RESEARCH PAPER

Site-specific glycosylation profile of influenza A (H1N1) hemagglutinin through tandem mass spectrometry

Esteban Cruz^a, Joel Cain^b, Ben Crossett^c, and Veysel Kayser^a

^aFaculty of Pharmacy, The University of Sydney, Sydney NSW, Australia; ^bSchool of Life and Environmental Sciences, The University of Sydney, Sydney NSW, Australia; ^cMass Spectrometry Core Facility, The University of Sydney, Sydney NSW, Australia

ABSTRACT

The study of influenza virus evolution in humans has revealed a significant role of glycosylation profile alterations in the viral glycoproteins – hemagglutinin (HA) and neuraminidase (NA), in the emergence of both seasonal and pandemic strains. Viral antigenic drift can modify the number and location of glycosylation sites, altering a wide range of biological activities and the antigenic properties of the strain. In view of the key role of glycans in determining antigenicity, elucidating the glycosylation profiles of influenza strains is a requirement towards the development of improved vaccines. Sequence-based analysis of viral RNA has provided great insight into the role of glycosite modifications in altering virulence and pathogenicity. Nonetheless, this sequence-based approach can only predict potential glycosylation sites. Due to experimental challenges, experimental confirmation of the occupation of predicted glycosylation sites has only been carried out for a few strains. Herein, we utilized HCD/CID-MS/MS tandem mass spectrometry to characterize the site-specific profile of HA of an egg-grown H1N1 reference strain (A/New Caledonia/20/1999). We confirmed experimentally the occupancy of glycosylation sites identified by primary sequence analysis and determined the heterogeneity of glycan structures. Four glycosylation sequons on the stalk region (N28, N40, N304 and N498) and four on the globular head (N71, N104, N142 and N177) of the protein are occupied. Our results revealed a broad glycan microheterogeneity, i.e., a great diversity of glycan compositions present on each glycosite. The present methodology can be applied to characterize other viruses, particularly different influenza strains, to better understand the impact of glycosylation on biological activities and aid the improvement of influenza vaccines.

ARTICLE HISTORY

Received 14 June 2017
Revised 27 August 2017
Accepted 6 September 2017

KEYWORDS

antigenic site; CID; glycosite modification; H1N1; HCD; hemagglutinin; influenza glycosylation; mass spectrometry

Introduction

Influenza viruses undergo a high rate of antigenic drift, leading to gradual antigenic modifications that are responsible for the persistent emergence of seasonal influenza strains. Occasional antigenic shift (viral reassortments) can also lead to pandemic outbreaks that pose a serious public health threat.^{1,2} The production of immunological memory after a primary exposure (through infection or vaccination) to influenza is essential to trigger an accelerated and efficacious immune response to subsequent infection.^{3,4} For this reason, it is crucial to elucidate the mechanisms that prompt such response and determine the viral antigens that elicit the generation of immunological memory.

The selection of annual strains for influenza vaccines relies on detailed characterization of the genetic and antigenic features of circulating viruses.⁵ Due to their surface exposure, the envelope glycoproteins NA and HA play a prominent role in host immune cell recognition and are considered to be the main antigenic determinants in the virus.^{6,7} Influenza A viruses are thus further classified into subtypes according to the antigenic variants of their surface glycoproteins – 18 HA (H1–H18) and 11 NA (N1–N11) subtypes.⁸ HA is the most abundant protein in the viral envelope and is consequently also the focal point of virus surveillance.⁹

The antigenic sites in HA are comprised mainly of polypeptide regions on the globular head; however, it has been demonstrated that the presence of glycans in the proximity of these sites can affect its biological activity, thereby altering immune cell recognition and receptor binding specificity.^{10,11} HA undergoes N-linked glycosylation (no O-glycosylation has been reported), whereby glycans are attached to asparagine residues within the consensus sequence Asn-Xaa-Ser (Xaa can be any amino acid except proline).^{12,13} Hemagglutinin assembles as a homotrimer that displays a surface-exposed globular head formed by part of the HA1 chain, whereas the stalk region is comprised mostly of α -helix coils and a transmembrane domain from HA2.⁹ Both the stalk and the head region are often heavily glycosylated, and glycan attachment can affect a wide spectra of biological properties, such as immunogenicity, virulence and receptor specificity.^{14,15} Overall, glycan attachment on the stalk region is highly conserved, and glycans on this area play a critical role in correct protein folding and membrane transport.^{16,17} Conversely, the globular head of HA exhibits a considerably higher rate of variation. Most antigenic sites are found on the HA head, therefore modifications on this region usually impair immune recognition.^{18,19}

Influenza subtype H1 has undergone extensive alteration over time in the number and position of glycans attached, mostly on the globular head.²⁰ These modifications are associated with adaptation mechanisms caused by antigenic drift, whereby novel virus subtypes avoid host cell immune recognition by masking antigenic regions through the variation in the localization or number of glycosites.^{20,21}

The consensus sequons required for N-glycosylation make it possible to analyze glycosylation occupancy profiles by searching for potential acceptor sites in the primary sequence of proteins. Sequence-based analysis of potential glycosylation sites on H1 has revealed that in the early stages of virus evolution after the outbreak of the 1918 pandemic, the H1N1 virus subtype predominantly increased the number of glycosylation sites on its globular head. However, following 1950, the number of sites remained somewhat constant and the alteration in position became the prominent feature.²⁰ Strikingly, the H1 of the 2009 swine flu pandemic virus resembles the 1918 pandemic H1 not only in its antigenic epitopes, but also in that they both lack glycosylation sites near the Sa antigenic site.^{22,23} This absence of shielding glycans was an important factor contributing to the pathogenicity of the 1918 pandemic strain.¹⁴

Although sequence-based glycosylation analysis has substantially contributed to the understanding of virus evolution, the information obtained through this method is only limited, as the sequons it identifies are not necessarily occupied.²⁴ For this reason, there remains the need to confirm site occupancy experimentally. Moreover, it is important to assess whether differences in glycoforms exist among strains, and determine if the potential variations have an impact on biological properties. To this end, mass spectrometry based methods, especially those based on collision-induced dissociation (CID), higher-energy collision dissociation (HCD), electron-capture dissociation (ECD) and electron-transfer dissociation (ETD), are particularly suited to analyze protein posttranslational modifications.^{25,26} However, due to the experimental challenges that N-glycoproteomic analysis poses, characterization of site-specific glycosylation profiles has only been carried out for a few strains.

In this study, we employed HCD/CID tandem mass spectrometry to map glycosylation sites and characterize the glycan microheterogeneity (i.e., the subset of glycan structures on each particular position within the protein) of an egg-grown A/New Caledonia/20/1999 H1N1 reference strain. We confirmed the occupancy of 8 glycosylation sites on this reference strain – 4 on the stalk and 4 on the globular head. This profile was consistent to the one obtained through sequence-based analysis using the NetNGlyc server for prediction of glycosylation potential. Moreover, a great diversity of glycan compositions were found on those positions close to the Sa antigenic site, which could have important implications on the antigenicity of the strain.

