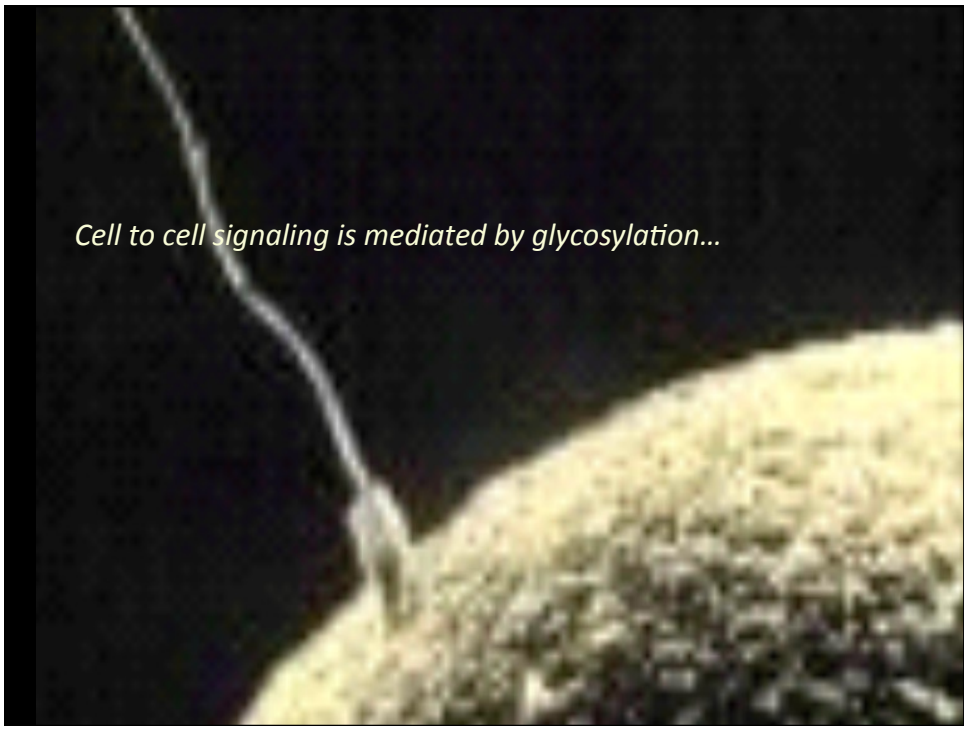
Bioinformatics of Glycans and Glycoproteins

Marshall Bern
Palo Alto Research Center
Protein Metrics

parc
A Xerox Company

PROTEIN METRICS



Outline

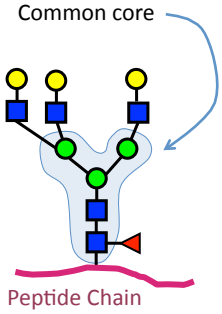
1) Background – Glycomics vs. Proteomics bioinformatics
2) Automatic detached N-glycan analysis from MS1 – Cartoonist
3) Semi-automatic glycan analysis from MS1 and MS2 – GlycoWorkbench
4) Automatic glycan analysis from MS2 – SimGlycan
5) Glycopeptide analysis from MS2 – Byonic

Outline

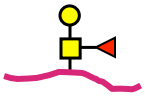
1) Background – Glycomics vs. Proteomics bioinformatics
2) Automatic detached N-glycan analysis from MS1 – Cartoonist
3) Semi-automatic glycan analysis from MS1 and MS2 – GlycoWorkbench
4) Automatic glycan analysis from MS2 – SimGlycan
5) Glycopeptide analysis from MS2 – Byonic

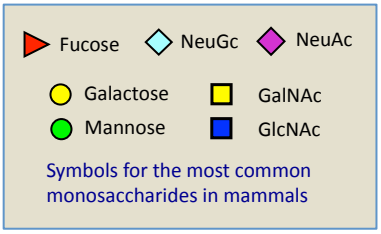
Two Types of Glycosylation

- N-linked Glycosylation
  - Attached to Asparagine
  - NX{S/T} motif, X ≠ P
    (rare case – NXC)
  - 1200 – 12,000 Da
  - Common core



- O-linked Glycosylation
  - Attached to Serine or Threonine
    (rare case – Tyrosine)
  - No consensus motif
  - 200 – 2000 Da
  - At least 7 cores





Proteomics vs. Glycomics

Peptides	Glycans
1. 19 different residue masses Total peptide mass uninformative	1. ~5 monosaccharide masses Total mass informative
2. Linear sequence	2. Branched structure
3. Only one type of peptide bond	3. Various types of linkages
4. Any sequence is valid	4. Structures follow certain patterns
5. Good databases exist	5. Databases have poor coverage

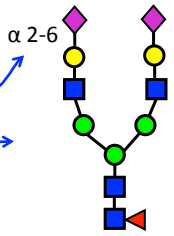
= Easier

= Harder

Proteomics vs. Glycomics

Peptides	Glycans
1. 19 different residue masses Total peptide mass uninformative	1. ~5 monosaccharide masses Total mass informative
2. Linear sequence	2. Branched structure
3. Only one type of peptide bond	3. Various types of linkages
4. Any sequence is valid	4. Structures follow certain patterns
5. Good databases exist	5. Databases have poor coverage

Different levels of glycan identification:
- Linkage information
- Cartoon (branching topology)
- Composition



4 HexNAc, 5 Hex, 1 Fuc, 1 NeuAc

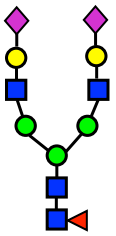
Proteomics vs. Glycomics

Peptides	Glycans
1. ~~19 different residue masses~~ Total peptide mass uninformative	1. ~~~5 monosaccharide masses~~ Total mass informative
2. Linear sequence	2. Branched structure
3. Only one type of peptide bond	3. Various types of linkages
4. Any sequence is valid	4. Structures follow certain patterns
5. Good databases exist	5. Databases have poor coverage

- ~1000 tryptic peptides in human proteome with mass 2351 Daltons
- ~20 in single sample:

LGLLVFPYTHQNWEVQYSR
EVTLHLLPGEQLLC[+57]EASTVLK
LQEEMLQREEAENTLQSFR
etc.

2351 Daltons
(not permethylated)



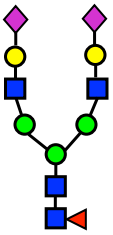
Proteomics vs. Glycomics

Peptides	Glycans
1. ~~19 different residue masses~~ Total peptide mass uninformative	1. ~~~5 monosaccharide masses~~ Total mass informative
2. Linear sequence	2. Branched structure
3. Only one type of peptide bond	3. Various types of linkages
4. Any sequence is valid	4. Structures follow certain patterns
5. Good databases exist	5. Databases have poor coverage

- ~1000 tryptic peptides in human proteome with mass 2351 Daltons
- ~20 in single sample:

LGLLVFPYTHQNWEVQYSR
EVTLHLLPGEQLLC[+57]EASTVLK
LQEEMLQREEAENTLQSFR
etc.

2351 Daltons
(not permethylated)

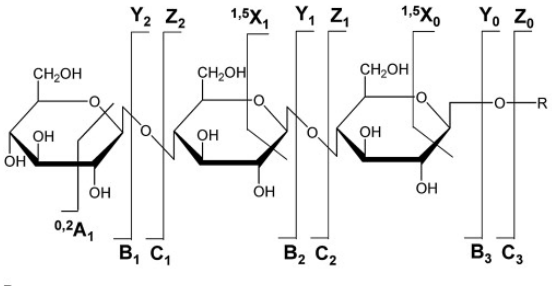


→ Proteomics uses MS2 for identification
Glycomics can use either MS1 or MS2 or both

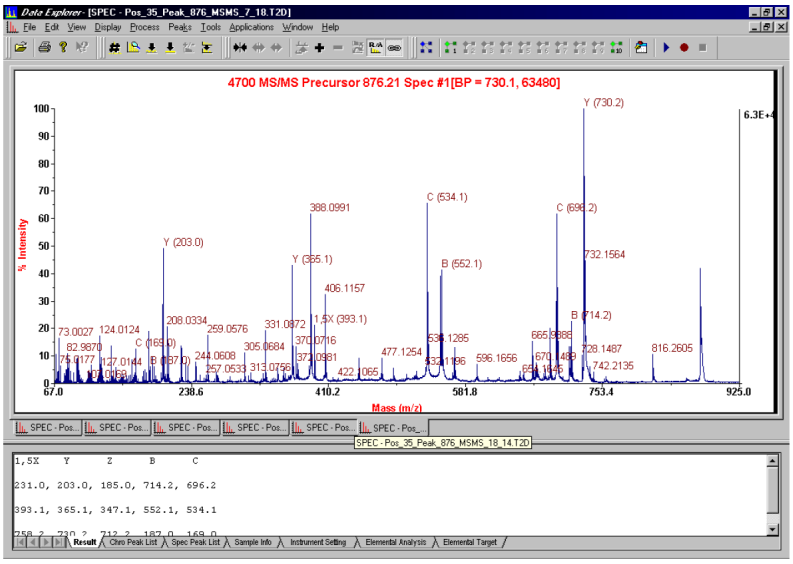
How do glycans fragment in MS2 ?

Domon-Costello nomenclature

- Draw the glycan with reducing end to the right
- Cut before ┼O – left half = B-ion, right half = Y-ion
- Cut after –O┼ left half = C-ion, right half = Z-ion
- Internal fragments can be YY, ZZ, etc.
- Cross-ring = X ^{1,5}X is most common



How do glycans fragment in MS2 ?



From AB Sciex brochure

Outline

1) Background – Glycomics vs. Proteomics bioinformatics
2) Automatic detached N-glycan analysis from MS1 – Cartoonist
3) Semi-automatic glycan analysis from MS1 and MS2 – GlycoWorkbench
4) Automatic glycan analysis from MS2 – SimGlycan
5) Glycopeptide analysis from MS2 – Byonic

Cartoonist Software

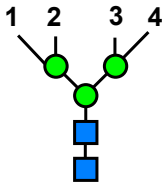
- Developed at Xerox PARC, mostly by David Goldberg (currently Ebay fellow)
- Glycomics guidance from Anne Dell, Stuart Haslam at Imperial College
- Platform-independent Java software, Free!, bundled with PARC mass spectrum browser
- Works well for MALDI-TOF profiles of detached, permethylated, mammalian N-glycans
- Works to some extent for other types of data

Cartoonist uses an "expert systems" approach

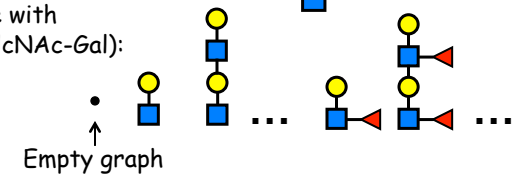
1. Use **graph grammar** to generate a library of all plausible cartoons (topology only, no linkage information!)
   ~145,000 N-glycan cartoons in library (bigger than current databases)
2. Find the **set** (typically 1 – 20) matching each MS peak
3. Use organism- and tissue-specific **hints** to select most likely cartoon from each set

N-Glycan Graph Grammar

1. Start with N-glycan core

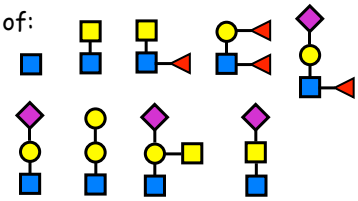


2. Fill antennae with lactosamine (GlcNAc-Gal):



Empty graph

3. Cap with one of:



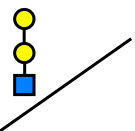
4. Optionally add:

