# *Mass Spectrometry-Based Proteomic Applications in Cell/Scaffold Products*

Daniel Martin
Institute for Systems Biology

# What is proteomics?

"Proteomics includes not only the identification and quantification of proteins, but also the determination of their localization, modifications, interactions, activities, and, ultimately, their function."

-Stan Fields in *Science*, 2001.

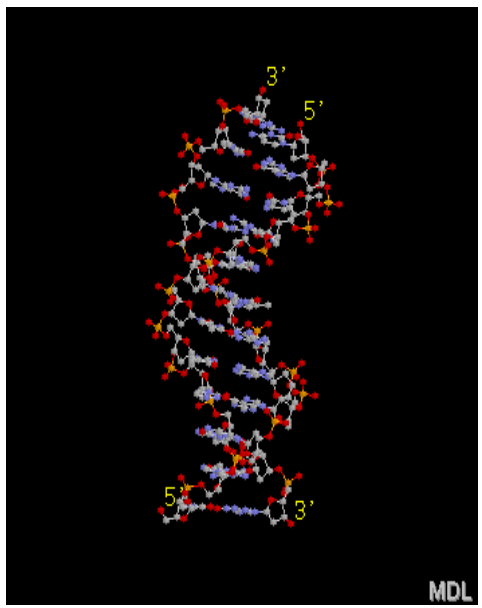
# The genome and the proteome: a comparison

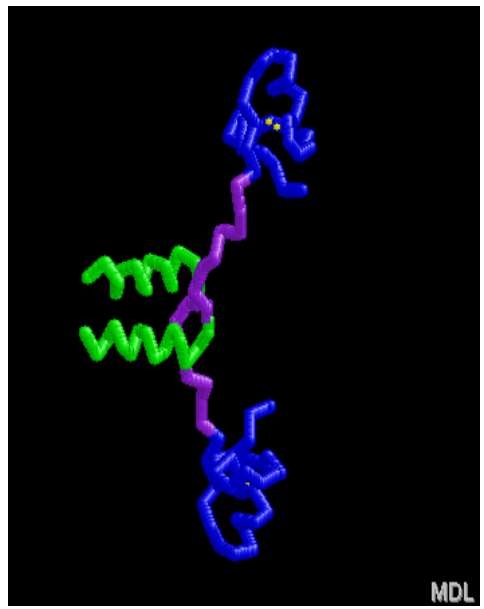
## Genome

- ➔ static
- ➔ able to amplify (PCR)
- ➔ homogeneous
- ➔ no variability in amount

## Proteome

- ➔ dynamic – condition dependent
- ➔ ***no amplification***
- ➔ non-homogenous
- ➔ high variability in amount (>$10^9$)



# Proteomics technologies and methods

- Two-dimensional gel electrophoresis
- mass spectrometry
- protein chips
- yeast 2-hybrid
- phage display
- antibody engineering
- high-throughput protein expression
- high-throughput X-ray crystallography

# Mass Spectrometry based proteomics: What it is and what it isn't

What it is:

- A highly powerful tool for protein identification and quantification
- Complementary to other technologies and analysis methods

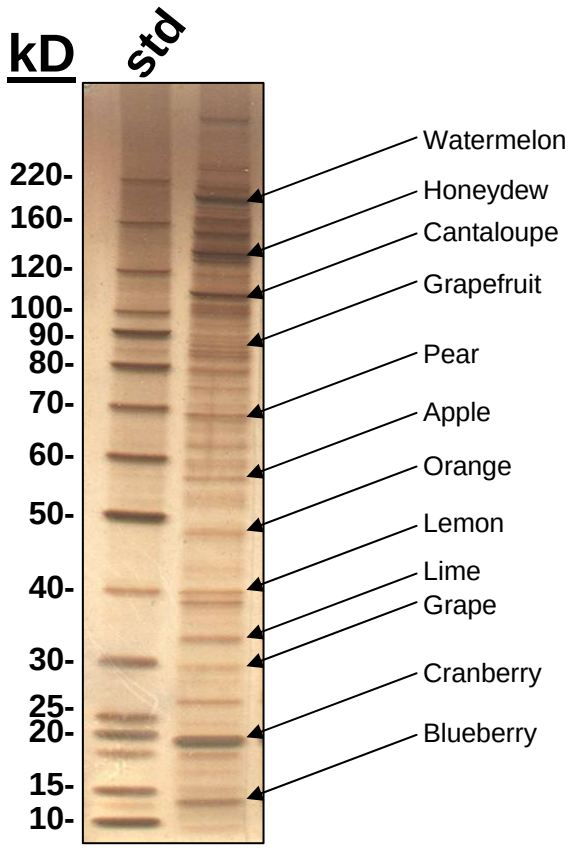
What it is not:

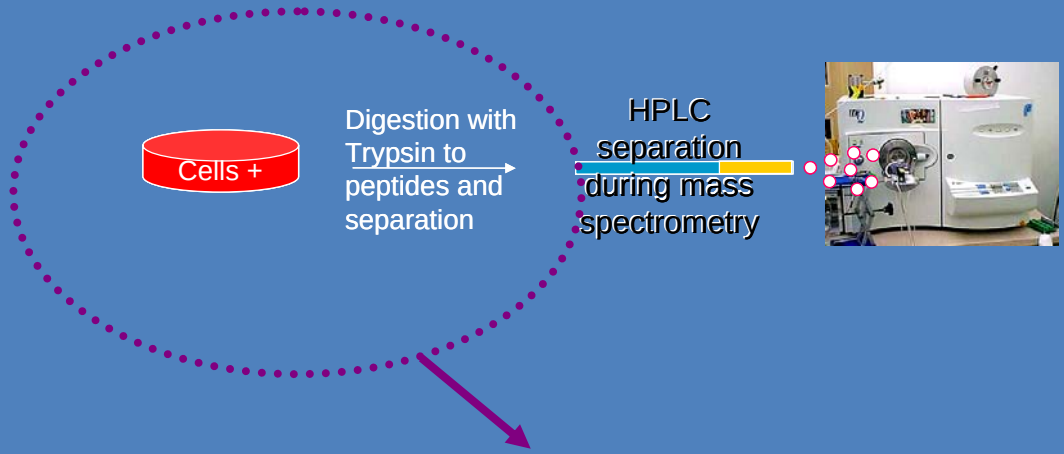
- Magic
- Able to give all the answers
- Simple (relatively speaking)
- Cheap

# Mass Spectrometry based proteomics: What can we measure?

- proteins in mixtures
- quantitative analysis of protein expression
- post-translational modifications:
  - phosphorylation (a challenge)
  - glycosylation (present/absent)
- protein interactions

# How to think about spectrometry based proteomics






Protein, MW = 10,000 +

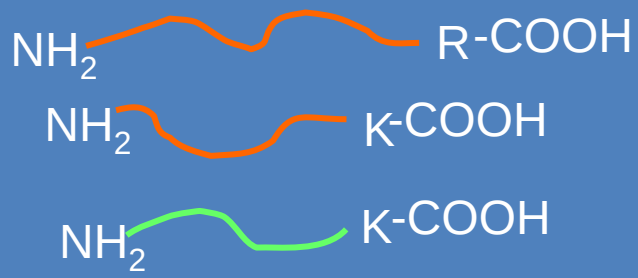
digest into peptides

Trypsin

N-TH**K**.NCPHIVVGTPG**R**.IPD-C

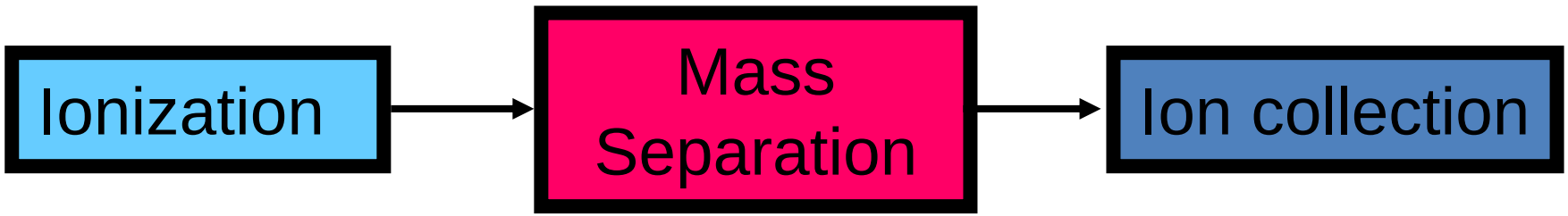
cleaves C-teminal side of arginine (R) and lysine (K)

