2 $O$-glycan are <1.1% (Figure 3C, supplemental Table S4).
The glycosylation sites Ser63 and Ser64 were identified based on the ETD spectrum of the Pronase- and exoglycosidases-treated glycopeptide $\alpha$EVSSLPT$_{67}$ carrying two non-sialylated core 1 O-glycans (Figure 2B). The diagnostic ions c$_4$ and c$_6$ exclude Thr67 as glycosylation site and confirm the occupation of Ser63 and Ser64. The stepping-energy CID spectrum and lower stepping-energy CID spectrum of this glycopeptide provided further evidence that the glycopeptide contains two individual core 1 O-glycans instead of one core 2 O-glycan (supplemental Figure S6). The stepping-energy CID spectrum features lower mass range oxonium ions that indicate only GalNAc-containing O-glycans (approximately equal signal intensities for $m/z$ 138 ([HexNAc-CH$_6$O$_3$]$^+$) + $m/z$ 168 ([HexNAc-2H$_2$O]$^+$) compared to $m/z$ 126 ([HexNAc-C$_2$H$_6$O$_3$]$^+$) + $m/z$ 144 ([HexNAc-C$_2$H$_4$O$_2$]$^+$) (Figure 2A). Accordingly, both stepping-energy CID and lower stepping-energy CID spectra contained no diagnostic oxonium ions that would indicate core 2 O-glycans such as HexNAc$_2$ or HexNAc$_2$Hex$_1$.

The O-glycosylation site Thr67 was identified from the ETD spectrum of the Proteinase K- and exoglycosidases-treated glycopeptide $\alpha$SLPTTD$_{69}$ (D indicates the deamidated Asn caused by PNGase F treatment) with a non-sialylated core 1 O-glycan (Figure 2C). The fragment ions c$_4$ and z$_5$ identify the O-glycosylation site Thr67.

The O-glycosylation sites Thr71 and Thr72 were identified by the Pronase- and exoglycosidases-treated glycopeptides. The glycopeptide $\gamma$TTN$_{73}$ at $m/z$ 452.1883 [M+2H]$^{2+}$ was observed with a single HexNAc together with a HexNAc$_1$Hex$_1$, due to incomplete galactosidase digestions (Figure 2E). The peptide sequence could be confirmed via the exact mass of the peptide ion at $m/z$ 335.1562 [M+H]$^+$ as well as peptide fragment ions. This peptide sequence is, however, not unique and could also result from $\delta$TTN$_{69}$. Similarly, glycopeptides with the peptide sequences TTNS were found, which could as well correspond to the peptides $\gamma$TTNS$_{74}$ or $\delta$TTNS$_{70}$. Ultimately, the same sample also contained a glycopeptide with the unique peptide sequence $\gamma$TTNSAT$_{76}$ confirming the glycosylation sites Thr71 and Thr72, since no further cleavage products were found to support the other sequence (supplemental Table S1C). The lower mass range oxonium ions of these glycopeptides featured close to equal signal intensities for $m/z$ 138
([HexNAc-CH₆O₃]⁺) + m/z 168 ([HexNAc-2H₂O]⁺) compared to m/z 126 ([HexNAc-C₂H₆O₃]⁺) + m/z 144 ([HexNAc-C₂H₄O₂]⁺), indicating the presence of two GalNAc-containing core 1 O-glycans only instead of one core 2 O-glycan (Figure 2E, supplemental Figure S7).

The O-glycosylation site Thr76 was identified from a Pronase-treated glycopeptide 75ATK₇₇, carrying a monosialylated core 1 structure. The peptide and glycan portion were identified by the exact mass and several b- and y-ions as well as B- and Y-ions (Figure 2D).

The de-N-glycosylated tryptic O-glycopeptides with the peptide sequence 5₃MLFVEPILEVSSLPTDSTTNSATK₇₇ (D indicates deamidation after PNGase F treatment) were analyzed in order to determine the glycan microheterogeneity of this region, containing the glycosylation sites Ser63, Ser64, Thr 67, Thr71, Thr72, Thr76. Although the tryptic glycopeptide contains six O-glycosylation sites, the observed fragmentation spectra indicated two simultaneously occupied glycosylation sites (supplemental Figure S8), featuring mono- and non-sialylated core 1 O-glycans. Relative quantitation revealed that approx. 50% of the glycopeptides contain one monosialylated core 1 O-glycan and 40% one monosialylated core 1 and one non-sialylated core 1 O-glycan (Figure 3D, supplemental Table S4).

