

# ***Intro to Proteomics/Proteomics Workflow***

# 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.

# Aren't DNA Arrays Enough?

**NO !**

- Proteins are the primary effectors in cellular function.
- RNA levels may not correlate well with protein expression due to translational regulation, protein degradation, splicing, nuclear transport, post-translational modification.
- Proteins are the target of most pharmacologic therapies.

# The genome and the proteome: a comparison

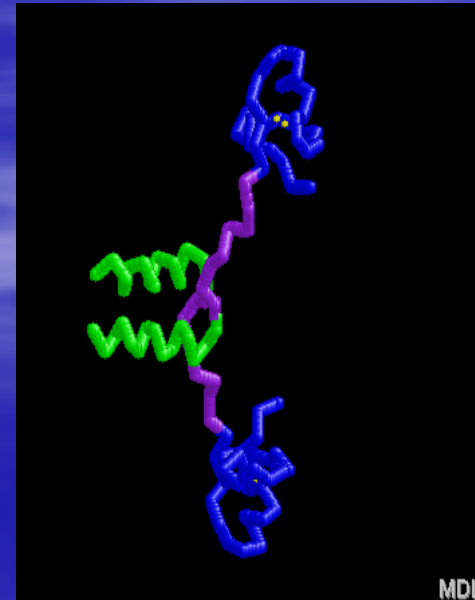
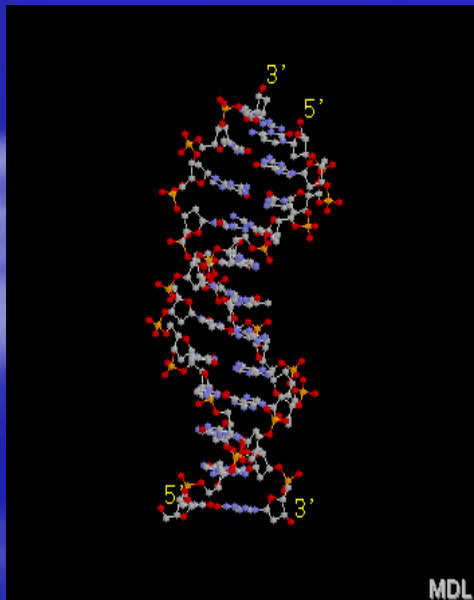
---

## Genome

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

## Proteome

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





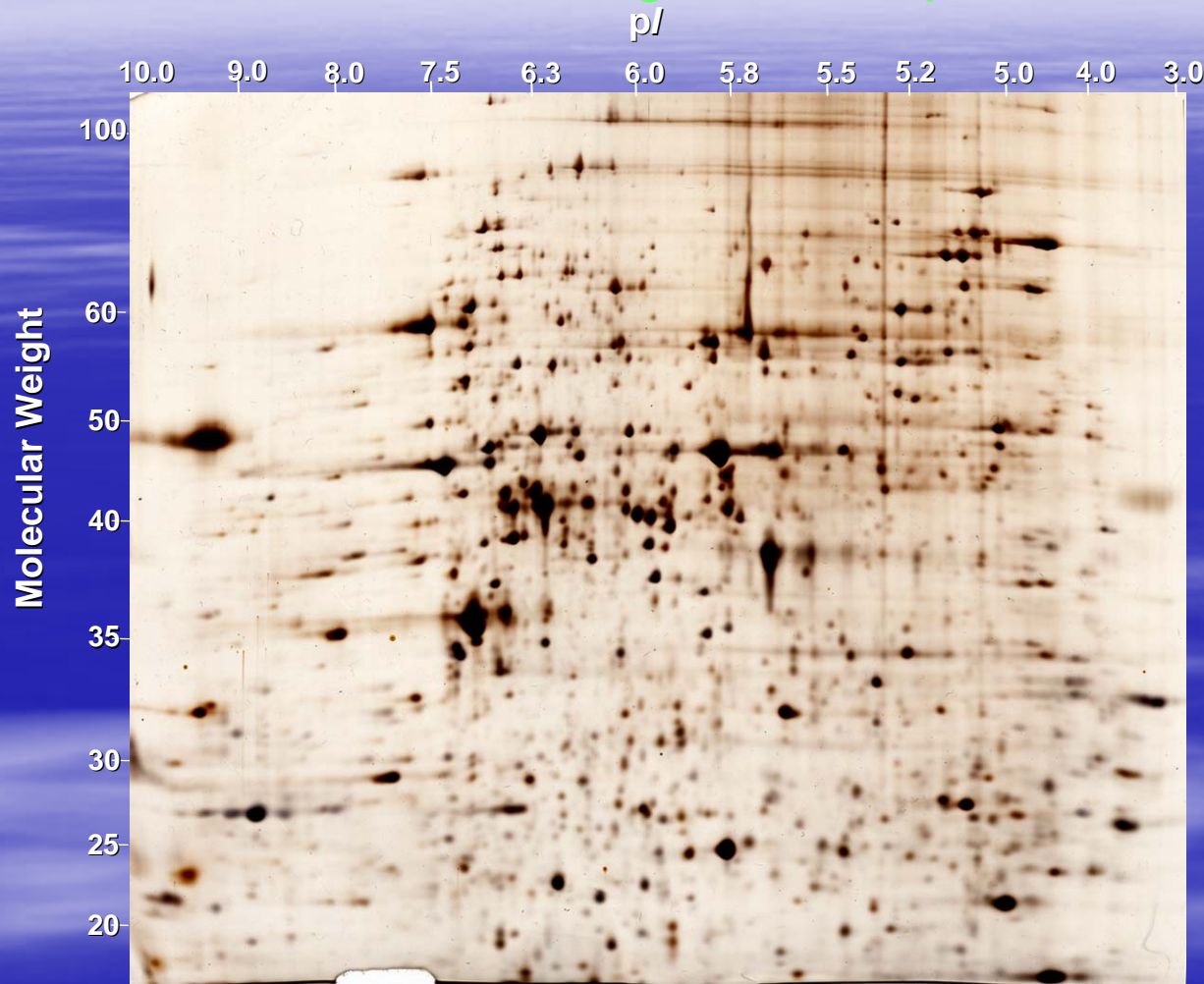
# Proteomics technologies and methods

---

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

# Proteome analysis: historical perspective

## Two-dimensional gel electrophoresis



Gygi, et. al. 1999, *Molecular and Cellular Biology* 19:1720

Compare Multiple 2D Gels

1949--Linus Pauling describes the difference in mobility between normal and sickle hemoglobin.

Large scale purification

Edman degradation

Clone protein

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

---

### What it is:

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

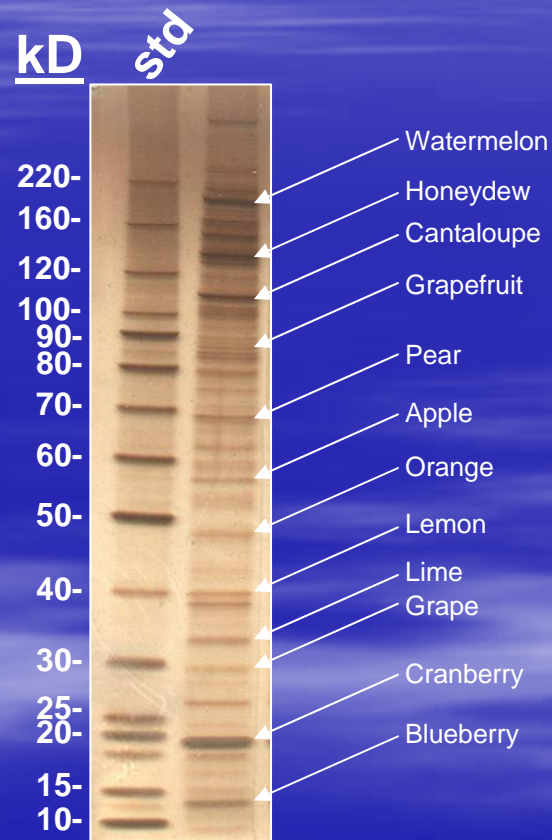
### What it is not:

Magic

Able to give all the answers

Simple (relative to running a gel)

# How to think about spectrometry based proteomics



Dynamic  
Range =  $10^3$

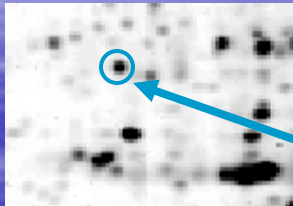


# What can mass spectrometry tell me?

---

- proteins in mixtures
- quantitative analysis of protein expression
- post-translational modifications:
  - phosphorylation
  - N- or C- terminal modifications
  - glycosylation
- protein interactions

## 2DE based Proteomics--Today



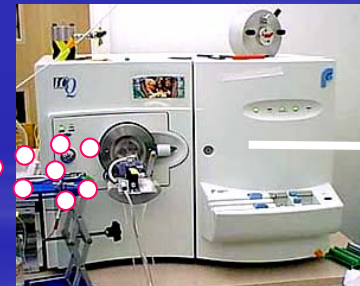
Compare gels to find spots of interest

Cut out the spot of interest for in-gel digest

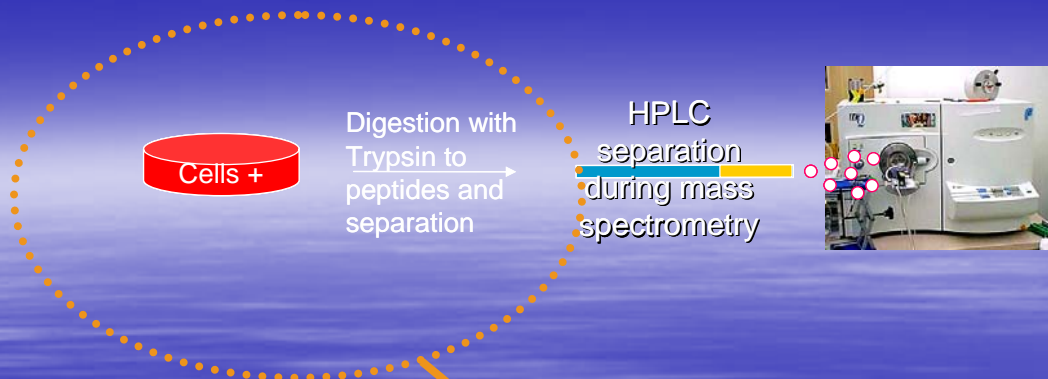
Identify peptides by mass spectrometry

Drawbacks: cannot identify low abundance proteins and some insoluble proteins. Time and labor intensive.

