Site-Specific Glycosylation Profiles for Recombinant Hemagglutinin and Neuraminidase from Different Sources

Zachary C. Goecker, Meghan C. Burke, Yi Liu, Yuri A. Mirokhin, Dmitrii V. Tchekhovskoi, Stephen E. Stein

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

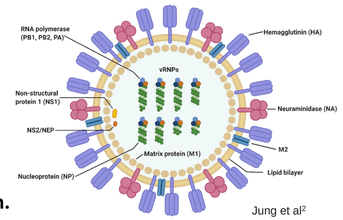
NIST National Institute of Standards and Technology U.S. Department of Commerce | Mass Spectrometry Data Center

Overview

- Glycopeptide abundance distribution spectra (GADS)[1] were created for influenza A recombinant glycoproteins hemagglutinin and neuraminidase (H5N1) to evaluate glycan variation.
- Variation in glycan abundances and distribution was most pronounced between different vendors and least pronounced between biological replicates of the same protein and same vendor.
- Glycan distribution was skewed toward higher mass (sialylated complex) glycans for Sino Biological and toward lower mass glycans (oligomannose and hybrid) for US Biological and BioVision.

Introduction

Influenza is a viral disease responsible for the death of ~500,000 people annually worldwide. Perhaps the most complex component of this virus is the N-linked glycan shield on its external proteins, hemagglutinin and neuraminidase. These proteins attach to host cells leading to infection and are the main antigens of interest for vaccine development and antiviral treatments. Their distributions of N-linked glycans at multiple sites are highly complex and serve to evade antibodies and perhaps serve other purposes. However, the accurate measurement of these site-specific glycan distributions and determination of their variability. However, the accurate measurement of these site-specific glycan distributions and determination of their variability presents a significant analytical challenge. In 2020, Chang et al successfully compared glycans from hemagglutinin variants.[2] This work attempts to develop methods for accurate and reproducible measurement of these distributions as well as for their detailed representation.



Jung et al[2]

Methodology

Recombinant proteins:

Hemagglutinin strain A/Hong Kong/483/1997(H5N1) and neuraminidase strain A/Thailand/1(KAN-1)/2004(H5N1) were obtained from Sino Biological, BioVision, and US Biological. 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:

MS_Piano and NIST MS Search were used for data annotation and visualization. NIST developed software was used for glycopeptide and GADS library creation.

Digital reprint

Results and Discussion

Replicate variation

Table 1. Missing glycans between replicates

Protein	Sequon	Peptide	Glycan	Glycan mass	% base peak
Hemagglutinin	39	KnVTVTHAQDILER/3	G6H3F	1850.69	10%
	170	NVVVVLKKnSTYPTIKR/4	G2H5	1216.42	28%
	181	RSYnNTNQE/3	G4H5F	1768.64	16%
	302	CQTPMGAInSSMPFHNIHPLTIGECPK/5	G4H5FS	2059.73	88%
			G4H5F	1768.64	12%
		CQTPMGAInSSMPFHNIHPLTIGECPK/4	G6H5FS2	2919.04	18%
			G2H6	1379.48	10%
			G4H4F	1606.59	20%
		CQTPMGAInSSMPFHNIHPLTIGECPKYVK/5	G6H4F	2012.75	12%
Neuraminidase	126	TFFLTQGALLNDKHSnGTVK/4	G6H3F3	2142.81	38%
	215	CACVnGSCFTVMTDGPSNGQASHKIFK/4	G4H5FS	2059.73	18%

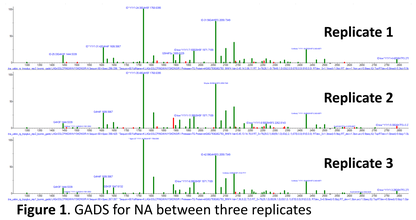


Figure 1. GADS for NA between three replicates

Protein	Sequon	Match score
Hemagglutinin	39	894 ± 84
	170	935 ± 4
	181	925 ± 76
	302	893 ± 69
	500	813 ± 73
Neuraminidase	68	869 ± 72
	126	928 ± 36
	215	929 ± 26

Table 2. Average match scores per protein glycosylation site

Discussion:
To determine baseline variation, glycan profiles were compared between three biological replicates. Glycan distribution was similar between replicates (Fig 1), with few missing glycans between replicates (Table 1) and average match scores around 900/999 (Table 2).

Lot variation

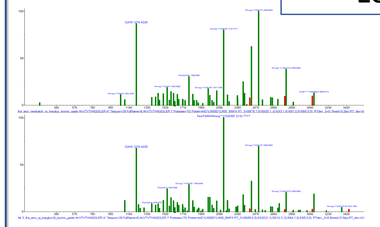


Figure 3. GADS for HA between 2 lots

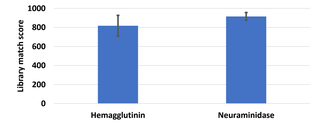


Figure 4. Inter-lot match scores

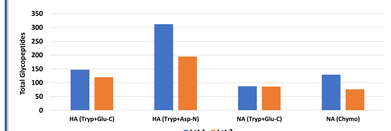


Figure 5. Total glycopeptides between lots

Discussion:
Separate production lots were compared for both Sino Biological and BioVision. Glycan distribution patterns were similar (Fig 3), as well as the inter-lot match scores between identical peptides (Fig 4). However, the total number of glycopeptides did fluctuate (Fig 5).

