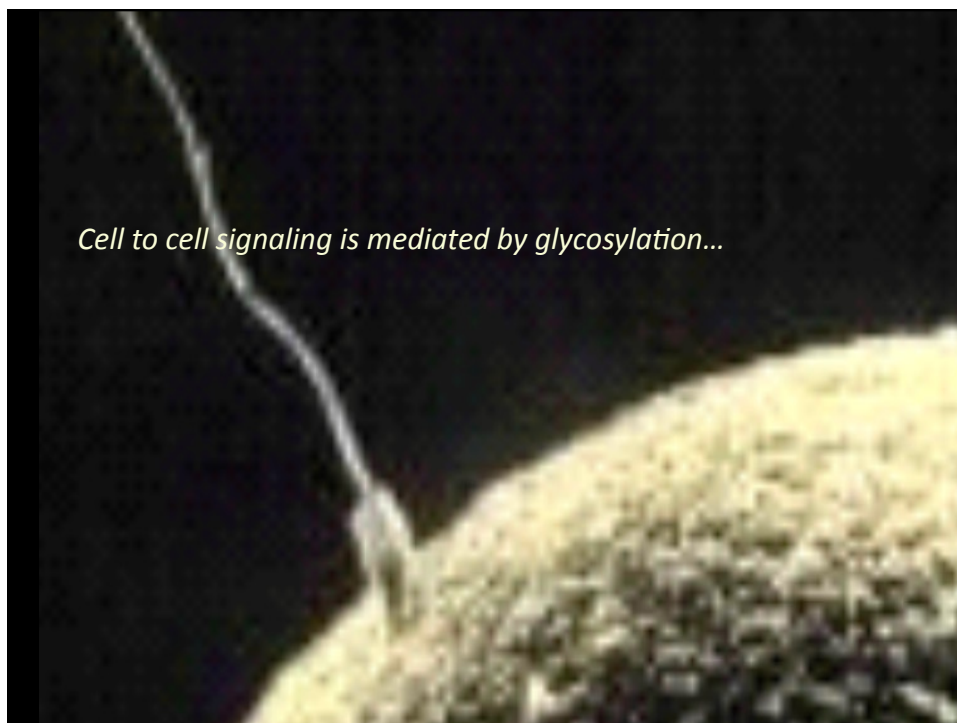


Bioinformatics of Glycans and Glycoproteins

Marshall Bern
Palo Alto Research Center
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

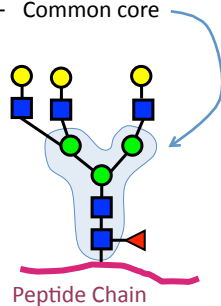
Outline

- 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

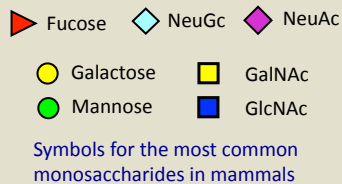
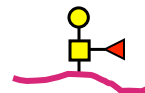
N-linked Glycosylation

- 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



ABRF Course

5

Proteomics vs. Glycomics

Peptides

1. 19 different residue masses
Total peptide mass uninformative
2. Linear sequence
3. Only one type of peptide bond
4. Any sequence is valid
5. Good databases exist

Glycans

1. ~5 monosaccharide masses
Total mass informative
2. Branched structure
3. Various types of linkages
4. Structures follow certain patterns
5. Databases have poor coverage

= Easier

= Harder

ABRF Course

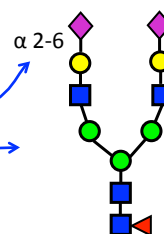
6

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

ABRF Course 7

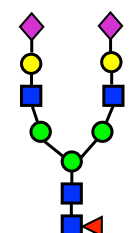
Proteomics vs. Glycomics

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:

LGLLVFPYTHQNWVEVQYSR
EVTLLHLLPGEQLLC(+57)EASTVLK
LQEEMLQREEAENTLQSFR
etc.

2351 Daltons
(not permethylated)



ABRF Course 8

Proteomics vs. Glycomics

Peptides

- 19 different residue masses
Total peptide mass uninformative
- Linear sequence
- Only one type of peptide bond
- Any sequence is valid
- Good databases exist

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

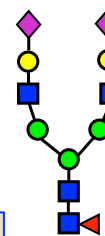
LGLLVFPYTHQNWEVQYSR
EVTLLPLPGEQLLC(+57)EASTVLK
LQEMLQREEAENTLQSFRR
etc.

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

Glycans

- ~5 monosaccharide masses
Total mass informative
- Branched structure
- Various types of linkages
- Structures follow certain patterns
- Databases have poor coverage

2351 Daltons
(not permethylated)



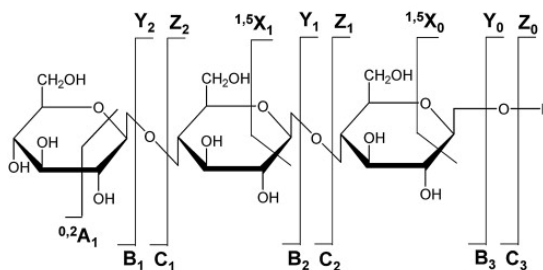
ABRF Course

9

How do glycans fragment in MS2 ?

Domon-Costello nomenclature

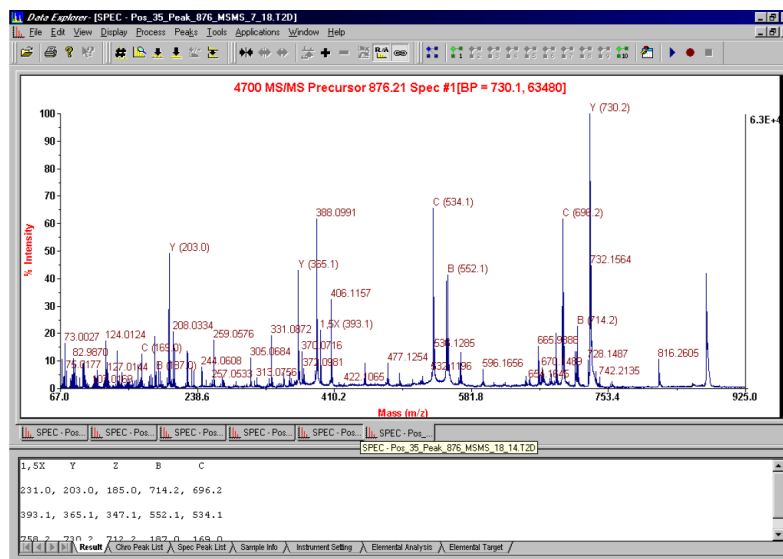
- Draw the glycan with reducing end to the right
- Cut before \vdash O – left half = B-ion, right half = Y-ion
- Cut after $-$ O \vdash left half = C-ion, right half = Z-ion
- Internal fragments can be YY, ZZ, etc.
- Cross-ring = X 1,5 X is most common



ABRF Course

10

How do glycans fragment in MS2 ?



From AB Sciex brochure

ABRF Course

11

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

ABRF Course

12

Cartoonist Software

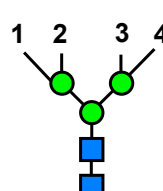
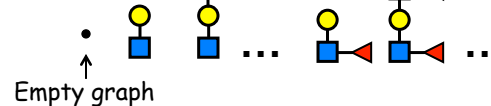
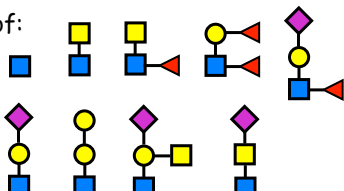


- Developed at Xerox PARC, mostly by [David Goldberg](#) (currently Ebay fellow)
- Glycomics guidance from [Anne Dell](#), [Stuart Haslam](#) at Imperial College
- Platform-independent Java software, [Free!](#), bundled with PARC mass spectrum browser
- Works well for MALDI-TOF profiles of detached, 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):
 
3. Cap with one of:
 
4. Optionally add:
 - Bisecting GlcNAc,
 - Base fucose,
 - NeuAc → NeuGc

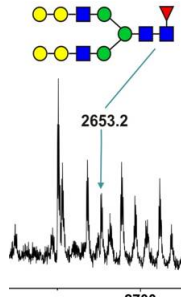
ABRF Course 15

Example – 2653 Daltons

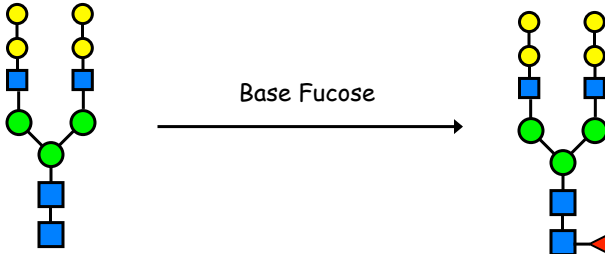
No lactosamines: • 

GlcNAc-Gal-Gal caps: 