The analysis of Pronase-treated samples after N-glycan release and subsequent glycan trimming with exoglycosidases also revealed a highly O-glycosylated peptide region with up to 16 occupied O-glycosylation sites. Using C18-RP-LC-ESI-QTOF-MS/MS analysis, three different glycopeptide clusters eluted close to each other featuring three peptide portions (Figure 4A). Two of them only contained occupied O-glycosylation sites, Thr82-Ser121 and Asp81-Ser121. For the latter one, Asp81 was identified as deamidated Asn due to PNGase F treatment. The two peptides with only occupied O-glycosylation sites contained 18 potential O-glycosylation sites of which up to 16 were found to be occupied (Figure 4A).

The glycopeptide at m/z 1437.0252 [M+5H]⁵⁺ covered 14 HexNAc residues (₁₀D₄DTDEPTQPTTEPTQPTIQPTQPTQLPTDSPTQPTTGS₁₂₁; D indicates the deamidated Asn caused by PNGase F treatment) (Figure 4B). The peptide mass (m/z 2169.5083 [M+2H]²⁺) was determined
from the tandem MS spectrum and the sequence was confirmed by several b- and y- ions (Figure 4B). A lower-stepping-energy CID fragmentation spectrum with a zoom-in of the peptide + HexNAc fragments of this highly glycosylated peptide further proved the glycopeptide identity (Figure 4C). Since the glycopeptide spectrum also contains a lower intensity signal at \( m/z \) 407.1627 indicative for HexNAc\(_2\), it might be possible that a small portion of these glycopeptides also contains core 2 O-glycans. Evaluating the low mass range oxonium ions of the different O-glycopeptides and combined N-/O-glycopeptides it can be seen that the N-glycan- and thus GlcNAc-containing N-/O-glycopeptide shows slightly higher signal intensities for \( m/z \) 138 ([HexNAc-CH\(_4\)O\(_3\)]\(^+\) + \( m/z \) 168 ([HexNAc-2H\(_2\)O])\(^+\) compared to \( m/z \) 126 ([HexNAc-C\(_2\)H\(_4\)O\(_3\)]\(^-\) + \( m/z \) 144 ([HexNAc-C\(_2\)H\(_4\)O\(_2\))\(^-\) (supplemental Figure S9C). In contrast the corresponding O-glycopeptides show close to equal ratios of \( m/z \) 138 + \( m/z \) 168 compared to \( m/z \) 126 + \( m/z \) 144, indicating mainly GalNAc-containing O-glycopeptides (supplemental Figure S9A/B). Thus, we conclude that the up to 16 present HexNAc residues attached to the glycan portion

\[ ^{81} \text{DTTDEPTTQPTTQPTQPTTQLPTDSPTQTTGS}_{121} \] (D indicates the deamidated Asn caused by PNGase F treatment) represent up to 16 occupied O-glycosylation sites, with possible smaller amounts of core 2 O-glycans.

In total, 26 O-glycosylation sites, carrying mainly core 1 O-glycans, were identified on the N-terminal domain of human C1-Inh (Figure 5). Nine of them (Ser31, Thr47, Thr48, Ser63, Ser64, Thr67, Thr71, Thr72 and Thr76) were detected with their exact location. Additionally, one O-glycosylation site was identified as either Thr27 or Ser28, next to a heavily glycosylated mucin-type O-glycan region spanning from Thr82 to Ser121 with up to 16 O-glycosylation sites.

**Released N-glycan analysis of C1-Inh**

To get an overview of the overall C1-Inh N-glycosylation on all six sites, released N-glycans were analyzed using linkage-specific sialic acid derivatization and MALDI-TOF-MS analysis (Figure 6,
supplemental Table S2). The majority of the N-glycan structures were sialylated species, among which a diantennary, α2,6-linked disialylated N-glycan was the most abundant one (Hex$_3$HexNAc$_4$(α2,6)N-acetylneuraminic acid (NeuAc$_2$)) (supplemental Figure S10).

Approximately 90% of C1-Inh N-glycans contained at least one α2,6-linked NeuAc and about 30% featured α2,3-linked NeuAc. Approximately 25% of the N-glycans were monofucosylated and MS/MS analysis identified fucoses to be mainly core-linked. Multi-fucosylation (two or more fucoses), indicative of antenna fucosylation, was less than 5%. C1-Inh N-glycans were almost completely galactosylated (95%) and the majority of the structures were diantennary glycans (80%), next to tri- or tetraantennary glycans (below 20%).

Site-specific N-glycopeptide analysis of C1-Inh

In order to identify N-glycosylation sites and site-specific glycan microheterogeneity, proteolytic treatment was carried out using trypsin, Proteinase K and Pronase. The samples were analyzed by C18-RP-LC-ESI-QTOF-MS/MS and C18-PGC-LC-ESI-QTOF-MS/MS using stepping-energy CID. All six potential N-glycosylation sites of C1-Inh were identified and are listed in supplemental Table S3 and representative tandem MS spectra of each site are shown in Figure 1C and supplemental Figure S11/S12. The corresponding glycoforms of each N-glycosylation site are depicted in Figure 5 and are elaborated in more detail in the following.

