Journal of proteome research

Article

pubs.acs.org/jpr

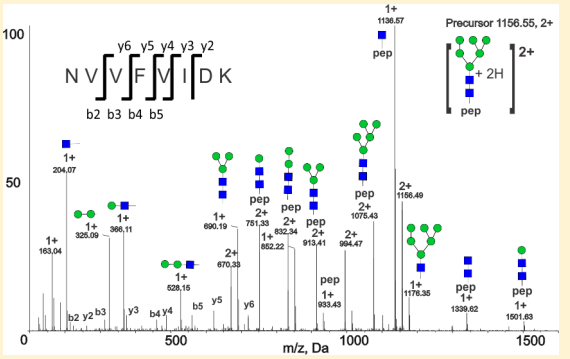
Site-Specific Glycan Microheterogeneity of Inter-Alpha-Trypsin Inhibitor Heavy Chain H4

Kevin Brown Chandler,[†,‡] Zuzana Brnakova,[‡] Miloslav Sanda,[‡] Shuo Wang,[†,‡] Stephanie H. Stalnaker,[§] Robert Bridger,[§] Peng Zhao,[§] Lance Wells,[§] Nathan J. Edwards,[†] and Radoslav Goldman*[,†,‡]

[†]Department of Biochemistry and Molecular & Cellular Biology, Georgetown University, Washington, D.C. 20057, United States
[‡]Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, D.C. 20057, United States
[§]Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602-4712, United States

Supporting Information

ABSTRACT: Inter-alpha-trypsin inhibitor heavy chain H4 (ITIH4) is a 120 kDa acute-phase glycoprotein produced primarily in the liver, secreted into the blood, and identified in serum. ITIH4 is involved in liver development and stabilization of the extracellular matrix (ECM), and its expression is altered in liver disease. In this study, we aimed to characterize glycosylation of recombinant and serum-derived ITIH4 using analytical mass spectrometry. Recombinant ITIH4 was analyzed to optimize glycopeptide analyses, followed by serum-derived ITIH4. First, we confirmed that the four ITIH4 N-X-S/T sequons (N81, N207, N517, and N577) were glycosylated by treating ITIH4 tryptic/GluC glycopeptides with PNGaseF in the presence of ^{18}O water. Next, we performed glycosidase-assisted LC−MS/MS analysis of ITIH4 trypsin-GluC glycopeptides enriched via hydrophilic interaction liquid chromatography to characterize ITIH4 N-glycoforms. While microheterogeneity of N-glycoforms differed between ITIH4 protein expressed in HEK293 cells and protein isolated from serum, occupancy of N-glycosylation sites did not differ. A fifth N-glycosylation site was discovered at N274 with the rare nonconsensus NVV motif. Site N274 contained high-mannose N-linked glycans in both serum and recombinant ITIH4. We also identified isoform-specific ITIH4 O-glycoforms and documented that utilization of O-glycosylation sites on ITIH4 differed between the cell line and serum.



KEYWORDS: *inter-alpha-trypsin inhibitor heavy chain H4, ITIH4, glycopeptide MS/MS, isoform-specific glycosylation, site occupancy, nonconsensus N-glycosylation, O-glycosylation, N-glycosylation*

INTRODUCTION

Inter-alpha-trypsin inhibitor heavy chain H4 (ITIH4) is a 120 kDa serum glycoprotein secreted primarily by the liver.[1] ITIH4 is one of five human inter-alpha-trypsin inhibitor heavy chain (ITIH) proteins, a family involved in stabilization of the extracellular matrix (ECM).[2−4] Unlike other inter-alpha-trypsin heavy chains, ITIH4 lacks the consensus sequence required for hyaluronan-mediated covalent linkage to bikunin (a peptide with proteinase inhibitory activity) and therefore does not possess intrinsic trypsin-inhibitory activity.[5] ITIH4 circulates in human blood at a concentration of approximately 100 μg/mL.[6] It is a positive acute phase protein regulated by IL-6 and implicated in liver development and regeneration.[7−11] ITIH4 can be cleaved by plasma kallikrein into two fragments that undergo additional proteolytic processing into an 85 kDa N-terminal fragment and a 35 kDa C-terminal fragment.[12−14] The plasma kallikrein cleavage site is found within the proline-rich region (PRR) of ITIH4. ITIH4 also contains a von Willebrand factor A (vWF) domain, which may mediate interactions with other proteins.[1,2] Altered expression and protein levels of ITIH4 are found in a number of cancers including ovarian[15,16] and lung cancer[17] and in hepatic fibrosis.[18,19] Serum ITIH4 proteolytic fragments resulting from proteolytic processing of the PRR are also of interest due to their cancer biomarker potential.[20−22]

Protein glycosylation is a common co-/post-translational modification that may affect protein solubility, structure, function, and proteolytic processing and can mediate interactions with glycan binding proteins involved in protein turnover, inflammatory responses, and/or adhesion.[23−25] The glycosylation state of extracellular matrix glycoproteins can also impact cell adhesion and cell−cell communication.[26] In addition, studies have demonstrated that glycosylation can provide protection from proteolytic cleavage.[27] Covalently linked carbohydrates account for approximately 20% of the molecular weight of the mature ITIH4 protein.[12,14] ITIH4 is N-glycosylated at N81, N207, N517, and N577,[28−31] and there is also evidence that ITIH4 is O-glycosylated within the PRR and near the plasma kallikrein cleavage site.[12,32] Characterization of ITIH4 glycoforms may aid in determining if ITIH4 stability is affected by glycosylation,

Received: February 15, 2014
Published: June 2, 2014

offer clues about its role in extracellular matrix stability and as an acute phase protein, and offer insight into potential interactions of ITIH4 with glycan binding proteins involved in inflammation and adhesion. Previous reports indicate that all four canonical N-glycosylation sites (N-X-S/T) of ITIH4 are glycosylated in human serum, but site occupancy and microheterogeneity of its glycoforms remain undefined.[27−30] In the current study, we aimed to characterize glycosylation of recombinant and serum-derived ITIH4 through glycosidase-assisted LC−MS/MS analysis of ITIH4 trypsin-GluC glycopeptides enriched via HILIC chromatography and complementary detached glycan analyses. Recombinant overexpressed ITIH4 was used for initial optimization of analytical methods, followed by purification and analysis of serum-derived ITIH4. In this study, we compared N- and O-glycosylation in recombinant ITIH4 overexpressed in HEK293 cells and ITIH4 isolated from human serum and identified a nonconsensus N-glycosylation site. This is to our knowledge the first characterization of site-specific N- and O-glycoforms of ITIH4.

EXPERIMENTAL PROCEDURES

Purification of ITIH4 from Human Serum

Human serum was obtained from five healthy individuals enrolled under protocols approved by the Georgetown University Institutional Review Board. Upon thawing, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche, Mannheim, Germany) was added to serum samples to prevent proteolytic degradation of ITIH4. All subsequent steps were performed at 4 °C, in the presence of protease inhibitors. Proteins were precipitated from pooled serum samples via a step gradient of ammonium sulfate, filtered in Amicon Ultra Centrifugal 30 kDa MWCO filters (EMD Millipore, Billerica, MA, USA), and washed with TBS (20 mM Tris, pH 7.5, 150 mM NaCl). Fractions were monitored for the presence of ITIH4 via Western blot using anti-ITIH4 primary antibody C2 (Santa Cruz Biotechnology, Dallas, TX, USA). ITIH4-containing fractions, precipitated at 40− 55% ammonium sulfate saturation, were applied to columns with Cibacron Blue F3G-A ligand covalently bound to agarose beads (HiTrap Blue HP, 5 mL; GE Healthcare, Little Chalfont, UK) and eluted using a NaCl step gradient. ITIH4 fractions, eluted between 450 mM and 1 M NaCl, were desalted and separated via reversed-phase chromatography on a ProSwift RP-1S column (4.6 × 50 mm; Dionex, Sunnyvale, CA, USA) at a flow rate of 1 mL/min using the following gradient: 0−10 min, 1−30% solvent B; 10−60 min, 30−60% solvent B (solvent A: 2% ACN with 0.08% TFA; solvent B: 98% ACN with 0.05% TFA). Proteins were quantified using a spectrophotometer (A280), and the purity of ITIH4 fractions was assessed by SDS-PAGE stained with Coomassie Blue.

Recombinant ITIH4 Production

Plasmid RC219753, containing a Myc-DDK-tagged cDNA clone of *Homo sapiens* Inter-alpha-trypsin inhibitor heavy chain H4 transcript variant 1 in pCMV6-Entry vector, was obtained from OriGene (Rockville, MD, USA). The clone and resulting protein sequence differ from the canonical ITIH4 sequence (Q14624) at residues 85 (I to N) and 698 (P to T). HEK293 cells were stably transfected with the plasmid DNA with MegaTran 1.0 (Origene, Rockville, MD, USA) in a 1:3 ratio according to manufacturer's instructions. After selection on 500 μg/mL geneticin (G418, Gibco, Life Technologies, Carlsbad, CA), stably transfected cells were tested for ITIH4 over-expression. The stably transfected culture used for recombinant ITIH4 purification was maintained in DMEM (Gibco, Life Technologies, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine, 1% nonessential amino acids (NEAA), and 250 μg/mL of G418. After confluence was reached, media were replaced with supplemented DMEM without FBS after repeated gentle washing with 20 mM HEPES buffer. Media were collected after 24 h of incubation, filtered in Amicon 100 kDa MWCO filters (EMD Millipore, Billerica, MA, USA), washed twice with PBS buffer, and then purified via reversed-phase chromatography on a ProSwift RP-1S column as described above.

Proteolytic Digests

Serum-derived and recombinant ITIH4 samples (25 μg) were suspended in 50 mM NH_4HCO_3, pH 7.8 (Sigma-Aldrich, St. Louis, MO, USA) with 0.05% RapiGest (Waters, Milford, MA, USA), reduced with 5 mM DTT, and alkylated with 15 mM iodoacetamide (Sigma-Aldrich, St. Louis, MO, USA). Next, samples were incubated with 0.5 U (1 μg) of endoproteinase GluC (Roche; Mannheim, Germany) at 25 °C for 18 h. Endoproteinase GluC was deactivated by heating the samples at 95 °C for 10 min followed by incubation with 0.008 U (0.2 μg) trypsin (Promega, Madison, WI) at 37 °C for 18 h. Independently, serial trypsin-chymotrypsin digests were performed, using the same conditions for the trypsin digest as described above. Trypsin was deactivated by heating to 95 °C for 10 min, followed by addition of chymotrypsin (Pierce/Thermo Fisher Scientific, Waltham, MA, USA) at a ratio of 1:50 to total protein, in 50 mM Tris-buffer, pH 8, with 2 mM $CaCl_2$. Samples were digested using a Barocycler (Pressure BioSciences, South Easton, MA, USA) at 50 °C with 99 cycles of alternating pressure (20 kPsi, 50 s, followed by 10 s at atmospheric pressure).

HILIC Chromatography of ITIH4 Glycopeptides

Twenty micrograms of each ITIH4 trypsin/GluC digest was separated on an XBridge HILIC column (5 μm, 2.1 × 100 mm; Waters, Milford, MA, USA) as described previously.[33] A 5 min isocratic equilibration at 90% ACN and 0.05% TFA was followed by a linear gradient from 90% to 50% acetonitrile with 0.05% TFA over 35 min. Dried fractions collected every 2 min were suspended in 20 μL of water with 0.1% formic acid, and 1 μL of each fraction was injected onto a Symmetry C18 (100 Å, 3 μm, 180 μm, 20 mm; Waters, Milford, MA, USA) trap column and a BEH C18 column (NanoAcquity, 300 Å 1.7 μm particles, 75 μm ID × 150 mm; Waters, Milford, MA, USA) coupled to a QStar Elite mass spectrometer (Applied Biosystems, Foster City, CA, USA) to monitor glycopeptides by nano-LC−ES-MS/MS. Trapping/washing was performed using 2% ACN, 0.1% formic acid at 15 μL/min flow rate. Separation was performed at a flow rate of 0.4 μL/min using the following gradient: 0 min 99% A, 14 min 55% A, 15 min 1% A, 17 min 1% A, 18 min 99% A, 30 min 99% A (solvent A, 0.1% formic acid in 2% ACN; solvent B, 0.1% formic acid in 98% ACN). Ion spray voltage was set to 2500 V, ion source gas (GS1) 20, declustering potential 60, and interface heater temperature 150 °C. Using data-dependent mode, after the survey scan (m/z 400−1800) the three most intense precursor ions were selected for collision-induced dissociation (CID). MS/MS spectra were recorded from m/z 150 to 2000. Dynamic exclusion was set at 20 s and 30 counts, and collision energy and accumulation time were set automatically as described previously.[34] MS/MS spectra were subsequently assessed for the presence of oxonium ions (m/z 204, 1^+, HexNAc; m/z 366, 1^+, Hex-HexNAc; m/z 274, 1^+, NeuAc − H_2O; m/z 292, 1^+, NeuAc), resulting from glycopeptide

dx.doi.org/10.1021/pr500394z | J. Proteome Res. 2014, 13, 3314−3329

fragmentation. HILIC fractions collected from 14 to 40 min, with MS/MS spectra containing oxonium ions, were pooled for further analysis.