Results

Prediction of N-glycosylation sites

Prior to tandem mass spectrometry analysis, the amino acid sequence of the A/New Caledonia/20/1999 H1N1 strain was scanned to identify potential N-glycosylation sequons using the

Figure 1. Location of potential N-glycosylation sites within A/New Caledonia/20/1999 H1. Monomeric H2 (PDB entry, 2WR3) was used as a template to generate the homology model of A/New Caledonia/20/1999 H1.

NetNGlyc 1.0 server. Fig. 1 shows the predicted sites and Table 1 displays the potential assigned for glycosylation for each sequon.

A total of 10 sequons were identified within the amino acid sequence of the strain using the NetNGlyc 1.0 server. Four of these sites are located on the globular head (N71, N104, N142 and N177), five on the stalk (N27, N28, N71, N304 and N498) and one on the cytoplasmic tail (N557). With the exception of asparagine N27, all other sequons were predicted to be glycosylated (Fig. 1). The case of N27 is exceptional as its adjacent N28 residue is also a potential acceptor site, but only N28 yields a positive result (≥ 0.5).

Mass spectrometry analysis of electrophoretically-fractionated viral proteins

The first step in sample preparation for mass spectrometry involved a preliminary fractionation of the viral proteins from a whole inactivated virus sample suspension through SDS-PAGE. This isolation step was performed in order to reduce sample complexity prior to nanoLC-MS/MS analysis.

Following gel extraction and trypsin digestion, bands 1–6 (Fig. 2) were subjected to a preliminary nanoLC-MS/MS analysis to identify the bands where HA was present. The whole inactivated virus samples used in this analysis are produced in embryonated chicken eggs, thus the Byonic search was conducted against the strain's proteome database and then against a chicken proteome database. This last step allowed for the detection of remnant chicken proteins from the purification steps of the whole virus stock suspension, which also improves the confidence of viral protein matches in this case.

HA was predominantly detected in band 4 (~60 kDa). This band also displayed the highest HA protein coverage (53%) among all the bands (Table S1). Subunits HA1 and HA2 were detected in bands 5 (~50 kDa) and 6 (~22 kDa), respectively,

Table 1. Estimated glycosylation potential of the glycosylation sequons identified using the NetNGlyc 1.0 server.

Residue number	Position	Sequence	Potential assigned
27	Stalk	NNST	0.4009
28	Stalk	NSTD	0.7964
40	Stalk	NVTV	0.7477
71	Side of head	NCSV	0.7470
104	Side of head	NGTC	0.6300
142	Top of head	NHTV	0.6911
177	Top of head	NLSK	0.7380
304	Stalk	NSSL	0.6707
498	Stalk	NGTY	0.5190
557	Cytoplasmic domain	NGSL	0.6833

and were subjected to glycopeptide analysis. Moreover, peptides from HA were identified in bands 1 and 2, likely due to the presence of hemagglutinin trimers and dimers. The overall protein coverage of HA across all bands was 71.5%.

Glycopeptide analysis

Bands 4–6 (Fig. 2) were then analyzed to characterize the glycosylation profile of hemagglutinin. Precursor glycopeptide ions subjected to HCD yield several diagnostic oxonium ions.

Figure 2. Reducing SDS-PAGE of whole inactivated A/New Caledonia/20/1999 H1N1 virus. The numbering on the virus lane indicates an arbitrarily-assigned order to identify the prominent protein bands obtained through viral fractionation that were subjected to nanoLC-MS/MS analysis.

Diagnostic ions at 204.08 m/z (HexNAc + H) and 366.14 m/z (Hex(1)HexNAc(1) + H) are typically present in HCD spectra of glycosylated peptides. The HCD spectra were used in this study to assign peptide identity. The higher dissociation energy of HCD allows for the fragmentation of the peptide backbone, whereby the presence of **b** and **y** ions in the spectrum is matched by Byonic to the theoretical fragmentation patterns of the tryptic peptides to score the PSMs (peptide spectral matches) and assign identity. Fig. 3A shows a representative HCD spectra displaying the diagnostic oxonium ions at 186.076, 204.087 and 366.140 along with **b** and **y** ions for the reported peptide. Other peptide + fragmented glycan ions (**Y** ions, e.g., peptide+HexNAc, peptide+HexNAc(2)) detected through HCD can also be instrumental to manually validate the peptide identity assigned by the search engine. The peptide backbone mass can be inferred from the trimannosyl ions detected in the HCD/CID-MS/MS spectrum pairs. The theoretical **Y** ions are then matched to the HCD/CID-MS/MS spectrum pair to validate the spectral matches. Byonic derives the mass of the glycan by calculating the mass difference between the precursor ion (peptide + glycan) and the identified tryptic peptide, which is then used to predict the composition of the oligosaccharide by searching this value against the mammalian glycan database within the engine. Byonic is thus able to predict glycoforms to the level of monosaccharide composition.

A complementary CID fragmentation of each precursor ion also allows for the validation of the predicted glycan composition from the Byonic server. Fig. 3B displays the CID spectrum of the same precursor ion as the one used for HCD in Fig. 3A. In this HCD/CID setup, CID causes predominantly the cleavage of glycosidic bonds, resulting in the sequential fragmentation of the glycan. The glycan information derived from low-energy CID is limited, since it does not generate cross-ring fragmentation to elucidate linkage information, nor does it allow to differentiate between glycan isomers in many cases or elucidate topology. Still, the sequential fragmentation enables the elucidation of glycan monosaccharide compositions, by which the prediction from the engine can be manually validated. Consequently, predicted glycan compositions from glycopeptides with a high byonic score for the HCD spectrum were subjected to manual interpretation.

Confirmation of glycosylated residues

Eight out of the nine predicted glycosylation sites were confirmed to be occupied. Four of these sites are on the stalk region (N28, N40, N304 and N498) and four on the globular head (N71, N104, N142 and N177). Fig. 4 displays the location of these sequons on a 3D structure homology model of the hemagglutinin monomer. Prediction for asparagine 557 was positive, however, no peptide was detected for the cytosolic domain nor the transmembrane domain of HA, suggesting that only the ectodomain of the protein can be characterized using this sample preparation. Inasmuch as these domains are unlikely to play an important role in immunogenic response – since they are not exposed to the surface to influence in host cell recognition and have not been related to virulence, obtaining glycosylation information for N557 was deemed unnecessary. Furthermore, it is not expected for glycosylation site N557 to

Figure 3. Representative CID and HCD spectra of H1 glycopeptides. (A) Annotated HCD spectrum of glycopeptide ESSWPNHHTVTGVSASC SHNGK. The spectrum exhibits the diagnostic oxonium ion at 204.08 m/z and the presence of b and y ions derived from backbone fragmentation. (B) Annotated CID spectrum of glycopeptide ESSWPNHHTVTGVSASC SHNGK with an attached glycan with a HexNAc(4)Hex(5)Fuc(1) composition. Sequential glycan fragmentation is predominant in the spectrum and allows for the validation of predicted glycan composition. The presence of a core fucose can be confirmed by the detection of ion $\text{Pep}+2\text{HexNAc-Fuc}$ in both the HCD and CID spectra.

be occupied given its location in the cytosolic region of the protein.