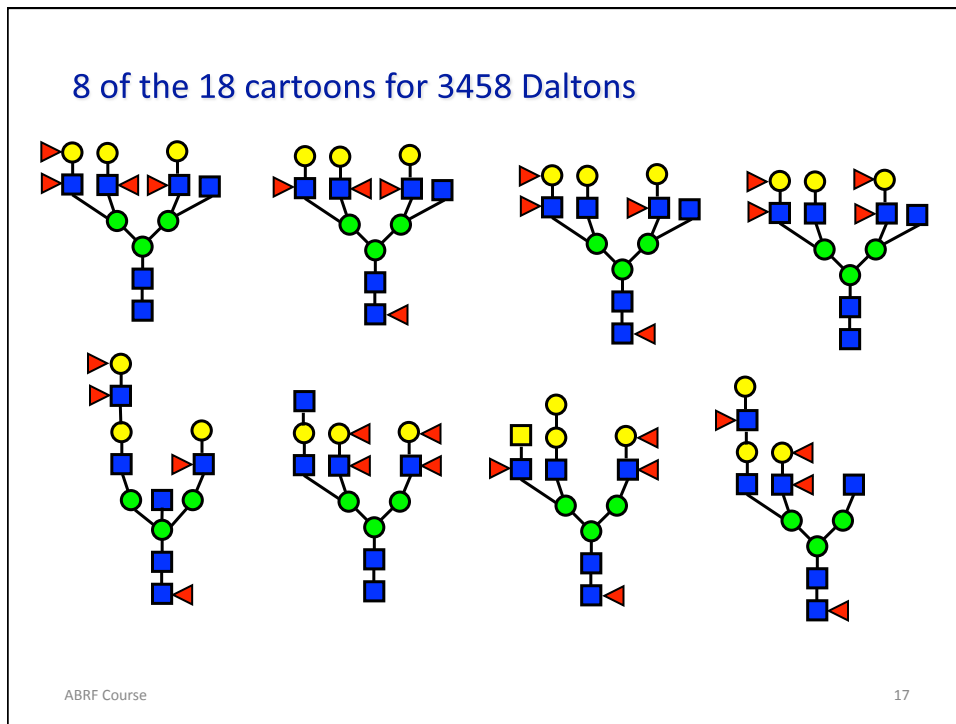
Core



Base Fucose



ABRF Course 16



Organism-Specific Hints (Penalties)

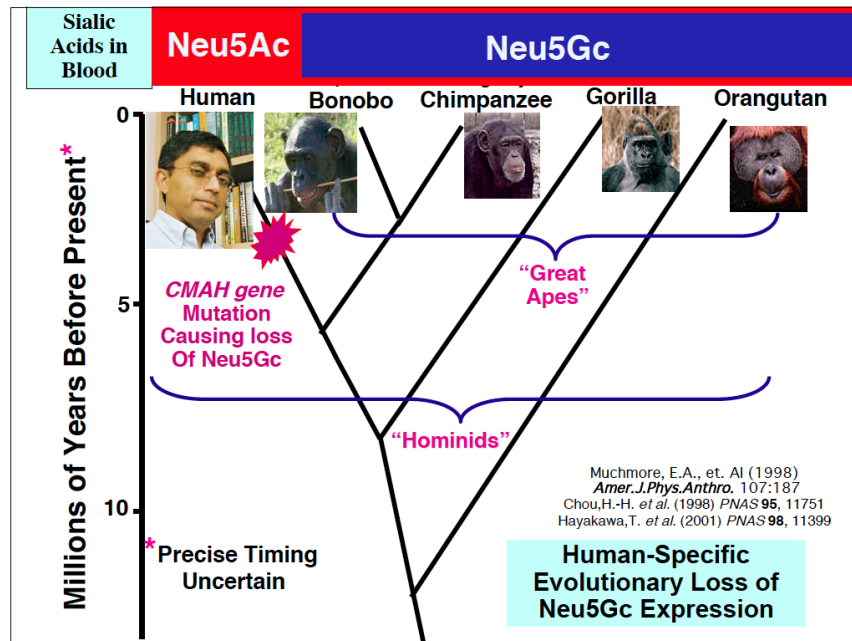
- Unlikely patterns for all mammals:
 - Bisecting GlcNAc
 - Uncapped terminal GlcNAc
 - Fucose on antennae, but not on base
- Specific to Mouse:
 - LacdiNAc, Sialyl Lewis-X, Multiple-fucose antenna
- Specific to Human:
 - Gal-alpha-Gal cap
 - NeuGc (Is this what makes us human??)

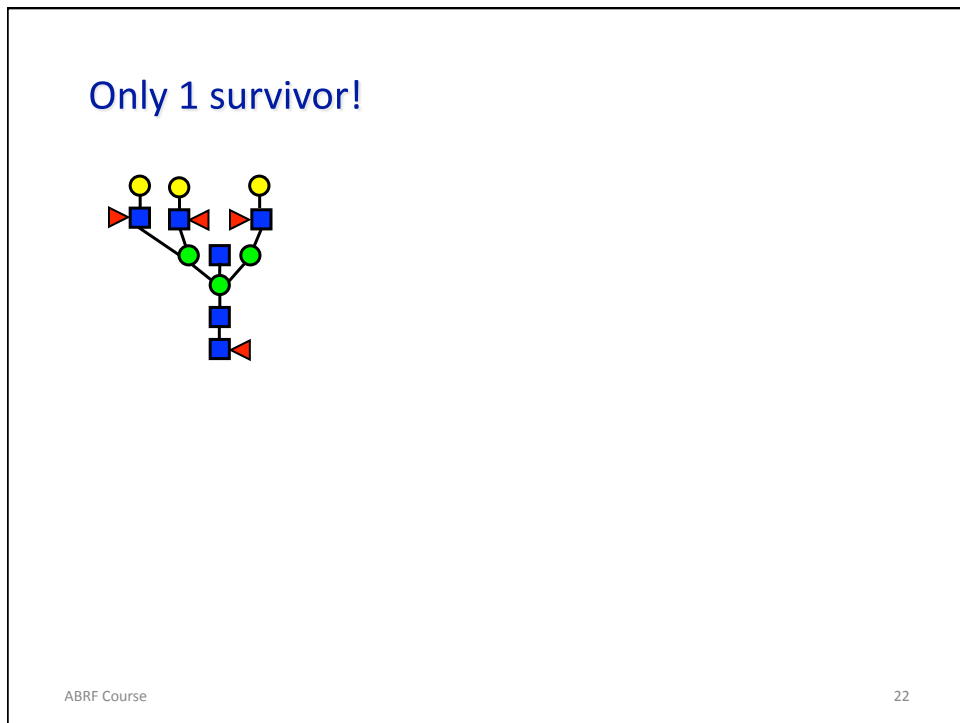
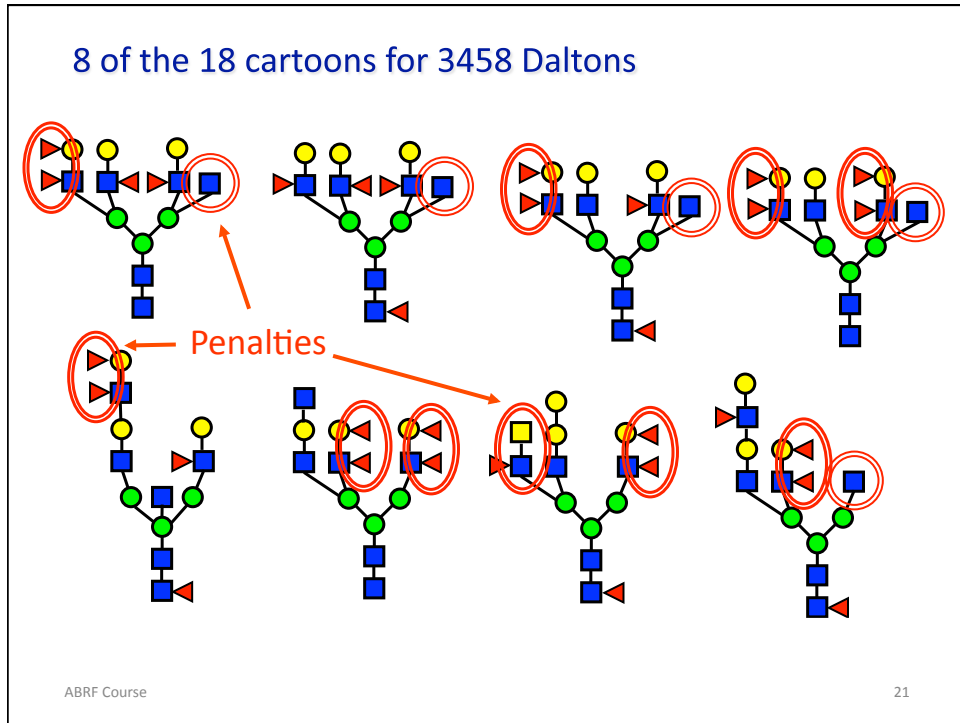