Peptides,
MW < 4,000



# Mass Spectrometry Primer

A mass spectrometer
measures mass to charge ratio
or m/z



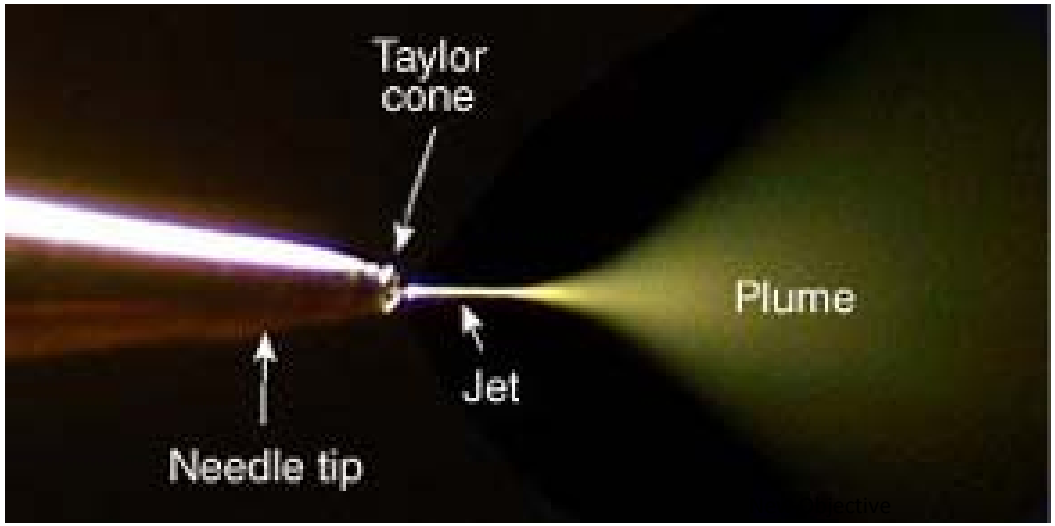

- MALDI
- Electrospray

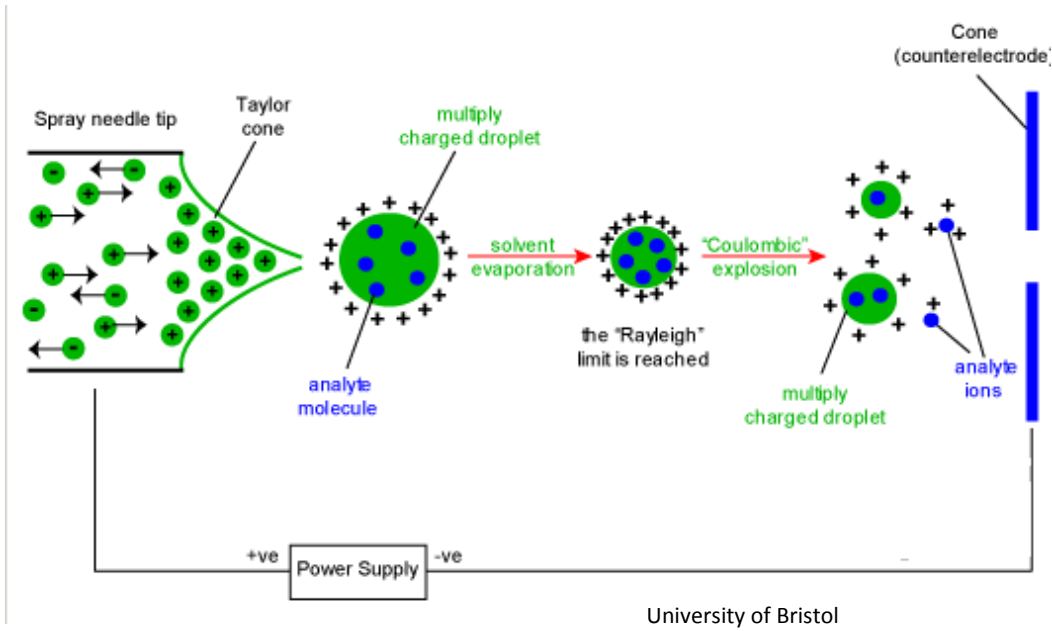
- quadrupole
- ion trap
- time-of-flight

- mass analysis

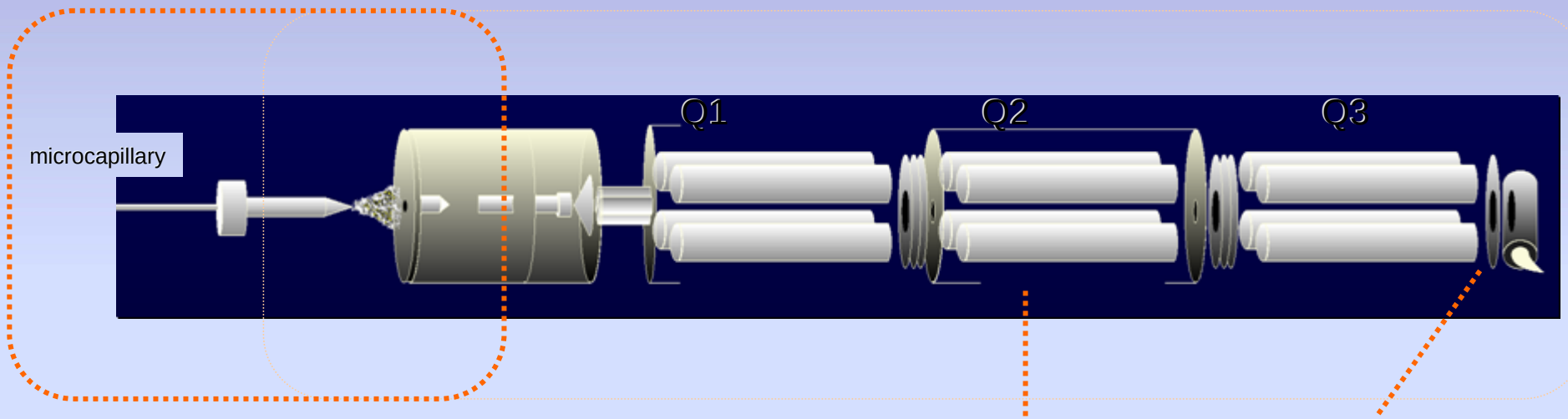
# Electrospray ionization (ESI)

High voltage placed on a fused silica column causes a spray of charged droplets which evaporate leaving charged peptides




# Triple Quadrupole Mass Spectrometer



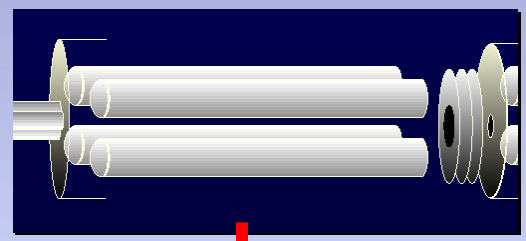

**electrospray source**
**Ionization**

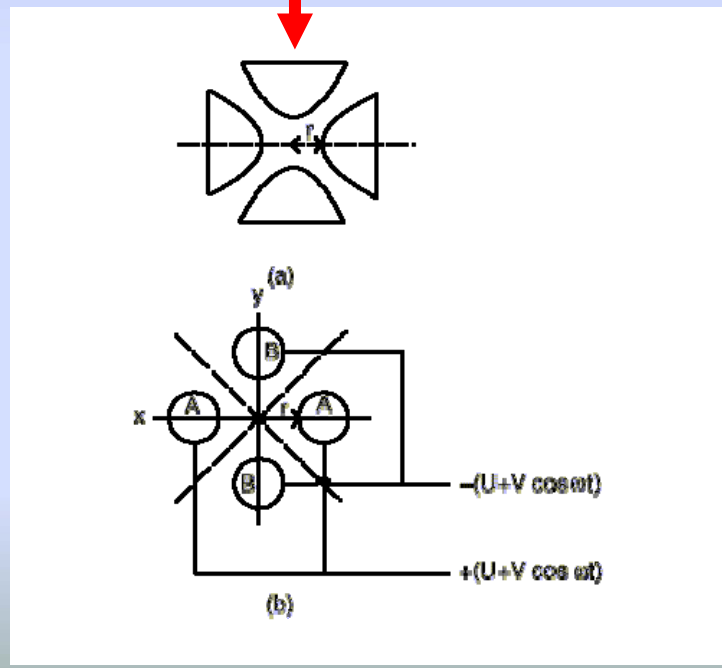
**collision cell**
**Mass Selection**

**Detector**
**Ion Collection**

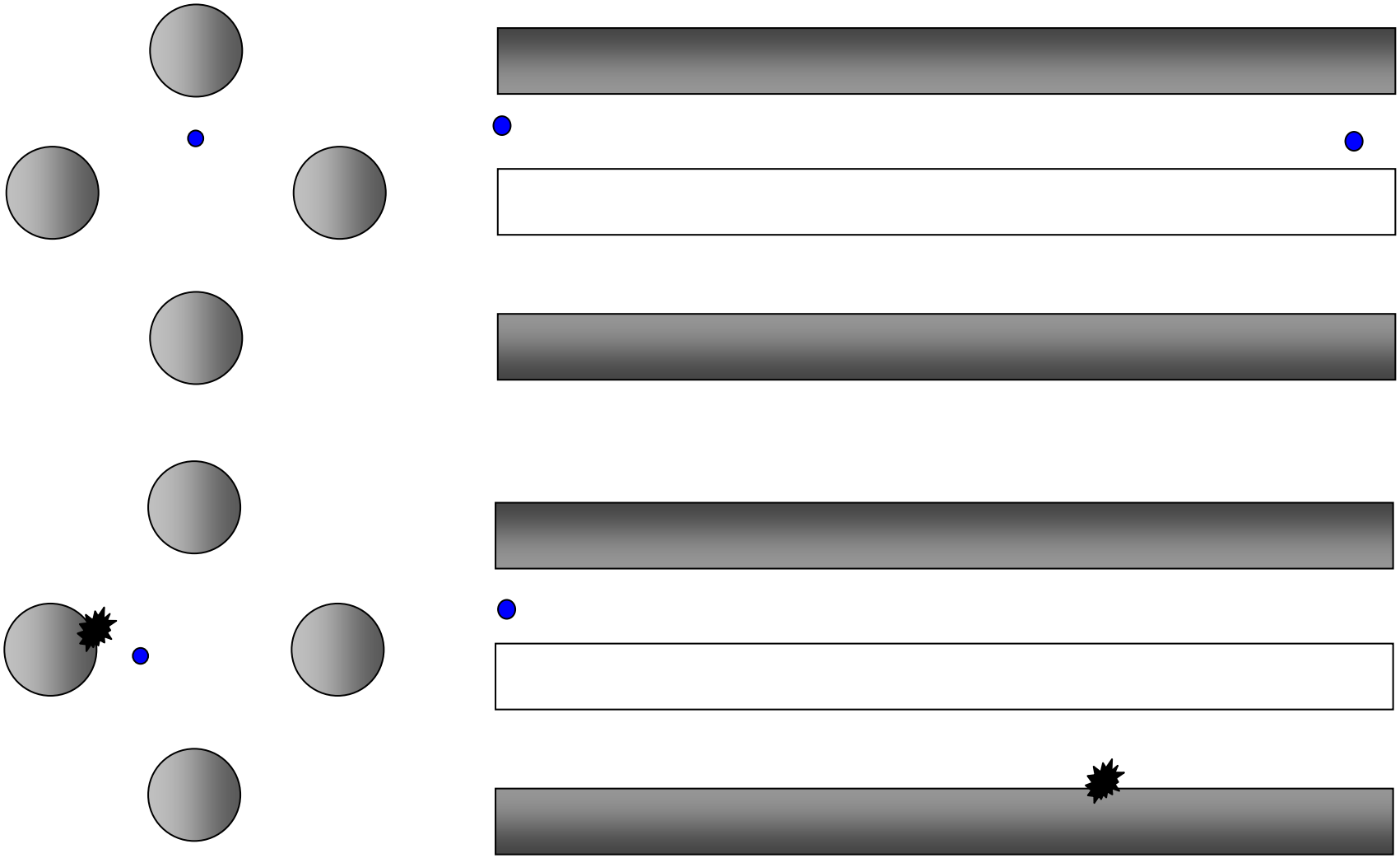
# Quadrupole Optics

- In a quadrupole mass spectrometer four (quad) parallel rods (poles) are arranged equidistantly from a central (imaginary) axis.
- Charged ions are injected along the central axis of the quadrupole assembly.
- Static and alternating (radio frequency) electric potentials are applied to opposite pairs of rods, creating a fluctuating electric field.

