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.

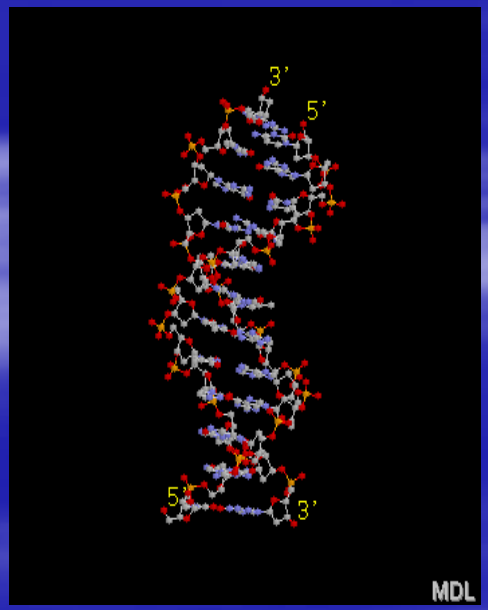
The genome and the proteome: a comparison

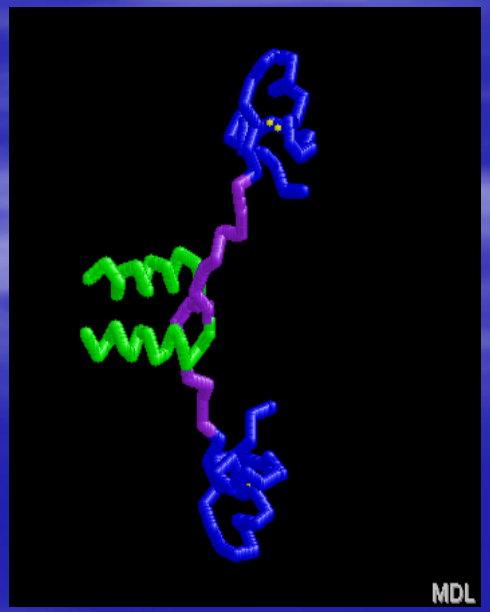
Genome

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

Proteome

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



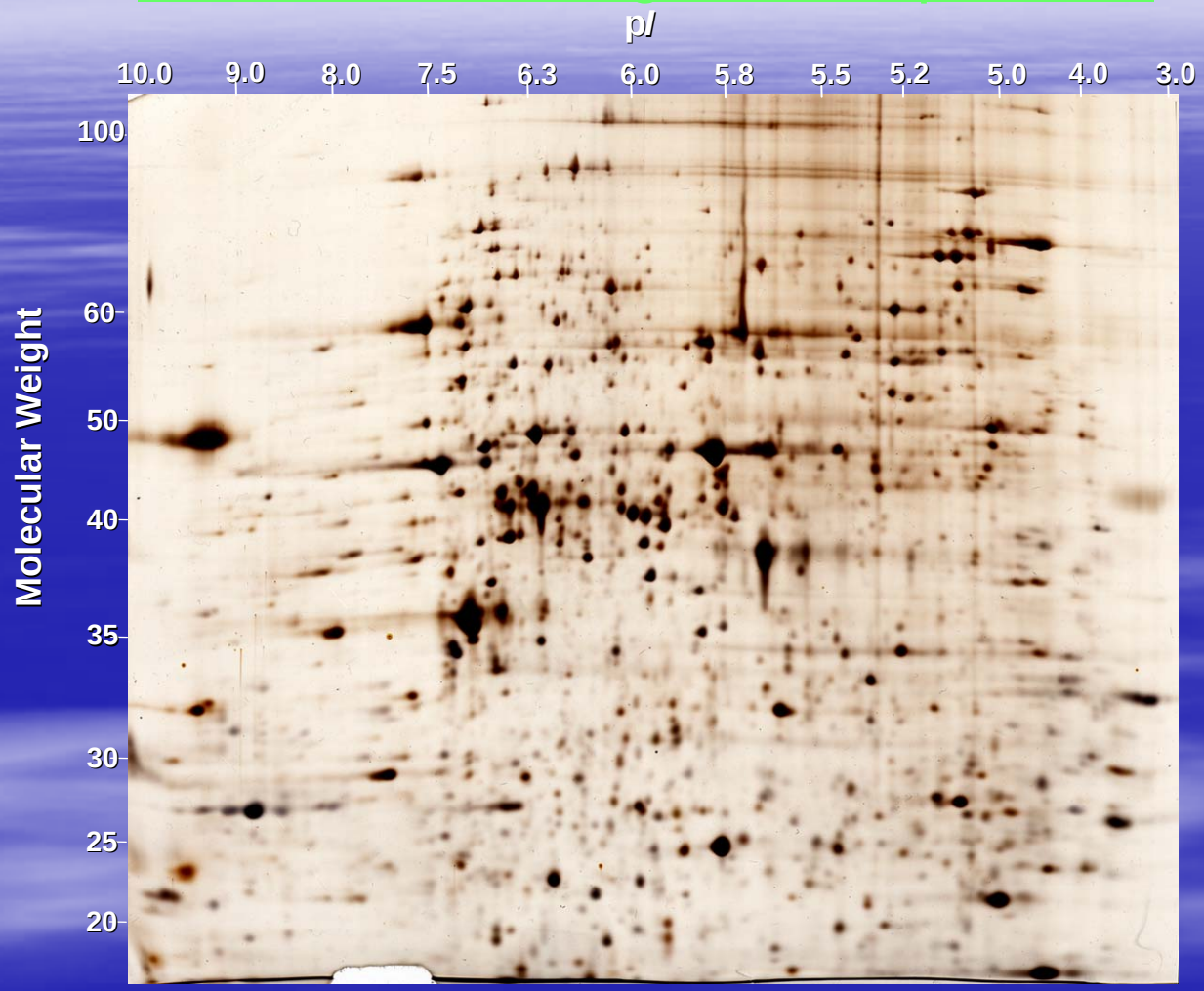


Proteomics technologies and methods

- 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

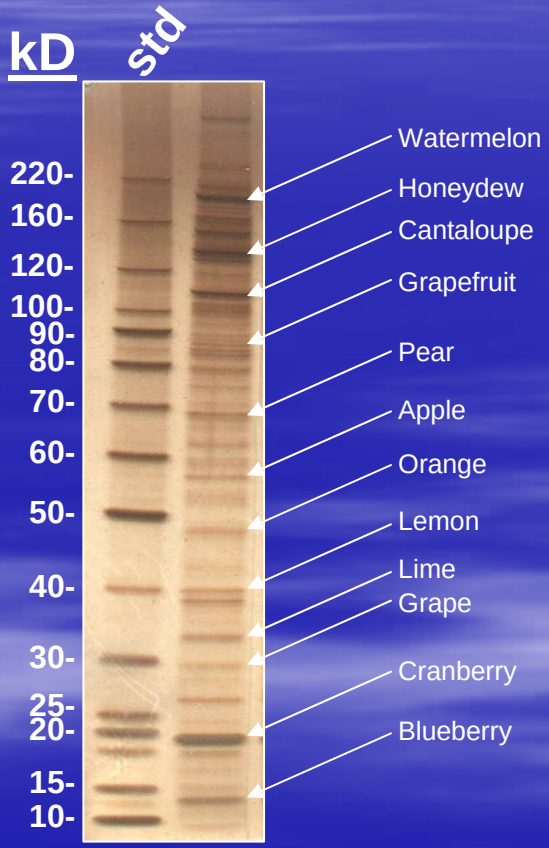
What it is not:

Magic

Able to give all the answers

Simple (relative to running a gel)

How to think about spectrometry based proteomics

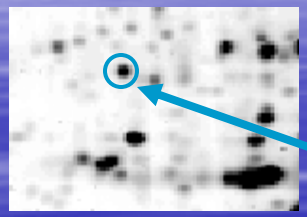


Dynamic Range $=10^3$

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.

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

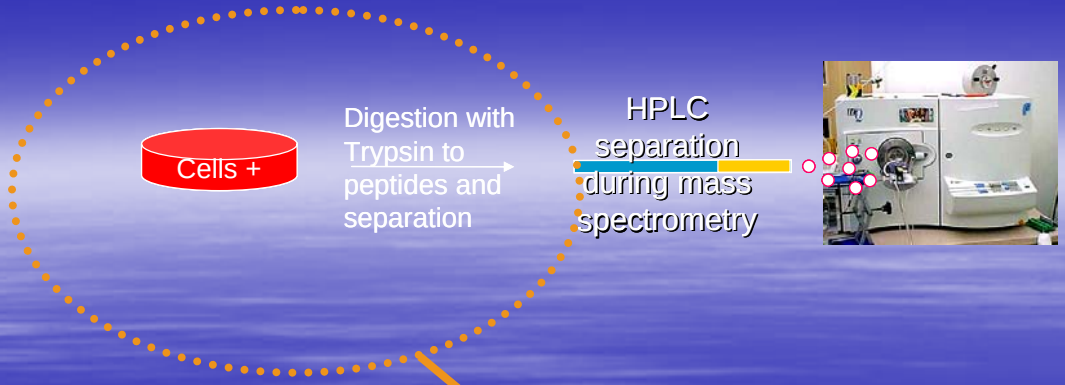
Cells +

Digestion with Trypsin to peptides and separation

HPLC separation during mass spectrometry

Mass spectrometry identifies proteins

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	12.38	5.06	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	5.48	1.10	4	M
SEC24B	Protein SAR1	5.39	3.65	1	C
KLK2	Glandular kallikrein 2 precursor	5.34	3.45	6	S
JAG1	Jagged 1 precursor	4.92	1.43	23	M
APLP2	Amyloid-like protein 2 precursor	4.17	2.06	8	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
EPHA1	Ephrin A1 receptor	2.97	0.41	4	M
[illegible]	[illegible]	2.95	1.00	2	M
IL6R	Interleukin-6 receptor alpha chain precursor	2.84	0.72	3	M
C10rf8	Protein C10rf8 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	0.46	2	M
MLP	MARCKS-related protein	2.38	0.98	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	2.24	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	0.48	10	S
CI-M/P	Cation-independent mannose-6-phosphate receptor	2.20	0.50	16	M
TNFRSF10	TNF receptor superfamily member 10D precursor	2.19	0.16	3	M
PLXNB1	Plexin-B1/SEP receptor precursor	2.18	0.51	2	M
SEMA4A	Semaphorin 4A precursor	2.18	0.40	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



