ThermoFisher SCIENTIFIC

Recent Developments in Thermo Scientific LSMS Proteomics Software

David Horn

Proteomics Software Strategic Marketing Manager

September 22, 2011

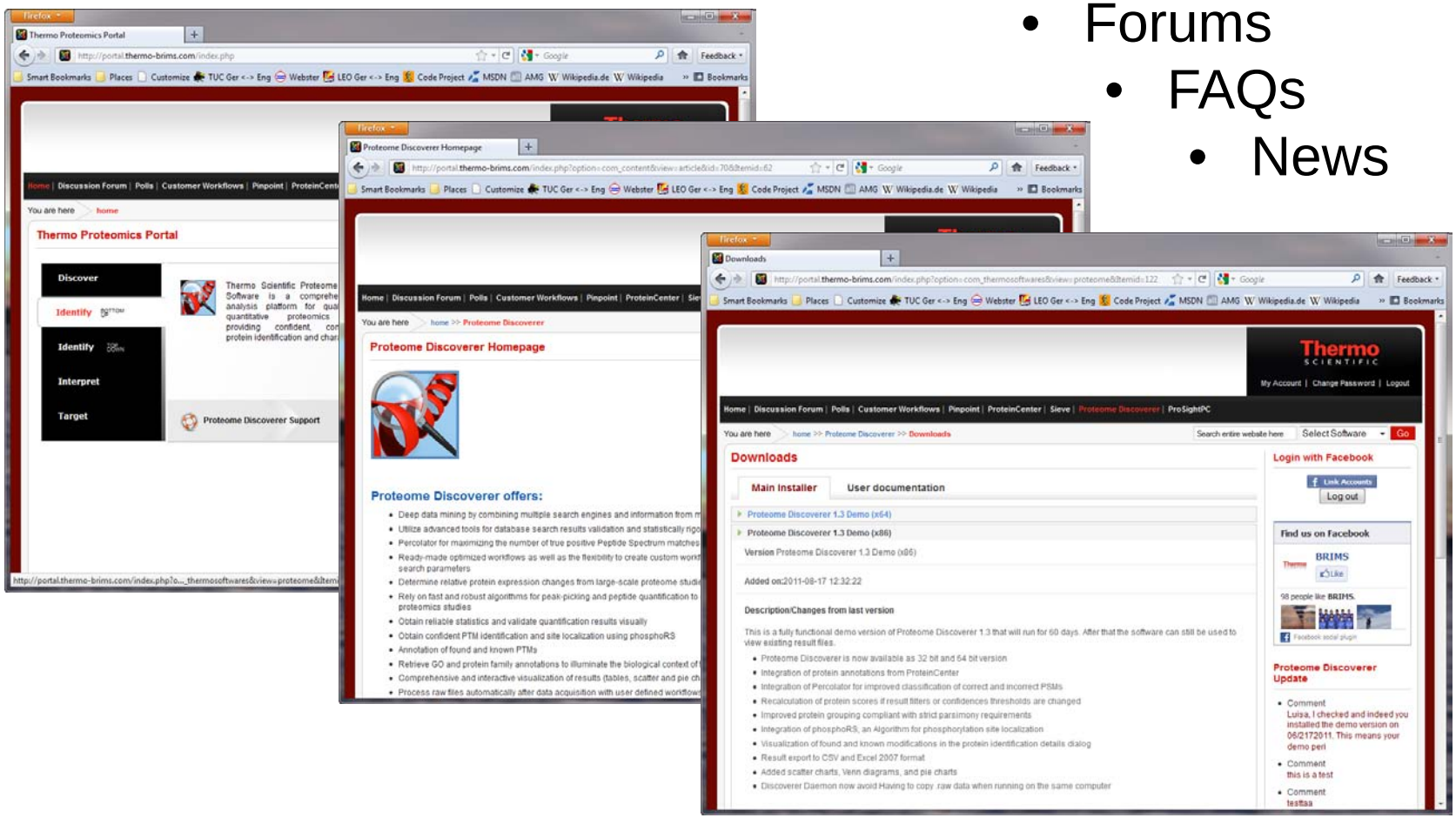
What's New In Proteomics Software For 2011?

- 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

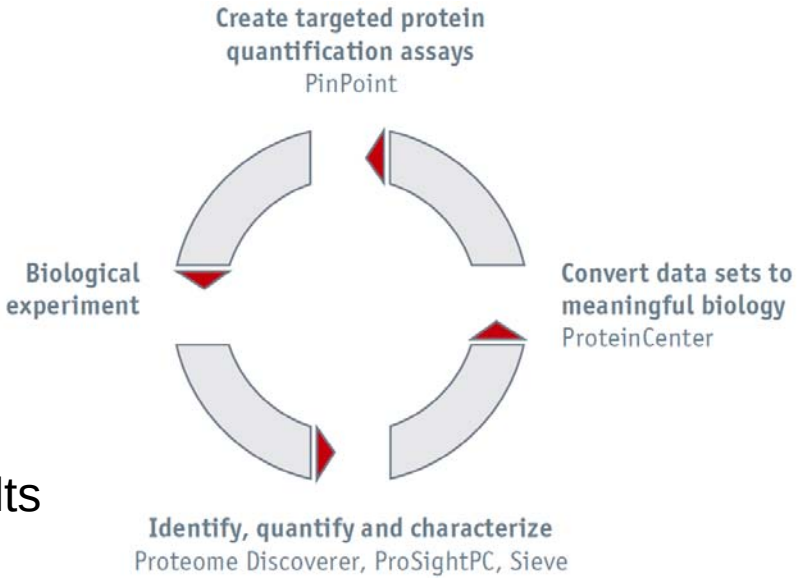


ThermoFisher SCIENTIFIC

ProteinCenter™

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

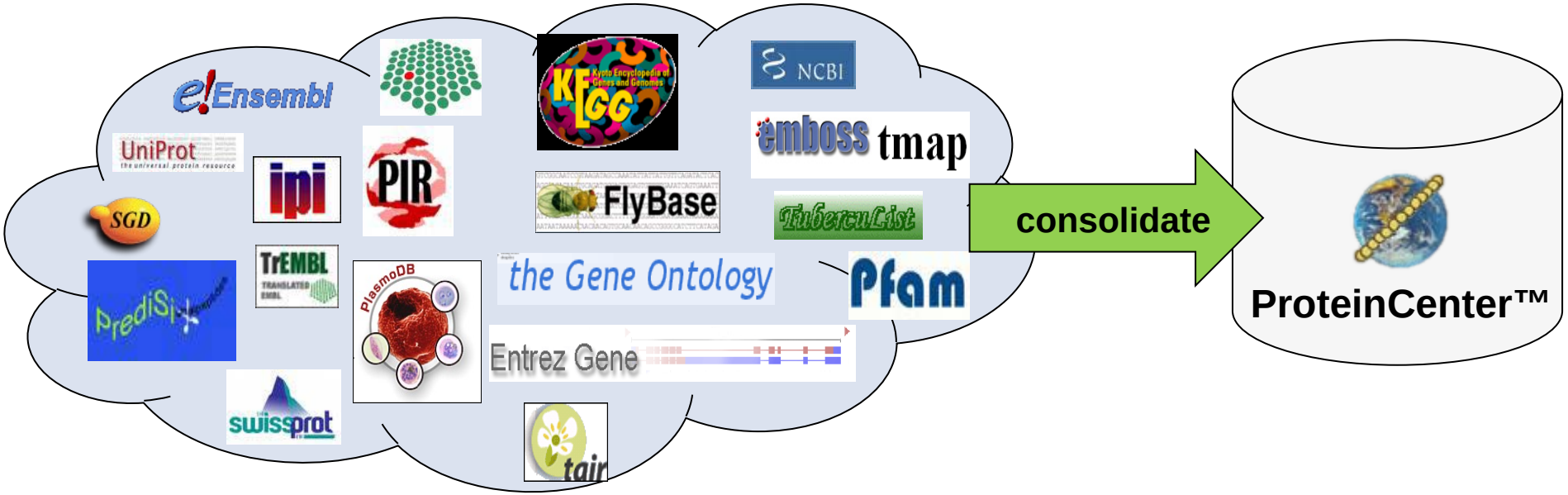
- **Key features**
  - Dataset comparison
  - Overrepresentation analysis (GO terms, Pfam, keywords, pathways, etc.)
  - Heat maps and profiling of quantitative proteomics data
  - KEGG pathway analysis
  - Comparison to previously published results



More information at:
http://www.proxeon.com/productrange/data_interpretation/introduction/index.html 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

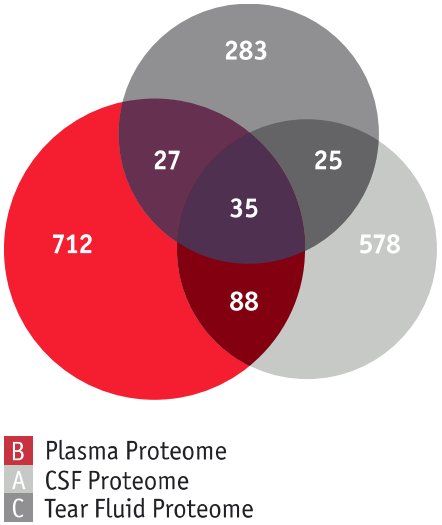


ThermoFisher SCIENTIFIC

Compare data sets

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

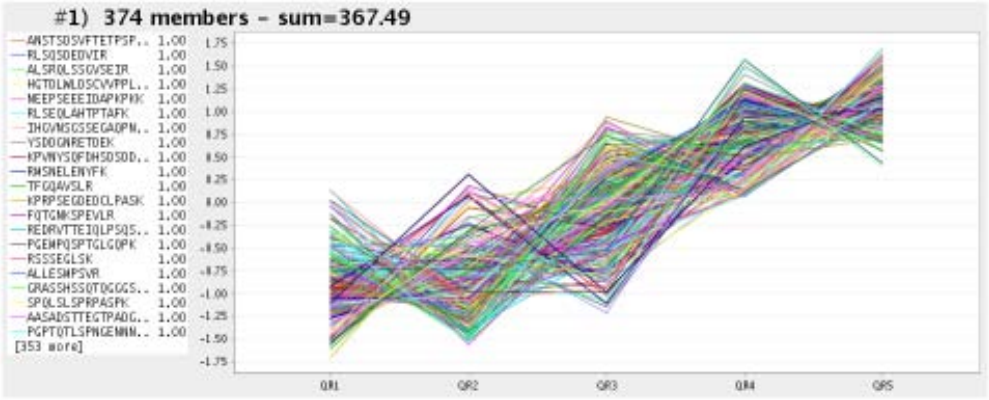
See in which data set each protein has been observed. Determine true overlaps.



