



Deborah Chang, Joshua A. Klein, Joseph Zaia
Center for Biomedical Mass Spectrometry

Download and install GlycReSoFr from github:

https://github.com/BostonUniversityCBMS/glycresoft_sqlalchemy

Agilent 6550 LC-tandem MS data on alpha-1-acid glycoprotein and influenza A virus are available for public download through the Pride archive:

<https://www.ebi.ac.uk/pride/archive/projects/PXD003498>

Data reference:

Khatri, K., Klein, J. A., White, M. R., Grant, O. C., Leymarie, N., Woods, R. J., Hartshorn, K. L., and Zaia, J. (2016) Integrated Omics and Computational Glycobiology Reveal Structural Basis for Influenza A Virus Glycan Microheterogeneity and Host Interactions, *Mol Cell Proteomics* 15, 1895-1912. PMC5083086.

Overview

1. Preprocess
raw data

2. Construct glycan
hypothesis

3. Construct glycopeptide
hypothesis

4. Search glycopeptide
sequences

Build Search Space

BUILD A GLYCAN SEARCH SPACE
BUILD A GLYCOPEPTIDE SEARCH
SPACE

Analyze Sample



SEARCH GLYCAN
COMPOSITIONS
SEARCH GLYCOPEPTIDE
SEQUENCES

View Results

VIEW DATABASE SEARCH
RESULTS

Sample	tryptic digest of A/Philippines/2/82 (H3N2)
Mass spectrometry	Agilent 6550 LC/MS with chip-cube ionization online HILIC enrichment, C18 separation
Database	combined Uniprot and IAV protein database PEAKS Studio 8.0
PTMs	Carbamidomethylation of Cys (fixed) Deamidation of Asn (variable) Oxidation of Met (variable)

Khatri, K. et al. *Molecular & Cellular Proteomics* (2016)

STEP 1: Preprocessing

- Profile to centroid
- Identify isotopic patterns and determine monoisotopic peak

Click the flask to add a new mzML file

Build Search Space

BUILD A GLYCAN SEARCH SPACE
BUILD A GLYCOPEPTIDE SEARCH SPACE

Analyze Sample



SEARCH GLYCAN COMPOSITIONS
SEARCH GLYCOPEPTIDE SEQUENCES

View Results

VIEW DATABASE SEARCH RESULTS

Add Sample To Workspace

Add mzML file here

SELECT MZML FILE

Please provide a file in mzML format

Sample Name

Provide a name to identify this sample

Preset Configurations

LC-MS/MS Glycoproteomics

Maximum Charge State

12

Start Processing Time

0.0

End Processing Time

256

Missed Peaks Permitted

1

✓ Fit MS/MS Features Only

MS¹ Parameters

Minimum Isotopic Score

35

Averagine

Glycopeptide

Background Reduction

5

Custom Formula

MSⁿ Parameters

Minimum Isotopic Score

10

Averagine

Peptide

Custom Formula

SUBMIT

STEP 2: Glycan hypothesis construction

- Combinatorial hypothesis, or
- Predetermined dataset from a text file

Build Search Space

BUILD A GLYCAN SEARCH SPACE
BUILD A GLYCOPEPTIDE SEARCH SPACE

Analyze Sample



SEARCH GLYCAN COMPOSITIONS
SEARCH GLYCOPEPTIDE SEQUENCES

View Results

VIEW DATABASE SEARCH RESULTS

Glycan Search Space Building



Hypothesis Name

Reduction Type

Derivatization Type

Native

Native

Custom Reduction (Formula)

Custom Substituent (Name or I

[PREGENERATED DATABASE](#)

[COMBINATORIAL HYPOTHE...](#)

[HYPOTHESIS FROM A TEXT ...](#)

[MERGE TWO HYPOTHESES](#)

Create a Combinatorial Search Space:

Note: This process can only produce *N*-glycans at this time

Monosaccharides

Residue Name	Lower Bound	Upper Bound
Hex	3	9
HexNAc	2	8
Fuc	0	4
NeuAc	0	4
Name	Bound	Bound