Trypsin digestion of C1-Inh allowed identification and characterization of glycosylation sites Asn238, Asn253 and Asn352. Glycopeptides were identified on the basis of their tandem MS spectra. Additional glycopeptides from these sites with glycoforms that follow the same biosynthetic pathway were identified by their exact mass on MS level. The relative abundances of the different glycoforms attached to the corresponding tryptic glycopeptides were extracted in an automated manner using the software package LaCytools (21). The relative intensities for each glycosylation site were normalized to the total intensity of all glycoforms attached to the same peptide moiety. (Figure 3A, supplemental Table S4).
For Asn238, 21 N-glycoforms were quantified, with HexNAc4Hex5NeuAc2 (52.0%) being the most abundant one. The vast majority of the N-glycans are diantennary (77%), followed by triantennary (21.4%) and tetraantennary ones (1.6%). While the glycans are almost completely sialylated (98.6%) only 20.1% are fucosylated.

The relative quantitation of 12 glycoforms of the glycosylation site Asn253 identified HexNAc4Hex5NeuAc2 (45.8%) as the most abundant structure. Around 84.1% of the N-glycans are diantennary, next to another 15.5% triantennary and less than 1% tetraantennary ones. Asn253 features almost complete sialylation and approximately 34.7% fucosylation.

For Asn352, six N-glycoforms were used for relative quantitation. Next to four diantennary sialylated N-glycans also two triantennary sialylated N-glycan were quantified. Again, HexNAc4Hex5NeuAc2 was the most abundant structure with 77.7%. Four of these N-glycans were fucosylated representing 9.6%.

The glycoforms present on Asn25 were identified by Pronase-treated glycopeptides. Pronase digestion was used for N-glycosylation site characterization to generate smaller peptide portions and thereby overcoming interference from the two O-glycosylation sites in close proximity (Ser31 and Thr27/Ser28) to Asn25. Site-specific elucidation of the N-glycoforms was facilitated by tandem mass spectrometry which allowed the distinction between O-glycan and N-glycan on this peptide moiety. Asn25 features di- and triantennary N-glycans with and without core fucosylation ranging from mono- to trisialylation. For relative quantitation, 18 different glycan compositions were detected attached to the tryptic glycopeptides around Asn25 (23NPNATSSSSQDPESLQDR40)(Figure 3B). To identify the contribution of the O-glycans, the de-N-glycosylated species of this peptide were analyzed after PNGase F treatment as discussed in the O-glycosylation analysis section (Figure 3B insert). As a result, it can be assumed that most of the relative signal intensities of the N-and O-glycosylated peptide contains one core 1 O-glycan. The most abundant glycan composition of the tryptic glycopeptide is HexNAc3Hex5NeuAc3, which is most likely the N-glycan HexNAc4Hex5NeuAc2 considering the presence of a monosialylated core 1 O-glycan. In total, approximately 68.9% of the glycans followed the composition HexNAc4Hex5 with different degrees of
sialylation and with and without fucosylation. Likewise, 26.4% had the composition HexNAc$_6$Hex$_7$, 4.1% HexNAc$_4$Hex$_5$ and 0.6% HexNAc$_3$Hex$_8$. Including the likely presence of a core 1 O-glycan structure in each of these compositions, the glycosylation profile is similar to the one of Asn238, Asn253 and Asn352.

Glycosylation site Asn69 was characterized by Proteinase K-generated glycopeptides. Three different glycopeptides containing the N-glycosylation site Asn69 could be detected after Proteinase K-treatment, featuring non-fucosylated, di- and triantennary structures with two and three sialic acids. Due to the location of Asn69 close to six neighboring O-glycosylation sites (Ser63, Ser64, Thr67, Thr71, Thr72, Thr76) further mapping of the microheterogeneity in a relative manner was not possible.

The sixth N-glycosylation site Asn81 was identified as part of a heavily O-glycosylated peptide region after Pronase-treatment, followed by PNGase F N-glycan release and exoglycosidase treatment as described for O-glycopeptide identification. Most likely due to the large amount of O-glycans, the N-glycan release in this region was not complete, allowing the detection of the glycopeptide cluster with the peptide portion $^\gamma$TANTTDEPTTQPTTQPTQPTQPTQPTTQLPTDSPTQPTTGS$_{121}$ featuring the glycan composition HexNAc$_{18-21}$Hex$_3$ (Figure 4A, supplemental Figure S12). The stepping-energy CID spectrum of this glycopeptide contained several diagnostic oxonium ions clearly indicating the presence of an N-glycan. However, due to the complexity of this glycopeptide and the application of exoglycosidases it was not possible to characterize the glycan microheterogeneity of Asn81.

