

# Recent Developments in Thermo ● Scientific LSMS Proteomics Software

David Horn

Proteomics Software Strategic Marketing  
Manager

September 22, 2011

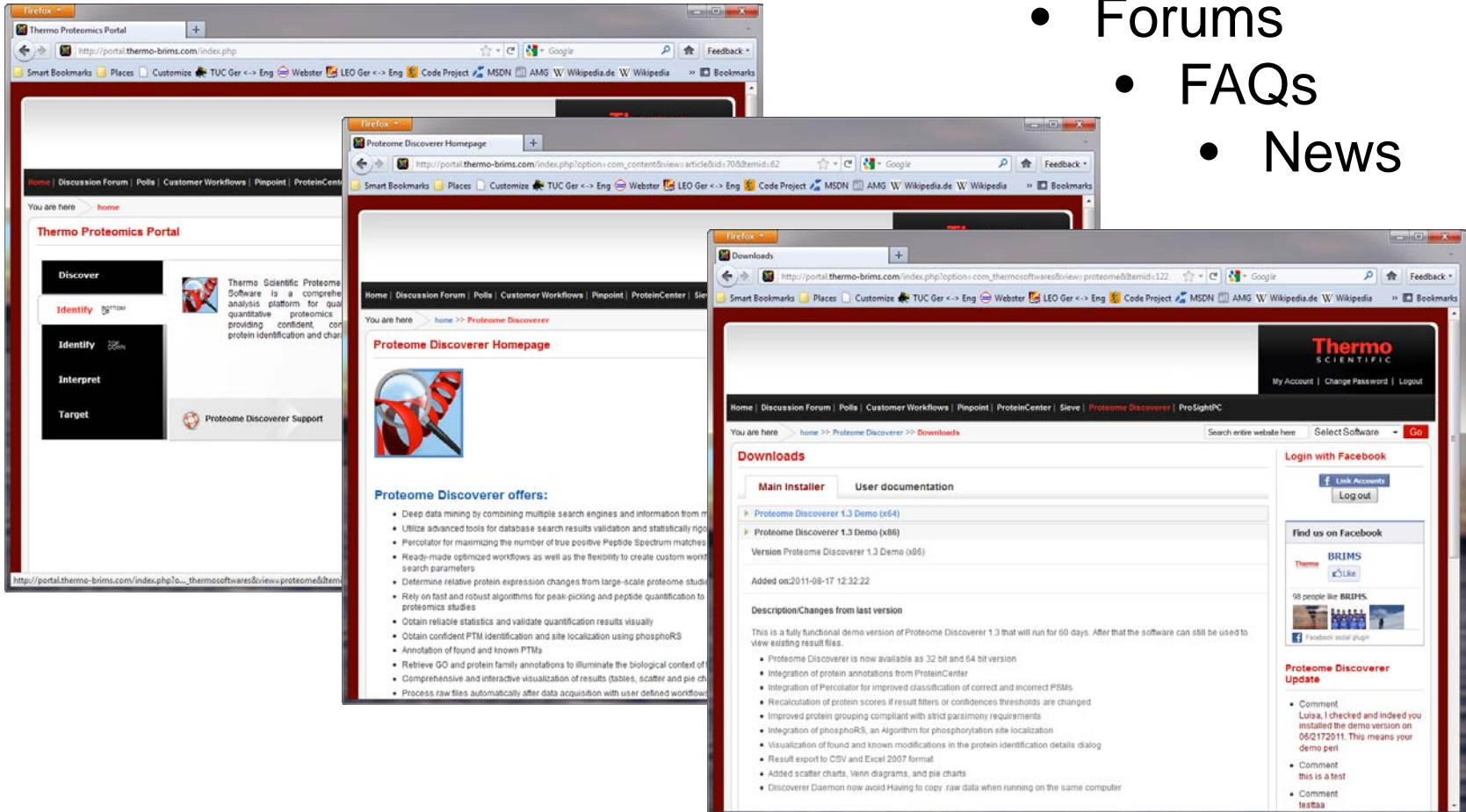
# What's New In Proteomics Software For 2011?

- **Thermo Scientific Proteomics Software Portal**
- **Proteome Discoverer 1.3**
  - Percolator, PhosphoRS, Annotation
- **ProteinCenter**
  - Released early 2011
- **SIEVE 1.3 SP2**
  - ROC analysis, improved statistics, support for new instruments
- **Pinpoint 1.2**
  - Improved HR/AM support, support for new instruments
- **SimGlycan**
  - Glycomics software
- **Protein Deconvolution 1.0**
  - ReSpect™ and Xtract for intact protein mass determination

# Thermo Scientific Proteomics Portal

- <http://portal.thermo-brims.com>

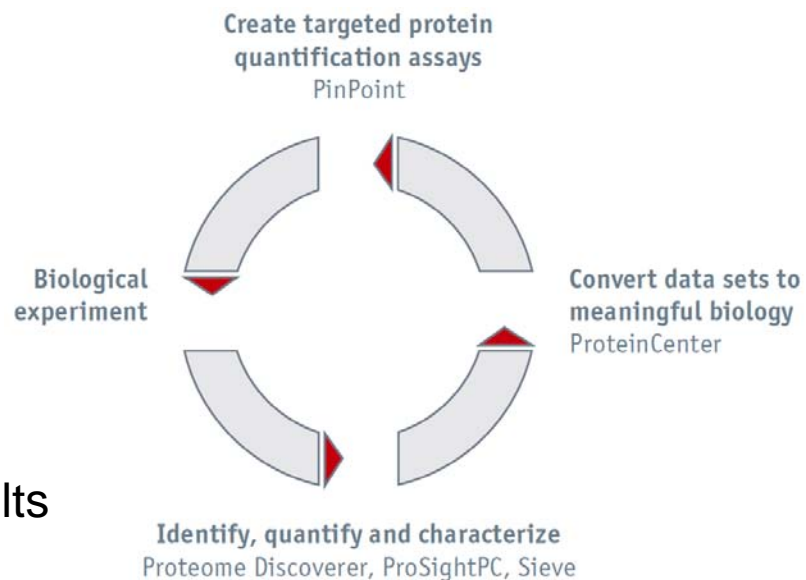
- Downloads
  - Help Resources
  - Forums
  - FAQs
  - News



- **Web server for biological interpretation of proteomics data**

- **Key features**

- Dataset comparison
- Overrepresentation analysis (GO terms, Pfam, keywords, pathways, etc.)
- Heat maps and profiling of quantitative proteomics data
- KEGG pathway analysis
- Comparison to previously published results

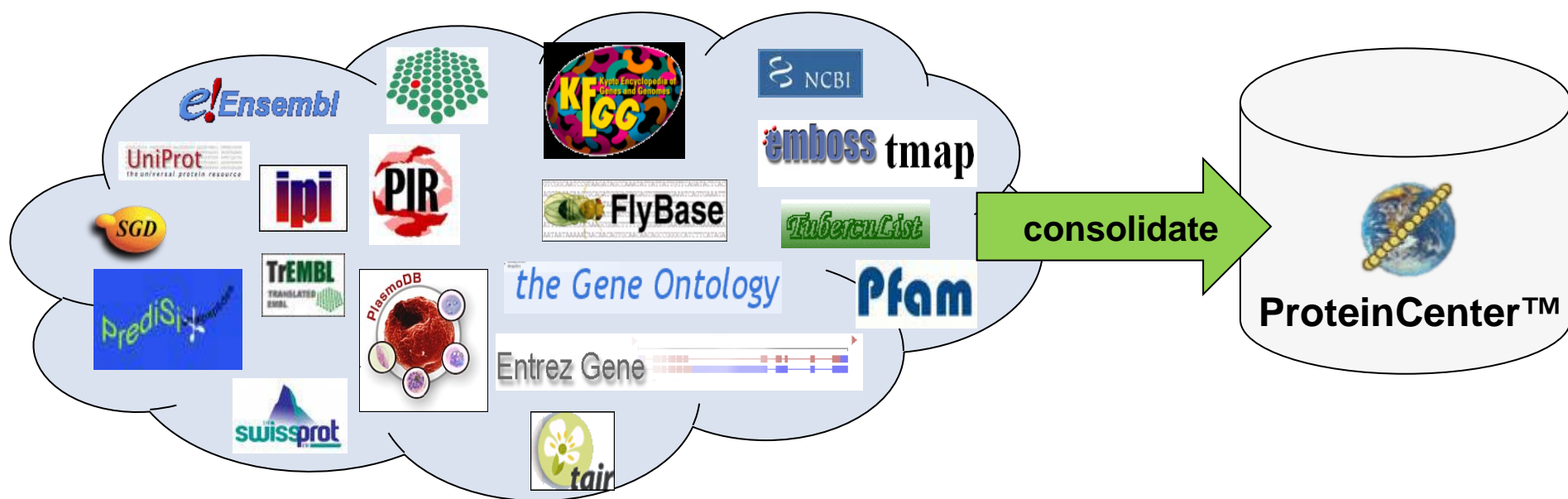


More information at:

[http://www.proxeon.com/productrange/data\\_interpretation/introduction/index.html](http://www.proxeon.com/productrange/data_interpretation/introduction/index.html) or the Thermo Scientific proteomics web portal

# ProteinCenter™

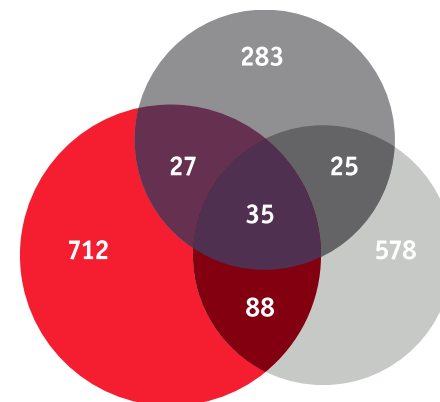
- Protein-centric data warehouse specifically designed for interpretation of proteomics data
- >16 million protein sequences from the major public protein databases distilled from 130 million accession codes from past and present versions
- The consolidated database is updated bi-weekly



# Compare data sets

Independent of input format, searched protein database, and database version

See in which data set each protein has been observed. Determine true overlaps.

