Variation of Site-Specific Glycosylation Profiles for Influenza Glycoproteins from Different Sources

Zachary Goecker, Meghan Burke, Concepcion Remoroza, Yi Liu, Yuri Mirokhin, Sergey Sheetlin,
Dmitrii Tchekhovskoi, Xiaoyu Yang, and Stephen Stein

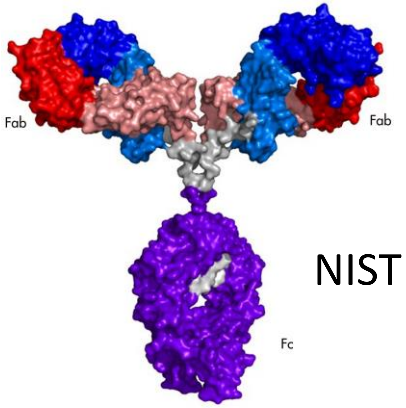
BMD Staff Seminar

May 17, 2022

National Institute of
Standards and Technology
U.S. Department of Commerce

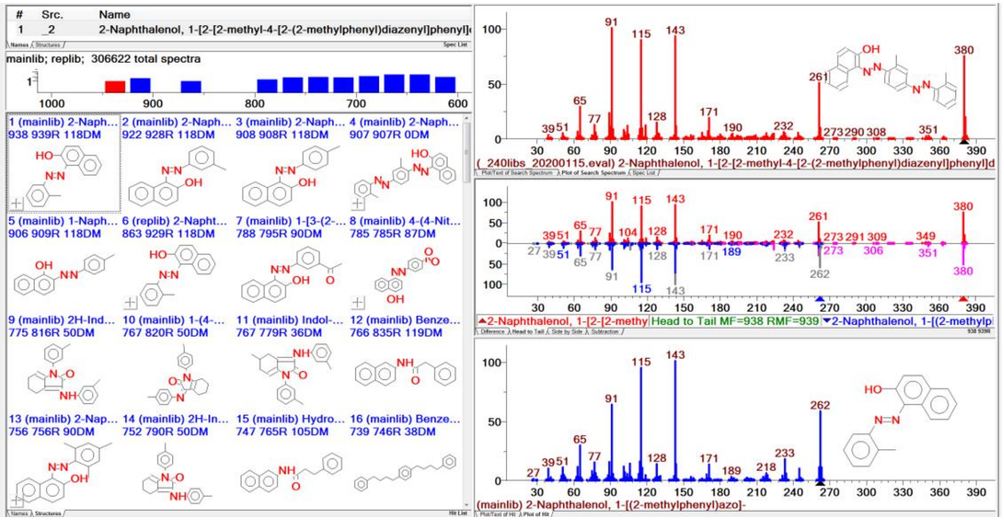
Mass Spectrometry
Data Center

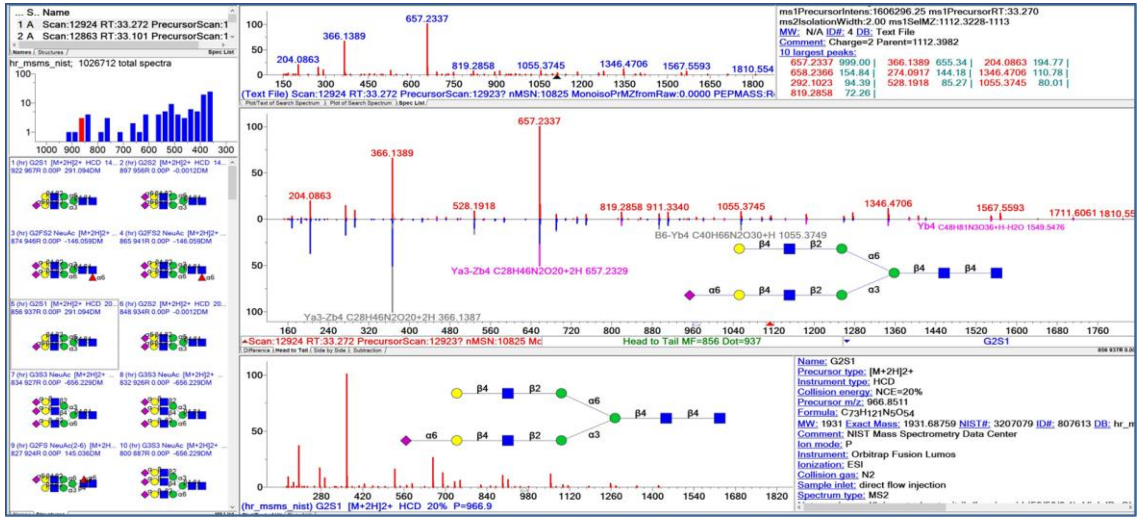
Why Glycopeptides at NIST?



NISTMab

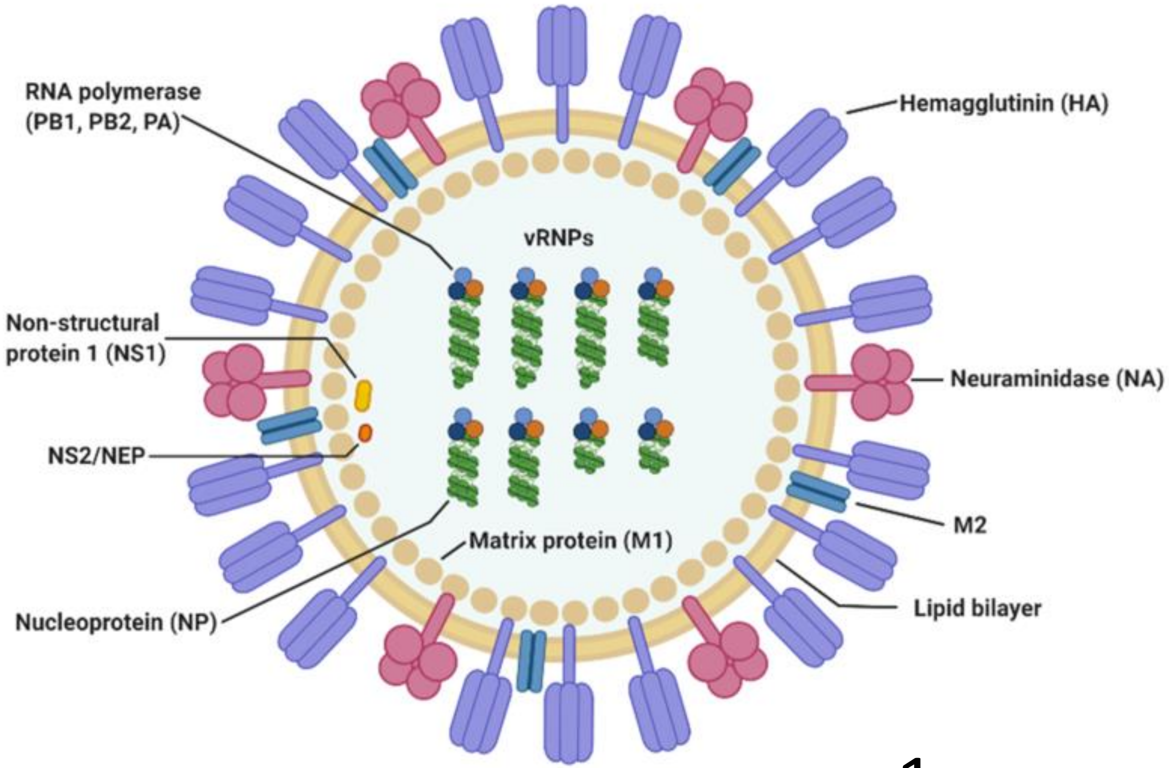
- Tandem libraries are growing.
- Glycopeptides prominent in biologics.
- Measurement of glycosylation is difficult and hard to reproduce.
- Variation in glycosylation pattern unknown. How reproducible?





Influenza Virus

- 10 proteins
- Hemagglutinin (HA) and neuraminidase (NA) transmembrane proteins.
- Diversity of strains arise through two mechanisms: point mutations in the viral genome or reassortment between two co-circulating strains.

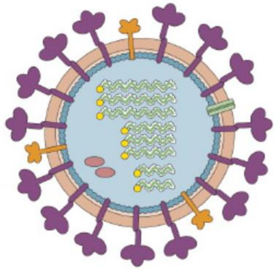


1

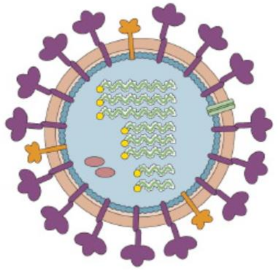
Influenza Vaccine

- Quadrivalent: 4 strains
  - 2 Influenza A strains
  - 2 Influenza B strains
- Embryonated chicken eggs
- Inactivated by rupturing membrane
- Adjuvants

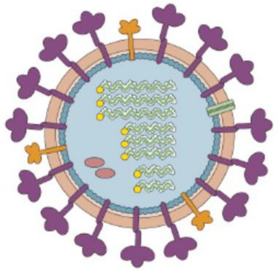
Influenza A (H3N2)



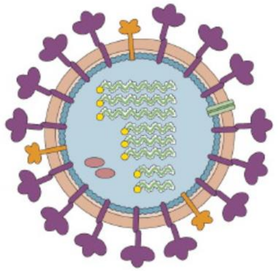
Influenza A (H1N1)



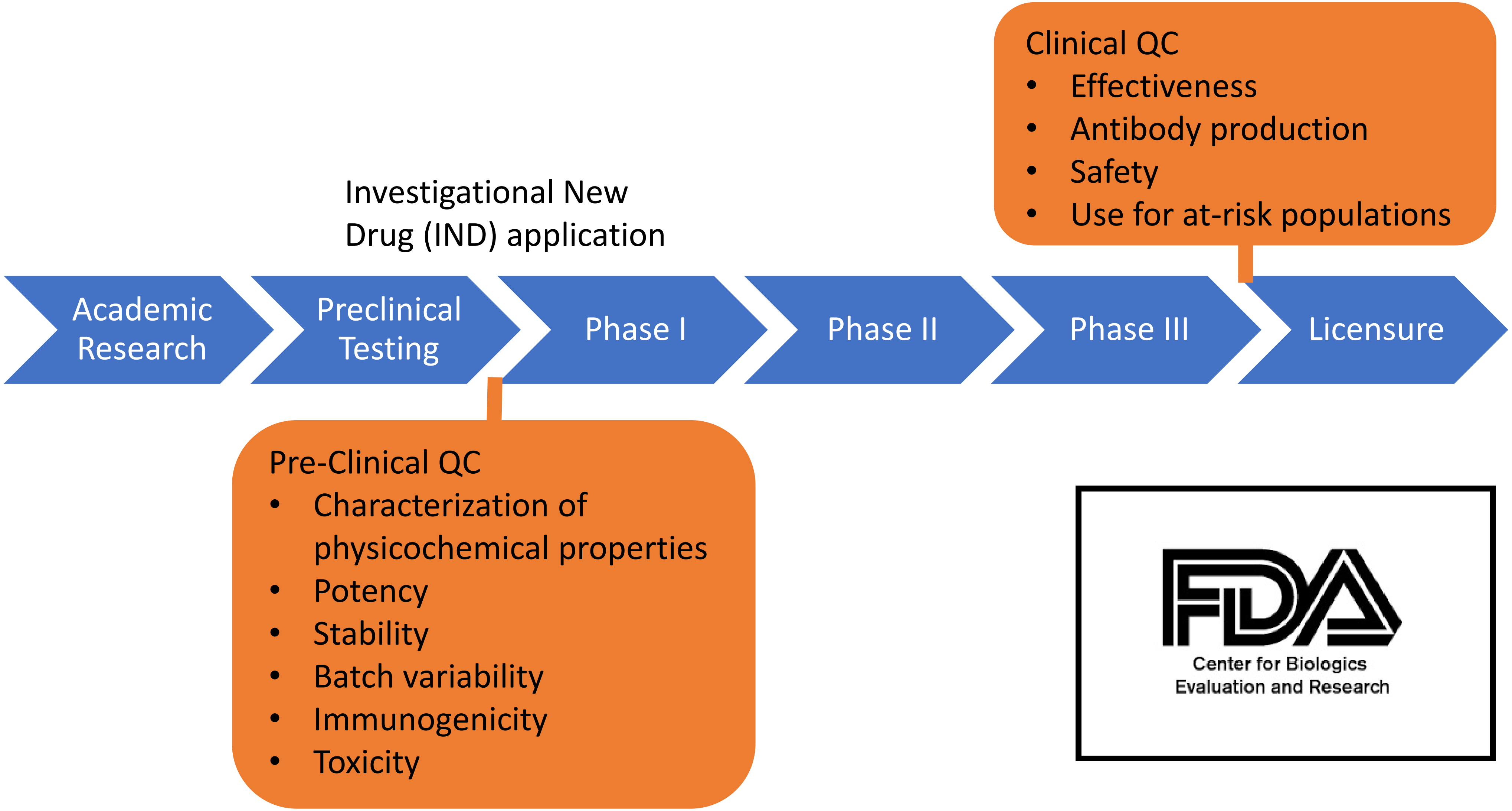
Influenza B (Victoria)