Constraints

Limit

Constrained Value

HexNAc

>

Fuc

Name

=

Name/Value

GENERATE

N-glycans

ID: 738fc91593b34955844e94861566656e:1

Search Mass

Error Tolerance (PPM)

10



	Composition	Theoretical Mass	
1	Hex 3 HexNAc 2	910.327779978	N-Glycan
2	Fuc 1 Hex 3 HexNAc 2	1056.38568878	N-Glycan
3	Hex 3 HexNAc 3	1113.4071525	N-Glycan
4	Fuc 1 Hex 3 HexNAc 3	1259.4650613	N-Glycan
5	Fuc 2 Hex 3 HexNAc 3	1405.5229701	N-Glycan
6	Hex 3 HexNAc 4	1316.48652502	N-Glycan
7	Fuc 1 Hex 3 HexNAc 4	1462.54443382	N-Glycan
8	Fuc 2 Hex 3 HexNAc 4	1608.60234262	N-Glycan
9	Fuc 3 Hex 3 HexNAc 4	1754.66025141	N-Glycan
10	Hex 3 HexNAc 5	1519.56589754	N-Glycan
11	Fuc 1 Hex 3 HexNAc 5	1665.62380634	N-Glycan
12	Fuc 2 Hex 3 HexNAc 5	1811.68171513	N-Glycan
13	Fuc 3 Hex 3 HexNAc 5	1957.73962393	N-Glycan
14	Fuc 4 Hex 3 HexNAc 5	2103.79753273	N-Glycan
15	Hex 3 HexNAc 6	1722.64527006	N-Glycan
16	Fuc 1 Hex 3 HexNAc 6	1868.70317886	N-Glycan
17	Fuc 2 Hex 3 HexNAc 6	2014.76108765	N-Glycan
18	Fuc 3 Hex 3 HexNAc 6	2160.81899645	N-Glycan
19	Fuc 4 Hex 3 HexNAc 6	2306.87690525	N-Glycan
20	Hex 3 HexNAc 7	1925.72464258	N-Glycan
21	Fuc 1 Hex 3 HexNAc 7	2071.78255137	N-Glycan
22	Fuc 2 Hex 3 HexNAc 7	2217.84046017	N-Glycan
23	Fuc 3 Hex 3 HexNAc 7	2363.89836897	N-Glycan
24	Fuc 4 Hex 3 HexNAc 7	2509.95627777	N-Glycan
25	Hex 3 HexNAc 8	2128.8040151	N-Glycan
26	Fuc 1 Hex 3 HexNAc 8	2274.86192389	N-Glycan
27	Fuc 2 Hex 3 HexNAc 8	2420.91983269	N-Glycan
28	Fuc 3 Hex 3 HexNAc 8	2566.97774149	N-Glycan
29	Fuc 4 Hex 3 HexNAc 8	2712.03565029	N-Glycan

6/5/2017

Glycan Search Space Building

Hypothesis Name

PREGENERATED DATABASE COMBINATORIAL HYPOTHE..

Read Glycan Composition From Text File:

SELECT FILE...

GENERATE

```
{Hex:8; HexNAc:2}      N-Glycan
{Fuc:1; Hex:8; HexNAc:7}      N-Glycan
{Hex:9; HexNAc:2}      N-Glycan
{Hex:5; HexNAc:2}      N-Glycan
{Hex:5; HexNAc:3}      N-Glycan
{Hex:5; HexNAc:4}      N-Glycan
{Fuc:1; Hex:5; HexNAc:4}      N-Glycan
{Fuc:2; Hex:5; HexNAc:4}      N-Glycan
{Fuc:3; Hex:5; HexNAc:4}      N-Glycan
{Hex:5; HexNAc:5}      N-Glycan
{Fuc:1; Hex:5; HexNAc:5}      N-Glycan
{Fuc:2; Hex:5; HexNAc:5}      N-Glycan
{Fuc:1; Hex:8; HexNAc:4; Neu5Ac:1}      N-Glycan
{Hex:6; HexNAc:2}      N-Glycan
{Hex:6; HexNAc:3}      N-Glycan
{Fuc:1; Hex:6; HexNAc:3}      N-Glycan
{Hex:6; HexNAc:4}      N-Glycan
{Hex:6; HexNAc:5}      N-Glycan
{Fuc:1; Hex:6; HexNAc:5}      N-Glycan
{Fuc:2; Hex:6; HexNAc:5}      N-Glycan
{Hex:7; HexNAc:2}      N-Glycan
{Hex:7; HexNAc:6; Neu5Ac:4}      N-Glycan
```

