Site-specific N-glycosylation analysis of human factor XI: Identification of a noncanonical NXC glycosite

Valegh Faid, Nouha Denguir, Violaine Chapuis, Nicolas Bihoreau and Guillaume Chevreux

Analytical Department, LFB Biotechnologies, Courtaboeuf, France

Human factor XI (hFXI) is a 160-kDa disulphide-linked homodimer zymogen involved in the coagulation cascade. Its deficiency results in bleeding diathesis referred to as hemophilia C. hFXI bears five N-glycosylation consensus sites per monomer, N$_{72}$, N$_{108}$, N$_{335}$ on the heavy chain and N$_{432}$, N$_{473}$ on the light chain. This study reports the first in-depth glycosylation analysis of hFXI based on advanced MS approaches. Hydrophilic interaction LC and MS characterization and quantification of the N-glycans showed that the two major forms are complex biantennary mono-H$_{2}$,2,6-sialylated (A$_{2}$S$_{1}$, 20%) and bis-H$_{2}$,2,6-sialylated structures (A$_{2}$S$_{2}$, 66%). Minor triantennary structures (A$_{3}$S$_{3}$F, 223c 1.5%; A$_{3}$S$_{3}$, 223c 2%) were also identified. MS analyses of intact hFXI revealed full occupation of two of the three heavy-chain glycosites and almost full-site occupancy of the light chain. Analysis of hFXI glycopeptides by LC-MS/MS enabled site-specific glycan profiling and occupancy. It was evidenced that N$_{335}$ was not glycosylated and that N$_{72}$ and N$_{108}$ were fully occupied, whereas N$_{432}$ and N$_{473}$ were occupied at about 92 and 95%, respectively. We also identified a new glycosite of the noncanonical format NXC at N$_{145}$, occupied at around 5%. These data provide valuable structural information useful to understand the potential roles of N-glycosylation on hFXI function and could serve as a structural reference.

Keywords: Coagulation factor XI / Glycoproteomics / Glycosylation / Mass spectrometry / NXC glycosite

1 Introduction

Coagulation is triggered in response to the formation of a tissue factor and activated factor VII complex following vascular injury. This complex activates factor X that acts on prothrombin to generate trace amounts of thrombin, sufficient to activate factor V and factor VIII and to convert factor XI (FXI) into activated FXI (FXIa) by a positive feedback. This leads to an increase in both factor IX (FIX) and factor X activation and thus to the generation of a substantial amount of thrombin required to ensure clot formation. FXI deficiency, also known as hemophilia C, is a rare hereditary coagulation disorder. Its prevalence is estimated at approximately 1:1 000 000. This disease is characterized by bleeding episodes of varying degrees following trauma or surgical procedures [1]. The treatment involves substitution with fresh frozen plasma or human FXI (hFXI) concentrate such as Hemoleven$^\text{®}$.

hFXI is a 160-kDa zymogen made of two identical monomers of 607 amino acids, linked by a [C$_{321}$-C$_{321}$] disulphide bond. Each monomer has four homologous domains in their N-terminal part known as Apple domains, A$_{1}$, A$_{2}$, A$_{3}$, and A$_{4}$, constituting the heavy chain (HC). These domains con-
tain binding sites for platelets [2], heparin [3], prothrombin [4], thrombin [5], high molecular weight kininogen (HMWK) [6], FIX [7], activated factor XII FXIIa [8], and Gp1b protein [9, 10]. The C-terminal part of hFXI contains the catalytic domain that belongs to the trypsin-like serine protease family, characterized by its triad of amino acids H432, D462, and S557. The protein can be activated by FXIIa, thrombin or by its activated form (FXIa). FXI activation corresponds to the cleavage between R470 and I470 generating the N-terminal HC of 47 kDa and the C-terminal light chain of 33 kDa, which remain linked by a [C162-C482] disulphide bond. It has been reported that hFXI dimerization is necessary for its full activation. Activation can generate an intermediate containing an activated monomer and a nonactivated one [11], the first activated monomer can then activate the second one by a transactivation mechanism [12].

hFXI has five consensus sites for N-glycosylation NXS/T per monomer (X being any amino acids except Pro). Three sites are located on the HC at positions N108 (A1 domain), N108 (A2 domain), and N135 (A4 domain) and two sites are located on the light chain at positions N432 and N673, within the catalytic domain. Three glycosites, N108, N332, and N673, are conserved on the human plasma kallikrein) with which hFXI shares almost 58% sequence homology [13]. However, no information is available so far about the role played by the N-glycosylation on the modulation of hFXI physical/biological properties. In this report, we carried out the first in-depth MS-based characterization of hFXI N-glycosylation. After peptidyl-N-glycosidase (PNGase) F release, N-glycans were identified in their native forms using MALDI-TOF MS and quantified by HILIC–UPLC/FD, which is hydrophilic interaction LC-ultraperformance LC coupled with fluorescence detection, after fluorescent labeling. N-glycosylation occupancy was assessed by MALDI-TOF MS on the intact glycoprotein before and after PNGase F treatment. The average number of occupied glycosites was calculated taking into account the mass of the major N-glycan structure previously determined. After activation, hFXIa HC and light chain were profiled by LC-MS to get additional information about N-glycosylation occupancy of each chain. Finally, the protein was digested by Lys-C and trypsin for a site-specific glycosylation analysis of hFXI by LC-MS/MS providing a precise determination of glycosite heterogeneity and occupancy.

2 Materials and methods

2.1 Source of hFXI

hFXI was derived from Hemoleven® (LFB Biomédicaments), the only available hFXI concentrate produced and commercialized in France, after an additional RP-HPLC step to remove excipients. RP-HPLC purification was carried out using a 250 × 4.6 mm Vydac 214TP C4 column at flow rate and temperature of 0.8 mL/min and 40°C, respectively, using 0.1% v/v TFA aqueous solution as mobile phase A and 0.1%

TFA ACN solution as mobile phase B. After equilibration in 5% B, elution was performed by applying a linear gradient from 5 to 95% B in 25 min. Detection was done at 280 nm. The fractions were collected and vacuum-dried.