Overall, the site-specific glycopeptide-based N-glycoform distribution was in good agreement with the total N-glycoform pattern determined by MALDI-TOF-MS of released N-glycans.
Discussion

In this study, we present for the first time a comprehensive site-specific glycoproteomic analysis of human C1-Inh by analyzing N- and O-glycopeptides next to released N-glycans with multi-methodological mass spectrometric approaches.

Remarkably, we were able to detect up to 26 O-glycosylation events on C1-Inh. Of these, nine were assigned to a specific site in the N-terminal domain (Ser31, Thr47, Thr48, Ser63, Ser64, Thr67, Thr71, Thr72, Thr76). Another O-glycan could be assigned to either Thr27 or Ser28, leaving some ambiguity.

With respect to the other detected O-glycans, we were able to identify a heavily O-glycosylated region (Thr82-Ser121) with up to 16 occupied O-glycosylation sites out of 18 possible sites by tandem MS. This shows overall very high O-glycosylation site occupancy of this mucin-type region.

Our findings are perfectly in line with previous studies on C1-inhibitor O-glycosylation, but go way beyond those reports: O-glycosylation sites Thr47 and Thr48 have been reported earlier in human plasma, cerebrospinal glycoproteins and human urine containing core-1 O-glycans (10, 13, 25) and were confirmed in our study. Likewise, Ser31 was confirmed from human plasma (13). Furthermore, Ser64 and Ser71 are known from amino acid sequencing studies (11) and were for the first time confirmed by our mass spectrometric study. The occupation of the sites Thr83, Thr88, Thr92 and Thr96 has previously been revealed by amino acid sequencing, and in addition Bock et al. also suggested the O-glycosylation of Thr99, Thr106, Thr107, Thr111, Thr115, Thr118 and Thr119, however, they could not provide unequivocal evidence (11). Additionally, the peptide portions $^{113}$SPTQPTTGSF$_{121}$ and $^{113}$SPTQPTTGSFCGPVTL$_{128}$ have been recently identified with two and one O-glycan attached, respectively. These findings are largely in-line with our mass spectrometric data which demonstrated occupation of up to 16 of the 18 possible O-glycosylation sites in the region Thr$_{82}$-Ser$_{121}$. Hence, the MS/(MS) analysis of Thr$_{82}$-Ser$_{121}$ demonstrated for the first time the excessive O-glycosylation of this region. Notably, to the best of our knowledge the O-glycosylation sites Ser63, Thr67, Thr72, Thr76 as well as the O-glycosylation region Thr27/Ser28 are reported for the first time. In a recent study on plasma
O-glycosylation also Ser36 was identified as an O-glycosylation site (13), which was not confirmed by our data using the purified glycoprotein. However, we cannot exclude the glycosylation of this site.

In the current study, we mainly observe mono-, di-, and non-sialylated core 1 O-glycans and smaller amounts of core 2 O-glycans for the glycosylation sites Thr47/Thr48. It might be also possible that the other O-glycosylation sites carry smaller amounts of core 2 O-glycans, which were not detected.

These findings can be only compared to Ser31, Thr47 and Thr48, since these sites have been identified before by MS (10, 13, 25). However, in these studies the glycopeptides were captured using hydrazine chemistry, which included acid hydrolysis to release O-glycopeptides, or after sialidase treatment of the glycopeptides. Using the hydrazine chemistry approach confirmed that the O-glycopeptides contained sialic acid but further information about the glycan microheterogeneity was lost. In all three studies only core 1-containing glycopeptides were identifies.

The N-glycosylation sites of C1-Inh have been mapped by mass spectrometry (25–31), as well as other analytical approaches in various studies (4, 8, 12, 32). In agreement with literature, we confirmed the glycosylation of all six N-glycosylation sites. Perkins et al. and Strecker et al., reported that C1-Inh carried disialylated complex-type N-glycans (8, 12) and small amounts of tri- and tetraantennary glycan structures (8). This is in accordance with the major structure detected in our study. Similarly, small amounts of tri- and tetra-antennary glycan structures have been detected by Perkins et al. Hitherto, knowledge on site-specific N-glycosylation of C1-Inh has been very limited. A previous report detected solely diantennary structures on Asn352 (31), whereas we also detected small amounts of less than 3% of triantennary glycan structures at this site.