Text file format:

{monosaccharide:count[; monosaccharide:count]} N-Linked/O-Linked/GAG-Linker

Phil82_glycomics_NG_6520

ID: d0bc4aa733e840fa84a42aab64421d4f:1

Search Mass

Error Tolerance (PPM)

10

	Composition	Theoretical Mass	
1	Hex 8 HexNAc 2	1720.59189707	N-Glycan
2	Fuc 1 Hex 8 HexNAc 7	2882.04666847	N-Glycan
3	Hex 9 HexNAc 2	1882.64472049	N-Glycan
4	Hex 5 HexNAc 2	1234.43342682	N-Glycan
5	Hex 5 HexNAc 3	1437.51279933	N-Glycan
6	Hex 5 HexNAc 4	1640.59217185	N-Glycan
7	Fuc 1 Hex 5 HexNAc 4	1786.65008065	N-Glycan
8	Fuc 2 Hex 5 HexNAc 4	1932.70798945	N-Glycan
9	Fuc 3 Hex 5 HexNAc 4	2078.76589825	N-Glycan
10	Hex 5 HexNAc 5	1843.67154437	N-Glycan
11	Fuc 1 Hex 5 HexNAc 5	1989.72945317	N-Glycan
12	Fuc 2 Hex 5 HexNAc 5	2135.78736197	N-Glycan
13	Fuc 1 Hex 8 HexNAc 4 Neu5Ac 1	2563.90396742	N-Glycan
14	Hex 6 HexNAc 2	1396.48625023	N-Glycan
15	Hex 6 HexNAc 3	1599.56562275	N-Glycan
16	Fuc 1 Hex 6 HexNAc 3	1745.62353155	N-Glycan
17	Hex 6 HexNAc 4	1802.64499527	N-Glycan
18	Hex 6 HexNAc 5	2005.72436779	N-Glycan
19	Fuc 1 Hex 6 HexNAc 5	2151.78227659	N-Glycan
20	Fuc 2 Hex 6 HexNAc 5	2297.84018539	N-Glycan
21	Hex 7 HexNAc 2	1558.53907365	N-Glycan
22	Hex 7 HexNAc 6 Neu5Ac 4	3535.23822976	N-Glycan

6/5/2017

11

STEP 3: Glycopeptide database construction

Enumerate every possible combination of peptide, glycan composition set, and glycosylation site arrangement for every peptidoform, using:

- combinatorial hypothesis, or
- data-driven hypothesis from mzIdentML

Build Search Space

BUILD A GLYCAN SEARCH SPACE
BUILD A GLYCOPEPTIDE SEARCH SPACE

Analyze Sample



SEARCH GLYCAN COMPOSITIONS
SEARCH GLYCOPEPTIDE SEQUENCES

View Results

VIEW DATABASE SEARCH RESULTS

Glycopeptide Search Space Building

Hypothesis Name

Protein List

SELECT FILE

A list of protein sequences to digest and glycosylate *in-silico*, provided in Fasta format or mzIdentML.