2.2 Chemical and biochemical materials

PNGase F-based N-deglycosylation, 2-aminobenzamide (2-AB) labeling kits, 2-AB-bovine fetuin N-linked glycan library, and α,1,3/4-fucosidase from almond meal were purchased from Prozyme (San Leandro, CA, USA). A3S2 standards and α2,3-sialidase from Macrobdella Decora were from Merck-Millipore (Billerica, MA, USA). Acetic acid (CH3COOH), ammonium bicarbonate (NH4HCO3), sodium chloride (NaCl), urea, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), sodium cyanoborohydride (NaBH4), DTN, and iodoacetamide (IAA) were purchased from Sigma-Aldrich Chemical (St. Louis, MO, USA). Trihydroxyacetophenone (THAP) and protein/peptide calibration standards for MALDI-TOF MS were purchased from Bruker Daltonics (Bremen, Germany). Methanol (MeOH) and ACN were HPLC-grade reagent and purchased from Biosolve (Leenderweg, The Netherlands) and JT Baker (Philipsburg, NJ, USA), respectively. Sodium iodide (NaI) and [Glul]-fibrinopeptide B solutions were from Waters (Milford, MA, USA). Factor XIIa was purchased from Millipore Corp. (Billerica, MA, USA). All the aqueous solutions were prepared using ultrapure water (18.2 MΩ cm resistivity at 25°C, total organic carbon < 5 ppb).

2.3 SDS-PAGE analysis of hFXI

SDS-PAGE was performed under reducing condition using NuPAGENovex 4–12% Bis–Tris gel, Xcell SureLock Mini-Cell, and NuPAGE MES running buffer (Invitrogen, Carlsbad, CA, USA). One microgram of HPLC-collected protein material was loaded, the gel was CBB-stained and analyzed using an Image Scanner III (GE Healthcare, Piscataway, NJ, USA). The apparent molecular weights were calculated using Quantity One Software (Bio-Rad, Hercules, CA, USA).

2.4 N-Deglycosylation and purification of the released oligosaccharides

A 100 µg aliquot of purified hFXI was recovered in 47 µL of 50 mM Tris-HCl pH 8.0 and denatured in the presence of 0.2% w/v SDS and 100 mM β-mercaptoethanol at 95°C during 10 min. After cooling, 0.5 µL of a 15% w/v NP-40 solution was added to neutralize the excess of SDS and N-deglycosylation was carried out overnight at 37°C in presence of 8 mM PNGase F. The remaining glycosylamine forms were fully converted into reducing forms by incubating the sample with acetic acid 0.5% (final pH ~ 5.0) for 1 h at 37°C. The reducing N-glycans were purified.

© 2014 The Authors. PROTEOMICS published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim www.proteomics-journal.com
by SPE onto a 50 mg Hypersep Hypercarb column (ThermoFisher Scientific, San Jose, CA, USA) [14]. The SPE column was sequentially washed with 1 mL methanol and 2 × 1 mL of 0.1% v/v aqueous TFA. The oligosaccharides were dissolved in 200 μL of 0.1% v/v aqueous TFA, applied to the column and washed with 2 × 1 mL of 0.1% v/v aqueous TFA. Elution of glycans was performed by applying 2 × 500 μL of 25% v/v aqueous ACN containing 0.1% v/v TFA. The elute was vacuum-dried.

2.5 Fluorescent labeling of N-glycans by reductive amination with 2-AB

Purified glycans were reductively aminated with 2-AB by recovering dried glycans with 10 μL of 33% v/v acetic acid in DMSO containing 0.35 M 2-AB and 1 M NaNBH₃CN. The reaction was kept at 37°C for 16 h. The derivatives were finally purified on a 30 mg Oasis HLB cartridge (Waters, Milford, MA, USA). The cartridge was sequentially wetted and equilibrated with 2 × 1 mL of ACN. The derivatized oligosaccharides were tenfold diluted with ACN and loaded onto the cartridge. The bound material was washed by applying 2 × 1 mL of ACN and eluted by applying 2 × 400 μL of aqueous ACN 20% v/v. The eluate was vacuum-dried.

2.6 HILIC-UPLC/FD profiling of 2-AB-labeled hFXI N-glycans

Dried 2-AB-labeled glycans were recovered with 100 μL of ultrapure water and a 10 μL aliquot was threefold diluted with ACN before being injected on the HILIC column. HILIC-UPLC/FD profiling was carried out onto a 150 × 2.1 mm Acquity BEH amide column (Waters, Milford, MA, USA) equilibrated with ACN/ammonium formate 250 mM pH 4.4 (69:31; v/v) at a flow rate of 0.4 mL/min and at 47°C. After injection of 5 μL of glycan sample, the gradient was operated from 69 to 63% of ACN in 40 min. The structural assignment of the 2-AB-labeled N-glycan structures was achieved by matching their retention times with those of α2,3/α2,6-sialylated biantennary (A₂S₁) and bovine fetuin N-linked glycan standards. Confirmation of sialyl and fucosyl linkages was done using specific exoglycosidases, α2,3-sialidase from Macrobella Decora and α1,3/4-fucosidase from almond meal, respectively.

2.7 MALDI-TOF MS experiments

MALDI-TOF MS experiments were performed on an Autoflex II instrument (Bruker Daltonics) using the Flex Control 2.5 software. The mass range was comprised between m/z 1000 and 3500 for native glycan profiling and between m/z 10 000 and 200 000 for protein profiling. All mass spectra recorded represent accumulated mass spectra obtained from 3000 laser shots. External calibration of the mass spectra was carried out by using standard peptide/protein masses according to the manufacturer’s specifications. Negative-mode MALDI of native N-glycans and positive-mode MALDI of native/N-deglycosylated hFXI were performed by spotting 0.5 μL of a matrix solution (10 mg/mL THAP in a 50% v/v aqueous ACN) onto a polished stainless steel MALDI sample plate and mixed with 0.5 μL of samples (dissolved in a 0.1% v/v aqueous formic acid) and allowed to vacuum-dry.

