Measuring Variation in Glycosylation Among Influenza Vaccine Glycoproteins

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

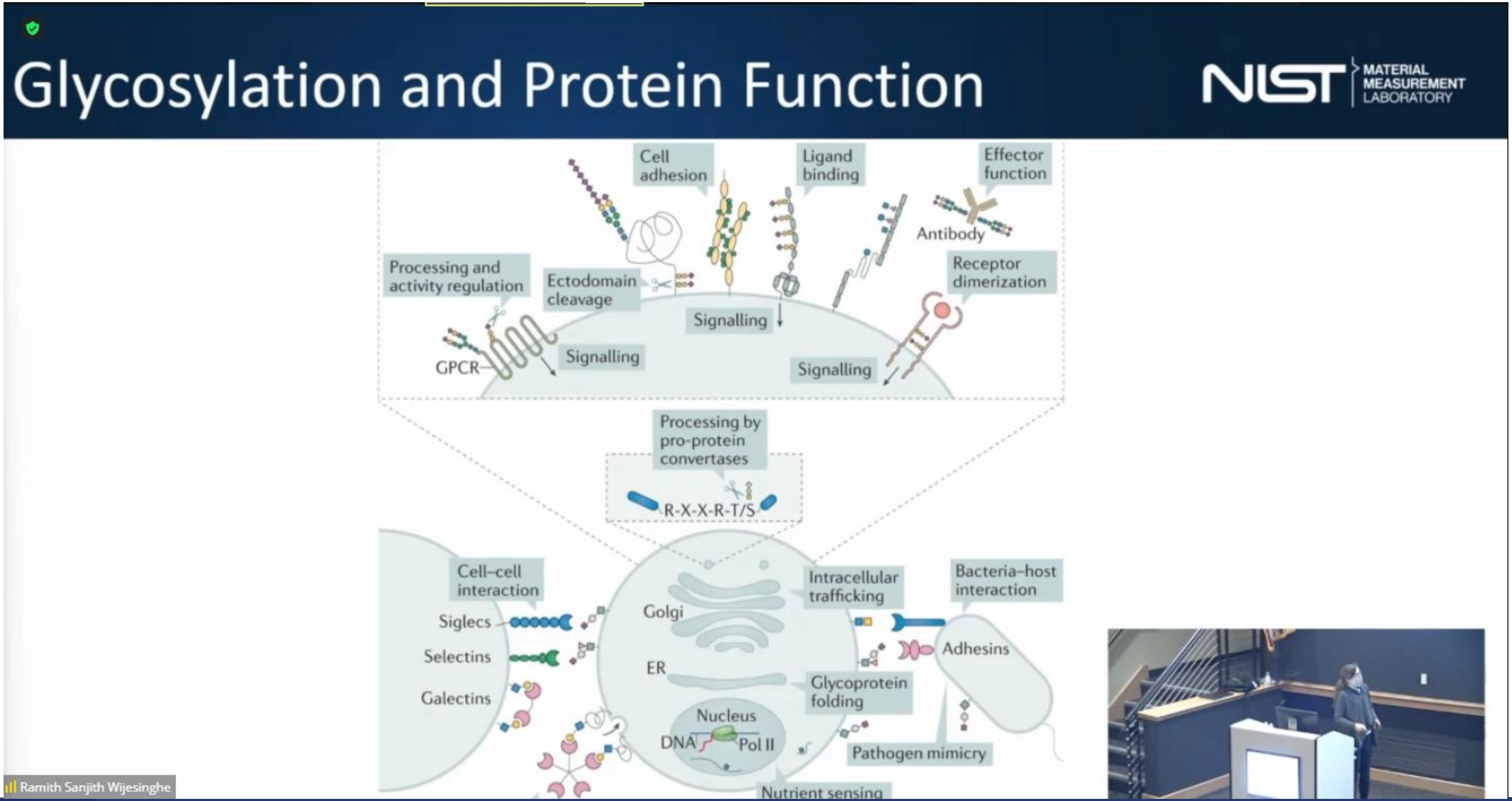
BMD Staff Seminar

Jan 31, 2023

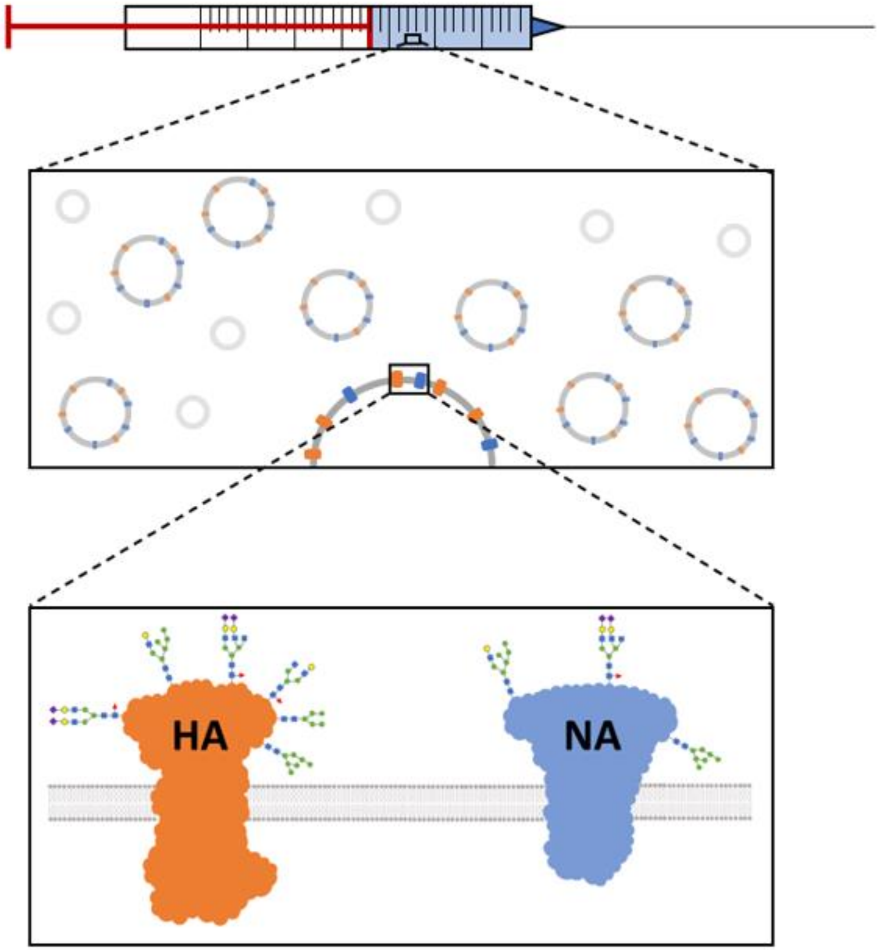
NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Mass Spectrometry
Data Center

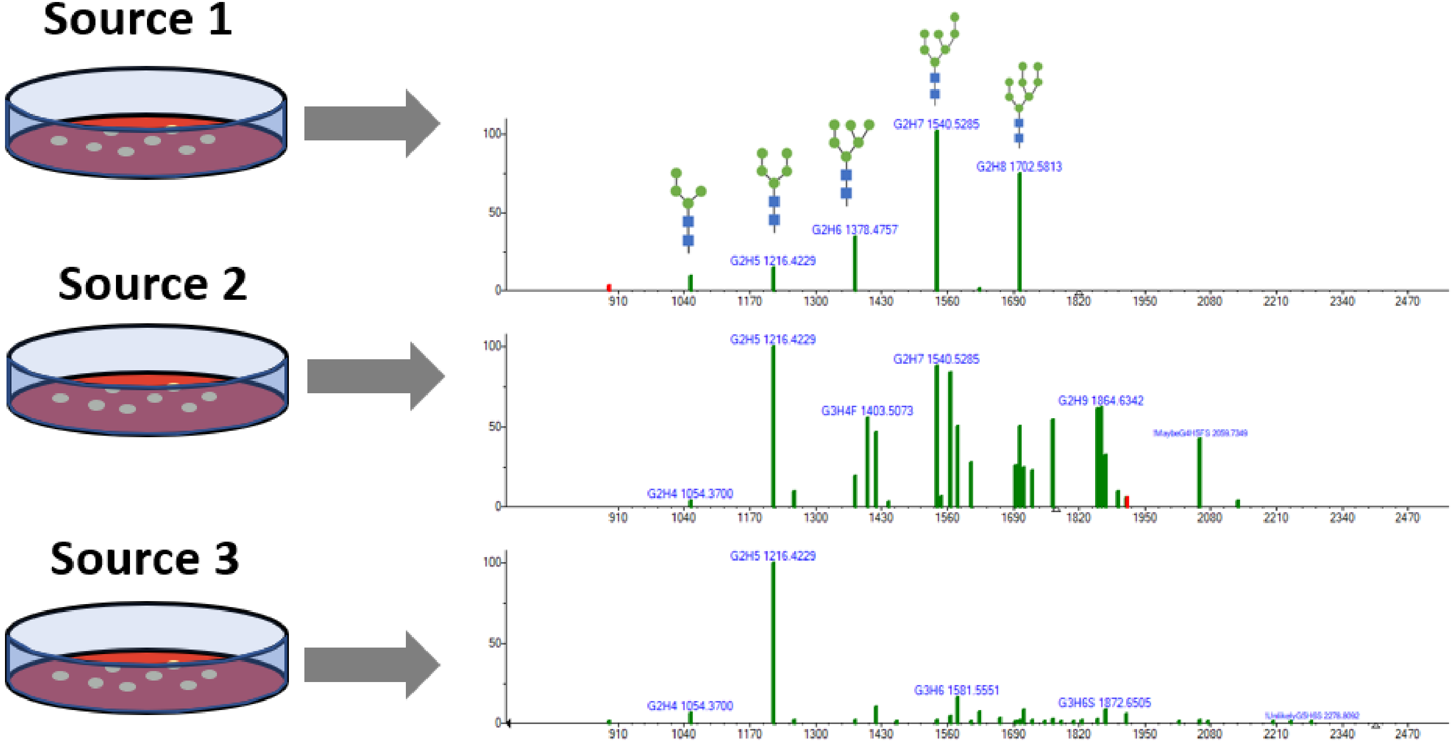
BMD Seminar Déjà Vu?



Glycosylation and Protein Function

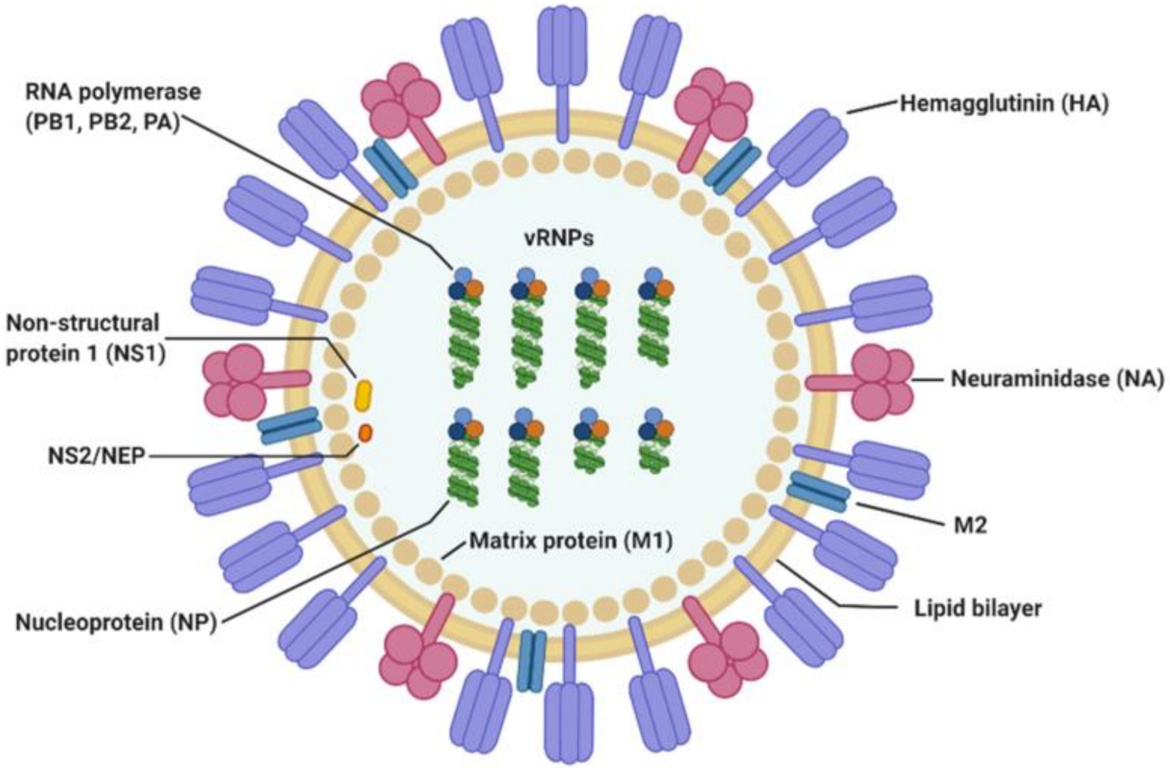


<u>G</u>lycopeptide <u>A</u>bundance <u>D</u>istribution <u>S</u>pectrum



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.



Jung et al., 2020

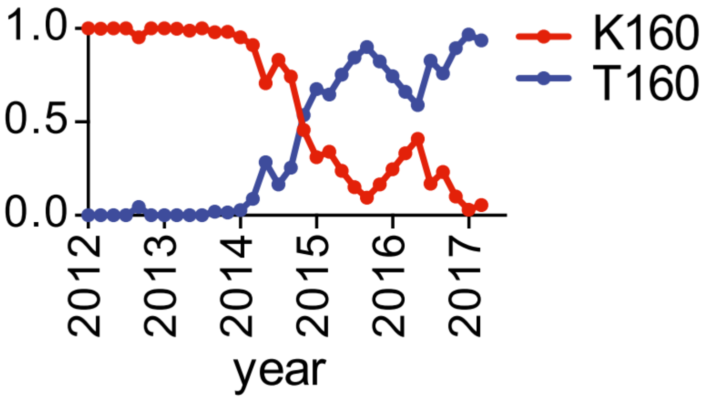
Why Glycosylation Matters For Vaccines

...THHLNFKYPAL...