## Mass spectrometry identifies proteins



| Gene      | Protein Name   | Avx/A | S.D. | Unique | Cell |
|-----------|--|-------|------|--------|------|
| TNFRSF10  | TNFR receptor superfamily member 10B precursor                 | 17.17 | 4.58 | 6      | M    |
| TRAF3     | TRAF3 precursor  | 17.02 | 4.58 | 6      | M    |
| TRAF6     | TRAF6 precursor  | 17.02 | 4.58 | 6      | M    |
| TRAF3     | Transferrin receptor protein 1                                 | 7.81  | 2.23 | M      |      |
| PLXK      | Transferrin receptor protein 1 precursor                       | 7.86  | 2.68 | M      |      |
| CAR3P     | Transmembrane protein 30A binding protein 3 precursor          | 6.53  | 1.62 | M      |      |
| LILR4     | Low-density lipoprotein receptor precursor                     | 5.65  | 1.64 | M      |      |
| TNFRSF12  | TNFR superfamily member 12 precursor                           | 5.63  | 1.70 | M      |      |
| TRAF5     | TRAF5  | 5.43  | 1.10 | M      |      |
| SLIT1IN   | SLIT1-ROBO1-15A unannotated protein                            | 3.99  | 1.93 | 1      |      |
| PLXK2     | Transferrin receptor 2 precursor                               | 3.99  | 1.93 | 1      |      |
| JAK1      | JAK1 precursor   | 4.92  | 1.43 | M      |      |
| ALP2P     | Amorphous lipid-like protein 2 precursor                       | 4.41  | 2.08 | M      |      |
| CD44V1    | CD44 variant protein 1 precursor                               | 4.18  | 2.08 | M      |      |
| LRR43     | Membrane phosphoprotein LRR43                                  | 3.59  | 1.39 | M      |      |
| NCX3      | Protein tyrosine phosphatase precursor                         | 3.40  | 0.66 | M      |      |
| NCX3P     | Protein tyrosine phosphatase precursor A precursor             | 3.40  | 0.66 | M      |      |
| LCAM      | Ligand precursor   | 3.24  | 0.33 | M      |      |
| ALX3P     | ALX3 precursor   | 3.24  | 0.33 | M      |      |
| POW128S   | Protoporphyrin oxidoreductase precursor                        | 3.13  | 1.23 | M      |      |
| POW128S   | Protoporphyrin oxidoreductase precursor                        | 2.97  | 1.41 | M      |      |
| ILKAP     | ILKAP precursor  | 2.58  | 1.10 | M      |      |
| ILKAP     | Interleukin-1 receptor alpha chain precursor                   | 2.84  | 0.72 | M      |      |
| CD180     | Protein C 10B precursor  | 2.86  | 1.10 | M      |      |
| IL14      | Interleukin-14 precursor                                       | 2.86  | 1.10 | M      |      |
| PLAB1     | Prostate differentiation factor                                | 2.47  | 0.34 | M      |      |
| NCX1      | Protein tyrosine phosphatase precursor                         | 2.38  | 0.66 | M      |      |
| NCX1P     | Protein tyrosine phosphatase precursor A precursor             | 2.38  | 0.66 | M      |      |
| BAM       | Brain 2-methylglutamate precursor                              | 2.34  | 0.48 | M      |      |
| NCX1P     | Protein tyrosine phosphatase precursor A precursor             | 2.30  | 0.66 | M      |      |
| NG2       | Neogenin precursor   | 2.27  | 0.53 | M      |      |
| PCAP      | Procalcitonin polypeptide precursor, contains Saposin A        | 2.26  | 0.35 | 1      |      |
| PCAP      | Procalcitonin polypeptide precursor                            | 2.26  | 0.35 | 1      |      |
| ATOX1     | Protein wavycoat protein ATOX1                                 | 2.21  | 0.22 | M      |      |
| NCAM10    | NCAM10 precursor   | 2.20  | 0.46 | M      |      |
| NCAM10    | Neural cell adhesion molecule in anionic-6-phosphate precursor | 2.20  | 0.46 | M      |      |
| TNFRSF10P | TNFR superfamily member 10B precursor                          | 2.19  | 0.10 | M      |      |
| NCX1P     | Protein tyrosine phosphatase precursor A precursor             | 2.15  | 0.66 | M      |      |
| SEMMA4    | Semaphorin 4A precursor  | 2.15  | 0.48 | M      |      |
| LDLR      | Low-density lipoprotein receptor precursor                     | 2.05  | 0.95 | M      |      |



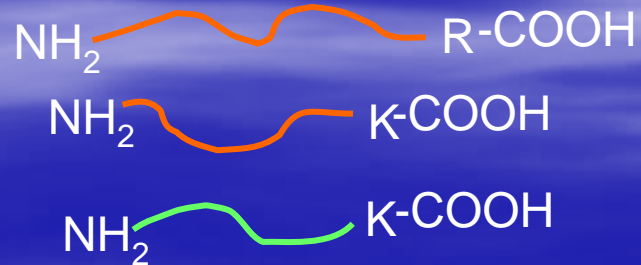
Protein, MW = 10,000 +



digest into peptides



Peptides,  
MW < 4,000



Trypsin

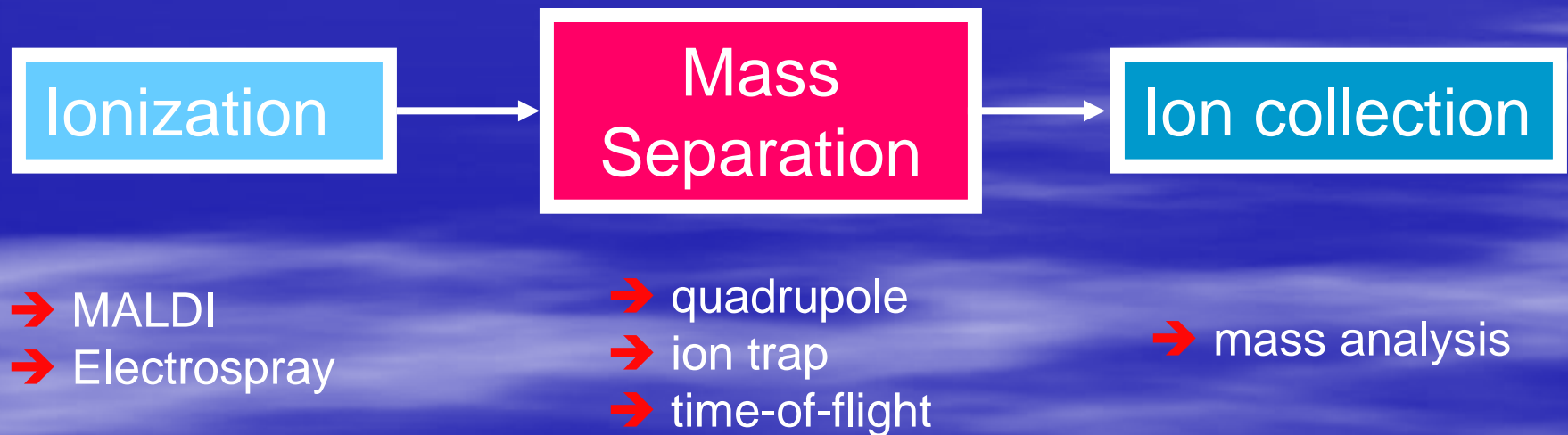
N-THK.NCPHIVVGTPGR.IPD-C

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



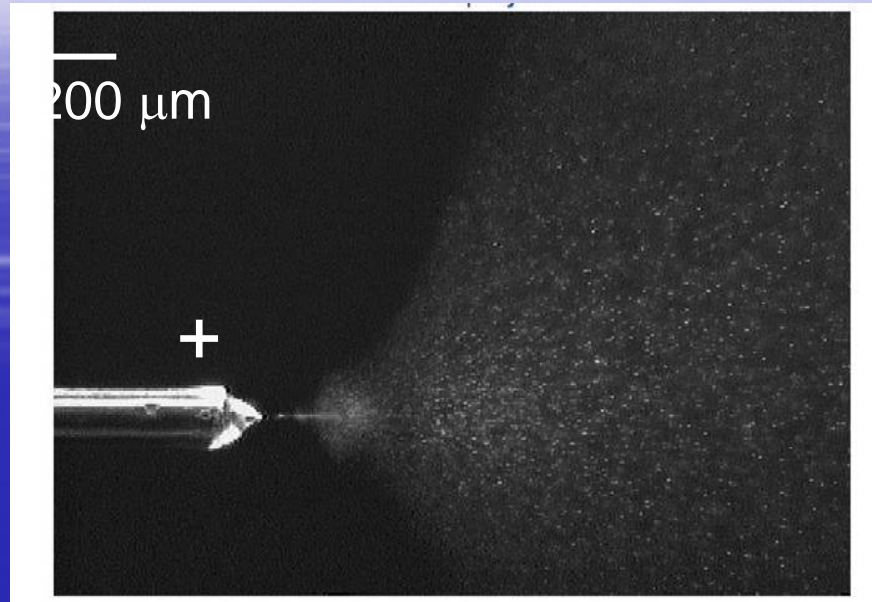
# Mass Spectrometry Primer

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



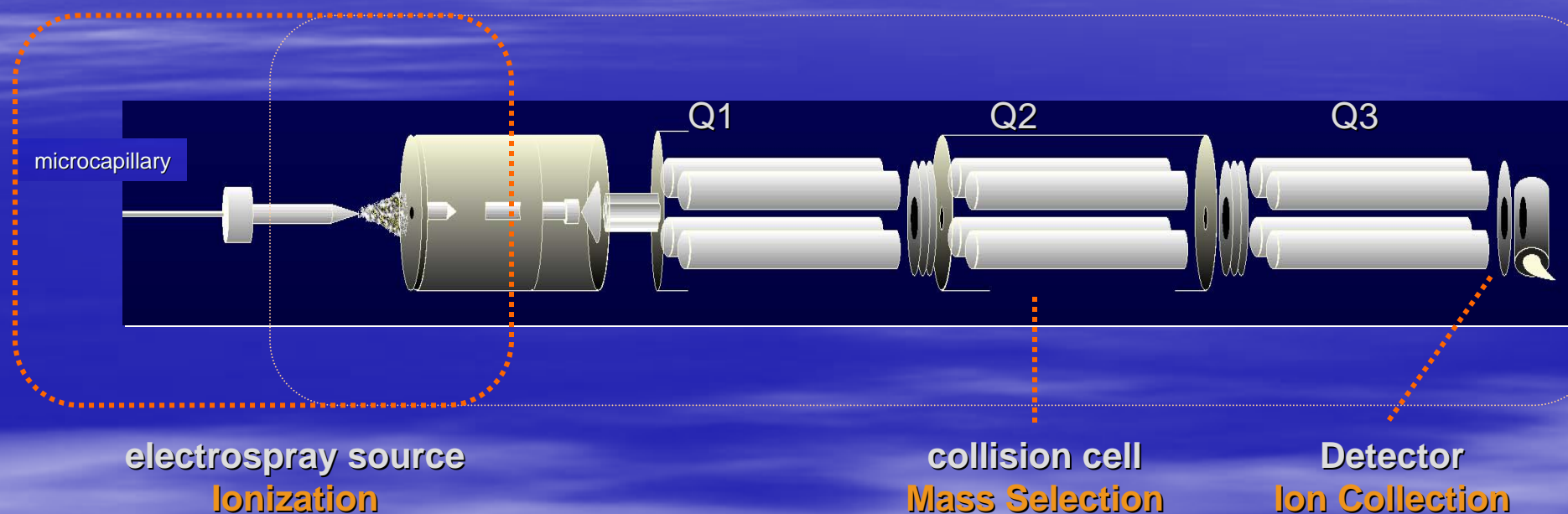
# Electrospray ionization (ESI)

---



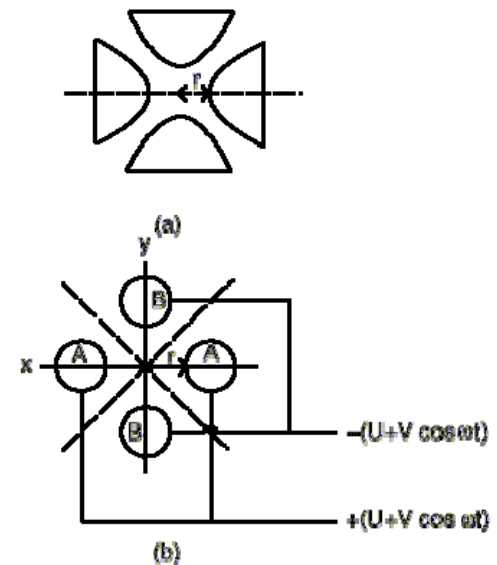
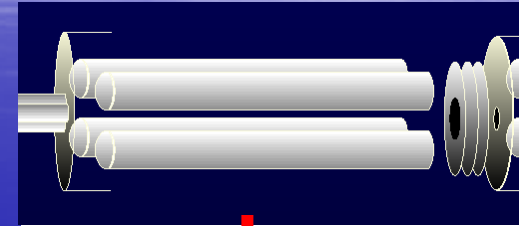
High voltage placed on a fused silica column causes a spray of charged droplets which evaporate leaving charged peptides