The role of C1-Inh glycosylation has been investigated and discussed with regards to its functional activity in a few studies. The high degree of sialylation can prolong serum half-life as desialylation of the protein resulted in faster clearance from blood in rabbits, supposedly due to the asialoglycoprotein receptor in the liver (33). However, deglycosylation of human C1-Inh with PNGase F, O-glycanase, or both had no detectable impact on its C1s protease inhibition activity (34, 35). Different observations were made for the
involvement of C1-Inh glycosylation in the inhibition of kallikrein. While the desialylated and de-N-glycosylated human C1-Inh retain its ability to complex kallikrein, the further O-glycanase-treated protein species lost this feature (35). These results suggest the contribution of N-terminal O-glycosylation in C1-Inh protease inhibition in the contact system. In contrary, studies investigating the protease inhibition activity of recombinant WT C1-Inh and constructs with deletion of the first 98 or 120 amino acid, representing the highly glycosylated N-terminal domain, showed the same interaction kinetics with C1s, kallikrein or factor XIIa (amino acid numbering including the signal peptide; recombinant expression in P. pastoris and E. coli) (36, 37).

The N-terminus has been described to be a “rod-like” domain, likely due to the presence of the glycans (8, 11, 38). In fact, the N-terminal domain, either due to its size and/or charge, has been suggested to interfere with the interaction of C1-Inh with cell surface bound kallikrein and factor XIIa (39, 40). It may be hypothesized that the extensive N- and O-glycosylation increases the overall and local size, charge, and hydrophilicity as well as thermodynamic stability of the N-terminal domain, thereby modifying the physicochemical properties of C1-Inh. Moreover, we found that the highly O-glycosylated region with up to 12-16 occupied sites was resistant to protease treatment, demonstrating that glycosylation can prevent C1-Inh from proteolytic degradation.

In conclusion, we here applied a powerful approach for in-depth site-specific characterization of C1-Inh glycosylation revealing 10 O-glycosylation sites carrying mainly core-1 type O-glycans, with five of them being novel. In addition, we identified a heavily O-glycosylated portion of C1-Inh spanning from Thr82-Ser121 with up to 16 O-glycans attached. Likewise, we covered all six N-glycosylation sites of C1-Inh by site-specific glycosylation analysis.

This extensive and specific information on C1-Inh glycosylation will help to better understand existing functional studies and it is essential for further targeted studies to investigate the role of glycosylation of this plasma glycoprotein. Furthermore we elucidated plasma-derived human C1-Inh glycosylation which will be important to further evaluate the glycosylation profiles of recombinantly expressed C1-Inh.
Acknowledgements

H. Mehmet Kayıli is grateful for financial support of scholarship program (2214) of TUBITAK and VU University Amsterdam. This project was further supported by the European Union (Seventh Framework Programme HighGlycan project, grant number: 278535). We would like to acknowledge Jonas Nilsson for the fruitful discussions about O-glycopeptide spectra interpretation.

Disclosure of Conflicts of Interest

RE, DW, SZ are employed by Sanquin.

References

1. Caliezi C, Wuillemin WA, Zeerleder S, et al (2000) C1-Esterase Inhibitor: An Anti-Inflammatory Agent and Its Potential Use in the Treatment of Diseases Other Than Hereditary Angioedema. Pharmacol Rev 52:91–112.
2. Bowen T, Cicardi M, Farkas H, et al (2010) 2010 International consensus algorithm for the diagnosis, therapy and management of hereditary angioedema. Allergy, asthma, Clin Immunol 6:24.
3. Davis AE, Lu F, Mejia P (2010) C1 inhibitor, a multi-functional serine protease inhibitor. Thromb Haemost 104:886–893.
4. Koles K, van Berkel PHC, Mannesse MLM, et al (2004) Influence of lactation parameters on the N-glycosylation of recombinant human C1 inhibitor isolated from the milk of transgenic rabbits. Glycobiology 14:979–986.
5. Koles K, van Berkel PHC, Pieper FR, et al (2004) N- and O-glycans of recombinant human C1 inhibitor expressed in the milk of transgenic rabbits. Glycobiology 14:51–64.
6. Bos IGA, De Bruin EC, Karuntu YA, et al (2003) Recombinant human C1-inhibitor produced in Pichia pastoris has the same inhibitory capacity as plasma C1-inhibitor. Biochim Biophys Acta 1648:75–83.
7. Lamark T, Ingebrigtsen M, Bjørnstad C, et al (2001) Expression of active human C1 inhibitor serpin domain in Escherichia coli. Protein Expr Purif 22:349–58.
8. Perkins SJ, Smith KF, Amatayakul S, et al (1990) Two-domain structure of the native and reactive centre cleaved forms of C1 inhibitor of human complement by neutron scattering. J Mol Biol 214:751–763.
9. Nuijens JH, Eerenberg-Belmer JM, Huijbregts CCM, et al (1989) Proteolytic inactivation of plasma C1 inhibitor in sepsis. J Clin Invest 84:443–450.
10. Halim A, Rüetschi U, Larson G, Nilsson J (2013) LC-MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins. J Proteome Res 12:573–584.