Protein, MW = 10,000 +

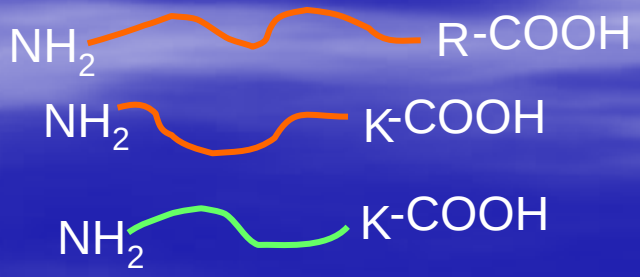
digest into peptides

Trypsin

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

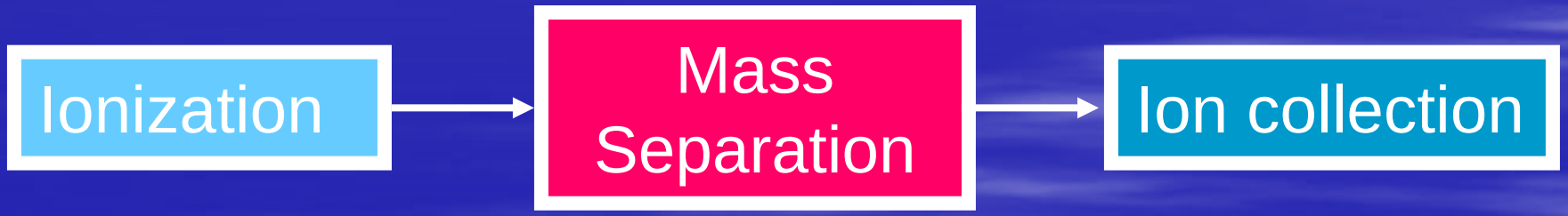
cleaves C-teminal side of arginine (R) and lysine (K)