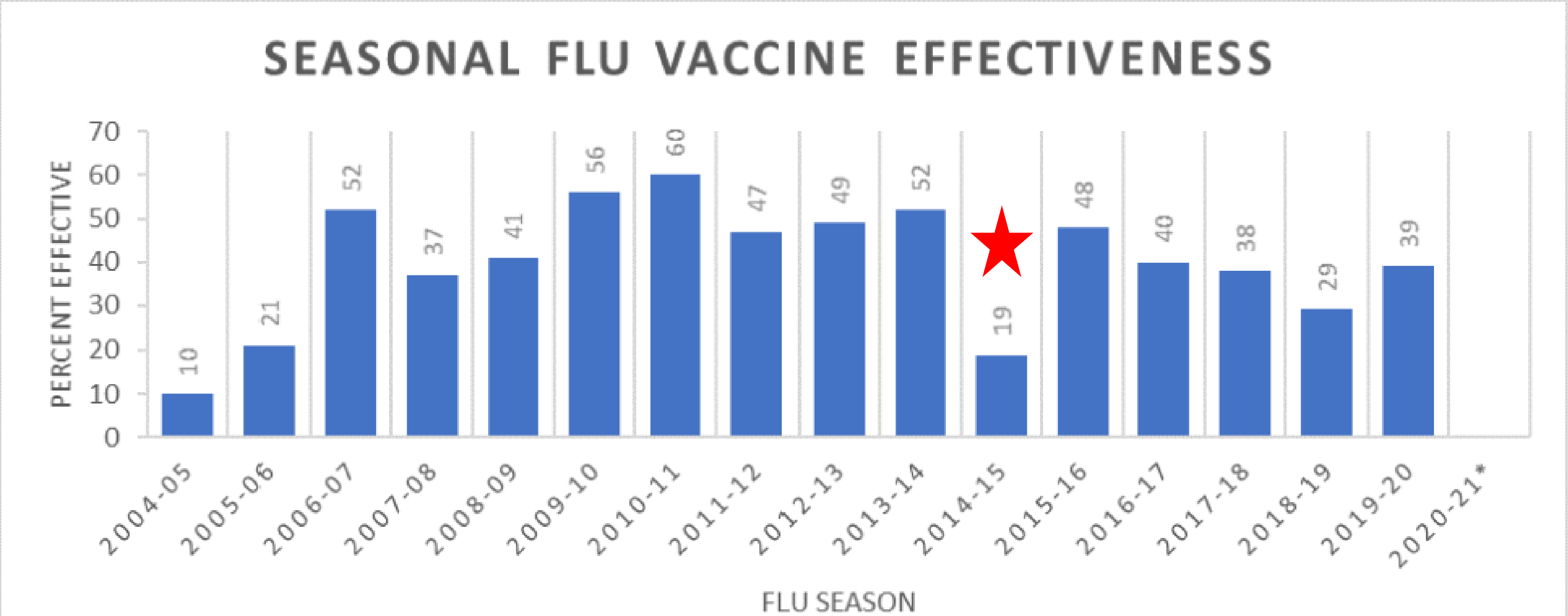
...THLNFTYPAL...

Sequon motif

NXT/S X≠P

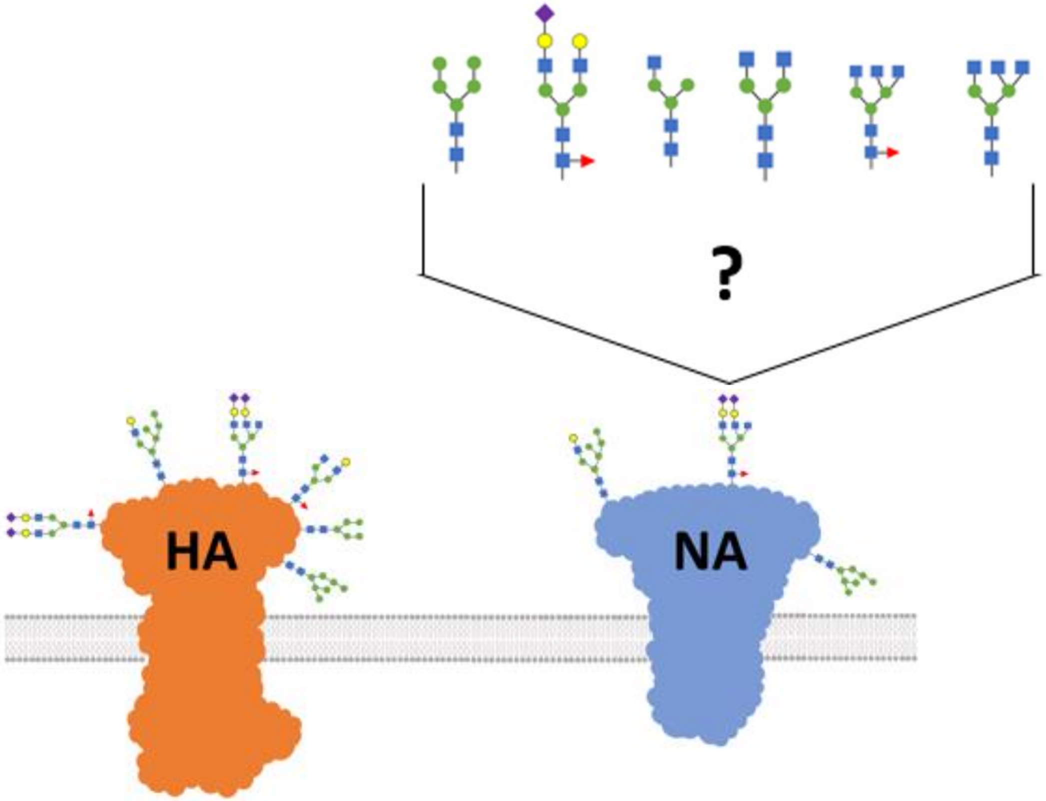


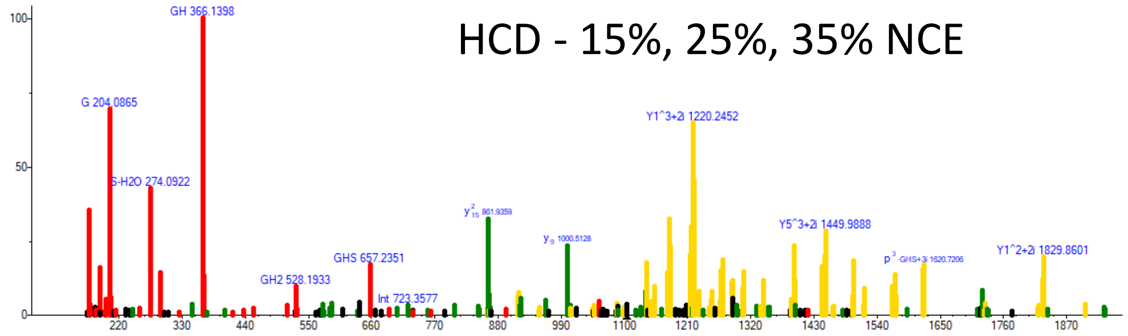
Zost et al., 2017



www.cdc.gov

Site-Specific Glycosylation



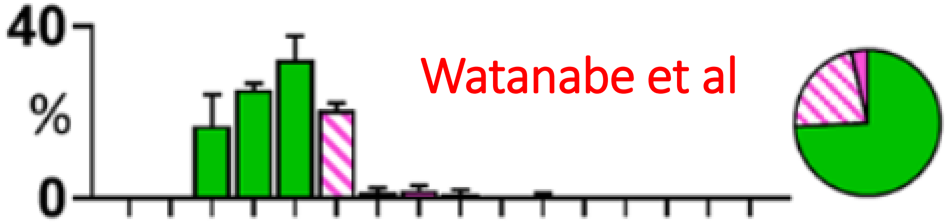


HCD - 15%, 25%, 35% NCE

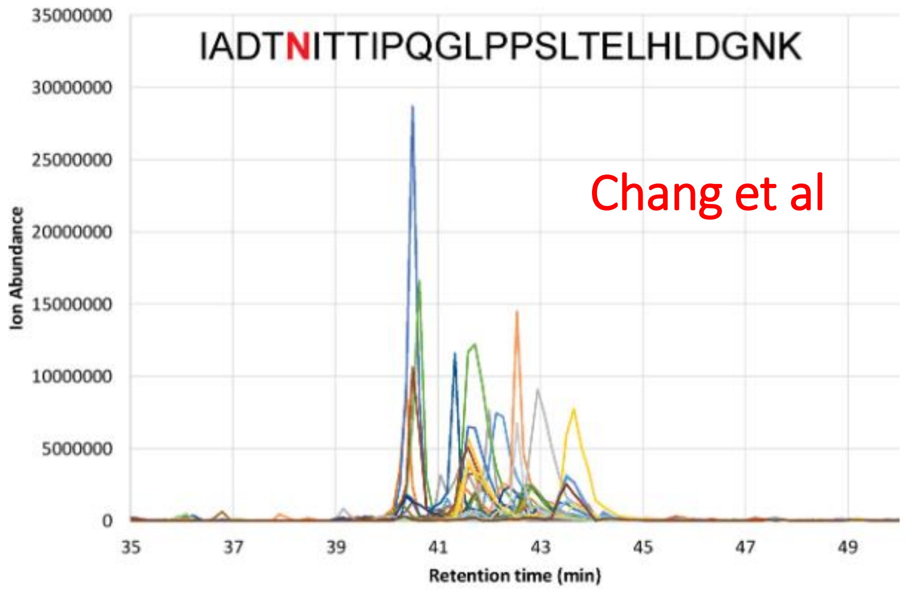
Oxonium ions
Peptide backbone ions
Glycopeptide ions

Visualizing Glycan Distributions

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



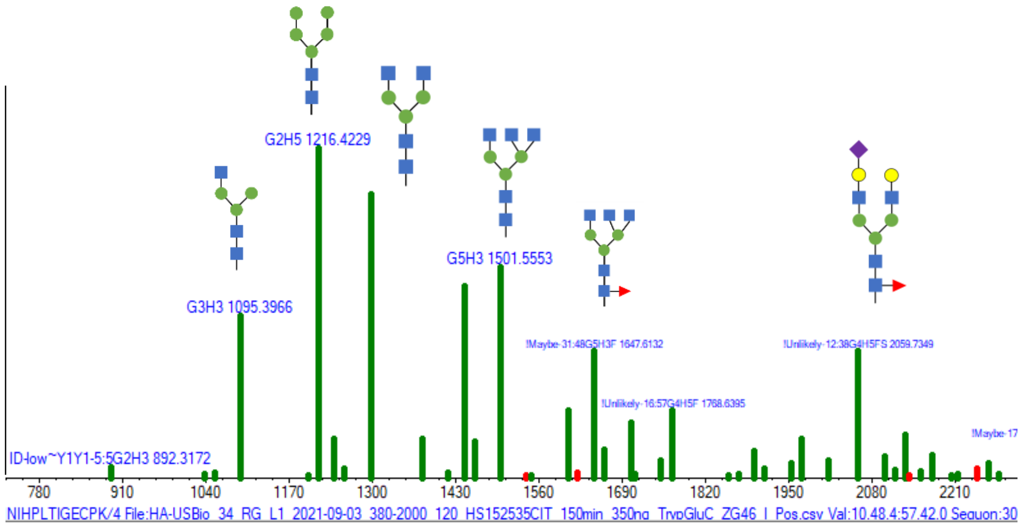
Watanabe et al



Chang et al

Visualizing Glycan Distributions

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

Data Validation

- Hyperscore
- Retention time
- Contingent ion trap scan
- Oxonium ions
- Number of instances
- Glycopeptide (Y) ion series
- MS1 XIC
- MS1 Isotopic envelope
- MS2 purity
- Precursor mass overlap

$$\text{hyperscore} = \log\left(N_b\,!\,N_y\,!\sum_{i=1}^{N_b} I_{\mathrm{b,i}} \sum_{i=1}^{N_y} I_{\mathrm{y,i}} \right)$$

