multi-methodological mass spectrometric N- and O-glycosylation analysis of human C1-inhibitor reveals extensive mucin-type O-glycosylation

Kathrin Stavenhagen\textsuperscript{1,2\ast}, H. Mehmet Kayıli\textsuperscript{2,3,4\ast}, Stephanie Holst\textsuperscript{1}, Carolien A. M. Koeleman\textsuperscript{3}, Ruchira Engel\textsuperscript{6,7}, Diana Wouters\textsuperscript{6,7}, Sacha Zeerleder\textsuperscript{6,7}, Bekir Salih\textsuperscript{5}, Manfred Wuhrer\textsuperscript{1,2}

\textsuperscript{1}Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands
\textsuperscript{2}Division of BioAnalytical Chemistry, VU University Amsterdam, Amsterdam, The Netherlands
\textsuperscript{3}Department of Chemistry, Çankırı Karatekin University, Çankırı, Turkey
\textsuperscript{4}Department of Nutrition and Dietetics, Karabuk University, Karabuk, Turkey
\textsuperscript{5}Department of Chemistry, Hacettepe University, Ankara, Turkey
\textsuperscript{6}Department of Immunopathology, Sanquin Research and Landsteiner Laboratory of the AMC, Amsterdam, The Netherlands
\textsuperscript{7}Department of Hematology, Academic Medical Center, University of Amsterdam, The Netherlands

\textsuperscript{\ast} Authors equally contributed to this work
6
Human C1-inhibitor (C1-Inh) is a serine protease inhibitor and a highly glycosylated plasma glycoprotein. Both the structural features and biological role of C1-Inh glycosylation are still largely unknown. Here, we performed for the first time a site-specific N- and O-glycosylation analysis of C1-Inh using mass spectrometry (MS)-based approaches. Various proteases were applied, partly in combination with PNGase F and exoglycosidase treatment, in order to analyze the (glyco)peptides by C18-porous graphitized carbon (PGC)-liquid chromatography (LC)-electrospray ionization (ESI)-quadrupole-time of flight (QTOF-)MS/MS and C18-reversed-phase (RP)-LC-ESI-ion trap (IT)-MS/MS. Fragmentation was achieved by stepping-energy collision-induced dissociation (CID) and electron-transfer dissociation (ETD). The analysis revealed an extensively O-glycosylated N-terminal region. Six novel O-glycosylation sites were identified, carrying core1-type O-glycans, next to one novel O-glycosylation site within the region Thr27-Ser28 and the confirmation of four known sites. In addition, we detected a heavily O-glycosylated portion of C1-Inh 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 matrix-assisted laser desorption ionization (MALDI)-time of flight (TOF)/TOF-MS/MS. Our data represents the first in-depth glycosylation characterization of C1-Inh, indicating its vast O-glycosylation. This structural information will form the basis to advance further functional studies on C1-Inh glycosylation.
Introduction

Protein glycosylation is one of the most common post-translational modifications with a large impact on biological processes, protein stability, and protein functions.\textsuperscript{1–3} It is estimated that 50% to 70% of the serum proteins are glycosylated.\textsuperscript{4} One of them is human C1-inhibitor (C1-Inh), 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. 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 sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (> 80kDa) due to its heavy glycosylation.\textsuperscript{5, 6} It has been reported that C1-Inh possesses six occupied N- and up to 24 O-glycosylation sites.\textsuperscript{3, 5, 7, 8} Of the latter ones eight have been identified with their exact location.\textsuperscript{3, 7} C1-Inh is one of the most heavily glycosylated proteins in human plasma,\textsuperscript{9} making it an analytical challenge to characterize.

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.\textsuperscript{7, 11} However, the biological role of C1-Inh carbohydrate groups is still largely unknown. To overcome this, a detailed site-specific glycan characterization is needed to set the basis for further functional studies to understand glycan involvement in C1-Inh function.

Glycosylation is a very complex modification, often requiring the application and combination of a variety of analytical approaches to achieve comprehensive protein glycosylation information. The analysis of released glycans is a valuable tool for defining the glycome of the whole protein, but lacks site-specific information. Therefore, it is often complemented by glycopeptide-centric approaches.\textsuperscript{14} Glycopeptide analysis can be hindered by a high sample complexity caused by the macro- and microheterogeneity of the protein, reflecting partial occupation of glycosylation sites and variations of glycans attached to one specific glycosylation site.\textsuperscript{15} However, recent analytical advances including improved workflows, protein/peptide enrichment strategies, and instrumental developments facilitated a better coverage of glycoproteomic approaches\textsuperscript{16, 17} for the analysis of single proteins\textsuperscript{18, 19} or complex mixtures.\textsuperscript{20, 21}

Here, we present for the first time an in-depth glycosylation analysis of C1-Inh using a panel of mass spectrometric approaches, including site-specific characterization of N- and O-glycosylation by C18-(PGC)-liquid chromatography (LC)-electrospray ionization (ESI)-quadrupole-time of flight (QTOF)-mass spectrometry (MS)/MS, C18-reversed-phase (RP)-LC-ESI-ion trap (IT)-MS/MS, as well as glycomic characterization of released N-glycans by matrix-assisted laser desorption ionization (MALDI)-time of flight (TOF)/TOF-MS/MS.
Materials – Methods

Chemicals and samples

Human C1-Inhibitor (Cetor®) was provided by Sanquin (Amsterdam, The Netherlands). Proteinase K from *Tritirachium album*, Pronase from *Streptomyces griseus*, 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 (Na$_2$HPO$_4$ x 2H$_2$O), sodium bicarbonate, dithiotreitol (DTT), ethanol, potassium dihydrogen phosphate, SDS, sodium chloride and trifluoric acid (TFA) were purchased from Merck (Darmstadt, Germany). 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) originated from Fluorochem (Hadfield, UK) and HPLC SupraGradient acetonitrile (ACN) from Biosolve (Valkenswaard, The Netherlands). Peptide-N-glycosidase F (PNGase F) was bought from Roche Diagnostics (Mannheim, Germany). Sequencing grade modified trypsin was purchased from Promega (Madison, WI) and Glyko sialidase A and Glyko beta-galactosidase were purchased from Prozyme (Hayward, CA). NuPAGE® Novex® 4-12% Bis-Tris gels, NuPAGE® MES SDS Running Buffer (20X), SimplyBlue™ SafeStain, and Novex® Sharp Pre-stained Protein Standard were from Life Technologies (Carlsbad, CA). Deionized water (R > 18.2 MΩ cm-1; PURELAB Ultra system; Elga Labwater, Ede, The Netherlands) was used for all preparations in this study.