# Triple Quadrupole Mass Spectrometer

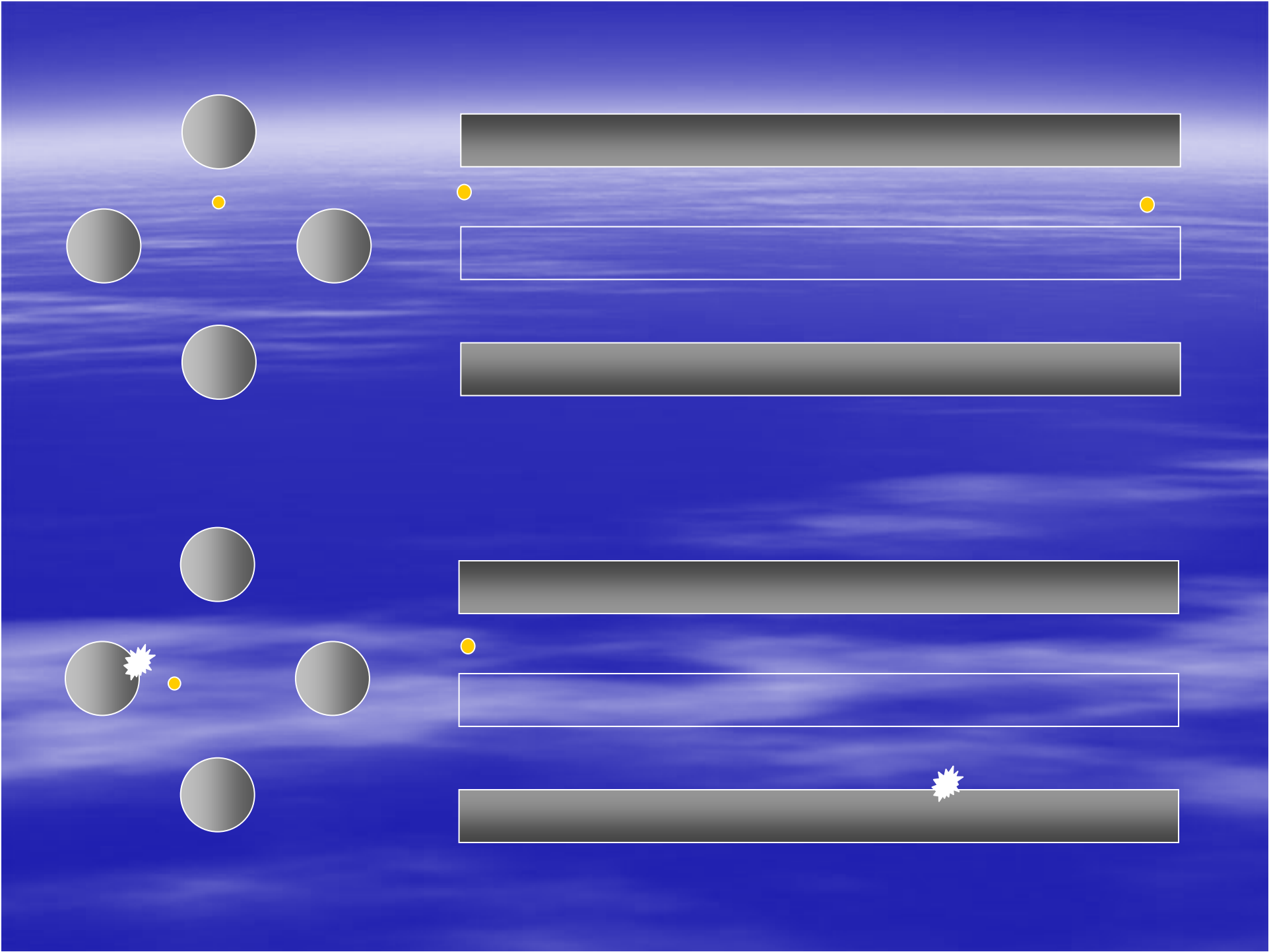


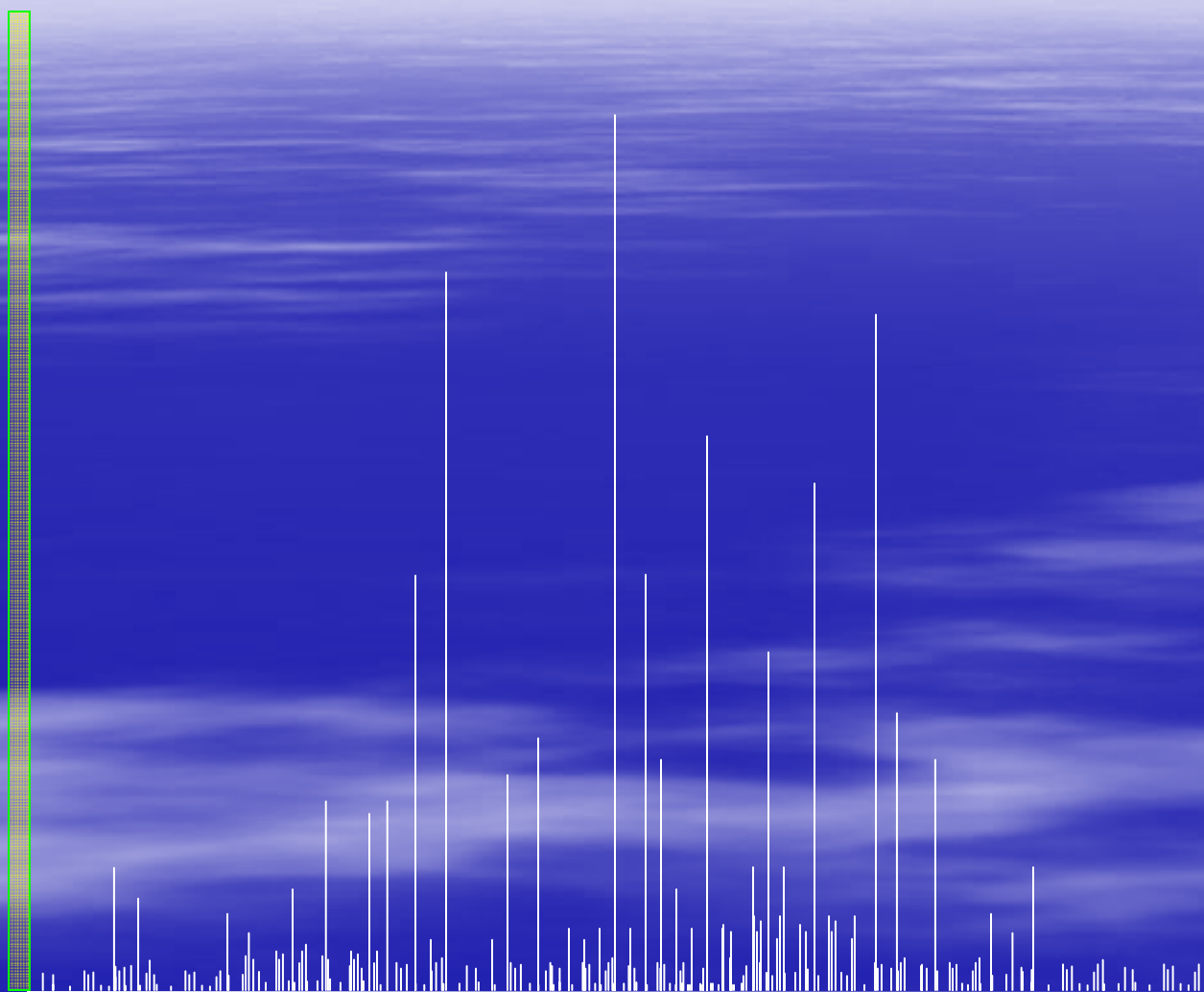
# Quadrupole Optics

- 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.