Peptide variation

Trypsin + Lys-C	Chymotrypsin
Trypsin + Asp-N	Chymotrypsin + Glu-C
Trypsin + Glu-C	alpha-Lytic
Trypsin + Chymotrypsin	

Figure 6. Protease combinations used

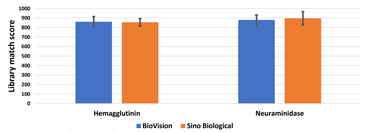


Figure 8. Inter-peptide match scores

SYnNTNQE
SYnNTNQEDLLVLWGIHHPNDAAEQTK
nNTNQEDLLVLW
IKRSYnNTNQEDLLV

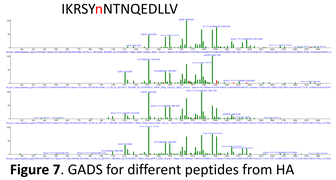


Figure 7. GADS for different peptides from HA

Discussion:
Glycopeptides from the same sequon but which are cleaved at different amino acids (Fig 6) were compared via GADS (Fig 7) and inter-peptide library match scores (Fig 8). Glycan distribution was highly similar (>800) regardless of peptide cleavage.

Vendor variation

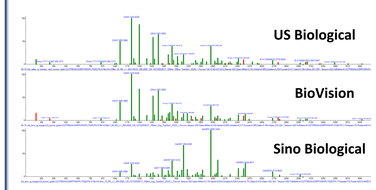


Figure 9. GADS for HA between three vendors

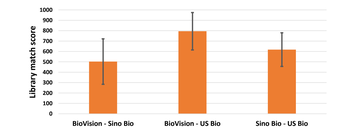


Figure 11. Library match scores between three vendors

HA – sequon 39 - KnVTVTHAQDILER/4	US Biological: G5H3 > G2H5 > G6H3	BioVision: G6H3 > G5H3 > G2H5	Sino Biological: G2H5 > G2H7 > G5H7F52
HA – sequon 39 - KnVTVTHAQDILER/3	G2H5 > G5H3 > G3H3	G2H5 > G6H3 > G5H3	G2H5 > G5H6F > G2H7
HA – sequon 181 - SYnNTNQE/3	G4H3F > G5H3F > G5H4F	G5H4F5 > G4H5FS > G5H3F5	G5H4F > G4H5F > G4H5FS
HA – sequon 181 - SYnNTNQE/2	G2H5 > G4H3F > G4H3	G4H3F > G4H3 > G2H5	G2H5 > G4H3F > G4H5F
HA – sequon 302 - CQTPMGAInSSMPFHNIHPLTIGECPK/4	G2H5 > G4H3 > G5H3	G2H5 > G4H3 > G5H3	G4H5FS > G4H5F > G5H6F
HA – sequon 302 - CQTPMGAInSSMPFHNIHPLTIGECPKYVK/5	G2H5 > G4H3 > G5H3	G2H5 > G4H3 > G5H3	G4H5FS > G4H5F > G2H5
HA – sequon 500 - SVKnGTYDYPQYSEE/3	G2H5 > G4H3 > G5H3	G4H3 > G2H5 > G5H3	G4H5FS > G2H5 > G5H6F5
NA – sequon 126 - TFFLTQGALLNDKHSnGTVK/4	G5H4F2 > G5H4F > G4H3F	G4H3F > G5H4F2 > G5H4F	G5H4F2 > G5H4F > G6H3F2

Figure 10. Top 3 abundant glycans for each glycopeptide

Discussion:
Glycoproteins from BioVision, Sino Biological, and US Biological were compared in terms of glycan distribution patterns (Fig 9), most abundant glycans identified (Fig 10), and library matching between identical peptides (Fig 11).

Conclusions

- **Variation due to vendor differences resulted in the lowest levels of library match (as low as 502 ± 218), whereas replicate, lot, and peptide variation resulted in high match scores (898 ± 73, 867 ± 101, and 870 ± 51 respectively).**
- **The most abundant glycans for BioVision and US Biological HA are G2H5, G5H3, and G4H3, whereas the most abundant glycans for Sino Biological are G2H5, G4H5FS, and G4H5F.**
- **Glycans that were present in hemagglutinin but not neuraminidase include G6H7FS3, G5H6FS2, and G5H6FS3. No glycans were identified in neuraminidase that were not also in hemagglutinin.**

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]Chang, D., Hackett, W. E., Zhong, L., Wan, X. F., & Zaia, J. (2020). Measuring site-specific glycosylation similarity between influenza A virus variants with statistical certainty. Molecular & Cellular Proteomics, 19(9), 1533-1545.
[3]Jung, H. E., & Lee, H. K. (2020). Host protective immune responses against influenza A virus infection. *Viruses*, *12*(5), 504.

Disclaimer: Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.