LC−MS/MS Analysis of Glycopeptides and Exoglycosidase-Treated Glycopeptides

ITIH4 trypsin-GluC digests, trypsin-chymotrypsin digests, HILIC-enriched fractions of both digests and exoglycosidase-treated glycopeptide fractions were separated by reversed phase chromatography (Tempo Eksigent-AB Sciex, Framingham, MA, USA) on a ChromXP C18-CL (3 μm, 120 Å, 180 μm, 20 mm) trap column and ChromXP C18-CL (3 μm, 120 Å, 75 μm, 150 mm) HPLC capillary chip column (Eksigent-AB Sciex) interfaced with a 5600 TripleTOF mass spectrometer (AB Sciex, Framingham, MA, USA). First, a 10 min trapping/washing step was performed using 2% ACN, 0.1% formic acid (FA) at a 3 μL/min flow rate followed by a 123 min gradient elution (solvent A: 2% ACN with 0.1% FA; solvent B: 100% ACN with 0.1% FA) at a flow rate of 300 nL/min, with the following conditions: 5−35% solvent B 0−90 min; 35−98% solvent B 90−93 min; 98% solvent B 93−100 min. As described previously, mass spectrometric conditions were set to ion spray voltage 2400 V, ion source gas (GS1) 13, declustering potential 100 and interface heater temperature 150 °C. The instrument was operated in data-dependent mode; after a survey scan (m/z 400−1600), the 20 most abundant precursor ions were selected for CID, with collision energy and accumulation time set automatically. MS/MS spectra were recorded from m/z 100 to 1600, with the dynamic exclusion time set to 6 s, and 150 counts, for two repeated precursors.[35] To remove sialic acid, we treated 1 μg of enriched glycopeptides with 5 units of 2/3,6,8 neuraminidase (NEB, Ipswich, MA, USA) in 50 mM sodium citrate, pH 6, with the goal of collapsing multiple sialylated glycoforms into fewer desialylated species. To elucidate the nature of fucose linkage on N-linked glycans, we treated 1 μg of enriched glycopeptides with 5 units of 2−3,6,8 neuraminidase (NEB, Ipswich, MA, USA) and 0.5 mU alpha 1−3,4 fucosidase (Sigma-Aldrich, St. Louis, MO, USA) in 50 mM sodium acetate, 5 mM $CaCl_2$, pH 5.5. Exoglycosidase-treated fractions were analyzed via LC−MS/MS methods, as described above.

MS/MS Data Interpretation

Peptide identification was performed with ProteinPilot software (AB Sciex, Framingham, MA, USA) using the Paragon algorithm[36] with 20 274 human protein sequences from the UniProtKB/SwissProt database (release 31-Oct-2012). Search parameters include trypsin/GluC or trypsin/chymotrypsin digest with up to one missed cleavage, up to one variable modification (oxidation of methionine), carbamidomethylation of cysteines (fixed), 0.1 Da mass tolerance for precursors and fragment ions. ProteinPilot results were filtered to retain peptides with ≥95% confidence and proteins with ≥3 peptides. MS/MS spectral data files (.wiff) underwent peak detection and output in mzXML format using the msconvert tool from ProteoWizard (v 2.1).[37] The GlycoPeptideSearch (GPS) tool was used to match glycopeptides to CID MS/MS spectral data sets[32,38] with a subset of human N-glycan structures from GlycomeDB that have been annotated to indicate consistency with known glycosylation enzymes (GlycO).[39,40] Specific and semispecific *in silico* trypsin and endoproteinase GluC digests (for analysis of N-glycosylation sites N81, N207 and N517) or trypsin and chymotrypsin digests (for analysis of N-glycosylation site N577) were applied to the human ITIH4 sequence from UniProt (Q14624). For this search, we also considered peptides with the two known ITIH4 sequence variants at residues I85 (I to N) and P698 (P to T). Glycopeptide-spectrum matches were filtered as previously described[32] with minor changes discussed below. Briefly, spectra were required to contain at least 1 oxonium ion with relative intensity ≥10% and two intact peptide ions with relative intensity ≥5%. In addition, a maximum precursor tolerance of 0.1 Da and an isotope-cluster score (ICScore) ≤ 20 were required for all peptide matches. A second set of GPS searches was performed on glycopeptide MS/MS data sets using lists of trypsin-GluC or trypsin-chymotrypsin peptides confirmed to be glycosylated based on labeling experiments with PNGaseF/$H_2{}^{18}O$. To calculate glycopeptide relative intensity, we submitted observed glycopeptide retention times and precursor m/z values to PeakView software (ABSciex, v1.2) and enforced a 0.5 min retention time window and 10 ppm mass tolerance. The intensity of each glycoform was then divided by the sum of all glycopeptide intensities for a given peptide substrate to determine the relative intensity of each glycopeptide glycoform.

Site-Occupancy Determination

Occupancy of ITIH4 N-glycosylation sites was quantified as previously described.[34] Briefly, 1 μg aliquots of ITIH4 trypsin-GluC and trypsin-chymotrypsin glycopeptides were heated at 95 °C for 10 min to deactivate proteases and then dried and resuspended in 10 μL of $H_2{}^{18}O$ (Cambridge Isotope Laboratories, Tewksbury, MA, USA) with 500 units of PNGaseF and 1× protease inhibitor cocktail and incubated overnight at 37 °C. Deglycosylated glycopeptides were then separated by reversed phase chromatography (Tempo Eksigent-AB Sciex, Framingham, MA, USA) on a ChromXP C18-CL (3 μm, 120 Å, 180 μm, 20 mm) trap column and ChromXP C18-CL (3 μm, 120 Å, 75 μm, 150 mm) HPLC capillary column (Eksigent-AB Sciex) interfaced with a 5600 TripleTOF mass spectrometer (AB Sciex, Framingham, MA, USA). First, a 10 min trapping/washing step was performed using 2% ACN with 0.1% FA at 3 μL/min, followed by a 45 min gradient elution (solvent A: 2% ACN with 0.1% FA; solvent B: 100% ACN with 0.1% FA) at a flow rate of 300 nL/min, as follows: 5% solvent B 0−1 min; 5−35% solvent B 1−16 min; 35−98% solvent B 16−18 min; 98% solvent B 18−24 min. Mass spectrometric conditions were as described above.

Detached N-Glycan Analysis

Recombinant ITIH4 (20 μg) and serum-derived ITIH4 (5 μg) were reduced (DTT) and denatured in lithium dodecyl sulfate (LDS) sample buffer by heating at 99 °C for 10 min and separated on a 4−12% Bis-Tris gel during a 35 min run in MES SDS buffer at 200 V and 120 mA. Proteins were visualized with Coomassie Blue; gel bands at 120 kDa were excised, destained using 50% acetonitrile with 25 mM NH_4HCO_3, and washed with water prior to incubation with 500 units of PNGaseF (NEB, Ipswich, MA, USA) in 50 mM sodium phosphate, pH 7.5 for 12 h at 37 °C. The supernatant containing detached N-glycans was removed, and gel bands were treated with trypsin in 10 mM NH_4HCO_3 overnight at 37 °C; protein identity was confirmed by MS/MS. Permethylation and analysis of detached N-glycans was performed as described previously with the following modifications.[41,42] Spectra were acquired on a 4800 MALDI-TOF/TOF (AB Sciex, Framingham, MA, USA) in positive ion mode as previously described.[43]

Detached O-Glycan Analysis

Twenty micrograms of recombinant ITIH4 underwent reductive beta-elimination.[44] Briefly, samples were treated with 1 M $NaBH_4$ in 50 mM NaOH for 18 h at 45 °C, neutralized by

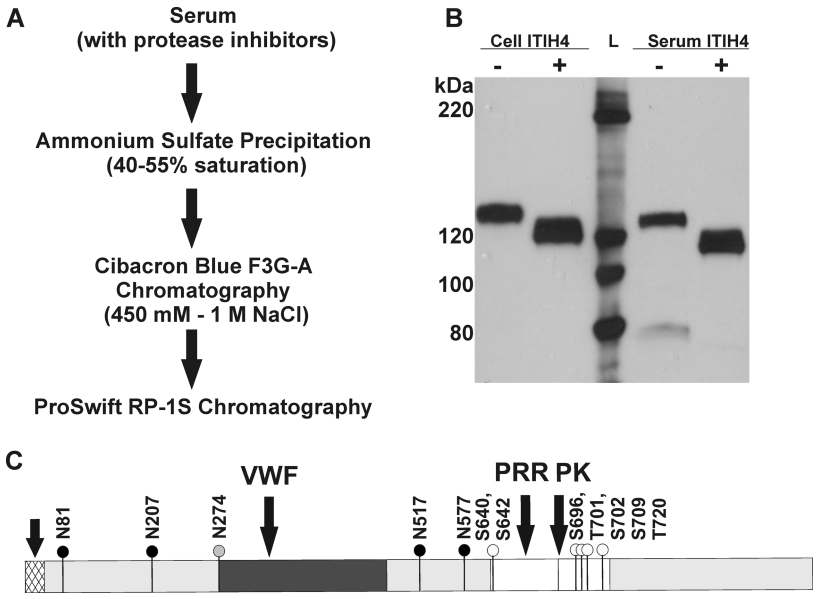


Figure 1. Purification of glycoprotein ITIH4 from human serum. (A) Purification schema of ITIH4 from serum. (B) Western blot (anti-ITIH4, SC-21987) of ITIH4 overexpressed and purified from HEK293 cells (Cell ITIH4) and from serum (serum ITIH4). Lane 1: Cell-derived ITIH4; lane 2: Cell ITIH4 treated with PNGaseF; lane 3: protein ladder with protein molecular weights indicated to the left of the blot; lane 4: serum ITIH4; lane 5: serum ITIH4 treated with PNGaseF. (C) Schema of ITIH4 glycosylation sites.

Table 1. ITIH4 Purification from Serum[a]

fraction	vol (μL)	tot prot (μg)	ITIH4 (μg)	fold purif	% yield
serum	1000	62000	123	1.0	100
40−55% ammonium sulfate	300	10360	103.2	5.0	83.9
cibacron (450 mM to 1 M NaCl)	390	1460	38.1	13.2	31.0
ProSwift RP-1S	100	109	26.8	123.9	21.8

[a]Enrichment and yield of ITIH4 from serum.

addition of 10% acetic acid, desalted using an AG50-X8 (Bio Rad) cartridge, and eluted with 5% acetic acid. Borate was removed by resuspending the dried sample in 9:1 methanol/acetic acid followed by drying under a stream of nitrogen at 37 °C four times. Detached glycans were permethylated, as described above. Permethylated glycans were dissolved in 1 mM NaOH/50% MeOH and directly infusion via nanospray emitter onto a tandem mass spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific, Waltham, MA, USA) with a flow of 0.4 μL/min for manual MS/MS analysis. Intact permethylated O-glycans were analyzed via FTMS from m/z 300−2000, and MS/MS spectra were acquired from m/z 50−2000.

O-Glycosylation Site Mapping

Serum ITIH4 (20 μg) was isolated on a 4−12% Bis-Tris gel under reducing conditions as described above. Protein bands were excised from the gel and, after destaining, treated with 500 units of PNGaseF overnight at 37 °C. The gel bands were washed by sonication in water. In-gel trypsin digestion was performed by adding 10 mM NH_4HCO_3 at pH 7.8 and 0.2 μg of trypsin followed by incubation at 37 °C overnight. Peptides were extracted with 50% ACN, desalted using a Sep-Pak TC18 cartridge (100 mg, Waters, Milford, MA, USA), and reacted for 1 h at room temperature with mTRAQ (AB Sciex, Framingham, MA, USA) reagent Δ4 label with 70% ethanol to boost the O-glycopeptide precursor charge prior to ETD analysis. mTRAQ Δ4-labeled peptides were separated via reversed-phase chromatography and analyzed on a linear ion trap mass spectrometer (LTQ Orbitrap XL) using targeted analysis of mTRAQ Δ4-labeled O-glycopeptides. Glycopeptide samples were loaded onto a 75 μm × 8.5 cm C18 reverse phase column (YM C GEL ODS-AQ120AS-5) using nitrogen pressure. Peptides were eluted over a 100 min linear gradient from 5 to 100% solvent B with a flow rate of 250 nL/min. The ion source capillary temperature was set to 200 °C with a capillary voltage of 46 V and a tube lens voltage of 120 V. Full MS scans were collected from m/z 300−2000, followed by ETD fragmentation of targeted analytes (90 ms reaction time) and collection of MS/MS spectra (m/z 200−2000).

RESULTS

Purification of ITIH4 from Cell Media and Human Serum

Recombinant ITIH4 was isolated from the cell media of HEK293 cells overexpressing ITIH4-cMyc-DDK protein by ultrafiltration followed by RP-1S chromatography. After these steps, ITIH4 exceeded 90% purity based on SDS-PAGE with Coommassie blue staining (data not shown). Proteomic analysis of a tryptic digest confirmed the identity of ITIH4 (71% sequence coverage) with minimal protein impurities. To purify ITIH4 from serum, we devised a three-step purification process to minimize biased selection of specific glycoforms by the chromatographic steps (Figure 1A). Ammonium sulfate precipitation resulted in a 5-fold enrichment of ITIH4 with a yield of 84% (Table 1). Further enrichment was achieved by Cibacron Blue chromatography (3-fold enrichment) followed by reversed-phase chromatography on a monolithic protein purification column (10-fold enrichment). The observed mass

dx.doi.org/10.1021/pr500394z | J. Proteome Res. 2014, 13, 3314−3329

Table 2. N-Glycosylation Occupancy of Serum and Recombinant ITIH4[a]

site	peptide(s)	*m/z* (charge) unlabeled	*m/z* (charge) labeled	serum ITIH4	cell ITIH4
N81	KAFITNF (T,G)	420.734 (2+)	422.240 (2+)	+	>80%
N207	STFMTNQLVDALTTWQNK (T,G)	700.014 (3+)	701.018 (3+)	>90%	>97%
	MTNQLVDALTTWQNK (T,C)	881.943 (2+)	883.449 (2+)		
N274	NVVFVIDK (T,G)	467.274 (2+)	468.780 (2+)	<1%	<1%
N517	LPTQNITFQTE (T,G)	646.330 (2+)	647.836 (2+)	>98%	>99%
N577	NQALNLSLAY (T,C)	553.796 (2+)	555.302 (2+)	>95%	>90%
	NQALNLSLAYSFVTPLTMVVTKPDDQE (T,G)	1027.853 (3+)	1028.857 (3+)		

[a]Relative occupancy was determined using extracted ion chromatograms from the MS survey scan of deglycosylated (PNGaseF, ^{18}O) and non-glycosylated peptides from serial digests with trypsin, endoproteinase GluC (T, GluC), or chymotrypsin (T, C).

shift of recombinant and serum ITIH4 on SDS-PAGE after treatment with PNGaseF confirmed that ITIH4 was heavily N-glycosylated (Figure 1B).