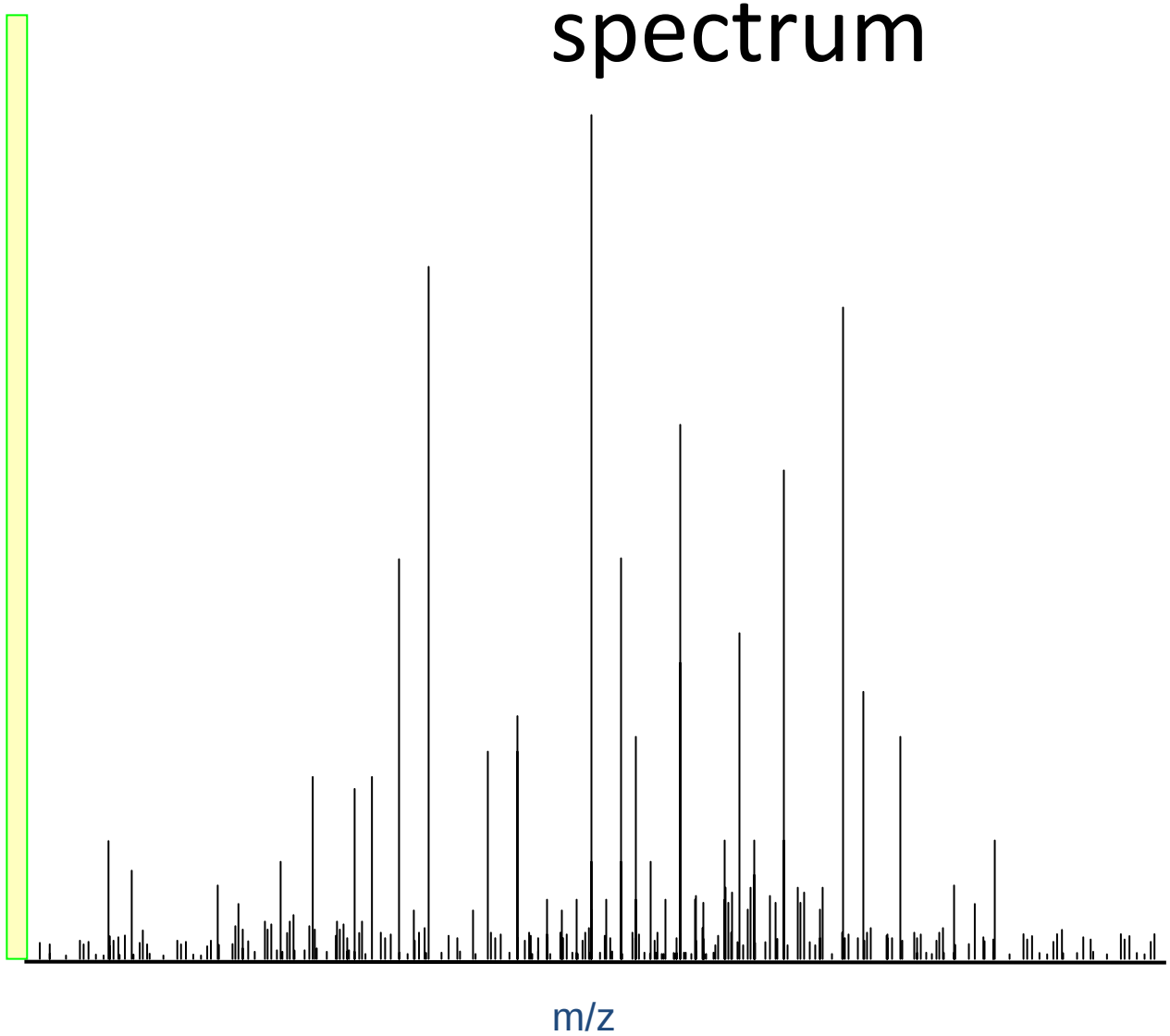




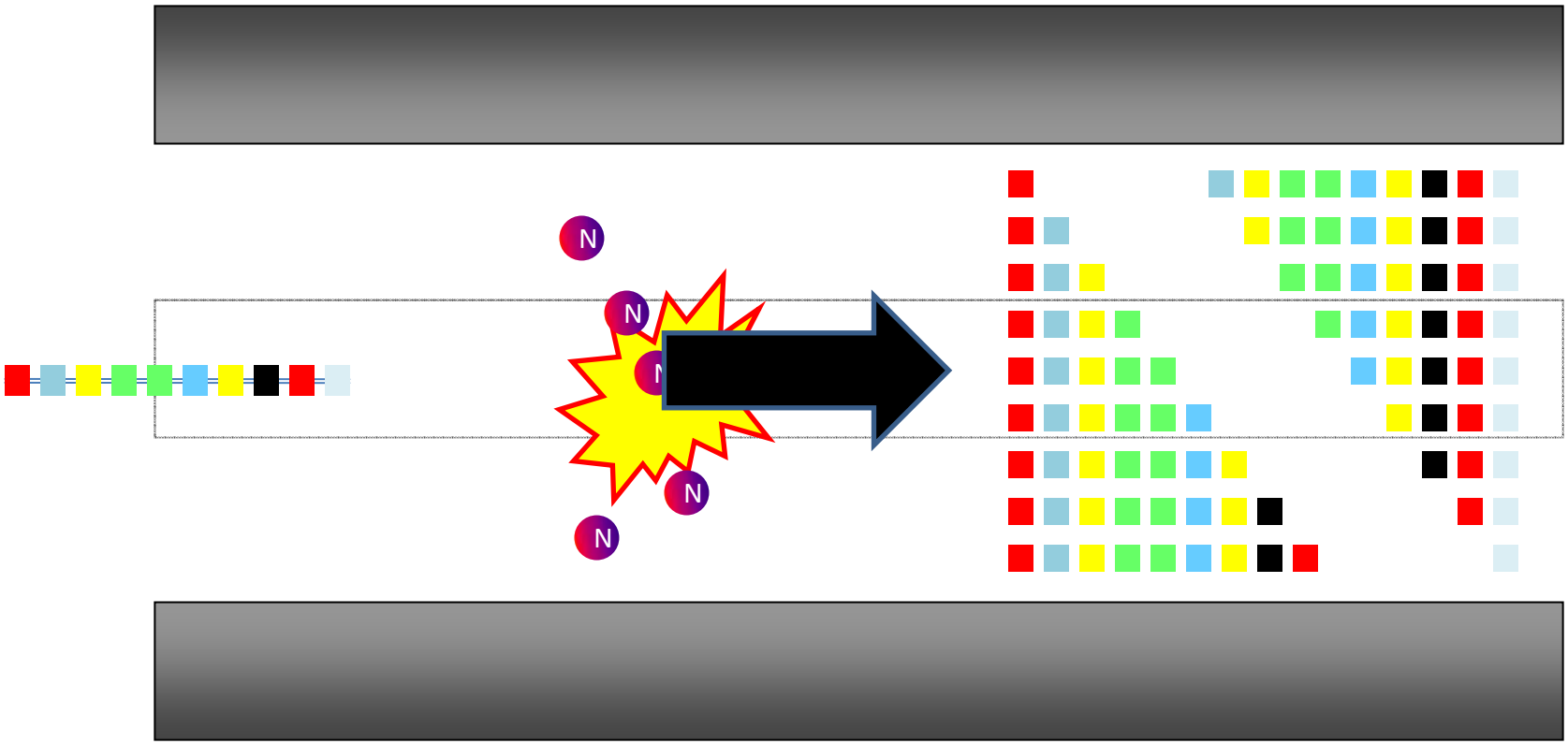
# A mass spectrometer can be used as a filter or “tuner”



# “Spinning the dial” generates a spectrum



# Mass Spectrometers identify peptides by fragmenting along the amide backbone



# Fragmentation occurs along the backbone revisited

+NSGDIVNLGSIAGR+

*b* *y*

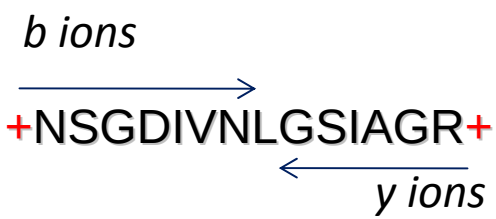
+N SGDIVNLGSIAGR+
+NS GDIVNLGSIAGR+
+NSG DIVNLGSIAGR+
+NSGD IVNLGSIAGR+
+NSGDIV NLGSIAGR+
+NSGDIVN LGSIAGR+
+NSGDIVNL GSIAGR+
+NSGDIVNLG SIAGR+
+NSGDIVNLGS IAGR+
+NSGDIVNLGSIA GR+
+NSGDIVNLGSIAG R+