SDS-polyacrylamide gel electrophoresis (PAGE)

Samples of C1-inh (10 µg) were diluted in water and supplemented with 4x Laemmli buffer containing 5% 2-mercaptoethanol. Proteins were reduced and denatured for 10 min at 95°C to 100°C. The samples were loaded onto a NuPAGE 4-12% gradient Bis–Tris gel together with 5 µl Novex Sharp Pre-stained Protein Standard and run with NuPAGE MES SDS running buffer for 50 min at 200 V. Proteins were visualized using SimplyBlue SafeStain.

Released N-glycan analysis

In-gel PNGase F treatment for the analysis of released N-glycans

In-gel digestions were performed as previously described$^{18}$ with minor modifications. Briefly, bands of interest were cut into pieces, and transferred to Eppendorf tubes. Gel pieces were sequentially washed with washing buffer (25 mM sodium bicarbonate, pH 8) and dehydrated with ACN. Proteins were reduced in-gel by incubating for 30 min at
56°C in reduction buffer (10 mM DTT in washing buffer) and dehydrated with ACN prior to in-gel cysteine alkylation for 20 min at RT in alkylation buffer (55 mM IAA in washing buffer) protected from light. Samples were repeatedly washed and dehydrated with washing buffer and ACN, respectively, until gel pieces were completely de-stained. Gel particles were dried in a centrifugal vacuum concentrator at 30°C for a maximum of 5 min. Subsequently, 20 µl to 30 µl PNGase F solution (2 U in 2% NP-40 and 2.5xPBS) were added. Samples were kept on ice for 1 h to let the gel absorb the buffer and enzymatic cleavage was allowed to take place overnight at 37°C. Next, the supernatant containing N-glycans was collected, 20 µl of water were added to the gel pieces and after 20 min of sonication the supernatant was added to the first extraction. The combined supernatants containing the N-glycans were further processed as described below.

**Derivatization of released N-glycans and hydrophilic interaction liquid chromatography (HILIC)-solid phase extraction (SPE) glycan enrichment**

Released N-glycans from the in-gel PNGase F treatment were derivatized by ethyl esterification as described in Reiding et al. with small adaptations. Shortly, 2 µl of released N-glycans were added to 20 µl of ethylation reagent (0.25 M EDC and 0.25 M HOBt in ethanol, 1:1) in quadruplicate, and incubated for 1 h at 37°C (in desiccator protected from evaporation). Subsequently, 20 µl ACN were added and the mixture was incubated at 20°C for 30 min. Samples were let to reach room temperature prior to glycan purification by HILIC-SPE modified from a protocol described previously. Pipette tips of 20 µl volume were packed with 3 mm cotton thread, washed by pipetting 3 x 20 µl water, followed by equilibration with 3 x 20 µl 85% ACN. Samples were loaded by carefully pipetting up-and-down for 30 times. Unbound material was removed by sequential pipetting of 3 x 20 µl 85% ACN/1% TFA and 3 x 20 µl 85% ACN, followed by elution of N-glycans in 10 µl water.

**MALDI-TOF/TOF-MS/MS analysis**

For mass spectrometric analysis, 5 µl of released, derivatized, and purified N-glycans were spotted onto an anchor chip MALDI target plate (Bruker Daltonics, Bremen, Germany) and co-crystallized with 1 µl of 5 mg/ml sDHB in 50% ACN/50% water containing 1 mM NaOH. Samples were allowed to dry at room temperature. MALDI-TOF-MS spectra were acquired using an UltraflxeXtreme mass spectrometer (Bruker Daltonics) in positive ion reflector mode, controlled by flexControl 3.4 software Build 119 (Bruker Daltonics). The instrument was externally calibrated using a peptide calibration kit (Bruker Daltonics). 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 (2000 Hz laser frequency, 200 shots per raster spot during complete random walk). Tandem MS (MALDI-TOF/TOF-MS/MS) was performed for structural elucidation via fragmentation in gas-off TOF/TOF mode.

**Data processing and evaluation of MALDI-TOF-MS spectra**

The N-glycan profiles obtained by MALDI-TOF-MS were internally re-calibrated on the following peaks m/z 1982.700 (hexose (Hex)₅ N-acetylhexosamine (HexNAc)α₂,6 N-acetylneuraminic acid (NeuAc)₂), m/z 2255.793 (Hex₅HexNAcα₂,3NeuAc₁ α₂,6NeuAc₁), m/z 2301.835 (Hex₅HexNAcα₂,6NeuAc₂), m/z 2447.893 (Hex₅HexNAcα₂,6NeuAc₃ fucose(Fuc)₁), m/z 2940.052 (Hex₆HexNAcα₂,3NeuAc₁α₂,6NeuAc₂), m/z
3086.113 \((\text{Hex,HexNAc,}\alpha_2,3\text{NeuAc,}\alpha_2,6\text{NeuAc,Fuc}_1)\), \(m/z\) 3532.227 \((\text{Hex,HexNAc,}\alpha_2,3\text{NeuAc,}\alpha_2,6\text{NeuAc})\), and \(m/z\) 3578.269 \((\text{Hex,HexNAc,}\alpha_2,3\text{NeuAc,}\alpha_2,6\text{NeuAc})\) using the cubic enhanced algorithm in FlexAnalysis Software (Version 3.3 build 65; Bruker Daltonics). Spectra were further smoothed (Savitzky Golay algorithm, peak width: \(m/z\) 0.06, 4 cycles) and baseline-corrected (Tophat algorithm). A composition list for targeted data extraction was generated comprising peaks with signal/noise>3, proper isotopic pattern as judged by visual inspection, and matching glycan composition determined with GlycoWorkbench 2.1 stable build 140 (European Carbohydrates DataBase project; http://www.eurocarbdb.org/) with accuracy of 0.3 Da.