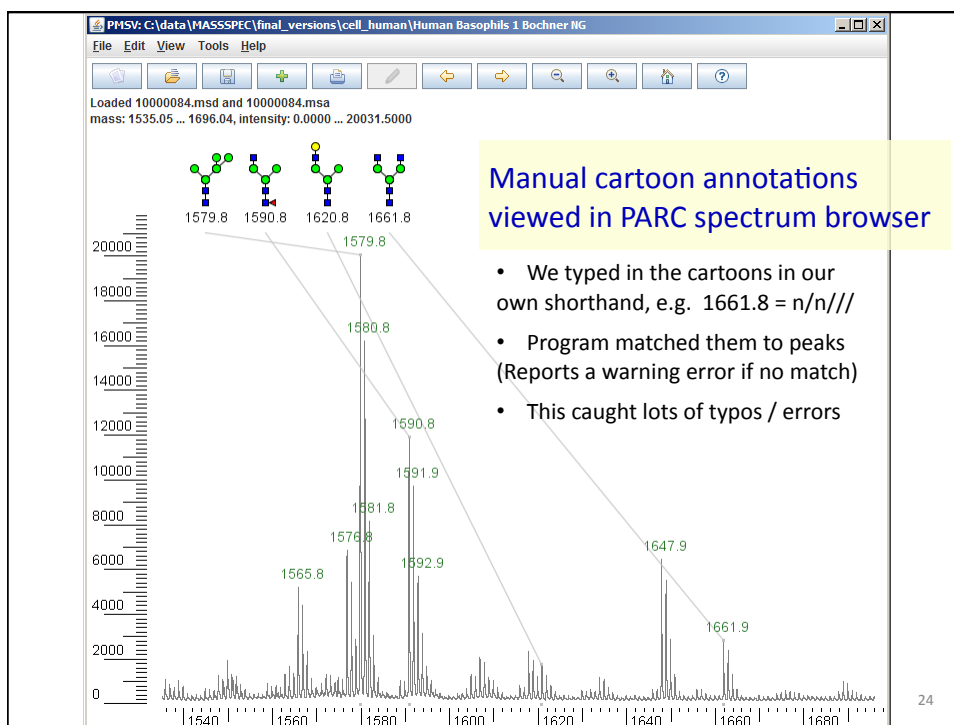
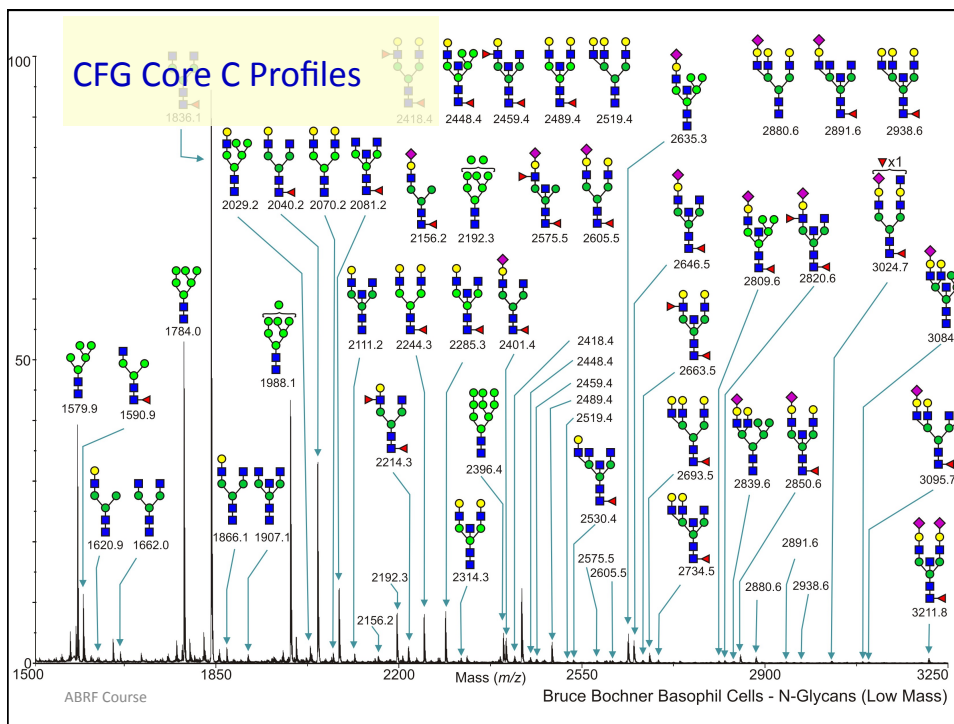
Organism-Specific Hints (Penalties)

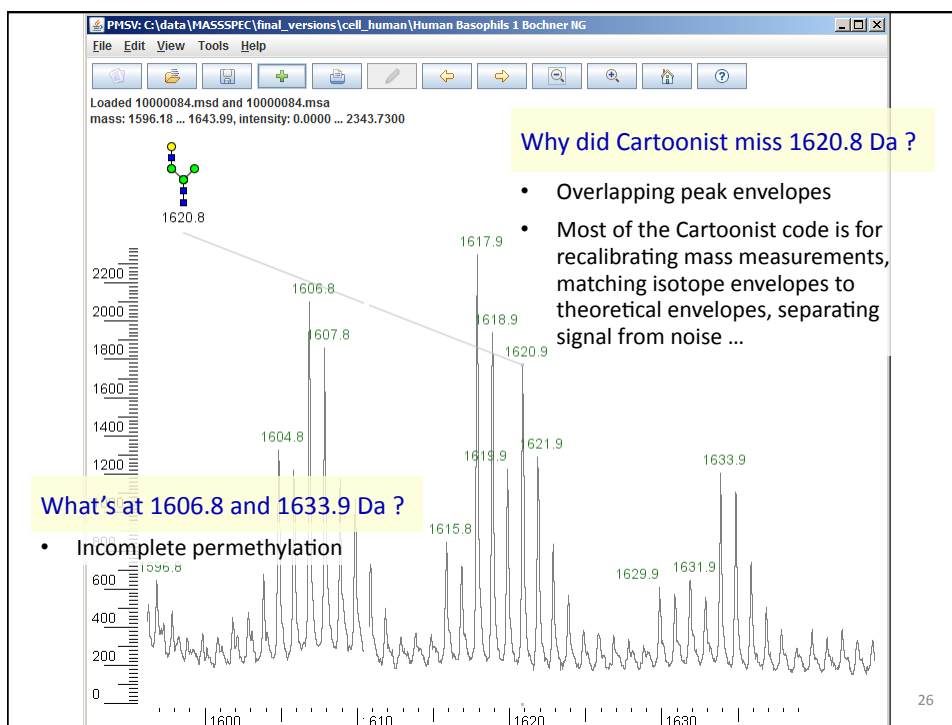
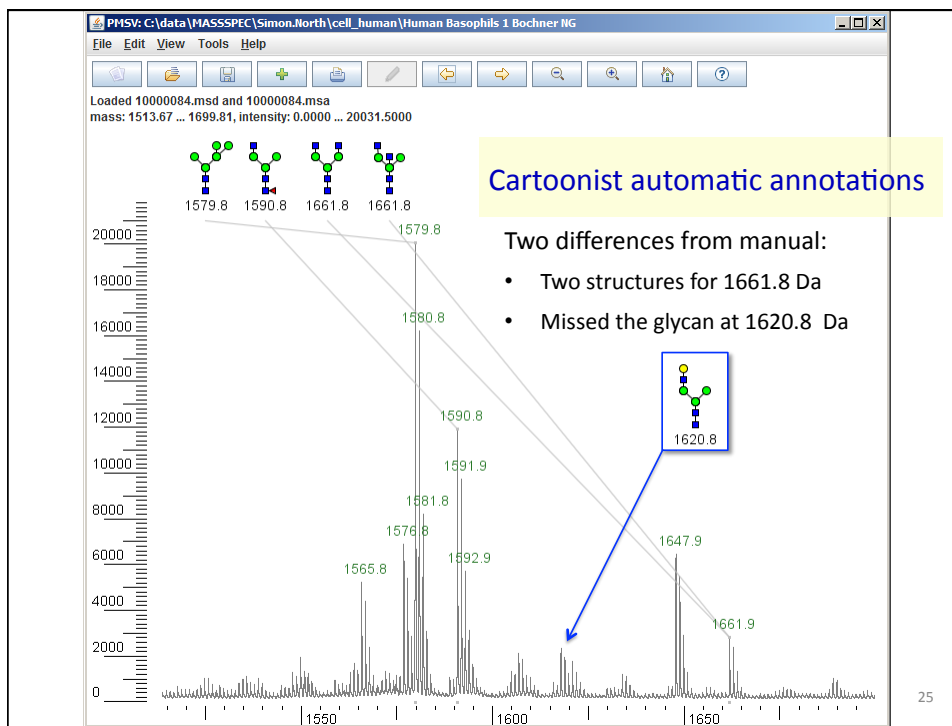
- Unlikely patterns for all mammals:
 - Bisecting GlcNAc
 - Uncapped terminal GlcNAc
 - Fucose on antennae, but
- Specific to Mous:
 - LacdiNAc, Sialyl
- Specific to Human:
 - Gal-alpha-Gal cap
 - NeuGc (Is this what makes us human??)

```
# 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
```



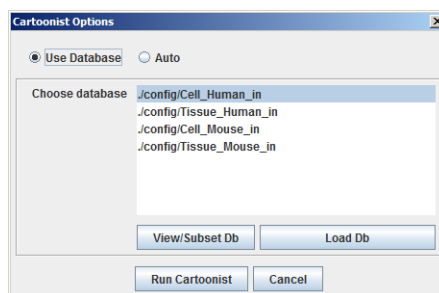
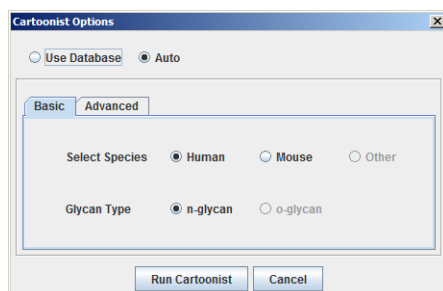






How to run Cartoonist?