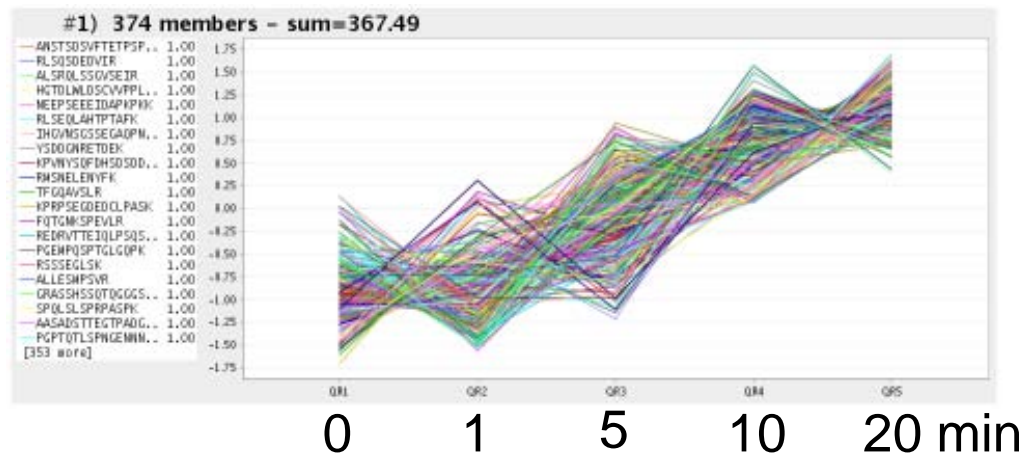


■ Plasma Proteome  
■ CSF Proteome  
■ Tear Fluid Proteome

Acc. Key	No	O	Description	S	Data set	Cluster	Gene	AA	AS	Fr	Tax	Molecular Functions	Cellular Components	Biological Processes	TM	SP	Pep
<a href="#">4502105</a>	<a href="#">3</a>		annexin IV	<input type="checkbox"/>		-	ANXA4	321		C	Hs				0		<a href="#">7</a>
<a href="#">32189392</a>	<a href="#">3</a>		peroxiredoxin 2 isoform a	<input type="checkbox"/>		-	PRDX2	198		C	Hs				1		<a href="#">6</a>
<a href="#">67461552</a>	<a href="#">3</a>		Erlin-1	<input type="checkbox"/>		-	ERLIN1	346			Hs				0		<a href="#">2</a>
<a href="#">42656431</a>	<a href="#">3</a>		similar to FKSG30	<input type="checkbox"/>		-	LOC389036	534			Hs				0		<a href="#">3</a>
<a href="#">20357529</a>	<a href="#">3</a>		guanine nucleotide-binding pro...	<input type="checkbox"/>		-	GNB2	340			Hs				0		<a href="#">5</a>
<a href="#">4507677</a>	<a href="#">3</a>		heat shock protein 90kDa beta,...	<input type="checkbox"/>		-	HSP90B1	803			Hs				1		<a href="#">8</a>
<a href="#">67089147</a>	<a href="#">3</a>		farnesyl-diphosphate farnesylt...	<input type="checkbox"/>		-	FDFT1	417			Hs				4		<a href="#">7</a>
<a href="#">IPI00412577.1</a>	<a href="#">3</a>		34 kDa protein	<input type="checkbox"/>		-	ANXA2	302			Hs				1		<a href="#">5</a>
<a href="#">33188452</a>	<a href="#">3</a>		peroxiredoxin 2 isoform b	<input type="checkbox"/>		-	PRDX2	147			Hs				0		<a href="#">6</a>
<a href="#">ENSP00000237530</a>	<a href="#">3</a>		-	<input type="checkbox"/>		-	RPN2	676			Hs				5		<a href="#">8</a>

# ProteinCenter example – Phosphopeptide profiling

- Taken from the poster “**Bioinformatics analysis of quantitative proteomics datasets using soft clustering algorithms**”
- EGF stimulation of HeLa cells, profiling performed in ProteinCenter



- Cluster 1: phosphopeptides that increase in abundance over time
- Over-representation analysis shows
  - MAPK, ErbB pathways overexpressed
  - Phosphorylation over-represented

## Over-represented KEGG pathways

Analysis data set: group 1 Reference data set: group 7

<<< 1 2 3 >>>

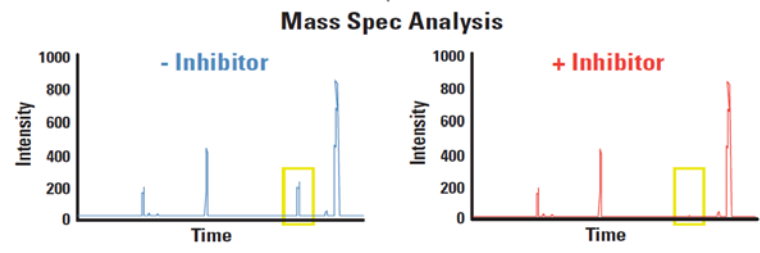
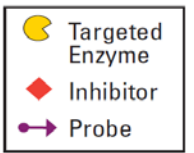
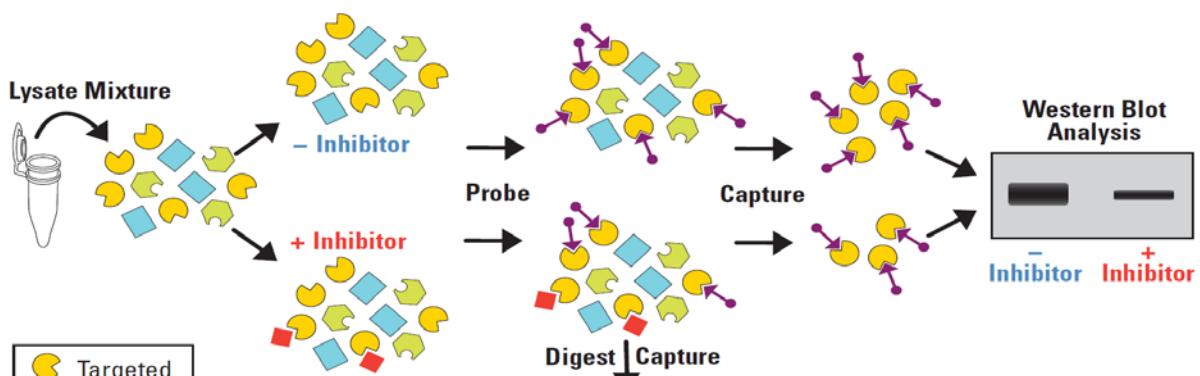
Description	Occurrence	Count	Ref. Count
<a href="#">MAPK signaling pathway (hsa)</a>	1	6	0
<a href="#">ErbB signaling pathway (hsa)</a>	1	5	0
<a href="#">Insulin signaling pathway (hsa)</a>	1	7	1
<a href="#">Gap junction (hsa)</a>	1	4	0
<a href="#">GnRH signaling pathway (hsa)</a>	1	4	0
<a href="#">Acute myeloid leukemia (hsa)</a>	1	4	0
<a href="#">Chronic myeloid leukemia (hsa)</a>	1	4	0



