ThermoFisher
SCIENTIFIC

Introduction in Proteomics

Overview

- Proteomics definition
- From genome to proteome
- Aminoacids and proteins
- What proteomics can do?
- Peptide fragmentation
- Proteomics and gel electrophoresis
- Post translational modifications
- DD NL MS3 for phosphorylation identification
- Intact protein analysis
- Lock mass
- Protein quantification
- RP-HPLC basics

Easy to remember

Proteins

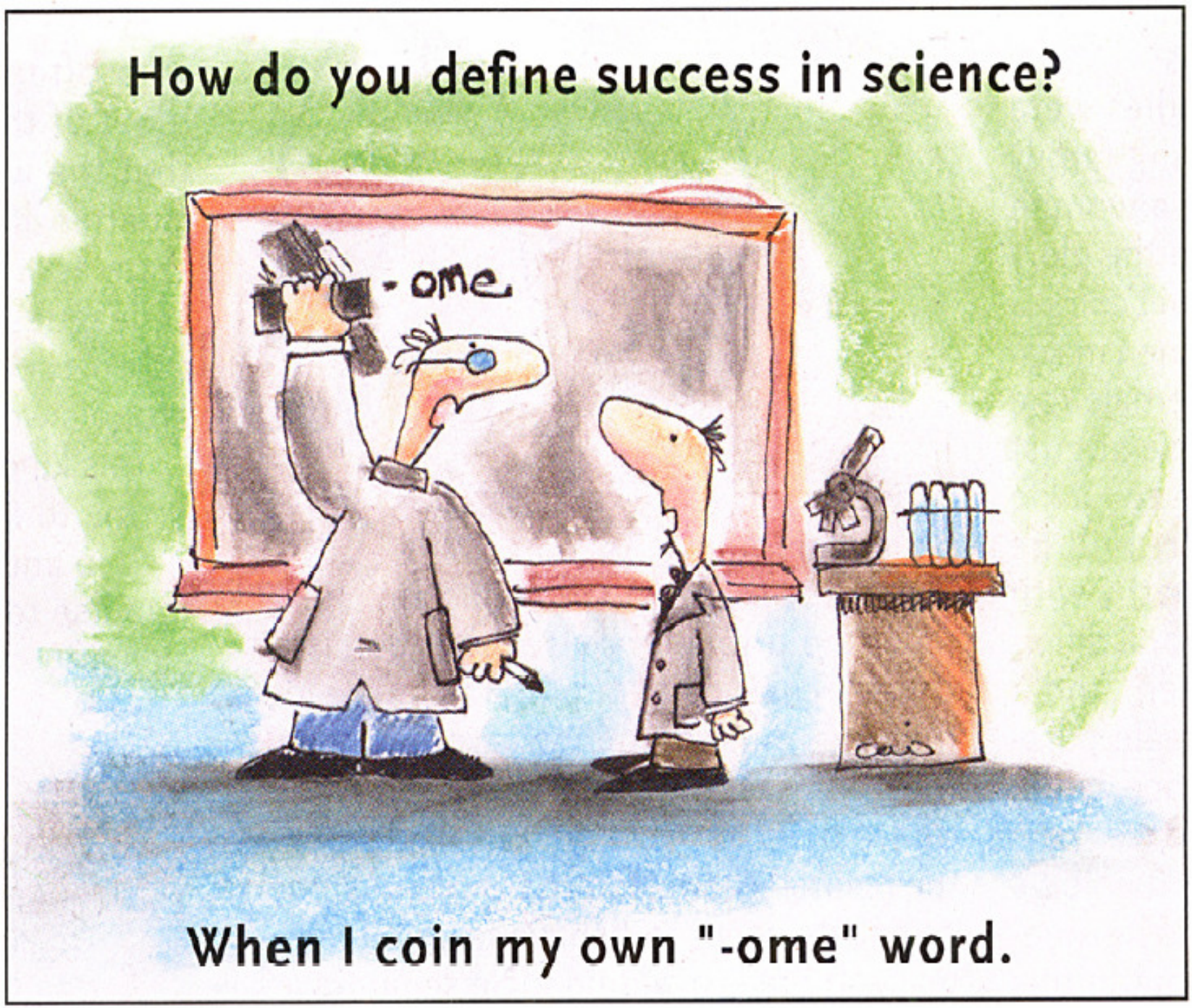
= molecules of **amino acids** that perform much of life's function

Proteome

= set of all proteins in a cell

Proteomics

= study of protein's structure & function


How do you define success in science?
-ome
When I coin my own "-ome" word.

Not so easy to remember

Proteomics represents the effort to establish the *identities*, *quantities*, *structures*, and *biochemical and cellular functions* of all proteins in an organism, organ, or organelle, and how these properties vary in space, time, or physiological state.

MCP 1.10 pg 675 National Research Council Steering committee

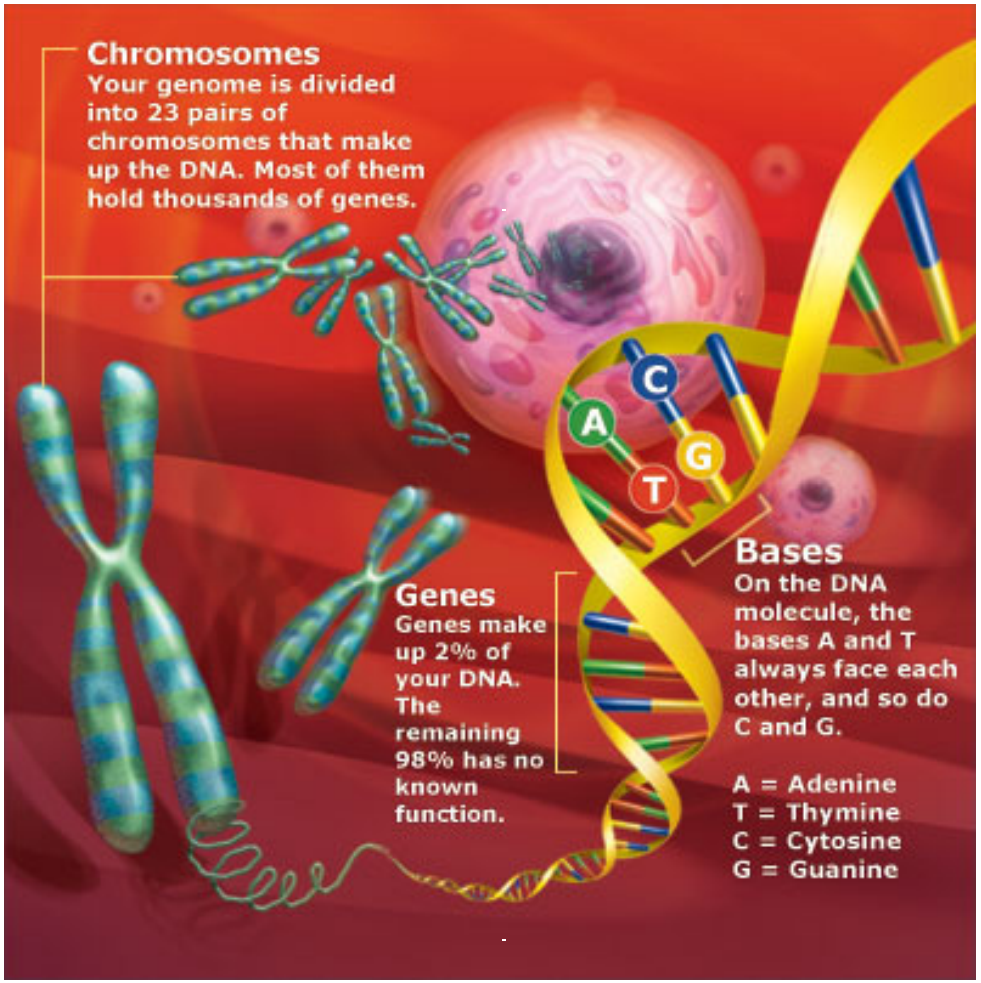
Proteomics

A true multidiscipline science

- Protein chemistry
- Mass spectrometry
- Genomics
- Bioinformatics
- Computer science
- Separation science

From Genome to Proteome

From cell to gene



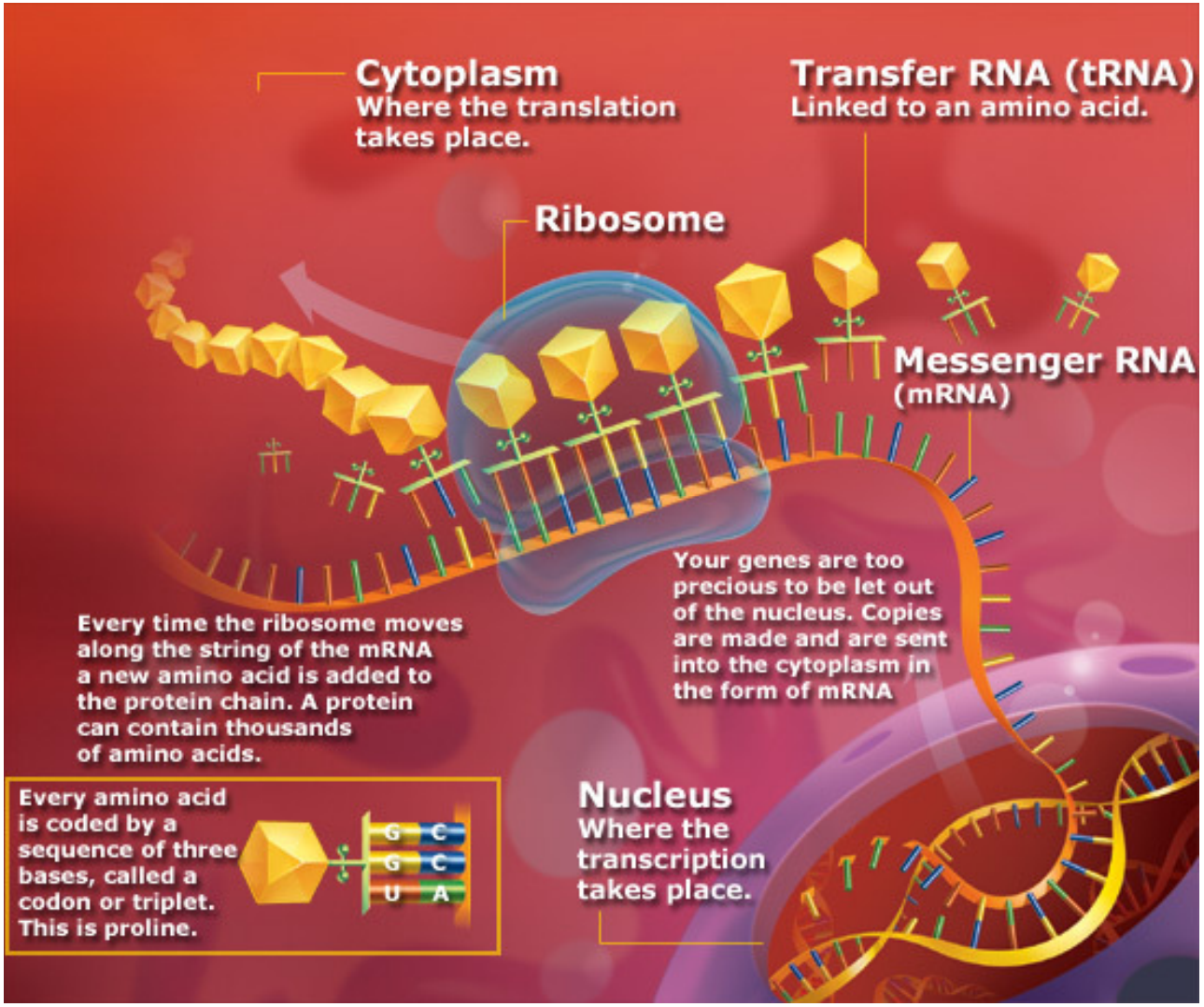
The human genome is located right in the heart of the cells, in the nucleus.

The genes, parts of the DNA (double helix), are the functional units of the genome.

They hold all the information necessary to create the proteins you need.

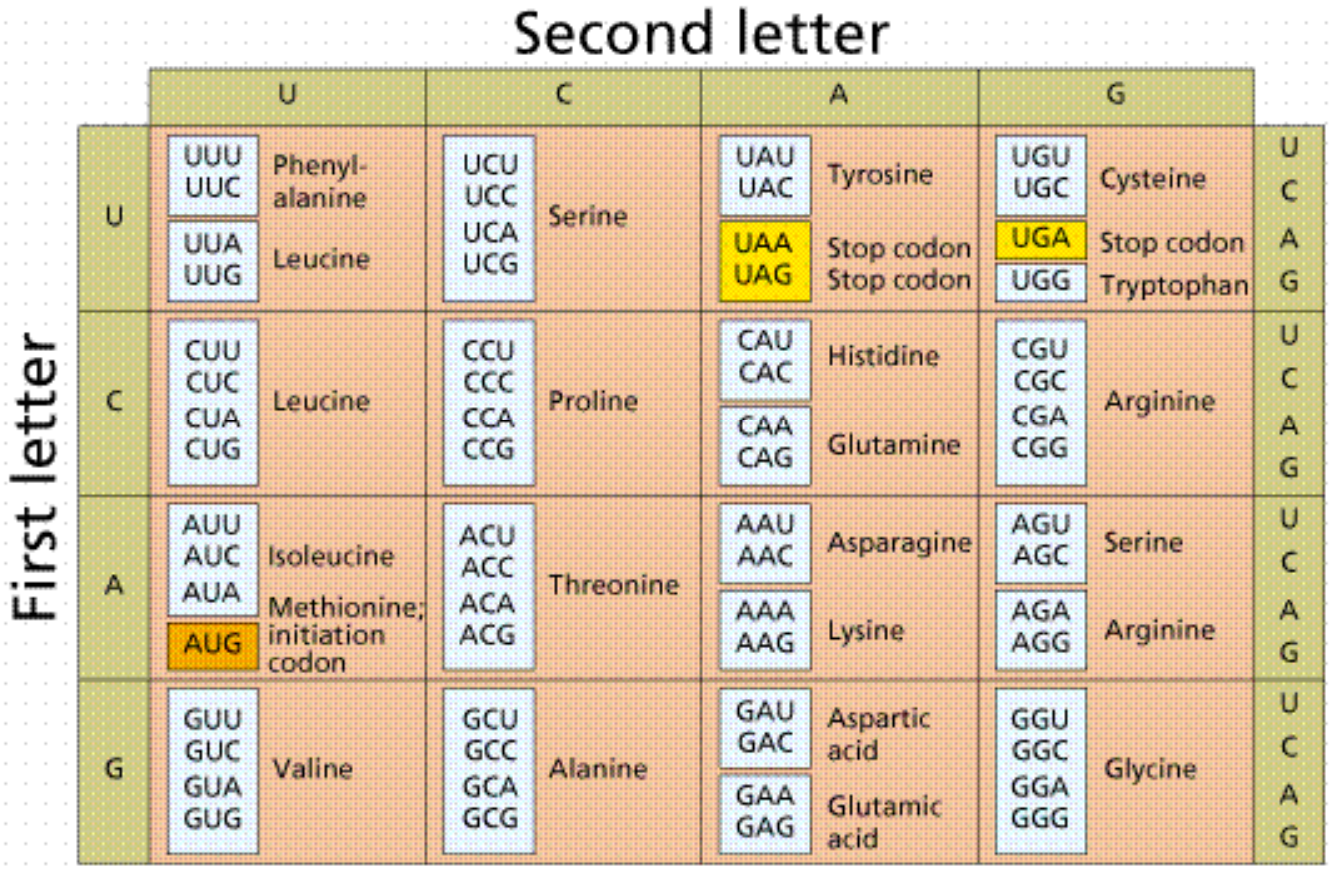
Genes are located along thread-like structures called chromos**ome**s.

Protein Synthesis



Protein synthesis is the *transcription* and *translation* of specific parts of DNA to form proteins.

Codons set amino acids that are used



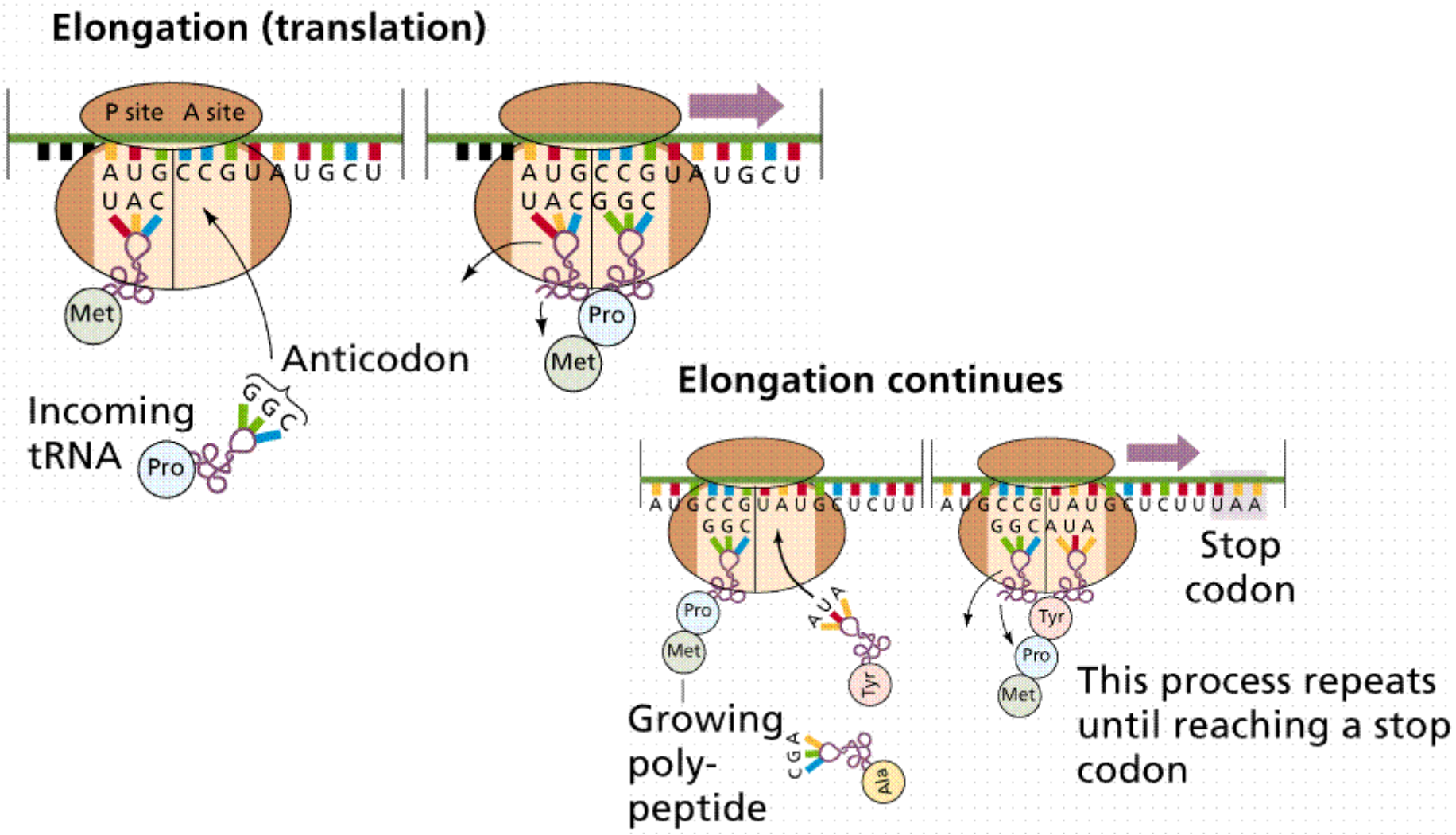
First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU, UUC: Phenylalanine UUA, UUG: Leucine	UCU, UCC, UCA, UCG: Serine	UAU, UAC: Tyrosine UAA: Stop codon UAG: Stop codon	UGU, UGC: Cysteine UGA: Stop codon UGG: Tryptophan	U C A G
C	CUU, CUC, CUA, CUG: Leucine	CCU, CCC, CCA, CCG: Proline	CAU, CAC: Histidine CAA, CAG: Glutamine	CGU, CGC, CGA, CGG: Arginine	U C A G
A	AUU, AUC, AUA: Isoleucine AUG: Methionine; initiation codon	ACU, ACC, ACA, ACG: Threonine	AAU, AAC: Asparagine AAA, AAG: Lysine	AGU, AGC: Serine AGA, AGG: Arginine	U C A G
G	GUU, GUC, GUA, GUG: Valine	GCU, GCC, GCA, GCG: Alanine	GAU, GAC: Aspartic acid GAA, GAG: Glutamic acid	GGU, GGC, GGA, GGG: Glycine	U C A G

One Genome, multiple Proteomes

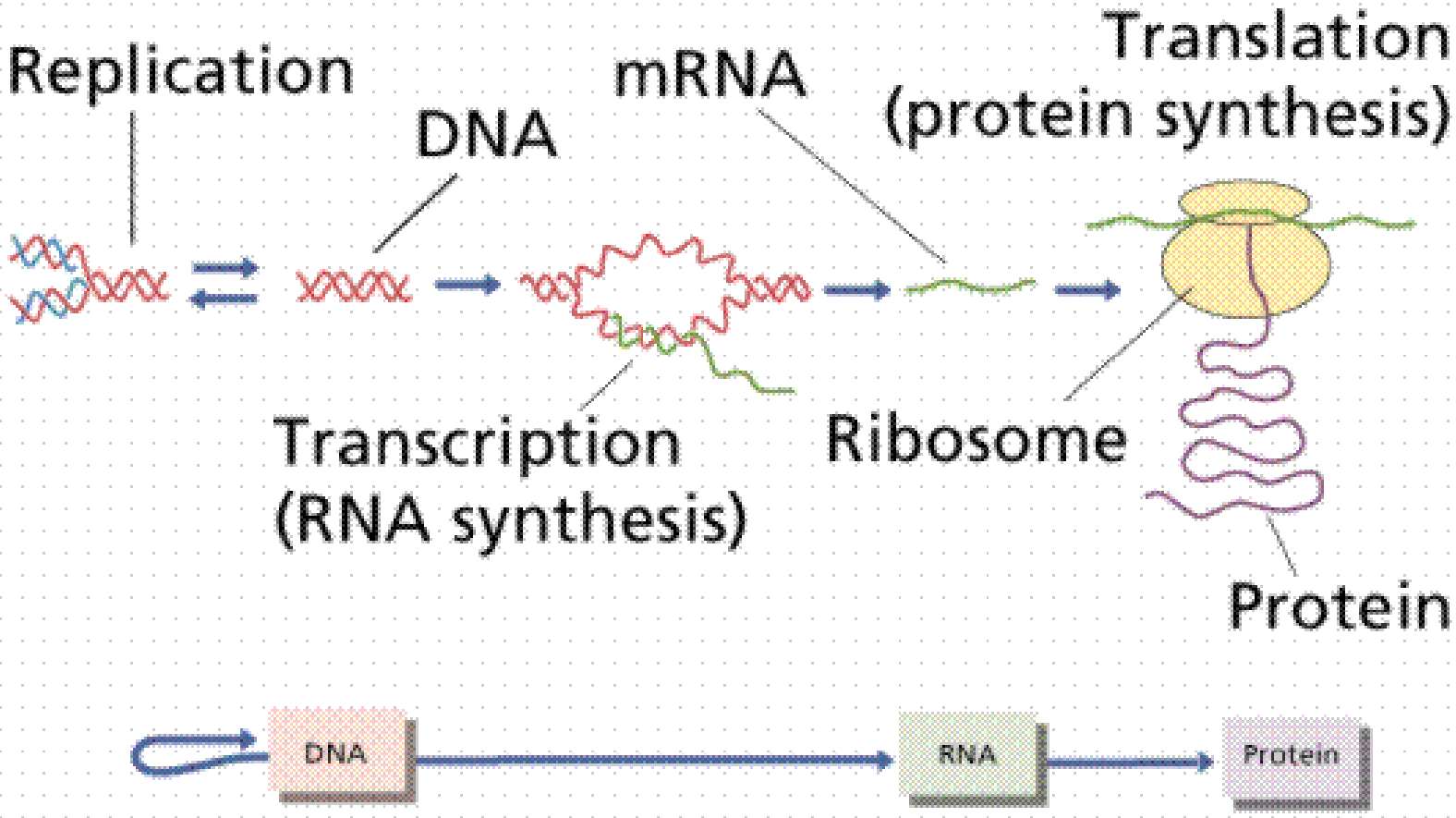
tagacgacct ggcccaacgc tgtgcccagt acaagaagga tggctgtgac ttcgccaaat ggcgttgtgt gctcaagatc ggcaagaaca ccccctccta ccaagctatc cttgagaatg ccaacgtact ggcacgctat gcgtccatct gccaatccca gcgcattgtg cccattgtagagcctgaggt gctgcctgat ggagatcacg accttgacag ggctcagaag gtcacagaga cagttctggc cgctgtgtac aaggcactca atgaccacca tgtcttcctg gagggcaccctcctgaagcc caacatggtg accgcaggac agtcctgctc caagaagtac aattatgagg acaacgctag agctacagtg ttggccctgt ccagaactgt gccagctgct gtccctggtgtgactttctt gtcaggaggt cagtcggagg aggatgcctc tgtcatttgg atgctatcaa caagatc

tagacgacct ggcccaacgc tgtgcccagt acaagaagga tggctgtgac ttcgccaaat ggcgttgtgt gctcaagatc ggcaagaaca ccccctccta ccaagctatc cttgagaatg ccaacgtact ggcacgctat gcgtccatct gccaatccca gcgcattgtg cccattgtagagcctgaggt gctgcctgat ggagatcacg accttgacag ggctcagaag gtcacagaga cagttctggc cgctgtgtac aaggcactca atgaccacca tgtcttcctg gagggcaccctcctgaagcc caacatggtg accgcaggac agtcctgctc caagaagtac aattatgagg acaacgctag agctacagtg ttggccctgt ccagaactgt gccagctgct gtccctggtgtgactttctt gtcaggaggt cagtcggagg aggatgcctc tgtcatttgg atgctatcaa caagatc

The building polypeptide chain



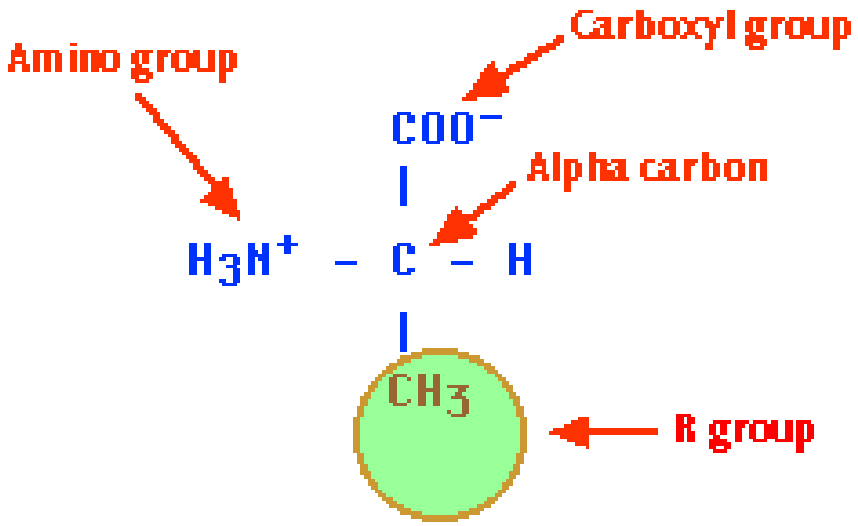
Central Dogma of Biology



genome transcriptome proteome

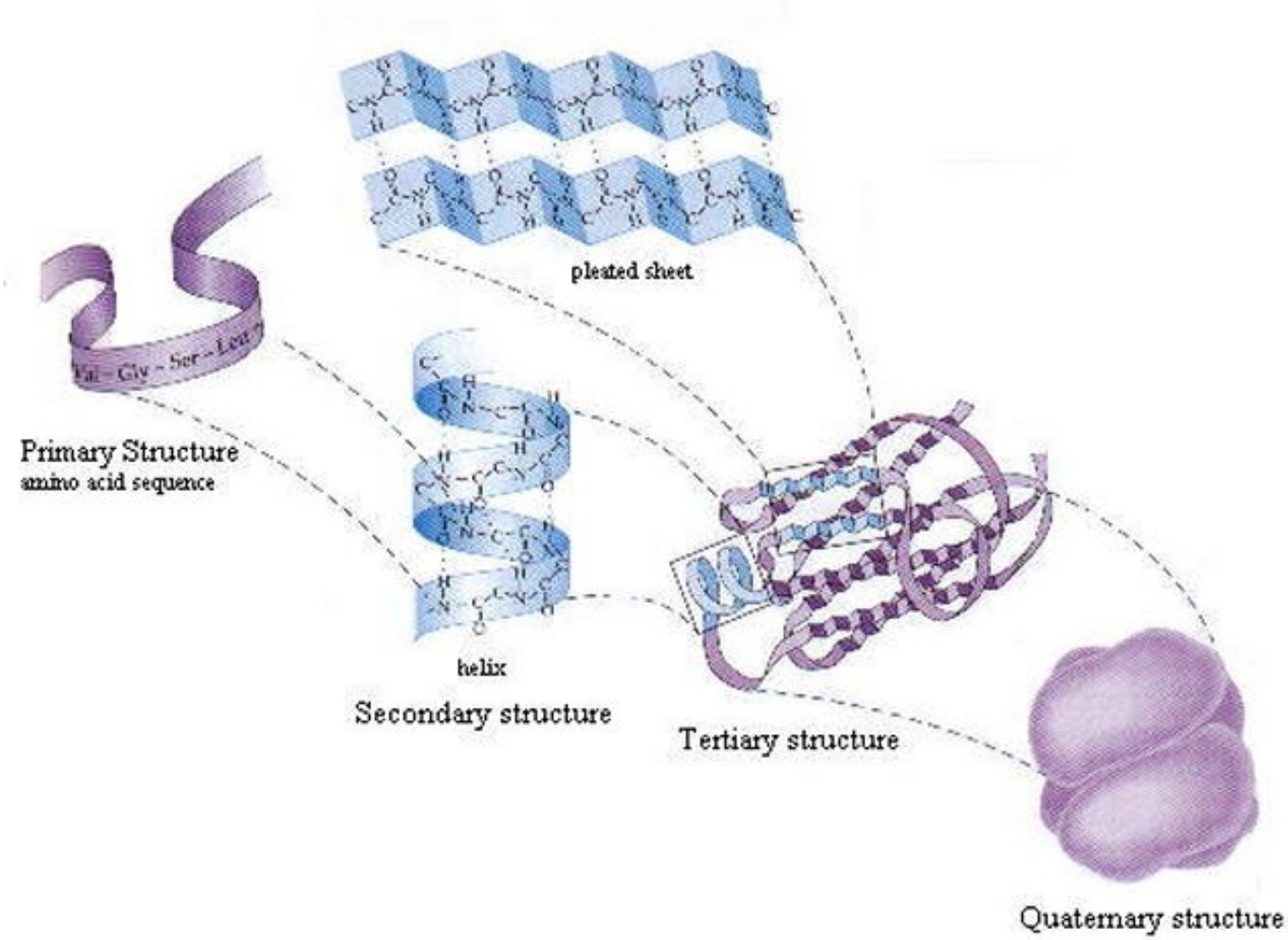
Amino acids

Amino acid



- The human body needs 20 amino acids to be able to make (or synthesize) its thousands of proteins.

Aminoacid structure



What are Proteins?

Proteins are strings of **amino acids** and are the **active elements of cells**.

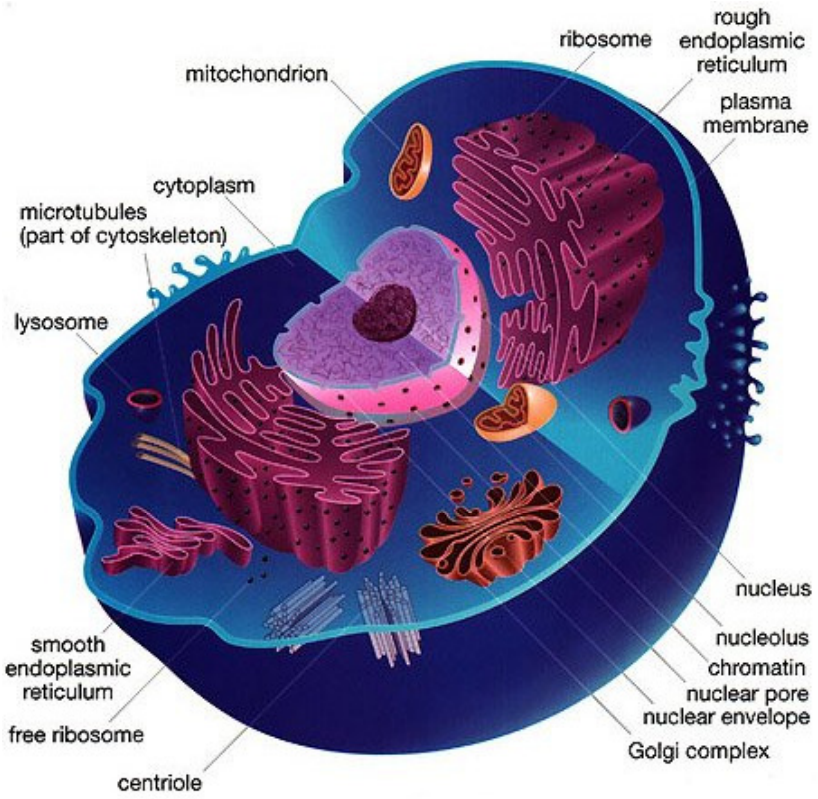
Where we find proteins?

Cells and body tissues, hormones, antibodies, and enzymes.

Protein functions?

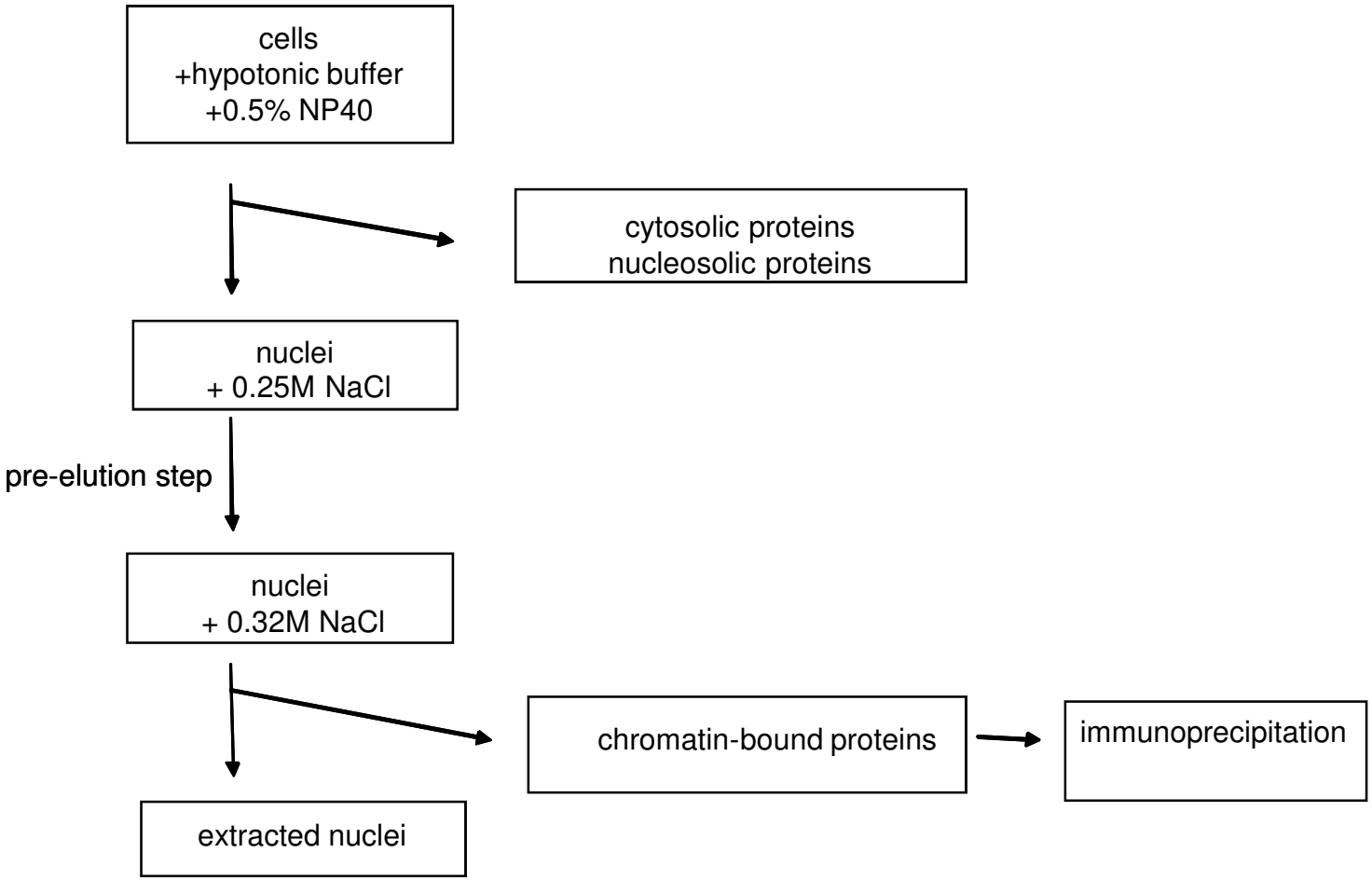
- structure providing proteins (hair and fingernails).
- help in digestion (stomach enzymes), detoxify poisons, or help fight disease.
- cell membranes proteins and are important in controlling how substances pass through these membranes.
- enzymes proteins are catalysts for biochemical reactions.
- antibodies proteins react with foreign substances to defend the body.

Sourse of Proteins



- Grow cells and blow them up (lyses)
- Dissect tissue sample, homogenize and lyse
- Synthesize

Sourse of Proteins (cell fractionation)



Protein Immunoprecipitation

Immunoprecipitation (IP) is the technique of precipitating a protein antigen out of solution using an antibody that specifically binds to that particular protein.

Biological fractionation

- Necessary!! To decrease complexity before analytical fractionation
  - Abundant protein depletion
  - Membrane fraction
  - Soluble fraction
  - Organellar fractionation
  - Affinity chromatography
  - Etc…

Separation techniques versus protein characteristics

- **Charge**

1. Ion exchange chromatography
2. Electrophoresis
3. Isoelectric focusing

- **Polarity**

1. Adsorption chromatography
2. Paper chromatography
3. RP chromatography
4. Hydrophobic interaction chromatography

- **Size**

1. Dialysis and ultrafiltration
2. Gel electrophoresis
3. Gel filtration chromatography (size exclusion chromatography)
4. Ultracentrifugation

- **Specificity**

1. Affinity chromatography

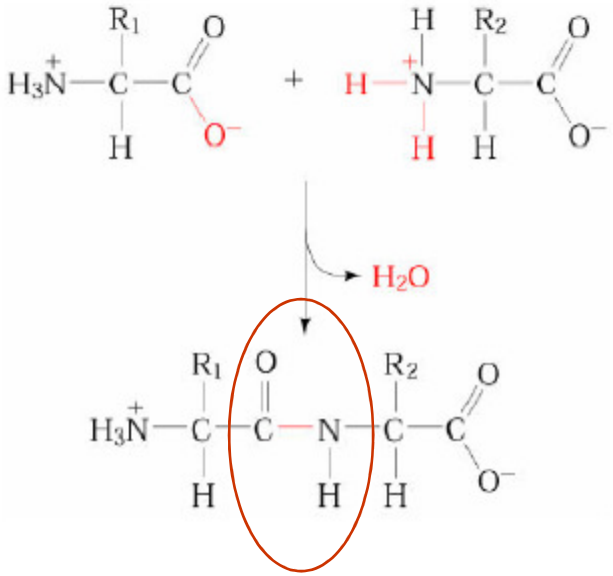
PI Protocol

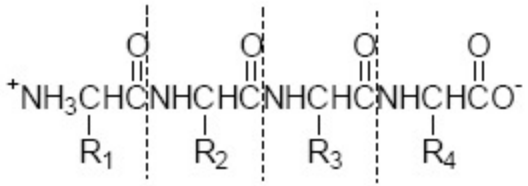
- Lyse cells and prepare sample for immunoprecipitation.
- Pre-clear the sample by passing the sample over beads that are not coated with antibody to soak up any proteins that non-specifically bind to the beads.
- Incubate solution with antibody against the protein of interest. Antibody can be attached to solid support before this step (direct method) or after this step (indirect method). Continue the incubation to allow antibody-antigen complexes to form.
- Precipitate the complex of interest, removing it from bulk solution.
- Wash precipitated complex several times. Spin each time between washes or place tube on magnet when using superparamagnetic beads and then remove supernatant. After final wash, remove as much supernatant as possible.
- Elute proteins from solid support (using low-pH or SDS sample loading buffer).
- Analyze complexes or antigens of interest. This can be done in a variety of ways:
  - SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) followed by gel staining.
  - SDS-PAGE followed by: staining the gel, cutting out individual stained protein bands, and sequencing the proteins in the bands by MALDI-Mass Spectrometry
  - Transfer and Western Blot using another antibody for proteins that were interacting with the antigen followed by chemiluminesent visualization.

Peptide bond formation

Two amino acids can undergo a condensation reaction to form a dipeptide. Further condensation reactions result in a polypeptide.