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

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
MAPK signaling pathway (hsa)		6	0
ErbB signaling pathway (hsa)		5	0
Insulin signaling pathway (hsa)		7	1
Gap junction (hsa)		4	0
GnRH signaling pathway (hsa)		4	0
Acute myeloid leukemia (hsa)		4	0
Chronic myeloid leukemia (hsa)		4	0

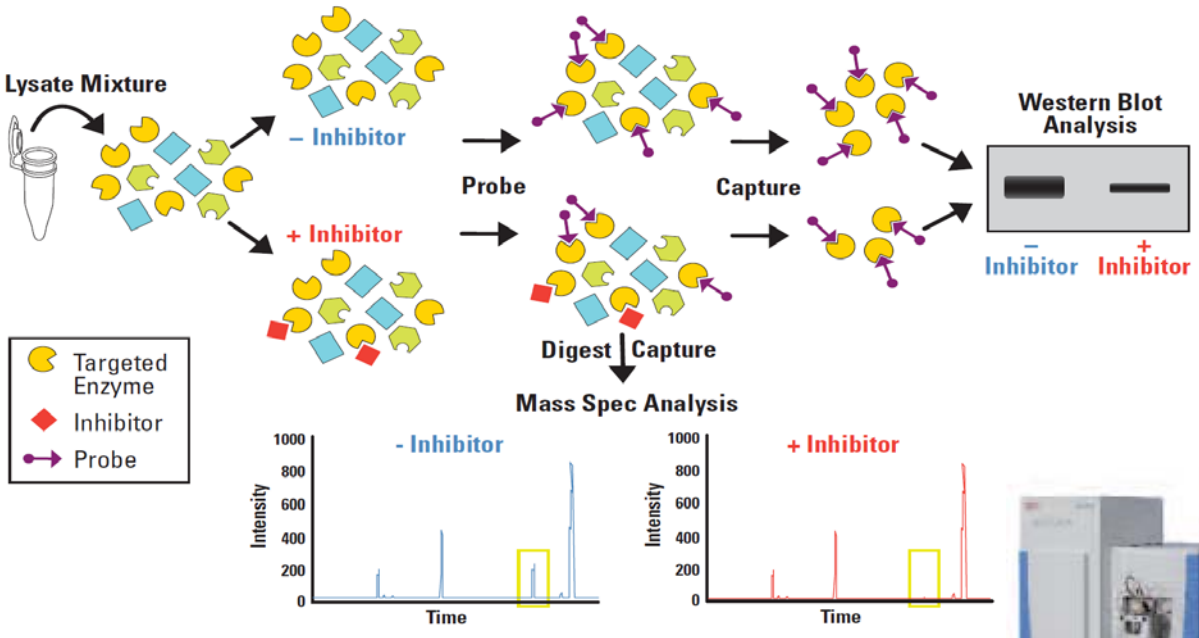
Proteome Discoverer + ProteinCenter + Pinpoint for kinase activity profiling