The calibrated spectra, together with the reviewed composition list were applied to an in-house developed software for automated data processing MassyTools 1.0,\(^{21}\) which extracted the area under the curve for each peak of the mass list. Background defined as the lowest area observed in a \(m/z\)-window of 20 Th around the analyte was subtracted from intensities of all isotopic peaks. Furthermore, peaks in the spectra with negative intensity values were set to an intensity value of zero and areas under the curve were normalized to a total relative intensity of 100% for each spectrum. At the end, only glycan compositions which have been confirmed by MS/MS and their directly related compositions (± one monosaccharide) were taken into account for relative quantitation. Next, derived glycosylation traits such as \(\alpha_2,6\)-sialylation, \(\alpha_2,3\)-sialylation, fucosylation, antenna fucosylation (defined as amount of fucoses equal or higher than two) were calculated by summing the relative intensities of the relevant peaks, and averaged within the four technical replicates. GraphPad Prism Version 5.04 (2010, GraphPad Software, Inc., La Jolla, CA) was used for visualization.

**N-and O-glycopeptide analysis**

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

For samples to be treated with exoglycosidases, de-N-glycosylated was achieved using an in-solution PNGase F 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**

Gel bands were prepared for in-gel digestions as described for the in-gel PNGase F treatment. In this case, 25 mM AmBiC was used as washing buffer and PNGase F treatment was replaced by proteolytic digestion using either 0.15 µg of trypsin, 1 µg of Proteinase K or 1 µg Pronase in 30 µl of 25 mM AmBiC. After overnight incubation at 37°C, the supernatant was collected and additional 20 µl of 25 mM AmBiC were added to the gel pieces and incubated at 37°C for 1 h. Both supernatants were combined and stored at -20°C until further analysis of the intact glycopeptides.

Except for glycopeptide samples that were further treated with exoglycosidases, all de-
N-glycosylated glycopeptides were obtained by in-gel PNGase F treatment followed by in-gel proteolytic digestion. In-gel PNGase F treatment was performed as described above. The gel pieces were then washed twice for 5 min with 100 µl 25 mM AmBiC, followed by 100 µl ACN. Gel particles were dried in a centrifugal vacuum concentrator (Eppendorf, Hamburg, Germany) at 30°C for a maximum of 5 min and in-gel proteolytic digestion was performed as described in the first part of this section.

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), and 1 µl of galactosidase (5 mU). The digestions were carried out overnight at 37°C.

C18-PGC-LC-ESI-QTOF-MS/MS
C18-PGC-LC-ESI-QTOF-MS/MS analysis was performed as described previously\textsuperscript{25} 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.\textsuperscript{25} 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. For a selected set of samples lower-stepping-energy was applied with the same settings as above, only with stepping energy is set to 60% and 80% each half of the time (instead of 80%-140%).

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 x 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% FA (v/v)) and solvent B (95% acetonitrile/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. To acquire data for relative quantitation the MS was operated in MS only mode. For 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). 3,4-hexanedione was used to form reagent radical ions. The acquisition of ETD spectra was performed in a targeted manner using an inclusion list for auto MS/MS. The parameters were set as followed: 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.

Data evaluation of LC-MS/MS spectra
DataAnalysis 4.2 software (Bruker Daltonics) was used to analyze glycopeptides of C1-Inh by manually scanning for glycan oxonium ions (m/z 204.0866, [HexNAc]+H; m/z 366.1395, [Hex,HexNAc]+H; m/z 657.2349, [Hex,HexNAc,NeuAc]+H; m/z 528.1923, [Hex,HexNAc]+H). The defined glycopeptide spectra were analyzed manually to identify the glycan structure and the mass of the peptide backbone as described previously. The determined peptide moiety masses were then searched using the ExPASY FindPept software tool (http://www.expasy.org/tools/findpept.html), which includes a list covering theoretical peptide sequences generated by C1-Inh digestion. The search parameters for the Findpept tool were set with a mass tolerance of 0.01 Da, selecting the enzyme settings as “no enzyme”, and IAA treated cysteines. Glycopeptide fragment spectra obtained by stepping-energy CID using QTOF-MS/MS featured
characteristic peptide b- and y-ions next to glycan-derived B- and Y-type ions and thus, allowing direct confirmation of the peptide backbone. The resulting MS/MS spectra were further analyzed to manually match b- and y-ions.

For relative quantitation signal intensities of all tryptic glycopeptides and partially also of mis-cleaved ones were extracted in an automated manner using LaCy tools (version 1.0.0.27). The analyte list was curated based on manual data analysis. 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.

Results

**SDS-PAGE analysis of C1-Inh**

C1-Inh was first analyzed by SDS-PAGE with and without PNGase F treatment, resulting in broad bands of apparent molecular masses of 65-75 kDa and 75-90 kDa, respectively (Suppl. Figure S1). This suggests an N-glycan contribution of approximately 15 kDa. The calculated molecular mass of the C1-Inh peptide portion is approximately 55 kDa. Thus, the diffuse bands of the PNGase F treated C1-Inh indicate an O-glycosylation contribution of more than 10 kDa.

**O-glycosylation analysis of C1-Inh**

C1-Inh was subjected to in-gel Pronase, Proteinase K and trypsin treatment for O-glycosylation site identification and characterization. To reduce sample heterogeneity and enhance O-glycosylation site identification a portion of these digestes were additionally de-N-glycosylated by PNGase F. Pronase and Proteinase K-treated samples were then further 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 HexNAc or HexNAc-Hex moiety attached to the O-glycosylation site and therefore, allowing more reliable site-specific analysis. Samples were analyzed by C18-PGC-LC-ESI-QTOF-MS/MS and C18-RP-LC-ESI-QTOF-MS/MS analysis applying different tandem MS modes to obtain more detailed glycopeptide information. Next to the targeted ETD mode, a combined higher- and lower-energy CID (stepping-energy CID) was applied, resulting in b- and y-type peptide fragments from elevated collision energies, in addition to B- and Y-type glycan-derived fragments, due to lower collision
Figure 1: 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 as OG or NG, respectively, 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. Green circle = mannose; yellow circle = galactose; blue square = N-acetylglucosamine; red triangle = fucose; purple diamond = N-acetylnuraminic acid.