- 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, ...**



ABRF Course

27

Using / creating databases

Select	My Code	Linear Code	Glycan Structure	Link
<input type="checkbox"/>	man5	M??(M??(M??)M??)M??GN??GN		See Cartoon carbNlink_35518_A
<input type="checkbox"/>	man6	M??M??(M??(M??)M??)M??GN??GN		See Cartoon carbNlink_35521_A
<input type="checkbox"/>	/n/H3//	GN??M??(M??(M??)M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/n/n//f	GN??M??GN??M??(M??)M??GN??GN		See Ca...
<input type="checkbox"/>	man7	M??M??(M??(M??)M??)M??GN??GN,M		See Ca...
<input type="checkbox"/>	/ng/H3//	A??GN??M??(M??(M??)M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/ng/n//f	A??GN??M??(A??GN??M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/ng/ng//	A??GN??M??(A??GN??M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/n/n//bf	GN??(A??GN??M??)M??(M??)M??GN??GN		See Ca...
<input type="checkbox"/>	man8	M??M??(M??(M??)M??)M??GN??GN,M,M		See Ca...
<input type="checkbox"/>	/ngt/m//	NJ??A??GN??M??(M??(M??)M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/ng/ng//f	A??GN??M??(A??GN??M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/ng/n//bf	GN??(A??GN??M??)M??(M??)M??GN??GN		See Ca...
<input type="checkbox"/>	man9	M??M??(M??(M??)M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/ngt/H3//	NJ??A??GN??M??(M??(M??)M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/ngs/ng//	NN??A??GN??M??(A??GN??M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/ng/H3//bf	GN??(A??GN??M??)M??(M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/ngt/ng//	NJ??A??GN??M??(A??GN??M??)M??GN??GN		See Ca...
<input type="checkbox"/>	/ng/ng/H3//	A??GN??(A??GN??)M??(M??(M??)M??)M??GN??GN		See Cartoon carbNlink_35519_A
<input type="checkbox"/>	/ng/ng//bf	GN??(A??GN??M??)M??(A??GN??M??)M??GN??GN		See Cartoon carbNlink_35456_A
<input type="checkbox"/>	man10	G??M??M??(M??(M??)M??)M??GN??GN		See Cartoon carbNlink_35528_A
<input type="checkbox"/>	/ngs/ng//f	NN??A??GN??M??(A??GN??M??)M??GN??GN		See Cartoon carbNlink_35474_A
<input type="checkbox"/>	/ngt/ng//f	NJ??A??GN??M??(A??GN??M??)M??GN??GN		See Cartoon carbNlink_35473_A
<input type="checkbox"/>	/ngs/ngg//	NN??A??GN??M??(A??GN??M??)M??GN??GN		See Cartoon carbNlink_35571_A
<input type="checkbox"/>	/ngg/ngg//f	A??A??GN??M??(A??A??GN??M??)M??GN??GN		See Cartoon carbNlink_35489_A
<input type="checkbox"/>	/ngt/ngg//	NJ??A??GN??M??(A??A??GN??M??)M??GN??GN		See Cartoon carbNlink_35570_A

ABRF Course

28

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

The screenshot displays the GlycoWorkbench software interface. At the top, there is a menu bar with options like Home, Edit, View, Structure, and Tools. Below the menu is a toolbar with various icons for adding and editing glycan components. The main workspace contains three glycan structures, each with its corresponding m/z value: m/z: 2809.3979, m/z: 2966.4718, and m/z: 3054.5242. To the right of the workspace is a 'PeakList' table with columns for 'Mass to charge', 'Intensity', 'Relative intensity', and 'Ion'. Below the workspace is a mass spectrum plot titled 'Spectrum 1/1, MS' showing relative intensity versus m/z ratio. The plot shows several peaks, with the most prominent ones corresponding to the m/z values of the structures in the workspace.

ABRF Course 31

GlycoWorkbench

This screenshot is similar to the one above but includes three blue callout boxes with white text. The first box, labeled 'Database of known structures', points to the 'PeakList' table on the right side of the interface. The second box, labeled 'Canvas for drawing glycan structures', points to the central workspace area where the glycan structures are displayed. The third box, labeled 'Spectrum (MS1)', points to the mass spectrum plot at the bottom of the interface.

