Journal of proteome research

Article

pubs.acs.org/jpr

Glycosylation Characterization of an Influenza H5N7 Hemagglutinin Series with Engineered Glycosylation Patterns: Implications for Structure−Function Relationships

Lisa M. Parsons,[†] Yanming An,[†] Robert P. de Vries,[‡] Cornelis A. M. de Haan,[§] and John F. Cipollo*,[†]

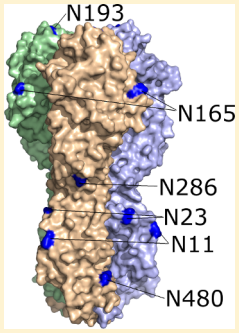
[†]Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland 20993, United States

[‡]Department of Medicinal Chemistry and Chemical Biology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584CG Utrecht, The Netherlands

[§]Virology Division, Department of Infectious Diseases & Immunology, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584CL Utrecht, The Netherlands

Supporting Information

ABSTRACT: The glycosylation patterns of four recombinant H5 hemagglutinins (HAs) derived from A/Mallard/Denmark/64650/03 (H5N7) have been characterized. The proteins were expressed in (i) HEK293T cells to produce complex glycoforms, (ii) HEK293T cells treated with *Vibrio cholera* neuraminidase to provide asialo-complex glycoforms, (iii) HEK293S GnTI(−) cells with predominantly the canonical $Man_5GlcNAc_2$ glycoform, and (iv) *Drosophila* S2 insect cells producing primarily paucimannose glycoforms. Previously, these HAs were used to investigate the effect of different glycosylation states on the immune responses in chicken and mouse systems. Evidence was found that high-mannose glycans diminished antibody response via DC-SIGN interactions. We performed two semiquantitative analyses including MALDI-TOF MS permethylation analysis of released glycans and LC−MS^E analysis of glycosylation site microheterogeneity. Glycosylation site occupancy was also determined by LC−MS^E. Our major findings include (1) decreasing complexity of glycosylation from the stem to the globular head, (2) absence of glycosylation at N^{10} and N^{193}, (3) complex glycans at N^{165} in HEK293T cell HA but high mannose glycans at this site in HEK293S and S2 cells, and (4) differences between the three-dimensional structures of H3 and H5 HAs that may explain glycan type preferences at selected sites. Biological implications of the findings are discussed.



KEYWORDS: *hemagglutinin, influenza, flu, mass spectrometry, GLYMPS, glycosylation, HEK293, S2*

INTRODUCTION

Influenza A virus is responsible for an estimated three to five million cases of serious illness and 250 000 to 300 000 deaths each year.[1] New influenza strains enter the human population through a zoonotic host. Typically, influenza that infects birds cannot infect humans, in part because of species-specific differences between host cell receptors.[2] Occasionally a virus acquires enough of the necessary mutations to interact with a new species.[3]

The hemagglutinin (HA) protein is responsible for initial binding of the virion to host-cell receptors containing primarily α2−3-linked (in birds) or α2−6-linked sialic acid (in humans). After receptor binding, HA mediates fusion between viral and host membranes and is a major target of the host's immune system[4] that is sufficient to provide protective immunity against influenza A virus (IAV) infection.[5] HA is an elongated trimeric protein that protrudes from the viral membrane via transmembrane helix regions and has a thick stem extending to a wider globular head domain. It is cleaved post-translationally by host proteases into two cysteine-linked peptides, HA1 and HA2. Glycosylation can modulate cleavage into HA1 and HA2, influence infectivity, aid in proper folding and stability, impact receptor specificity, and affect both innate and humoral immune responses of the host.[6]

In recent work, we studied H5 HAs derived from the strain A/Mallard/Denmark/64650/03 produced to bear different glycosylation patterns and examined the differences in immunologic effects in chicken and mouse systems.[7] Here we performed glycosylation analysis on four of those strains to better define their glycosylation states in order to allow a better understanding of how different glycosylation states may lead to alternative biological consequences. The H5 HAs examined include (i) those carrying complex glycans produced in HEK293T cells; (ii) those carrying complex glycans produced in HEK293T cells followed by treatment with *Vibrio cholera* neuraminidase (VCNA); (iii) those containing $Man_5GlcNAc_2$ moieties derived from HEK293S GnTI(−) cells that did not contain *N*-acetylglucosamine transferase I (GnTI)[8] and, (iv) *Drosophila* S2 insect cell-produced HA proteins carrying paucimannosidic *N*-glycans. We used semiquantitative permethylation analysis to reveal the overall distribution of glycans released from the HAs. We also developed, qualified, and used a

Received: February 26, 2016
Published: November 8, 2016

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

nanoLC−MS^E method for semiquantitation of glycan distributions at each glycosylation site. The resultant data were mapped to the modeled three-dimensional surface of H5 to reveal the positions of HA-specific glycosylation patterns and antigenic sites. Finally, we compared existing crystal structures of H5 and H3 HAs to examine differences that can explain site-specific glycan subtype preferences, which may influence interactions with the host immune system.

■ EXPERIMENTAL PROCEDURES

Chemicals and Reagents

RapiGest surfactant and Sep-Pak C18 cartridges were purchased from Waters Corporation (Milford, MA). Porous graphite carbon (PGC) cartridges were obtained from Thermo Fisher Scientific, Inc. (Waltham, MA). Sequencing grade modified trypsin was obtained from Promega Corp. (Madison, WI). Peptide *N*-Glycosidase A (PNGase A) was purchased from Roche Diagnostics Corporation (Indianapolis, IN). TSKgel Amide-80 was obtained from Tosoh Bioscience LLC (Montgomeryville, PA). Iodomethane, dimethyl sulfoxide (DMSO), sodium hydroxide pellets, and other chemicals were obtained from Sigma-Aldrich (St. Louis, MO). Water used in mass spectrometry applications was Optima LC−MS grade. All of the other solvents were HPLC grade, and all of the other reagents were ACS grade or higher.

Hemagglutinin

Four samples of purified recombinant influenza HA from A/Mallard/Denmark/64650/03 (H5N7) prepared as described by de Vries et al.[7] were obtained from the laboratory of Dr. C. A. M. de Haan. The samples consisted of HEK293T-grown HA with and without treatment with VCNA; HA grown in HEK293S GnTI(−) cells, which lack a functional *N*-acetylglucosamine transferase I (GnTI); and HA grown in *Drosophila* S2 cells. Table 1 contains a list of the origins, abbreviations, and glycans expected for each sample.

Table 1. Influenza HAs Examined in This Study

abbreviation	origin	expected *N*-glycans
HEK293T	HEK293T cells	complex glycans with sialic acid
HEK293T+VCNA	VCNA-treated HA protein from HEK293T cells	complex glycans without sialic acid
HEK293S	HEK293S GnTI(−) cells	$Man_5GlcNAc_2$
S2	*Drosophila* S2 cells	paucimannose

Glycopeptide Preparation

Sample purity, glycosylation state, and concentration of glycoproteins were estimated by SDS-PAGE. Each HA sample (100 μg) was dried and resuspended in 100 μL of 50 mM ammonium bicarbonate with 0.1% RapiGest and 5 mM dithiothreitol (DTT). The samples were incubated at 60 °C for 30 min and cooled to room temperature. Iodoacetamide was added to a final concentration of 15 mM, and the samples were placed in the dark for 30 min at room temperature. Trypsin was added at an enzyme:protein ratio of 1:20 (w/w), and the samples were incubated at 37 °C for 18 h. To degrade the RapiGest, trifluoroacetic acid (TFA) (99.9% purity) was added in sufficient quantity (generally 0.5 μL) to bring the pH below 2. The samples were incubated for 45 min at 37 °C and then centrifuged for 20 min at 13 000 rpm to remove the insoluble materials. The remaining supernatants, containing a mixture of peptides and glycopeptides, were vacuum-dried for further analysis.

Enrichment of Glycopeptides

Intact glycopeptides were enriched from the peptide/glycopeptide mixture by solid-phase extraction (SPE) using TSKgel Amide 80 hydrophilic interaction liquid chromatography (HILIC) resin as described previously.[9] Briefly, 200 mg of Amide-80 resin (400 μL of wet resin) was placed in a Supelco fritted 1 mL column, washed with 1 mL of 0.1% TFA in water, and conditioned with 1 mL of 80% acetonitrile (ACN)/0.1% TFA/water. Dried trypsin-processed samples were resuspended in 50 μL of 75% ACN/0.1% TFA/water, applied to the resin, and allowed to absorb for several minutes. The columns were then washed three times with 1 mL of the conditioning buffer. Glycopeptides were eluted with 1 mL of 60% ACN containing 0.1% TFA followed by 1 mL of 40% ACN/0.1% TFA/water. The eluents were combined, dried by vacuum centrifugation, and resuspended in 25 μL of 0.1% TFA/water for analysis by nanoLC−MS^E.

N-Glycan Release

PNGase A was used to enzymatically remove glycans from the peptide/glycopeptide mixture to obtain nonglycosylated and deglycosylated peptides and free glycans. Dried, trypsinized samples were resuspended in 50 mM ammonium acetate (pH 5.0). The pH of the prepared samples was tested with pH strips to verify that the pH was between 5.0 and 5.5, and the volume was adjusted to 50 μL accordingly. The samples were boiled for 10 min to destroy any residual trypsin activity and then cooled, and 0.1 milliunit of PNGase A was added. After incubation for 3 h at 37 °C, an additional 0.05 milliunit of PNGase A was added, and the samples were incubated for an additional 13 h at 37 °C.

Purification of Deglycosylated Peptides and Free *N*-Glycans

PNGase A-released glycans were enriched using C18 SPE cartridges preconditioned with 2 mL of ethanol followed by 2 mL of water. After the samples were applied, the columns were washed five times with 1 mL of water, and then the peptides were eluted with 2 mL of 100% isopropanol. The load and wash fractions, containing free oligosaccharides, were pooled and further purified over PGC columns to obtain free glycans as described below. To obtain the nonglycosylated and deglycosylated peptides, each eluate fraction was dried in a SpeedVac concentrator and resuspended in 25 μL of 0.1% TFA in water for analysis by nanoLC−MS^E without further processing.

To prepare free glycans, 1-butanol was added to the pooled load and wash fractions to a final concentration of 1% (v/v). The samples were loaded onto 100 mg PGC columns prepared first by sequential washes with 1 mL of 100% ACN, 1 mL of 60% ACN in water, 1 mL of 30% ACN in water, and 1 mL of water. All of the solutions contained 0.1% TFA. The loaded columns were washed three times with 1 mL of 0.1% TFA in water and then eluted with 30% ACN/0.1% TFA/water followed by 60% ACN/0.1% TFA/water. The eluents were pooled and dried in glass vials by rotary evaporation.

Permethylation of Free *N*-Glycans

Samples were permethylated following the protocol of Ciucanu and Costello[10] with modifications as follows. Twenty sodium hydroxide pellets were ground by mortar and pestle to a fine powder. The powder was poured into a 20 mL glass vial and

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

vortexed with 5 mL of DMSO. Each dried sample was resuspended in 100 μL of the NaOH/DMSO slurry and mixed with 50 μL of iodomethane (ICH_3). The mixtures were shaken at 300 rpm at room temperature for 1 h, after which 100 μL of the NaOH/DMSO slurry and 50 μL of iodomethane were added, and the reactions were allowed to proceed for an additional hour. This process was repeated once more for a total of three times. To neutralize the samples, 400 μL of water and then 400 μL of chloroform were added, followed by vortexing for 30 s. The samples were centrifuged at 2000*g* for 2 min, and the top aqueous fraction was discarded. Addition of 400 μL of water, centrifugation, and removal of the top layer was repeated five times or until the pH of the aqueous layer was neutral as indicated by pH paper. To dry the sample and remove weakly volatile side products, the samples were placed in a rotary evaporation device for 2 h at room temperature.