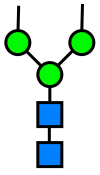
Bisecting GlcNAc,
Base fucose,
NeuAc → NeuGc

Example – 2653 Daltons

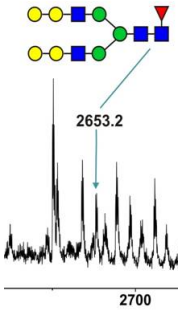
No lactosamines: •

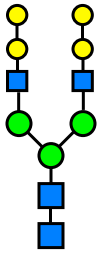
GlcNAc-Gal-Gal caps:



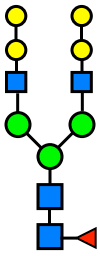


Core

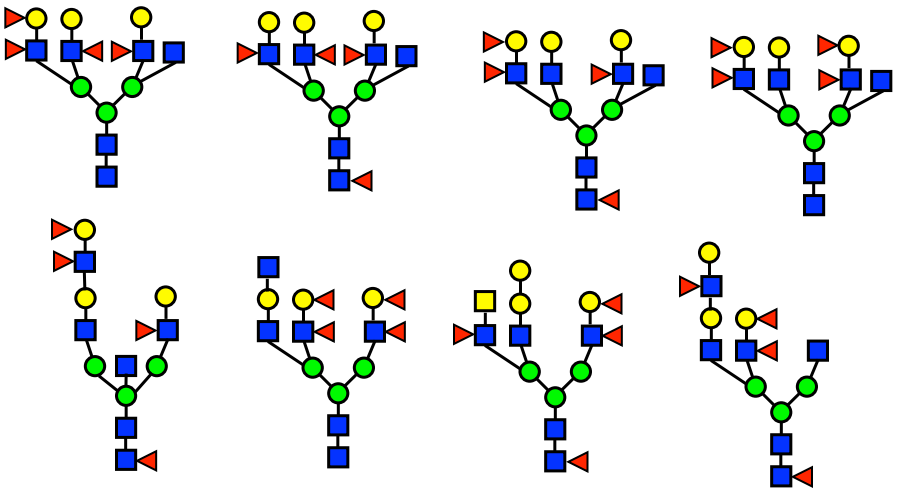




Base Fucose



8 of the 18 cartoons for 3458 Daltons



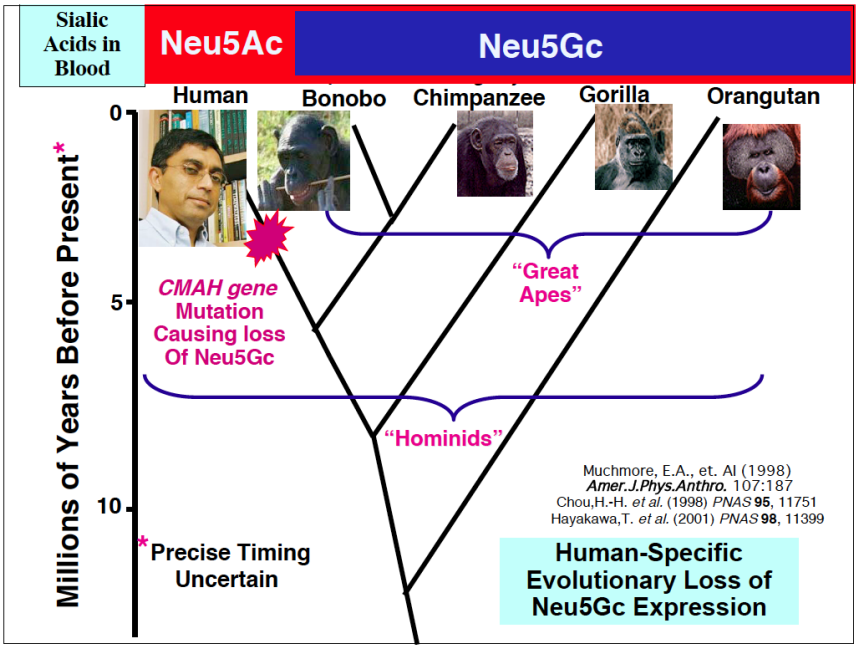
Organism-Specific Hints (Penalties)

- Unlikely patterns for all mammals:
  - Bisecting GlcNAc
  - Uncapped terminal GlcNAc
  - Fucose on antennae, but not on base
- Specific to Mouse:
  - LacdiNAc, Sialyl Lewis-X, Multiple-fucose antenna
- Specific to Human:
  - Gal-alpha-Gal cap
  - NeuGc (Is this what makes us human??)

Organism-Specific Hints (Penalties)

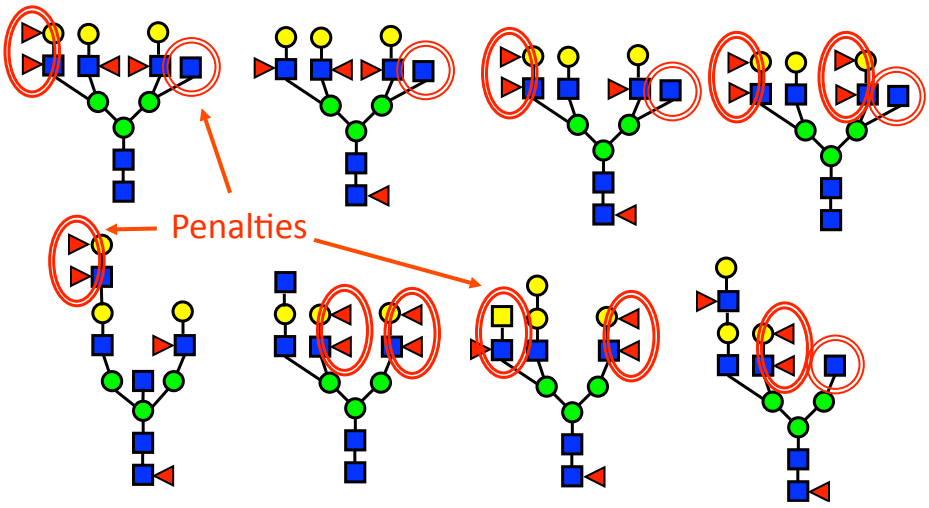
- Unlikely patterns for all mammals:
  - Bisecting GlcNAc
  - Uncapped terminal GlcNAc
  - Fucose on antennae, but
- Specific to Mou
  - LacdiNAc, Sialyl
- Specific to Human:
  - Gal-alpha-Gal cap
  - NeuGc (Is this what makes us human??)

```
# for Human
#
D_STACKED          3
D_MAN4             1
D_BISECTED         1
D_MULTIPLE_FUCOSE  1
D_MULTIPLE_SIALIC  1
D_SIAL_LEWIS_X     1
D_NO_BASE_FUC      1
D_HAS_5_ANTENNAE   1
D_LACDINAC         3
D_BAD_HYBRID       1
D_NAKED_GLCNAC     1
D_SDA              1
D_ONE_ANTENNAE     3
D_SIAL_LACDINAC    3
D_NAKED_FUC_GLCNAC 3
D_HAS_GAL          0
D_LACTOSAMINE      0
D_ONE_ARM          3
D_GLCNAC_ON_GAL    0
D_GAL_ALPHA_GAL    5
D_NEUGC            5
```