ABRF Course 32

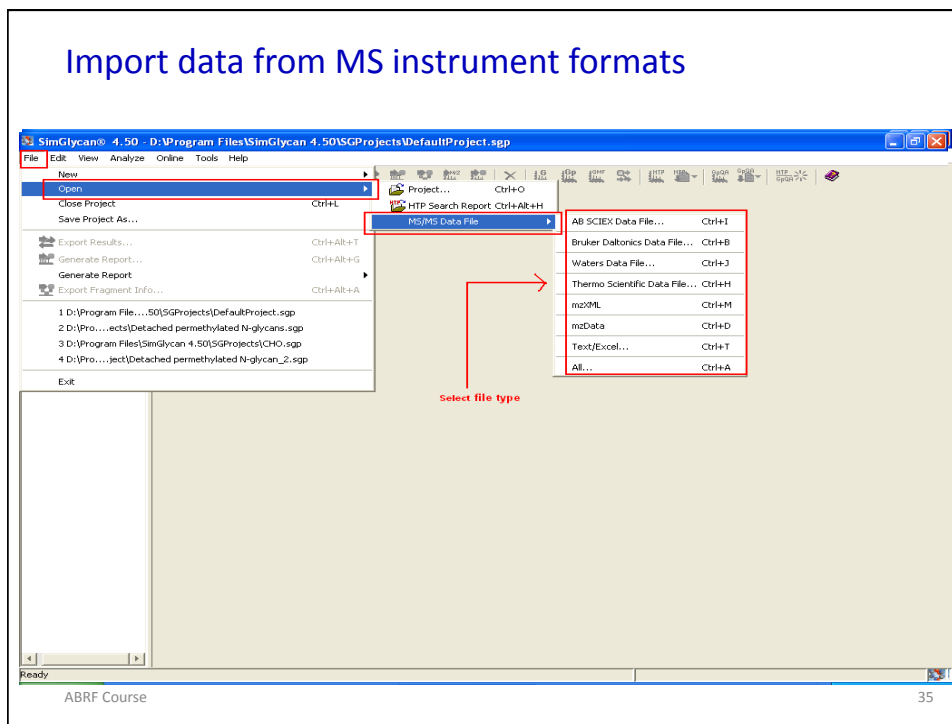
Outline

- 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

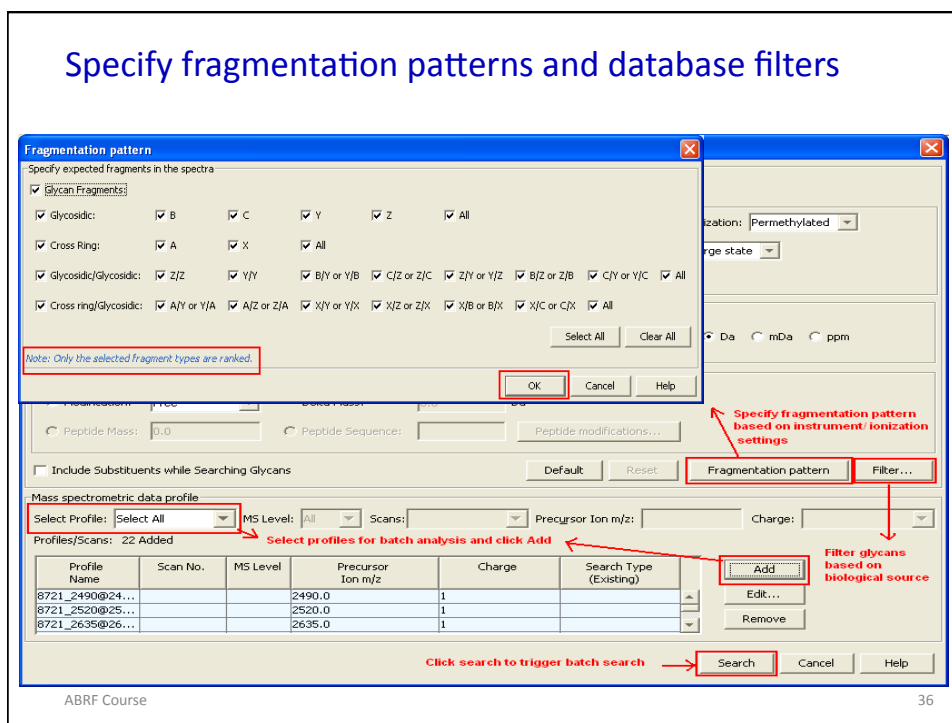
SimGlycan

- 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

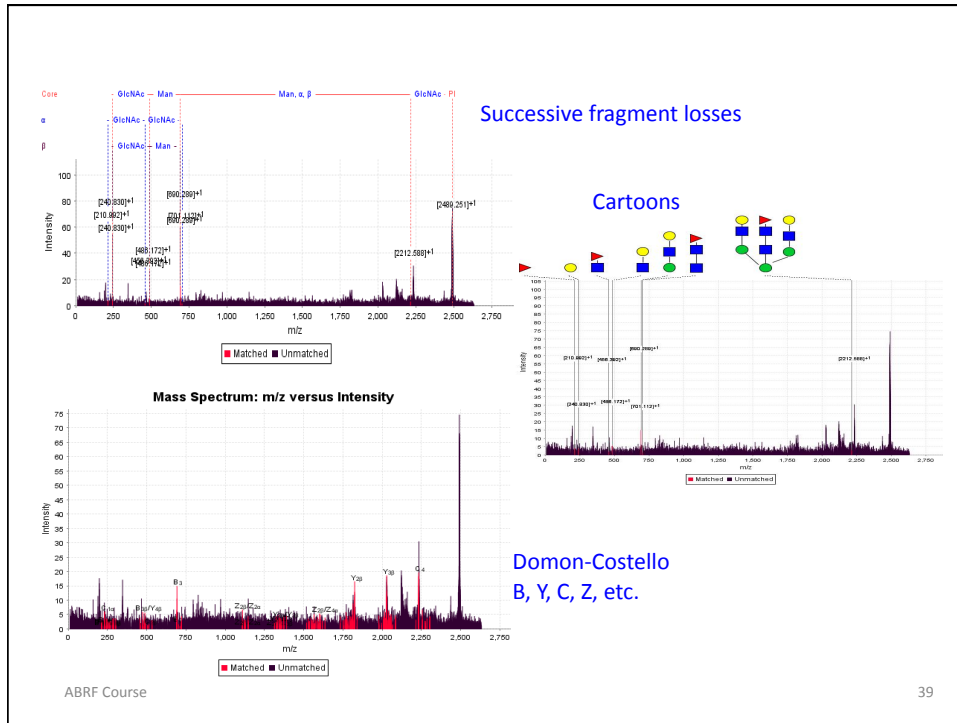
The screenshot displays the SimGlycan 4.50 interface. The main window shows a glycan structure (G04570) with a sequence: Gal (b1-4) GlcNAc (b1-2) Man (a1-3) [Fuc (a1-6) GlcNAc (b1-4) GlcNAc (b1-4)] [Gal (b1-4) GlcNAc (b1-2) Man (a1-6)] Man (b1-4) GlcNAc. The structure is a branched N-glycan with Galactose, GlcNAc, and Mannose residues. A table on the right lists annotated spectra with their proximity scores. The top entry is 1-4)GlcNAc(b1-4)] [Gal(b1-4... with a score of 84.039. Below the structure is a table with the following data:

Name	Value
Glycan Sequence	Gal(b1-4)GlcNAc(b1-2)Man(a1-3)[Fuc(a1-6)GlcNAc(b1-4)GlcNAc(b1-4)] [Gal(b1-4)GlcNAc(b1-2)Man(a1-6)]Man(b1-4)GlcNAc
Composition	(G)8)2 (F)0)1 (GlcNAc)5 (Man)3
Chemical Formula (Underivatized)	C ₇₂ H ₁₂₂ O ₄₂ N ₆
Glycan Mass	2466.26155501
Carbohydrate Mass	2466.26155501
Precursor m/z (Charge State)	2489.251323201 (1)
Class	Glycoprotein/N-Glycan, Epitope, Glycolipid, Glycosphingolipid
Reaction	

The bottom status bar shows 'Ready' and 'Total no. of hits : 13'.

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

The screenshot shows the 'Plot Spectrum' dialog box in SimGlycan 4.50. The 'Annotation type' section has three radio buttons: 'Successive Fragment Loss', 'Symbolic Representation', and 'Domon Costello Nomenclature'. The 'Domon Costello Nomenclature' option is selected. The 'Show peaks of selected vs experimental spectrum' section has four radio buttons: 'Matched', 'Matched and Unmatched', 'Selected', and 'All'. The 'Matched' option is selected. The 'Show Peptide Fragments' section has two checkboxes: 'Loss of Water Molecules' and 'Loss of Ammonia Molecules'. The 'Show Glycan Fragments' section has several checkboxes for glycan fragment types: Glycosidic (B, C, Y, Z, All), Cross Ring (A, X, All), Glycosidic/Glycosidic (Z/Z, Y/Y, B/Y or Y/B, C/Z or Z/C, Z/Y or Y/Z, B/Z or Z/B, C/Y or Y/C, All), and Cross ring/Glycosidic (A/Y or Y/A, A/Z or Z/A, X/Y or Y/X, X/Z or Z/X, X/B or B/X, X/C or C/X, All). The 'Color picker' section has two color swatches: 'Matched' (red) and 'Un-Matched' (black). The 'Note' at the bottom says: 'To annotate a peak, drag cursor over it.' The 'OK' button is highlighted with a red box. The status bar at the bottom shows 'Ready' and 'Theoretical Fragments: % Cross ring Match: Single Cross ring Match: 100 Cross ring/Glycosidic Match: 99.16'.



Customize Report: Filter, Sort, Present Results

Specify G and Gp Report Contents

Filtering options

- Threshold Precursor Intensity: \geq 0.0
- Threshold percent (%) match for including the results in the report
 - % Glycosidic Match
 - Single Glycosidic Match: \geq 50.0 %
 - Glycosidic/Glycosidic Match: \geq 50.0 %
 - % Cross ring Match
 - Single Cross ring Match: \geq 0.0 %
 - Cross ring/Glycosidic Match: \geq 0.0 %
 - Threshold Proximity Score: \geq 75.0

Organize Report

- Remove scans/precursor m/z values with no result
- Sort Results
 - Sort By: Precursor m/z
 - Then By: MS Level
 - Order: Ascending
- Normalize abundance of analytes based on an internal standard
 - Internal Standard: Abundance: Amount: $\mu\text{g/ml}$
 - Internal Standard Intensity Normalized Intensity Relative Quantity

ABRF Course 40

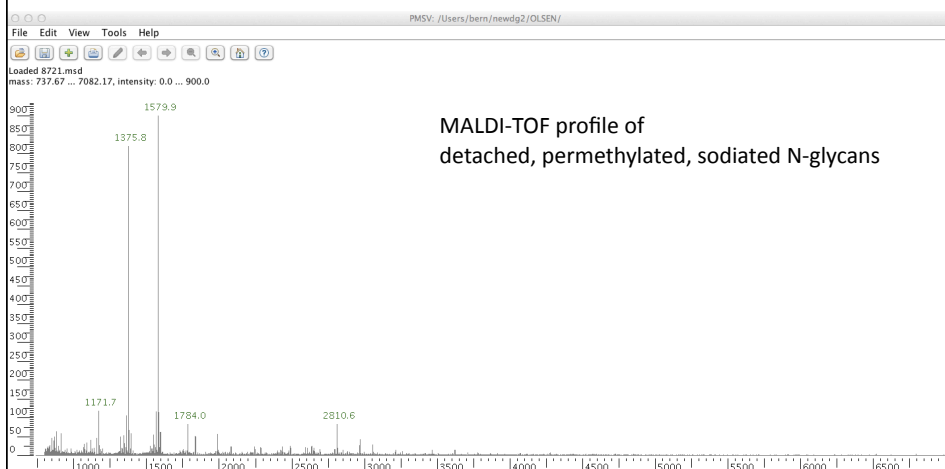
Outline

- 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

ABRF Course

41

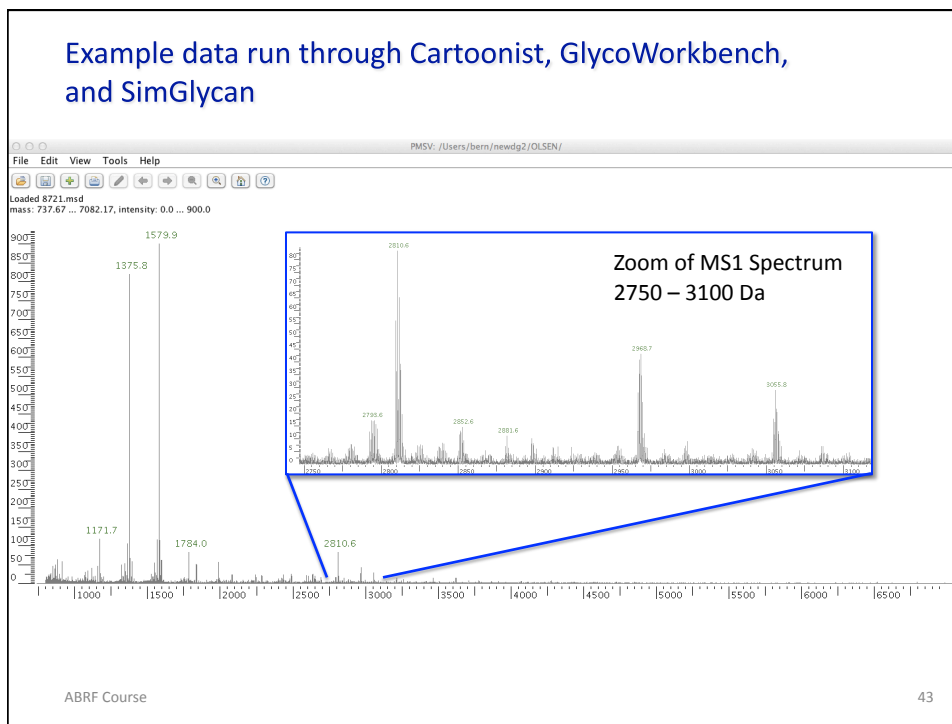
Example data run through Cartoonist, GlycoWorkbench, and SimGlycan