Select Modifications

Name	Target	Formula	Mass
2-dimethylsuccinyl	C	C6H8O4	144.042259
2-monomethylsuccinyl	C	C5H6O4	130.026609
2-nitrobenzyl	Y	C7H5N1O2	135.032028
2-succinyl	C	C4H4O4	116.010959
2HPG	R	C16H10O5	282.052824
3-deoxyglucosone	R	C6H8O4	144.042259
3-phosphoglyceryl	K	C3H5O6P1	167.982375
3sulfo	N	C7H4O4S1	183.983029
4-ONE	C	C9H14O2	154.09938

Search by name

+ CONSTANT

+ VARIABLE

- SELECTION

CREATE CUSTOM

Constant

Carbamidomethyl C2H3N1O157.0214

Variable

Add fasta file or
mzIdentML here

Glycopeptide Search Space Building

Enzymatic Digest

caspace 3
caspace 2
caspace 1
factor xa
caspace 7
caspace 6
caspace 5
caspace 4
glutamyl endopeptidase
trypsin

Missed Cleavages Allowed

2

Glycan Definitions

Text File of Glycan Structures or Compositions

SELECT FILE...

Maximum Number of Glycosylations
per Peptide

1

GENERATE

List of glycan structures/compositions to attach to
each protein.

Or

Select a Glycan Hypothesis or Sample
Analysis

Select glycan hypothesis created
in the previous step here

STEP 4: Glycopeptide sequence search

- Match preprocessed data to database of glycopeptides
- Extract all chromatograms in MS1 dimension
- Map MS/MS spectra onto chromatograms
- Estimate FDR by using target-decoy with a monotonic q-value

Identified glycopeptides are composed from matched chromatogram with mapped MS/MS spectra above q-value threshold

Build Search Space

BUILD A GLYCAN SEARCH SPACE
BUILD A GLYCOPEPTIDE SEARCH SPACE

Analyze Sample



SEARCH GLYCAN COMPOSITIONS
SEARCH GLYCOPEPTIDE SEQUENCES

View Results

VIEW DATABASE SEARCH RESULTS

Match Hypothesis Against Tandem Samples



Select one or more samples

20140903_05
20140903_05_phil-82

Choose a Hypothesis

Phil82_peptides_1_1_0.mzid

Select a glycopeptide
hypothesis from STEP 3

MS ¹ Mass PPM Error Tolerance	Peak Grouping PPM Error Tolerance	MS ² Mass PPM Error Tolerance
10	15	20
q-Value Threshold	Minimum Oxonium Threshold	
0.05	0.05	
SUBMIT		

Phil82_peptides_1_1_0.mzid at 20140903_05 (0)



MS¹ Mass Accuracy: 10.0 PPM
Grouping Tolerance: 15.0 PPM
MSⁿ Mass Accuracy: 20.0 PPM
Minimum q-Value For Clustering: 0.05

Target Database: [Phil82_peptides_1_1_0.mzid](#)
Sample Searched: [20140903_05](#)

ID: ed965f6f-6c34-4f51-b9ce-f6be8a0be816



Protein Name	Glycopeptides Matched	Glycosylation Site Count	Search by Scan ID
AFG99160 A/Philippines/2/1982	47	10	
ACN50246 A/Boshehr/388/2006	43	9	
AAA18781 A/Philippines/2/1982	41	10	
ADJ41805 A/Philippines/2/1982	41	10	