Glycan microheterogeneity profile

Table 2 shows the glycan monomeric compositions found through HCD/CID-MS/MS on each glycosylation site within H1. Overall, glycan heterogeneity was quite broad, especially for sites N142 and N498.

The structural information obtained through CID fragmentation, albeit limited, still allows to discriminate between high mannose, complex or hybrid glycans in some cases. A high mannose structure can be assigned to the monosaccharide compositions in the form $\text{HexNAc}(2)\text{Hex}(5-12)$, as long as the sequential fragmentation confirms the neutral loss of only hexose residues in the antennae from the glycopeptide Y ion series. Complex and hybrid glycans can be distinguished when diagnostic Y ions are detected, such as $\text{Pep}+\text{HexNAc}(2)\text{Hex}(4)$, for compositions with more than 2 HexNAc residues. In such cases, at least one of the HexNAc residues must be in the antennae, thus ruling out a high mannose glycan. Additionally, the $\text{Pep}+\text{HexNAc}(2)\text{Hex}(4)$ fragment confirms that one hexose must be attached to one glucose residue from the trimannosyl core. Both these features combined can only occur in hybrid glycans, since only HexNAc residues are attached to the try-mannosyl core in complex glycans. Fig. 3B shows a representative CID spectrum with two diagnostic ions ($\text{Pep}+\text{HexNAc}(3)$

$\text{Hex}(5)+3$ and $\text{Pep}+\text{HexNAc}(3)\text{Hex}(5)\text{Fuc}(1)+3$) for a hybrid structure.

Other diagnostic ions shed further structural information. A 512 m/z peak in the lower m/z range of the CID spectra corresponds to a $\text{Hex}(1)\text{HexNAc}(1)\text{Fuc}(1)$ glycan B ion, characteristic of an outer arm fucose. Conversely, core fucosylation can be determined from the detection of Y ions such as $\text{Pep}+\text{HexNAc}(1)\text{Fuc}(1)$ and $\text{Pep}+\text{HexNAc}(2)\text{Fuc}(1)$. The latter fragments can also be observed in some HCD spectra to further confirm fucose attachment to the chitobiose core. Special consideration was given to glycans with a mass consistent with isobaric glycan compositions containing either $\text{Fuc}+\text{NeuGc}$ or $\text{Hex}+\text{NeuAc}$. In these cases, not all spectra contained sufficient information to derive the correct composition. Nonetheless, diagnostic ions (NeuAc (292.10 m/z), $\text{NeuAc-H}_2\text{O}$ (274.09 m/z), NeuGc (308.09 m/z), $\text{NeuGc-H}_2\text{O}$ (290.09 m/z)) were present in some HCD spectra and were used to identify the correct composition. Moreover, glycan B fragments corresponding to HexNAc-Gal-NeuAc (657 m/z) or HexNAc-Gal-NeuGc (673 m/z) detected in CID were also present in some spectra.

Residues N142 and N498 displayed a great variety of glycan monomeric compositions. Conversely, for sites N28, N40, N71 and N104, only a very limited number of compositions were detected. Overall, complex and hybrid glycans were predominantly found. Nonetheless, high mannose glycans are also present in N142 and N177, both located on the top of the globular head. Fucosylation was observed in most glycopeptides and was confirmed by the presence of fucose residues in the CID

Figure 4. Location of confirmed glycosylation sites in relation to the most relevant antigenic sites on A/New Caledonia/20/1999 H1. (A) View of the globular head of H1 from the top. (B) View of H1 from the side to highlight the distribution of glycosites along the stalk region and the side of the globular head. Monomeric H2 (PDB entry, 2WR3) was used as a template to generate the homology model of A/New Caledonia/20/1999 H1.

spectra or diagnostic ions in the HCD spectra. Core fucoses are predominant, but outer arm fucoses were also found in some glycoforms. Highly bulky complex glycans were found on N498.

Discussion

Sequence-based analysis has previously been employed by others to study the variation in the glycosylation profile of several strains during virus evolution.^{27,28} Although N-glycosylation prediction contributes greatly to the understanding of glycosite variation across strains, this method only provides a prediction that relies on primary sequence context, and the actual occupancy of glycosites must be confirmed experimentally. Moreover, it is known that the location of the sequon can affect glycan maturation, thereby resulting in different sets of glycoforms for each site depending on the accessibility of N-glycosyltransferases.²⁹ To this end, we used mass spectrometry analysis to confirm the occupancy of glycosites predicted by the NetNGlyc 1.0 Server for the A/New Caledonia/20/1999 H1 protein. Others have shown that hemagglutinin from egg-grown viruses and human cell lines display the same occupancy profile.³⁰ Therefore, the results obtained in this study are relevant to the understanding of influenza virus adaptation in humans.

The occupancy profile obtained through HCD/CID-MS/MS of tryptic glycopeptides was consistent with the prediction by

the NetNGlyc 1.0 server. All eight sequons confirmed experimentally had been predicted by the server. Indeed, N28 (and not N27), was shown to be glycosylated as was predicted. The transfer of N-glycan precursors to asparagine residues in endoplasmic reticulum (ER) bound ribosomes is affected by conformational constraints as well as inaccessibility due to steric hindrance from nearby residues. To this end, since N28 is occupied, it is unlikely that an oligosaccharyltransferase can act upon asparagine N27. The positive prediction for site N557 was not accurate, due to the location of the residue within the cytoplasmic domain of the protein. Nonetheless, such inaccuracies can easily be identified to prevent misleading interpretations when performing sequence-based analyses.

Glycosites N142 and N177 are of particular relevance for the study of immunogenicity and virulence. These residues are both located on the receptor binding site (RBS) and the attachment of oligosaccharides can mask the antigenic Sa region. Remarkably, both acceptor sites were absent in the 1918 and 2009 pandemic strains. Wei *et al.* showed that mice display cross-neutralization of 1918 and 2009 pandemic viruses after vaccination with the 1918 strain, whereas such antibodies did not protect against seasonal strains – providing evidence that the introduction of N-glycosylation sites on the vicinity of the RBS can mask antigenic recognition by pandemic antibodies.³¹

A vast heterogeneity of glycan compositions was found on N142 (Table 2), which is even more notable taking into account that the predicted compositions can encompass several glycan isoforms; meaning that the actual heterogeneity is presumably broader. Among the predicted oligosaccharides attached to N142 are glycan compositions consistent with complex, high mannose and hybrid glycans. The vast complexity encountered on this asparagine residue can be attributed to the high surface exposure, providing accessibility for N-glycan processing and maturation (Fig. 4).

In the case of glycosite N177, a much narrower variety of glycoforms was detected. However, the intensities for these glycopeptides were much lower compared to those glycopeptides containing site N142. Therefore, many glycopeptides for N177 could have been suppressed and missed in the analysis, which constitutes a limitation of the method. Still, the MS/MS analysis predicted the occupancy of high mannose, hybrid and complex glycoforms.