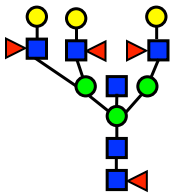


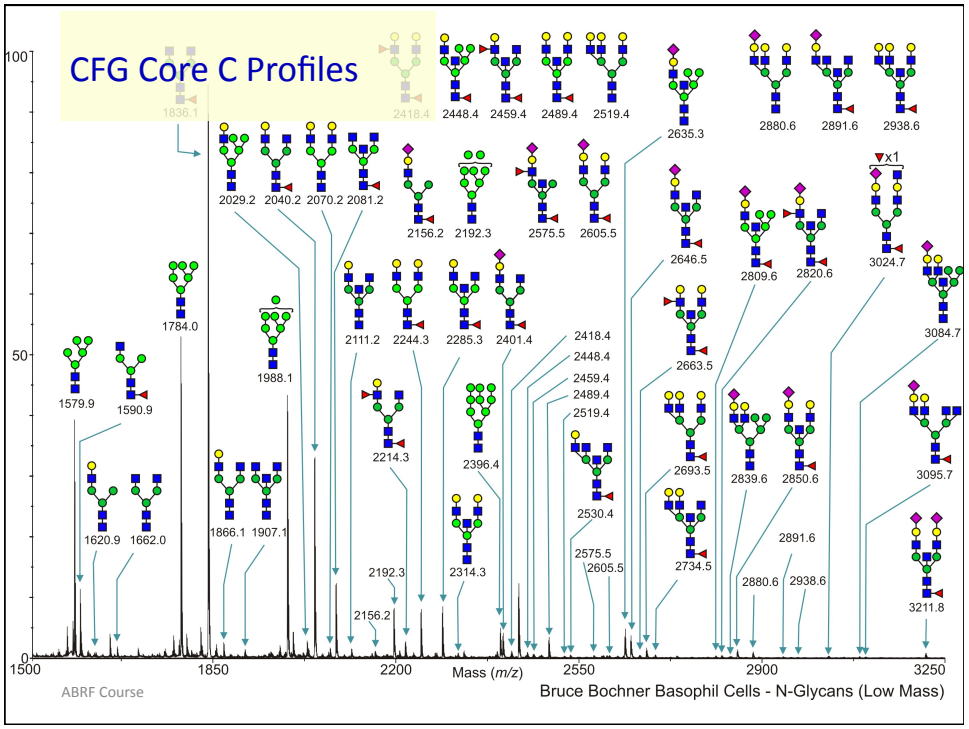
From a talk by Ajit Varki, UCSD

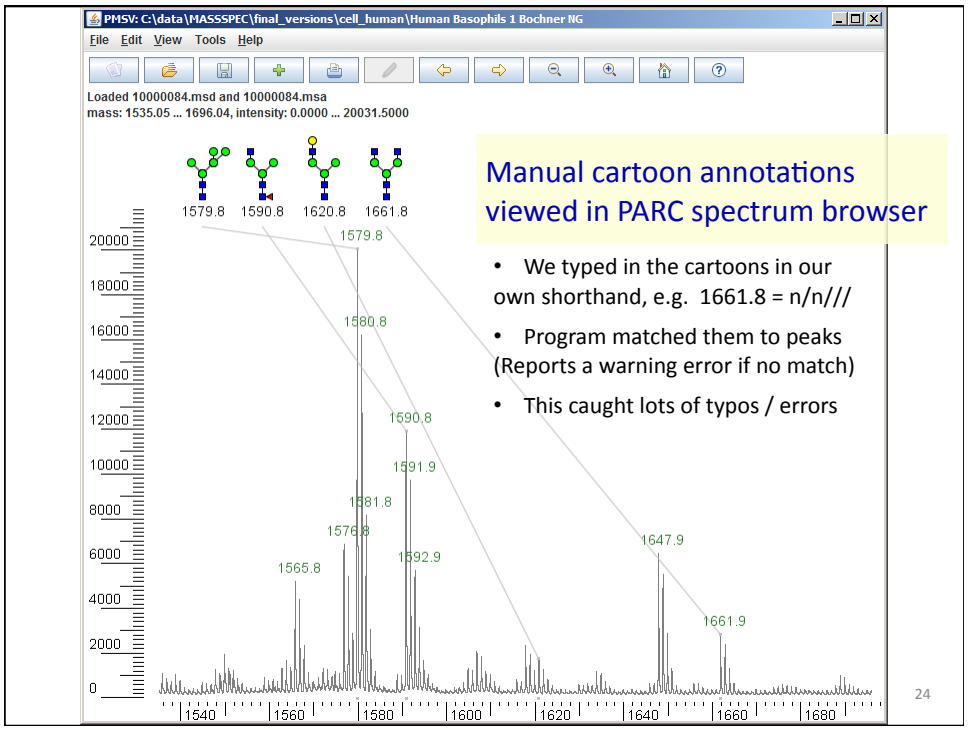
8 of the 18 cartoons for 3458 Daltons

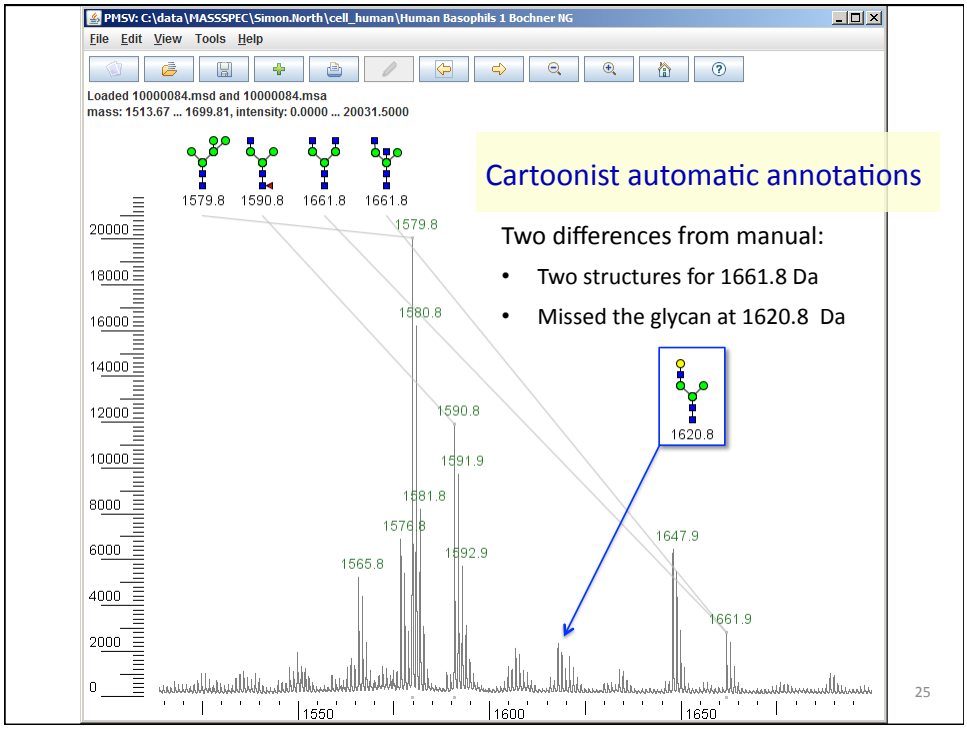


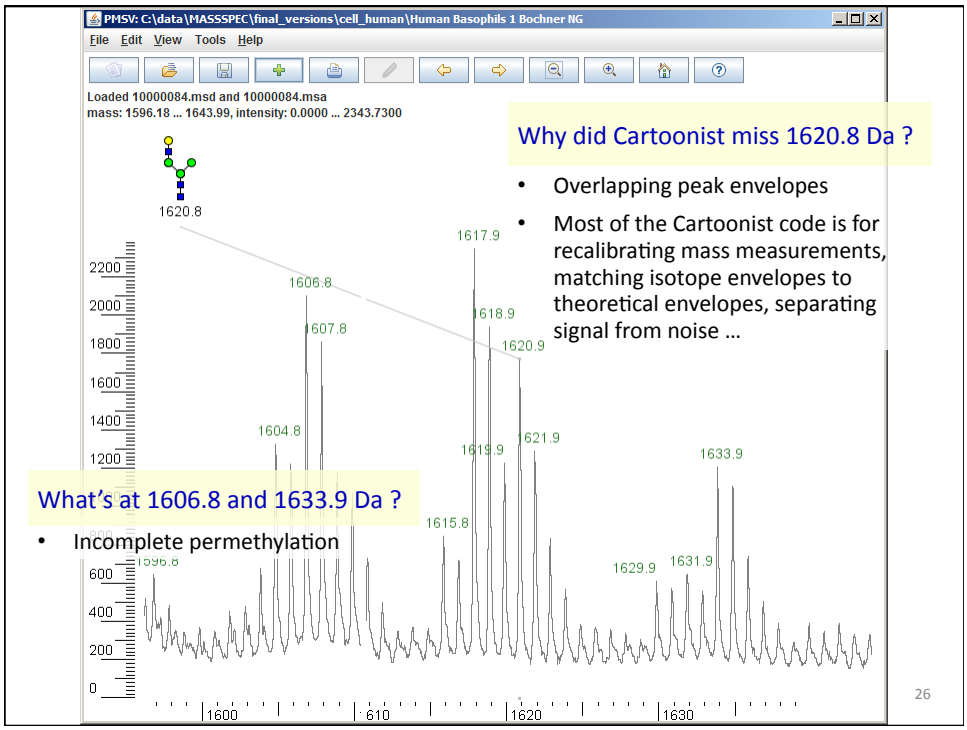
Only 1 survivor!




CFG Core C Profiles
100
50
0
1500
1850
2200
2550
2900
3250
Mass (m/z)
ABRF Course
Bruce Bochner Basophil Cells - N-Glycans (Low Mass)
1579.9
1590.9
1620.9
1662.0
1784.0
1866.1
1907.1
1988.1
2029.2
2040.2
2070.2
2081.2
2111.2
2156.2
2192.3
2214.3
2244.3
2285.3
2314.3
2396.4
2401.4
2418.4
2448.4
2459.4
2489.4
2519.4
2530.4
2575.5
2605.5
2635.3
2646.5
2663.5
2693.5
2734.5
2809.6
2820.6
2839.6
2850.6
2880.6
2891.6
2938.6
3024.7
3084.7
3095.7
3211.8
x1

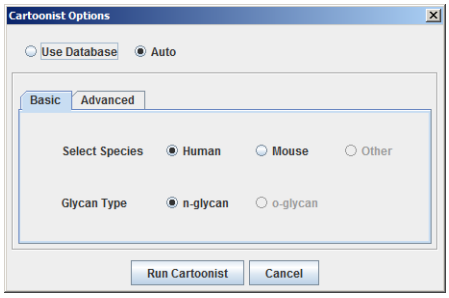

PMSV: C:\data\MASSSPEC\final_versions\cell_human\Human Basophils 1 Bochner NG
File Edit View Tools Help
Loaded 10000084.msd and 10000084.msa
mass: 1535.05 ... 1696.04, intensity: 0.0000 ... 20031.5000
1579.8
1590.8
1620.8
1661.8
1580.8
1590.8
1591.9
1581.8
1576.8
1565.8
1592.9
1647.9
1661.9
Manual cartoon annotations viewed in PARC spectrum browser
• We typed in the cartoons in our own shorthand, e.g. 1661.8 = n/n///
• Program matched them to peaks (Reports a warning error if no match)
• This caught lots of typos / errors
24

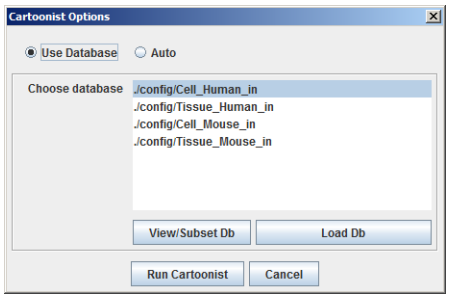

PMSV: C:\data\MASSSPEC\Simon.North\cell_human\Human Basophils 1 Bochner NG
File Edit View Tools Help
Loaded 10000084.msd and 10000084.msa
mass: 1513.67 ... 1699.81, intensity: 0.0000 ... 20031.5000
1579.8 1590.8 1661.8 1661.8
Cartoonist automatic annotations
Two differences from manual:
• Two structures for 1661.8 Da
• Missed the glycan at 1620.8 Da
1620.8
1579.8
1580.8
1590.8
1591.9
1581.8
1576.8
1565.8
1592.9
1647.9
1661.9
1550
1600
1650
25


PMSV: C:\data\MASSSPEC\final_versions\cell_human\Human Basophils 1 Bochner NG
File Edit View Tools Help
Loaded 10000084.msd and 10000084.msa
mass: 1596.18 ... 1643.99, intensity: 0.0000 ... 2343.7300
1620.8
Why did Cartoonist miss 1620.8 Da ?
• Overlapping peak envelopes
• Most of the Cartoonist code is for recalibrating mass measurements, matching isotope envelopes to theoretical envelopes, separating signal from noise …
1617.9
1606.8
1618.9
1607.8
1620.9
1604.8
1621.9
1619.9
1633.9
1615.8
1629.9
1631.9
What's at 1606.8 and 1633.9 Da ?
• Incomplete permethylation
1600
610
1620
1630
26

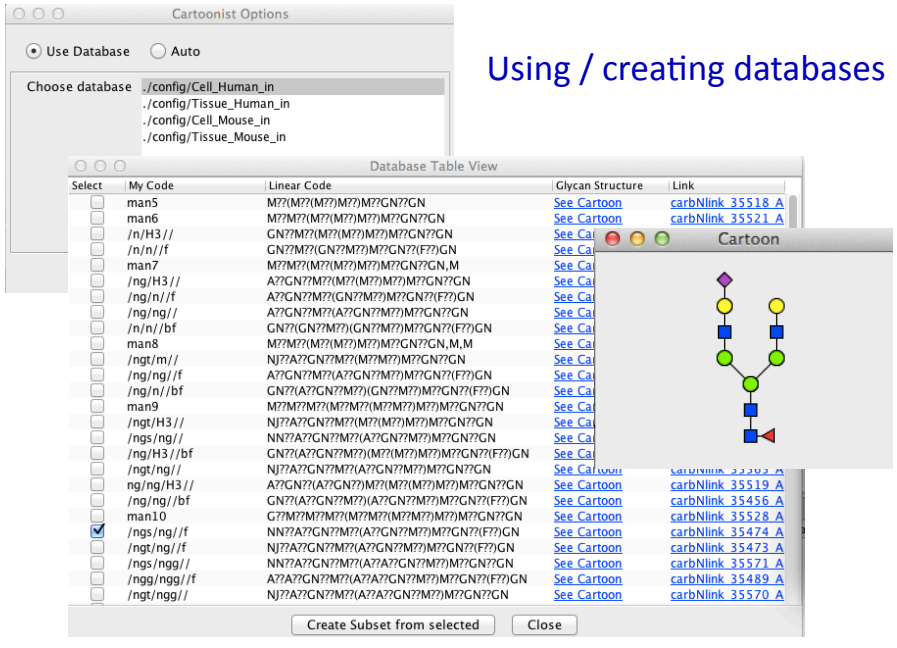
How to run Cartoonist?