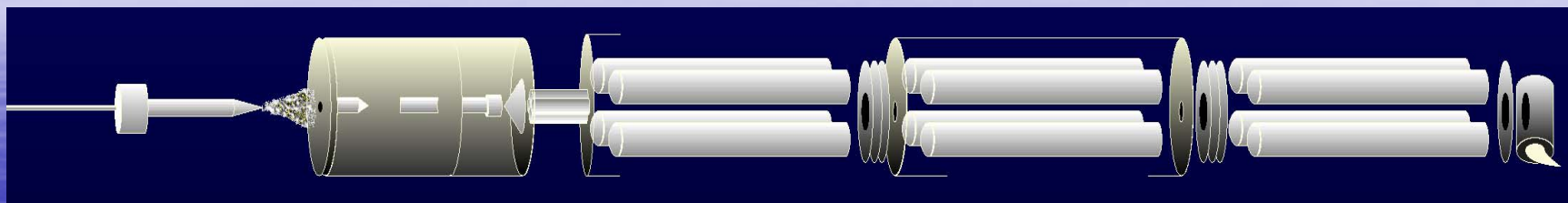




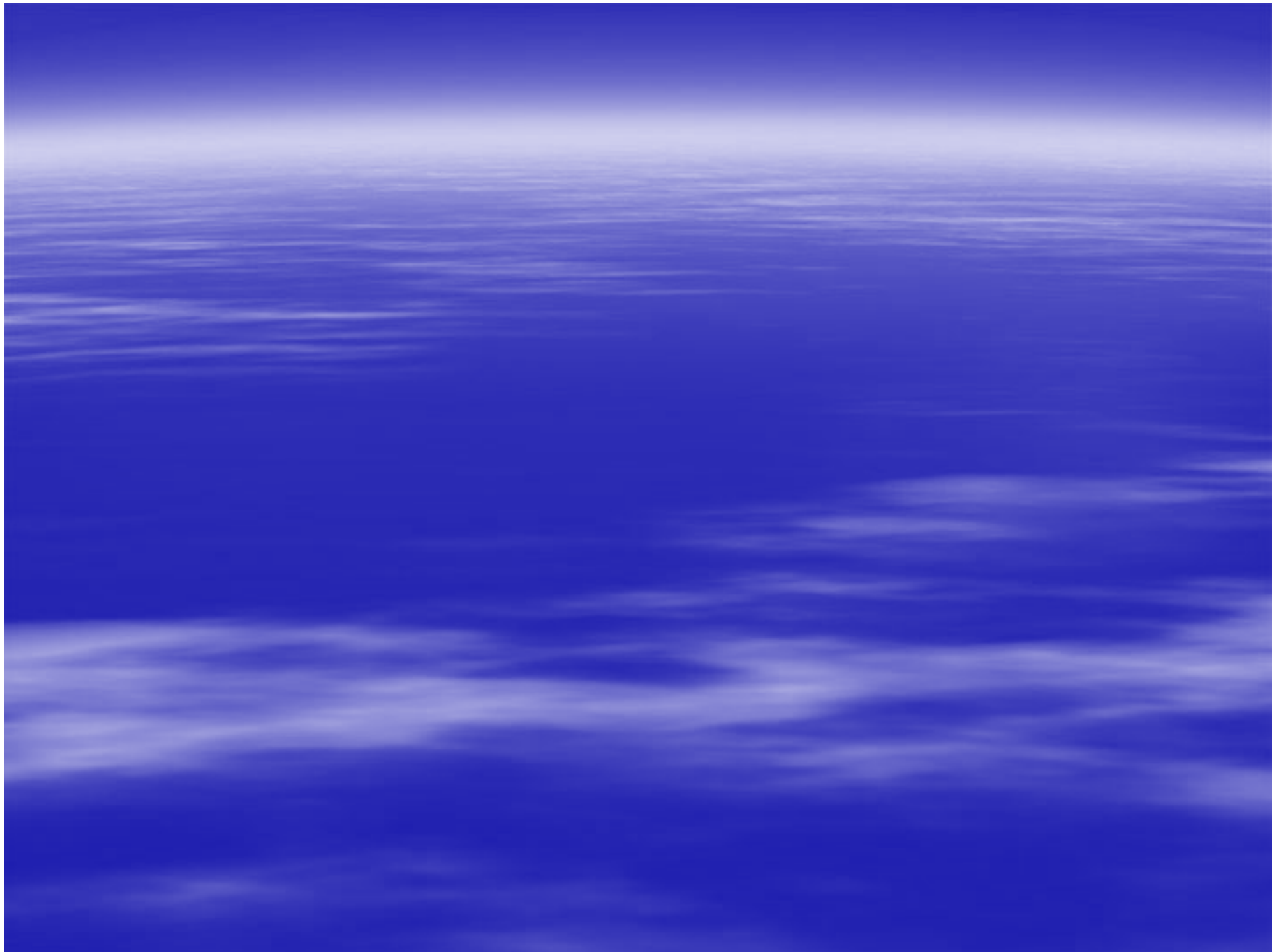


$m/z$

## Quadrupole Optics cont.

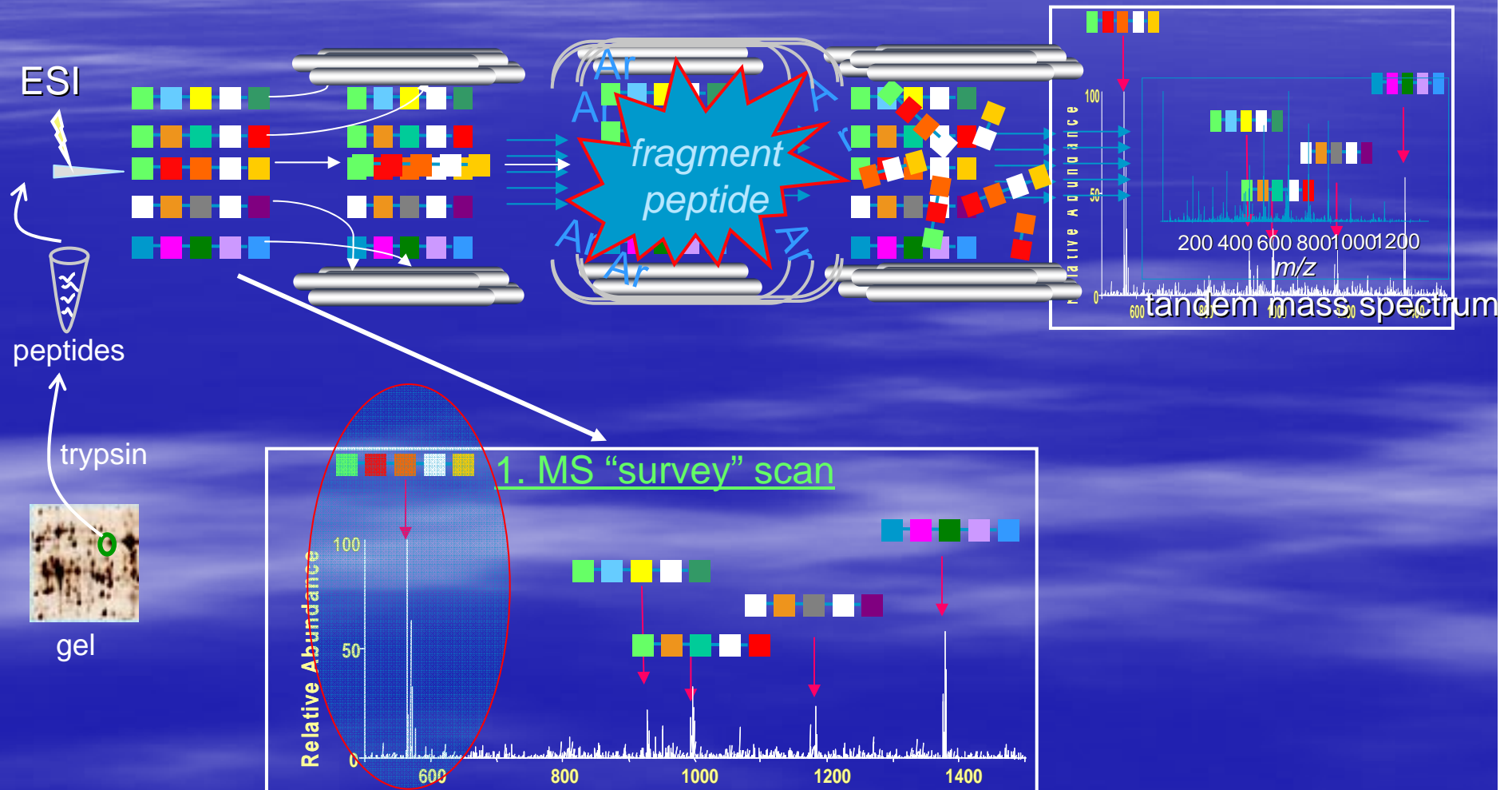


- The quadrupole can function in a second mode called tandem mass spectrometry or MS/MS.
- A particular peak is chosen from a MS scan and the first quad allows only that  $m/z$  to pass into the second quad.
- The second quad accelerates the species through a voltage causing collisions with an inert gas present.
- If the ion is a peptide, the collisions cause bond breakage selectively at the **amide** bonds
- The charged fragments enter the third quad which performs a MS scan generating a unique pattern associated with the fragments and thus the parent peptide
- These fragments can be deconvoluted to give a peptide sequence

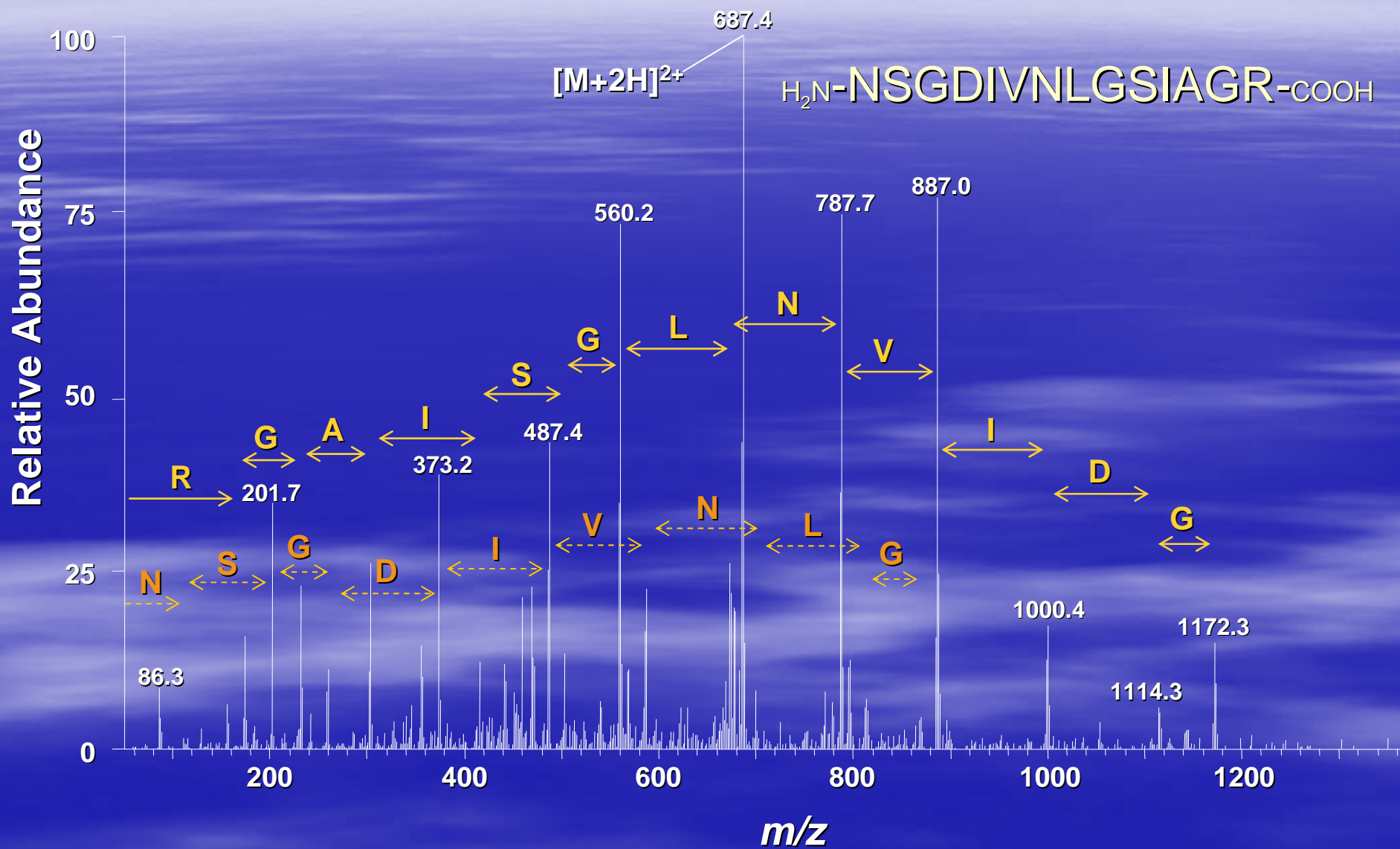
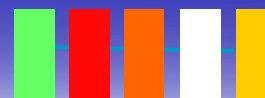




# Tandem mass spectrometry



# Sample Peptide



# 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 out from all others



# SEQUEST output file

INTERACT by J.Eng - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: http://198.107.152.2/cgi-bin/displayions\_html5?Dta=/data/search/dan/4serum/interact.htm What's Related

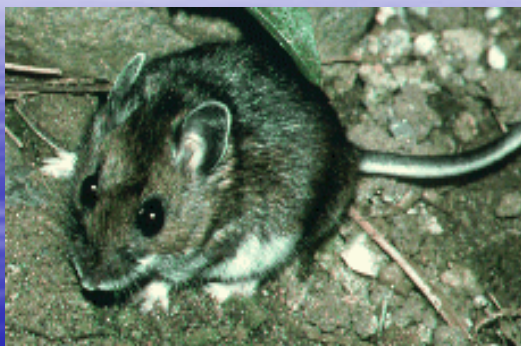
HealthLinks: Ca Google Yahoo! Finance Index of /data/ fishersci.com Hotmail Inbox dmartin@systems Calendar

FILE: /data/search/dan/4serum/interact-data.htm Tryptic: ☐ J.Eng 04/2000 ☐ LastUndo  
 Strict: ☐ DelRows: ☐  ☐ RestoreOrig  
 Text1:  Text2:  XCorr:  dCn:  Sp:  InclAA:  MarkAA:  ☐ SortProt  
☐ SortPep

|      |                      |               |        |        |        |   |         |               |     |                          |
|------|----------------------|---------------|--------|--------|--------|---|---------|---------------|-----|--------------------------|
| 1408 | ./serum4.1554.1554.3 | 3333.6 (-1.2) | 6.3209 | 0.361  | 1252.5 | 1 | 38/ 104 | SW:ALBU HUMAN | +9  | K.SHC*IAEVENDEMPADLPSLA  |
| 1514 | ./serum4.1702.1702.3 | 2834.2 (+0.3) | 6.8501 | 0.473  | 1660.2 | 1 | 33/ 84  | SW:A2MG HUMAN |     | R.SLFTDLEAENDVLHC*VAFVVP |
| 1008 | ./serum4.1219.1219.3 | 2430.6 (-1.2) | 5.9068 | 0.448  | 1643.3 | 1 | 32/ 72  | SW:TRFE HUMAN | +4  | K.SDNC*EDTPEAGYFAVAVVK.K |
| 1119 | ./serum4.1302.1302.3 | 2416.6 (+1.2) | 5.3039 | 0.362  | 2199.8 | 1 | 32/ 60  | SW:CFHD HUMAN | +4  | K.C*YFPYLENGYNQNYGR.K    |
| 1108 | ./serum4.1294.1294.3 | 2297.6 (-0.4) | 4.9588 | 0.359  | 1437.1 | 1 | 31/ 68  | SW:CO4 HUMAN  | +6  | R.GC*GEQTMIIYLAPTAAASR.Y |
| 1112 | ./serum4.1298.1298.2 | 2297.6 (-0.6) | 4.8816 | 0.430  | 1407.7 | 1 | 24/ 34  | SW:CO4 HUMAN  | +6  | R.GC*GEQTMIIYLAPTAAASR.Y |
| 1067 | ./serum4.1266.1266.2 | 2113.4 (-0.8) | 4.8355 | 0.345  | 2241.4 | 1 | 22/ 28  | SW:ALBU HUMAN | +11 | R.PC*FSALEVDETYVPK.E     |
| 1070 | ./serum4.1267.1267.3 | 2113.4 (+0.5) | 4.6168 | 0.345  | 1260.1 | 1 | 31/ 56  | SW:ALBU HUMAN | +11 | R.PC*FSALEVDETYVPK.E     |
| 1078 | ./serum4.1272.1272.3 | 2016.2 (-1.4) | 4.5652 | 0.268* | 2251.6 | 1 | 27/ 48  | SW:ALBU HUMAN | +11 | K.QNC*ELFEQLGEYK.F       |
| 764  | ./serum4.1031.1031.3 | 1703.9 (-0.5) | 4.5637 | 0.305  | 1894.7 | 1 | 27/ 44  | SW:HPT1 HUMAN | +9  | K.SC*AVAEYGVYVK.V        |
| 1379 | ./serum4.1527.1527.2 | 2004.3 (+2.6) | 4.5351 | 0.335  | 1023.4 | 1 | 21/ 26  | SW:A1BG HUMAN |     | R.C*EGPIPDVTFELLR.E      |
| 1071 | ./serum4.1268.1268.2 | 2113.4 (+2.0) | 4.5312 | 0.370  | 1722.4 | 1 | 21/ 28  | SW:ALBU HUMAN | +11 | R.PC*FSALEVDETYVPK.E     |
| 1295 | ./serum4.1456.1456.2 | 2234.6 (-0.5) | 4.4580 | 0.431  | 1087.2 | 1 | 22/ 32  | SW:A1BG HUMAN |     | K.VTLTC*VAPLSGVDFQLR.R   |
| 985  | ./serum4.1202.1202.2 | 2422.7 (+1.9) | 4.4538 | 0.295  | 1550.0 | 1 | 19/ 28  | SW:TRFE HUMAN | +4  | K.LC*MGSGNLNC*EPNNK.E    |
| 1158 | ./serum4.1338.1338.2 | 2465.8 (+1.8) | 4.3237 | 0.320  | 791.6  | 1 | 19/ 34  | SW:KNH HUMAN  | +2  | K.LGQSLDC*NAEVYVVPWEK.K  |
| 1120 | ./serum4.1303.1303.2 | 2416.6 (-0.7) | 4.2952 | 0.335  | 1041.9 | 1 | 21/ 30  | SW:CFHD HUMAN | +4  | K.C*YFPYLENGYNQNYGR.K    |
| 877  | ./serum4.1112.1112.2 | 1936.1 (+2.9) | 4.2935 | 0.272  | 1246.6 | 1 | 18/ 26  | SW:TRFE HUMAN | +4  | R.FDEFFSEGC*APGSK.K      |
| 1166 | ./serum4.1344.1344.2 | 2441.7 (+0.4) | 4.2876 | 0.328  | 652.8  | 1 | 24/ 40  | SW:A2MG HUMAN | +1  | K.AGAFCLSEDAGLGISSTASLR  |
| 1116 | ./serum4.1300.1300.2 | 2297.6 (+2.4) | 4.2768 | 0.386  | 1346.4 | 1 | 22/ 34  | SW:CO4 HUMAN  | +6  | R.GC*GEQTMIIYLAPTAAASR.Y |
| 1306 | ./serum4.1468.1468.3 | 2814.1 (-0.7) | 4.2237 | 0.002  | 658.8  | 1 | 27/ 84  | SW:CERU HUMAN | +3  | R.MYSVNGYTFGSLPGLSMC*AED |
| 1126 | ./serum4.1307.1307.2 | 2416.6 (+2.3) | 4.1085 | 0.269  | 1092.9 | 1 | 21/ 30  | SW:CFHD HUMAN | +4  | K.C*YFPYLENGYNQNYGR.K    |
| 1023 | ./serum4.1230.1230.2 | 2015.3 (+2.2) | 4.1073 | 0.396  | 1189.6 | 1 | 21/ 28  | SW:IGJ HUMAN  | +1  | K.C*YTAVPLVYGGETK.M      |
| 1012 | ./serum4.1222.1222.3 | 2015.3 (-0.3) | 4.0929 | 0.164  | 1386.5 | 1 | 30/ 56  | SW:IGJ HUMAN  | +1  | K.C*YTAVPLVYGGETK.M      |
| 1291 | ./serum4.1454.1454.2 | 2234.6 (+1.7) | 4.0300 | 0.258  | 882.3  | 1 | 20/ 32  | SW:A1BG HUMAN |     | K.VTLTC*VAPLSGVDFQLR.R   |
| 1294 | ./serum4.1455.1455.3 | 2234.6 (+1.0) | 4.0248 | 0.267  | 1238.3 | 1 | 29/ 64  | SW:A1BG HUMAN |     | K.VTLTC*VAPLSGVDFQLR.R   |
| 1153 | ./serum4.1330.1330.3 | 2297.6 (+0.2) | 4.0169 | 0.224  | 1330.8 | 1 | 31/ 68  | SW:CO4 HUMAN  | +6  | R.GC*GEQTMIIYLAPTAAASR.Y |
| 1162 | ./serum4.1342.1342.2 | 2465.8 (-0.6) | 3.9660 | 0.295  | 930.3  | 1 | 20/ 34  | SW:KNH HUMAN  | +2  | K.LGQSLDC*NAEVYVVPWEK.K  |
| 1013 | ./serum4.1223.1223.2 | 2430.6 (-0.4) | 3.9568 | 0.268  | 631.8  | 1 | 20/ 36  | SW:TRFE HUMAN | +4  | K.SDNC*EDTPEAGYFAVAVVK.K |
| 1484 | ./serum4.1662.1662.2 | 2529.9 (-0.1) | 3.9566 | 0.205  | 526.7  | 1 | 18/ 36  | SW:TRFE HUMAN | +2  | R.SAGWNIPIGLLYC*DLPEPR.K |