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

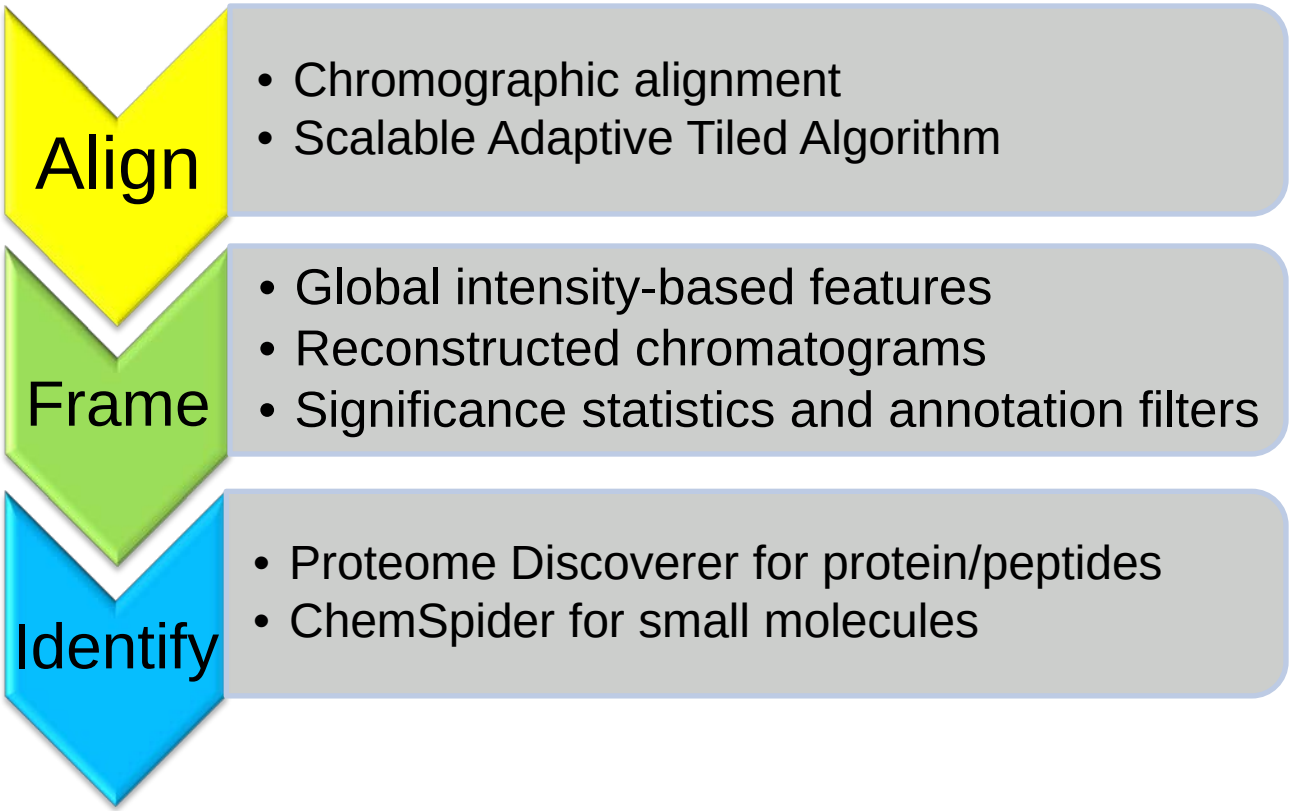
150 proteins, all kinases

ThermoFisher SCIENTIFIC

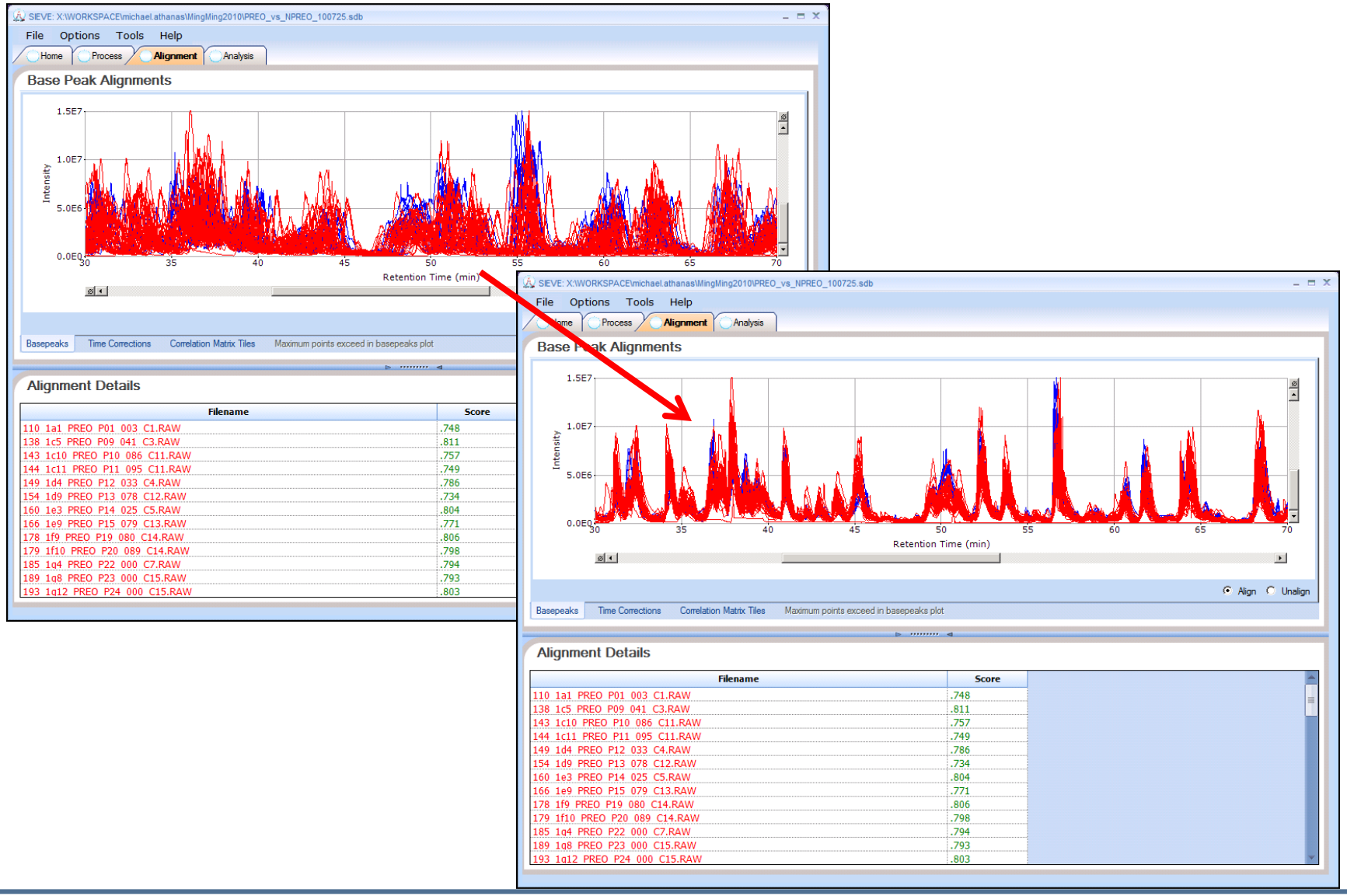
SIEVE

- **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

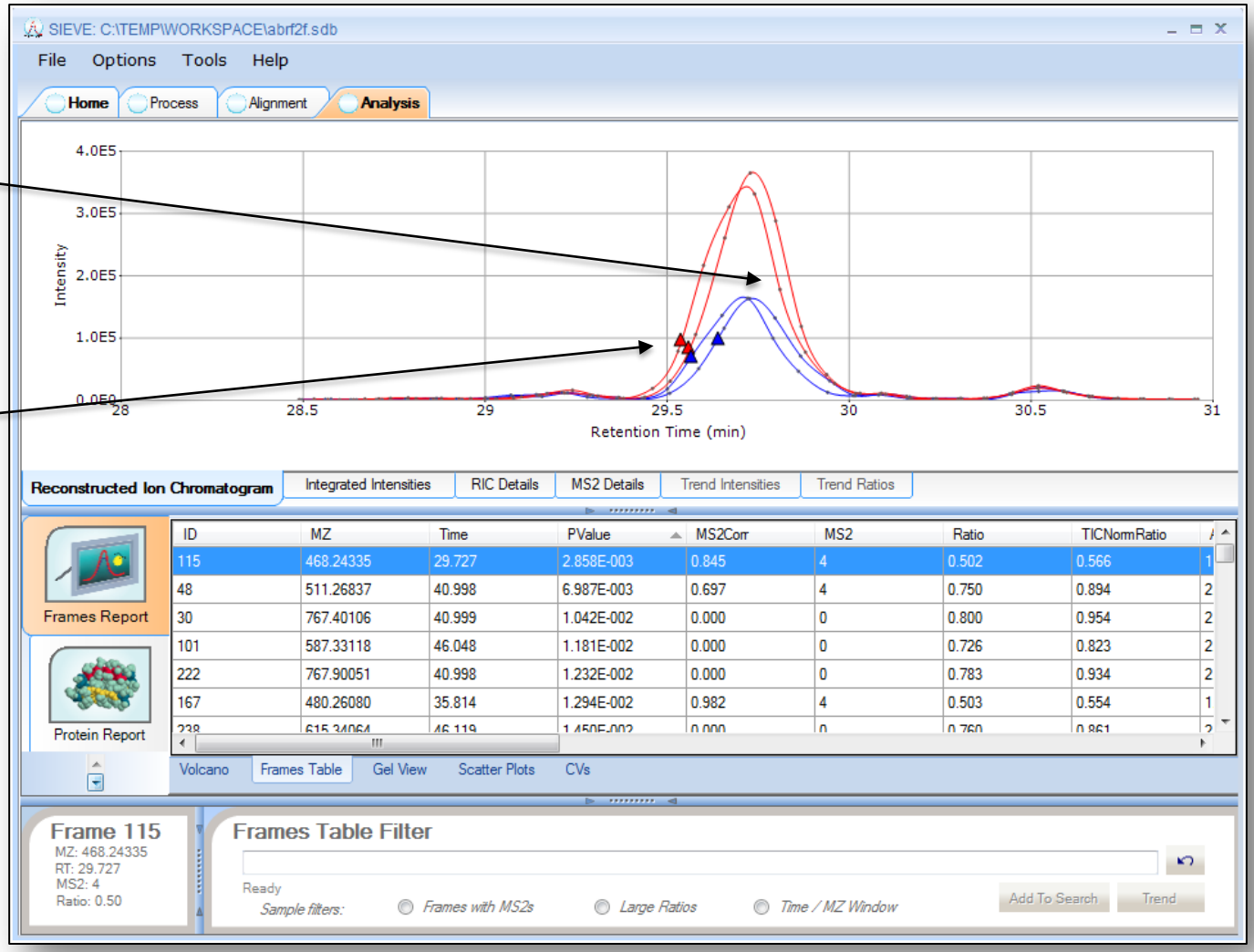


ThermoFisher SCIENTIFIC

Viewing frames in SIEVE

Peak shape based upon full scans.

Triangles represent MS2 scans.

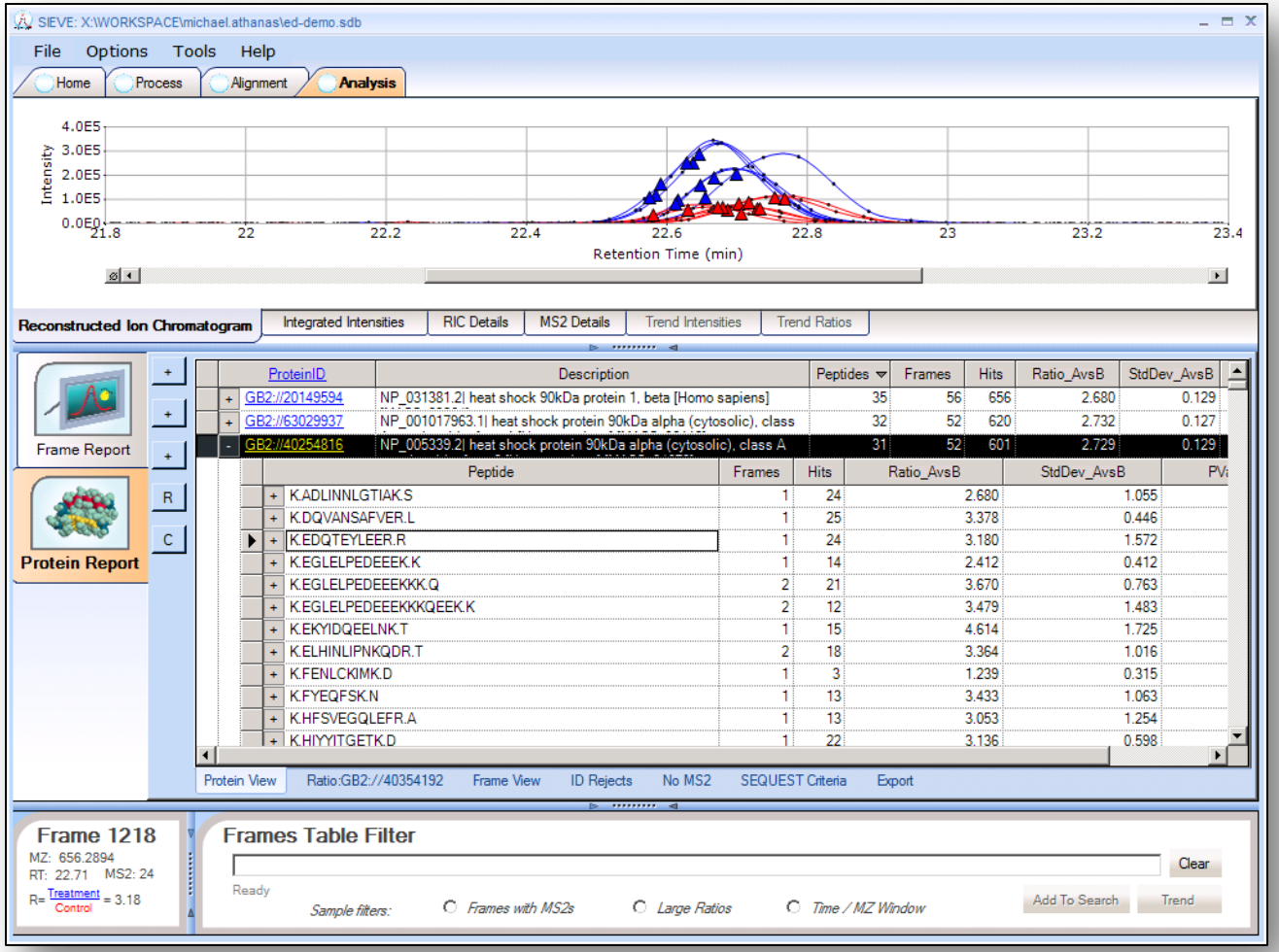


ThermoFisher SCIENTIFIC

SIEVE Protein Report

- 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?

Proteins→Peptides→

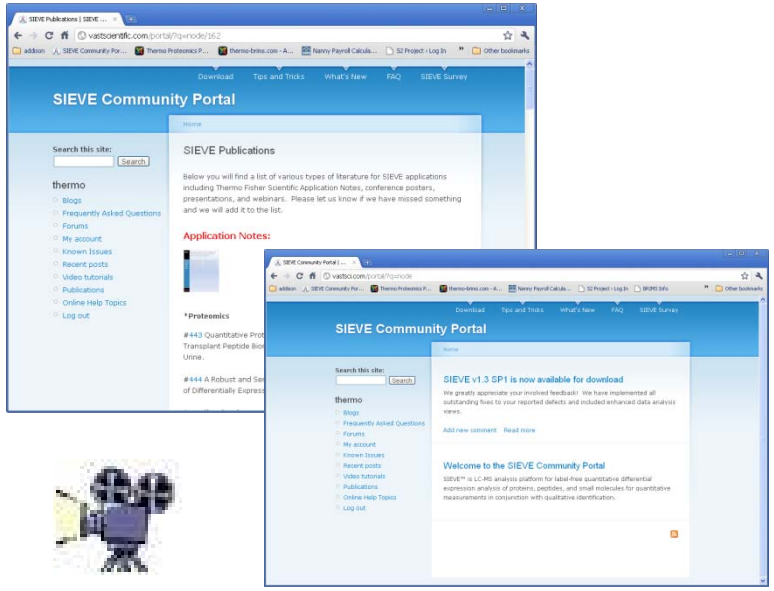


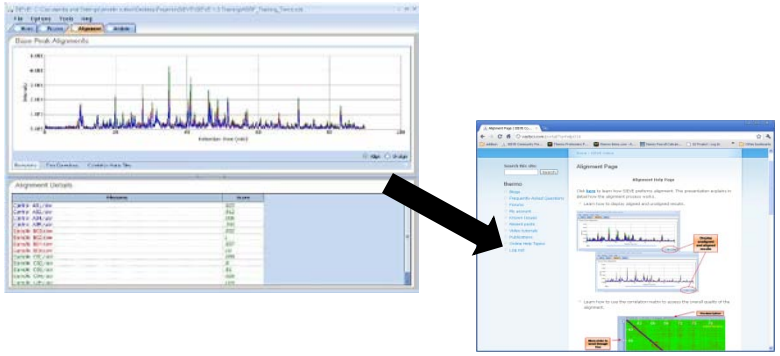
ThermoFisher SCIENTIFIC

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.