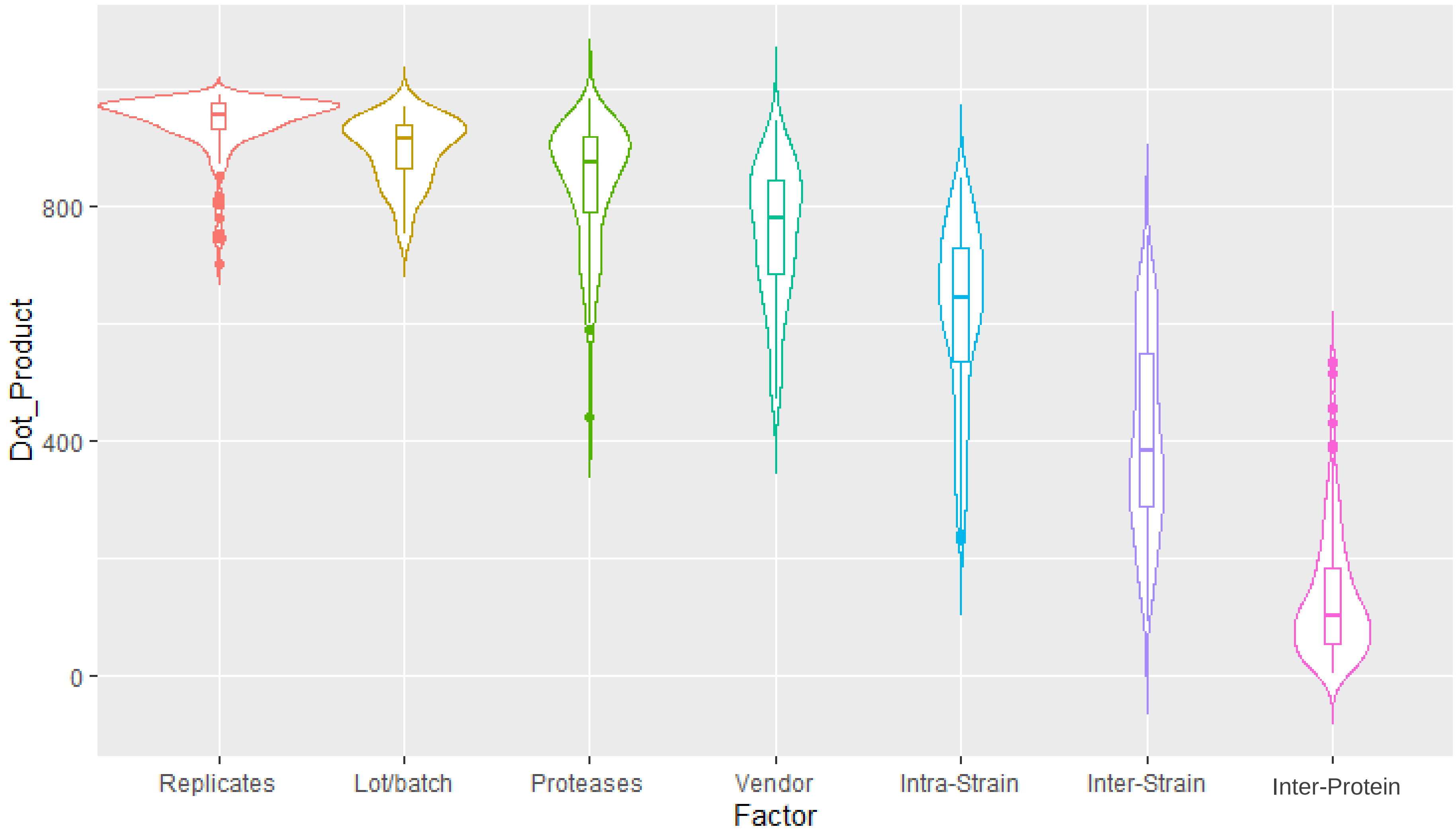
Recombinant Proteins

TABLE 1 Recombinant proteins analyzed for site-specific glycosylation

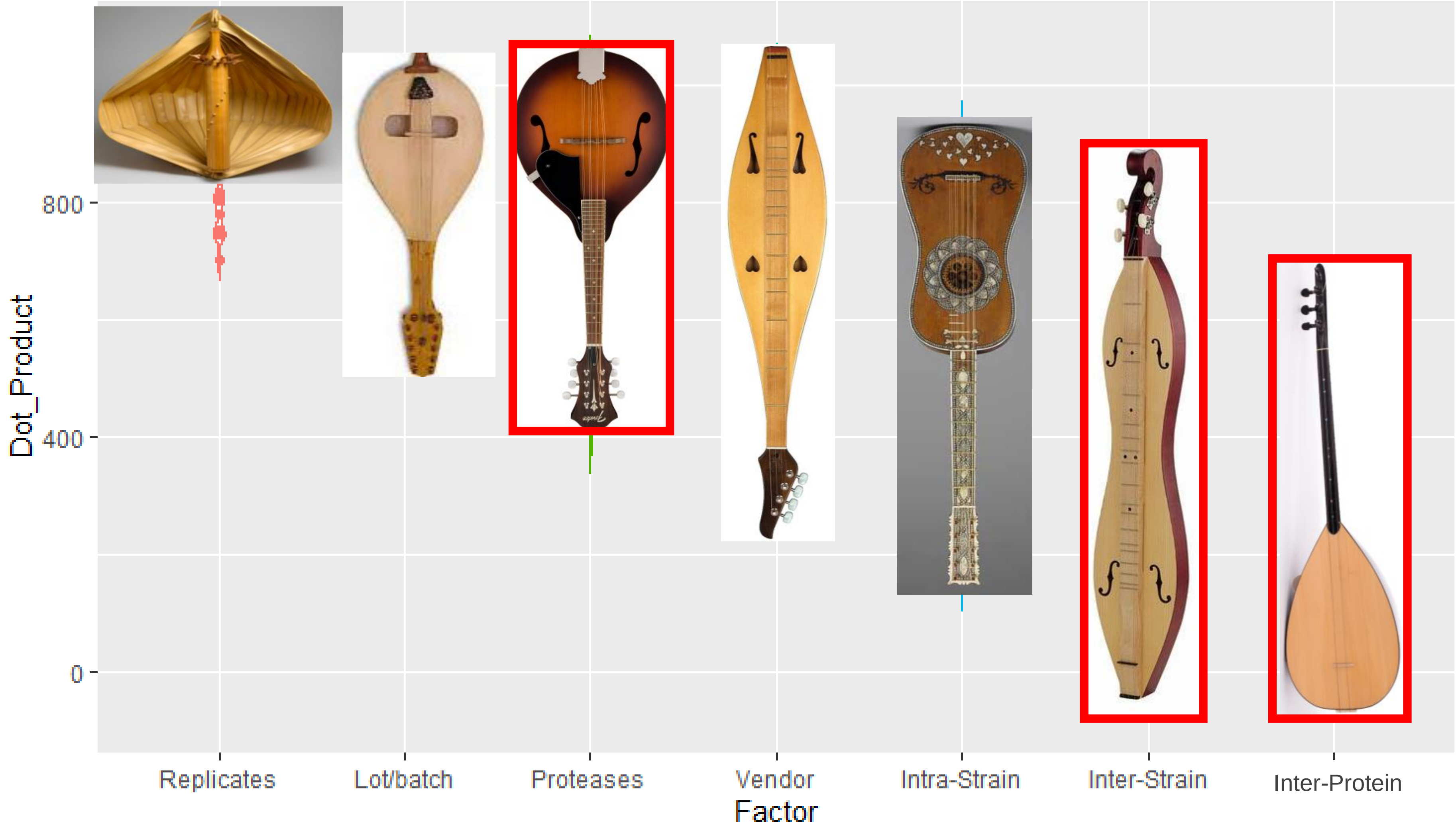
Abbreviation	Protein	Strain	Subtype	Vendors	Number of sequons	Protein mass* (kDa)
HA-CA09	**HA**	**A/California/04/2009**	**H1N1**	**Creative Biomart**	**8**	**63**
HA-NC99	**HA**	**A/New Caledonia/20/1999**	**H1N1**	**Sino Biological**	**10**	**63**
HA-JP57	**HA**	**A/Japan/305/1957**	**H2N2**	**Creative Biomart**	**8**	**63**
HA-HK14	**HA**	**A/Hong Kong/485197/2014**	**H3N2**	**Biovision**	**13**	**64**
HA-HK97	**HA**	**A/Hong Kong/483/1997**	**H5N1**	**BioVision, US Biological, Sino Biological**	**8**	**64**
NA-AZ08	**NA**	**A/Arizona/13/2008**	**H1N1**	**Sino Biological**	**9**	**52**
NA-TH04	**NA**	**A/Thailand/1(KAN-1)/2004**	**H5N1**	**BioVision, US Biological, Sino Biological**	**3**	**49**
NA-NL03	**NA**	**A/Netherlands/219/2003**	**H7N7**	**Creative Biomart**	**11**	**52**

*Unglycosylated

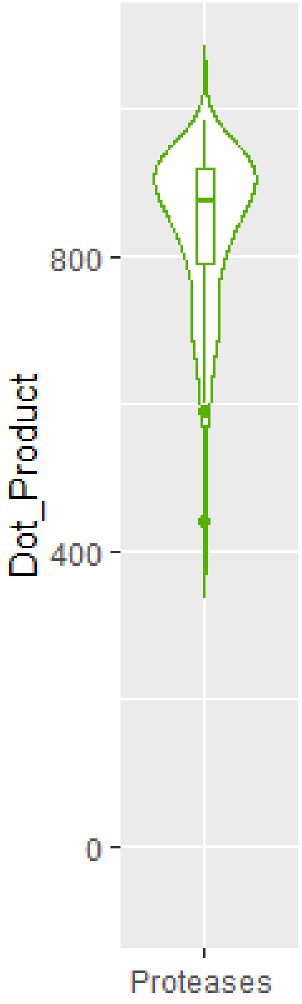
Review - Variation in glycosylation profile



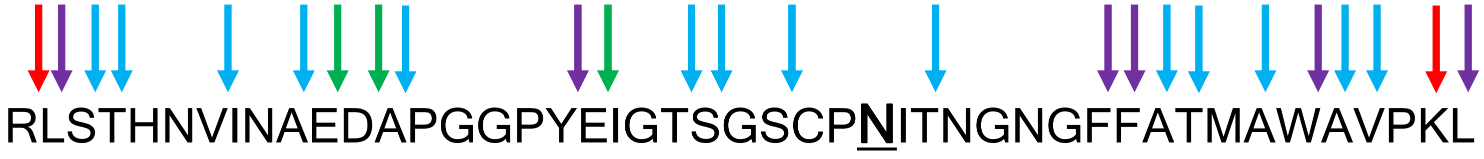
Review - Variation in glycosylation profile



Variation From Different Protease Digestions

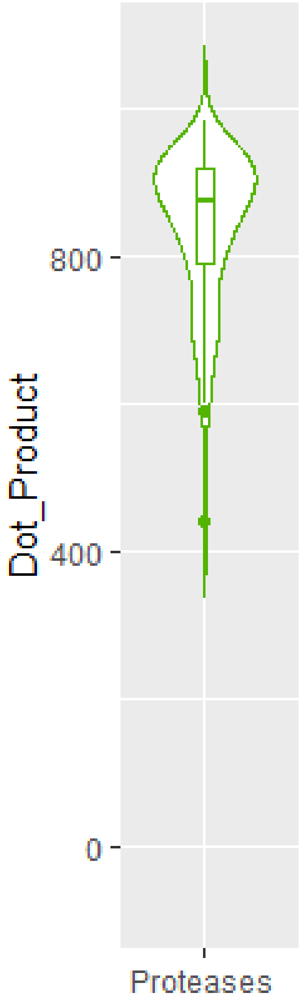


Protease Combination	Amino Acid Cleavage Sites
Trypsin + Lys-C	KR
Trypsin + Glu-C	KRED
Trypsin + Chymotrypsin	KRFWYL
Chymotrypsin + Glu-C	FWYLED
Chymotrypsin	FWYL
Alpha-lytic	TASV

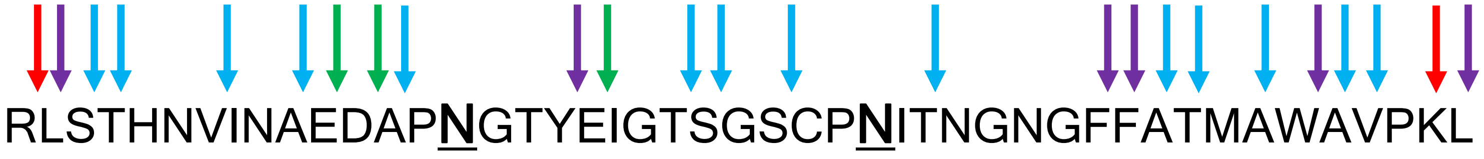


RLSTHNVINAEDAPGGPYEIGTSGSCPNITNGNGFFATMAWAVPKL

Variation From Different Protease Digestions

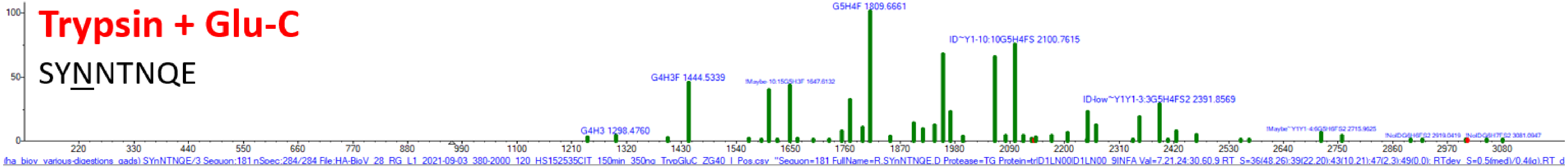


Protease Combination	Amino Acid Cleavage Sites
Trypsin + Lys-C	KR
Trypsin + Glu-C	KRED
Trypsin + Chymotrypsin	KRFWYL
Chymotrypsin + Glu-C	FWYLED
Chymotrypsin	FWYL
Alpha-lytic	TASV



RLSTHNVINAEDAPNGTYEIGTSGSCPNITNGNGFFATMAWAVPKL

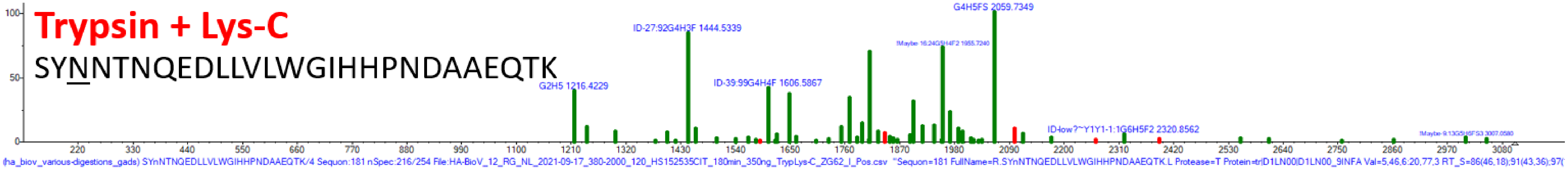
Variation From Different Protease Digestions



Trypsin + Glu-C

SYNNTNQE

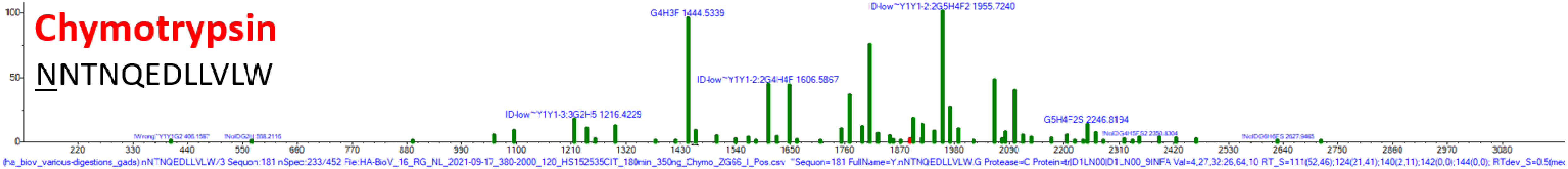
(ha_biov_various-digestions_gads) SYnNTNQE/3 Sequon:181 nSpec:284/284 File:HA-BioV_28_RG_L1_2021-09-03_380-2000_120_HS152535CIT_150min_350ng_TrypGluC_ZG40_I_Pos.csv "Sequon=181 FullName=R.SYnNTNQE.D Protease=TG Protein=tr|D1LN00|D1LN00_9INFA Val=7,21,24:30,60,9 RT_S=36(48,26);39(22,20);43(10,21);47(2,3);49(0,0); RTdev_S=0.5(med)/0.4(q),RT_d