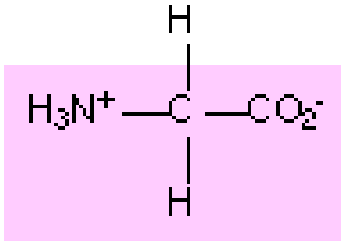
R1-COOH + R2-NH2 → R1-CO-NH-R2 + H2O



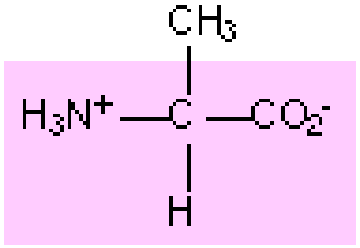


Amino acids are linked with the peptide bond, amide bond

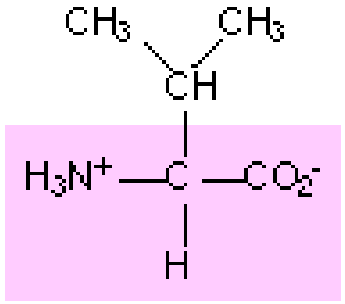
The aliphatic amino acids



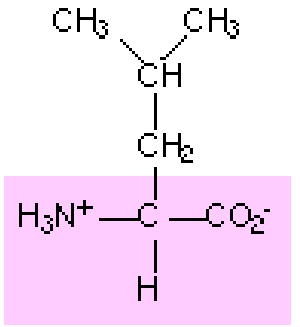
Glycine, Gly, G



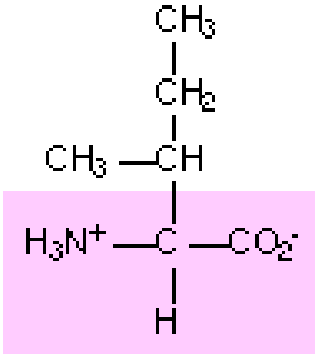
Alanine, Ala, A



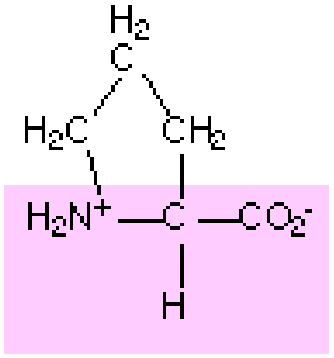
Valine, Val, V



Leucine, Leu, L

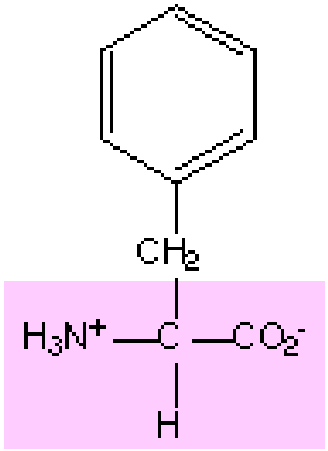


Isoleucine, Ile, I

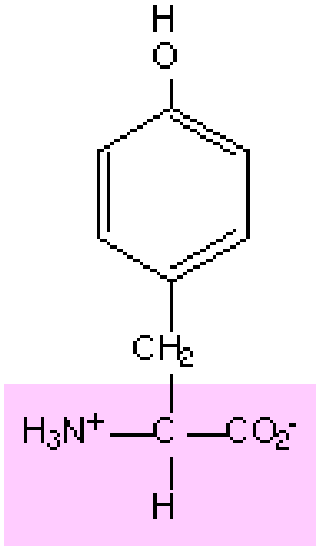


Proline, Pro, P

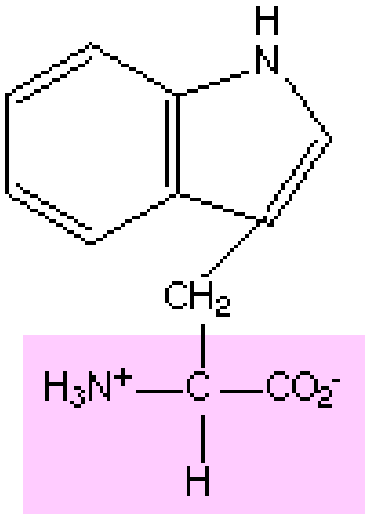
The aromatic amino acids



Phenylalanine, Phe, F

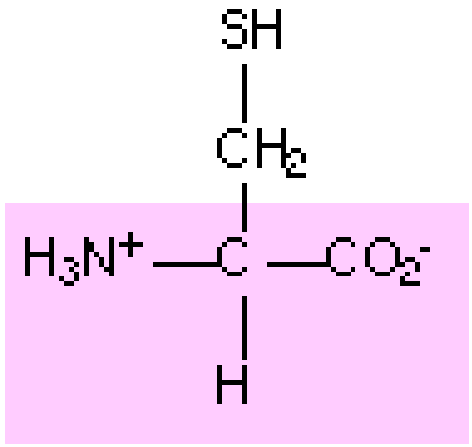


Tyrosine, Tyr, Y

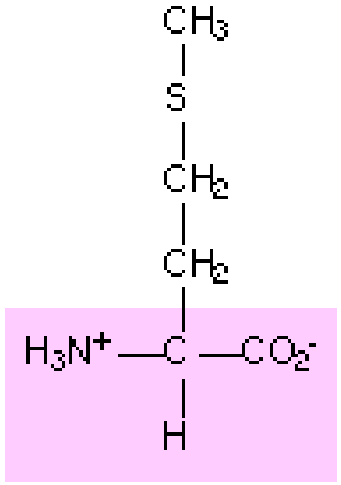


Tryptophan, Trp, W

The sulfur containing amino acids

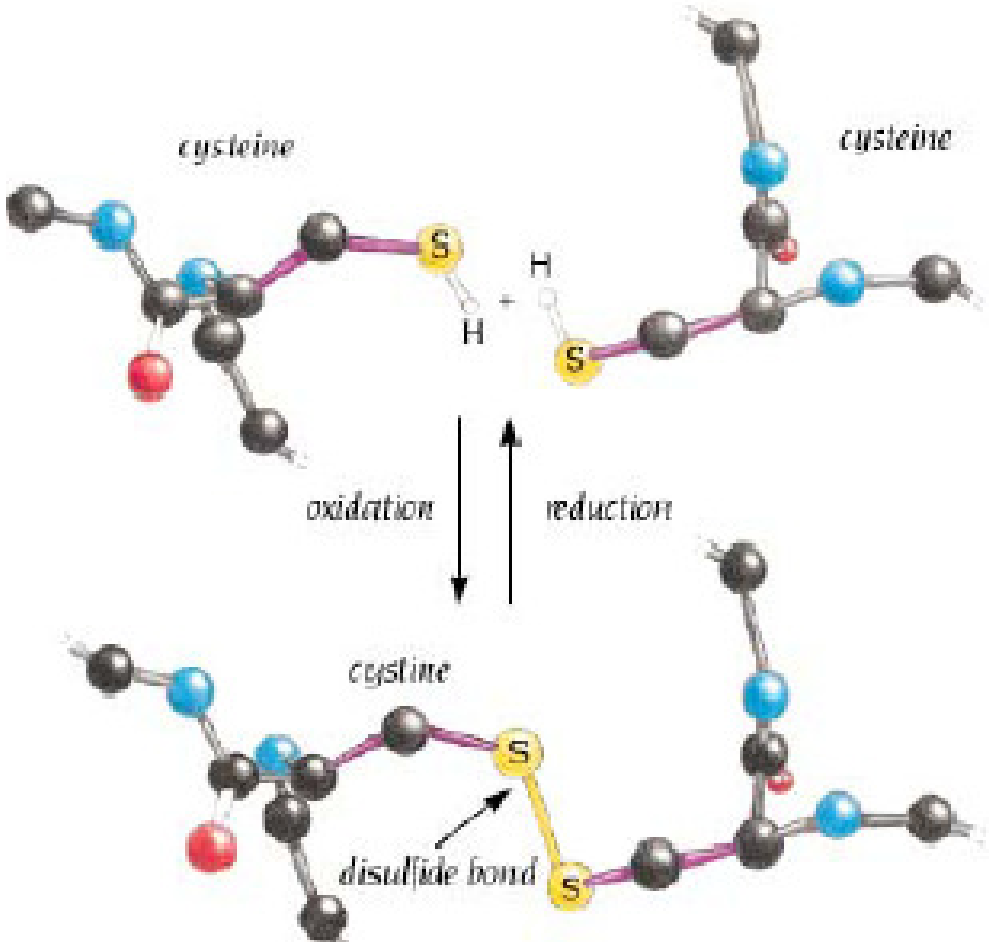


Cysteine, Cys, C



Methionine, Met, M

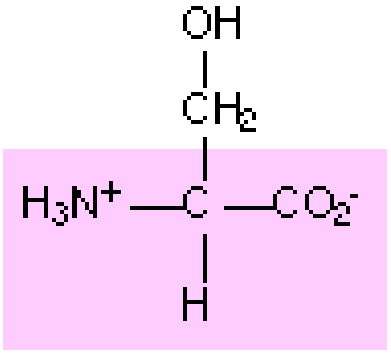
S-S bond



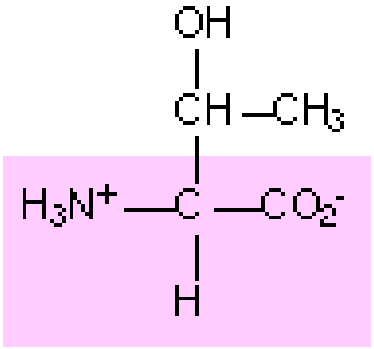
2Cysteines/oxidation=disulfide bond (only in extracellular and not intracellular proteins)

Disulfide bonds stabilize protein structure by providing crosslink

The hydroxyl amino acids

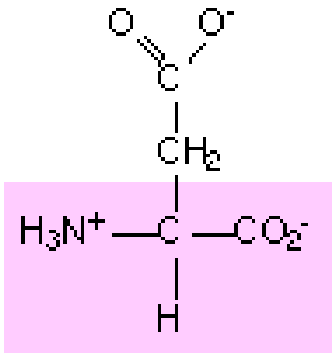


Serine, Ser, S

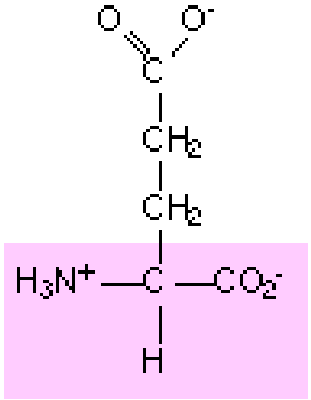


Threonine, Thr, T

The acidic amino acids

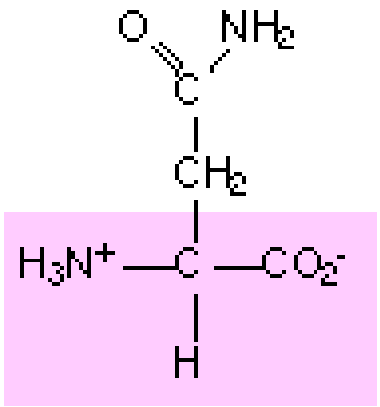


Aspartate, Asp, D

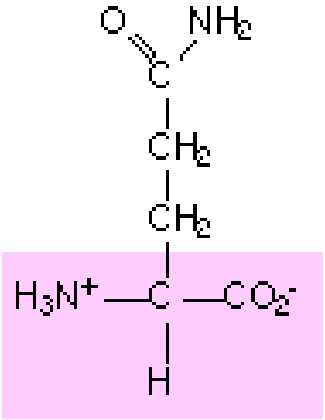


Glutamate, Glu, E

The amide amino acids

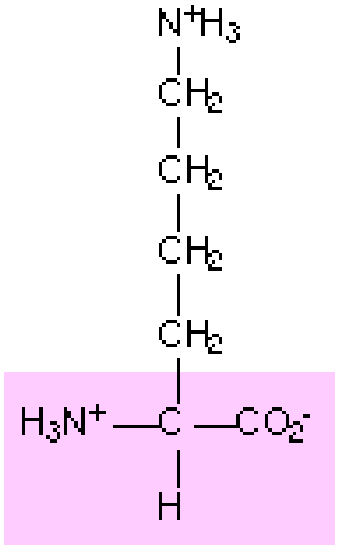


Asparagine, Asn, N

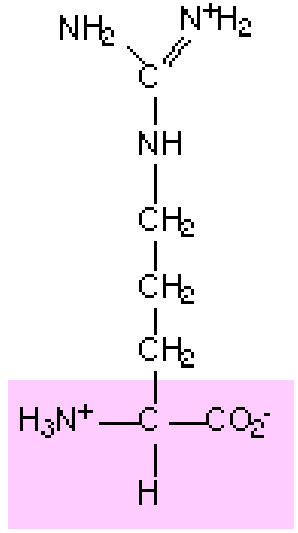


Glutamine, Gln, Q

The basic amino acids

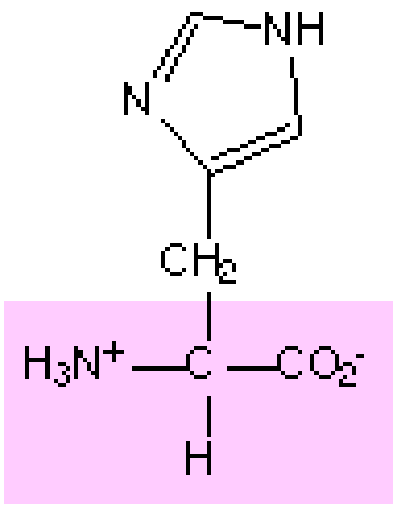


Lysine, Lys, K



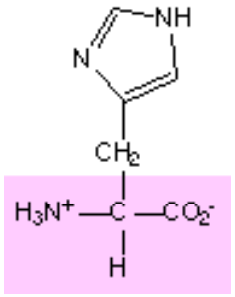
Arginine, Arg, R

The imidazole amino acid

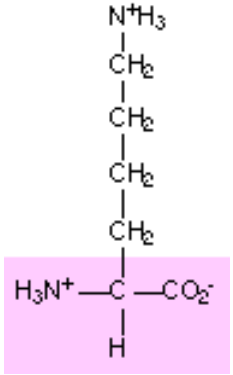


Histidine, His, H

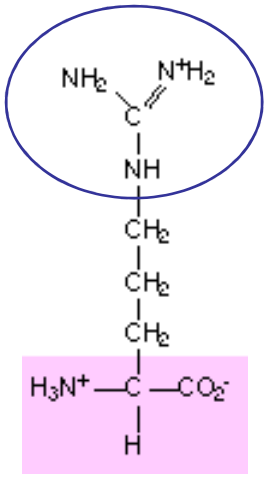
Amino acids that accept a positive charge



Histidine
H
His



Lysine
K
Lys



Arginine
R
Arg

Aminoacids clasification

Side chains of aminoacids

Responsible for many of the uniques properties of proteins

- charged or polar groups provide interesting catalytic groups

- nonpolar amino acids - a protein folding issue

Amino acid	Abbreviated names		M_r
Nonpolar, aliphatic R groups			
Glycine	Gly	G	75
Alanine	Ala	A	89
Valine	Val	V	117
Leucine	Leu	L	131
Isoleucine	Ile	I	131
Methionine	Met	M	149
Aromatic R groups			
Phenylalanine	Phe	F	165
Tyrosine	Tyr	Y	181
Tryptophan	Trp	W	204
Polar, uncharged R groups			
Serine	Ser	S	105
Proline	Pro	P	115
Threonine	Thr	T	119
Cysteine	Cys	C	121
Asparagine	Asn	N	132
Glutamine	Gln	Q	146
Positively charged R groups			
Lysine	Lys	K	146
Histidine	His	H	155
Arginine	Arg	R	174
Negatively charged R groups			
Aspartate	Asp	D	133
Glutamate	Glu	E	147

Monoisotopic and Average Mass

Amino acid	3LC	SLC	Average	Monoisotopic
Glycine	Gly	G	57.0519	57.02146
Alanine	Ala	A	71.0788	71.03711
Serine	Ser	S	87.0782	87.02303
Proline	Pro	P	97.1167	97.05276
Valine	Val	V	99.1326	99.06841
Threonine	Thr	T	101.1051	101.04768
Cysteine	Cys	C	103.1388	103.00919
Leucine	Leu	L	113.1594	113.08406
Isoleucine	Ile	I	113.1594	113.08406
Asparagine	Asn	N	114.1038	114.04293
Aspartic acid	Asp	D	115.0886	115.02694
Glutamine	Gln	Q	128.1307	128.05858
Lysine	Lys	K	128.1741	128.09496
Glutamic acid	Glu	E	129.1155	129.04259
Methionine	Met	M	131.1926	131.04049
Histidine	His	H	137.1411	137.05891
Phenyalanine	Phe	F	147.1766	147.06841
Arginine	Arg	R	156.1875	156.10111
Tyrosine	Tyr	Y	163.1760	163.06333
Tryptophan	Trp	W	186.2132	186.07931

What Proteomics can do?

Types of Experiments; key questions of Proteomics

Protein separation
In order identify each protein in the mixture.

Protein identification
Mass spectrometry
Antibody based assays can also be used, but are unique to one sequence motif.
Edman degradation used to confirm sequence when MS unavailable.

Protein quantification
Gel-based methods, differential staining of gels with fluorescent dyes (difference gel electrophoresis).
Gel-free, various tagging or chemical modification methods, label free methods.

Protein sequence analysis
Searching databases.

Structural proteomics
High-throughput determination of protein structures in three-dimensional space. Methods are x-ray crystallography and NMR spectroscopy.

Interaction proteomics
The investigation of protein interactions on the atomic, molecular and cellular levels.

Protein modification
Phosphoproteomics and glycoproteomics.

Cellular proteomics
The goal is to map location of proteins and protein-protein interactions during key cell events. Uses techniques such as X-ray Tomography and optical fluorescence microscopy.

Methods for protein analysis

High resolution mass spectrometry methods

Top down methods

Intact proteins

Bottom up methods

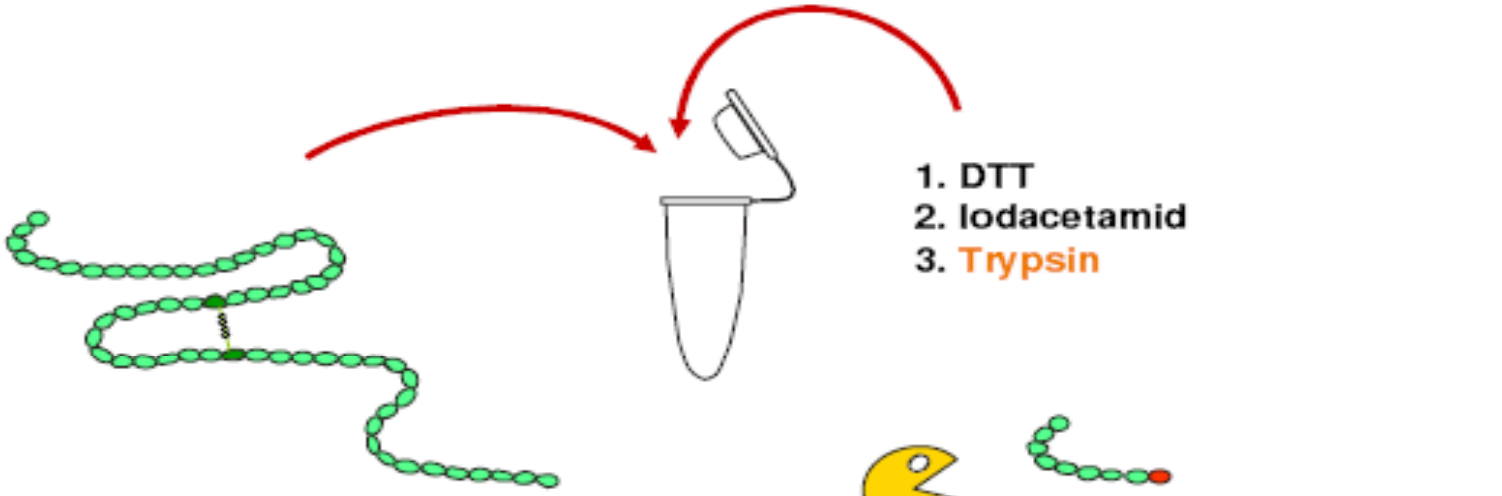
Digested peptides

De Novo Sequencing

Gel based methods

One/Two-dimensional gel electrophoresis

Protein digestion



- Reduction of cysteines by DTT
- Alkylation of cysteins by IAA
- Digestion of the protein by trypsin

(cleaves after lysine and arginine)

Why Trypsin

- Very Specific R/ K/ except R/P K/P
- Very Active
- Can buy it in a very pure form
  - Promega, Sigma, Princeton Scientific
- Resistant to digesting itself
- K and R residues in an average protein are optimally spaced
  - Peptides are usually 600-3000 Da
- For electrospray, fragments are at least doubly charged because of C term basic AA's K and R

Chemical and enzymatic cleavage reagents

Chemical reagents	***Cleavage sites***
Cyanogen bromide	after M
BNPS-skatole or DMSO + HCl	after W
Acid hydrolysis	D/P then random
Endopeptidases	Cleavage sites
Trypsin	after K/R
Endoproteinase Lys-C	after K
Endoproteinase Asp-N	before D
Endoproteinase Arg-C	after R
Chymotrypsin	after F/W/Y/L
Pepsin	after F/W/Y/L
Thermolysin	before L/I/M/F/W

Proteomics and Mass Spectrometry

Methods for protein analysis

High resolution mass spectrometry methods

Top down methods

Intact proteins

Bottom up methods

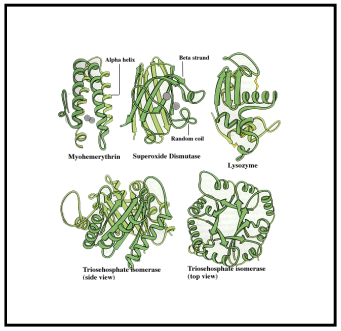
Digested peptides

De Novo Sequencing

Gel based methods

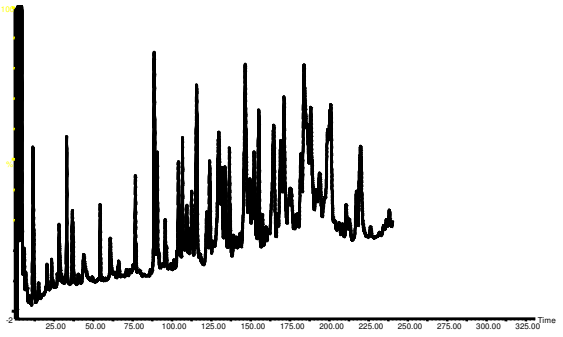
One/Two-dimensional gel electrophoresis

Schematic representation of the top-down approach

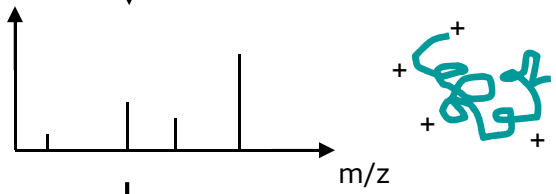


Protein mixture

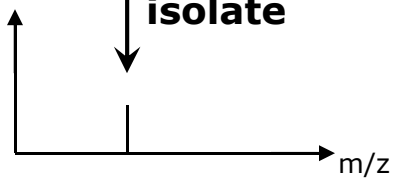
RP-HPLC



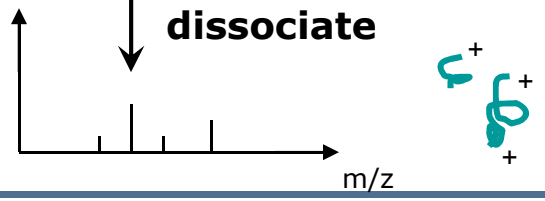
ESI-FTICR MS



isolate



dissociate

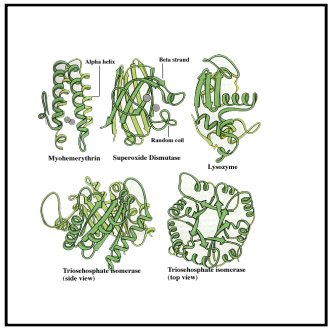


Compare experimental deconvoluted spectra

with predicted spectra

Protein identification

Schematic representation of the bottom-up approach



Protein mixture

1) Denaturation
2) Reduction
3) Alkylation
4) Trypsin digestion

Peptide fragments

HPLC

Mass spectrometer

MS

Peptide mass fingerprinting

m/z

Search

Mascot
ProFound
MS-Fit

Database

Search

Peptide sequence tagging

G L F E

m/z

MS/MS

Selected ion

Mascot
XTandem
Sequest

Protein identification

Mass Spectrometry based protein identification

Peptide Mass Fingerprinting (PMF)

- Determine m/z of the peptide ions only (MS)

Product Ion Scanning

- Determine the m/z of the peptide ions (Parent ions)
- Fragment peptide ions
- Determine m/z of fragments (Product Ions)

Peptide Mass Fingerprinting

Isolated protein

Digestion

MALDI MS

999.99
1111.11
1222.22
1333.33
1444.44
1555.55
1666.66

Database Query (compare with list of 'in-silico' digests)

999.99
1111.11
1222.22
1333.33
1444.44
1555.55
1666.66

Finger Print limitations

- You need a mass spectrometer capable of reasonable accurate masses
- Genome must be pretty small
  Yeast or smaller for good results
- Mixtures of two or more proteins can be a problem

Finger Print Advantages

- Usually can give you better coverage
- Very Fast
- Easy
- Good preliminary screening

Product Ion Scanning

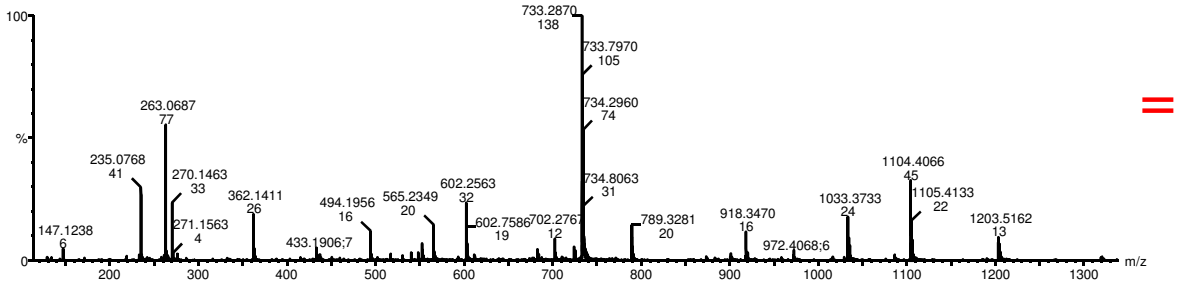
- Digest Protein with trypsin
- Determine the m/z of a peptide ion
  - ESI, MALDI
- Isolate the peptide ion from any other ions
- Fragment the peptide ion
- Determine mass of fragments
- Obtain amino acid sequence data from fragments

SEARCHING....

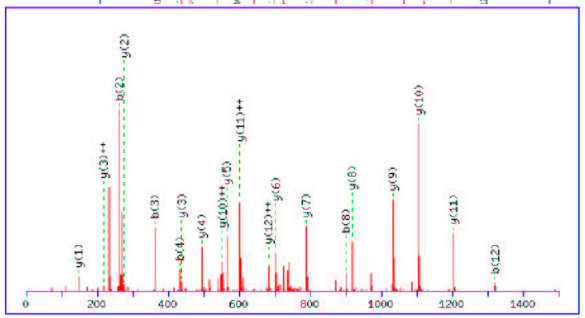
NR Database approx

1 Million protein sequences

50 million tryptic peptide sequences



=



Time = 15 seconds

Thermo SCIENTIFIC

Computer Programs

Thermo Scientific application specific software

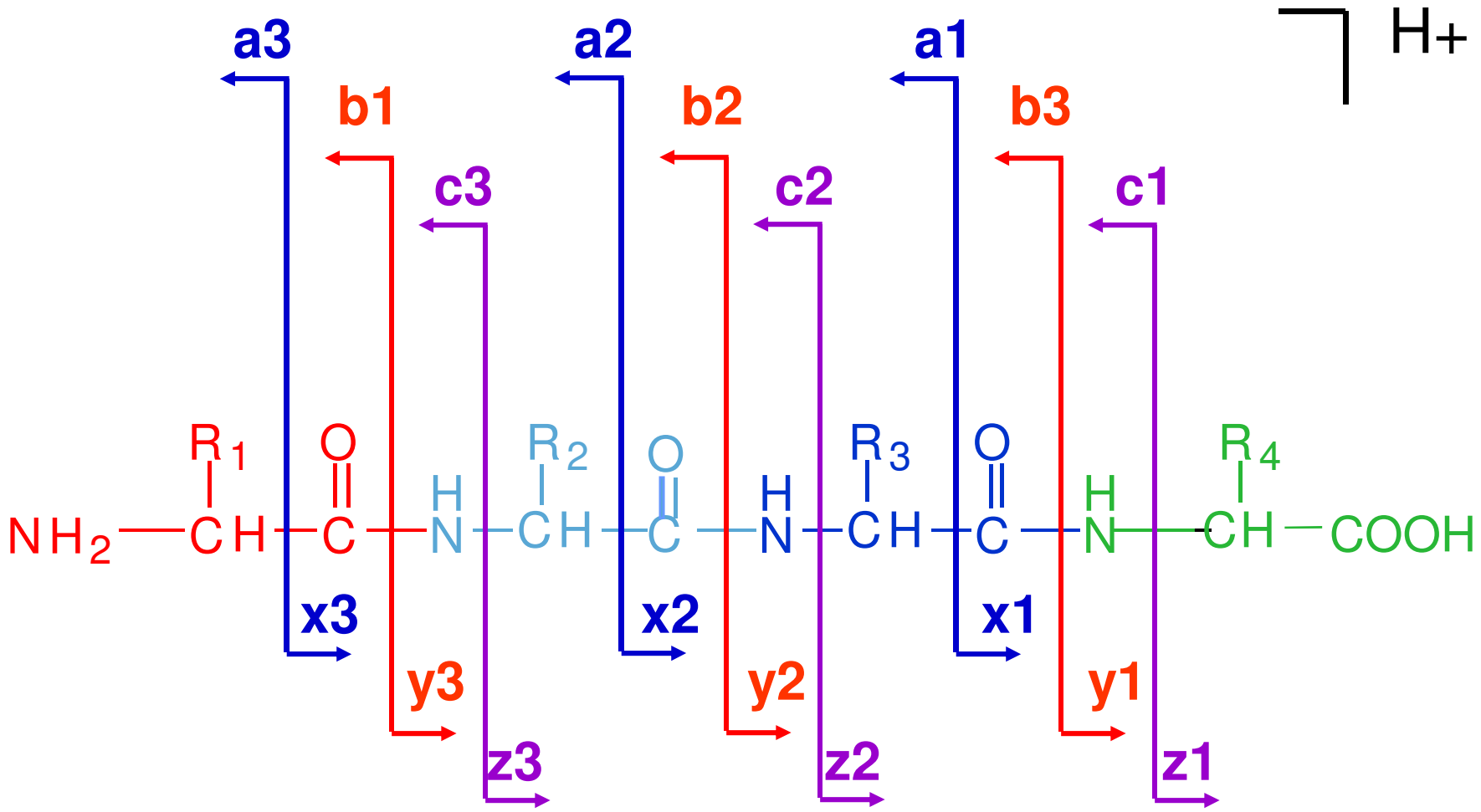
- Xcalibur data system (operating platform, Protein Calculator, Xtract)
- BioWorks protein identification software
- Proteome Discoverer
- Mass Frontier (structure of compounds)
- MetWorks (drug methabolism software)
- SIEVE (differential expression software)

Search Engines

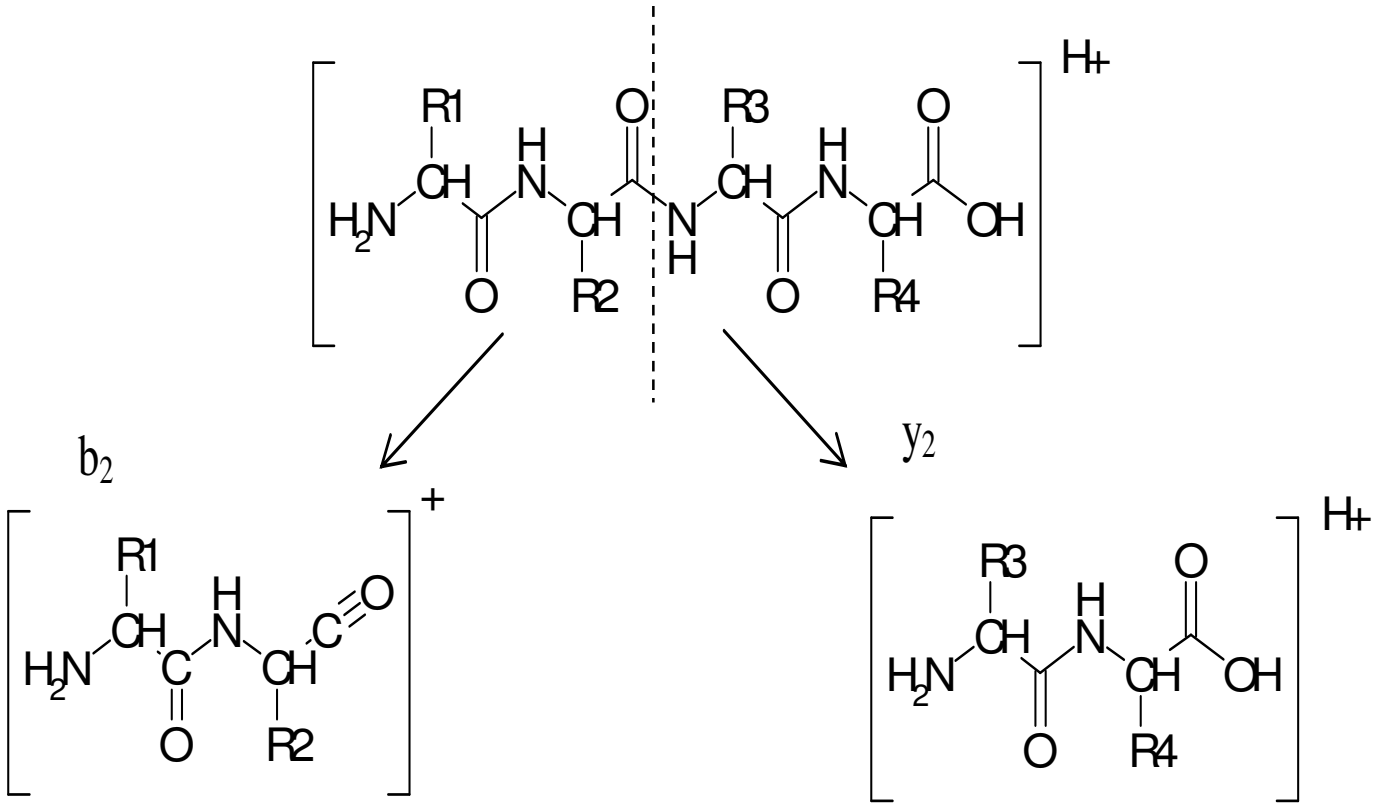
- 4 general Types
  - Automated De novo sequencing
    - Peaks
    - Lutefisk XP
  - Peptide Sequence tags
    - Guten Tag
  - Cross Correlation
    - SEQUEST
  - Probability Based
    - Mascot
    - xTandem 2
    - OMSSA
    - PROTE_PROBE

Peptide Fragmentation

Roepstorff Nomenclature for Possible Peptide Fragments



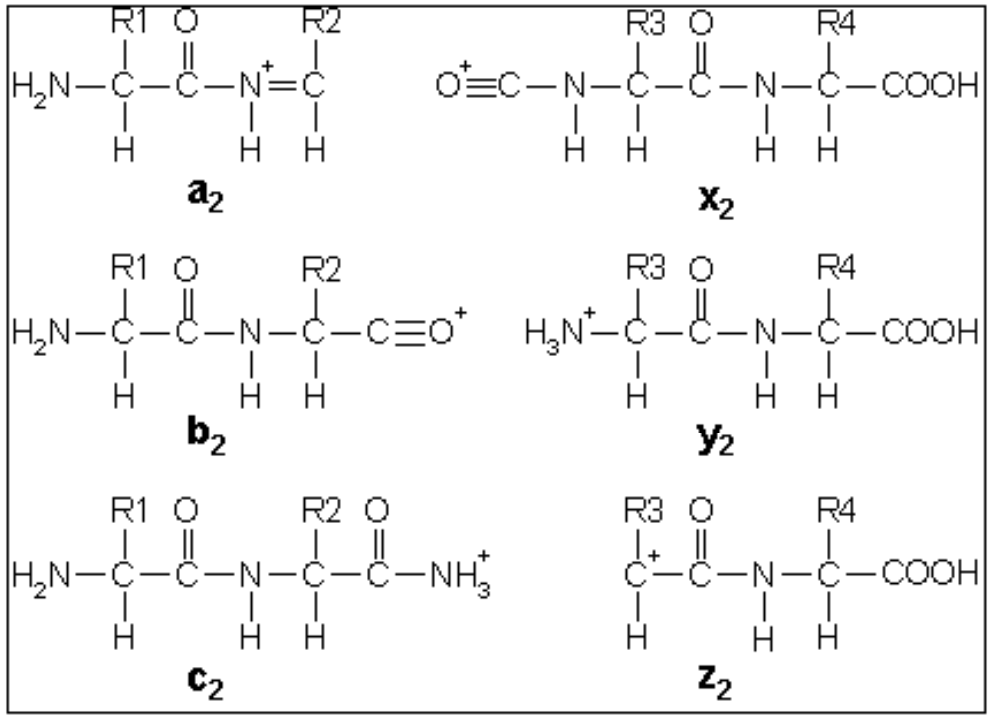
Fragmentation scheme of a tetrapeptide showing the formation of b and y ions



Peptide Fragmentation (Low-Energy Collision induced fragmentation)

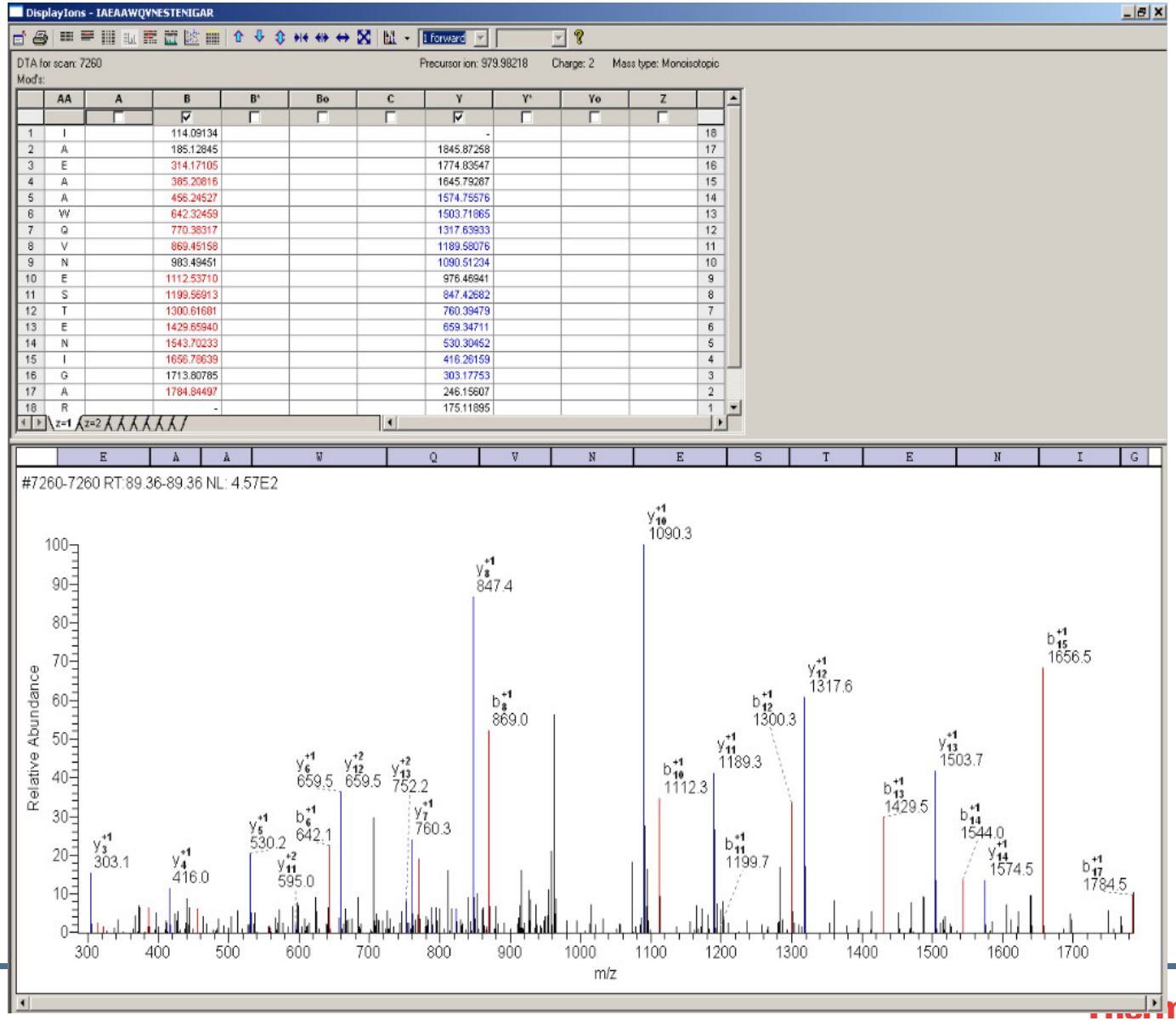
- Under low energy dissociation conditions, peptides primarily fragment at the C-N bond.
- If the charge is retained on the N terminal fragment, the ion is classed as either *a*, *b* or *c*.
- If the charge is retained on the C terminal, the ion type is either *x*, *y* or *z*.
- A subscript indicates the number of residues in the fragment.
- The loss of CO from the b ion is known as a-type ion
- In addition, peaks are seen for ions which have lost ammonia (-17 Da) denoted *a**, *b** and *y** and water (-18 Da) denoted *a°*, *b°* and *y°*.

The structures of the six singly charged sequence ion

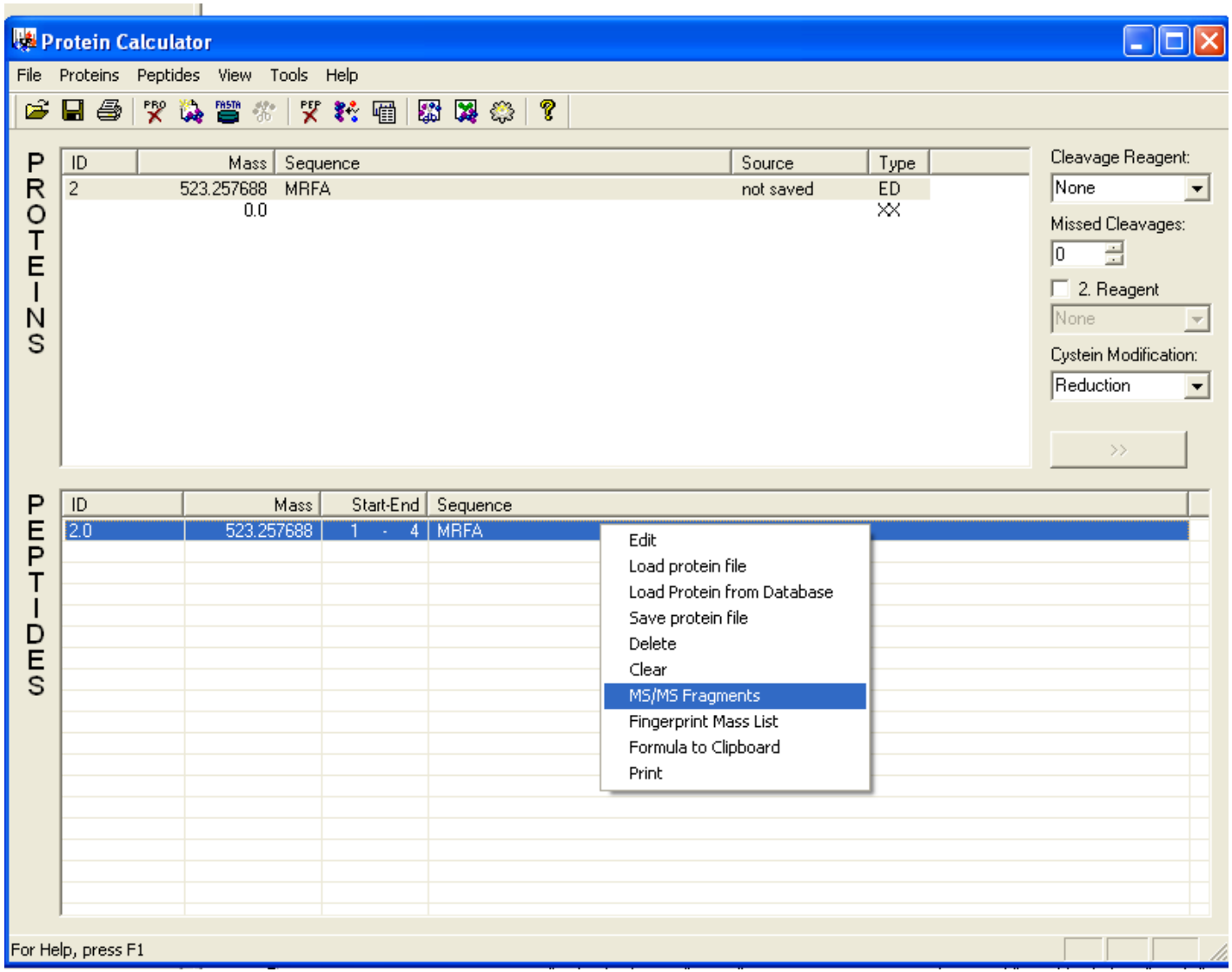


http://www.matrixscience.com/help/fragmentation_help.html

Example of *y* and *b* ions in BioWorks™

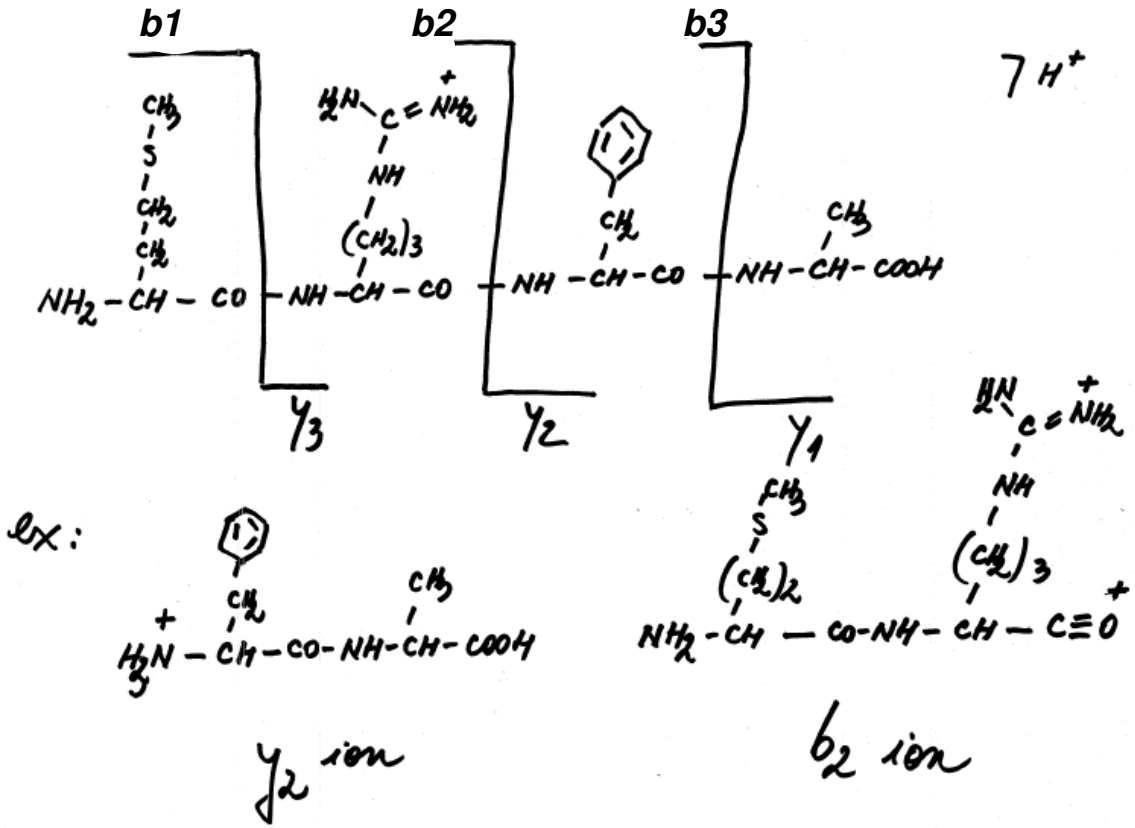


MRFA fragmentation



To better undersand

M R F A



HCD vs. CID on MRFA

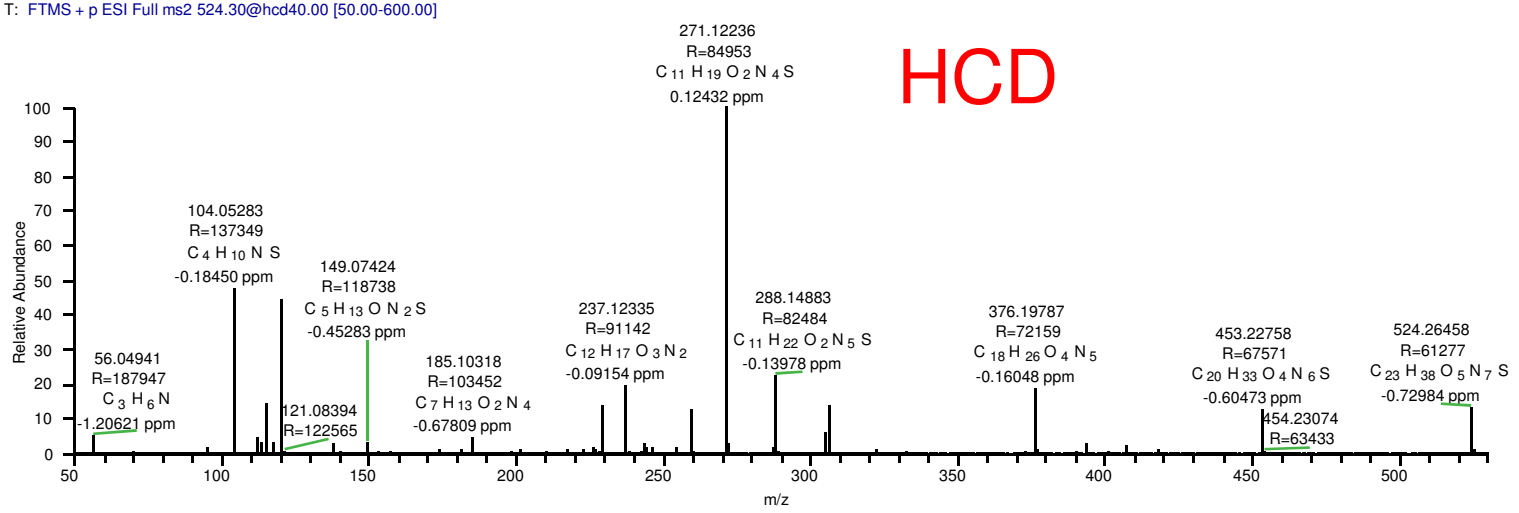
E:\MRFA_CID_HCD

Sorted Fragments Peptid-ID: 2.0

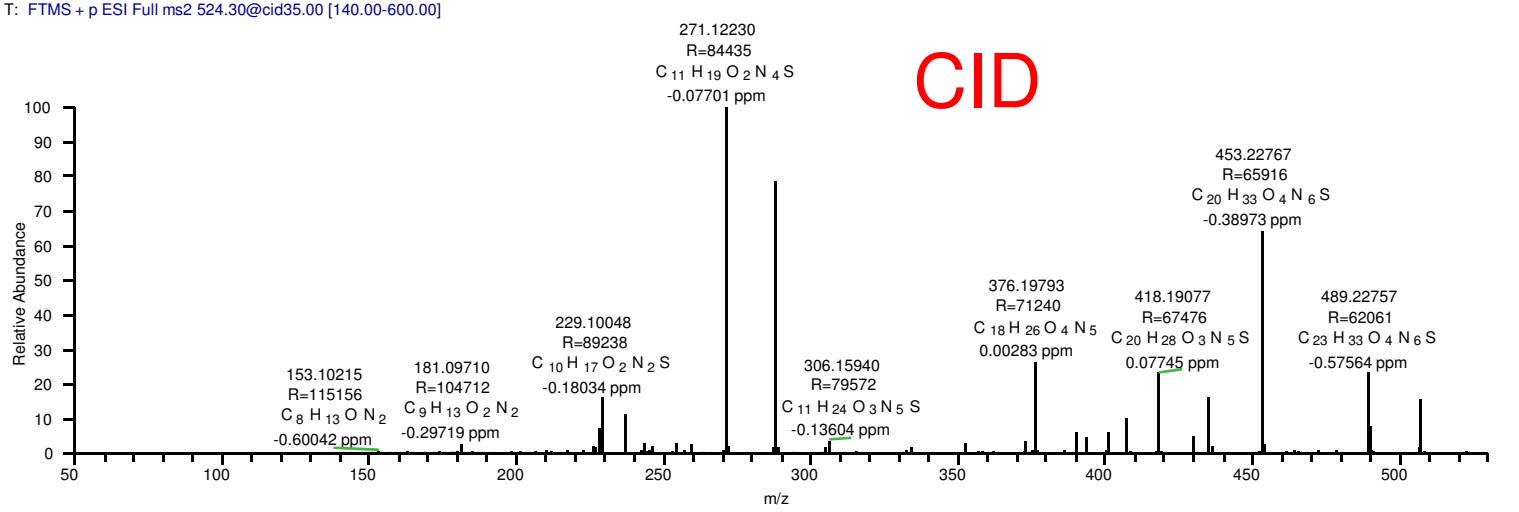
Mass	Charge	Type	Sequence
44.049476	1	I 4[M+H]+	A
72.044390	1	Y0 1[M+H]+	A
73.028406	1	Y* 1[M+H]+	A
90.054955	1	Y 1[M+H]+	A
104.052847	1	I 1[M+H]+	M
114.037197	1	B0 1[M+H]+	M
115.021212	1	B* 1[M+H]+	M
120.080776	1	I 3[M+H]+	F
129.113473	1	I 2[M+H]+	R
132.047761	1	B 1[M+H]+	M
219.112804	1	Y0 2[M+H]+	FA
220.096820	1	Y* 2[M+H]+	FA
237.123369	1	Y 2[M+H]+	FA
270.138308	1	B0 2[M+H]+	MR
271.122323	1	B* 2[M+H]+	MR
288.148872	1	B 2[M+H]+	MR
375.213915	1	Y0 3[M+H]+	RFA
376.197931	1	Y* 3[M+H]+	RFA
393.224480	1	Y 3[M+H]+	RFA
417.206721	1	B0 3[M+H]+	MRF
418.190737	1	B* 3[M+H]+	MRF
435.217286	1	B 3[M+H]+	MRF