ThermoFisher SCIENTIFIC

SIEVE 2.0—Revolutionary Breakthrough for Metabolomics

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

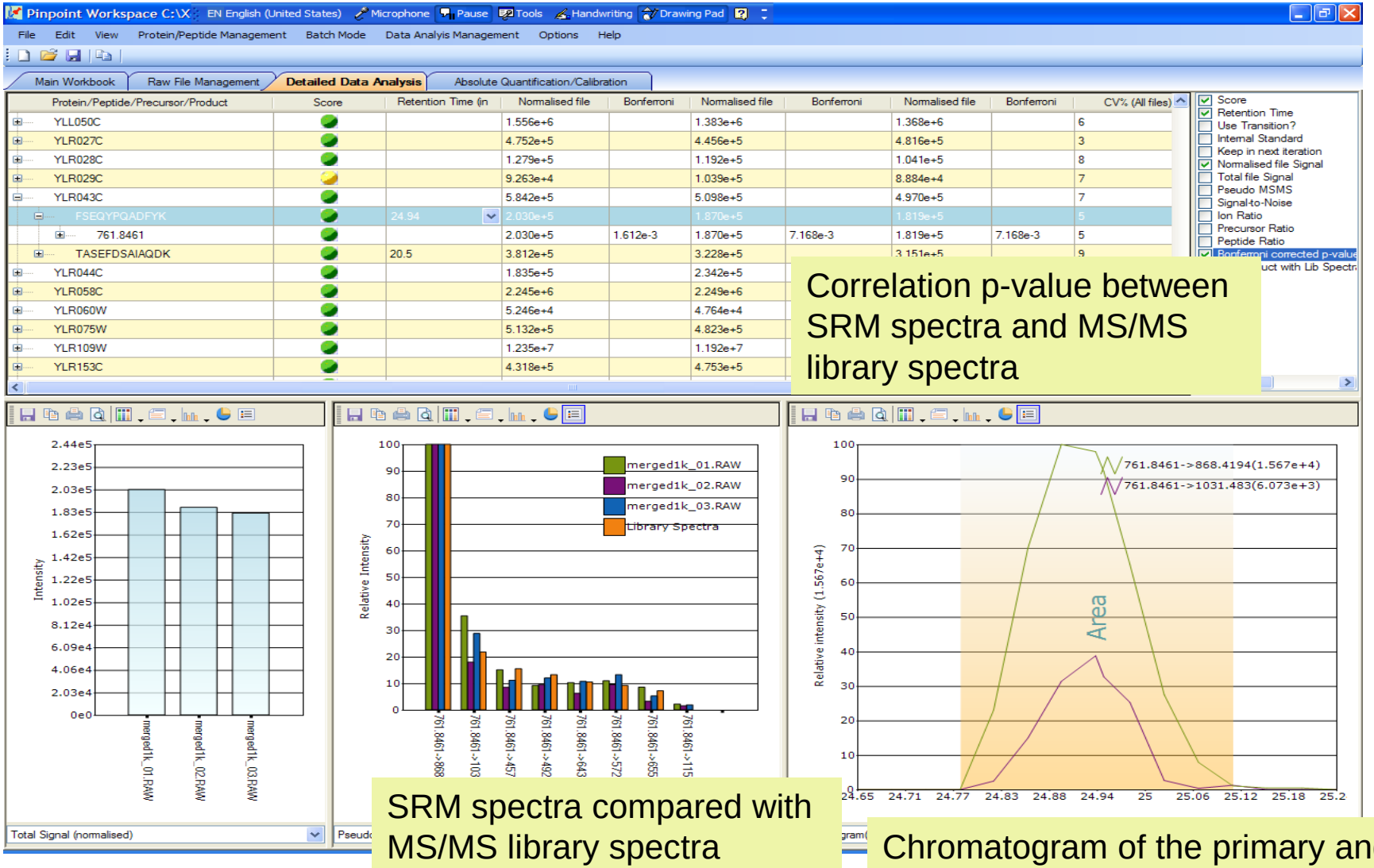
Features	SIEVE 1.3	SIEVE 2.0
Component Elucidation	Framing by Intensity	Chromatographic peak detection
Alignment	Aligns TIC	Aligns Components
Background correction to reduce false positives		✓
Grouping of adducts, dimers, isotopes		✓
Look for missing peaks		✓
Statistical tools	✓	✓
Graphical displays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

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



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

SRM spectra compared with MS/MS library spectra

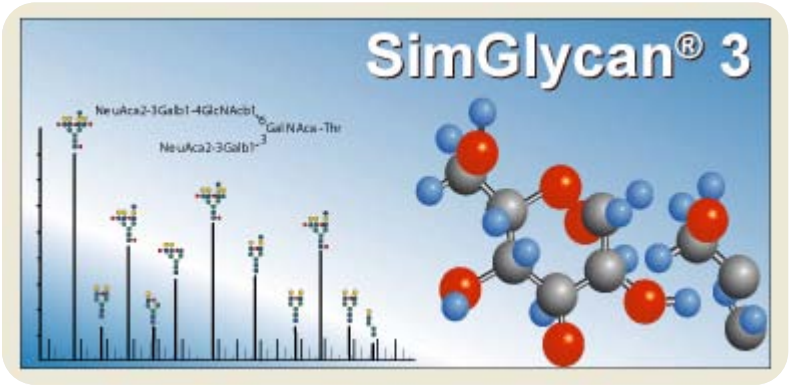
Chromatogram of the primary and secondary transitions. Area calculated based on primary

SimGlycan® 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^n data (Unique to ion traps)
  - Largest commercially-available glycan database (over 9500 structures)
  - Automatically process up to 1500 input spectra