http://198.107.152.2/cgi-bin/displayions\_html5?Dta=/data/search/dan/4serum/serum4/serum4.1662.1662.2.dta&MassType=0&NumAxis=1&D

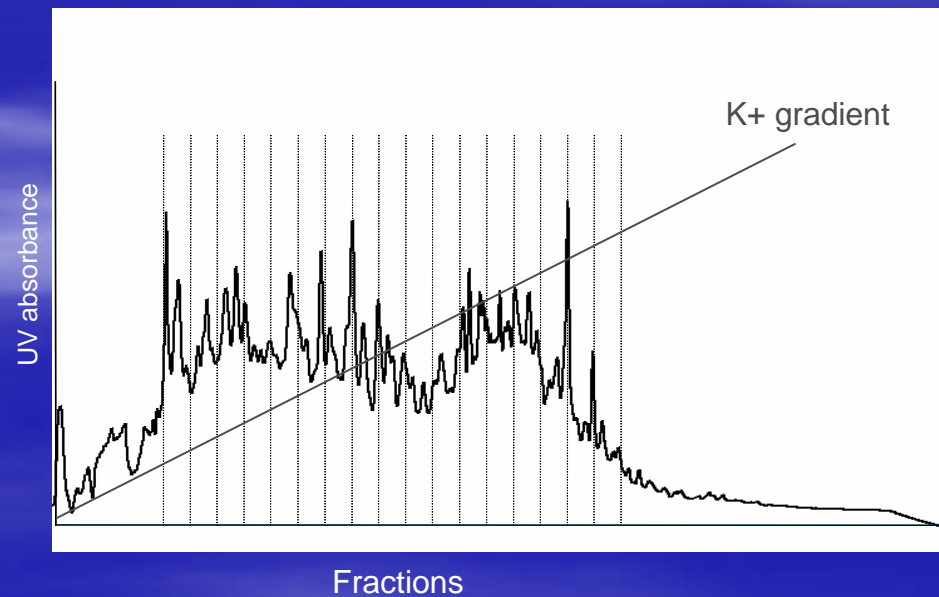
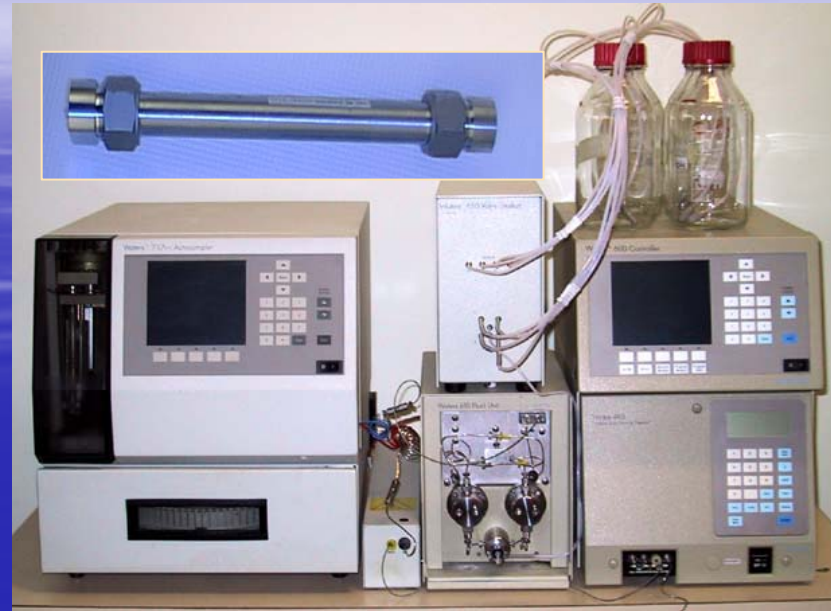
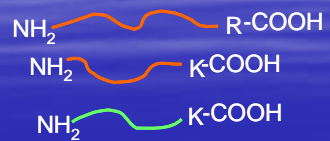
# Mixture analysis – the need for separations



- Mass spectrometers have limited peak capacities requiring separation and fractionation prior to analysis
- Separation methods include:
  - gels
  - liquid chromatography
  - affinity chromatography
  - immunochromatography
  - selective enrichment by covalent chemistry

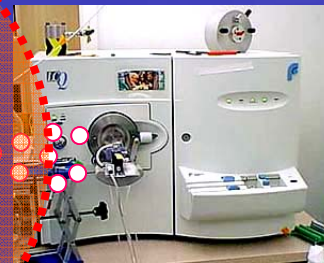
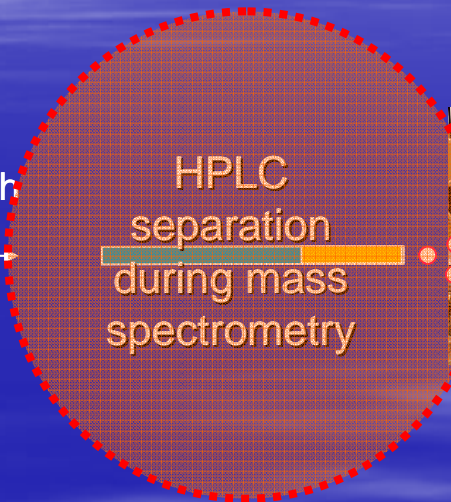
# Separation of peptides by electrostatic charge using cation exchange

In acidic conditions, tryptic peptides are positively charged





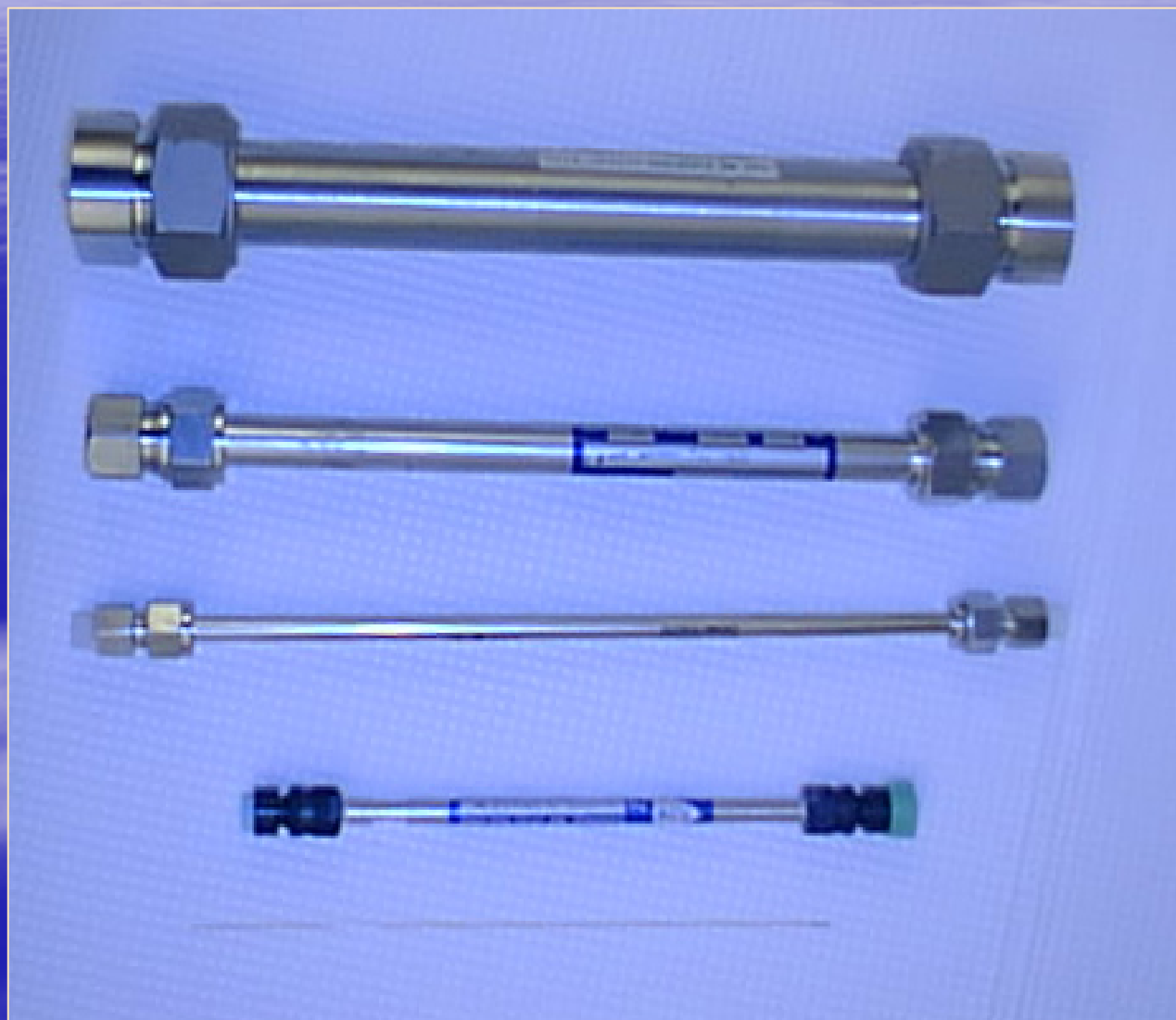
Digestion with  
Trypsin to  
peptides and  
separation