11. Bock SC, Skriver K, Thogersen HC, et al (1986) Human C1 Inhibitor: Primary Structure, cDNA Cloning, and Chromosomal Localization. Biochemistry 25:4292–4301.

12. Strecker G, Ollierhartmann MP, Vanhalbeek H, et al (1985) Primary Structure Elucidation of Carbohydrate Chains of Normal C1-Esterase Inhibitor (C1-Inh) by 400-MHz H-1-Nmr Study. C R Acad Sc Paris 301:571–576.

13. King SL, Joshi HJ, Schjoldager KT, et al (2017) Characterizing the O-glycosylation landscape of human plasma, platelets, and endothelial cells. Blood Adv 1:429–442.

14. Perkins SJ (1993) Three-dimensional structure and molecular modelling of C1-inhibitor. Behring Inst Mitt 93:63–80.

15. Dwek RA (1996) Glycobiology: Toward Understanding the Function of Sugars. Chem Rev 96:683–720.

16. Bieberich E (2014) Synthesis, processing, and function of N-glycans in N-glycoproteins. adv Neurobiol 9:47–70.

17. Plomp R, Hensbergen PJ, Rombouts Y, et al (2014) Site-specific N-glycosylation analysis of human immunoglobulin E. J Proteome Res 13:536–546.

18. Stavenhagen K, Plomp R, Wuhrer M (2015) Site-specific protein N- and O-glycosylation analysis by a C18-porous graphitized carbon-liquid chromatography-electrospray ionization mass spectrometry approach using Pronase treated glycopeptides. Anal Chem 87:11691–11699.

19. Hinneburg H, Stavenhagen K, Schweiger-Hufnagel U, et al (2016) The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. J Am Soc Mass Spectrom 27:507–519.

20. Halim A, Westerlind U, Pett C, et al (2014) Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. J Proteome Res 13:6024–6032.

21. Jansen BC, Falck D, de Haan N, et al (2016) LaCyTools – a targeted LC-MS data processing package for relative quantitation of glycopeptides. J Proteome Res 15:2198–2210.

22. Reiding KR, Blank D, Kuijper DM, et al (2014) High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. Anal Chem 86:5784–5793.

23. Selman MHJ, Hemayatkar M, Deelder AM, Wuhrer M (2011) Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides. Anal Chem 83:2492–2499.

24. Jansen BC, Reiding KR, Bondt A, et al (2015) MassyTools: A High-Throughput Targeted Data Processing Tool for Relative Quantitation and Quality Control Developed for Glycomic and Glycoproteomic MALDI-MS. J Proteome Res 14:5088–5098.

25. Halim A, Nilsson J, Rüetschi U, et al (2012) Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. Mol Cell proteomics 11:M111.013649.

26. Bunkenborg J, Pilch BJ, Podtelejnikov A V., Wisniewski JR (2004) Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. Proteomics 4:454–465.
27. Liu T, Qian WJ, Gritsenko M a., et al (2005) Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. J Proteome Res 4:2070–2080.

28. Chen R, Jiang X, Sun D, et al (2009) Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. J Proteome Res 8:651–661.

29. Jia W, Lu Z, Fu Y, et al (2009) A strategy for precise and large scale identification of core fucosylated glycoproteins. Mol Cell Proteomics 8:913–923.

30. Zhang H, Li X-J, Martin DB, Aebersold R (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat Biotechnol 21:660–6.

31. Nilsson J, Rüetschi U, Halim A, et al (2009) Enrichment of glycopeptides for glycan structure and attachment site identification. Nat Methods 6:809–811.

32. Schoenberger OL (1992) Characterization of carbohydrate chains of C1-inhibitor and of desialylated C1-inhibitor. FEBS Lett 314:430–434.

33. Minta JO (1981) The role of sialic acid in the functional activity and the hepatic clearance of C1-INH. J Immunol 126:245–249.

34. Reboul A, Prandini MH, Colomb MG (1987) Proteolysis and deglycosylation of human C1 inhibitor. Effect on functional properties. BiochemJ 244:117–121.

35. Ghannam A, Sellier P, Fain O, et al (2016) C1 Inhibitor as a glycoprotein: The influence of polysaccharides on its function and autoantibody target. Mol Immunol 71:161–165.

36. Bos IGA, Lubbers YTP, Roem D, et al (2003) The functional integrity of the serpin domain of C1-inhibitor depends on the unique N-terminal domain, as revealed by a pathological mutant. J Biol Chem 278:29463–29470.

37. Coutinho M, Aulak KS, Davis AE 3rd (1994) Functional analysis of the serpin domain of C1 inhibitor. J Immunol 153:3648–3654.