N-Glycan Site Occupancy of Serum and Cell-Derived ITIH4

To confirm that the four canonical N-glycosylation motifs in ITIH4 were glycosylated, we treated ITIH4 glycopeptides (enriched using HILIC chromatography) with PNGaseF in the presence of $H_2{}^{18}O$. To avoid ^{18}O incorporation into the peptide C-terminus during trypsin digestion, we inactivated proteases by heating and addition of protease inhibitors prior to PNGaseF/^{18}O labeling. ^{18}O-Labeled *b*- and *y*-ions in the peptide MS/MS spectra enabled residue-specific resolution of the N-glycosylation sites. Peptides containing the four consensus sequence N-glycosylation sites were detected with the expected +3 Da mass shift. The MS/MS spectra of ^{18}O-labled peptides confirm the expected glycosylation of the N81, N207, N517, and N577 sites within the NXS/T motifs. We also detected a fifth peptide, NVVFVIDK, with a +3 Da shift, in both recombinant and serum ITIH4. The *b*- and *y*-ions in the MS/MS spectrum of the (N > D/^{18}O)VVFVIDK peptide definitively confirm that the +3 Da shift is located at N274. Figure 1C shows a schematic of the confirmed ITIH4 glycosylation sites.

To provide a quantitative estimate of occupancy at each ITIH4 N-glycosylation site, we applied this same PNGaseF/$H_2{}^{18}O$ isotopic labeling strategy to ITIH4 digests, as previously described.[45,46] Importantly, while the N-glycosylation site confirmation experiments above were performed on enriched glycopeptide fractions, these occupancy experiments were performed on unprocessed (non-HILIC enriched) ITIH4 digests. This enabled us to estimate the glycosylation site occupancy by comparing the intensity of an ^{18}O-labeled peptide (which was formerly glycosylated) with the unlabeled (and therefore nonglycosylated) peptide containing the same N-glycosylation site. Quantification of the non-glycosylated and corresponding deglycosylated peptides in extracted ion chromatograms (XIC) of ITIH4 digests treated with PNGaseF/$H_2{}^{18}O$ shows that at least three of the four consensus glycosylation sites (N207, N517, and N577) are highly occupied (>90%) in serum-derived ITIH4 (Table 2). We observe two labeled (+3 Da) semispecific peptides containing site N81, AFIT(N > D/^{18}O)F (residues 77−82) and KAFIT(N > D/^{18}O)F (residues 76−82), in ITIH4 trypsin-GluC digests treated with PNGaseF/$H_2{}^{18}O$. We were unable to confidently quantify the occupancy of site N81 in serum ITIH4 because the non-glycosylated peptides containing site N81 are below the limit of detection in serum-derived ITIH4 digests. In recombinant ITIH4, site N81 is partially occupied (>80%) and sites N207, N517, and N577 are highly occupied (>90%). Occupancy of the noncanonical NVV motif is low (<1%) in both serum and recombinant ITIH4. These observations are also confirmed by glycopeptide MS/MS data for both recombinant and serum-derived ITIH4 (Figure 2). These results demonstrate that N-glycosylation site occupancy of the five ITIH4 *N*-glycosylation sites is similar in serum and recombinant ITIH4.

N-Glycan Microheterogeneity in Recombinant ITIH4

To achieve site-specific characterization of N- and O-glycan microheterogeneity, we performed serial proteolysis (GluC-trypsin and trypsin-chymotrypsin) of ITIH4, enriched glycopeptides using HILIC chromatography, and analyzed glycopeptides via LC−ESI-MS/MS before and after treatment with neuraminidase and fucosidase. All observed glycoforms are summarized in Table 3. In recombinant ITIH4, biantennary sialylated glycans were the dominant glycoforms at N81; triantennary, fucosylated, and high-mannose forms were also observed. At sites N207, N517, and N577, complex fucosylated, asialo-glycans were the dominant glycoforms, though some sialylated glycoforms were also observed. At N207, all detected glycoforms were fucosylated, and several glycopeptides with bifucosylated N-linked glycans were detected. We observed the greatest microheterogeneity at N517, with nonfucosylated, monofucosylated, and bifucosylated bi-, tri-, and tetra-antennary N-linked glycans. At N577 both sialylated and asialo-glycoforms were observed, as well as singly fucosylated N-linked glycans. Analysis of detached, permethylated N-linked glycans from recombinant ITIH4 corroborates the glycan compositions matched to glycopeptides from CID MS/MS. The detached glycan data also confirmed the abundance of fucosylated glycoforms—fucosylated N-linked glycans accounted for 89% of the relative intensity of permethylated glycans analyzed via MALDI-TOF MS (Supplementary Figure 1, Supporting Information). Double-fucosylated N-linked glycans contributed 3% of the N-glycan relative intensity. Fully sialylated N-linked glycans accounted for only 13% of the signal, confirming glycopeptide results that also demonstrated that fully sialylated glycans are a minor component of recombinant ITIH4 N-glycoforms. Notably, we observed the M5 N-linked glycan in MALDI-TOF spectra, representing 0.1% of the relative intensity, consistent with our previous observation that high-mannose glycopeptides are very low in abundance. We used optimized methods for analysis of recombinant ITIH4 to study site-specific glycoforms in serum-derived ITIH4.

N-Glycan Microheterogeneity in Serum-Derived ITIH4

Analysis of serum-derived ITIH4 glycopeptides showed that all canonical glycosylation sites in serum-derived ITIH4 contained complex sialylated N-linked glycans (Table 3). Analysis of detached, permethylated N-linked glycans from serum-derived ITIH4 was consistent with glycopeptide results (data not shown). Glycosylation sites N81 and N207, located closest to the N-terminus of the protein, carried primarily bi- and triantennary

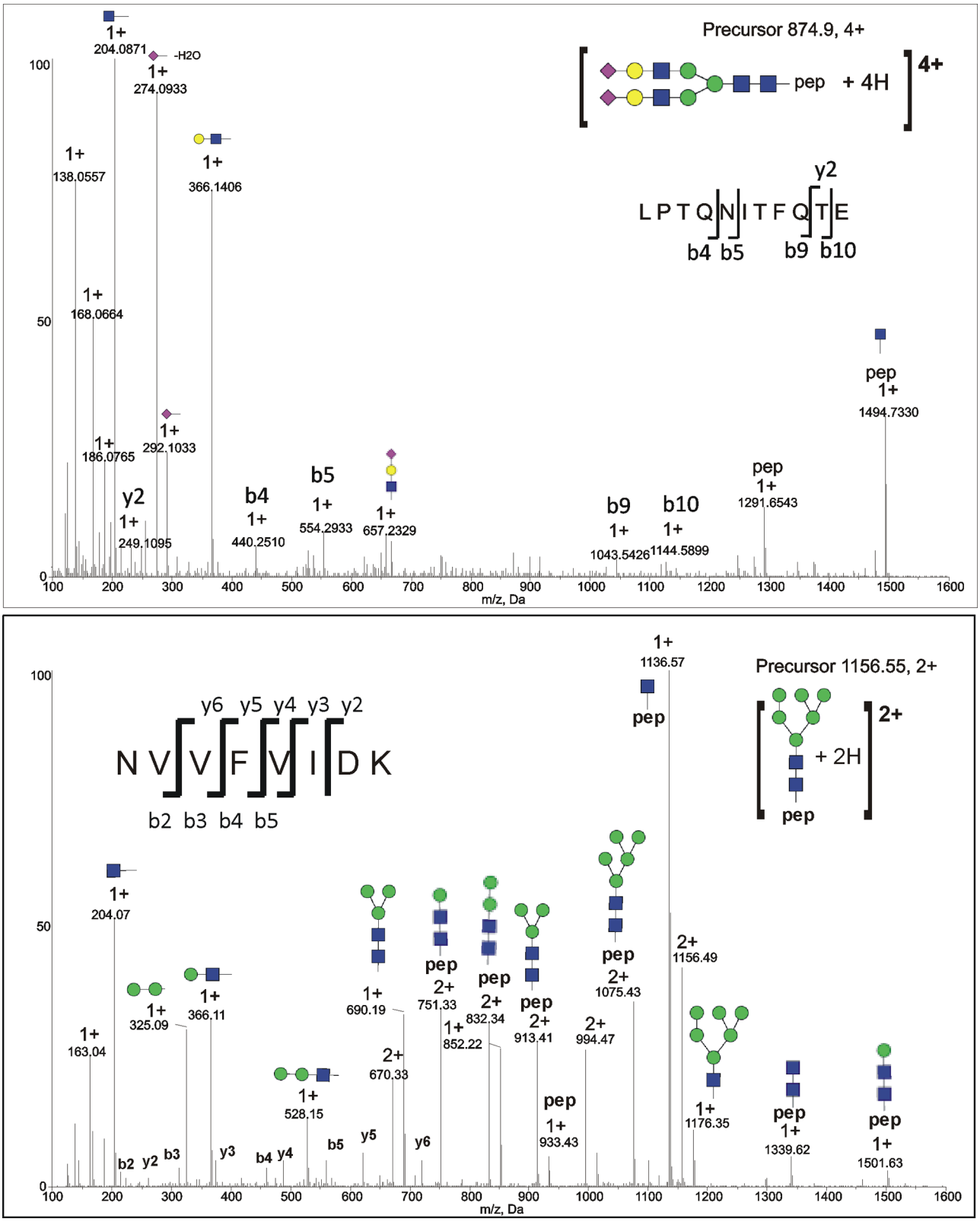


Figure 2. CID spectra of selected ITIH4 *N*-glycopeptides. (Top) CID spectrum of [LPTQNITFQTE + A2G2S2 + 4H]$^{4+}$ (site N517) glycopeptide from serum ITIH4. (Bottom) CID spectrum of noncanonical glycosylation site N274, peptide [NVVFVIDK + HexNAc2Hex6 + 2H]$^{2+}$.

sialylated glycans; a higher degree of microheterogeneity was present at the N517 and N577 sites with additional fucosylated and tetra-antennary sialylated glycans observed in addition to the dominant bi- and triantennary sialylated forms. We detected the highest number of glycoforms, and highly branched tetra-antennary glycans, at the N517 site. A CID fragmentation spectrum of [LPTQNITFQTE + A2G2S2 + 4H]$^{4+}$ (m/z 874.9) glycopeptide is shown in Figure 2A. Oxonium ions, including [HexNAc + H]$^{1+}$ at m/z 204, [NeuAc − H_2O + H]$^{1+}$ at m/z 274, [Neu5Ac + H]$^{1+}$ at m/z 292, [HexNAc-Hex + H]$^{1+}$ at m/z 366

dx.doi.org/10.1021/pr500394z | J. Proteome Res. 2014, 13, 3314−3329

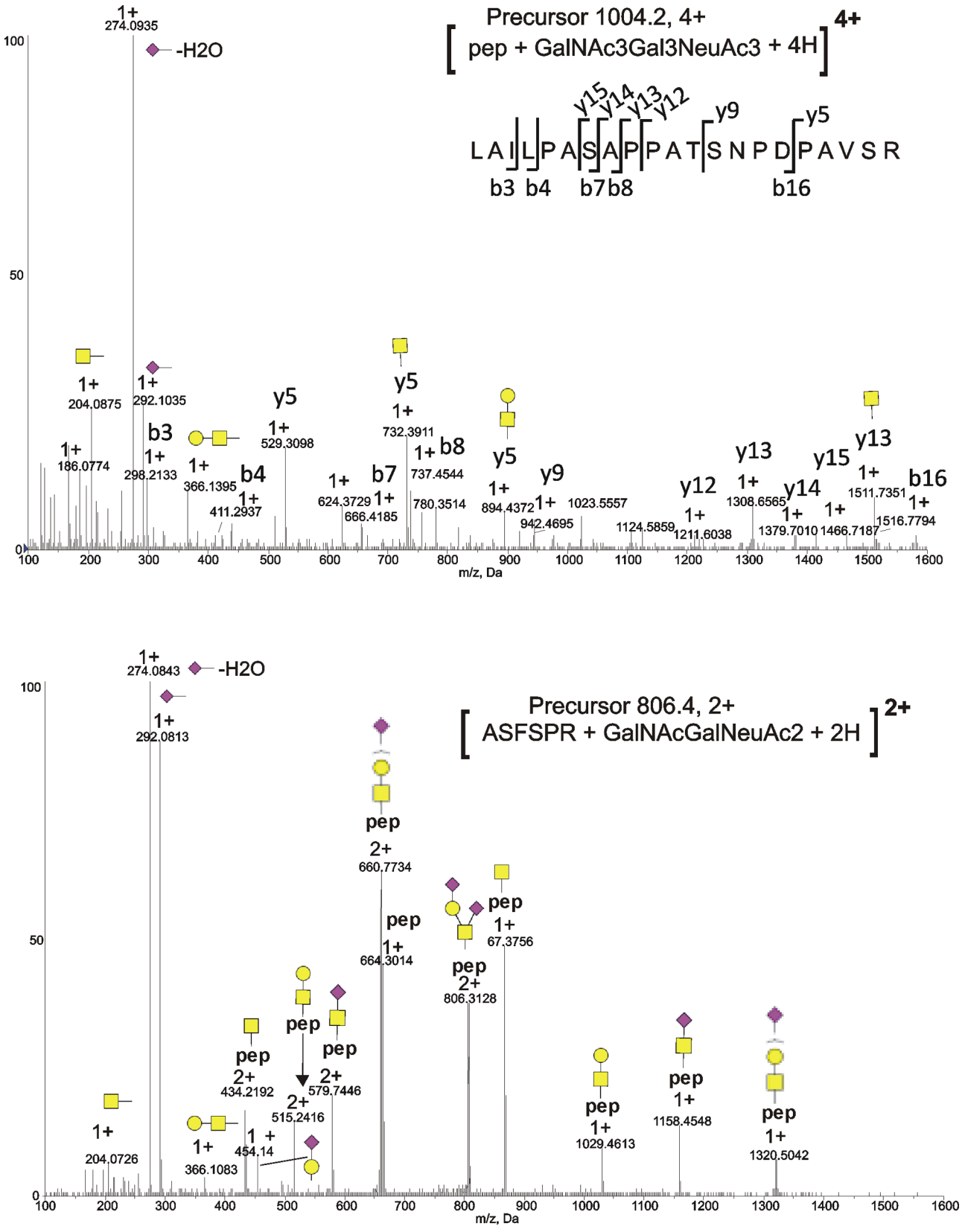


Figure 3. CID Spectra of selected ITIH4 O-glycopeptides. (Top) Spectrum of O-glycopeptide [LAILPASAPPATSNPDPAVSR + $HexNAc_3Hex_3NeuAc_3$ + 4H]$^{4+}$. (Bottom) Spectrum of O-glycopeptide [ASFSPR + $HexNAc_1Hex_1NeuAc_1$ + 2H]$^{2+}$.