- Go to Tools menu in PARC mass spectrum viewer
- Auto makes up the cartoons according to "expert system" rules
- Use Database matches a cartoon list to peaks (the same way we digitized CFG profiles) Best for O-glycans, non-mammalian N-glycans, …





Using / creating databases



Outline

1) Background – Glycomics vs. Proteomics bioinformatics

2) Automatic detached N-glycan analysis from MS1 – Cartoonist

3) Semi-automatic glycan analysis from MS1 and MS2 – GlycoWorkbench

4) Automatic glycan analysis from MS2 – SimGlycan

5) Glycopeptide analysis from MS2 – Byonic

GlycoWorkbench

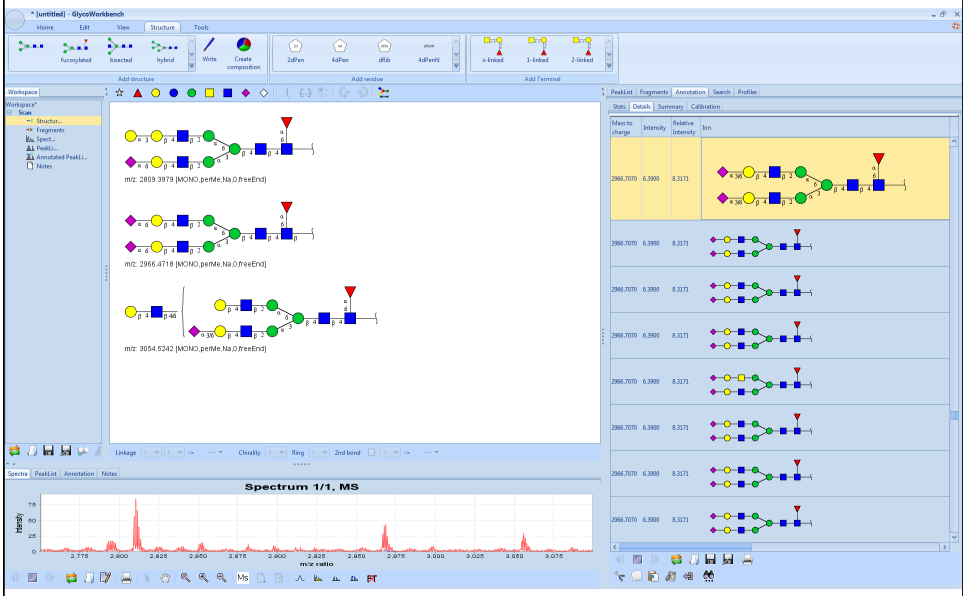
- Developed at Imperial College
  (Primary Authors: Alessio Ceroni, Kai Maass, David Damerell)

- Java code for Windows, Mac OS, and Linux. **Free!**

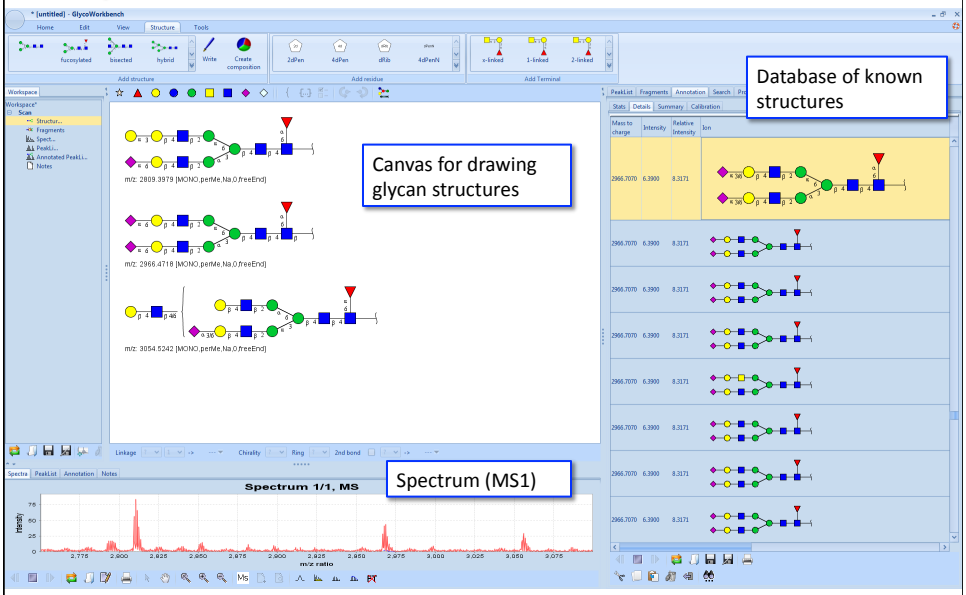
http://www.softpedia.com/get/Science-CAD/GlycoWorkbench.shtml

- Functions:
  - Match MS1 peaks to glycan structures in a database (GlycomeDB)
    (Caveat: Matches everything! No expert system knowledge)
  - Match MS2 peaks to fragments of a selected structure
  - Draw new structures and calculate masses

GlycoWorkbench



GlycoWorkbench



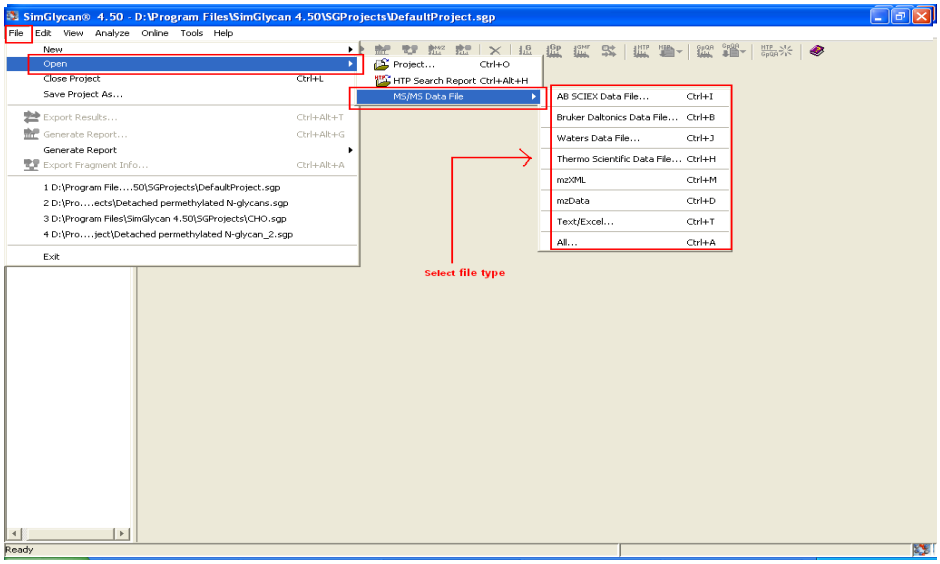
Outline

1) Background – Glycomics vs. Proteomics bioinformatics
2) Automatic detached N-glycan analysis from MS1 – Cartoonist
3) Semi-automatic glycan analysis from MS1 and MS2 – GlycoWorkbench
4) Automatic glycan analysis from MS2 – SimGlycan
5) Glycopeptide analysis from MS2 – Byonic

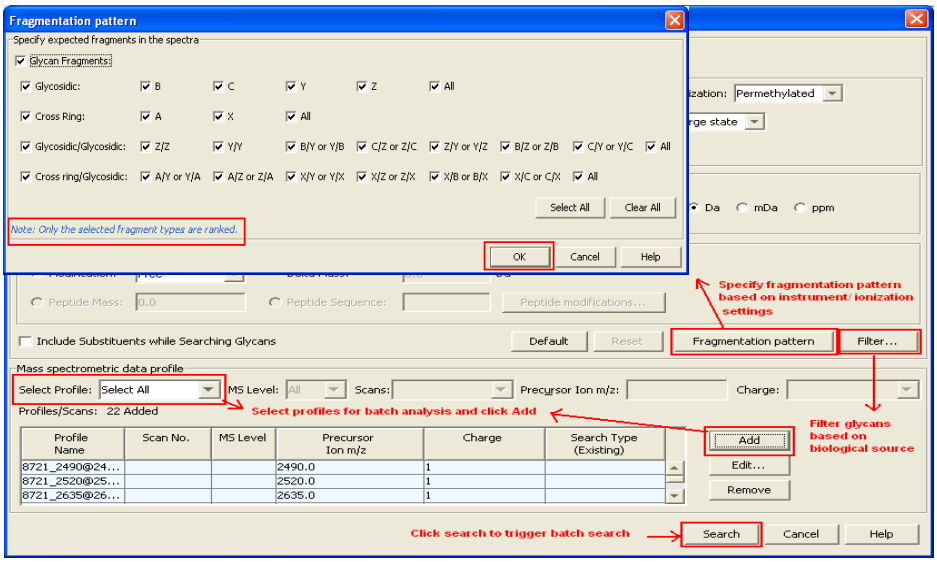
SimGlycan

- Commercial software from Premier Biosoft (Palo Alto, CA)
- Annual license, Approx. $10K / year
- Functions:
  - Identifies glycan structures from MS2 fragmentation spectra
  - Structures are chosen from databases of known glycans (e.g., GlycomeDB)
  - High-throughput: can run a batch of 10,000 MS2 spectra
  - Scoring algorithm: which structure has highest percentage of its fragments matching MS2 peaks ?

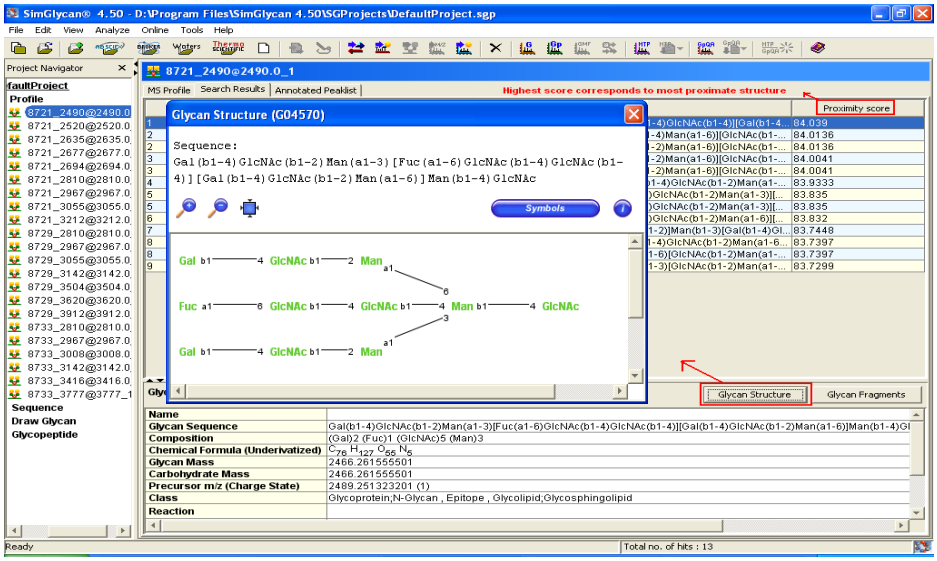
Import data from MS instrument formats



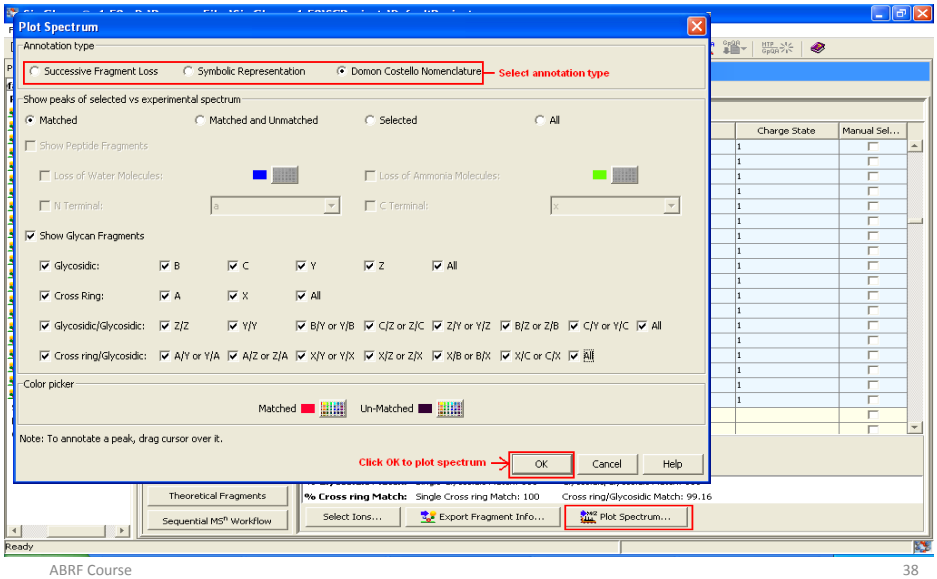
Specify fragmentation patterns and database filters