# Y ions

+NSGDIV

NLGSIAGR+

NLGSIAGR



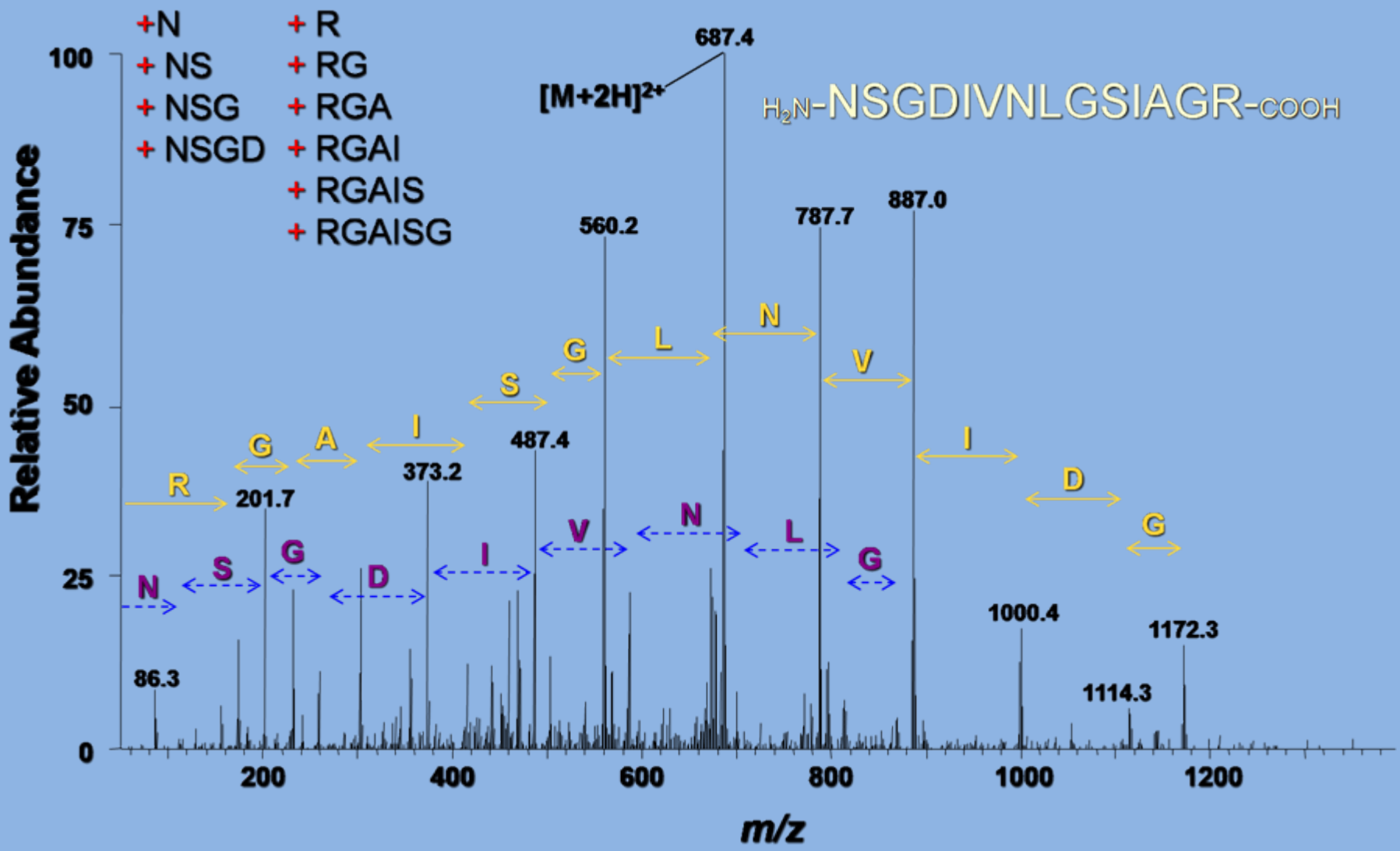

+N
+ NS
+ NSG
+ NSGD
+ NSGDI
+ NSGDIV
+ NSGDIVN
+ NSGDIVNL
+ NSGDIVNLG
+ NSGDIVNLGS
+ NSGDIVNLGSI
+ NSGDIVNLGSIA
+ NSGDIVNLGSIAG

*b ions*

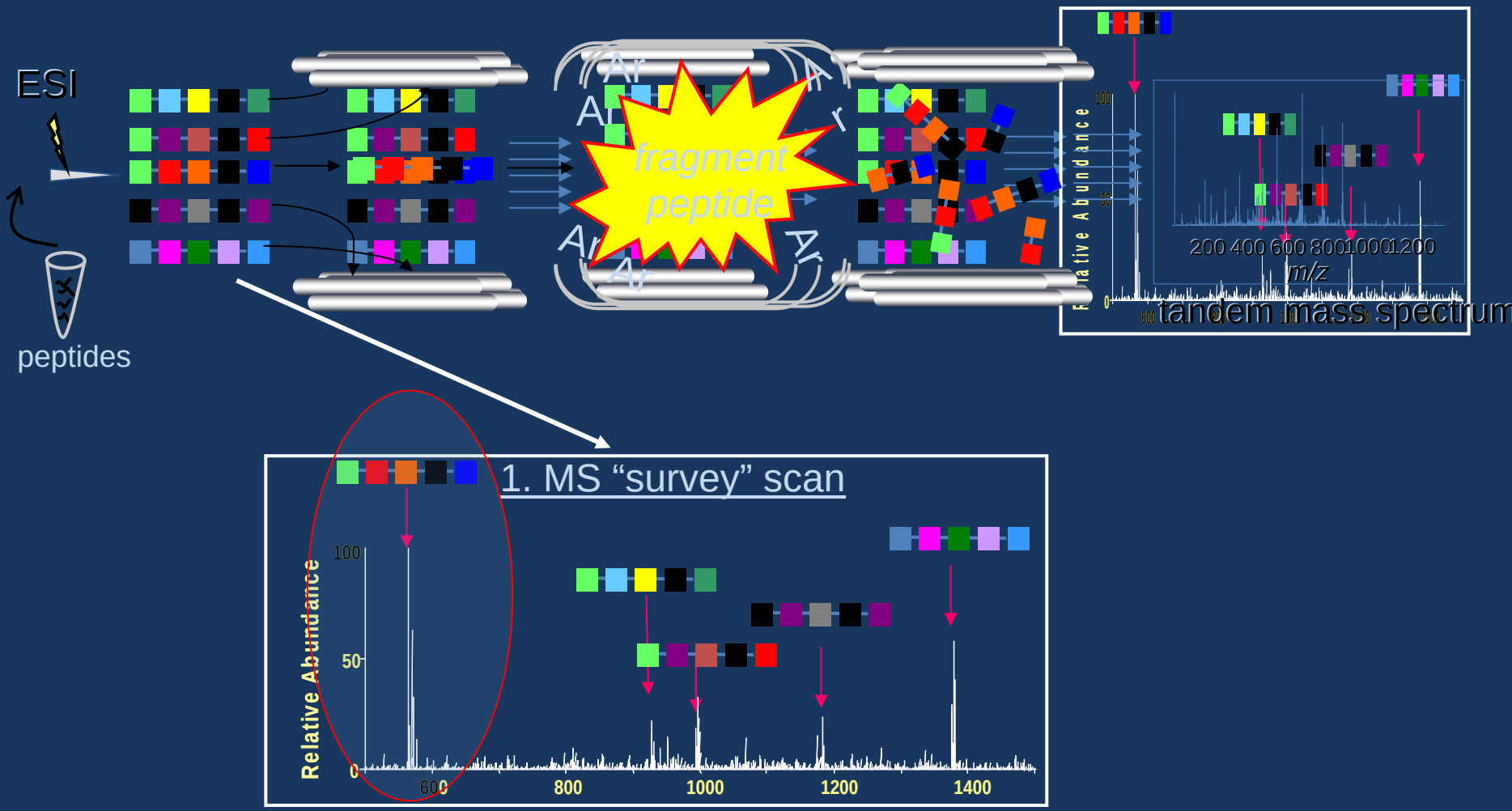
+ R
+ RG
+ RGA
+ RGAI
+ RGAIS
+ RGAISG
+ RGAISGL
+ RGAISGLN
+ RGAISGLNV
+ RGAISGLNVI
+ RGAISGLNVID
+ RGAISGLNVIDG
+ RGAISGLNVIDGS

*y ions*

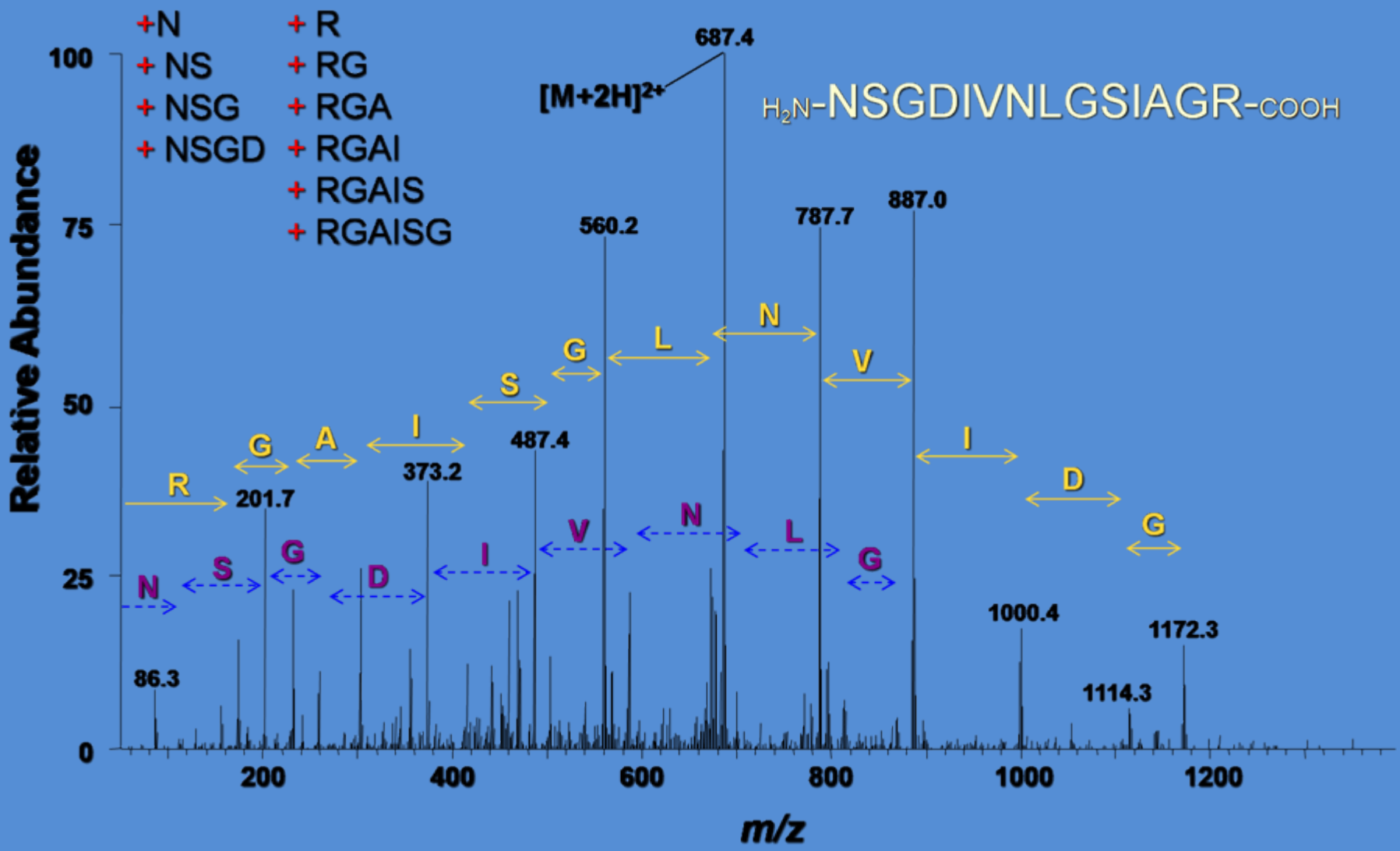
# The fragment “ladder” allows identification



