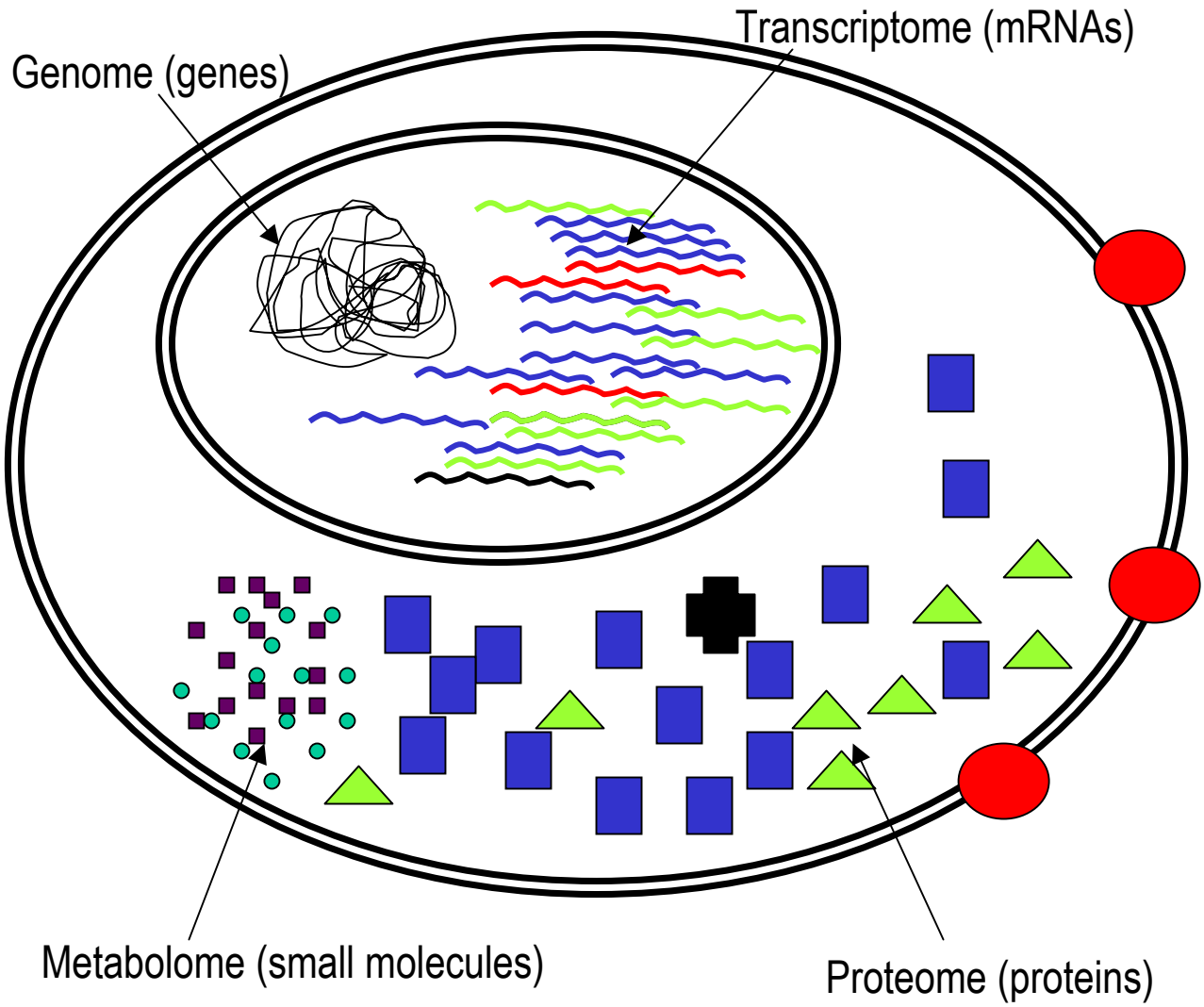
BIOC4004 - Industrial Biochemistry
Lecture 19 - Mon Mar 15, 04

Topics for the Day:

- Proteomics
- 2-D gels revisited
- Proteome profiling
  - the old way
  - ICAT
- Mass Spectrometry
  - MALDI-TOF
  - ESI

What's going on in a (eukaryotic) Cell ?



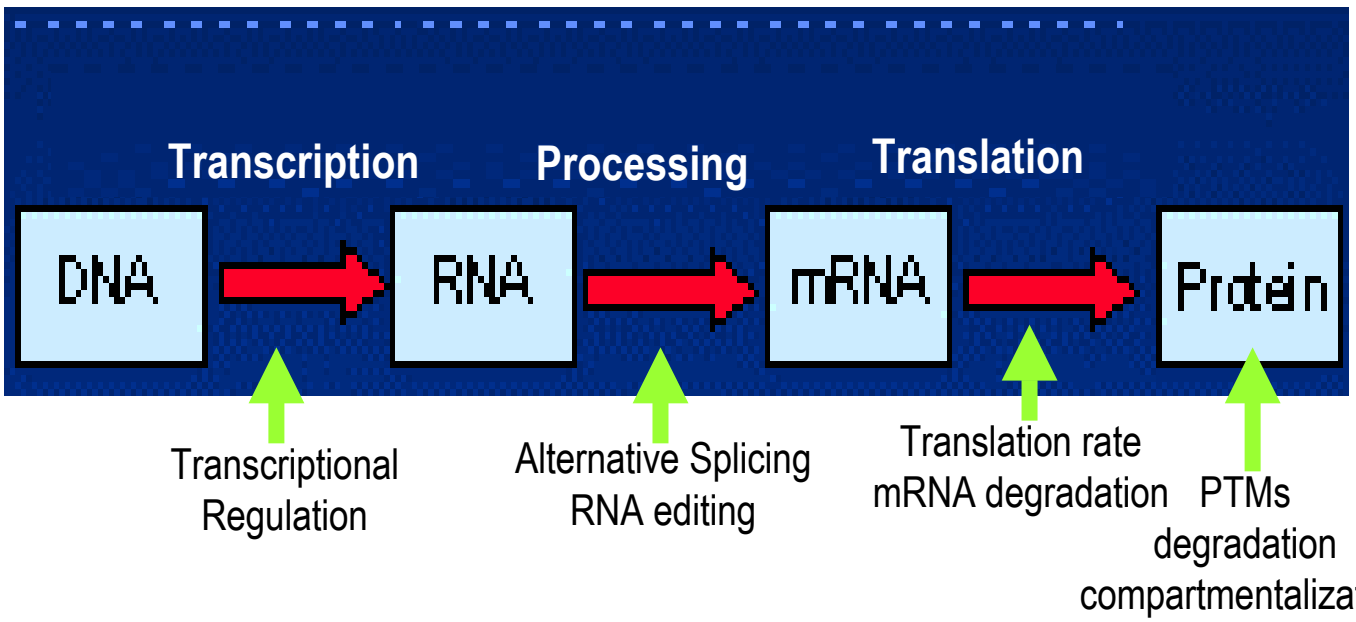
- A cell is just a bag full of the by-products of "life processes"
- "Life processes" are dynamic; by-products are in state of flux
- By studying and quantifying these by-products, we can gain a better understanding of "life processes"
    - What's going on ?
    - When is it happening ?
    - How much of it is happening ?
    - How does it differ from other cell-types ?
    - What happens when these processes go awry ?