2.8 LC-MS analysis of FXIIa-generated hFXI light and heavy chains

One hundred micrograms of hFXI were reconstituted in 60 μL of 50mM HEPES and 10 μL of 1M NaCl and incubated in the presence of 5 μL of FXIIa (1.9 mg/mL) for 16 h at room temperature [7]. The product obtained was vacuum-dried then placed in contact with 25 μL of denaturing buffer (8 M urea in 0.4M NH₂HCO₃) for 15 min at room temperature. After denaturation, the protein was reduced by 5 μL of 100 mM DTT for 30 min at 55°C. Around 30 μg of protein material was injected onto a C4 RP column (150 × 2.1 mm, 5 μm) temperature-controlled to 40°C. The mobile phase A was 0.1% v/v formic acid in water, while mobile phase B consisted of 0.1% v/v formic acid in ACN. Elution was done by applying a linear gradient of mobile phase B at a flow rate of 350 μL/min from 10 to 55% in 28 min. The eluate was analyzed online using a SynaptG2S HDMS mass spectrometer (Waters, Manchester, UK) operated in the positive mode at a resolution of 20 000 full width at half maximum. ESI source settings were as follows: desolvation gas flow rate and temperature, 800 L/h and 150°C respectively; source temperature, 150°C; capillary voltage, 3 kV; cone voltage 33 V. Calibration of the TOF analyzer was done using NaI cluster ions over the mass range 300–3000 and [Glu1]-fibrinopeptide B was used as a lock mass. All MS measurements were done over the mass range m/z 300–3000 with a scan time of 1 s. Data collection and processing were done by MassLynx 4.1 software (Waters). Deconvolution of mass spectra was carried out using the MaxEnt 1 algorithm.

2.9 LC-MS/MS analyses of hFXI-derived peptides/glycopeptides

Two hundred micrograms of hFXI were reduced, alkylated, and digested with trypsin or Lys-C, separately, at an enzyme-to-substrate ratio of 1/50 w/w at 37°C for 16 h. Digests were acidified to pH 2 using TFA; samples were stored at −20°C prior to analysis. A part of the tryptic/Lys-C-digested peptides was purified on a Hypersep C18 SPE column (ThermoFisher Scientific, San Jose, CA, USA). The purified peptides were vacuum-dried and recovered in 100 μL of 50 mM Tris-HCl buffer pH 8.0 and deglycosylated using PNGase F (final activity 0.25 mU/μL) overnight at 37°C. Around 50 μg of N-deglycosylated peptides were injected onto a BEH300 C18 RP column (150 × 2.1 mm, 1.7 μm), temperature-controlled
to 60°C. Mobile phase A was 0.1% formic acid in water, while mobile phase B consisted of 0.1% formic acid in ACN. Elution was done by applying a linear gradient of mobile phase B at a flow rate of 300 μL/min from 5 to 50% in 104 min. The eluate was analyzed online using a SynaptG2S HDMS mass spectrometer (Waters) operated in the positive-ion mode at a resolution of 20 000 full width at half maximum. ESI source settings, calibration of the TOF analyzer, and lock-mass correction procedures were the same as described in Section 2.8. ESI spectra deconvolution was done using the MaxEnt 3.0 algorithm. Data-dependent MS/MS acquisitions (DDA) included one MS survey scan (m/z 300–3000, 1 s) and one MS/MS scan (m/z 50–3000, 1 s). This method provides ion chromatogram peaks with a high-dot definition in both MS and MS/MS chromatographic traces, necessary first to obtain representative combined mass spectra of peptide glycoforms across their RT range and secondly to accurately integrate extracted ion chromatograms (XICs) of ions of interest for quantitation of site occupancies. Interpretation of MS/MS data was done using the PEAKS Studio 5.0 software from Bioinformatics solutions.

3 Results and discussion

3.1 Purity assessment of hFXI

Purified hFXI, analyzed by SDS-PAGE under reducing conditions (see Supporting Information Fig. 1), appeared as a single major band at 68 kDa, consistent with the mass of one monomer. Purity of hFXI was estimated at about 90%, which was sufficient to characterize the N-glycosylation of the protein.

3.2 MALDI-TOF MS and HILIC/FD profiling of total hFXIN-glycan chains

The PNGase F-released N-glycans were analyzed by MALDI-TOF MS in linear negative-ion mode. The linear mode has the advantage to dramatically limit post-source fragmentation of labile groups such as sialic acids or sulfation and to provide a semiquantitative evaluation of the N-glycan population. The mass spectrum obtained from the total pool of hFXIN-glycans is presented in Fig. 1A. Glycan structures were identified by comparing experimental masses with average theoretical ones. This profile shows that the major N-glycans are nonfucosylated sialylated biantennary complex-type structures A₃S₂ and A₃S₂ at m/z 1930.8 and 2222.4, respectively. Sialylated triantennary structures, A₃S₁ and A₃S₁, at m/z 2587.3 and 2879.2 are also retrieved in lower abundance. Trace amounts of their fucosylated counterparts are also detected (see m/z 2077.1, 2368.2, 2733.2, and 3025.3).