Trypsin + Lys-C

SYNNTNQEDLLVLWGIHHPNDAAEQTK

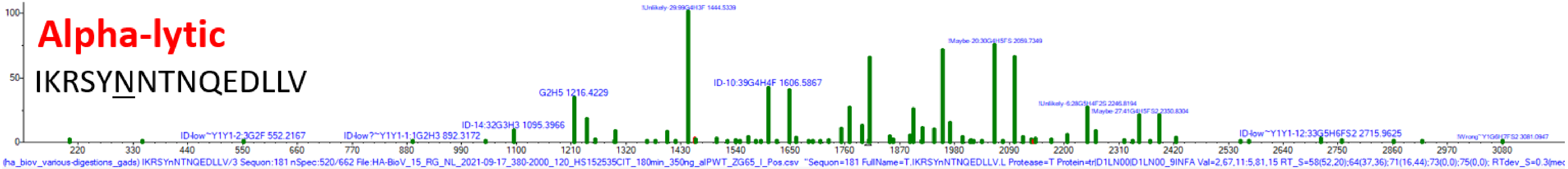
(ha_biov_various-digestions_gads) SYnNTNQEDLLVLWGIHHPNDAAEQTK/4 Sequon:181 nSpec:216/254 File:HA-BioV_12_RG_NL_2021-09-17_380-2000_120_HS152535CIT_180min_350ng_TrypLys-C_ZG62_I_Pos.csv "Sequon=181 FullName=R.SYnNTNQEDLLVLWGIHHPNDAAEQTK.L Protease=T Protein=tr|D1LN00|D1LN00_9INFA Val=5,46,6:20,77,3 RT_S=86(46,18);91(43,36);97(



Chymotrypsin

NNTNQEDLLVLW

(ha_biov_various-digestions_gads) nNTNQEDLLVLW/3 Sequon:181 nSpec:233/452 File:HA-BioV_16_RG_NL_2021-09-17_380-2000_120_HS152535CIT_180min_350ng_Chymo_ZG66_I_Pos.csv "Sequon=181 FullName=Y.nNTNQEDLLVLW.G Protease=C Protein=tr|D1LN00|D1LN00_9INFA Val=4,27,32:26,64,10 RT_S=111(52,46);124(21,41);140(2,11);142(0,0);144(0,0); RTdev_S=0.5(me

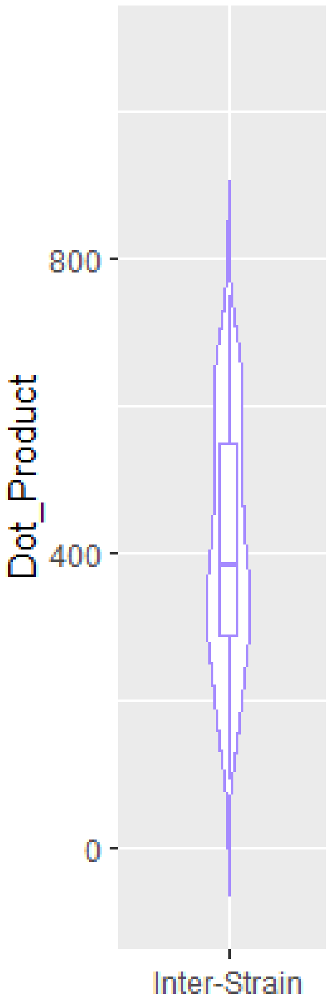


Alpha-lytic

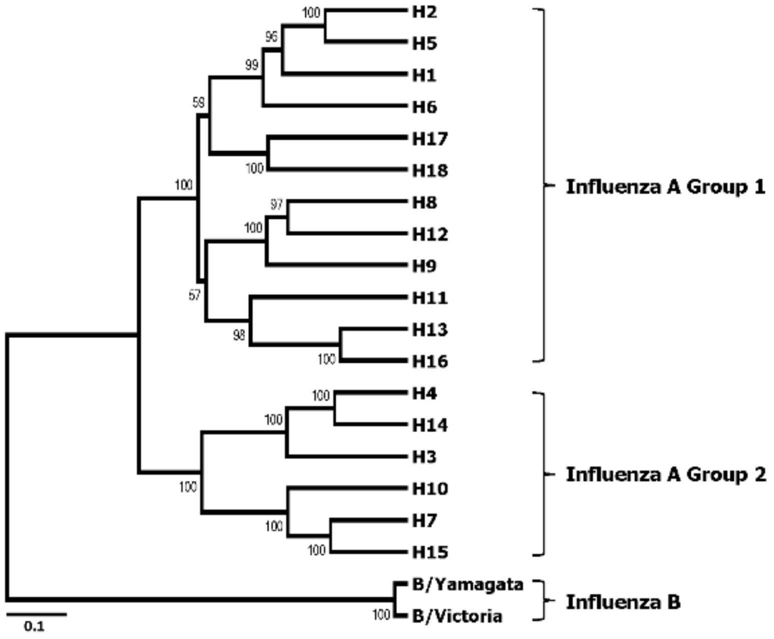
IKRSYNNTNQEDLLV

(ha_biov_various-digestions_gads) IKRSYnNTNQEDLLV/3 Sequon:181 nSpec:520/662 File:HA-BioV_15_RG_NL_2021-09-17_380-2000_120_HS152535CIT_180min_350ng_alPWT_ZG65_I_Pos.csv "Sequon=181 FullName=T.IKRSYnNTNQEDLLV.L Protease=T Protein=tr|D1LN00|D1LN00_9INFA Val=2,67,11:5,81,15 RT_S=58(52,20);64(37,36);71(16,44);73(0,0);75(0,0); RTdev_S=0.3(me

Variation From Different Strains



Protein	Strain	Subtype
HA	**A/California/04/2009**	**H1N1**
HA	**A/New Caledonia/20/1999**	**H1N1**
HA	**A/Japan/305/1957**	**H2N2**
HA	**A/Hong Kong/485197/2014**	**H3N2**
HA	**A/Hong Kong/483/1997**	**H5N1**
NA	**A/Arizona/13/2008**	**H1N1**
NA	**A/Thailand/1(KAN-1)/2004**	**H5N1**
NA	**A/Netherlands/219/2003**	**H7N7**



```
MK-TIIALSYILCLVFAQKIPGNDNSTATLCLGHHAVENGTIVKTITNDRIEVTNATELV
MKAKLLVLLCTFT----------ATYADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLL
MKAILVVLLYTFA----------TANADTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLL
-MEKIVLLLATVS----------LVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDIL
--MAIIYLILLFT----------AVRGDQICIGYHANNSTEKVDTILERNVTVTHAKDIL
    :: :   .                 :* *  :  .   *.*  : .: **    :

QNSSIGEICDSPH------QILDGENCTLIDALLGDPQCDGFQN-KKWDLFVERS-KAYS
EDSHNGKLCLLKGI-----APLQLGNCSVAGWILGNPECELLISKESWSYIVETPNPENG
EDKHNGKLCKLRGV-----APLHLGKCNIAGWILGNPECESLSTASSWSYIVETPSSDNG
ERTHNGKLCDLNGV-----KPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKASPAND
EKTHNGKLCKLNGI-----PPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDG
     . :.                .*.  .  *. * *         . : *      .

NCYPYDVPDYASLRSLVATSGTLE---FNNE----SFNWTGVTQ-NGTSSACIRR-SSSS
TCYPGYFADYEELREQLSSVSSFERFEIFPK----ESSWPNHTV-TGVSASCSHN-GKSS
TCYPGDFIDYEELREQLSSVSSFERFEIFPK----TSSWPNHDSNKGVTAACPHA-GAKS
LCYPGNFNDYEELKHLLSRINHFEKIQIIPK-----SSWSNHDASSGVSSACPYL-GKSS
LCYPGSFNDYEELKHLLSSVKHFEKVKILPK-----DRWTQHTTT-GGSRACAVS-GNPS
 *:*    :  .::         *   :  .               *  :  :*    .  .

FFSRLNWLTHLNYTY-----PALNVTMPNNEQFDKLYIWGVHHPGTDKDQIFLYAQSSGR
FYRNLLWLTGKNGLY-----PNLSKSYVNNKEKEVLVLWGVHHPPNIGNQRALYHTENAY
FYKNLIWLVKKGNSY-----PKLSKSYINDKGKEVLVLWGIHHPSTSADQQSIYQNADTY
FFRNVVWLIKKNSTY-----PTIKRSYNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTY
FFRNMVWLTKKGSDY-----PVAKGSYNNTSGEQMLIIWGVHHPNDETEQRTLYQNVGTY
*:  : *                .         : : :**.*          ::

IT-VSTKRSQQAVIPNIGSRPRI-----RDIPSRISIYWTIVKPGDILLINSTGNLIAPR
VS-VVSSHYSRRFTPEIAKRPKV-----RDQEGRINYYWTLLEPGDTIIFEANGNLIAPW
VF-VGSSRYSKKFKPEIAIRPKV-----RDQEGRMNYYWTLVEPGDKITFEATGNLVVPR
IS-VGTSTLNQRLVPEIATRPKV-----NGQSGRIEFFWTILKPNDAINFESNGNFIAPE
VS-VGTSTLNKRSTPEIATRPKV-----NGQGGRMEFSWTLLDMWDTINFESTGNLIAPE
     :.        :*.  *           .*:   : : .  .   :    ..: *
```

```
GYFKI-RSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPF-QNVNRITYGACPRYVKHST
YAFALSRGFGSGIITSNAPMDECDAKCQTPQGAINSSLPF-QNVHPVTIGECPKYVRSAK
YAFAMERNAGSGIIISDTPVHDCNTTCQTPKGAINTSLPF-QNIHPITIGKCPKYVKSTK
YAYKIVKKGDSTIMKSELEYGNCNTKCQTPMGAINSSMPF-HNIHPLTIGECPKYVKSNR
YGFKISKRGSSGIMKTEGTLENCETKCQTPLGAINTTLPF-HNVHPLTIGECPKYVKSEK
  :       * ::  .              *.: .  *:  . :  : * ** :*:

LKLATGMRNVPEKQ----TRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLKS
LRMVTGLRNIPSIQ----SRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSGYAADQKS
LRLATGLRNIPSIQ----SRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSGYAADLKS
LVLATGLRNAPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADQES
LVLATGLRNVPQIE----SRGLFGAIAGFIEGGWQGMVDGWYGYHHSNDQGSGYAADKES
* :..* :  *  :     **:*******:*.** **: **:*:   . .* . *** :*

TQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLV
TQNAINGITNKVNSVIEKMNTQFTAVGKEFNKLERRMENLNKKVDDGFLDIWTYNAELLV
TQNAIDEITNKVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWTYNAELLV
TQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLDVWTYNAELLV
TQKAFDGITNKVNSVIEKMNTQFEAVGKEFSNLERRLENLNKKMEDGFLDVWTYNAELLV
** *:: :. ::* :      ::  :   :..:. .: :*:: ::*   *  : : ** *

ALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDHN
LLENERTLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNNECMESVKNGTYDYP
LLENERTLDYHDSNVKNLYEKVRSQLKNNAKEIGNGCFEFYHKCDNTCMESVKNGTYDYP
LMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECMESVKNGTYDYP
LMENERTLDFHDSNVKNLYDKVRMQLRDNVKELGNGCFEFYHKCDDECMNSVKNGTYDYP
 :.*:  ::  *..:  *  * :  *  .. ::*****:  ***:: *:  :   *::

VYRDEALNNRFQIKGVELK--SGYKDWILWI-SFAISCFLLCVALLGFIMWACQKGNIRC
KYSEESKLNREKIDGVKLE--SMGVYQILAIYSTVASSLVLLVSLGAISFWMCSNGSLQC
KYSEEAKLNREEIDGVKLE--STRIYQILAIYSTVASSLVLVVSLGAISFWMCSNGSLQC
QYSEEARLNREEISGVKLE--SMGTYQILSLYSTVASSLALAIMVAGLSLWMCSNGSLQC
KYEEESKLNRNEIKGVKLS--SMGVYQILAIYATVAGSLSLAIMMAGISFWMCSNGSLQC
 :   :     :* ...*.  .     **   : . ..: : : :  : .:  ....: *
```

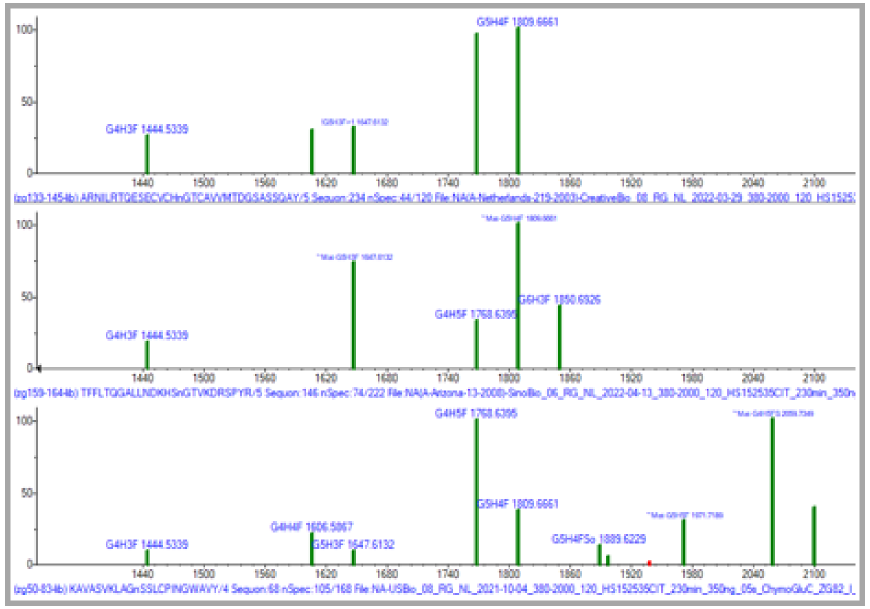
A

		*	*		*			*	*					*			*					*	*	*
HA A/NewCaledonia/20/1999		27	28		40			70	104			142			176							303	497	556
HA A/HongKong/485197/2014	24			38		54	61	79		138	142		149	174		181			262		301		499	
HA A/California/04/2009		27	28		40				104											293		304	498	557
HA A/HongKong/483/1997		26	27		39									170			181					302	500	559
HA A/Japan/305/1957		25	26		38												179	180				300	494	553