What's is Proteomics ?

- the study of a cell's entire protein content
    - identification of protein diversity
        - alternative splicing
        - post-translational modifications
    - quantification of protein levels (when possible)
    - comparison of proteomes between different cell types (?)

Why Proteomics ?

- Ultimately, proteins are the ones doing the work !!!!
    - Protein content give insight into cell processes
    - We need to study them **directly**
- Many different factors play into the protein content of a cell
    - many different levels of regulation
    - pre- and post-translational factors
- There isn't necessarily a correlation between mRNA levels and protein levels

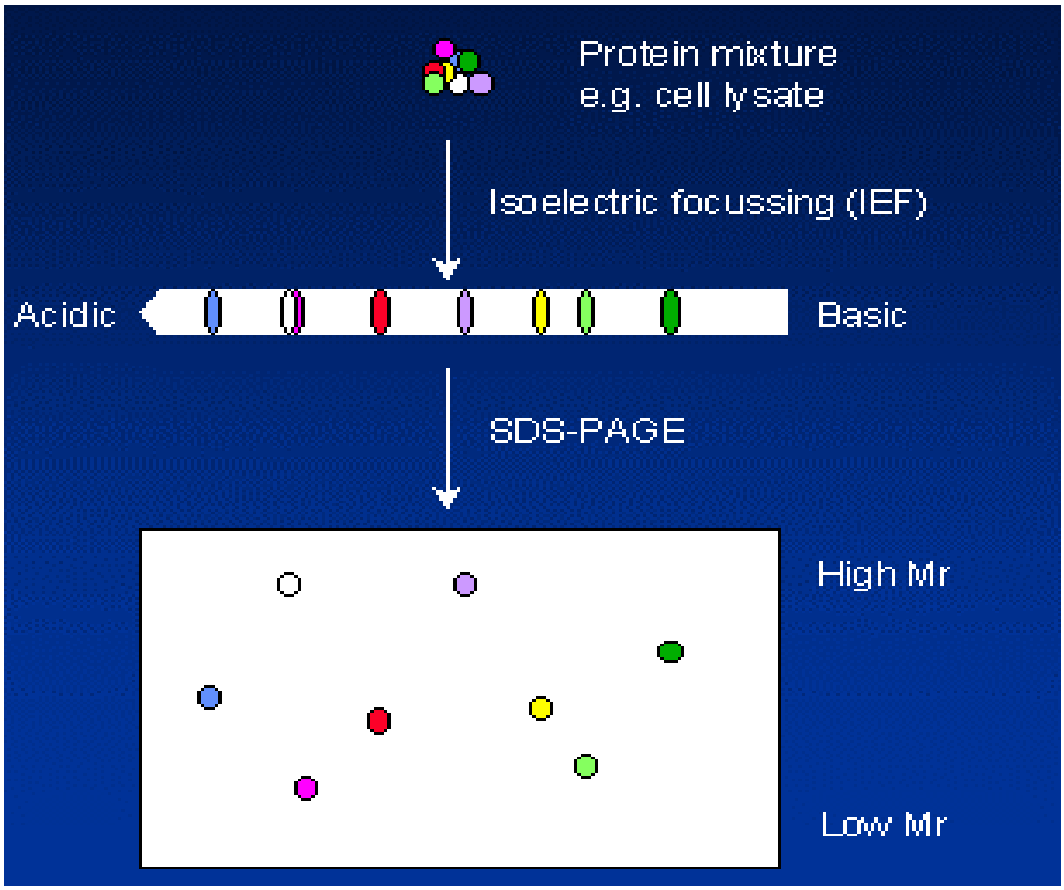


Proteomics Requirements

If we want to study a cell's protein content (proteome) we need:

- a means to extract proteins from the cells of interest
- a means to resolve the protein mixture into individual components
- a means to identify (and quantify) the individual components