Peptides,
MW < 4,000



Mass Spectrometry Primer

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

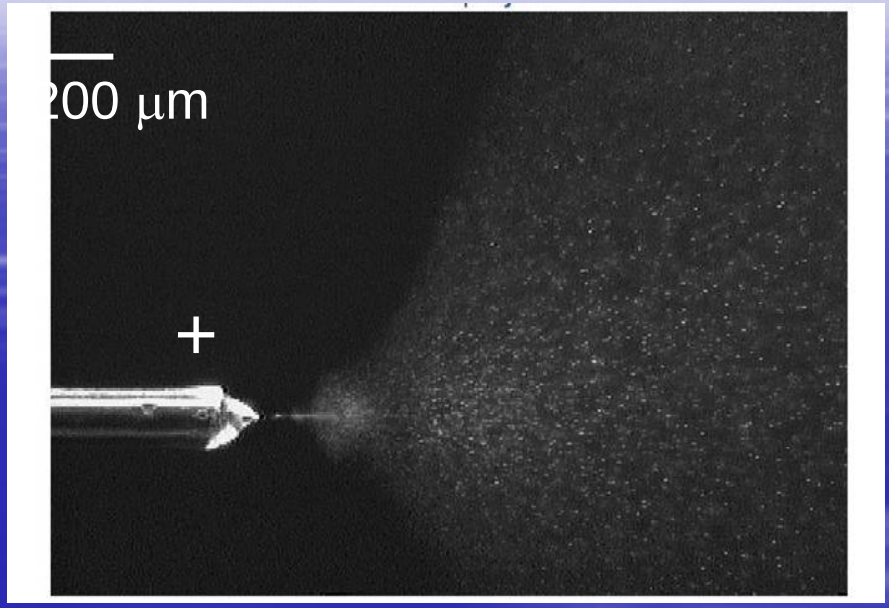


- MALDI
- Electrospray

- quadrupole
- ion trap
- time-of-flight

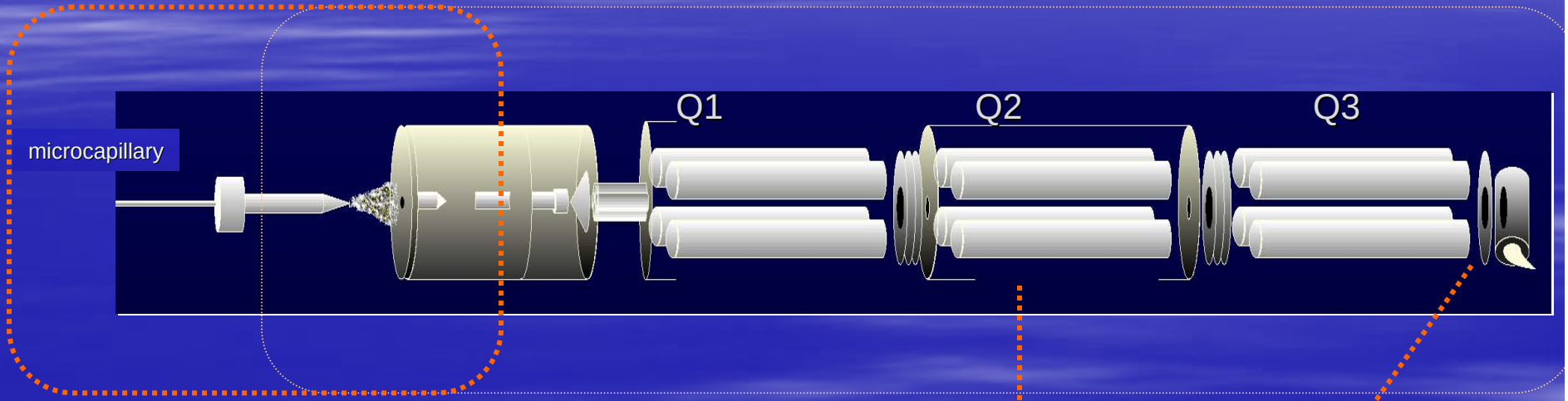
- mass analysis

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

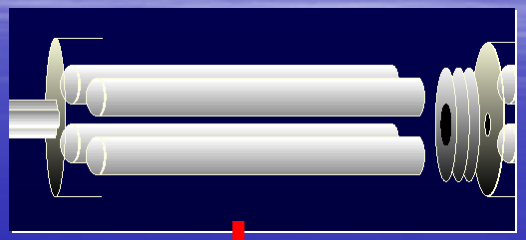


electrospray source
Ionization

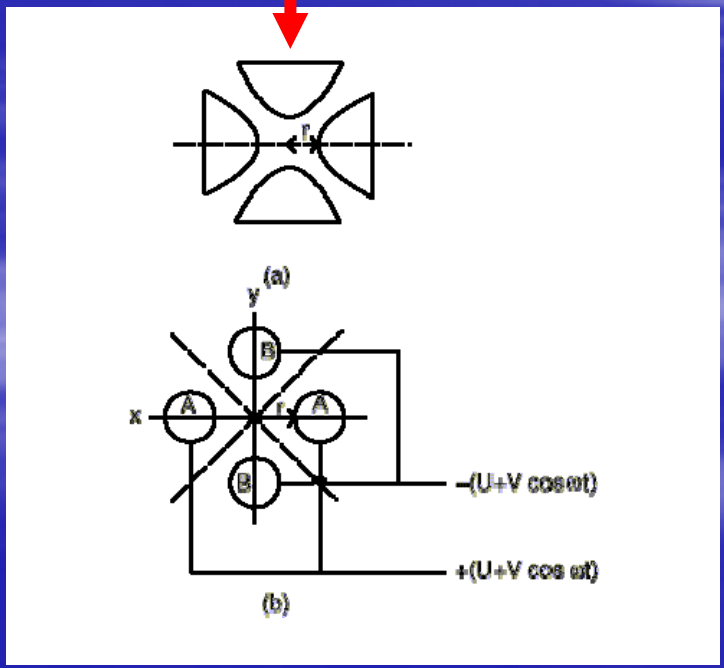
collision cell
Mass Selection

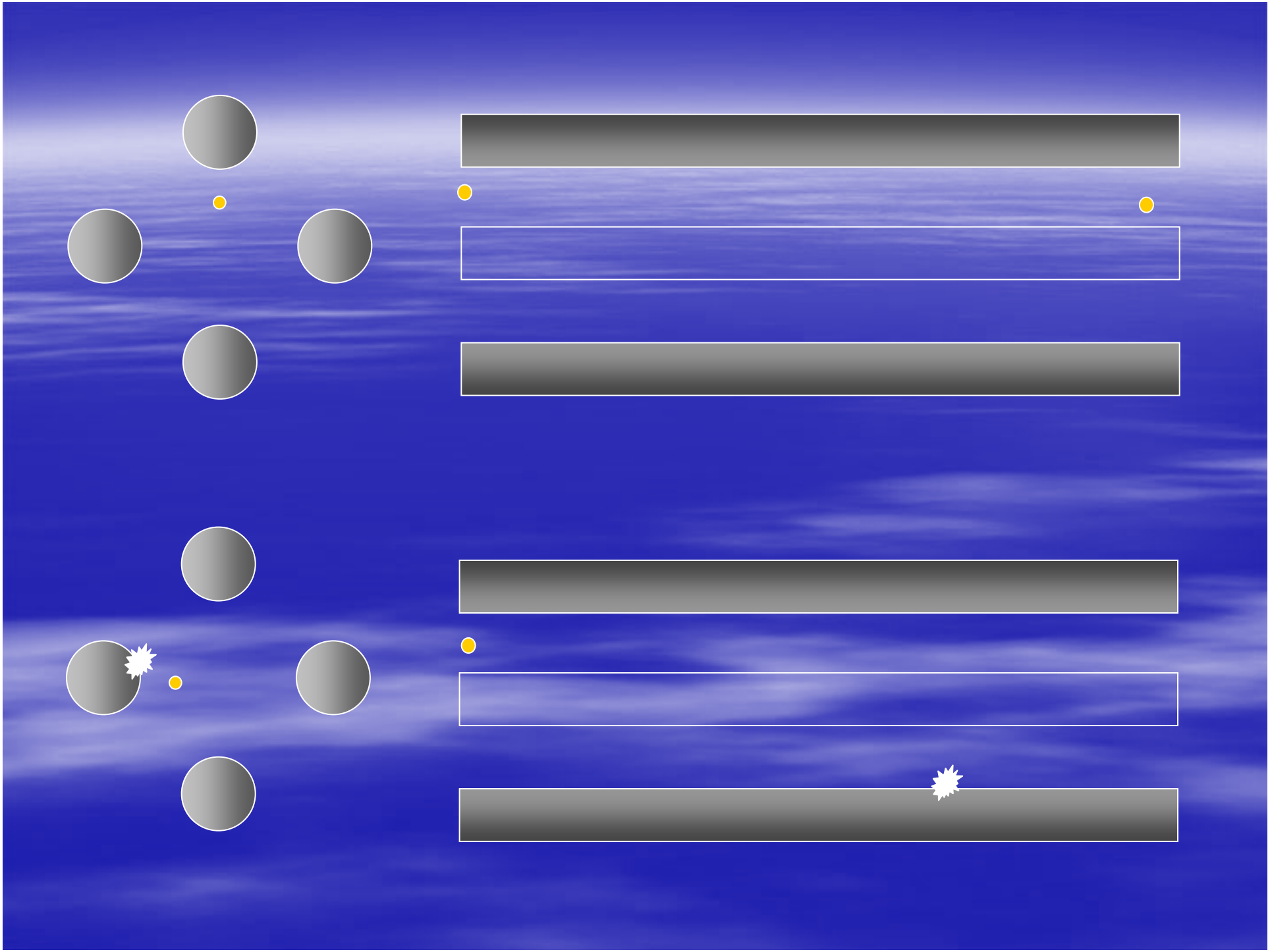
Detector
Ion Collection

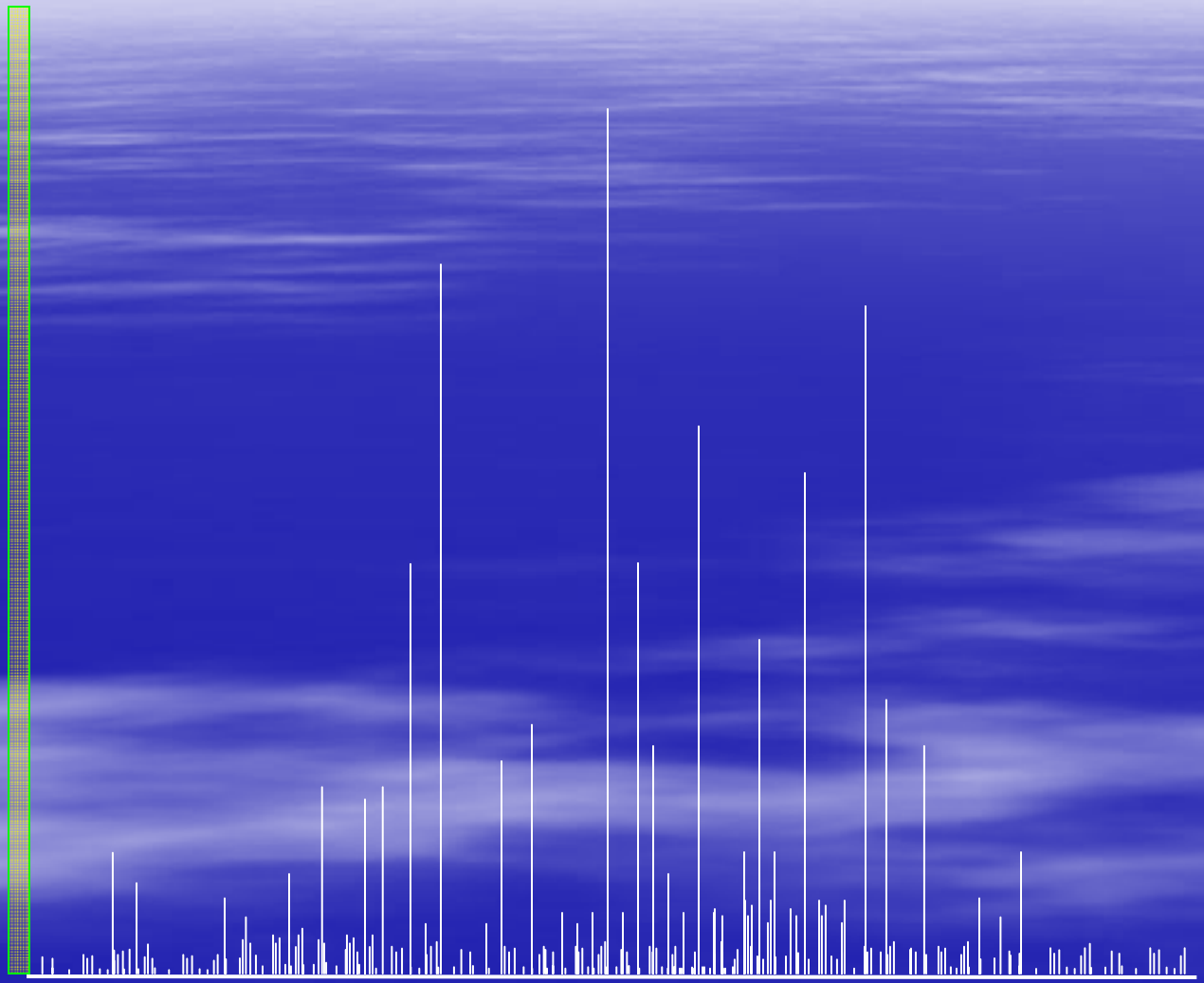
Quadrupole Optics



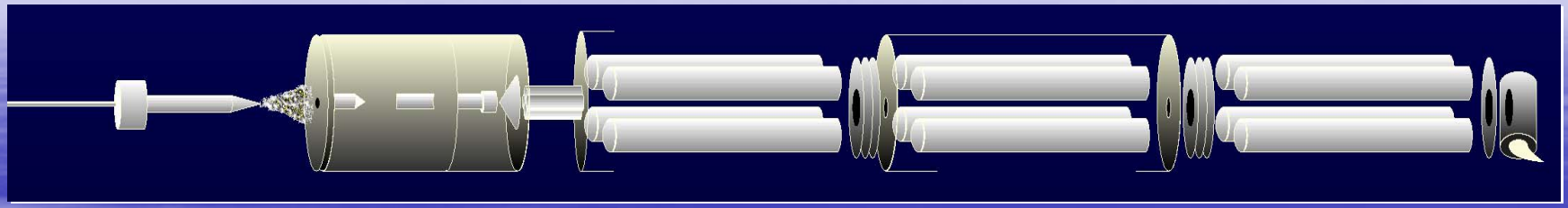
- In a quadrupole mass spectrometer four (quad) parallel rods (poles) are arranged equidistantly from a central (imaginary) axis.
- Charged ions are injected along the central axis of the quadrupole assembly.
- Static and alternating (radio frequency) electric potentials are applied to opposite pairs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