# Tandem mass spectrometry: “Shotgun Proteomics”



# The fragment “ladder” allows identification



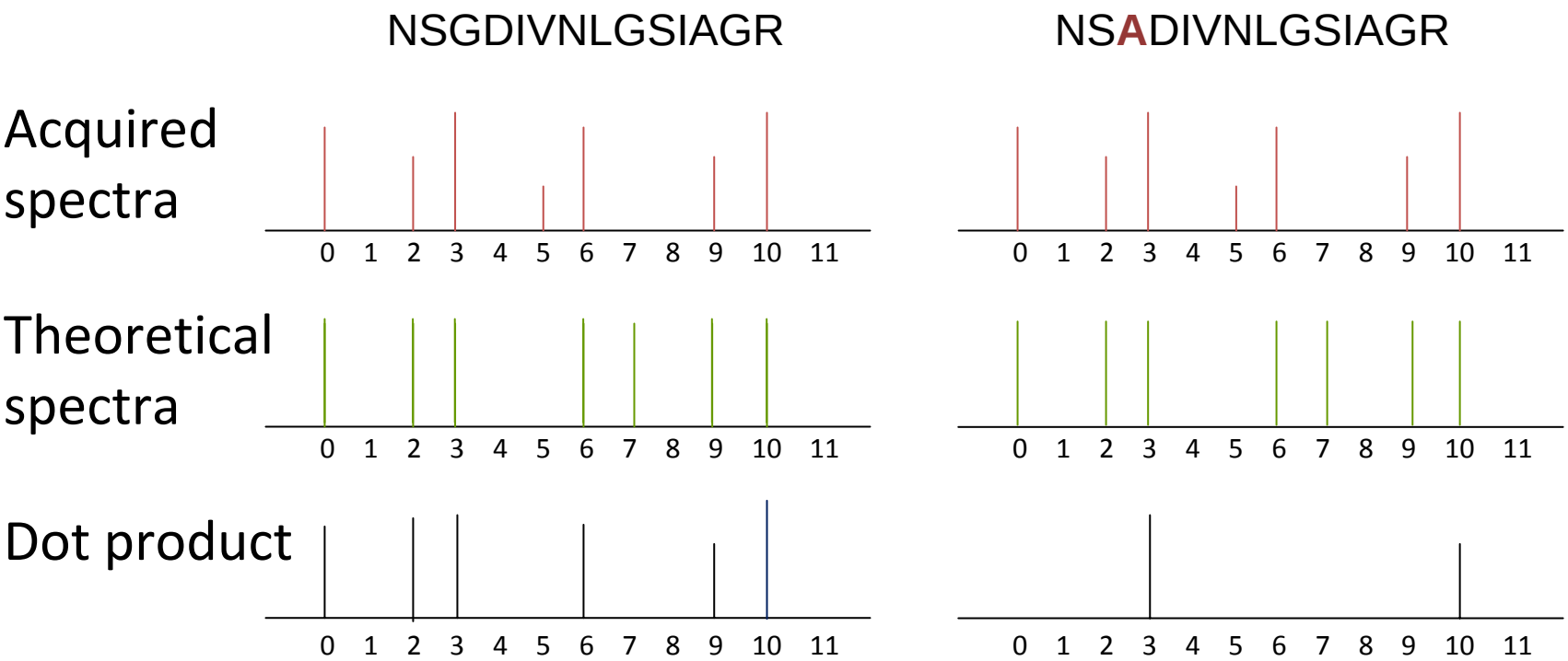
# Making an identification by database searching using SEQUEST

- SEQUEST is a search program that assigns a peptide sequence to a spectra by comparing it to virtual spectra from a protein database

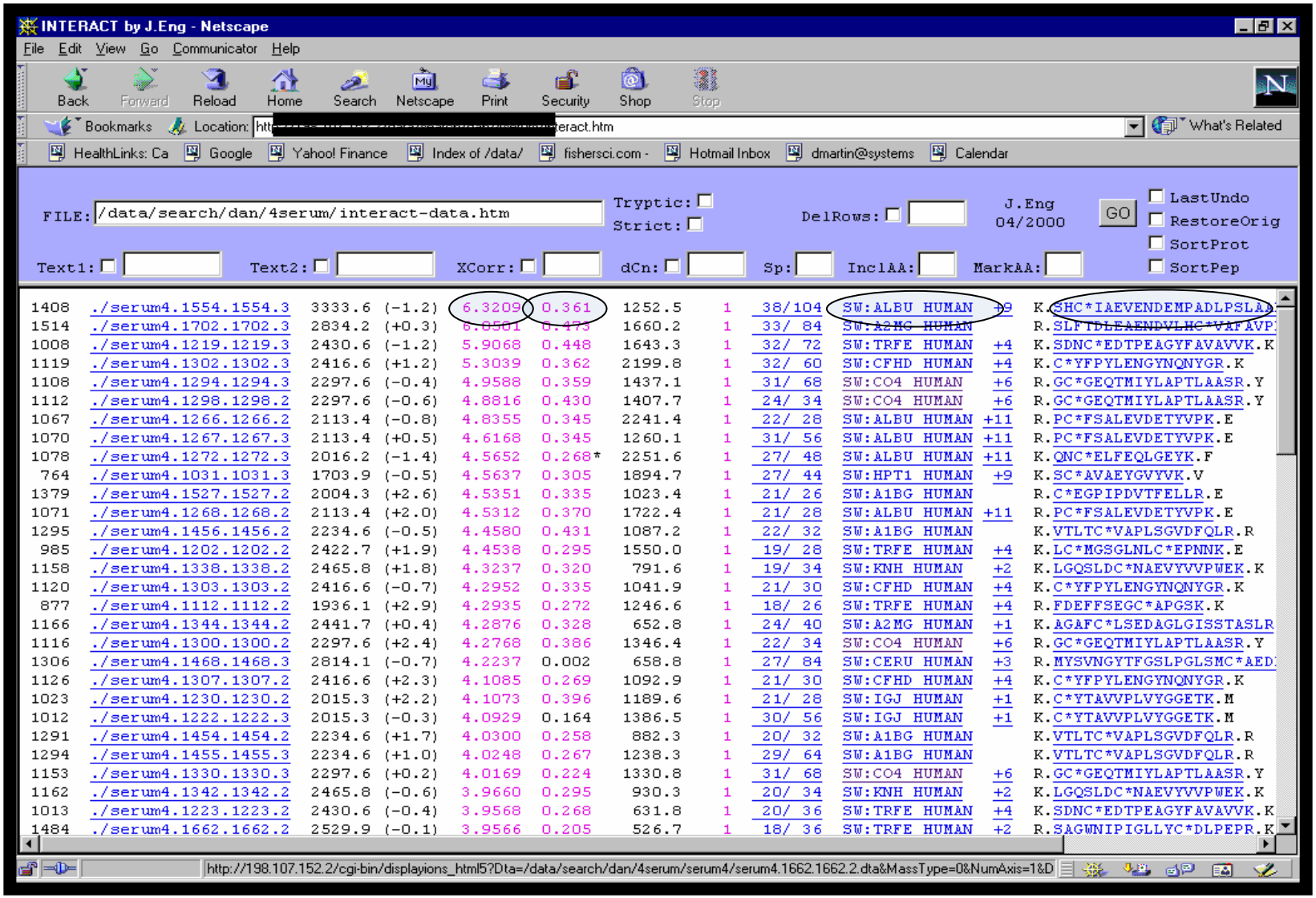
# SEQUEST Example

1. An MS/MS scan of m/z 750 and charge 2+➔ the molecular weight is 1500 Da
2. SEQUEST searches a protein database starting at the first amino acid to find all possible peptides that weight 1500 +/- 1.5
3. SEQUEST fragments each virtually and compares to the experimental spectra.
4. For a good spectra, ***one peptide stands beats out all others***

# Scoring a “match”

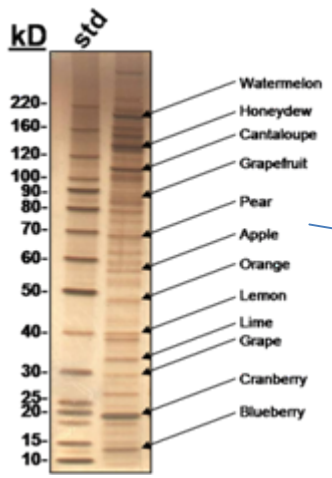


# SEQUEST output file



# How to think about spectrometry based proteomics

## In the end you get a list!




Analysis of an IP of the DNA polymerase II gives 400 proteins

Analysis of cultured yeast yields many thousands of proteins in a day's work

The same can be said for analysis mammalian cells

Be careful what you wish for

# Limitations:

- Dynamic Range $10^{3\text{-}4}$
- Complexity limits analysis
  - MS based proteomics is a sampling method
- Quantification is usually relative rather than absolute

# What does this mean?

Mass spectrometry based proteomics:

Is good at:

- In depth analysis of pure samples across 3-4 logs of concentration
  - Cells, organelles, IP's

Is not good at:

- Samples with dominant proteins and high dynamic ranges
  - Serum, plasma, csf, urine
  - *Tissue culture media with serum added*

# What can mass spectrometry based proteomics do for me?

# Biomarkers!

# Biomarkers

- Cellular markers that will tell us a story
  - Markers of desired tissues-identity
  - Markers of scaffold health/consistancy
  - Markers of host response