OK

MRFA_CID_HCD #183-198 RT: 3.01-3.27 AV: 16 NL: 1.93E7
T: FTMS + p ESI Full ms2 524.30@hcd40.00 [50.00-600.00]



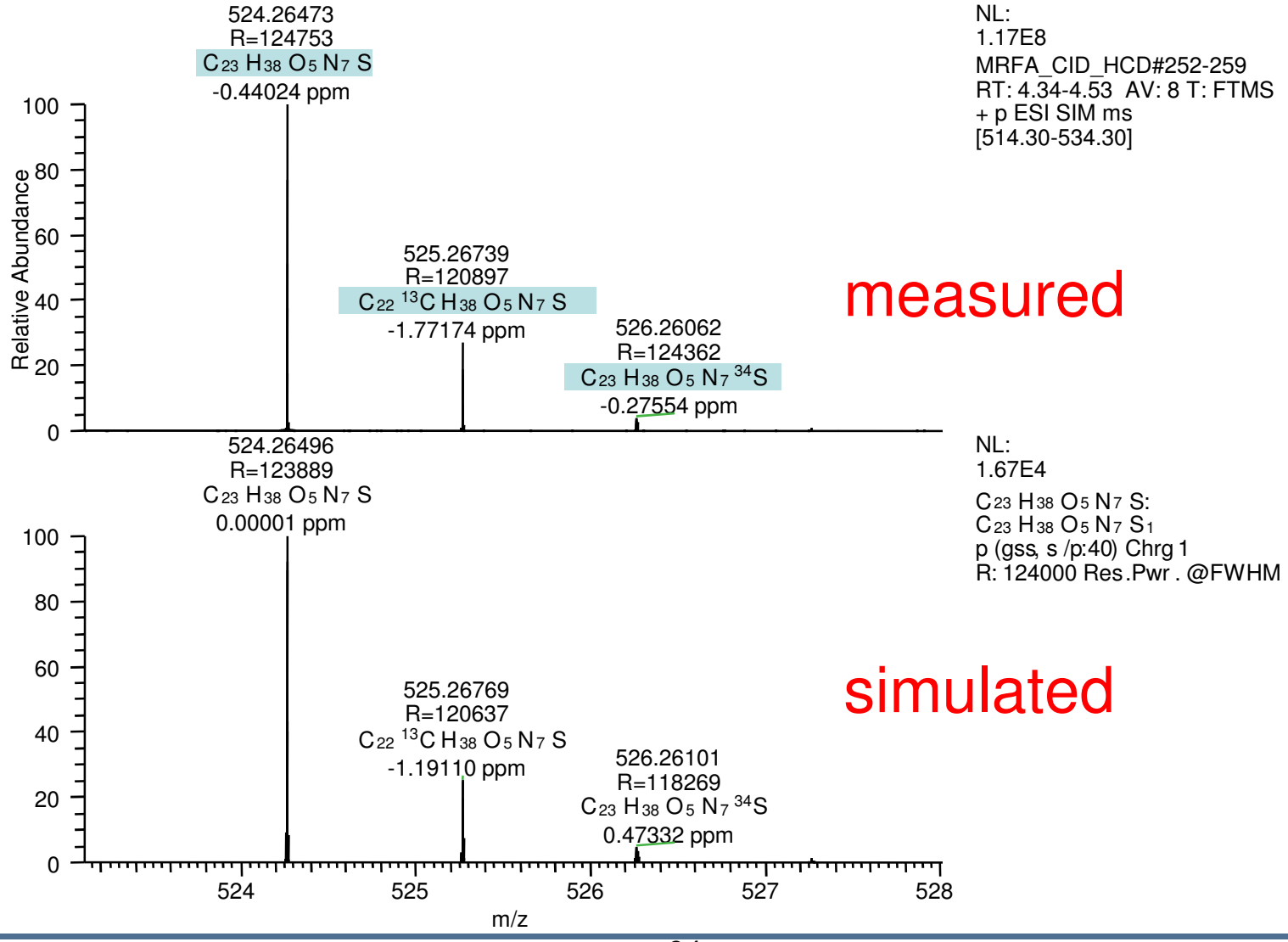
MRFA_CID_HCD #133-152 RT: 2.18-2.49 AV: 20 NL: 3.56E7
T: FTMS + p ESI Full ms2 524.30@cid35.00 [140.00-600.00]



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

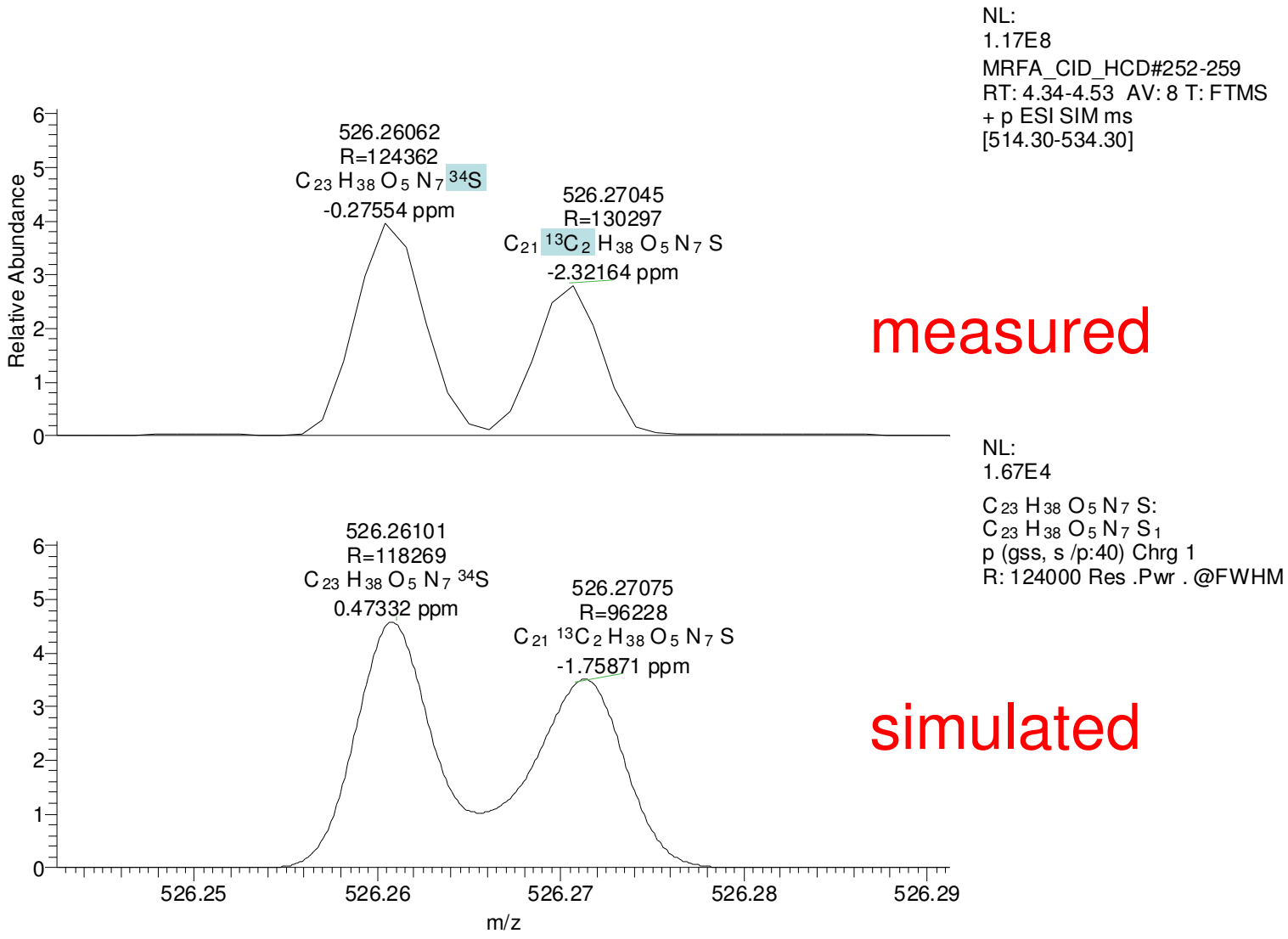
Elemental composition on MRFA



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

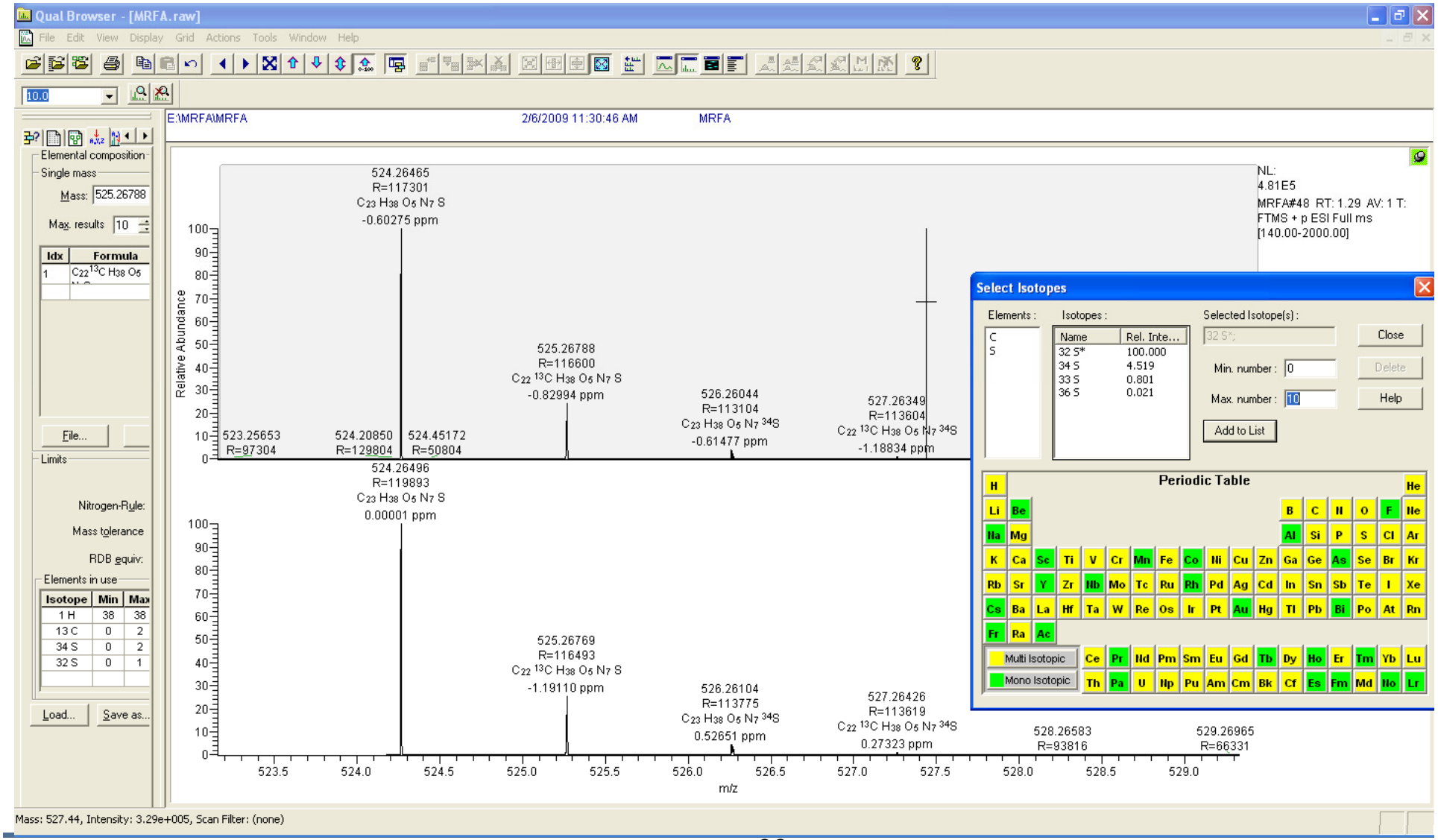
Isotopic distribution of 526.26 on MRFA



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

Isotopic contribution of ^{13}C and ^{34}S in MRFA



Thermo SCIENTIFIC

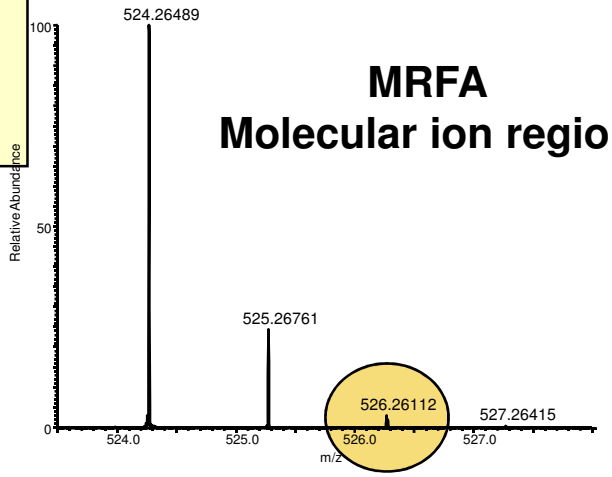
ThermoFisher SCIENTIFIC

Detection Time & Mass Resolution

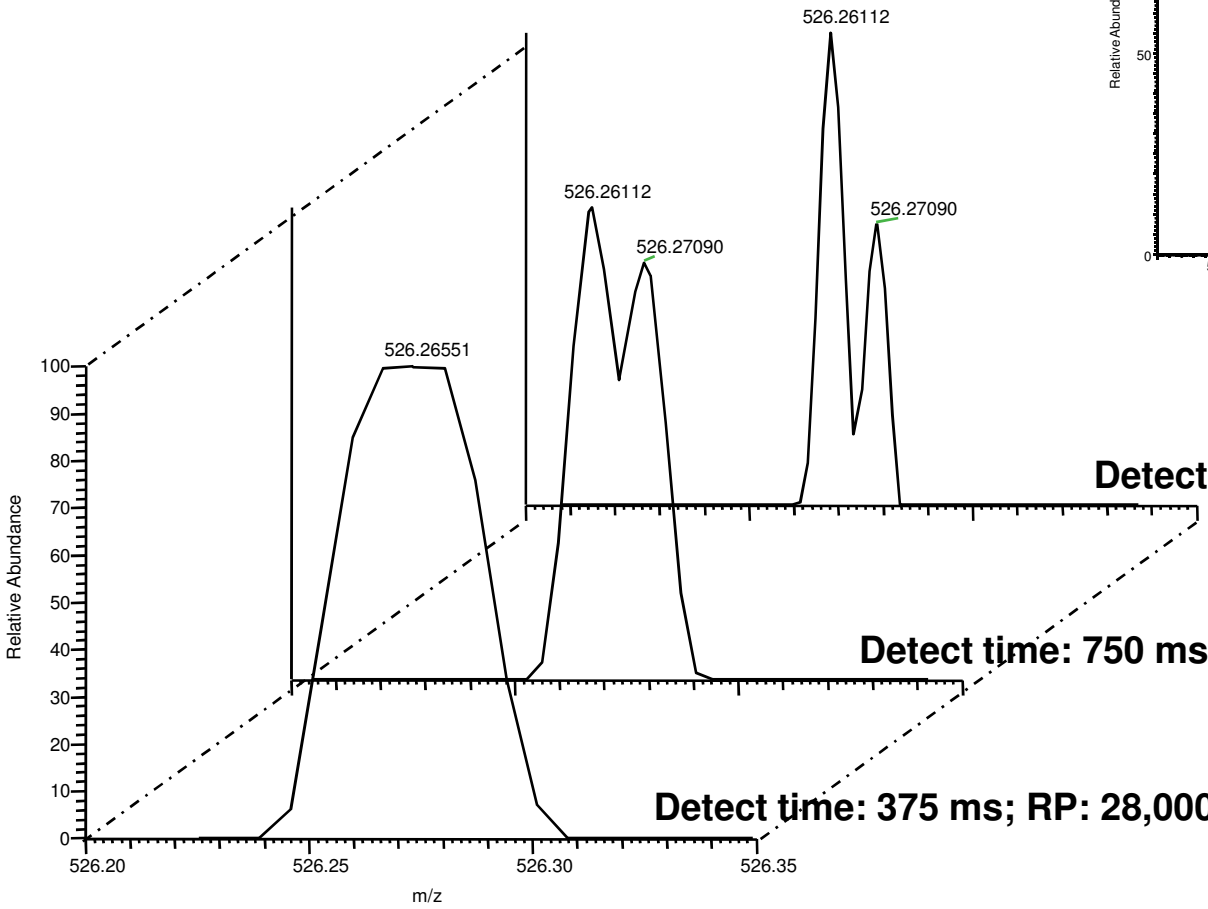
$^{13}C_2$ and ^{34}S isotopes

$C_{23}H_{38}N_7O_5{}^{34}S$ 526.2607 u $^{13}C_2C_{21}H_{38}N_7O_5S$ 526.2716 u

Δm = 0.0109 RP ($m/\Delta m$) = **48,000**



MRFA
Molecular ion region



Detect time: 1500 ms; RP: 95,000

Detect time: 750 ms; RP: 55,000

Detect time: 375 ms; RP: 28,000

Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

„Freedom of fragmentation“ (and detection)

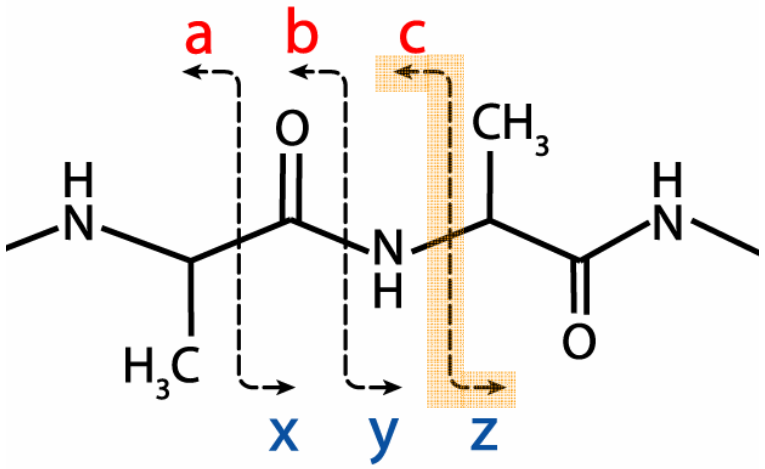
- Combination of three different and complementary fragmentation techniques **CID, HCD, and ETD** for sequence assignments with absolute confidence.

- Most comprehensive solution for
  - complex PTM analysis
  - intelligent sequencing of peptides
  - top-down and middle-down analysis
  - protein quantitation via stable isotope labelling such as iTRAQ™ or label-free quantitation

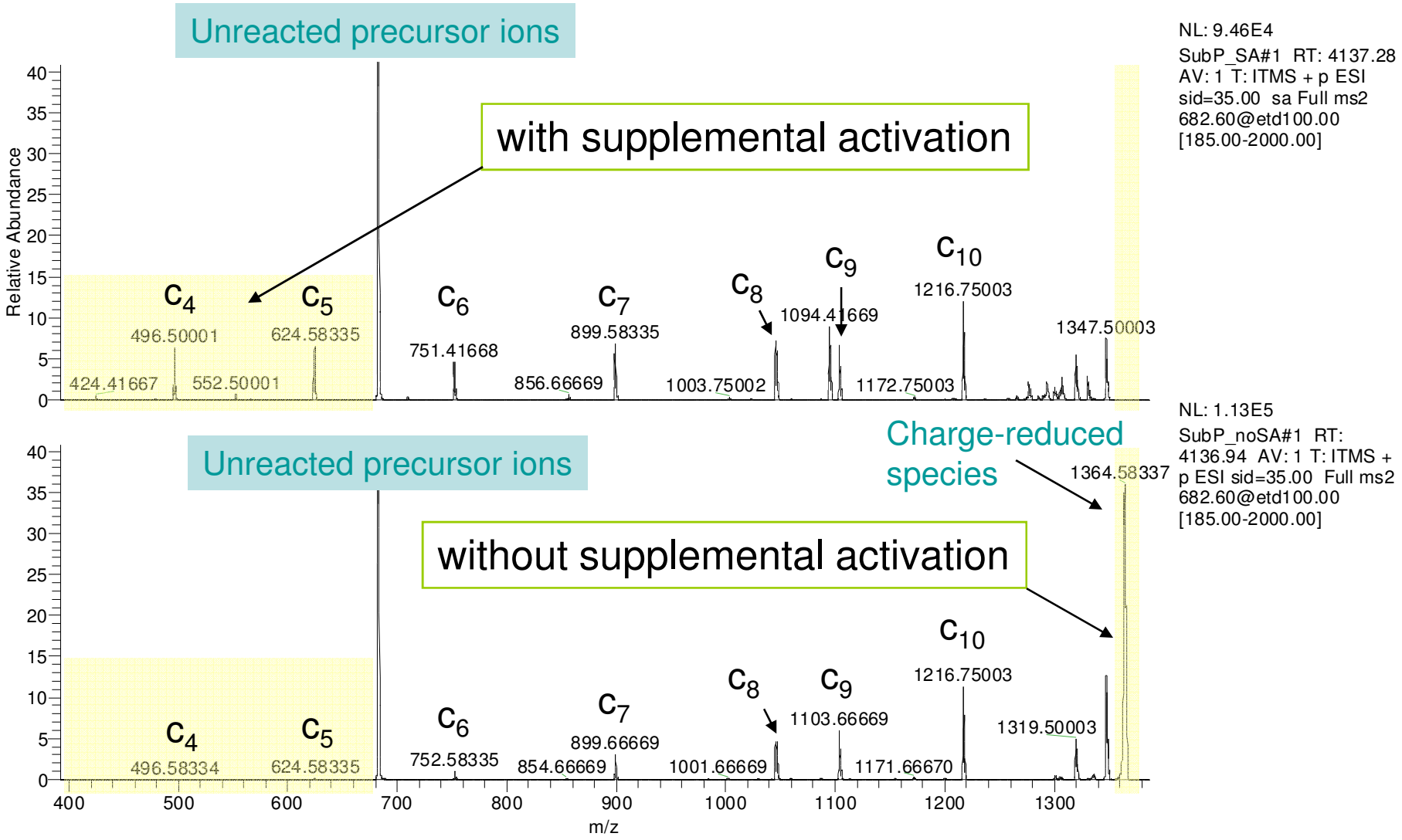
Electron Transfer Dissociation (ETD)

$$[M + 2H]^{2+} + A^{-} \rightarrow [M + 2H]^{+\bullet} + A$$

$$[M + 2H]^{+\bullet} \longrightarrow \underbrace{[c+H]^{+}}_{c} + \underbrace{[z+H]^{+\bullet}}_{z^{\bullet}}$$



Importance of supplemental activation for ETD



Example: Substance P

ETD vs CID

- ETD has no low mass cut off and provides for a complete interpretation of the spectrum.

- ETD enables sequencing of larger peptides.

- Peptide sequence information from CID and ETD spectra are complementary.

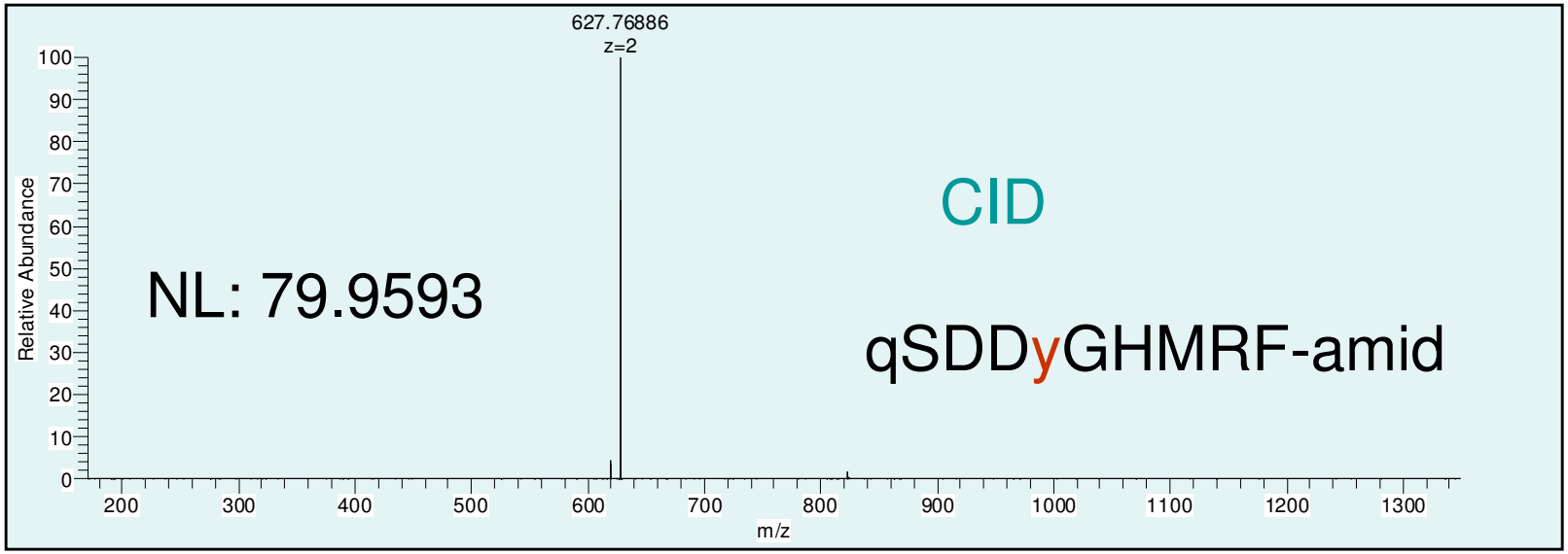
- CID/ETD toggle method improves sequence coverage and increases protein ID confidence.

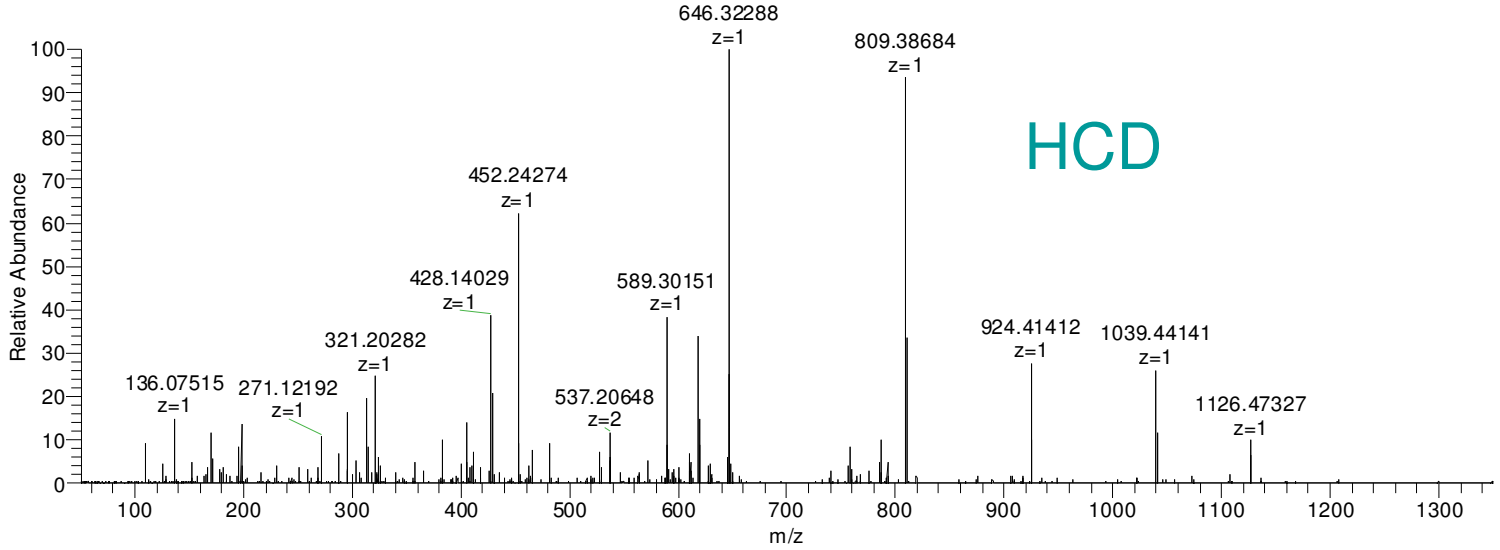
- With ETD, high quality spectra are obtained for >2+ precursor ions.

- Supplemental activation improves ETD of 2+ ions.

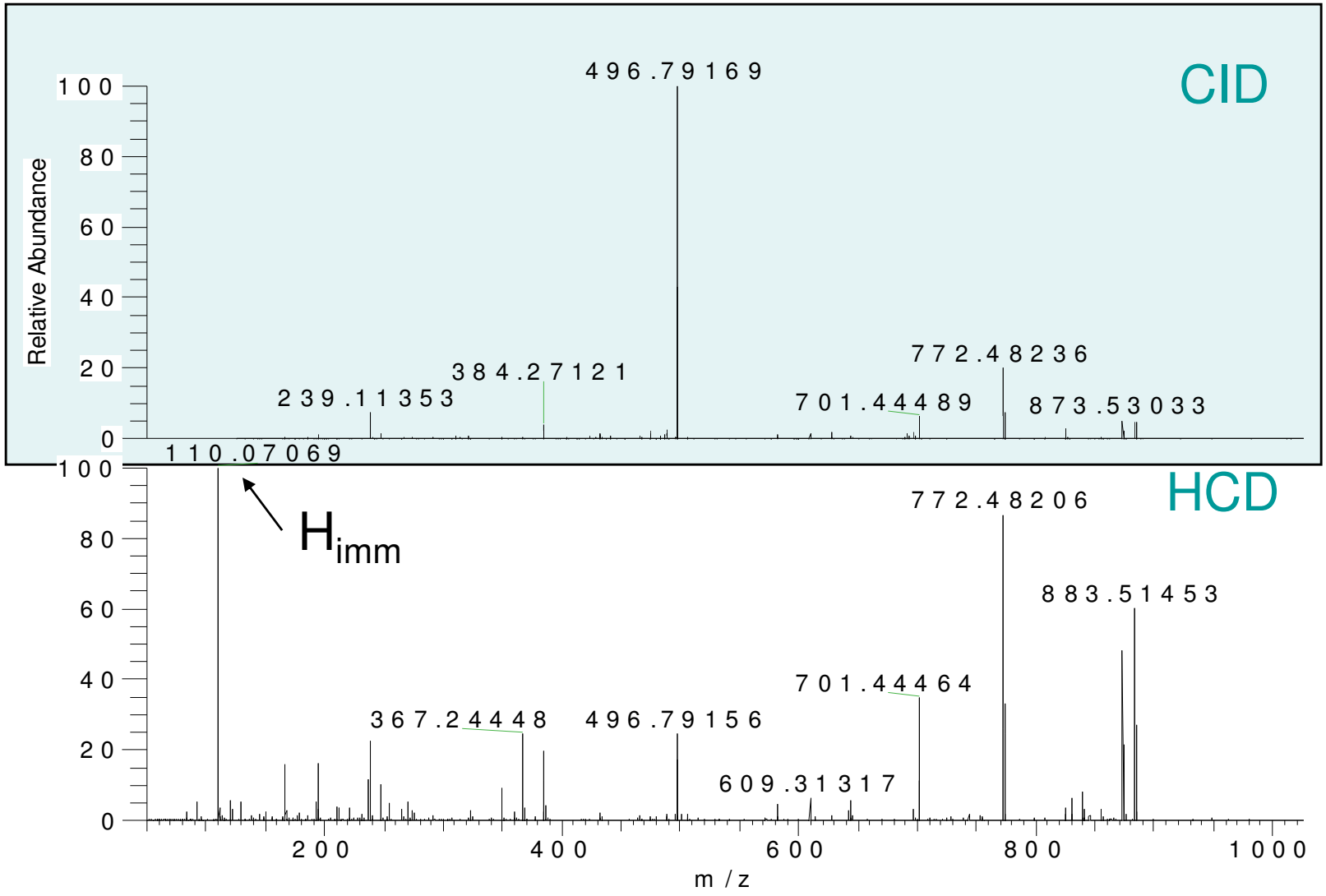
- Digestion with LysC, AspN and GluC may provide additional benefits for proteome analysis with ETD.

CID and HCD of 667.7485^{2+} (Sulfatation)





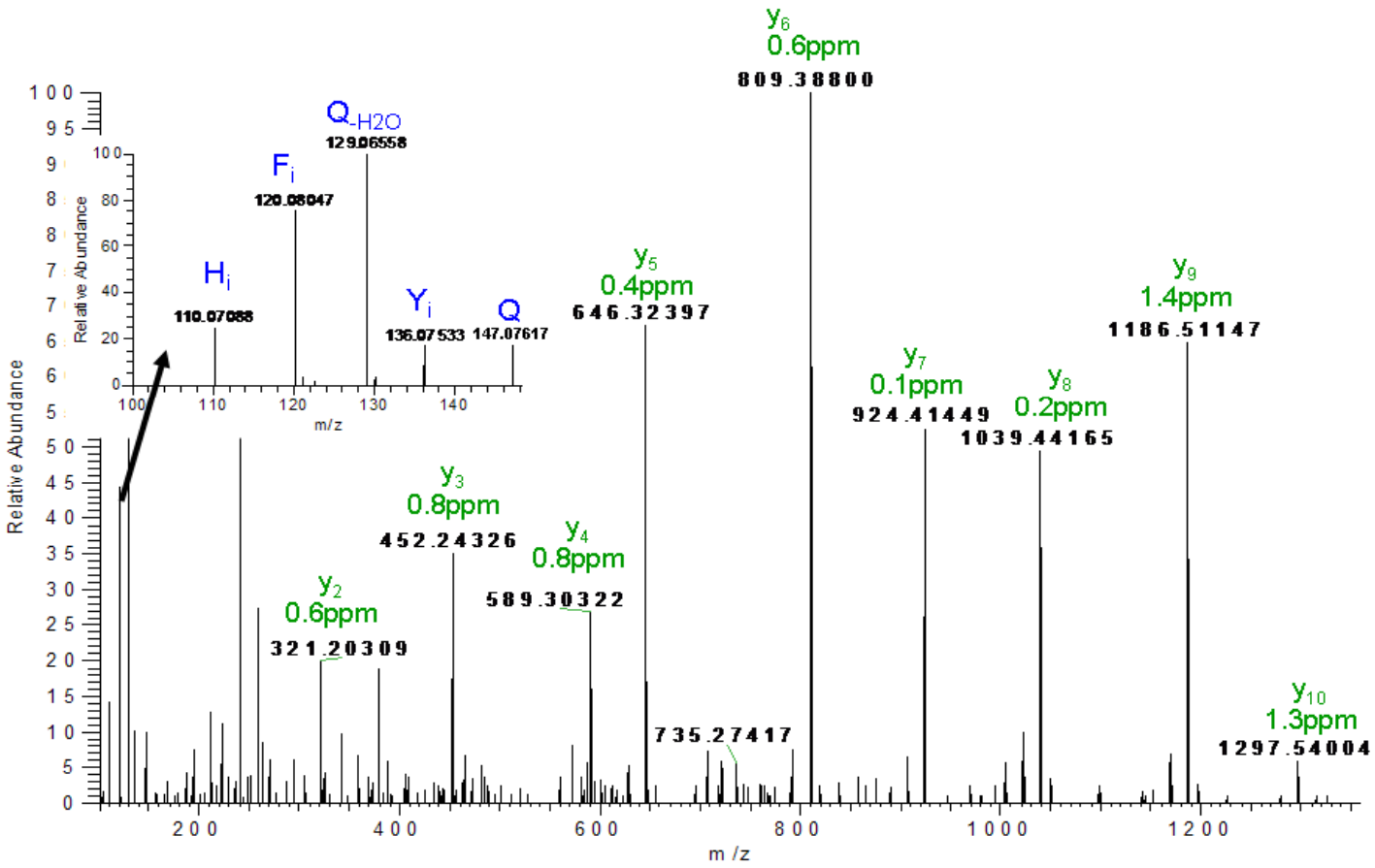
HCD vs CID **HTAGFIPRL-amide**



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

HCD MS/MS Spectrum of Perisulfakinin



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

Methods for protein analysis

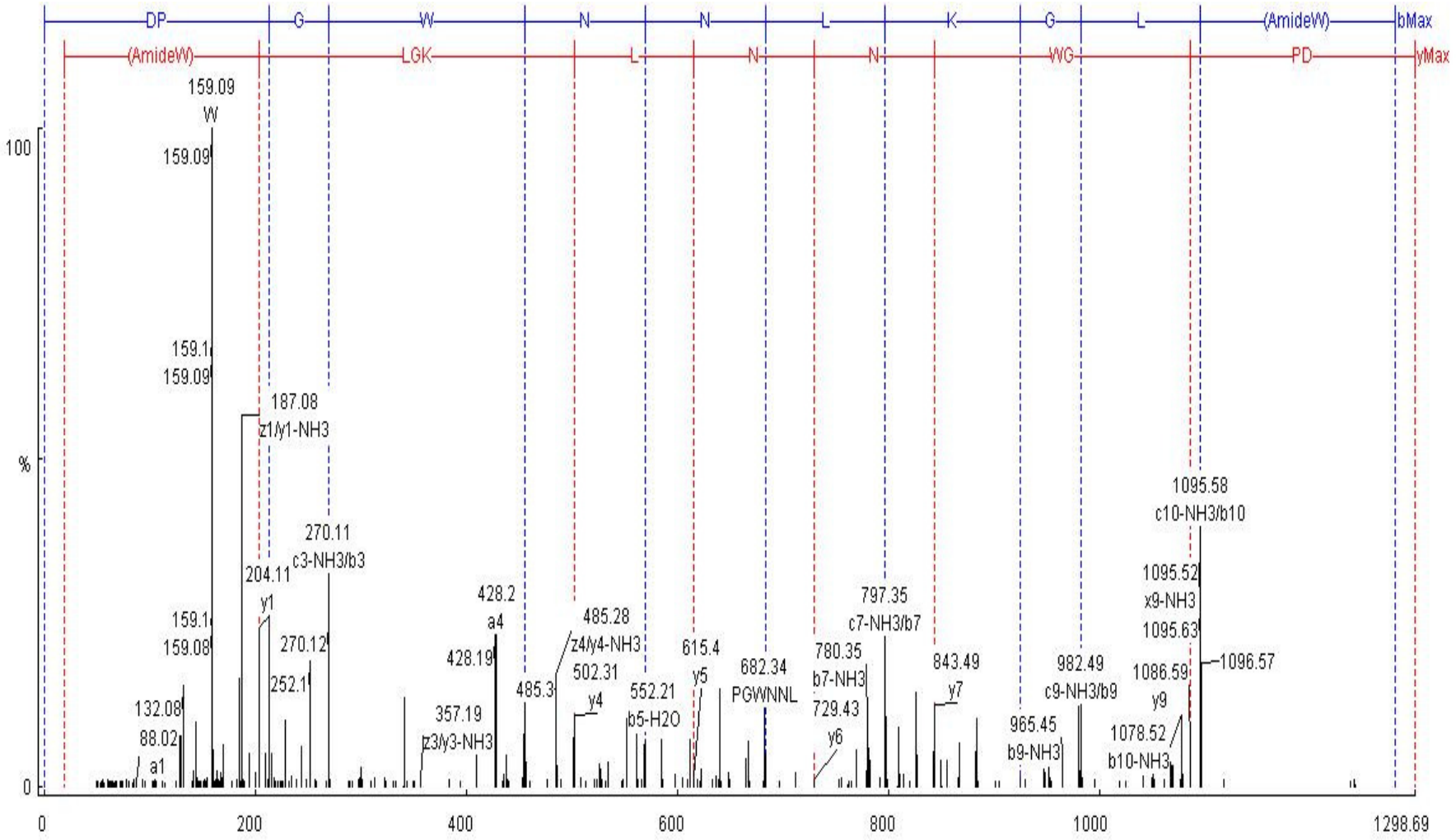
High resolution mass spectrometry methods
Top-down methods
Intact proteins

Bottom-up methods
Digested peptides

De Novo Sequencing

Gel based methods
One/Two-dimensional gel electrophoresis

Automated *de novo* Sequencing of $DPGWNNLKGLW_{amide}$ using PEAKS™ Software



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

Problems !!!

- You have two ladders of overlapping masses
- Cannot tell which is a *b* ion or which is a *y* ion
- Incomplete fragmentation
- Chemical noise
- Mass accuracy of some instruments are not good enough to determine R groups
  - Glutamine and lysine differ by 0.036 .
  - Phenylalanine and oxidized methionine differ by 0.033
  - Gly + Val = 156.090 u and Arg = 156.101

Big Problem

- Can take you a very long time to “sequence” a “good” product ion spectra without a computer
  - 30 minutes if your good
  - 1-2 days to never if you are not
- One experiment can generate 10,000 MS/MS spectra

Beside the amount...

- Proteins due to their nature are very complex
  eg. cysteine – cysteine cross-linking, tertiary structure formation and post-translational modifications)

- Proteins can be very large in size and hard to handle.

Decrease complexity

- **Separate proteins (before digestion)**
  - 2-d electrophoresis
  - 1-d electrophoresis
  - Chromatography (RP, cation/anion exchange, size excision)
  - Immunoafinity
  - Molecular weight-centrifugation, ultrafiltration, solvent precipitation

- **Separate peptides (after digestion)**
  - Multidimensional chromatography (MUDPIT)

Reducing agents – detergents (SDS, CHAPS), salts (urea, guandine), acids (formic)

Break sulfur bridges – DTT and use iodoacetamide to keep them from reforming

Heat

Proteomics and Gel Electrophoresis

Methods for protein analysis

High resolution mass spectrometry methods
Top down methods
Intact proteins

Bottom up methods
Digested peptides

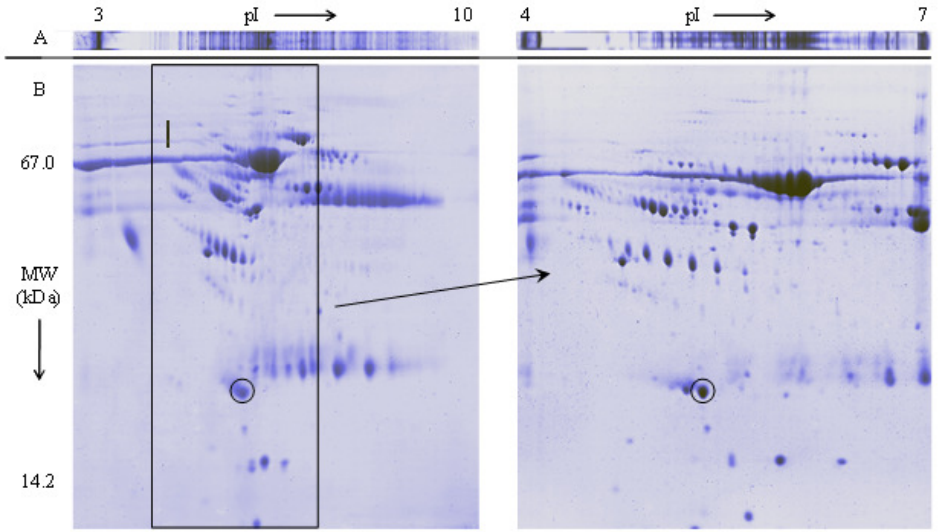
De Novo Sequencing

Gel based methods
One/Two-dimensional gel electrophoresis

2-D electrophoresis

2-DE is a powerful separation technique, which allows simultaneous resolution of thousands of proteins.

2-D electrophoresis



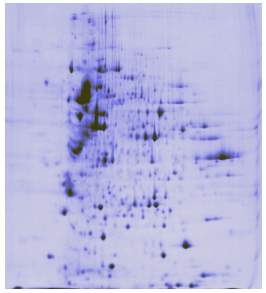
Disadvantages

- Modest detection limit
- High abundance proteins
- Protein bias (pI)
- Difficult to automate
- Labor intensive
- Requires many more mass spec runs

Advantages

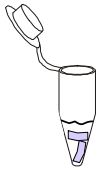
- High resolution separation
- Can get quantitation by staining
- Good "**snapshot**" of the Proteome!