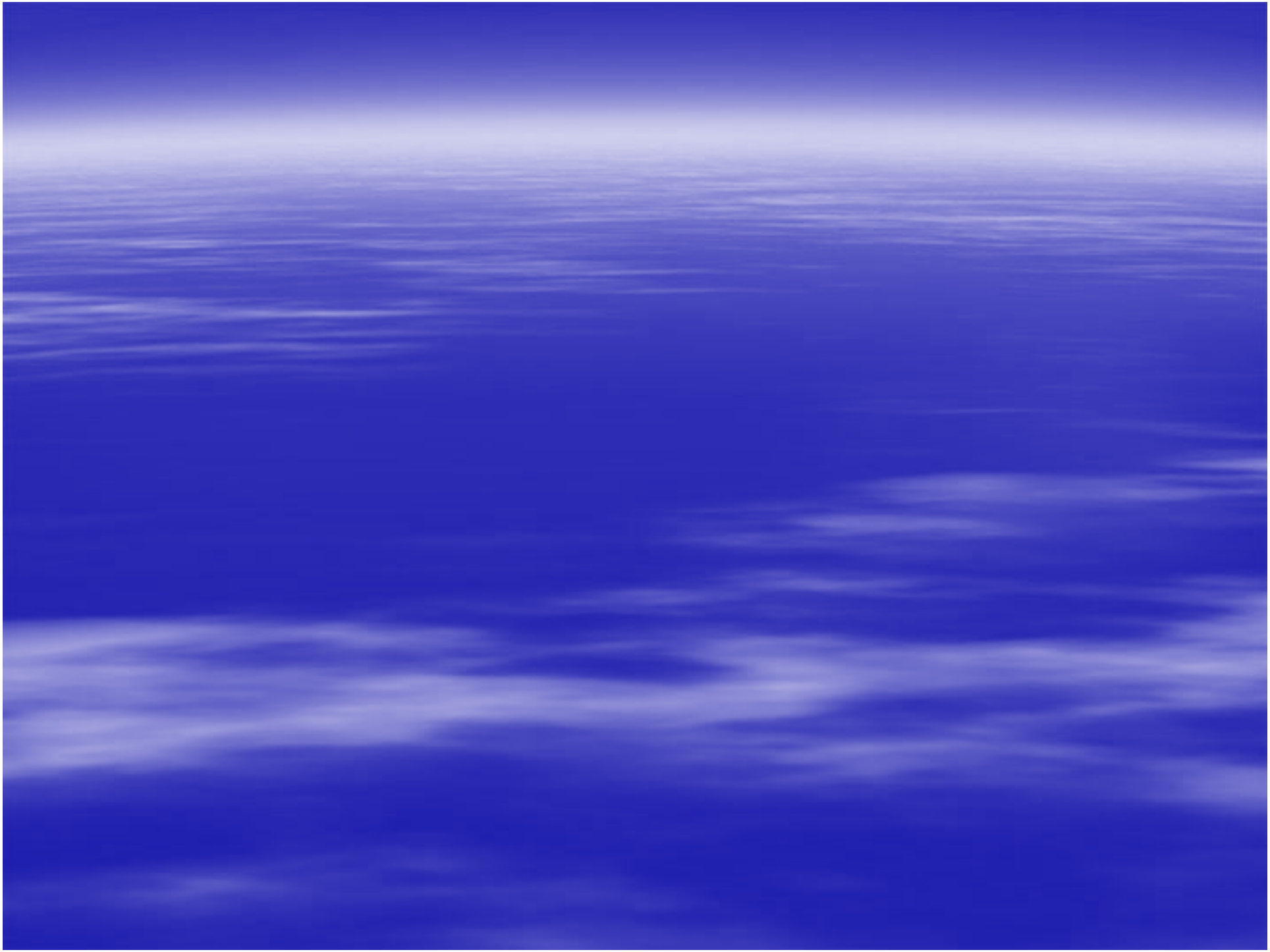



m/z

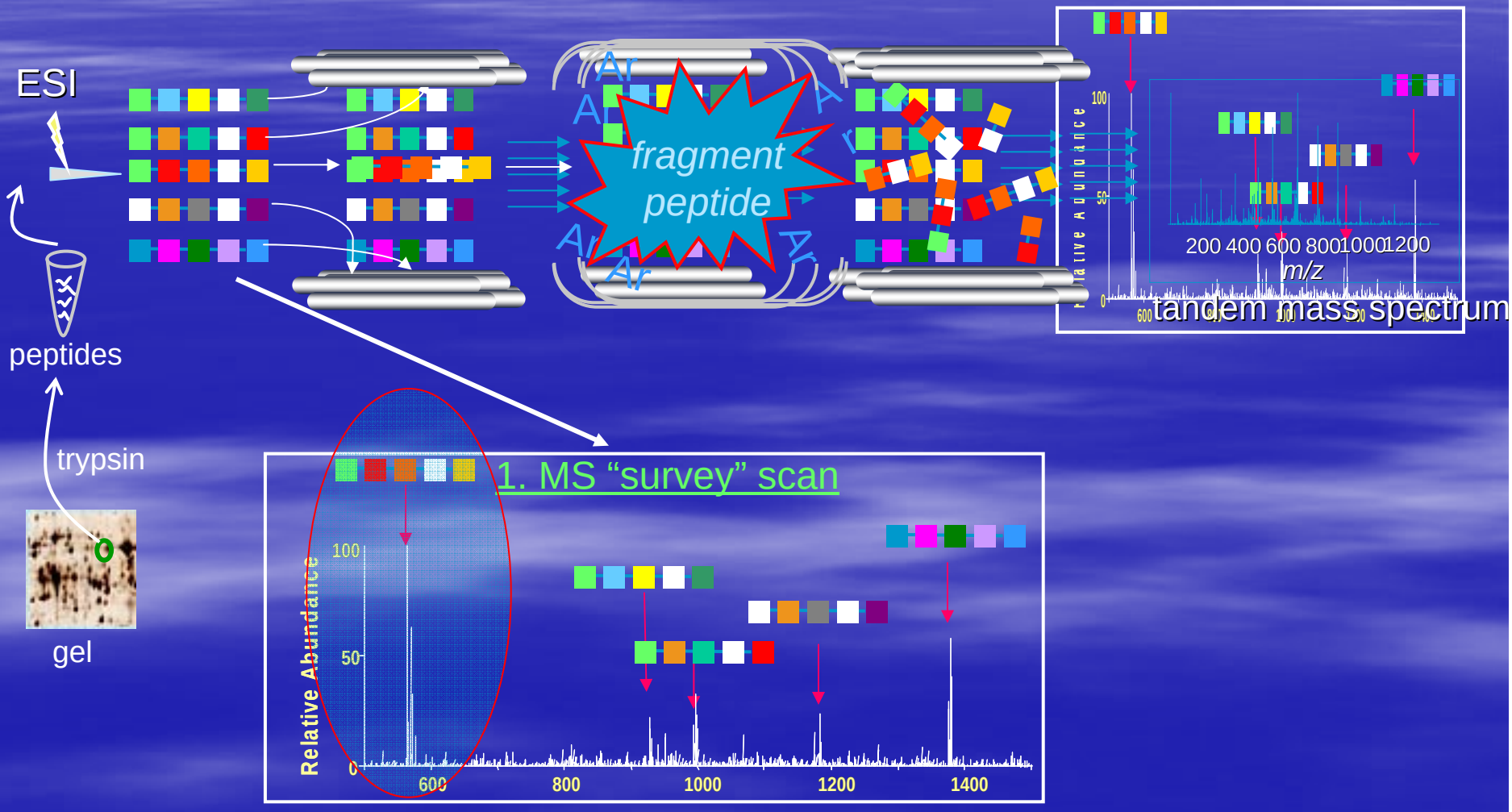
Quadrupole Optics cont.



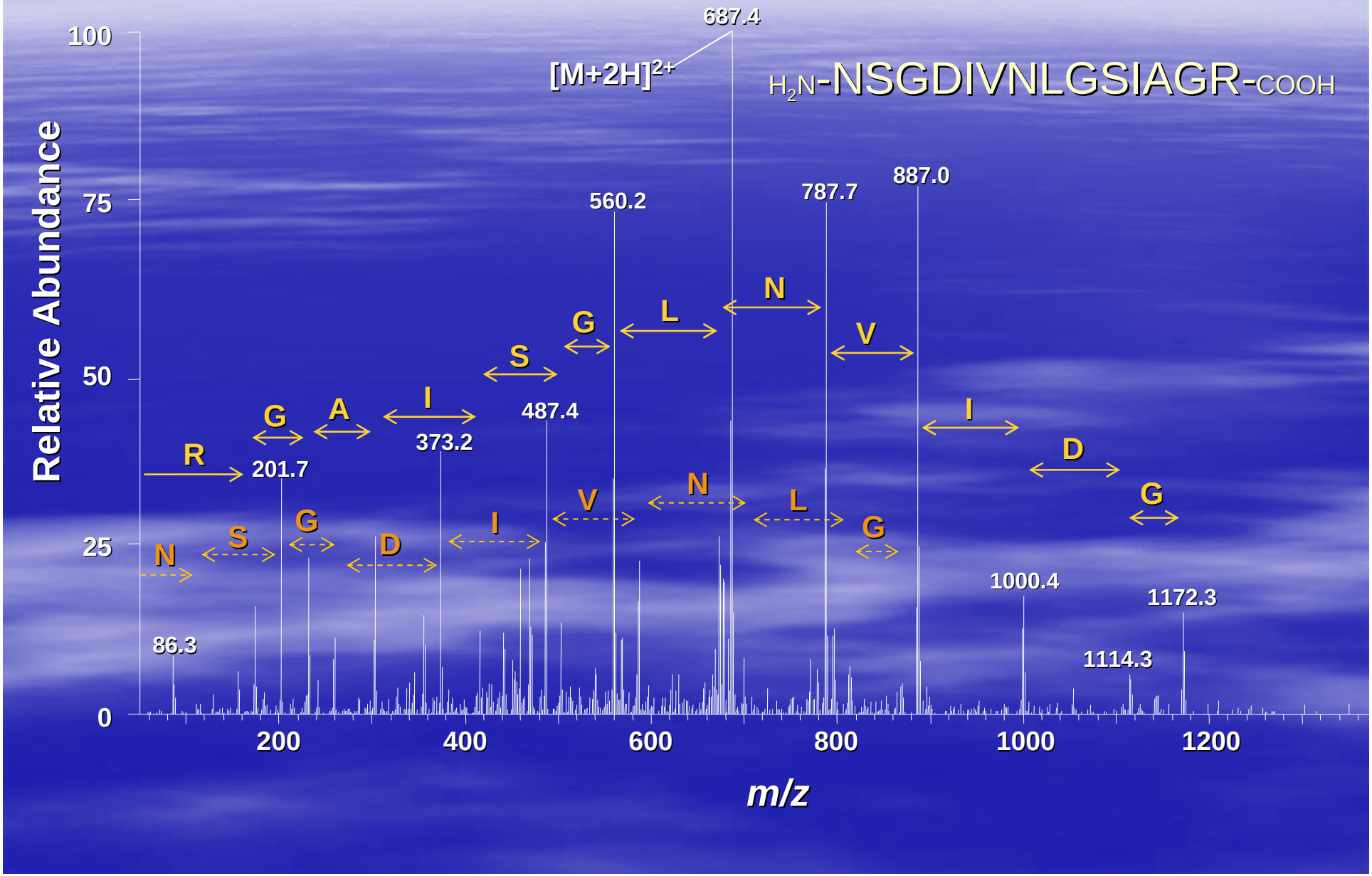
- The quadrupole can function in a second mode called tandem mass spectrometry or MS/MS.
- A particular peak is chosen from a MS scan and the first quad allows only that m/z to pass into the second quad.
- The second quad accelerates the species through a voltage causing collisions with an inert gas 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