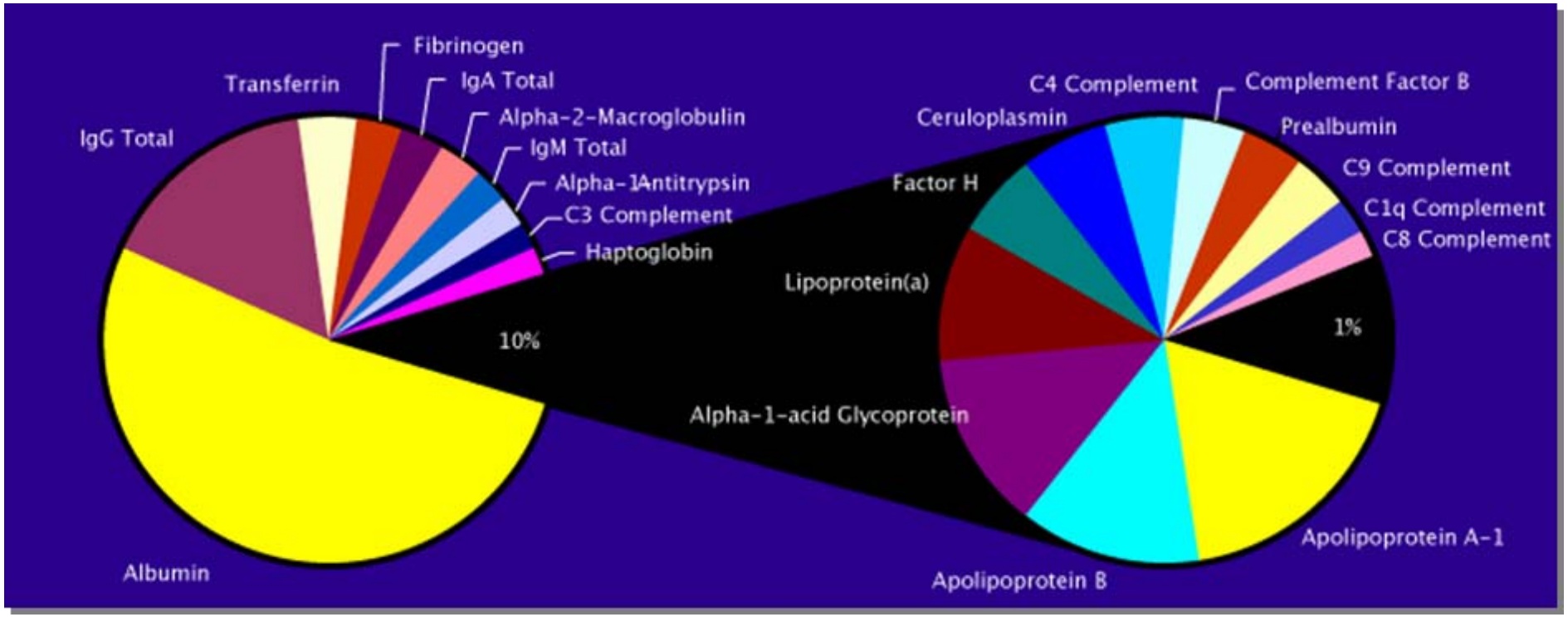
# Proteomics and Biomarkers

Expectations have been high

# Proteomics is being used to discover new biomarkers

- "Clinical" based discovery strategies
  - Paired clinical samples are probed to determine disease specific differences
  - Example: Compare serum, urine, CSF
- "Target based" validation strategies
  - Tissue or cells are studied to identify targets for later validation in clinical samples or model systems

# The Serum Challenge I: A Few Very Abundant Proteins




Serum albumin represents >50% total serum protein itself

22 most abundant serum proteins represent 99% total protein

**“Discovery” of low abundance biomarkers is a big challenge!**

# Analysis of Cells

Protein lists can be assembled-
(you got your wish)

- Many thousands will be seen
- All are potential markers
- This may satisfy some (based on the presentions yesterday)

- If I don’t know which proteins are the markers, how many can I follow through time?

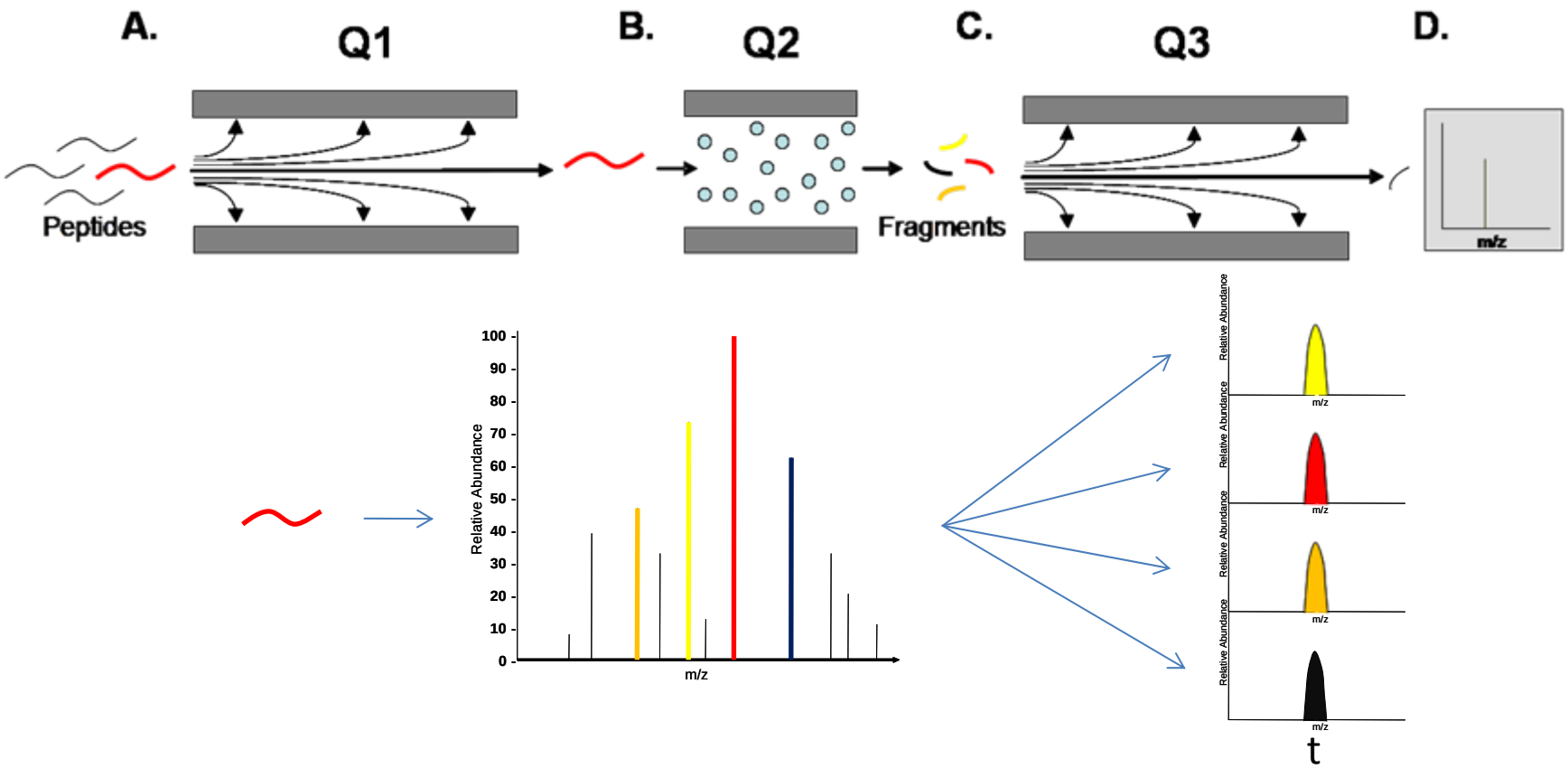
# Targeted Proteomics

# I have a candidate list

**Atlas based proteomics**

- The goal: generate and measure a minimal set of "peptide transitions" that completely and non-redundantly represent a proteome
- Uses a variant of mass spectrometry based proteomics called Multiple Reaction Monitoring (MRM)
  - Higher dynamic range
  - Higher sampling speed
  - It is still not magic

# Multiple Reaction Monitoring Mass Spectrometry (MRM-MS)

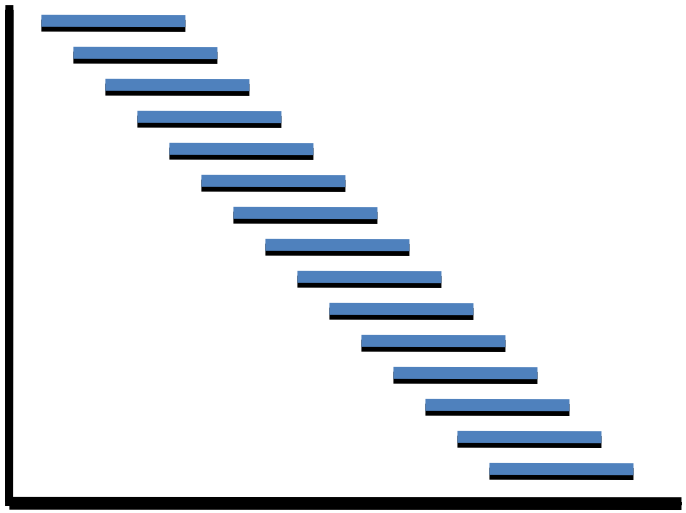



Each measurement “transition” can be made in 10 msec

A list of 100 can be cycled once per second

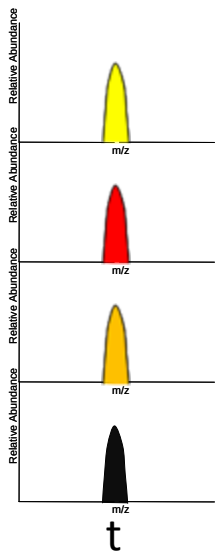
Elution over 20 seconds will have 20 points to generate a quantifiable peak