N-glycan chains, labeled at their reducing end with a fluorescent tag (2-AB), were then analyzed and quantified using HILIC-UPLC/FD. Identification of each peak was done by comparing their RTs with those from N-glycan standards and using specific exoglycosidases as described in Section 2.6. The HILIC profile, presented on Fig. 1B, shows that the major N-glycan structure, eluted at 16.3 min, is a bisialylated biantennary glycan, A₃S₂, that accounts for 66.1% of the total N-glycan population. The second major structure, eluted at 11.6 min, is a monosialylated biantennary glycan (A₃S₁) accounting for 19.8%. Besides these biantennary glycans, minor triantennary structures contributing to 5.4% of the total N-glycan population were observed. The fucosylation level is less than 3% for the biantennary structures and about 40% for the triantennary forms. Moreover, HILIC provides selectivity toward many branching/linkage N-glycan isomers, that is it discriminates core-fucosylated from distally fucosylated structures and α2,3- from α2,6-sialylated glycans by their RTs [15]. The major peak eluted at 16.3 min corresponds to the bis-α2,6-sialylated biantennary N-glycan (A₃S₂), the mono-and bis-α2,3-sialylated N-glycan standards, which eluted earlier, were not detected for hFXI. The peak at 18.4 min corresponds to a core-fucosylated A₃S₁F structure. The peak at 26.5 min corresponds to a A₃S₁ glycan with two α2,6- and one α2,3-linked sialic acids, whereas the fully α2,6-sialylated triantennary standard, which elutes later, was not detected for hFXI. Two isomeric distally monofucosylated trisialylated triantennary structures, evidenced at 29.5 and 30.0 min, accounted for about 1.5% of the glycan population. No distally fucosylated A₃S₁F isomer was evidenced. This result is in accordance with the preference of α1,3-fucosyltransferases to fucosylate exclusively N-acetylgalcosamine from nonsialylated or α2,3-sialylated N-acetyllactosamine antennae and with the impossibility for human α2,3/6-sialyltransferases to act on preassembled Lewis-derived acceptors [16]. Finally, hFXIN-glycans display a high sialylation level (>85%) as expected for the human blood glycoproteins [17], suggesting that the purification process preserved the integrity of the native protein N-glycosylation.

3.3 MALDI-TOF MS of intact versus N-deglycosylated hFXI

After having identified the major hFXIN-glycan structures, we investigated the global N-glycosylation occupancy over the entire glycoprotein. For this purpose, we measured the average mass of the protein by MALDI-TOF MS, before and after deglycosylation with PNGase F (see Fig. 2). As shown in Fig. 2A, two main signals at m/z 152.4 and 76.6 kDa are detected, corresponding respectively to singly and doubly charged ions. The observed mass of 152.4 kDa is consistent with the mass of a cysteinylated (+103 Da) hFXI dimer, glycosylated eight times mainly by a bisialylated biantennary N-glycan A₃S₂ (glycoprotein theoretical mass: 153.8 kDa). The observed mass using the doubly charged species (2 × 76.6 = 153.2 kDa) was close to the theoretical one. This result suggests that only four of the five consensus sites per hFXI monomer are actually occupied. Figure 2B shows the
MALDI-TOF mass spectrum of N-deglycosylated hFXI with two peaks visible at m/z 68.1 and 136.1 kDa, corresponding to doubly and singly charged ions, respectively. The average experimental mass of 136.1 kDa is consistent with a nonglycosylated, cysteinylated hFXI dimer (theoretical average mass: 136.2 kDa). Considering these two experiments, a mass-loss of 17.0 kDa is measured after PNGase F treatment. Assuming an average mass of 2.2 kDa corresponding to the major hFXI N-glycan (A₂S₂), the mass loss of 17.0 kDa suggests the occupation of eight N-glycosylation sites. In addition, this result also indicates that no other PTM, giving a significant measurable mass shift such as O-glycosylation, is present.

3.4 Glycosylation analysis of heavy and light chains by LC-MS

To get accurate information about N-glycosylation microheterogeneity and site occupancy, hFXI was activated with FXIIa to release, after reduction of the disulfide bridges, its heavy (E₁-R₃₆₉) and light (I₁₇₀-V₆₀₇) chains, which were further profiled by LC-MS. The deconvoluted ESI-TOF mass spectra of reduced hFXI heavy (A) and light (B) chains are shown in Fig. 3. The HC (see Fig. 3A) gives a major signal at 45 619 Da corresponding to the [E₁-R₃₆₉] sequence modified by A₂S₂ structures on the two glycosites (theoretical mass: 45 619 Da). This glycoform accounts for about 60% of the total HC population, considering a consistent ESI response independently of the glycan type. Its monofucosylated counterpart was detected at 45 767 Da (10%), with less mass accuracy (+2 Da). The mass of 45 327 Da corresponds to the HC bearing both A₂S₂ and A₃S₃ structures (~18%). A minor glycoform with A₂S₂/A₁S₁ structures was also detected at 46 276 Da (~10%), as well as its monofucosylated and undersialylated counterparts at 46 422 Da (~3%) and 45 984 Da (A₂S₂/A₁S₁ and/or A₂S₂/A₁S₁; ~3%), respectively. The mass at 45 124 Da corresponds to a C-terminal [E₁-K₃₆₅] truncated form of the major HC glycoform (A₂S₂/A₂S₂), probably occurring during sample digestion with FXIIa. Figure 3B shows the mass spectrum of hFXI light chain with
two series of signals corresponding to the monoglycosylated light chain (see masses at 28 749 and 29 044 Da) and the doubly glycosylated LC (see masses from 30 666 to 31 905 Da). The masses of 28 749 and 29 044 Da are consistent with the [I_{370}-V_{607}] light chain sequence modified at one site by A_{2S_1} and A_{2S_2} structures, respectively. Considering their intensities, these two monoglycosylated light chain glycoforms represented about 10% of the total light chain population suggesting that one and/or the two light chain consensus glycosite(s) (N_{432} and/or N_{473}) is/are partially occupied at a minimum of 90%. Masses detected at 30 958 and 31 249 Da are consistent with the [I_{370}-V_{607}] light chain sequence modified at the two sites by A_{2S_1}/A_{2S_2} and A_{2S_2}/A_{2S_2} structures, respectively. These forms account for about 34 and 45% of the total light chain forms, respectively. Finally, the minor signal measured at 30 666 Da corresponds to glycan composition(s) A_{2S_1}/A_{2S_1} and/or A_{2S_2}/A_{2S_2}, estimated at ~5% whereas those at 31 613 and 31 905 Da correspond to compositions A_{2S_2}/A_{2S_1} (~3%) and A_{2S_2}/A_{2S_2} (~3%).