The presence of terminal sialic acid residues in several of the predicted glycoforms attached to both N142 and N177 could play an important role in the immunogenicity of the strain. Ohuchi *et al.* suggested that the presence of terminal NeuAc and NeuGc on HA glycans in the vicinity of the receptor binding sites may impair binding to sialic acid-containing receptors.³² Although the effect of the presence of sialic acid residues on immune cell recognition in this case is unknown, the bulkiness and negative charge of NeuAc and NeuGc would likely further impair an immune response. Overall, terminal sialylation of both cell surfaces and glycoproteins has been shown to decrease antigenicity.^{33,34} Therefore, the attachment of negatively charged terminal residues in embryonated-egg vaccine production could be a drawback in terms of HA antigenicity compared to other recombinant hemagglutinin based vaccines, where sialylation can be controlled or non-existent, as is the case in the baculovirus expression system.³⁵

Table 2. Microheterogeneity of A/New Caledonia/20/1999 H1.

Position of glycan attachment within the protein sequence	Predicted glycan composition	Ion detected (m/z)	Glycan mass (Da)	$\Delta_{m/z}$ (Da)
Site 1: Asparagine 28 (NSTD) Stalk	HexNAc(5)Hex(3)	1,389.59 (+6)	1,501.55	−0.3369
Site 2: Asparagine 40 (NVTV) Stalk	HexNAc(5)Hex(8)Fuc(1)	1,474.63 (+3)	2,457.87	0.0079
	HexNAc(5)Hex(6)	988.68 (+4)	1,987.71	0.00022
Site 3: Asparagine 71 (NCSV) Side of the head	HexNAc(3)Hex(5)NeuGc(1)	1,564.69 (+3)	1,726.59	−0.36158
Site 4: Asparagine 104 (NGTC) Side of the head	HexNAc(7)Hex(3)Fuc(1)	1,389.85 (+4)	2,053.77	0.27418
	HexNAc(2)Hex(3)	1,081.43 (+4)	892.31	−0.001997
Site 5: Asparagine 142 (NHTV) Top of the head	HexNAc(6)Hex(6)Fuc(1)	1,149.22 (+4)	2,336.85	0.00096
	HexNAc(6)Hex(5)	1,054.67 (+4)	2,028.74	0.24239
	HexNAc(6)Hex(4)NeuAc(1)	1,104.64 (+4)	2,157.78	0.25827
	HexNAc(6)Hex(4)Fuc(1)	1,068.44 (+4)	2,012.74	0.24967
	HexNAc(6)Hex(4)	953.60 (+5)	1,866.68	0.00167
	HexNAc(6)Hex(3)Fuc(1)NeuAc(1)	1,466.57 (+3)	2,141.78	−0.35536
	HexNAc(6)Hex(3)Fuc(2)	1,046.42 (+4)	1,996.75	−0.00844
	HexNAc(6)Hex(3)	991.16 (+4)	1,704.63	0.00043
	HexNAc(5)Hex(6)Fuc(2)	1,134.72 (+4)	2,279.82	0.00087
	HexNAc(5)Hex(6)Fuc(1)	1,098.45 (+4)	2,133.77	0.00237
	HexNAc(5)Hex(6)	1,061.70 (+4)	1,987.71	−0.24352
	HexNAc(5)Hex(5)Fuc(3)	1,130.96 (+4)	2,263.83	0.00054
	HexNAc(5)Hex(5)Fuc(2)	1,094.45 (+4)	2,117.77	0.00123
	HexNAc(5)Hex(5)Fuc(1)	1,057.94 (+4)	1,971.71	0.0035
	HexNAc(5)Hex(5)	1,021.42 (+4)	1,825.66	−0.00022
	HexNAc(5)Hex(4)NeuAc(1)	1,035.43 (+4)	1,954.70	−0.24191
	HexNAc(5)Hex(4)Fuc(1)	1,356.23 (+3)	1,809.66	0.00052
	HexNAc(5)Hex(4)	980.91 (+4)	1,663.60	−0.00032
	HexNAc(5)Hex(4)Fuc(2)	1,955.72 (+4)	1,930.69	0.00039
	HexNAc(5)Hex(3)Fuc(1)NeuGc(1)	1,404.24 (+3)	1,954.70	−0.32956
	HexNAc(5)Hex(3)Fuc(1)	976.91 (+4)	1,647.61	−0.00067
	HexNAc(4)Hex(7)	1,402.13 (+3)	1,948.68	0.32814
	HexNAc(4)Hex(6)NeuGc(1)	1,450.25 (+4)	2,091.72	0.33191
	HexNAc(4)Hex(6)NeuAc(1)	1,444.92 (+3)	2,075.72	0.33262
	HexNAc(4)Hex(6)Fuc(2)	1,084.19 (+4)	2,076.75	0.00188
	HexNAc(4)Hex(6)Fuc(1)	1,047.93 (+4)	1,930.69	0.2478
	HexNAc(4)Hex(6)	1,324.20 (+3)	1,784.63	0.00147
	HexNAc(4)Hex(5)NeuGc(1)	1,047.18 (+4)	1,926.67	0.00257
	HexNAc(4)Hex(5)NeuAc(1)	1,367.21 (+3)	1,931.67	0.3165
	HexNAc(4)Hex(5)Fuc(3)	1,079.95 (+4)	2,060.75	0.00052
	HexNAc(4)Hex(5)Fuc(2)	1,043.68 (+4)	1,914.69	0.00044
	HexNAc(4)Hex(5)Fuc(1)	1,007.41 (+4)	1,768.63	0.24923
	HexNAc(4)Hex(5)	970.65 (+4)	1,622.58	−0.00094
	HexNAc(4)Hex(4)Fuc(2)NeuGc(1)	1,061.69 (+4)	2,059.73	−0.24548
	HexNAc(4)Hex(4)Fuc(2)NeuAc(1)	1,057.69 (+4)	2,043.74	−0.24382
	HexNAc(4)Hex(4)Fuc(1)NeuAc(1)	1,021.17 (+4)	1,897.68	−0.24484
	HexNAc(4)Hex(4)NeuAc(1)	1,003.16 (+4)	1,751.62	0.2533
	HexNAc(4)Hex(4)Fuc(2)	1,003.17 (+4)	1,752.64	0.00223
	HexNAc(4)Hex(4)Fuc(1)	901.58 (+5)	1,606.58	0.00127
	HexNAc(4)Hex(4)	929.89 (+4)	1,460.52	0.00082
	HexNAc(4)Hex(3)Fuc(3)	981.16 (+4)	1,736.64	−0.00291
	HexNAc(4)Hex(3)Fuc(1)	926.14 (+4)	1,444.53	0.00019
	HexNAc(3)Hex(6)Fuc(1)NeuGc(1)	1,431.24 (+3)	2,034.70	0.32942
	HexNAc(3)Hex(6)Fuc(1)NeuAc(1)	1,402.56 (+3)	2,018.70	0.33363
	HexNAc(3)Hex(6)NeuGc(1)	1,037.41 (+4)	1,888.64	0.24807
	HexNAc(3)Hex(6)Fuc(1)	996.91 (+4)	1,727.61	0.00239
	HexNAc(3)Hex(6)	1,280.18 (+3)	1,581.55	−0.01264
	HexNAc(3)Hex(5)Fuc(1)	956.39 (+4)	1,565.56	0.00357
	HexNAc(3)Hex(5)	919.8 (+4)	1,419.50	−0.00045
	HexNAc(3)Hex(4)Fuc(2)	915.88 (+4)	1,549.56	0.00085
	HexNAc(3)Hex(4)Fuc(1)	915.64 (+4)	1,403.50	0.00039
	HexNAc(3)Hex(3)Fuc(1)	875.37 (+4)	1,241.45	−0.00017
	HexNAc(3)Hex(3)	838.85 (+4)	1,095.39	0.00015
	HexNAc(2)Hex(5)	869.11 (+4)	1,216.42	0.00035
	HexNAc(2)Hex(6)	909.62 (+4)	1,378.47	0.00027
	HexNAc(2)Hex(7)	950.13 (+4)	1,540.52	−0.00177
	HexNAc(2)Hex(8)	1,320.51 (+4)	1,702.58	−0.01245
Site 6: Asparagine 177 (NLSK) Top of the head	HexNAc(5)Hex(4)Fuc(1)NeuAc(1)	1,553.65 (+2)	2,100.76	−0.0034
	HexNAc(4)Hex(6)	1,166.17 (+3)	1,784.63	0.00055
	HexNAc(2)Hex(8)	1,354.56 (+2)	1,702.58	0.00196
	HexNAc(2)Hex(7)	1,273.53 (+2)	1,540.52	0.00127
	HexNAc(2)Hex(6)	1,193.00 (+2)	1,378.47	0.0015
	HexNAc(2)Hex(5)	1,111.98 (+2)	1,216.42	0.00069
	HexNAc(2)	585.06 (+4)	406.15	−0.00064
Site 7: Asparagine 304 (NSSL) Stalk	HexNAc(6)Hex(4)	1,657.73 (+4)	1,866.68	0.25101
	HexNAc(6)Hex(3)	1,621.22 (+4)	1,704.63	0.25377