# Liquid chromatography miniaturization for ESI-MS

**decrease size, increase sensitivity**



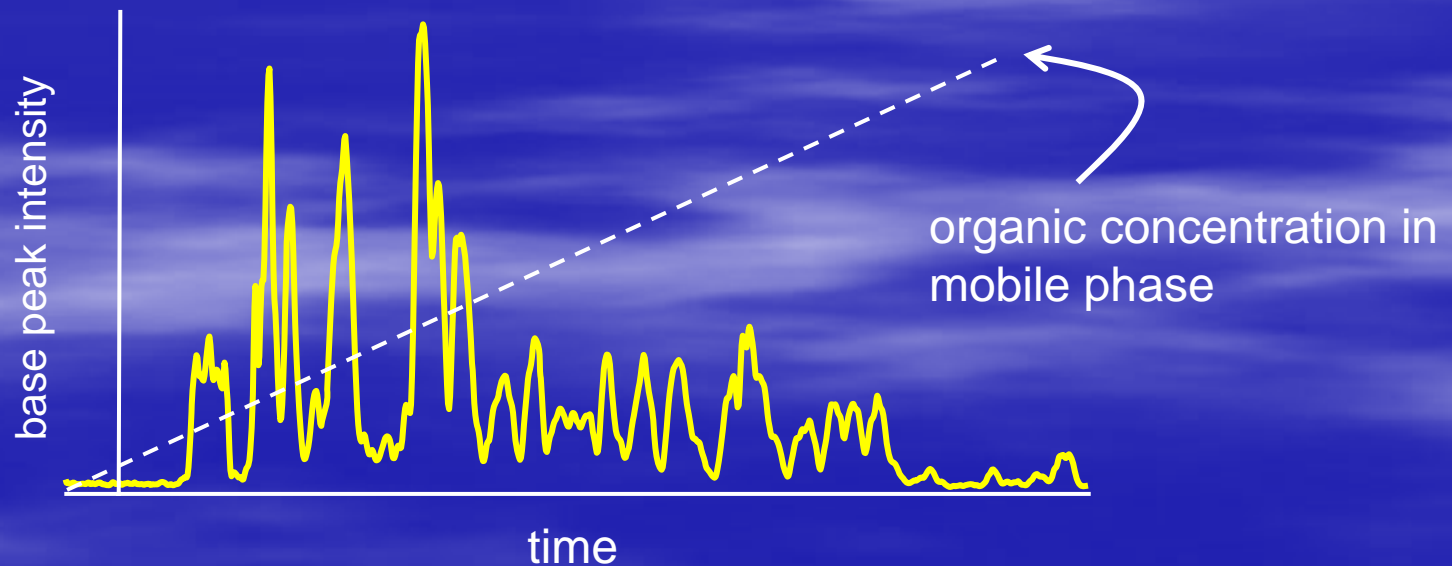
| <u>Size</u>  | <u>Sensitivity</u>  |
|---|--|
| 2.1 cm  | ----   |
| 1.0 cm  | 4.4 fold   |
| 4.6 mm  | 21 fold  |
| 1.0 mm  | 441 fold   |
| 50 µm   | 176,400 fold   |



## microcapillary LC-MS

---

- reverse-phase separation of peptides – allows for a gradual introduction of peptides into the mass spectrometer.
- removal of contaminants through washing



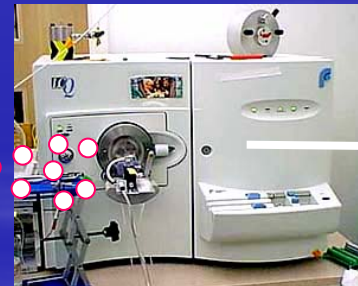
# A New Paradigm--Mass spectrometry based Proteomics--Without Electrophoresis

Mass spectrometry  
identifies proteins



Digestion with  
Trypsin to  
peptides and  
separation

HPLC  
separation  
during mass  
spectrometry



| Gene     | Protein Name                                       | Area  | S.D. | Unique | Cell |
|----------|--|-------|------|--------|------|
| TNFRSF10 | TNF receptor superfamily member 10B precursor      | 17.17 | 4.58 | 6      | M    |
| CD4      | CD4 precursor                                      | 12.88 | 5.08 | 5      | M    |
| TPRC     | Transferrin receptor protein 1                     | 7.81  | 2.27 | 3      | M    |
| PLTR     | Functional adhesion molecule 1 precursor           | 7.36  | 1.98 | 4      | M    |
| CD40     | CD40-like protein 3 precursor                      | 7.03  | 2.72 | 3      | S    |
| LDLR     | Low-density lipoprotein receptor precursor         | 6.57  | 1.64 | 12     | M    |
| ACM12    | Acyl-coenzyme A precursor                          | 6.58  | 1.40 | 1      | S    |
| CD44     | CD44 precursor                                     | 5.93  | 1.10 | 4      | M    |
| SCCHN    | Scch1-1152, unannotated protein                    | 5.39  | 3.63 | 1      | V    |
| CD44     | CD44-like protein 2 precursor                      | 5.24  | 3.42 | 6      | S    |
| CD44     | CD44-like protein 3 precursor                      | 4.92  | 1.53 | 23     | M    |
| CD44     | CD44-like protein 4 precursor                      | 4.17  | 2.08 | 9      | M    |
| CD44     | CD44-like protein 5 precursor                      | 3.89  | 1.08 | 22     | S    |
| CD44     | CD44-like protein 6 precursor                      | 3.59  | 0.88 | 37     | S    |
| CD44     | CD44-like protein 7 precursor                      | 3.48  | 0.72 | 3      | S    |
| CD44     | CD44-like protein 8 precursor                      | 3.24  | 0.33 | 2      | C    |
| CD44     | CD44-like protein 9 precursor                      | 3.21  | 0.38 | 7      | M    |
| CD44     | CD44-like protein 10 precursor                     | 3.13  | 1.49 | 3      | M    |
| CD44     | CD44-like protein 11 precursor                     | 2.97  | 0.41 | 4      | M    |
| CD44     | CD44-like protein 12 precursor                     | 2.85  | 1.00 | 2      | M    |
| CD44     | CD44-like protein 13 precursor                     | 2.85  | 0.10 | 2      | M    |
| CD44     | CD44-like protein 14 precursor                     | 2.83  | 0.74 | 10     | M    |
| CD44     | CD44-like protein 15 precursor                     | 2.47  | 0.33 | 6      | S    |
| CD44     | CD44-like protein 16 precursor                     | 2.45  | 0.48 | 2      | M    |
| CD44     | CD44-like protein 17 precursor                     | 2.45  | 0.93 | 4      | C    |
| CD44     | CD44-like protein 18 precursor                     | 2.34  | 0.48 | 14     | S    |
| CD44     | CD44-like protein 19 precursor                     | 2.33  | 0.36 | 27     | M    |
| CD44     | CD44-like protein 20 precursor                     | 2.17  | 0.01 | 5      | M    |
| CD44     | CD44-like protein 21 precursor, contains Saposin A | 2.16  | 0.35 | 1      | S    |
| CD44     | CD44-like protein 22 precursor                     | 2.13  | 1.69 | 15     | M    |
| CD44     | CD44-like protein 23 precursor                     | 2.11  | 1.22 | 3      | GOI  |
| CD44     | CD44-like protein 24 precursor                     | 2.10  | 0.46 | 10     | S    |
| CD44     | CD44-like protein 25 precursor                     | 2.10  | 0.88 | 15     | M    |
| CD44     | CD44-like protein 26 precursor                     | 2.10  | 0.10 | 3      | M    |
| CD44     | CD44-like protein 27 precursor                     | 2.10  | 0.51 | 2      | M    |
| CD44     | CD44-like protein 28 precursor                     | 2.10  | 0.01 | 4      | M    |
| CD44     | CD44-like protein 29 precursor                     | 2.09  | 0.08 | 1      | M    |
| CD44     | CD44-like protein 30 precursor                     | 2.09  | 0.52 | 3      | GOI  |

Now all elements of the system  
make sense

# Quantitative Proteomics

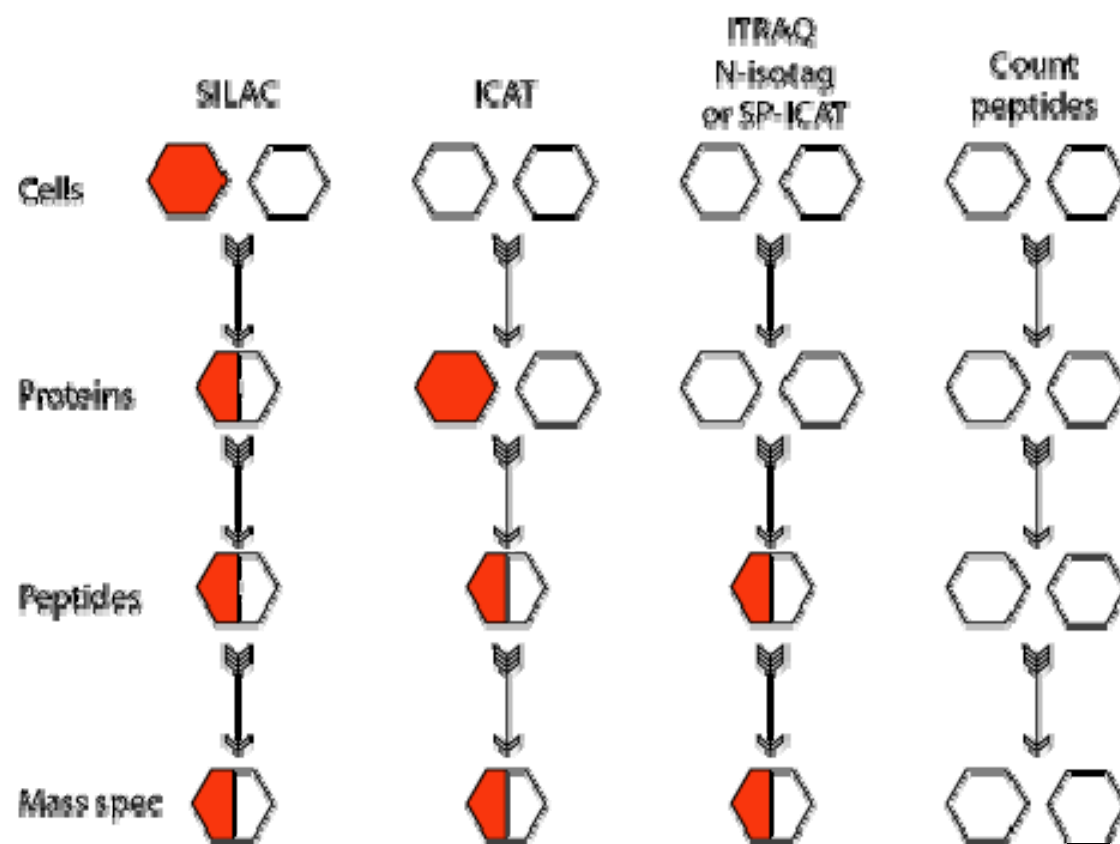
# Using stable heavy isotopes to perform quantification

- Because a mass spectrometer measures mass, *isotopic modifications can allow one to distinguish two samples.*
- This is done using stable isotopes such as N15, C13 and Deuterium.
- Isotopes can be introduced at various stages of an experiment