Approach used to identify a spot from a 2-D gel analysis



2-D gel

Excision of protein spot



Gel spot

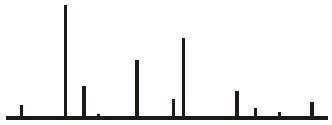
Digestion Extraction

Peptide fragments

Mass spectrometer

MS

Peptide mass fingerprinting

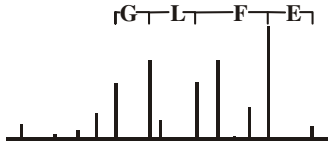


Search

Database (NR, EST)

Search

Peptide sequence tagging

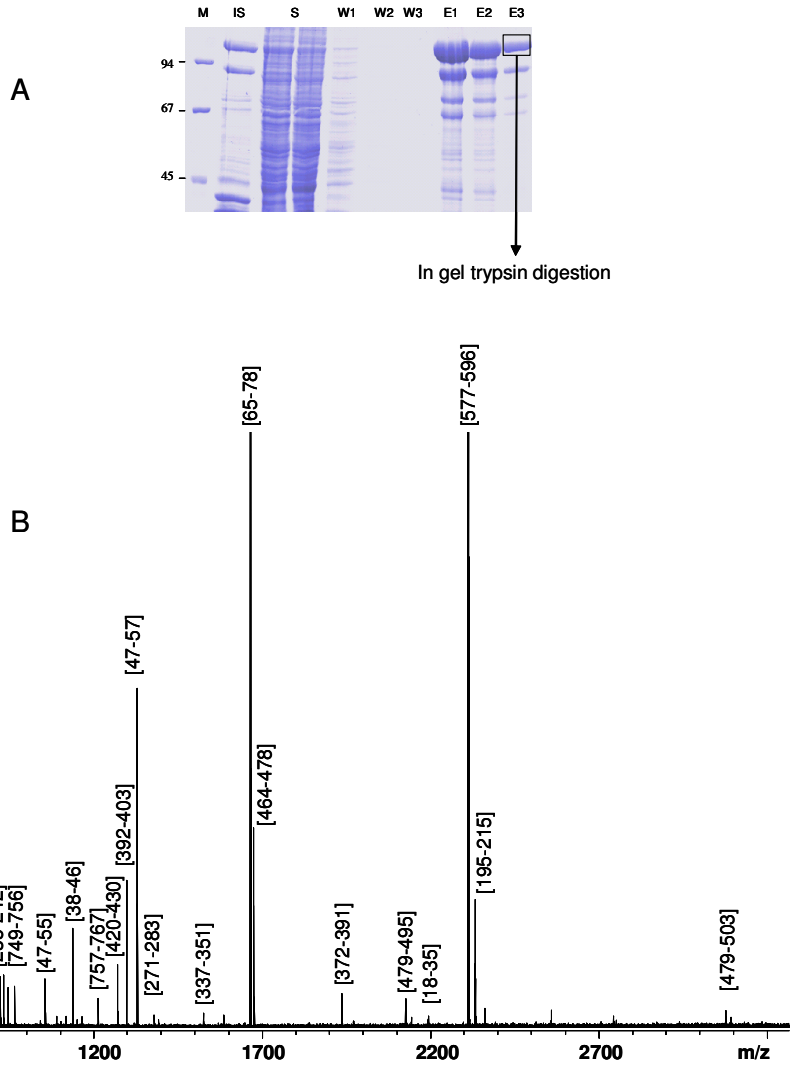


MS/MS

Selected ion

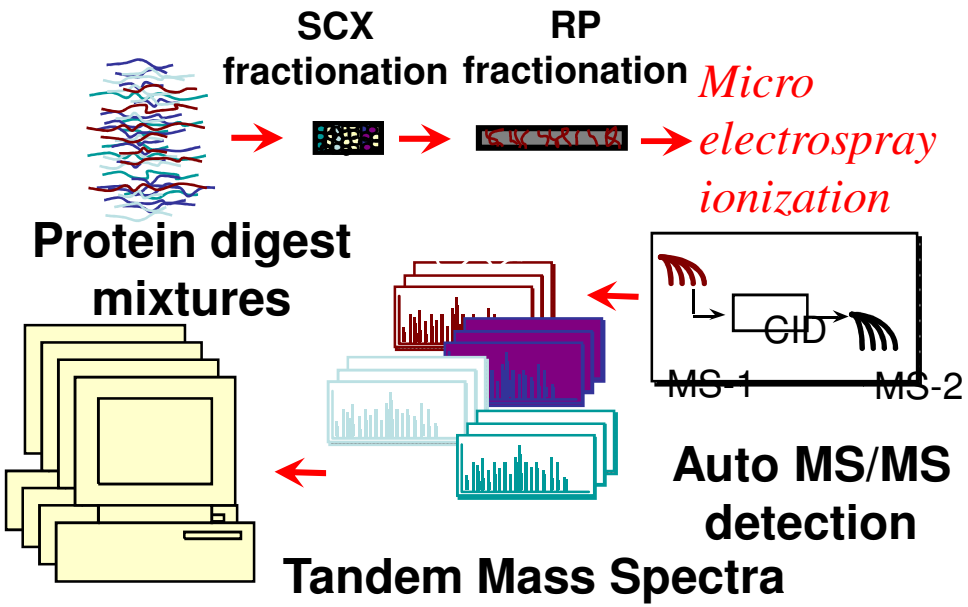
Protein identification

1-D electrophoresis



- **Disadvantages**
  - Difficult to automate
  - Very limited expression profiling
- **Advantages**
  - No staining necessary
  - High protein recovery
  - Decreased runs needed
    - 2-50 proteins can be identified per band or section

Typical MUDPIT Preparation (Multidimensional Protein Identification Technology)



Database search protein I.D.

- **Advantages**
  - Dynamic Range 10^5
  - Sensitivity! Low abundance proteins
  - Minimized protein bias
  - Highly automated
- **Disadvantages**
  - Poor isoform & modification distinction
  - Still overwhelms the mass spectrometer
  - Does not give you a very good "snapshot"

Post Translational Modification

After Translation

Post-translational modifications

- Phosphorylation – Ser, Thr, Tyr
- Methylation – Lys
- Oxidation – Met
- Deamidation – Asn, Gln
- Glycosylation

PTM Scanning

- Can search data for unexpected modifications from the unimod database

  www.unimod.org

Ways to look for PTM's

- **Mass spectrometry**
  - Good for PTM's with a stable mass
    - Phosphorylation
    - Methylation
    - Acetylation
  - Difficult to analyze large modifications
    - Large carbohydrates

- **Edman Sequencing**
    - Phosphorylation

- **Radio-labeling**
  - Very sensitive
  - Usually hard to get the identity or AA with PTM

The Trinity of Protein Phosphorylation Analysis

I. **Identify the kinase responsible for phosphorylation**

- Predict potential phosphorylation sites with the Scansite program (http://scansite.mit.edu)
- Gel kinase assay

II. Identify the sites of phosphorylation

- Separate phosphoproteins by SDS-PAGE
- Digest in-gel the phosphoproteins to generate (phospho)peptides
- MS/MS analysis

III. **Identify the function of phosphorylation**

- Generate mutants forms of protein for *in vitro* testing

Most important part of finding PTM's

- Protein Coverage!

Problems with Mass spec approach

- **Cannot prove a negative**
  - Just because you do not see your PTM does not mean it is not there!

Some common PTM's

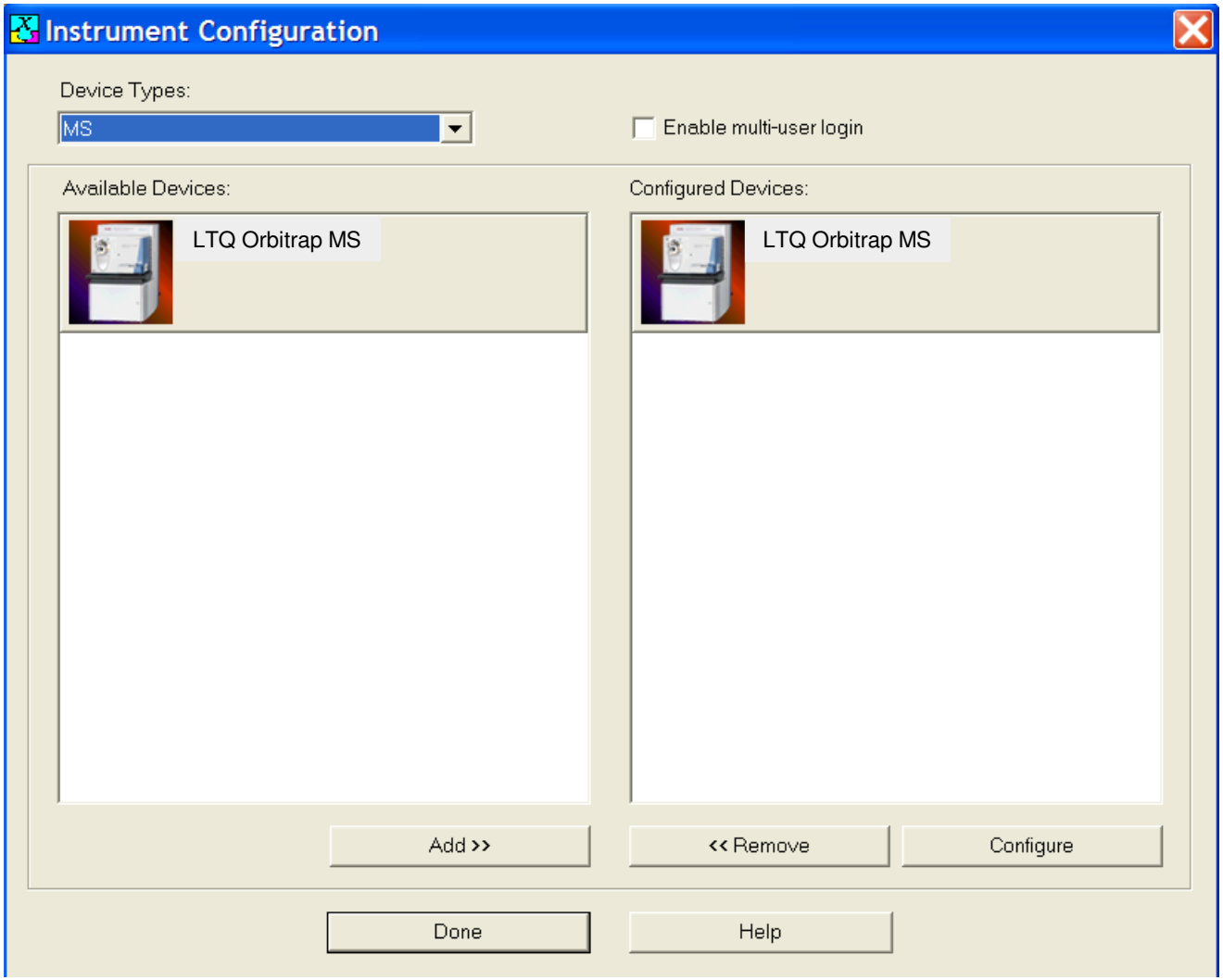
Table 1. Some common and important post-translational modifications

PTM type	ΔMass[a] (Da)	Stability[b]	Function and notes
Phosphorylation			Reversible, activation/inactivation of enzyme activity, modulation of molecular interactions, signaling
pTyr	+80	+++	
pSer, pThr	+80	+/++	
Acetylation	+42	+++	Protein stability, protection of N terminus. Regulation of protein–DNA interactions (histones)
Methylation	+14	+++	Regulation of gene expression
Acylation, fatty acid modification			Cellular localization and targeting signals, membrane tethering, mediator of protein–protein interactions
Farnesyl	+204	+++	
Myristoyl	+210	+++	
Palmitoyl	+238	+/++	
etc.			
Glycosylation			
N-linked	>800	+/++	Excreted proteins, cell–cell recognition/signaling
O-linked	203, >800	+/++	O-GlcNAc, reversible, regulatory functions
GPI anchor	>1,000	++	Glycosylphosphatidylinositol (GPI) anchor. Membrane tethering of enzymes and receptors, mainly to outer leaflet of plasma membrane
Hydroxyproline	+16	+++	Protein stability and protein–ligand interactions
Sulfation (sTyr)	+80	+	Modulator of protein–protein and receptor–ligand interactions
Disulfide bond formation	–2	++	Intra- and intermolecular crosslink, protein stability
Deamidation	+1	+++	Possible regulator of protein–ligand and protein–protein interactions, also a common chemical artifact
Pyroglutamic acid	–17	+++	Protein stability, blocked N terminus
Ubiquitination	>1,000	+/++	Destruction signal. After tryptic digestion, ubiquitination site is modified with the Gly-Gly dipeptide
Nitration of tyrosine	+45	+/++	Oxidative damage during inflammation

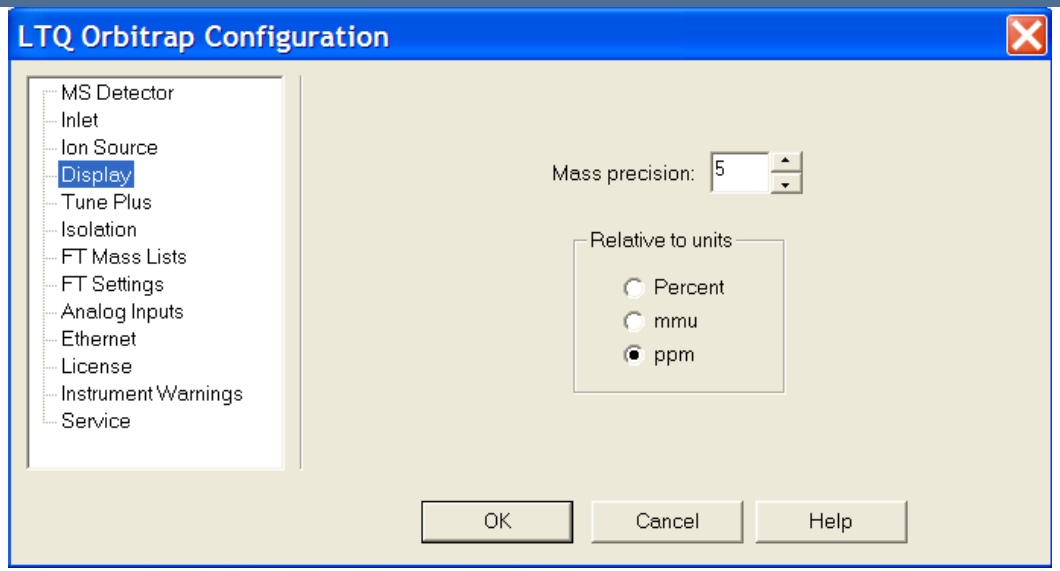
From Jensen et al. nature march 2003 vol 21

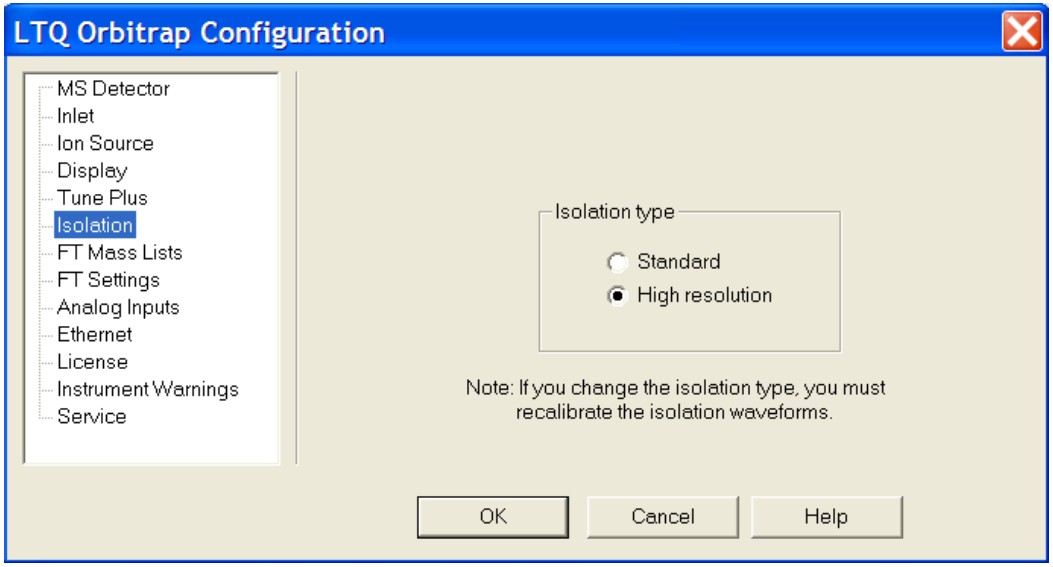
Data Dependent NL MS3 method setup for phopshorylation identification

Instrument configuration settings



Configuration window: important settings





Thermo SCIENTIFIC

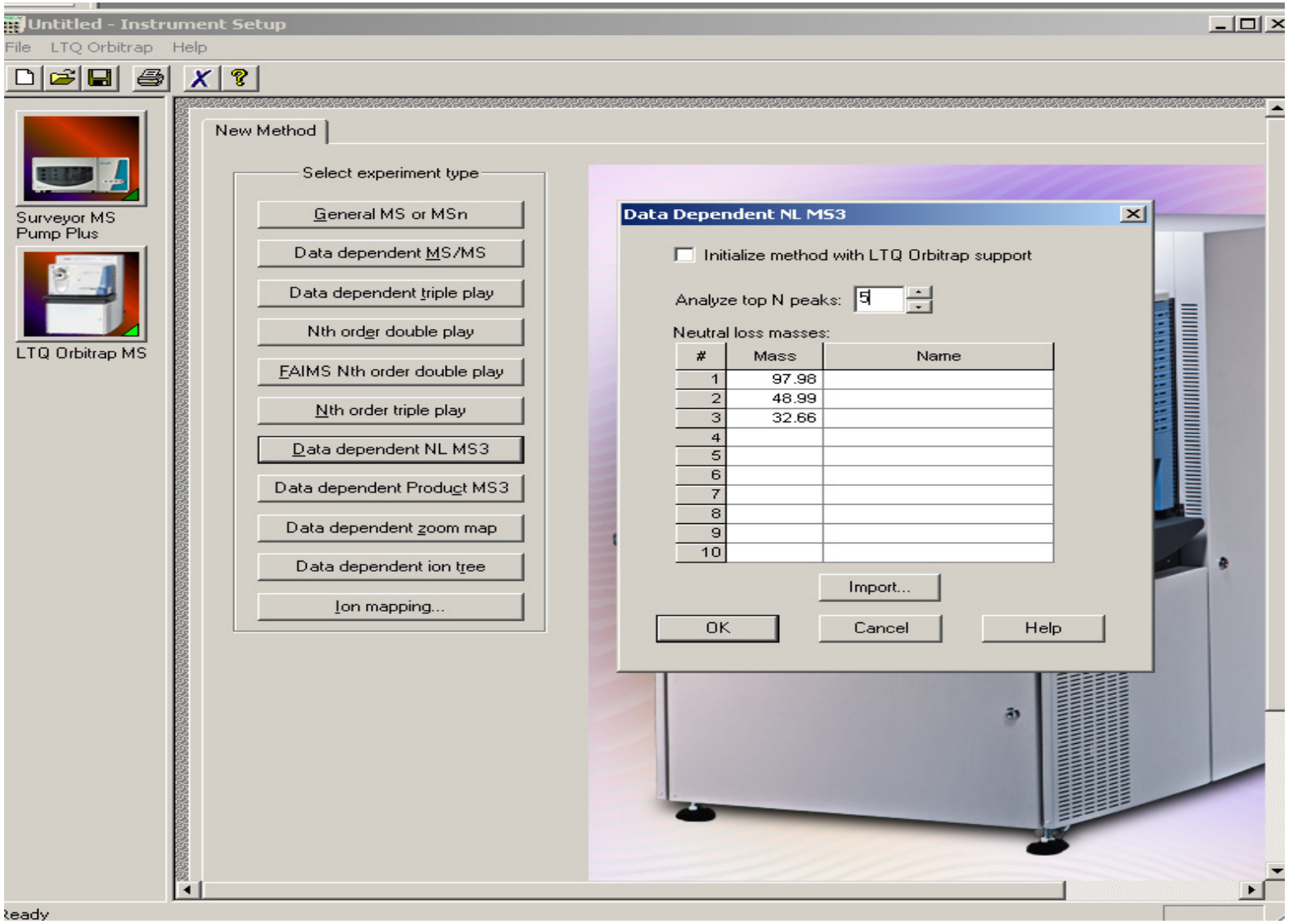
ThermoFisher SCIENTIFIC

Data Dependent NL MS3

Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

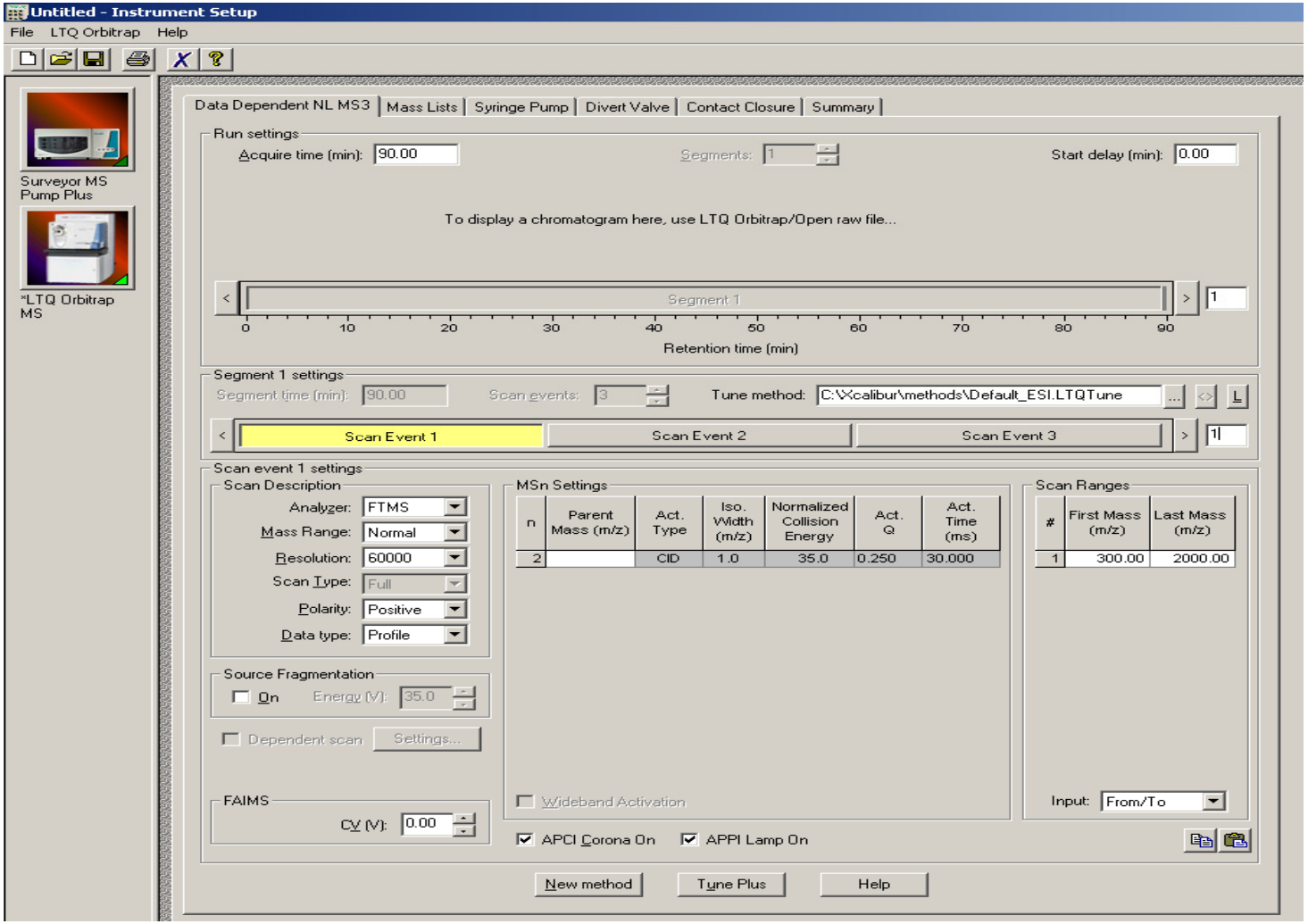
NL in MS2 trigger MS3



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

Scan events

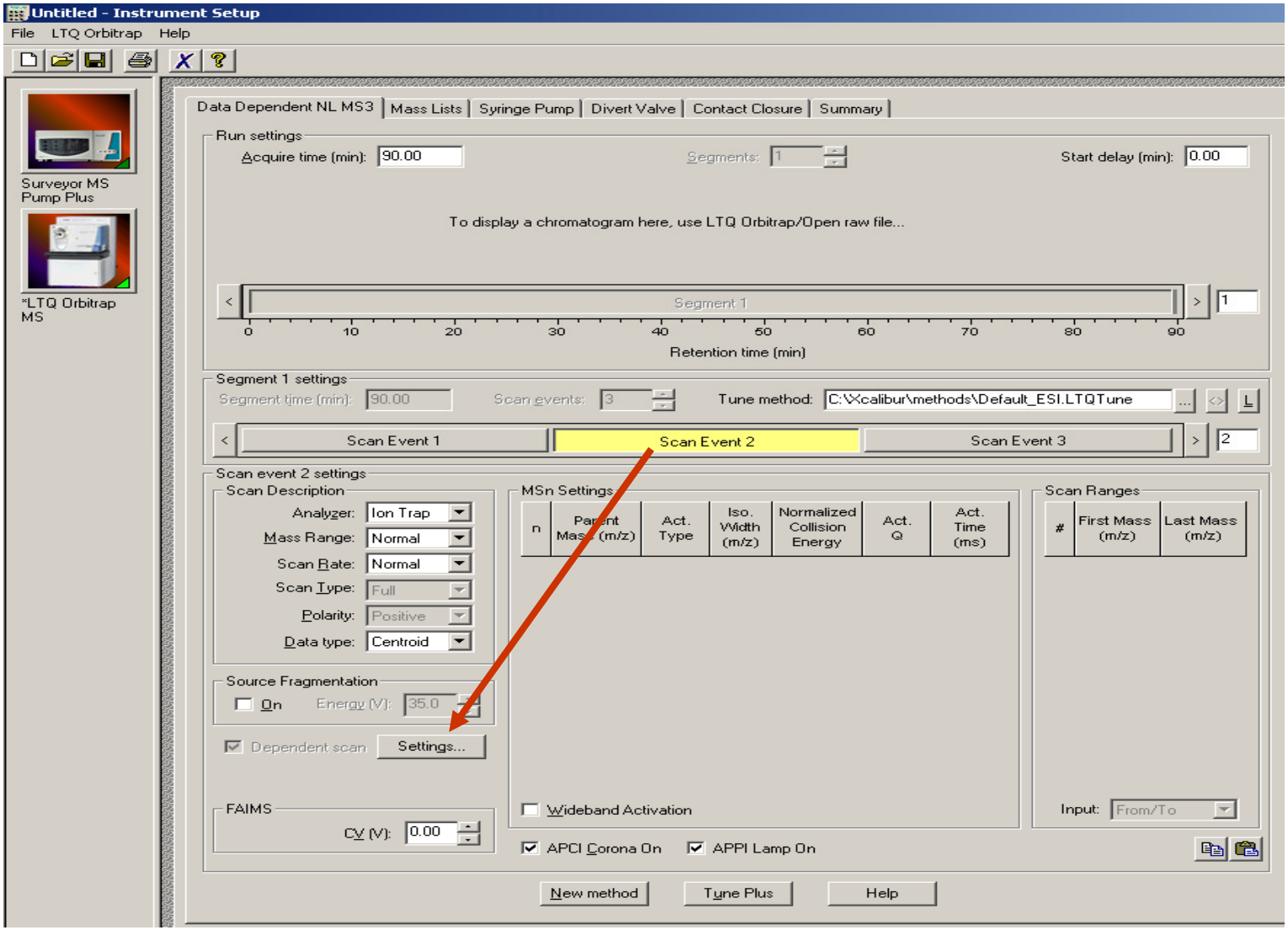


1. First scan event is the reference scan (usually a FTMS full scan)

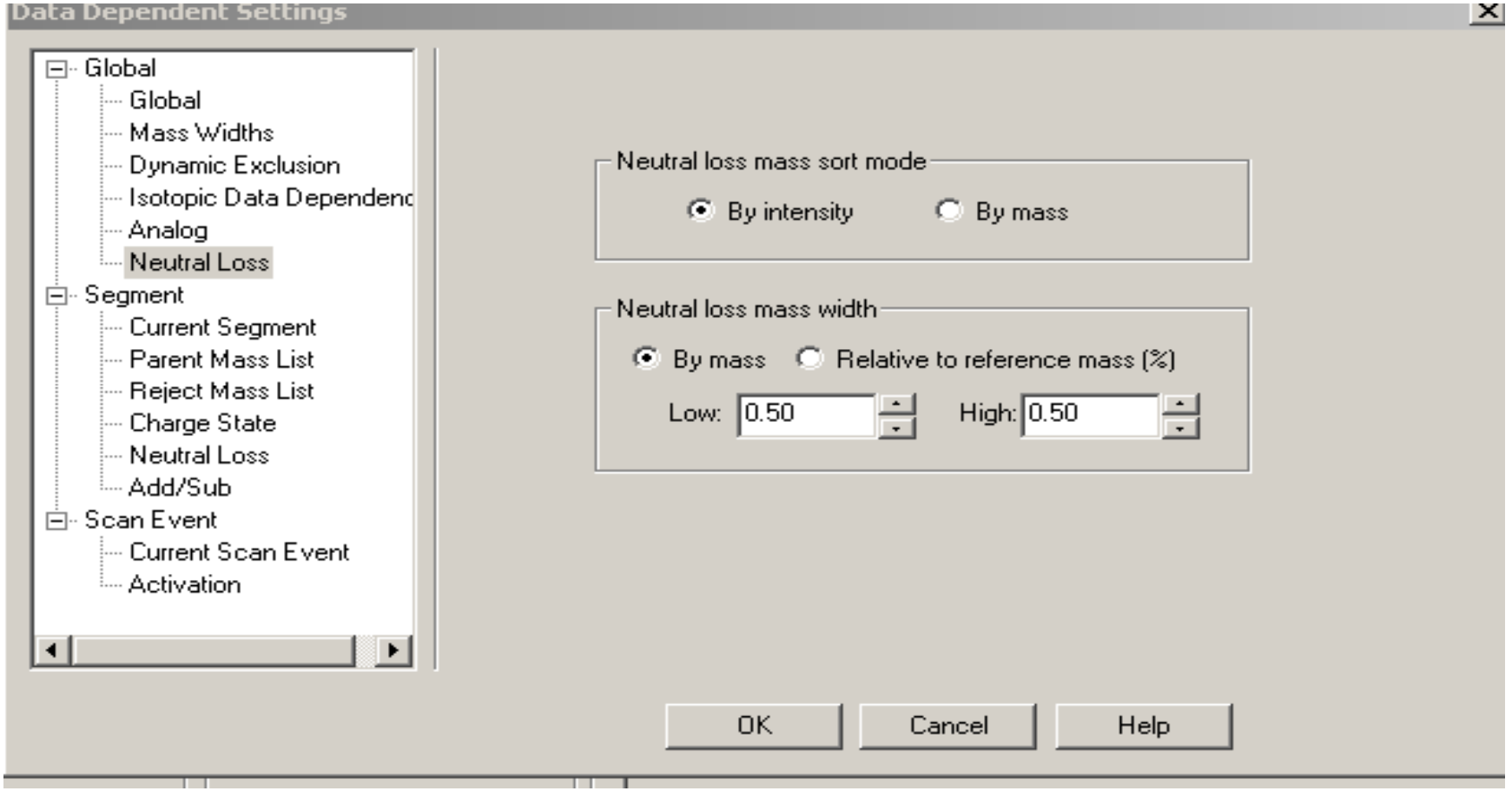
2. All subsequent scans events may be dependent on scan event 1.

3. A scan depending on a MS scan will produce a MS/MS spectrum. A scan depending on a MS/MS scan will produce a MS^3 spectrum

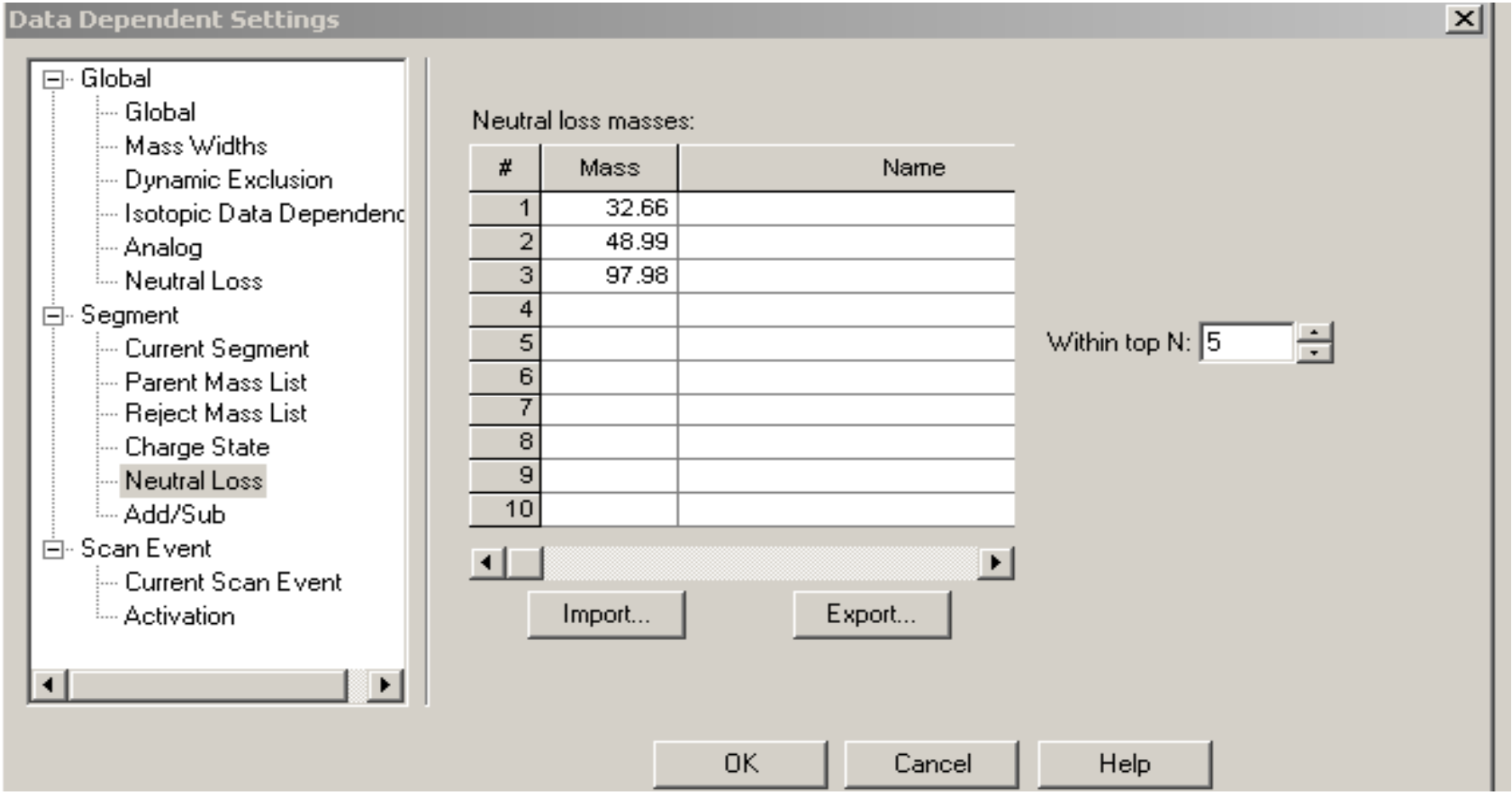
Scan event 2



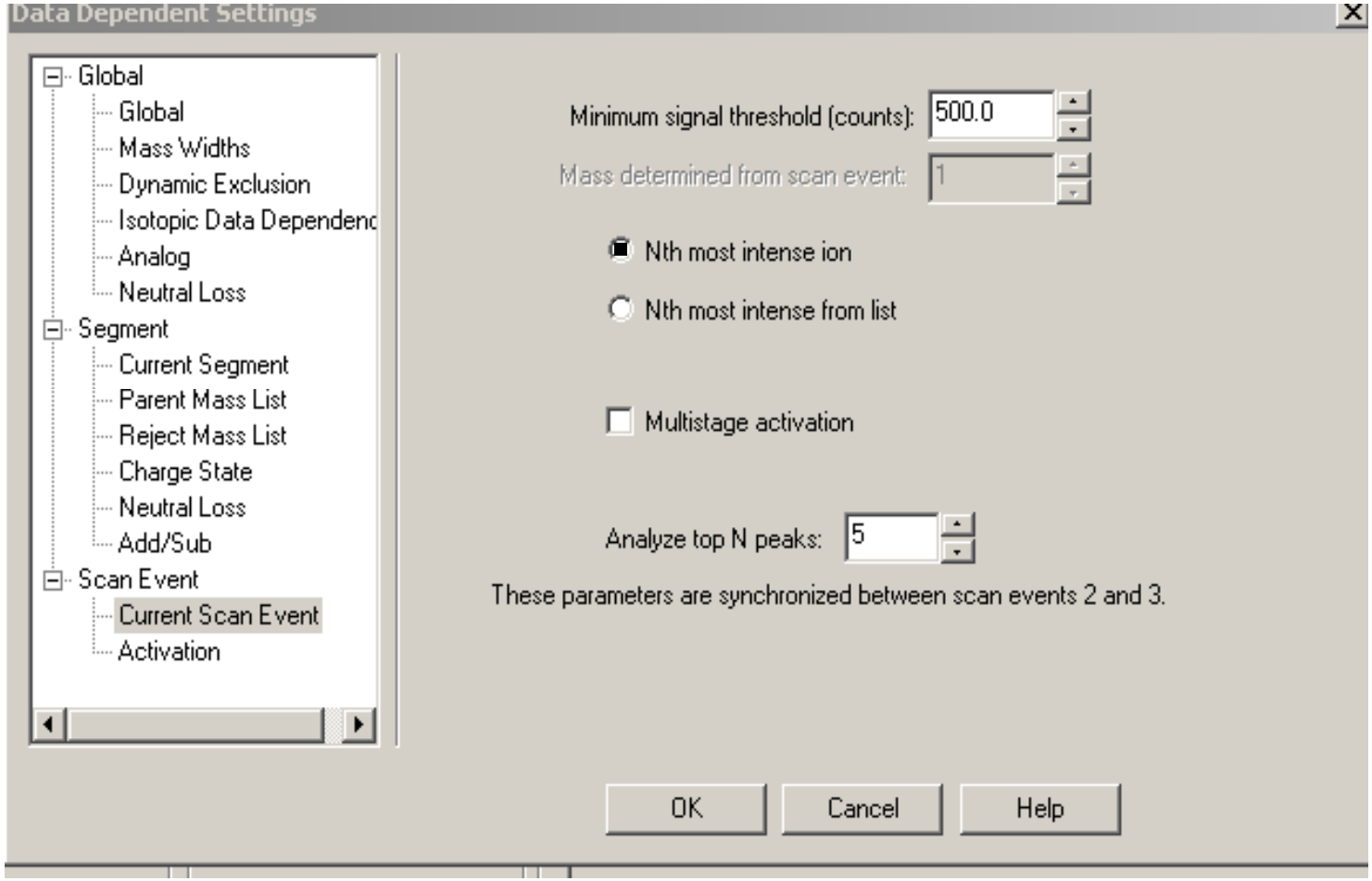
Scan event 2. Settings



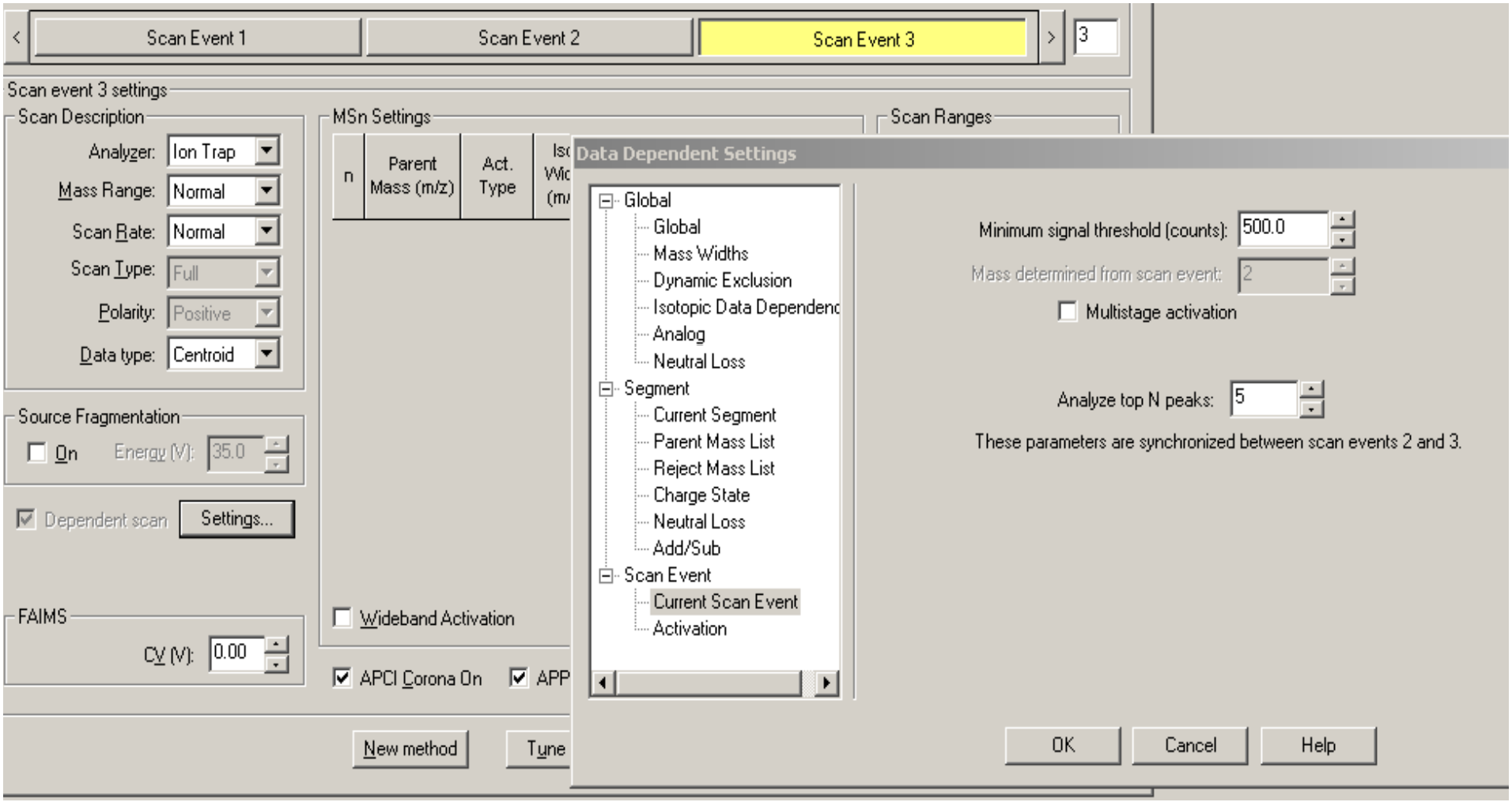
NL masses



Scan event 2 determined from Scan event 1



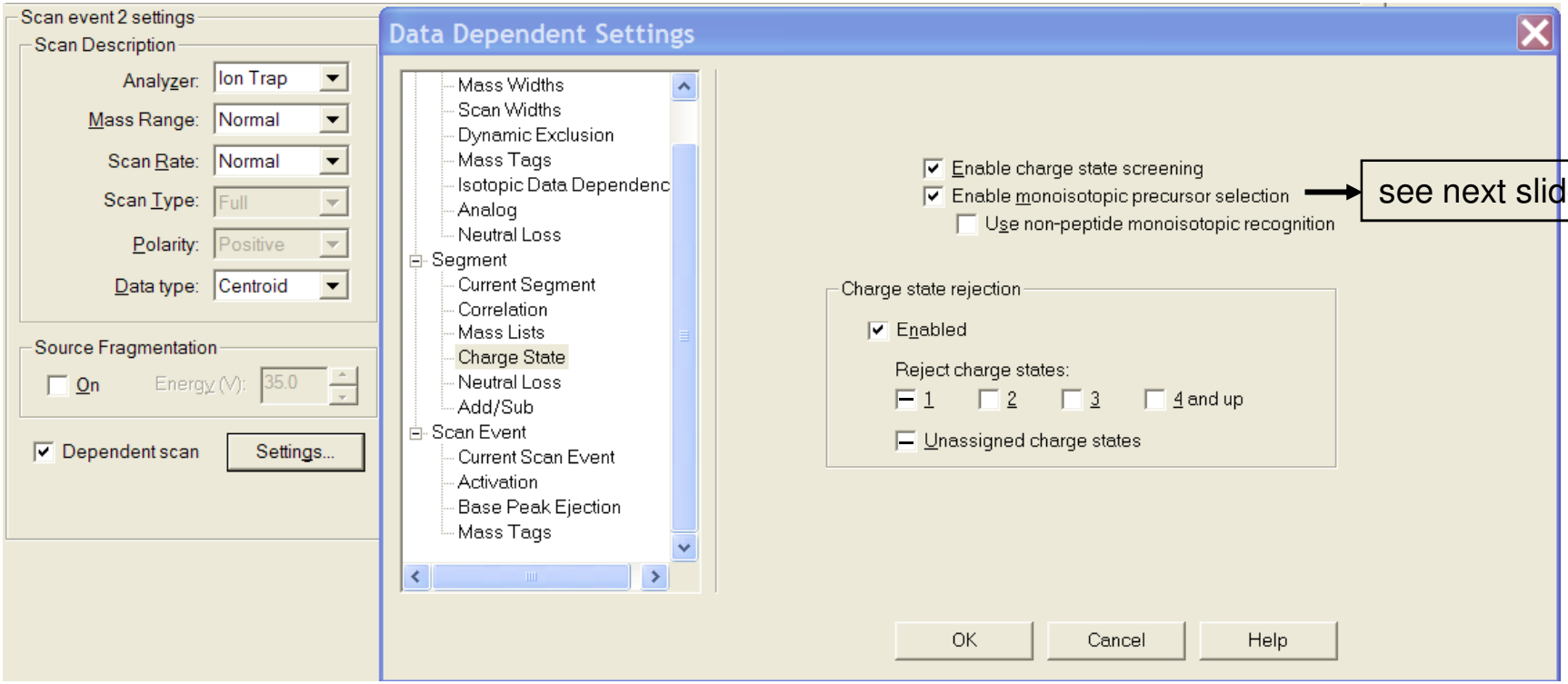
Scan event 3 (MS3 of the peaks which lost phosphate)



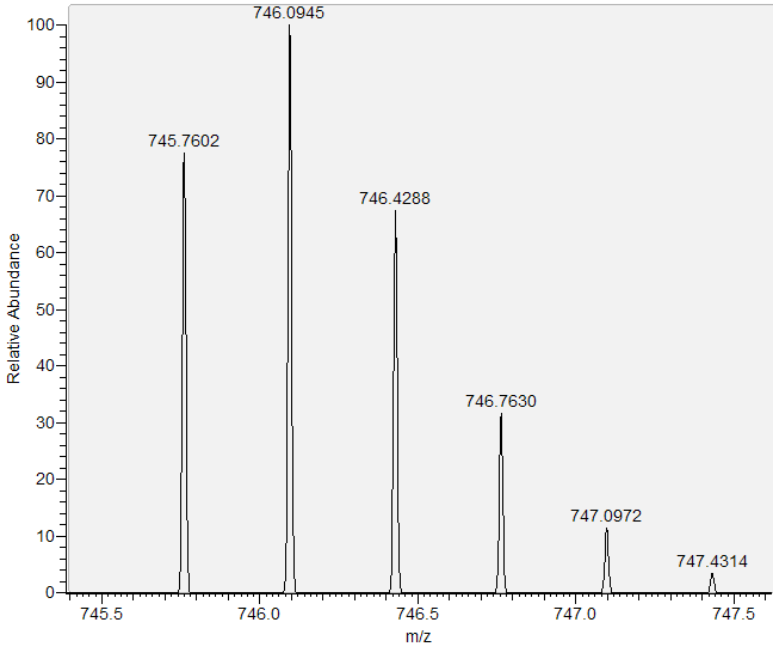
Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