These results reveal that among the three consensus heavy chain glycosites, at positions N_{72}, N_{108}, and N_{335}, only two are fully glycosylated while for the light chain, the two sites are mostly occupied. Furthermore, the relative composition of glycan structures found on the heavy chain versus the light chain differs significantly, the light chain being less efficiently sialylated. The partial N-glycan occupancy of one and/or two light chain glycosites could be correlated to their position in the C-terminus of the glycoprotein, known to be statistically less efficiently glycosylated by the oligosaccharyltransferase (OST) [18]. Finally, the high mass accuracy obtained in this experiment confirms that hFXI did not bear any other PTMs.

### 3.5 Site-specific N-glycosylation analysis by LC-MS/MS

hFXI glycopeptides, obtained after digestion with either trypsin or Lys C, were analyzed by LC-MS/MS to determine the site-specific occupancy and distribution of the N-glycan structures. The glycopeptides were retrieved by extracting characteristic oxonium ions HexNAC+ (m/z 204, HexNAC+ is N-acetylhexosamine), HexHexNAC+ (m/z 366), and NeuAcHexHexNAC+ (m/z 657) from the TIC of data-dependent MS/MS spectra. Since glycopeptides can be eluted with different RT depending on their relative hydrophilicities, MS spectra of each glycopeptide were iteratively combined to select RT windows giving reproducible profiles with constant peak-to-peak relative intensities. Figure 4 presents six combined deconvoluted ESI MS spectra of trypsin-/Lys C-generated peptides of interest bearing each one consensus or nonconsensus glycosite (Supporting Information Table 1 lists experimental/theoretical masses of glycoforms corresponding to these peptides). The tryptic peptide [V_{71}N_{72}RTAAISGYSFK] of the HC was modified by seven different glycans (see Fig. 4A). The major glycoform (~60%) detected at 3618.4639 Da corresponds to A_{2S_2} structure (~16 ppm). Trace amounts of its fucosylated counterpart (A_{2S_2}F) were also detected at mass 3764.3219 Da (~1%). Glycopeptides modified by trisialylated triantennary structures were also found at masses 4274.7080 Da (A_{3S_2}, ~20%) and 4420.7759 Da (A_{3S_3}F, ~15%). The latter A_{3S_3}F glycoform may correspond to a triantennary structure distally fucosylated, as seen from the HILIC–UPLC analysis of the released N-glycans (see Fig. 1B), indicating a preferential fucosylation...
of triantennary structures in distal position. Other minor glycoforms (less than 5%) are observed at masses 3983.8342 and 4129.9025 Da and correspond to the hexasialylated counterparts of the structures described in Section 3.4. This glycosite is positioned on the A1 Apple domain of hFXI structure, close to the E$_{66}$ residue, which is known to be important for thrombin binding [19]. It is also interesting to note that N$_{72}$ is significantly more sialylated than the other glycosites.

Figure 4B shows the mass spectrum of the glycopeptide $[G_{104}INYN_{108}SSVAK_{111}]$, mainly glycosylated by an A$_2$S$_2$ structure, giving a mass of 3257.3040 Da (~90%). The remaining 10% of the species were identified as minor biantennary (A$_2$S$_1$ and A$_2$S$_2$F) and triantennary (A$_3$S$_2$ and A$_3$S$_3$) structures. N$_{108}$ is positioned on the A2 domain of the HC, known to bind hFIX and high molecular weight kininogen [20]. The glycosylated peptide $[V_{426}YSGILN_{432}QSEIK_{437}]$, presented in Fig. 4C displays a very similar profile with a major A$_2$S$_2$ structure representing about 95% of the glycoforms and almost the same remaining biantennary and triantennary structures.

Figure 4D displays the mass spectrum of the glycopeptide $[L_{468}ETTVN_{473}YTDSQRPICLPSK_{486}]$, from the light chain. N$_{473}$ is the most heterogeneous modified glycosite with both A$_2$S$_1$ and A$_2$S$_2$ structures as the two major forms, detected respectively at 4135.7910 (~35%) and 4426.8931 Da (~50%). A wide range of minor structures (~15%) were also identified, consisting of other biantennary and triantennary structures.
Figure 4. Deconvoluted ESI-TOF MS profiles of trypsin-/Lys-C-generated peptides/glycopeptides of interest showing N-glycan heterogeneity at glycosites N\textsubscript{72} (A), N\textsubscript{108} (B), N\textsubscript{432} (C), N\textsubscript{473} (D), N\textsubscript{335} (E), N\textsubscript{145} (F). Mass spectra were processed by MassLynx 4.1 using MaxEnt3 algorithm.

with variable sialylation and fucosylation. $N_{473}$ is the less sialylated glycosite (sialylation level ~70%) and this result confirms the lower sialylation of the light chain comparing to the HC. Interestingly, $N_{132}$ and $N_{473}$ are two conserved glycosites retrieved on the human plasma kallikrein catalytic domain that shares 58% sequence homology with hFXI. In spite of the conservation of these two glycosylation sites, the idea that these PTMs have a biological role is not certain. Indeed, Tang et al. [21] described abaculovirus/Sf9 system-produced unglycosylated kallikrein that retained in vitro activity even though all three glycosite-related asparagines were mutated to glutamates.

Figure 4E shows the mass spectrum of the peptide $[L_{332}SSN_{335}GSPTK_{340}]$, a unique mass of 890.4584 Da was detected, corresponding to the nonmodified peptide (−0.7 ppm). The absence of glycosylation at this site was due to the presence of a proline at the Y-position of the consensus sequence NXS/TY, which is known to hinder N-glycosylation [22, 23]. Interestingly, human complement factor H (hCFH) shares the same tetrapeptide [N\textsubscript{199}GSP\textsubscript{202}] that was demonstrated to be unoccupied [24]. We assume a protective effect of proline at the Y-position of this consensus site by preserving structural and functional integrities of hFXI. Indeed, this glycosylation site position would not be compatible with the maintenance of structural and functional integrities of the glycoprotein, as it is located at proximity of the dimer interface domain. The presence of a glycan at this position may hinder both the conformational integrity of the A4 domain and the protein dimerization, required for normal hFXI biological activity [25].