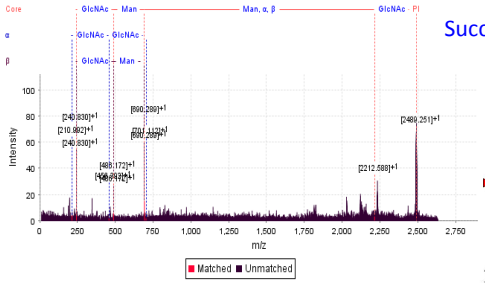


Results: Identified glycans, related info, and annotated spectra



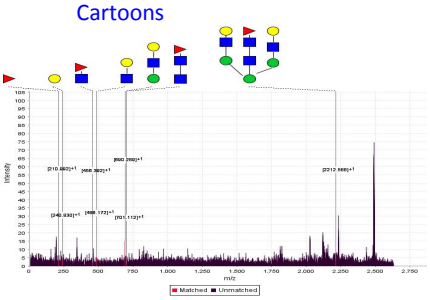
Three ways to annotate MS2 spectra:
Successive fragment losses, cartoons, or Domon-Costello



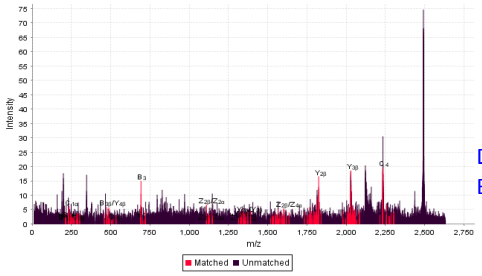


Successive fragment losses

Cartoons

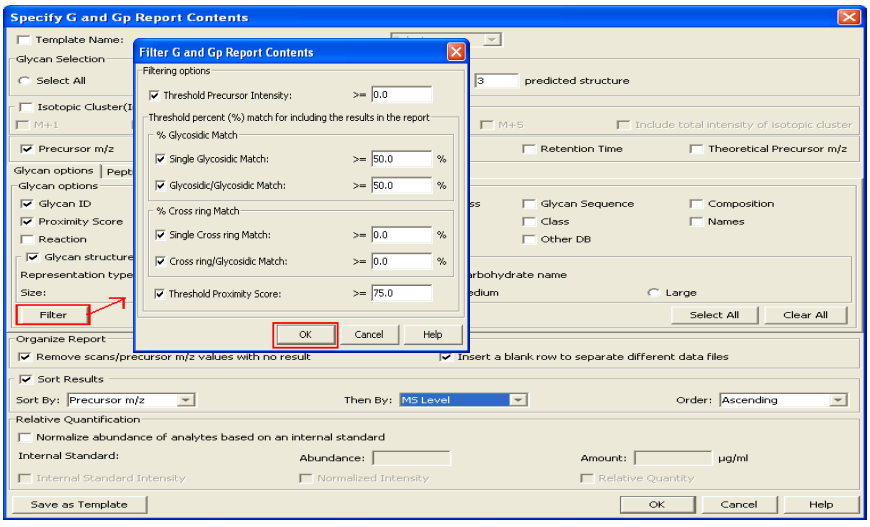


Mass Spectrum: m/z versus Intensity



Domon-Costello
B, Y, C, Z, etc.

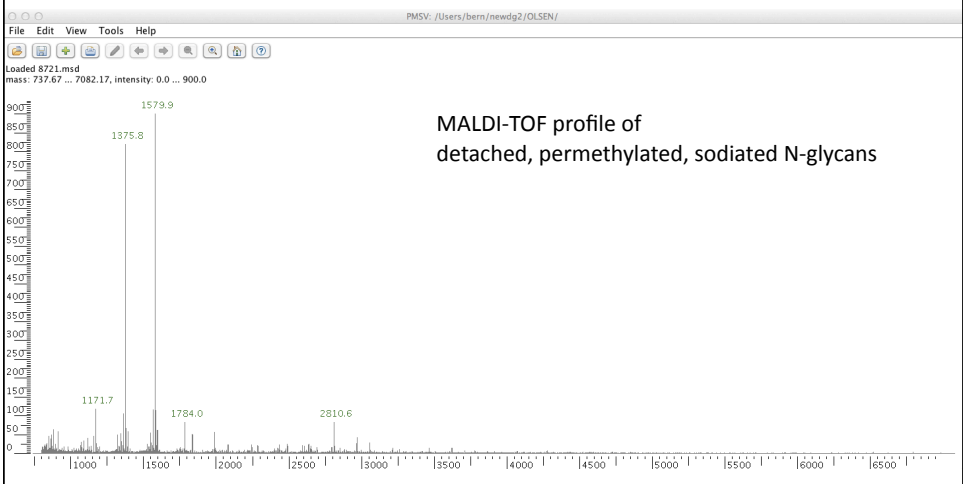
Customize Report: Filter, Sort, Present Results



Outline

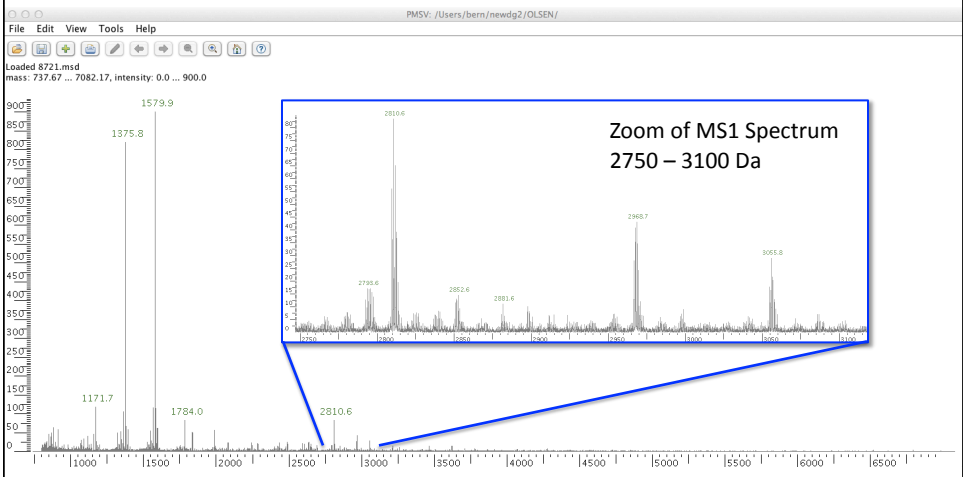
1) Background – Glycomics vs. Proteomics bioinformatics
2) Automatic detached N-glycan analysis from MS1 – Cartoonist
3) Semi-automatic glycan analysis from MS1 and MS2 – GlycoWorkbench
4) Automatic glycan analysis from MS2 – SimGlycan
5) Glycopeptide analysis from MS2 – Byonic

Example data run through Cartoonist, GlycoWorkbench, and SimGlycan

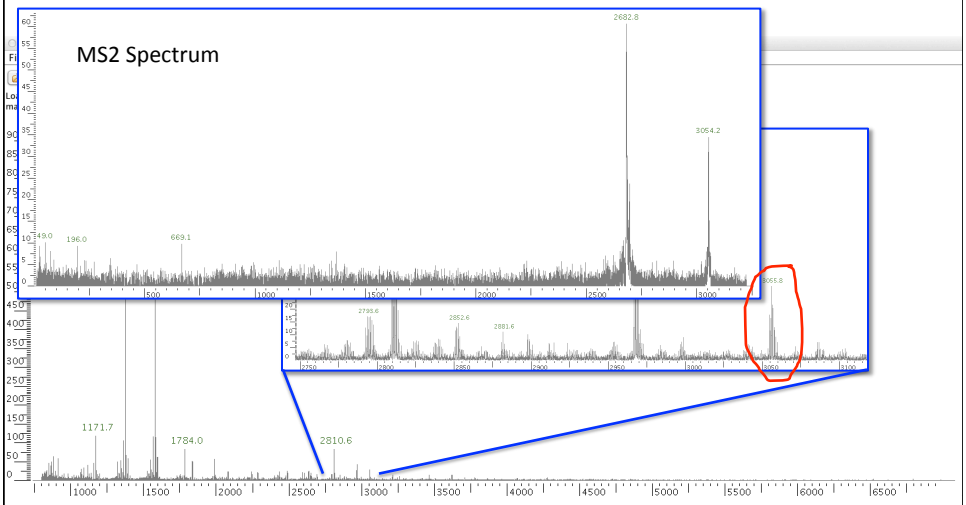


MALDI-TOF profile of
detached, permethylated, sodiated N-glycans

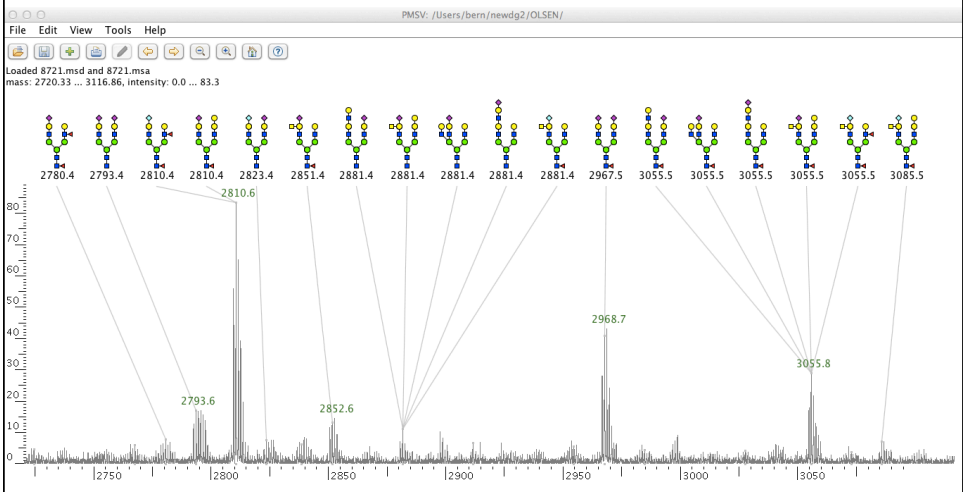
Example data run through Cartoonist, GlycoWorkbench, and SimGlycan



MALDI – TOF/TOF spectrum on selected peak (3055 Da)