Influenza B (Yamagata)



Quality Control of Influenza Vaccines



Investigational New Drug (IND) application

Academic Research

Preclinical Testing

Phase I

Phase II

Phase III

Licensure

Clinical QC
- Effectiveness
- Antibody production
- Safety
- Use for at-risk populations

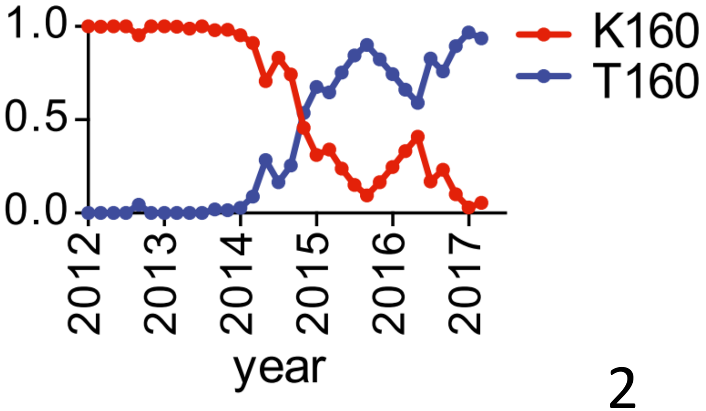
Pre-Clinical QC
- Characterization of physicochemical properties
- Potency
- Stability
- Batch variability
- Immunogenicity
- Toxicity

FDA
Center for Biologics
Evaluation and Research

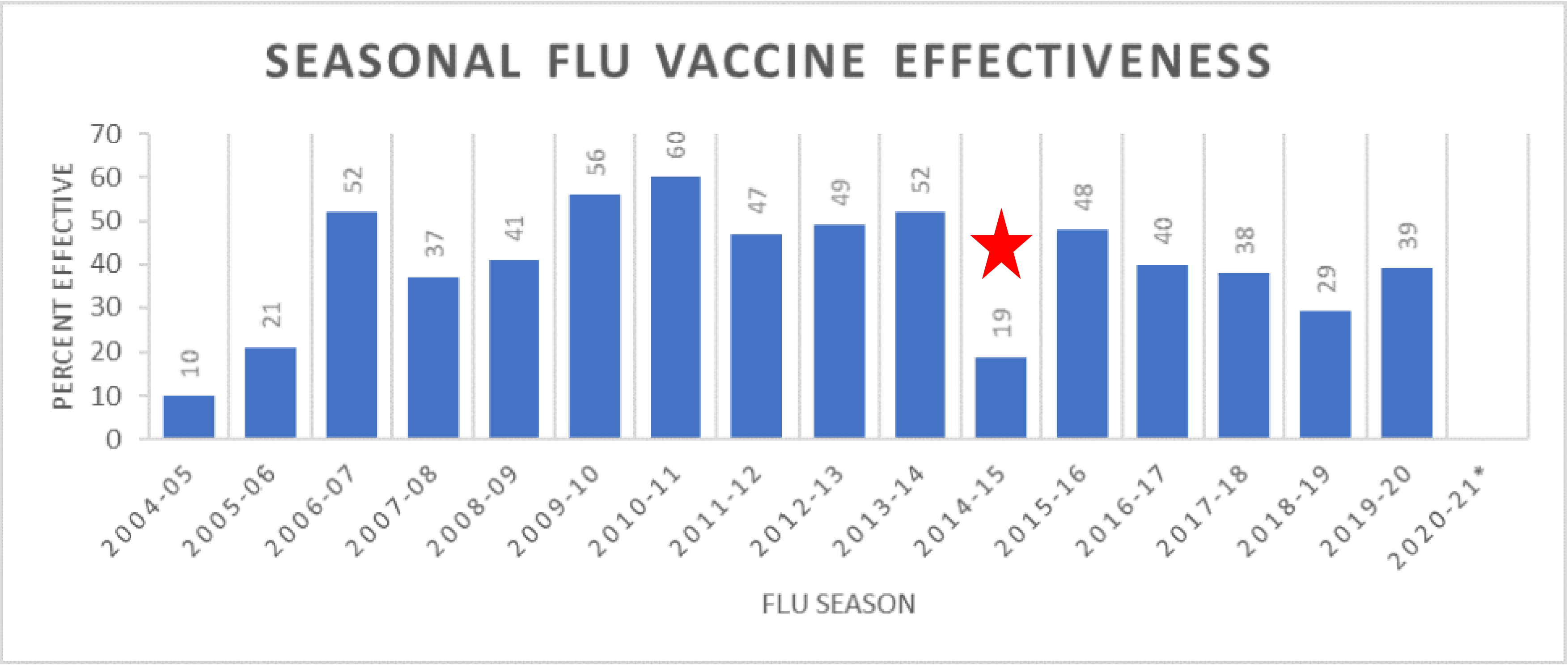
Why Glycosylation Matters For Vaccines

…THHLNFKYPAL…
…THLNFTYPAL…

Sequon motif
NXT/S X≠P

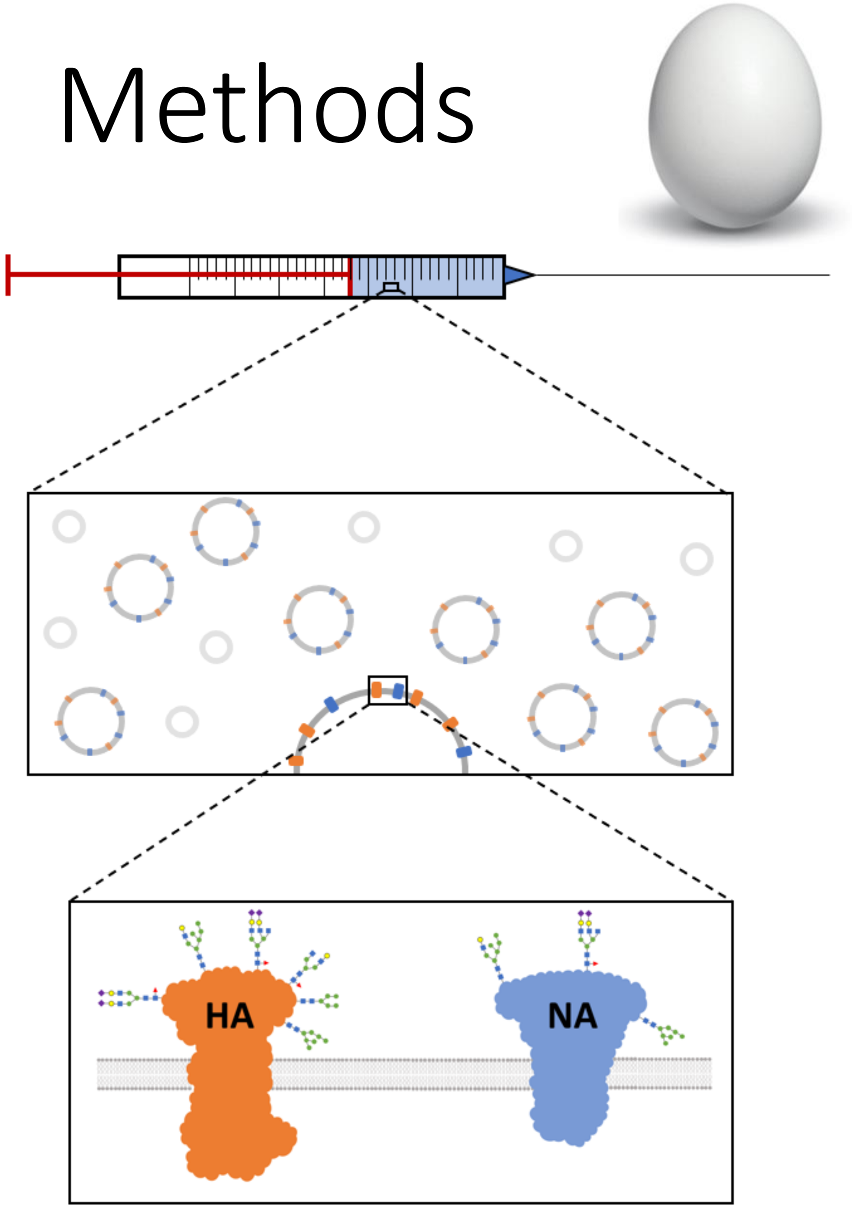


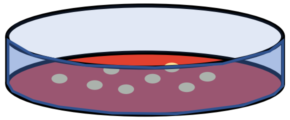
2



3

Methods

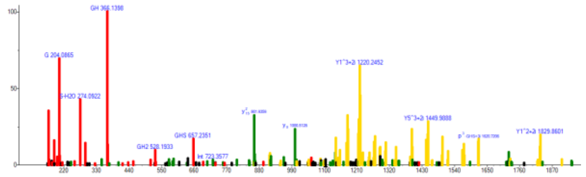




HEK293

Reduction
Alkylation
Digestion
Purification

HCD - 15%, 25%, 35% NCE

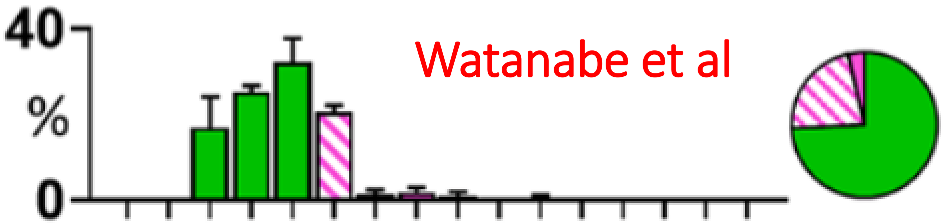


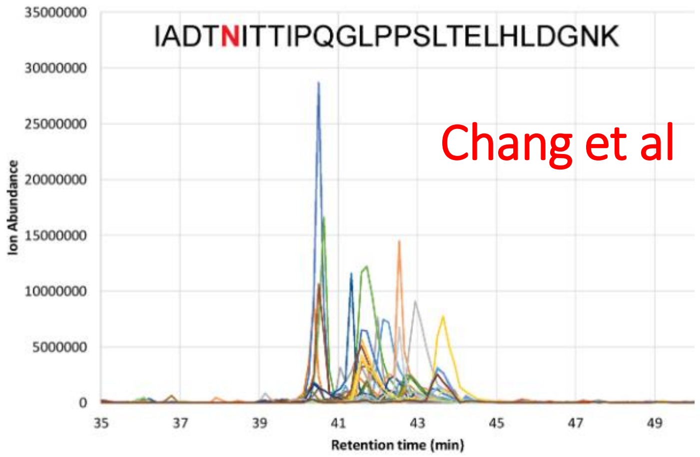
Oxonium ions
Peptide backbone ions
Glycopeptide ions

