Variation of Site-Specific Glycosylation Profiles for Influenza Glycoproteins from Recombinant Sources and Vaccines

Zachary C. Goecker, Meghan C. Burke, Yi Liu, Yuri A. Mirokhin, Sergey L. Sheetlin, Xiaoyu Yang, Dmitrii V. Tchekhovskoi, and Stephen E. Stein

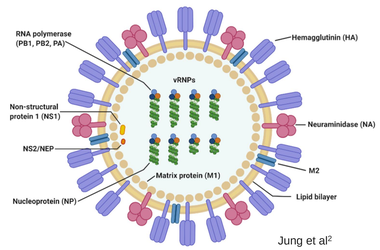
Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

Overview

- Glycopeptide abundance distribution spectra (GADS)[1] were created for influenza A recombinant glycoproteins and commercial influenza monovalent and quadrivalent vaccines.
- Variation in glycan abundances and distribution was most pronounced between different strains and least pronounced between biological replicates of the same protein and same vendor.
- Adjacent sequons may be isolated via a non-specific protease or EThcD fragmentation and GADS for commercial influenza vaccines can be partially elucidated with current methods.

Introduction

Influenza is a viral disease responsible for the death of ~500,000 people annually worldwide. As a result, vaccination against influenza is one of the most common immunizations, taken by about 50% of the U.S. population annually. On average, the effectiveness of the influenza vaccine is only 40%. One of the possible factors contributing this low efficacy is the variation of the glycan distributions on the external proteins, hemagglutinin and neuraminidase. The accurate measurement of these site-specific glycan distributions and determination of their variability presents a significant analytical challenge. This work develops methods for both accurate and reproducible measurements of these distributions to reduce uncertainty in vaccine development.



Jung et al[2]

Methodology

Recombinant proteins:

Nine hemagglutinin and neuraminidase strains were obtained from Sino Biological, BioVision, US Biological, and Creative Biomart. All proteins are from recombinant expression in HEK293 cell line. Multiple lots were ordered from these vendors.

Sample processing:

Purified recombinant protein was digested using 0.1% RapiGest surfactant in 50 mM ABC, 20 mM DTT, and 55 mM IAA. Protein was digested using the following protease combinations: Trypsin+Glu-C, Trypsin+Lys-C, Trypsin+Asp-N, Trypsin+Chymotrypsin, Chymotrypsin, alpha-Lytic protease, and Glu-C+Chymotrypsin. Digests were purified and desalted using a MonoSpin column procedure.

Instrumental analysis:

Digests were analyzed on a Thermo Scientific Fusion Lumos mass spectrometer using an optimized stepped HCD energy-ion trap fragmentation sequence.

Software:

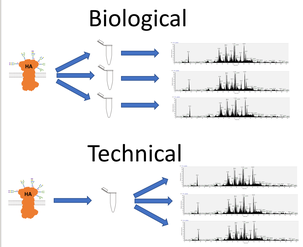
NIST MS_Piano and NIST MS Search were used for annotation and visualization. NIST developed software was used for glycopeptide and GADS library creation. GADS pseudo-spectra give the abundances of different glycans from the same peptide sequence.

Digital reprint

Results and Discussion

Variation in Glycosylation Patterns

Replicate Variation



Protein	Sequon	Average match score
Hemagglutinin	39	963
	170	967
	181	939 ± 40
	302	946 ± 44
	500	926 ± 27
Neuraminidase	68	939 ± 14
	126	961 ± 24
	215	927 ± 68
Total		946 ± 38

Protein	Sequon	Average match score
Hemagglutinin	39	980 ± 2
	170	913 ± 60
	181	968 ± 16
	302	884 ± 92
	500	884 ± 27
Neuraminidase	68	973 ± 6
	126	945 ± 21
	215	959 ± 11
Total		939 ± 62

Figure 1. Biological and technical replicate match scores

Lot Variation

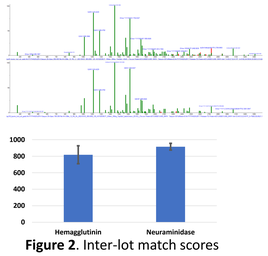


Figure 2. Inter-lot match scores

Vendor Variation

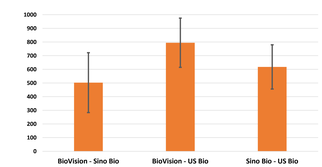


Figure 3. Library match scores between three vendors

Protease Variation

SYnNTNQE
SYnNTNQEDLLVLWGIHHPNDAAEQTK
nNTNQEDLLVLW
IKRSYnNTNQEDLLV **Average match: 819 ± 106**