Monoisotopic precursor selection toggle…

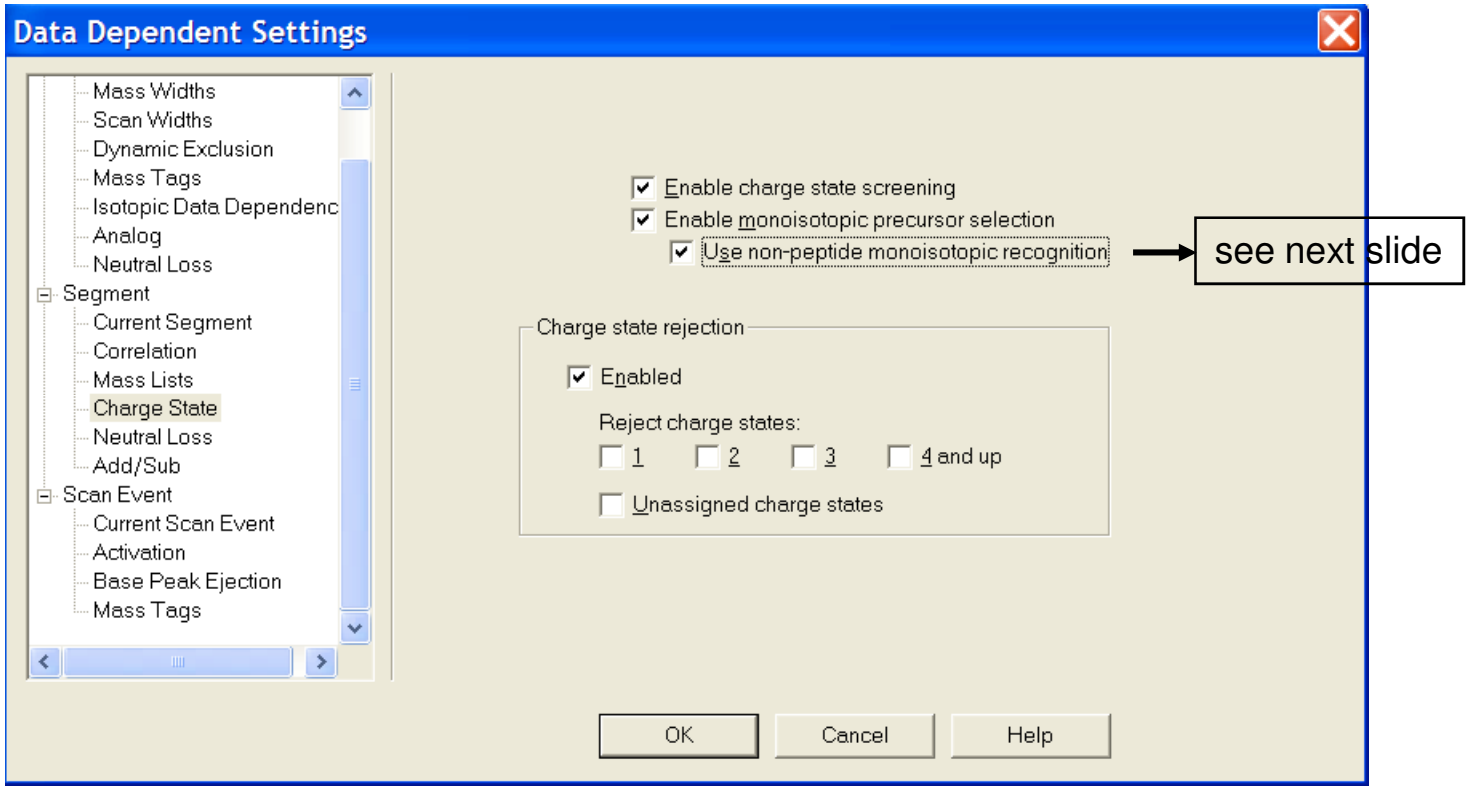


… And What It Does



- Problem: monoisotope signal is lower abundant than first isotope ⇒ wrong precursor ion mass would be used for db searches ⇒ wrong results
- Monoisotope Toggle reads the charge state…
- …calculates a theoretical pattern for the measured m/z…
- …determines the correct monoisotope signal ⇒ correct results from db search

Non-Peptide Monoisotopic Recognition…



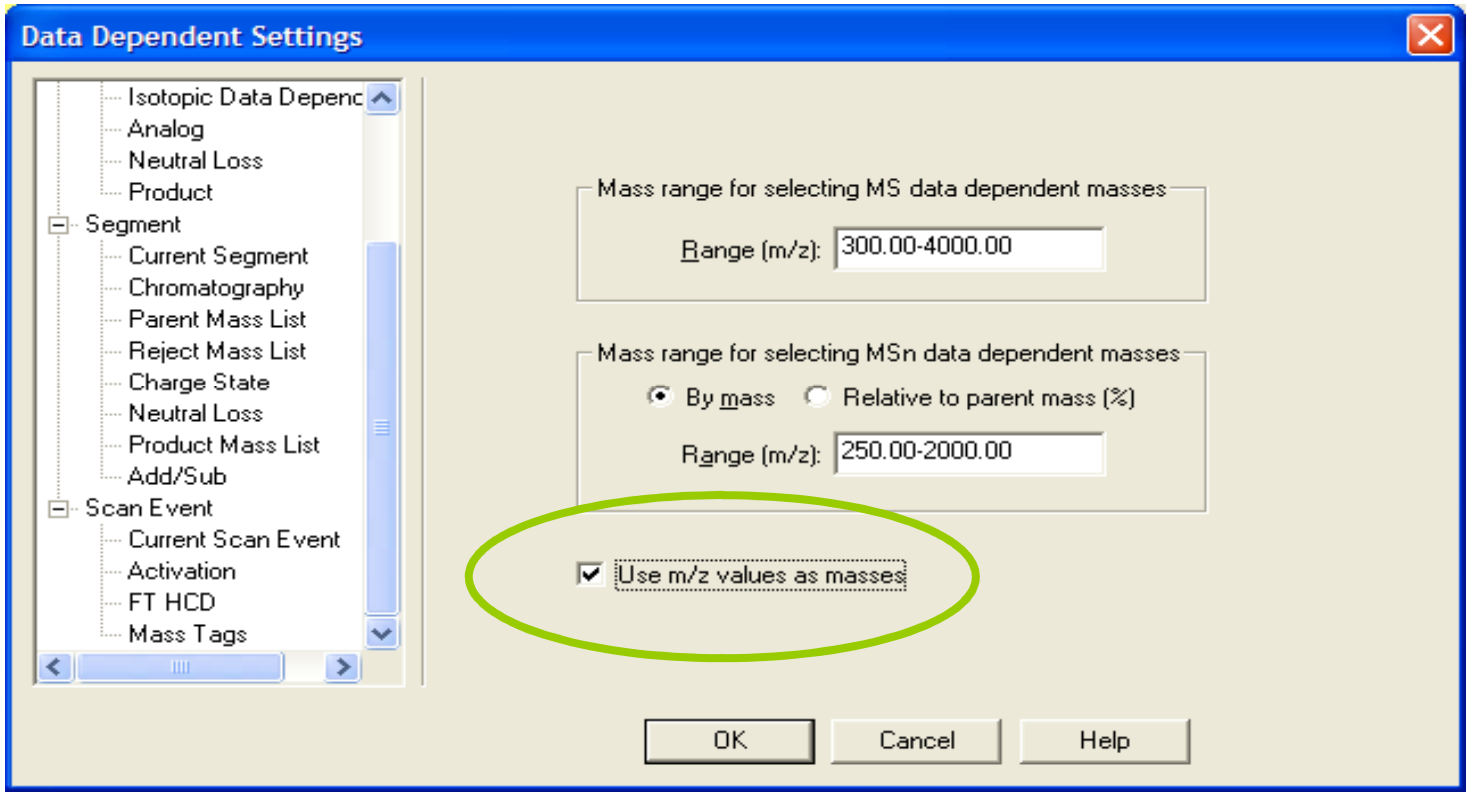
…And What It Does

- determines a charge state (isotope distance) for a detected signal cluster
- looks for a signal left from the most abundant signal
- this signal needs to have at least 60% intensity of the highest one
- the toggle will only work if the isotope cluster follows a ‚standard intensity deviation' → ‚Cl' or ‚Br' containing cluster will not work

New features

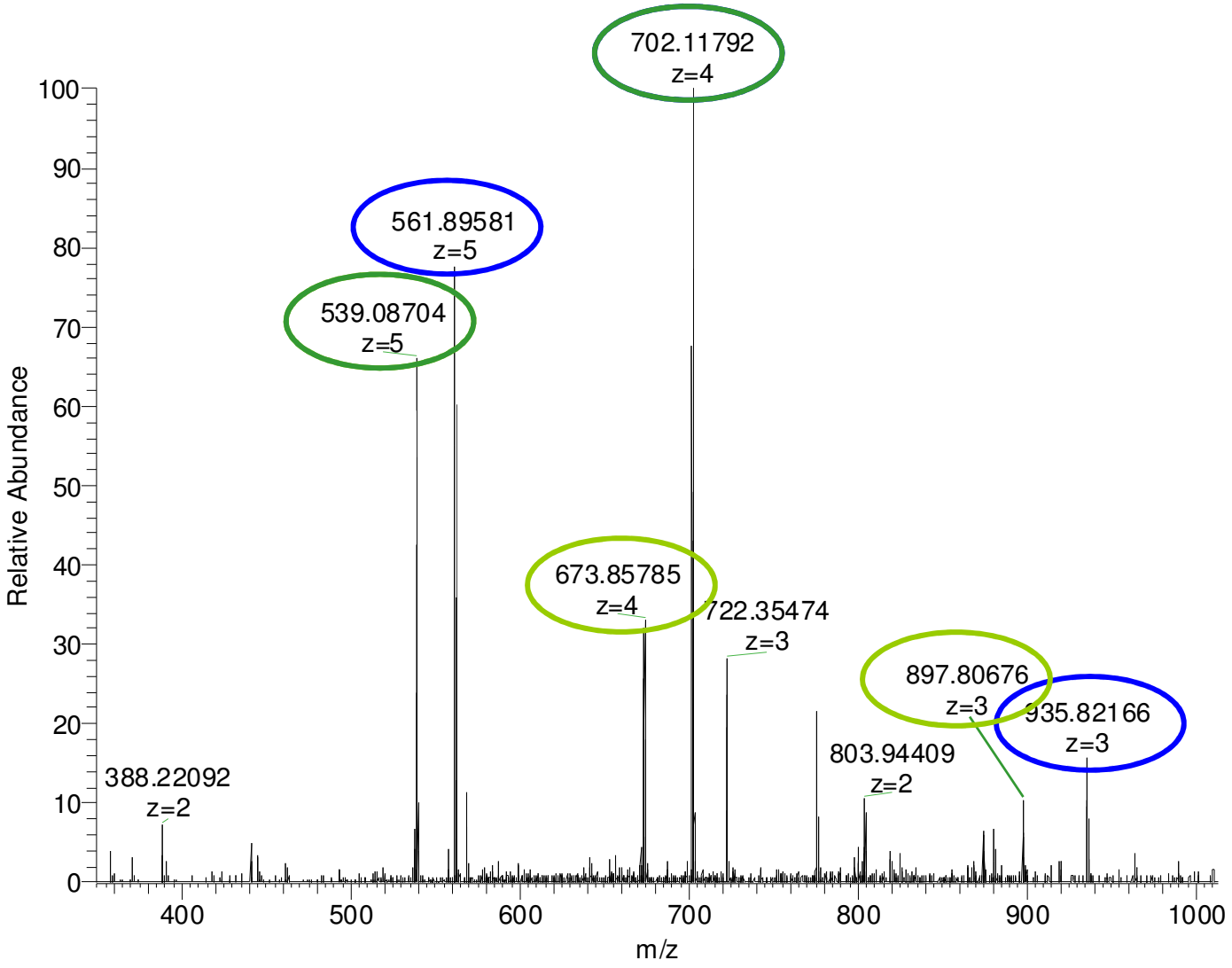
- **Use m/z values as Mass**
  - Triggers only one charge state out of a CS distribution
- **Chromatography Trigger**
  - Peak apex triggering
- **NL Triggered HCD**
  - Performs HCD fragmentation on precursors, which carry modifications, which result in CID spectra with NL

“Use m/z values as Mass”



- triggers only the most intense charge state of a charge state distribution

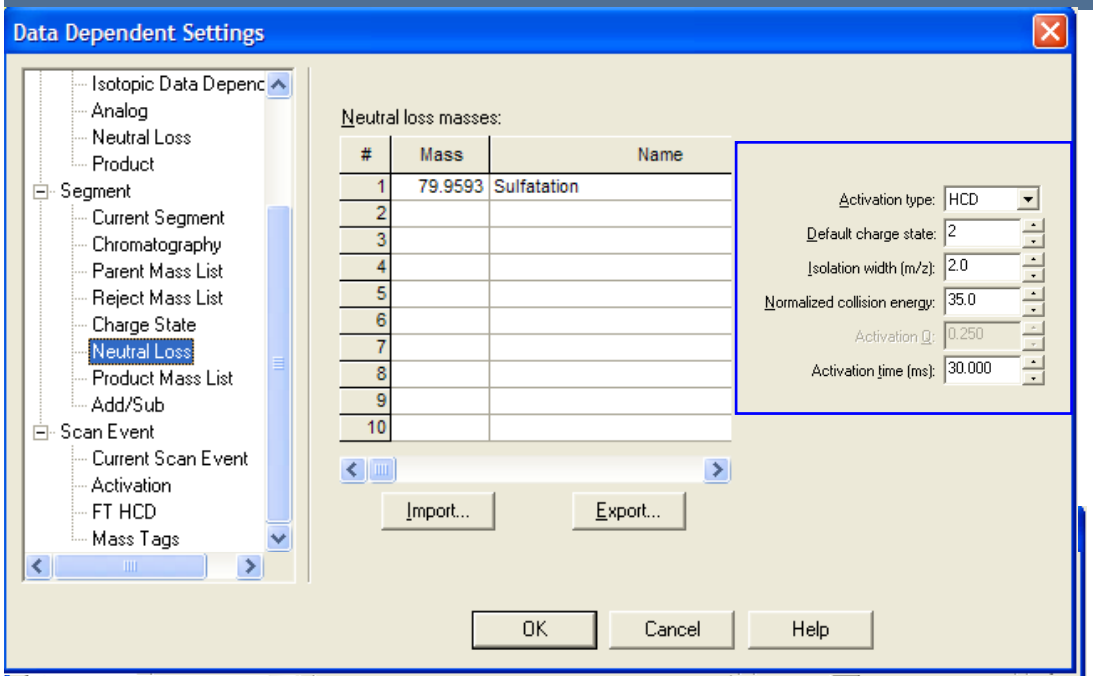
“Use m/z values as Mass”

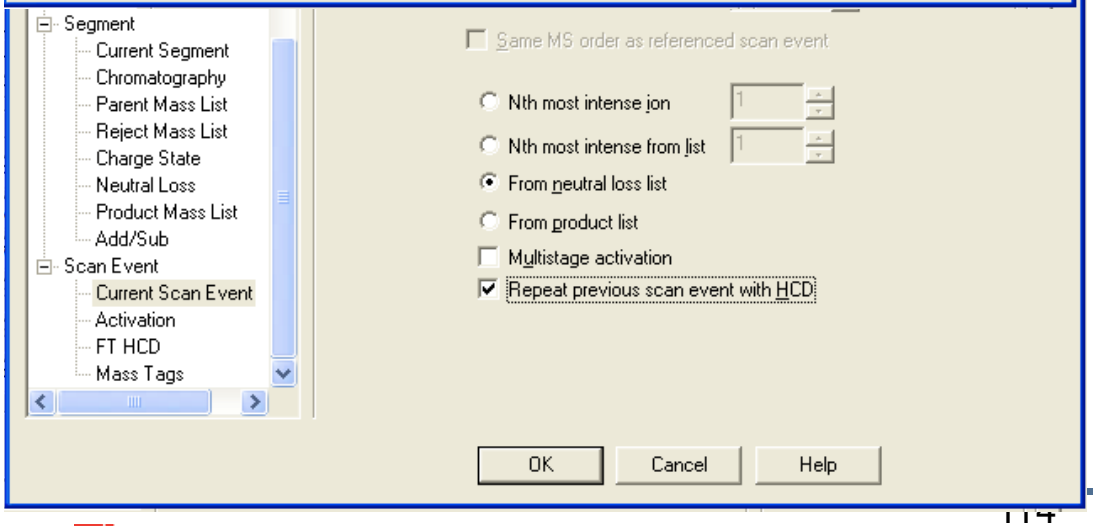


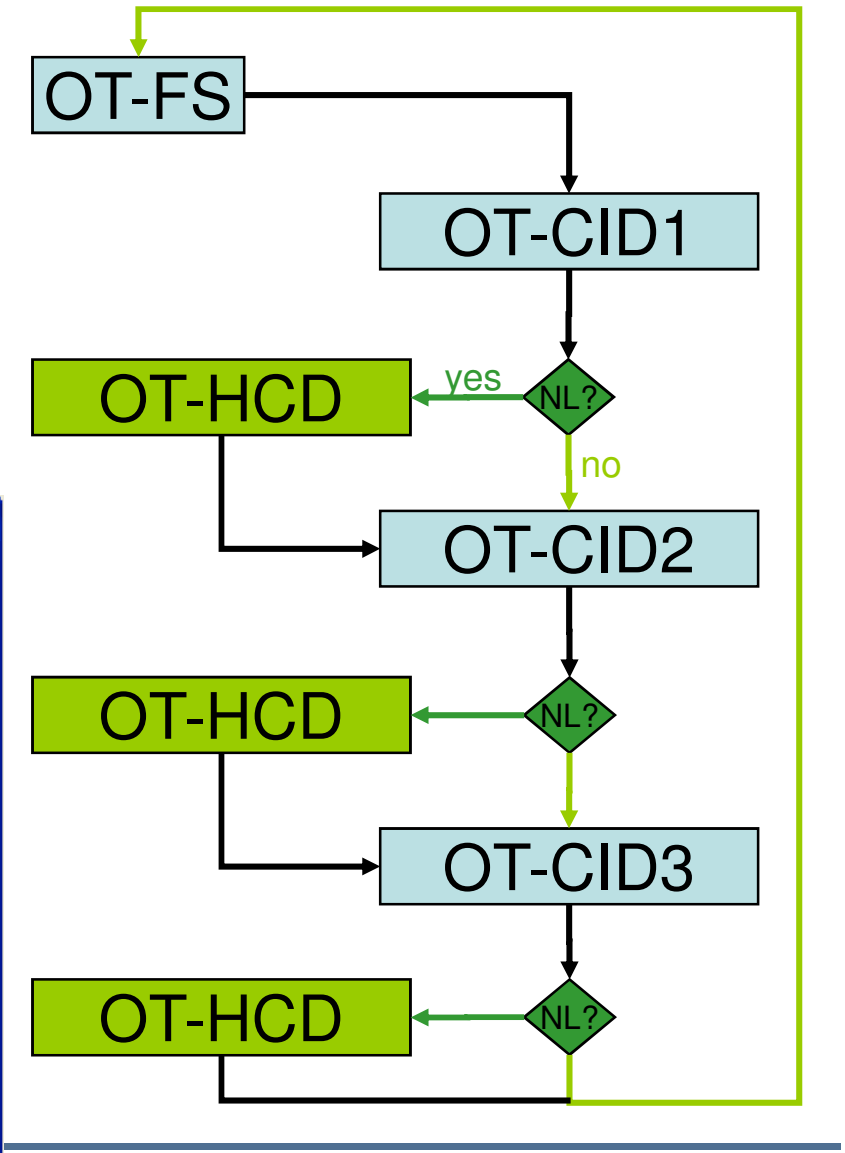
Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

Neutral Loss Triggered HCD

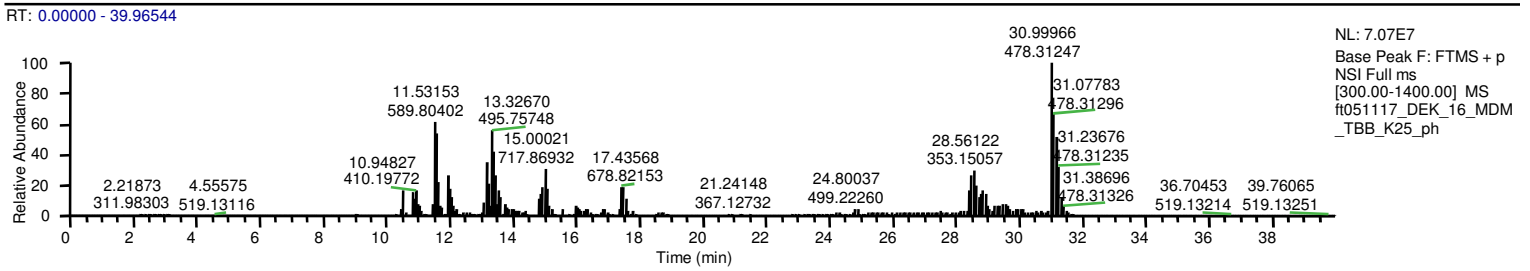




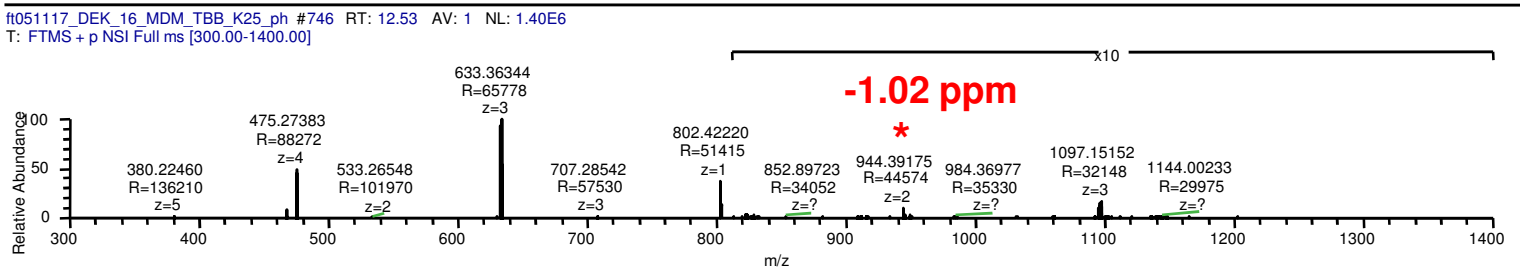


Identification of phosphorylated Ser-303 in DEK protein using NLMS3 method

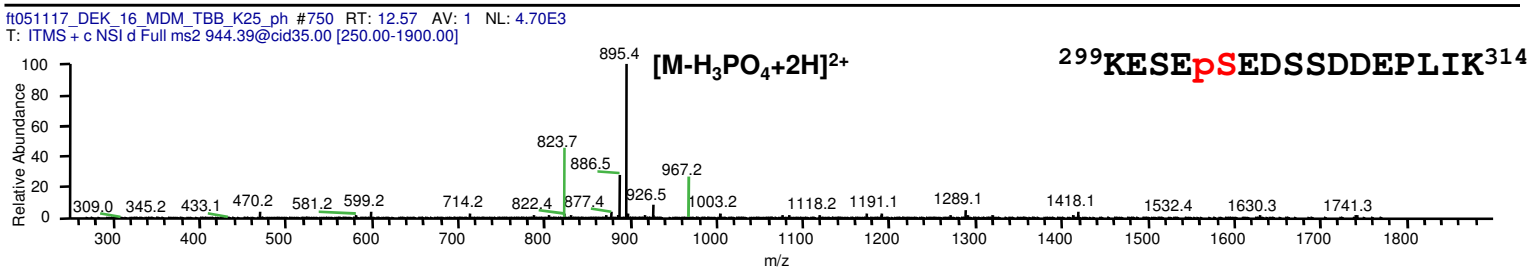
Base Peak Chromatogram



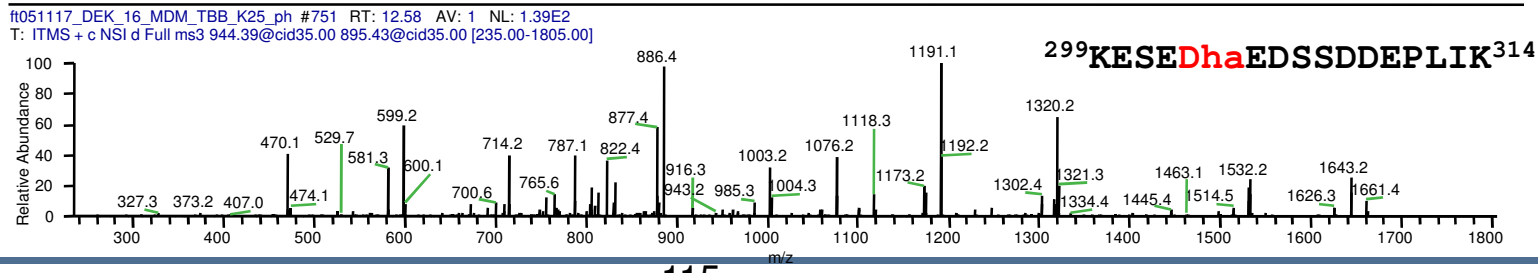
FTMS



ITMS2 (944.39)



ITMS3 (895.43)



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

Identification of phosphorylated Ser-303 in DEK protein using NLMS3 method and Bioworks

	AA	A	B	B*	Bo	C	Y	Y*	Yo	Z	
1	K	101.10732	129.10224	112.07569	111.09167		-	-	-		16
2	E	230.14992	258.14483	241.11828	240.13427		1759.67934	1742.65279	1741.66878		15
3	S	317.18195	345.17686	328.15031	327.16630		1630.63675	1613.61020	1612.62618		14
4	E	446.22454	474.21945	457.19290	456.20889		1543.60472	1526.57817	1525.59415		13
5	S*	613.22290	641.21781	624.19126	623.20725		1414.56213	1397.53558	1396.55156		12
6	E	742.26549	770.26041	753.23386	752.24984		1247.56377	1230.53722	1229.55320		11
7	D	857.29243	885.28735	868.26080	867.27678		1118.52117	1101.49463	1100.51061		10
8	S	944.32446	972.31938	955.29283	954.30881		1003.49423	986.46768	985.48367		9
9	S	1031.35649	1059.35141	1042.32486	1041.34084		916.46220	899.43565	898.45164		8
10	D	1146.38343	1174.37835	1157.35180	1156.36778		829.43018	812.40363	811.41961		7
11	D	1261.41038	1289.40529	1272.37874	1271.39473		714.40323	697.37668	696.39267		6
12	E	1390.45297	1418.44788	1401.42134	1400.43732		599.37629	582.34974	581.36572		5
13	P	1487.50573	1515.50065	1498.47410	1497.49008		470.33370	453.30715	452.32313		4
14	L	1600.58980	1628.58471	1611.55816	1610.57415		373.28093	356.25438	355.27037		3
15	I	1713.67386	1741.66878	1724.64223	1723.65821		260.19687	243.17032	242.18630		2
16	K	-	-	-	-		147.11280	130.08626	129.10224		1

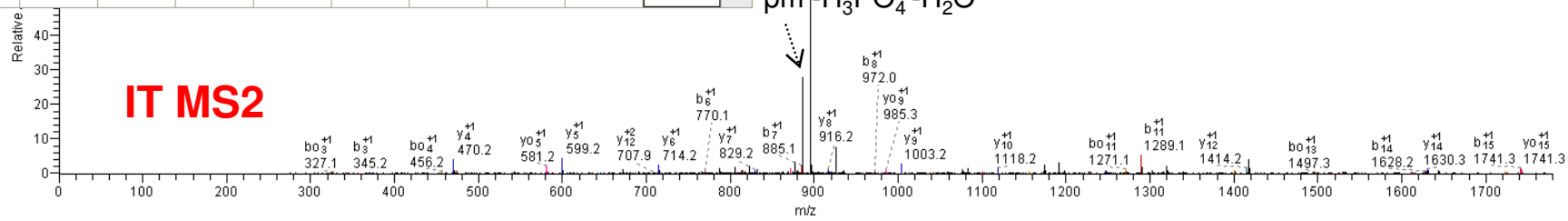
Mods: (STY* +79.96633) (STY# -18.01056)

pm -H_3PO_4

pm -H_3PO_4 -H_2O

^{299}KESEpSEDSSDDEPLIK^{314}

IT MS2



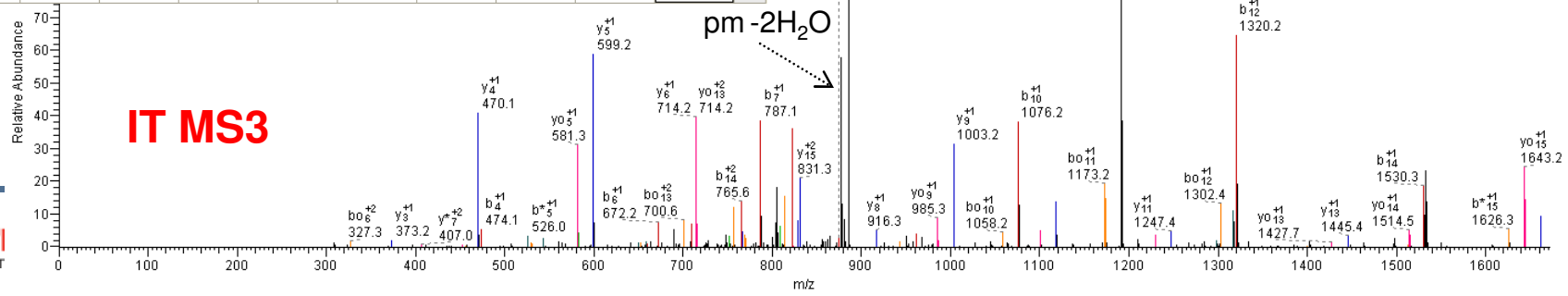
	AA	A	B	B*	Bo	C	Y	Y*	Yo	Z	
1	K	101.10732	129.10224	112.07569	111.09167		-	-	-		16
2	E	230.14992	258.14483	241.11828	240.13427		1661.70245	1644.67590	1643.69189		15
3	S	317.18195	345.17686	328.15031	327.16630		1532.65986	1515.63331	1514.64929		14
4	E	446.22454	474.21945	457.19290	456.20889		1445.62783	1428.60128	1427.61726		13
5	S#	515.24601	543.24092	526.21437	525.23036		1316.58524	1299.55869	1298.57467		12
6	E	644.28860	672.28352	655.25697	654.27295		1247.56377	1230.53722	1229.55320		11
7	D	759.31554	787.31046	770.28391	769.29989		1118.52117	1101.49463	1100.51061		10
8	S	846.34757	874.34249	857.31594	856.33192		1003.49423	986.46768	985.48367		9
9	S	933.37960	961.37452	944.34797	943.36395		916.46220	899.43565	898.45164		8
10	D	1048.40654	1076.40146	1059.37491	1058.39089		829.43018	812.40363	811.41961		7
11	D	1163.43349	1191.42840	1174.40185	1173.41784		714.40323	697.37668	696.39267		6
12	E	1292.47608	1320.47099	1303.44445	1302.46043		599.37629	582.34974	581.36572		5
13	P	1389.52884	1417.52376	1400.49721	1399.51319		470.33370	453.30715	452.32313		4
14	L	1502.61291	1530.60782	1513.58127	1512.59726		373.28093	356.25438	355.27037		3
15	I	1615.69697	1643.69189	1626.66534	1625.68132		260.19687	243.17032	242.18630		2
16	K	-	-	-	-		147.11280	130.08626	129.10224		1

^{299}KESEDhaEDSSDDEPLIK^{314}

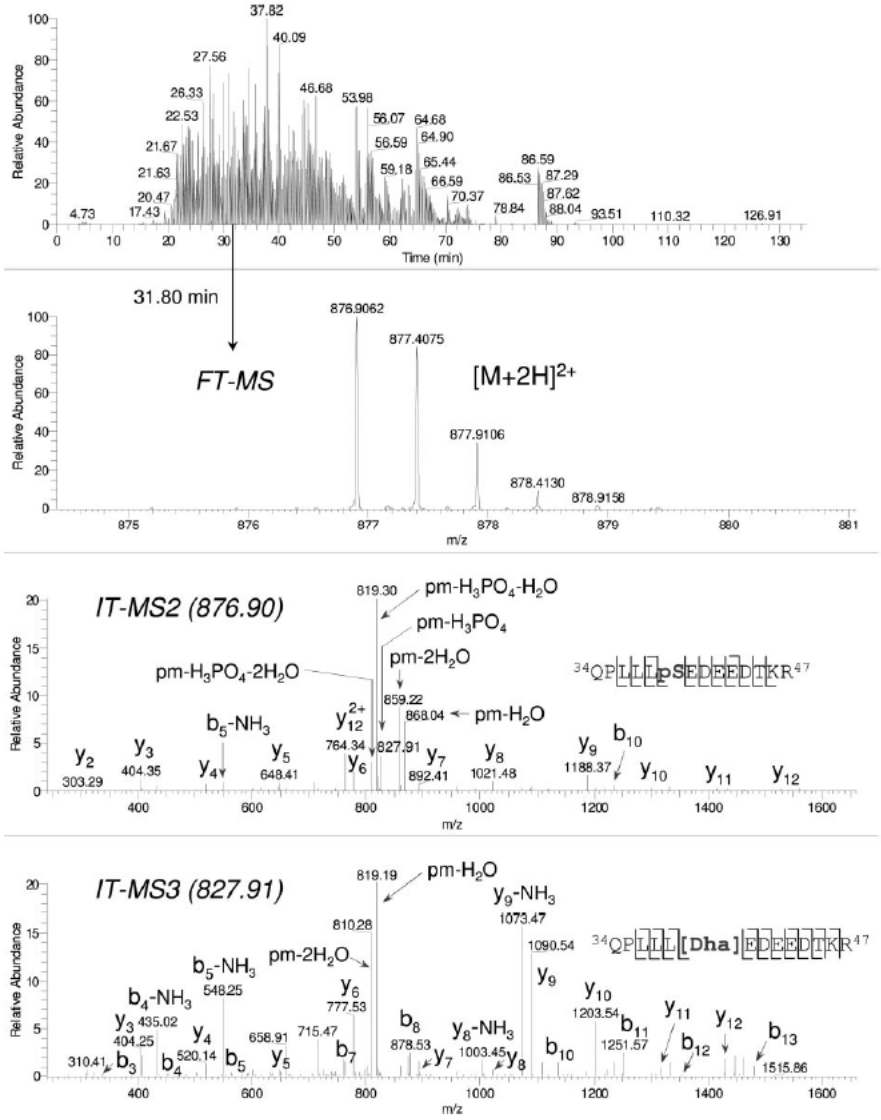
pm -H_2O

pm -$2H_2O$

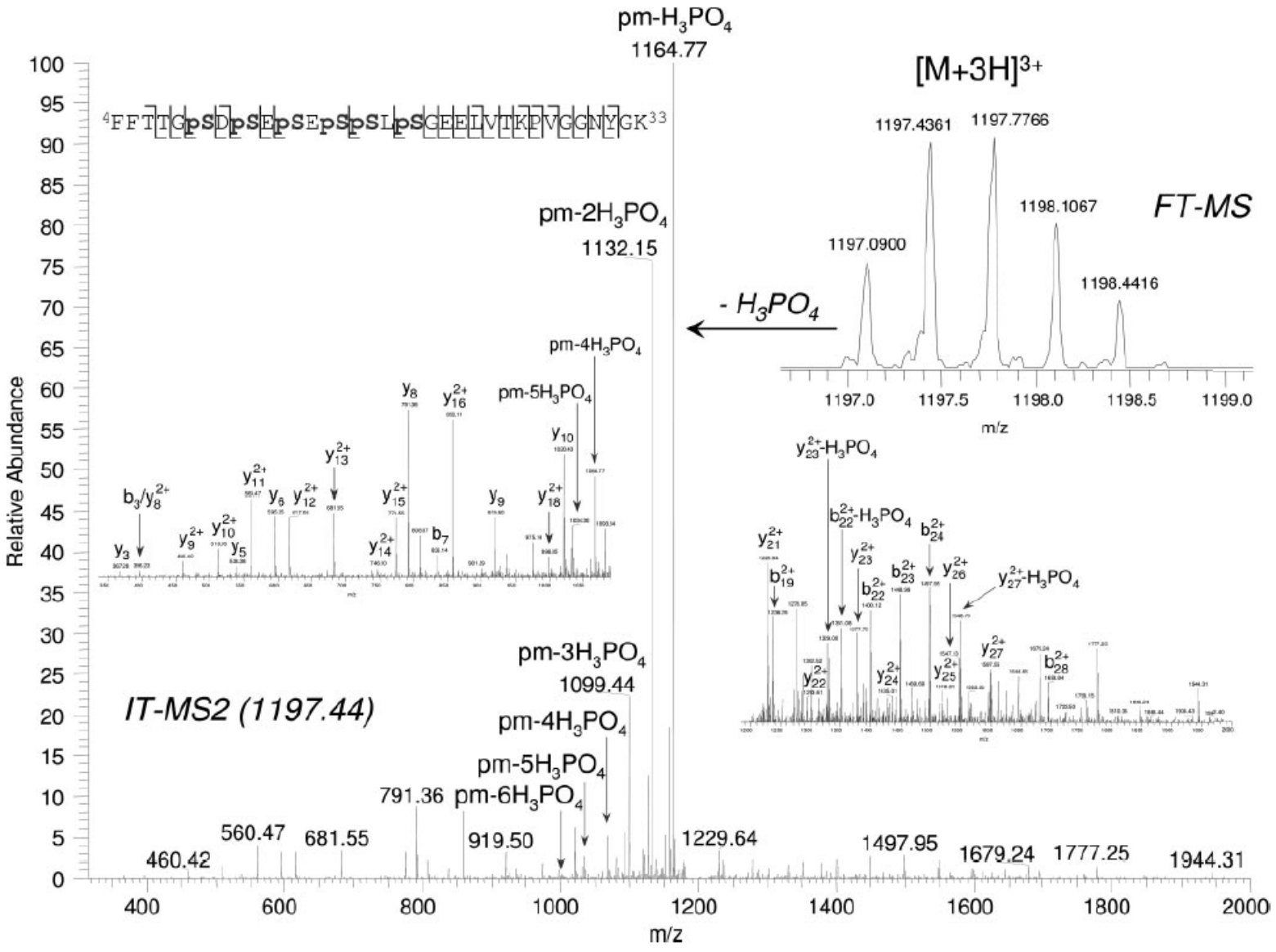
IT MS3



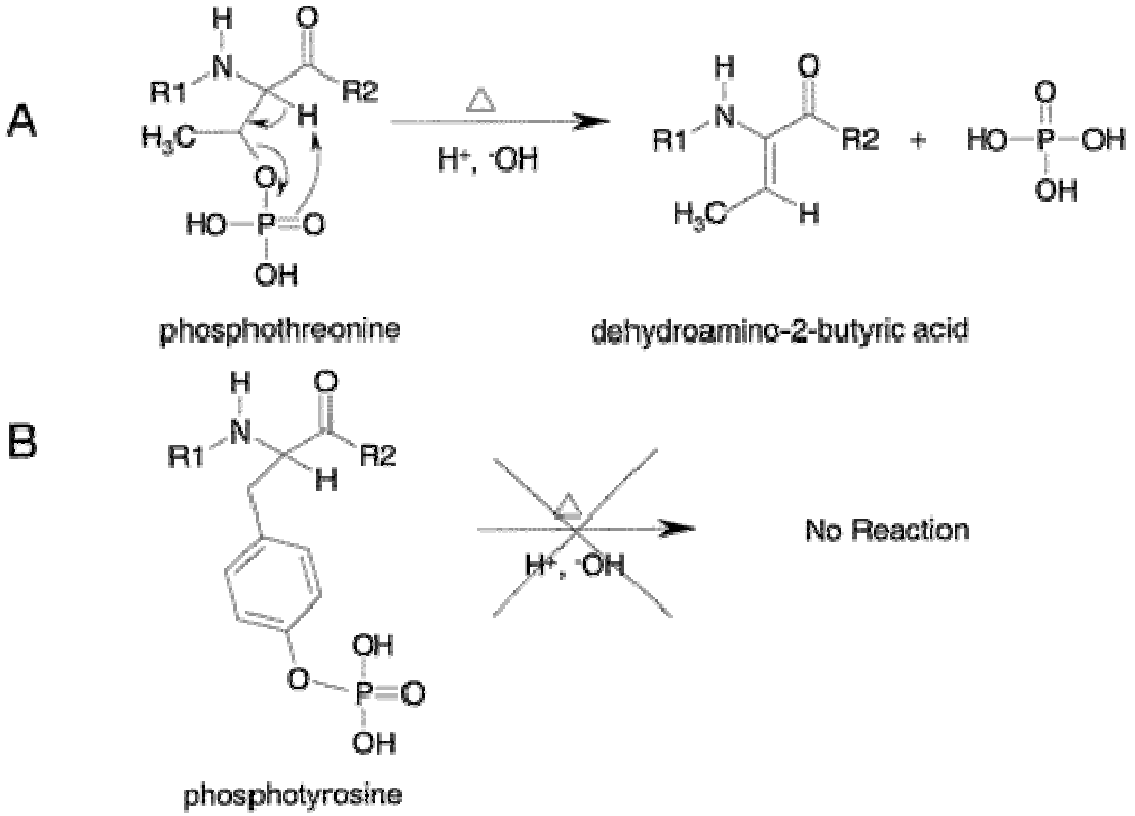
Identification of the phosphorylation site *Ser-39* from eIF3c subunit using neutral loss/MS3 method



Multiphosphorylation of the N-terminal part of eIF3c



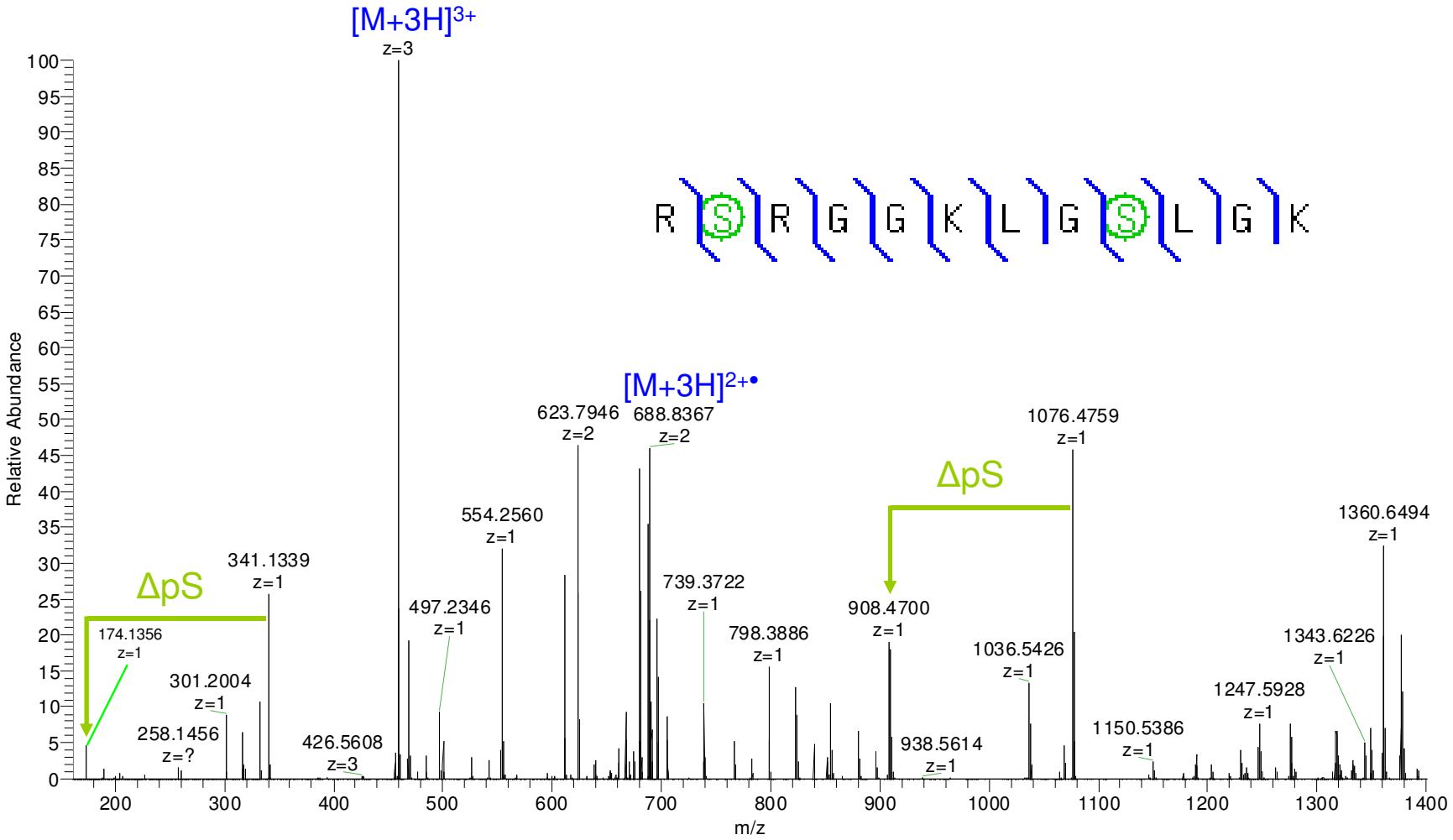
Neutral Loss Scanning



(A) elimination of phosphate from phosphothreonine to produce dehydroamino-2-butyric acid (scheme is also valid for phosphoserine where dehydroalanine forms) and phosphoryc acid

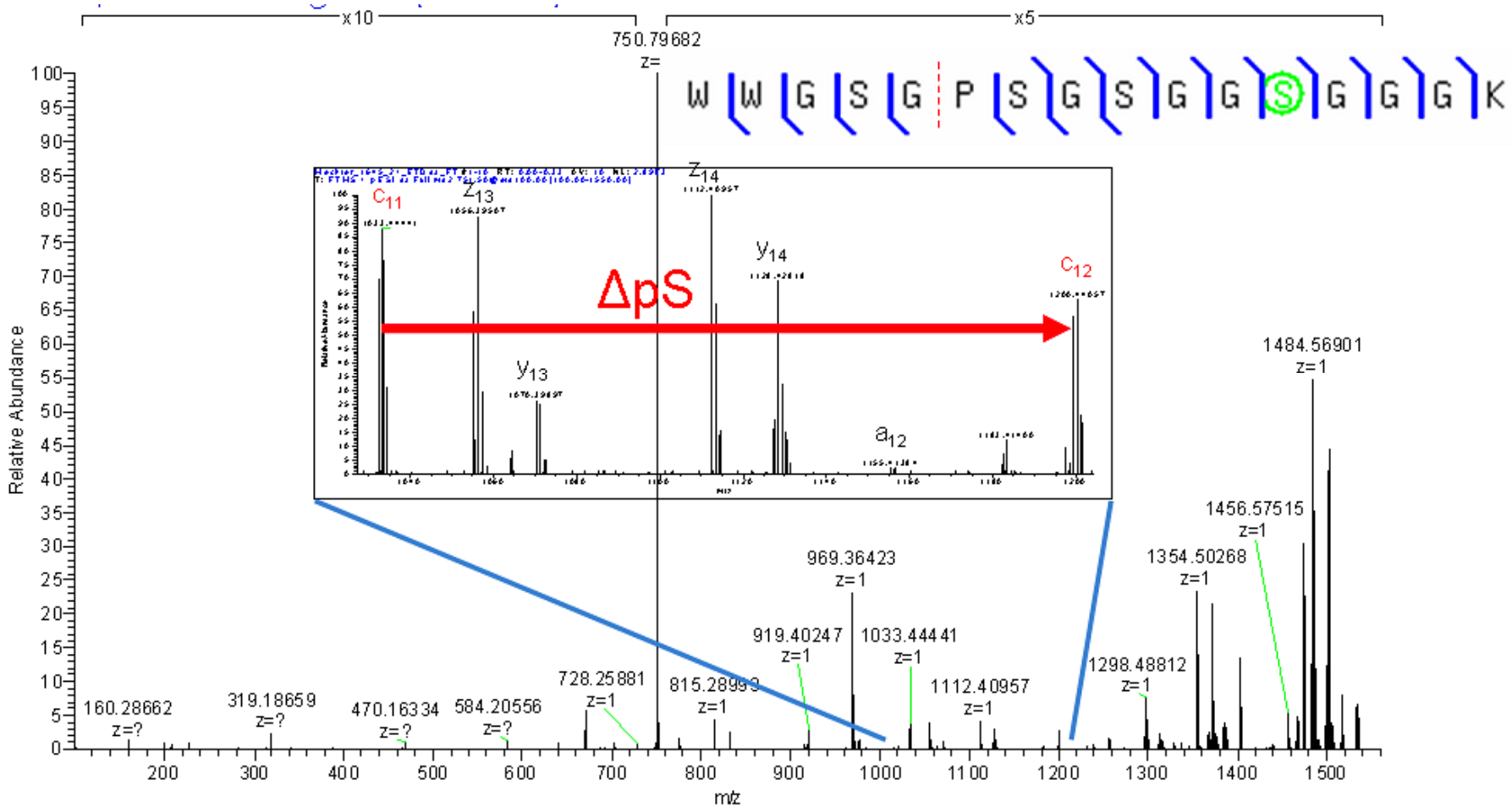
(B) absence of the same mechanistic pathway for b-elimination of phosphate from phosphotyrosine.