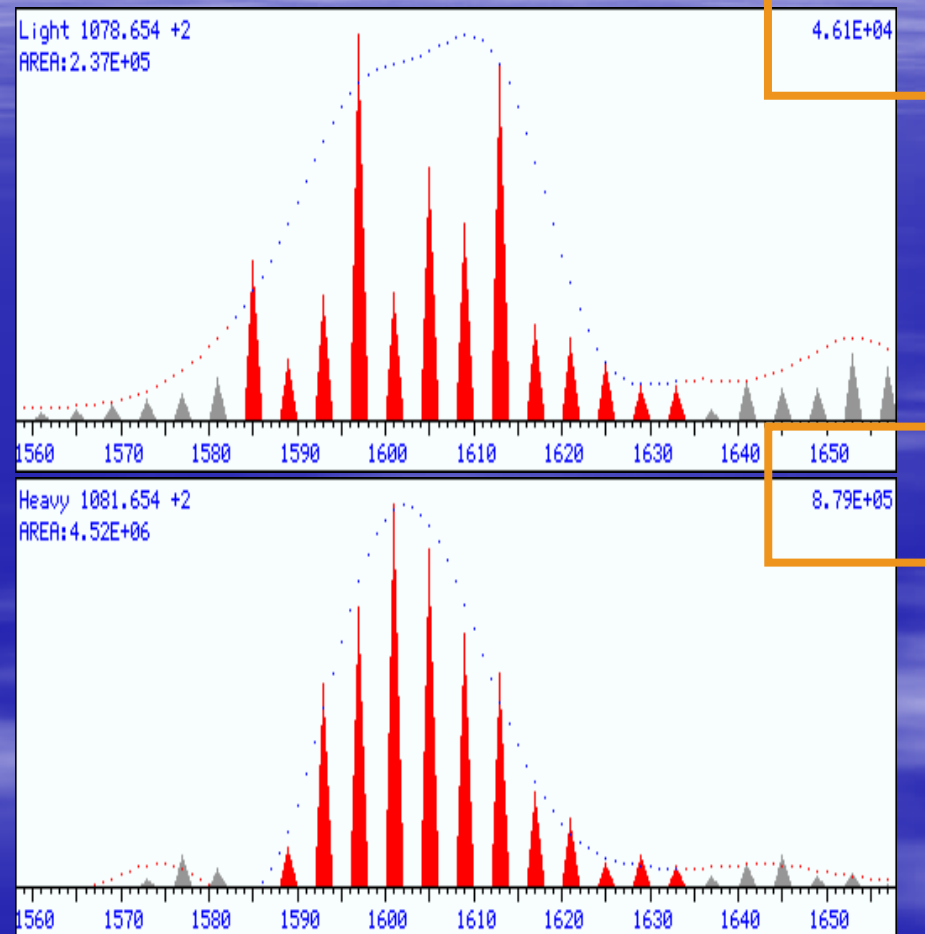
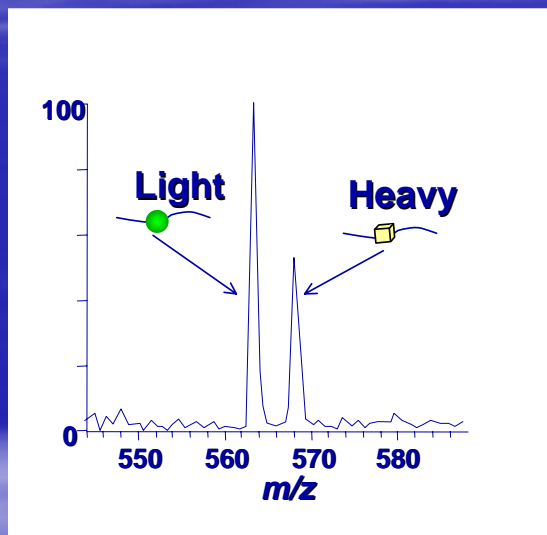
# Methods of Stable Isotope Incorporation

- Metabolic Labeling- with a heavy amino acid
  - SILAC (stable isotope labeling with amino acids in cell culture) uses Lysine or Arginine - 2004
- Chemical Modification
  - Cysteine
    - ICAT–Cysteine labeling of PROTEINS with capture of only the modified peptides - 2002
    - SP-ICAT a cysteine labeling of peptides on solid phase - 2005
  - Amine based
    - ITRAQ (2004) and N-Isotag (?) –amine labeling of peptides





For each peptide a ratio of heavy to light is calculated.

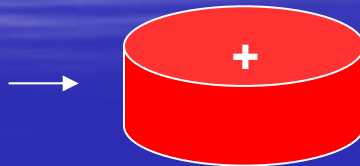


Light : Heavy  
1 : 19.07

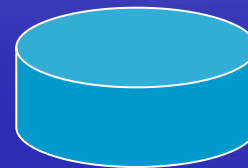
# SILAC

Media w/o Lys and Arg  
FCS dialyzed to remove  
amino acids

Heavy  
Lysine,  
Arginine  
or both

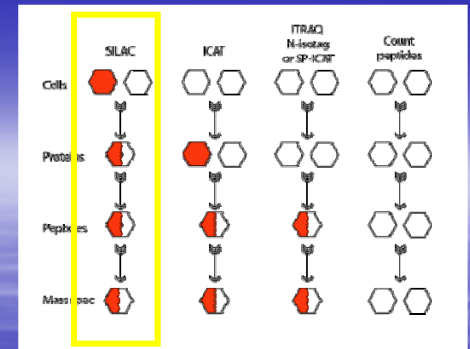
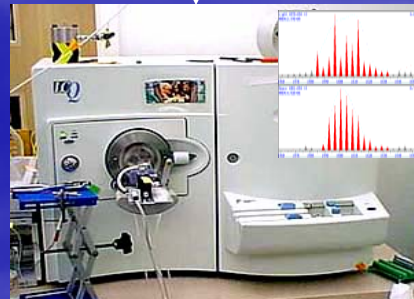


Stimulus  
of interest



Control

Lyse cells  
together



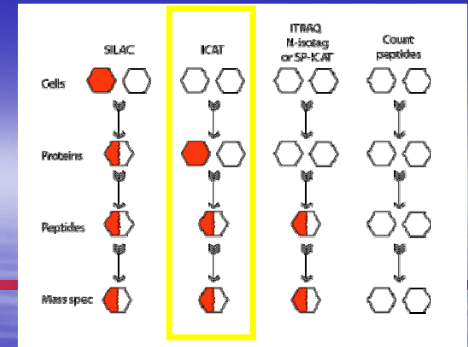
Unmodified  
amino acids

Simple

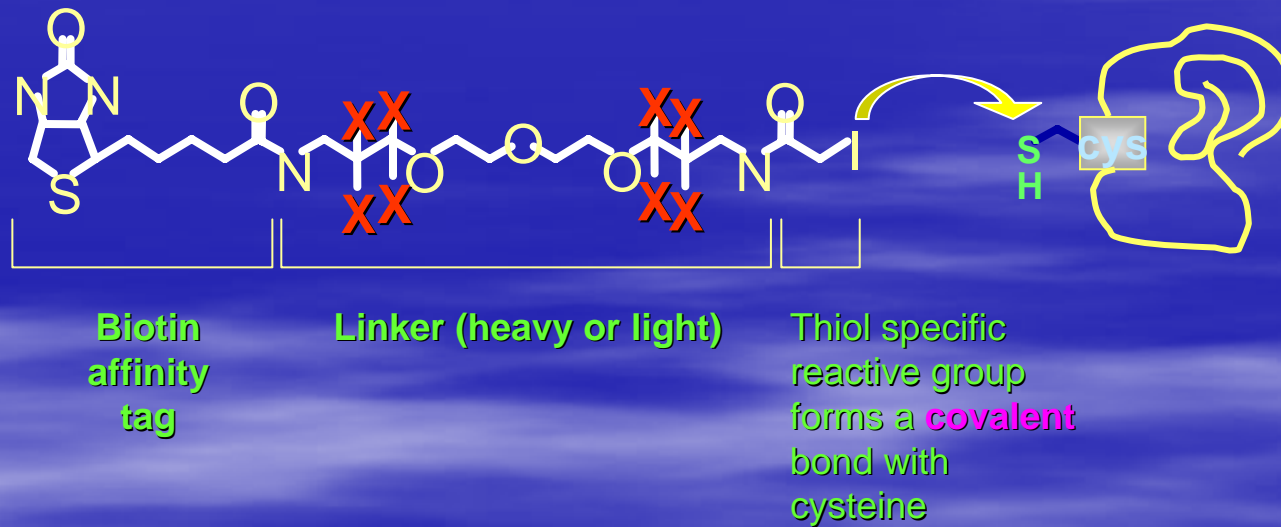
Straightforward

Dialysis of FCS may be a  
problem for some cell lines

# Isotope Coded Affinity Tags (ICAT)

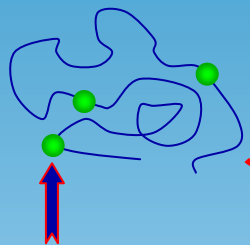


ICAT Reagents: Heavy reagent: d8-ICAT (X=deuterium)  
Light reagent: d0-ICAT (X=hydrogen)

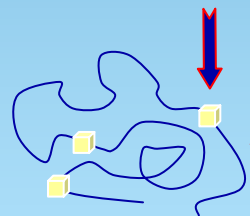


# Protein Quantification and Identification by the ICAT Strategy

**Mixture 1**

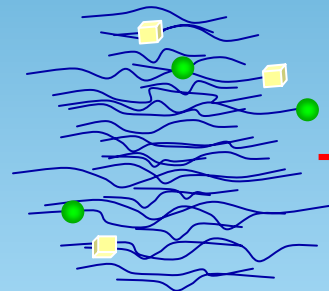


**ICAT-labeled  
cysteines**



**Mixture 2**

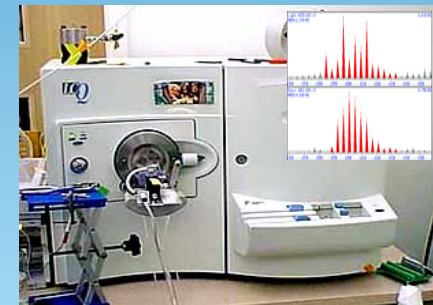
**Proteins**



**Combine and  
proteolyze**

**Affinity  
separation  
with Avidin**

**Peptides**



More complicated  
Potential for losses  
Some proteins lack cys



How is this technology  
implemented?

Large  
scale



Medium  
scale



Small scale



# Global Proteomics

- Big project.
- Costly
- Requires extensive resources for follow up
- Large potential payoffs for novel discoveries

Big project example:

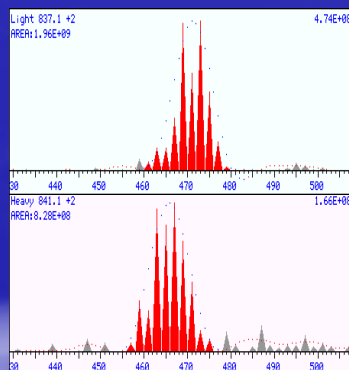
Analysis of proteins released from  
androgen stimulated or unstimulated  
cultured prostate cancer cells

Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium.  
Cancer Res. 2004 Jan 1;64(1):347-55.

- This project took approximately 12 months to complete.
- If performed beginning today the proteomic portion could be finished in 3-4 months.
  - Automation of mass spectrometers
  - Automation of data analysis using PeptideProphet and ProteinProphet
  - Automation of quantification using ASAPratio