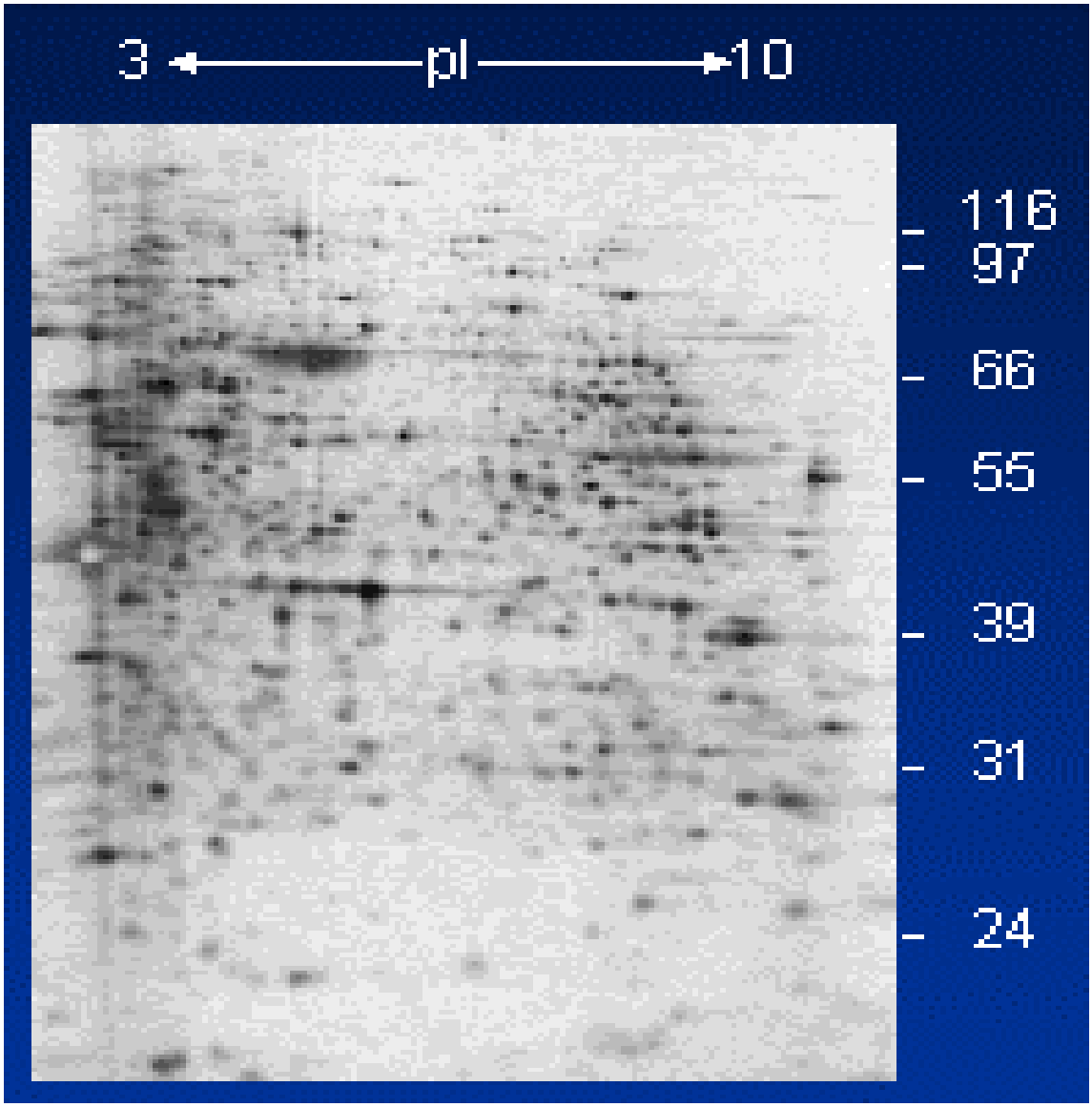
Proteomics and 2-D gel electrophoresis



2-D gel electrophoresis is a perfect protein separation technique for proteomics research:

- perform cell lysis & protein extraction
- 1st dimension: separation according to pI
- 2nd dimension: separation according to MW

Real 2-D gel electrophoresis data....



- gels are fixed and stained (colloidal coomassie blue, silver stain, SYPRO ruby)
- digitized image produced by scanning laser densitometry

- 2-D data is nice and messy !!!!
  - PTMs lead to streaky spots
    - (eg. phosphorylation changes protein charge and size)

- Reproducibility is a challenge, but good labs can (generally) attain it

Real 2-D gel electrophoresis problems

Resolution & Sensitivity of Detection

- 2-D gels can resolve ~ a few thousand different proteins
- A typical cell can produce tens of thousands of proteins
- Some proteins are produced in **very tiny** amounts

The more proteins you want to look at a time, the worse the resolution and detection

The Solution: get more selective !!!!!!!

Protein Resolution

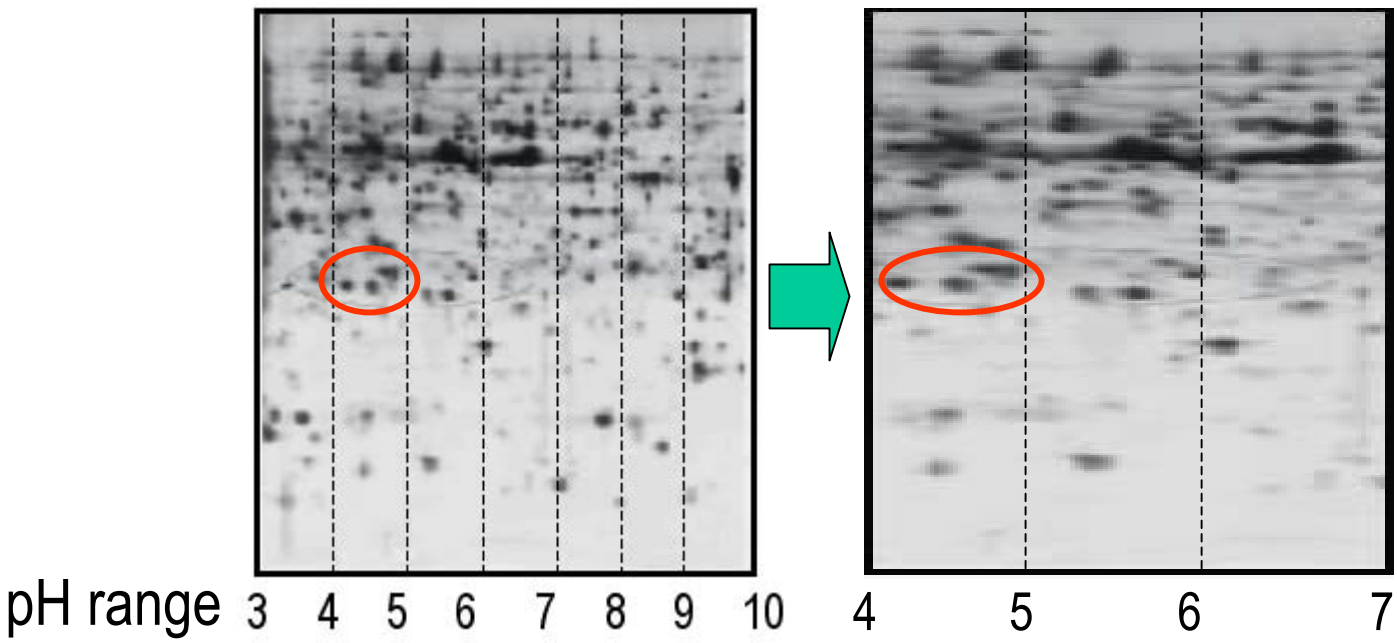
- Sample fractionation
  - different extraction techniques to target specific types of proteins
    - eg. phosphoproteins can be targeted
  - cell fractionation techniques to target different cell compartments
    - eg. cytosol vs membrane vs nucleus vs mitochondria
- Narrow range IEF
  - use IEF with narrower pH range so you can focus on a narrower set of proteins --> better resolution

Protein Detection

- staining (silver, (fluorescent) dyes)
- autoradiography (metabolic labeling)
- fluorescent labeling
- Lectins (glycosylated proteins)
- Antibodies