# Scheduled MRMs using a Waters Triple Quadrupole



time

- 32 segments over a 30 minute run
- 32 ions / seg
- **1,024** transitions

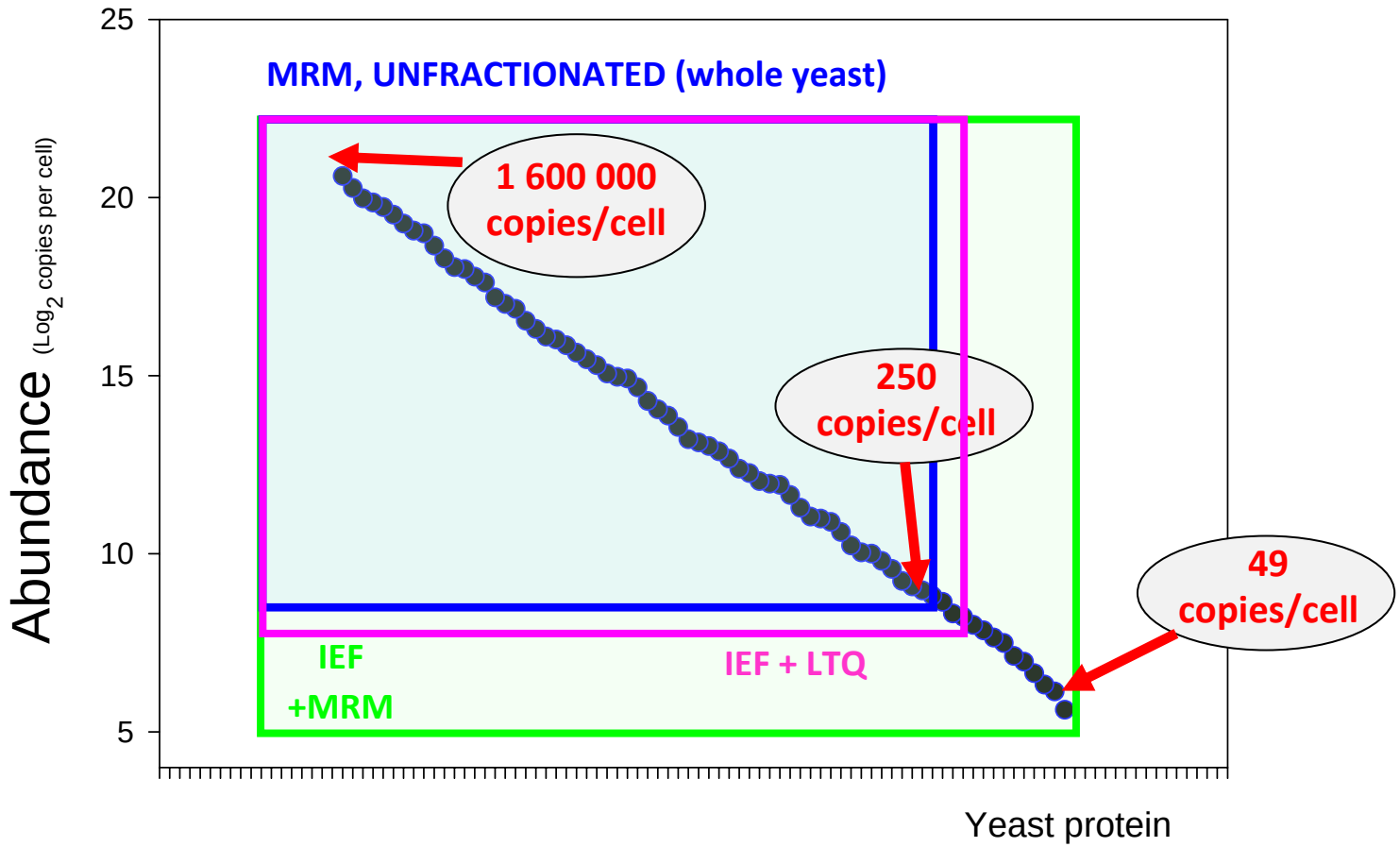



t

100 proteins per run at a coverage of three peptides per protein and three fragments per peptide.

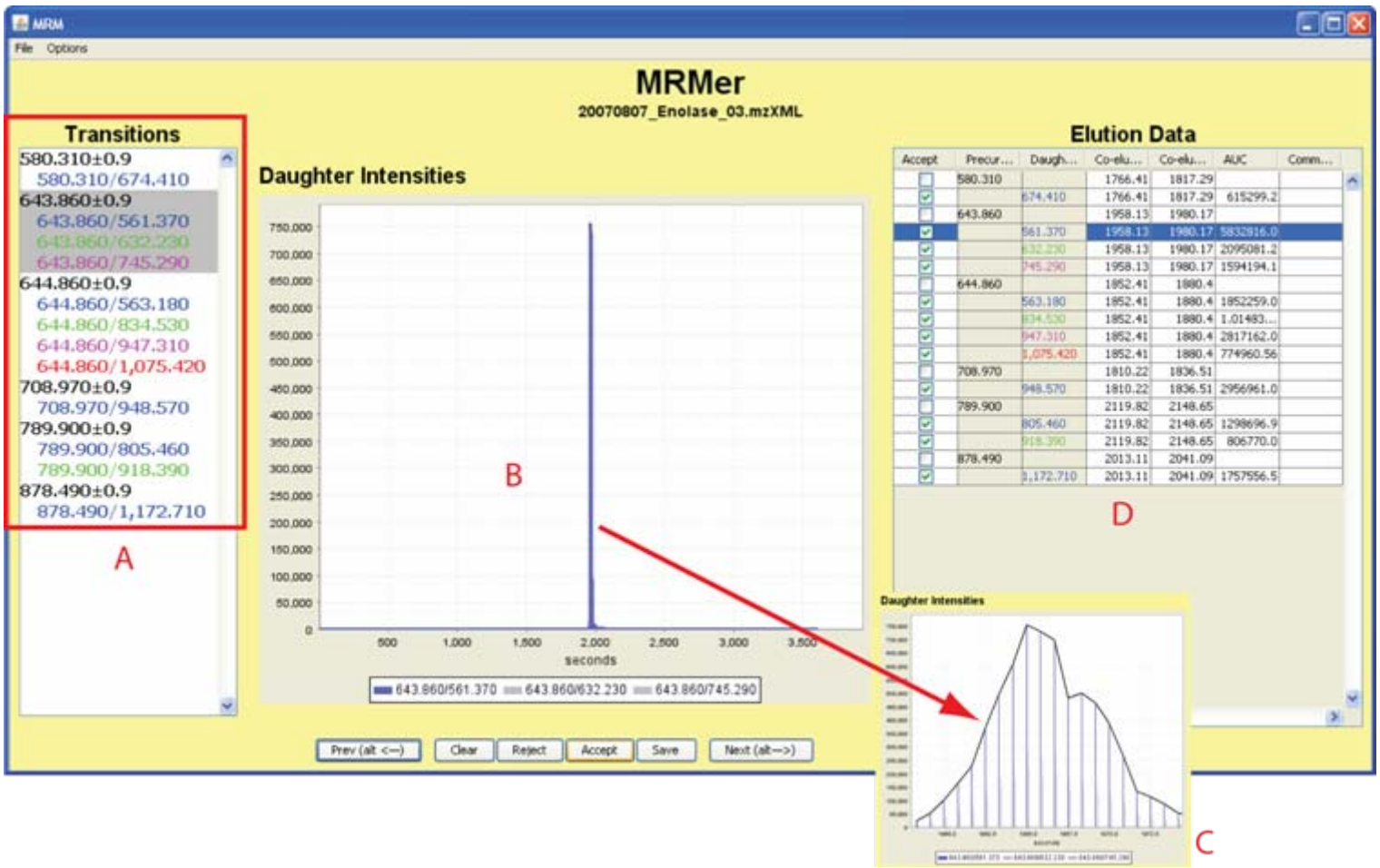
1 run/hr -> full coverage of 6000 proteins in 3 days

# Targeted Analysis of Yeast Proteins by MRM improves sensitivity




Picotti P. et al submitted

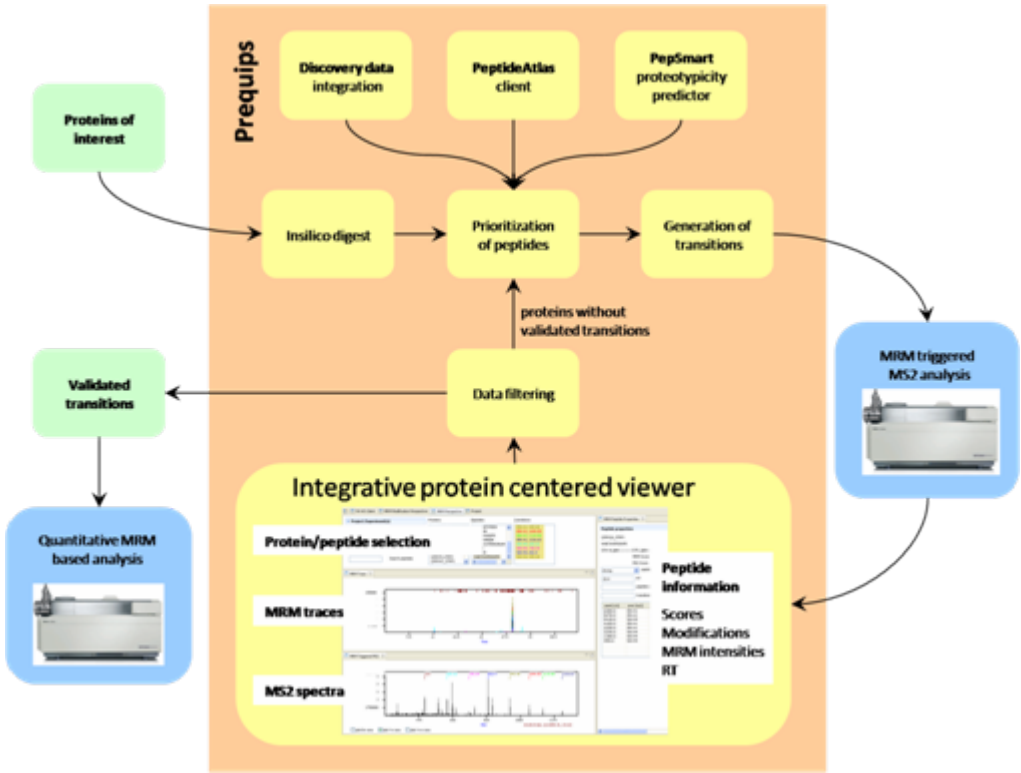
# We have developed software to deal with large MRM analysis



# PeptideAtlas: Different builds

| Build | # Exps | # MS Runs | Searched Spectra | ID P>0.9 | Distinct Peptides | Distinct Proteins |
|---|---|---|---|---|---|---|
| Human All | 90 | 1517 | 3.3 M | 334 k | 35,391 | 8000 |
| Drosophila | >100 | >1500 | ~10 M | 480k | 100,000 | > 10.000 |
| Human Plasma | 40 | 39,659 | >14 M | 660 k | 31,953 | ~3000 |
| Yeast | 46 | 1326 | 4.1 M | 536 k | 35 k | > 4000 |

INSTITUTE FOR Systems Biology



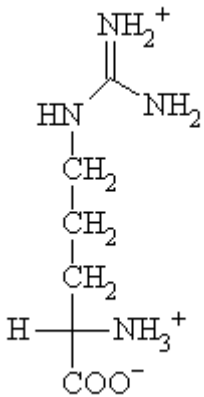

In the near future, we will take the atlases and build (and verify) peptides that represent the entire proteome to generate a “MRM transition atlas”

# Implications:

- Shotgun proteomics, with dependency on duty cycle and inherent dynamic range limits, may be superseded by targeted studies
- This may increase the ability to interrogate for a large number of targets in a biological sample
- This may help (but not solve) issues of measuring rare proteins in biofluids (serum, urine, CSF)