# Proteome Discoverer + ProteinCenter + Pinpoint for kinase activity profiling

Peterman, *et al*  
ASMS poster  
MP666

Sample Preparation  
Data Acquisition  
Data Processing



Unbiased database matching  
Proteome Discoverer



800 proteins, requires  
extensive manual curation



Biological Organization  
Protein Center

Targeted Protein Quantitation  
Pinpoint 1.2



150 proteins, all  
kinases

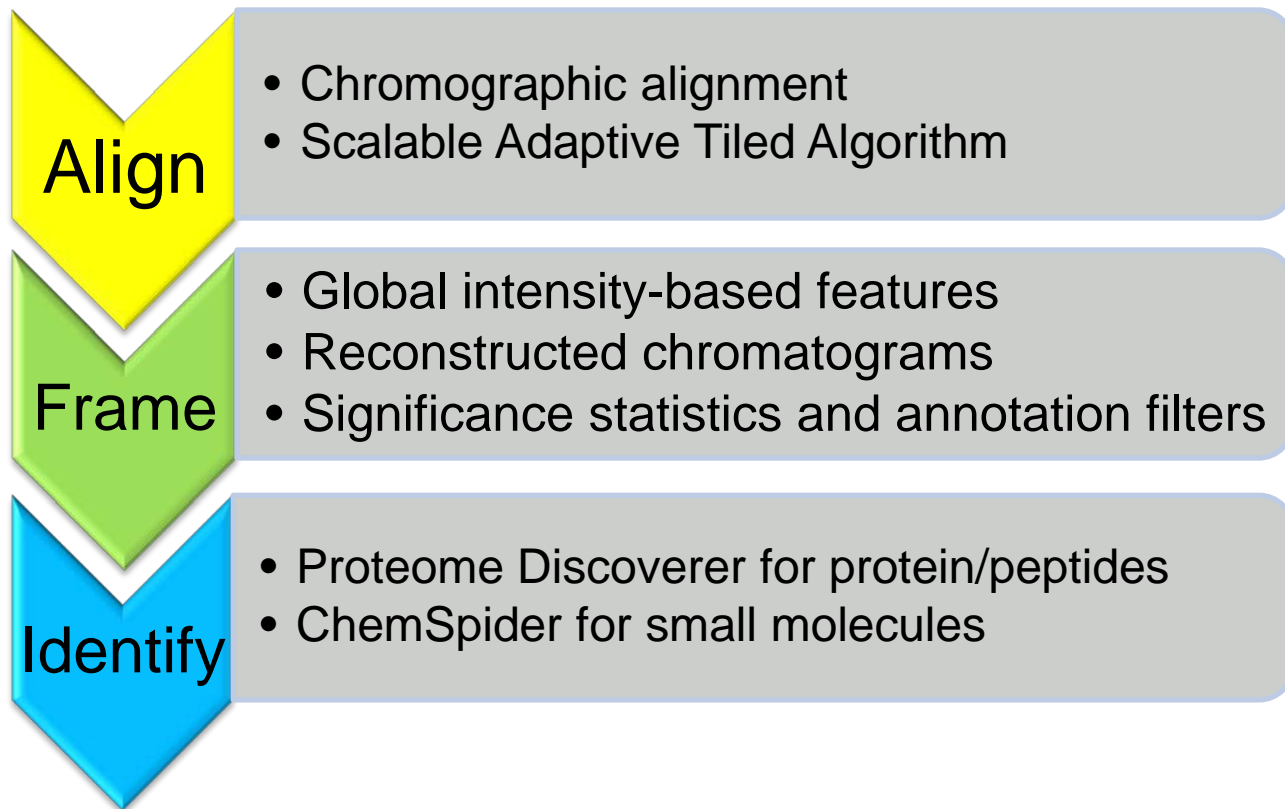


# SIEVE

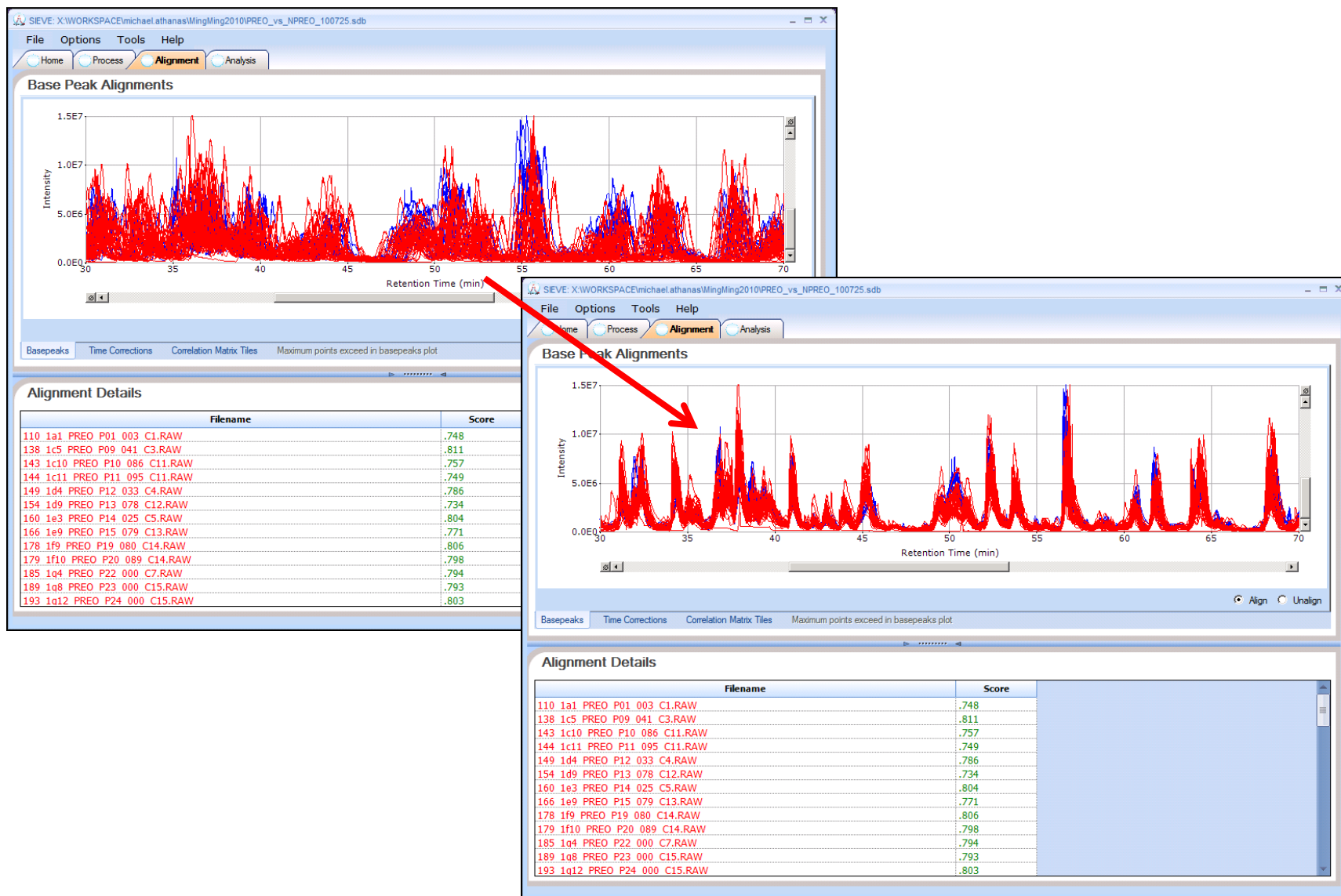
---

- **Label-free quantification software for high resolution accurate mass (HR/AM) proteomics and small molecule data**
- **Current Version: SIEVE 1.3 SP2**
- **Key features**
  - Several types of analyses, including A vs. B, time course, receiver operating characteristics (ROC) curves, non-differential single class analysis
  - Powerful chromatographic alignment algorithm
  - Powerful statistical analysis, including principal components analysis (PCA) and K-means clustering
  - Improved integration with Proteome Discoverer
  - Results can be exported to ProteinCenter and Pinpoint

# SIEVE Processing Workflow



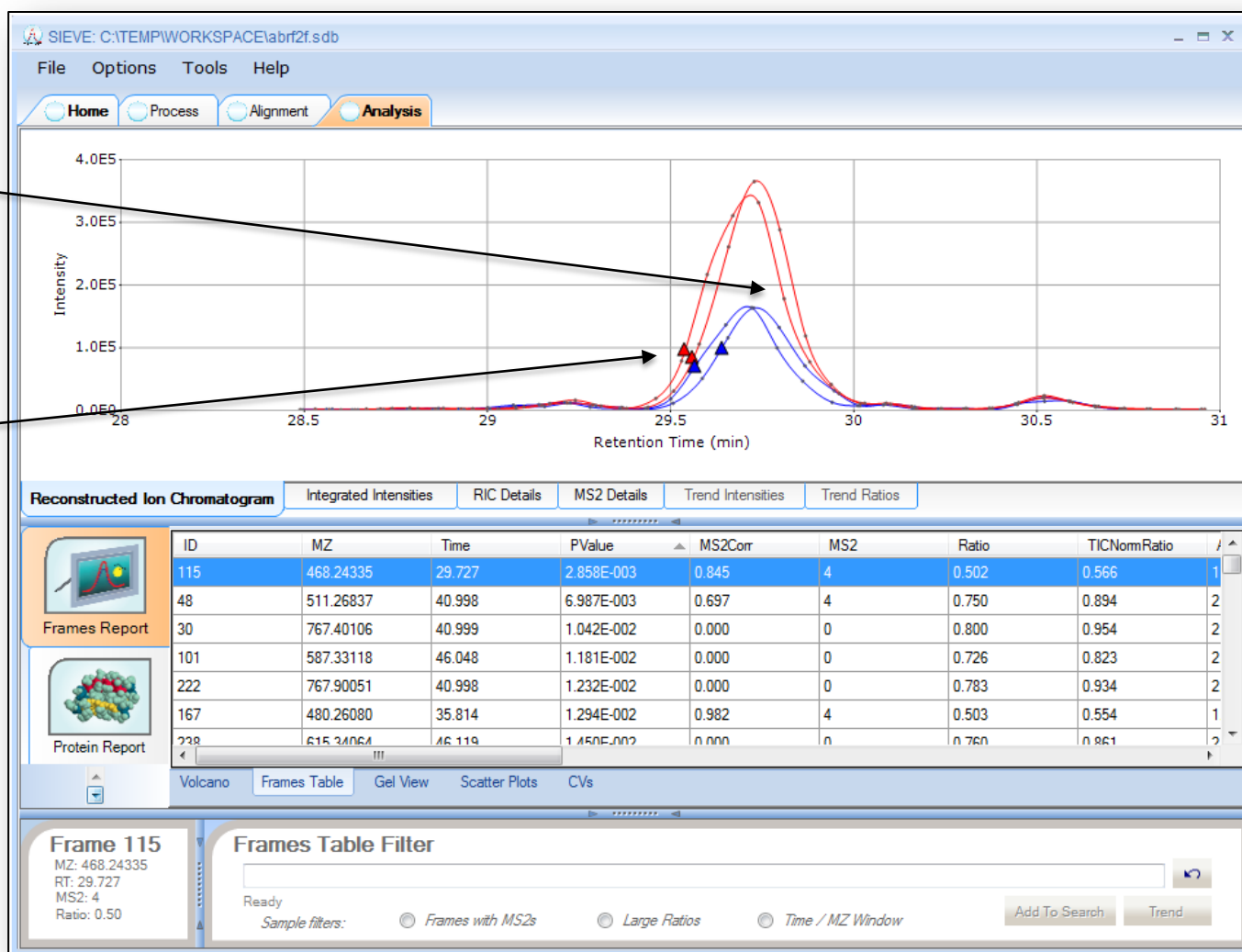
# Chromatographic Alignment



# Viewing frames in SIEVE

Peak shape based upon full scans.

Triangles represent MS2 scans.