Narrow Range IEF

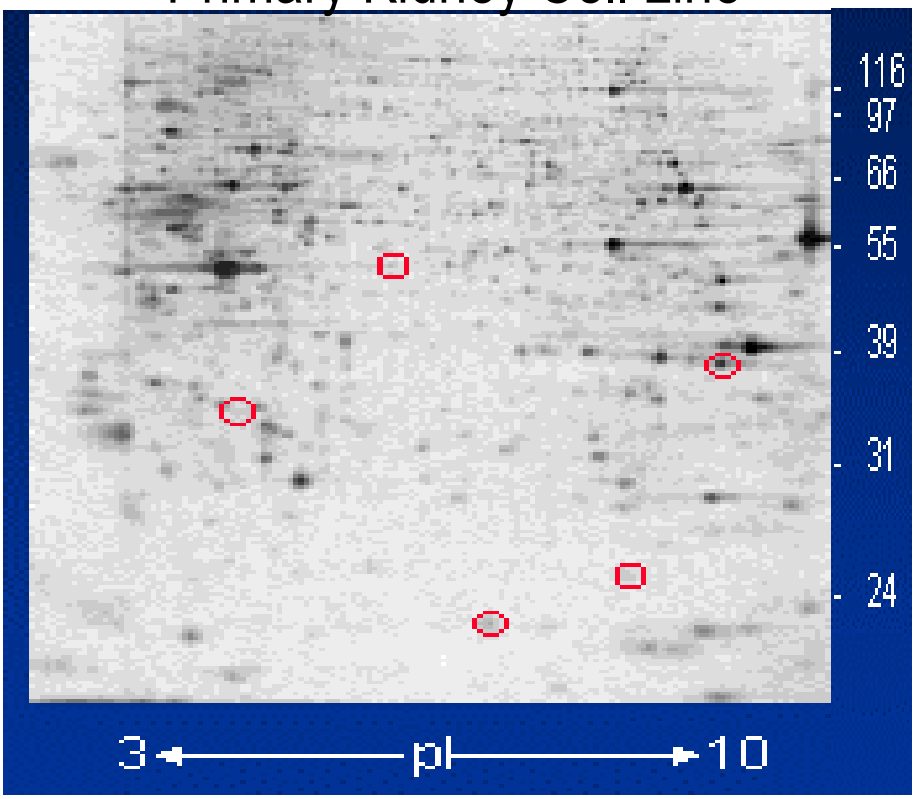
Let's say we're interested in proteins that have a pI of ~ 5



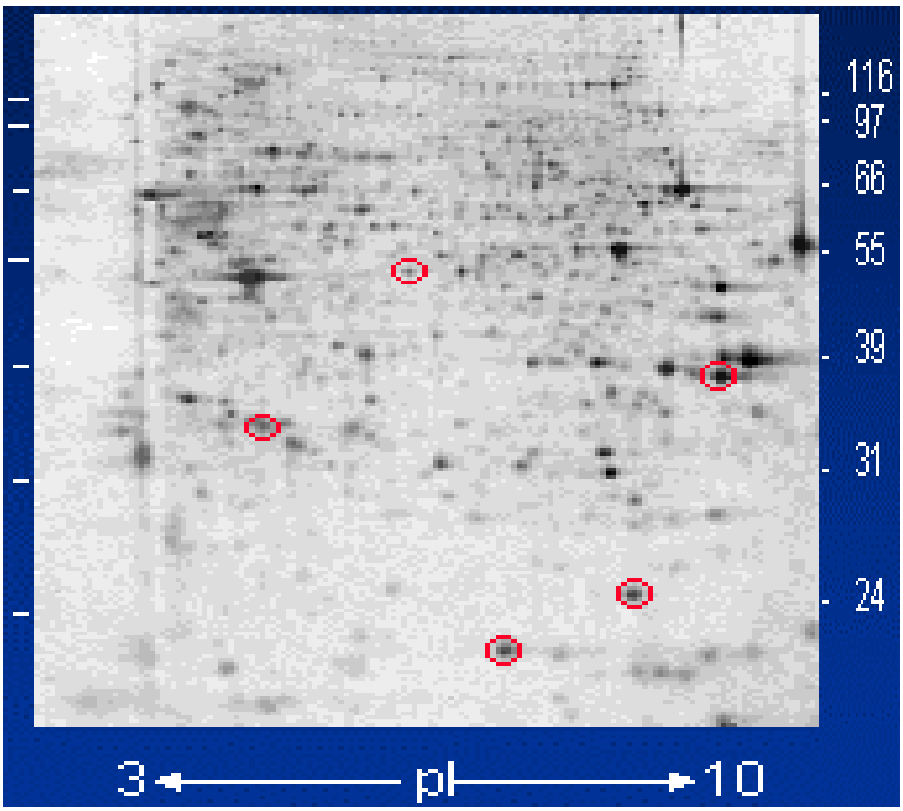
- by using a narrow range IEF separation, we zoom in on the area of interest
- there is a practical limitation on the amount of protein that can be resolved in the 2nd dimension
- the proteins within the range of interest “spread out" along the 1st dimension; all others migrate to cathode or anode
    - better resolution (same "gel size" used to separate a narrower band of proteins)
    - better detection (load more protein at the start)
        - most of the proteins won't remain in IEF gel

What we're really interested in: Proteome profiling

Primary Kidney Cell Line

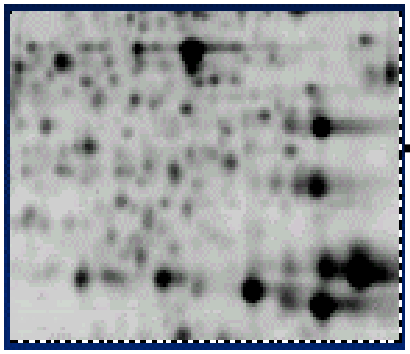


Renal Cell Carcinoma

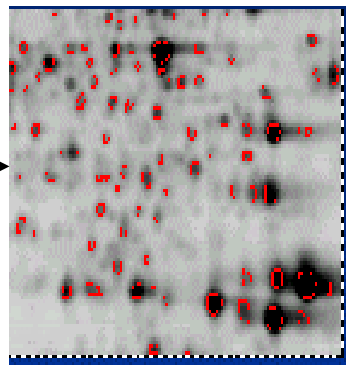


- We can compare proteomic profiles of different cell types and find out fundamental differences between them
- Need to compare 2-D gels
- Need to make sure gels are "comparable"
  - always a challenge
- If we can identify differentially expressed proteins, we can go on to determine their cellular role
  - finite time, need to focus on the "good stuff"

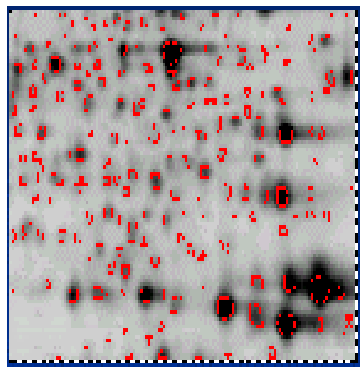
Proteome profiling: thank god for computers !!! Sort of...