38. Odermatt E, Berger H, Sano Y (1981) Size and shape of human C1-Inhibitor. FEBS Lett 131:283–285.

39. Ravindran S, Grys TE, Welch RA, et al (2004) Inhibition of plasma kallikrein by C1-inhibitor: role of endothelial cells and the amino-terminal domain of C1-inhibitor. Thromb Haemost 92:1277–1290.

40. Schousboe I (2003) Binding of activated factor XII to endothelial cells affects its inactivation by the C1-esterase inhibitor. Eur J Biochem 270:111–118.

Figure legends

Figure 1: Mass spectrometric elucidation of C1-Inh O-glycosylation. Representative O-glycopeptide mass spectra are shown. (A) Stepping-energy CID spectrum of a Pronase-treated glycopeptide with the peptide sequence 23NPNATS28 carrying a monosialylated core 1 structure and a monosialylated diantennary N-glycan, confirming the N-glycosylation site Asn25 and indicating either Thr27 or Ser28 to be O-
glycosylated; (B) The Pronase-generated glycopeptide 30SSQDPESL37 with a monosialylated core 1 glycan attached, confirming the O-glycosylation site Ser31. The sample was analyzed by ion trap-MS/MS using ETD (experimental details in supplemental data). # indicates the loss of an acetyl group; (C) Stepping-energy CID spectrum of a Pronase-treated glycopeptide with the peptide sequence 46ATTV49 with two monosialylated core 1 structures attached, confirming the O-glycosylation sites Thr47 and Thr48. The zoom-in shows the lower mass oxonium ions. Close to equal intensities of m/z 138 ([HexNAc-CH₆O₃]+) + m/z 168 ([HexNAc-2H₂O]+) compared to m/z 126 ([HexNAc-C₂H₆O₃]+) + m/z 144 ([HexNAc-C₂H₄O₂]+) are diagnostic for an only GalNAc-containing glycan portion attached to the peptide (20).

Glycan depiction: blue square = N-acetylglucosamine, yellow square = N-acetylgalactosamine, yellow circle = galactose, green circle = mannose, red triangle = fucose, purple diamond = N-acetylneuraminic acid

**Figure 2:** Mass spectrometric elucidation of C1-Inh O-glycosylation. Representative O-glycopeptide mass spectra are shown. (A) Lower mass oxonium ion region of a stepping-energy CID spectrum of the glycopeptide 61EVSSLPT67 carrying two HexNAc₁Hex₁ structures. Almost equal intensities of m/z 138 + m/z 168 compared to m/z 126 + m/z 144 are diagnostic for an only GalNAc-containing glycan portion attached to the peptide, confirming the presence of two core 1 O-glycans; (B) A Proteinase K-treated glycopeptide after exoglycosidase treatment with the peptide sequence 61EVSSLPT67 carrying two HexNAc₁Hex₁ structures, confirming the O-glycosylation sites Ser63 and Ser64. The sample was analyzed by QTOF-MS/MS using ETD. (C) The Proteinase K-treated glycopeptide after PNGase F and exoglycosidase treatment with the peptide sequence 64SLPTTD₆⁹ (D indicated deamidation after PNGase F treatment) and a HexNAc₁Hex₁ O-glycan attached, confirming the O-glycosylation site Thr67. The sample was analyzed by QTOF-MS/MS using ETD. (D) Stepping-energy CID spectrum of a Pronase-treated glycopeptide with the peptide sequence 75ATK₇₇ carrying a monosialylated core 1 structure, confirming the O-glycosylation site Thr76; (E) Stepping-energy CID spectra of a Pronase-treated
glycopeptide after exoglycosidase treatment with the peptide sequence $^{71}$TTN$^{73}$ carrying two monosialylated core 1 structure, confirming the O-glycosylation site Thr71 and Thr72. The zoom in shows the lower mass oxonium ions. An almost equal ratio of $m/z$ 138 + $m/z$ 168 compared to $m/z$ 126 + $m/z$ 144 is diagnostic for an only GalNAc-containing glycan portion attached to the peptide, confirming the presence of two core 1 O-glycans.

**Figure 3:** Site-specific relative quantification of $N$- and $O$-glycoforms present at C1-Inh. The signal intensities were normalized by the sum of all glycopeptides at each site containing the same peptide moiety. The average intensities and standard deviations of the glycoforms were calculated based on the analysis of triplicates. (A) Relative distribution of the tryptic $N$-glycopeptides attached to Asn238, Asn253 and Asn352. (B) Relative distribution of the tryptic glycopeptide $^{23}$NPNATSSSSQDPESLQDR$^{40}$ containing $N$-glycans attached to Asn25 and $O$-glycans attached to Thr27/Ser28 and/or Ser31. The insert diagram contains the $O$-glycopeptide distribution after enzymatic de-$N$-glycosylation using PNGase F. (C) Relative distribution of the tryptic and miss-cleaved $O$-glycopeptides covering the sites Thr47, Thr48. (D) Relative distribution of the tryptic $O$-glycopeptides covering the sites Ser63, Ser64, Thr67, Thr71, Thr72, Thr76. (D indicated deamidation after PNGase F treatment.)