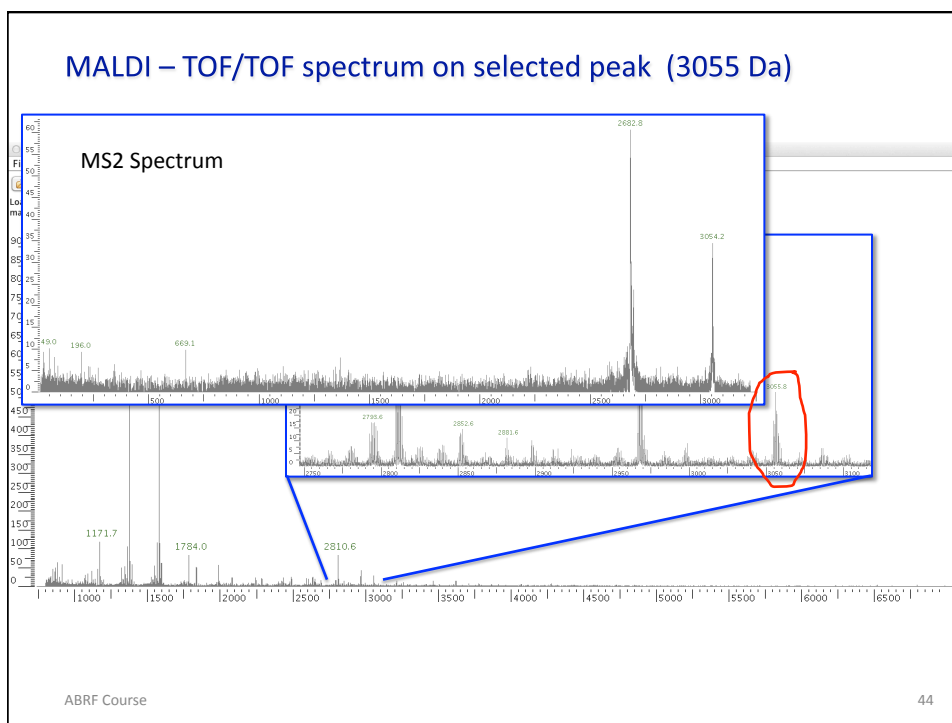
ABRF Course

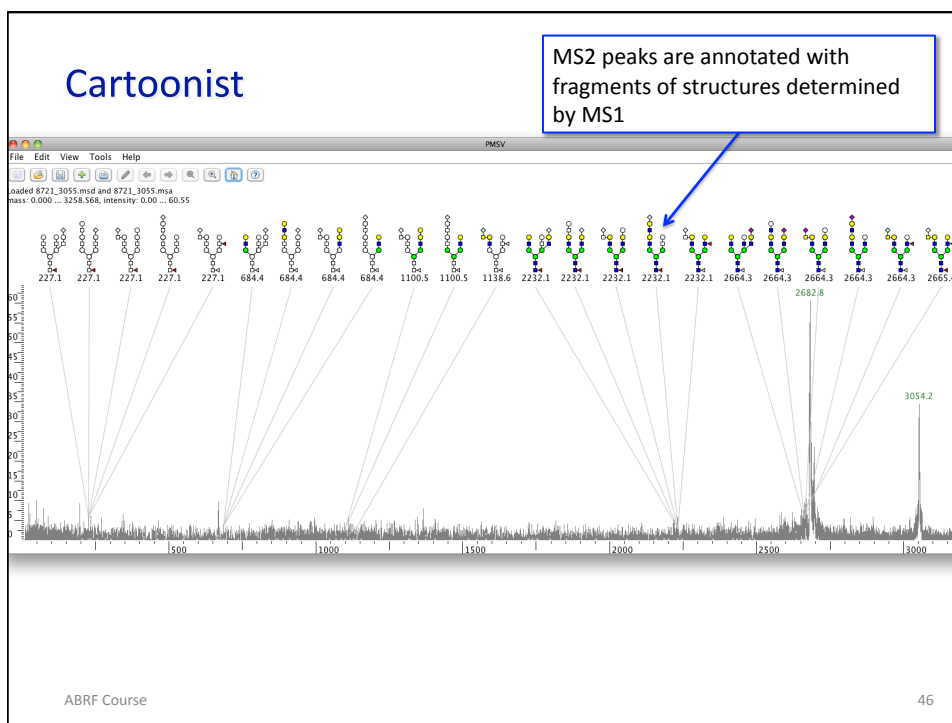
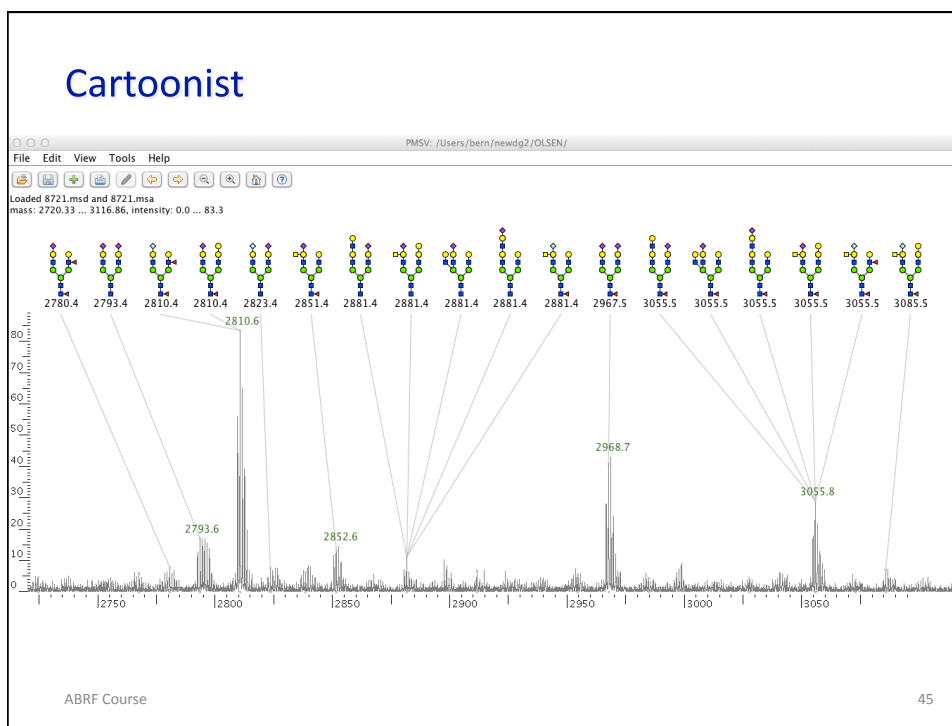
42

Example data run through Cartoonist, GlycoWorkbench, and SimGlycan



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





GlycoWorkbench

The screenshot displays the GlycoWorkbench interface. At the top, there is a menu bar with options like Home, Edit, View, Structure, and Tools. Below the menu is a toolbar with various icons for adding and editing glycan components. The main workspace is divided into several panels:

- Structure Editor:** Shows a central glycan structure with various colored nodes (yellow, blue, green, red) representing different sugar residues.
- Structure List:** A table on the right side lists several glycan structures with their corresponding mass-to-charge ratios (m/z) and relative intensities. The first structure is highlighted in yellow.
- Mass Spectrum:** A plot at the bottom left shows the mass spectrum for the selected structure, with peaks labeled at m/z 2793.6, 2822.6, 2988.7, and 3055.8.

A callout box with a blue border and arrow points to the structure list, containing the text: "Gives lots of annotations - User chooses structure".

ABRF Course 47

GlycoWorkbench

This screenshot is similar to the one on slide 47, showing the GlycoWorkbench interface. The main difference is the zoomed-in view of the mass spectrum plot at the bottom left. The plot shows several peaks, with the most prominent ones labeled at m/z 2793.6, 2822.6, 2988.7, and 3055.8. The structure list on the right is also visible, with the first structure highlighted in yellow. A callout box with a blue border and arrow points to the structure list, containing the text: "Gives lots of annotations - User chooses structure".

ABRF Course 48

GlycoWorkbench

After the user chooses structure, the software will match MS2 peaks to fragments

The screenshot shows the GlycoWorkbench interface. At the top, there's a menu bar with options like Home, Edit, View, Structure, and Tools. Below the menu is a toolbar. The main workspace contains a glycan structure diagram with various colored nodes (yellow, green, blue, red) and linkages. Below the structure is a mass spectrum plot titled "Spectrum 1/1, MS" showing intensity versus m/z. To the right, there's a table of MS2 fragments with columns for Fragment, Type, m/z, ions, Neutral Exchanges, and Fragment Mass. A blue arrow points from the text box to the table.