and [HexNAc-Hex-Neu5Ac + H]$^{1+}$ at *m/z* 657 in the spectrum are diagnostic of glycopeptide fragmentation and offer clues regarding glycan topology; intact peptide peaks including [LPTQNITFQTE + H]$^{1+}$ at *m/z* 1291.65, and [LPTQNITFQTE + HexNAc + H]$^{1+}$ at *m/z* 1494.73 enable the assignment of peptide identity and the determination of the N-linked glycan

dx.doi.org/10.1021/pr500394z | J. Proteome Res. 2014, 13, 3314−3329

Table 3. (A) Site-Specific N- and O-Glycan Microheterogeneity of Recombinant ITIH4 and (B) Site-Specific N- and O-glycan Microheterogeneity of Serum ITIH4[a]

(A)

AFITNF (N81)

composition	Oxford	*m/z* (z)	%
HexNAc2Hex5	M5	964.90 (2+)	4
HexNAc2Hex6	M6	1045.93 (2+)	2
HexNAc4Hex3	A2	670.96 (3+)	2
HexNAc3Hex5		711.29 (3+)	2
HexNAc4Hex4	A2G1	724.97 (3+)	1
HexNAc3Hex6		765.31 (3+)	2
HexNAc4Hex5	A2G2	778.99 (3+)	14
HexNAc5Hex4	A3G1	792.68 (3+)	6
HexNAc3Hex5NeuAc1		808.33 (3+)	4
HexNAc4Hex5dHex1	FA2G2	827.67 (3+)	3
HexNAc5Hex4dHex1	FA3G1	841.35 (3+)	4
HexNAc5Hex5	A3G2	846.69 (3+)	3
HexNAc4Hex5NeuAc1	A2G2S1	876.02 (3+)	42
HexNAc4Hex5NeuAc2	A2G2S2	973.05 (3+)	10
HexNAc5Hex6NeuAc3	A3G3	900.70 (3+)	1

ALTTWQNK (N207)

composition	Oxford	*m/z* (z)	%
HexNAc4Hex3dHex1	FA2	802.69 (3+)	5
HexNAc4Hex4dHex1	FA2G1	856.70 (3+)	3
HexNAc4Hex5dHex1	FA2G2	910.72 (3+)	13
HexNAc5Hex4dHex1	FA3G1	924.4 (3+)	7
HexNAc3Hex7Fuc1		951.03 (3+)	5
HexNAc4Hex5dHex2	FA2F1G2	959.41 (3+)	2
HexNAc5Hex4dHex2	FA3F1G1	973.08 (3+)	12
HexNAc5Hex5dHex1	FA3G2	978.42 (3+)	4
HexNAc6Hex3Fuc2		986.76 (3+)	2
HexNAc4Hex5NeuAc1dHex1	FA2G2S1	1007.75 (3+)	14
HexNAc5Hex4NeuAc1dHex1	FA3G1S1	1021.43 (3+)	10
HexNAc3Hex7NeuAc1dHex1		1048.08 (3+)	7
HexNAc5Hex4NeuAc1dHex2	FA3F1G1S1	1070.11 (3+)	6
HexNAc4Hex5NeuAc2dHex1	FA2G2S2	1104.79 (3+)	10

NVVFVIDK (N274)

composition	Oxford	*m/z* (z)	%
HexNAc2Hex5	M5	717.33 (3+)	20
HexNAc2Hex6	M6	771.34 (3+)	62
HexNAc2Hex7	M7	825.36 (3+)	13
HexNAc2Hex8	M8	879.38 (3+)	5

LPTQNITFQTE (N517)

composition	Oxford	*m/z* (z)	%
HexNAc4Hex3dHex1	FA2	912.73 (3+)	3
HexNAc4Hex4dHex1	FA2G1	966.75 (3+)	2
HexNAc5Hex3dHex1	FA3	980.43 (3+)	1
HexNAc4Hex4NeuAc1	A2G1S1	761.58 (4+)	<1
HexNAc4Hex5dHex1	FA2G2	1020.77 (3+)	4
HexNAc5Hex4dHex1	FA3G1	776.09 (4+)	<1
HexNAc4Hex4NeuAc1dHex1	FA2G1S1	1063.78 (3+)	1
HexNAc4Hex5dHex1	FA2F1G2	1069.46 (3+)	1
HexNAc5Hex4dHex1	FA3G1	1034.43 (3+)	2
HexNAc5Hex4dHex2	FA3F1G1	1083.13 (3+)	3
HexNAc5Hex4NeuAc1dHex1	FA3G1S1	1131.48 (3+)	9
HexNAc5Hex5dHex1	FA3G2	1088.46 (3+)	3
HexNAc4Hex5NeuAc1dHex1	FA2G2S1	1117.80 (3+)	26
HexNAc5Hex5NeuAc1dHex1	FA3G2S1	889.37 (4+)	2
HexNAc5Hex5dHex2	FA3F1G2	1137.15 (3+)	<1

Table 3. continued

LPTQNITFQTE (N517)

composition	Oxford	*m/z* (z)	%
HexNAc5Hex6dHex1	FA3G3	857.11 (4+)	3
HexNAc4Hex5NeuAc1dHex2	FA2F1G2S1	1166.49 (3+)	<1
HexNAc4Hex5NeuAc2dHex1	FA2G2S2	911.38 (4+)	5
HexNAc5Hex5NeuAc2dHex1	FA3G2S2	962.15 (4+)	<1
HexNAc5Hex6NeuAc1dHex1	FA3G3S1	929.89 (4+)	9
HexNAc5Hex6NeuAc1dHex2	FA3F1G3S1	966.39 (4+)	2
HexNAc5Hex6NeuAc2dHex1	FA3G3S2	1002.66 (4+)	7
HexNAc5Hex6NeuAc3dHex1	FA3G3S3	1075.43 (4+)	1
HexNAc6Hex5NeuAc1dHex1	FA4G2S1	940.14 (4+)	2
HexNAc6Hex7dHex1	FA4G4	948.40 (4+)	1
HexNAc6Hex7NeuAc1dHex1	FA4G4S1	1021.17 (4+)	3
HexNAc6Hex7NeuAc2dHex1	FA4G4S2	1093.94 (4+)	5
HexNAc6Hex7NeuAc4dHex1	FA4G4S4	1239.49 (4+)	<1

NQALNLSLAY (N577)

composition	Oxford	*m/z* (z)	%
HexNAc4Hex3dHex1	FA2	851.04 (3+)	5
HexNAc4Hex4dHex1	FA2G1	905.06 (3+)	2
HexNAc4Hex5dHex1	FA2G2	959.07 (3+)	4
HexNAc4Hex5NeuAc1dHex1	FA2G2S1	1056.11 (3+)	17
HexNAc4Hex5NeuAc2dHex1	FA2G2S2	865.11 (4+)	4
HexNAc5Hex4dHex1	FA3G1	972.73 (3+)	3
HexNAc5Hex4NeuAc1dHex1	FA3G1S1	1069.78 (3+)	13
HexNAc5Hex6NeuAc1dHex1	FA3G3S1	883.62 (4+)	7
HexNAc5Hex6NeuAc2dHex1	FA3G3S2	956.39 (4+)	7
HexNAc5Hex6NeuAc3dHex1	FA3G3S3	1029.17 (4+)	3
HexNAc6Hex5NeuAc1dHex1	FA4G2S1	893.87 (4+)	3
HexNAc6Hex5NeuAc2dHex1	FA4G2S2	966.65 (4+)	2
HexNAc6Hex7NeuAc1dHex1	FA4G4S1	974.90 (4+)	7
HexNAc6Hex7NeuAc2dHex1	FA4G4S2	1047.67 (4+)	10
HexNAc6Hex7NeuAc2dHex2	FA4F1G4S2	1084.18 (4+)	3
HexNAc6Hex7NeuAc3dHex1	FA4G4S3	1120.45 (4+)	5
unk	unk	1096.44 (3+)	5

IPKPEASFSPR (S640, S642)

composition	*m/z* (z)	%
HexNAc1Hex1NeuAc1	628.98 (3+)	4
HexNAc1Hex1NeuAc2	726.01 (3+)	91
HexNAc1Hex4NeuAc1	791.01 (3+)	5

GPDVLTATVSGK (T506, T508, S510)

composition	*m/z* (z)	%
HexNAc1Hex1NeuAc1	600.96 (3+)	3
HexNAc1Hex1NeuAc2	697.99 (3+)	21
HexNAc2Hex2	625.63 (3+)	10
HexNAc2Hex2NeuAc1	722.66 (3+)	35
HexNAc2Hex2NeuAc 1dHex1	771.35 (3+)	2
HexNAc2Hex2NeuAc2	819.70 (3+)	18
HexNAc2Hex3NeuAc 2dHex1	922.36 (3+)	3
HexNAc1Hex4NeuAc1	762.99 (3+)	5
HexNAc3Hex1NeuAc1	736.34 (3+)	3

LAILPASATPATSNPDPAVSR

composition	*m/z* (z)	%
HexNAc1	751.40 (3+)	17
HexNAc1Hex1	805.41 (3+)	10
HexNAc1Hex1NeuAc2	999.48 (3+)	13
HexNAc2	819.08 (3+)	14
HexNAc2Hex2	927.12 (3+)	6
HexNAc2Hex2NeuAc1	1024.16 (3+)	4
HexNAc2Hex2NeuAc2	841.14 (4+)	2
HexNAc3Hex3NeuAc3	1005.19 (4+)	6

dx.doi.org/10.1021/pr500394z | J. Proteome Res. 2014, 13, 3314−3329

Table 3. continued

LAILPASATPATSNPDPAVSR			
composition		*m/z* (z)	%
HexNAc3Hex3NeuAc4		1077.97 (4+)	6
HexNAc3Hex4		886.78 (3+)	8
HexNAc4Hex4		878.15 (4+)	4
HexNAc4Hex4NeuAc3		1096.49 (4+)	3
HexNAc4Hex4NeuAc4		1169.26 (4+)	2
HexNAc5Hex3		888.41 (3+)	4
(B)			
KAFITNF (N81)			
composition	Oxford	*m/z* (z)	%
HexNAc4Hex5NeuAc2	A2G2S2	1015.76 (3+)	95
HexNAc5Hex6NeuAc3	A3G3S3	926.12 (4+)	5
ALTTWQNK (N207)			
composition	Oxford	*m/z* (z)	%
HexNAc4Hex5NeuAc2	A2G2S2	1056.10 (3+)	100
NVVFVIDK (N274)			
composition	Oxford	*m/z* (z)	%
HexNAc2Hex5	M5	717.32 (3+)	14
HexNAc2Hex6	M6	771.34 (3+)	86
LPTQNITFQTE (N517)			
composition	Oxford	*m/z* (z)	%
HexNAc4Hex5NeuAc1	A2G2S1	802.08 (4+)	3
HexNAc4Hex5NeuAc2	A2G2S2	874.86 (4+)	72
HexNAc4Hex5NeuAc2dHex1	FA2G2S2	911.38 (4+)	3
HexNAc5Hex6NeuAc1	A3G3S1	893.36 (4+)	<1
HexNAc5Hex6NeuAc2dHex1	FA3G3S2	1002.67 (4+)	1
HexNAc5Hex6NeuAc3	A3G3S3	1038.92 (4+)	13
HexNAc5Hex6NeuAc3dHex1	FA3G3S3	1075.43 (4+)	7
HexNAc5Hex6NeuAc3dHex2	FA3F1G3S3	1111.95 (4+)	1
HexNAc6Hex7NeuAc2	A4G4S2	1057.43 (4+)	<1
HexNAc6Hex7NeuAc3	A4G4S3	1130.22 (4+)	<1
NQALNLSLAY (N577)			
composition	Oxford	*m/z* (z)	%
HexNAc4Hex5NeuAc2	A2G2S2	1104.46 (3+)	65
HexNAc5Hex6NeuAc3	A3G3S3	992.65 (4+)	28
HexNAc5Hex6NeuAc3dHex1	FA3G3S3	1029.17 (4+)	7
IPKPEASFSPR (S640, S642)			
composition		*m/z* (z)	%
HexNAc1Hex1NeuAc1		628.98 (3+)	89
HexNAc1Hex1NeuAc2		726.01 (3+)	11
LAILPASAPPATSNPDPAVSR (P698) (S696, T701, T702, S709)			
composition		*m/z* (z)	%
HexNAc2Hex2NeuAc2		840.14 (4+)	36
HexNAc2Hex2NeuAc3		912.91 (4+)	5
HexNAc3Hex3NeuAc3		1004.20 (4+)	50
HexNAc3Hex3NeuAc4		1076.98 (4+)	6
HexNAc4Hex4NeuAc4		1168.26 (4+)	1
LAILPASATPATSNPDPAVSR (T698) (S696, T698, T701, T702, S709)			
composition		*m/z* (z)	%
HexNAc1Hex1NeuAc1		902.44 (3+)	7
HexNAc2Hex2NeuAc2		841.14 (4+)	27
HexNAc3Hex3NeuAc2		932.42 (4+)	2
HexNAc3Hex3NeuAc3		1005.20 (4+)	54
HexNAc3Hex3NeuAc4		1077.97 (4+)	6
HexNAc4Hex4NeuAc4		1169.27 (4+)	4

[a]Glycan composition, putative structure (Oxford notation), mass to charge (*m/z*), charge, and glycopeptide relative intensities are shown.

composition; and peptide *b* and *y* ions (y_2: *m/z* 249.11, 1^+; b_4: *m/z* 440.25, 1^+; b_5: *m/z* 554.29, 1^+; b_9: *m/z* 1043.54, 1^+; and b_{10}: *m/z* 1144.59, 1^+) corroborate the identity of the peptide.