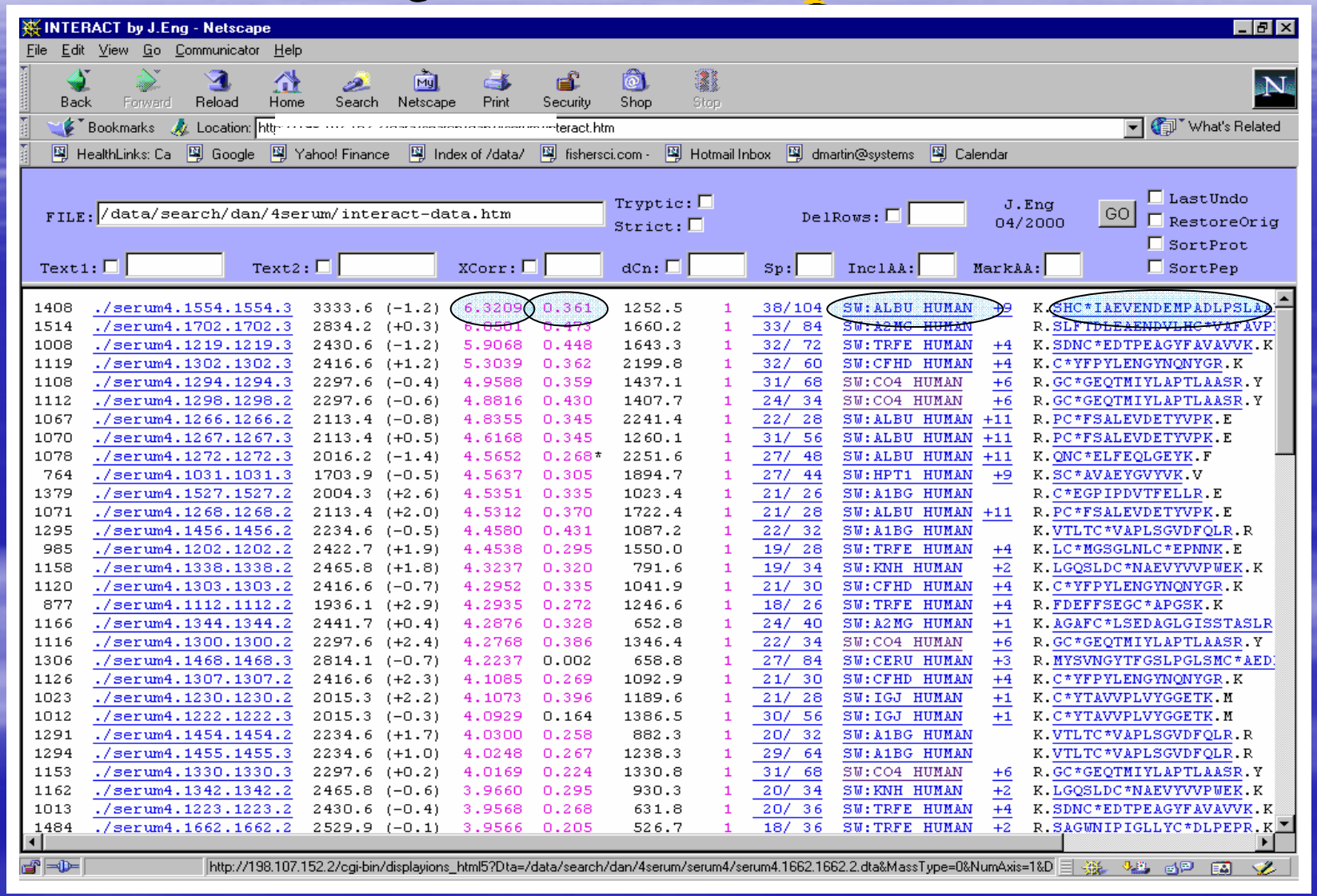
Making an identification by database searching using SEQUEST

- 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



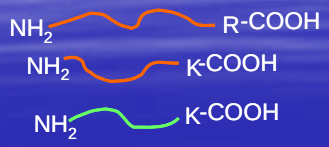
Mixture analysis – the need for separations

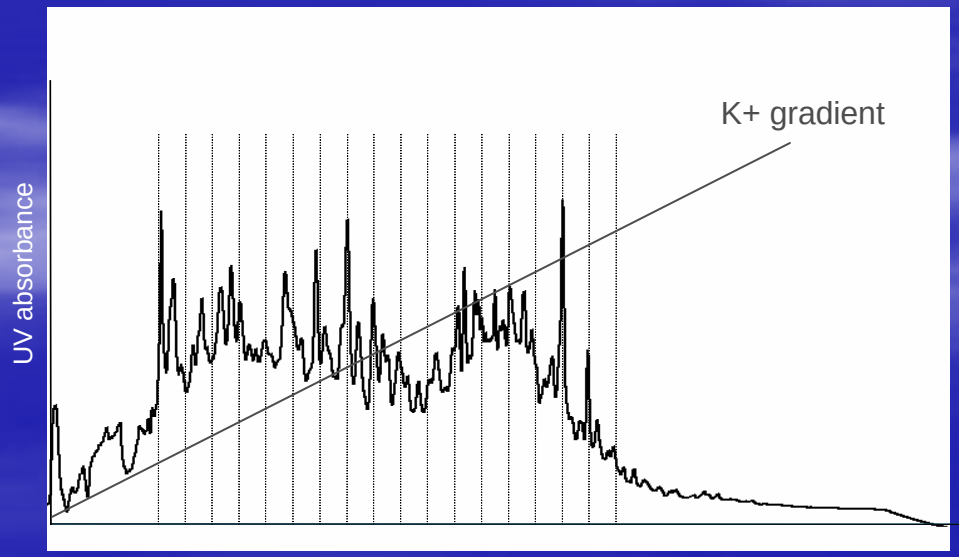
- Mass spectrometers have limited peak capacities requiring separation and fractionation prior to analysis