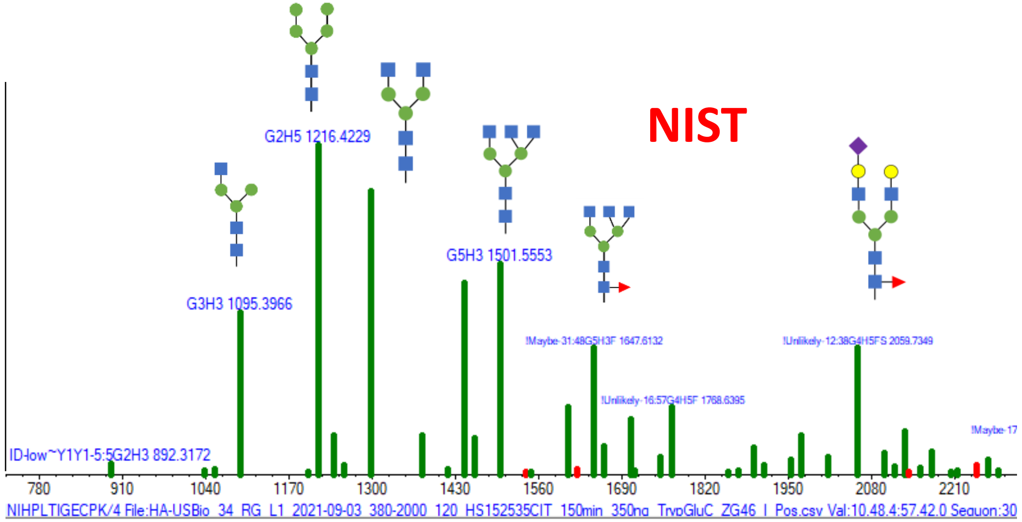
Methods – Data Processing

1. Glycopeptide search using Byonic software
2. Tandem library creation and spectral validation via create_glycopeptide_lib.exe and MS_Piano.exe.
3. GADS creation using make-gads.exe

Glycopeptide Abundance Distribution Spectrum

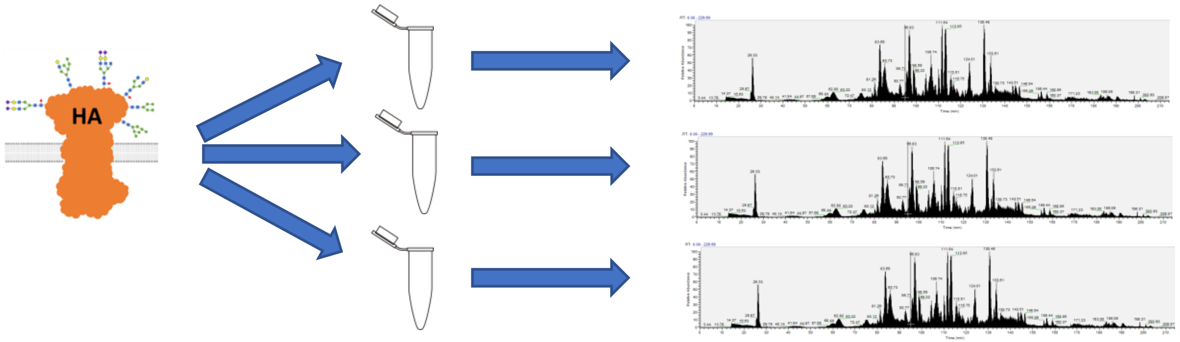

Watanabe et al


Chang et al


NIST

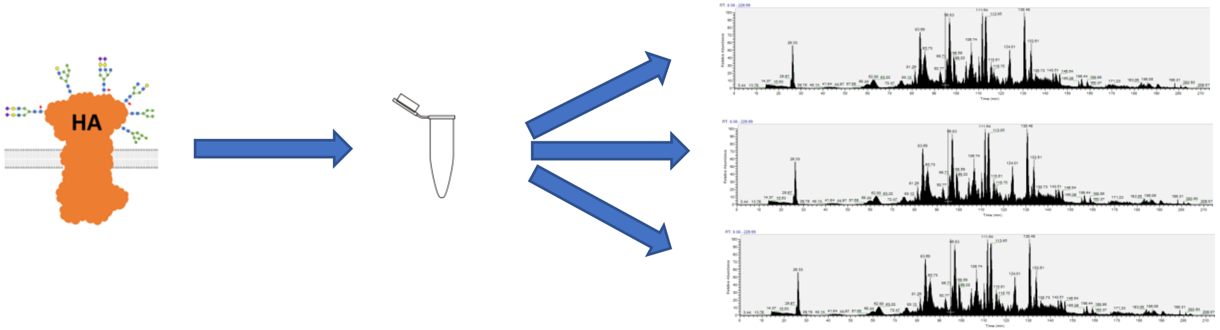
Variation in glycosylation profile - Replicate

Biological Replicate



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

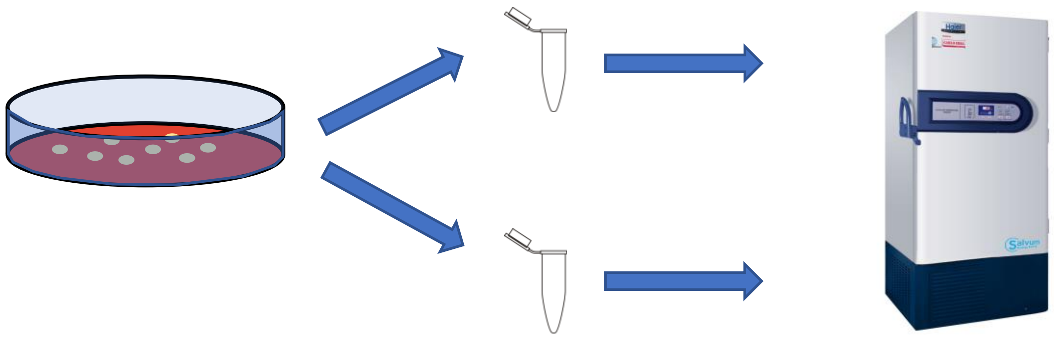
Technical Replicate



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

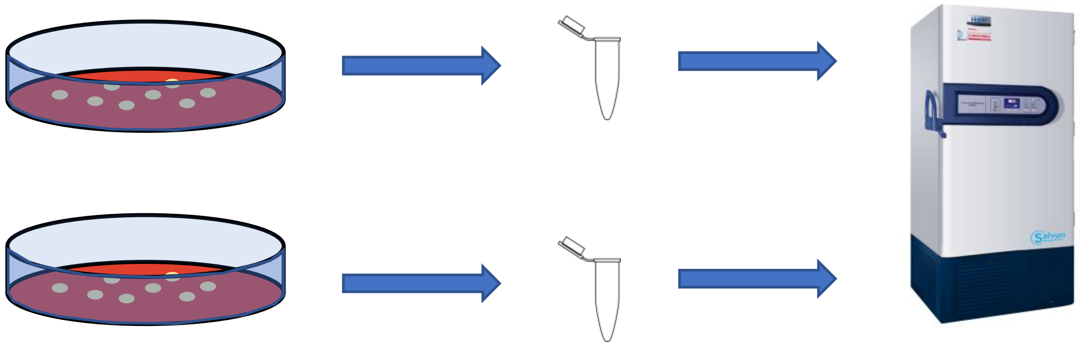
Variation in glycosylation profile - Lot

Batch Variation



Protein	Sequon	Average match score
Hemagglutinin	39	889 ± 75
	170	843 ± 38
	181	942 ± 19
	302	909 ± 45
	500	872 ± 69
Total		897 ± 57

Lot Variation



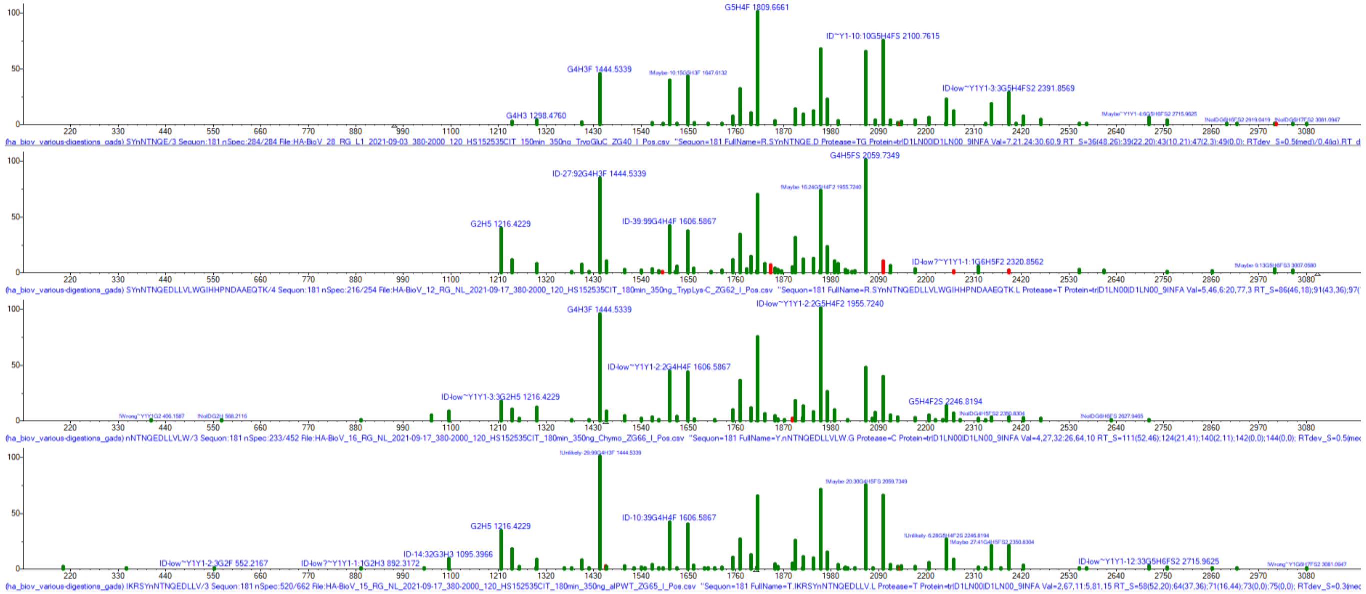
Protein	Sequon	Average match score
Neuraminidase	68	903 ± 39
	126	852 ± 72
	215	910 ± 48
Total		891 ± 57