# SIEVE Protein Report

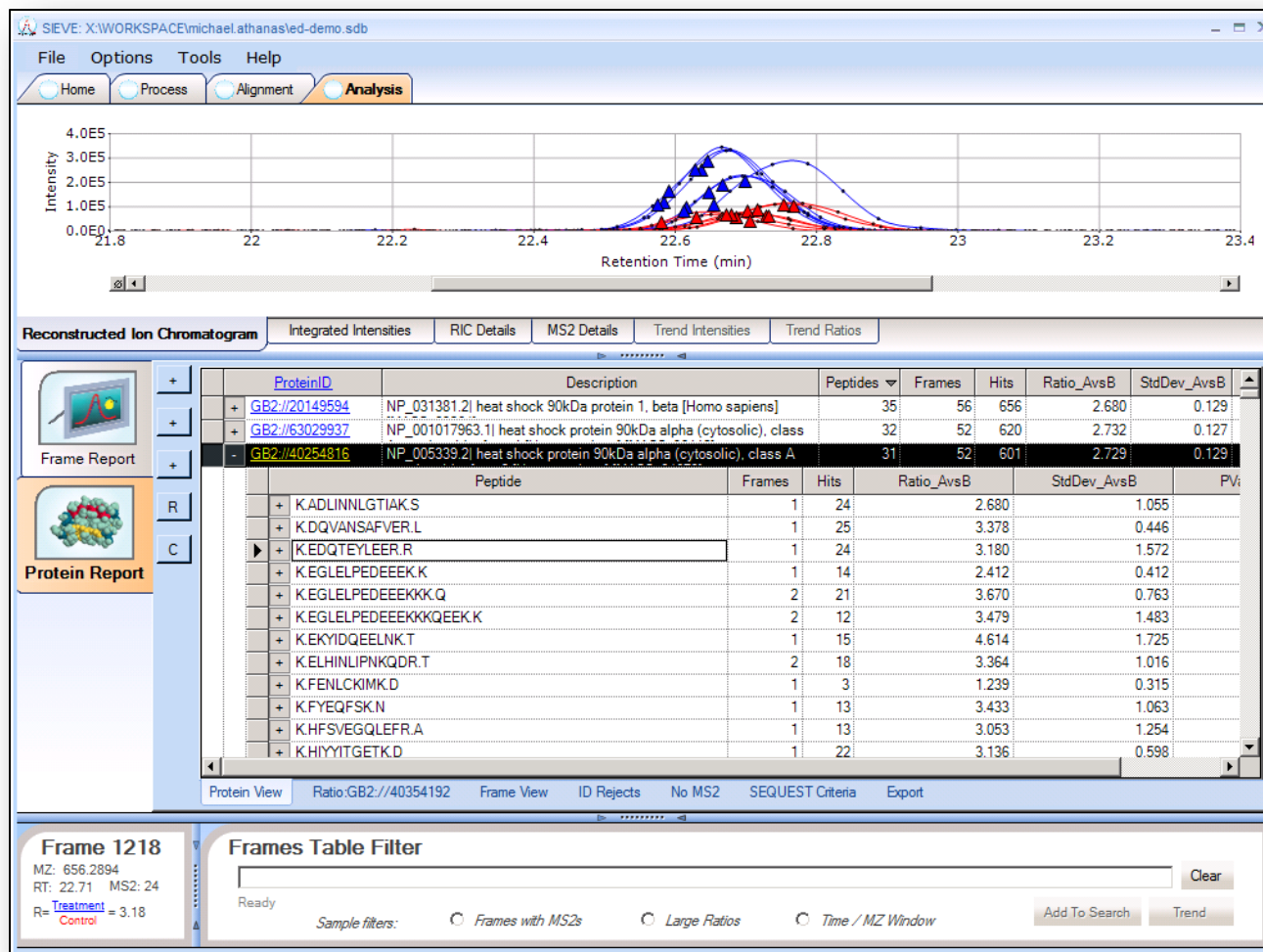
## Proteins → Peptides →

- Which proteins and peptides are found in my differential experiment?

- At what ratios?

- How many distinct peptides?

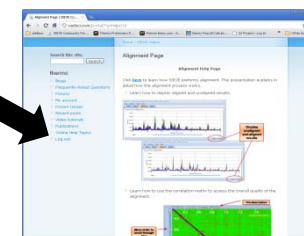
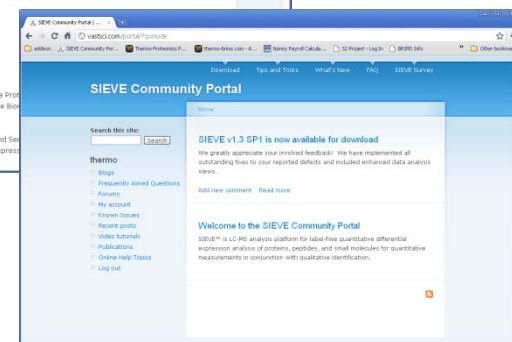
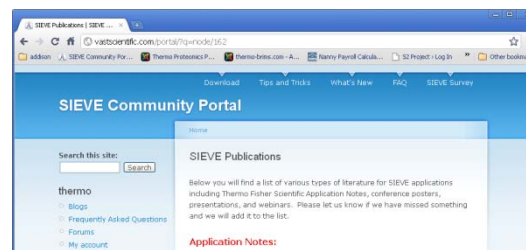
- What is the probability that the protein is differentially expressed?



# SIEVE community portal

<http://sieve.vastsci.com>

- Demo software available for instant download 30 day free trial license, after license expires software becomes a viewer.
- Publication page with direct links to application notes, poster, presentation and peer reviewed publications.
- Direct links in software to portal for topic specific help, which allows for continuous updating
- Over 30 record video tutorials to guide the user through the software.



# SIEVE 2.0—Revolutionary Breakthrough for Metabolomics

*SIEVE 2.0 accelerates mining of Metabolomics datasets with optimized component detection and dramatically decreased false positives.*

<b>Features</b>	<b>SIEVE 1.3</b>	<b>SIEVE 2.0</b>
<b>Component Elucidation</b>	<b>Framing by Intensity</b>	<b>Chromatographic peak detection</b>
<b>Alignment</b>	<b>Aligns TIC</b>	<b>Aligns Components</b>
<b>Background correction to reduce false positives</b>		✓
<b>Grouping of adducts, dimers, isotopes</b>		✓
<b>Look for missing peaks</b>		✓
<b>Statistical tools</b>	✓	✓
<b>Graphical displays</b>	✓	✓
<b>Database searches (local &amp; on-line)</b>	✓	✓



# Pinpoint

---

- **Software for SRM and HR/AM targeted quantitative proteomics**
- **Key features**
  - Targeted relative and absolute peptide quantification
  - Support for all Thermo MS platforms (TSQ, Orbitrap, ion trap)
  - Q Exactive, OT Elite support in Pinpoint 1.2
  - Accepts input data directly from Proteome Discoverer and SIEVE
  - Retention time prediction trained by peptide retention time standards kit
  - Streamlined processing from method development to data processing
  - Peptide mapping tool for analysis of recombinant proteins and biopharmaceuticals

# Automated Data Processing in Pinpoint – Verification Tools

Pinpoint Workspace C:\X\ EN English (United States) Microphone Pause Tools Handwriting Drawing Pad

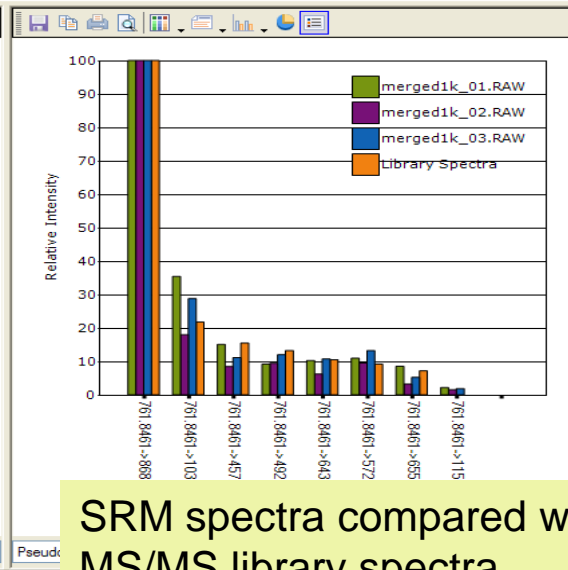
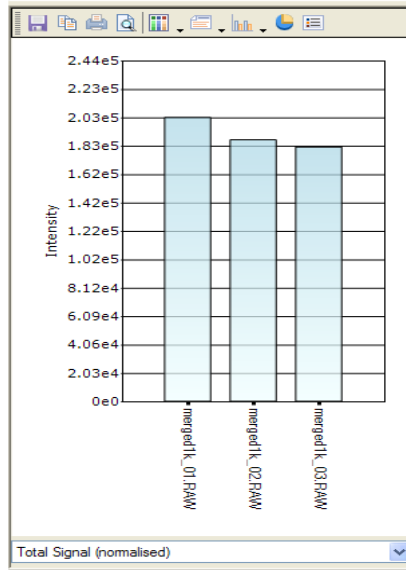
File Edit View Protein/Peptide Management Batch Mode Data Analysis Management Options Help

Main Workbook Raw File Management **Detailed Data Analysis** Absolute Quantification/Calibration