(Continued on next page)

Site 8: Asparagine 498 (NGTY) Stalk

HexNAc(4)Hex(5)NeuAc(1)	1,686.49 (+4)	1,913.67	−0.24401
HexNAc(4)Hex(4)NeuGc(1)	1,649.97 (+4)	1,767.61	−0.25517
HexNAc(3)Hex(6)NeuAc(1)	1,659.22 (+4)	1,872.64	0.25161
HexNAc(3)Hex(4)Fuc(1)NeuAc(1)	1,614.70 (+4)	1,694.60	0.2467
HexNAc(8)Hex(8)Fuc(1)	1,342.18 (+3)	3,067.11	−0.00606
HexNAc(7)Hex(7)Fuc(1)NeuAc(2)	1,414.88 (+3)	3,284.17	0.3354
HexNAc(6)Hex(6)	1,050.08 (+3)	2,190.79	0.00079
HexNAc(6)Hex(5)Fuc(1)NeuAc(2)NeuGc(1)	1,341.19 (+3)	3,064.07	0.02
HexNAc(6)Hex(3)Fuc(2)NeuGc(1)	1,471.89 (+3)	2,303.84	−0.34158
HexNAc(5)Hex(7)	1,415.21 (+3)	2,149.76	−0.33692
HexNAc(5)Hex(5)Fuc(1)NeuAc(1)	1,458.23 (+3)	2,262.81	−0.33147
HexNAc(5)Hex(5)NeuGc(1)	1,409.55 (+3)	2,132.75	−0.32933
HexNAc(5)Hex(5)Fuc(1)	1,465.08 (+2)	1,971.71	0.00223
HexNAc(5)Hex(5)	1,307.84 (+3)	1,825.66	0.00225
HexNAc(5)Hex(4)Fuc(1)NeuAc(1)	1,401.21 (+3)	2,100.76	−0.32961
HexNAc(5)Hex(4)Fuc(1)NeuGc(1)	1,409.55 (+3)	2,116.75	−0.32763
HexNAc(5)Hex(4)NeuGc(1)	1,384.87 (+3)	1,970.69	−0.00099
HexNAc(5)Hex(3)Fuc(1)NeuGc(1)	1,355.53 (+3)	1,945.70	−0.32857
HexNAc(4)Hex(6)	1,293.83 (+3)	1,784.63	−0.00351
HexNAc(4)Hex(5)Fuc(1)NeuGc(1)	1,390.54 (+4)	2,075.72	−0.32938
HexNAc(4)Hex(5)NeuGc(1)	1,366.19 (+3)	1,929.67	0.00128
HexNAc(4)Hex(5)Fuc(3)	1,386.20 (+3)	2,060.75	−0.00084
HexNAc(4)Hex(5)Fuc(2)	1,006.74 (+3)	1,914.69	0.00054
HexNAc(4)Hex(5)Fuc(1)	1,288.83 (+3)	1,768.63	−0.00049
HexNAc(4)Hex(5)	1,101.78 (+3)	1,622.58	0.00172
HexNAc(4)Hex(4)Fuc(2)NeuAc(1)	1,385.21 (+3)	2,043.74	−0.32697
HexNAc(4)Hex(4)Fuc(1)NeuGc(1)	1,360.86 (+3)	1,913.67	0.32938
HexNAc(4)Hex(4)Fuc(1)NeuAc(1)	1,336.52 (+3)	1,987.68	−0.33167
HexNAc(4)Hex(4)Fuc(1)	943.87 (+4)	1,606.58	−0.00836
HexNAc(3)Hex(6)Fuc(1)NeuGc(1)	1,382.18 (+3)	2,034.70	−0.34415
HexNAc(3)Hex(6)Fuc(1)	1,343.03 (+2)	1,727.61	0.00136
HexNAc(3)Hex(5)NeuGc(1)	1,298.49 (+3)	1,726.59	−0.00482
HexNAc(3)Hex(4)Fuc(1)NeuAc(1)	1,287.83 (+2)	1,694.60	0.00233
HexNAc(3)Hex(4)NeuAc(1)	1,239.15 (+3)	1,548.54	0.00166
HexNAc(3)Hex(4)Fuc(2)	1,239.15 (+3)	1,549.56	−0.01108
HexNAc(3)Hex(4)	1,107.95 (+2)	1,257.44	0.00098
HexNAc(2)Hex(8)	1,330.49 (+2)	1,702.58	−0.02006

Glycosites N71 and N104 are located on the side of the globular head (Fig. 4). In the case of N71, only one glycan with a composition consistent with a non-fucosylated hybrid glycan was found (Table 2). For N104, one glycan with a complex-like composition was detected. Others have proposed that the occupancy of site N104 might be sterically hindered by glycan attachment on N71 due to their close proximity (Fig. 4).²⁰ Moreover, it has been argued that since both glycosites are able to shield antigenic site Ca2, glycosylation on N71 could render the occupancy of N104 unnecessary to mask antigenic site Ca2.²⁰ It was confirmed that both N71 and N104 are indeed glycosylated in egg-grown viruses. It was unexpected, however, to find such a limited set of glycan compositions in these sites, in contrast to other positions such as N142, N177 and N498. As discussed previously, these results are potentially due to a limited detection of the glycopeptides containing these acceptor sites. However, the close proximity of both residues also suggests that glycan processing might be impaired due to steric hindrance.