Variation in glycosylation profile - Proteases

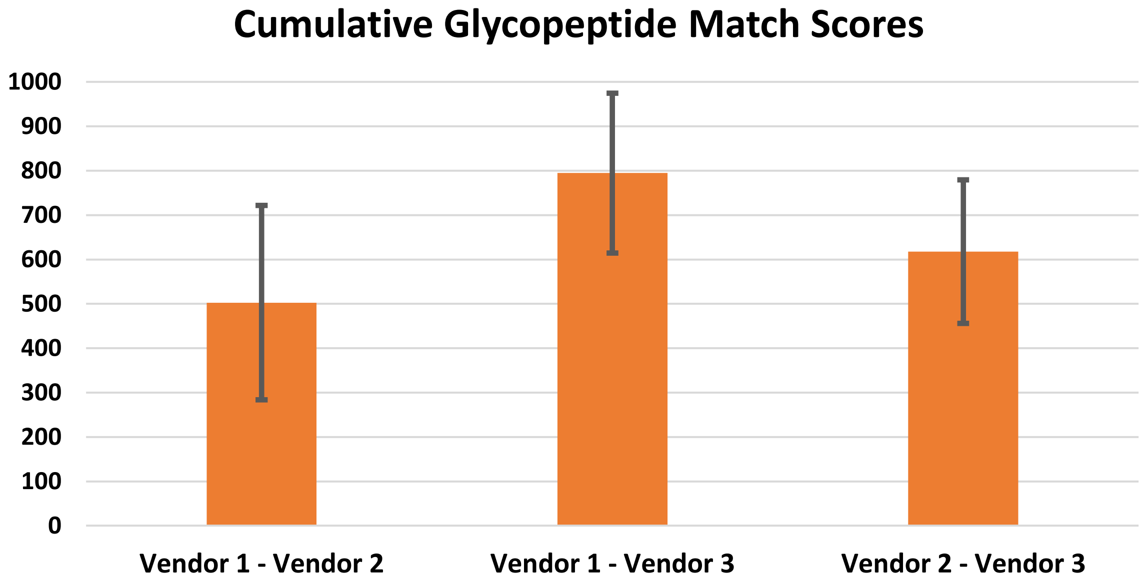
- Trypsin + Lys-C
- Trypsin + Glu-C
- Trypsin + Chymotrypsin
- Chymotrypsin + Glu-C
- Chymotrypsin
- Alphalytic protease

Average match: 819 ± 106

SYnNTNQE
SYnNTNQEDLLVLWGIHHPNDAAEQTK
nNTNQEDLLVLW
IKRSYnNTNQEDLLV



Variation in glycosylation profile - Vendor



Vendors

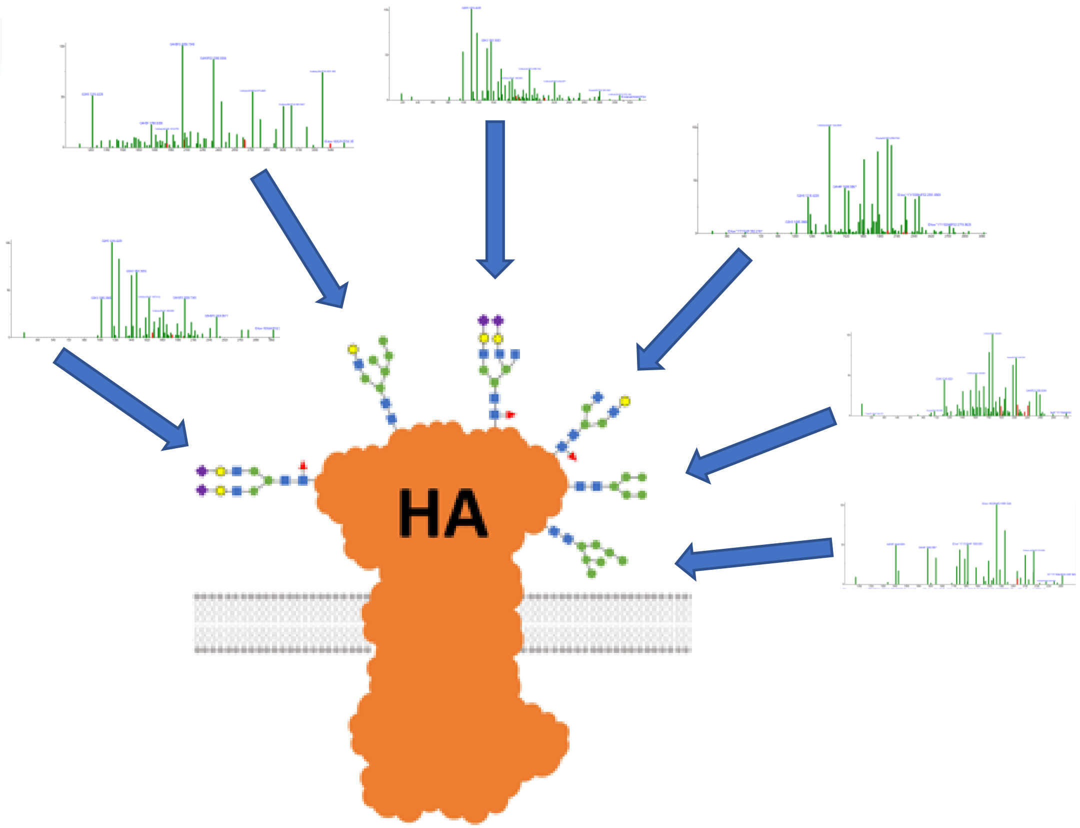
- BioVision
- Sino Biological
- US Biological

Strains

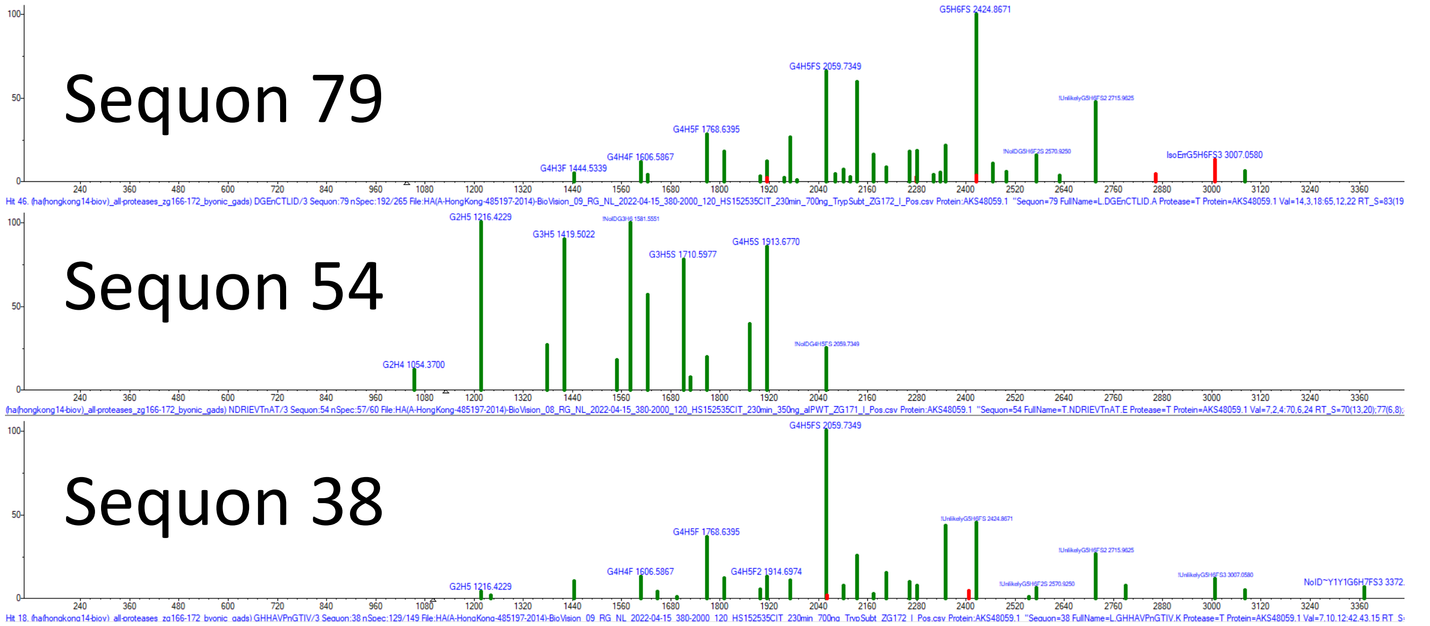
- HA (A/Hong Kong/483/1997) H5N1
- NA (A/Thailand/1(KAN-1)/2004) H5N1