More  
abundant  
with  
androgen  
addition



| Gene     | Protein Name  | A+/A- | S.D. | Unique | Cell |
|----------|---|-------|------|--------|------|
| TNFRSF10 | TNF receptor superfamily member 10B precursor           | 17.17 | 4.58 | 6      | M    |
| BSG      | Basigin precursor                                       | 12.88 | 5.02 | 5      | M    |
| TFRC     | Transferrin receptor protein 1                          | 7.81  | 2.27 | 3      | M    |
| F11R     | Junctional adhesion molecule 1 precursor                | 7.36  | 1.68 | 4      | M    |
| IGFBP3   | Insulin-like growth factor binding protein 3 precursor  | 7.08  | 2.72 | 3      | S    |
| LDLR     | Low-density lipoprotein receptor precursor              | 6.57  | 1.64 | 12     | M    |
| ANGPT2   | Angiopoietin-2 precursor                                | 6.08  | 1.40 | 1      | S    |
| IGSF8    | EWI2  | 5.43  | 1.10 | 4      | M    |
| SCOTIN   | ??Scotin-BAB71152. unnamed protein                      | 5.39  | 3.63 | 1      | ?    |
| KLK2     | Glandular kallikrein 2 precursor                        | 5.34  | 3.42 | 6      | S    |
| JAG1     | Jagged 1 precursor                                      | 4.92  | 1.43 | 23     | M    |
| APLP2    | Amyloid-like protein 2 precursor                        | 4.17  | 2.06 | 9      | M    |
| SPINT1   | Kunitz-type protease inhibitor 1 precursor              | 3.89  | 1.08 | 22     | S    |
| LRIG1    | Membrane glycoprotein LIG-1                             | 3.59  | 1.39 | 3      | M    |
| KLK3     | Prostate specific antigen precursor                     | 3.50  | 0.66 | 37     | S    |
| VEGF     | Vascular endothelial growth factor A precursor          | 3.49  | 0.24 | 3      | S    |
| LGMN     | Legumain precursor                                      | 3.24  | 0.33 | 2      | C    |
| ALCAM    | CD166 antigen precursor                                 | 3.21  | 0.38 | 7      | M    |
| KIAA1265 | Hypothetical protein KIAA1265                           | 3.13  | 1.24 | 3      | M    |
| ACVR1B   | activin A type IB receptor                              | 2.97  | 0.41 | 4      | M    |
| PLXNB3   | Plexin-B3   | 2.95  | 1.00 | 2      | M    |
| IL6R     | Interleukin-6 receptor alpha chain precursor            | 2.84  | 0.72 | 3      | M    |
| C1orf8   | Protein C1orf8 precursor                                | 2.66  | 0.10 | 2      | M    |
| ST14     | Suppressor of tumorigenicity 14                         | 2.60  | 0.74 | 10     | M    |
| PLAB     | Prostate differentiation factor                         | 2.47  | 0.34 | 6      | S    |
| NOTCH2   | Neurogenic locus notch homolog protein 2 precursor      | 2.45  | 0.40 | 2      | M    |
| MLP      | MARCKS-related protein                                  | 2.38  | 0.93 | 4      | C    |
| B2M      | Beta-2-microglobulin precursor                          | 2.34  | 0.48 | 14     | S    |
| NRP1     | Neuropilin-1 precursor                                  | 2.33  | 0.36 | 27     | M    |
| NEO1     | Neogenin precursor                                      | 2.27  | 0.03 | 5      | M    |
| PSAP     | Proactivator polypeptide precursor, contains: Saposin A | 2.26  | 0.35 | 1      | S    |
| CDH1     | E-cadherin  | 2.23  | 1.69 | 15     | M    |
| ATOX1    | Copper transport protein ATOX1                          | 2.21  | 0.22 | 3      | GOL  |
| ADAM10   | ADAM10  | 2.20  | 0.46 | 10     | S    |
| IGF2R    | Cation-independent mannose-6-phosphate receptor         | 2.20  | 0.60 | 16     | M    |
| TNFRSF10 | TNF receptor superfamily member 10D precursor           | 2.19  | 0.10 | 3      | M    |
| PLXNB1   | Plexin-B1/SEP receptor precursor                        | 2.18  | 0.51 | 2      | M    |
| SEMA4A   | Semaphorin 4A precursor                                 | 2.18  | 0.48 | 4      | M    |
| VLDLR    | Very low-density lipoprotein receptor precursor         | 2.09  | 0.05 | 1      | M    |
| GALNT1   | Polypeptide N-acetylgalactosaminyltransferase           | 2.09  | 0.52 | 3      | GOL  |

# Medium Scale and Phosphorylation Proteomics

The key to medium scale proteomics is a mechanical or affinity based preparation that reduces the complexity of the starting material.

Examples: centrifugation, precipitation.

# Peroxisome membrane

## Quantitative mass spectrometry reveals a role for the GTPase Rho1p in actin organization on the peroxisome membrane

Marcello Morali,<sup>1</sup> Jennifer J. Smith,<sup>1</sup> Sunhee Jung,<sup>1</sup> Eugene Yi,<sup>1</sup> Alexey I. Nesvizhskii,<sup>1</sup> Rowan H. Christmas,<sup>1</sup> Ramsey A. Salem,<sup>1</sup> Yuen Yi C. Tam,<sup>2</sup> Andrei Fagarasanu,<sup>2</sup> David R. Goodlett,<sup>1</sup> Ruedi Aebersold,<sup>1</sup> Richard A. Radubinski,<sup>2</sup> and John D. Aitchison<sup>1,2</sup>

<sup>1</sup>Section for Systems Biology, Seattle, WA 98103  
<sup>2</sup>Department of Cell Biology, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

**W**e have combined classical subcellular fractionation with large-scale quantitative mass spectrometry to identify proteins that enrich specifically with peroxisomes of *Saccharomyces cerevisiae*. In two complementary experiments, isotope-coded affinity tags and tandem mass spectrometry were used to quantify the relative enrichment of proteins during the purification of peroxisomes. Mathematical modeling of the data from 306 quantified proteins led to a prioritized list of 70 candidates whose enrichment scores indicated a high likelihood of them being peroxisomal. Among these pro-

teins, eight novel peroxisome-associated proteins were identified. The top novel peroxisomal candidate was the small GTPase Rho1p. Although Rho1p has been shown to be tethered to membranes of the secretory pathway, we show that it is specifically recruited to peroxisomes upon their induction in a process dependent on its interaction with the peroxisome membrane protein Pax25p. Rho1p regulates the assembly state of actin on the peroxisome membrane, thereby controlling peroxisome membrane dynamics and biogenesis.

### Introduction

Although the complete sequence of a genome provides a blueprint for the protein inventory of an organism, understanding the dynamic and responsive organization of a proteome remains a major challenge. Within eukaryotic cells, subcellular organelles are the most obvious level of organization, constituting assemblies of localized proteins that impart efficiency and control over the biochemical functions performed by the proteome. Recent advances that have increased the sensitivity and throughput of mass spectrometry (MS) have made possible the identification of proteins in samples of complexity on the order of organelles. However, the use of MS to comprehensively define organelle protein content is still a formidable undertaking. The polydispersity within organelle classes resulting from biological diversity and the limited resolving power of sub-fractionation techniques contribute to the notorious problem of

organelle contamination by proteins from other cellular compartments. Moreover, the levels of different proteins in an organelle fraction can vary over several orders of magnitude, resulting in highly represented proteins, or even contaminants, dominating the mass spectrometric analysis.

The issue of sample complexity has been addressed at both the prefractionation and instrumentation levels (for review see Aebersold and Mann, 2003). Likewise, various biochemical methods, including serial purification, immunodepletion, and free flow electrophoresis, have been applied to reduce contaminants (for review see Bruet et al., 2003). Although these methods improve sample purity, they remain unable to discriminate between bona fide organelle constituents and residual contaminants.

The problem of contaminants in isolated organelles is not new to the postgenomic era. Classically, de Duve (1992) defined one constituent of a subcellular fraction not as the proteins present in the fraction but rather as the proteins that specifically enrich in that fraction relative to other fractions, a designation that requires knowledge of relative protein abundances. The application of these principles of fractionation analysis to high-throughput proteomics can, in effect, address the issue of contaminating proteins. However, traditional MS is not well

“We have combined classical subcellular fractionation with large-scale quantitative mass spectrometry to identify proteins that enrich specifically with peroxisomes of *Saccharomyces cerevisiae*.”

Downloaded from www.jcb.org on February 3, 2006

The online version of this article includes supplemental material.  
Correspondence to John D. Aitchison: jda@systemsbiology.org.  
Abbreviations used in this paper: AP, affinity-purified peroxisomal membrane; DsRed, *Drosophila* sp. red fluorescent protein; ICA7, isotope-coded affinity tag; MS, mass spectrometry; pIC50/MS/MS, microcapillary liquid chromatography/MS; microarray comparative hybridization; PIS, peroxisomal targeting signal; RFP, monomeric DsRed; S25, *Saccharomyces cerevisiae* Database.

© The Rockefeller University Press. 0021-9545/06/271112-11\$15.00  
The Journal of Cell Biology, Vol. 174, No. 4, December 18, 2006 1100–1112  
http://www.jcb.org/cgi/doi/10.1083/jcb.200608110

Supplemental material can be found at:  
http://www.jcb.org/cgi/content/full/jcb.200608110

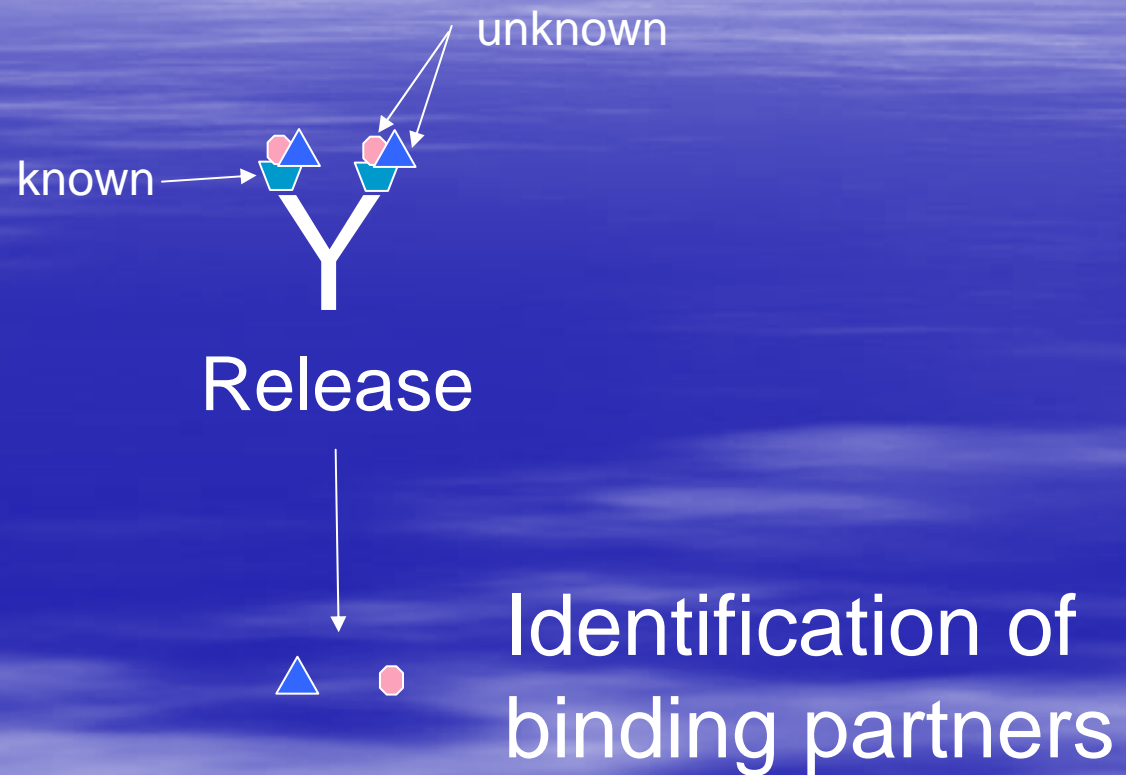
# Micro Proteomics



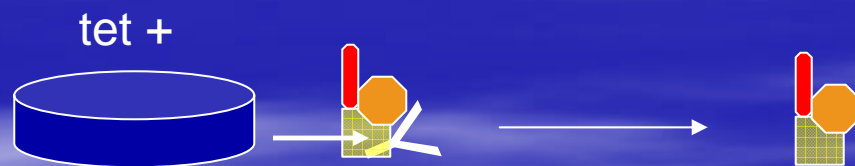
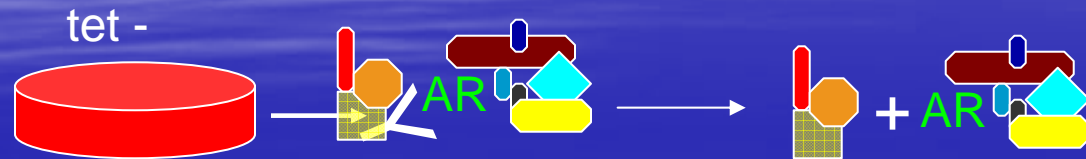
# Micro-Proteomics

- Focus on a particular protein complex
- Affinity based purification
- Takes advantage of existing technology
- Results more approachable for an individual investigator

# Micro-Proteomics



# Stable tet-off FLAG-AR Hela cell line



Big Ratios  
define  
proteins of  
interest

# Conclusions 1

- Proteomics is rapidly advancing:
  - Relative quantification is here.
  - Large scale experiments are becoming easier with better automation tools, BUT they generate vast amounts of data and consume significant resources.
  - Medium and small scale projects can be approached by an individual investigator here and now.
  - Phosphorylation is observable but methodology is still under development.

# Conclusions 2

- Mass spectrometers are fantastic
- The results you get out are determined by what you put in.
- The results you get out are determined by what you put in.
- The results you get out are determined by what you put in.
- Problems are most often NOT the result of poor instrument performance
- Your results depend on the following:
  - The purity and cleanliness of you preparation
  - The complexity of your sample compared to the dynamic range and duty cycle of the instrument