Automated spot detection

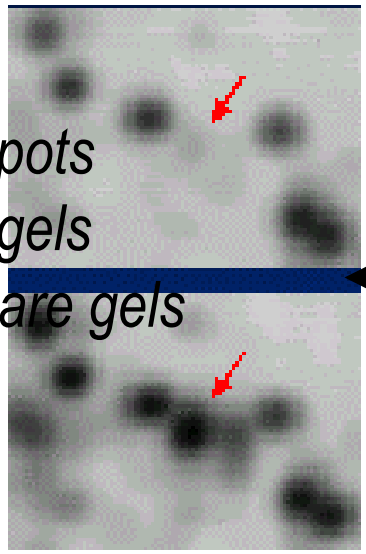


- too stringent
- missing spots



- not stringent enough
- too many "fake" spots

- *find spots*
- *align gels*
- *compare gels*



You would like to see this result across a few replicates

- make sure it is a biologically relevant difference
- not a technical error or a biological quirk

Next: Identifying a spot !!!!!

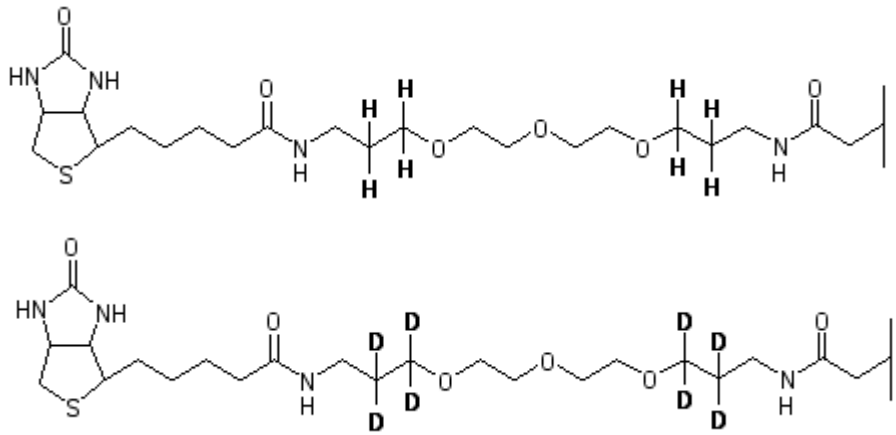
- You have several pieces of information regarding the protein:
  - MW
  - pI
  - tissue, organism
  - cellular localization ?
- Run against Dbase of 2-D PAGE proteome maps (@ EXPASY)
  - Swiss 2-D PAGE Dbase
- What if you're working on something not on the Dbase ?

2-D gel comparisons are problematic…

- *lots of experimental noise makes the comparisons difficult*
- *ideally, run both samples on the same gel*
- *can't run the two samples on the same gel…or can you ?*

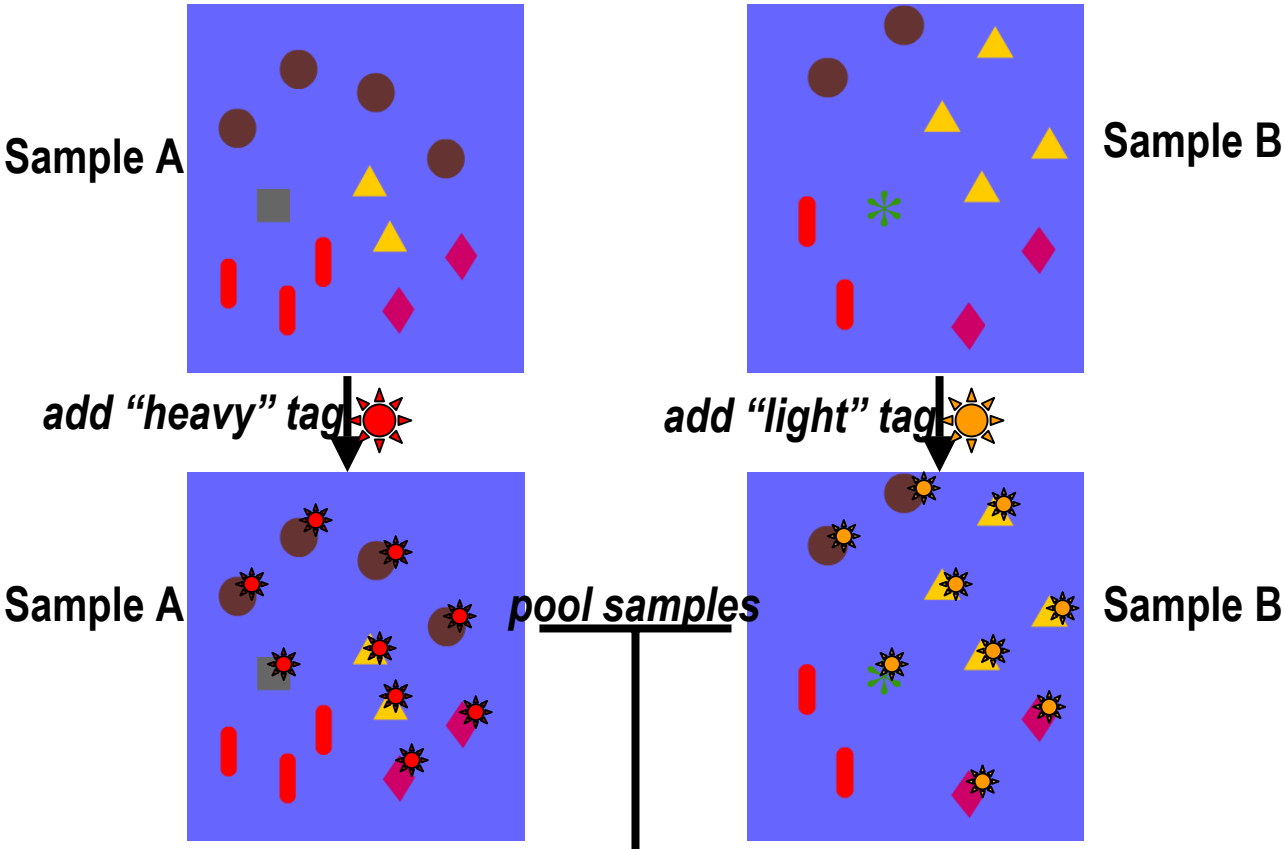
ICAT (Isotope-Coded Affinity Tags)