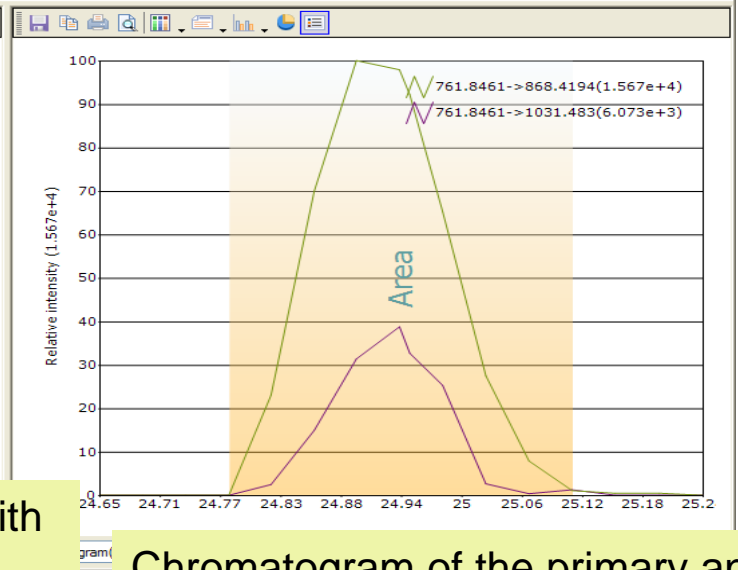
Protein/Peptide/Precursor/Product	Score	Retention Time (in)	Normalised file	Bonferroni	Normalised file	Bonferroni	Normalised file	Bonferroni	CV% (All files)
YLL050C	●		1.556e+6		1.383e+6		1.368e+6		6
YLR027C	●		4.752e+5		4.456e+5		4.816e+5		3
YLR028C	●		1.279e+5		1.192e+5		1.041e+5		8
YLR029C	●		9.263e+4		1.039e+5		8.884e+4		7
YLR043C	●		5.842e+5		5.098e+5		4.970e+5		7
FSEQYPQADFYK	●	24.94	2.030e+5		1.870e+5		1.819e+5		5
761.8461	●		2.030e+5	1.612e-3	1.870e+5	7.168e-3	1.819e+5	7.168e-3	5
TASEFDLSAIAQDK	●	20.5	3.812e+5		3.228e+5		3.151e+5		9
YLR044C	●		1.835e+5		2.342e+5				
YLR058C	●		2.245e+6		2.249e+6				
YLR060W	●		5.246e+4		4.764e+4				
YLR075W	●		5.132e+5		4.823e+5				
YLR109W	●		1.235e+7		1.192e+7				
YLR153C	●		4.318e+5		4.753e+5				

Score  
 Retention Time  
 Use Transition?  
 Internal Standard  
 Keep in next iteration  
 Normalised file Signal  
 Total file Signal  
 Pseudo MSMS  
 Signal-to-Noise  
 Ion Ratio  
 Precursor Ratio  
 Peptide Ratio  
 Bonferroni corrected p-value  
 Product with Lib Spectr

Correlation p-value between SRM spectra and MS/MS library spectra



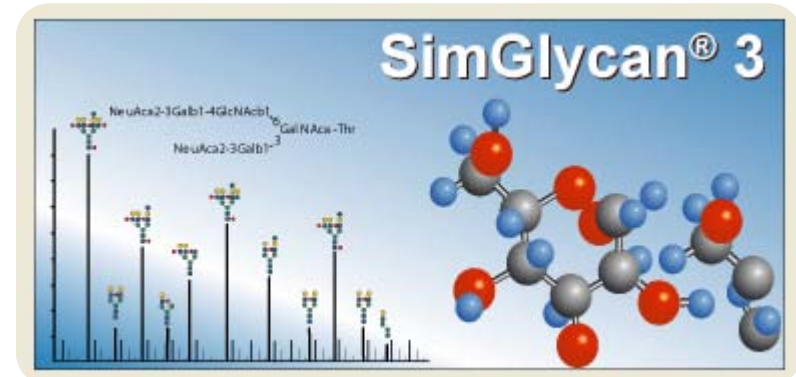
SRM spectra compared with MS/MS library spectra



Chromatogram of the primary and secondary transitions. Area calculated based on primary

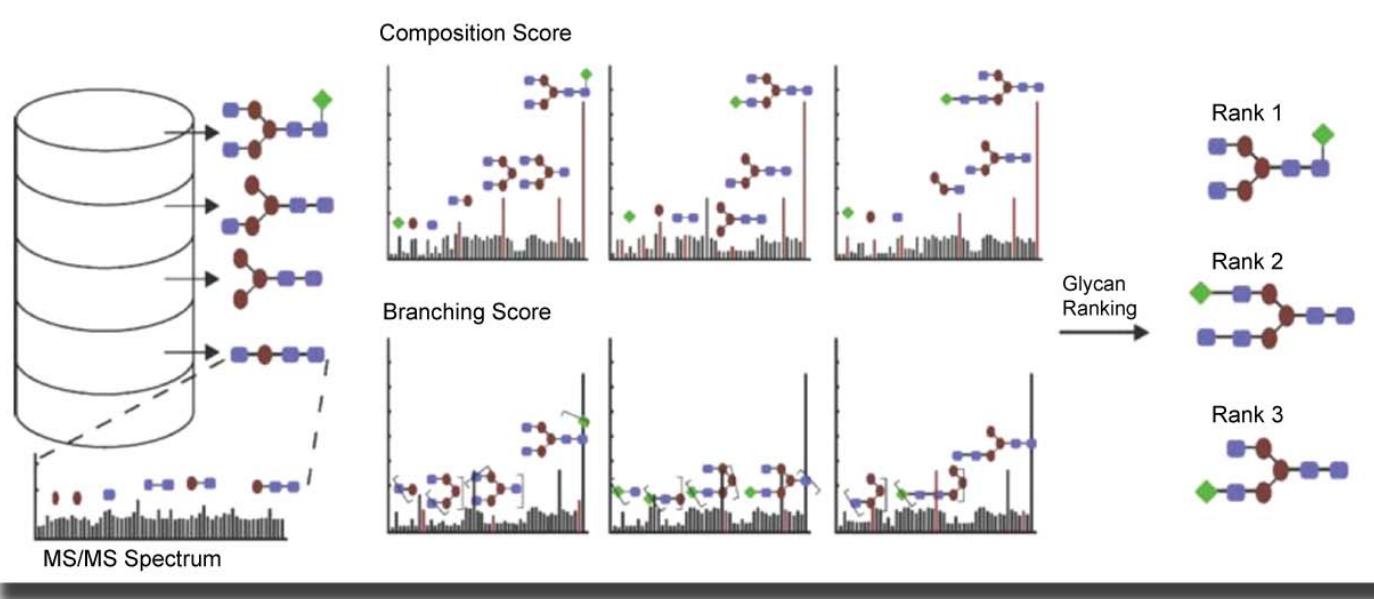
# SimGlycan<sup>®</sup> Software for Glycomics

- Automated software package for glycomics analysis – the most comprehensive software package available for glycan structural elucidation
- Key Features
  - Structural analysis/identification of glycans
    - Characterization of N-linked glycans from antibodies
  - Structural characterization of MS<sup>n</sup> data (**Unique to ion traps**)
  - Largest commercially-available glycan database (over 9500 structures)
  - Automatically process up to 1500 input spectra
- Available through Premier BioSoft



# How SimGlycan™ Works

- Database search approach



*Methods Mol Biol.* **2010**,600,269-81

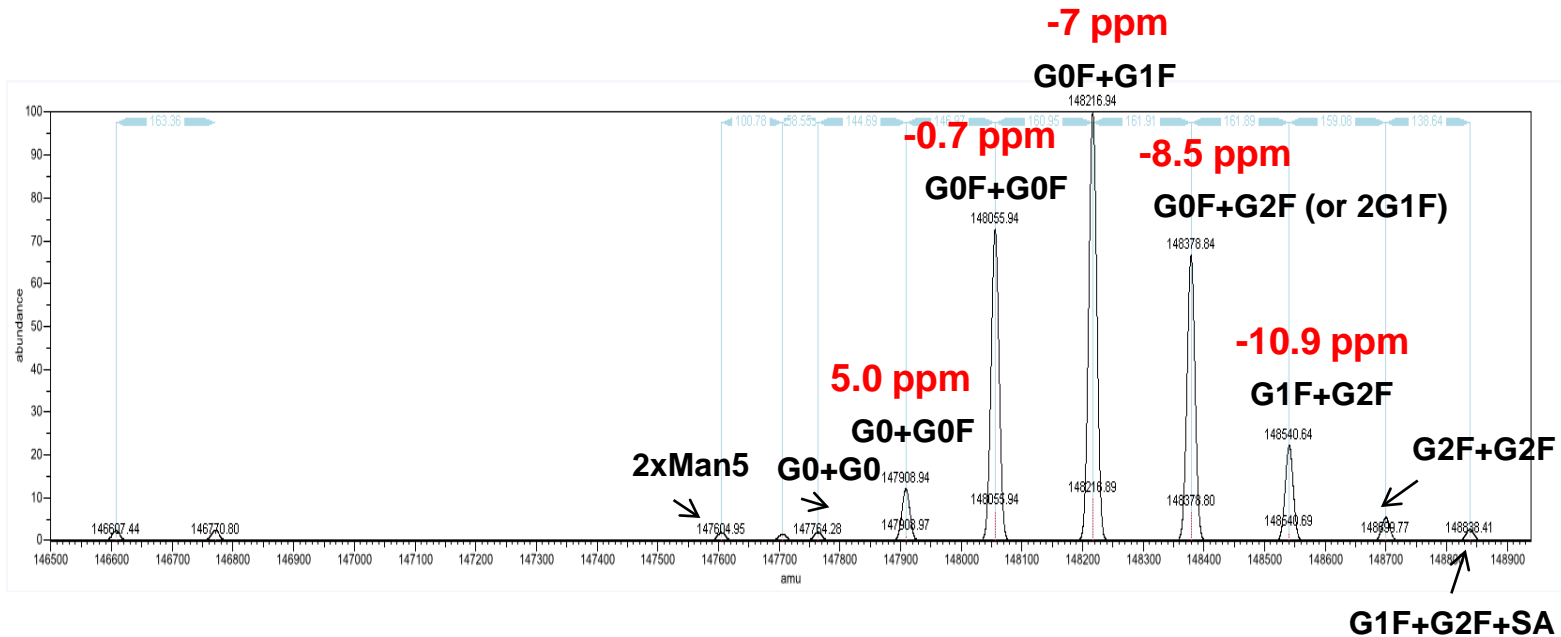
- MS/MS data are searched against SimGlycan's own database of theoretical fragmentation of over 9650 glycans
- Each proposed structure is assigned a score to reflect how closely it matches with the experimental data based on composition and branching
- Other relevant biological information for the proposed glycan structures such as the glycan class, reaction, pathway, and enzyme are also made available via interactive links