Non-Canonical N-Glycosylation

Consistent with our observation that glycosylation was present at N274 based on PNGaseF/^{18}O labeling (N > D/^{18}O), we also detected four high-mannose glycoforms of the NVVFVIDK peptide, including M5, M6, M7, and M8 glycoforms in recombinant ITIH4, and M5 and M6 glycoforms in serum-derived ITIH4. The CID spectrum of serum-derived glycopeptide NVVFVIDK + M6 (Figure 2B) is representative of CID MS/MS spectra from both serum and recombinant ITIH4 and offers definitive evidence of glycosylation at this non-canonical site. The spectrum contains oxonium ions including $[HexNAc + H]^{1+}$ at *m/z* 204, $[Hex\text{-}Hex + H]^{1+}$ at *m/z* 325, $[HexNAc\text{-}Hex + H]^{1+}$ at *m/z* 366, and $[HexNAc\text{-}Hex\text{-}Hex + H]^{1+}$ at *m/z* 528 consistent with a high-mannose glycan and a strong peptide *b*- and *y*-ion series (b_2-b_5 and y_2-y_6) leading to a confident assignment. In addition, a series of glycopeptide Y ions including $[NVVFVIDK + HexNAc_2Hex_1 + 2H]^{2+}$ at *m/z* 751.33, $[NVVFVIDK + HexNAc_2Hex_2 + 2H]^{2+}$ at *m/z* 832.34, $[NVVFVIDK + HexNAc_2Hex_3 + 2H]^{2+}$ at *m/z* 913.41, $[NVVFVIDK + HexNAc_2Hex_4 + 2H]^{2+}$at *m/z* 994.47, and $[NVVFVIDK + HexNAc_2Hex_5 + 2H]^{2+}$ at *m/z* 1075.43, consisting of the intact peptide and partially fragmented glycan, and B ions including $[HexNAc_1Hex_3 + H]^{1+}$ at *m/z* 690.19, $[HexNAc_1Hex_6 + H]^{1+}$ at *m/z* 1176.35 consisting of the fragmented glycan, are also present in the spectrum. ITIH4 glycopeptides from the recombinant ITIH4 demonstrate a greater variety of the high-mannose glycoforms (M5-M8) (Table 3), but the glycoforms in serum are limited to high-mannose forms as well.

Core- and Outer-Arm Fucosylation of ITIH4 N-Linked Glycans

On the basis of glycopeptide fragmentation data, evidence suggests that most fucosylated recombinant ITIH4 glycoforms are core fucosylated while most fucosylated serum ITIH4 glycoforms are outer-arm fucosylated. The CID spectra of fucosylated N-glycopeptides from recombinant ITIH4 contained $[HexNAc\text{-}Fuc + H]^{1+}$ peaks at *m/z* 350, and peptide-specific peaks equivalent to the mass of the peptide with HexNAc-Fuc, which suggests that these glycopeptides contain core fucosylated glycans.[34,37] However, we also detected $[HexNAc_1Hex_1Fuc_1 + H]^{1+}$ ions at *m/z* 512 in a minority of spectra from recombinant ITIH4, suggesting minor amounts of outer-arm fucosylation may also be present. Because rearrangement of fucose during fragmentation from core to outer arm is unlikely,[47] a likely explanation is that the glycopeptides of recombinant ITIH4 represent a mixture of outer-arm and core fucosylated peptides. In serum-derived ITIH4, we only observed fucosylated glycopeptide ions at sites N517 and N577. However, we observed both $[HexNAc_1Hex_1Fuc_1 + H]^{1+}$ at *m/z* 512 (outer-arm), $[HexNAc\text{-}Fuc + H]^{1+}$ at *m/z* 350 (core), and $[HexNAc_1Hex_1Fuc_1Neu5Ac_1 + H]^{1+}$ at *m/z* 803 in MS/MS spectra consistent with the presence of N-glycans with sialylated outer-arm linked fucose. This suggests that serum ITIH4, like recombinant ITIH4, contains both core- and outer arm-fucosylated N-linked glycans.

In vertebrates, fucose is linked to the N-glycan core via an $\alpha 1{-}6$ linkage, whereas outer-arm fucoses are linked via $\alpha 1{-}3$ linkage. Therefore, treatment with $\alpha(1{-}3,4)$ fucosidase should remove outer-arm fucose residues but should not cleave core fucose. In recombinant ITIH4 most fucosylated glycopeptides

dx.doi.org/10.1021/pr500394z | J. Proteome Res. 2014, 13, 3314−3329

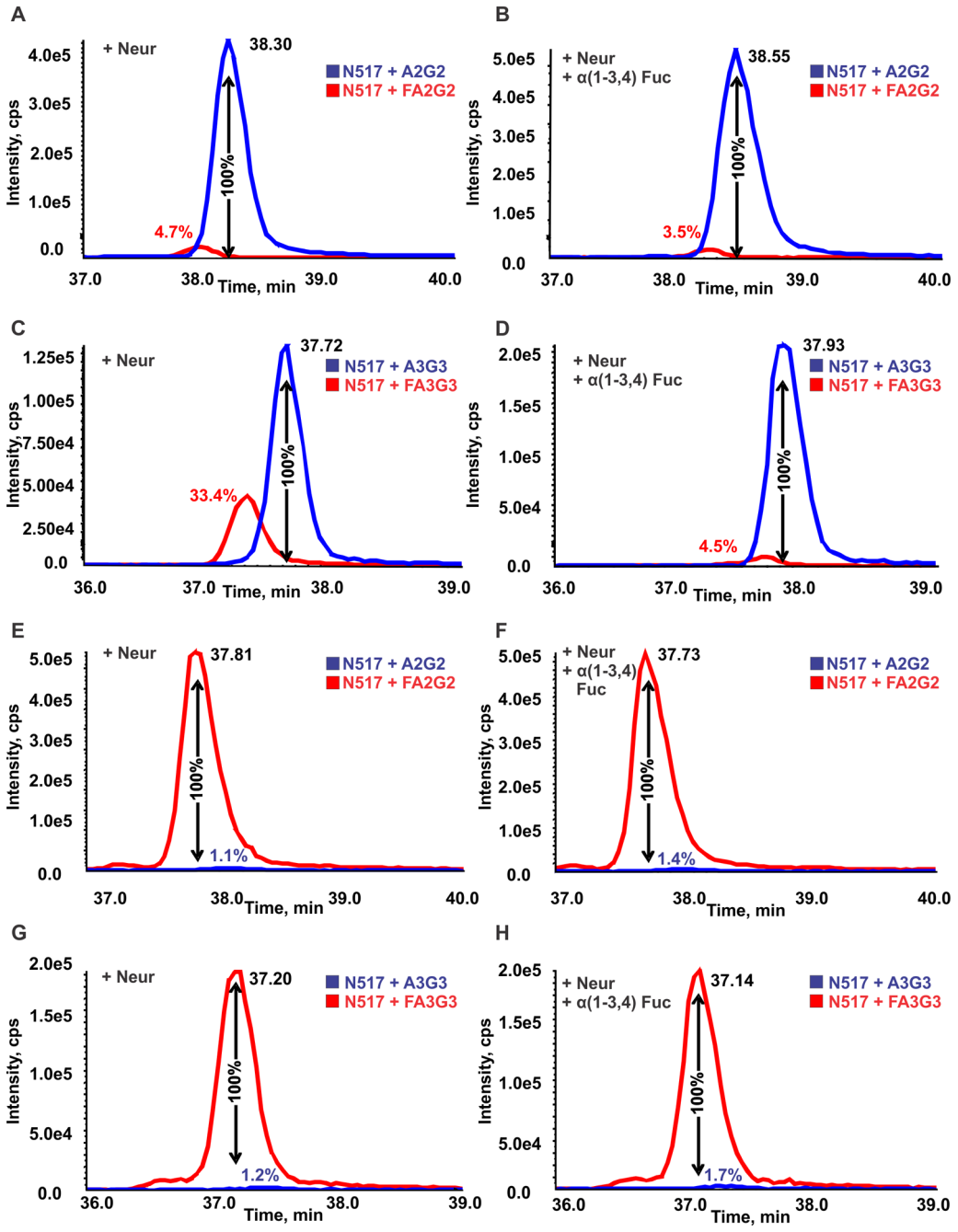


Figure 4. Serum and recombinant ITIH4 glycopeptides treated with neuraminidase, $\alpha(1-3,4)$ Fucosidase. Extracted ion chromatograms of (A) serum-derived glycopeptides LPTQNITFQTE (site N517) + A2G2 and + FA2G2 after treatment with $\alpha(2-3,6,8)$ neuraminidase. (B) Serum-derived glycopeptides LPTQNITFQTE + A2G2 and + FA2G2 after treatment with $\alpha(2-3,6,8)$ neuraminidase and $\alpha(1-3,4)$ fucosidase. (C) Serum-derived glycopeptides LPTQNITFQTE + A3G3 and + FA3G3 after treatment with neuraminidase. (D) Serum-derived glycopeptides LPTQNITFQTE + A3G3 and + FA3G3 after treatment with neuraminidase and $\alpha(1-3,4)$ fucosidase. (E) Recombinant ITIH4 glycopeptides LPTQNITFQTE + A2G2 and + FA2G2 after treatment with $\alpha(2-3,6,8)$ neuraminidase. (F) Recombinant ITIH4 glycopeptides LPTQNITFQTE) + A2G2 and +FA2G2 after treatment with $\alpha(2-3,6,8)$ neuraminidase and $\alpha(1-3,4)$ fucosidase. (G) Serum recombinant ITIH4 glycopeptides LPTQNITFQTE + A3G3 and + FA3G3 after treatment with neuraminidase. (H) Recombinant ITIH4 glycopeptides LPTQNITFQTE + A3G3 and + FA3G3 after treatment with neuraminidase and $\alpha(1-3,4)$ fucosidase.

were resistant to $\alpha(1-3,4)$ fucosidase, indicating that recombinant ITIH4 fucosylated glycoforms are core fucosylated. We observed the near complete disappearance of three doubly fucosylated glycopeptides ALTTWQNK + FA2BF1G1, LPTQNITFQTE + FA2F1G2, and LPTQNITFQTE + FA2BF1G1 in recombinant ITIH4 digests after $\alpha(1-3,4)$ fucosidase treatment. This presumably occurred due to the conversion of these forms to singly fucosylated forms after removal of $\alpha 1-3$ linked fucose on the outer arms of the glycans. These results confirmed that recombinant ITIH4 glycopeptides were highly core-fucosylated, but some also contained outer-arm fucosylation. The ratio of the fucosylated to nonfucosylated forms of recombinant ITIH4 glycopeptides LPTQNITFQTE + A2G2 and LPTQNITFQTE + A3G3 remained virtually unchanged after treatment with $\alpha(1-3,4)$ fucosidase in recombinant ITIH4 (Figure 4E–H); the peak area of the unglycosylated

dx.doi.org/10.1021/pr500394z | J. Proteome Res. 2014, 13, 3314–3329

LPTQNITFQTE + A2G2 glycopeptide is 1.4% of the LPTQNITFQTE + FA2G2 precursor in untreated glycopeptides, versus 1.7% after treatment. A similar pattern was observed for LPTQNITFQTE + A3G3 and LPTQNITFQTE + FA3G3, with the unfucosylated form representing <2% of the fucosylated form before and after fucosidase treatment, indicating that recombinant ITIH4 is predominantly core-fucosylated.

In serum-derived ITIH4, fucosylated biantennary glycoforms are predominantly core-fucosylated while triantennary fucosylated glycoforms contain higher levels of outer-arm fucosylation. The ratio of the fucosylated glycopeptide LPTQNITFQTE + FA2G2 to the unfucosylated LPTQNITFQTE + A2G2 glycopeptide was 4.7% prior to treatment with $\alpha(1-3,4)$ fucosidase (Figure 4A). We observed very little change in the ratio of the intensities of serum-derived ITIH4 glycopeptides LPTQNITFQTE + FA2G2 and LPTQNITFQTE + A2G2 before and after treatment with $\alpha(1-3,4)$ fucosidase (Figure 4A,B), indicating that this glycopeptide contains core-linked fucose. In contrast, we observed that the precursor intensity of LPTQNITFQTE + FA3G3 was 33% of the intensity of the unfucosylated form LPTQNITFQTE + A3G3 prior to fucosidase treatment, whereas after treatment this ratio dropped to 4.5% (Figure 4C,D). Therefore, the fucosylated triantennary glycopeptide containing site N517 is a mixture of outer-arm fucosylated and core-fucosylated forms. On the basis of the $\alpha(1-3,4)$ fucosidase treatment the predominant glycoform is outer-arm fucosylated and a small amount (4.5%) of the core-fucosylated glycoform is also present. In conclusion, core fucosylation dominates the N517 *biantennary* glycan, while outer arm fucosylation represents a majority of the *triantennary* fucosylated glycoforms in serum ITIH4, and at site N577 the majority of fucosylated glycoforms contain outer-arm fucosylated glycans.