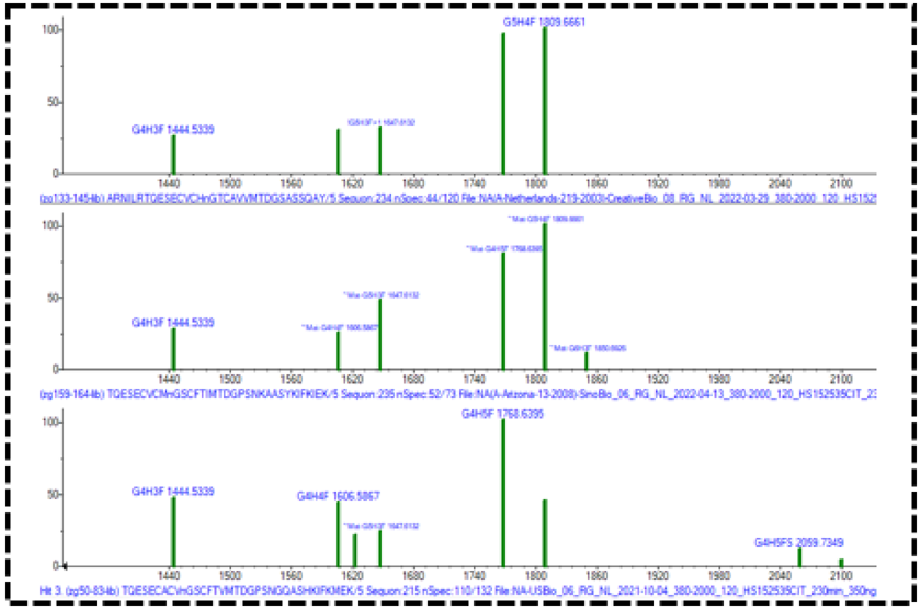
B

					*					*	*		*			
NA A/Netherlands/219/2003	32		47	56	57		67	68		87	145	200	234	401		
NA A/Arizona/13/2008		44			58	63			70	88	146		235		434	455
NA A/Thailand/1(KAN-1)/2004										68	126		215			

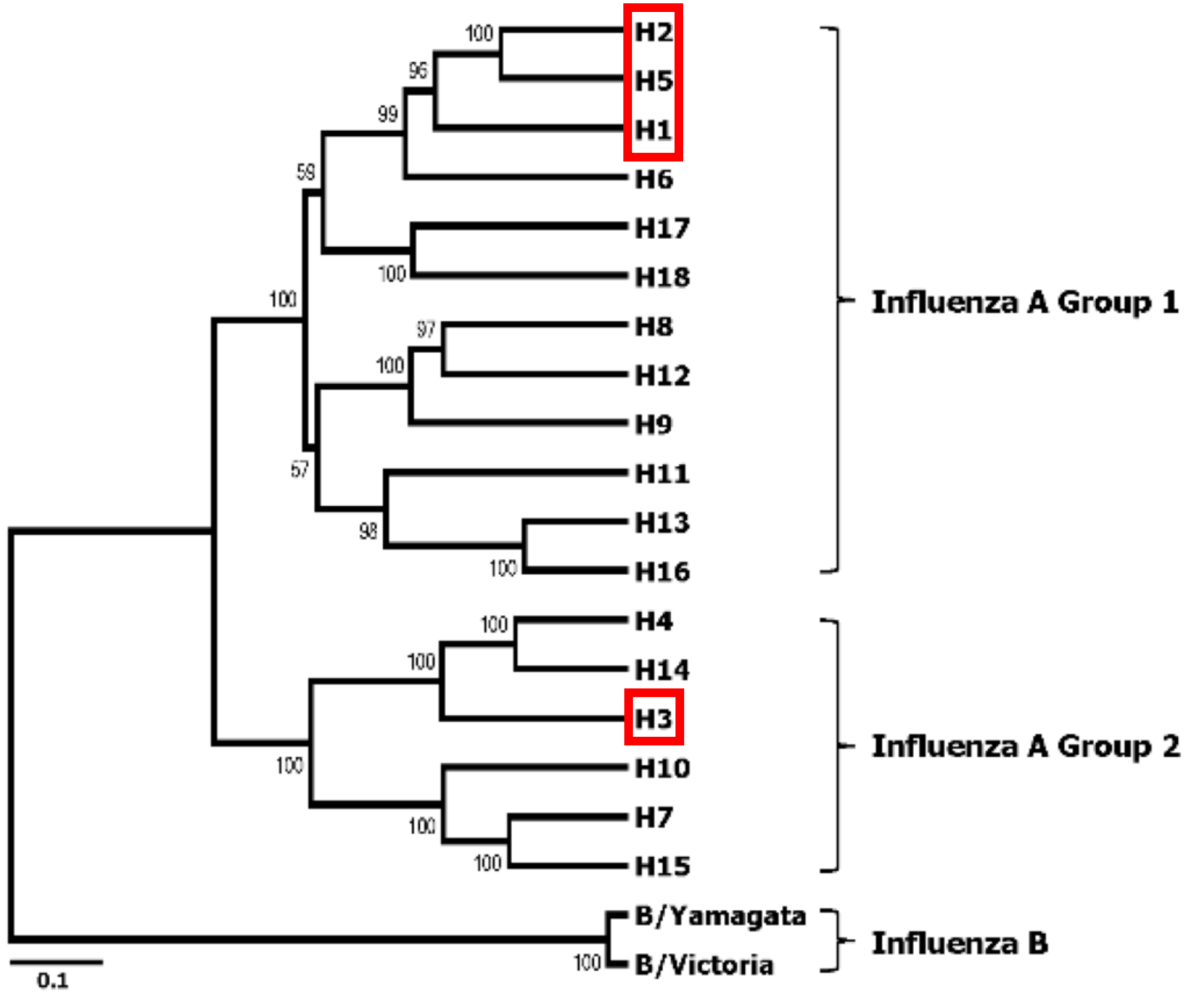
C



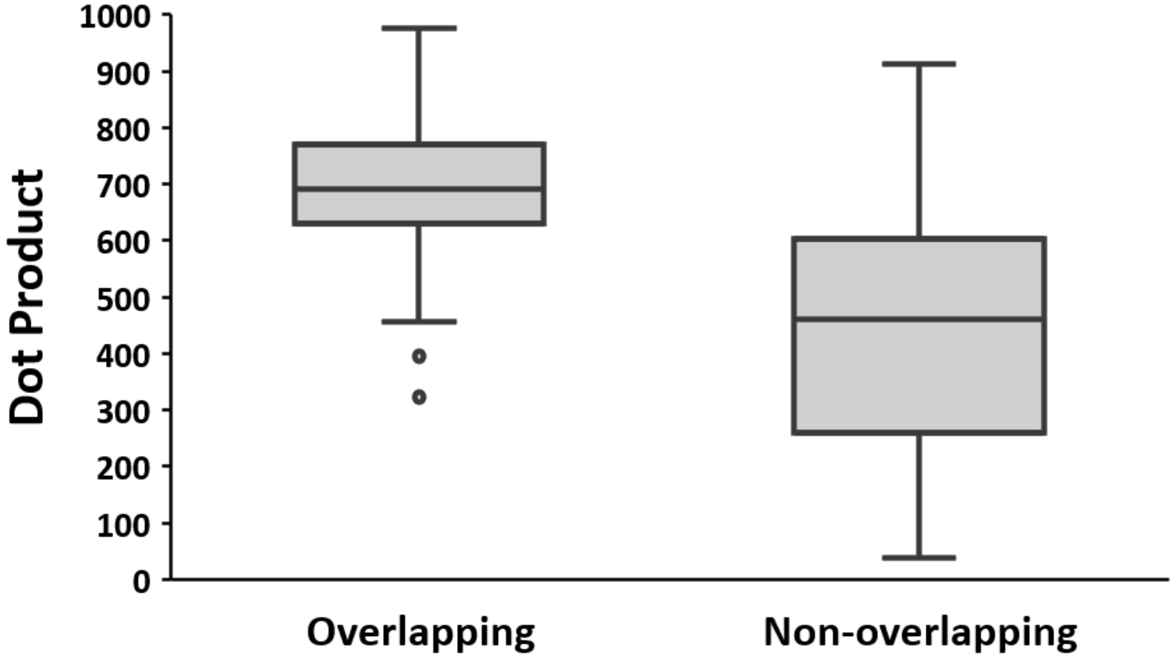
D



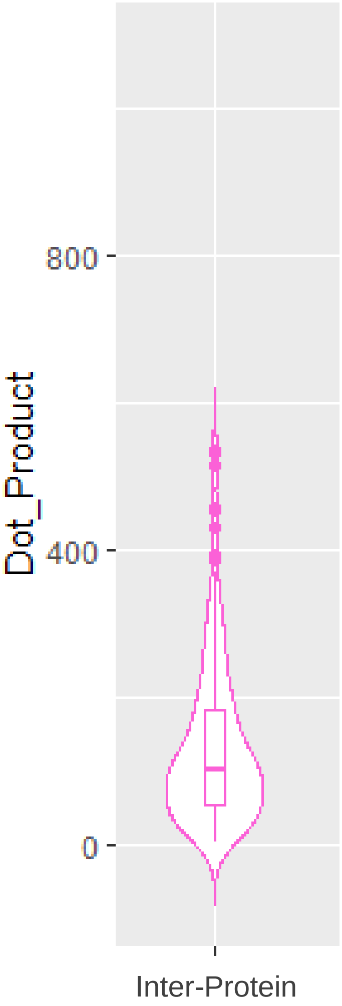
Conserved Regions Have Conserved Glycosylation Distribution