- Available through Premier BioSoft

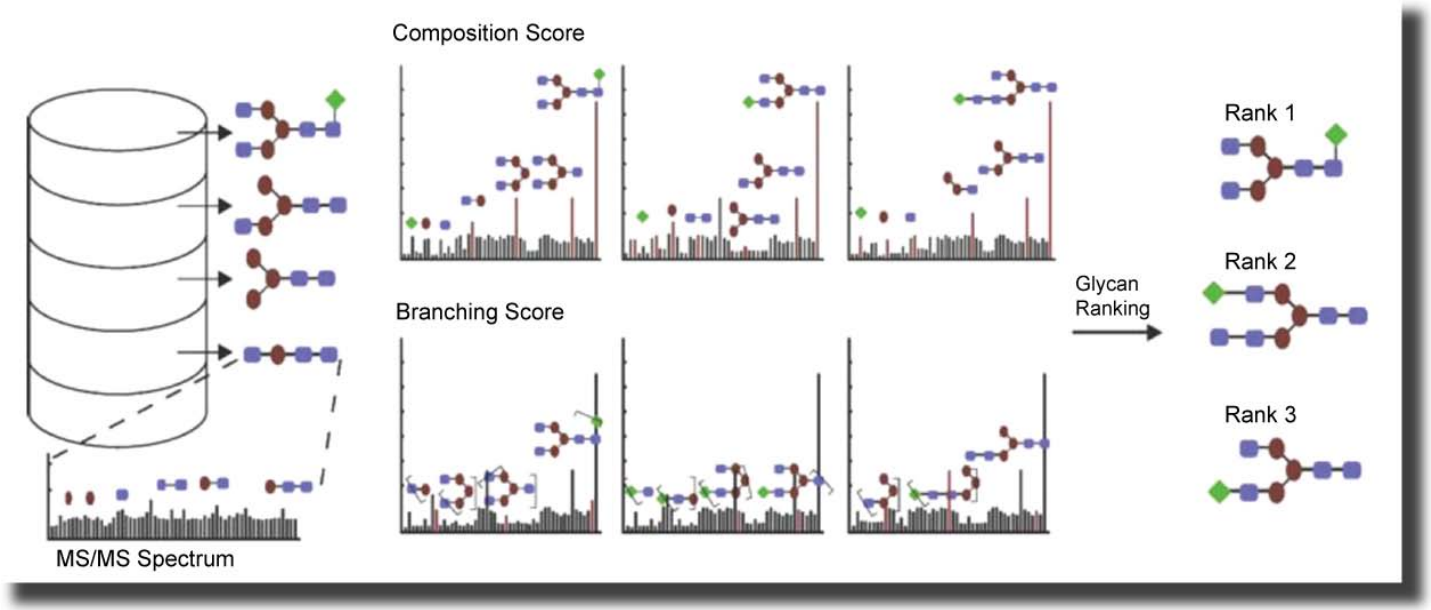


PREMIER Biosoft International

ThermoFisher SCIENTIFIC

How SimGlycan™ Works

- Database search approach



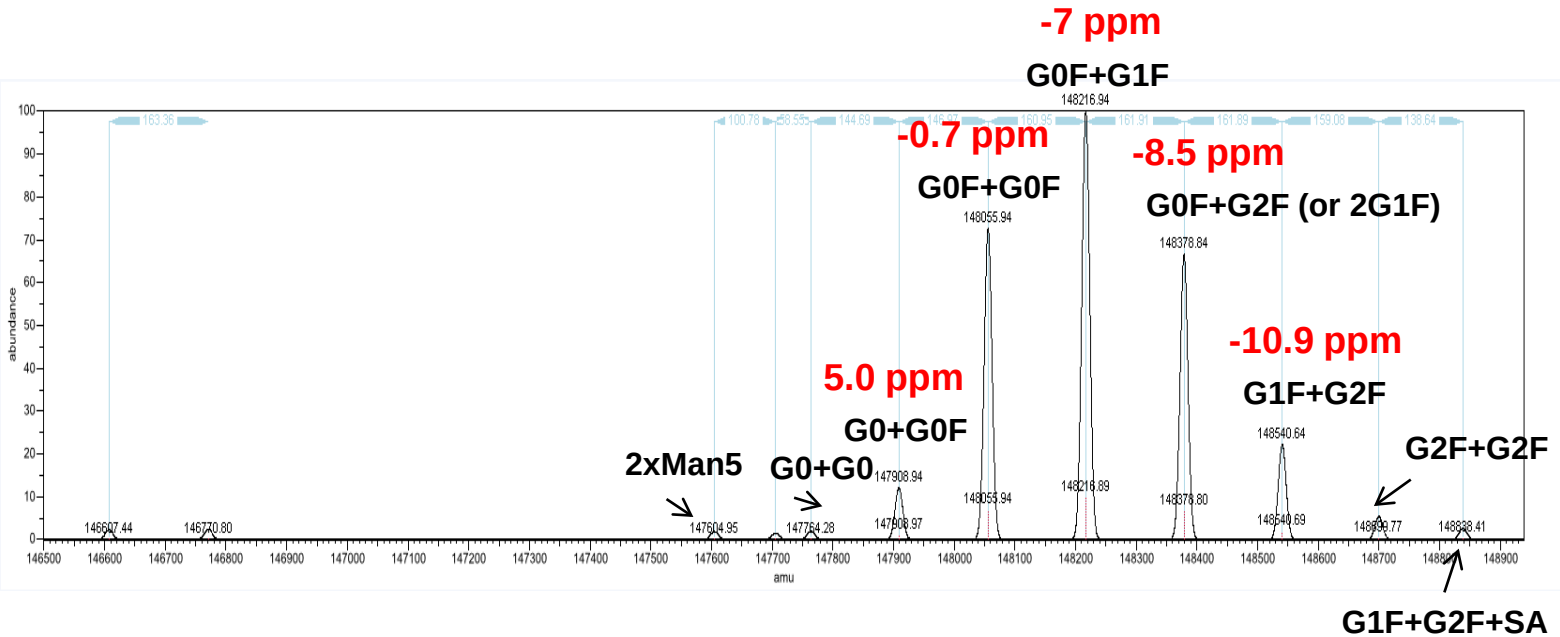
Methods Mol Biol. **2010**,600,269-81

- 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

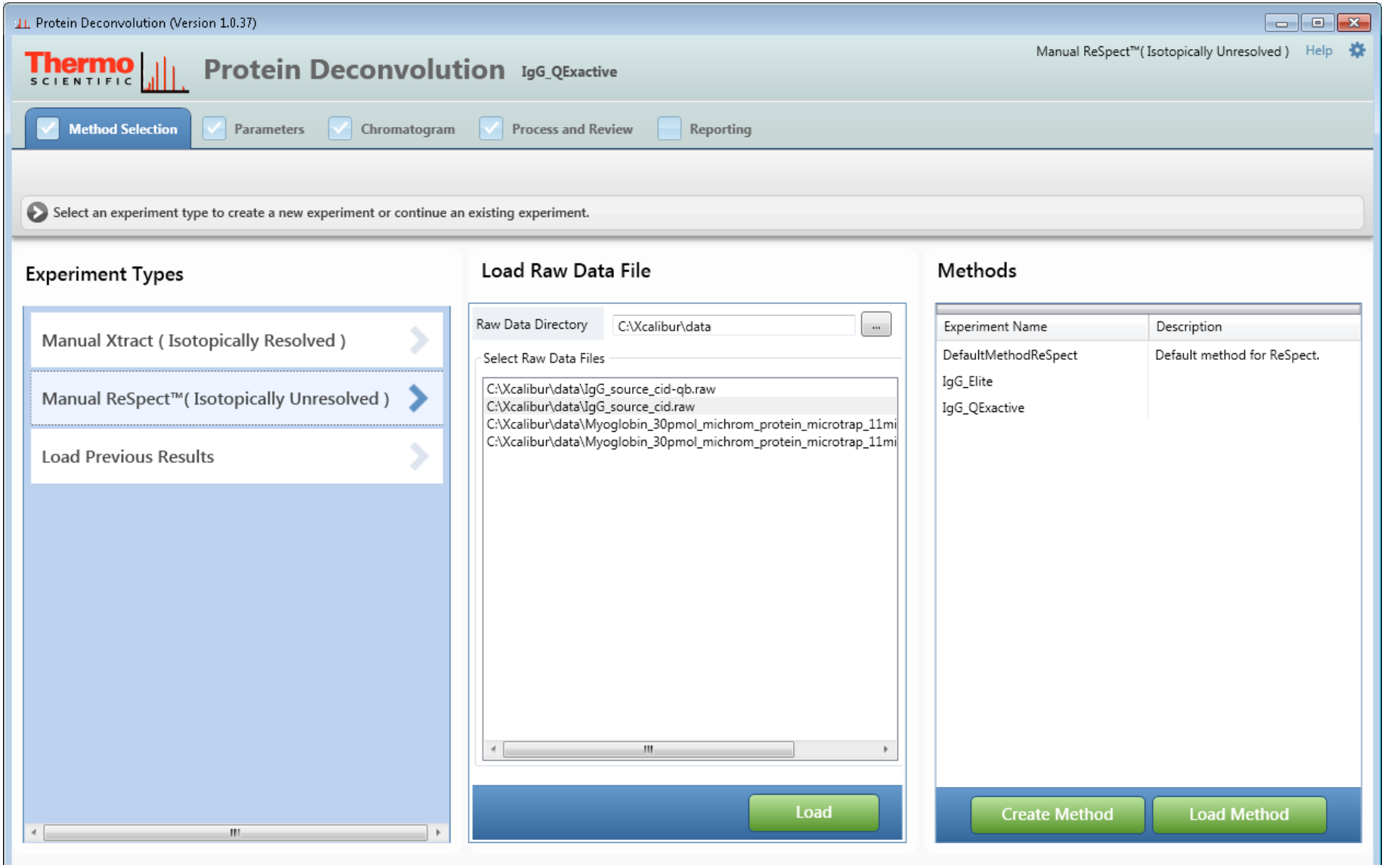
ReSpect™ Deconvolution of Q Exactive IgG Data



		ppm mass measurement errors				
RAW file	Q Exactive	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

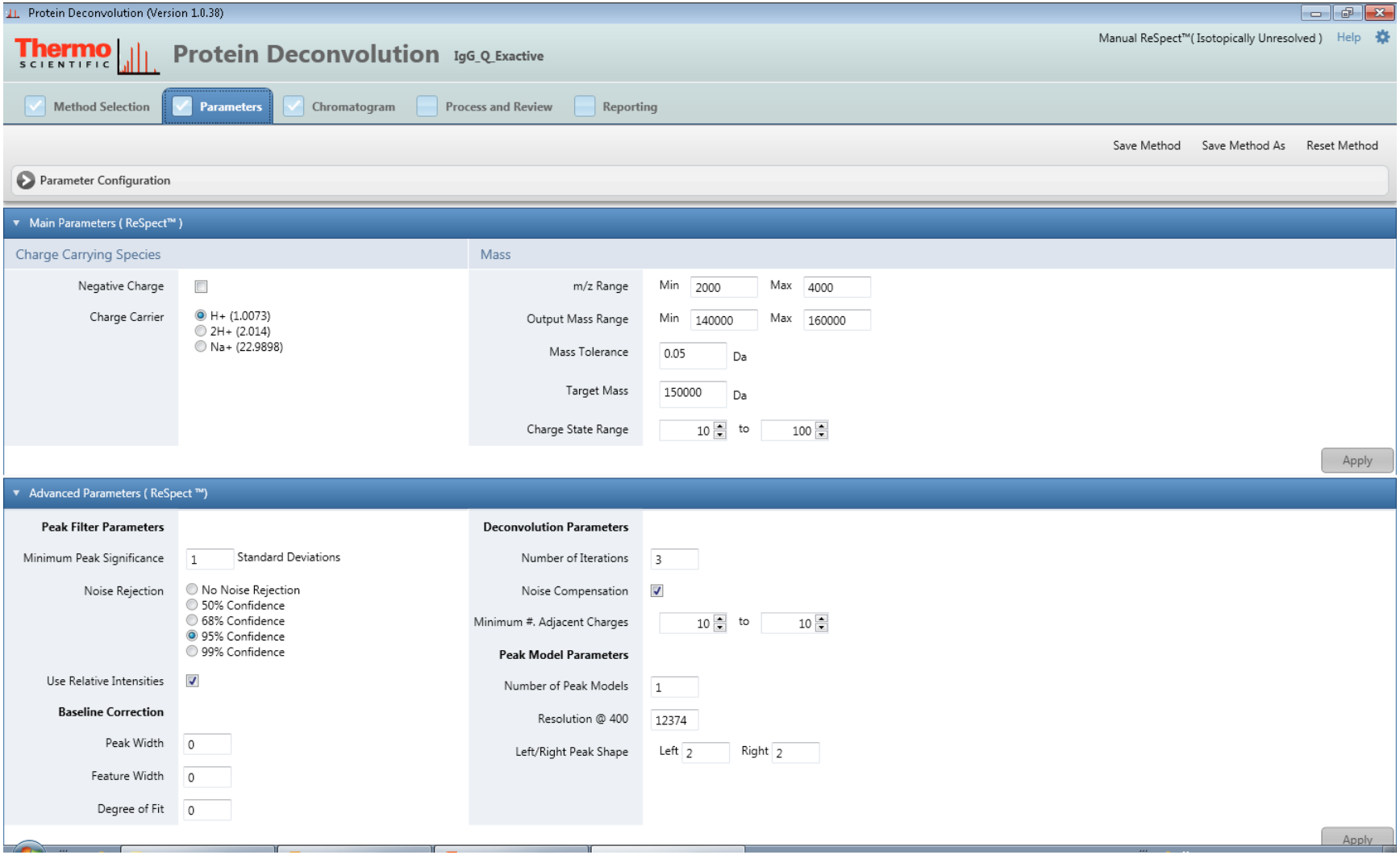
-6.9 +/- 6.4 ppm mass tolerance for above measurements