O-Glycopeptide Microheterogeneity in Recombinant ITIH4

We also characterized several O-glycopeptides in recombinant and serum-derived ITIH4. Recombinant ITIH4 is glycosylated on overlapping peptides IPKPEASFSPR (residues 634−644) and ASFSPR (residues 639−644) containing potentially glycosylated residues S640 and S642. We identified two major glycoforms of this peptide, including $HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1HexNAc_1NeuAc_2$, and several minor glycoforms (Table 3). To complement these analyses, we selectively detached O-glycans from recombinant ITIH4 via reductive beta-elimination and then performed MS and MS/MS analyses on detached, permethylated O-glycans (Supplementary Figure 1B, Supporting Information). We detected simple O-glycan structures, including $HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1HexNAc_1NeuAc_2$, equivalent to the glycan compositions on IPKPEASFSPR and ASFSPR. Therefore, we suggest that it is likely that only one of the potential O-glycosylation sites on this peptide (either S640 or S642) is glycosylated. We also identified a second glycopeptide, GPDVLTATVSGK, spanning residues 501−512, with multiple glycoforms (Table 3), including $HexNAc_1Hex_1NeuAc_2$ and $HexNAc_2HexNAc_2NeuAc_2$. Because we detect only simple O-glycans in our detached glycan analysis, and we detect glycopeptides with higher compositions on glycopeptide GPDVLTATVSGK, we suggest that at least two of the potentially O-glycosylated residues on this peptide (T506, T508, S510) are glycosylated with simple O-glycans. We also detected two fucosylated O-glycopeptides on GPDVLTATVSGK from recombinant ITIH4. We did not detect fucosylated O-glycans in our detached glycan analyses, suggesting that fucosylated O-glycans were below the limit of detection.

The recombinant ITIH4 protein sequence differs from the canonical sequence in UniProt (Q14624) at residues 85 (I to N) and 698 (P to T). We detect multiple glycoforms of peptide LAILPASA*T*PATSNPDPAVSR (residues 690−710). Nonsialylated O-glycopeptides including $HexNAc_1$, $HexNAc_2$, $HexNAc_1Hex_1$, $HexNAc_2Hex_2$, $HexNAc_3Hex_4$ and $HexNAc_4Hex_4$ account for over half of the glycoforms (based on relative intensity) of this peptide; we also detect compositions consistent with simple sialylated O-glycans including $HexNAc_1Hex_1NeuAc_2$, $HexNAc_3Hex_3NeuAc_3$, and $HexNAc_4Hex_4NeuAc_4$ (Table 3A). In addition, we detect a third (partially overlapping) glycopeptide, PAVSRVMNMK + $HexNAc_1HexNAc_1NeuAc_2$ in trypsin-chymotrypsin digests of recombinant ITIH4. This peptide contains a single potential glycosylation site at S709. We have identified multiple O-glycoforms on three different peptides in trypsin-GluC and trypsin-chymotrypsin digests of recombinant ITIH4.

O-Glycopeptide Microheterogeneity in Serum-Derived ITIH4

We also characterized O-glycopeptides in serum ITIH4 by CID MS/MS and detached O-glycan analyses. Glycopeptide IPKPEASFSPR + $HexNAc_1Hex_1NeuAc_1$, and the doubly sialylated $[IPKPEASFSPR + HexNAc_1Hex_1NeuAc_2 + 3H]^{3+}$ form at m/z 726.01 with one missed GluC cleavage (at glutamic acid) were observed in both serum and recombinant ITIH4. An alternately cleaved glycopeptide, ASFSPR, consistent with cleavage by GluC at the glutamic acid residue, is observed with the same glycan compositions, including $[ASFSPR + HexNAc_1NeuAc_1 + 2H]^{2+}$ m/z 579.78, $[ASFSPR + HexNAc_2Hex_2 + 2H]^{2+}$ at m/z 653.66, $[ASFSPR + HexNAc_1Hex_1NeuAc_1 + 2H]^{2+}$ at m/z 660.79, and $[ASFSPR + HexNAc_1Hex_1NeuAc_2 + 2H]^{2+}$ at m/z 806.36, as summarized in Table 3B. A CID fragmentation spectrum of $[ASFSPR + HexNAc_1Hex_1NeuAc_2 + 2H]^{2+}$ (m/z 806.36) in Figure 3B shows multiple glycopeptide Y ions with the intact peptide and a partially fragmented glycan attached to the peptide backbone. There are two serine residues in this peptide (S640, S642), each representing a potential site of glycan attachment. On the basis of CID fragmentation alone, we are unable to determine if glycosylation is restricted to a single site. However, the compositions of serum-derived ITIH4 detached O-glycans (Supplementary Figure 1C, Supporting Information) suggest that the peptide is glycosylated at a single site.

In ITIH4 purified from pooled human serum we detect glycoforms of both LAILPASA*P*PATSNPDPAVSR and LAILPASA*T*PATSNPDPAVSR, indicating that both the canonical ITIH4 sequence and the P698 to T variant are present in the sample. The glycan compositions observed on LAILPASA*P*PATSNPDPAVSR in serum-derived ITIH4 include $HexNAc_2Hex_2NeuAc_2$, $HexNAc_2Hex_2NeuAc_3$, $HexNAc_3Hex_3NeuAc_3$, $HexNAc_3Hex_3NeuAc_4$, and $HexNAc_4Hex_4NeuAc_4$ and are similar to those observed on the *T*698 variant peptide (Table 3). Despite the difference in sequence, the most common glycoforms ($HexNAc_2Hex_2NeuAc_2$ and $HexNAc_3Hex_3NeuAc_3$) associated with both peptides have similar compositions and relative intensities in serum ITIH4 (Table 3). Potential O-glycosylated residues include S696, T701, S702, and S709. We detected mostly simple O-glycan structures in detached O-glycan analyses of serum-derived ITIH4 (Supplementary Figure 1, Supporting Information). We also detected more complex O-glycans in detached O-glycan analyses, but these represented minor components of the mixture. The most likely explanation for these observations is that the observed glycan

compositions, including $HexNAc_3Hex_3NeuAc_3$, and $HexNAc_3Hex_3NeuAc_4$, represent the sum of multiple smaller O-glycans divided among three glycosylated residues in the peptide. In the glycopeptide CID MS/MS spectrum of [LAILPASAPPATSNPDPAVSR + $HexNAc_3Hex_3NeuAc_3$ + 4H]$^{4+}$ (m/z 1004.2) from serum-derived ITIH4 shown in Figure 3A the y_5 ion (m/z 529.31, 1$^+$) as well as the y_5 ion with HexNAc (m/z 732.39, 1$^+$) and HexNAc-Hex (m/z 894.44, 1$^+$) are clearly present. This strongly supports that one of the sites of glycosylation is S709, which is further corroborated by evidence of glycosylation on S709 in recombinant ITIH4. We also observe these ions (y_5 at m/z 529.31, 1$^+$; y_5 + HexNAc at m/z 732.39, 1$^+$; and y_5 + HexNAc-Hex at m/z 894.44, 1$^+$) in the P698 to T variant glycopeptides in serum and recombinant ITIH4, indicating that in all cases S709 is glycosylated.

O-Glycosylation Site Determination

To further explore the sites of O-glycosylation in serum-derived ITIH4, we used targeted electron-transfer dissociation (ETD) on ITIH4 glycopeptides. As opposed to most of our CID analyses, ETD of glycopeptides results in fragmentation of the peptide backbone rather than of glycosidic linkages. Peptide c- and z-ions resulting from ETD fragmentation of glycopeptides retain unfragmented carbohydrate moieties, which can be used to determine the site(s) of glycosylation. We labeled the serum-derived ITIH4 glycopeptides with mTRAQ reagents to boost the charge, treated them with a broad-specificity neuraminidase to remove sialic acids, and then subjected them to ETD (Supplementary Figure 4, Supporting Information). On the basis of the MS2 spectrum of glycopeptide IPKPEASFSPR, S640 is O-glycosylated, while S642 is not. This is supported by the presence of c_7, c_8, and c_9 and z_5, z_6, and z_7 ions with HexNAc-Hex in the fragmentation spectrum, as well as the absence of this modification on z_3, z_4, c_5, and c_6 ions.

ETD fragmentation of the desialylated, mTRAQ-labeled LAILPASAPPATSNPDPAVSR + $HexNAc_2Hex_2$ peptide from serum ITIH4 indicates that the selected glycopeptide is glycosylated at S696 and T701, based on c- and z-ions in the MS2 spectrum (Supplementary Figure 4, Supporting Information). We were unable to obtain information on the desialylated LAILPASAPPATSNPDPAVSR + $HexNAc_3Hex_3$ glycopeptide, so we are not able to corroborate the CID results that also indicate that S709 is O-glycosylated. On the basis of the available information, the most likely explanation is that LAILPASAPPATSNPDPAVSR is O-glycosylated at S696, T701 and S709. We detect similar glycan compositions associated with peptides LAILPASA*T*PATSNPDPAVSR and LAILPASA*P*PATSNPDPAVSR in serum ITIH4, suggesting that the sequence variant is not glycosylated differently.

DISCUSSION

ITIH4 is a secreted glycoprotein that is up-regulated by IL-6 as part of the acute phase response to turpentine-induced inflammation in rats and in response to infection in humans.[8,9,11] Glycosylation of ITIH4 is of considerable interest due to the biomarker potential of its proteolytic products and glycoforms in several forms of cancer, the involvement of ITIH4 in liver development and stabilization of the ECM, and observed changes in ITIH4 expression and glycosylation in ovarian and other cancers.[11,15−17,20] Glycosylation can mediate protein interactions, influence protein stability, and impact protein function. Our results confirm previous reports that ITIH4 is heavily glycosylated; it is therefore plausible that glycosylation modifies ITIH4 structure and function. While microheterogeneity of glycoforms differs between ITIH4 protein expressed in HEK293 cells and protein isolated from serum, occupancy of N-glycosylation sites does not differ. However, utilization of O-glycosylation sites differs between the cell line and serum.

Canonical ITIH4 N-Glycosylation Sites are Highly Occupied

We have determined occupancy of each glycosylation site quantitatively by treatment of the N-glycopeptides by PNGaseF in the presence of $H_2{}^{18}O$. Our results show that at least three of the four canonical N-glycosylation sites (N207, N517, and N577) are highly occupied >90% occupancy) in serum-derived ITIH4, and a similarly high degree of glycosylation was observed on the recombinant ITIH4. We also confirm that site N81 is glycosylated. N-Glycosylation site occupancy is thought to depend on the primary sequence of the protein because the transfer of precursor glycans from dolichol donors to asparagine residues occurs prior to protein folding.[48,49] The observed similarity in occupancy between recombinant and serum-derived ITIH4 is consistent with this hypothesis and suggests that cell culture might be a representative model for determination of site occupancy of protein glycoforms. However, variations in cell culture conditions, including changes in temperature, pH, and the availability of metal ions, reportedly affect N-glycosylation site occupancy.[50] It is interesting to note that the noncanonical glycosylation site at position N274 has low occupancy in both recombinant and serum-derived ITIH4.

Site-Specific Microheterogeneity of ITIH4 N-Glycoforms

Contrary to site occupancy, the microheterogeneity of N-glycoforms differs substantially between recombinant and serum-derived ITIH4. On the basis of CID MS/MS of untreated and exoglycosidase-treated glycopeptides and MALDI-TOF analysis of detached, permethylated N-glycans, we conclude that complex, core fucosylated, asialo-glycoforms are dominant in recombinant ITIH4. Fully sialylated forms represent a minor component of all N-glycoforms at all sites except N81. This is also the only canonical glycosylation site with detectable high mannose glycans. Core-fucosylated glycans at N81 represent a minor component of the total microheterogeneity. This is in contrast to the glycoforms at the three other canonical glycosylation sites (N207, N517, and N577) of recombinant ITIH4, where core-fucosylated forms dominate. We also observed more highly branched N-linked glycans, including tetra-antennary glycans, at N517 and N577.

Of the four canonical sites, we observed the highest microheterogeneity of glycoforms at N517. This may be due in part to the efficiency of our analytical workflow for this glycosylation site; for example, the LPTQNITFQTE peptide is the only proteolytic product we observe at the N517 N-glycosylation site, but we observe semispecific proteolytic cleavage (peptides AFINTF, KAFITNF) near the N81 site. The observation of multiple proteolytic products at site N81 may limit our ability to detect minor glycoforms. However, in trypsin-GluC digests we observe only one proteolytic product containing site N207, and in trypsin-chymotrypsin digests we only observe one proteolytic product containing site N577. Even at these sites we observe fewer glycoforms than at N517, which suggests that the site-specific differences in microheterogeneity are likely of biological or structural origin. However, we cannot rule out the possibility that differences in peptide ionization may also play a role in these observations.

Serum-derived ITIH4 N-linked glycopeptides demonstrate higher levels of sialylation, lower levels of fucosylation,

differences in fucose linkage, and less microheterogeneity compared to recombinant ITIH4. Complex, sialylated N-glycans are present at all four canonical N-glycosylation sites of serum-derived ITIH4. MALDI-analysis confirmed the predominance of sialylated N-linked glycans and also corroborated the presence of minor fucosylated glycoforms. On the basis of our site-specific characterization of glycopeptides from serum ITIH4, it appears that most fucosylated glycoforms are restricted to sites N517 and N577, although we also detected a triantennary fucosylated N-glycan at N81 after treatment with a broad-specificity neuraminidase. We also detected more branching (tri- and tetra-antennary glycoforms) at N517 and N577 than at N81 and N207, a pattern that is consistent with recombinant ITIH4. This is consistent with the fact that site specificity of N-glycan structures, including branching, depends on the structure of the mature protein.[51,52] Processing of N-linked glycans by exoglycosidases and glycosyltransferases, including extension and capping, take place primarily in the Golgi apparatus after the protein is already folded; access of glycosyltransferases to N-linked glycans is thought to be influenced by the protein structure.[53] Therefore, processes governed by protein structure should have similar outcomes in recombinant and serum-derived ITIH4 if the enzymes are present in active form. This is consistent with our findings of site-specific N-glycan branching and suggests that activity of neuraminidases is lower and activity of FUT8 is higher in the HEK293 compared to human liver, the major source of the serum derived ITIH4. This conclusion has to be viewed, however, with caution given many unknown factors including distribution and stability of the glycoforms *in vivo*.