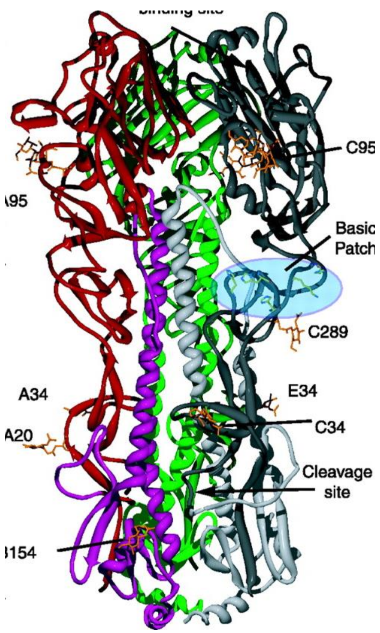


Jang et al., 2014

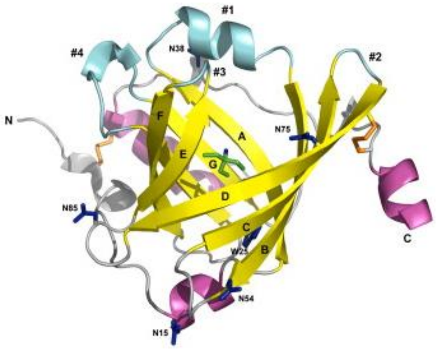


Variation from different proteins



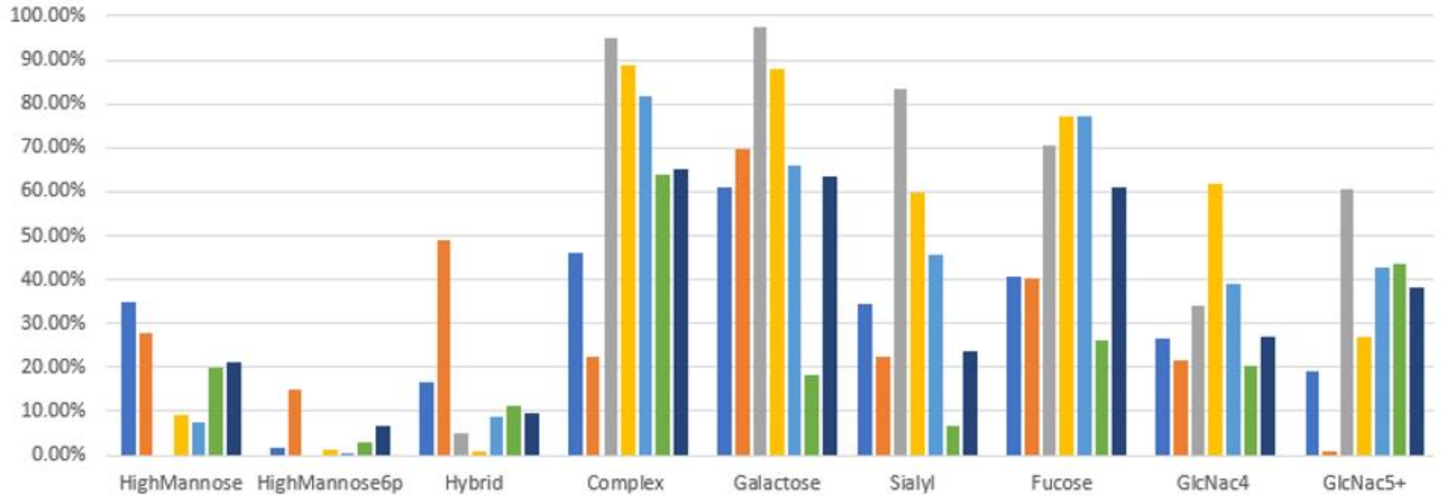


Hemagglutinin

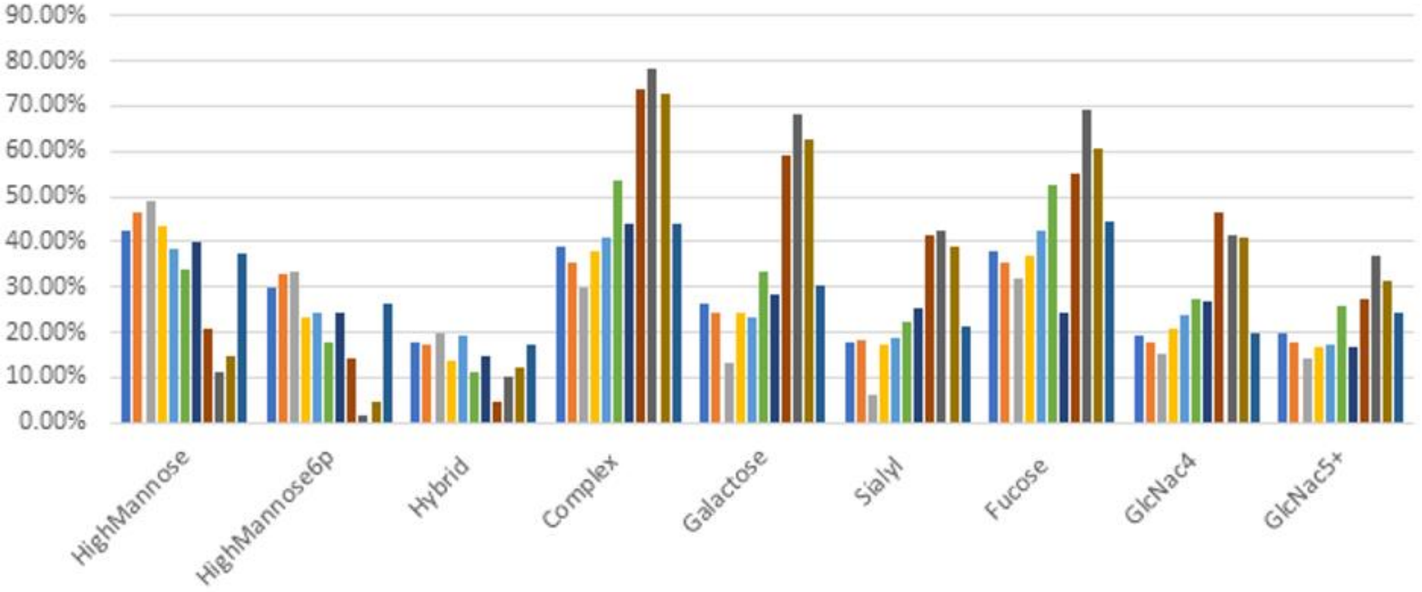


A1AG

Inter-Protein Comparison



Influenza Hemagglutinin



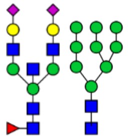
Sars-CoV-2 Spike protein

Review - 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

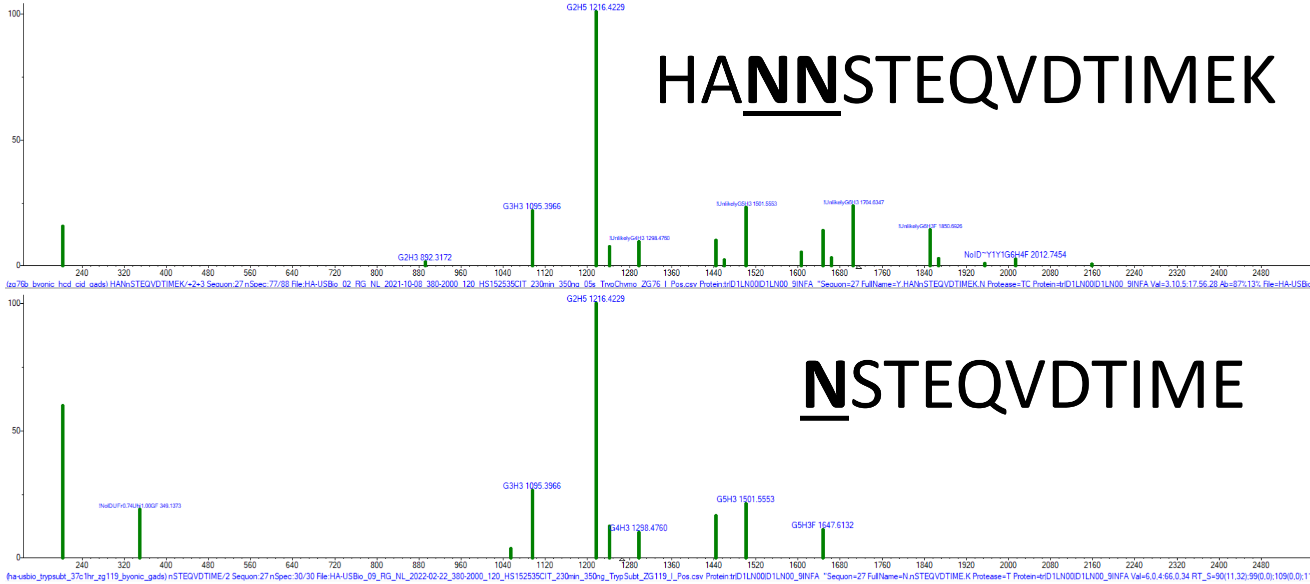
NNSS, NNTT, NNST, NNTS



…PIDETEQGSYNNTSGEQMLIIWGVHHR…

Isolation of Adjacent Sequons - Subtilisin

- *Bacillus subtilis* – extracellular serine endopeptidases



HA**NN**STEQVDTIMEK

NSTEQVDTIME

DQICIGYHA**N** ???

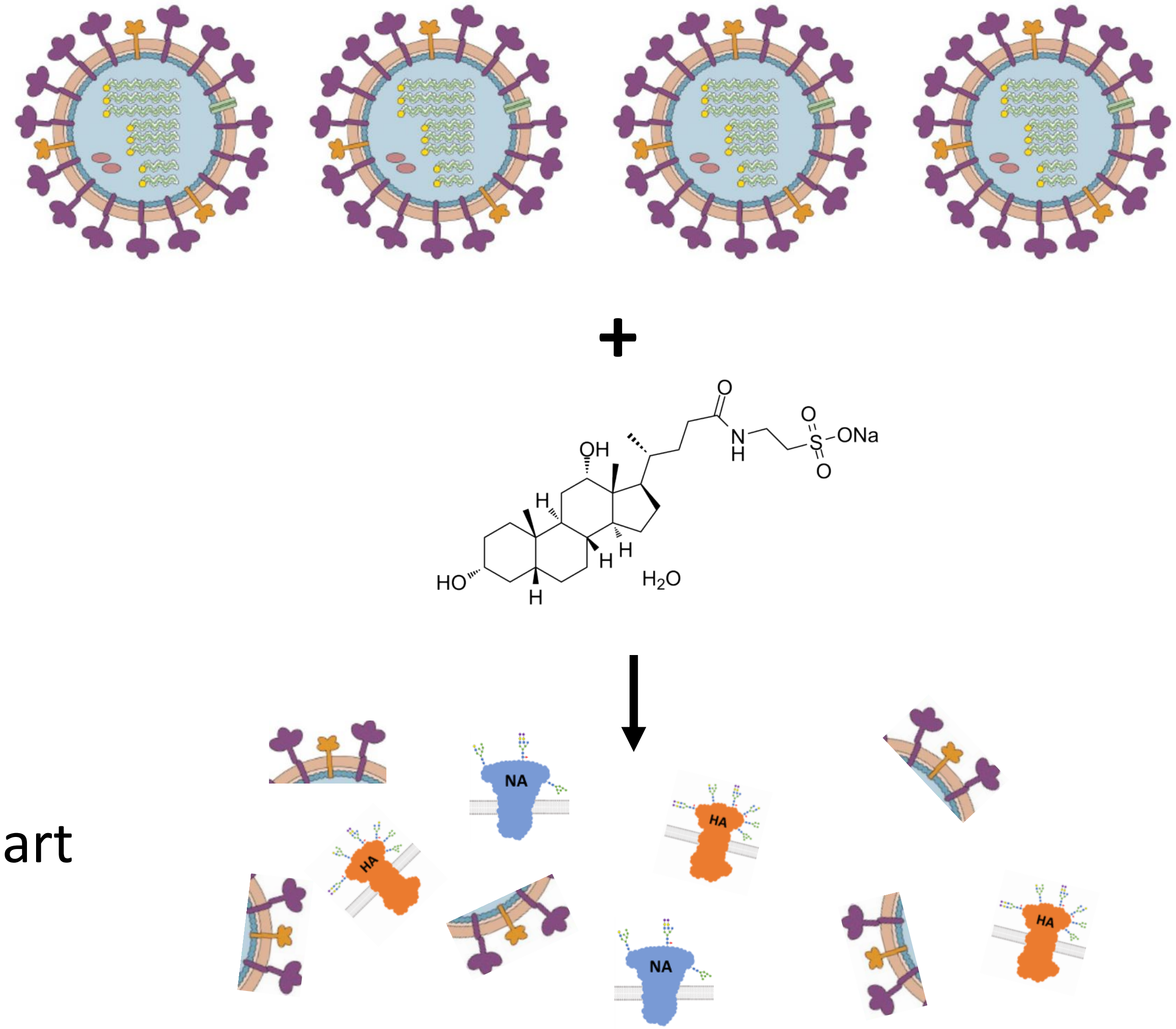
Isolation of Adjacent Sequons - Subtilisin

- *Bacillus subtilis* – extracellular serine endopeptidases

- Three more proteases from *Bacillus subtilis* were obtained
  - Esperase
  - Savinase
  - Neutrase

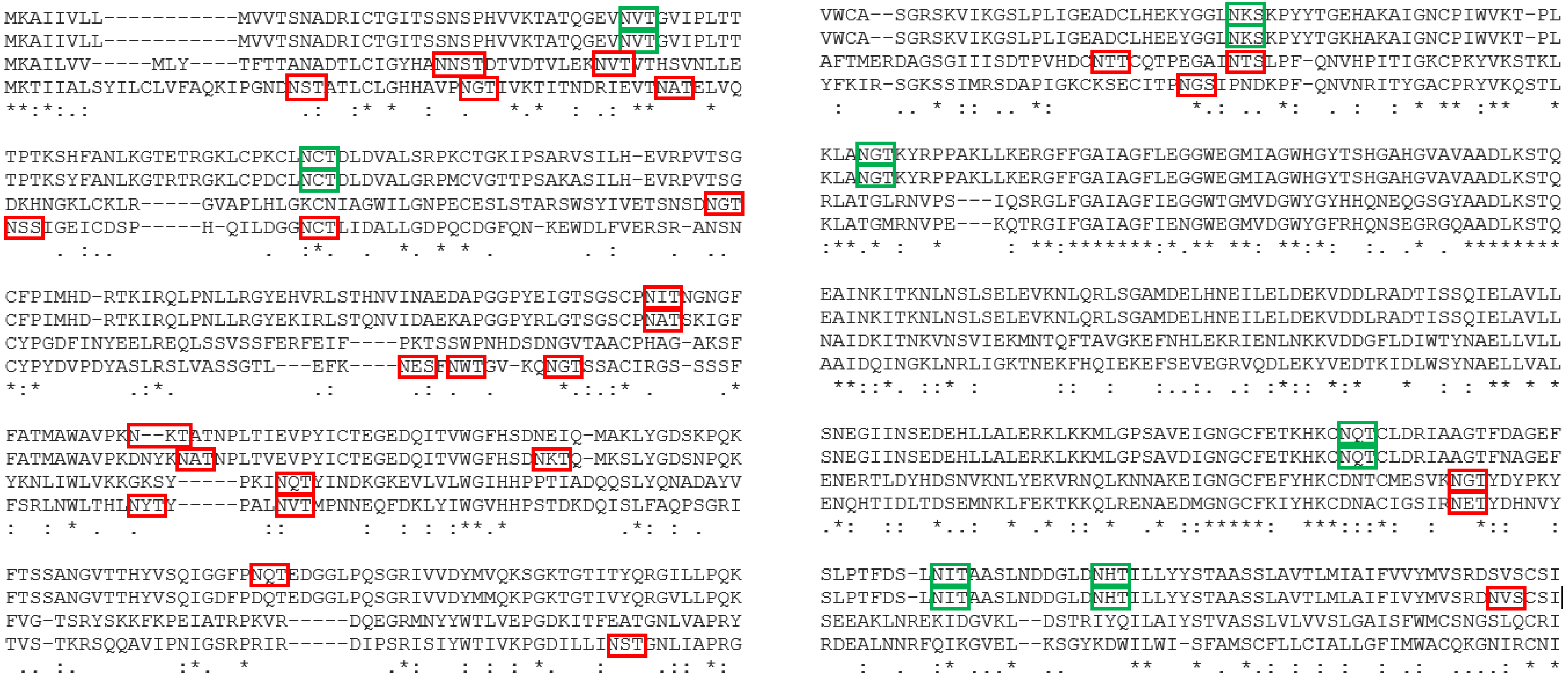
Influenza Vaccines

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



- Afluria Quadrivalent
- 3 Monovalent vaccines from NIBSC
- 3 Monovalent vaccines from Creative Biomart

Quadrivalent Vaccine Challenges

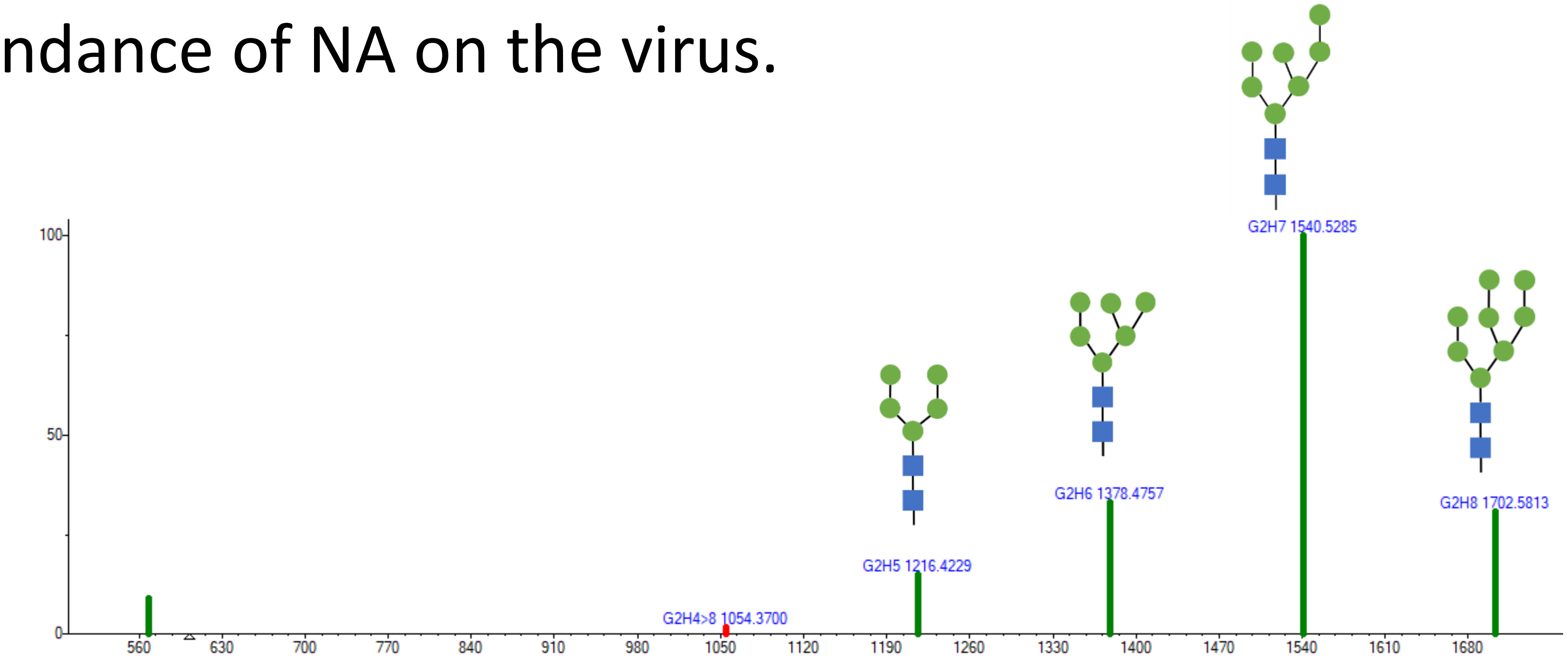


```
MKAIIVLL-----------MVVTSNADRICTGITSSNSPHVVKTATQGEVNVTGVIPLTT
MKAIIVLL-----------MVVTSNADRICTGITSSNSPHVVKTATQGEVNVTGVIPLTT
MKAILVV-----MLY----TFTTANADTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLE
MKTIIALSYILCLVFAQKIPGNDNSTATLCLGHHAVPNGTIVKTITNDRIEVTNATELVQ
**:*:.:                 .:  :* *  :  .   *.*  : .: **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

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
   :  .  :* ...*  ..     **   * . *.: : : :  : .:  ....: * *
```