Further MS analysis revealed the presence of two unexpected minor N-glycopeptides with deconvoluted
masses of 2674.1055 Da and 2965.2046 Da, presenting compositions consistent with NI/LCLLK amino acid sequences, glycosylated by monosialylated (A3S2, –3.7 ppm) and bisialylated (A3S2, –2.1 ppm) biantennary N-glycans, respectively (data not shown). This peptide is not unique in hFXI sequence that contains two similar peptides [N145ICLLK150] and [N235ICLLK240]. In both cases, this new glycosite is noncanonical with a composite NXC sequence, with a Cys residue instead of Ser/Thr commonly found at the third position of canonical N-glycosylation sites of the NXS/T format. To determine which of the two sites is N-glycosylated, we carried out an enzymatic digestion of hFXI using Lys-C followed by an LC-MS/MS analysis. Figure 4F shows the mass spectrum of the glycosylated peptide [S114AQECQERTDDVHCHFYTATQFPSLEHNR145ICLLK150]. This peptide displays two masses at 6567.8218 and 6858.8975 Da corresponding to the peptide sequence glycosylated by monosialylated (A3S2, –3.3 ppm) and bisialylated (A3S2, –0.3 ppm) biantennary N-glycans, respectively. Despite the presence of a cysteine at the N+2-position of the asparagine, the N-glycan profile remains comparable to those of N108 and N142. Glycosylation of such a site has already been reported on various mammalian glycoproteins: human α1T glycoprotein [26], human von Willebrand factor (vWF) [27, 28], human CD69 [29], murine/human fetal antigen 1 [30, 31], human alpha-lactalbumin [32], bovine/human plasma-derived and recombinant human protein C [33, 34], recombinant human epidermal growth factor receptor [35] and plasma-derived and recombinant human tissue plasminogen activator (t-PA) [36]. Furthermore, a recent glycoproteomic study on mouse tissues and blood plasma reported that about 1.3% of glycosylated sequences are of the NXC type, while canonical glycosites NXS/T accounted for more than 96%, the rest being rare noncanonical glycosites, e.g. for example NG or NXV [37]. The 3D structure analysis of hFXI (PDB ID 2F83, [38]) suggests that the presence of N-glycans on this N145 may hinder the A2 domain unfolding: particularly by keeping the α-helix [A315–Thr121] away from the antiparallel β-sheet formed by the seven β-strands characterizing the Apple domains, with the putative consequence of disturbing interaction with hFIX.

### 3.6 Site-specific N-glycan occupancy analysis by LC-MS/MS

Extraction of masses corresponding to the naturally nonglycosylated variants of the five glycosylated peptides was carried out to assess their site occupancy. As seen in the previous section, no glycoform of the tryptic peptide [L322SSN326GSPTK340] has been detected, indicating that N325 is unoccupied. This was consistent with the LC-MS analysis of hFXI HC that showed that only two of the three glycosites were occupied. The absence of MS signals corresponding to naked tryptic/Lys-C peptides bearing N32 and N108 suggested that these sites are fully occupied. Positive signals were extracted for peptides bearing N145 (HC), N332 (LC), and N473 (LC), suggesting partial site occupancies. LC-MS data of the light chain indicated variable site occupancies at N145 and/or N473 with doubly glycosylated forms representing about 90% of the light chain variants, the remaining 10% corresponding to monoglycosylated variants. In order to estimate the percentage of occupancy at each glycosite, the glycopeptides were deglycosylated by PNGase F. This cleavage leads to deamidation of asparagine into aspartic acid resulting in a difference of RTs and a +0.984 Da mass shift between the naked and glycosylated peptides. Estimation of the site occupancy was done by dividing the specific area of XIC peak corresponding to the deamidated peptide, by the sum of XIC areas of deamidated and native peptides. HC: heavy chain; LC: light chain.

### Table 1. Site occupancies of the six consensus and nonconsensus hFXI glycosites

| Position | Peptide sequence | Occupancy [%] |
|----------|------------------|---------------|
| N72 (HC) | 1VNRTAAISGYSFK63 | 100           |
| N108 (HC) | 1GNYNSVAV153 | 100           |
| N145 (HC) | 14SADECOERCTDDVHCHF | 5 |
| N335 (HC) | 333LSSNGSPTK340 | 0 |
| N432 (LC) | 432VYSGILNQSEK437 | 92 |
| N473 (LC) | 468LETQNYTDQSRPCLPK486 | 95 |

Reduced and alkylated hFXI was digested using Lys-C or trypsin. Peptide mixture was treated with PNGase F and the digest was analyzed by LC-MS/MS. For each glycosite, percentage occupancy was estimated by dividing the XIC area of the corresponding PNGase F-deamidated peptide by the sum of XIC areas of deamidated and native peptides. HC: heavy chain; LC: light chain.
an uncommon NXC glycosylated site were identified and all
characterized as being partially occupied. However, only few
studies provided an estimation of their percentages of oc-
cupancy. Occupancies at N71 of human alpha lactalbumin
and at N142 of murine fetal antigen 1 were measured, respec-
tively, at 1 and 5% [31, 32]. Jiang et al. [36] reported 35–60% site occupancies at N146 of various recombinant human t-PA.
Gil et al. reported 74–78% site occupancies at the N129 of
both plasma-derived and recombinant human protein C pro-
duced in transgenic pig milk [34]. With only 5% occupancy,
N145 of hFXI will be a new example of this kind of atypical gly-
cosylation site. Previous studies showed that synthetic pep-
tides bearing such a glycosite were low-affinity substrates for
OST as evidenced by the poor glycosylation efficiency [43,44].
Moreover, N-glycosylation by OST and disulfide bond forma-
tion are two cotranslational processes occurring at the same
time on neosynthesized glycoproteins in the ER that can com-
pete, leading to reduced availability of the cysteine residue for
binding to OST.