# Protein Deconvolution 1.0

---

- Workflow software for intact protein mass determination
- User interface “borrowed” from small molecule software package
- Includes 2 deconvolution algorithms:
  - ReSpect™ for isotopically unresolved proteins (e.g. IgG)
  - Xtract for isotopically resolved proteins
- As of today, the software is feature complete and undergoing testing
- Target release date: Late October, early November

# ReSpect™ Deconvolution of Q Exactive IgG Data



RAW file	Q Exactive	ppm mass measurement errors				
		G0+G0F	G0F+G0F	G0F+G1F	G0F+G2F	G1F+G2F
1	1	-10.5	0.7	-10.5	-13.8	-18.0
2	1	-3.2	-4.3	-6.9	3.2	N/A
3	1	-11.6	-1.1	-8.8	-11.2	-12.0
4	1	5.1	-5.0	-2.6	5.1	5.6
5	2	-14.3	3.0	-6.9	-5.4	-5.9
6	2	-8.6	-2.2	-12.2	-12.5	-12.9
7	2	-14.3	-6.6	-12.3	-14.8	-10.1

-6.9 +/- 6.4 ppm mass tolerance for above measurements

# Protein Deconvolution 1.0

Protein Deconvolution (Version 1.0.37)

Thermo SCIENTIFIC Protein Deconvolution IgG\_QExactive Manual ReSpect™ ( Isotopically Unresolved ) Help

Method Selection  Parameters  Chromatogram  Process and Review  Reporting

Select an experiment type to create a new experiment or continue an existing experiment.

### Experiment Types

- Manual Xtract ( Isotopically Resolved )
- Manual ReSpect™ ( Isotopically Unresolved )**
- Load Previous Results

### Load Raw Data File

Raw Data Directory: C:\Xcalibur\data

Select Raw Data Files

- C:\Xcalibur\data\IgG\_source\_cid-qb.raw
- C:\Xcalibur\data\IgG\_source\_cid.raw
- C:\Xcalibur\data\Myoglobin\_30pmol\_michrom\_protein\_microtrap\_11mi
- C:\Xcalibur\data\Myoglobin\_30pmol\_michrom\_protein\_microtrap\_11mi

Load

### Methods

Experiment Name	Description
DefaultMethodReSpect	Default method for ReSpect.
IgG_Elite	
IgG_QExactive	

Create Method Load Method



# Protein Deconvolution Method Parameters

Protein Deconvolution (Version 1.0.38) Manual ReSpec™ (Isotopically Unresolved) Help

Thermo SCIENTIFIC Protein Deconvolution IgG\_Q\_Exactive

Method Selection  **Parameters**  Chromatogram  Process and Review  Reporting

Save Method Save Method As Reset Method

Parameter Configuration

▼ Main Parameters ( ReSpec™ )

Charge Carrying Species	Mass
Negative Charge <input type="checkbox"/>	m/z Range Min 2000 Max 4000
Charge Carrier <input checked="" type="radio"/> H+ (1.0073) <input type="radio"/> 2H+ (2.014) <input type="radio"/> Na+ (22.9898)	Output Mass Range Min 140000 Max 160000
	Mass Tolerance 0.05 Da
	Target Mass 150000 Da
	Charge State Range 10 to 100

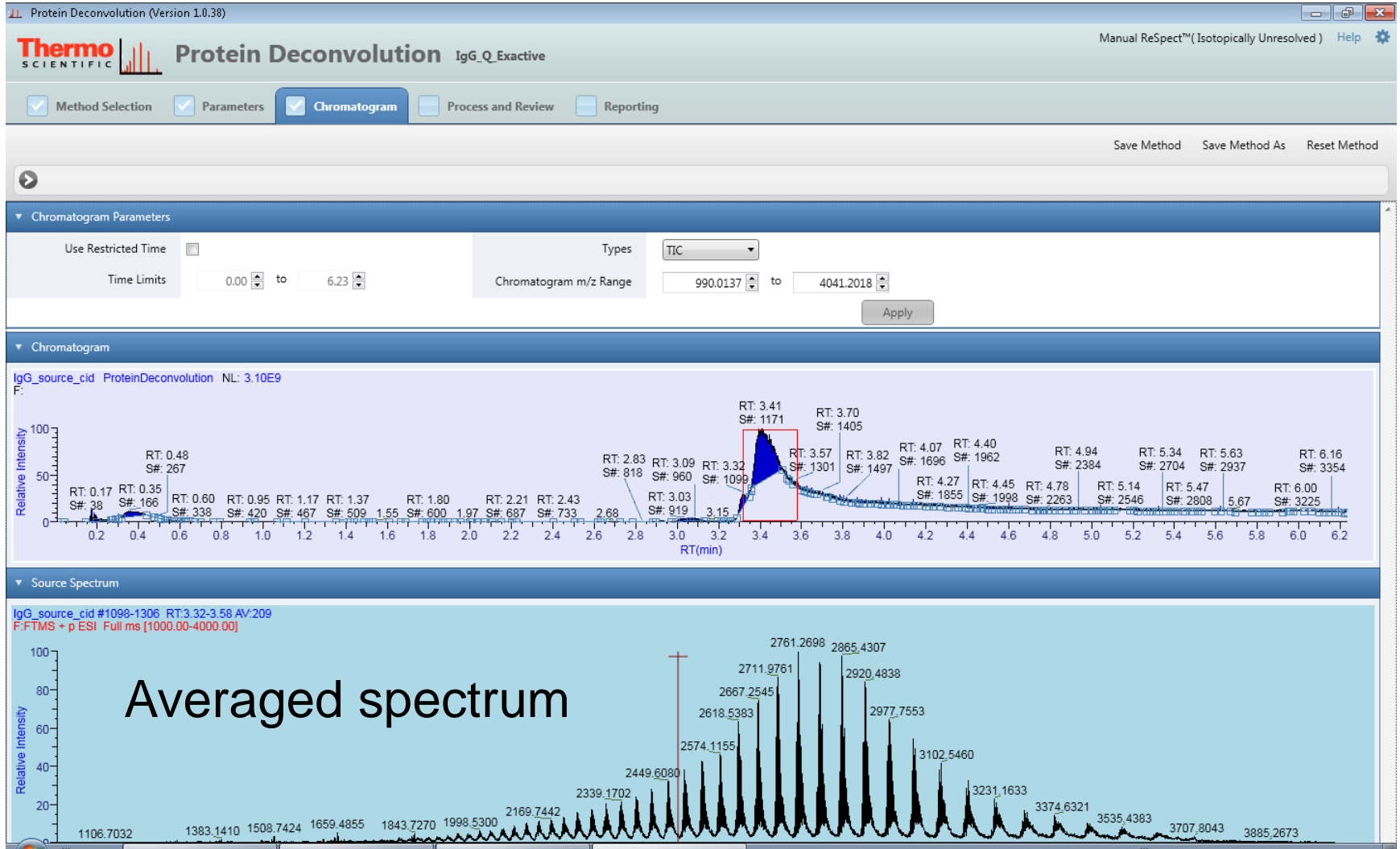
Apply

▼ Advanced Parameters ( ReSpec™ )

<b>Peak Filter Parameters</b>	<b>Deconvolution Parameters</b>
Minimum Peak Significance 1 Standard Deviations	Number of Iterations 3
Noise Rejection <input type="radio"/> No Noise Rejection <input type="radio"/> 50% Confidence <input type="radio"/> 68% Confidence <input checked="" type="radio"/> 95% Confidence <input type="radio"/> 99% Confidence	Noise Compensation <input checked="" type="checkbox"/>
Use Relative Intensities <input checked="" type="checkbox"/>	Minimum #. Adjacent Charges 10 to 10
<b>Baseline Correction</b>	<b>Peak Model Parameters</b>
Peak Width 0	Number of Peak Models 1
Feature Width 0	Resolution @ 400 12374
Degree of Fit 0	Left/Right Peak Shape Left 2 Right 2

Apply

# Protein Deconvolution – Chromatogram Tab



# Protein Deconvolution - ReSpect™ Deconvolution



# Protein Deconvolution - Report

Protein Deconvolution (Version 1.0.38) Manual ReSpec™ (Isotopically Unresolved) Help

Thermo SCIENTIFIC Protein Deconvolution IgG\_Q\_Extactive

Method Selection Parameters Chromatogram Process and Review **Reporting**

1 / 19 146% Find

## ProteinDeconvolution Report

Created: 9/22/2011 9:54:51 AM

Sample Information	
File Name	C:\XCALIBUR\Intact\IgG_source_cid.raw
Instrument Method	C:\Xcalibur\Intact\mab_SS_1.meth
Vial	CStk1-01:23
Injection volume (µL)	40
Sample Weight	0
Sample volume (µL)	0
ISTD Amount	0
Dil Factor	1

Chromatogram Parameters	
Use Restricted Time	False
Time Range	0.004648833 - 6.2300942
Type	TIC
Chromatogram m/z Range	990.013713569749 - 4041.20183719198

Source Chromatogram

[IgG\\_source\\_cid ProteinDeconvolution](#) NL: 3.10E9

# ProSightPC 2.0

---

- **Software for top-down protein identification and characterization**
- **Created by Neil Kelleher's group at Northwestern University**
  
- **Key Features**
  - Top-down and bottom-up protein identification
  - Iterative sequence matching using Sequence Gazer™
  - Xtract/THRASH for data reduction
  - UniProt flatfile database support
  - Multiplexed spectral ID
  - Error tolerant identification of modifications, truncations, etc
  
- ProSightPC 2.0 SP1 out soon
- ProSightPC 2.1 in development (faster and more confident searches)
- Planning a ProSightPC node for Proteome Discoverer

# Conclusions

---

- Thermo offers the most comprehensive and powerful suite of software applications in proteomics.
- The software applications work synergistically to extract information from our complex datasets.
- We can work collaboratively with 3<sup>rd</sup> parties and scientists to make new algorithms and software available.
- In 2011, we will have released or launched Proteome Discoverer 1.3, SIEVE 1.3 (+ 2 service packs), SIEVE 2.0, Pinpoint 1.1, Pinpoint 1.2, ProteinCenter, ProSightPC 2.0 SP1, SimGlycan, and Protein Deconvolution 1.0. (8 major software releases, 3 service packs!)
- Thermo Scientific is dedicated to provide the best possible proteomics software to our customers.