Additionally, C18-RP-LC-ESI-IT-MS/MS was performed in targeted ETD mode. The C18-PGC-LC-ESI-QTOF-MS/MS approach has recently been developed by us and facilitates detection of more hydrophilic glycopeptides that often result from Pronase and Proteinase K treatment and are not retained using C18-RP-LC only. In total, 27 O-glycosylation sites were identified on human C1-Inh, all present on the N-terminal domain of the protein (Figure 1). Ten of them (Ser31, Thr47, Thr48, Ser51, Ser63, Ser64, Thr67, Thr71, Thr72 and Thr76) were detected with their exact
location. Additionally, one O-glycosylation site was identified either Thr27 or Ser28, next to a heavily O-glycosylated region between Asn81-Ser121 with up to 16 occupied O-glycosylation sites. A list of all O-glycopeptides found in this study including glycan composition and peptide sequence is given in Suppl. Tables S1. A detailed identification of all O-glycosylation sites is described below.

**Determination of the C1-Inh O-glycosylation sites**

Stepping-energy CID O-glycopeptide spectra were classified by scanning the MS/MS data for the characteristic oxonium ion at \( m/z \) 204.0867 [HexNAc\_Hex\_γ+H]. Manual spectra annotation was performed for O-glycopeptide identification based on the exact mass and glycan-derived B- and Y-ions as well as peptide-derived b- and y-ions.\(^{26}\) The O-glycosylation site Thr76 was identified from a Pronase-treated glycopeptide at \( m/z \) 488.2123 [M+2H]\(^2+\), carrying a monosialylated core 1 structure. From the tandem MS spectrum the peptide ion \( m/z \) 319.1951 [M+H]\(^+\) was deduced, corresponding to the sequence \_75ATK\_77, which was confirmed by the exact mass and peptide fragment ions (Figure 2A).

![Figure 2: Stepping-energy fragmentation spectra of various glycopeptides confirming O-glycosylation sites of C1-Inh. MS/MS spectra are shown for (A) a Pronase-treated glycopeptide with the peptide sequence \_75ATK\_77 carrying a monosialylated core 1 structure, confirming the O-glycosylation site Thr76; (B) a tryptic glycopeptide with the peptide sequence \_45VATTVIS\_52 with three monosialylated core 1 structures attached, confirming the O-glycosylation sites Thr47, Thr48 and Ser51; (C) a Pronase-treated glycopeptide with the peptide sequence \_23NPNAT\_28 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; (D) a Pronase-treated glycopeptide after exoglycosidase treatment with the peptide sequence \_71TTN\_73 carrying two monosialylated core 1 structure, confirming the O-glycosylation site Thr71 and Thr72.](image-url)
Three O-glycosylation sites at Thr47, Thr48 and Ser51 were observed in close proximity to each other, occurring on the same glycopeptides. The tryptic glycopeptide $\text{VATTVISK}_{45}^\text{Thr}^47$, carrying three core 1 monosialylated O-glycopeptides, was observed at $m/z$ 929.7329 $[M+3H]^3$ (Figure 2B). The peptide fragment ion $m/z$ 818.4993 $[M+H]^+$ was deduced from the tandem MS spectrum and the peptide sequence was confirmed by several b- and y-ions. Since only the glycan fragment ion HexNAc$_1$Hex$_1$NeuAc$_1$ and not HexNAc$_1$Hex$_1$NeuAc$_2$ was observed, it was concluded that all glycosylation sites of this glycopeptide carry the same monosialylated O-glycan. However, different glycopeptides covering the same glycosylation sites also show that at least one of them is also partially occupied by non-sialylated core 1 O-glycans. Likewise, two of these glycosylation sites were found to carry a disialylated core 1 structures (Suppl. Table S1B), showing the microheterogeneity of these glycosylation sites.

Another O-glycosylation site was identified either on Thr27 or Ser28. The Pronase-treated glycopeptide at $m/z$ 1058.3974 $[M+3H]^3$ with the peptide portion $\text{NPNATS}_{23}^\text{Ser}^28$ was observed with an N-glycan attached to Asn25 (HexNAc$_4$Hex$_5$NeuAc$_1$) and a monosialylated core 1 O-glycan (Figure 2C). The remaining N-glycan portion is due to an incomplete in-solution N-glycans release. The peptide portion was identified by the exact mass and several b- and y-ions. The diagnostic glycan-derived fragment peptide $\text{HexNAc}_1\text{Hex}_1\text{NeuAc}_1$ confirms the presence of the monosialylated O-glycan, however, the exact position was not located since there are two theoretical O-glycosylation sites present, Thr27 and Ser28. The analysis of tryptic de-N-glycosylated O-glycopeptides revealed the presence of non-, mono- and disialylated core 1 O-glycans attached to Thr27/Ser28 and Ser31 (Suppl. Table S1B).