Protein Deconvolution 1.0



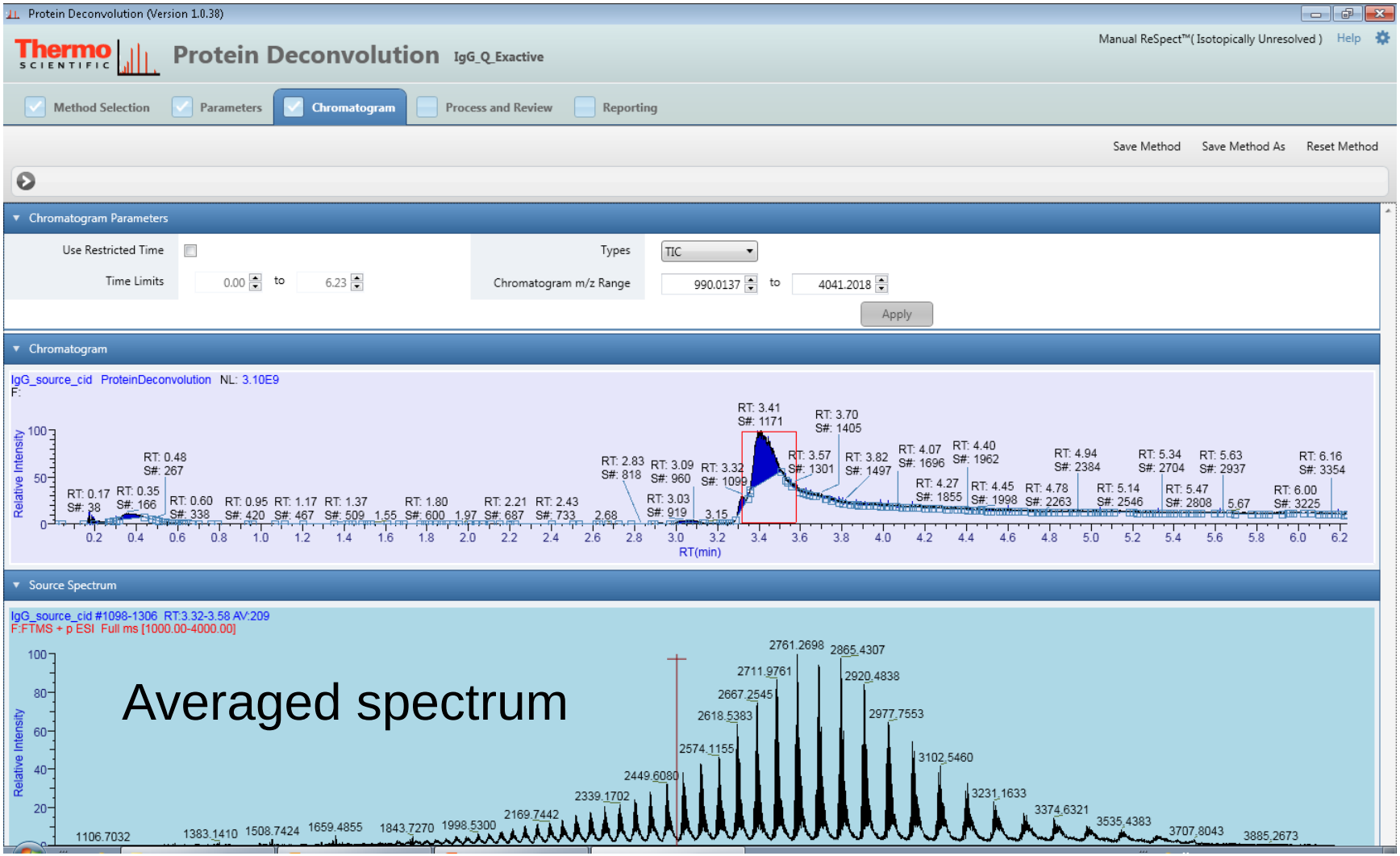
ThermoFisher SCIENTIFIC

Protein Deconvolution Method Parameters



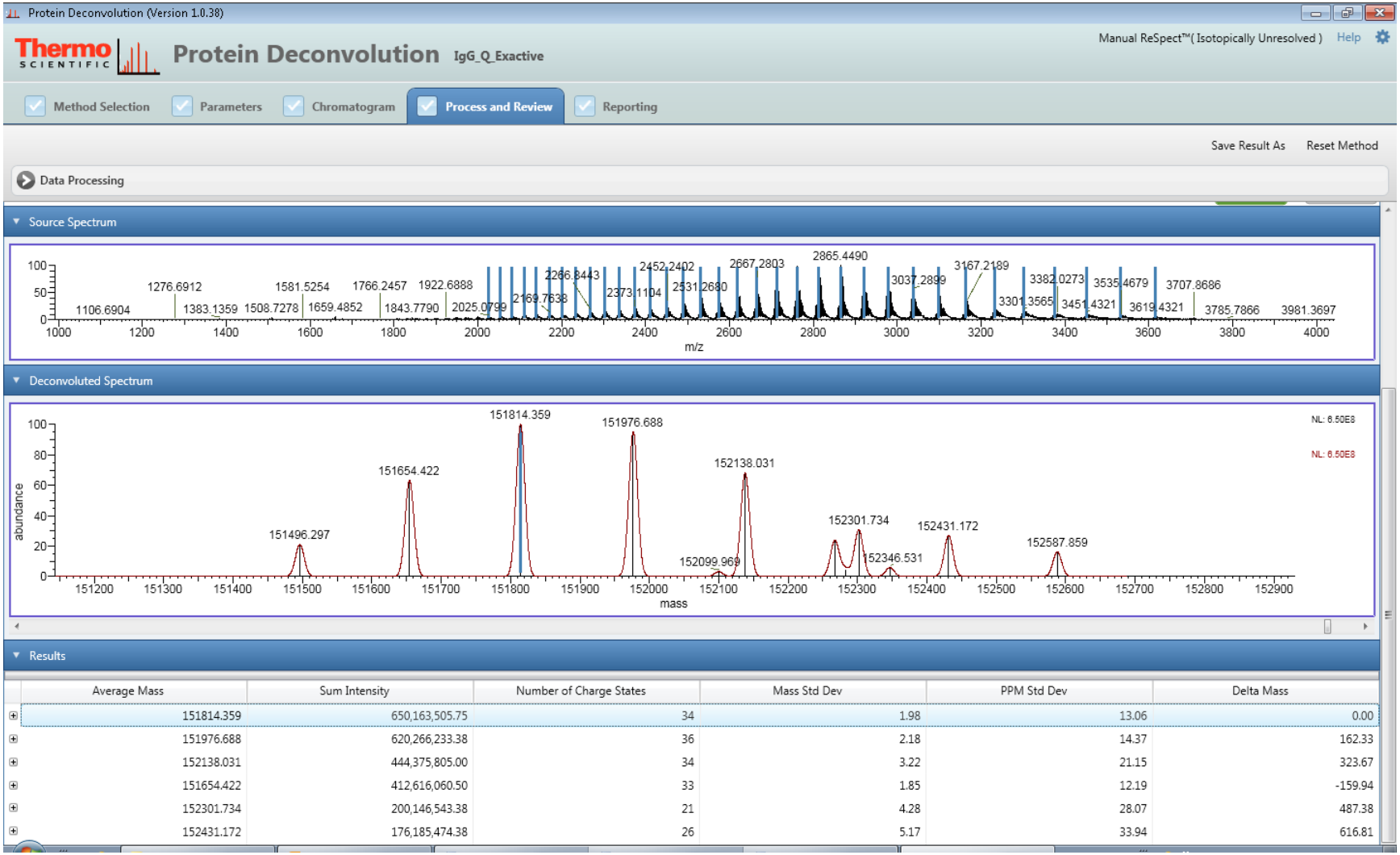
ThermoFisher SCIENTIFIC

Protein Deconvolution – Chromatogram Tab



ThermoFisher SCIENTIFIC

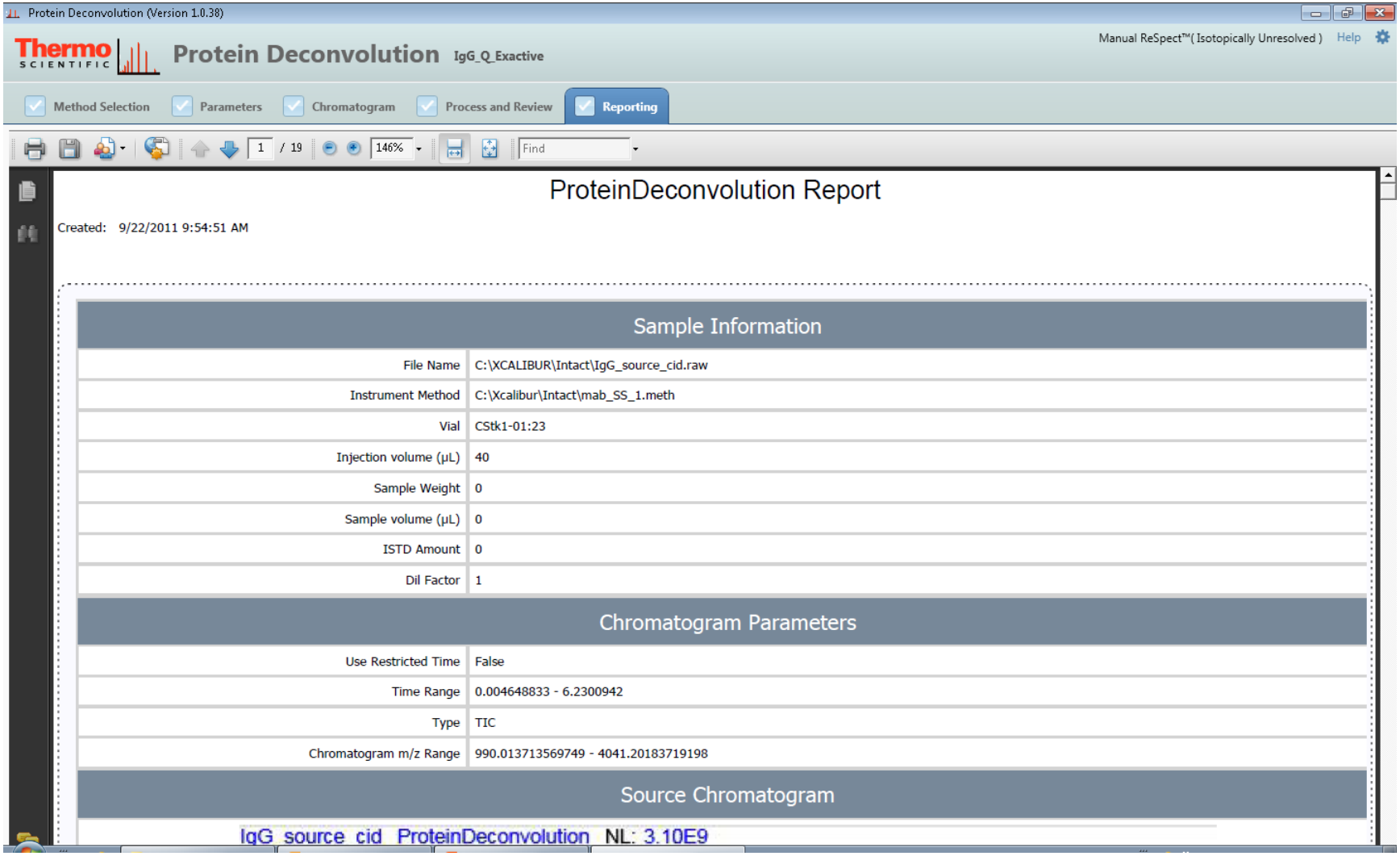
Protein Deconvolution - ReSpect™ Deconvolution



Average Mass	Sum Intensity	Number of Charge States	Mass Std Dev	PPM Std Dev	Delta Mass
151814.359	650,163,505.75	34	1.98	13.06	0.00
151976.688	620,266,233.38	36	2.18	14.37	162.33
152138.031	444,375,805.00	34	3.22	21.15	323.67
151654.422	412,616,060.50	33	1.85	12.19	-159.94
152301.734	200,146,543.38	21	4.28	28.07	487.38
152431.172	176,185,474.38	26	5.17	33.94	616.81

ThermoFisher SCIENTIFIC

Protein Deconvolution - Report



ThermoFisher SCIENTIFIC

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

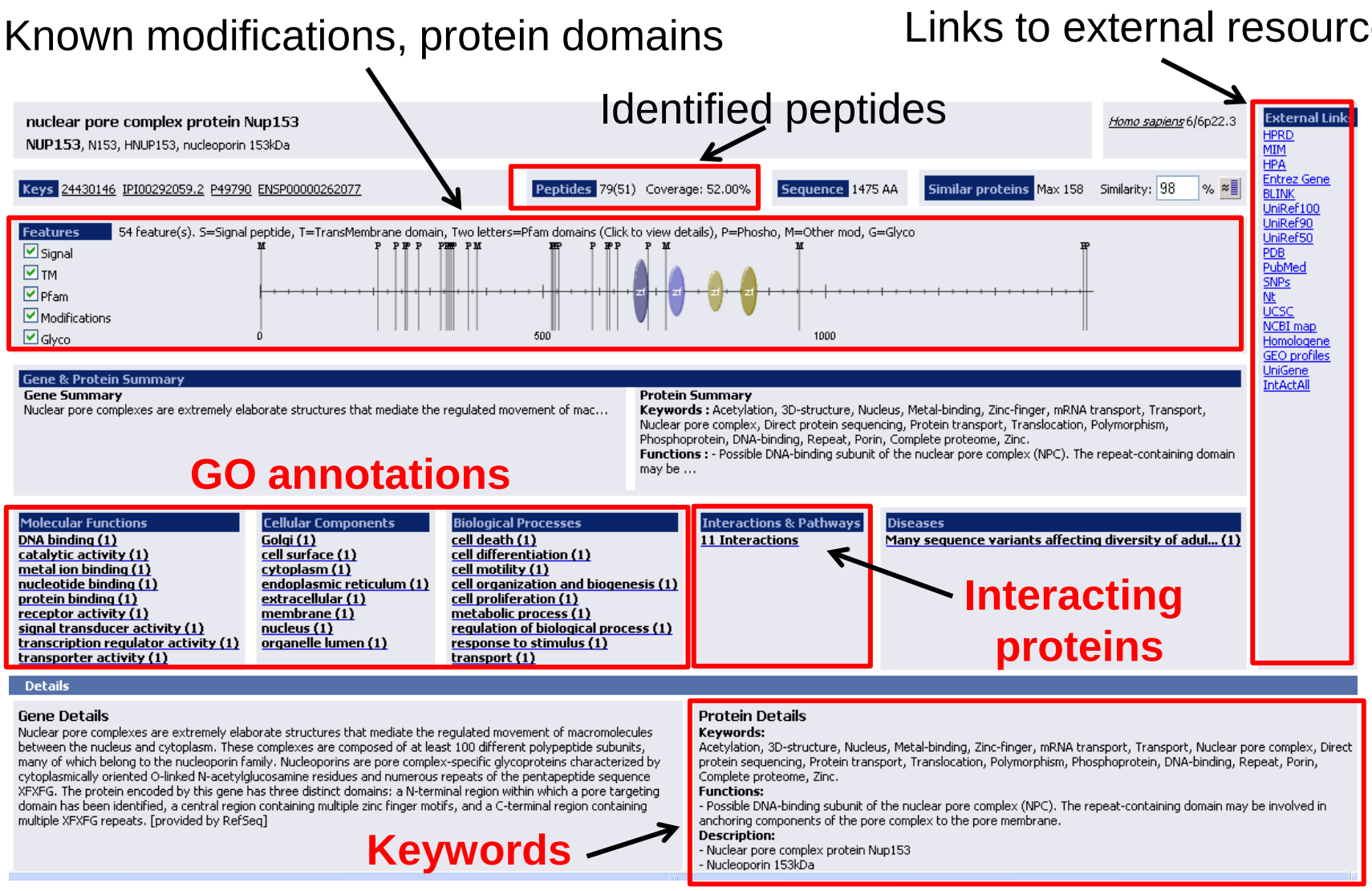
Conclusions

- 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 3rd 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.