● **Supplementary slides**



# ProteinCard highlights

Known modifications, protein domains

Links to external resources

Identified peptides

**nuclear pore complex protein Nup153**  
NUP153, N153, HNUP153, nucleoporin 153kDa

Keys: 24430146 IPI00292059.2 P49790 ENSP00000262077

Peptides: 79(51) Coverage: 52.00%

Sequence: 1475 AA

Similar proteins: Max 158 Similarity: 98 %

External Link: HPRD, MIM, HPA, Entrez Gene, BLINK, UniRef100, UniRef90, UniRef50, PDB, PubMed, SNPs, Nt, UCSC, NCBI map, Homologene, GEO profiles, UniGene, IntActAll

**Features** 54 feature(s). S=Signal peptide, T=TransMembrane domain, Two letters= Pfam domains (Click to view details), P=Phosho, M=Other mod, G=Glyco

- Signal
- TM
- Pfam
- Modifications
- Glyco

**Gene & Protein Summary**  
**Gene Summary**  
Nuclear pore complexes are extremely elaborate structures that mediate the regulated movement of mac...

**Protein Summary**  
**Keywords:** Acetylation, 3D-structure, Nucleus, Metal-binding, Zinc-finger, mRNA transport, Transport, Nuclear pore complex, Direct protein sequencing, Protein transport, Translocation, Polymorphism, Phosphoprotein, DNA-binding, Repeat, Porin, Complete proteome, Zinc.  
**Functions:** - Possible DNA-binding subunit of the nuclear pore complex (NPC). The repeat-containing domain may be ...

**GO annotations**

Molecular Functions	Cellular Components	Biological Processes
DNA binding (1)	Golgi (1)	cell death (1)
catalytic activity (1)	cell surface (1)	cell differentiation (1)
metal ion binding (1)	cytoplasm (1)	cell motility (1)
nucleotide binding (1)	endoplasmic reticulum (1)	cell organization and biogenesis (1)
protein binding (1)	extracellular (1)	cell proliferation (1)
receptor activity (1)	membrane (1)	metabolic process (1)
signal transducer activity (1)	nucleus (1)	regulation of biological process (1)
transcription regulator activity (1)	organelle lumen (1)	response to stimulus (1)
transporter activity (1)		transport (1)

**Interactions & Pathways**  
11 Interactions

**Diseases**  
Many sequence variants affecting diversity of adul... (1)

**Interacting proteins**

**Details**

**Gene Details**  
Nuclear pore complexes are extremely elaborate structures that mediate the regulated movement of macromolecules between the nucleus and cytoplasm. These complexes are composed of at least 100 different polypeptide subunits, many of which belong to the nucleoporin family. Nucleoporins are pore complex-specific glycoproteins characterized by cytoplasmically oriented O-linked N-acetylglucosamine residues and numerous repeats of the pentapeptide sequence XFxFG. The protein encoded by this gene has three distinct domains: a N-terminal region within which a pore targeting domain has been identified, a central region containing multiple zinc finger motifs, and a C-terminal region containing multiple XFxFG repeats. [provided by RefSeq]

**Protein Details**  
**Keywords:** Acetylation, 3D-structure, Nucleus, Metal-binding, Zinc-finger, mRNA transport, Transport, Nuclear pore complex, Direct protein sequencing, Protein transport, Translocation, Polymorphism, Phosphoprotein, DNA-binding, Repeat, Porin, Complete proteome, Zinc.  
**Functions:** - Possible DNA-binding subunit of the nuclear pore complex (NPC). The repeat-containing domain may be involved in anchoring components of the pore complex to the pore membrane.  
**Description:**  
- Nuclear pore complex protein Nup153  
- Nucleoporin 153kDa

**Keywords**

# SimGlycan™ Results Page

Scan1@1784.638\_1 (ABRFglycans031009\_DHBdma\_sample12\_2)

MS Profile Search Results Annotated Peaklist

Glycan Rank	Glycan ID	Glycan Sequence	Composition Score	Branching Pattern Score	User Assigned Rank	User Comment
1	G00199	Gal(b1-4)GlcNAc(b1-2)Man...	1	1		
1	G00274	Gal(b1-4)GlcNAc(b1-2)Man...	1	1		

Glycan Rank
Glycan ID
Glycan Sequence
Composition Score
Branching Pattern Score

Additional biological information such as class, reaction, pathway and enzyme

Glycan Structure

Glycan Structure (G00199)

Sequence:  
Gal (b1-4) GlcNAc (b1-2) Man (a1-3) [ Gal (b1-4) GlcNAc (b1-2) Man (a1-6) ] Man (b1-4) GlcNAc (b1-4) GlcNAc

Carbohydrate Name 1

Glycan Information

Name	Glycan Sequence
Glycan Sequence	Gal(b1-4)GlcNAc(b1-2)Man(a1-3)[Gal(b1-4)GlcNAc(b1-2)Man(a1-6)]Man(b1-4)GlcNAc(b1-4)GlcNAc
Composition	(Gal)2 (GlcNAc)4 (Man)3 (Asn)1
Glycan Mass	1640.6 (Mass of 'Asn' excluded)
Carbohydrate Mass	1640.6
Precursor Ion m/z	1784.64
Class	Glycoprotein, N-Glycan
Reaction	
Pathway	
Enzyme	

# SIEVE Identify

Two Internal Identification Algorithms:

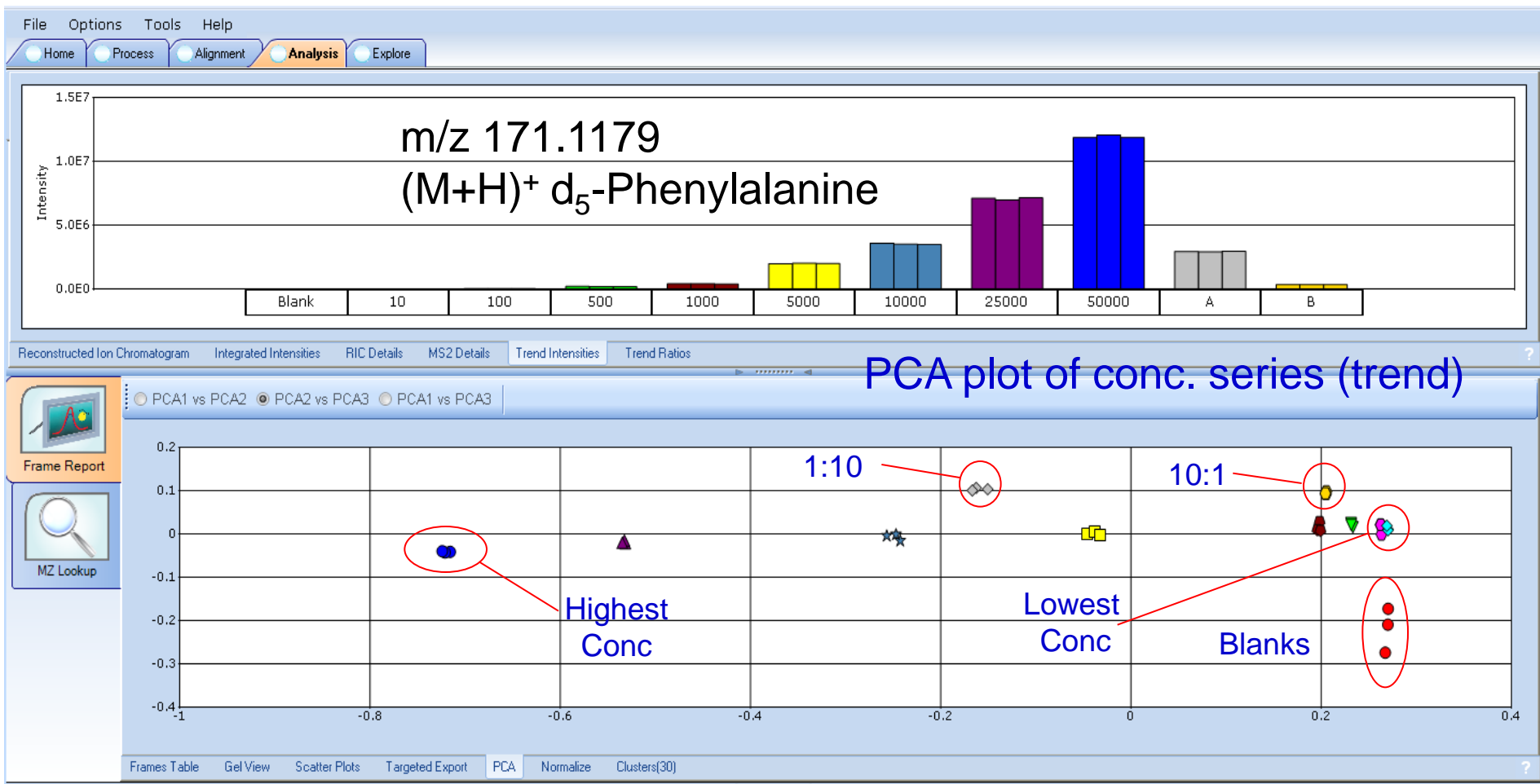
- Proteomics: SEQUEST + Protein Report
- Interoperable with Proteome Discoverer
- Small Molecule: ChemSpider