The detection of highly bulky complex glycans on N498 (e.g., HexNAc(8)Hex(8)Fuc(1) and HexNAc(7)Hex(7)Fuc(1)NeuAc(2)) suggests the presence of poly-N-acetylactosamine groups, which are commonly found on membrane-proximal glycosites.^{36,37} The close proximity of N498 to the viral membrane (Fig. 4B) does not seem to decrease the degree of glycan maturation on this asparagine residue, resulting in a high complexity of structures.

In this mass spectrometry setup, some tryptic glycopeptides produce considerably higher intensities than others. This causes

that more information on glycan microheterogeneity can be gathered on those glycopeptides that ionize with a higher efficiency and whose *m/z* values are better detected under these experimental conditions, whereas only limited information can be derived from other regions. This was likely the cause of the limited glycopeptide detection for sites N28, N40, N71 and N104. Unfortunately, this drawback prevents an appropriate comparison of glycan microheterogeneity among glycosites. Obtaining a more comprehensive profile is required to better understand the role of glycan structures on the biological properties of the virus. Still, these results allowed to confirm glycan attachment on these sites.

An advantage of utilizing reducing SDS-PAGE as a preliminary fractionation step is that analysing HA0, HA1 and HA2 separately provides a more comprehensive characterization of glycan microheterogeneity, owing to the fact that the lower overall presence of non-glycosylated peptides in HA1 or HA2 isolated fractions (compared to HA0) could allow the detection of certain glycopeptides with low total ion current (TIC). In fact, most of the distinct glycan compositions detected on site N498 come from the analysis of HA2 and were not present in the analysis of HA0 (band 4 in Fig. 2). This is likely due to the large presence of non-glycosylated peptides in HA0, which can suppress the signal of N498 glycopeptides.

The experimental confirmation of glycosylation sites is highly important in the study of influenza viral evolution. Furthermore, there is evidence that egg and mammalian cell derived HA display the same occupancy profile.³⁰ Consequently, characterizing further egg-derived strains with this

method would also yield site occupancy information relevant to other approved production platforms, such as cell-based and recombinant technologies. Currently, egg-based vaccines comprise the overwhelming majority of available influenza vaccine products. However, the FDA has recently approved the cell-based Flucelvax (Seqirus), produced in Madin-Darby Canine Kidney (MDCK) epithelial cells, and the recombinant Flublok (Protein Sciences Corporation), produced in insect cells.^{38–40} It is noteworthy to mention that the glycan structures found on egg-derived viruses can differ significantly to those present on viruses that replicate in alternative cell substrates.^{30,41,42}

The results presented in this study confirmed the glycan occupancy profile of A/New Caledonia/20/1999 hemagglutinin, validated the prediction by the NetNGlyc 1.0 Server and characterized the site-specific structural heterogeneity of the attached glycans. The mass spectrometry approach employed was able to furnish information on sequon occupancy and glycan structures in a single run. This efficient strategy can be extended to study further influenza glycoproteins in order to attain deeper knowledge of the role of N-glycosylation in viral adaptation mechanisms. Site-specific mass spectrometry methods, however, typically provide only partial information on the structure of the oligosaccharides. To this regard, additional work should be carried out to elucidate more detailed structural features to assess the impact of glycoforms on parameters such as antigen masking, immunogenicity, SA receptor binding affinity and viral membrane resistance to surfactant treatment. The latter is highly relevant for split-virus vaccine manufacturing, where a surfactant is employed to disrupt the viral membrane. Gaining further insight into the glycosylation characteristics of influenza viruses can therefore potentially improve the manufacturing process of split-virus vaccines, the selection of strains for the composition of seasonal vaccines and the development of recombinant products.

Materials and methods

N-glycosylation sequon analysis of A/New Caledonia/20/1999 HA

The amino acid sequence for the A/New Caledonia/20/1999 HA was obtained from the National Institute of Allergy and Infectious Disease (NIAID) Influenza Research Database (IRD)⁴³ through the web site at <http://fludb.org>. NetNGlyc 1.0 Server²⁴ was used for the prediction of potential glycosylation sites. NetNGlyc 1.0 Server utilizes several artificial neural networks that are able to predict the probability of an N-glycosylation motif to be occupied by analyzing the adjacent primary sequence. The server sets a default threshold of 0.5, whereby a higher value indicates a predicted glycosylated residue.

Homology modelling of A/New Caledonia/20/1999 HA

The homology model of A/New Caledonia/20/1999 HA was created with Prime (Schrödinger) using H2 (PDB entry 2WR3, chain A) as a template. The molecular structures were generated using VMD 1.9.3.

Hemagglutinin isolation through polyacrylamide electrophoresis and in-gel trypsin digestion

Egg-grown A/New Caledonia/20/1999 virus samples were provided by Sanofi-Pasteur (PA-US). Viral proteins were separated by means of reducing SDS-PAGE by running whole virus samples (suspended in PBS) on a 10% home-made bis-acrylamide gel. The virus suspension was diluted with SDS-PAGE sample buffer containing 1% 2-Mercaptoethanol and 2% SDS. The samples were boiled at 95 °C for 10 minutes before loading onto the gel. The gel was stained with Coomassie Brilliant Blue R-250 overnight, and the bands corresponding to HA0 monomers, HA1 and HA2 were excised from the gel. The bands were subsequently destained using 40% (v/v) acetonitrile and dried using a vacuum concentrator prior to incubation with 5 µL 12 mg/mL trypsin (Promega) at 4 °C for 1 hour. 20 µL of NH₄HCO₃ 50 mM pH 6.8 were added and incubated overnight to extract the proteins from the gel.

Analysis of tryptic peptides and glycopeptides by nanoRPLC-MS/MS

The tryptic peptide fraction was first desalted using a C18 ZipTip (Millipore) and eluted with 80% (v/v) acetonitrile. The desalted peptides were then dried by vacuum centrifugation and resuspended in 0.1% formic acid (FA) prior to loading onto the nanoRPLC column. The glycopeptide solution was loaded onto an in-house packed 20 cm × 75 µm Reprosil-Pur C18AQ (3 µm, 120 Å; Dr. Maisch GmbH) column/emitter using an easy nanoLC II HPLC (Proxeon). HPLC separation was carried out over 140 minutes at a flow rate of 250 nL/min using a 0–40% solvent B gradient, where solvent A consists in 0.1% (v/v) formic acid and solvent B is 90% (v/v) acetonitrile and 0.1% formic acid. Instrument parameters were set up as follows: source voltage = 2.0 kV, S-lens RF level = 68%, and capillary temperature = 275 °C. MS analysis was performed with an Orbitrap Velos Pro MS (Thermo Scientific). The initial MS scan was collected in the Orbitrap mass analyzer (300–1,700 m/z; MS AGC = 1×10^6) with a resolution of 30,000 at 300 m/z. The three most intense precursors were then selected for fragmentation using either data-dependent higher-energy collisional dissociation (HCD) or collision induced (CID) fragmentation. HCD parameters were as follows: activation time = 0.1 ms, resolution = 7,500, maximum injection time = 500 ms, dynamic exclusion = enabled with repeat count 1, normalized energy = 45, exclusion duration = 60 s, default charge state = 2, and MSn AGC $2 \times 10,000$. CID parameters as follows: activation time = 10 ms, maximum injection time = 300 ms, dynamic exclusion = enabled with repeat count 1, normalized energy = 35, exclusion duration = 30 s, default charge state = 2, and MSn AGC = 2×10^4 .