To further reduce sample heterogeneity, Pronase- and Proteinase K-treated samples were subjected to exoglycosidases treatment with sialidase and galactosidases, which should trim down the glycan portion to a single HexNAc per glycosylation site. This allowed identification of the two glycosylation sites Thr71 and Thr72. The glycopeptide $\text{TTP}_{71}^\text{Thr}^73$ 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 2D). 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 $\text{TTNS}_{67}^\text{Asp}^70$. Similarly, glycopeptides with the peptide sequences TTNS were found, which could as well correspond to the peptides $\text{TTNS}_{74}^\text{Asp}^70$, or $\text{TTNS}_{70}^\text{Asp}^70$. Ultimately, the same sample also contained a glycopeptide with the unique peptide sequence $\text{TTNSAT}_{76}^\text{Thr}^77$, confirming the glycosylation sites Thr71 and Thr72, since no further cleavage products were found to support the other sequence (Suppl. Table S1C). 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 12 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 3A). 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 nature of the third heavily glycosylated peptide is discussed later in the N-glycosylation section.
Figure 3: 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\(_1\) = TANTTDEPTTQPTTEPTQPTIQPTQPTQLPTDSPTQPTTG\(_S\); Pep2\(_2\) = DTDEPTQPTTEPTQPTIQPTQPTQLPTDSPTQPTTG\(_S\); Pep3\(_3\) = TTDEPTQPTTEPTQPTIQPTQPTQLPTDSPTQPTTG\(_S\) (D indicates the deamidated Asn caused by PNGase F treatment) (B) Stepping–energy CID spectrum of Pep2 with 14 HexNAc residues, indicating 14 occupied O-glycosylation sites. (C) Stepping–energy CID spectrum of the same glycopeptide as in panel B using lower–energy stepping–energy CID with a focus on the glycan–derived Y–ions.
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 3A). The glycopeptide at m/z 1437.0252 [M+5H]^5+ covered 14 HexNAc residues (\(\text{DTTDEPTQPTTEPTTQPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQLDSPTQPTTGS}\)). D indicates the deamidated Asn caused by PNGase F treatment (Figure 3B). 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 3B). 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 3C). Since O-glycosylation sites do not have a common consensus sequence it is hard to identify them in CID mode when several possible sites are present on one glycopeptide. Thus, we applied ETD-fragmentation using QTOF and IT instruments to a list of targeted glycopeptides (indicated in Suppl. Table S1) to identify the exact location of the glycosylation sites, as described in the following. QTOF-ETD fragmentation was applied to the Proteinase K-, PNGase F- and exoglycosidases-treated glycopeptide at m/z 731.8255 [M+2H]^2+, which contains two HexNAcHex, O-glycans attached to the peptide sequence \(61\text{EVSSLPT}_{67}\) (Figure 4A).

\[\text{Figure 4: ETD fragmentation spectra of various glycopeptides confirming O-glycosylation sites of C1-Inh. MS/MS spectra are shown for (A) a Proteinase K-treated glycopeptide after exoglycosidase treatment with the peptide sequence }\text{EVSSLPT}_{67}\text{ carrying two HexNAcHex, structures, confirming the O-glycosylation sites Ser63 and Ser64. The sample was analyzed using QTOF-MS/MS. (B) The Proteinase K-treated glycopeptide after PNGase F and exoglycosidase treatment with the peptide sequence }\text{SLPTD}_{69}\text{ (D indicated deamidation after PNGase F treatment.) and HexNAcHex, O-glycan attached, confirming the O-glycosylation site Thr67. The sample was analyzed by QTOF-MS/MS. (C) The Pronase-generated glycopeptide }\text{SSQDPESL}_{37}\text{ with a monosialylated core 1 glycan attached, confirming the O-glycosylation site Ser31. The sample was analyzed by ion trap-MS/MS.}\]
The diagnostic ions c4 and c6 exclude Thr67 as glycosylation site and confirm the occupation of Ser63 and Ser64. The same approach was used to identify the O-glycosylation site Thr67 (Figure 4B). The Proteinase K-treated glycopeptide m/z 499.7259 [M+2H]²⁺ carries one HexNAc₁Hex₁ O-glycan attached to the peptide sequence ⁶₄SLPTTD₆₉ (D indicates the deamidated Asn caused by PNGase F treatment). The fragment ions c₄ and z₅ identify the O-glycosylation site Thr67.

The glycosylation site Ser31 was identified by the Pronase-treated glycopeptide m/z 759.78 [M+2H]²⁺ using IT-ETD fragmentation (experimental procedure is provided in the Supplementary Information). A monosialylated core 1 O-glycan was detected attached to the peptide portion ᵃ₀SSQDПESL₆₇ (Figure 4C). The fragment ions z₈ and z₇ with the O-glycan attached were observed next to c₄ and c₅ peptide + glycan fragments identifying Ser31 as glycosylation site. The analysis of the tryptic de-N-glycosylated O-glycopeptides showed that the sites Ser31 and Thr27/Thr28 together carry core 1 O-glycans in their non-, mono-, and disialylated form. However, except for the monosialylated species, which was present on both sites, it could not be determined which site carries which glycan.

### N-glycosylation analysis of C1-Inh

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

Six potential N-glycosylation sites are present on C1-Inh (Uniprot P05155). The glycosylation sites of Asn25, Asn69, and Asn81 are located on the N-terminal domain of C1-Inh, while the serpin domain contains Asn238, Asn252, and Asn352. To get an overview of the overall C1-Inh N-glycosylation, released N-glycans were profiled using linkage-specific sialic acid derivatization and MALDI-TOF-MS analysis (Figure 5). The derivatization stabilizes N-acetylneuraminic acids (NeuAc) in MALDI-TOF-MS measurements and allows discrimination of NeuAc linkage (between α₂,3- and α₂,6-linked NeuAc) by inducing mass shifts of -18.01 Da due to lactonization of α₂,3-linked NeuAc and +28.03 Da due to ethyl esterification of α₂,6-linked NeuAc. Glycan structures were identified based on MS/MS data as well as literature (Suppl. Figure S2, Suppl. 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 (Hex₅HexNAc₄(α₂,6)NeuAc₂; 2301.835 Da [M+Na]+; Figure 5).

Approximately 90% of C1-Inh N-glycans contained at least one α₂,6-linked NeuAc. Besides that, about 30% featured α₂,3-linked NeuAc containing N-glycans with mixed structures (Figure 5 inset). Approximately one quarter of the N-glycans were monofucosylated and MS/MS analysis identified them to be mainly core linked, as indicated by a loss of fucose with 146 Da from the precursor prior to a loss of the reducing end N-acetylgalactosamine (GlcNAc; -221 Da) (Suppl. Table S1). Multi-fucosylation (two or more fucoses), indicative of antenna fucosylation, was less than 5% (Figure 5 inset). 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%; Figure 5 inset).
Figure 5: MALDI-TOF-MS analysis of C1-Inh N-glycans after PNGase F release and linkage-specific sialic acid derivatization. Glycans were assigned on the basis of MS/MS data (Suppl. Figure S3, Suppl. Table S1) and literature. 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. drk grey circle = mannose; light grey circle = galactose; dark grey square = N-acetylglucosamine; triangle = fucose.