✓ Fuc	✓ Hex	✓ HexNAc	✓ Neu5Ac
0 : 3	0 : 9	0 : 7	0 : 4

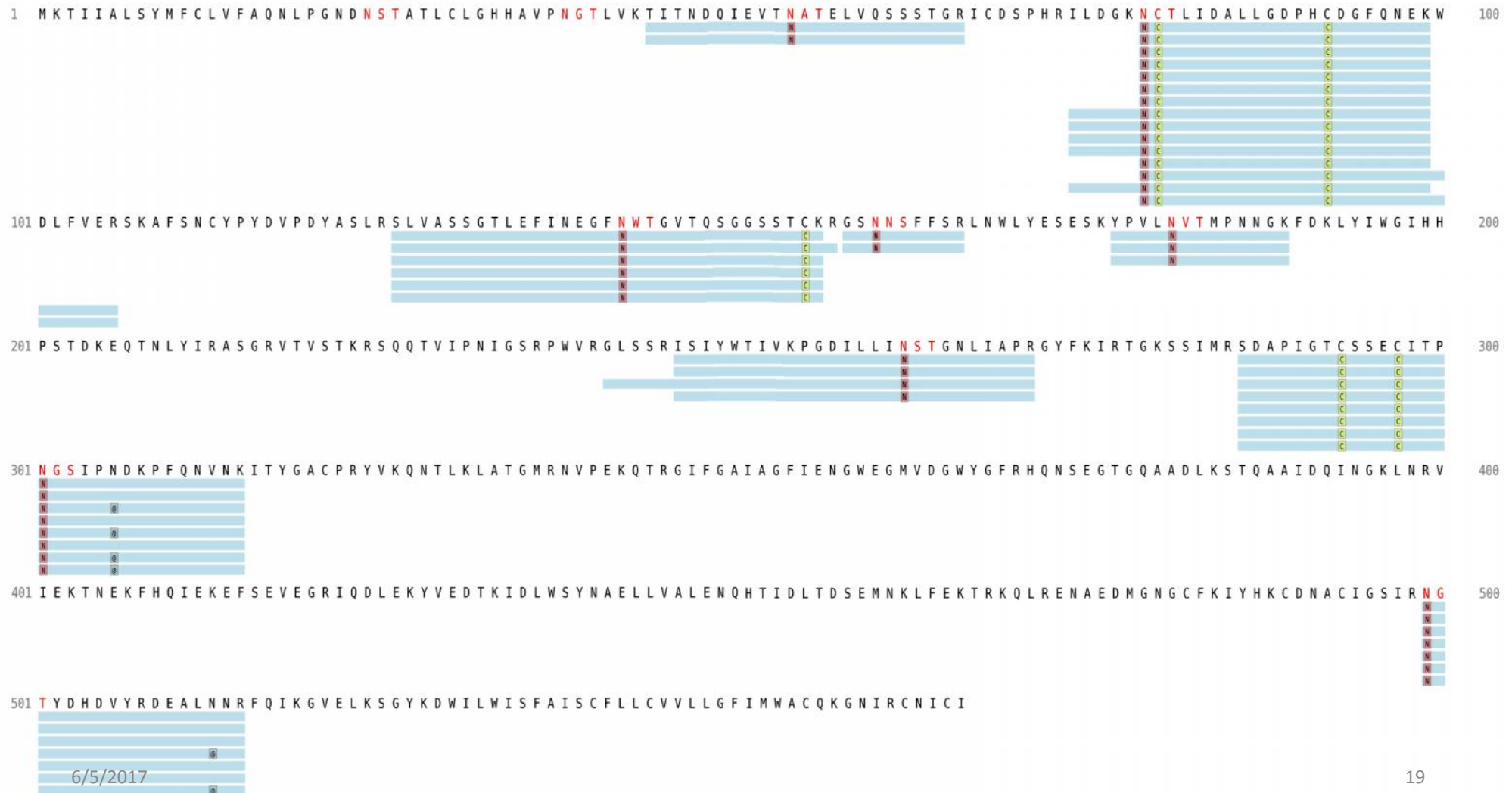
>AFG99160|A/Philippines/2/1982

MKTIIALSYMFLVFAQNLPGNDNSTATLCLGHHAVP~~N~~GTLVKTIITNDQIEVT~~N~~ATELVQSSSTGRICDSPHRILDGK~~N~~C
 TLIDALLGDPHCDGFQNEKWDLFVERSKAFSNCYPYDVPDYASLRSLVASSGTLEFINEGF~~N~~WTGVTQSGGSSTCKRGS~~N~~
 NSFFSRLNWLYESESKYPVL~~N~~VTMPNNGKFDKLYIWGIHHPSTDKEQTNLYIRASGRVTVSTKRSQQTVIPNIGSRPWVR
 GLSSRISIIYWTIVKPGDILL~~N~~STGNLIAPRGYFKIRTGKSSIMRSDAPIGTCSECTIP~~N~~GSIPNDKPFQNVNKITYGA
 CPRYYKQNTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGRHQNSEGTGQAADLKSTQAAIDQINGKLN RV
 IEKTEK~~N~~FKH~~N~~QIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTDSEM~~N~~NKLFKTRKQLRENAEDMGN
 GCFKIYHKCDNACIGSIR~~N~~GTYDHDVYRDEALNNRFQIKGVELKSGYKDWILWISFAISCFLLCVVLLGFIMWACQKGNI
 RCNICI

OVERVIEW

GLYCOPEPTIDES

SITE DISTRIBUTION



OVERVIEW

GLYCOPEPTIDES

SITE DISTRIBUTION

Observed Mass	Sequence	Tandem Score	q-value	Profile Score	Volume
4731.2828	ISIIYWTIVKPGDILLIN(N)STGNLIAPR Hex 9 HexNAc 2	152.2075	0.0	0.0000	7.79986e+07
3310.3762	YPVLN(N)VTMPNNGK Hex 9 HexNAc 2	144.9729	0.0	0.1493	7.23460e+06