Average match score: 638 ± 218

Variation in glycosylation profile – Same Strain



Average Match Score: 617 ± 151

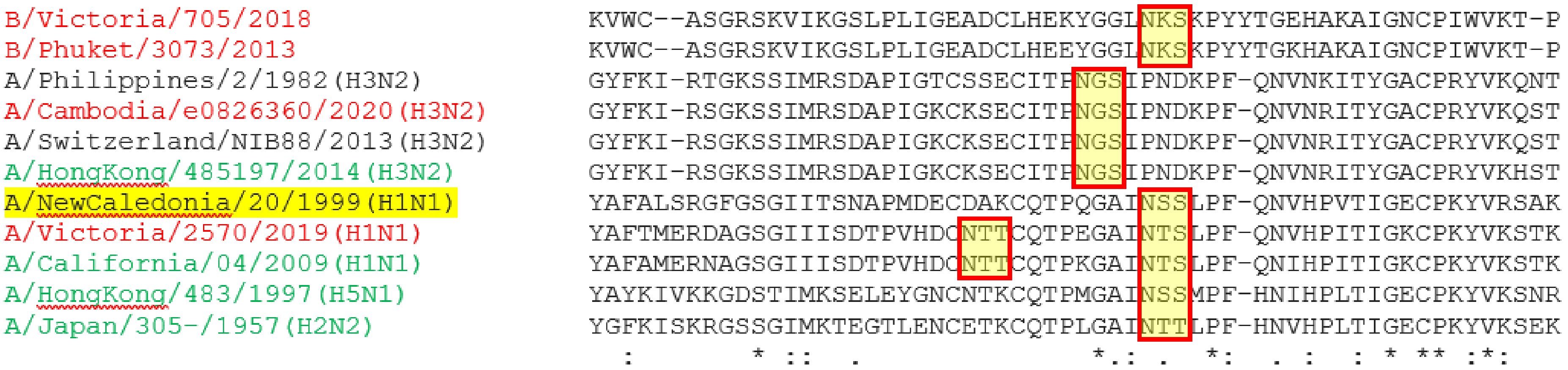


Variation in glycosylation profile – Different Strains

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

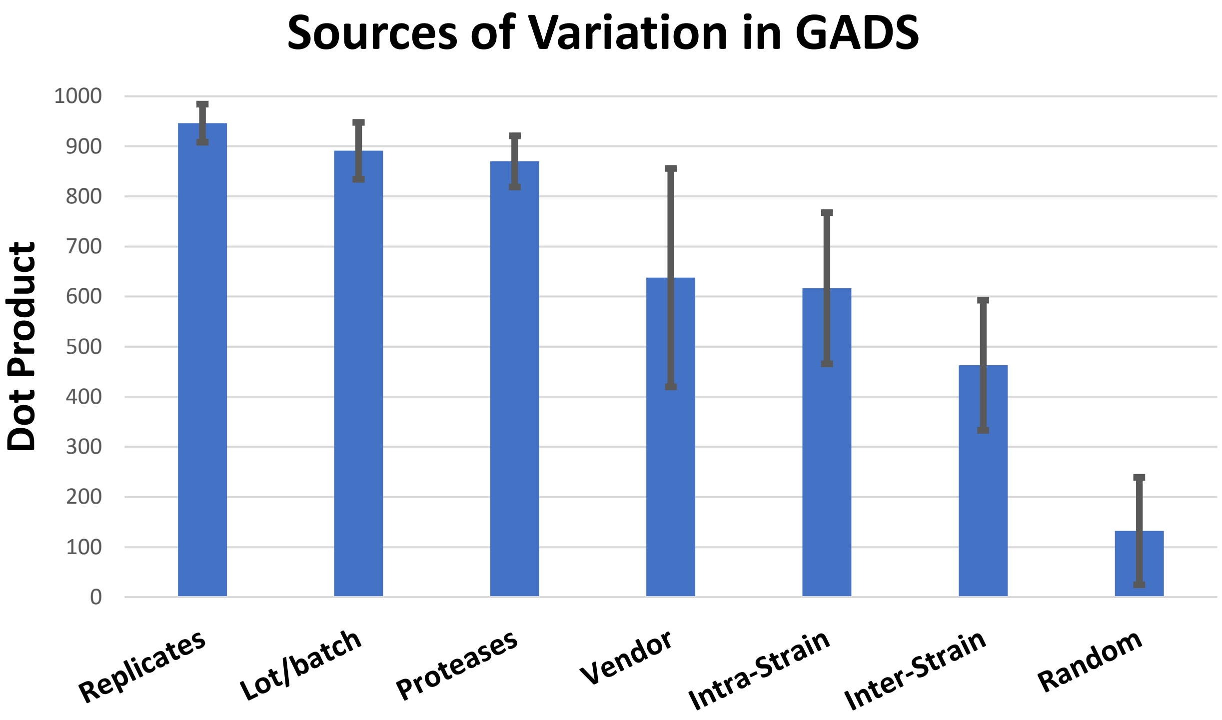
*Unglycosylated

Average Match Score:
463 ± 130



```
B/Victoria/705/2018              KVWC--ASGRSKVIKGSLPLIGEADCLHEKYGGLNKSKPYYTGEHAKAIGNCPIWVKT-P
B/Phuket/3073/2013               KVWC--ASGRSKVIKGSLPLIGEADCLHEEYGGLNKSKPYYTGKHAKAIGNCPIWVKT-P
A/Philippines/2/1982(H3N2)       GYFKI-RTGKSSIMRSDAPIGTCSSECITPNGSIPNDKPF-QNVNKITYGACPRYVKQNT
A/Cambodia/e0826360/2020(H3N2)   GYFKI-RSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPF-QNVNRITYGACPRYVKQST
A/Switzerland/NIB88/2013(H3N2)   GYFKI-RSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPF-QNVNRITYGACPRYVKQST
A/HongKong/485197/2014(H3N2)     GYFKI-RSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPF-QNVNRITYGACPRYVKHST
A/NewCaledonia/20/1999(H1N1)     YAFALSRGFGSGIITSNAPMDECDAKCQTPQGAINSSLPF-QNVHPVTIGECPKYVRSAK
A/Victoria/2570/2019(H1N1)       YAFTMERDAGSGIIISDTPVHDCNTTCQTPEGAINTSLPF-QNVHPITIGKCPKYVKSTK
A/California/04/2009(H1N1)       YAFAMERNAGSGIIISDTPVHDCNTTCQTPKGAINTSLPF-QNIHPITIGKCPKYVKSTK
A/HongKong/483/1997(H5N1)        YAYKIVKKGDSTIMKSELEYGNCNTKCQTPMGAINSSMPF-HNIHPLTIGECPKYVKSNR
A/Japan/305-/1957(H2N2)          YGFKISKRGSSGIMKTEGTLENCETKCQTPLGAINTTLPF-HNVHPLTIGECPKYVKSEK
                                   :       * ::  .              *.: .  *:  . :  : * ** :*:
```

Summary of Variation



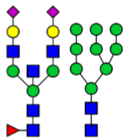
- Glycosylation of replicates is most similar and between strains is least similar.
- More deviation associated with lower match scores.
- “Random” is measured between different proteins from different vendors. (A1AG vs HA)

Isolation of Adjacent Sequons

- Site-specific glycosylation methods require a single sequon per peptide. Different proteases are used to maximize isolation.
- Influenza, HIV, Ebola, Herpesvirus, and MERS have adjacent sequons

NXT/S X≠P

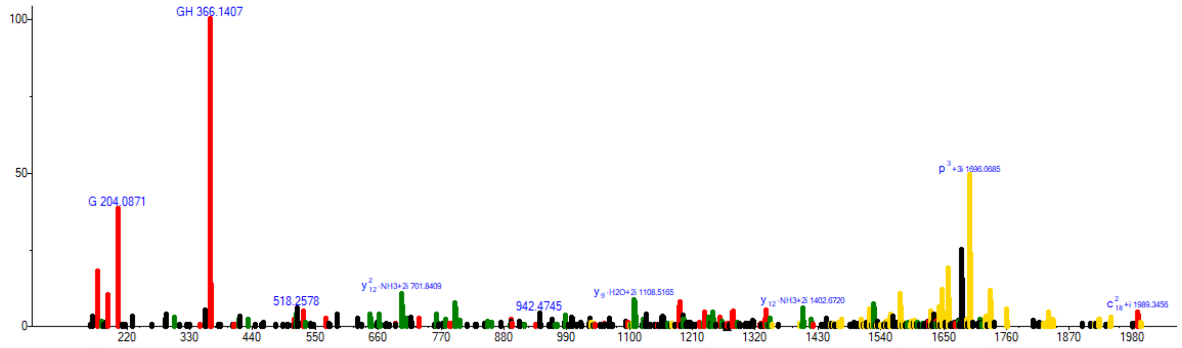
NNSS, NNTT, NNST, NNTS

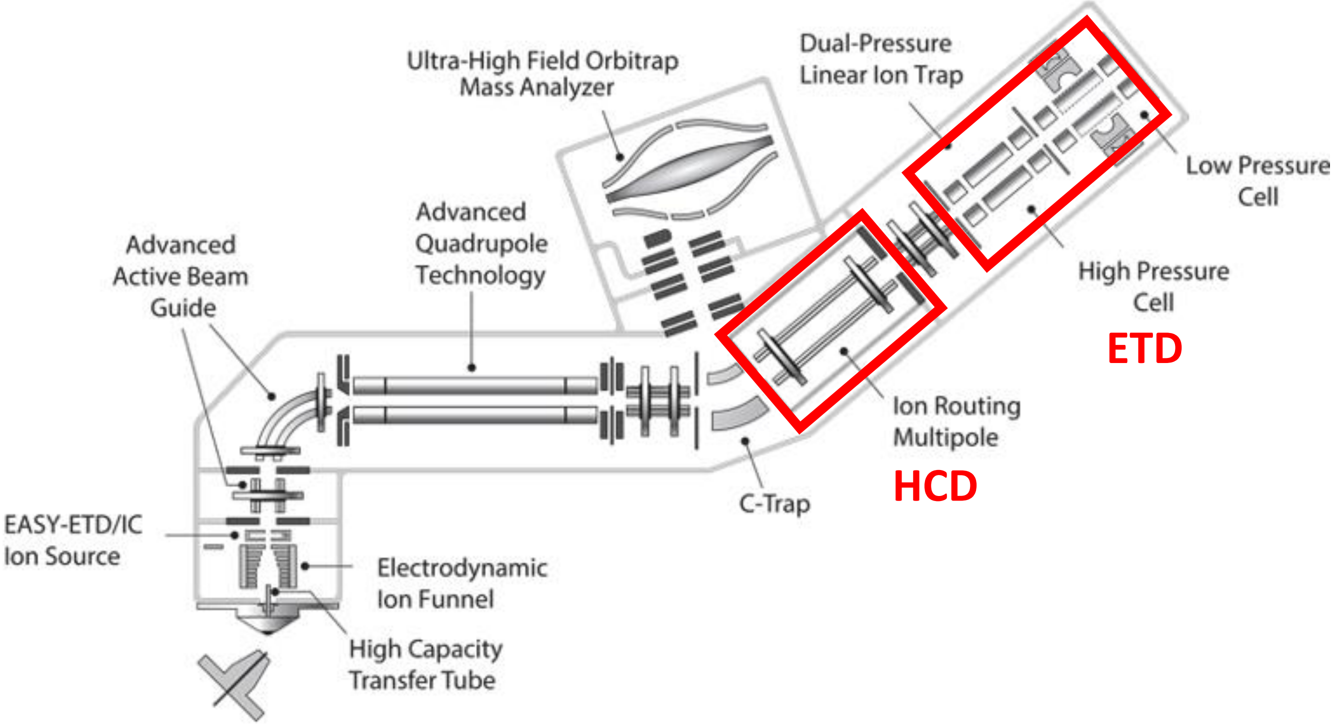


…PIDETEQGSYNNTSGEQMLIIWGVHHR…

Isolation of Adjacent Sequons - EThcD

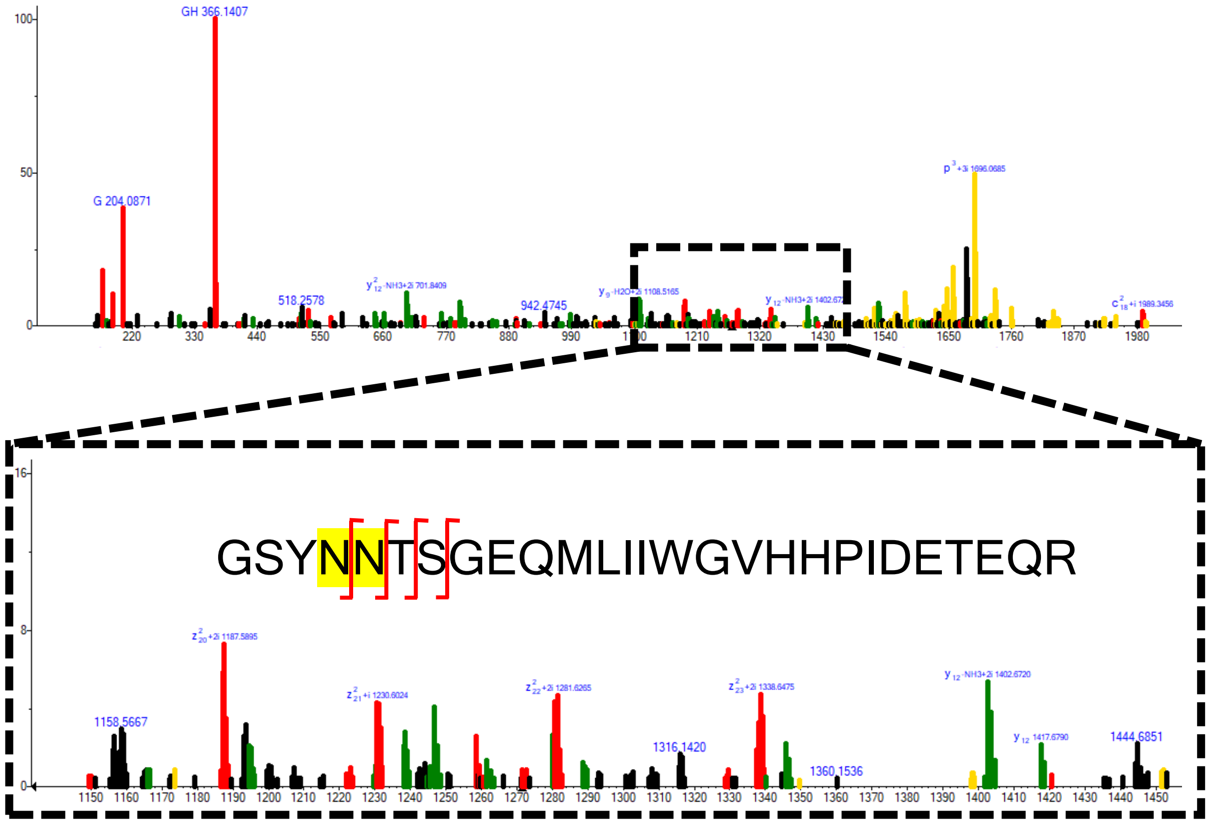
- ETD (Electron Transfer Dissociation)
- HCD (High-Energy Collisional Dissociation)