Site-specific N-glycopeptide analysis of C1-Inh
In order to identify N-glycosylation sites and glycan microheterogeneity, proteolytic treatment was carried out using trypsin, Proteinase K and Pronase. The samples were analyzed by C18-RP-LC-ESI-IT-MS/MS and C18-PGC-LC-ESI-QTOF-MS/MS and glycopeptide spectra were classified and identified as described for the O-glycopeptides. All six potential N-glycosylation sites of C1-Inh were identified and are listed in Suppl. Table S3. The corresponding glycoforms are depicted in Figure 1. Trypsin digestion of C1-Inh allowed identification and characterization of glycosylation sites Asn238, Asn253 and Asn352. Glycopeptides were identified based on 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 (Suppl. Table S3A). These three N-glycosylation sites carry a similar repertoire of glycans with di-, tri- and tetaantennary structures, with and without core fucosylation and different degrees of sialylation from non- to disialylation (Figure 1, Suppl. Table S3A). Asn238 was also observed with glycans carrying three and four sialic acids. Representative stepping-energy CID tandem MS spectra of these three sites are shown in Suppl. Figure S3B-D. The different glycoforms present on Asn25 were identified by Pronase-treated glycopeptides. Since there are two O-glycosylation sites in close proximity, Pronase digestion was used to generate smaller peptide portions with a maximum of one extra O-glycosylation site, which allowed better N-glycosylation site characterization (Figure 2C and Suppl. Table S3C).
Site-specific elucidation of the N-glycoforms was only possible for glycopeptide precursors that were selected for tandem MS. Thus, fewer glycoforms as compared to Asn238, Asn253 and Asn352 were identified including di- and triantennary N-glycans with and without core fucosylation ranging from mono- to trisialylation (Figure 1, Suppl. Table S3C).

Similarly, glycosylation site Asn69 was characterized by Proteinase K-treated glycopeptides. This glycosylation site is surrounded by six O-glycosylation sites, which does not allow the detection after trypsin digestion due to the glycopeptide
microheterogeneity. However, three different glycopeptides containing the 
N-glycosylation site Asn69 could be detected after Proteinase K-treatment. The 
N-glycans were non-corefucosylated, di- and triantennary structures with two and three 
sialic acids (Figure 1, Suppl. Figure S3A, Suppl. Table S3C).
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 
\( \gamma \),TANTTDEPTTQPTTEPTQPTQPTQPTQPLPTDSPTQPTTGS\(_{121}\) featuring the glycan 
composition HexNAc\(_{19}\)Hex\(_{3}\) (Figure 3A). The stepping-energy CID spectrum of this 
peptide at \( m/z \) 1476.4602 [M+6H]+ with the glycan portion HexNAc\(_{19}\)Hex\(_{3}\) contains 
several peptide fragment ions to confirm the peptide sequence Thr79-Ser121 (Figure 
6A). Diagnostic oxonium ions clearly indicate the presence of an N-glycan, such as \( m/z \) 
528.1889 (HexNAc\(_{1}\)Hex\(_{2}\)), \( m/z \) 690.2411 (HexNAc\(_{1}\)Hex\(_{3}\)), \( m/z \) 893.3202 (HexNAc\(_{2}\)Hex\(_{3}\)) 
and \( m/z \) 1096.3969 (HexNAc\(_{3}\)Hex\(_{3}\)). A lower-energy CID fragmentation spectrum with 
a zoom-in of the peptide + HexNAc fragments of this highly glycosylated peptide further 
proved the identity of this glycopeptide (Figure 6B). However, due to the complexity of 
this glycopeptide and the application of exoglycosidases it is not possible to characterize 
the glycan microheterogeneity of Asn81.

**Relative quantification of C1-Inh N- and O-glycan microheterogeneity**

The relative abundances of the different glycoforms attached to the corresponding 
glycopeptides were extracted in an automated manner using the software package 
LaCytools.\(^{27}\) The relative intensities for each glycosylation site were normalized to the 
total intensity of all glycoforms attached to the same peptide moiety (Suppl. Table S4, 
Figure 7). For the N-glycosylation sites Asn238, Asn253 and Asn352 the glycopeptide 
profiles were directly obtained from their tryptic glycopeptides (Figure 7A, Suppl. 
Table S4). For Asn238, 22 different N-glycoforms were identified and 21 of them were 
quantified. With approximately 52.0% HexNAc\(_{4}\)Hex\(_{5}\)NeuAc\(_{2}\) is the most abundant one 
(Figure 7A, Suppl. Table S3A). The vast majority of the N-glycans are diantennary 
(77.0%), followed by triantennary (21.4%) and tetraantennary ones (1.6%). While 
the glycans are almost completely sialylated (98.6%) only 20.1% are fucosylated. These 
results are in accordance with the released N-glycan profiles.
The glycosylation site Asn253 was identified with 12 different glycoforms of which 
11 could be quantified in a relative manner (Figure 7A, Suppl. Table S4). Also for 
this glycosylation site the most abundant structure is HexNAc\(_{3}\)Hex\(_{3}\)NeuAc\(_{2}\) (45.8%). 
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.
Ten N-glycoforms were identified for Asn352, but only six of them could be used for 
relative quantitation due to the low intensity of the remaining glycopeptides. Next to 
four diantennary sialylated N-glycans also two triantennary sialylated N-glycan was
Figure 7: 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 of the glycoforms were calculated and the standard deviation is based on multiple analyses with different proteolytic enzymes. (A) Relative distribution of the tryptic N-glycopeptides attached to Asn238, Asn253 and Asn352. (B) Relative distribution of the tryptic glycopeptide \( ^{23} \text{NPNATSSQDPSLQDR} \) containing N-glycans attached to Asn25 and O-glycans attached to Thr27/Ser28 and/or Ser31. The inset 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 and Ser51. (D) Relative distribution of the tryptic O-glycopeptides covering the sites Ser63, Ser64, Thr67, Thr71, Thr72, Thr76. (Deamidation indicated after PNGase F treatment)
quantified. Again HexNAc₄Hex₃NeuAc₂ was the most abundant structure with 77.7%. Four of these N-glycans were fucosylated representing 9.6%.