MALDI-TOF Analysis of Permethylated *N*-Glycans

A MALDI plate was prespotted with 1 μL of a matrix solution (20 mg/mL 2,5-dihydroxybenzoic acid (DHB) in 20% ACN, 1 mM sodium acetate, 80% water) and dried. Permethylated *N*-glycans were suspended in 20% acetonitrile/water and mixed 1:1 (v/v) with matrix solution on top of the dried matrix, and the plate was dried again. Samples were analyzed with a Perseptive Biosystems Voyager DE RF MALDI-TOF mass spectrometer in positive ion reflectron mode. Three spots were made per sample. Data were collected from scans of 30 laser shots and summed for 300 scans per spot. Maltooligosaccharide was used as an external calibrant. The MS data were processed using DataExplorer (Perseptive Biosystems).

Reversed-Phase NanoLC−MS^E Analysis of Glycopeptides and Peptides

Each prepared glycopeptide or deglycosylated/nonglycosylated peptide sample was injected three separate times onto a C18 column (BEH nanocolumn, 100 μm i.d. × 100 mm, 1.7 μm particles, Waters Corporation) for nanoLC−MS^E analysis. A Waters nanoAcquity UPLC system was used for automatic sample loading and flow control. Samples were loaded onto the column prepared with 3% ACN/97% water and, after a 1 min load, eluted with a 47 min gradient from 3 to 50% ACN at a flow rate of 0.4 μL/min. All of the chromatography solutions included 0.1% formic acid. The eluent flowed to an uncoated 20 μm i.d. PicoTip Emitter (New Objective Inc., Woburn, MA). The mass spectrometer was a Waters SYNAPT G2 HDMS system (Waters Corp. Milford, MA). The applied source voltage was 3000 V. Data were collected in positive polarity mode using data-independent MS^E acquisition, which consists of a starting 4 V scan followed by a scan ramping from 15 to 45 V in 0.9 s. To calibrate internally, every 30 s, 200 fmol/μL glufibrinopeptide B in 50% methanol/1% acetic acid/water was injected through the lockmass channel at a flow rate of 500 nL/min. Initial calibration of the mass spectrometer was performed using a 5 mM sodium formate infusion and tuned for a minimum resolution of 20 000 fwhm.

Data Analysis for Peptide and Glycopeptide Identification

NanoLC−MS^E data were processed using BiopharmaLynx 1.3x (Waters) and GLYMPS (in-house software) to identify specific glycans on each peptide. A database was created for BiopharmaLynx based on glycans previously identified on HAs grown in HEK293T cells and on glycans known to exist in HEK293T cells. The search settings included trypsin digest with up to 1 missed cleavage, fixed cysteine carbamidomethylation, variable methionine oxidation, and variable *N*-glycan modifications based on the library. BiopharmaLynx searches for the peptide plus the parent mass but does not attempt to assign or verify the parent assignment with glycopeptide fragments. In addition, because each potential glycan must be entered manually along with the mass and linkage information, it is possible that a glycan will not be assigned either because of database error or because it was not present in the library. To address these issues, we created a program called GLYMPS, which uses the processed data from BiopharmaLynx to assign the glycopeptide identities. The peptide sequence and the mode of protein hydrolysis is supplied to GLYMPS to provide the software with a list of potential peptides. The process starts with the core fragments such as peptide + GlcNAc, then each larger fragment is assigned in stepwise fashion until the parent mass is reached. Peptide fragments (with and without carbohydrate modification) and oxonium ions are also assigned. Output includes a labeled spectrum. GLYMPS was used to verify and correct the assignments made by BiopharmaLynx. The mass accuracy was set to 35 ppm in GLYMPS. Assignments were based on (1) the presence of a core fragment (peptide, peptide + HexNAc, peptide + $HexNAc_2$, peptide + $dHex_1HexNAc_1$, or peptide + $Hex_1HexNAc_2$), (2) the presence of three or more peptide fragments, (3) the presence of three or more assigned glycopeptide fragments; (4) the requirement that the assignment is made in at least two out of three injections, and (5) the existence of the glycan in GlycoSuiteDB.

Structural Comparison of H5 and H3 Hemagglutinin

A search of the flu database (www.fludb.org) for nonduplicated H5 or H3 sequences in complete segments returned 2880 H5 sequences and 6726 H3 sequences. The H3 sequences were further constrained to only those with tryptophan at position 222, giving 2984 sequences. Consensus sequences for each group were calculated from alignments of all of the sequences using tools in the flu database. Alignment between the H5 and H3 consensus sequences was performed with BLAST.[11]

Van der Waals and hydrogen bond comparisons were made on existing crystal structures in the Protein Data Bank (PDB) (www.pdb.org). The structures were chosen on the basis of sequence homology to the H5 consensus sequence and sequence homology to the 1968 H3 220-loop. They were further selected on the basis of a resolution of 2.5 Å or better and no ligand or antibody near the head region. MakeMultimer[12] was used to generate the trimer structures from the BIOMT transformation matrices in the PDB files as needed. To calculate hydrogen bonds and identify conserved residues at the glycan−protein interface, each structure was analyzed using LigPlot.[13] To more easily tally these interactions, an in-house Python script was used to measure and count the van der Waals distances between the glycans and the protein.

■ RESULTS

The glycosylation patterns of the four different hemagglutinin preparations were characterized as follows: (1) semiquantitative permethylation profiling analysis of free glycans performed using MALDI-TOF MS, (2) site occupancy analysis performed using nanoLC−MS^E, and (3) semiquantitative site-specific glycoform identification by nanoLC−MS^E. The workflow diagram is shown in Figure 1. HA samples were reduced, alkylated with iodoacetamide, and trypsinized according to standard procedures (see Experimental Procedures). Amide 80

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

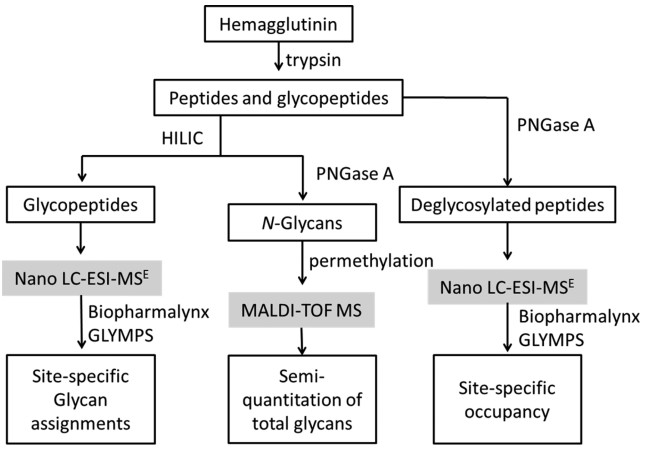


Figure 1. Workflow diagram of the glycoprotein analysis approach used in this study.

SPE-enriched glycopeptide samples were used for analyses 1 and 2 above. Our enrichment process has been qualified to be nonbiased, as previously reported.[9] Glycopeptide nanoLC–MS^E data were processed using BiopharmaLynx and GLYMPS software. Permethylation profile data were processed by GlycoMod or in-house developed software. I-Tasser[14] was used to model the three-dimensional representations of HAs, and Glycam[15] was used to model the three-dimensional representations of HAs bearing select glycans modeled to the protein surface. Although MS glycan data do not differentiate between diastereomers, the glycosylation repertoire of the cell lines used here is well-known. Therefore, where possible, compositions are inferred from this point forward.

It should be noted that released glycan and glycopeptide analyses examine very different chemical entity classes, one being released and permethylated glycans and the other being protein-linked glycans. The ionization methods used are also different, and the latter utilizes LC separation prior to detection. Complete correlation cannot be expected. As

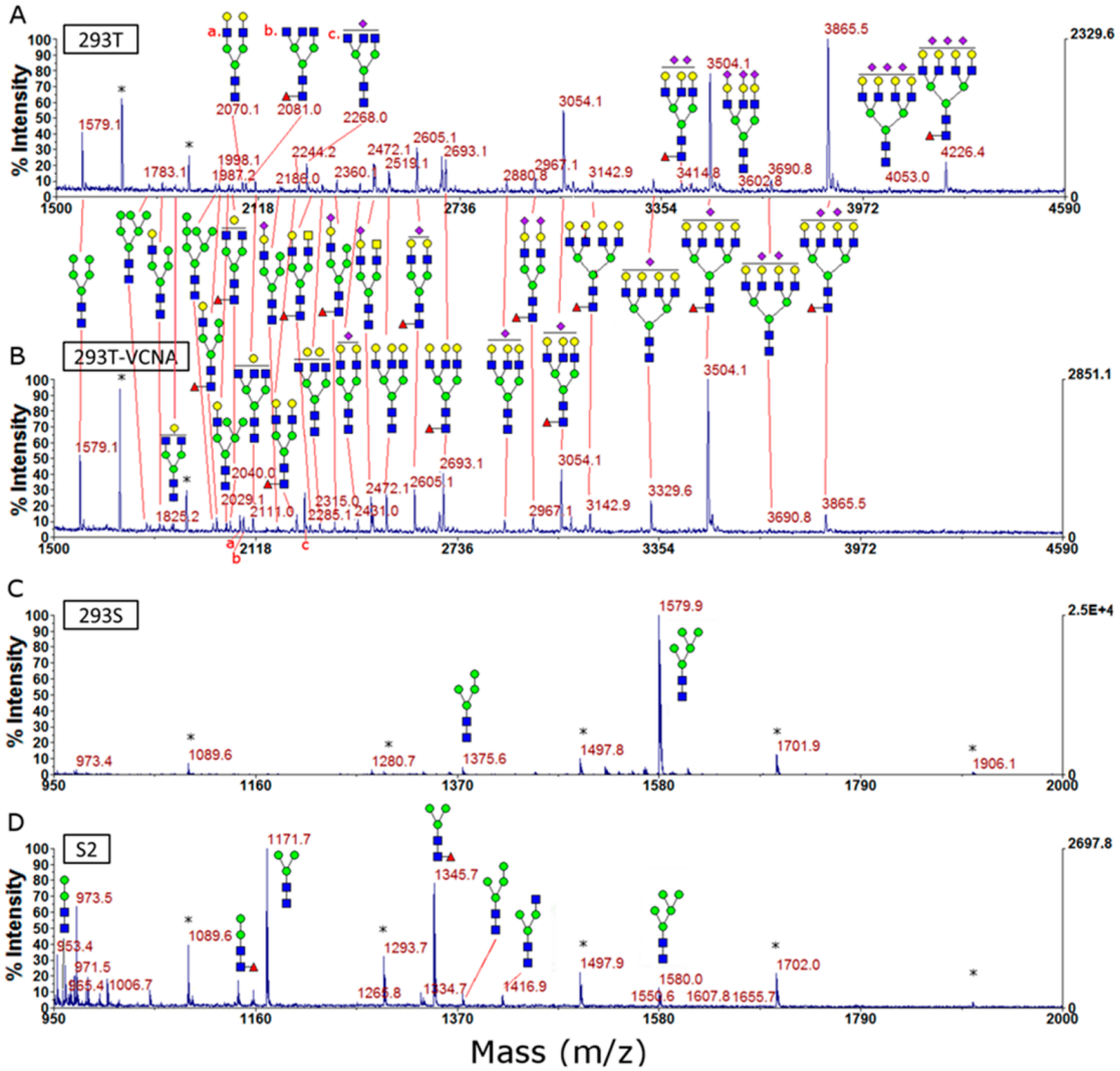


Figure 2. MALDI-TOF analysis of free glycans for samples (A) HEK293T, (B) HEK293T+VCNA, (C) HEK293S, and (D) S2. Maltohexose contaminants are marked with asterisks. Glycan structures are based on previously published structures in the GlycoSuite database. Blue squares, *N*-acetylglucosamine; green circles, mannose; yellow circles, galactose; red triangles, fucose; purple diamonds, sialic acid. Because of spectral crowding, small red letters (a, b, and c) have been used to label three of the glycans that appear in both B and A.