LTQ Orbitrap XL ETD for Phosphopeptide Analysis RpSRGGKLGpSLGK



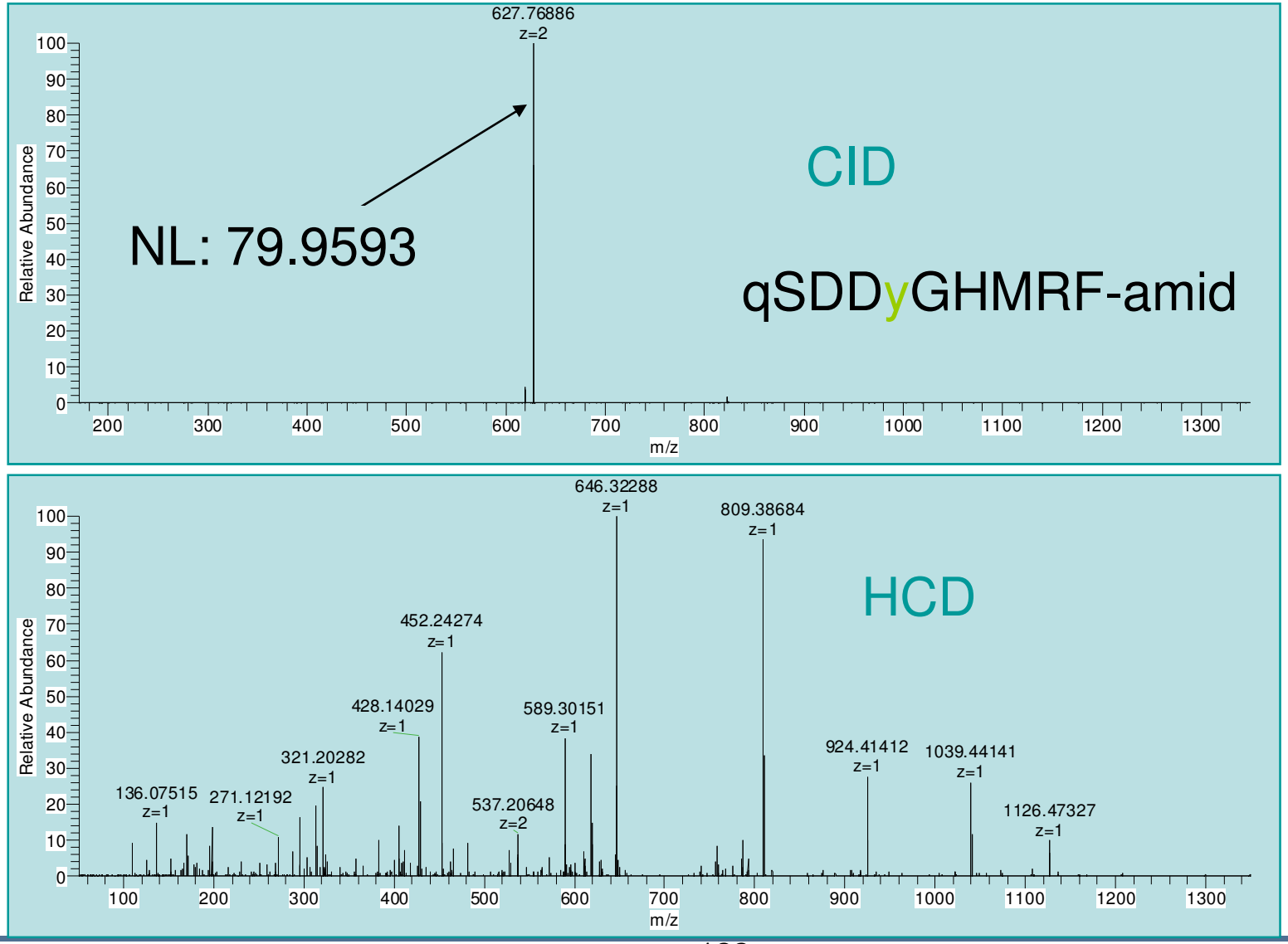
Phosphopeptide standards courtesy of Karl Mechtler, IMP Vienna, Austria

LTQ Orbitrap XL ETD for Phosphopeptide Analysis using ETD



Phosphopeptide standards courtesy of Karl Mechtler, IMP Vienna, Austria

CID and HCD of 667.7485^{2+} (Sulfatation)



Thermo SCIENTIFIC

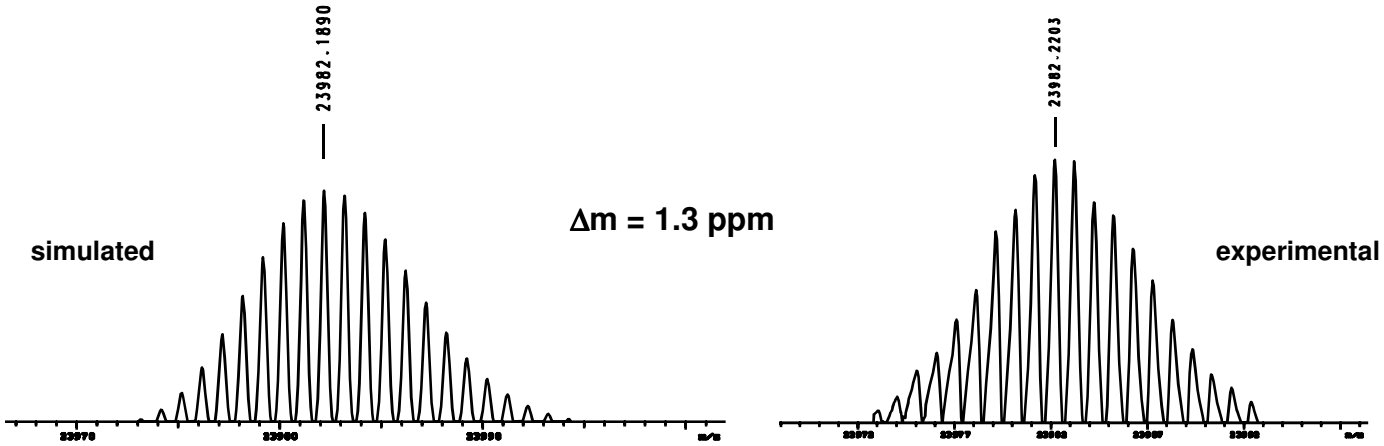
ThermoFisher SCIENTIFIC

Intact protein analysis

Analysis of intact phosphoproteins reveals the number of phosphorylation sites

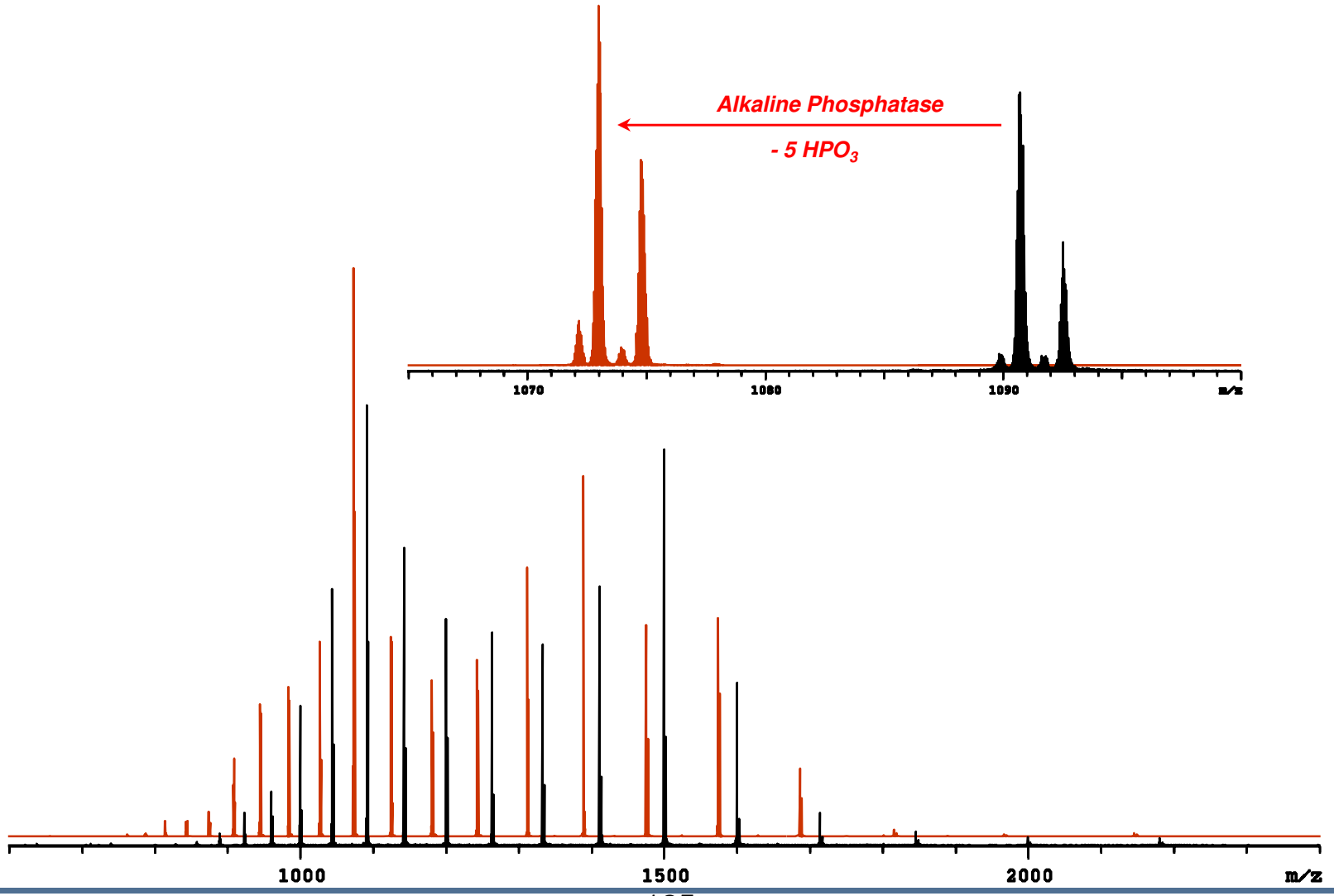
Beta-casein (5x phosphorylated)

$C_{1080}H_{1697}N_{268}O_{325}P_5S_6$



^{1}RELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQ
TEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQT
PVVVPPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFT
ESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQ
SVLSLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGP
VRGPFPIIV^{209}

Dephosphorylation by AP treatment reveals the number of phosphorylation sites



Amino acid sequence of bovine Carbonic Anhydrase II

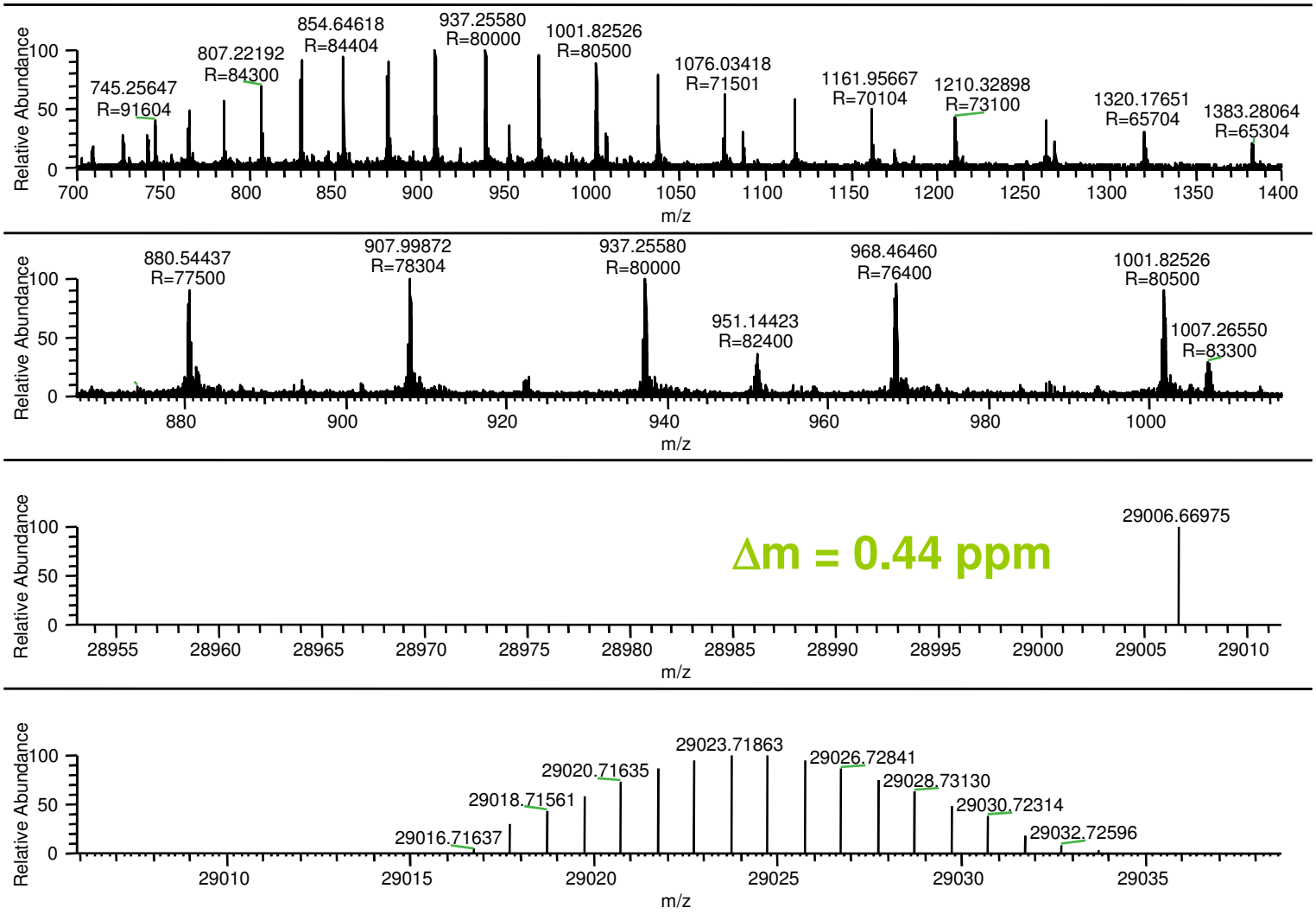
Ac–SHHWGYGKHNGPEHWHKDFPIANGERQSPVDIDTKAVVQDPALKPLALVY
GEATSRRMVNNGHSFNVEYDDSQDKAVLKDGPLTGTYRLVQFHFHWGSSD
DQGSEHTVDRKKYAAELHLVHWNTKYGDFGTAAQQPDGLAVVGVFLKVGD
ANPALQKVLDALDSIKTKGKSTDFPNFDPGSLLPNVLDYWTYPGSLTTPP
LLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLK
NRQVRGFPK

Sum Formula: $C_{1312}H_{1996}N_{358}O_{384}S_3$

Monoisot. Mass: 29,006.682670

N-Terminus: N-Acetylation

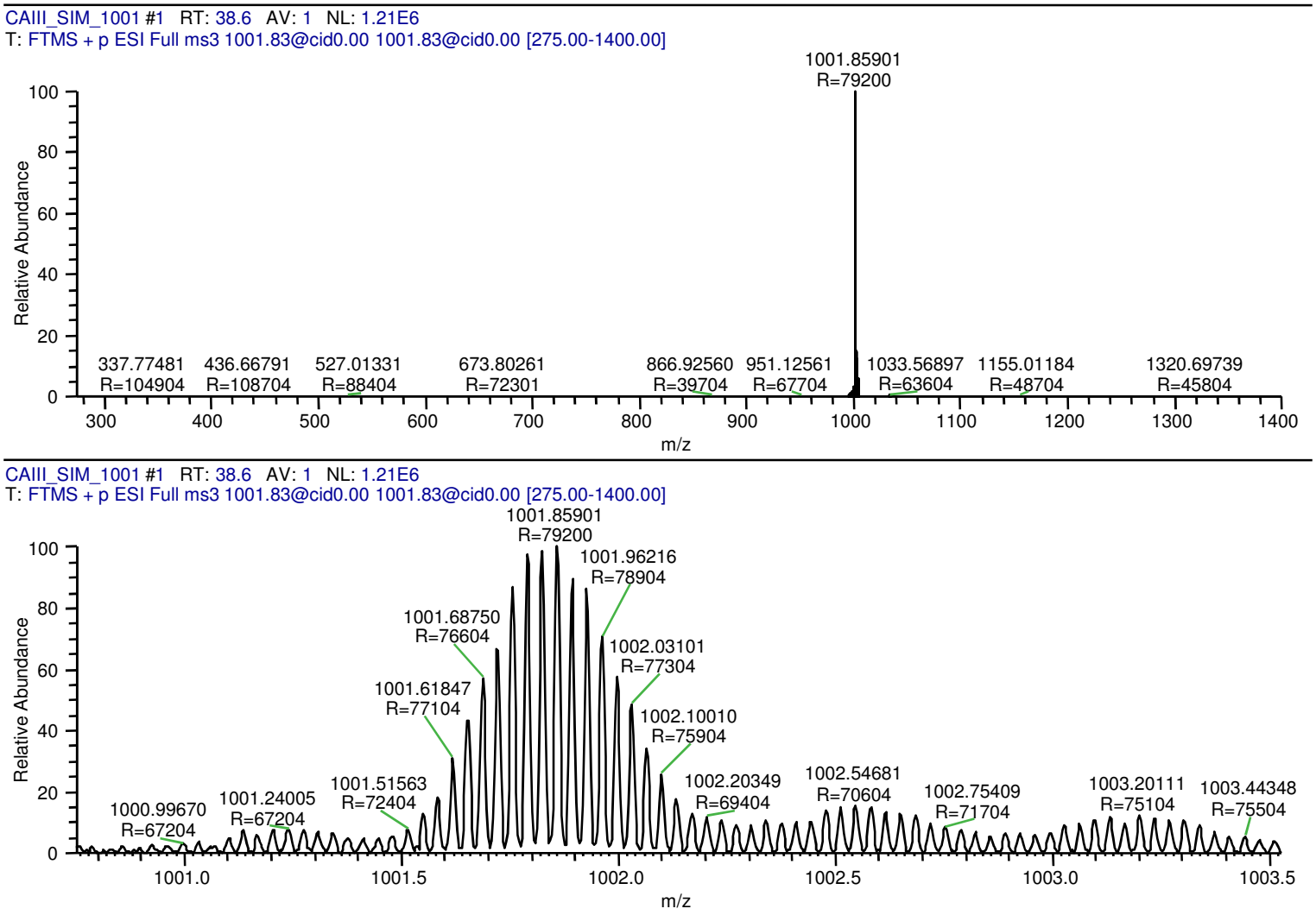
Full FTMS of the Bovine Carbonic Anhydrase II



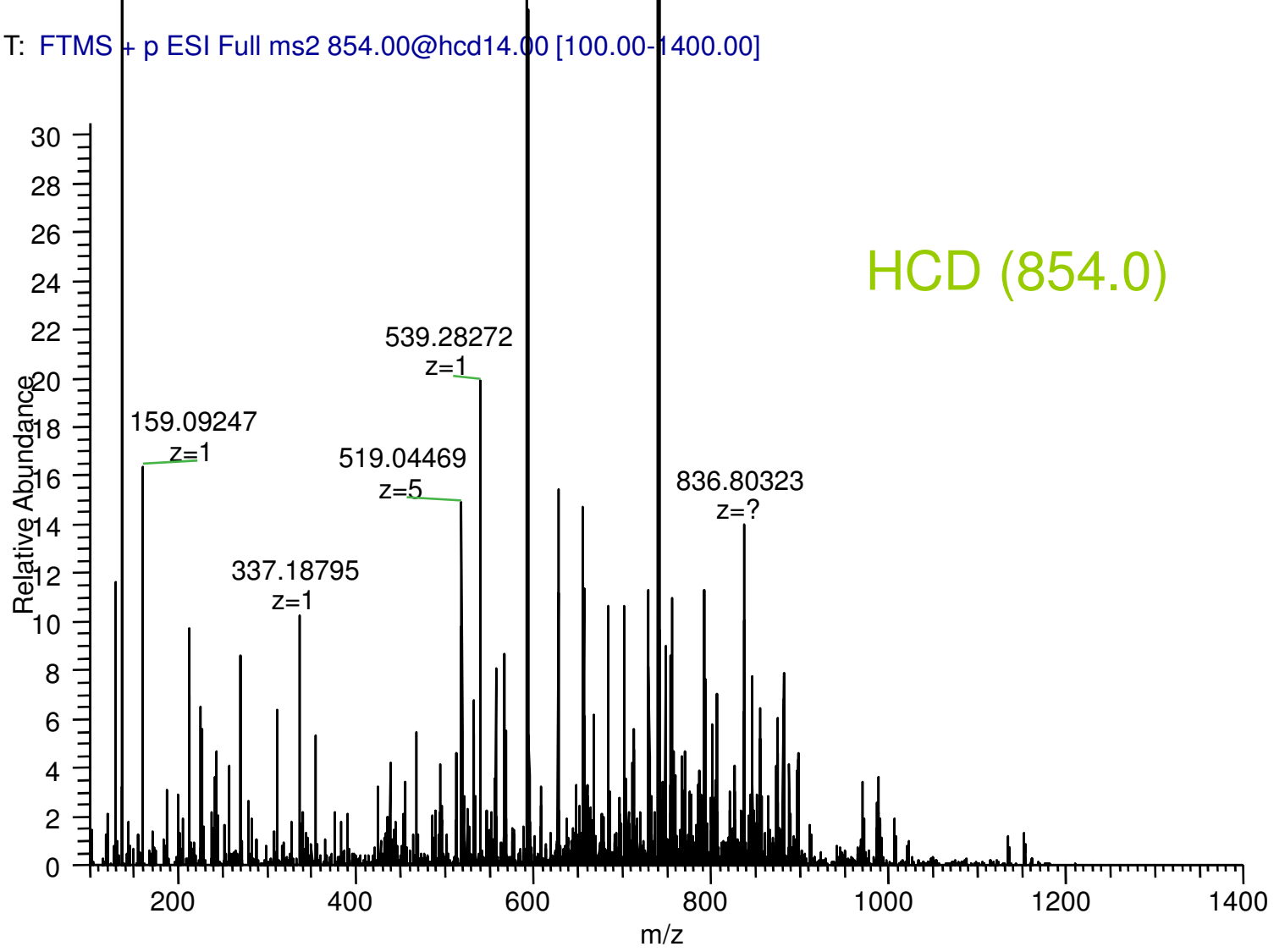
Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

Isolation of the $[M+29H]^{29+}$ ion of the Bovine Carbonic Anhydrase II



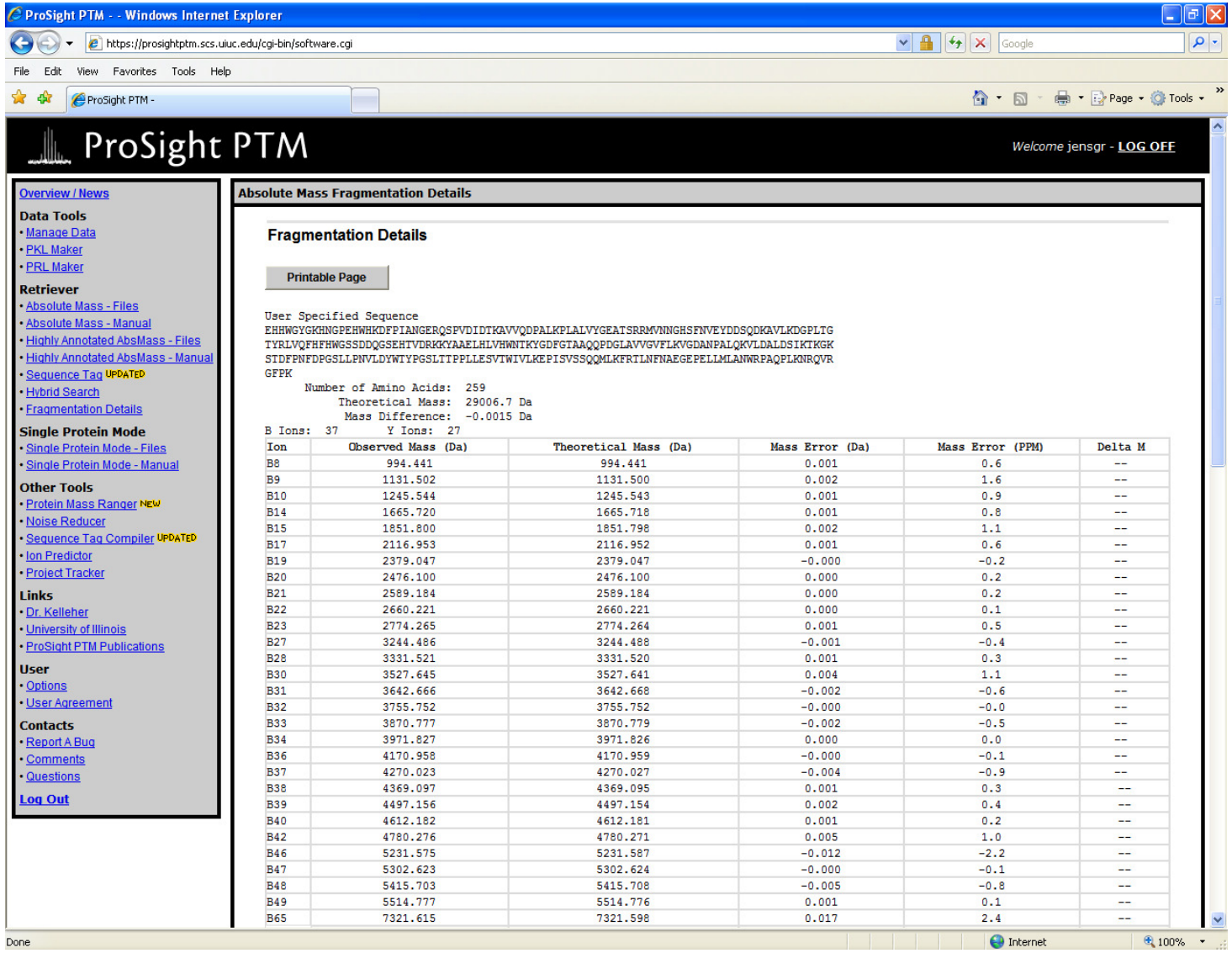
Higher Energy Collision Dissociation (HCD) of the $[M+34H]^{34+}$ ion



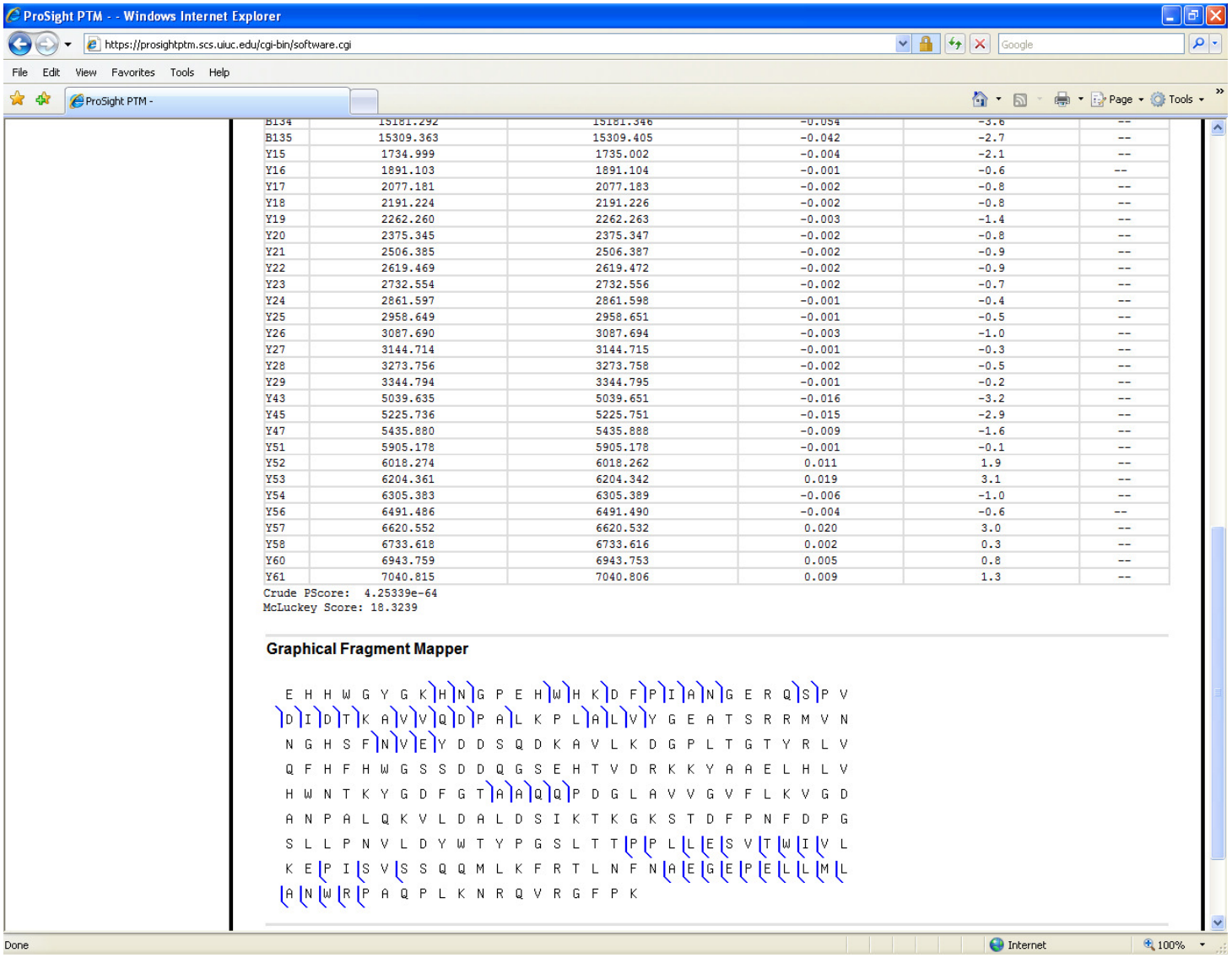
Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

HCD Fragmentation Details using ProSight PTM



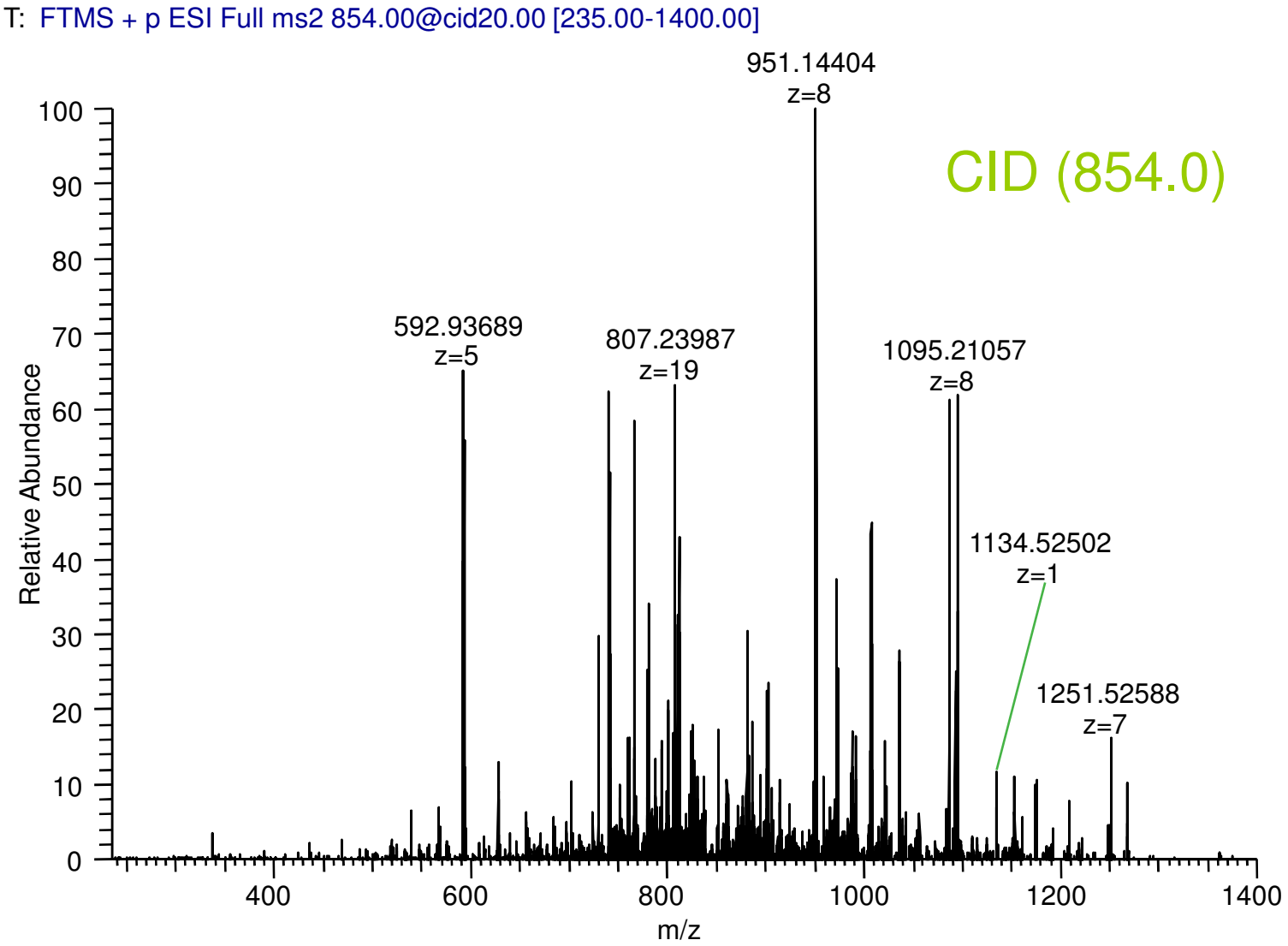
HCD Fragmentation Details using ProSight PTM



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

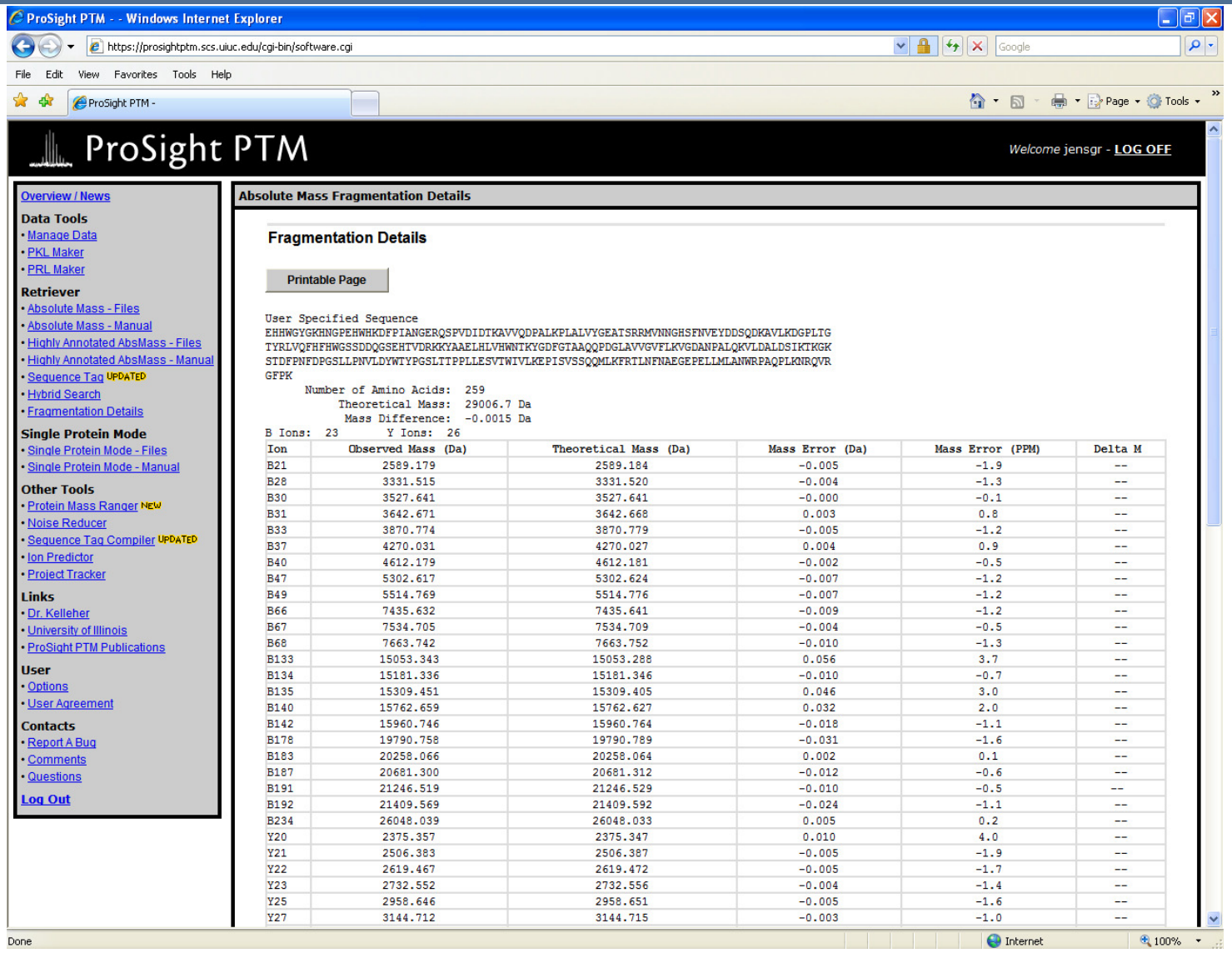
Collision Induced Dissociation (CID) of the $[M+34H]^{34+}$ ion



Thermo SCIENTIFIC

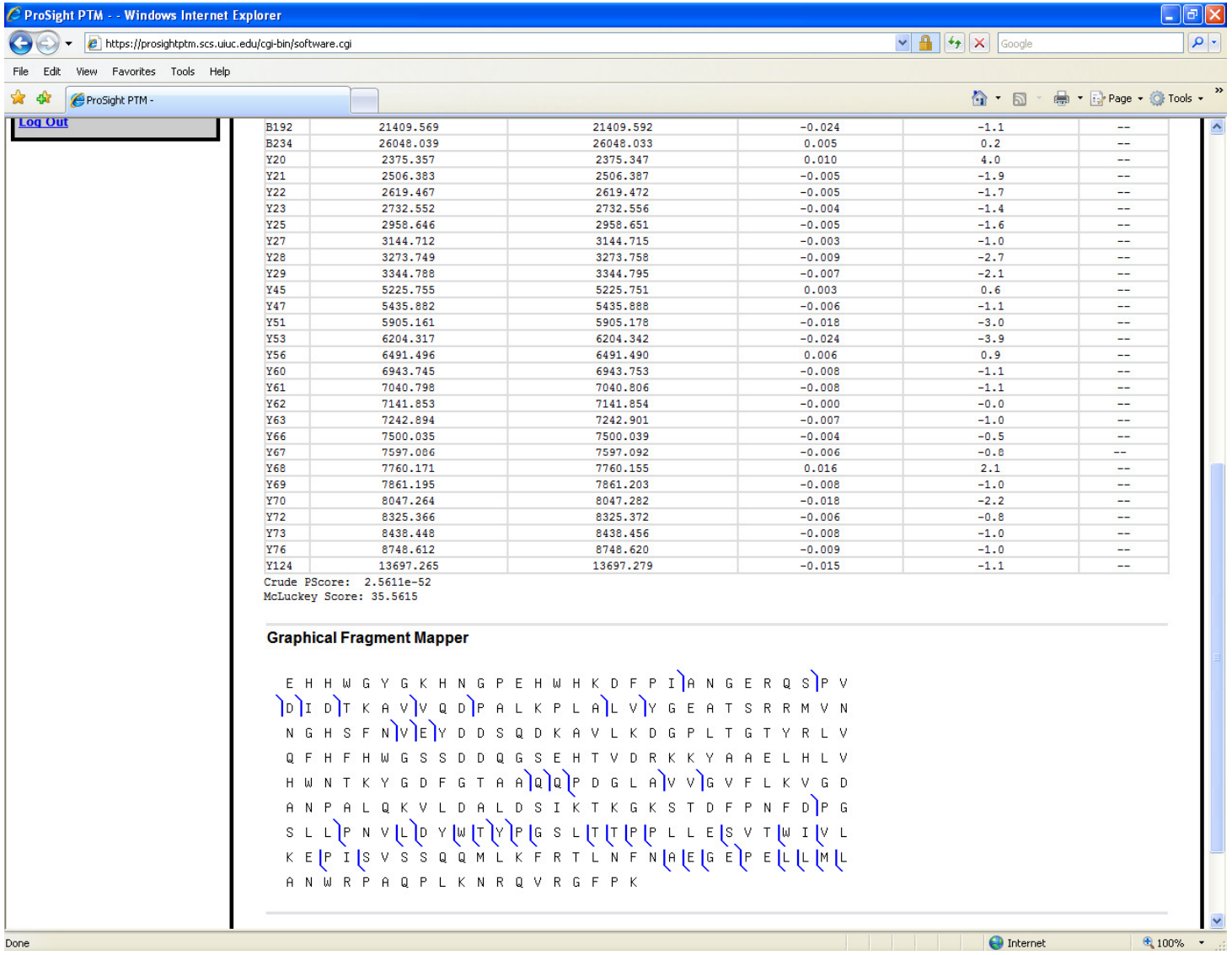
ThermoFisher SCIENTIFIC

CID Fragmentation Details using ProSight PTM



Ion	Observed Mass (Da)	Theoretical Mass (Da)	Mass Error (Da)	Mass Error (PPM)	Delta M
B21	2589.179	2589.184	-0.005	-1.9	--
B28	3331.515	3331.520	-0.004	-1.3	--
B30	3527.641	3527.641	-0.000	-0.1	--
B31	3642.671	3642.668	0.003	0.8	--
B33	3870.774	3870.779	-0.005	-1.2	--
B37	4270.031	4270.027	0.004	0.9	--
B40	4612.179	4612.181	-0.002	-0.5	--
B47	5302.617	5302.624	-0.007	-1.2	--
B49	5514.769	5514.776	-0.007	-1.2	--
B66	7435.632	7435.641	-0.009	-1.2	--
B67	7534.705	7534.709	-0.004	-0.5	--
B68	7663.742	7663.752	-0.010	-1.3	--
B133	15053.343	15053.288	0.056	3.7	--
B134	15181.336	15181.346	-0.010	-0.7	--
B135	15309.451	15309.405	0.046	3.0	--
B140	15762.659	15762.627	0.032	2.0	--
B142	15960.746	15960.764	-0.018	-1.1	--
B178	19790.758	19790.789	-0.031	-1.6	--
B183	20258.066	20258.064	0.002	0.1	--
B187	20681.300	20681.312	-0.012	-0.6	--
B191	21246.519	21246.529	-0.010	-0.5	--
B192	21409.569	21409.592	-0.024	-1.1	--
B234	26048.039	26048.033	0.005	0.2	--
Y20	2375.357	2375.347	0.010	4.0	--
Y21	2506.383	2506.387	-0.005	-1.9	--
Y22	2619.467	2619.472	-0.005	-1.7	--
Y23	2732.552	2732.556	-0.004	-1.4	--
Y25	2958.646	2958.651	-0.005	-1.6	--
Y27	3144.712	3144.715	-0.003	-1.0	--

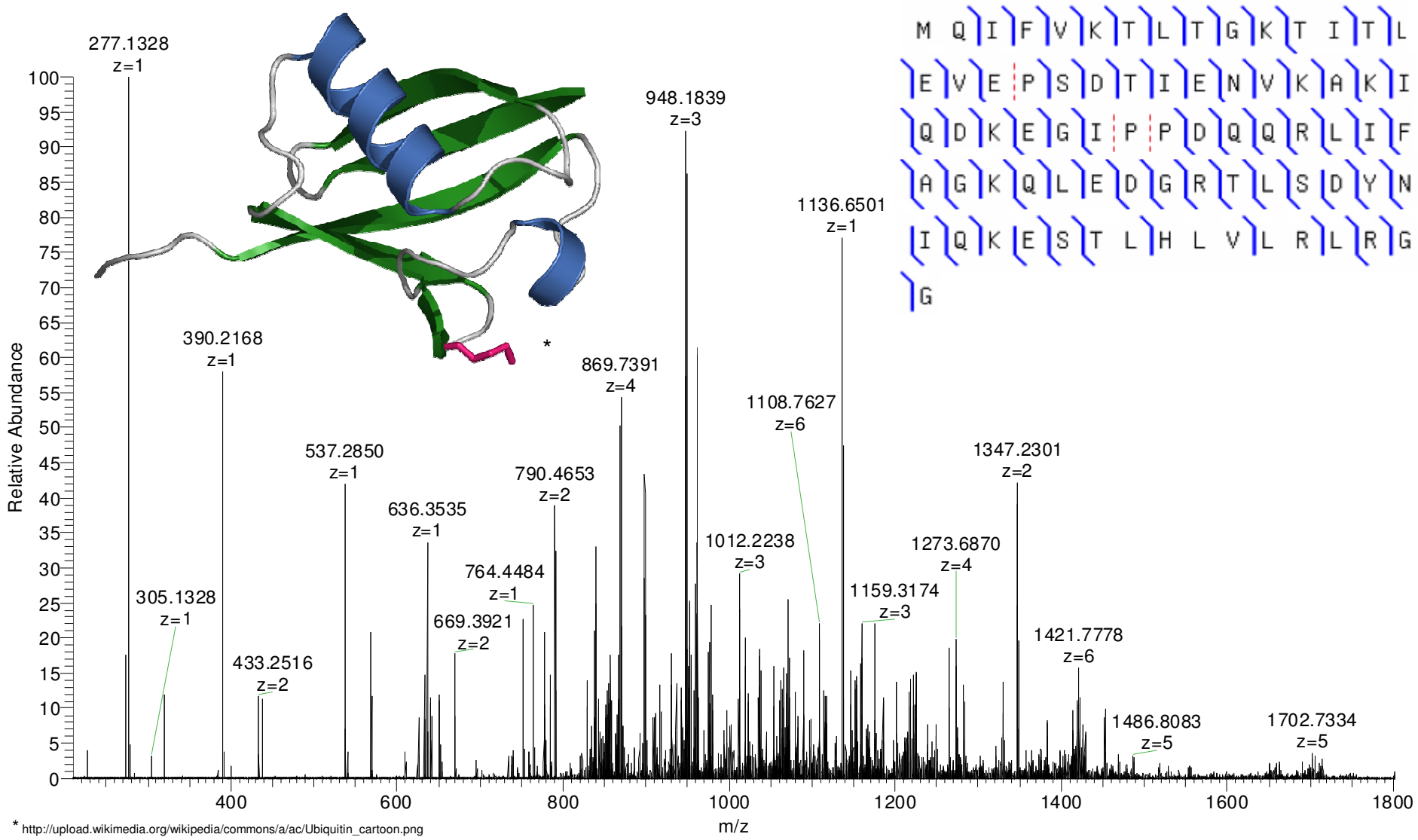
CID Fragmentation Details using ProSight PTM