# Quantitative Implications

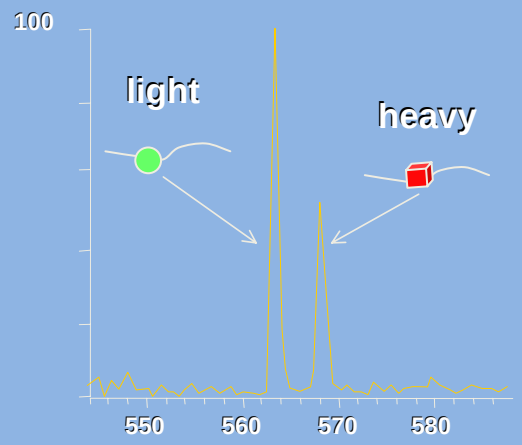
NSGDIVNLGSIAGR

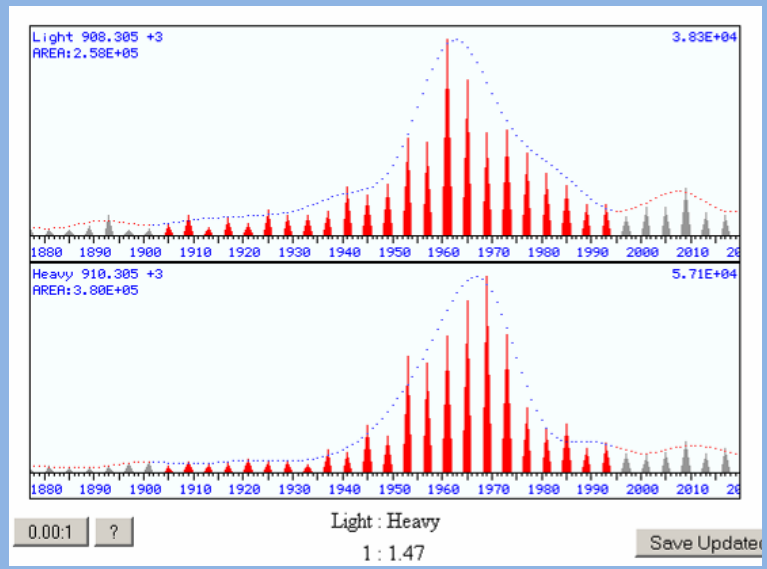


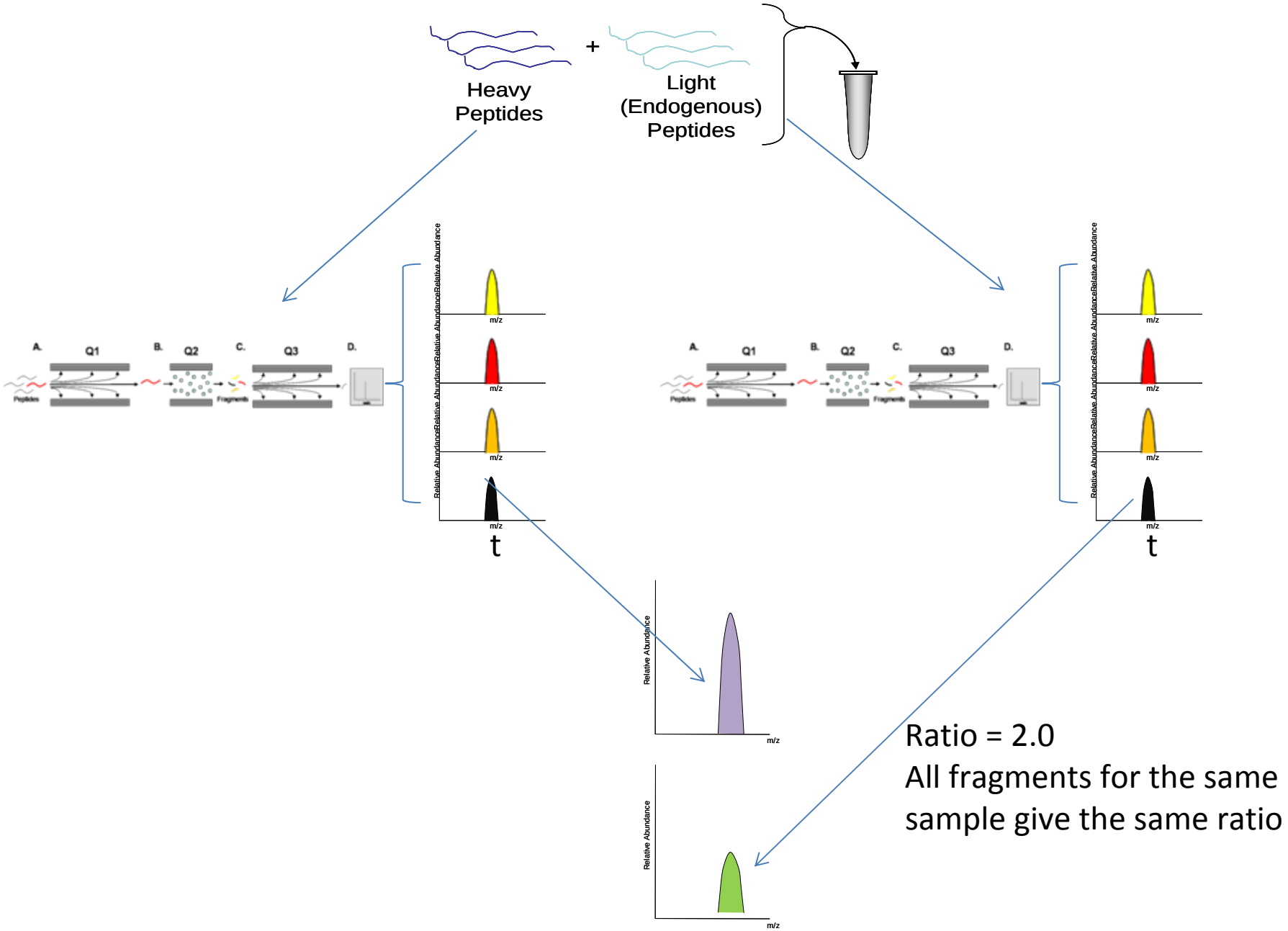

$^{13}C_6{}^{15}N_4$

# Isotopic modification strategies

- Mass spectrometers measure mass-hence isotopically different peptides can be compared.
- Single ion chromatograms for each isotope can be compared to determine relative quantities of each.



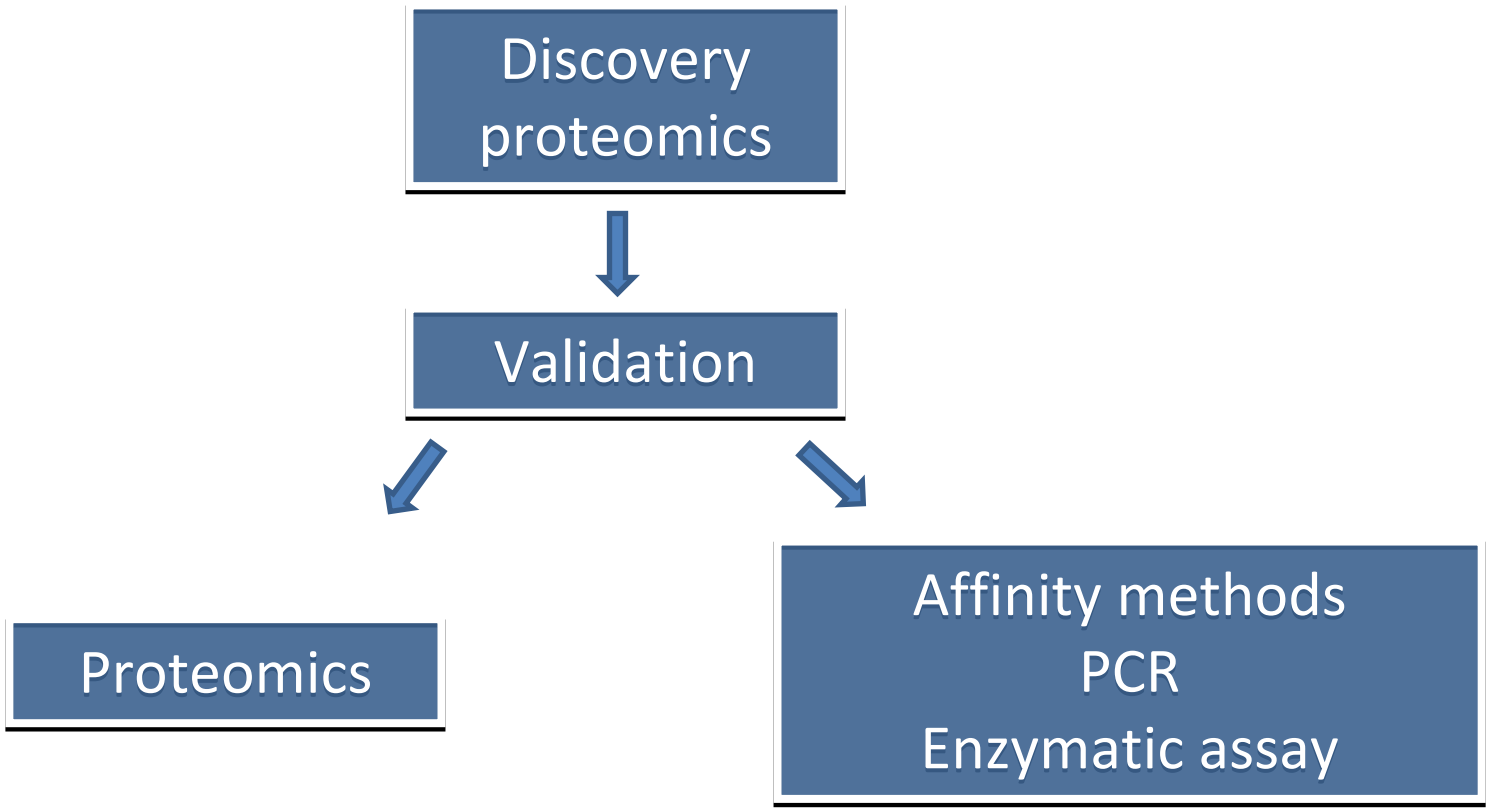




Heavy Peptides
+
Light (Endogenous) Peptides
A.
Q1
B.
Q2
C.
Q3
D.
Peptides
Fragments
Relative Abundance
m/z
t
Ratio = 2.0
All fragments for the same sample give the same ratio

# How will this play out?

- There may come a day when every one of the unique peptide targets can be paired with an isotopically heavy peptide to determine absolute concentration.

# *Mass Spectrometry-Based Proteomic Applications in Cell/Scaffold Products*



# *Mass Spectrometry-Based Proteomic Applications in Cell/Scaffold Products*

- Discover proteins for in-vitro assays during product development and during production
  - Monitor cells
- Monitor processes at the protein level with targeted mass spectrometry (MRM) or multi-dimensional affinity reagent panel (antibody chip)
  - Monitor growth media during production with targeted analysis/affinity reagent
- Monitor for target protein presence after implantation (likely via affinity reagents)

# Proteomics is just a tool

- You can’t screw a lightbulb in with a hammer
- There were a lot of if’s yesterday
  - If I could measure xxxxxx