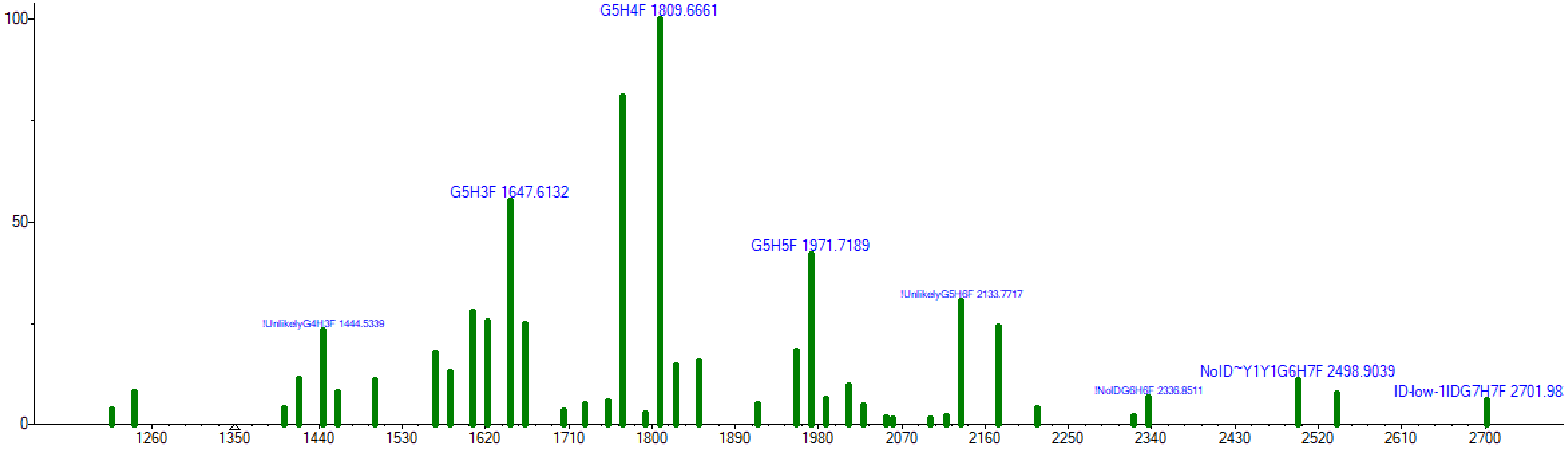
Afluria GADS

- 8→23 of the 40 sequons for the four hemagglutinin proteins identified. 27 are non-overlapping
- Three major GADS classes: high mannose, G4F complex, G5F complex
- No GADS were identified for NA proteins, probably due to a lower abundance of NA on the virus.



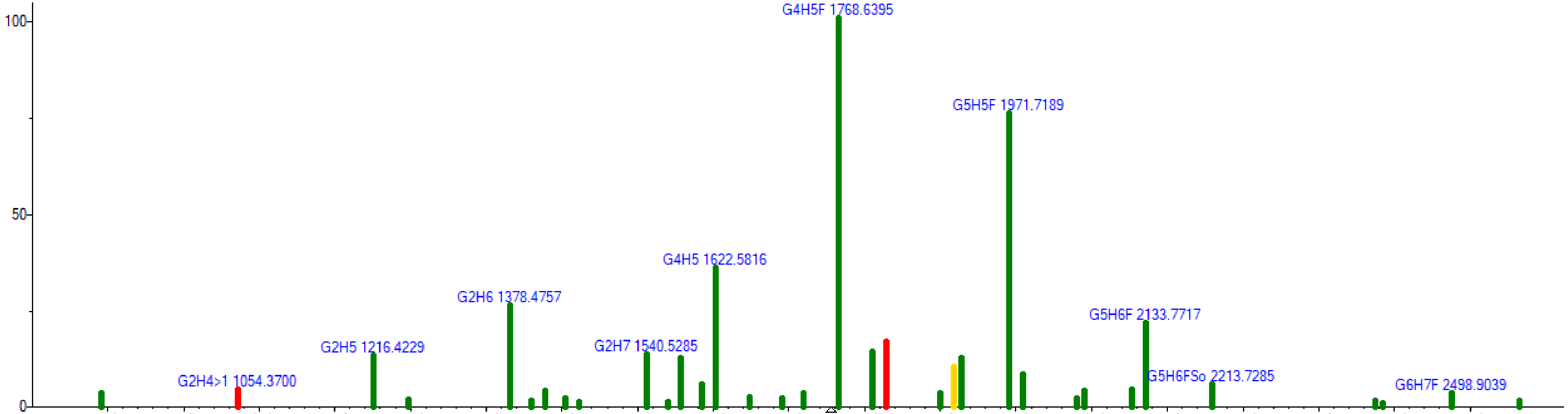
Afluria GADS

- 8→23 of the 40 sequons for the four hemagglutinin proteins identified. 27 are non-overlapping
- Three major GADS classes: high mannose, G4F complex, G5F complex
- No GADS were identified for NA proteins, probably due to a lower abundance of NA on the virus.



Afluria GADS

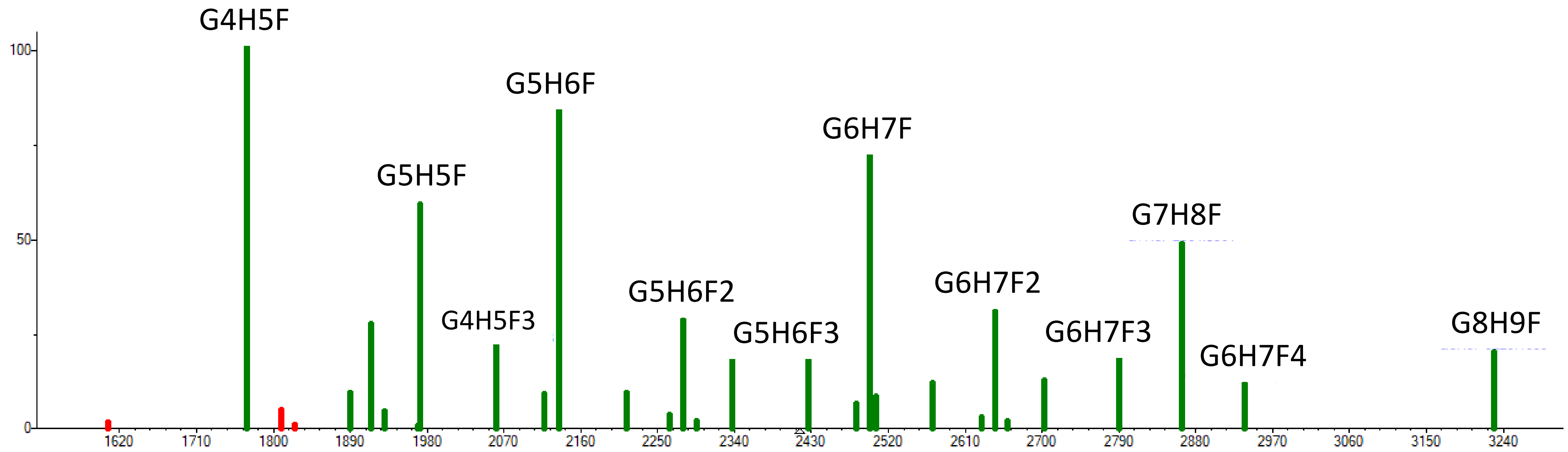
- 8→23 of the 40 sequons for the four hemagglutinin proteins identified. 27 are non-overlapping
- Three major GADS classes: high mannose, G4F complex, G5F complex
- No GADS were identified for NA proteins, probably due to a lower abundance of NA on the virus.



(afluria_msfragger_tryplysc_zg188_rep2_gads) CLnCTDLDVALSRPK/3-RTLow Sequon:74 nSpec:93/99 File:Afluria-vaccine_03_Dial-RG_NL_2022-07-25_380-2000_120_HS152535CIT_230min_2ug_TrypLysC_ZG188-rerun_I_Pos.msf.tsv Protein:EPI1629575_he

Monovalent vaccine GADS

- 90% sequons detected
- Three major GADS classes: high mannose, G4/G5, G4/G5 + high mannose
- Again, no major GADS identified for NA.



(zg271-276_filtered_gads) CLGHHAVPnGTLVKTITNDQIE/4 Sequon:38 nSpec:78/122 File:NIBSC-Vaccine(A-Philippines-2-1982)_13_Cyr_NL_2022-12-08_380-2000_120_HS152535CIT_230min_3ug_ChymoGluC_ZG274_I_Pos.msf.tsv Protein:ADJ41805.1_hemagglut

Open Search for Vaccine Modifications

Mass shift (Da)	Modification
15.99	Oxidation
28.03	Ethylation
42.00	Acetylation
72.02	Carboxyethyl
113.08	Acetyl Deoxyhypusine
145.03	?
200.11	?
226.16	?

← Formalin

← Formalin

← B-Propiolactone

Next Steps

- Three more commercial vaccines were purchased for the 2022-2023 flu season
  - Flublok – Recombinant vaccine
  - Afluria – Egg based vaccine
  - Flucelvax – Cell culture based vaccine
  - How does glycosylation change… year to year

    … strain to strain

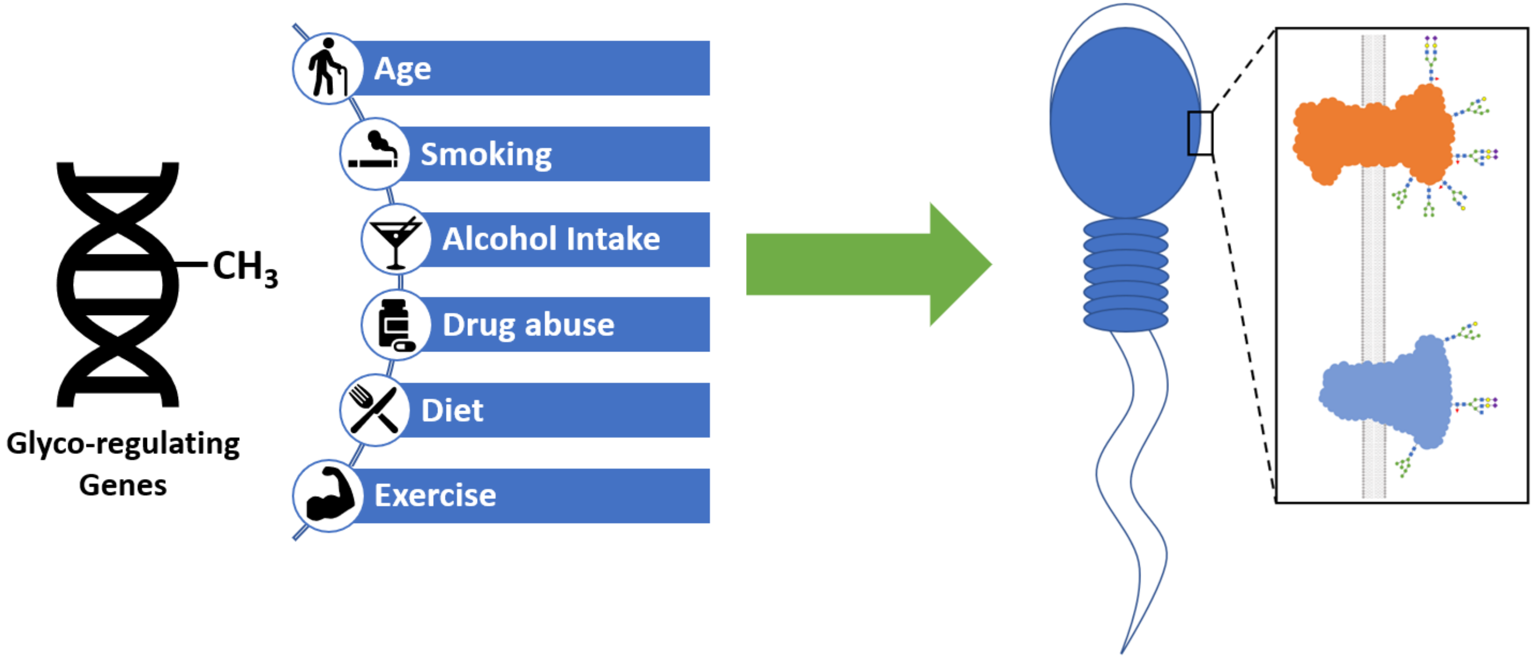
    … between sources

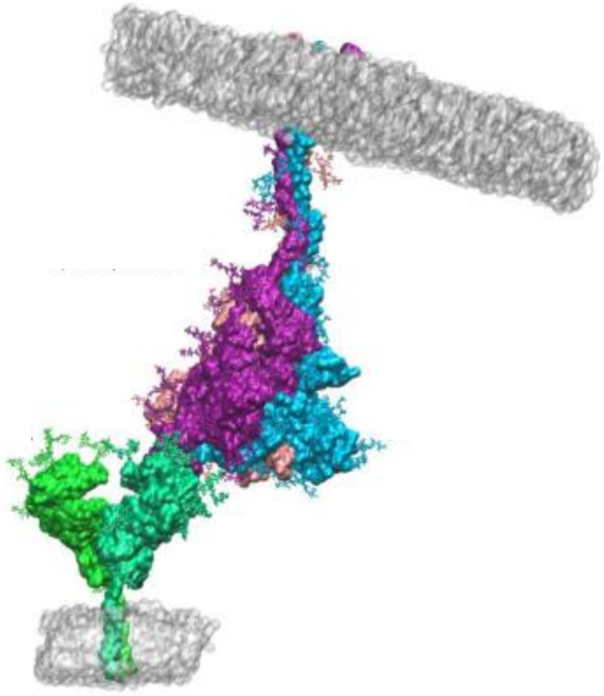
Next Steps

- Characterize forensic evidence using glycosylation.
- Classification of sperm and egg glycosylation for determining likelihood of fertilization.