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

ETD spectrum of Ubiquitin, 12+ charge state, with Orbitrap detection



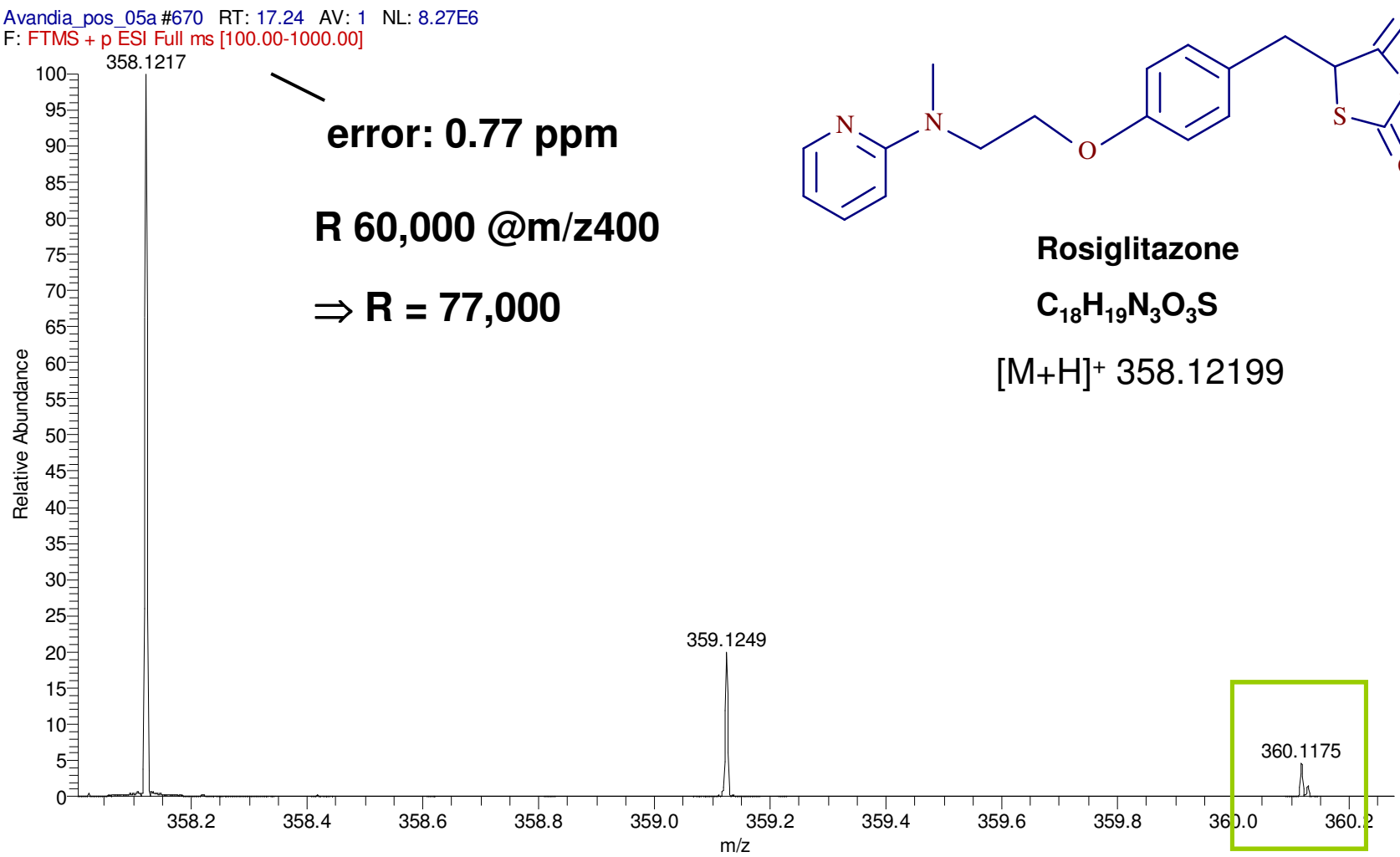
Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

FUNDAMENTAL: Mass Accuracy

$$\text{Parts per million (PPM)} = \frac{[\text{Mass}_{\text{theor}} - \text{Mass}_{\text{exp}}]}{\text{Mass}_{\text{theor}}} \times 10^6$$

Mass Accuracy and Resolving Power



error: 0.77 ppm

R 60,000 @m/z400

⇒ R = 77,000

Rosiglitazone

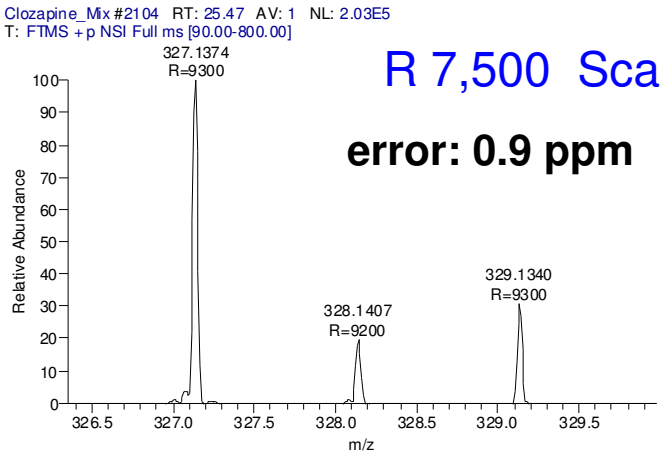
$C_{18}H_{19}N_3O_3S$

$[M+H]^+$ 358.12199

Thermo SCIENTIFIC

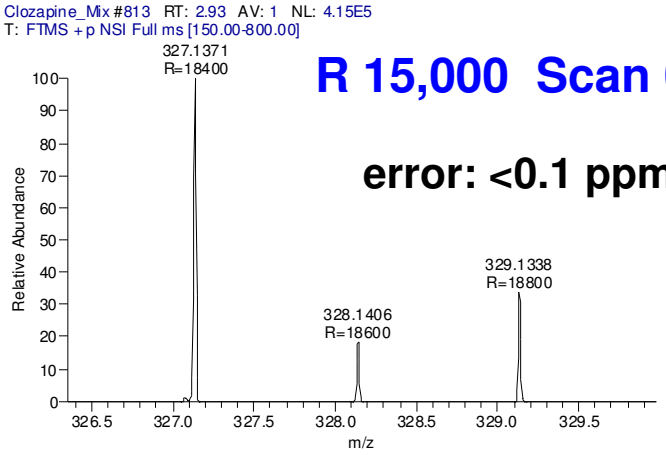
ThermoFisher SCIENTIFIC

Mass Accuracy and Resolution: Clozapine



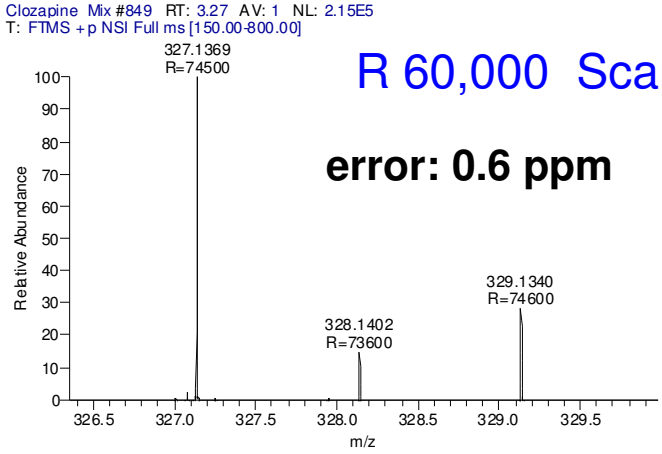
R 7,500 Scan 0.3 sec

error: 0.9 ppm



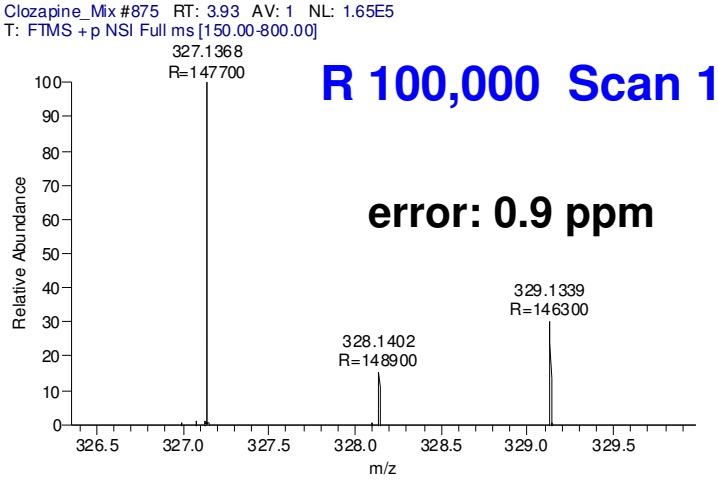
R 15,000 Scan 0.3 sec

error: <0.1 ppm



R 60,000 Scan 1 sec

error: 0.6 ppm



R 100,000 Scan 1.8 sec

error: 0.9 ppm

Different scan speeds and resolution – Mass accuracy < 1ppm

Mass to charge (m/z)

Mass spectrometers do not measure mass, but m/z!

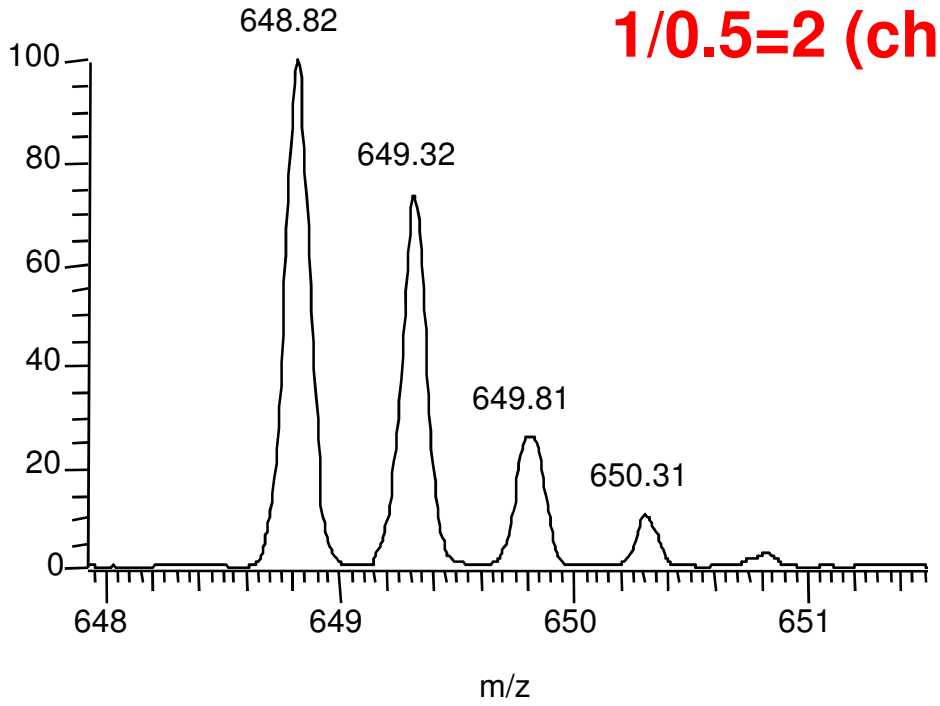
- If a peptide with a mass of 700 Da gets ionized and acquires 2 protons at pH 3.0
- The mass spectrometer sees

  700+2 = 702/2 = 351.00
- This mass is referred to as $[M+2H]2^{+}$

Determining Charge State

649.32-648.82=0.5

1/0.5=2 (charge state)

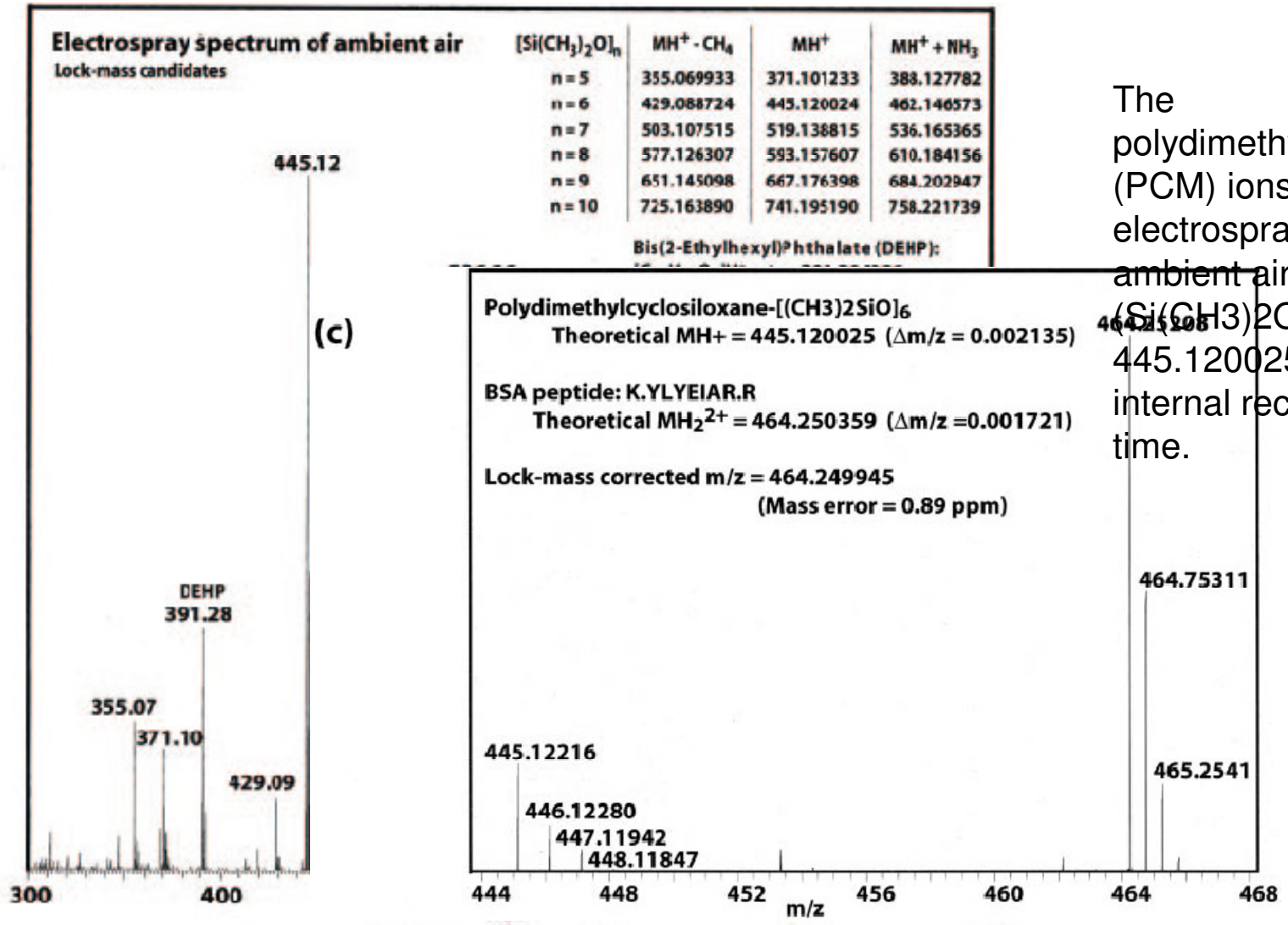


m/z= 648.82

(648.82*2)-2 = M 1295.64

Lock Mass

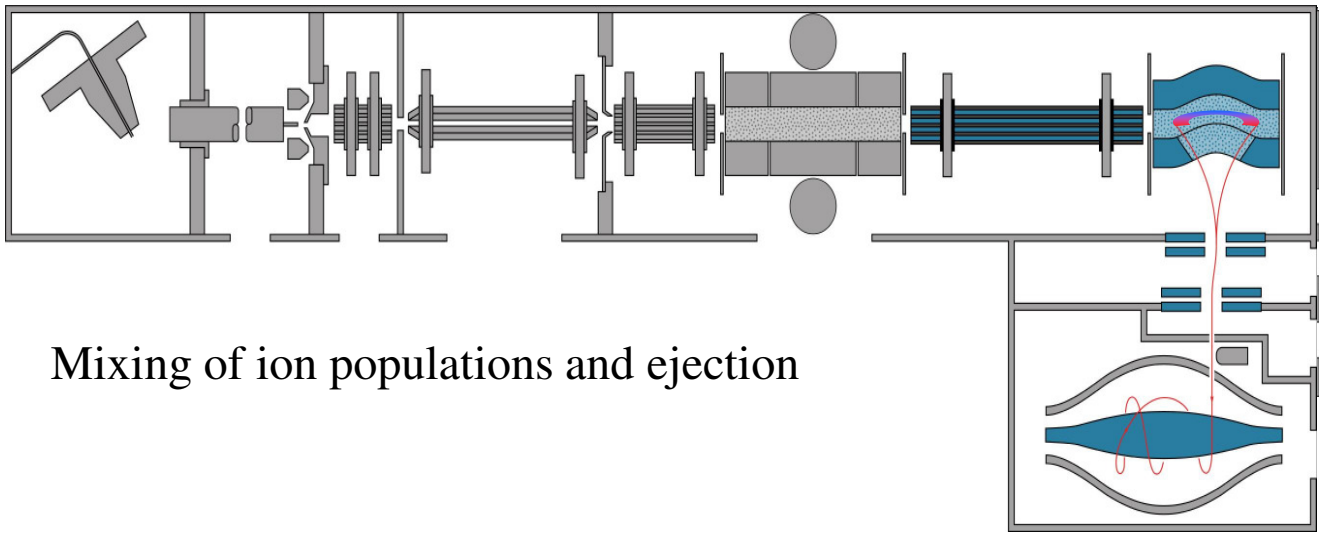
Ions which can be used as lock masses during online LC-MS analyses



The polydimethylcyclosiloxane (PCM) ions generated in the electrospray process from ambient air (protonated (Si(CH3)2O))6; *m/z* 445.120025) were used for internal recalibration in real time.

Olsen, Mann et al. *Mol. Cell. Proteomics* 2005, **4**: 2010-2021
"Parts per million mass accuracy on an orbitrap mass spectrometer via lock-mass injection into a C-trap."

Internal calibration = Lock Mass



Mixing of ion populations and ejection

Olsen, Mann et al. *Mol. Cell. Proteomics* 2005, **4**: 2010-2021
"Parts per million mass accuracy on an orbitrap mass spectrometer via lock-mass injection into a C-trap."

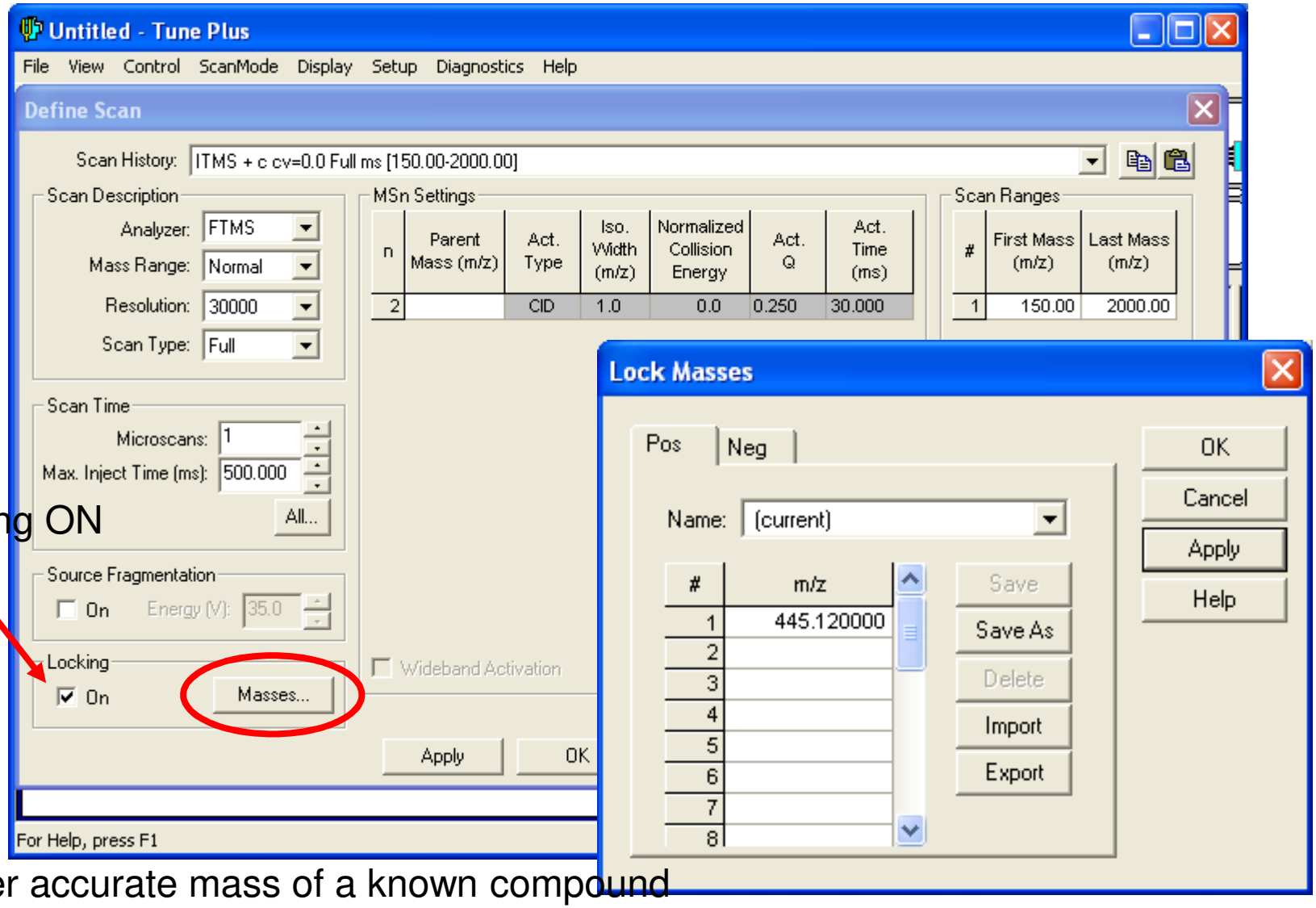
Internal Calibration

- Lock masses are isolated in the LTQ with a reduced inject time.
- The C-Trap is filled with them. The reduced inject time adjusts the lock mass intensity to appr. 5%.
- The LTQ is filled for a second time using the full inject time.
- Ions from the second filling are moved to the C-trap and mixed with the previous ion population of the calibrant.
- Ions are shot into the orbitrap.
- After the transient acquisition and data processing an existing calibration function is adjusted based on the measured frequencies of the lock masses.

Getting the most from your lock mass

- For a specific compound of interest use a lock mass close in mass
- Note: if you have a lock mass at low m/z but your compound has high m/z, it might be better to use external calibration
- **Generally: we recommend using 1 lock mass only**
- If you put more lock masses, you need fill time for each of them, and it does not necessarily improve the result
- If no lock mass is found the system applies the last external calibration
- Thus, you should keep your orbitrap externally well calibrated anyway!

Internal Calibration: Tune Setup

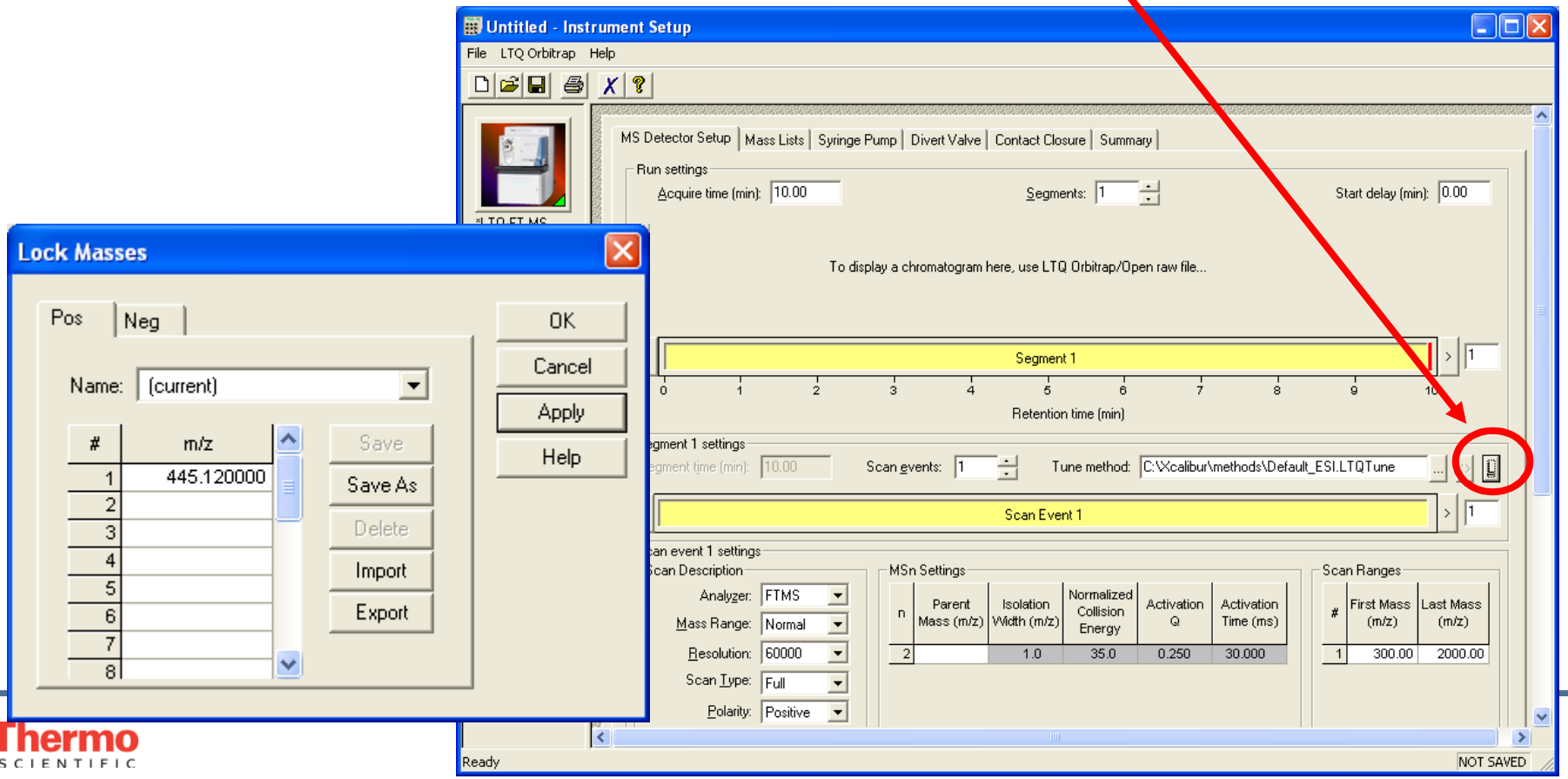


1. Locking ON

2. Enter accurate mass of a known compound

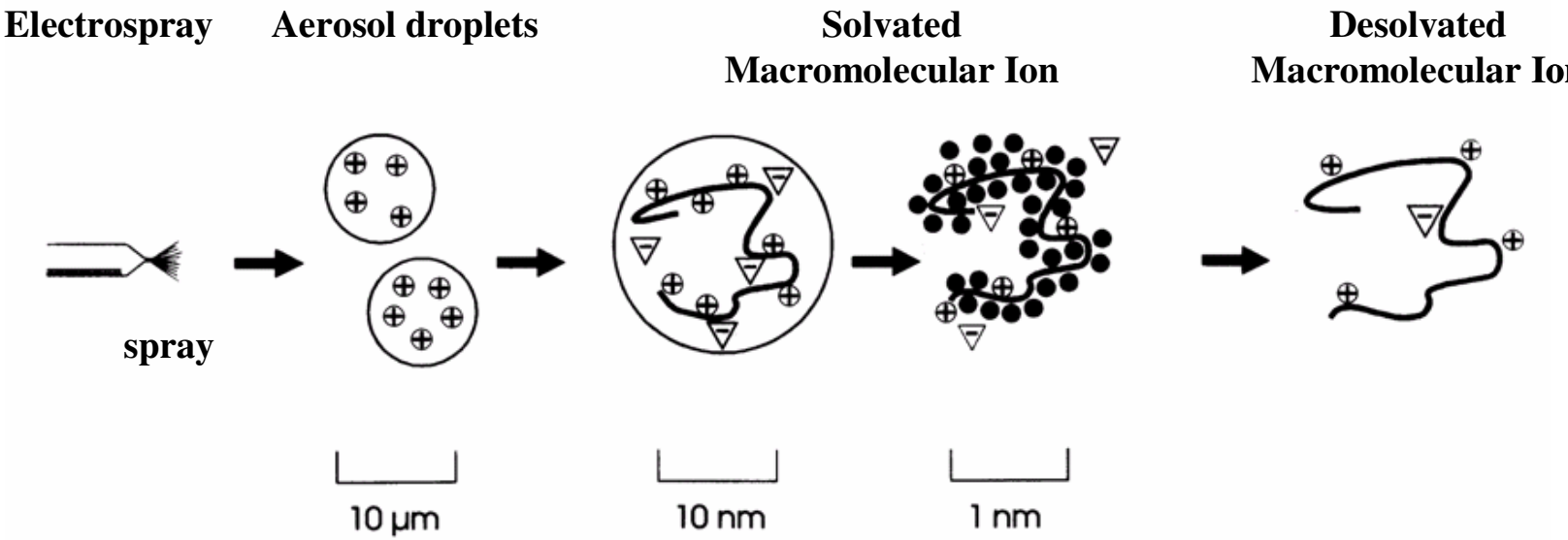
Internal Calibration: Method Setup

- Enable lock mass in the current method '**L**'
- Can edit lock masses
- A lock mass from the 'method' take preference over the lock mass from 'Tune' file



Thermo SCIENTIFIC

Ionisation and desorption in ESI mass spectrometry



Ions are formed by spraying the solution from a fine needle into an electric field at atmospheric pressure.

The droplets pick up charges, in the form of protons; solvent evaporates as the droplets enter the lower pressure regions of the instrument leaving “naked” intact peptides with varying numbers of charges (z =1,2,3,4…n).

Protein Quantification

Why Quantitate on the protein level

- There can be a big disconnect between the mRNA level and the protein level
  - Protein Turn over
  - Post-translational modifications

Quantitation in Proteomics

- Gel based
- LC-MS based
- Label free approaches
- Stable isotope labelling

Differential proteomics

Comparison of the relative amounts of proteins between two or more biological samples

- Diseased vs healthy
- Larva vs pupa vs adult
- Sepsis - die vs live

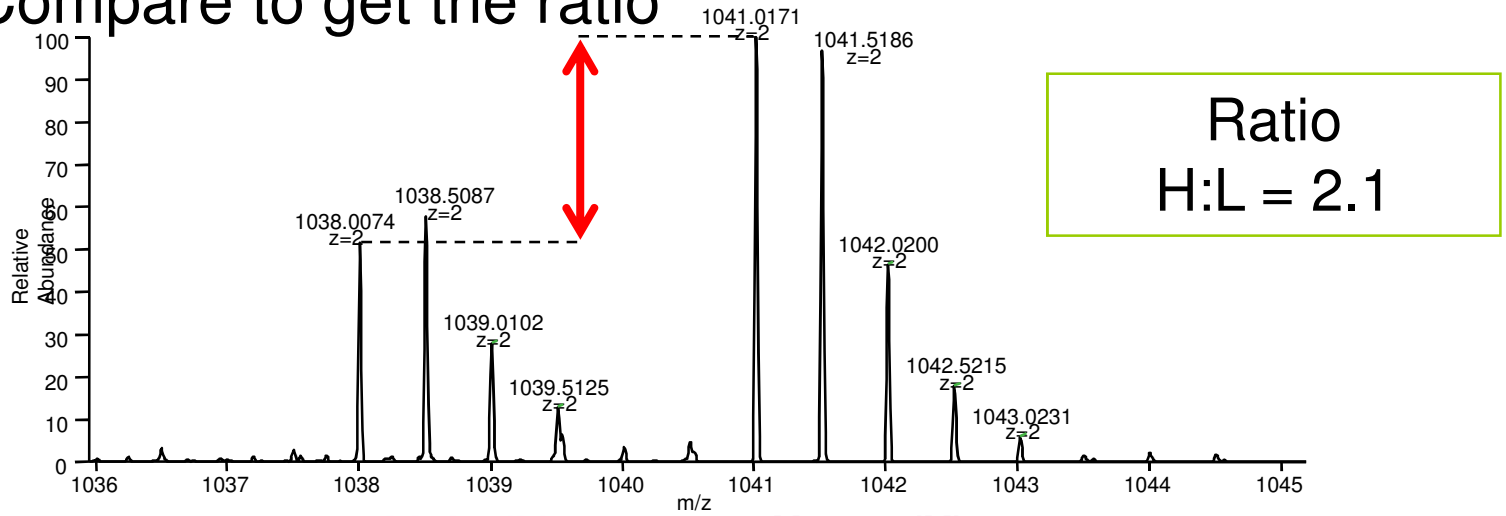
Labeling methods

Allow to:

- Samples to be run in tandem
- Labeled and unlabeled peptides to elute together (LC)
- Labeled and unlabeled peptides ionize similarly (ESI)

Quantitation with Isotope Labels

- Incorporate a ‘mass tag’ into protein or peptide
  - Create **heavy** and light version
- Mix the labeled samples
- Analyse together
- Use mass spectrometer to differentiate the heavy from light versions
- Compare to get the ratio



Ratio
H:L = 2.1

Light (L) **Heavy (H)**

General idea: Differential proteomics

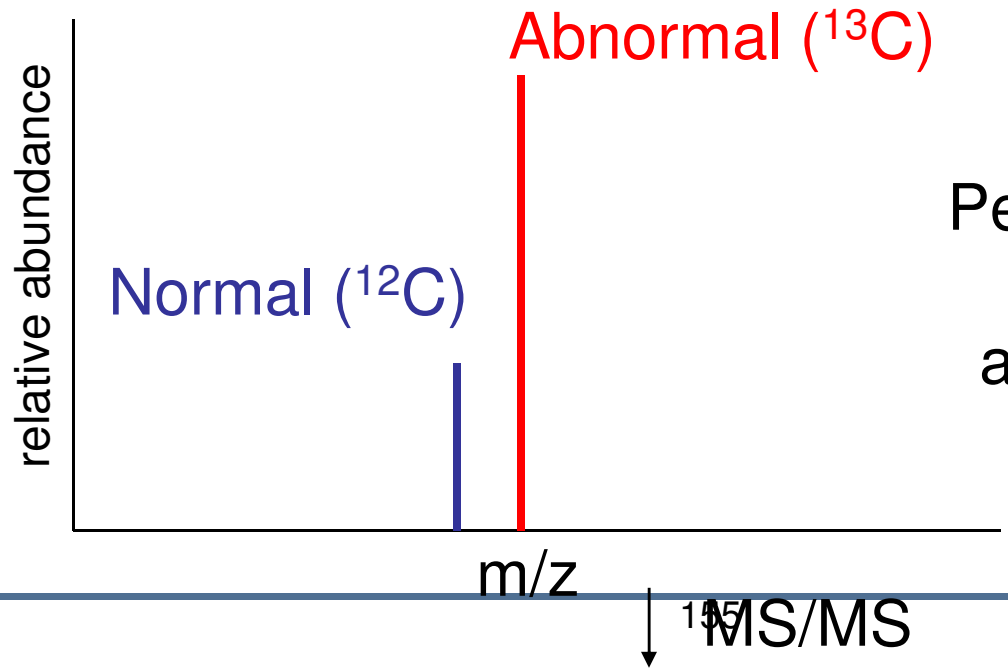
Enzymatic digestion (Trypsin)

↓

LC/MS/MS

In the LC, heavy and light co-elute

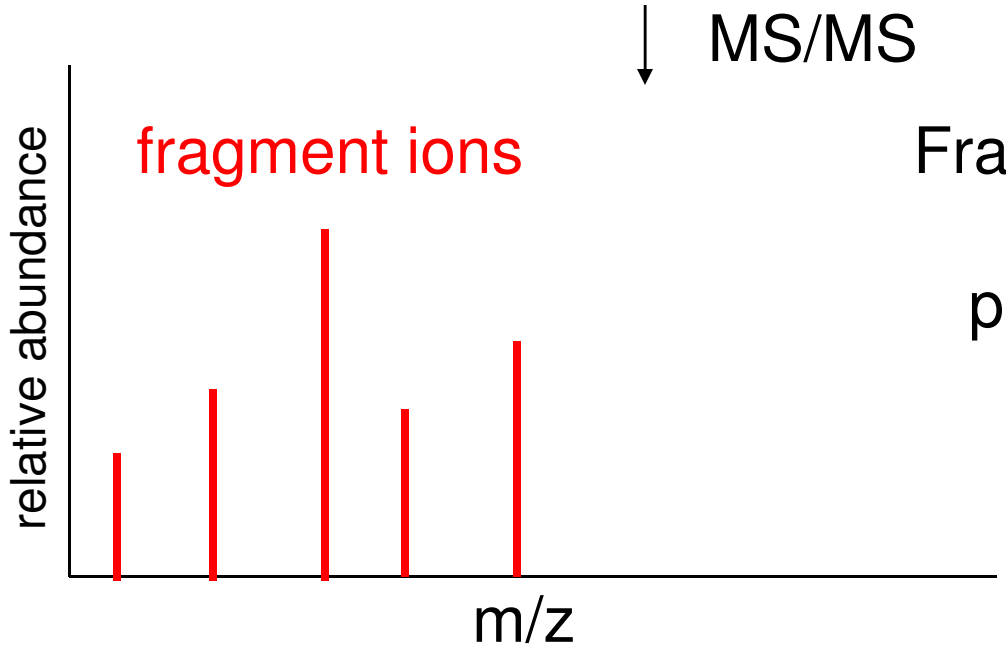
↓ MS



Peptide found to be up regulated in abnormal sample

↓ ^{155}MS/MS

General idea: Differential proteomics



Fragment ion masses help to identify peptide sequence belonging to a specific protein

Protein determined to be up regulated in abnormal sample

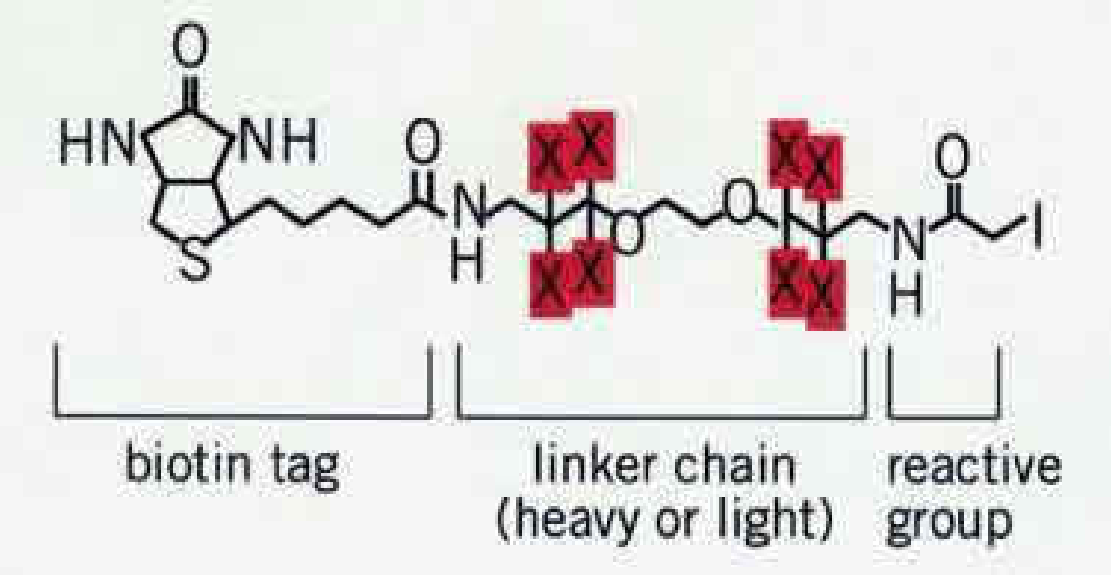
Stable isotope labeling methods

- **I**sotope-**C**oded **A**ffinity **T**agging (**ICAT**)
  - Hydrogen vs deuterium (8 Da separation)
  - ^{13}C vs ^{12}C (9 Da separation)

- **I**sotope **T**agging for **R**elative and **A**bsolute protein **Q**uantitation (**iTRAQ**)
  - Four isobaric reagents
  - Compare quantity of reporter molecule in MS/MS

- **S**table **I**sotope **L**abeling of **A**mino acids in **C**ell culture (**SILAC**)
  - Cell growth in stable isotope-enriched media (^{13}C Glucose, ^{15}N Ammonium, ^{2}H Water)

Isotope-Coded Affinity Tagging (ICAT)



X = Hydrogen (Light)

or

Deuterium (Heavy)

8 Da difference

Used to affinity capture reacted cysteine containing peptides (avidin)

Binds to and modifies cysteine residues (alkylation)

http://dir.niehs.nih.gov/proteomics/emerg2.htm

ICAT

- Analyze only the peptides containing cysteine
    - Reduce complexity of sample
    - Cys relative abundance 2.5%
- Problems with getting the hydrogen and deuterium labeled -
    - peptides to co-elute
- Changed linker chain so that carbons were isotopically labeled (^{13}C vs ^{12}C, 9 Da separation)
    - ^{13}C and ^{12}C peptides co-elute
- Reducing complexity also means missing a lot of peptides and reduction in confidence in protein identification

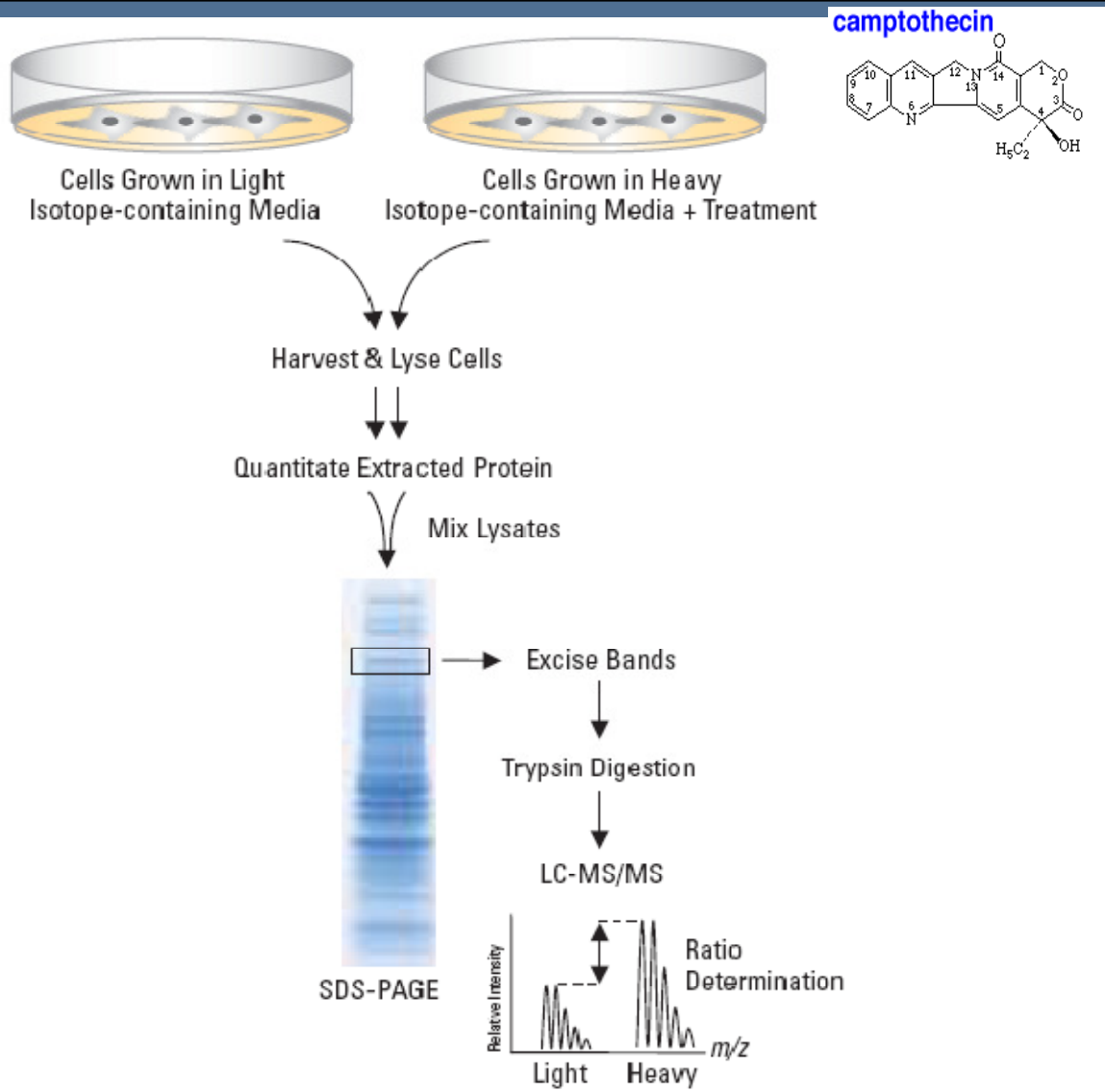
Stable **I**sotope **L**abeling of **A**mino acids in **C**ell culture (**SILAC**)

- SILAC uses *in vivo* metabolic incorporation of "heavy" ^{13}C- or ^{15}N-labeled amino acids into proteins followed by mass spectrometry-based analysis.
- 'Kits' commercially available

Mann, M. (2006). Functional and quantitative proteomics using SILAC. *Nature Reviews*. 7:952-959.

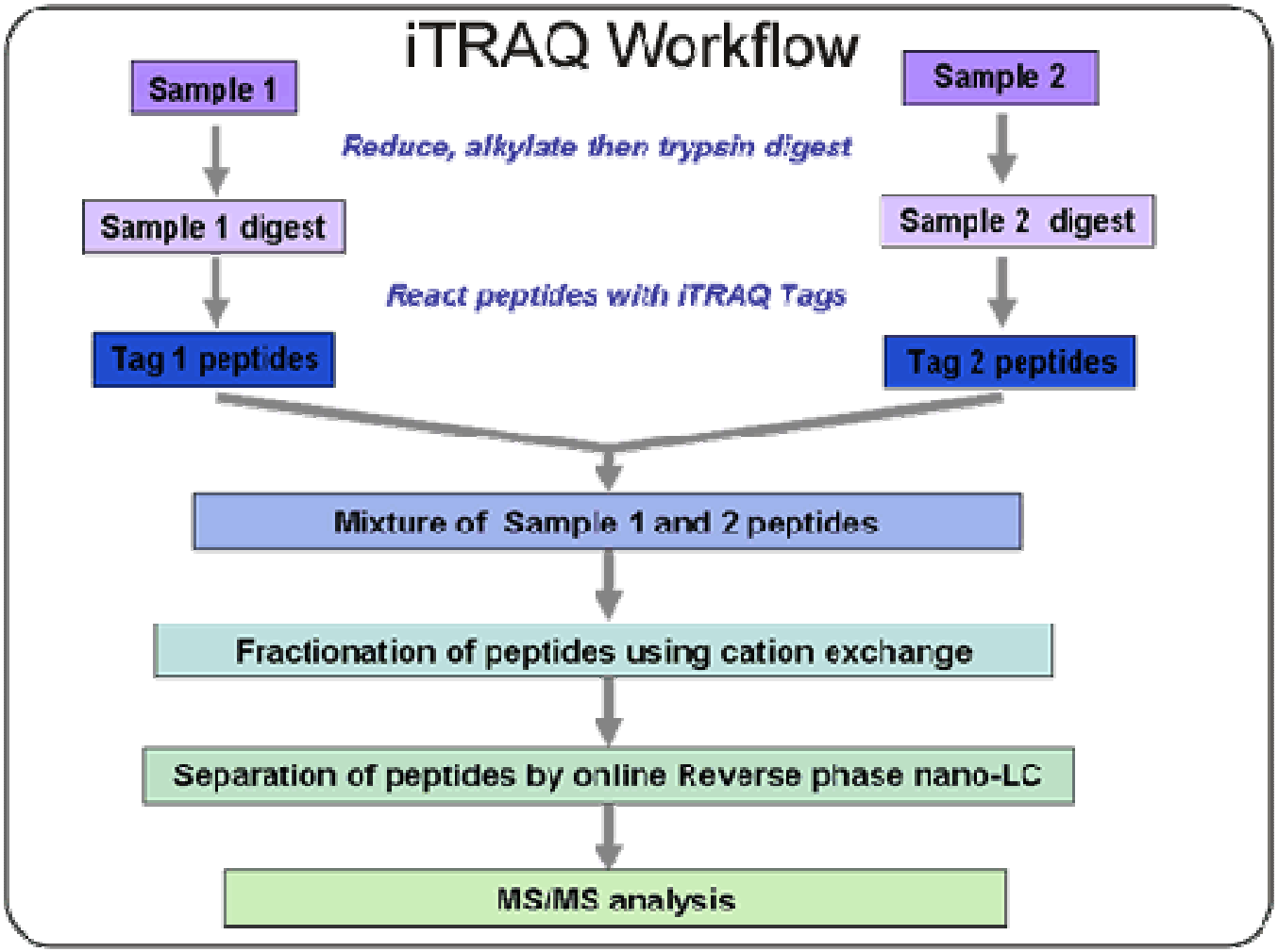
Everly, P.A., *et al*. (2004).Quantitative cancer proteomics: Stable isotope labeling with amino acids (SILAC)
as a tool for prostate cancer research. *Mol & Cell Proteomics*. 3.7: 729-735.