4 Concluding remarks

In an attempt to improve the current knowledge of hFXI
structure, in this article, we describe the first full characteriza-
tion of hFXI N-glycosylation including its microheterogeneity
and site occupancy using advanced MS techniques. Among
the five canonical glycosities, only four were shown to be occu-
pied (HC, N72 and N108; LC, N432 and N473). C-terminal light
chain glycosites N412 and N473 display high but not full occu-
pancies (92 and 95%) whereas HC glycosites N72 and N108
are fully glycosylated. All occupied glycosites are positioned
at key areas of the glycoprotein that binds ligands involved
in the coagulation cascade. N-glycans attached close to these
areas may play direct or indirect roles in the modulation of
ligands recruitment. N72 was found to be the most acidic site
whereas N473 was the less sialylated. N108 (HC) was shown to
be unoccupied due to the presence of a proline at Y-position.
Furthermore, hFXI can now be added to the list of glyco-
proteins bearing a noncanonical glycosite of the NXC type,
with its N145 occupied at only 5%. These structural data could
serve as a reference, useful to understand the potential roles
of N-glycosylation on hFXI physicochemical and functional
properties.

The authors would like to thank Marie Romanin for her sup-
port.

Potential conflict of interest: All authors are employees of LFB
Biotechnologies subsidiary of LFB SA group, commercializing the
FXI concentrate (Hemoleven®) used for this study.

5 References

[1] Bolton-Maggs, P. H., Factor XI deficiency and its manage-
ment. Haemophilia 2000, 6 Suppl 1, 100–109.
[2] Baglia, F. A., Jameson, B. A., Walsh, P. N., Identification and
characterization of a binding site for platelets in the Apple
3 domain of coagulation factor XI. J. Biol. Chem. 1995, 270,
6734–6740.
[3] Ho, D. H., Badellino, K., Baglia, F. A., Walsh, P. N., A binding
site for heparin in the Apple 3 domain of factor XI. J. Biol.
Chem. 1998, 273, 16382–16390.
[4] Baglia, F. A., Badellino, K. O., Ho, D. H., Dasari, V. R. et al.,
A binding site for the kringle II domain of prothrombin in
the Apple 1 domain of factor XI. J. Biol. Chem. 2000, 275,
31954–31962.
[5] Baglia, F. A., Walsh, P. N., A binding site for thrombin in the
Apple 1 domain of factor XI. J. Biol. Chem. 1996, 271, 3652–
3658.
[6] Baglia, F. A., Jameson, B. A., Walsh, P. N., Fine mapping of
the high molecular weight kinogen binding site on blood
coagulation factor XI through the use of rationally designed
synthetic analogs. J. Biol. Chem. 1992, 267, 4247–4252.
[7] Di Scipio, R. G., Kurachi, K., Davie, E. W., Activation of human
factor IX (Christmas factor). J. Clin. Invest. 1978, 61, 1528–
1538.
[8] Baglia, F. A., Jameson, B. A., Walsh, P. N., Identification and
characterization of a binding site for factor XIIa in the Apple
4 domain of coagulation factor XI. J. Biol. Chem. 1993, 268,
3838–3844.
[9] Baglia, F. A., Gailani, D., Lopez, J. A., Walsh, P. N., Identifica-
tion of a binding site for glycoprotein Ibalpha in the Apple
3 domain of factor XI. J. Biol. Chem. 2004, 279, 45470–45476.
[10] Baglia, F. A., Shrimpton, C. N., Emsley, J., Kitagawa, K. et al.,
Factor XI interacts with the leucine-rich repeats of glycopro-
tein Ibalpha on the activated platelet. J. Biol. Chem. 2004,
279, 49323–49329.
[11] Smith, S. B., Verhamme, I. M., Sun, M. F., Bock, P. E. et al.,
Characterization of novel forms of coagulation factor XIa:
independence of factor XIa subunits in factor IX activation.
J. Biol. Chem. 2008, 283, 6696–6705.
[12] Wu, W., Sinha, D., Shikov, S., Yip, C. K. et al., Factor XI ho-
modimer structure is essential for normal proteolytic activa-
tion by factor XIIa, thrombin, and factor XIa. J. Biol. Chem.
2008, 283, 18655–18664.
[13] Fujikawa, K., Chung, D. W., Hendrickson, L. E., Davie, E. W.,
Amino acid sequence of human factor XI, a blood coagula-
tion factor with four tandem repeats that are highly ho-
mologous with plasma prekallikrein. Biochemistry 1985, 24,
2417–2424.
[14] Packer, N. H., Lawson, M. A., Jardine, D. R., Redmond, J. W.,
A general approach to desalting oligosaccharides released
from glycoproteins. Glycoconjug. J. 1998, 15, 737–747.
[15] Guile, G. R., Rudd, P. M., Wing, D. R., Prime, S. B., Dwek,
R. A., A rapid high-resolution high-performance liquid chro-
matographic method for separating glycan mixtures and an-
alyzing oligosaccharide profiles. Anal. Biochem. 1996, 240,
210–226.
[16] Grabenhorst, E., Nimtz, M., Costa, J., Conradt, H. S., In
vivo specificity of human alpha1,3/4-fucosyltransferases II-
VII in the biosynthesis of LewisX and SialylLewisX motifs on
complex-type N-glycans. Coexpression studies from bhek-21
cells together with human beta-trace protein. *J. Biol. Chem.* 1998, 273, 30985–30994.

[17] Morelle, W., Flahaut, C., Michalski, J. C., Louvet, A. et al., Mass spectrometric approach for screening modifications of total serum N-glycome in human diseases: application to cirrhosis. *Glycobiology* 2006, 16, 281–293.

[18] Bano-Polo, M., Baldin, F., Tamborero, S., Marti-Renom, M. A. et al., N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thrsequon. *Protein. Sci.* 2011, 20, 179–186.

[19] Yun, T. H., Baglia, F. A., Myles, T., Navaneetham, D. et al., Thrombin activation of factor XI on activated platelets requires the interaction of factor XI and platelet glycoprotein Ib alpha with thrombin anion-binding exosites I and II, respectively. *J. Biol. Chem.* 2003, 278, 48112–48119.