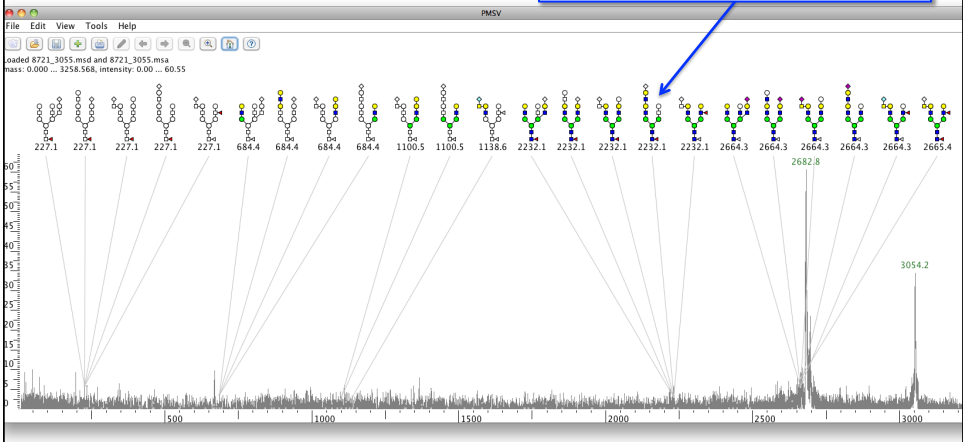


Cartoonist

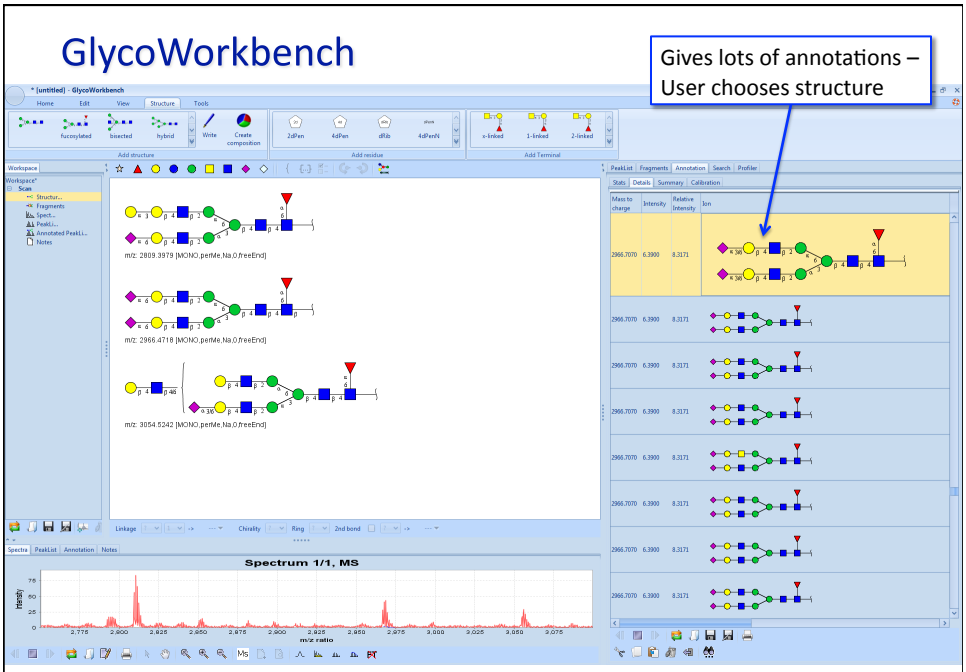


Cartoonist

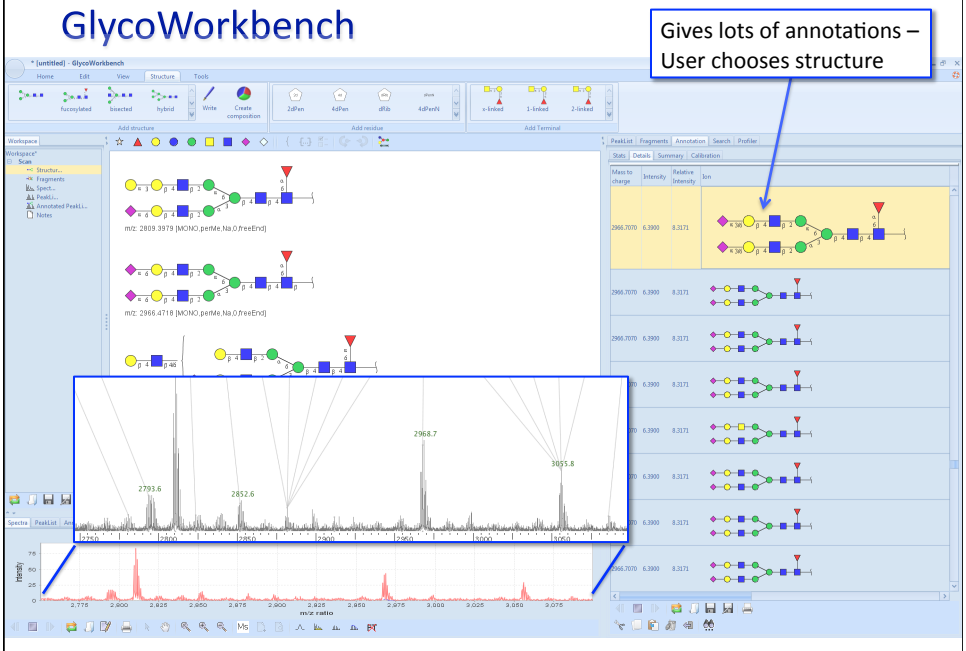
MS2 peaks are annotated with fragments of structures determined by MS1



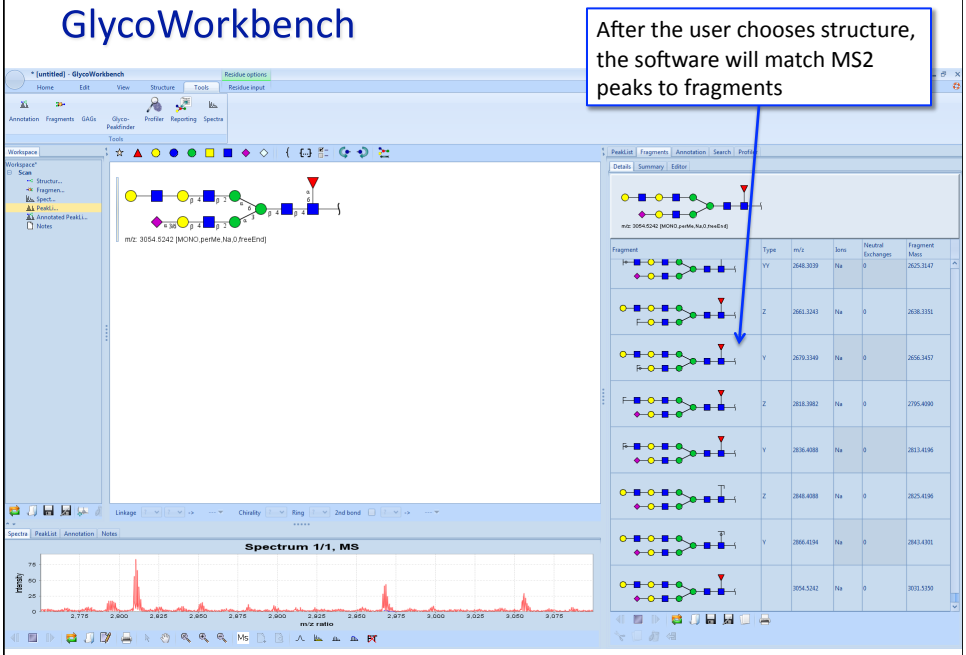
GlycoWorkbench



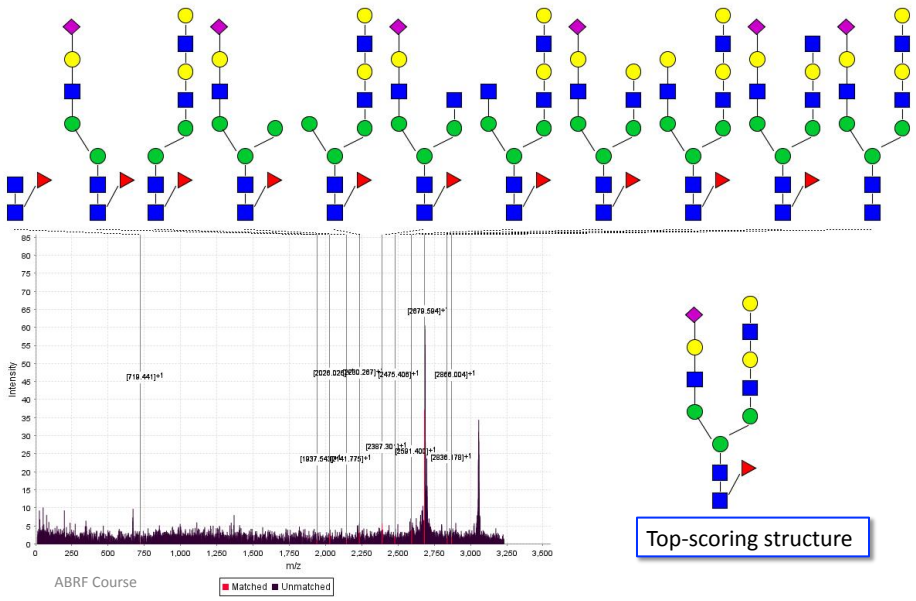
GlycoWorbench



GlycoWorkbench

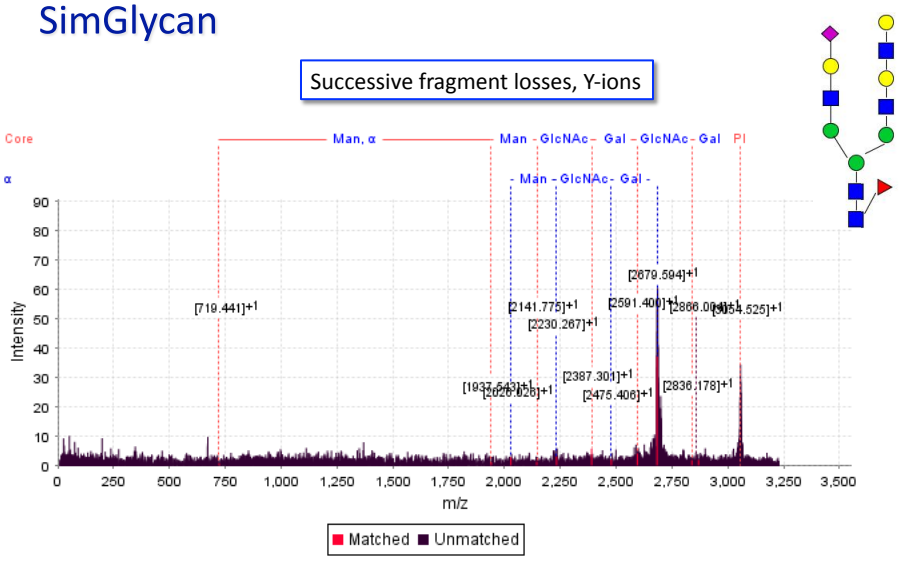


SimGlycan



SimGlycan

Successive fragment losses, Y-ions



Outline

1) Background – Glycomics vs. Proteomics bioinformatics
2) Automatic detached N-glycan analysis from MS1 – Cartoonist
3) Semi-automatic glycan analysis from MS1 and MS2 – GlycoWorkbench
4) Automatic glycan analysis from MS2 – SimGlycan
5) Glycopeptide analysis from MS2 – Byonic