Gygi *et al., 1999. Nature Biotechnology 17: 994 - 999*

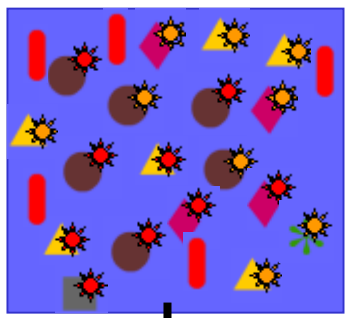


- "Heavy" and "Light" tags: defined MW difference
- tags can be coupled to proteins on Cysteine residues
- tags are biotinylated and can be captured using avidin magbeads or avidin affinity matrix

Proteome profiling using ICAT



"light" and "heavy" peptide pairs are of interest



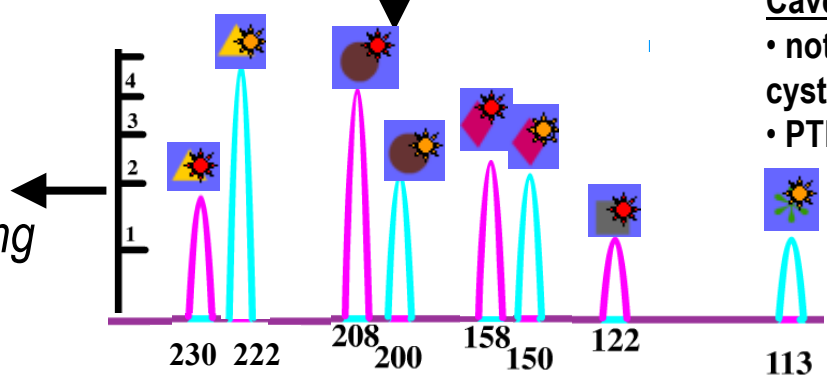
- Protease digestion
  - peptides
- Affinity capture of tagged peptides
- Separation & "quantitation" of peptides using LC
- ID of "peptide pairs" using MS

Caveats:

- **not all proteins label; cysteine rare**
- **PTMs can screw things up**

why ?

MS sequencing

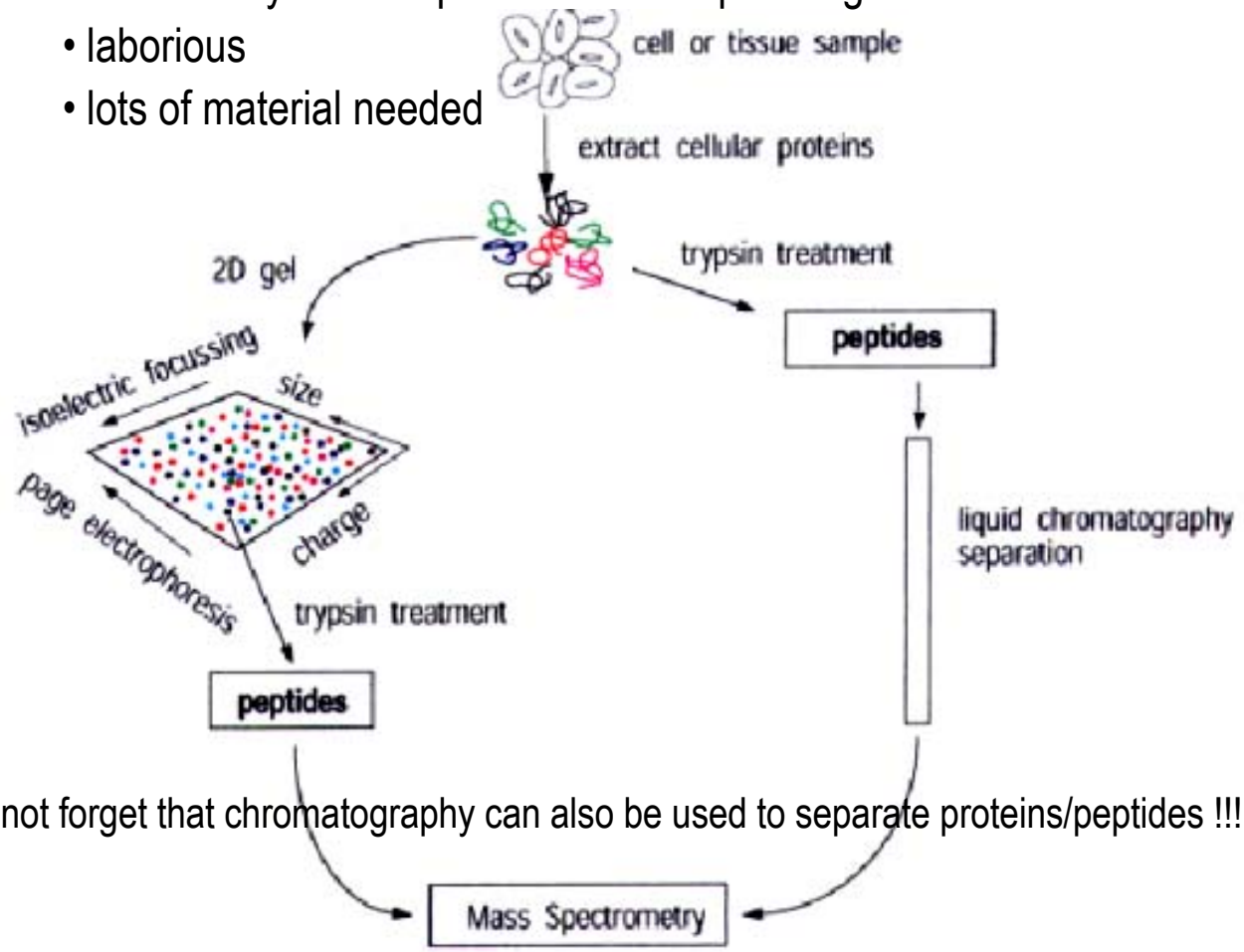


Mass Spectrometry and Proteomics

• Need to go from spot (or band) on a gel (or a chromatography peak) to a protein ID

• Sequence info ? A bonus !!!

- In the "olden days": used protein microsequencing
  - laborious
  - lots of material needed



Let's not forget that chromatography can also be used to separate proteins/peptides !!!

What is Mass Spec ?