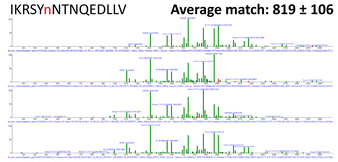


Figure 4. GADS from different peptides of the same sequon

Inter-strain Variation

Average Match Score: 463 ± 130

Protein	Strain	Subtype	Number of sequons	Protein mass* (kDa)
HA	A/California/04/2009	H1N1	8	63
NA	A/California/04/2009	H1N1	8	52
HA	A/Hong Kong/483/1997	H5N1	8	64
HA	A/Hong Kong/485197/2014	H3N2	13	64
HA	A/Japan/305/1957	H2N2	8	63
HA	A/New Caledonia/20/1999	H1N1	10	63
NA	A/Arizona/13/2008	H1N1	9	52
NA	A/Netherlands/219/2003	H7N7	11	52
NA	A/Thailand/1(KAN-1)/2004	H5N1	3	49

*Unglycosylated

Figure 5. GADS comparison between nine influenza strains

Intra-strain Variation

Average Match Score: 617 ± 151

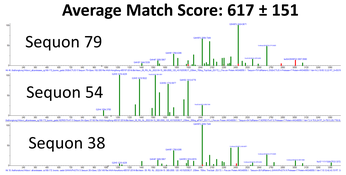


Figure 6. Match scores between different glycosylation sites within the same protein

Isolating Adjacent Sequons

EThcD

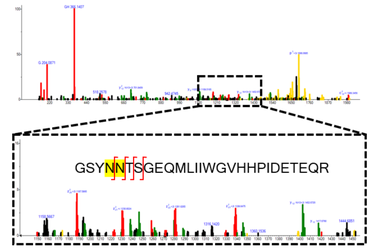


Figure 7. Tandem mass spectrum of glycopeptide containing the adjacent sequon. One sequon is unoccupied as shown by the z ion series. However, not all z ions in the series are present.

Subtilisin non-specific protease

…NNST…
…NNTT…
…NNTS…
…NNSS…
} Adjacent sequon motifs

ANNSTDTVDTVL
HANNSTDTVDTVL
HANNSTDTVDTV
nSTDTVDTVL
nSTDTVDTVLEK

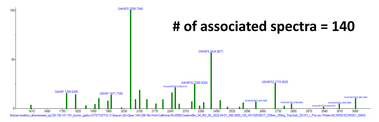


Figure 8. The non-specific protease subtilisin may be used to cleave between two adjacent asparagines, regardless of glycan occupancy. GADS are obtained which contain many supporting tandem spectra.

Influenza Vaccines

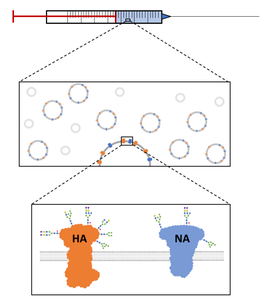


Figure 9. Hemagglutinin and neuraminidase are two glycoproteins found in influenza vaccines.

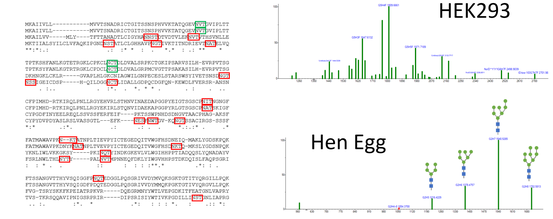


Figure 10. Sequence alignment of the four hemagglutinin strains found in the annual quadrivalent flu shot. Many sequons share a sequence with other strains, both in confined (red) and expansive (green) regions.

Figure 11. Typical GADS for recombinant hemagglutinin from HEK293 cell line are complex, whereas typical GADS from hemagglutinin from hen eggs are high-mannose.

Conclusions

- **Inter-strain variation resulted in the lowest levels of library match (463 ± 130), followed by intra-strain variation (617 ± 151), and vendor variation (638 ± 218). However, variation between replicates (946 ± 38), lots (891 ± 57), and different protease combinations (870 ± 51) resulted in high match scores.**
- **Adjacent sequons may be isolated using EThcD fragmentation or non-specific cleavage using the protease subtilisin and most glycans are high-mannose in quadrivalent vaccines.**
- **Results herein demonstrate the advance of site-specific glycosylation methods in determining glycosylation patterns for influenza vaccines, which could be implemented in vaccine QC metrics.**

References and Disclaimers

[1]Remoroza, C. A., Burke, M. C., Liu, Y., Mirokhin, Y. A., Tchekhovskoi, D. V., Yang, X., & Stein, S. E. (2021). Representing and Comparing Site-Specific Glycan Abundance Distributions of Glycoproteins. Journal of Proteome Research, 20(9), 4475-4486.
[2]Jung, H. E., & Lee, H. K. (2020). Host protective immune responses against influenza A virus infection. *Viruses*, *12*(5), 504.