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

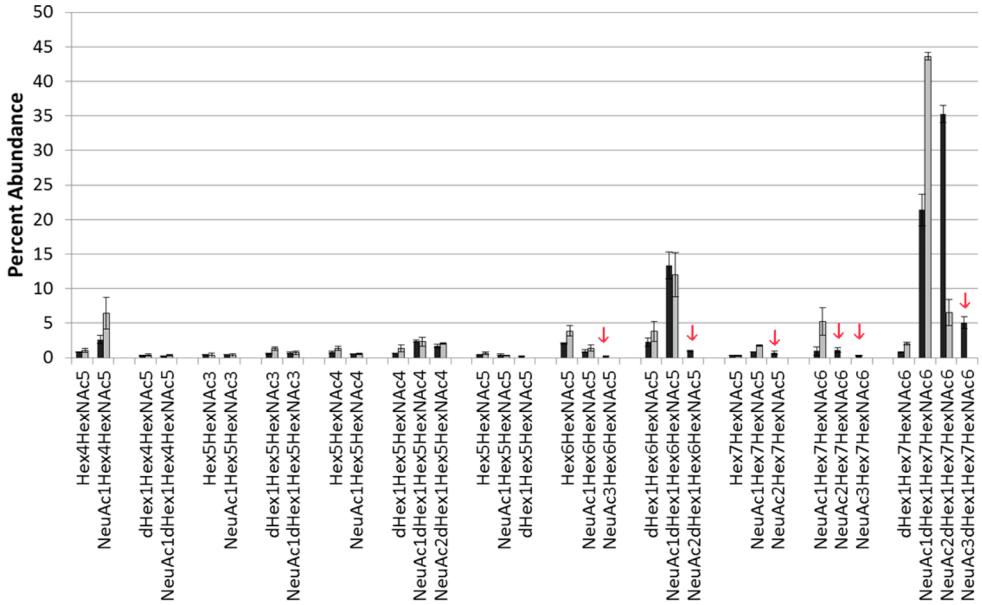


Figure 3. Comparison of MALDI-measured percent abundances of sialylated glycans from the HEK293T (black bars) and HEK293T+VCNA (gray bars) samples. Arrows indicate sialylated glycans that could not be found in the HEK293T+VCNA sample.

shown in Table S2, more glycan compositions were detected by MALDI-TOF MS. However, there was a high degree of overlap since 84% of the glycan compositions were detected in both analyses. Glycans that did not completely overlap between glycan and glycoproteomic pairs can be rationalized as biosynthetically related to detected glycans (see Table S2).

Permethylation Analysis of HA Glycans by MALDI-TOF MS

Semiquantitative analysis of glycans released from each HA was performed by permethylation profiling using MALDI-TOF MS as described in Experimental Procedures. Assigned representative MALDI spectra of all four samples are shown in Figure 2. Permethylated glycans detected for each HA are listed in Table S1 along with standard deviations derived from triplicate measurements. Glycans derived from HA produced in HEK293T cells without or with VCNA treatment contained almost exclusively complex glycans. The percent abundances of HA glycans derived from HEK293T and HEK293T+VCNA sources are compared in Figure 3. HEK293T+VCNA-derived HA glycans (Figures 2B and 3) had less of the trisialylated peaks than the untreated sample. More signal intensity was seen for the highly sialylated forms in the untreated sample at m/z 4226.4 ($NeuAc_3Fuc_1Gal_4Man_3GlcNAc_6$), m/z 4053.0 ($NeuAc_3Gal_4Man_3GlcNAc_6$), and m/z 3602.8 ($NeuAc_3Gal_3Man_3GlcNAc_5$) compared with the treated sample (Figures 2A and 3). VCNA treatment did not remove all sialylation. For example, the peak at m/z 3865.5 ($NeuAc_2Fuc_1Gal_4Man_3GlcNAc_6$, containing two sialic acids) was still present, albeit much weaker, and the peak at m/z 3504.1 ($NeuAc_1Fuc_1Gal_4Man_3GlcNAc_6$, with one sialic acid) increased in overall abundance compared with the untreated sample. The increase in the less glycosylated form is due to partial desialylation of compositionally related glycans, which then combine and appear as intensity at the common lower position in the spectrum.

In the original study, glycan samples from HA derived from HEK293T cells without or with VCNA treatment were labeled by 2-AA, and MALDI-TOF MS data were collected in negative mode.[7] While sialic acid substitutions can be stabilized under these conditions through loss of some acidic carbonyl protons during ionization, some losses can still occur, which can lead to underaccounting of sialyl substitutions. In this study we permethylated the glycans, which protects against sialic acid loss by C1 carbonyl substitution, thus preventing intramolecular hydrolysis by removal of the hydrolytic mobile proton via methyl esterification.[16] In this way the sialic acid residues are stabilized, allowing for a more accurate accounting of sialyl substitutions. While more sialylation was found in the present study, the general trend remains the same. Here we estimate that approximately 30% of the sialic acid was removed by VCNA on the treated samples.

HEK293S GnTI(−) cell-derived HA had the expected $Man_5GlcNAc_2$ with trace amounts of $Man_4GlcNAc_2$ detected, as shown in Figure 2C. These compositions are consistent with the respective high-mannose glycans. Figure 2D shows representative MALDI spectra of the glycans from the HA produced in *Drosophila* S2 cells. The major peaks in this sample were $Man_3GlcNAc_2$ and $Fuc_1Man_3GlcNAc_2$, with lower intensities produced for $Man_4GlcNAc_2$, $Man_5GlcNAc_2$, $Man_3GlcNAc_3$, $Man_2GlcNAc_2$, and $Fuc_1Man_2GlcNAc_2$, consistent with paucimannose glycans, which have been reported previously in S2 cells.[17]

Site-Specific Occupancy

Deglycosylated peptides were analyzed by nanoLC−MS^E to estimate percent occupancy of glycosylation sites. A mass difference of 0.985 Da is produced by the conversion of Asn to Asp (Asp*) through the action of PNGase A during glycan release.[18] The intensity of the peptide bearing the converted Asp* expressed as a percent of Asp* + Asn provides the estimated percent occupancy for any given glycopeptide with some caveats. An apparent Asp* can arise through spontaneous deamidation in proteins and is most significant when a glycine follows the sequon asparagine.[19] This glycine substitution pattern was found in only one glycopeptide (containing site N^{480}). Therefore, spontaneous deamination was likely not a major source of error in our analysis, with the possible exception of site N^{480}. PNGase A can be selective, and

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

Table 2. N-Glycosylation Sites of the Hemagglutinin Protein from A/Mallard/Denmark/64650/03 (H5N7) and Their Glycosylation Potentials As Predicted with NetNGlyc

site	peptide position[a]	theoretical sequence	NetNGlyc potential
10	1−22	DQICIGYHANNSTEQVDTIMEK	0.3793
11	1−22	DQICIGYHANNSTEQVDTIMEK	0.7826
23	23−35	NVTVTHAQDILEK	0.7219
165	163−189	SYNNTNQEDLLVLWGIHHPNDAAEQMK	0.6297
193	190−208	LYQNPTTYVSVGTSTLNQR	0.7293[b]
286	278−304	CQTPMGAINSSMPFHNIHPLTIGECPK	0.6020
480	480−493	NGTYDYPQYSEEAR	0.5839

[a]Numbering based on the mature protein, which lacks the first 16 amino acids. [b]This sequon contains a proline and is not expected to be glycosylated.

downstream manipulation, such as C18 capture, can cause some purification bias during sample preparation. It also should be noted that like pairs of peptides identical in sequence and composition except for the presence of either an Asp or Asn will not have identical ionization potentials. The glycopeptide sequences are shown in Table 2. The results are summarized in Table 3.

Table 3. Site-Specific Occupancies

	sample			
site	HEK293T	HEK293T+VCNA	HEK293S	S2
N11	100[a]	100	100	100
N23	61.9[b] ± 4.7	30.5[b] ± 3.2	93.9 ± 2.4	100
N165	100	100	100	100
N193	1.1 ± 0.02	3.7 ± 0.5	1.6 ± 0.1	1.8 ± 0.2
N286	100	100	100	100
N480	86.9 ± 7.6	83.5 ± 7.5	96.3 ± 0.9	85.6 ± 1.5

[a]No nonglycosylated peptides could be found for values of 100%, and thus, the standard deviation could not be calculated. [b]Peptides with very low signal intensity.

HA produced by HEK293T cells without or with subsequent VCNA treatments gave essentially the same site occupancy results, which was expected since the samples differed only by the addition of the VCNA treatment. Sites N^{11}, N^{165}, and N^{286} were estimated to be fully occupied. Conversely, site N^{193} was estimated to be occupied at less than 4%, with possibly a fraction due to spontaneous deamidation. As will be discussed later, no glycopeptides containing this site were detected by nanoLC−MS^E analyses. Site N^{193} is next to a proline and is therefore not expected to be glycosylated at significant levels.[20] The signal intensities for HEK293T and VCNA-treated HEK293T at site N^{23} were below 3:1 making occupancy estimates difficult. Experiments performed in triplicate for this site for HEK293 HAs without VCNA and with VCNA ranged from 31 to 62%, respectively, for the two samples. However, as will be seen later, glycopeptide analysis produced strong signal intensities for a range of glycoforms at this site, hinting that this site is mostly occupied. The data indicated that no occupancy was observed for site N^{10}.

HEK293S cell HA sites N^{11}, N^{23}, N^{165}, N^{286}, and N^{480} were all estimated to be occupied at greater than 90%. No occupancy was observed for sites N^{10} and N^{193}. Similarly, the S2-derived HA was also nearly fully occupied at sites N^{11}, N^{23}, N^{165}, N^{286}, and N^{480}, with no or trace occupancies detected for sites N^{10} or N^{193}, respectively. Overall, all of the HAs' five sites were highly glycosylated. The site occupancy results are summarized in Table 3.

Site-Specific Glycan Assignments

The glycopeptides were analyzed by nanoLC−MS^E. The data were processed and analyzed using BiopharmaLynx (Waters) in conjunction with GLYMPS. Approximately 85% of the glycans identified in HEK293T without or with VCNA by LC−MS were also found in the released glycan scans performed by MALDI-TOF MS. Nearly all of the LC−MS-assigned glycopeptide-resident glycans were also found in the MALDI-TOF MS permethylation analysis data for all of the other HAs with the exception of $Fuc_1Man_5GlcNAc_2$, a low-abundance peak seen only in the LC−MS data of S2-derived HA. While the MALDI-TOF and LC−MS paired experiments did not completely overlap for HEK293T without or with VCNA, all of the detected glycans were synthetically related. As the chemical properties of the analytes and the analytical methods used in their detection were different, some differences are expected. Across the four samples, 84% of the glycoforms were copresent in the two analyses.