Analysis of MS/MS spectra of intact N-glycopeptides

The MS data was processed with Proteome Discoverer 2.0 (Thermo Scientific), with HCD scans searched using Byonic (Protein Metrics Inc.). The search was performed against an influenza A virus (A/New Caledonia/20/1999) and Chicken protein sequence databases using the following settings: Full

trypsin specificity with a maximum of two missed cleavages, an MS tolerance of 20 ppm and an MS2 tolerance of 0.05 Da. A total of 3 common1 modifications (deamidation of asparagine and glutamine – +0.984016, cysteine propionamide – +71.037114 and asparagine glycosylation) were allowed and 1 total rare modification (oxidation of methionine – +15.994915 Da). The N-glycan database (309 mammalian no sodium) available in the Byonic search engine was chosen for N-linked modification. HCD spectra were manually inspected to confirm the matched glycopeptide fragmentation pattern and the presence of diagnostic HexNAc oxonium ions. CID spectra were manually inspected to validate the predicted glycan composition by manual annotation.

Disclosure of potential conflicts of interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors would like to thank Sanofi-Pasteur for providing the virus samples. We would also like to thank Ben Parker and Stuart Cordwell for valuable discussions.

Funding

The authors would like to thank Sanofi-Pasteur for financial support. We thank the Faculty of Pharmacy of The University of Sydney for financial contribution. EC acknowledges the Ministry of Science, Technology and Telecommunications of the Republic of Costa Rica for postgraduate scholarship.

References

- [1] Taubenberger JK, Kash JC. Influenza Virus Evolution, Host Adaptation, and Pandemic Formation. *Cell Host & Microbe*. 2010;7(6):440-51. doi:10.1016/j.chom.2010.05.009
- [2] Nelson MI, Holmes EC. The evolution of epidemic influenza. *Nat Rev Genet*. 2007;8(3):196-205. doi:10.1038/nrg2053. PMID:17262054
- [3] Potter CW, Oxford JS. Determinants of immunity to influenza infection in man. *Br Med Bull*. 1979;35(1):69-75. doi:10.1093/oxfordjournals.bmb.a071545. PMID:367490
- [4] Mancini N, Solforosi L, Clementi N, De Marco D, Clementi M, Burioni R. A potential role for monoclonal antibodies in prophylactic and therapeutic treatment of influenza. *Antiviral Res*. 2011;92(1):15-26. doi:10.1016/j.antiviral.2011.07.013. PMID:21798290
- [5] Stohr K, Bucher D, Colgate T, Wood J. Influenza virus surveillance, vaccine strain selection, and manufacture. *Methods Mol Biol*. 2012;865:147-62. doi:10.1007/978-1-61779-621-0_9. PMID:22528158
- [6] Johansson BE, Bucher DJ, Kilbourne ED. Purified influenza virus hemagglutinin and neuraminidase are equivalent in stimulation of antibody response but induce contrasting types of immunity to infection. *J Virol*. 1989;63(3):1239-46. PMID:2915381
- [7] Johansson BE, Moran TM, Bona CA, Popple SW, Kilbourne ED. Immunologic response to influenza virus neuraminidase is influenced by prior experience with the associated viral hemagglutinin. II. Sequential infection of mice simulates human experience. *J Immunol*. 1987;139(6):2010-4. PMID:3624874
- [8] Virk RK, Gunalan V, Tambyah PA. Influenza infection in human host: Challenges in making a better influenza vaccine. *Expert Rev Anti Infect Ther*. 2016;14(4):365-75. doi:10.1586/14787210.2016.1155450. PMID:26885890
- [9] Sriwilaijaroen N, Suzuki Y. Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. *Proc Jpn Acad Ser B Phys Biol Sci*. 2012;88(6):226-49. doi:10.2183/pjab.88.226. PMID:22728439
- [10] Wang CC, Chen JR, Tseng YC, Hsu CH, Hung YF, Chen SW, Chen CM, Khoo KH, Cheng TJ, Cheng YS, et al. Glycans on influenza hemagglutinin affect receptor binding and immune response. *Proc Natl Acad Sci U S A*. 2009;106(43):18137-42. doi:10.1073/pnas.0909696106. PMID:19822741
- [11] Wanzeck K, Boyd KL, McCullers JA. Glycan shielding of the influenza virus hemagglutinin contributes to immunopathology in mice. *Am J Respir Crit Care Med*. 2011;183(6):767-73. doi:10.1164/rccm.201007-1184OC. PMID:20935106
- [12] Chen W, Zhong Y, Qin Y, Sun S, Li Z. The Evolutionary Pattern of Glycosylation Sites in Influenza Virus (H5N1) Hemagglutinin and Neuraminidase. *PLoS ONE*. 2012;7(11):e49224. doi:10.1371/journal.pone.0049224. PMID:23133677
- [13] Kim JI, Park MS. N-linked glycosylation in the hemagglutinin of influenza A viruses. *Yonsei Med J*. 2012;53(5):886-93. doi:10.3349/ymj.2012.53.5.886. PMID:22869469
- [14] Sun X, Jayaraman A, Maniprasad P, Raman R, Houser KV, Pappas C, Zeng H, Sasisekharan R, Katz JM, Tumpey TM. N-Linked Glycosylation of the Hemagglutinin Protein Influences Virulence and Antigenicity of the 1918 Pandemic and Seasonal H1N1 Influenza A Viruses. *J Virol*. 2013;87(15):8756-66. doi:10.1128/JVI.00593-13. PMID:23740978
- [15] de Vries RP, de Vries E, Bosch BJ, de Groot RJ, Rottier PJ, de Haan CA. The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity. *Virology*. 2010;403(1):17-25. doi:10.1016/j.virol.2010.03.047. PMID:20441997
- [16] Zhang X, Chen S, Yang D, Wang X, Zhu J, Peng D, Liu X. Role of stem glycans attached to haemagglutinin in the biological characteristics of H5N1 avian influenza virus. *J Gen Virol*. 2015;96(Pt 6):1248-57. doi:10.1099/vir.0.000082. PMID:25667326
- [17] Roberts PC, Garten W, Klenk HD. Role of conserved glycosylation sites in maturation and transport of influenza A virus hemagglutinin. *J Virol*. 1993;67(6):3048-60. PMID:8497042
- [18] Herve PL, Lorin V, Jouvion G, Da Costa B, Escriviou N. Addition of N-glycosylation sites on the globular head of the H5 hemagglutinin induces the escape of highly pathogenic avian influenza A H5N1 viruses from vaccine-induced immunity. *Virology*. 2015;486:134-45. doi:10.1016/j.virol.2015.08.033. PMID:26433051
- [19] Medina RA, Stertz S, Manicassamy B, Zimmermann P, Sun X, Albrecht RA, Uusi-Kerttula H, Zagordi O, Belshe RB, Frey SE, et al. Glycosylations in the globular head of the hemagglutinin protein modulate the virulence and antigenic properties of the H1N1 influenza viruses. *Sci Transl Med*. 2013;5(187):187ra70. doi:10.1126/scitranslmed.3005996. PMID:23720581
- [20] Sun S, Wang Q, Zhao F, Chen W, Li Z. Glycosylation Site Alteration in the Evolution of Influenza A (H1N1) Viruses. *PLoS ONE*. 2011;6(7):e22844. doi:10.1371/journal.pone.0022844. PMID:21829533
- [21] Tate MD, Job ER, Deng YM, Gunalan V, Maurer-Stroh S, Reading PC. Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses*. 2014;6(3):1294-316. doi:10.3390/v6031294. PMID:24638204
- [22] Xu R, Ekiert DC, Krause JC, Hai R, Crowe JE, Wilson IA. Structural basis of pre-existing immunity to the 2009 H1N1 pandemic influenza virus. *Science (New York, NY)*. 2010;328(5976):357-60. doi:10.1126/science.1186430
- [23] Igarashi M, Ito K, Yoshida R, Tomabechi D, Kida H, Takada A. Predicting the Antigenic Structure of the Pandemic (H1N1) 2009 Influenza Virus Hemagglutinin. *PLoS ONE*. 2010;5(1):e8553. doi:10.1371/journal.pone.0008553. PMID:20049332
- [24] NetNGlyc. NetNGlyc 1.0 Server Available: <http://www.cbs.dtu.dk/services/NetNGlyc/>. Accessed 22 may, 2016.
- [25] Doll S, Burlingame AL. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem Biol*. 2015;10(1):63-71. doi:10.1021/cb500904b. PMID:25541750
- [26] Thaysen-Andersen M, Packer NH. Advances in LC-MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N-