3148.3205	YPVLN(N)VTMPNNGK Hex 8 HexNAc 2	142.9709	0.0	0.9545	5.94425e+06
3997.6446	N(N)C(C)TLIDALLGDPHC(C)DGFQNEK Hex 6 HexNAc 3	136.2318	0.0	0.9633	1.98568e+07
3632.5006	N(N)C(C)TLIDALLGDPHC(C)DGFQNEK Hex 5 HexNAc 2	135.2552	0.0	0.9662	8.54205e+07
3956.6096	N(N)C(C)TLIDALLGDPHC(C)DGFQNEK Hex 7 HexNAc 2	135.2195	0.0	0.9658	2.52317e+07
4241.7457	N(N)C(C)TLIDALLGDPHC(C)DGFQNEK Hex 5 HexNAc 5	128.4610	0.0	0.9652	3.49300e+07
4949.0884	SDAPIGTC(C)SSEC(C)ITPN(N)GSIPNDKPFQNVNK Hex 8 HexNAc 2	127.6049	0.0	0.9495	9.19823e+06
2986.2669	YPVLN(N)VTMPNNGK Hex 7 HexNAc 2	127.1043	0.0	0.9496	9.39625e+05
3794.5537	N(N)C(C)TLIDALLGDPHC(C)DGFQNEK Hex 6 HexNAc 2	125.3592	0.0	0.7708	2.21185e+07
4387.8035	N(N)C(C)TLIDALLGDPHC(C)DGFQNEK Fuc 1 Hex 5 HexNAc 5	125.2955	0.0	0.9557	8.48322e+06
4143.6981	N(N)C(C)TLIDALLGDPHC(C)DGFQNEK Fuc 1 Hex 6 HexNAc 3	119.4886	0.0	0.9574	9.95420e+06
5111.1442	SDAPIGTC(C)SSEC(C)ITPN(N)GSIPNDKPFQNVNK Hex 9 HexNAc 2	118.4171	0.0	0.9473	5.62079e+06
4482.9180	ILDGKN(N)C(C)TLIDALLGDPHC(C)DGFQNEK Hex 7 HexNAc 2	102.7344	0.0	0.7411	3.55297e+06