Example MS^E spectra for each HA's N^{23} site are shown in Figure 4 (see Table 2 for sequence and position). The parent mass, which includes the peptide plus the full glycan, is indicated at the far right in each spectrum, with the assigned glycopeptide fragments decreasing in size from right to left revealing glycan composition and sequence. Glycopeptide fragment assignments were made by GLYMPS. Peptide fragments assigned by BiopharmaLynx are indicated by red dashed lines and were also confirmed by GLYMPS analysis. Oxonium ions (also assigned by GLYMPS) were found in the low-mass region of the spectra. Oxonium ions were present in many of the assigned spectra, but in certain ones, particularly those of low abundance or those with only Hex and HexNAc, few to none were found. It was also noted that spectra that had an abundance of glycosidic fragments tended to have fewer parent and oxonium ions (compare Figure 4A−D), as might be expected on the basis of each ion's characteristic competitive decomposition pathways and the standardized instrument conditions used. We also noted that about 75% of the methionines were converted to *S*-carboxyamidomethylmethionine through a reaction with iodoacetamide during alkylation.[21] This modification results in an increase of 58 Da to the methionine, detected as a 57 Da increase because of the positive charge imparted by the carboxyamidomethylation, as is typical of such peptides, leading to detection of the M^+ peptide ion form.[21] Parent ions were detected with the modification intact, but upon exposure to higher collision energies, α-(methylthio)acetamide is lost, resulting in a neutral loss of 105 Da, as can be seen in Figure 5.[21] GLYMPS was used to find and assign glycopeptides with *S*-carboxymethylmethionines, which aided in systematic identification of this modification. This

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

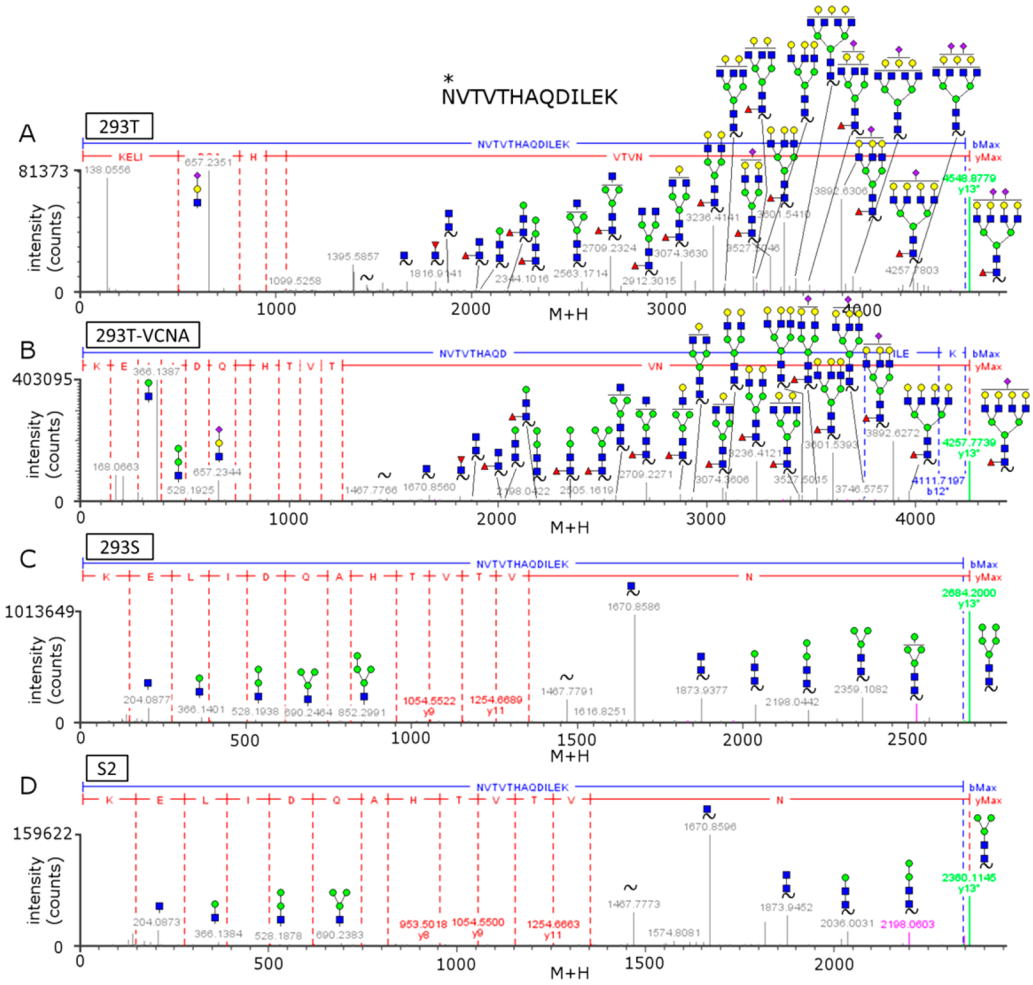


Figure 4. Example of glycosylation at site N^{23} in each sample. The full glycan attached at each site is shown to the far right. Red lines are y ions, and blue lines are b ions. The green line in each denotes the parent glycopeptide ion. Gray lines are fragments unassigned by BiopharmaLynx. Glycopeptide and oxonium ion assignments made by GLYMPS have been added to the spectra. Monosaccharide symbols: blue squares, GlcNAc; green circles, mannose; yellow circles, hexose; red triangles, deoxyhexose; purple diamonds, sialic acid; black squiggles, unfragmented peptide.

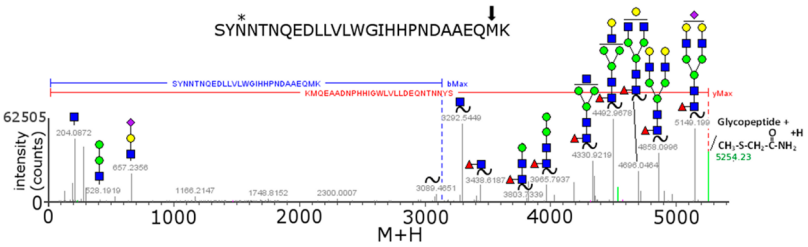


Figure 5. Example spectrum demonstrating changes caused by the reaction of methionine with iodoacetamide. The peptide containing site N^{165} (top) has a single methionine (highlighted with an arrow). Colors are as described in Figure 4. Peptide fragments y3, y6, y9, y13, y15, y16, y18, y23, y24, b5, and b9 are present but could not be assigned and labeled by BiopharmaLynx because of the complicated fragmentation. The parent mass of 5254.23 Da is the peptide (3136.5 Da) + glycan (2059.7 Da) + acetamide (57 Da) + H^+ (1 Da). When fragmentation begins, a fragment of 105 Da breaks off, resulting in a glycopeptide of 5149.2 Da with a peptide that now has a mass (M + H) of 3089.46 Da. Glycopeptides, peptide fragments, and oxonium ions were assigned by GLYMPS.

modification has been noted previously under circumstances leading to overalkylation but has not been reported in glycopeptides to our knowledge.

Each glycosylation site can contain a number of different glycoforms. Since the peptide moiety for each glycopeptide is the same and the glycan mass differences are small (~300 to 700 Da), the signal abundances detected by LC−MS^E under standard conditions should be comparable when neutral but not sialylated glycans are present. This provides a means for estimating the abundance of non-sialyl glycoforms at each glycosylated site. Data from separate experiments can be normalized by converting the signal abundances to percent abundances. This is calculated by dividing each individual glycoform signal by the total of all of the glycoform signals in an experiment. Similar estimates have been made previously for *O*-glycopeptides by Wada et al.[22] In that study, the authors compared glycopeptides from produced IgA1 clinical samples bearing a range of glycans from $Hex_2HexNAc_3$ to

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

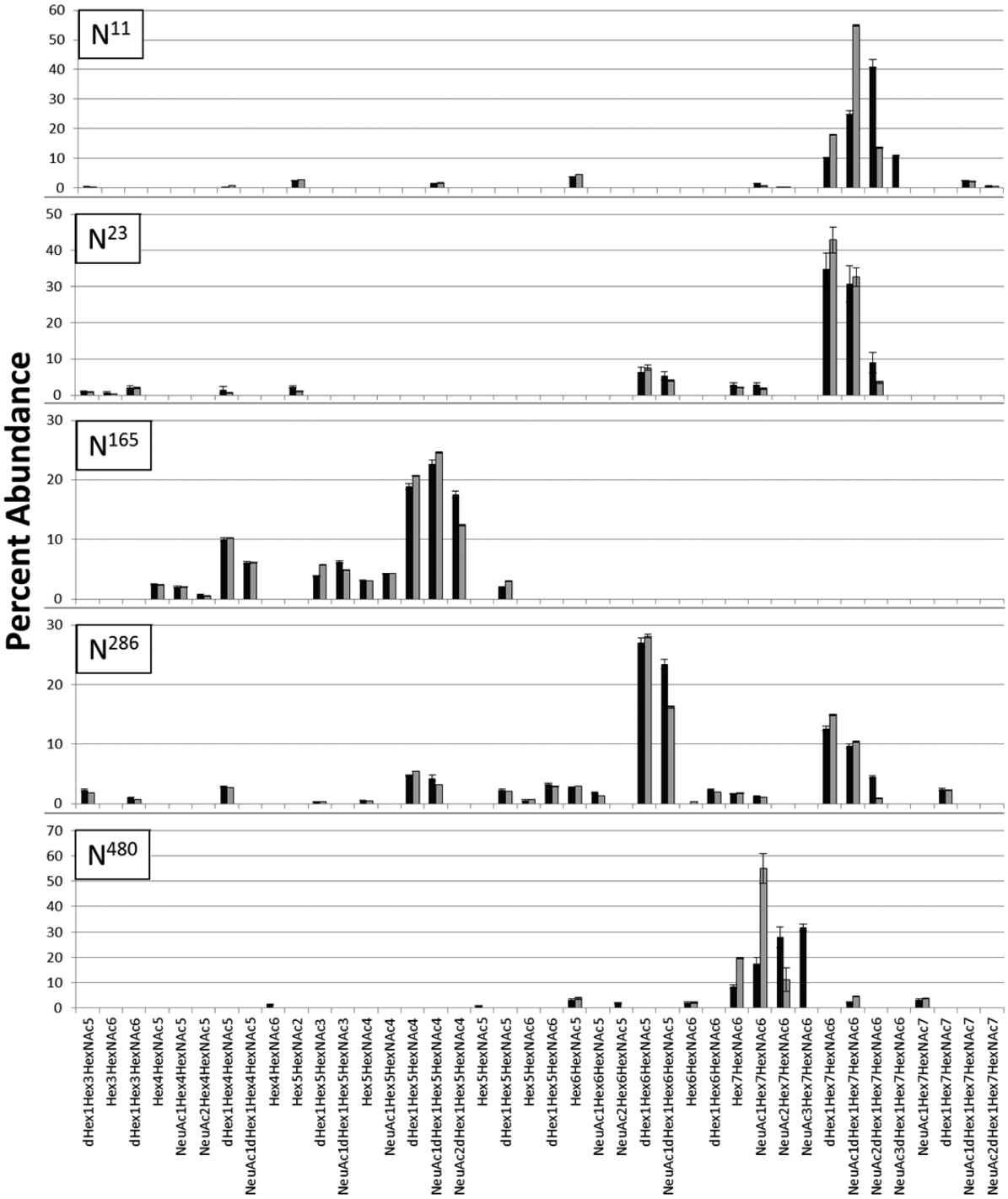


Figure 6. Percent abundances of LC−MS data for each site. Black bars represent HEK293T data and gray bars HEK293T+VCNA data. Bars are averages of three sets of data. Error bars represent one standard deviation.