The N-glycosylation site Asn25 occurred in the tryptic peptide \( _{25} \text{NPNATSSSSQDPESLQDR} \), that also contains the O-glycosylation site Ser31 and one O-glycosylation site, which is either Thr27 or Ser28. Fragmentation spectra of the tryptic glycopeptide indicate at least partial presence of O-glycans next to the N-glycan. Thus, the relative quantitation of the glycan moiety present on this glycopeptide is indicated as composition due to a mixture of N- and O-glycans (Figure 7B). In total, 18 different glycan compositions were identified for the tryptic peptide around Asn25 and were quantified in a relative manner. To identify the nature of the O-glycan microheterogeneity the de-N-glycosylated species of this peptide was analyzed after PNGase F treatment (Figure 7B inset). Around 81.7% of the O-glycopeptide contains only one monosialylated core 1 O-glycan and in total around 91.1% show O-glycosylation of only one of the two O-glycosylation sites with core 1 glycan carrying no, one or two sialic acids. Considering that fact, it can be assumed that most of the relative signal intensities of the N- and O-glycosylated peptide contains one core 1 O-glycan. However, also the non-O-glycosylated peptide species after de-N-glycosylation \( _{25} \text{NPDATSSSSQDPESLQDR} \); D indicates deamidation after PNGase F treatment) was found (data not shown), indicating that not all N-glycopeptides carry also O-glycans. With this knowledge about the O-glycan portion it is possible to get further insights into the N-glycan moiety of Asn25. The most abundant glycan composition of the tryptic glycopeptide is HexNAc₄Hex₅NeuAc₂, which is most likely 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 HexNAc₅Hex₆, 4.1% HexNAc₆Hex₅ and 0.6% HexNAc₇Hex₄. 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.

The N-glycosylation site Asn69 is located in close proximity of six O-glycosylation sites, which hinders its site-specific analysis. It was not possible to detect the tryptic glycopeptide, due to the vast microheterogeneity. Also relative quantitation based on Pronase-treated glycopeptides was not considered useful, since this approach only revealed three different glycoforms (Suppl. Table S3C) as a result of the microheterogeneity and thus biased proteolytic cleavage. Similarly, also the tryptic peptide around Asn81 contained several O-glycosylation sites that only allowed the detection of the site but not the mapping of the microheterogeneity.

Overall, the glycopeptide-based N-glycoform distribution of the four N-glycosylation sites was in good agreement with the total N-glycoform pattern determined by MALDI-TOF-MS of released N-glycans (Figure 1). Further, based on the released N-glycan analysis it can be concluded that the vast majority of the observed sialylated N-glycoforms contain \( \alpha 2,6 \)-linked sialic acid.

The O-glycopeptide samples were analyzed after enzymatic de-N-glycosylation. The tryptic glycopeptide \( _{45} \text{VATTVISK} \) was identified next to the mis-cleaved glycopeptide species \( _{41} \text{GEGKVATTVISK} \) (Figure 7C and Suppl. Table S3B). The completely cleaved glycopeptide mainly contains one monosialylated (56.0%) or disialylated core 1 O-glycans (14.8%). In total, this glycopeptide contains three potential O-glycosylation sites, which is in 72.7% of the cases occupied on one site, followed by 26.8% on two and
0.5% three on sites. However, also the mis-cleaved glycopeptide species was detected, carrying mainly two occupied glycosylation sites (98.9%) next to 1.1% full occupation. The different relative glycosylation of these two glycopeptides indicates a bias in tryptic digestion based on the glycosylation profile of the corresponding sites.

Similarly, the tryptic O-glycopeptide with the sequence $\text{MLFVEPILEVSSLPTD}$ST$^{D}$TNSATK$_{77}$ (D indicates deamidation after PNGase F-treatment) was analyzed (Figure 7D and Suppl. Table S3B). The peptide sequence contains up to six identified O-glycosylation sites, however, the relative quantitation only revealed simultaneous occupation of up to three sites. The major glycoforms is one monosialylated core 1 O-glycan (49.5%), followed by the presence of one monosialyted and one nonsialylated core 1 O-glycan at the same time (39.8%). Only smaller amounts of three occupied O-glycosylation sites were found (6.5%). O-glycopeptides with more than three occupied sites might have been not present or not detected due to their low signal intensities.

Discussion

In this study, we present a comprehensive glycoproteomic analysis of human C1-Inh based on two different approaches: firstly, N-glycomic analysis of released N-glycans was carried out with linkage-specific sialic acid derivatization followed by MALDI-TOF-MS analysis. Secondly, site-specific glycosylation analysis was performed on N- and O-glycopeptide level through the analysis of proteolytically cleaved C1-Inh. For this, various proteases were used, as well as enzymatic de-N-glycosylation and exoglycosidase treatment. Different MS-based strategies were employed, including C18-RP-LC and C18-PGC-LC setups followed by ESI-MS/MS with stepping-energy CID and ETD. By combining these workflows, we gained extensive information with respect to site-specific N- and O-glycosylation of C1-Inh.

In the current study on C1-Inh the use of two broad-specificity proteases – Pronase and Proteinase K – resulted in smaller peptide portions and reduced sample complexity for tandem MS analysis. Proline-, serine-, threonine-, and alanine-rich domains of extracellular proteins and protein portions can be highly O-glycosylated. Thus, O-glycosylation sites are often located close to each other, which may result in multiple glycosylation sites per peptide, which further complicates site-specific analysis. While this is particularly the case when highly specific proteases are used, cleavage by broad-specificity proteases mainly results in short peptide portions with a lower number of O-glycans. O-glycosylation site coverage was increased by additional exoglycosidase treatment using sialidase and galactosidase to further reduce sample heterogeneity. This strategy has previously been shown to be suitable for CID-based mass spectrometric characterization of O-glycopeptides that bear core 1 O-glycans with or without the presence of sialic acid residues. Besides that, the use of an integrated C18-PGC-LC platform in combination with Proteinase K and Pronase treatment enabled specifically the detection of glycopeptides with a smaller peptide backbone, which are too hydrophilic to be retained by C18 only. Combining these methods improved the identification and
location of glycans on C1-inh and lead to discovery of seven novel O-glycosylation sites. Remarkably, we were able to detect a total of 27 O-glycosylation events on C1-inhibitor. Of these, 10 were assigned to a specific site in the N-terminal domain (Ser31, Thr47, Thr48, Ser51, 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 16 O-glycans detected, we were able to demonstrate by MS/(MS) a heavily O-glycosylated region (Thr82-Ser121) with up to 16 occupied O-glycosylation sites out of 18 possible sites, showing that this region shows on average a very high O-glycosylation site occupancy.