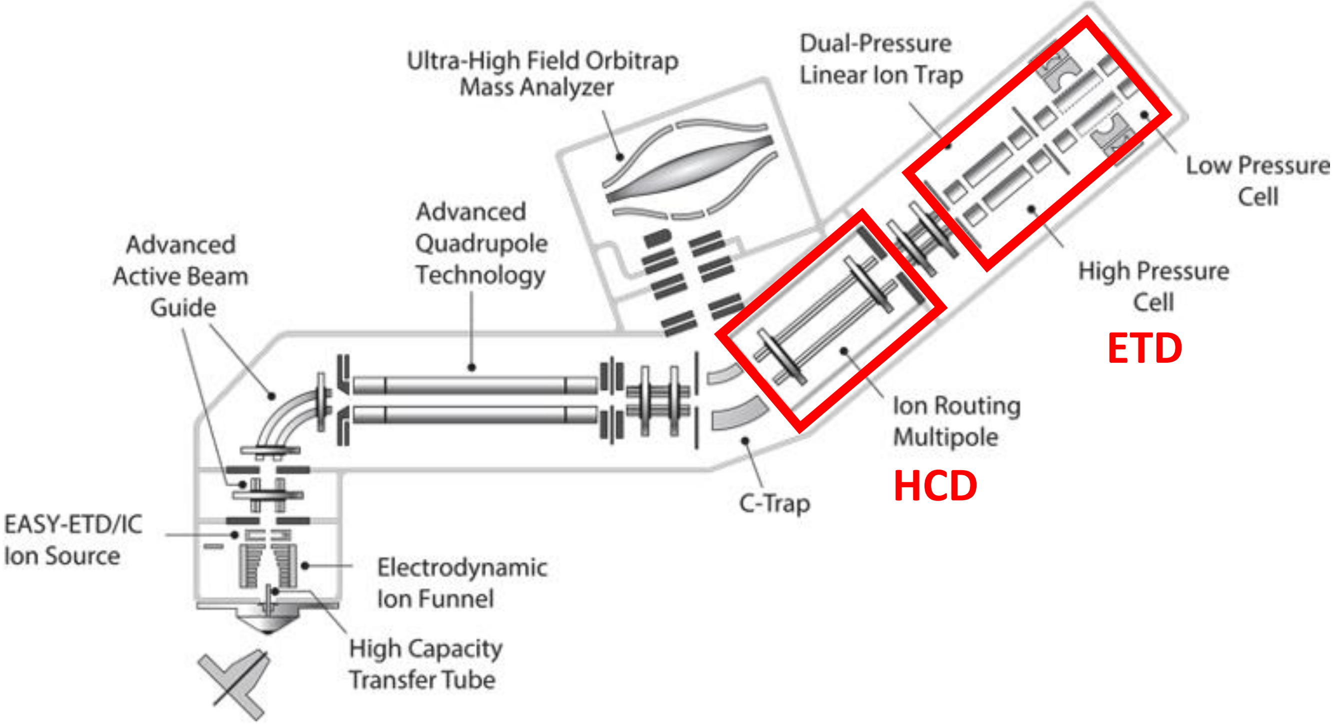




Isolation of Adjacent Sequons - EThcD

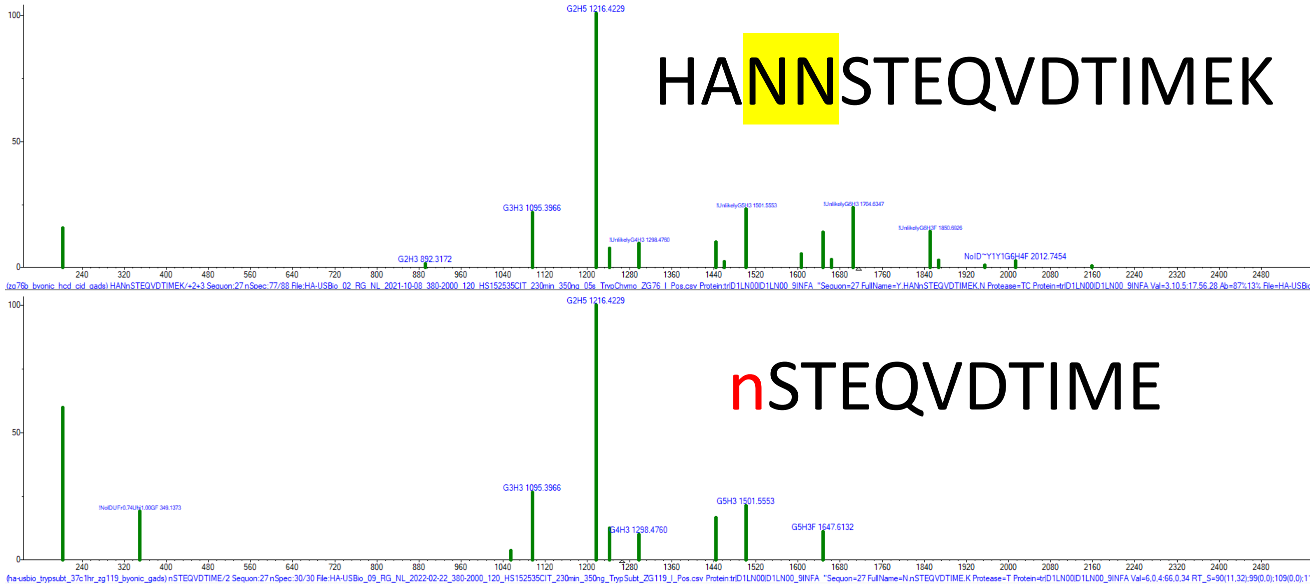
- ETD (Electron Transfer Dissociation)
- HCD (High-Energy Collisional Dissociation)





Isolation of Adjacent Sequons - Subtilisin

- *Bacillus subtilis* – extracellular serine endopeptidases



HANNSTEQVDTIMEK

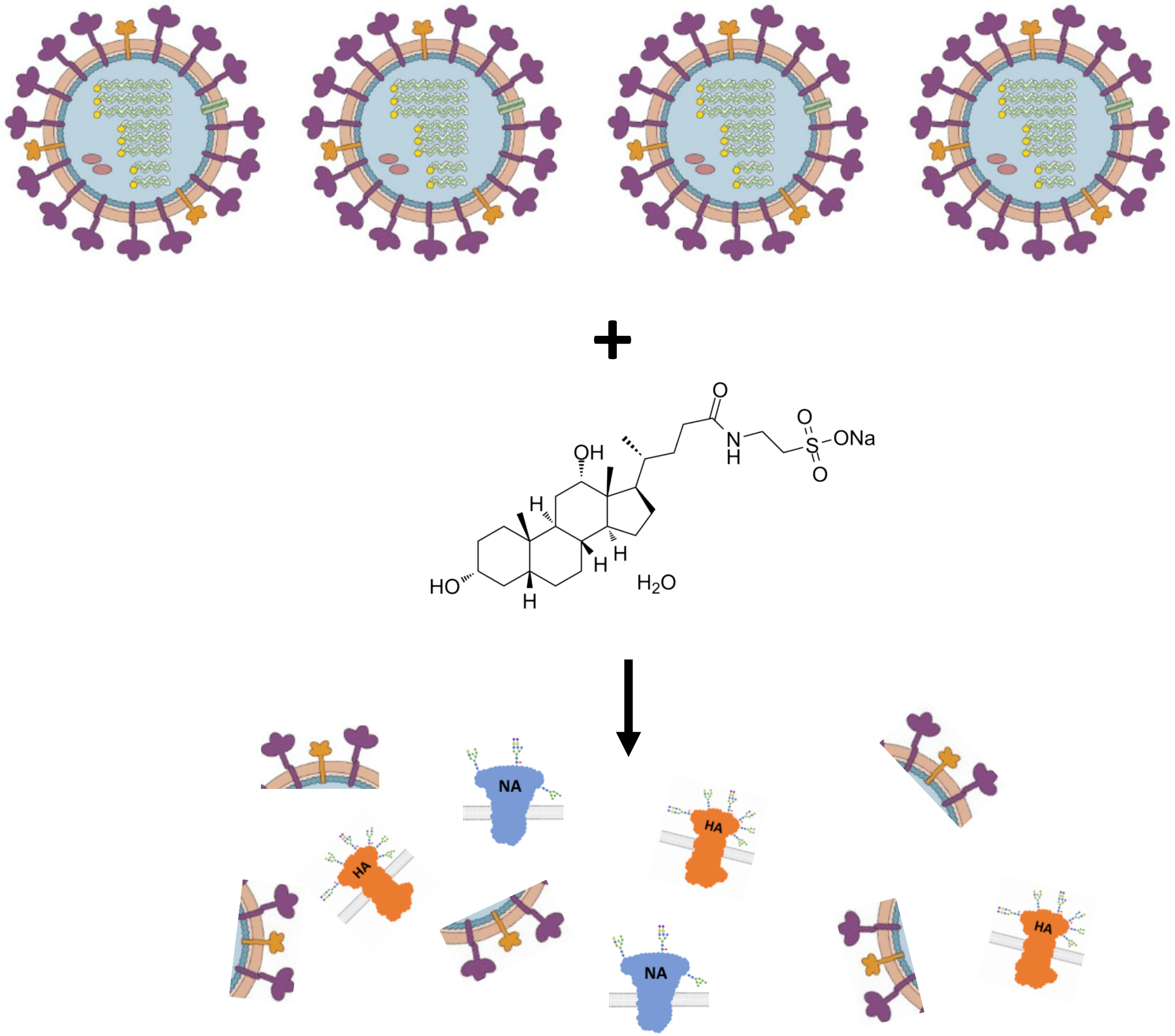
nSTEQVDTIME

DQICIGYHAn ???

Next Steps: Influenza Vaccines

- Afluria Quadrivalent
    - Split virion
    - 60 mg HA per dose (15 mg per strain)
    - Thimerosal
    - Mercury
    - Sodium Chloride
    - Sodium Phosphate
    - Potassium Phosphate
    - Potassium Chloride
    - Calcium Chloride
    - Sodium Taurodeoxycholate
    - Ovalbumin
    - Sucrose………