• MS determines mass-to-charge ratio (m/z) of individual molecules
  - molecules must be ionized in order to have charge
  - m/z effectively gives you the molecular weight of the ion

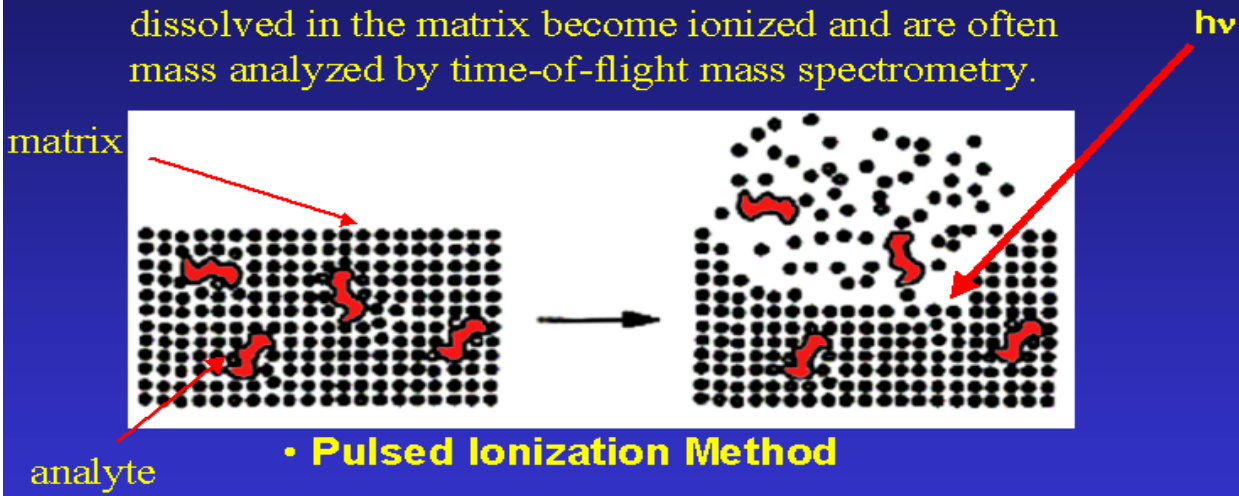
• m/z determined from ion "time-of-flight" in electric/magnetic field
  - time required to reach ion detector

• Ion Traps are used to selectively look at individual ion species

MALDI-TOF Mass Spectrometry

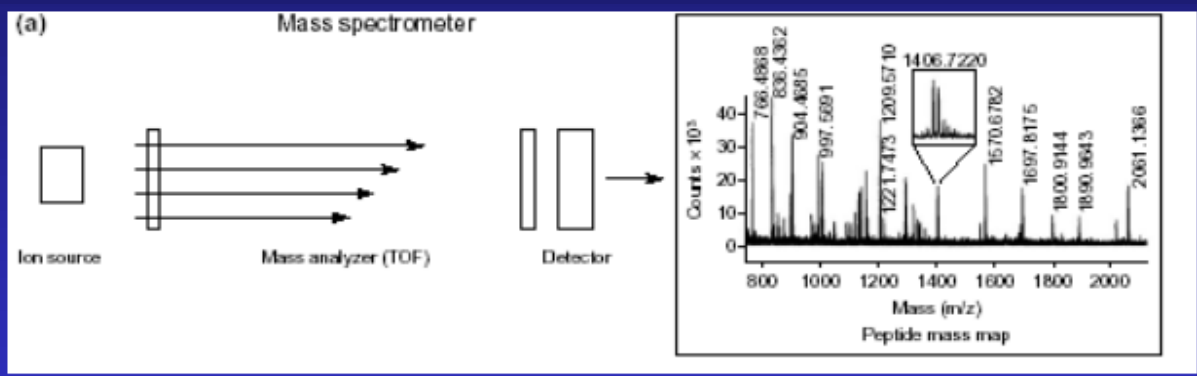
MALDI – matrix-assisted laser desorption/ionization

A pulsed laser beam is absorbed by matrix that strongly absorbs the laser wavelength. Analytes dissolved in the matrix become ionized and are often mass analyzed by time-of-flight mass spectrometry.



- **Pulsed Ionization Method**
- **Soft Ionization of Large Molecules**
- **Singly Charged Ions**

Time-of-flight mass spectrometers are frequently used with MALDI.