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 cerebrospinal glycoproteins and human urine containing core-1 O-glycans and were confirmed in our study. Furthermore, Ser64 and Ser71 are known from amino acid sequencing studies and were for the first time confirmed by our mass spectrometric study. In addition, the occupation of sites Thr83, Thr88, Thr92 and Thr96 which has previously been revealed by amino acid sequencing. Bock et al. also suggested the O-glycosylation of Thr99, Thr106, Thr107, Thr111, Thr115, Thr118 and Thr119, however, they could not provide unequivocal evidence. Their findings are largely in-line with our mass spectrometric data which demonstrated occupation of at least 16 of the 18 possible O-glycosylation sites in the region Thr82-Ser121. Hence, the MS/(MS) analysis of Thr82-Ser121 demonstrated for the first time the excessive O-glycosylation of this region. Notably, to the best of our knowledge the O-glycosylation sites Ser31, Ser51, Ser63, Thr67, Thr72, Thr76 as well as the O-glycosylation region Thr27/Ser28 are reported for the first time.

The analysis of this glycan-rich region Thr82-Ser121 on a site-specific level was hampered due to steric hindrance caused by the heavy glycosylation of C1-Inh that prohibited further proteolytic cleavages in this region, demonstrating some limitations of the chosen analytical approach. Likewise, ETD fragmentation for site analysis provided spectra of insufficient quality for O-glycosylation site analysis.

In the current study we referred to core 1 O-glycans only, since we could not find evidence in the MS/MS spectra for any other O-glycan core. Furthermore, core 1 O-glycans have also been reported in literature to be the major O-glycan core of serum glycoproteins. We observed O-glycosylation sites occupied with mono-, di-, and non-sialylated core 1-type O-glycans. These findings can be only compared to Thr47 and Thr48, since these sites have been identified before by MS. However, in both studies the glycopeptides were captured using hydrazine chemistry, which included acid hydrolysis to release O-glycopeptides. Using this approach confirmed that the O-glycopeptides contained sialic acid but further information about the glycan microheterogeneity was lost.

The N-glycosylation sites of C1-Inh have been mapped by mass spectrometry, as well as other analytical approaches in various studies. 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 and small amounts of tri- and tetraantennary glycan structures. 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, whereas we also detected small amounts of less than 1% 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. Deglycosylation of C1-Inh with PNGase F, O-glycanase, or both had no detectable impact on its protease inhibition activity. Neither had most of the highly glycosylated N-terminal domain of C1-Inh, as shown by the deletion of the first 98 or 120 N-terminal amino acids. 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. In fact, the N-terminus has been described to be a “rod-like” domain, likely due to the presence of the glycans. 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 proteases. Moreover, we found that the highly O-glycosylated region with 12-16 occupied sites was resistant to protease treatment, demonstrating that glycosylation can prevent C1-Inh from proteolytic degradation.

In conclusion, we demonstrate here a novel, powerful method applied for in-depth site-specific characterization of C1-Inh glycosylation revealing 11 O-glycosylation sites carrying core-1 type O-glycans, with seven 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. We demonstrated that the use of several proteases, including non-specific proteases like Pronase and Proteinase K, in combination with novel LC-MS approaches largely improves identification of glycosylation sites. The use of non-specific proteases has shown its benefit especially for glycan-rich protein areas by reducing sample complexity, which could be even more reduced by exoglycosidase digestion. This newly generated 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.

**Acknowledgements**

H. Mehmet Kayılı 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).

**Associated content**

Complete supplemental information will be online available.
C18-LC-ESI-IT-MS/MS
Pronase-treated (glyco)peptides of C1-Inh obtained from in-gel digestions were diluted 10 times with water and 4 µl were injected into a C18-LC-ESI-IT-MS/MS system consisting of an Ultimate 3000 RSLCnano system (Thermo Scientific, Sunnyvale, CA) coupled to an amazonSpeed ion trap (Bruker Daltonics). Data analysis was performed using DataAnalysis 4.2 (Bruker Daltonics).

Samples were loaded on a pre-column (Acclaim PepMap100 C18 column, 100 µm × 2 cm, 5 µm, 100 Å, Thermo Scientific) prior to separation on an Acclaim PepMap RSLC nano-column (75 µm × 15 cm, 2 µm, 100 Å, Thermo Scientific). A flow rate of 300 nL/min was applied in a multistep linear gradient (t = 0-5 min, c(B) = 3%; t = 35 min, c(B) = 27%; t = 40-45 min, c(B) = 70%; t = 46-58 min, c(B) = 3% with solvent A (0.1% FA in water) and solvent B (95% ACN and 5% water).

All analyses were performed in positive ion mode using a CaptiveSpray 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 1300 V. MS acquisition was performed in enhanced resolution mode within a m/z-range from m/z 400 to m/z 2000 or m/z 600 to m/z 1800. The maximum accumulation time was 200 ms, the ICC target was set to 200000 and SPS to m/z 1100 at a trap drive level of 100%. ETD MS/MS spectra were acquired from m/z 140 to m/z 2200 for a selected list of precursors. The ICC target was set to 100000.

Figure S1: SDS-PAGE analysis of reduced human C1-Inhibitor; (A) native C1-Inh, (B) C1-Inh after PNGase F treatment.
Figure S3: Representative stepping-energy fragmentation spectra confirming four of the six N-glycosylation sites of C1-Inh. MS/MS spectra are shown for (A) a Pronase-treated glycopeptide with the peptide sequence SLPTTNST, carrying a monosialylated core 1 O-glycan and a diantennary disialylated N-glycan, confirming the N-glycosylation site Asn69; (B) a tryptic glycopeptide with the peptide sequence DTFVNASR, with a diantennary disialylated N-glycan, confirming the N-glycosylation sites Asn238; (C) a tryptic glycopeptide with the peptide sequence VLSNNSDANLINTWVK, with a diantennary disialylated N-glycan, confirming the N-glycosylation sites Asn253; (D) a tryptic glycopeptide with the peptide sequence VGQLQLSHNLSLVILVPQNLK, with a diantennary disialylated N-glycan, confirming the N-glycosylation sites Asn352.