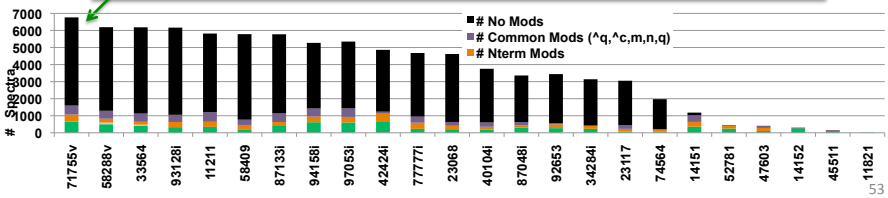
Byonic Software

- Developed at Xerox PARC and Protein Metrics Inc, 2005 – 2013
  (Primary Authors: Marshall Bern and Yong J. Kil)
- About $10K perpetual license, $3K annual renewal
- Will be released as a node in Thermo Fisher's Proteome Discoverer 2.0
- Functions:

  - Peptide / Protein identification from MS2 data

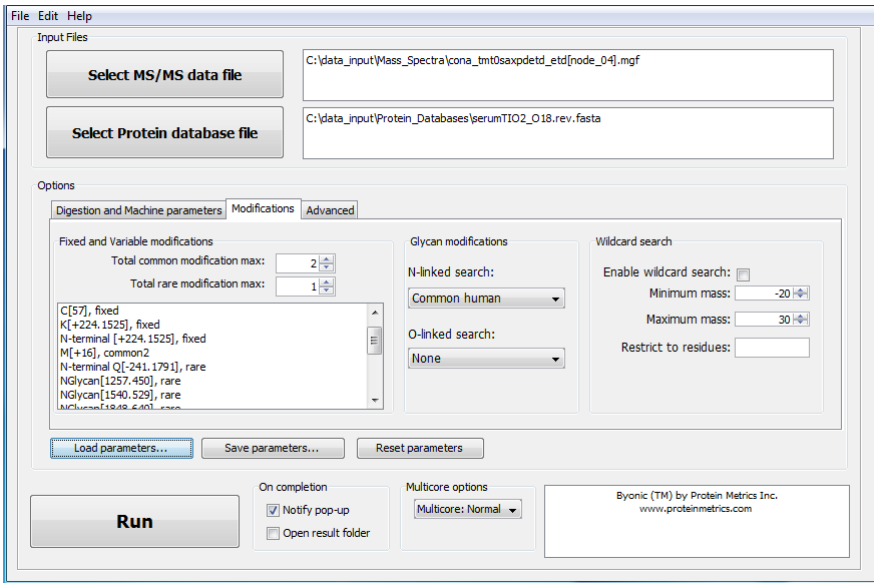
  - Glycopeptide identification

ABRF iPRG 2012 Study on Modified Peptide Identification
Byonic outperformed Mascot, OMSSA, X!Tandem, ProteinPilot, SpectrumMill, Andromeda, … on every measure (# IDs, # consensus IDs, # spikes, # modifications, site localization, FDR, …)

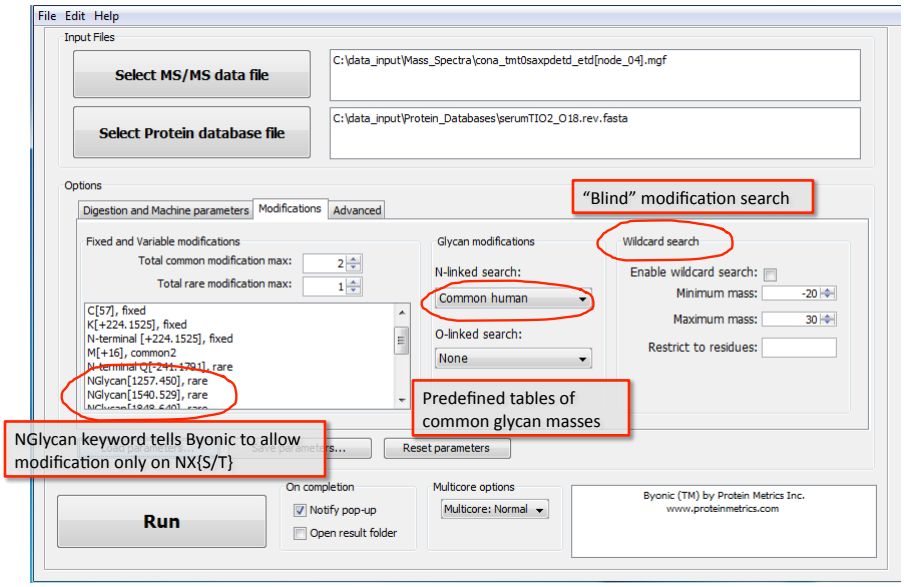


53

Byonic Input



Byonic Input



Product-Dependent ETD on Orbitrap

Rosa Viner et al,
Thermo Fisher

- HCD (or QTOF) gives mainly glycan fragments (e.g., 204.087 Da for HexNAc) ETD gives peptide fragments, but almost no glycan fragmentation
- **Idea:** Glycan peak in HCD scan triggers an ETD scan
- **Idea:** TMT^0 (Tandem mass tag) for adding charge and improving ETD spectra

Research Article

Increasing the Productivity of Glycopeptides Analysis by Using Higher-Energy Collision Dissociation-Accurate Mass-Product-Dependent Electron Transfer Dissociation

Julian Saba, Sucharita Dutta, Eric Hemenway, and Rosa Viner

Thermo Fisher Scientific, San Jose, CA 95134, USA

Correspondence should be addressed to Julian Saba, julian.saba@thermofisher.com

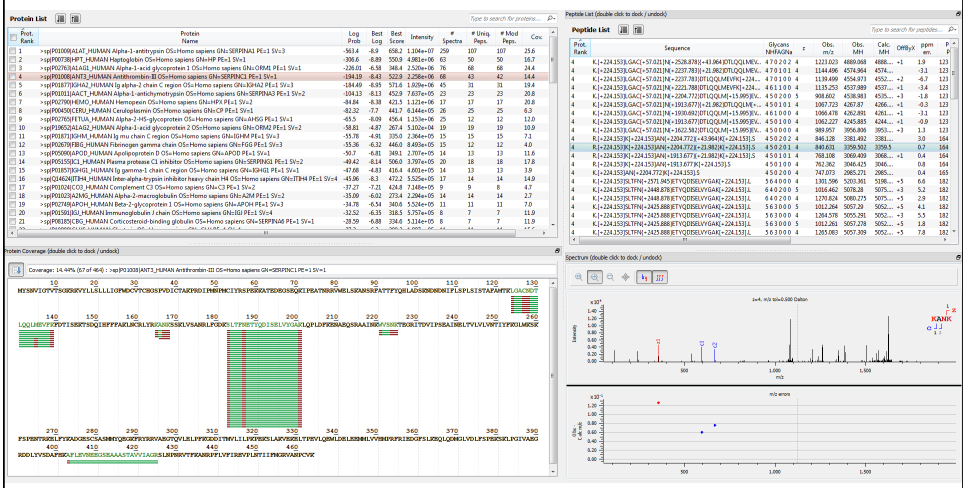
Received 6 February 2012; Accepted 15 March 2012

Academic Editor: Qiangwei Xia

Currently, glycans are attracting attention from the scientific community as potential biomarkers or as posttranslational modifications (PTMs) of therapeutic proteins. However, structural characterization of glycoproteins and glycopeptides remains analytically challenging. Here, we report on the implementation of a novel acquisition strategy termed higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation (HCD-PD-ETD) on a hybrid linear ion trap-orbitrap

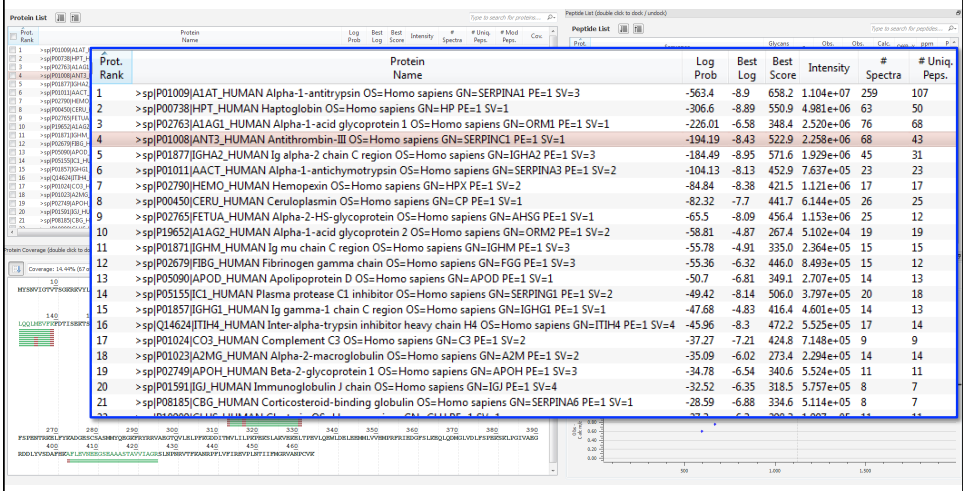
Glycoproteomics Example

- ConA-enriched blood serum, TMT0, SAX, HCD-pd-ETD on Orbitrap Velos
- Search allowed common modifications + ~400 predefined N-glycan masses



Byonic Glycoproteomics Search

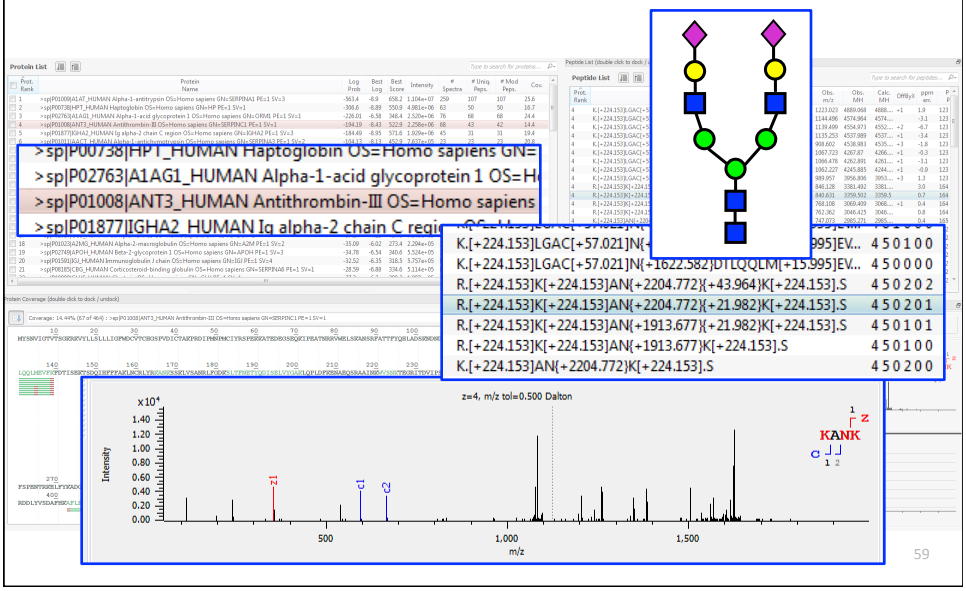
2535 ETD spectra ➔ ~70 glycoproteins, ~110 N-glyco sites, ~700 distinct glycopeptides