- 67 sequons within all HA and NA proteins

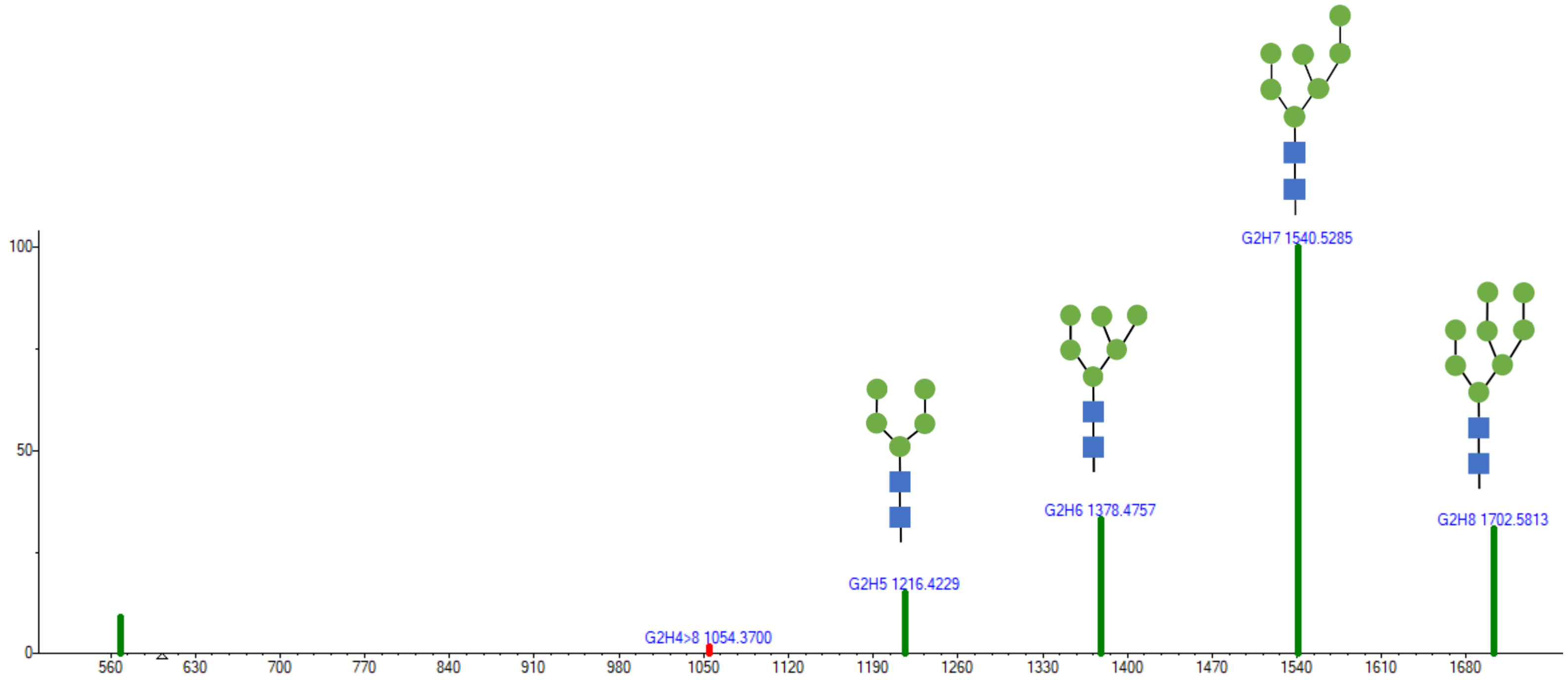


Influenza Vaccines – Challenges and Solutions

```
MKAIIVLL------------MVVTSNADRICTGITSSNSPHVVKTATQGEVNVTGVIPLTT
MKAIIVLL------------MVVTSNADRICTGITSSNSPHVVKTATQGEVNVTGVIPLTT
MKAILVV-----MLY----TFTTANADTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLE
MKTIIALSYILCLVFAQKIPGNDNSTATLCLGHHAVPNGTIVKTITNDRIEVTNATELVQ
**:*:.:                 .:  :* *  :  .   *.*  : .: **    *

TPTKSHFANLKGTETRGKLCPKCLNCTDLDVALSRPKCTGKIPSARVSILH-EVRPVTSG
TPTKSYFANLKGTRTRGKLCPDCLNCTDLDVALGRPMCVGTTPSAKASILH-EVRPVTSG
DKHNGKLCKLR-----GVAPLHLGKCNIAGWILGNPECESLSTARSWSYIVETSNSDNGT
NSSIGEICDSP-----H-QILDGGNCTLIDALLGDPQCDGFQN-KEWDLFVERSR-ANSN
    . :..            .  :*.  .  *. * * .       . :    .  ..

CFPIMHD-RTKIRQLPNLLRGYEHVRLSTHNVINAEDAPGGPYEIGTSGSCPNITNGNGF
CFPIMHD-RTKIRQLPNLLRGYEKIRLSTQNVIDAEKAPGGPYRLGTSGSCPNATSKIGF
CYPGDFINYEELREQLSSVSSFERFEIF----PKTSSWPNHDSDNGVTAACPHAG-AKSF
CYPYDVPDYASLRSLVASSGTL---EFK----NESFNWTGV-KQNGTSSACIRGS-SSSF
*:*       .:*.           .:      .: .  .     *.:.:* .     .*

FATMAWAVPKN--KTATNPLTIEVPYICTEGEDQITVWGFHSDNEIQ-MAKLYGDSKPQK
FATMAWAVPKDNYKNATNPLTVEVPYICTEGEDQITVWGFHSDNKTQ-MKSLYGDSNPQK
YKNLIWLVKKGKSY------PKINQTYINDKGKEVLVLWGIHHPPTIADQQSLYQNADAYV
FSRLNWLTHLNYTY------PALNVTMPNNEQFDKLYIWGVHHPSTDKDQISLFAQPSGRI
:  : * .  .          ::      :  : : :**.*         .*: : .

FTSSANGVTTHYVSQIGGFPNQTEDGGLPQSGRIVVDYMVQKSGKTGTITYQRGILLPQK
FTSSANGVTTHYVSQIGDFPDQTEDGGLPQSGRIVVDYMMQKPGKTGTIVYQRGVLLPQK
FVG-TSRYSKKFKPEIATRPKVR-----DQEGRMNYYWTLVEPGDKITFEATGNLVAPRY
TVS-TKRSQQAVIPNIGSRPRIR-----DIPSRISIYWTIVKPGDILLINSTGNLIAPRG
 .. :.        :*.  *           .*:   : : : *.   :    .:: *:

VWCA--SGRSKVIKGSLPLIGEADCLHEKYGGLNKSKPYYTGEHAKAIGNCPIWVKT-PL
VWCA--SGRSKVIKGSLPLIGEADCLHEEYGGLNKSKPYYTGKHAKAIGNCPIWVKT-PL
AFTMERDAGSGIIISDTPVHDCNTTCQTPEGAINTSLPF-QNVHPITIGKCPKYVKSTKL
YFKIR-SGKSSIMRSDAPIGKCKSECITPNGSIPNDKPF-QNVNRITYGACPRYVKQSTL
 :    .. * :: .. *:           *.: .. *:  . :  : * ** :**   *

KLANGTKYRPPAKLLKERGFFGAIAGFLEGGWEGMIAGWHGYTSHGAHGVAVAADLKSTQ
KLANGTKYRPPAKLLKERGFFGAIAGFLEGGWEGMIAGWHGYTSHGAHGVAVAADLKSTQ
RLATGLRNVPS---IQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSGYAADLKSTQ
KLATGMRNVPE---KQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLKSTQ
:**.* :  *     : **:*******:*.** **: **:*:  :. .* . ********

EAINKITKNLNSLSELEVKNLQRLSGAMDELHNEILELDEKVDDLRADTISSQIELAVLL
EAINKITKNLNSLSELEVKNLQRLSGAMDELHNEILELDEKVDDLRADTISSQIELAVLL
NAIDKITNKVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWTYNAELLVLL
AAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRVQDLEKYVEDTKIDLWSYNAELLVAL
 **::*. ::* :      ::  :   :..:. .: :*:: *:*   *  : : ** * *

SNEGIINSEDEHLLALERKLKKMLGPSAVEIGNGCFETKHKCNQTCLDRIAAGTFDAGEF
SNEGIINSEDEHLLALERKLKKMLGPSAVDIGNGCFETKHKCNQTCLDRIAAGTFNAGEF
ENERTLDYHDSNVKNLYEKVRNQLKNNAKEIGNGCFEFYHKCDNTCMESVKNGTYDYPKY
ENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNETYDHNVY
.*:  ::  *..:  * .* :: *  .* ::*****:  ***:::*:  :   *::   :

SLPTFDS-LNITAASLNDDGLDNHTILLYYSTAASSLAVTLMIAIFVVYMVSRDSVSCSI
SLPTFDS-LNITAASLNDDGLDNHTILLYYSTAASSLAVTLMLAIFIVYMVSRDNVSCSI
SEEAKLNREKIDGVKL--DSTRIYQILAIYSTVASSLVLVVSLGAISFWMCSNGSLQCRI
RDEALNNRFQIKGVEL--KSGYKDWILWI-SFAMSCFLLCIALLGFIMWACQKGNIRCNI
   :  .  :* ...*  ..     **   * . *.: : : :  : .:  ....: * *
```

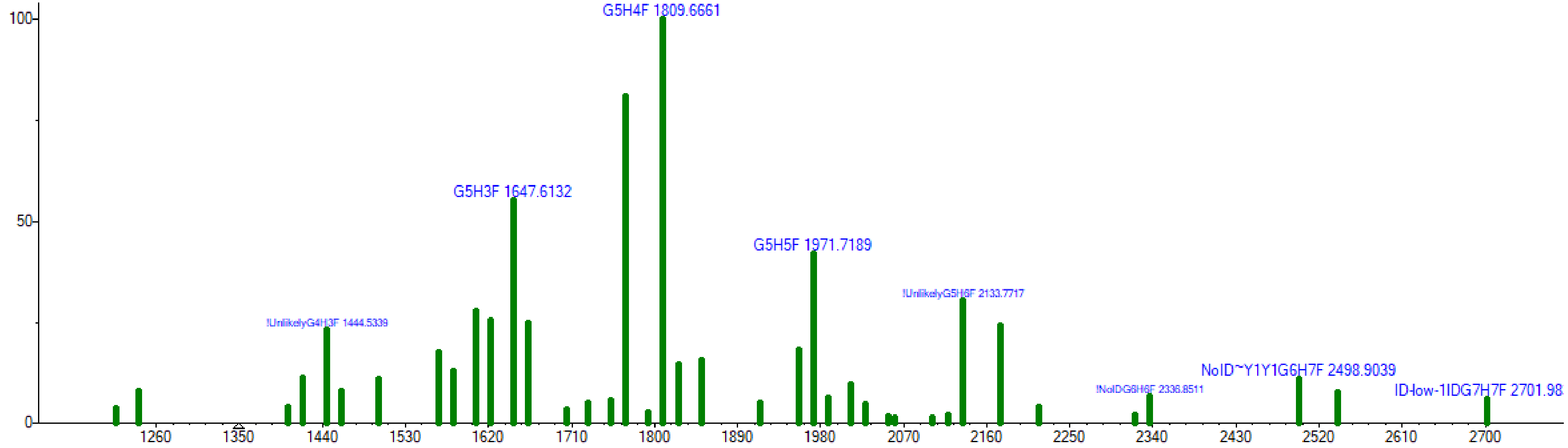
Afluria GADS

- ~8 of the 40 sequons for the four hemagglutinin proteins identified.
- The GADS are majority high-mannose.
- No GADS were identified for NA proteins, probably due to a lower abundance of NA on the virus.



Afluria GADS

- ~8 of the 40 sequons for the four hemagglutinin proteins identified.
- The GADS are majority high-mannose.
- No GADS were identified for NA proteins, probably due to a lower abundance of NA on the virus.



Conclusions

- Glycosylation of replicates is most similar and between strains is least similar.
- Adjacent sequons may be isolated using EThcD fragmentation or non-specific cleavage using the protease subtilisin.
- Most glycans are high-mannose in egg-based quadrivalent vaccines.
- Next steps:
  - Optimize methods for influenza vaccine digestion.
  - Build reference MS2 and GADS libraries
  - Apply methods to other viral glycoproteins (HIV, Herpesvirus, Ebola…)

Acknowledgments

Software

- Yuri Mirokhin
- Sergey Sheetlin
- Dmitrii Tchekhovskoi
- Xiaoyu Yang
- Guanghui Wang
- Stephen Stein

Lab mentoring

- Yi Liu
- Jane Zhang

Data acquisition/analysis

- Meghan Burke Harris
- Connie Remoroza

References

[1]Jung, H. E., & Lee, H. K. (2020). Host protective immune responses against influenza A virus infection. *Viruses*, *12*(5), 504.

[2]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.

[3]https://www.cdc.gov/flu/vaccines-work/effectiveness-studies.htm

[4]Zost, S. J., Parkhouse, K., Gumina, M. E., Kim, K., Perez, S. D., Wilson, P. C., ... & Hensley, S. E. (2017). Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. *Proceedings of the National Academy of Sciences*, *114*(47), 12578-12583.