Detection of Non-Canonical N-Glycosylation at N274

A small percentage of peptides (<1%) containing the non-canonical N-glycosylation site NVV at N274 were occupied in both serum and recombinant ITIH4. Noncanonical N-glycosylation has recently been described in immunoglobulins and a recent report also describes high-mannose glycoforms at noncanonical N-glycosylation sites on several murine proteins.[54−57] Using CID MS/MS, we found high-mannose N-glycopeptides on the NVVFVIDK peptide in both serum and cell-derived ITIH4. As we find only high-mannose N-glycans at this site, its accessibility to glycosidases and glycosyltransferases during protein maturation in the Golgi may be limited.[53] The N274 site is near the beginning of the von Willebrand factor type A (vWF) domain of ITIH4, which spans residues 272−432. Protein vWF domains are typically involved in protein−protein interactions, and glycosylation at this site could potentially affect ITIH4 interactions with other proteins.

Newly Described O-Glycopeptides of ITIH4

Both recombinant and serum ITIH4 are glycosylated on peptides IPKPEASFSPR and ASFSPR containing serine residues S640 and S642. Peptides LAILPASA*P*PATSNPDPAVSR and LAILPASA*T*PATSNPDPAVSR from two ITIH4 variants are glycosylated in serum and have similar associated glycan compositions. The presence of an additional threonine in the variant form does not impact the glycan compositions associated with the peptide. In recombinant ITIH4, LAILPASATPATSNPDPAVSR contains smaller and less sialylated glycan compositions compared to serum ITIH4. Peptide GPDVLTATVSGK is glycosylated exclusively in recombinant ITIH4. Initiation of O-GalNAc glycosylation through the transfer of GalNAc to serine or threonine residues is controlled by a family of at least 20 UDP-GalNAc:polypeptide GalNAc-transferases (GalNAc-Ts). Differences in glycosylation site- and tissue-specificity are expected to be more pronounced for O-GalNAc glycosylation compared to N-glycosylation. Our observation of site-specific differences in O-glycosylation between serum-derived and recombinant ITIH4 is consistent with this observation. We observe the same glycan compositions ($HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$) on peptides IPKPEASFSPR and ASFSPR in serum and recombinant ITIH4. In serum ITIH4, we detected glycosylation at S640 but not at S642. Glycosylation of this peptide in different contexts suggests that glycosylation at this site may play an important role in ITIH4 stability or function.

Serum ITIH4 peptide LAILPASAPPATSNPDPAVSR (residues 691−710) is O-glycosylated at three sites, on residues S696, T701, and S709, and we have characterized multiple glycoforms of this peptide. Our glycopeptide data indicate that this peptide contains glycan compositions as large as $HexNAc_3Hex_3NeuAc_4$. When we analyzed detached O-glycans in serum and recombinant ITIH4, we observed $HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$, and additional low-abundance glycoforms. Together, this suggests that S696, T701, and S709 are occupied with simple core-1 O-glycans. While we do not observe this peptide in recombinant ITIH4, we detect the PAVSRVMNMK (residues 706−715) + $HexNAc_1Hex_1NeuAc_2$ glycopeptide (containing S709) in trypsin-chymotrypsin digests of recombinant ITIH4. This suggests that S709 is glycosylated in both serum and recombinant ITIH4, while additional glycosylation occurs in serum ITIH4. We also observe multiple glycoforms of peptide GPDVLTATVSGK (with potentially glycosylated residues T506, T508, and S510) in recombinant ITIH4, but not in serum-derived ITIH4, and we detected two fucosylated O-glycopeptides on GPDVLTATVSGK from recombinant ITIH4 but did not detect any fucosylated O-glycopeptides in serum-derived ITIH4.

In addition to the glycopeptides discussed above, evidence of ITIH4 glycosylation on residues 719−725 has been reported in a study of glycopeptides from urinary protein digests.[31] However, we do not observe any glycopeptides at this site; this may be due to the different sources of ITIH4 or possibly reflect proteolytic resistance of the ITIH4 sequence. It is also possible that high levels of O-glycosylation in this region of ITIH4 may protect ITIH4 from proteolysis in its native environment but could also prevent us from observing glycopeptides if high-density glycosylation blocks access of trypsin, endoproteinase Glu-C, and chymotrypsin to proteolytic sites in this region. The ITIH4 sequence is atypical in that it contains regions without frequent sites for cleavage of trypsin, chymotrypsin, and other common proteases. This makes bottom-up analyses challenging and necessitates the use of multiple proteases for observation of different glycopeptides.

We detected several glycoforms of glycopeptide IPKPEASFSPR in serum and recombinant ITIH4, and we have confirmed that S640 is glycosylated in serum ITIH4. This sequence is specific to ITIH4 isoform 1 and therefore contributes to isoform-specific glycosylation that may influence ITIH4 stability or interaction with the ECM. On the basis of expression studies, ITIH4 is primarily synthesized in the liver,[1] and four isoforms of ITIH4 have been predicted based on mRNA and/or cDNA sequencing experiments.[12,58,59] Three ITIH4 isoforms have been described in adult liver tissue, while the fourth was found in fetal human liver. Isoform 1 was the first to be described and has been selected as the "canonical" sequence in UniProt. In this discussion, all references to amino acid residue locations are made in reference to ITIH4 isoform 1. Isoforms 2−4 are missing

small regions (all in the C-terminal half of the protein) compared to isoform 1. Isoform 2 also has a unique sequence "ACPSCSRSRAPAVPA" starting at residue 727. Specifically, isoforms 2−4 do not contain residues 621−650 of ITIH4 isoform 1, and intriguingly, our proteomic analyses of serum-derived ITIH4 show that the peptide IPKPEASFSPR (and ASFSPR with one missed GluC cleavage at E) from this region of ITIH4 is O-glycosylated. Indeed, glycoprotein interactions with the ECM are frequently modulated by the presence of O-glycans and isoform-specific O-glycosylation could have important implications for ECM stability. We have also selected ITIH4 isoform 1 for overexpression in HEK293 cells and confirm that this region is glycosylated in the cell line as well, and identical glycan compositions are present in serum and recombinant ITIH4.

In addition to detecting isoform-specific glycosylation, we also detected several O-glycopeptides that flank the proline-rich region of ITIH4, which may also have an impact on ITIH4 stability or interactions with other proteins, or the ECM. ITIH4 can undergo cleavage by plasma kallikrein between R688 and R689,[12] and may undergo additional proteolytic degradation after this initial cleavage based on detection of native peptides in serum and plasma.[20,21] ASFSPR (residues 639−644) and LAILPASA(P/T)PATSNPDPAVSR (residues 690−710), both O-glycopeptides described in our current study, flank the plasma kallikrein cleavage site and the potentially bioactive peptide. Due to the potential of glycosylation to impact proteolytic processing, these findings deserve further investigation.

CONCLUSION

In summary, we have used LC−MS/MS with CID and complementary techniques to characterize the site-specific glycosylation of serum and recombinant ITIH4. We described both N-glycopeptides, which were occupied by complex N-linked glycans, and singly and multiply sialylated O-glycopeptides, including previously unreported sites. Serum ITIH4 N-glycans were highly sialylated, while the major forms in recombinant ITIH4 were highly core-fucosylated and were not sialylated. Canonical N-glycosylation sites in serum and recombinant ITIH4 shared similar occupancy, site-specific branching, and degrees of microheterogeneity, and differed in the degree of core-fucosylation and sialylation. In addition, we characterized a noncanonical glycosylation site at N274 in serum and recombinant ITIH4 with the motif "NVV", and described ITIH4 isoform-specific O-glycosylation. We also observed differences in the sites of O-glycosylation between serum and recombinant ITIH4. The two serum ITIH4 variants contain heavily glycosylated peptides spanning residues 690−710. While this region is also glycosylated in recombinant ITIH4, glycan compositions associated with this peptide were smaller and less sialylated than the equivalent peptide in serum ITIH4. On the basis of our findings in serum ITIH4 the T698 variant does not affect the O-glycan composition of this peptide. We also determined that in LAILPASAPPATSNPDPAVSR is glycosylated on residues S696, T701, and S709. A second peptide spanning residues 634−644 is glycosylated in recombinant ITIH4 but not in serum-derived ITIH4. The function of these glycoforms is not known at present, but further study of ITIH4 glycosylation in the context of development and disease could offer important insight into the role that ITIH4 plays in liver development, ECM stability, and progression of diseases including ovarian cancer and liver disease. In this manuscript we characterize ITIH4 N-glycoforms and novel O-glycoforms; while microheterogeneity of glycoforms differs between ITIH4 protein expressed in HEK293 cells and protein isolated from serum, occupancy of N-glycosylation sites does not differ. However, utilization of O-glycosylation sites differs between the cell line and serum. This suggests that cell lines are likely a reasonable model for studies of site occupancy of N-glycosylation sites.

ASSOCIATED CONTENT

Supporting Information

Supplementary Figure 1: spectra of detached, N- and O-linked glycans from serum and recombinant derived ITIH4. Supplementary Figure 2: proteomic analyses and peptide identification of serum and recombinant ITIH4. Supplementary Figure 3: extracted ion chromatogram (XIC) plots of ITIH4 glycopeptides treated with PNGaseF/H2 ^{18}O for site occupancy determination. Supplementary Figure 4: electron transfer dissociation (ETD) spectra of de-sialylated, mTRAQ-labeled glycopeptides IPKPEASFSPR + HexNAcHex and LAILPASAPPATSNPDPAVSR + $HexNAc_2Hex_2$. Supplementary Figure 5: Table of all observed recombinant and serum-derived ITIH4 glycopeptides from CID MS/MS spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: rg26@georgetown.edu. Phone: 202-687-9869. Fax: 202-687-1988.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by U01 CA168926 and RO1 CA135069 awarded to R.G. and CCSG Grant NIH P30 CA51008 to the Lombardi Comprehensive Cancer Center supporting the Proteomics and Metabolomics Shared Resource. This work was also supported in part by P41 GM103490 (L.W. senior investigator). K.B.C. was supported, in part, by a Graduate Research Fellowship from the National Science Foundation. N.J.E. was supported, in part, by NIH/NCI/CPTI Grant CA126189.

ABBREVIATIONS

CID, collision-induced dissociation; ECM, extracellular matrix; ETD, electron transfer dissociation; FA, formic acid; Hex, hexose; Hex, *N*-acetylhexosamine; HILIC, hydrophilic interaction liquid chromatography; ITIH4, interalpha-trypsin inhibitor heavy chain H4; LC, liquid chromatography; MWCO, molecular weight cut off; PNGaseF, peptide-*N*-glycosidase F; PRR, proline-rich region; TBS, Tris-buffered saline; TFA, trifluoroacetic acid; vWF, von Willebrand Factor A domain

REFERENCES

(1) Tobe, T.; Saguchi, K.; Hashimoto, K.; Miura, N. H.; Tomita, M.; Li, F.; Wang, Y.; Minoshima, S.; Shimizu, N. Mapping of human inter-alpha-trypsin inhibitor family heavy chain-related protein gene (ITIHL1) to human chromosome 3p21 → p14. *Cytogenet. Cell Genet.* **1995**, *71*, 296−298.

(2) Salier, J. P.; Rouet, P.; Raguenez, G.; Daveau, M. The inter-alpha-inhibitor family: from structure to regulation. *Biochem. J.* **1996**, *315* (Pt 1), 1−9.