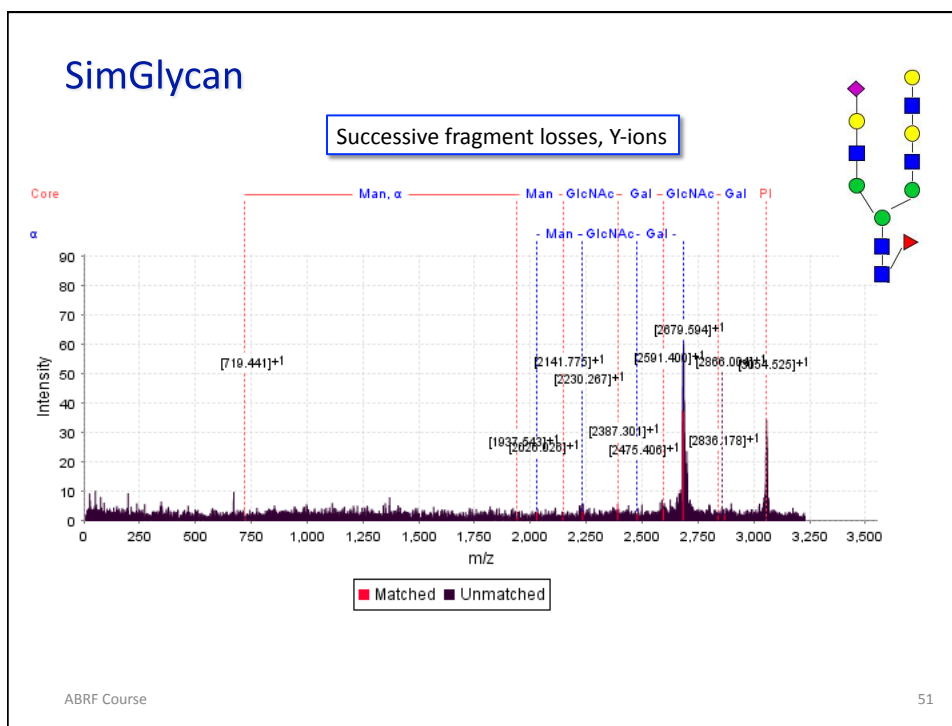
Fragment	Type	m/z	ions	Neutral Exchanges	Fragment Mass
[Glycan structure]	YY	3648.3039	Na	0	2625.3147
[Glycan structure]	Z	2661.3243	Na	0	2628.3351
[Glycan structure]	Y	2670.3349	Na	0	2626.3457
[Glycan structure]	Z	2652.2862	Na	0	2705.4098
[Glycan structure]	Y	2638.4088	Na	0	2613.4288
[Glycan structure]	Z	2648.4088	Na	0	2625.4288
[Glycan structure]	Y	2660.4284	Na	0	2643.4301
[Glycan structure]		3554.5342	Na	0	3021.5330

ABRF Course 49

SimGlycan

The screenshot shows the SimGlycan interface. At the top, there's a library of glycan structures represented by colored nodes and linkages. Below the library is a mass spectrum plot showing intensity versus m/z. A legend at the bottom indicates "Matched" (red square) and "Unmatched" (black square). A blue box highlights a specific glycan structure, labeled "Top-scoring structure".

ABRF Course Matched ■ Unmatched



Outline

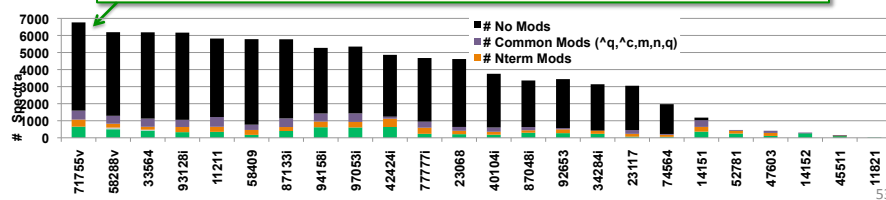
- 1) Background – Glycomics vs. Proteomics bioinformatics
- 2) Automatic detached N-glycan analysis from MS1 – Cartoonist
- 3) Semi-automatic glycan analysis from MS1 and MS2 – GlycoWorkbench
- 4) Automatic glycan analysis from MS2 – SimGlycan
- 5) Glycopeptide analysis from MS2 – Byonic

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, XITandem, ProteinPilot, SpectrumMill, Andromeda, ... on every measure (# IDs, # consensus IDs, # spikes, # modifications, site localization, FDR, ...)



Byonic Input

File Edit Help

Input Files

Select MS/MS data file: C:\data_input\Mass_Spectra\cona_tmt0saxpdetd_etc[node_04].mgf

Select Protein database file: C:\data_input\Protein_Databases\perumTIO2_O18.rev.fasta

Options

Digestion and Machine parameters | Modifications | Advanced

Fixed and Variable modifications

Total common modification max: 2

Total rare modification max: 1

C[S7], fixed
K[+224.1525], fixed
N-terminal [+224.1525], fixed
M[+16], common2
N-terminal Q[-241.1791], rare
NGlycan[1257.450], rare
NGlycan[1540.529], rare
MGlycan[1840.240], rare

Glycan modifications

N-linked search: Common human

O-linked search: None

Wildcard search

Enable wildcard search:

Minimum mass: -20

Maximum mass: 30

Restrict to residues:

Load parameters... Save parameters... Reset parameters

Run

On completion: Notify pop-up Open result folder

Multicore options: Multicore: Normal

Byonic (TM) by Protein Metrics Inc.
www.proteinmetrics.com

Byonic Input

The screenshot shows the Byonic software interface with the following configuration:

- Input Files:**
 - Select MS/MS data file: C:\data_input\Mass_Spectra\cona_tmt0saxpdetd_etc[node_04].mgf
 - Select Protein database file: C:\data_input\Protein_Databases\serumTIO2_018.rev.fasta
- Options - Advanced:**
 - Blind modification search:** Enabled (checked).
 - Wildcard search:** Enabled (checked).
 - Minimum mass:** -20
 - Maximum mass:** 30
 - Restrict to residues:** (empty)
 - Fixed and Variable modifications:**
 - Total common modification max: 2
 - Total rare modification max: 1
 - Selected keyword: **NGlycan** (circled in red)
 - Glycan modifications:**
 - N-linked search: **Common human** (circled in red)
 - O-linked search: None
 - Predefined tables of common glycan masses:** (circled in red)
- Run:** (circled in red)

ABRF Course 55

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:** TMT⁰ (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

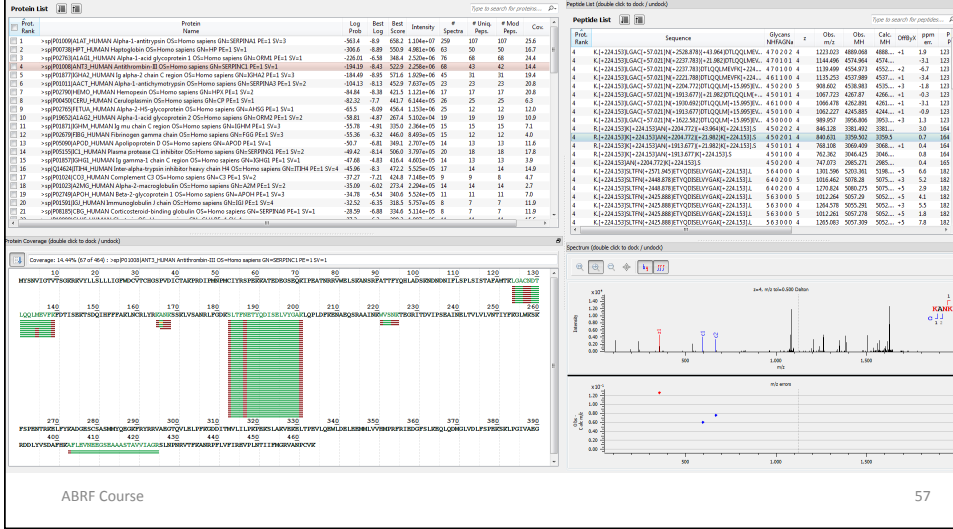
Copyright © 2012 Julian Saba et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