$Hex_5HexNAc_6$. To test whether LC−MS intensities of *N*-glycopeptides can be used in this way, a simple comparison was performed. RNase B has a single glycosylation site populated with high-mannose glycans. It is known that MALDI-TOF MS analysis of permethylated glycans is highly reproducible and semiquantitative.[23] We compared the released glycan permethylation profile of RNase B detected by MALDI-TOF MS analysis to the RNase B glycoforms detected on its single glycopeptide DSSTSAASSSNYCNQMMKSRNLTK by LC−MS^E. The average intensities of MALDI-TOF data from three separate spots and the average intensities of LC−MS^E data from three independent BiopharmLynx-processed experiments were converted to percent abundances and plotted (see Figure S1). The results showed that the two methods were comparable to within 20% of peak height or less in relative abundances of each composition, supporting that the two methods are similarly predictive of abundances of glycoforms bearing neutral glycans. Therefore, we concluded that the LC−MS^E method can be used for semiquantitative occupancy estimates for neutral glycoforms at each glycosylation site. Others have also used this method in glycopeptide quantitation produced from prostate-specific antigen in a multilaboratory comparative

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

study.[24] Where comparisons are made in samples derived from HEK293 cells containing both neutral and sialyl glycans, the comparisons only serve as an index of abundance.

Using the same LC−MS^E method, we analyzed all four HA samples. The HEK293T- and HEK293T+VCNA-derived HAs' site-by-site percent abundances are plotted in Figure 6. The glycan abundances from the HEK293T and HEK293T+VCNA samples overlap as expected, as can be seen in the overall glycan distributions detected by MALDI-TOF-MS. Desialylation by VCNA was incomplete at the glycopeptide level since some lower charged forms remained in the treated HEK293T HA sample compared with the untreated sample. Desialylation appeared most efficient at sites N^{11} and N^{480}, as evidenced by the most dramatic differences in sialylated forms compared with the other sites. Desialylation was less complete at sites N^{23}, N^{165}, and N^{286}. Both N^{11} and N^{480} are present at the stem region, which may be more accessible to the VCNA enzyme.

Glycopeptide analysis revealed a regionally specific pattern of glycan complexity. This is demonstrated in Figure 7 where the most abundant glycan detected at each glycosylation site in HEK293T, HEK293S, and S2 HAs is indicated next to the HA monomer's three-dimensional representation. The predominant form of HEK293S HA at all sites is $Man_5GlcNAc_2$, but for S2 and especially HEK293T HAs, the largest and most complex glycans were found near the bottom of the stem region of HA, the area nearest the transmembrane region in the intact virus. Smaller complex glycans were observed near the head region at site N^{165}, the only occupied glycosylated site on the head. Medium-sized complex forms were seen at N^{286} in the HA1 segment about midway toward the globular head.

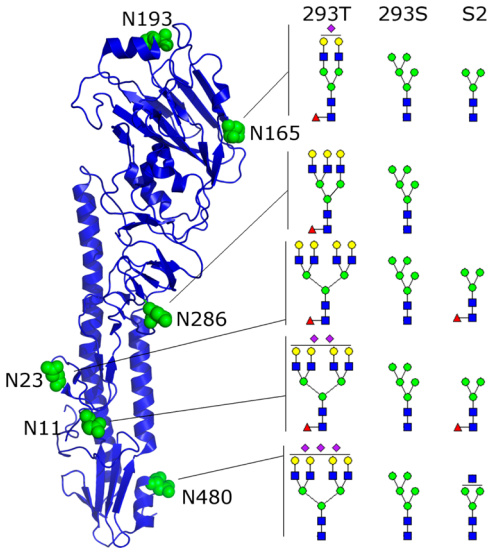


Figure 7. Predominant glycans found at each glycosylation site. The blue structure is a backbone ribbon representation of H5 modeled using I-TASSER[14] and drawn using PyMOL.[48] Green balls are glycosylated or potentially glycosylated asparagines. Glycan structures are the most predominant forms of each family determined from the data presented in Figures 6 and 8. Monosaccharide symbols are as described in Figure 4. Abbreviations are 293T for HEK293T and 293S for HEK293S GnTI(−).

The glycosylation complexities of the HEK293S and S2 cells' HA samples were simplified because of a nonfunctional GnTI and natural potential, respectively, as previously discussed. In a comparison of all glycans at each site, not just the predominant one, HA site N^{480}, near the bottom of the stem region, had the largest proportion of maximally processed glycoforms in both HAs (Figure 8). In HEK293S, HA compositions were observed that were consistent with trimming of the Man5GlcNAc2 glycan to smaller forms, with some core fucosylated forms also observed. The S2 HA sample also had maximal complexity at site N^{480}, which was observed as compositions consistent with processing to larger forms of the core $Man_3GlcNAc_2$ with added GlcNAc and Fuc.

The site with the second most diverse glycans in all four HAs was N^{23}, with essentially the same but less varied patterns described for N^{480}. Similarly, N^{11} also preferred $Fuc_1Man_3GlcNAc_2$, indicating processing heterogeneity in S2 cells. It is noteworthy that N^{165} was occupied by mannosyl glycans in HEK293S and S2 cells while in HEK293 without or with VCNA it was occupied by complex glycans. This could be significant since in other HAs, such as H3 and H1, this site is a target for SP-D lung surfactant, where it is occupied by high-mannose glycans.[25] The S2 HA glycopeptides containing N^{286} produced MS^E spectra with a signal-to-noise ration (S/N) less than 3, preventing accurate quantitation of glycoforms and failure of some spectra to meet our five inclusion criteria. However, close manual examination of the spectra did indicate that paucimannose forms were present.

Comparative Molecular Modeling of H3 and H5 HA

Like H5, H3 HA also has a glycosylation site at N^{165}, but unlike H5, which has complex glycoforms at N^{165}, only high-mannose glycoforms have been found on that site in H3 HA.[25a,26] Consensus protein sequences and available structures were compared in order to better understand this observation. Comparison of the consensus sequences for the H5 and H3 proteins shows that the N^{165} glycosylation site in H3 is shifted four amino acids upstream compared with the site in H5 (Figure 9A,D). This places the H3 N^{165} glycosylation site at the beginning of a β-strand rather than the end as in H5 and alters the site's position in relation to the 220 loop (215−223 in H5 and 219−227 in H3). This shift results in a potential association of H3 N^{165} glycoforms with the 220 loop, compared with the lack of association in H5 proteins (Figure 9A). As can be seen in the representative crystal structures of H3 and H5 shown in Figure 9B,C, the H3 glycoform appears to pack against W222 in the loop, whereas the H5 glycoform is well-separated from the loop.

For a more detailed comparison, the van der Waals and hydrogen-bonding interactions between the glycoforms and the 220 loop in crystal structures of several H3 and H5 proteins were compared. The crystal structures studied were limited to those with better than 2.5 Å resolution and without mutations or ligands near the region of interest. For the H3 proteins, an additional constraint was the presence of tryptophan at position 222, as this position was conserved until 2003.[27] Similar packing between glycoforms and the 220 loop can be observed in more recent H3 HA structures containing an arginine at position 222 (not shown). The structures used are listed in Table 4. Ligplot, a program to automatically plot protein−ligand interactions,[13] was used to access likely interactions between existing glycoforms and the 220 loop in the structures. In addition, van der Waals interactions were measured according to the distances calculated by Alvarez.[28] Figure 9E is a Ligplot representation of the H3 interactions in A/Aichi/2/1968 (PDB entry 2VIU),[29] and Figure 9F shows the H5 interactions in A/Duck/Singapore/3/97 (PDB entry 1JSM).[30]

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

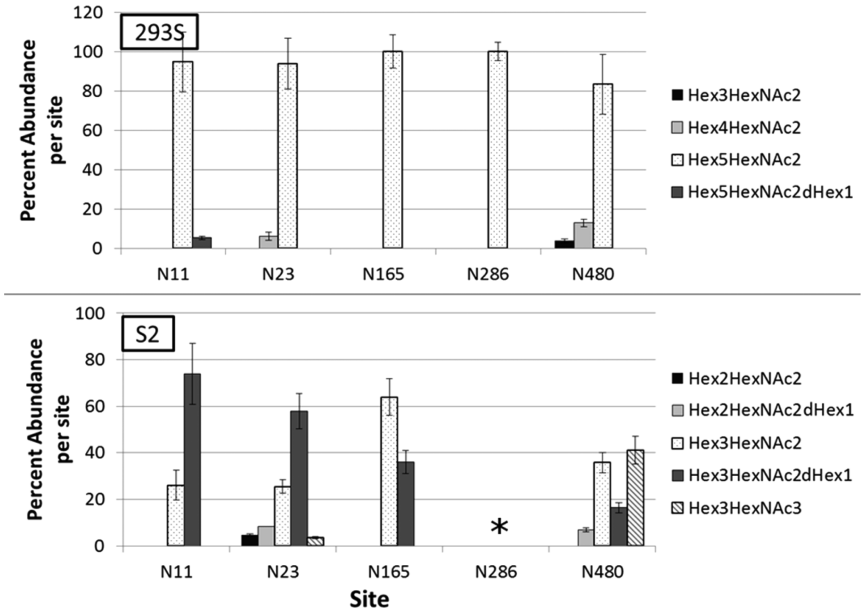


Figure 8. Percent abundances of LC−MS data at each site in the HEK293S and S2 samples. Bars represent the percent abundances of the glycans listed in the legends per site, not per sample. Bars are averages of three sets of data. Error bars represent one standard deviation. * indicates that paucimannose forms were detected but with S/N less than 3:1, preventing quantitation.