ChemSpider:

- Open web service API
- Loosely managed database collection
- Accurate mass → ID Candidates



# Metabolomics Workflow with SIEVE 2.0



# ProSightPC “Error Tolerant Search”

- ProSightPC is primarily for top-down protein identification, but it can be used for bottom-up and middle-down protein identification as well
- ProSightPC has some unusual and powerful features :
  - Xtract/THRASH deconvolution:
    - All peaks are converted from  $m/z$  to mass lists
  - Multiplexed identification:
    - Able to identify multiple species from a single MS/MS spectrum
  - Error tolerant search
    - Use wide precursor tolerance (>500 Da) while using narrow fragment tolerance (10 ppm)
    - Can identify any type of modification, including sequence truncations, post-translational modifications, single nucleotide polymorphisms
  - Annotated modifications in UniProt are indexed into the protein databases, automatically searching previously known modifications

# Error tolerant PTM identification using ProSightHT

- ProSightHT is an iterative search method for maximizing identifications in a dataset while maximizing the confidence in the results:

**High Throughput Wizard**

Running Highthroughput Logic  
Select a repository to load results to, and select/create a search tree with the desired search parameters

Repository: Ecoli\_Mods | New Repository | Search Tree Name: Ecoli\_DeltaM | Save

Experiment Filter

Min # fragments: 6

Max # fragments: 500

Min Intact Mass: 750

Monoisotopic

**Level 1 Search**  
Add Search: Ecoli Biomarker  
Conditions  
Success: load | Failure: run search  
Category: good

**Level 2 Search**  
Add Search: Ecoli Delta M  
Conditions  
Success: load | Failure: load  
Category: good | bad

Search 1: Accurate mass (10 ppm) precursor tolerance

If no confident match, run search 2

Search 2: Error tolerant "Delta M" search with 300 Da precursor tolerance

Cancel | < Prev | Next > | Finish

# Results of ProSightHT on O-GlcNAc HCD/ETD data

“Delta M”	Number of ID’s	Inferred Modification
0	2046	Exact matches
+57 Da	470	<b>Overalkylation (N-term or K)</b>
-48 Da	121	<b>Dethiomethyl (M)</b>
-17 Da	53	Pyroglutamic acid
+1 Da	42	Deamidation
+114 Da	30	<b>2x overalkylation</b>
+126 Da (125.90)	29	<b>Iodo-Y</b>
+22 Da	27	Sodium
+203 Da	27	<b>HexNAc</b>
+9 Da	19	?
+80 Da	19	<b>Phosphorylation</b>

- These PTM’s are identified without any prior knowledge of their existence



# ProSightPC

List of experiments

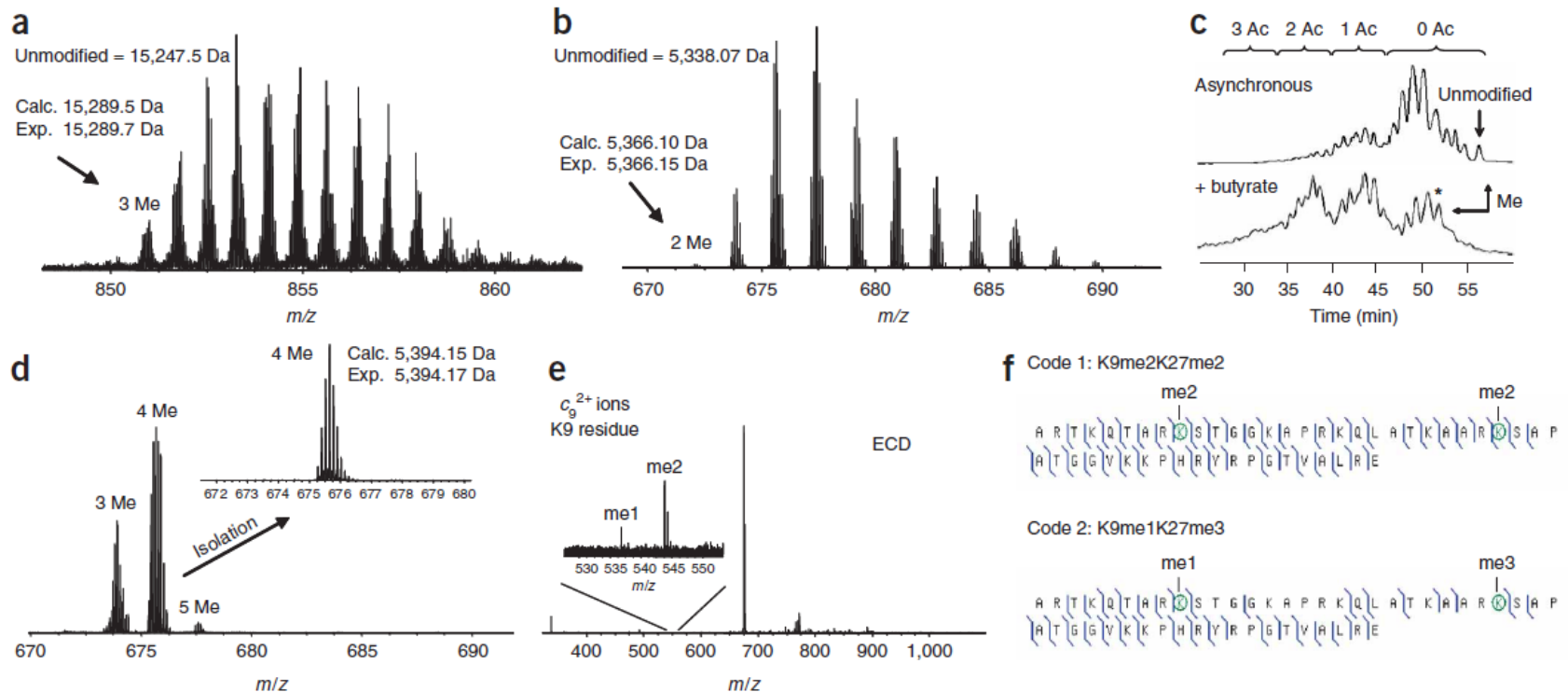
The screenshot displays the ProSightPC software interface. At the top, a menu bar includes File, Edit, View, Experiment Tools, Databases, ProSightHT, Tools, and Help. Below the menu is a toolbar with various icons. The main window is divided into several sections:

- List of experiments:** A table with columns: Exp ID, Search ID, Marked, Search Type, Pending Search, Best Expectatio..., Matching Forms, Largest Precursor (Mo..., and Exp Comment. It lists 10 experiments with IDs ranging from 935 to 7560.
- Search status:** A panel on the right with the text "Search status".
- Sequence Gazer™:** A central panel showing search parameters: Precursor Mass Type (Mono or Avg), Fragment Tolerance (10 Da ppm), Mass Type (Mono or Avg), and Δ m (On Off). It also displays Scores: P Score (1.38E-71), Expectation (4.86E-65), and PDE (155.0000). A note states: "NOTE: RED text denotes current selection".
- 51% Fragments Explained:** A large percentage displayed in a box with Rescore, Save, and Cancel buttons.
- Mass Difference:** A diagram showing a difference of 79.9688 Da (26132.2000 ppm) between an Observed mass of 3060.1640 and a Theoretical mass of 2980.2000.
- Sequence:** A protein sequence: b1 - T[D]N[A]G[D]Q[H-G-G-G[G-G]G[G]G[A]G[A]A[G]G[G]G - y11. A second sequence is shown below: b26 [G]G-E[N]Y-D-D[P-H-K - y1.
- Matching Fragments Table:** A table with columns: ID, Name, m/z, Monoisotopic Mass, Monoisotopic Theoretical Mass, Error (Da), Error (ppm), and Δ m. It lists 14 matching fragments.
- Threonine Information:** A panel on the right showing: Position: N:1 C:35, Amino Acid: T, RESID: none, Start PTM: None. Below this are PTM Choices: None (selected), Custom (0), and a list of PTMs including GPI-anchor, Dehydration, Riboflavin phosphorylation, Cysteinylolation, Sulfation, and Phosphorylation.

Results

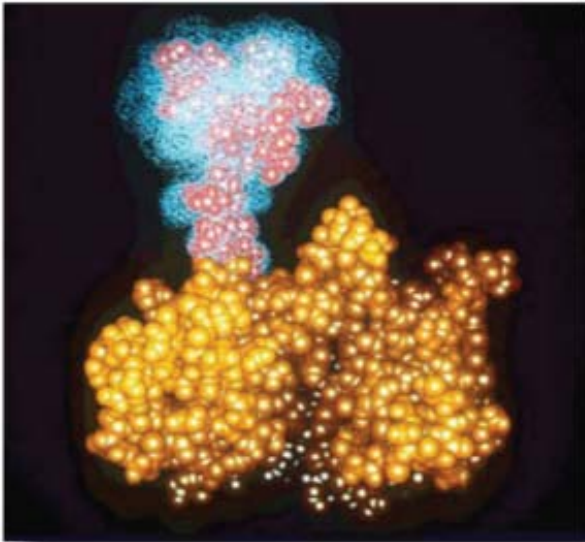
# ProSightPC Identification of Histone proteins

Figure 1 from Garcia *et al*, Nature Methods 2007



# Glycomics vs Glycoproteomics

- **Glycoproteomics** is the study of the profile of glycosylated proteins, i.e. the glycoproteome, in a biological system
- **Glycomics** is the study of glycome (repertoire of glycans)



- SimGlycan™ is for the **Glycomics Market**, focusing on released glycans