Glycopeptide Sequence

N(N-

Glycosylation)C(Carbamidomethyl)TLIDALLGDPHC(Carbamidomethyl)DGFQNEK

Hex 6 HexNAc 3

Start Position: 78 End Position: 99

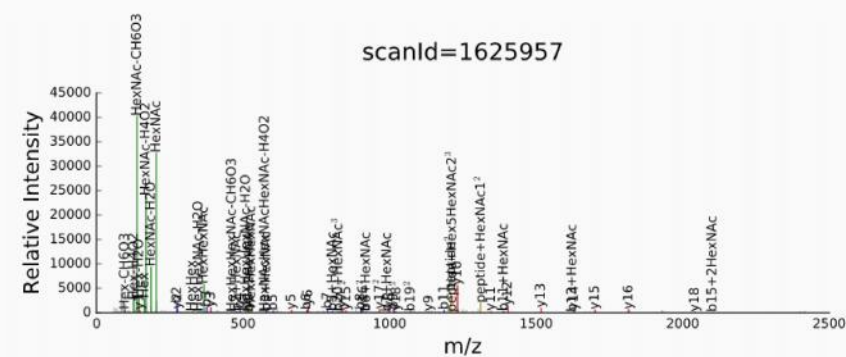
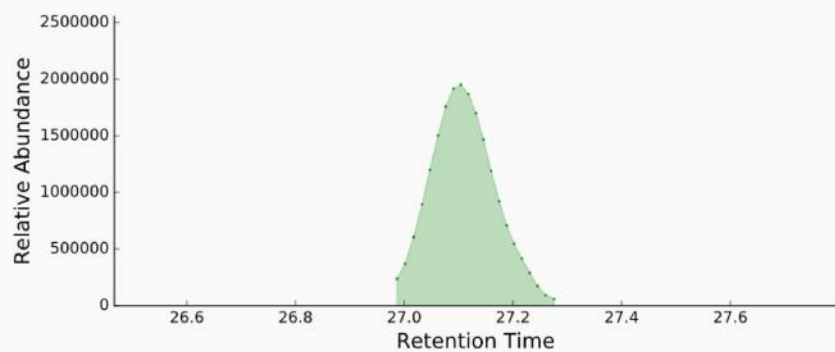
Observed Mass (PPM Error): 3997.64461822

(-1.014e-06)