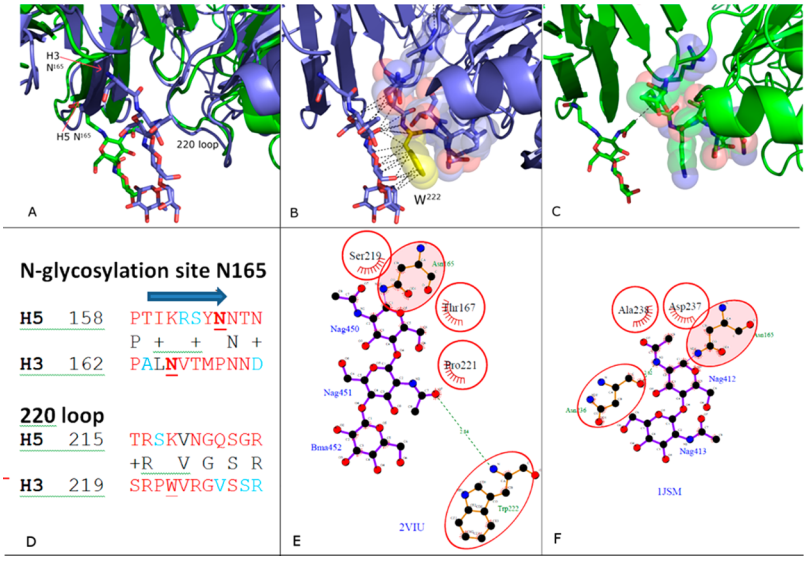


Figure 9. Comparison of site N^{165} in H3 and H5 hemagglutinin. (A−C) PDB structures 1JSM (H5) in green and 4FNK (H3) in blue. Figures were drawn with PyMOL.[48] (A) Overlay of the H3 and H5 proteins aligned using the A subunit of each structure. N^{165} residues are indicated with red arrows. Glycans are shown as sticks, and the protein is displayed in cartoon form. (B) H3 structure only, with a space-filling/stick model of the 220 loop. W^{222} is shown in yellow. (C) H5 structure only, with a space-filling/stick model of the 220 loop. In (B) and (C), van der Waals contacts between the glycans and the 220 loop are shown as dashed lines. (D) Sequence comparisons of H3 and H5 HA proteins in the regions of the N^{165} glycosylation sites and the 220 loop. W222 and both N^{165} sites are underlined. Amino acids in red are >90% conserved, those in blue are <80% conserved, and those in black are between 90 and 80% conserved. Numbering is based on common numbering in the H3 and H5 crystal structures. The location of the β-strand at site N^{165} in both structures is indicated by a blue arrow above the H5 sequence. (E, F) Ligplot[13] representations of conserved contacts between core GlcNAc residues and the protein at the N^{165} sites in H3 and H5 hemagglutinin proteins. N^{165} is highlighted with a pink oval. Hydrophobic contacts are represented by red arcs with spokes. Hydrogen bonds are shown as green dashed lines. Amino acids making contacts in two or more structures are shown as red ovals. (E) Site from H3 in structure 2VIU. (F) Site from H5 in structure 1JSM. The residue name for *N*-acetylglucosamine is Nag, and that for mannose is Bma.

An inter-subunit hydrogen bond exists between W222 and the second GlcNAc O7 in all four of the H3 crystal structures studied, and an intersubunit hydrogen bond between S219 and the first GlcNAc N2 atom exists in three of the structures. By comparison, all of hydrogen bonds to H5 glycoforms are intrasubunit and only one structure, A/Vietnam/1194/2004 (PDB entry 4BGW),[31] has a hydrogen bond to the second GlcNAc (GlcNAc2 O7 to N236). The first three sugars of all four H3 proteins made an average total of 22 inter-subunit van der Waals contacts to the 220 loop, whereas only two of the H5 structures had inter-subunit van der Waals contacts and only from their first GlcNAc (to P217 in 1JSM and to S216 in

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

Table 4. H3 and H5 Hemagglutinin Structures Used to Compare the Glycosylation Site at N^{165}; The Numbers of van der Waals (VDW) and Hydrogen-Bonding (HB) Contacts between the Glycans at Site N^{165} in the Crystal Structures and the Neighboring 220 Loop Are Also Given

H3 Hemagglutinin					
				no. of intersubunit contacts	
virus	ref	PDB entry	glycan resid[a]	VDW	HB
A/Aichi/2/1968	29	2VIU	450−452	28	1
A/Hong Kong/a/1968	43	4FNK	404−407	21	2
A/Equine/Richmond/2007	44	4UO0	431−434, 437	19	2
A/Port Chalmers/1/1973	45	4WE5	404−406	20	2
average				22	1.75
H5 Hemagglutinin					
				no. of intrasubunit contacts	
virus	ref	PDB entry	glycan resid[a]	VDW	HB
A/Duck/Singaport/3/1997	30	1JSM	412−413	1	0
A/Vietnam/1194/2004	31	4BGW	1322−1323	0	0
A/Indonesia/5/2005	46	4K62	601	0	0
A/Egypt/n03072/2010	47	4KW1	401−403	2	0
average				0.75	0

[a]"glycan resid" refers to the residue identification numbers for each sugar in the glycoform at N^{165} in the crystal structures.

4KW1). On the basis of these observations, we hypothesize that the interaction of the glycoforms in the H3 hemagglutinins with the 220 loop may limit further processing of the glycoforms in the Golgi apparatus by fixing the sugars in a configuration unfavorable for modification. However, in H5, the limited contacts to the first GlcNAc do not prevent further processing in the Golgi apparatus, allowing the creation of complex glycans as shown in the study presented here.

■ DISCUSSION

We have studied the glycosylation patterns of recombinant A/Mallard/Denmark/64650/03 (H5N7) HA produced using conditions to yield four different glycosylation patterns. These include HAs produced in (i) HA HEK293T cells with complex glycoforms, (ii) HEK293T cells treated with VCNA to provide asialo-complex glycoforms, (iii) HEK293S GnTI(−) cells with predominantly the canonical $Man_5GlcNAc_2$ glycoform, and (iv) *Drosophila* S2 insect cells bearing primarily paucimannose glycoforms. A summary of the strains and their expected glycan composition classes can be found in Table 1. Here we extended our previous study[7] by providing a more comprehensive structural analysis of these HAs. Quantitation of overall glycoforms and site-specific glycoforms has been performed. Site occupancies have been estimated, and we have modeled the glycosylation patterns on a site-by-site basis to the three-dimensional HA structure for each HA studied here, revealing glycan orientation relative to reported antigenic sites as well as other structural observations that may impact glycosylation states.

There are seven potential glycosylation sites in the H5N7 HA studied in this paper. NetNGlyc predicted glycosylation at sites N^{11}, N^{23}, N^{165}, N^{286}, and N^{480}, while sites N^{10} and N^{193} were predicted to lack glycans. Site N^{10} is embedded in another sequon that is glycosylated at N^{11}, while N^{193} contains proline, which often prevents sequon usage.[20] All five sites predicted to be occupied were found to be glycosylated in all of the HAs in this study at high percentages. The NetNGlyc predictions for H5N7 are similar to the predictions of our recent analysis of two H5N1 HAs derived from A/Vietnam/1203/2004 and A/bar-headed goose/Qinghai/14/2008, but the results were somewhat different.[32] In the H5N1 strains, we found that site N^{193} was occupied in insect-cell-grown HA while N^{286} was not occupied in HEK293T or egg-derived forms. In H5N7 HAs, however, N^{193} is not occupied in any cell system but N^{286} is occupied in all of them. Thus, sites N^{193} and N^{286} appear to be variably glycosylated depending on the cell system used. The reason for this is not clear. There are some minor differences present in the primary sequences of H5N1 and H5N7 HAs, but most are conservative substitutions and are distant from the glycosylation sites. N^{193} contains proline in the X position of the sequon NX(S/T) motif, which generally prohibits its use. N^{286} contains S in the S/T position of the motif, which reduces the glycosylation efficiency compared with T.[33] Therefore, both of these sites likely exist on the edge of glycosylation probability.

The general patterns of glycosylation for the five HAs studied here followed the expected trends for each cell type (293T, 293S, and S2) and enzymatic (VCNA) or genetic (GnTI oblation) manipulation used in their production. HEK293T cells produced HA bearing essentially exclusively complex-type *N*-glycans ranging from biantennary to tetraantennary in composition, and the majority contained one or more sialic acids. Treatment with VCNA removed approximately 30% of the sialic acids. Since HEK293T-cell-derived HAs without or with VCNA are used to study the role of desialylation on host receptor binding, our findings tell a cautionary tale for interpretation of these studies since significant sialylation can remain. The HEK293S-derived HA contained primarily $Man_5GlcNAc_2$, as expected, with low-abundance shorter, high-mannose forms as well. The *Drosophila* S2 cell-derived HA was modified with a range of insect glycoforms primarily of the compositions $Man_3GlcNAc_2$ and $Fuc_1Man_3GlcNAc_2$ with low-abundance forms of larger paucimannose glycans and those with addition of one GlcNAc, all of which are expected for this cell line.[34]

There was a common trend in site-specific glycan complexity in the HAs from different cell lines: the extent of glycan processing decreased in the direction of stem to globular head. This pattern was most striking in the HEK293T-derived HAs since these cells produce the greatest amounts of glycan

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

complexity (Figure 7). The most highly processed glycans were found at glycosylation sites closest to the base of the stem region, at sites N^{11}, N^{480}, and N^{23}. At these sites there were large amounts of tri- and tetraantannary glycans. Midway up the stem at N^{286}, medium-sized complex glycans were seen, with high amounts of tri- and biantennary forms and lesser amounts of tetraantennary forms. The glycans with the lowest complexity, consisting primarily of biantennary glycans, were seen at N^{165}, the only site on the globular head. The HAs from other cell sources followed essentially the same pattern of complexity in accordance with their cell-specific glycosylation potentials. A retrospective analysis of our glycosylation study of A/bar-headed goose/Qinghai/14/2008 H5N1 HA grown in HEK293T cells (see Figure S2) showed the same stem-to-head glycan processing pattern. Glycosylation site N^{154}, an additional head domain site on the H5N1 HAs, also followed the stem-to-head complexity trend since it had low complexity with glycan compositions similar to those at site N^{165}. The pattern has also been noted in a range of H1N1 strain HAs produced both recombinantly and in viral particles.[35] Protein structure, domain swapping, membrane association, presence of glycosylation processing enzymes, and sequon distance from the C-terminus can influence the presence and/or complexity of glycosylation.[36] Considering that the stem-to-head glycosylation trend is seen over a range of HAs produced under various conditions in a range of cell lines, it is likely that protein structure is the dominant factor determining the observed pattern.

In our previous study of A/Mallard/Denmark/64650/03 (H5N7)-derived HAs,[7] we immunized both chickens and mice with HAs to examine the impact that different glycosylation states have on immune response as interpreted from the following: host antibody production, each HA's ability to bind DC-SIGN (a host mannose lectin-based immune receptor), each HA's ability to bind fetuin (a surrogate host receptor), and each HA's HAI titer. Also studied were the effects of said glycosylation differences on dendritic cell activity. We found that HA proteins carrying terminal mannose moieties, present in high-mannose and paucimannose glycans, induce significantly lower HAI antibody titers than HA proteins carrying complex glycans. The high-mannose HA forms were found to bind more strongly to DC-SIGN than those bearing complex glycoforms. This suggests that antibody response (as measured by reduced GpG-induced production of interferon α) is downregulated through high-mannose HA binding to DC-SIGN on dendritic cells. Our findings in this study confirm the high-mannose status of the HEK293S- and S2-derived HAs and the exclusive presence of complex glycans on the HEK293T HAs. Clearly, the immune response of the host can be affected by the glycosylation status of the H5 HA antigen.

Surfactant Protein D (SP-D) is another mannose-binding molecule that is important for lung clearance of influenza.[37] HA glycosylation site N^{165} is indispensable for interaction with SP-D in many HAs, including H3 and H1 HAs, and requires that N^{165} be occupied by high-mannose glycans.[26,37,38] Glycosylation sites close by N^{165} are also important for SP-D interaction and viral lung clearance, as indicated in H3N2 viruses that demonstrate decreased pathogenicity with gain of glycosylation sites at N^{133} and N^{246}.[39] In our recent study of the A/Hong Kong/1/68-derived recombinant H3 HA series, engineered to have sequential addition of historically relevant glycosylation sites, the HAs were occupied by high-mannose glycans at both N^{133} and N^{246}, increasing the interaction with SP-D and facilitating clearance from the lung.[25b,39] Conversely, studies that have investigated SP-D activity toward H5 HAs that have glycosylation site N^{165} have not found strong interactions with SP-D.[40] Hartshorn et al.[40c] investigated the SP-D sensitivity of A/Vietnam/1203/04 H5N1 and found it to be resistant. This HA contains N^{165} and N^{154}, both proposed to be in the SP-D binding region of HA. Both H5 and H3 HAs have glycosylation sites that would predict SP-D activity promoting virus clearance. So, why are H3 viruses cleared while H5 viruses are not?