**Figure 4:** C18-PGC-LC-ESI−QTOF−MS/MS analysis of Pronase-generated $O$-glycopeptides with multiple glycosylation sites after PNGase F $N$-glycan release and exoglycosidase treatment with sialidase and galactosidase. (A) Extracted ion chromatograms of three different glycopeptide clusters spanning from Thr79 to Ser121 with different glycan compositions attached. Pep1=$^{79}$TANTTDEPTTTQPTTEPTQTIQPTQPTTQLPDSPTQPTTGS$^{121}$; Pep2=$^{82}$TTDEPTTTQPTTEPTQTIQPTQPTTQLPDSPTQPTTGS$^{121}$; Pep3=$^{81}$DTTDEPTTTQPTTEPTQTIQPTQPTTQLPDSPTQPTTGS$^{121}$ (D indicates the deamidated Asn caused by PNGase F treatment) (B) Stepping-energy CID spectrum of Pep3 with 14 HexNAc residues,
indicating up to 14 occupied O-glycosylation sites. (C) Stepping-energy CID spectrum of the same glycopeptide as in panel B using lower-energy stepping-energy CID (stepping energy is set to 60% and 80% each half of the time (instead of 80%-140%) with a focus on the glycan-derived Y-ions.

**Figure 5:** A schematic representation of the N- and O-glycosylation of C1-Inh with the serpin (blue letters) and N-terminus domain (black letters). All determined glycosylation sites are in bold letters depicted with their corresponding glycoforms. O- and N-glycosylation sites that have been identified after exoglycosidase treatment are marked with a *, since their microheterogeneity was not detected. Possible O-glycosylation sites that were not identified with their exact location are indicated in bold and italic letters.

**Figure 6. Total released N-glycan profiling of human C1-Inh.** MALDI-TOF-MS analysis of C1-Inh N-glycans was performed after PNGase F release and linkage-specific sialic acid derivatization. Glycans were assigned on the basis of MS/MS data and literature. N-acetylleuraminic acid (NeuAc) in α2,3- and α2,6-linkage are represented by a purple diamond pointing to the left and right, respectively. The insert shows the relative quantification of N-glycan features from quadruplets including antenna-fucosylation (defined as amount of fucoses higher than two) and antennarity.
Figures

Figure 1
Figure 2

A: GalNAc-containing O-glycans only
[M+2H]^2+ 731.8265
stepping-energy CID
33.6-58.8 eV
QTOF

B: [M+2H]^2+ 731.8255
ETD
QTOF

C: [M+2H]^2+ 499.7259
ETD
QTOF

D: [M+2H]^2+ 488.2123
stepping-energy CID
QTOF

E: GalNAc-containing O-glycans only
[M+2H]^2+ 452.1883
stepping-energy CID
32.2-32.5 eV
QTOF
Figure 5

1  MASRLTLTLV LLLLAGDRA SSSPNAFSS SSQDPSLQDR

41 GEKKVAVTIV SKMLFVEPIL EVSSLPFTNS TINSATKITA
12 12-16 x HexNac --> up to 16 O-glycosylation sites
81 NTSSSTTQP TTSTPTSTT QPTPTQTLE TDSTQPPTTS
122 FCPGPVTLG SDLESHTEDA VLSDLALVFES LKLYHAFSAM
161 KKVETNMAFS PFSIASLTTQ VLLGAGENTK TNLESILSYVP
201 KDFTCVHQAL KGFTGTKVTS VSQIFHSPDL AIRDTFVNAS
241 RTLYSSSPRV LSWNSDALE LINTWVAKNT NNKISRLLDS
281 LPSDTRLVLL NAIYLSAKWK TTFDPKTRM EPFHFKNSVI
321 KVPWMNSKYY PVAHIPQTLK KKVQGQLQLS HNLSLVILVP
361 QNLKHREDM EQLAPSQVFSK AMIEKLEMSK FPOPTLTLFR
401 IKVTTSQDML SIMEKLEFFD FSYDLNCGL TEPQLQVSA
441 MQHQTVELELT ETGVEAAASA AISVARTLLV FEVQCPFVF
481 LHQQHKEFV PMSRVYDPRA

Peptide - Signal peptide
Peptide - N-terminus
Peptide - SERPIN domain
5 - O-glycosylation confirmed after exoglycosidase treatment
NG* - confirmed N-glycosylation site
N/S/T - confirmed glycosylation site
S/T - possibly occupied O-glycosylation site
S/T/L1122 - glycosylation site region
Figure 6