- and O-glycoproteome. *Biochim Biophys Acta, Proteins Proteomics*. 2014;1844(9):1437-52. doi:10.1016/j.bbapap.2014.05.002
- [27] Das SR, Hensley SE, David A, Schmidt L, Gibbs JS, Puigbò P, Ince WL, Bennink JR, Yewdell JW. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. *Proc Natl Acad Sci U S A*. 2011;108(51):E1417-E22. doi:10.1073/pnas.1108754108. PMID:22106257
- [28] Das SR, Puigbo P, Hensley SE, Hurt DE, Bennink JR, Yewdell JW. Glycosylation focuses sequence variation in the influenza A virus H1 hemagglutinin globular domain. *PLoS Pathog*. 2010;6(11):e1001211. doi:10.1371/journal.ppat.1001211
- [29] Thaysen-Andersen M, Packer NH. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology*. 2012;22(11):1440-52. doi:10.1093/glycob/cws110. PMID:22798492
- [30] An Y, Rininger JA, Jarvis DL, Jing X, Ye Z, Aumiller JJ, Eichelberger M, Cipollo JF. Comparative Glycomics Analysis of Influenza Hemagglutinin (H5N1) Produced in Vaccine Relevant Cell Platforms. *J Proteome Res*. 2013;12(8):3707-20. doi:10.1021/pr400329k. PMID:23848607
- [31] Wei C-J, Boyington JC, Dai K, Houser KV, Pearce MB, Kong W-P, Yang Z-y, Tumpey TM, Nabel GJ. Cross-Neutralization of 1918 and 2009 Influenza Viruses: Role of Glycans in Viral Evolution and Vaccine Design. *Sci Transl Med*. 2010;2(24):24ra1-ra1. doi:10.1126/scitranslmed.3000799
- [32] Ohuchi M, Ohuchi R, Feldmann A, Klenk HD. Regulation of receptor binding affinity of influenza virus hemagglutinin by its carbohydrate moiety. *J Virol*. 1997;71(11):8377-84. PMID:9343193
- [33] Schauer R. Sialic acids as antigenic determinants of complex carbohydrates. *Adv Exp Med Biol*. 1988;228:47-72. doi:10.1007/978-1-4613-1663-3_2. PMID:2459931
- [34] Schauer R. Sialic acids as regulators of molecular and cellular interactions. *Curr Opin Struct Biol*. 2009;19(5):507-14. doi:10.1016/j.sbi.2009.06.003. PMID:19699080
- [35] Shi X, Jarvis DL. Protein N-Glycosylation in the Baculovirus-Insect Cell System. *Curr drug targets*. 2007;8(10):1116-25. doi:10.2174/138945007782151360. PMID:17979671
- [36] Mir-Shekari SY, Ashford DA, Harvey DJ, Dwek RA, Schulze IT. The glycosylation of the influenza A virus hemagglutinin by mammalian cells. A site-specific study. *J Biol Chem*. 1997;272(7):4027-36. doi:10.1074/jbc.272.7.4027. PMID:9020110
- [37] Fukuda M, Guan JL, Rose JK. A membrane-anchored form but not the secretory form of human chorionic gonadotropin-alpha chain acquires polylectosaminoglycan. *J Biol Chem*. 1988;263(11):5314-8. PMID:2451668
- [38] Soema PC, Kompier R, Amorij J-P, Kersten GFA. Current and next generation influenza vaccines: Formulation and production strategies. *Eur J Phar Biopharm*. 2015;94:251-63. doi:10.1016/j.ejpb.2015.05.023
- [39] Manini I, Domnich A, Amicizia D, Rossi S, Pozzi T, Gasparini R, Panatto D, Montomoli E. Flucelvax (Optaflu) for seasonal influenza. *Expert Rev Vaccines*. 2015;14(6):789-804. doi:10.1586/14760584.2015.1039520. PMID:25968069
- [40] Yang LP. Recombinant trivalent influenza vaccine (flublok((R))): A review of its use in the prevention of seasonal influenza in adults. *Drugs*. 2013;73(12):1357-66. doi:10.1007/s40265-013-0103-6. PMID:23928902
- [41] Zhang S. Comparative Characterization of the Glycosylation Profiles of an Influenza Hemagglutinin Produced in Plant and Insect Hosts. *Proteomics*. 2012;12(8):1269-88. doi:10.1002/pmic.201100474. PMID:22577028
- [42] Butler M, Spearman M. The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr Opin Biotechnol*. 2014;30:107-12. doi:10.1016/j.copbio.2014.06.010. PMID:25005678
- [43] Squires RB, Noronha J, Hunt V, García-Sastre A, Macken C, Baumgarth N, Suarez D, Pickett BE, Zhang Y, Larsen CN, et al. Influenza Research Database: An integrated bioinformatics resource for influenza research and surveillance. *Influenza Other Respir Viruses*. 2012;6(6):404-16. doi:10.1111/j.1750-2659.2011.00331.x. PMID:22260278