Epigenetic Change in DNA

Change in Semen Glycoforms





Conclusions

- Glycosylation profiles between replicates is most similar and between different proteins is least similar.
- Homologous sequence regions between different influenza strains have similar glycosylation distribution compared to non-conserved regions.
- Most glycans are high-mannose or mono-fucoylated complex in egg-based quadrivalent vaccines.
- Next steps:
  - Assess more subtilisin proteases ability to isolate adjacent sequons
  - Digest 2022-2023 vaccines and compare glycan distributions
  - Build reference MS2 and GADS libraries

Acknowledgments

Post-doc Adviser

- Stephen Stein

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

Administrative support

- Heather Blache
- Sarah Huber
- Bill Wallace

References

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

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.

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

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.

Jang, Y. H., & Seong, B. L. (2014). Options and obstacles for designing a universal influenza vaccine. Viruses, 6(8), 3159-3180.

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

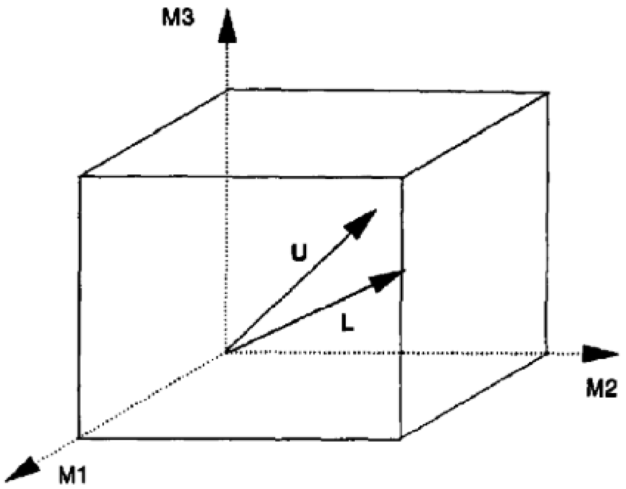


Figure 1. Vector representation of a hypothetical three-peak unknown (U) and library (L) mass spectrum in three-dimensional space (peaks have mass M1, M2, and M3).

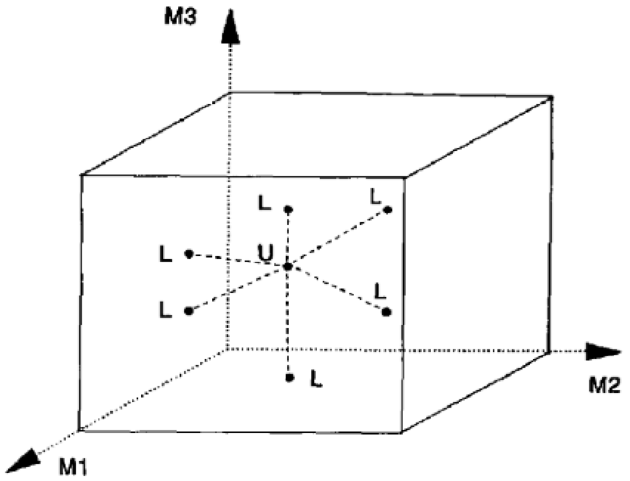
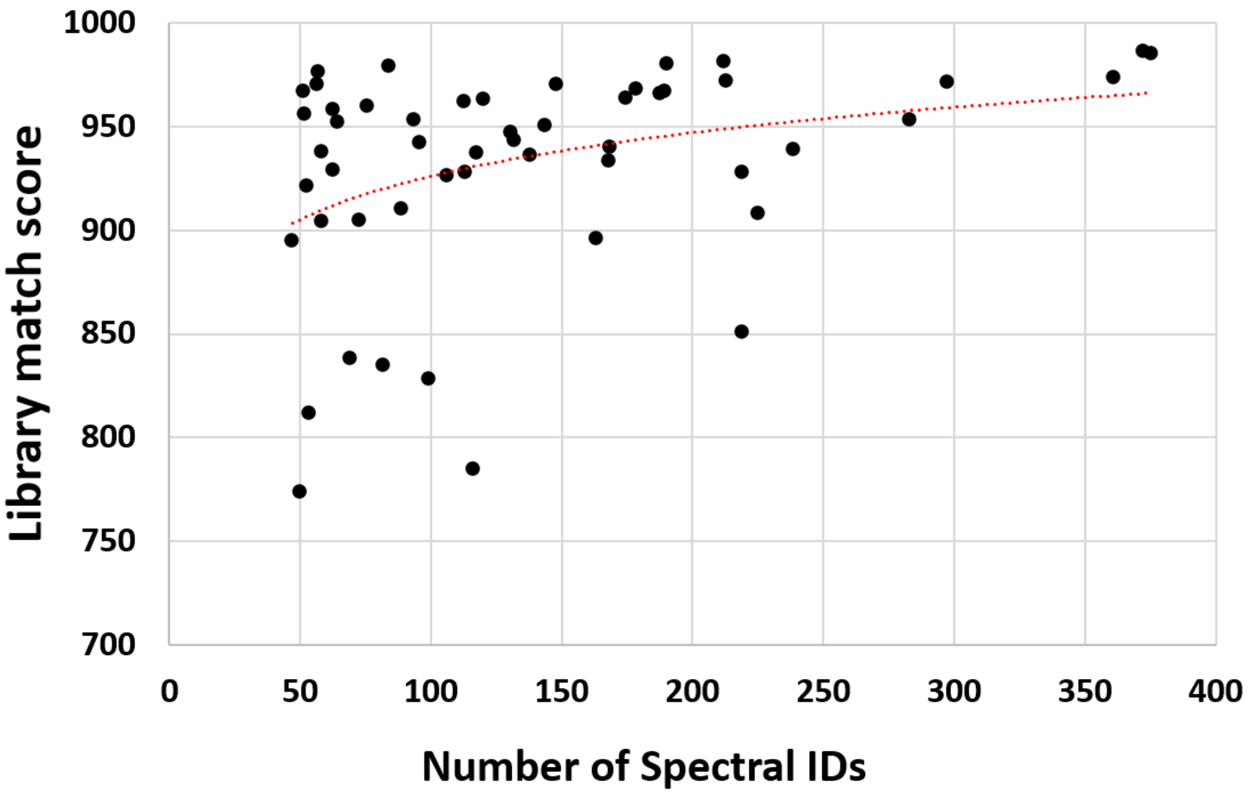


Figure 2. Point representation of library search results (L) for a hypothetical three-peak unknown (U) spectrum (masses M1, M2, and M3)

Supplemental Slides

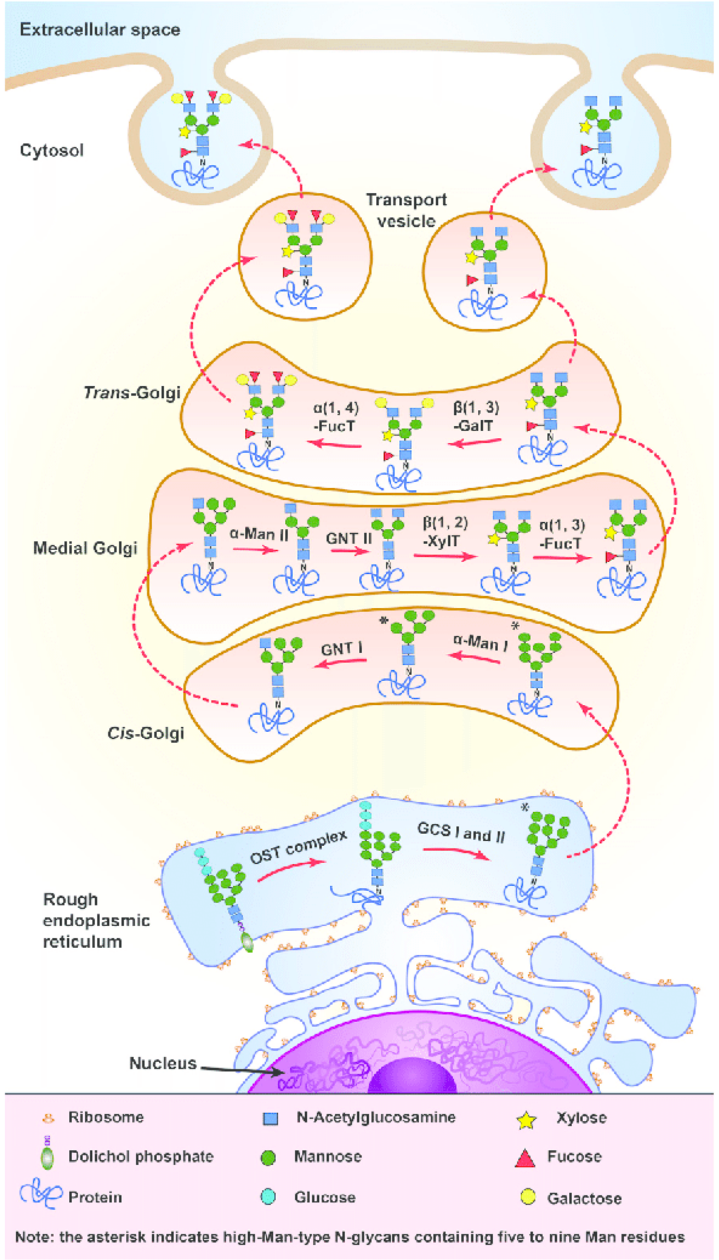


Supplemental Slides

Glycobiology

- In viruses
  - Immune evasion
  - Host cell attachment

- In mammalian cells
  - Protein folding
  - Protein stabilization
  - Communication
  - Function

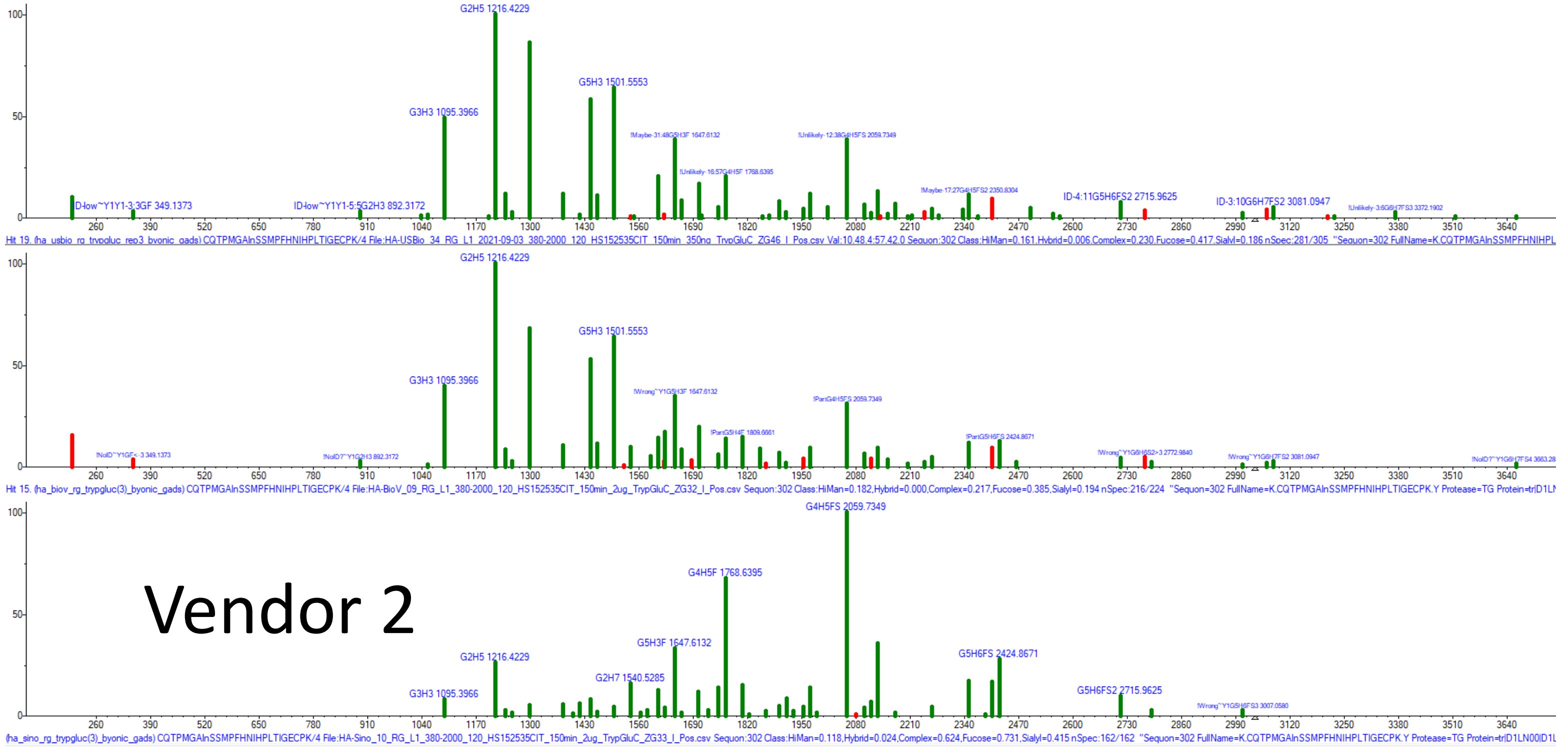


Note: the asterisk indicates high-Man-type N-glycans containing five to nine Man residues

Supplemental Slides

Vendor Comparison

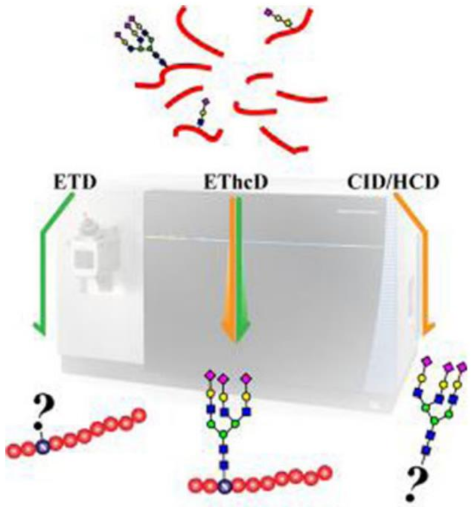
Glycan distribution was skewed toward higher mass (sialylated complex) glycans for vendor 2 and toward lower mass glycans (oligomannose and hybrid) for vendor 1 and vendor 3.

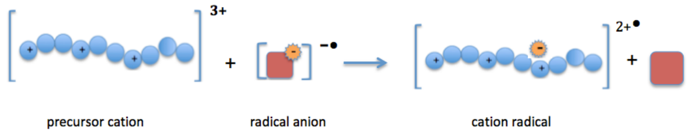


Vendor 2

Supplemental Slides

EThcD





Fluoranthene

Stepped HCD

Peptide bond
Glycosidic bond
N-glycosidic bond