(3) Bost, F.; Diarra-Mehrpour, M.; Martin, J. P. Inter-alpha-trypsin inhibitor proteoglycan family–a group of proteins binding and stabilizing the extracellular matrix. *Eur. J. Biochem.* **1998**, *252*, 339–346.
(4) Chen, L.; Mao, S. J.; McLean, L. R.; Powers, R. W.; Larsen, W. J. Proteins of the inter-alpha-trypsin inhibitor family stabilize the cumulus extracellular matrix through their direct binding with hyaluronic acid. *J. Biol. Chem.* **1994**, *269*, 28282–28287.
(5) Saguchi, K.; Tobe, T.; Hashimoto, K.; Sano, Y.; Nakano, Y.; Miura, N. H.; Tomita, M. Cloning and characterization of cDNA for inter-alpha-trypsin inhibitor family heavy chain-related protein (IHRP), a novel human plasma glycoprotein. *J. Biochem.* **1995**, *117*, 14–18.
(6) Choi-Miura, N. H. Quantitative measurement of the novel human plasma protein, IHRP, by sandwich ELISA. *Biol. Pharm. Bull.* **2001**, *24*, 214–217.
(7) Gonzalez-Ramon, N.; Alava, M. A.; Sarsa, J. A.; Pineiro, M.; Escartin, A.; Garcia-Gil, A.; Lampreave, F.; Pineiro, A. The major acute phase serum protein in pigs is homologous to human plasma kallikrein sensitive PK-120. *FEBS Lett.* **1995**, *371*, 227–230.
(8) Daveau, M.; Jean, L.; Soury, E.; Olivier, E.; Masson, S.; Lyoumi, S.; Chan, P.; Hiron, M.; Lebreton, J. P.; Husson, A.; Jegou, S.; Vaudry, H.; Salier, J. P. Hepatic and extra-hepatic transcription of inter-alpha-inhibitor family genes under normal or acute inflammatory conditions in rat. *Arch. Biochem. Biophys.* **1998**, *350*, 315–323.
(9) Pineiro, M.; Alava, M. A.; Gonzalez-Ramon, N.; Osada, J.; Lasierra, P.; Larrad, L.; Pineiro, A.; Lampreave, F. ITIH4 serum concentration increases during acute-phase processes in human patients and is up-regulated by interleukin-6 in hepatocarcinoma HepG2 cells. *Biochem. Biophys. Res. Commun.* **1999**, *263*, 224–229.
(10) Harraghy, N.; Mitchell, T. J. Isolation and characterization of the promoter and partial enhancer region of the porcine inter-alpha-trypsin inhibitor heavy chain 4 gene. *Clin. Diagn. Lab. Immunol.* **2005**, *12*, 1336–1339.
(11) Bhanumathy, C. D.; Tang, Y.; Monga, S. P.; Katuri, V.; Cox, J. A.; Mishra, B.; Mishra, L. Itih-4, a serine protease inhibitor regulated in interleukin-6-dependent liver formation: role in liver development and regeneration. *Dev. Dyn.* **2002**, *223*, 59–69.
(12) Nishimura, H.; Kakizaki, I.; Muta, T.; Sasaki, N.; Pu, P. X.; Yamashita, T.; Nagasawa, S. cDNA and deduced amino acid sequence of human PK-120, a plasma kallikrein-sensitive glycoprotein. *FEBS Lett.* **1995**, *357*, 207–211.
(13) Pu, X. P.; Iwamoto, A.; Nishimura, H.; Nagasawa, S. Purification and characterization of a novel substrate for plasma kallikrein (PK-120) in human plasma. *Biochim. Biophys. Acta* **1994**, *1208*, 338–343.
(14) Choi-Miura, N. H.; Sano, Y.; Oda, E.; Nakano, Y.; Tobe, T.; Yanagishita, T.; Taniyama, M.; Katagiri, T.; Tomita, M. Purification and characterization of a novel glycoprotein which has significant homology to heavy chains of inter-alpha-trypsin inhibitor family from human plasma. *J. Biochem.* **1995**, *117*, 400–407.
(15) Abdullah-Soheimi, S. S.; Lim, B. K.; Hashim, O. H.; Shuib, A. S. Patients with ovarian carcinoma excrete different altered levels of urine CD59, kininogen-1 and fragments of inter-alpha-trypsin inhibitor heavy chain H4 and albumin. *Proteome Sci.* **2010**, *8*, 58.
(16) Zhang, Z.; Bast, R. C., Jr.; Yu, Y.; Li, J.; Sokoll, L. J.; Rai, A. J.; Rosenzweig, J. M.; Cameron, B.; Wang, Y. Y.; Meng, X. Y.; Berchuck, A.; Van Haaften-Day, C.; Hacker, N. F.; de Bruijn, H. W.; van der Zee, A. G.; Jacobs, I. J.; Fung, E. T.; Chan, D. W. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **2004**, *64*, 5882–5890.
(17) Heo, S. H.; Lee, S. J.; Ryoo, H. M.; Park, J. Y.; Cho, J. Y. Identification of putative serum glycoprotein biomarkers for human lung adenocarcinoma by multilectin affinity chromatography and LC-MS/MS. *Proteomics* **2007**, *7*, 4292–4302.
(18) Yang, L.; Rudser, K. D.; Higgins, L.; Rosen, H. R.; Zaman, A.; Corless, C. L.; David, L.; Gourley, G. R. Novel biomarker candidates to predict hepatic fibrosis in hepatitis C identified by serum proteomics. *Dig. Dis. Sci.* **2011**, *56*, 3305–3315.
(19) Gangadharan, B.; Antrobus, R.; Dwek, R. A.; Zitzmann, N. Novel serum biomarker candidates for liver fibrosis in hepatitis C patients. *Clin. Chem.* **2007**, *53*, 1792–1799.
(20) Fung, E. T.; Yip, T. T.; Lomas, L.; Wang, Z.; Yip, C.; Meng, X. Y.; Lin, S.; Zhang, F.; Zhang, Z.; Chan, D. W.; Weinberger, S. R. Classification of cancer types by measuring variants of host response proteins using SELDI serum assays. *Int. J. Cancer* **2005**, *115*, 783–789.
(21) van den Broek, I.; Sparidans, R. W.; Schellens, J. H.; Beijnen, J. H. Liquid chromatography/tandem mass spectrometric method for the quantification of eight proteolytic fragments of ITIH4 with biomarker potential in human plasma and serum. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 2915–2928.
(22) van den Broek, I.; Sparidans, R. W.; van Winden, A. W.; Gast, M. C.; van Dulken, E. J.; Schellens, J. H.; Beijnen, J. H. The absolute quantification of eight inter-alpha-trypsin inhibitor heavy chain 4 (ITIH4)-derived peptides in serum from breast cancer patients. *Proteomics Clin. Appl.* **2010**, *4*, 931–939.
(23) Ashwell, G.; Morell, A. G. The role of surface carbohydrates in the hepatic recognition and transport of circulating glycoproteins. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1974**, *41* (0), 99–128.
(24) Ciechanover, A.; Schwartz, A. L.; Lodish, H. F. Sorting and recycling of cell surface receptors and endocytosed ligands: the asialoglycoprotein and tranferrin receptors. *J. Cell Biochem.* **1983**, *23* (1–4), 107–30.
(25) Barthel, S. R.; Wiese, G. K.; Cho, J.; Opperman, M. J.; Hays, D. L.; Siddiqui, J.; Pienta, K. J.; Furie, B.; Dimitroff, C. J. Alpha 1,3 fucosyltransferases are master regulators of prostate cancer cell trafficking. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (46), 1949–6.
(26) Bartolazzi, A.; Nocks, A.; Aruffo, A.; Spring, F.; Stamenkovic, I. Glycosylation of CD44 is implicated in CD44-mediated cell adhesion to hyaluronan. *J. Cell Biol.* **1996**, *132* (6), 1199–208.
(27) Russell, D.; Oldham, N. J.; Davis, B. G. Site-selective chemical protein glycosylation protects from autolysis and proteolytic degradation. *Carbohydr. Res.* **2009**, *344*, 1508–1514.
(28) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **2003**, *21*, 660–666.
(29) Liu, T.; Qian, W. J.; Gritsenko, M. A.; Camp, D. G.; Monroe, M. E.; Moore, R. J.; Smith, R. D. Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J. Proteome. Res.* **2005**, *4*, 2070–2080.
(30) Chen, R.; Jiang, X.; Sun, D.; Han, G.; Wang, F.; Ye, M.; Wang, L.; Zou, H. Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J. Proteome. Res.* **2009**, *8*, 651–661.
(31) Bunkenborg, J.; Pilch, B. J.; Podtelejnikov, A. V.; Wisniewski, J. R. Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* **2004**, *4*, 454–465.
(32) Halim, A.; Nilsson, J.; Ruetschi, U.; Hesse, C.; Larson, G. Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol. Cell Proteomics* **2012**, *11*, M111.
(33) Chandler, K. B.; Pompach, P.; Goldman, R.; Edwards, N. Exploring Site-Specific N-Glycosylation Microheterogeneity of Haptoglobin Using Glycopeptide CID Tandem Mass Spectra and Glycan Database Search. *J. Proteome. Res.* **2013**, *12*, 3652–3666.
(34) Sanda, M.; Pompach, P.; Brnakova, Z.; Wu, J.; Makambi, K.; Goldman, R. Quantitative liquid chromatography-mass spectrometry-multiple reaction monitoring (LC-MS-MRM) analysis of site-specific glycoforms of haptoglobin in liver disease. *Mol. Cell Proteomics* **2013**, *12*, 1294–1305.
(35) Pompach, P.; Brnakova, Z.; Sanda, M.; Wu, J.; Edwards, N.; Goldman, R. Site-specific glycoforms of haptoglobin in liver cirrhosis and hepatocellular carcinoma. *Mol. Cell Proteomics* **2013**, *12*, 1281–1293.
(36) Shilov, I. V.; Seymour, S. L.; Patel, A. A.; Loboda, A.; Tang, W. H.; Keating, S. P.; Hunter, C. L.; Nuwaysir, L. M.; Schaeffer, D. A. The Paragon Algorithm, a next generation search engine that uses sequence

temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell Proteomics* **2007**, *6*, 1638−1655.
(37) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics.* **2008**, *24*, 2534−2536.
(38) Pompach, P.; Chandler, K. B.; Lan, R.; Edwards, N.; Goldman, R. Semi-automated identification of N-Glycopeptides by hydrophilic interaction chromatography, nano-reverse-phase LC-MS/MS, and glycan database search. *J. Proteome. Res.* **2012**, *11*, 1728−1740.
(39) Ranzinger, R.; Herget, S.; Wetter, T.; von der Lieth, C. W. GlycomeDB - integration of open-access carbohydrate structure databases. *BMC. Bioinformatics.* **2008**, *9*, 384.
(40) Ranzinger, R.; Frank, M.; von der Lieth, C. W.; Herget, S. Glycome-DB.org: a portal for querying across the digital world of carbohydrate sequences. *Glycobiology* **2009**, *19*, 1563−1567.
(41) Kuster, B.; Wheeler, S. F.; Hunter, A. P.; Dwek, R. A.; Harvey, D. J. Sequencing of N-linked oligosaccharides directly from protein gels: in-gel deglycosylation followed by matrix-assisted laser desorption/ionization mass spectrometry and normal-phase high-performance liquid chromatography. *Anal. Biochem.* **1997**, *250*, 82−101.
(42) Kang, P.; Mechref, Y.; Novotny, M. V. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 721−734.
(43) Bekesova, S.; Kosti, O.; Chandler, K. B.; Wu, J.; Madej, H. L.; Brown, K. C.; Simonyan, V.; Goldman, R. N-glycans in liver-secreted and immunoglogulin-derived protein fractions. *J. Proteomics* **2012**, *75*, 2216−2224.
(44) Stalnaker, S. H.; Hashmi, S.; Lim, J. M.; Aoki, K.; Porterfield, M.; Gutierrez-Sanchez, G.; Wheeler, J.; Ervasti, J. M.; Bergmann, C.; Tiemeyer, M.; Wells, L. Site mapping and characterization of O-glycan structures on alpha-dystroglycan isolated from rabbit skeletal muscle. *J. Biol. Chem.* **2010**, *285*, 24882−24891.
(45) Liu, Z.; Cao, J.; He, Y.; Qiao, L.; Xu, C.; Lu, H.; Yang, P. Tandem 18O stable isotope labeling for quantification of N-glycoproteome. *J. Proteome. Res.* **2010**, *9*, 227−236.
(46) Wang, D.; Hincapie, M.; Rejtar, T.; Karger, B. L. Ultrasensitive characterization of site-specific glycosylation of affinity-purified haptoglobin from lung cancer patient plasma using 10 mum i.d. porous layer open tubular liquid chromatography-linear ion trap collision-induced dissociation/electron transfer dissociation mass spectrometry. *Anal. Chem.* **2011**, *83*, 2029−2037.
(47) Wuhrer, M.; Koeleman, C. A.; Hokke, C. H.; Deelder, A. M. Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 1747−1754.
(48) Kasturi, L.; Eshleman, J. R.; Wunner, W. H.; Shakin-Eshleman, S. H. The hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence N-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. *J. Biol. Chem.* **1995**, *270*, 14756−14761.
(49) Shakin-Eshleman, S. H.; Spitalnik, S. L.; Kasturi, L. The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. *J. Biol. Chem.* **1996**, *271*, 6363−6366.
(50) Gawlitzek, M.; Estacio, M.; Furch, T.; Kiss, R. Identification of cell culture conditions to control N-glycosylation site-occupancy of recombinant glycoproteins expressed in CHO cells. *Biotechnol. Bioeng.* **2009**, *103*, 1164−1175.
(51) Kornfeld, R.; Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **1985**, *54*, 631−664.
(52) Weisshaar, G.; Hiyama, J.; Renwick, A. G. Site-specific N-glycosylation of human chorionic gonadotrophin–structural analysis of glycopeptides by one- and two-dimensional 1H NMR spectroscopy. *Glycobiology* **1991**, *1*, 393−404.
(53) Thaysen-Andersen, M.; Packer, N. H. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology* **2012**, *22*, 1440−1452.
(54) Valliere-Douglass, J. F.; Kodama, P.; Mujacic, M.; Brady, L. J.; Wang, W.; Wallace, A.; Yan, B.; Reddy, P.; Treuheit, M. J.; Balland, A. Asparagine-linked oligosaccharides present on a non-consensus amino acid sequence in the CH1 domain of human antibodies. *J. Biol. Chem.* **2009**, *284* (47), 32493−506.
(55) Valliere-Douglass, J. F.; Eakin, C. M.; Wallace, A.; Ketchem, R. R.; Wang, W.; Treuheit, M. J.; Balland, A. Glutamine-linked and non-consensus asparagine-linked oligosaccharides present in human recombinant antibodies define novel protein glycosylation motifs. *J. Biol. Chem.* **2010**, *285* (21), 16012−22.
(56) Trinidad, J. C.; Schoepfer, R.; Burlingame, A. L.; Medzihradszky, K. F. N- and O-glycosylation in the murine synaptosome. *Mol. Cell. Proteomics* **2013**, *12*, 3474−88.
(57) 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*, 897−907.
(58) Kroczynska, B.; King-Simmons, L.; Alloza, L.; Alava, M. A.; Elguindi, E. C.; Blond, S. Y. BIP co-chaperone MTJ1/ERDJ1 interacts with inter-alpha-trypsin inhibitor heavy chain 4. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 1467−1477.
(59) Ota, T.; et al. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat. Genet.* **2004**, *36*, 40−45.