Spectrum Matches: 5

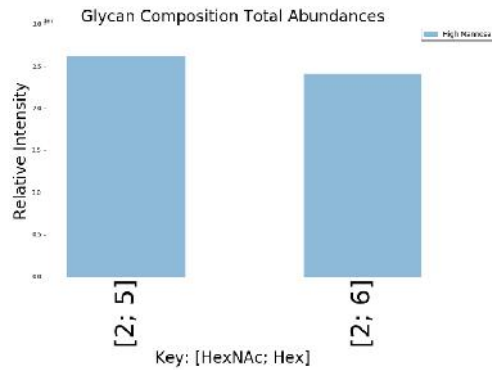
Charge States: 3, 4, 5

Profile Score	Tandem Score	TDA q-value	Volume
0.9633	136.2318	0.0000	1.9857e+07

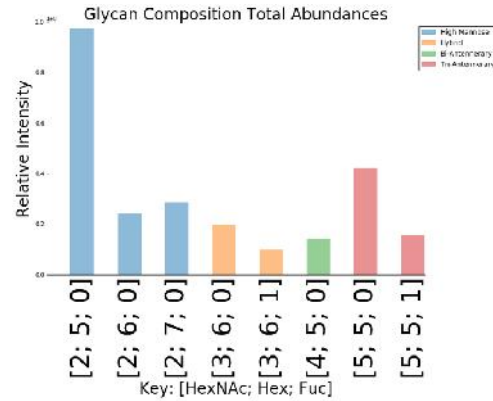


OVERVIEW

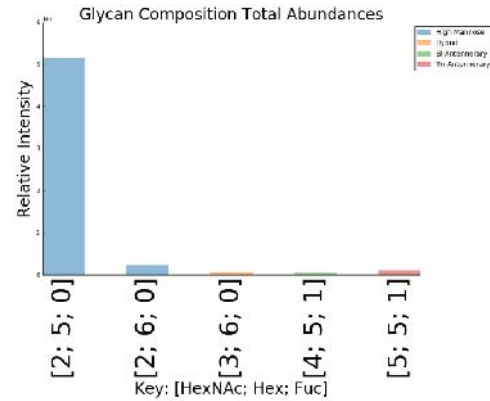
Site 53



Site 78

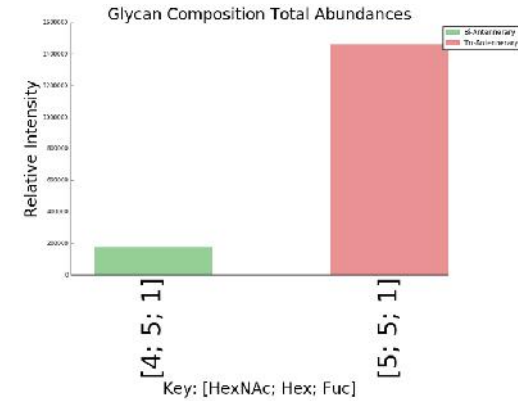


Site 141

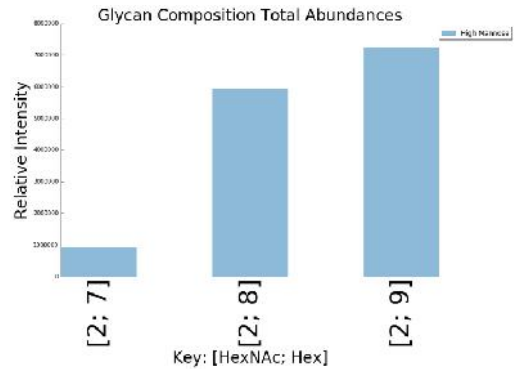


SITE DISTRIBUTION

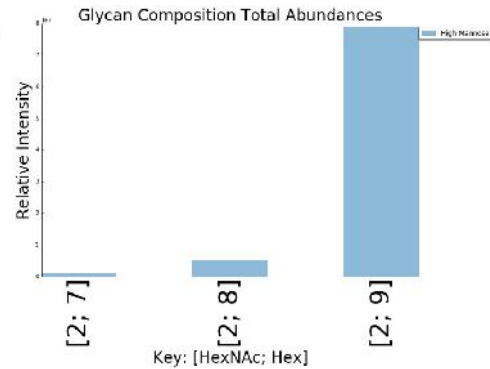
Site 159



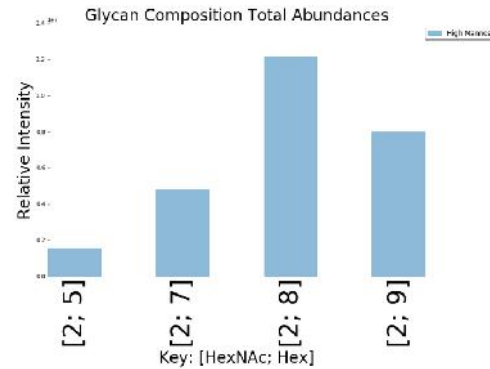
Site 180



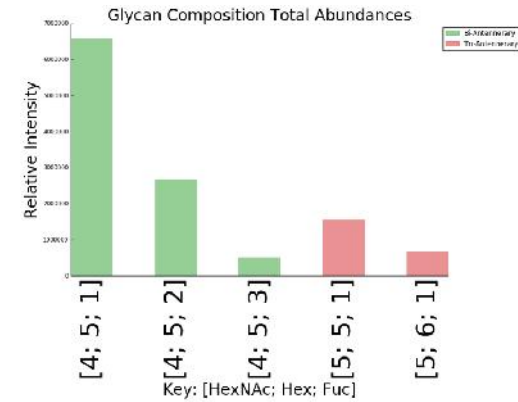
Site 261



Site 300



Site 498



Export Identifications As Text Files:

- ☐ glycopeptides (csv)
- ☐ glycopeptide spectrum matches (csv)
- ☐ mzIdentML (mzid 1.1.0)
- ☐ associated glycans (txt)