Prot. Rank	Protein Name	Log Prob	Best Log	Best Score	Intensity	# Spectra	# Uniq. Peps.
1	>sp|P01009|A1AT_HUMAN Alpha-1-antitrypsin OS=Homo sapiens GN=SERPINA1 PE=1 SV=3	-563.4	-8.9	658.2	1.104e+07	259	107
2	>sp|P00738|HPT_HUMAN Haptoglobin OS=Homo sapiens GN=HP PE=1 SV=1	-306.6	-8.89	550.9	4.981e+06	63	50
3	>sp|P02763|A1AG1_HUMAN Alpha-1-acid glycoprotein 1 OS=Homo sapiens GN=ORM1 PE=1 SV=1	-226.01	-6.58	348.4	2.520e+06	76	68
4	>sp|P01008|ANT3_HUMAN Antithrombin-III OS=Homo sapiens GN=SERPINC1 PE=1 SV=1	-194.19	-8.43	522.9	2.258e+06	68	43
5	>sp|P01877|IGHA2_HUMAN Ig alpha-2 chain C region OS=Homo sapiens GN=IGHA2 PE=1 SV=3	-184.49	-8.95	571.6	1.929e+06	45	31
6	>sp|P01011|AACT_HUMAN Alpha-1-antichymotrypsin OS=Homo sapiens GN=SERPINA3 PE=1 SV=2	-104.13	-8.13	452.9	7.637e+05	23	23
7	>sp|P02790|HEMO_HUMAN Hemopexin OS=Homo sapiens GN=HPX PE=1 SV=2	-84.84	-8.38	421.5	1.121e+06	17	17
8	>sp|P00450|CERU_HUMAN Ceruloplasmin OS=Homo sapiens GN=CP PE=1 SV=1	-82.32	-7.7	441.7	6.144e+05	26	25
9	>sp|P02765|FETUA_HUMAN Alpha-2-HS-glycoprotein OS=Homo sapiens GN=AHSG PE=1 SV=1	-65.5	-8.09	456.4	1.153e+06	25	12
10	>sp|P19652|A1AG2_HUMAN Alpha-1-acid glycoprotein 2 OS=Homo sapiens GN=ORM2 PE=1 SV=2	-58.81	-4.87	267.4	5.102e+04	19	19
11	>sp|P01871|IGHM_HUMAN Ig mu chain C region OS=Homo sapiens GN=IGHM PE=1 SV=3	-55.78	-4.91	335.0	2.364e+05	15	15
12	>sp|P02679|FIBG_HUMAN Fibrinogen gamma chain OS=Homo sapiens GN=FGG PE=1 SV=3	-55.36	-6.32	446.0	8.493e+05	15	12
13	>sp|P05090|APOD_HUMAN Apolipoprotein D OS=Homo sapiens GN=APOD PE=1 SV=1	-50.7	-6.81	349.1	2.707e+05	14	13
14	>sp|P05155|IC1_HUMAN Plasma protease C1 inhibitor OS=Homo sapiens GN=SERPING1 PE=1 SV=2	-49.42	-8.14	506.0	3.797e+05	20	18
15	>sp|P01857|IGHG1_HUMAN Ig gamma-1 chain C region OS=Homo sapiens GN=IGHG1 PE=1 SV=1	-47.68	-4.83	416.4	4.601e+05	14	13
16	>sp|Q14624|ITIH4_HUMAN Inter-alpha-trypsin inhibitor heavy chain H4 OS=Homo sapiens GN=ITIH4 PE=1 SV=4	-45.96	-8.3	472.2	5.525e+05	17	14
17	>sp|P01024|CO3_HUMAN Complement C3 OS=Homo sapiens GN=C3 PE=1 SV=2	-37.27	-7.21	424.8	7.148e+05	9	9
18	>sp|P01023|A2MG_HUMAN Alpha-2-macroglobulin OS=Homo sapiens GN=A2M PE=1 SV=2	-35.09	-6.02	273.4	2.294e+05	14	14
19	>sp|P02749|APOH_HUMAN Beta-2-glycoprotein 1 OS=Homo sapiens GN=APOH PE=1 SV=3	-34.78	-6.54	340.6	5.524e+05	11	11
20	>sp|P01591|IGJ_HUMAN Immunoglobulin J chain OS=Homo sapiens GN=IGJ PE=1 SV=4	-32.52	-6.35	318.5	5.757e+05	8	7
21	>sp|P08185|CBG_HUMAN Corticosteroid-binding globulin OS=Homo sapiens GN=SERPINA6 PE=1 SV=1	-28.59	-6.88	334.6	5.114e+05	8	7

Byonic Glycoproteomics Search

2535 ETD spectra ➔ ~70 glycoproteins, ~110 N-glyco sites, ~700 distinct glycopeptides

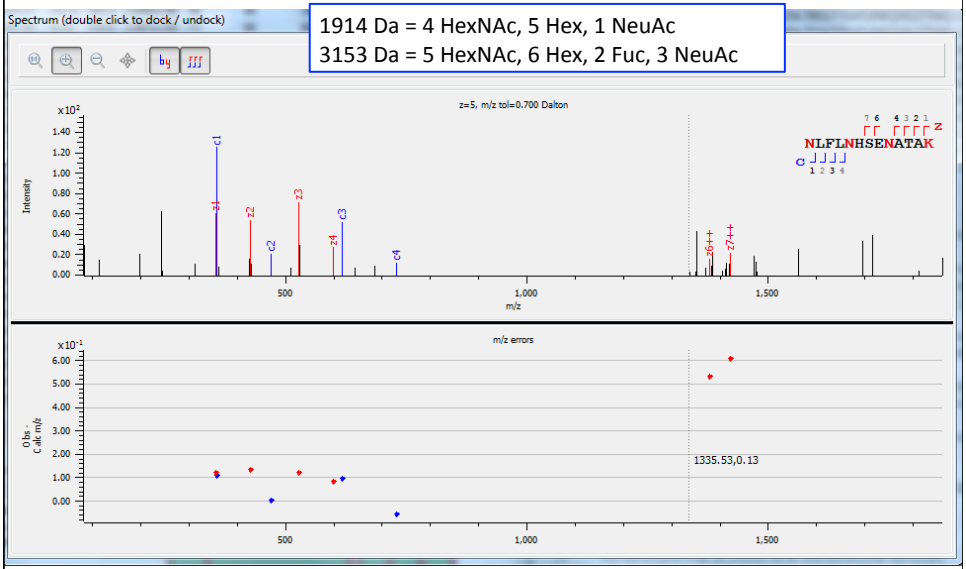


59

ETD spectrum of a peptide with two N-glycans

N[+224]LFN[+3153]HSEN[+1914]ATAK[+224]$^{5+}$ haptoglobin

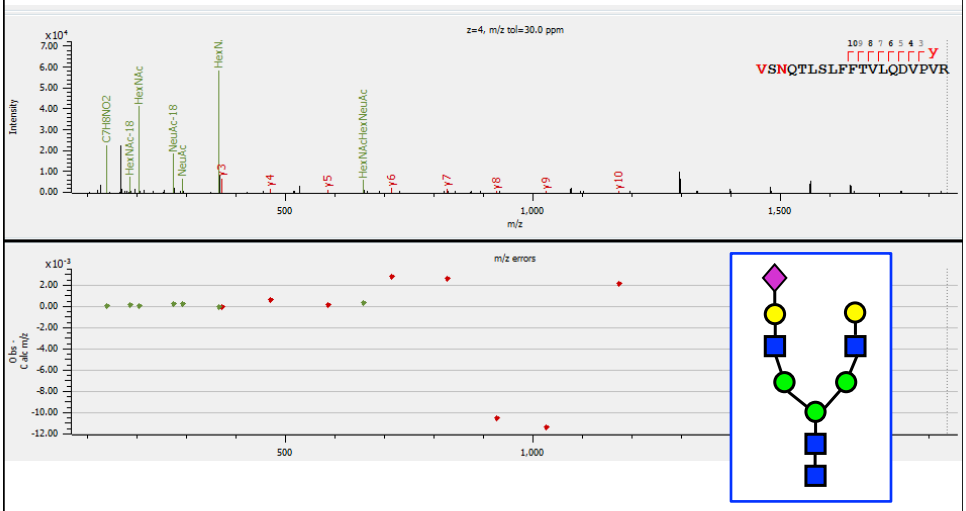
1914 Da = 4 HexNAc, 5 Hex, 1 NeuAc
3153 Da = 5 HexNAc, 6 Hex, 2 Fuc, 3 NeuAc



ABRF Course 60

HCD Spectrum

[+224]VSN[+1914]QTLSLFFTVLQDVPVR^{4+}



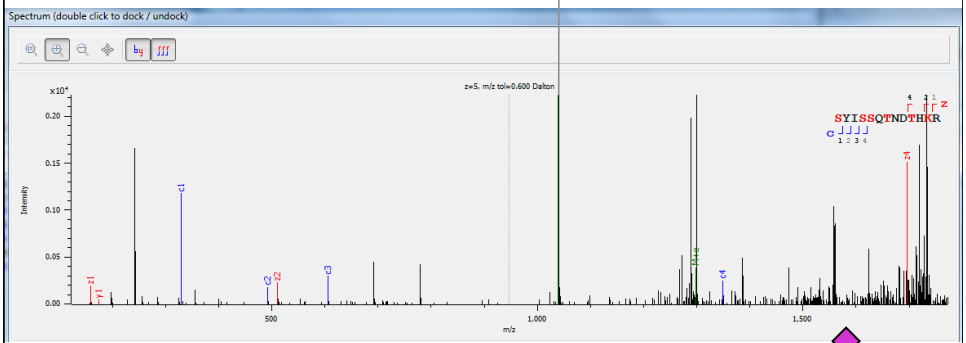
ETD spectrum of a peptide with four O-Glycans

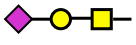
Glycophorin-A, TMT0 labeled, HCD-pd-ETD, Orbitrap Elite
Chia-Wei Lin, KH Khoo, *Academia Sinica, Taiwan*

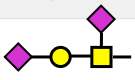
Note: O-Glycans fly off in HCD, so ETD is needed for site localization

No peaks in between

[+224]SYIS[+656]S[+656]QT[+947]NDT[+947]HK[+224]R^{5+}



656 Da 

947 Da 

Summary

1) Background – Glycomics vs. Proteomics bioinformatics

   Glycomics is much less developed, especially O-glycosylation

2) Automatic detached N-glycan analysis from MS1 – Cartoonist

   Strength: "Thinks" like a glycomics expert Weakness: Little support for MS2

3) Semi-automatic glycan analysis from MS1 and MS2 – GlycoWorkbench

   Strength: Complete package Weakness: Not high-throughput

4) Automatic glycan analysis from MS2 – SimGlycan

   Strength: Commercial support Weakness: Requires MS2

5) Glycopeptide analysis from MS2 – Byonic

   Strength: All-round good proteomics tool
   Only tool for glycoproteomics

   Weakness: Glycoproteomics sees only the most abundant species

Acknowledgments

- Grant support for development of Cartoonist: NIH R01 GM085718
- Grant for development of Byonic: NIH R21 GM094557, NSF Computing Innovations Fellowship
- Slides and help with SimGlycan: Sanjib Meitei, Arun Apte, Premier Biosoft
- Current Cartoonist team: Doron Kletter, Alex Brito, Mudita Singhal, PARC

References

- Cartoonist: Goldberg D, Sutton-Smith M, Paulson J, Dell A. Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra. Proteomics. 2005; 5:865–75.
- GlycoWorkbench: Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. J Proteome Res. 2008; 7:1650–9.
- SimGlycan: Apte A, Meitei NS, Bioinformatics in Glycomics: Glycan Characterization with Mass Spectrometry Data using SimGlycan in *Functional Glycomics: Methods and Protocols*, Editor: Jianjun Li, Springer Protocols
- GlycomeDB: Ranzinger R, Herget S, von der Leith C-W, Frank M. GlycomeDB– a unified database for for carbohydrate structures, Nucleic Acids Research, Jan 2011, D373-D376.
- Byonic: Bern M, Cai Y, Goldberg D. Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. Anal. Chem. 2007; 79(4), 1393-1400.