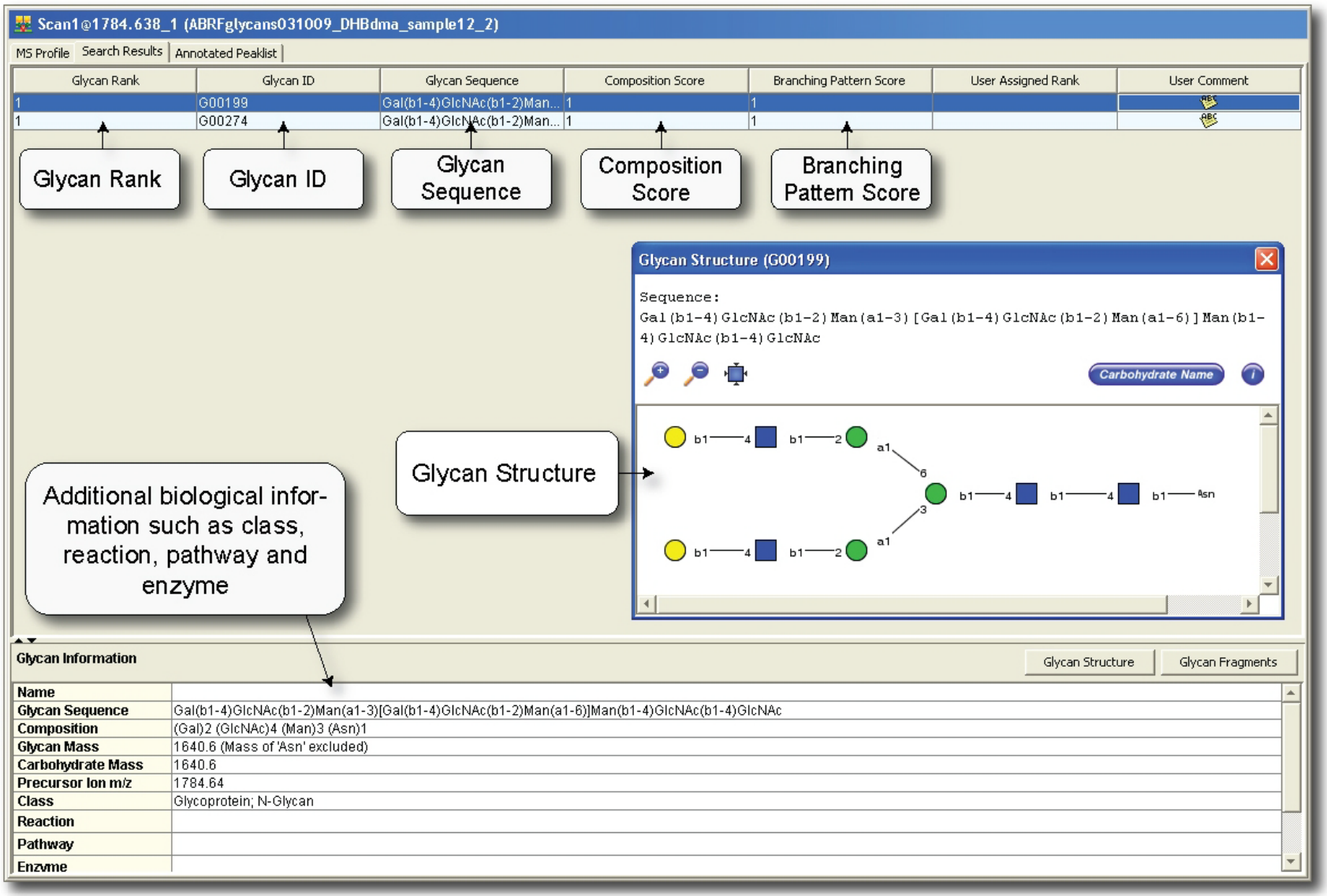
ThermoFisher SCIENTIFIC

Supplementary slides

ProteinCard highlights



SimGlycan™ Results Page



ThermoFisher SCIENTIFIC

SIEVE Identify

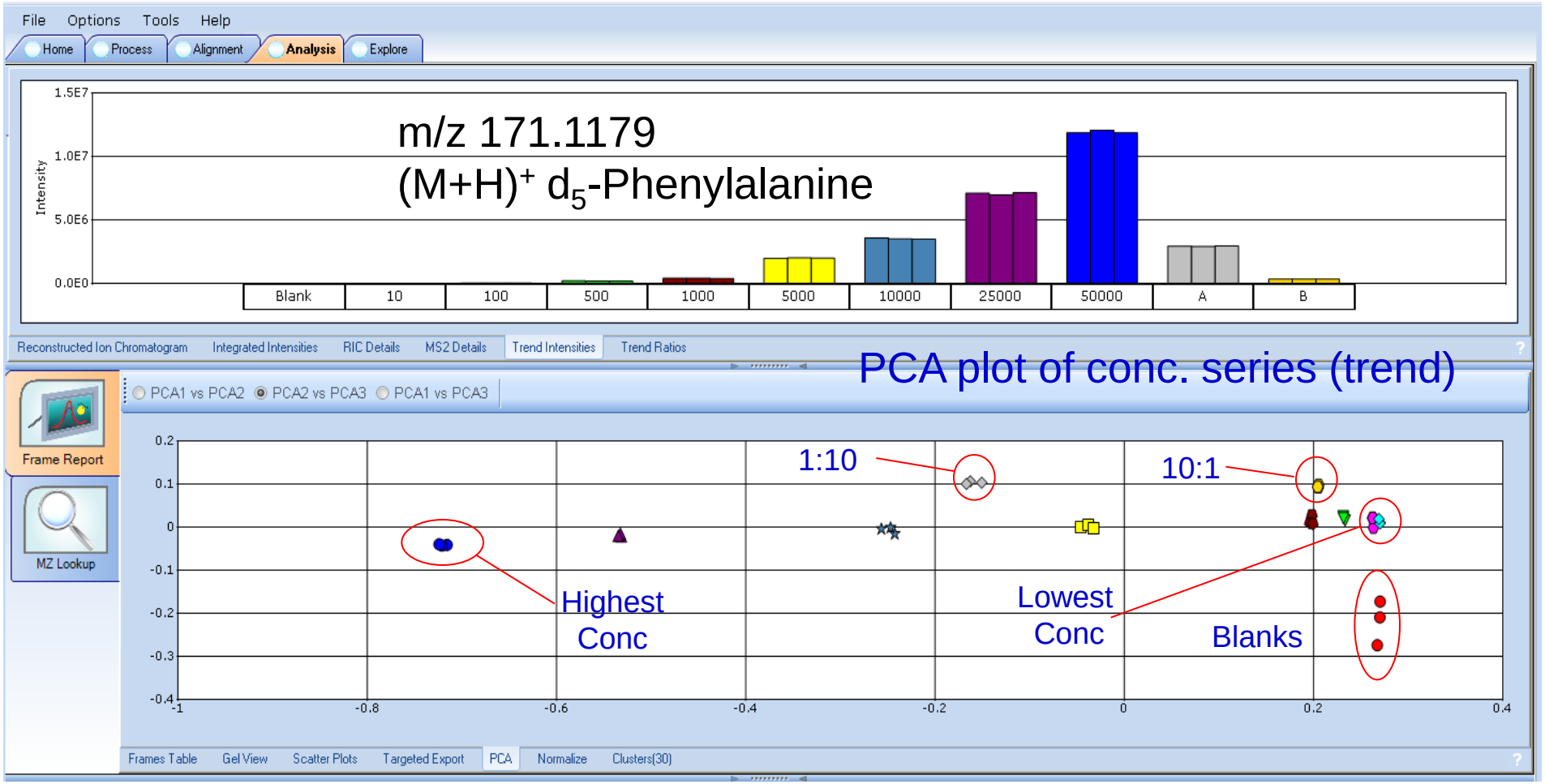
Two Internal Identification Algorithms:

- 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



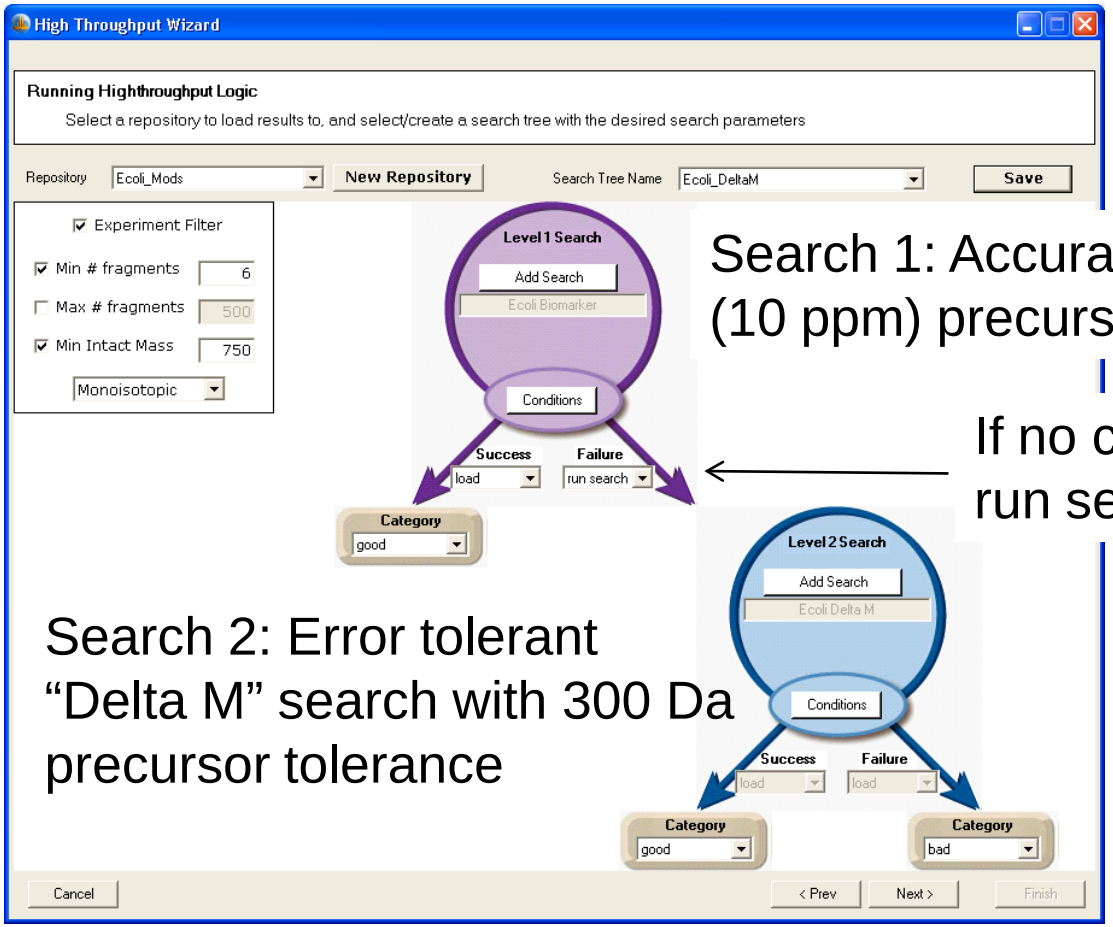
ThermoFisher SCIENTIFIC

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:



Results of ProSightHT on O-GlcNAc HCD/ETD data

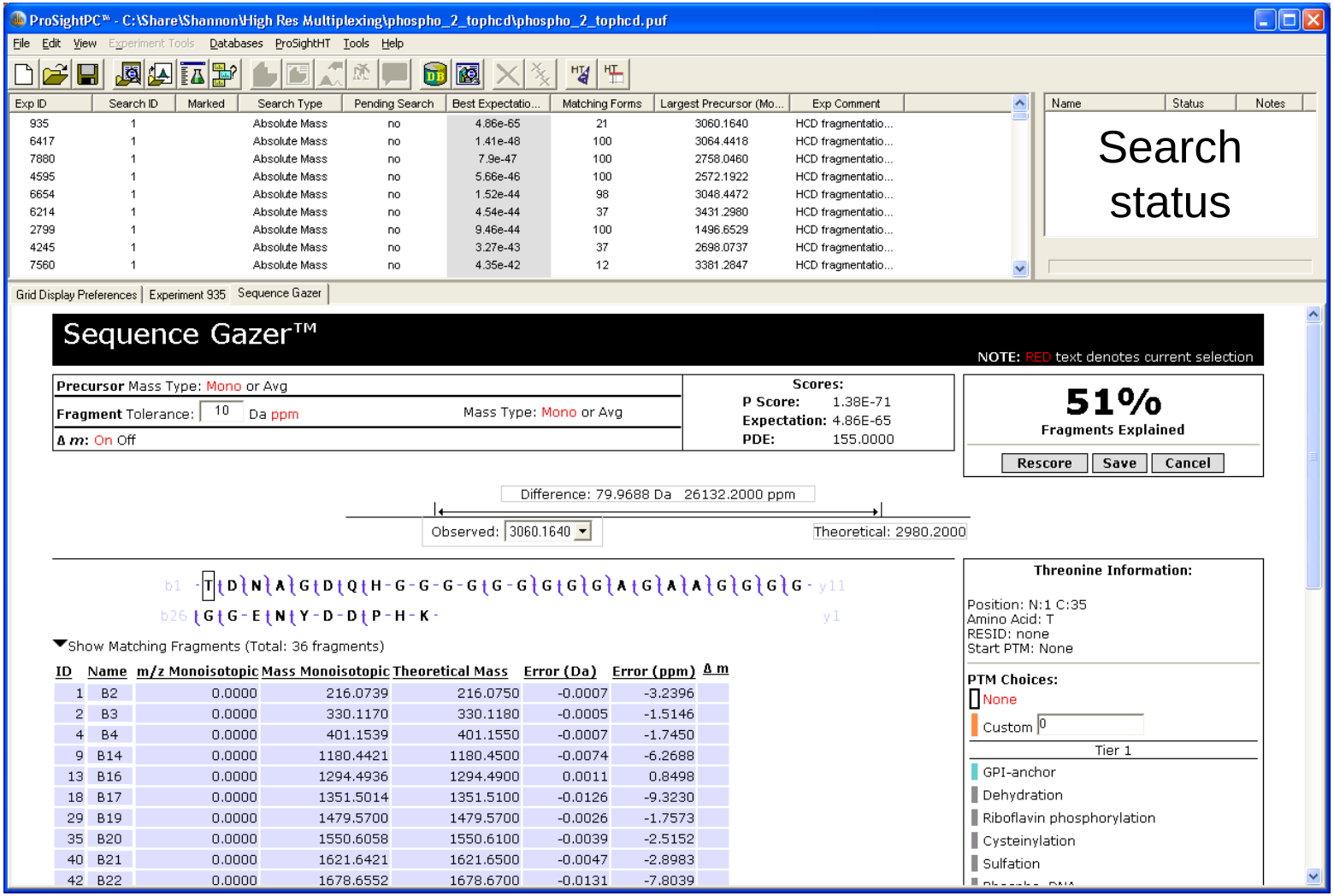
"Delta M"	Number of ID's	Inferred Modification
0	2046	Exact matches
+57 Da	470	**Overalkylation (N-term or K)**
-48 Da	121	**Dethiomethyl (M)**
-17 Da	53	Pyroglutamic acid
+1 Da	42	Deamidation
+114 Da	30	**2x overalkylation**
+126 Da (125.90)	29	**Iodo-Y**
+22 Da	27	Sodium
+203 Da	27	HexNAc
+9 Da	19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

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

ProSightPC

List of experiments

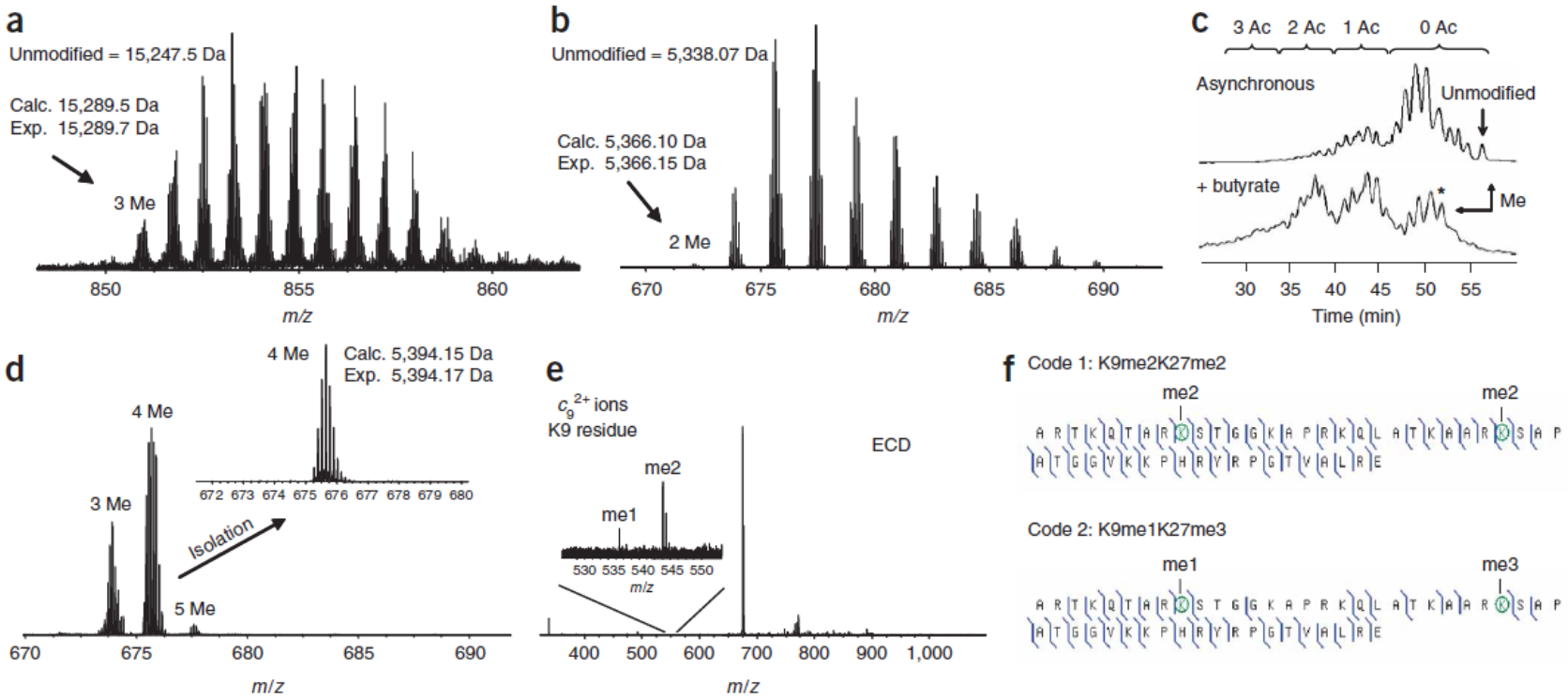
Results



ThermoFisher SCIENTIFIC

ProSightPC Identification of Histone proteins

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

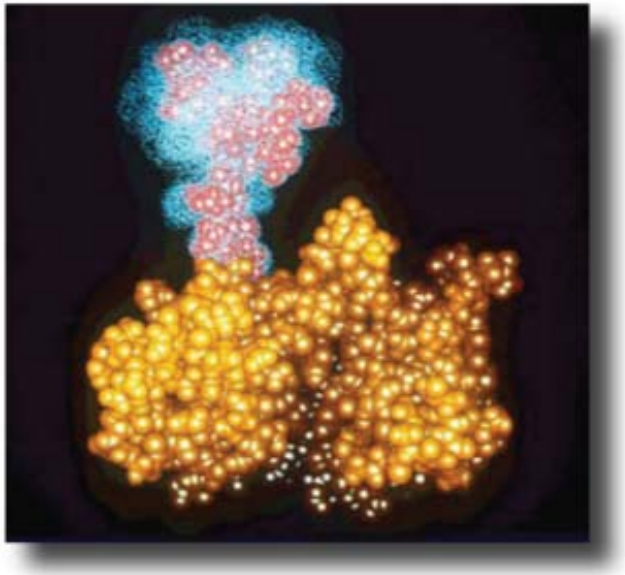


ThermoFisher SCIENTIFIC

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