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., & Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. Science, 369(6501), 330-333.

Chang, D., & Zaia, J. (2019). Why glycosylation matters in building a better flu vaccine. Molecular & Cellular Proteomics, 18(12), 2348-2358.

Questions?

zachary.goecker@nist.gov

Supplemental Slides

Dot Product

Cosine of the angle between spectra represented as vectors

$$\frac{(\Sigma W_L W_U)^2}{\Sigma W_L^2 \, \Sigma W_U^2}$$

W_L = Weighted intensity of library

W_U = Weighted intensity of unknown

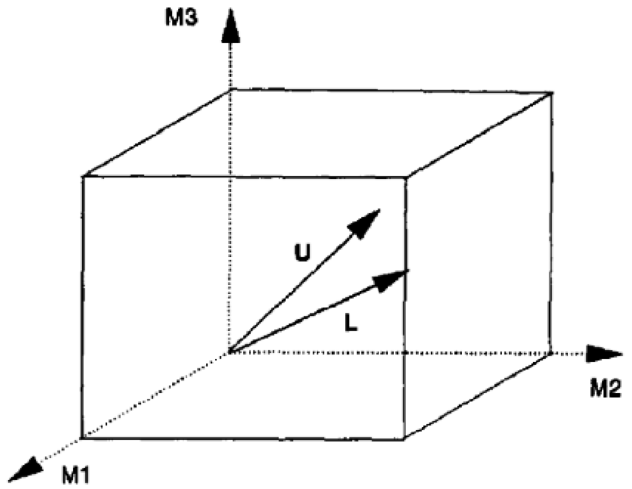


Figure 1. Vector representation of a hypothetical three-peak unknown (U) and library (L) mass spectrum in three-dimensional space (peaks have mass M1, M2, and M3).

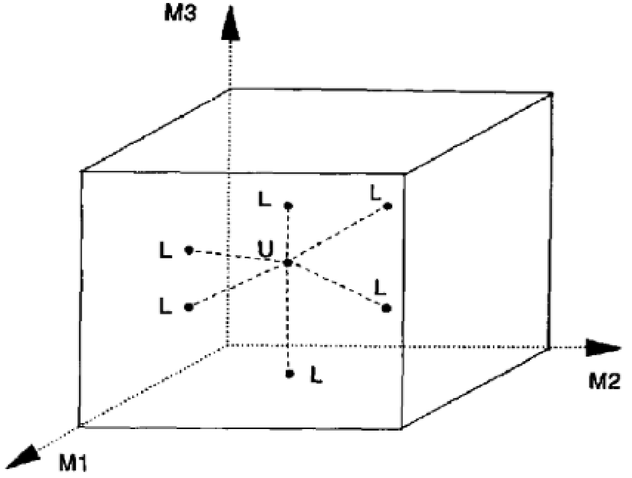
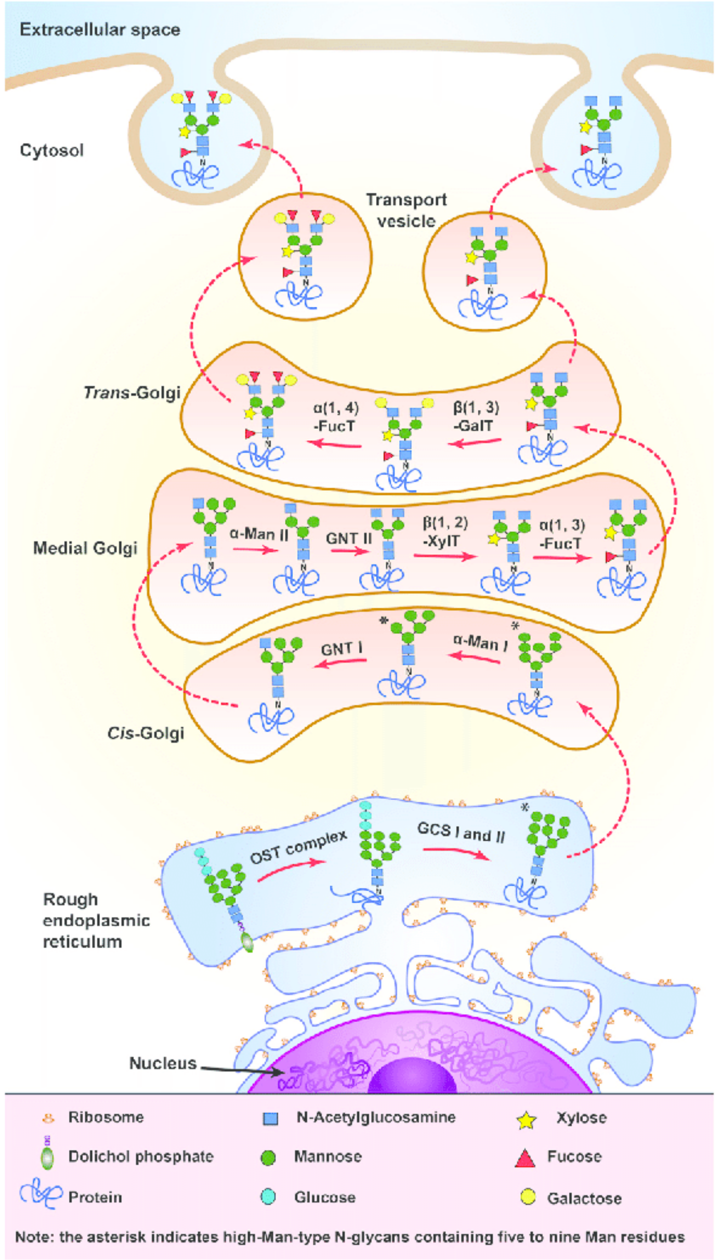


Figure 2. Point representation of library search results (L) for a hypothetical three-peak unknown (U) spectrum (masses M1, M2, and M3)

Supplemental Slides

Glycobiology

- In viruses
  - Immune evasion
  - Host cell attachment
- In mammalian cells
  - Protein folding
  - Protein stabilization
  - Communication
  - Function

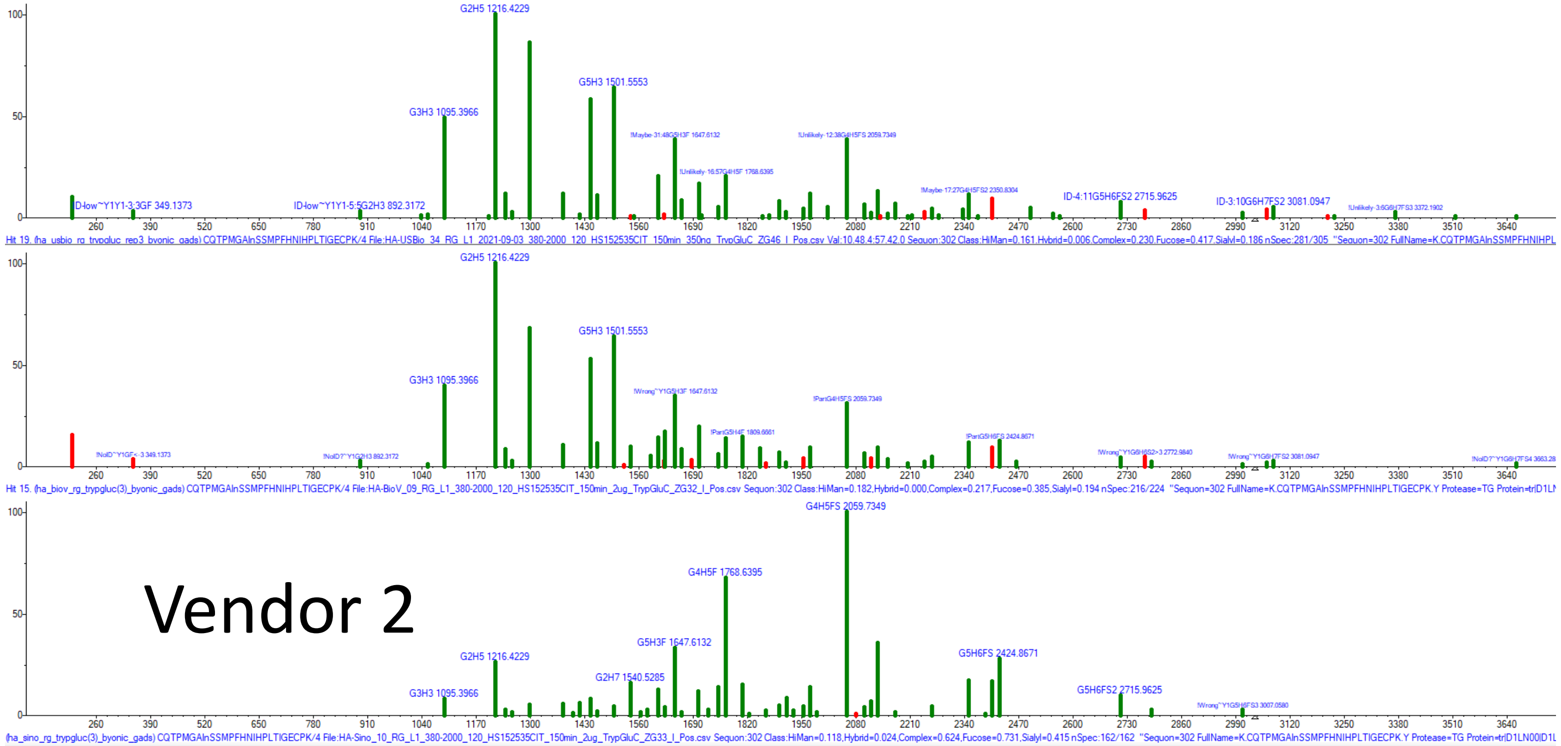


Note: the asterisk indicates high-Man-type N-glycans containing five to nine Man residues

Supplemental Slides

Vendor Comparison

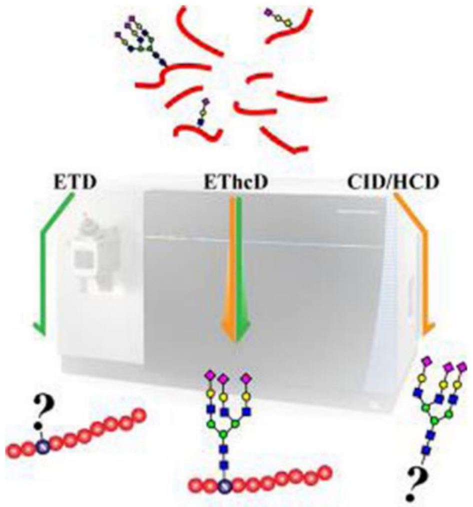
Glycan distribution was skewed toward higher mass (sialylated complex) glycans for vendor 2 and toward lower mass glycans (oligomannose and hybrid) for vendor 1 and vendor 3.

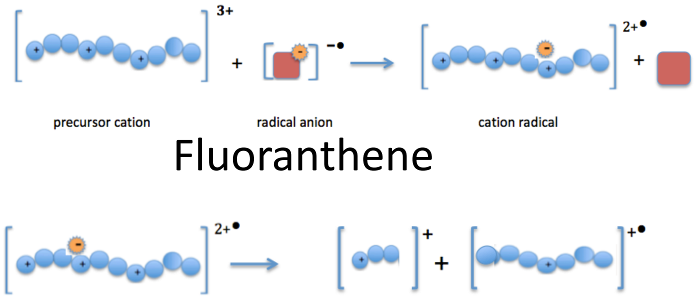


Vendor 2

Supplemental Slides

EThcD





Stepped HCD

Peptide bond
Glycosidic bond
N-glycosidic bond