SILAC workflow



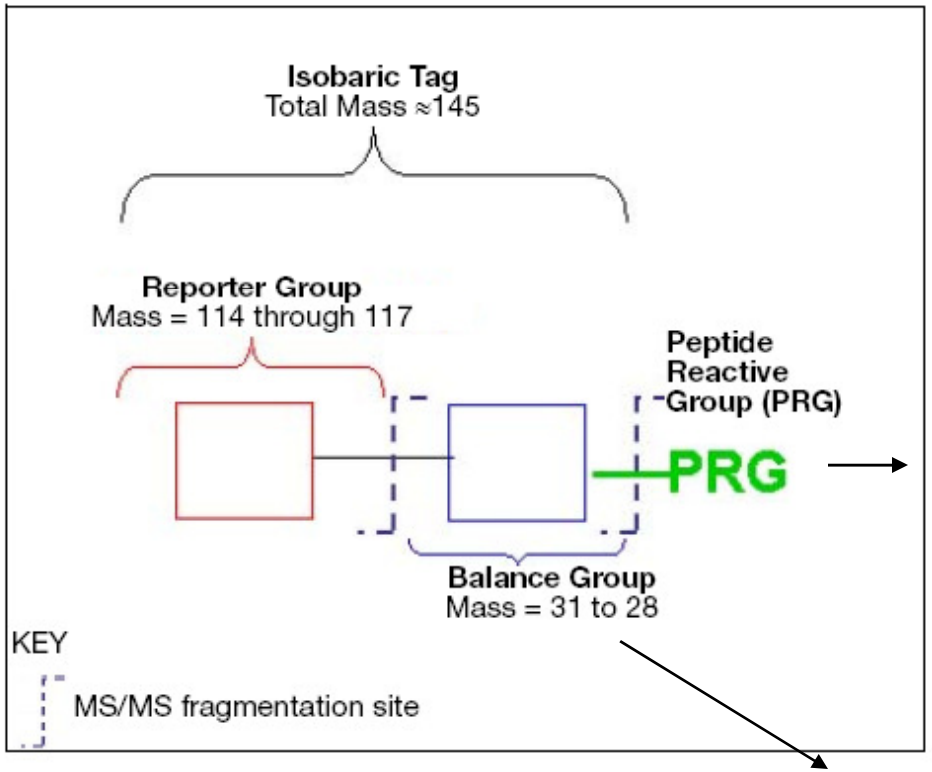
Mann, M. (2006). Functional and quantitative proteomics using SILAC. *Nature Reviews*. 7:952-959.

Isotope **T**agging for **R**elative and **A**bsolute protein **Q**uantitation (**iTraq**)



iTRAQ™ Reagent Structure

iTRAQ™ Reagents are non-polymeric, isobaric tagging reagents consisting of a reporter group, a balance group, and a peptide reactive group, as shown in Figure 1-2.

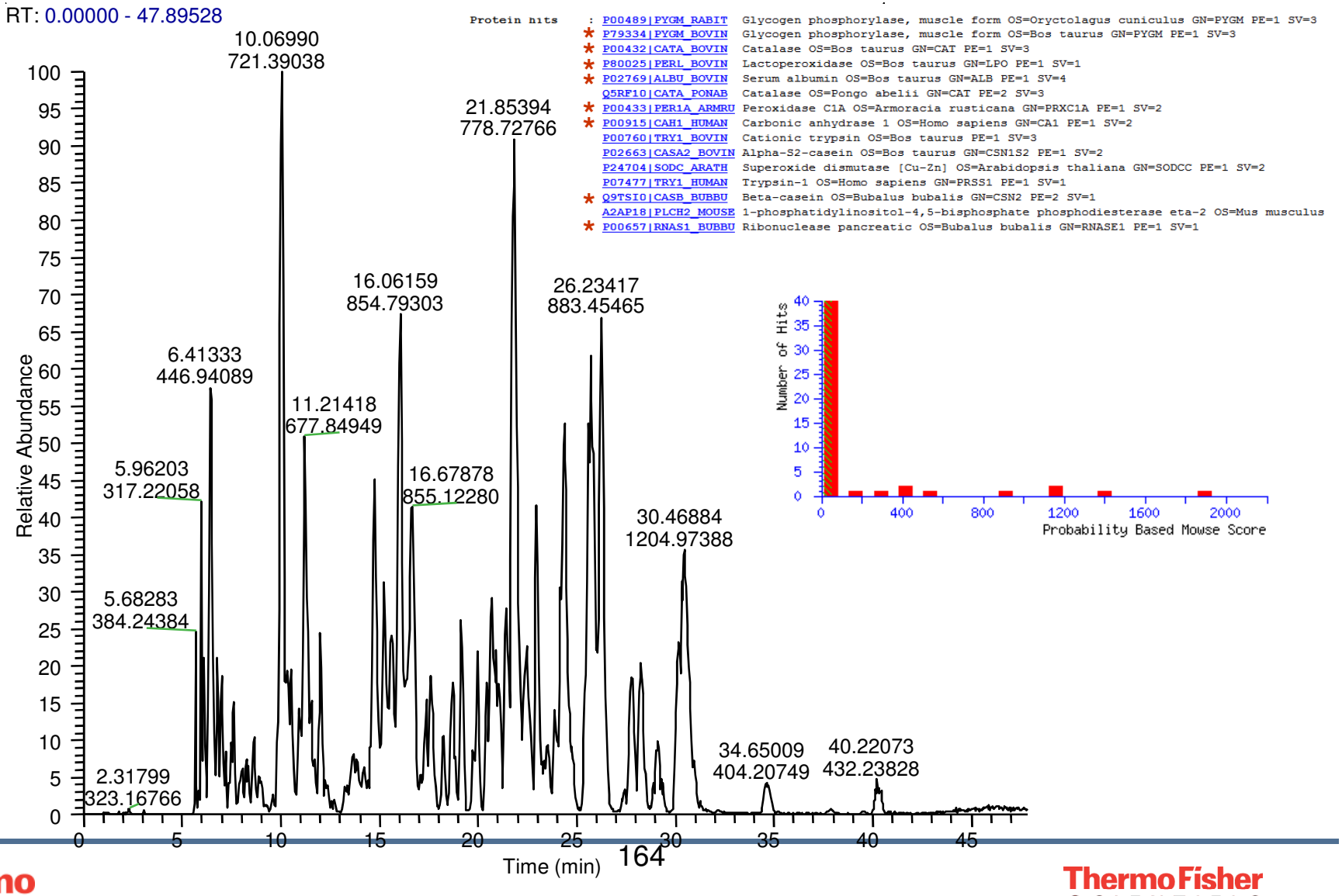


covalently links an iTRAQ™ reagent isobaric tag with each lysine side chain and N-terminal group of a peptide

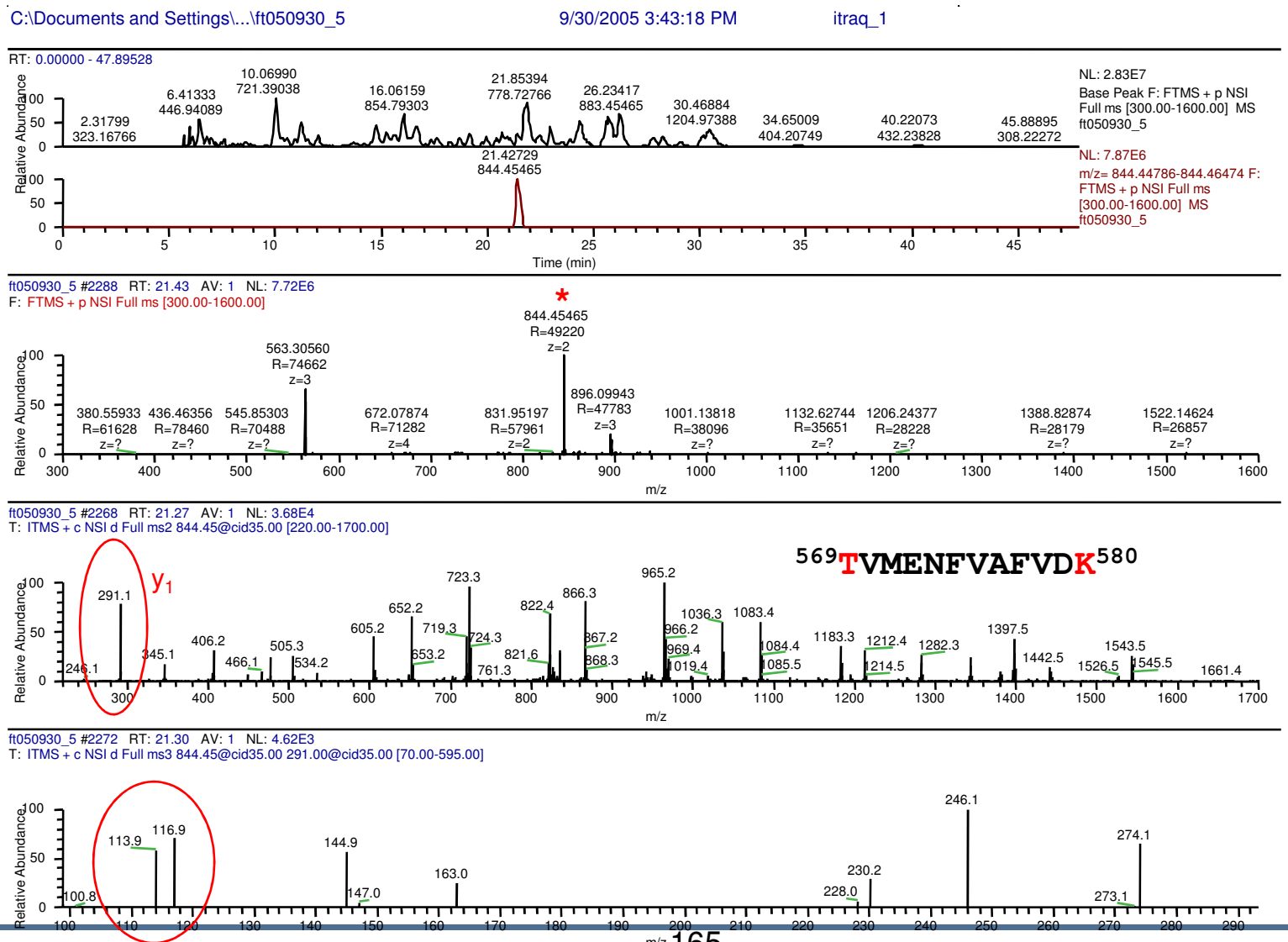
ensures that an iTRAQ™ reagent-labeled peptide, whether labeled with iTRAQ™reagent 114, 115, 116, or 117, displays at the same mass

Figure 1-2 iTRAQ™ Reagent structure

Base peak chromatogram and MASCOT identification of proteins from the PRG study sample



MS2 identification of an iTRAQ labeled tryptic peptide from BSA and quantification of the tag in a subsequent MS3 scan

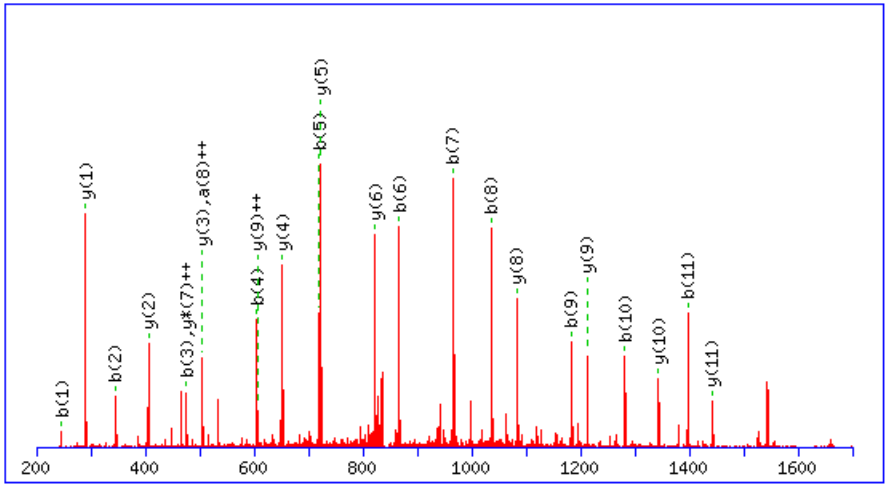


Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

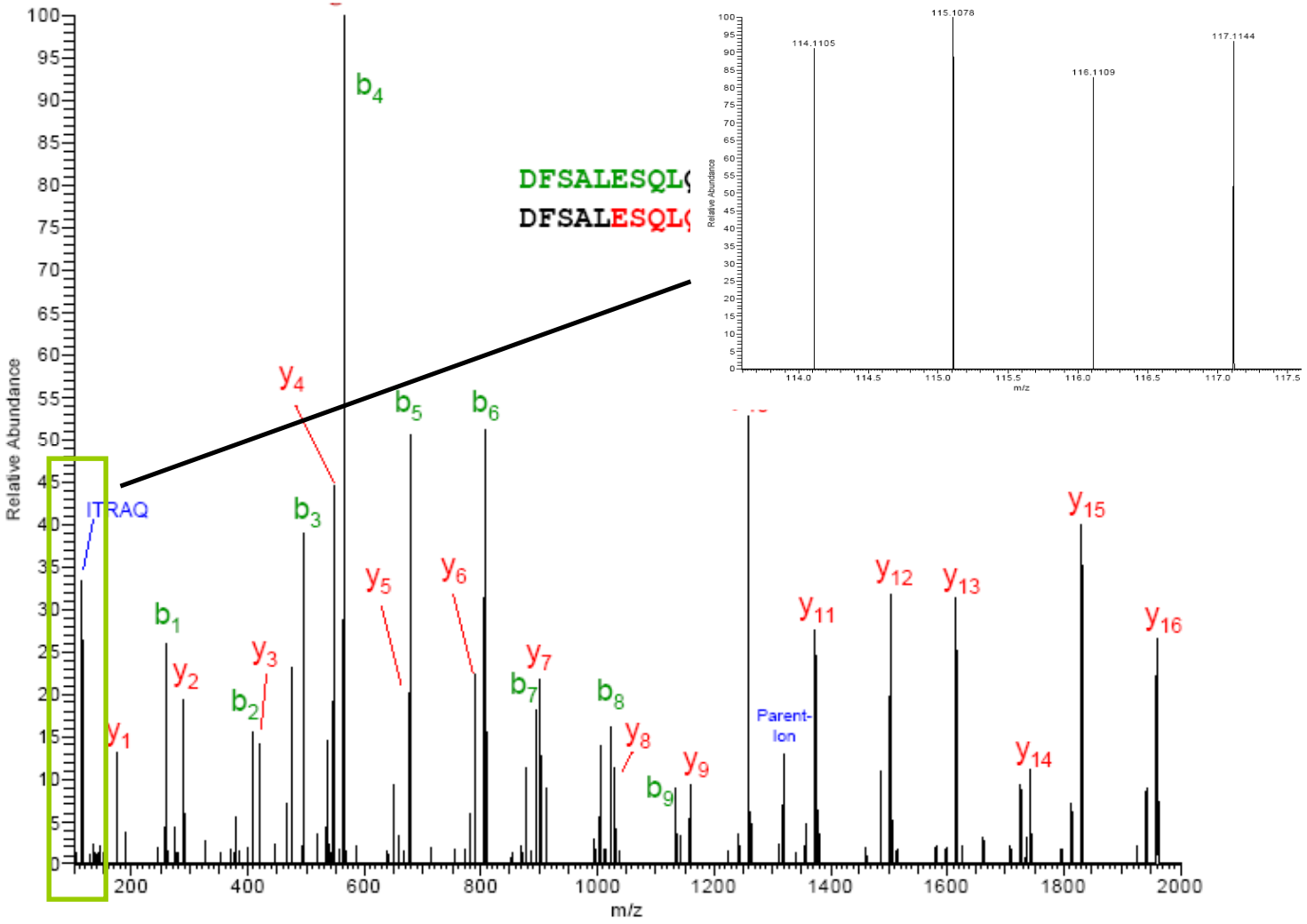
MS/MS identification of the iTRAQ labeled peptide

^{569}TVMENFVAFVDK^{580}



#	Immon.	a	a^{++}	a*	a*$^{++}$	b	b^{++}	b*	b*$^{++}$	Seq.	y	y^{++}	y*	y*$^{++}$	#
1	74.0600	218.1621	109.5847			246.1570	123.5821			T					12
2	72.0808	317.2305	159.1189			345.2254	173.1164			V	1442.7470	721.8771	1425.7205	713.3639	11
3	104.0528	448.2710	224.6391			476.2659	238.6366			M	1343.6786	672.3429	1326.6520	663.8297	10
4	102.0550	577.3136	289.1604			605.3085	303.1579			E	1212.6381	606.8227	1195.6116	598.3094	9
5	87.0553	691.3565	346.1819	674.3300	337.6686	719.3514	360.1794	702.3249	351.6661	N	1083.5955	542.3014	1066.5690	533.7881	8
6	120.0808	838.4249	419.7161	821.3984	411.2028	866.4199	433.7136	849.3933	425.2003	F	969.5526	485.2799	952.5260	476.7667	7
7	72.0808	937.4934	469.2503	920.4668	460.7370	965.4883	483.2478	948.4617	474.7345	V	822.4842	411.7457	805.4576	403.2324	6
8	44.0495	1008.5305	504.7689	991.5039	496.2556	1036.5254	518.7663	1019.4988	510.2531	A	723.4158	362.2115	706.3892	353.6982	5
9	120.0808	1155.5989	578.3031	1138.5723	569.7898	1183.5938	592.3005	1166.5672	583.7873	F	652.3786	326.6930	635.3521	318.1797	4
10	72.0808	1254.6673	627.8373	1237.6407	619.3240	1282.6622	641.8347	1265.6357	633.3215	V	505.3102	253.1588	488.2837	244.6455	3
11	88.0393	1369.6942	685.3508	1352.6677	676.8375	1397.6892	699.3482	1380.6626	690.8349	D	406.2418	203.6245	389.2153	195.1113	2
12	245.2094									K	291.2149	146.1111	274.1883	137.5978	1

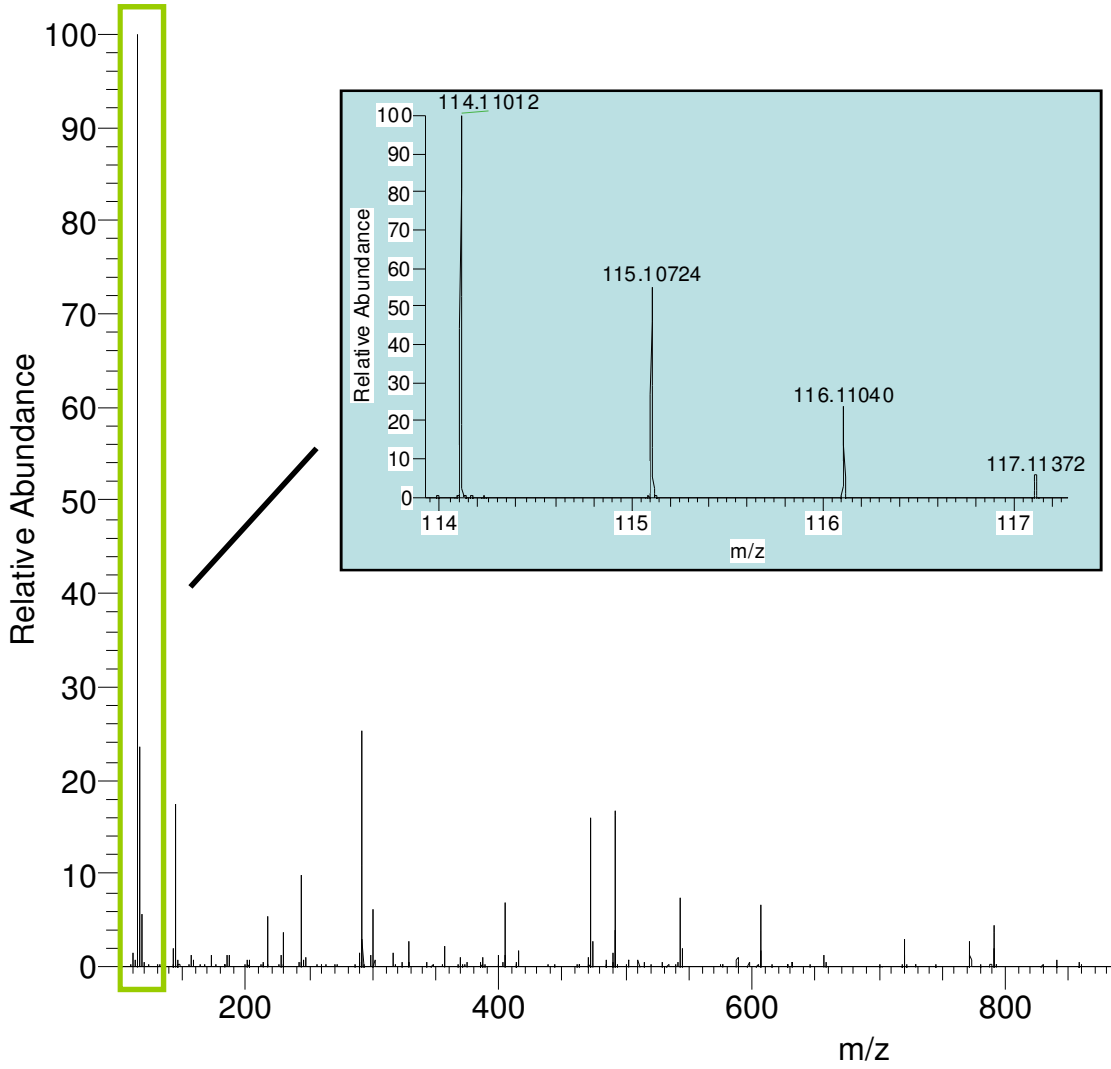
HCD MS/MS of iTRAQ labelled Peptide



Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

Carbonic Anhydrase (1 : 0.5 : 0.2 : 0.03)



#	b	Seq	y	#
1	244.1778	V		9
2	357.2618	L	1018.5901	8
3	472.2888	D	905.5060	7
4	543.3259	A	790.4791	6
5	656.4099	L	719.4420	5
6	771.4369	D	606.3579	4
7	858.4689	S	491.3310	3
8	971.5530	I	404.2989	2
9		K	291.2149	1

BLAIS PROTEOMICS CENTER

Thermo SCIENTIFIC

ThermoFisher SCIENTIFIC

HCD/iTRAQ

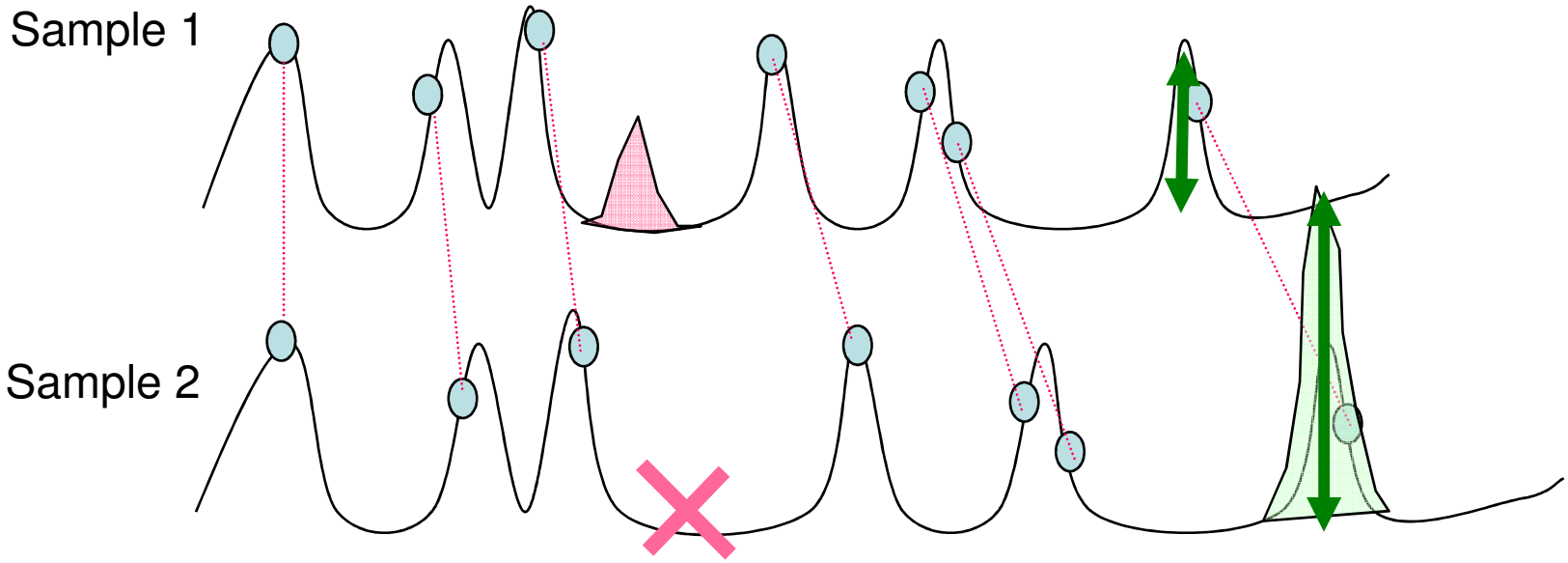
- No „low mass cutoff“ with HCD fragmentation
- Good quantitation in the low mass region
- Good sensitivity

➔No limits to detect reporter ions and fragments in one spectrum

➔Good peptide ID

➔Good quantitation

Label-Free Approach



- Align chromatographic traces (The quality of the alignment software is a key parameter in the comparative LC-MS procedure)
- Match peptide mass/charge state
- Quantify based on elution profile of each peptide
- Identify using MS/MS spectra

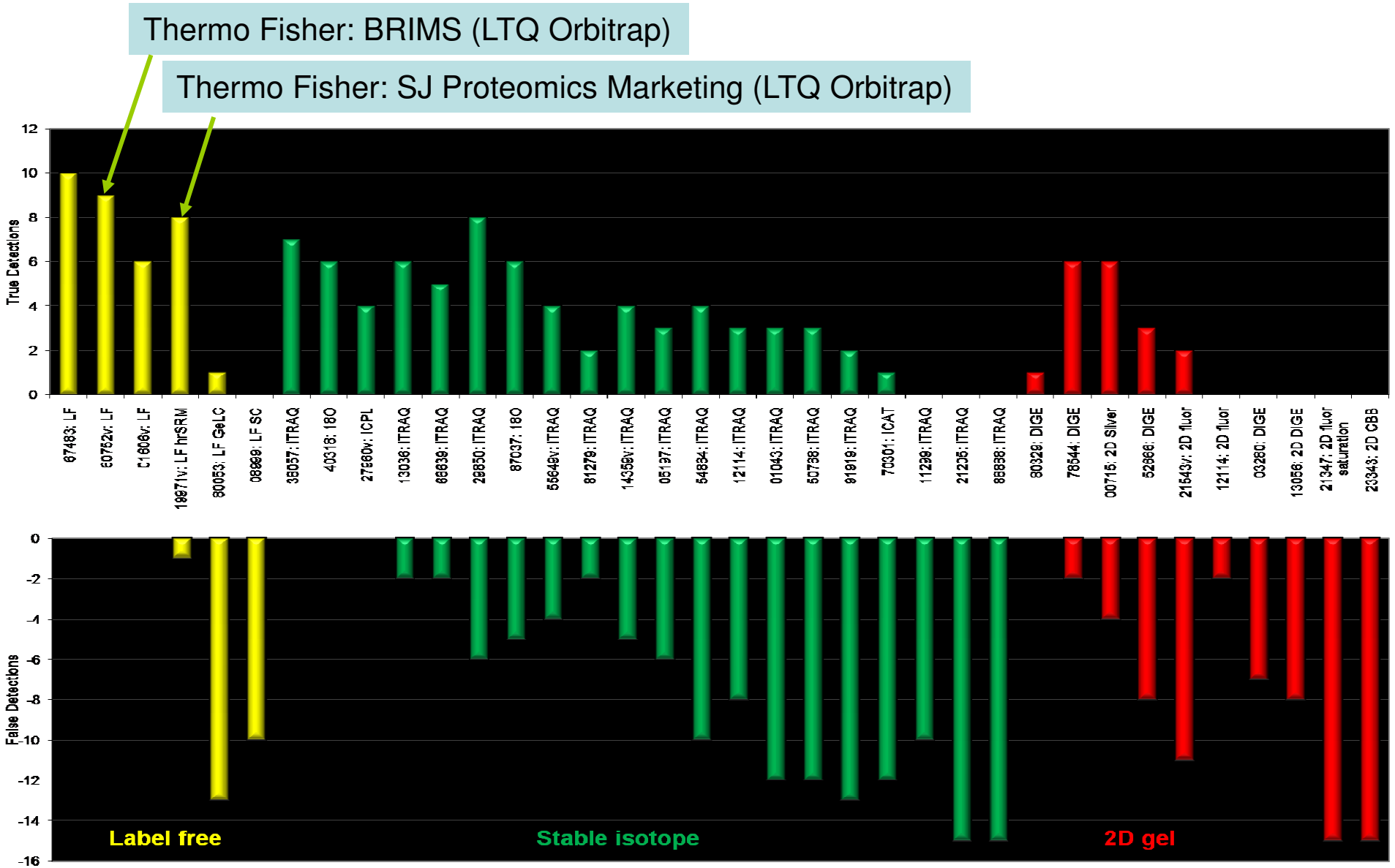
Table 1. Overview of LC-MS alignment software for proteomics solutions

Software name	Supplier / author	Database/ environment	availability	Functionality	website	reference
PLGS IdentityE Expression Informatics	Waters Corp	PLGS	commercial	f, h,i,b,a,r,s,l	http://www.waters.com/	[15, 16, 58, 59]
SIEVE	Thermo Scientific	BioWorks	commercial	p,h,v,b,a,r,s,l	http://www.thermo.com/	
DeCyderMS	GE Healthcare		commercial	f,h,i,b,v,a,r,s,l	http://www.gelifesciences.com/	[50]
Rosetta Elucidator	Rosetta Biosoft-ware		commercial	f,h,i,b,v,a,r,s,l	http://www.rosettabio.com/products/elucidator/default.htm	
MS-Xelerator	MsMetrix		commercial	f,l,i,b,a,r,s,	http://www.msmetrix.com	
MassView	SurroMed		custom	f,l,i,b,a,r,s,l		[61–63]
MetAlign	WUR		free for acad.	p,l,b,a,s	www.metalign.wur.nl	[33]
MzMine	VTT Finland		open source	f,h,v,a,r	http://mzmine.sourceforge.net/index.shtml	[42, 43]
MSight	SIB		open source	f,h,i,v,(a)	http://www.expasy.org/MSight/	[64]
MS Inspect	CPL (Fhcrc)	CPAS	open source	f,h,v,a,r (l,d)	http://proteomics.fhcrc.org/CPL/msinspect.html	[41, 65]
SpecArray	ISB /SPC	TPP	open source	f,h,i,v,a,r,s	http://tools.proteomecenter.org/SpecArray.php	[49]
PePPer	BROAD MIT	Genepattern	open source	h,a,r,s,l	http://www.broad.mit.edu/cancer/software/genepattern/desc/proteomics.html	[60, 66]
VIPER	PNNL	PRISM	open source	f,h,i,b,v,a,r,s,l,d	http://ncrr.pnl.gov/software	[57, 67]
OpenMS	Berlin Saarland Tubingen Univ.	TOPP	open source	(f,h,i,b,v,a,r,s,l,d)	www.openMS.de	[68–70]
SuperHirn	IMSB @ETH	Corra	open source	f,i,b,v,a,r,s	http://tools.proteomecenter.org/SuperHirn.php	[53]
CPM (continous profile models)	Listgarten/Emili	MatLab	free for acad.	l,a	http://www.cs.toronto.edu/~jenn/CPM/	[52]
Xalign	Purdue Univ	Xmass	upon request	(f,h),i,a,s	zhang100@purdue.edu	[71]
Fischer et.al.	ETH		not described	h,a	http://people.inf.ethz.ch/befische/	[46]
CRAWDAD	Washington Univ		upon request	f,l,i,a,r,s,l,d	http://proteome.gs.washington.edu/software/crawdad	
CHAMS	Inst Pasteur, Paris		web server	h,a,s	http://www.pasteur.fr/recherche/unites/BioIsys/chams/index.htm	[51, 72]
OBI-WARP	Univ. Texas		open source	a,r,l	http://bioinformatics.icmb.utexas.edu/obi-warp/	[73]
LCMSWARP	PNNL	PRISM	open source	h,a	http://ncrr.pnl.gov/software	[74]
LCMS2D	Albert Einstein College of Medicine				http://www.bioc.aecom.yu.edu/labs/angellab/	[75, 76]
PETAL	CPL (Fhcrc)	CPAS	open source	a	http://peiwang.fhcrc.org/research-project.html	[77, 78]

p/f: peak/feature detection; h/l: high/low resolution; i: de-isotoping; b: batch processing; v: LCMS 2-D visualization; a: alignment; r: result visualization; s: statistical analysis; l: link MS to MS/MS; d: results database

Comparative LC-MS: A landscape of peaks and valleys, Proteomics, 2008
Antoine H. P. America and Jan H. G. Cordewener

ABRF Competition PRG 2007

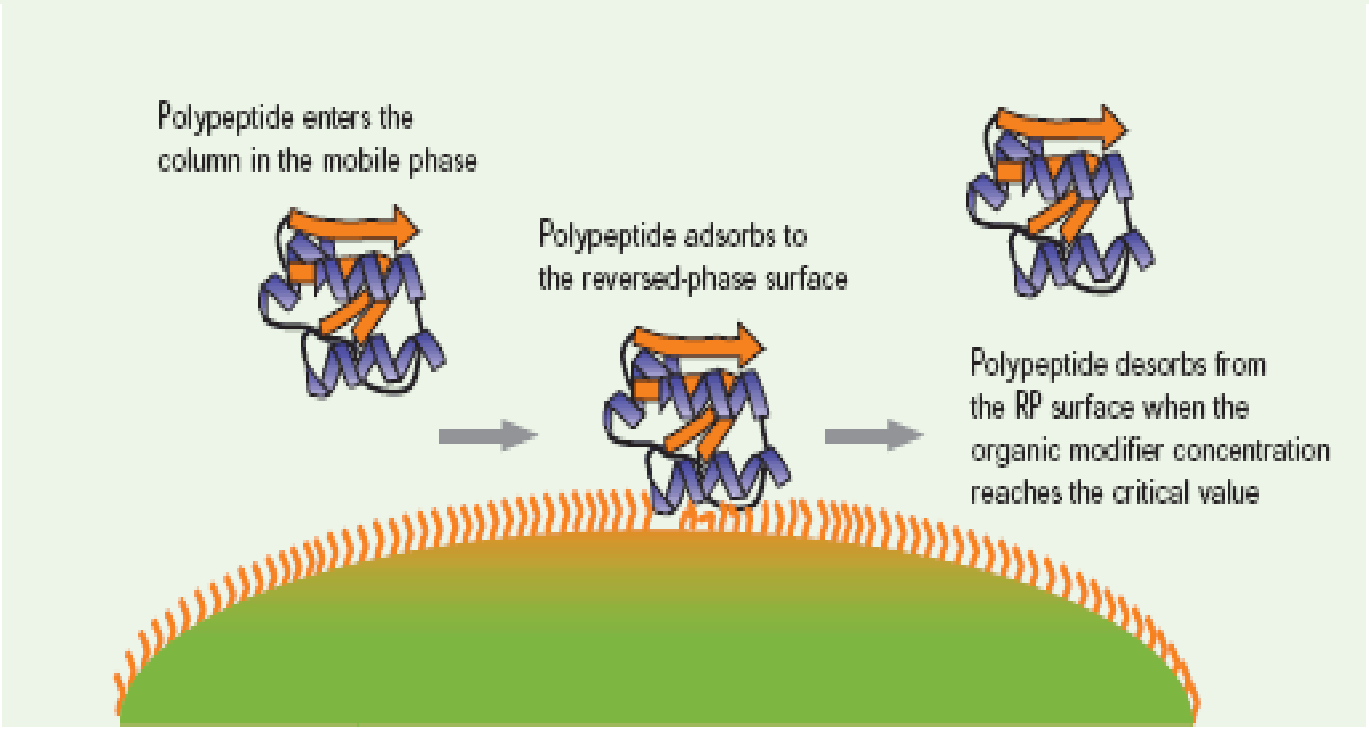


RP-HPLC basics

Analysis and purification of proteins and peptides by Reversed-Phase HPLC

- Separation of peptide fragments from enzymatic digests.
- Purification of natural and synthetic peptides.
- Study enzyme subunits and research cell functions.
- Analysis of protein therapeutic products.
- To verify conformation and to determine degradation products in intact proteins.

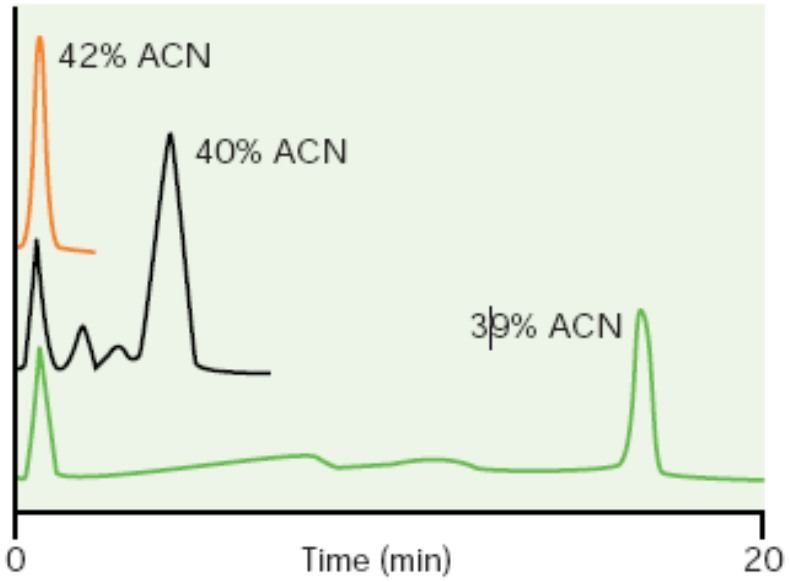
Mechanism of interaction between polypeptides and RP-HPLC columns



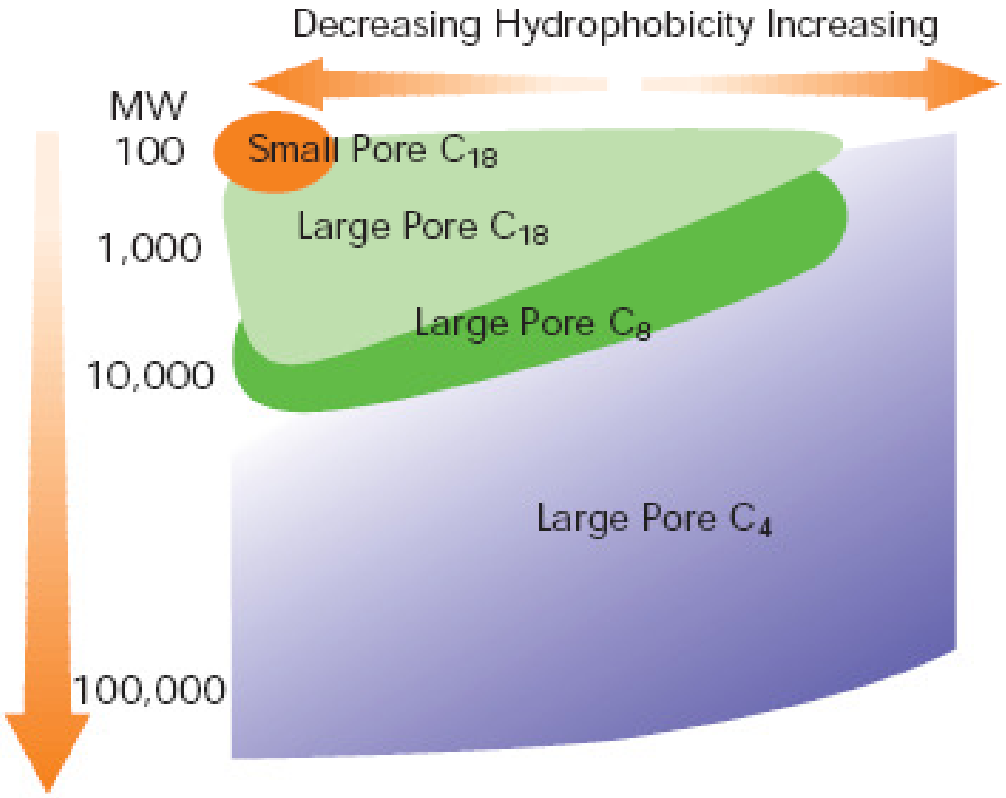
RP-HPLC separates polypeptides based on subtle differences in the “hydrophobic foot” of the polypeptide being separated.

Isocratic versus gradient elution

Shallow gradients can be used very effectively to separate similar polypeptides where isocratic separation would be impractical.



The role of the column in polypeptide separations by RP-HPLC



- The HPLC column provides the hydrophobic surface onto which the polypeptides adsorb.

- Columns consist of stainless steel tubes filled with small diameter, spherical adsorbent particles, generally composed of silica whose surface has been reacted with silane reagents to make them hydrophobic.

The role of the column in polypeptide separations by RP-HPLC

Adsorbent Pore Diameter

Polypeptides must enter a pore in order to be adsorbed and separated.

(pores of about 300 Å)

Some peptides (<~2,000 MW) may also be separated on particles of 100 Å pores.

Adsorbent Particle Size

Smaller diameter particles generally produce sharper peaks and better resolution.

Adsorbent Phase Type

Hydrophobic phase is usually a linear aliphatic hydrocarbon of eighteen (C18), eight (C8) or four (C4) carbons.

Column Dimensions: Length

- Column length does not significantly affect separation and resolution of proteins.

- Consequently, *short columns* of 5–15 cm length are often used for *protein separations*.

- Columns of *15–25 cm length* are often used for the separation of *synthetic and natural peptides and enzymatic digest maps.*

Column back-pressure

• Column back-pressure is directly proportional to the column length.

• In case of viscous solvents, such as isopropanol, shorter columns will result in more moderate back-pressures.

Column dimensions: diameter

When the diameter of an HPLC column is **reduced**

1. flow rate is **decreased**
2. solvent used **decreased**
3. detection sensitivity is increased (smaller amounts of polypeptide can be detected).

Very small diameter HPLC columns (**(75 µm Diameter)**
are particularly useful when coupling HPLC with mass spectrometry (LC/MS).

Columns, example

	Column Diameter (mm)	Typical Flow Rate (1)	Sample Capacity (2)	Maximum Practical Sample Load (3)
Capillary 	0.075 0.15 0.30 0.50	0.25 µL/min 1 µL/min 5 µL/min 10 µL/min	0.05 µg 0.2 µg 1 µg 2 µg	
Microbore 	1.0	25–50 µL/min	0.05–10 µg	
Narrowbore 	2.1	100–300 µL/min	0.2–50 µg	
Analytical 	4.6	0.5–1.5 mL/min	1–200 µg	10 mg
Semi-preparative 	10	2.5–7.5 mL/min	1,000 µg	50 mg
Preparative 	22	10–30 mL/min	5 mg	200 mg
Process 	50 100	50–100 mL/min 150–300 mL/min	25 mg 125 mg	1,000 mg 5,000 mg

Organic modifier

The organic modifiers solubilizes and desorbs the polypeptide from the hydrophobic surface.

Organic solvents used in LC/MS

Acetonitrile (ACN)

- It is volatile and easily removed from collected fractions.
- It has a low viscosity, minimizing column back-pressure.
- It has little UV adsorption at low wavelengths.
- It has a long history of proven reliability in RP-HPLC polypeptide separations.

Organic solvents used in LC/MS

Isopropanol

- Used for large or very hydrophobic proteins.

disadvantage of isopropanol is its high viscosity.

- To reduce the viscosity of isopropanol - use a mixture of 50:50 acetonitrile: isopropanol. Adding 1–3% isopropanol to acetonitrile has been shown to increase protein recovery in some cases.

Organic solvents used in LC/MS

Ethanol

- Elute hydrophobic, membrane-spanning proteins and is used in process purifications.

Methanol
Dichloromethane
Clorofprm
Hexane

Ion-pairing reagents and buffers

The ion-pairing reagents or buffers sets the eluent pH and interacts with the polypeptide to enhance the separation.

Ion-pairing reagents and buffers

Trifluoroacetic acid

- It is volatile and easily removed from collected fractions.
- It has little UV adsorption at low wavelengths.
- It has a long history of proven reliability in RP-HPLC polypeptide separations.
- Enhancement of chromatographyc resolution.
- Adverse effect on ions formation in LC/MS

Ion-pairing reagents and buffers

Heptafluorobutyric acid (HFBA)

- Is effective in separating basic proteins.

Triethylamine phosphate (TEAP)

- Has been used for preparative separations.

Formic acid (FA)

- In concentrations of 10 to 60%, has been used for the chromatography of very hydrophobic polypeptides.

Ion exchange chromatography orthogonal analytical techniques

The benefits of ion exchange chromatography

- high sample loading capacity.
- crude solutions can be loaded onto ion-exchange columns.
- addition of urea, acetonitrile or non-ionic detergents to break-up complexes.
- optimization of elution selectivity

The benefits of Reversed Phase chromatography

- a high degree of selectivity based on differences in hydrophobicity or molecular conformation.
- use of volatile buffers or ion-pairing agents.
- freedom from interferences by salt or buffers from ion exchange.

LC/MS additives and buffers summary

Protons donors

Acetic acid

Formic acid

Proton acceptors

Amonium hidroxid

Amonia solutions

Ion-Pair reagents

Trichloroaceticacid (low than 0.02% v/v)

Trifloroaceticacid

Triethylamine

Trimethylamine

Buffers

Amonium acetate

Amonium formate

How to make a gradient