- 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





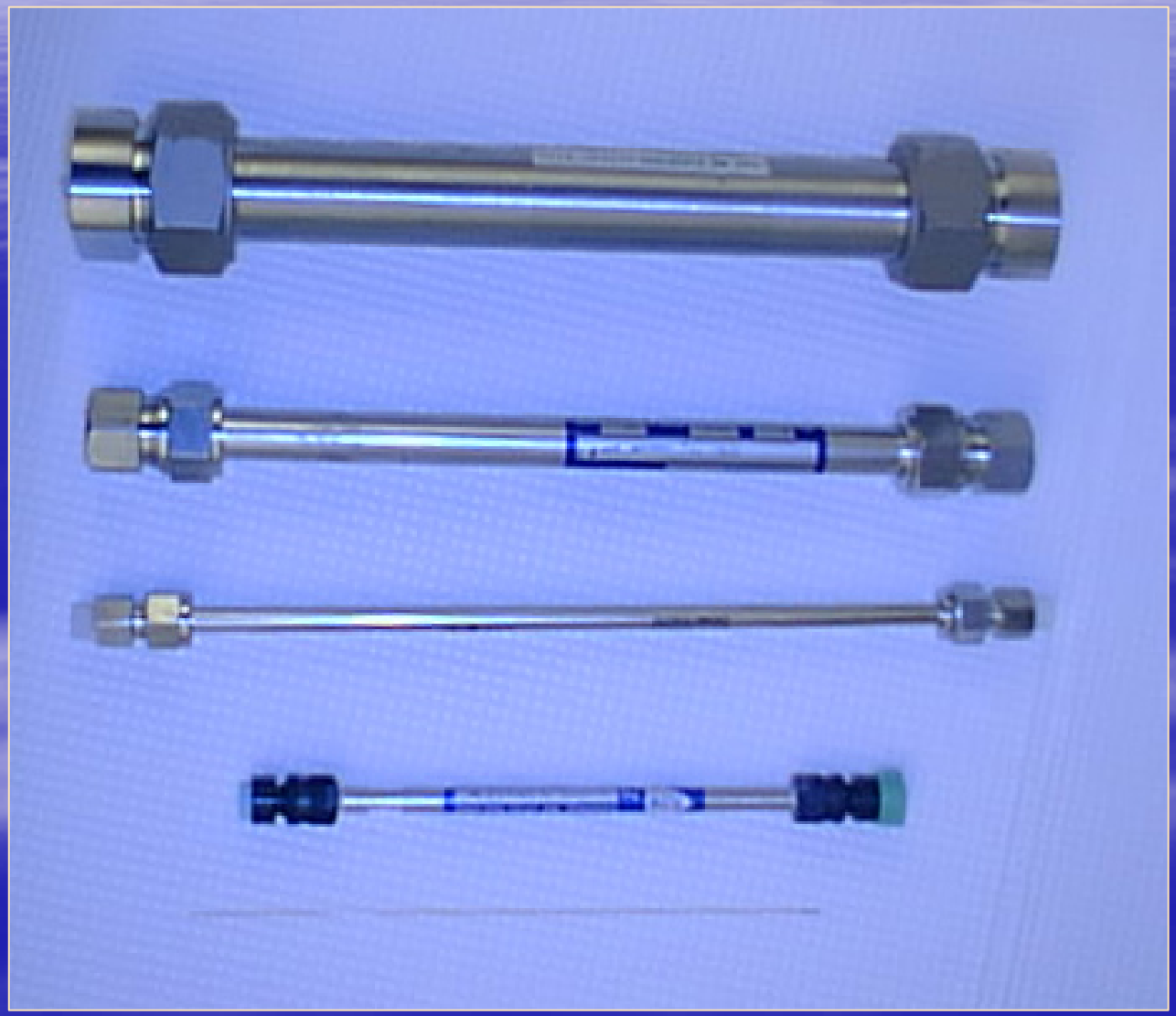
Cells +

Digestion with Trypsin to peptides and separation

HPLC separation during mass spectrometry

Liquid chromatography miniaturization for ESI-MS

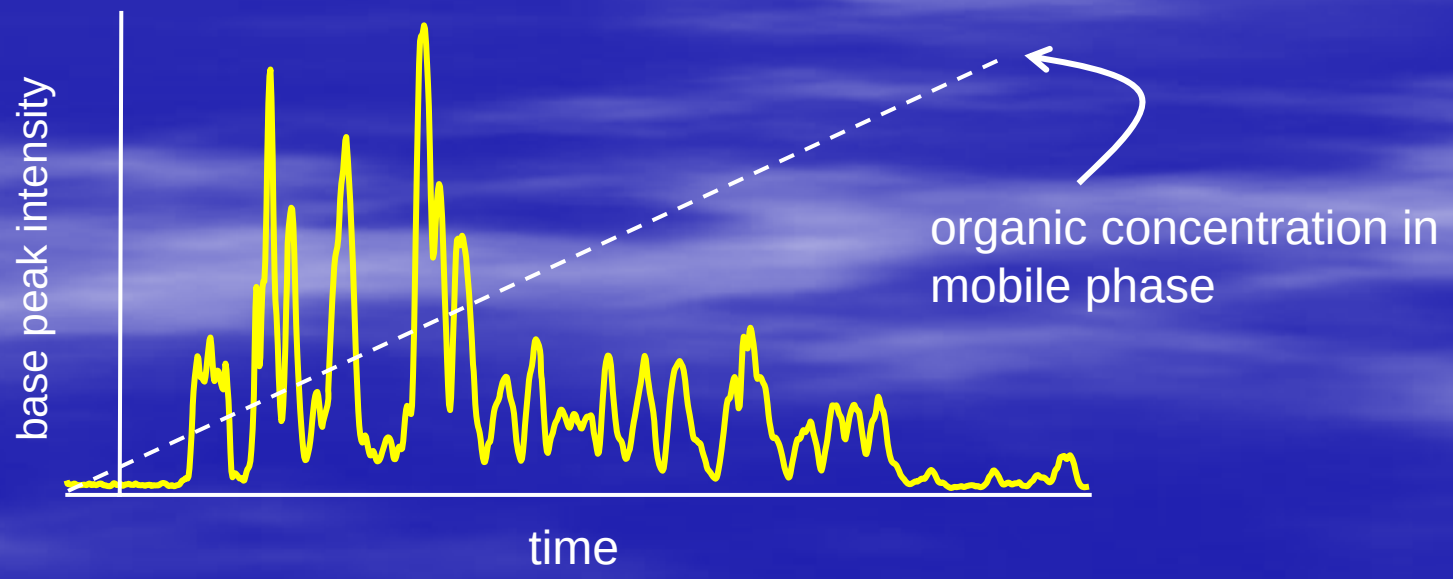
decrease size, increase sensitivity



Size ↓	Sensitivity ↑
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

Cells +

Digestion with Trypsin to peptides and separation

HPLC separation during mass spectrometry

Mass spectrometry identifies proteins

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	12.38	5.06	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	5.48	1.10	4	M
SECTM1	Prostate-RA87-1152; unnamed protein	5.39	3.65	1	S
KLK2	Glandular kallikrein 2 precursor	5.34	3.45	8	S
JAG1	Jagged 1 precursor	4.92	1.43	23	M
APLP2	Amyloid-like protein 2 precursor	4.17	2.06	8	M
SPINT1	Kunitz-type protease inhibitor 1 precursor	3.89	1.08	22	S
LRIG1	Membrane glycoprotein LIG-1	3.55	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
EPHA1	Ephrin A1 receptor	2.97	0.41	4	M
[illegible]	[illegible]	2.95	1.06	2	M
IL6R	Interleukin-6 receptor alpha chain precursor	2.84	0.72	3	M
C10rf8	Protein C10rf8 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	0.46	2	M
MLP	MARCKS-related protein	2.30	0.98	4	C
B2M	Beta-2-microglobulin precursor	2.24	0.48	14	S
NRP1	Neuropilin-1 precursor	2.23	0.36	27	M
NEO1	Neogenin precursor	2.27	0.03	5	M
PSAP	Proactivator polypeptide precursor, contains: Saposin A	2.24	0.33	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	0.48	10	S
CI-M/IGF2R	Cation-independent mannose-6-phosphate receptor	2.20	0.50	16	M
TNFRSF10	TNF receptor superfamily member 10D precursor	2.19	0.16	3	M
PLXNB1	Plexin-B1/SEP receptor precursor	2.18	0.51	2	M
SEMA4A	Semaphorin 4A precursor	2.18	0.40	4	M
VLDLR	very low-density lipoprotein receptor precursor	2.09	0.09	1	M
GALNT1	Polypeptide N-acetylgalactosaminyltransferase	2.09	0.52	3	GOL