We have studied a range of H5 and H3 HAs produced in host systems that are capable of producing complex, hybrid, and high-mannose glycans. These include HEK293, egg, and MDCK. While it is possible that high mannose forms are present at abundances below our limits of detection, we have observed only complex glycans at site N^{165} in H5 produced in MDCK cells (unpublished observations). Likewise, in all cases at site N^{165} we have found exclusively complex type in H5 HA[32] and exclusively high-mannose type in H3.[25b] Examples of the H5 complex-only situation can be seen in our publication, An et al., 2013, where both egg A/Vietnam/1203/04 H5N1 and HEK293 A/Bar-headed goose/Qinghai/14/2008 H5N1 HAs were as described. Our previous work (An et al.[32]) shows four different H3 HAs produced in MDCK cells derived from A/Hong Kong/68, all of which are high-mannose at N^{165}. Khatri et al.[35c] recently published glycosylation studies of another H3 HA, A/Philippines/2/82, produced in chicken eggs. Again, N^{165} was exclusively high-mannose. Therefore, we have observed cases in both H5 and H3 derived from both egg and MDCK that follow the HA-specific pattern of N^{165} glycan subtype, complex in H5 and high-mannose in H3. In the current study, HEK293 also appears to exclusively glycosylate N^{165} with complex glycans in H5 HA. That SP-D is dependent on high-mannose glycans at N^{165} explains why H3 viruses but not H5 viruses are efficiently cleared from the lung.

As shown here, at site N^{165} produced in cells not capable of producing complex and hybrid glycans (HEK293S and S2), only high-mannose and paucimannose types are seen. These HAs could be used to investigate whether the glycoforms at N^{165} of H5 viruses can serve as receptors for SP-D when high-mannose glycans are forced to be present via cell expression system selection. Since lectin-based innate immune factors may act differently on such HA antigens, such a strategy may be useful in vaccine design.

To better understand why the glycans at N^{165} in H5 are complex while those in H3 are high-mannose, we performed a comparative analysis of available X-ray crystal structures resolved at 2.5 Å or better. The N^{165} glycosite in H3 is four amino acids upstream compared with that of H5, placing it in a position recessed backward to the 220 loop and possibly less accessible to glycosylation enzymes. The first three sugars of the H3's N^{165} glycan make an average total of 22 inter-subunit van der Waals contacts to the 220 loop, whereas only two of the H5 structures had inter-subunit van der Waals contacts and only from their first GlcNAc. Inter-subunit hydrogen bonds to the second GlcNAc are found in H3, whereas none exist in the H5 structures.

As it is difficult to crystallize proteins with native glycans, the structures of proteins that lack them, such as those used here, may not be completely accurate and likely do not account for all of the significant energetic states. However, the similarities of the measurements among several proteins suggest that the H3 glycan is more constrained and less accessible to glycan processing enzymes in the Golgi apparatus compared with H5

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

HAs, resulting in high-mannose glycans in H3 and complex ones in H5.

CONCLUSION

In this paper, we have presented the first glycosylation analysis of an H5N7 influenza virus. Similar to H5N1 and H1N1 HAs, the glycosylation decreases in complexity from the stem to the head of the protein. The H5N7 HA studied here did not have head sites containing high-mannose glycans when grown in HEK293T cells and thus is less likely to be involved in SP-D- or DC-SIGN-based interactions. This situation may contribute to its pathogenicity and poor lung clearance if it successfully leaps to the human host. Like other H5 HAs, it has glycans present near known antigenic sites. The antigenic architecture of hemagglutinin H5 has recently been reviewed.[41] Antigenic sites have been located in regions containing or in the vicinity of four of the five glycosylation sites found to be occupied in the present study. These include sites N^{11}, N^{23}, N^{165}, and N^{286}. The majority of HA sequences used in those studies were derived from H5N1 viruses that were highly homologous to the H5 studied here. As glycans can mask epitopes, it may be of use to consider strategies to denude vaccine strains of glycosylation to maximize antigenic exposure, including pandemic ones as shown in recent reports.[42]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00175.

> Glycans found in each sample by MALDI-TOF MS (Table S1); comparison of glycoforms detected by MALDI-TOF MS of permethylated glycans and Q-TOF MSE glycopeptides (Table S2); comparison of glycan abundances of RNaseB from released permethylated glycans measured by MALDI-TOF MS and unlabeled glycopeptides measured by LC−MS (Figure S1); and percent abundances of LC−MS-measured glycopeptides at each site from previously studied goose/Qinghai/14/2008 H5N1 HA (Figure S2) (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: john.cipollo@fda.hhs.gov. Phone: 240-402-9445.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

R.P.d.V. was a recipient of VENI and Rubicon Grants from The Netherlands Organization for Scientific Research (NWO).

ABBREVIATIONS

Hex, hexose; Gal, galactose; Man, mannose; HexNAc, *N*-acetylhexosamine; GlcNAc, *N*-acetylglucosamine; dHex, deoxyhexose; Fuc, fucose; NeuAc, *N*-acetylneuraminic acid or sialic acid; HA, hemagglutinin; PGC, porous graphite carbon; HEK, human embryonic kidney; VCNA, *Vibrio cholera* neuraminidase; KIF, kifunensine

REFERENCES

(1) WHO Influenza (Seasonal). http://www.who.int/mediacentre/factsheets/fs211/en/ (accessed Nov 4, 2014).

(2) (a) Tzarum, N.; de Vries, R. P.; Zhu, X.; Yu, W.; McBride, R.; Paulson, J. C.; Wilson, I. A. Structure and receptor binding of the hemagglutinin from a human H6N1 influenza virus. *Cell Host Microbe* **2015**, *17* (3), 369−76. (b) Zhang, H.; de Vries, R. P.; Tzarum, N.; Zhu, X.; Yu, W.; McBride, R.; Paulson, J. C.; Wilson, I. A. A human-infecting H10N8 influenza virus retains a strong preference for avian-type receptors. *Cell Host Microbe* **2015**, *17* (3), 377−84.

(3) Imai, M.; Kawaoka, Y. The role of receptor binding specificity in interspecies transmission of influenza viruses. *Curr. Opin. Virol.* **2012**, *2* (2), 160−7.

(4) Mair, C. M.; Ludwig, K.; Herrmann, A.; Sieben, C. Receptor binding and pH stability - how influenza A virus hemagglutinin affects host-specific virus infection. *Biochim. Biophys. Acta, Biomembr.* **2014**, *1838* (4), 1153−68.

(5) Cox, M. M.; Patriarca, P. A.; Treanor, J. FluBlok, a recombinant hemagglutinin influenza vaccine. *Influenza Other Respir. Viruses* **2008**, *2* (6), 211−9.

(6) (a) Deshpande, K. L.; Fried, V. A.; Ando, M.; Webster, R. G. Glycosylation affects cleavage of an H5N2 influenza virus hemagglutinin and regulates virulence. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (1), 36−40. (b) Wagner, R.; Wolff, T.; Herwig, A.; Pleschka, S.; Klenk, H. D. Interdependence of hemagglutinin glycosylation and neuraminidase as regulators of influenza virus growth: a study by reverse genetics. *Journal of Virology* **2000**, *74* (14), 6316−23. (c) Hanson, S. R.; Culyba, E. K.; Hsu, T.-L.; Wong, C.-H.; Kelly, J. W.; Powers, E. T. The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (9), 3131−3136. (d) Wang, C.-C.; Chen, J.-R.; Tseng, Y.-C.; Hsu, C.-H.; Hung, Y.-F.; Chen, S.-W.; Chen, C.-M.; Khoo, K.-H.; Cheng, T.-J.; Cheng, Y.-S. E.; Jan, J.-T.; Wu, C.-Y.; Ma, C.; Wong, C.-H. Glycans on influenza hemagglutinin affect receptor binding and immune response. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (43), 18137−18142. (e) de Vries, R. P.; de Vries, E.; Bosch, B. J.; de Groot, R. J.; Rottier, P. J.; de Haan, C. A. The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity. *Virology* **2010**, *403* (1), 17−25.

(7) de Vries, R. P.; Smit, C. H.; de Bruin, E.; Rigter, A.; de Vries, E.; Cornelissen, L. A.; Eggink, D.; Chung, N. P.; Moore, J. P.; Sanders, R. W.; Hokke, C. H.; Koopmans, M.; Rottier, P. J.; de Haan, C. A. Glycan-dependent immunogenicity of recombinant soluble trimeric hemagglutinin. *Journal of Virology* **2012**, *86* (21), 11735−44.

(8) Reeves, P. J.; Callewaert, N.; Contreras, R.; Khorana, H. G. Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (21), 13419−13424.

(9) An, Y.; Cipollo, J. F. An unbiased approach for analysis of protein glycosylation and application to influenza vaccine hemagglutinin. *Anal. Biochem.* **2011**, *415* (1), 67−80.

(10) (a) Ciucanu, I.; Costello, C. E. Elimination of oxidative degradation during the per-O-methylation of carbohydrates. *J. Am. Chem. Soc.* **2003**, *125* (52), 16213−16219. (b) Ciucanu, I.; Kerek, F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **1984**, *131* (2), 209−217.

(11) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.

(12) Palmer, M. MakeMultimer.py. http://watcut.uwaterloo.ca/tools/makemultimer/index (accessed Feb 26, 2016).

(13) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1995**, *8*, 127−134.

(14) Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinf.* **2008**, *9*, 40.

(15) GLYCAM-Web. http://glycam.org (accessed Feb 26, 2016).