[20] Emsley, J., McEwan, P. A., Gailani, D., Structure and function of factor XI. *Blood* 2010, 115, 2569–2577.

[21] Tang, J., Yu, C. L., Williams, S. R., Springman, E. et al., Expression, crystallization, and three-dimensional structure of the catalytic domain of human plasma kallikrein. *J. Biol. Chem.* 2005, 280, 41077–41089.

[22] Gavel, Y., von Heijne, G., Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.* 1990, 3, 433–442.

[23] Mellquist, J. L., Kasturi, L., Spitalnik, S. L., Shakin-Eshleman, S. H., The amino acid following an asn-asn-ser/thrseron is an important determinant of N-linked core glycosylation efficiency. *Biochemistry* 1998, 37, 6833–6837.

[24] Fenaille, F., LeMignon, M., Grosell, C., Ramon, C. et al., Site-specific N-glycan characterization of human complement factor H. *Glycobiology* 2007, 17, 932–944.

[25] Gailani, D., Ho, D., Sun, M. F., Cheng, Q. et al., Model for a factor IX activation complex on blood platelets: dimeric conformation of factor Xla is essential. *Blood* 2001, 97, 3117–3122.

[26] Araki, T., Haupit, H., Hermentin, P., Schwick, H. G. et al., Preparation and partial structural characterization of alpha1T-glycoprotein from normal human plasma. *Arch. Biochem. Biophys.* 1998, 361, 250–256.

[27] Titani, K., Takio, K., Handa, M., Ruggeri, Z. M., Amino acid sequence of the von Willebrand factor-binding domain of platelet membrane glycoprotein Ib. *Proc. Natl. Acad. Sci. USA* 1987, 84, 5610–5614.

[28] Canis, K., McKinnon, T. A., Nowak, A., Haslam, S. M. et al., Mapping the N-glycome of human von Willebrand factor. *Biochem. J.* 2012, 447, 217–228.

[29] Vance, B. A., Wu, W., Ribaudo, R. K., Segal, D. M. et al., Multiple dimeric forms of human CD69 result from differential addition of N-glycans to typical (Asn-X-Ser/Thr) and atypical (Asn-X-cys) glycosylation motifs. *J. Biol. Chem.* 1997, 272, 23117–23122.

[30] Jensen, C. H., Krogh, T. N., Hojrup, P., Clausen, P. P. et al., Protein structure of fetal antigen 1 (FA1). A novel circulating human epidermal-growth-factor-like protein expressed in neuroendocrine tumors and its relation to the gene products of dlk and pg2. *Eur. J. Biochem.* 1994, 225, 83–92.

[31] Krogh, T. N., Bachmann, E., Teisner, B., Skjodt, K. et al., Glycosylation analysis and protein structure determination of murine fetal antigen 1 (mFA1)–the circulating gene product of the delta-like protein (dlk), preadipocyte factor 1 (Pref-1) and stromal-cell-derived protein 1 (SCP-1) cDNAs. *Eur. J. Biochem.* 1997, 244, 334–342.

[32] Giuffrida, M. G., Cavaletto, M., Giunta, C., Neuteboom, B. et al., The unusual amino acid triplet Asn-Ile-Cys is a glycosylation consensus site in human alpha-lactalbumin. *J. Protein Chem.* 1997, 16, 747–753.

[33] Stenflo, J., Fernlund, P., Amino acid sequence of the heavy chain of bovine protein C. *J. Biol. Chem.* 1982, 257, 12180–12190.

[34] Gil, G. C., Velander, W. H., VanCott, K. E., N-glycosylation microheterogeneity and site occupancy of an Asn-X-Cyssequence in plasma-derived and recombinant protein C. *Proteomics* 2009, 9, 2565–2567.

[35] Sato, C., Kim, J. H., Abe, Y., Saito, K. et al., Characterization of the N-oligosaccharides attached to the atypical Asn-X-Cys sequence of recombinant human epidermal growth factor receptor. *J. Biochem.* 2000, 127, 65–72.

[36] Jiang, H., Wu, S. L., Karger, B. L., Hancock, W. S., Characterization of the glycosylation occupancy and the active site in the follow-on protein therapeutic: TNK-tissue plasminogen activator. *Anal. Chem.* 2010, 82, 6154–6162.

[37] Zielinska, D. F., Gnad, F., Wisniewski, J. R., Mann, M., Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* 2010, 141, 987–997.

[38] Papagrigroriou, E., McEwan, P. A., Walah, P. N., Emsley, J., Crystal structure of the factor XI zymogen reveals a pathway for transactivation. *Nat. Struct. Mol. Biol.* 2006, 13, 557–558.

[39] Kasturi, L., Eshleman, J. R., Wunner, W. H., Shakin-Eshleman, S. H., The hydroxy amino acid in an Asn-X-Ser/Thrsequon can influence N-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. *J. Biol. Chem.* 1995, 270, 14756–14761.

[40] Shakin-Eshleman, S. H., Spitalnik, S. L., Kasturi, L., The amino acid at the X position of an N-linked glycoprotein is an important determinant of N-linked glycosylation efficiency. *J. Biol. Chem.* 1996, 271, 6363–6366.

[41] Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H. P. et al., Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms. *Proteomics* 2006, 6, 3369–3380.

[42] Bano-Polo, M., Baldin, F., Tamborero, S., Marti-Renom, M. A. et al., N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thrsequon. *Protein. Sci.* 2011, 20, 179–186.

[43] Bause, E., Breuer, W., Peters, S., Investigation of the active site of oligosaccharyltransferase from pig liver using synthetic tripeptides as tools. *Biochem. J.* 1995, 312, 979–985.

[44] Gerber, S., Lizak, C., Michaud, G., Bucher, M. et al., Mechanism of bacterial oligosaccharyltransferase: in vitro quantification of sequon binding and catalysis. *J. Biol. Chem.* 2013, 288, 8849–8861.

© 2014 The Authors. PROTEOMICS published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim www.proteomics-journal.com