Now all elements of the system make sense

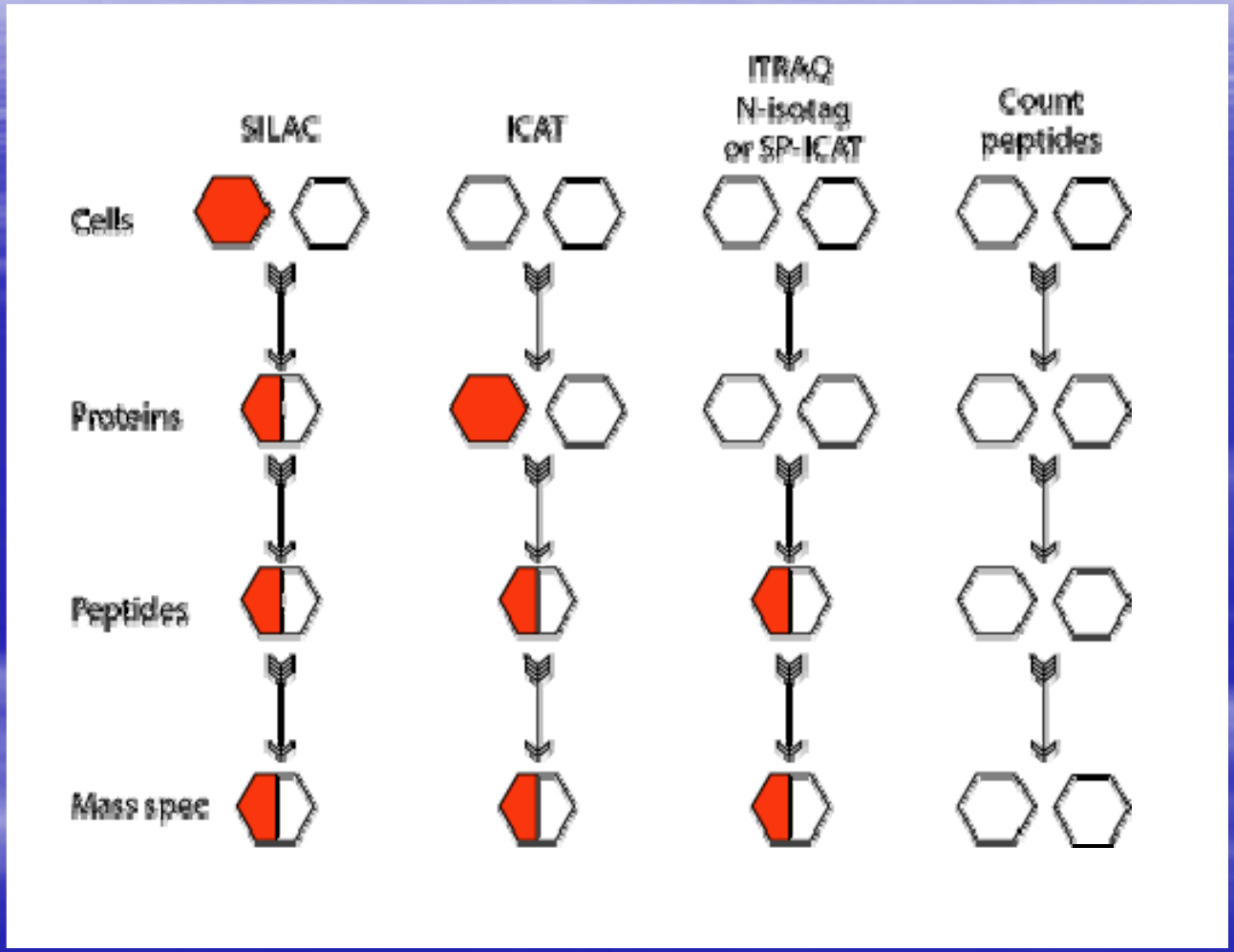
Quantitative Proteomics

Using stable heavy isotopes to perform quantification

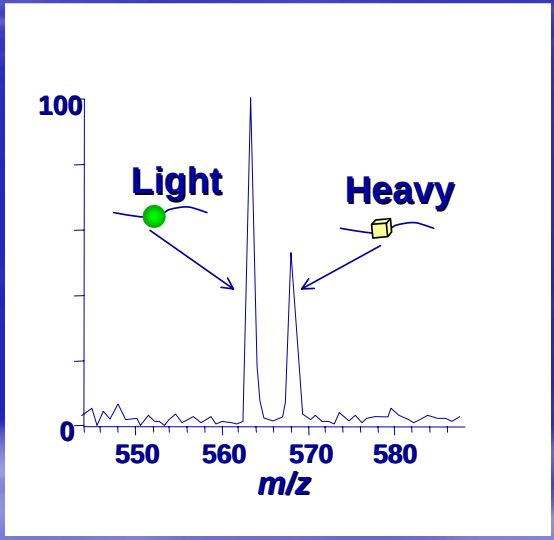
- Because a mass spectrometer measures mass, *isotopic modifications can allow one to distinguish two samples.*
- This is done using stable isotopes such as N15, C13 and Deuterium.
- Isotopes can be introduced at various stages of an experiment

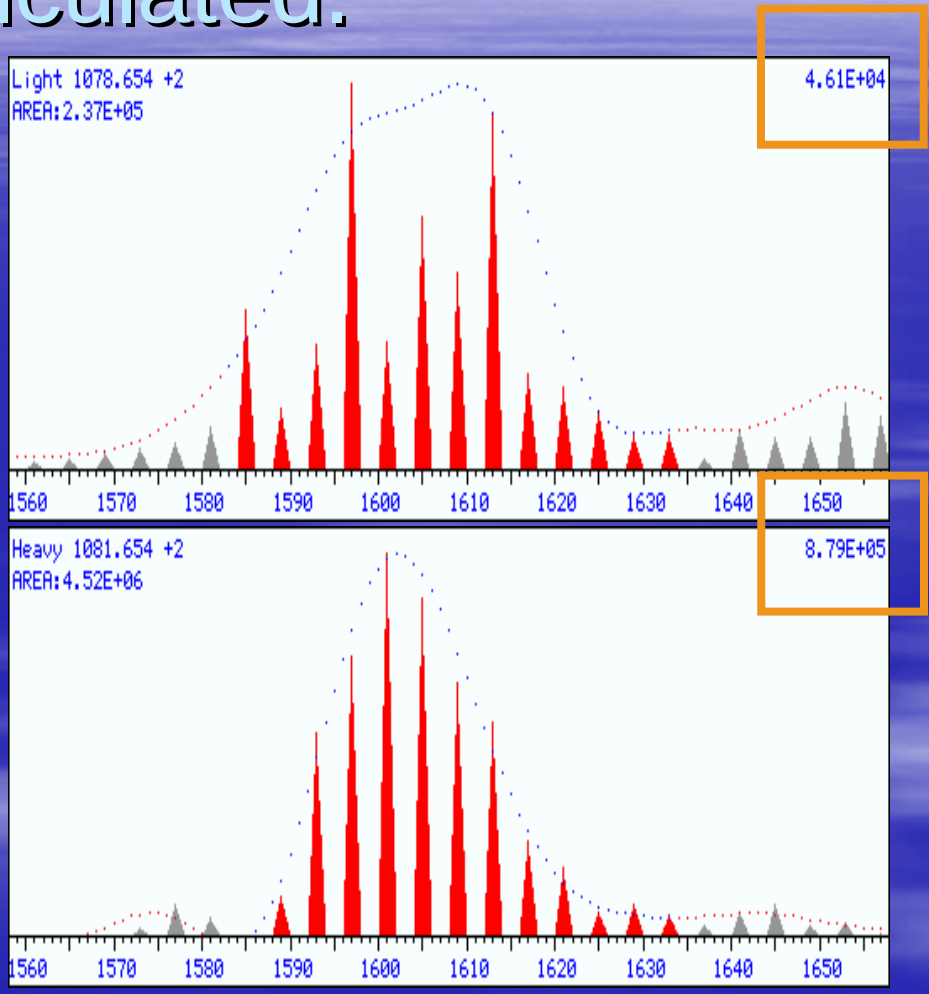
Methods of Stable Isotope Incorporation

- 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


SILAC
ICAT
ITRAQ
N-isotag
or SP-ICAT
Count
peptides
Cells
Proteins
Peptides
Mass spec

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

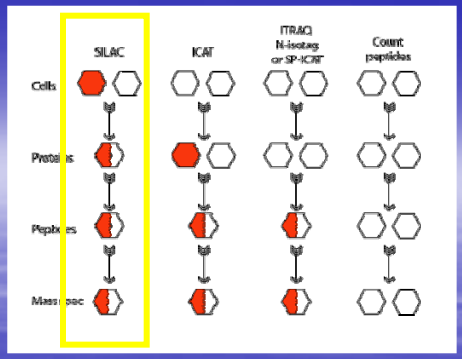




Light : Heavy

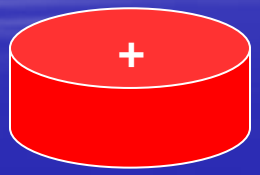
1 : 19.07

SILAC

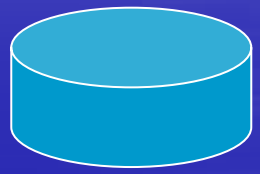


Media w/o Lys and Arg
FCS dialyzed to remove amino acids

Heavy Lysine, Arginine or both



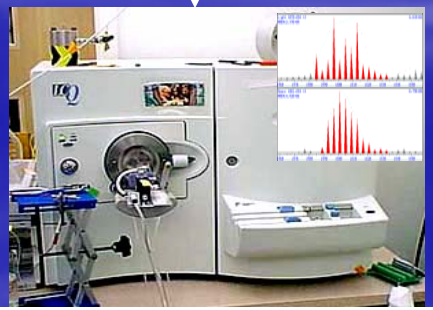
Stimulus of interest



Unmodified amino acids

Control

Lyse cells together

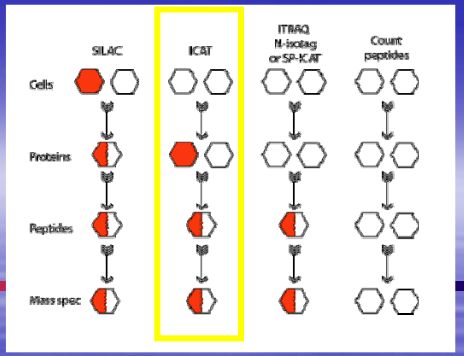


Simple

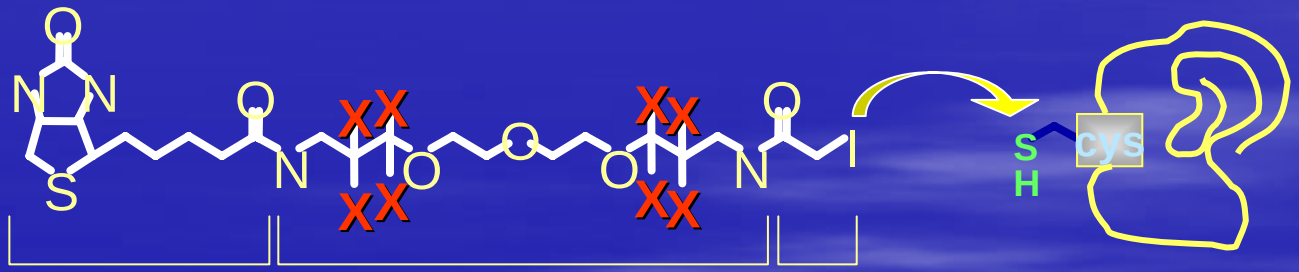
Straightforward

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)