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

(16) Ruhaak, L. R.; Zauner, G.; Huhn, C.; Bruggink, C.; Deelder, A. M.; Wuhrer, M. Glycan labeling strategies and their use in identification and quantification. *Anal. Bioanal. Chem.* **2010**, *397* (8), 3457−81.
(17) Kim, Y. K.; Shin, H. S.; Tomiya, N.; Lee, Y. C.; Betenbaugh, M. J.; Cha, H. J. Production and N-glycan analysis of secreted human erythropoietin glycoprotein in stably transfected Drosophila S2 cells. *Biotechnol. Bioeng.* **2005**, *92* (4), 452−461.
(18) Takahashi, N. Demonstration of a new amidase acting on glycopeptides. *Biochem. Biophys. Res. Commun.* **1977**, *76* (4), 1194−1201.
(19) Stephenson, R. C.; Clarke, S. Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. *J. Biol. Chem.* **1989**, *264* (11), 6164−70.
(20) Bause, E. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **1983**, *209* (2), 331−6.
(21) Lapko, V. N.; Smith, D. L.; Smith, J. B. Identification of an artifact in the mass spectrometry of proteins derivatized with iodoacetamide. *J. Mass Spectrom.* **2000**, *35* (4), 572−5.
(22) Wada, Y.; Tajiri, M.; Ohshima, S. Quantitation of saccharide compositions of O-glycans by mass spectrometry of glycopeptides and its application to rheumatoid arthritis. *J. Proteome Res.* **2010**, *9* (3), 1367−73.
(23) Viseux, N.; Hronowski, X.; Delaney, J.; Domon, B. Qualitative and quantitative analysis of the glycosylation pattern of recombinant proteins. *Anal. Chem.* **2001**, *73* (20), 4755−4762.
(24) Leymarie, N.; Griffin, P. J.; Jonscher, K.; Kolarich, D.; Orlando, R.; McComb, M.; Zaia, J.; Aguilan, J.; Alley, W. R.; Altmann, F.; Ball, L. E.; Basumallick, L.; Bazemore-Walker, C. R.; Behnken, H.; Blank, M. A.; Brown, K. J.; Bunz, S. C.; Cairo, C. W.; Cipollo, J. F.; Daneshfar, R.; Desaire, H.; Drake, R. R.; Go, E. P.; Goldman, R.; Gruber, C.; Halim, A.; Hathout, Y.; Hensbergen, P. J.; Horn, D. M.; Hurum, D.; Jabs, W.; Larson, G.; Ly, M.; Mann, B. F.; Marx, K.; Mechref, Y.; Meyer, B.; Moginger, U.; Neususs, C.; Nilsson, J.; Novotny, M. V.; Nyalwidhe, J. O.; Packer, N. H.; Pompach, P.; Reiz, B.; Resemann, A.; Rohrer, J. S.; Ruthenbeck, A.; Sanda, M.; Schulz, J. M.; Schweiger-Hufnagel, U.; Sihlbom, C.; Song, E.; Staples, G. O.; Suckau, D.; Tang, H.; Thaysen-Andersen, M.; Viner, R. I.; An, Y.; Valmu, L.; Wada, Y.; Watson, M.; Windwarder, M.; Whittal, R.; Wuhrer, M.; Zhu, Y.; Zou, C. Interlaboratory study on differential analysis of protein glycosylation by mass spectrometry: the ABRF glycoprotein research multi-institutional study 2012. *Mol. Cell. Proteomics* **2013**, *12* (10), 2935−51.
(25) (a) Skehel, J. J.; Stevens, D. J.; Daniels, R. S.; Douglas, A. R.; Knossow, M.; Wilson, I. A.; Wiley, D. C. A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81* (6), 1779−83. (b) An, Y.; McCullers, J. A.; Alymova, I.; Parsons, L. M.; Cipollo, J. F. Glycosylation Analysis of Engineered H3N2 Influenza A Virus Hemagglutinins with Sequentially Added Historically Relevant Glycosylation Sites. *J. Proteome Res.* **2015**, *14* (9), 3957−69.
(26) An, Y.; McCullers, J. A.; Alymova, I.; Parsons, L. M.; Cipollo, J. F. Glycosylation Analysis of Engineered H3N2 Influenza A Virus Hemagglutinins with Sequentially Added Historically Relevant Glycosylation Sites. *J. Proteome Res.* **2015**, *14*, 3957−69.
(27) Yang, H.; Carney, P. J.; Chang, J. C.; Guo, Z.; Villanueva, J. M.; Stevens, J. Structure and receptor binding preferences of recombinant human A(H3N2) virus hemagglutinins. *Virology* **2015**, *477*, 18−31.
(28) Alvarez, S. A cartography of the van der Waals territories. *Dalton Trans.* **2013**, *42*, 8617−36.
(29) Fleury, D.; Wharton, S. A.; Skehel, J. J.; Knossow, M.; Bizebard, T. Antigen distortion allows influenza virus to escape neutralization. *Nat. Struct. Biol.* **1998**, *5* (2), 119−23.
(30) Ha, Y.; Stevens, D. J.; Skehel, J. J.; Wiley, D. C. H5 avian and H9 swine influenza virus haemagglutinin structures: possible origin of influenza subtypes. *EMBO J.* **2002**, *21* (5), 865−75.
(31) Xiong, X.; Coombs, P. J.; Martin, S. R.; Liu, J.; Xiao, H.; McCauley, J. W.; Locher, K.; Walker, P. A.; Collins, P. J.; Kawaoka, Y.; Skehel, J. J.; Gamblin, S. J. Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature* **2013**, *497* (7449), 392−6.
(32) An, Y.; Rininger, J. A.; Jarvis, D. L.; Jing, X.; Ye, Z.; Aumiller, J. J.; Eichelberger, M.; Cipollo, J. F. Comparative glycomics analysis of influenza Hemagglutinin (H5N1) produced in vaccine relevant cell platforms. *J. Proteome Res.* **2013**, *12* (8), 3707−20.
(33) Gavel, Y.; von Heijne, G. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng., Des. Sel.* **1990**, *3* (5), 433−42.
(34) Shi, X.; Jarvis, D. L. Protein N-glycosylation in the baculovirus-insect cell system. *Curr. Drug Targets* **2007**, *8* (10), 1116−25.
(35) (a) Mir-Shekari, S. Y.; Ashford, D. A.; Harvey, D. J.; Dwek, R. A.; Schulze, I. T. The glycosylation of the influenza A virus hemagglutinin by mammalian cells. A site-specific study. *J. Biol. Chem.* **1997**, *272* (7), 4027−36. (b) Zhang, S.; Sherwood, R. W.; Yang, Y.; Fish, T.; Chen, W.; McCardle, J. A.; Jones, R. M.; Yusibov, V.; May, E. R.; Rose, J. K.; Thannhauser, T. W. Comparative characterization of the glycosylation profiles of an influenza hemagglutinin produced in plant and insect hosts. *Proteomics* **2012**, *12* (8), 1269−88. (c) Khatri, K.; Staples, G. O.; Leymarie, N.; Leon, D. R.; Turiak, L.; Huang, Y.; Yip, S.; Hu, H.; Heckendorf, C. F.; Zaia, J. Confident assignment of site-specific glycosylation in complex glycoproteins in a single step. *J. Proteome Res.* **2014**, *13* (10), 4347−55.
(36) (a) Rudd, P. M.; Dwek, R. A. Glycosylation: heterogeneity and the 3D structure of proteins. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32* (1), 1−100. (b) Gavel, Y.; Heijne, G. v. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng., Des. Sel.* **1990**, *3* (5), 433−442.
(37) Job, E. R.; Deng, Y. M.; Tate, M. D.; Bottazzi, B.; Crouch, E. C.; Dean, M. M.; Mantovani, A.; Brooks, A. G.; Reading, P. C. Pandemic H1N1 influenza A viruses are resistant to the antiviral activities of innate immune proteins of the collectin and pentraxin superfamilies. *J. Immunol.* **2010**, *185* (7), 4284−91.
(38) (a) Reading, P. C.; Pickett, D. L.; Tate, M. D.; Whitney, P. G.; Job, E. R.; Brooks, A. G. Loss of a single N-linked glycan from the hemagglutinin of influenza virus is associated with resistance to collectins and increased virulence in mice. *Respir. Res.* **2009**, *10*, 117. (b) Tate, M. D.; Brooks, A. G.; Reading, P. C. Specific sites of N-linked glycosylation on the hemagglutinin of H1N1 subtype influenza A virus determine sensitivity to inhibitors of the innate immune system and virulence in mice. *J. Immunol.* **2011**, *187* (4), 1884−94. (c) Reading, P. C.; Morey, L. S.; Crouch, E. C.; Anders, E. M. Collectin-mediated antiviral host defense of the lung: evidence from influenza virus infection of mice. *J. Virol.* **1997**, *71* (11), 8204−12. (d) Wilson, I. A.; Skehel, J. J.; Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. *Nature* **1981**, *289* (5796), 366−73. (e) Anders, E. M.; Hartley, C. A.; Jackson, D. C. Bovine and mouse serum beta inhibitors of influenza A viruses are mannose-binding lectins. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87* (12), 4485−9.
(39) Wanzeck, K.; Boyd, K. L.; McCullers, J. A. Glycan shielding of the influenza virus hemagglutinin contributes to immunopathology in mice. *Am. J. Respir. Crit. Care Med.* **2011**, *183* (6), 767−73.
(40) (a) Suptawiwat, O.; Boonarkart, C.; Chakritbudsabong, W.; Uiprasertkul, M.; Puthavathana, P.; Wiriyarat, W.; Auewarakul, P. The N-linked glycosylation site at position 158 on the head of hemagglutinin and the virulence of H5N1 avian influenza virus in mice. *Arch. Virol.* **2015**, *160* (2), 409−15. (b) Hillaire, M. L.; van Eijk, M.; van Trierum, S. E.; van Riel, D.; Saelens, X.; Romijn, R. A.; Hemrika, W.; Fouchier, R. A.; Kuiken, T.; Osterhaus, A. D.; Haagsman, H. P.; Rimmelzwaan, G. F. Assessment of the antiviral properties of recombinant porcine SP-D against various influenza A viruses in vitro. *PLoS One* **2011**, *6* (9), e25005. (c) Hartshorn, K. L.; Webby, R.; White, M. R.; Tecle, T.; Pan, C.; Boucher, S.; Moreland, R. J.; Crouch, E. C.; Scheule, R. K. Role of viral hemagglutinin glycosylation in anti-influenza activities of recombinant surfactant protein D. *Respir. Res.* **2008**, *9*, 65.

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412

(41) Velkov, T.; Ong, C.; Baker, M. A.; Kim, H.; Li, J.; Nation, R. L.; Huang, J. X.; Cooper, M. A.; Rockman, S. The antigenic architecture of the hemagglutinin of influenza H5N1 viruses. *Mol. Immunol.* **2013**, *56* (4), 705−19.
(42) Wang, C. C.; Chen, J. R.; Tseng, Y. C.; Hsu, C. H.; Hung, Y. F.; Chen, S. W.; Chen, C. M.; Khoo, K. H.; Cheng, T. J.; Cheng, Y. S.; Jan, J. T.; Wu, C. Y.; Ma, C.; Wong, C. H. Glycans on influenza hemagglutinin affect receptor binding and immune response. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (43), 18137−42.
(43) Ekiert, D. C.; Kashyap, A. K.; Steel, J.; Rubrum, A.; Bhabha, G.; Khayat, R.; Lee, J. H.; Dillon, M. A.; O'Neil, R. E.; Faynboym, A. M.; Horowitz, M.; Horowitz, L.; Ward, A. B.; Palese, P.; Webby, R.; Lerner, R. A.; Bhatt, R. R.; Wilson, I. A. Cross-neutralization of influenza A viruses mediated by a single antibody loop. *Nature* **2012**, *489* (7417), 526−32.
(44) Collins, P. J.; Vachieri, S. G.; Haire, L. F.; Ogrodowicz, R. W.; Martin, S. R.; Walker, P. A.; Xiong, X.; Gamblin, S. J.; Skehel, J. J. Recent evolution of equine influenza and the origin of canine influenza. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (30), 11175−80.
(45) Yang, H.; Carney, P. J.; Chang, J. C.; Guo, Z.; Villanueva, J. M.; Stevens, J. Structure and receptor binding preferences of recombinant human A(H3N2) virus hemagglutinins. *Virology* **2015**, *477*, 18−31.
(46) Zhang, W.; Shi, Y.; Lu, X.; Shu, Y.; Qi, J.; Gao, G. F. An airborne transmissible avian influenza H5 hemagglutinin seen at the atomic level. *Science* **2013**, *340* (6139), 1463−7.
(47) Shore, D. A.; Yang, H.; Balish, A. L.; Shepard, S. S.; Carney, P. J.; Chang, J. C.; Davis, C. T.; Donis, R. O.; Villanueva, J. M.; Klimov, A. I.; Stevens, J. Structural and antigenic variation among diverse clade 2 H5N1 viruses. *PLoS One* **2013**, *8* (9), e75209.
(48) *The PyMOL Molecular Graphics System*, version 1.6.9; Schrödinger, LLC: New York.

DOI: 10.1021/acs.jproteome.6b00175
J. Proteome Res. 2017, 16, 398−412