56

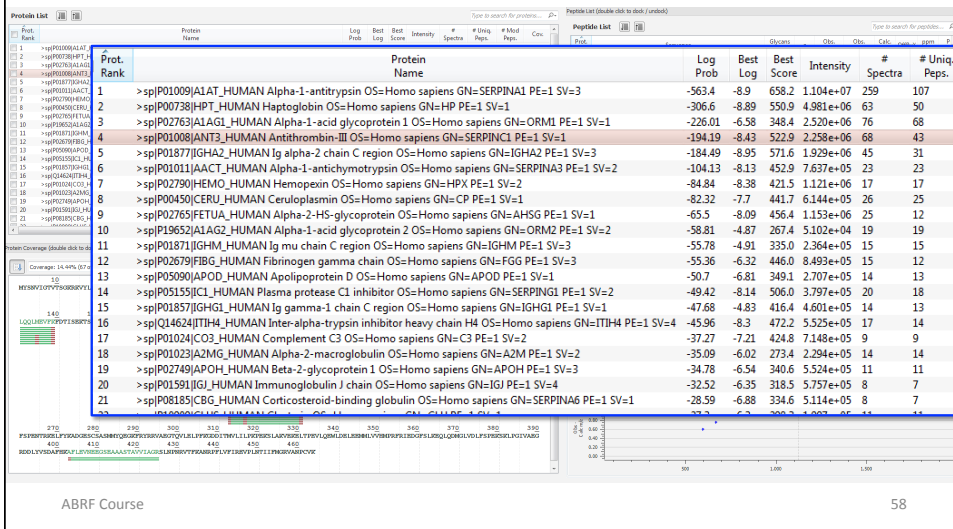
Glycoproteomics Example

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



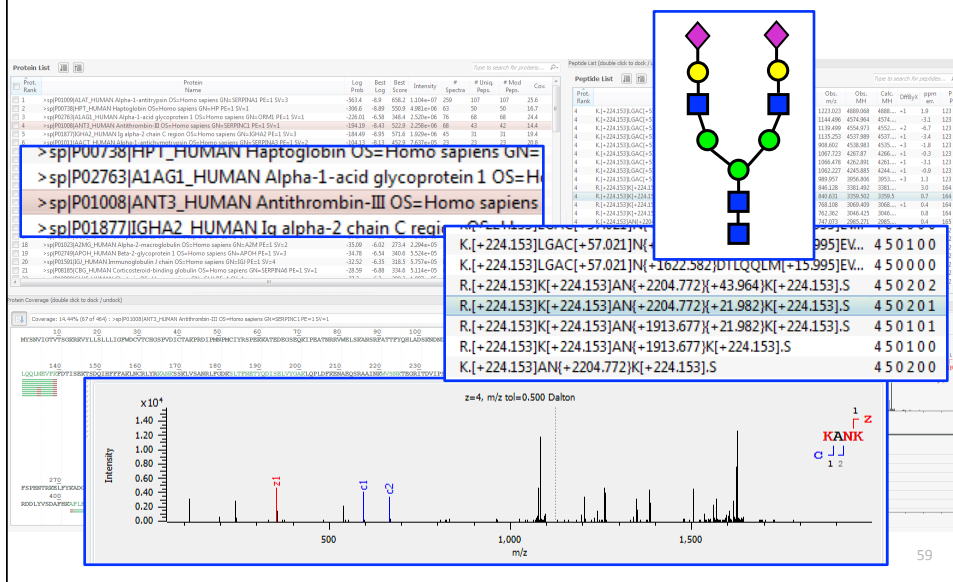
Byonic Glycoproteomics Search

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



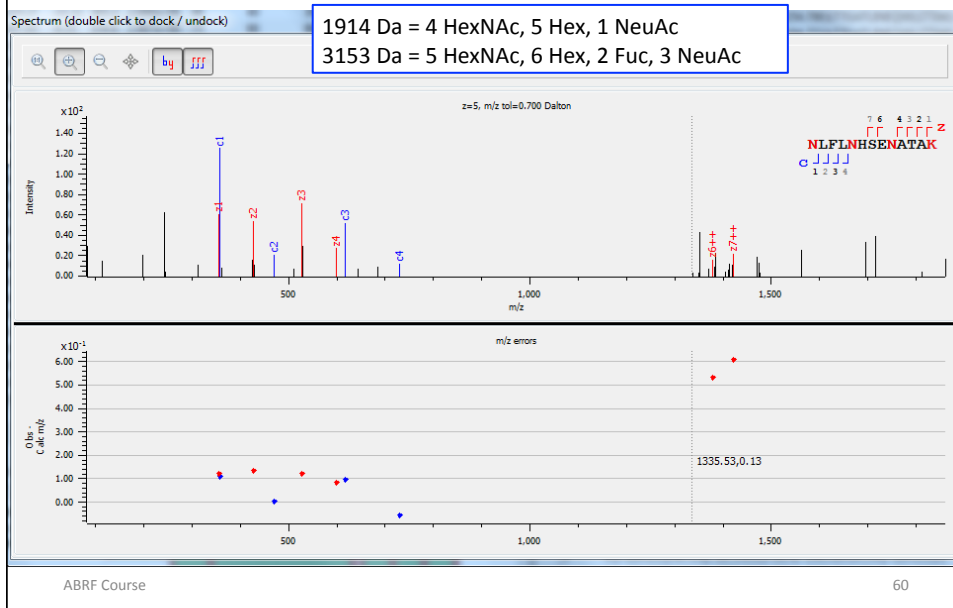
Byonic Glycoproteomics Search

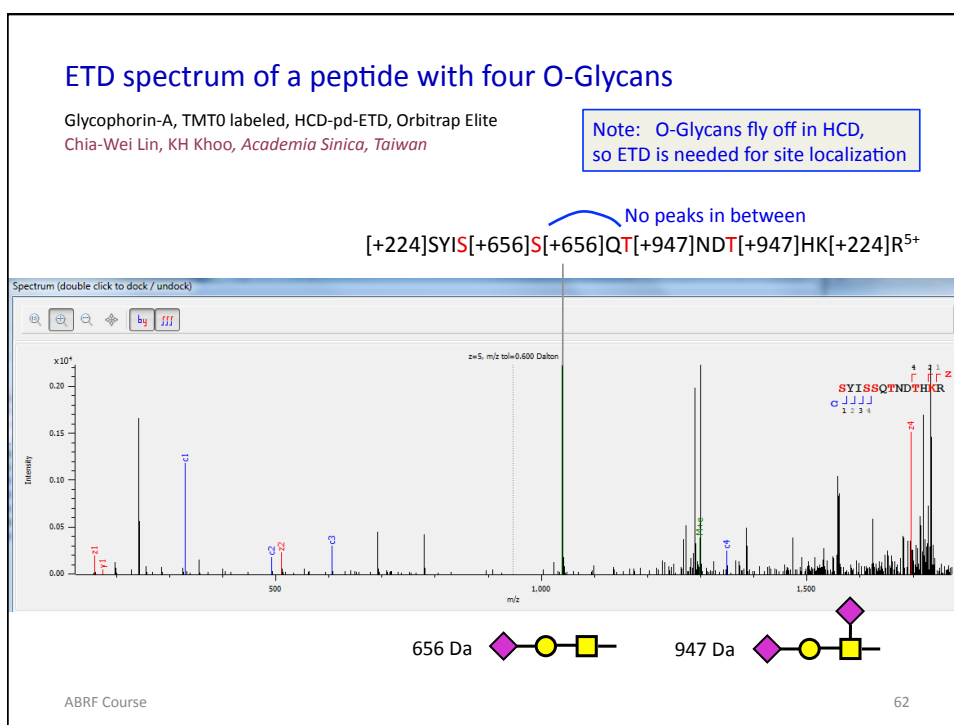
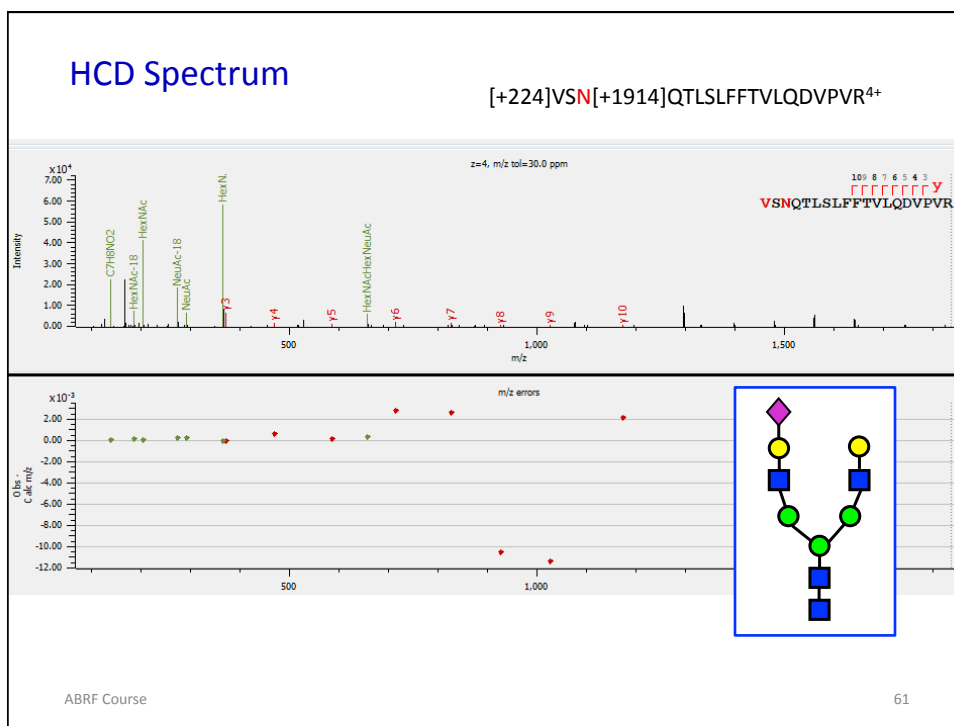
2535 ETD spectra → ~70 glycoproteins, ~110 N-glyco sites, ~700 distinct glycopeptides



ETD spectrum of a peptide with two N-glycans

N[+224]LFN[+3153]HSEN[+1914]ATAK[+224]⁵⁺ haptoglobin





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

ABRF Course

63

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 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.

ABRF Course

64