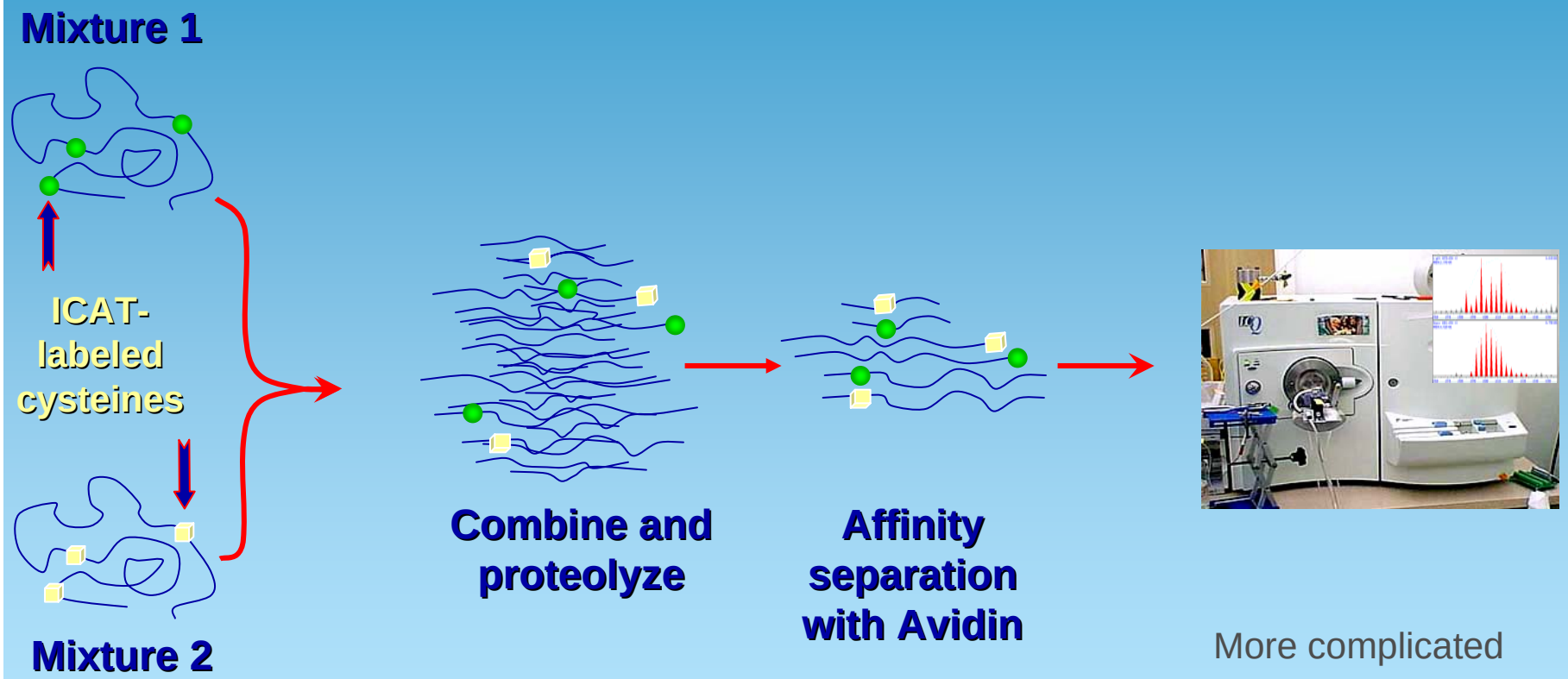


Biotin affinity tag

Linker (heavy or light)

Thiol specific reactive group forms a **covalent** bond with cysteine

Protein Quantification and Identification by the ICAT Strategy



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

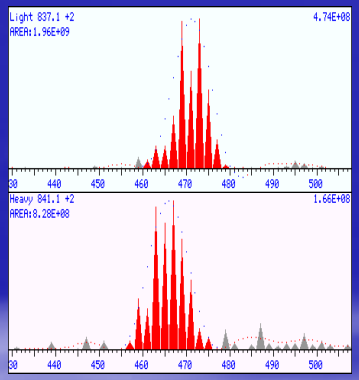
Big project example:

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
TNFRSF10I	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
TNFRSF10I	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

JCB: ARTICLE

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

Marcello Marelli,[1] Jennifer J. Smith,[1] Sunhee Jung,[1] Eugene Yi,[1] Alexey I. Nesvizhskii,[1] Rowan H. Christmas,[1] Ramsay A. Saleem,[1] Yuen Yi C. Tam,[2] Andrei Fagarasanu,[2] David R. Goodlett,[1] Ruedi Aebersold,[1] Richard A. Rachubinski,[2] and John D. Aitchison[1,2]

[1]Institute for Systems Biology, Seattle, WA 98103
[2]Department of Cell Biology, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

We 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 proteins, 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 Pex25p. 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 organellar protein content is still a formidable undertaking. The polydispersity within organelle classes resulting from biological diversity and the limited resolving power of subfractionation 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, immunoisolation, and free flow electrophoresis, have been applied to reduce contaminants (for review see Brunet 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 true constituents 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

The online version of this article includes supplemental material.

Correspondence to John D. Aitchison: jaitchison@systemsbiology.org

Abbreviations used in this paper: AP, affinity-purified peroxisomal membrane; DsRed, *Discosoma* sp. red fluorescent protein; ICAT, isotope-coded affinity tag; MS, mass spectrometry; μLC/ESI-MS/MS, microcapillary liquid chromatography/electrospray ionization tandem MS; PTS, peroxisomal targeting signal; RPP, nonaxenic DsRed; SGD, Saccharomyces Genome Database.

© The Rockefeller University Press $8.00
The Journal of Cell Biology, Vol. 167, No. 6, December 26, 2004 1099–1112
http://www.jcb.org/cgi/doi/10.1083/jcb.200408119

Supplemental Material can be found at:
http://www.jcb.org/cgi/content/full/jcb.200408119/DC1

THE JOURNAL OF CELL BIOLOGY

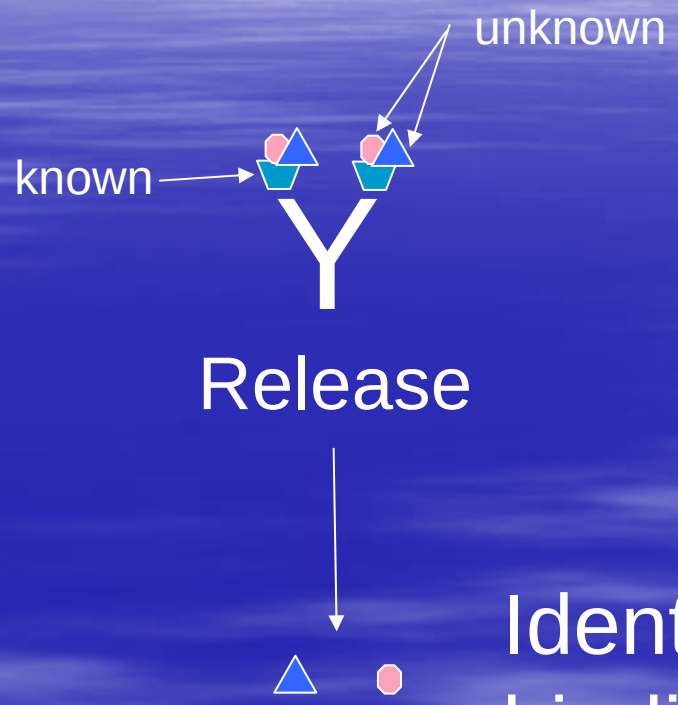
"We have combined classical subcellular fractionation with large-scale quantitative mass spectrometry to identify proteins that enrich specifically with peroxisomes of Saccharomyces cerevisiae."

Micro Proteomics

Micro-Proteomics

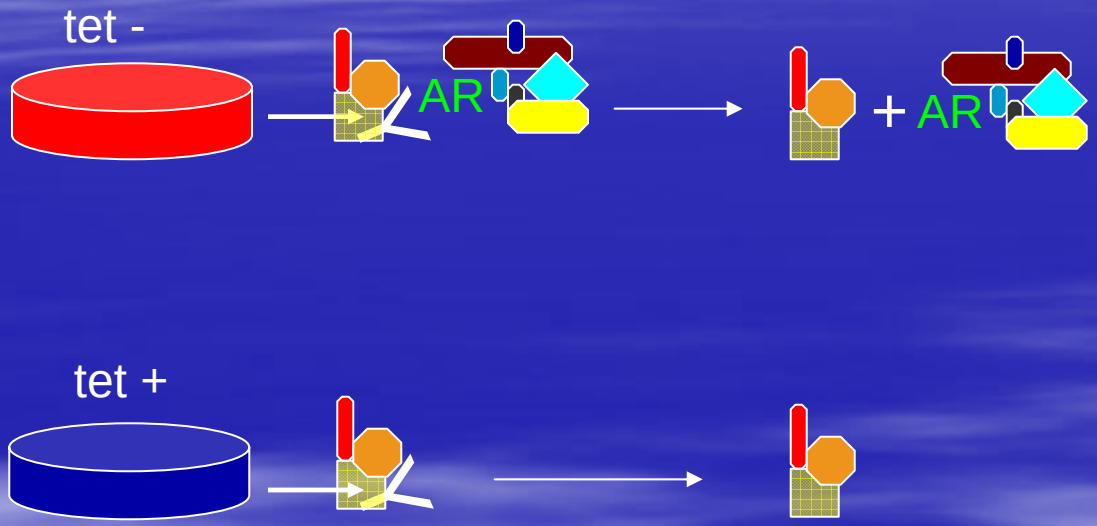
- Focus on a particular protein complex
- Affinity based purification
- Takes advantage of existing technology
- Results more approachable for an individual investigator

Micro-Proteomics



Identification of binding partners

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