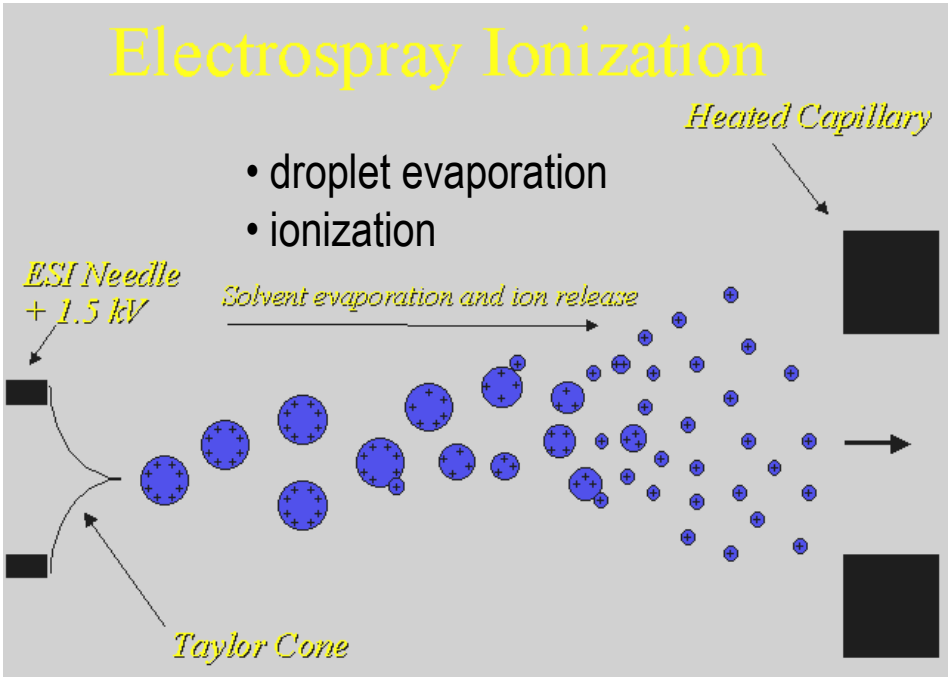


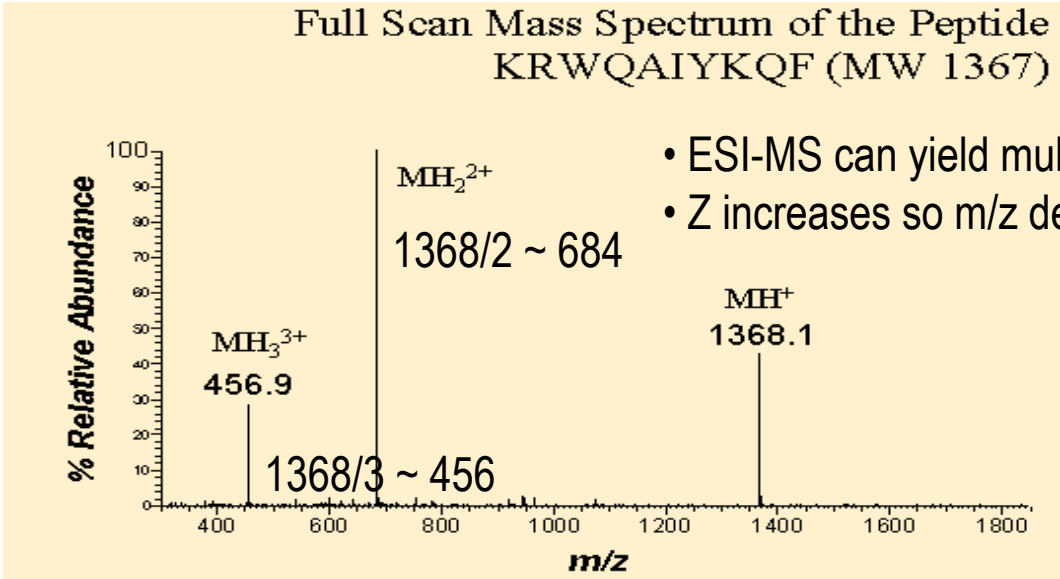
TOF is a foot race – the light ones have higher velocity and reach the detector first, since all ions have the same kinetic energy (KE=qV)

MALDI advantages:

- ionization due to ions generated from matrix, gentle on proteins
- waaaaaay higher sensitivity, little amounts required
- waaaaaay higher mass ranges can be examined (good, since proteins are heavy)
- method of choice for fast, relatively accurate peptide fingerprinting

Electrospray Ionization (ESI)-MS





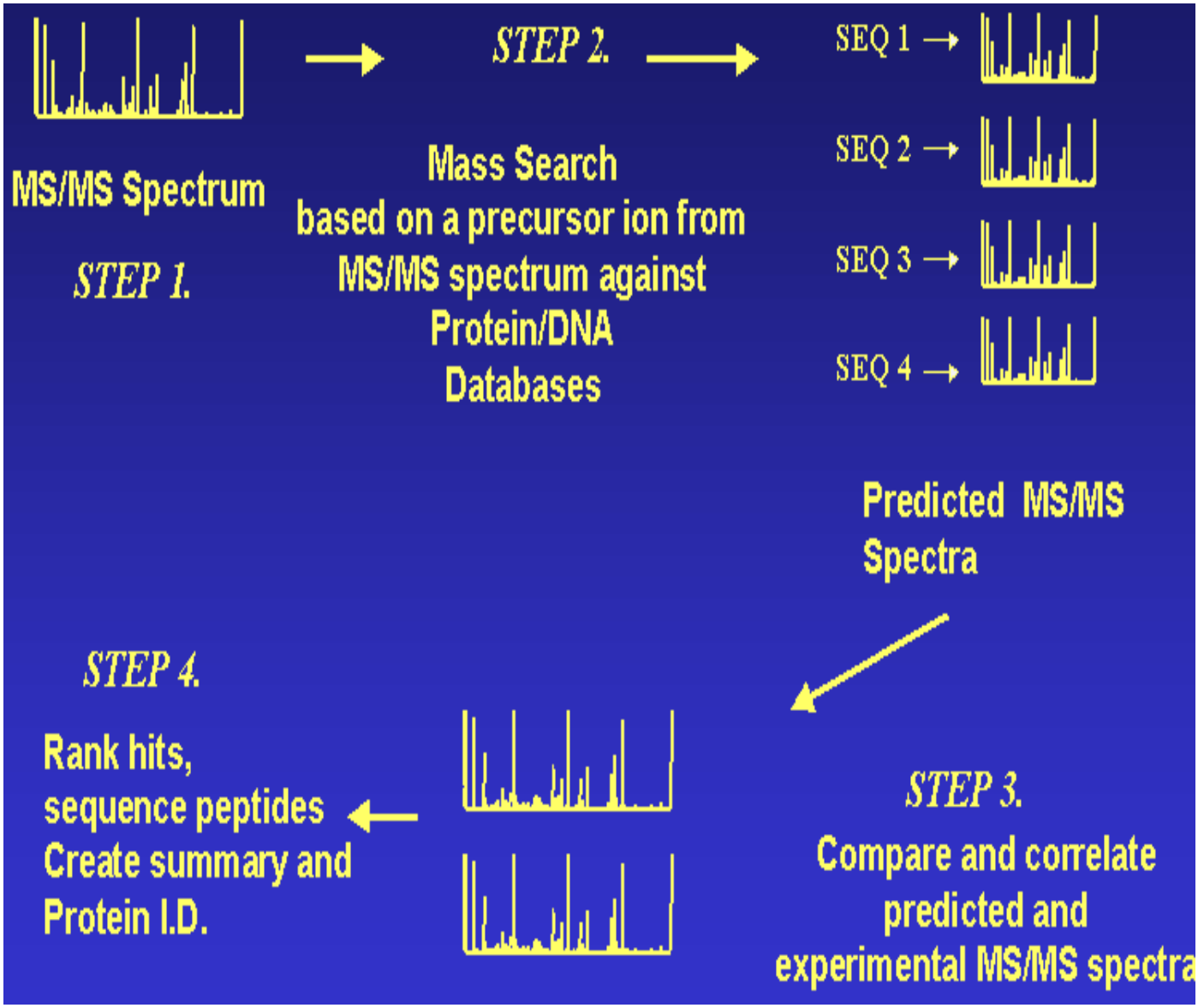
- ESI-MS can yield multiple ions
- Z increases so m/z decreases

- ESI-MS generates better spectra than MALDI-TOF MS

Protein Determination using ESI-MS: Peptide Fingerprinting

- Take protein, perform proteolytic cleavage: you get different peptides
- Any given peptide has a characteristic m/z
- Run against database of predicted m/z based on sequence data
    - look for best matches of various peptides to predicted proteins
    - the better the mass accuracy, the more stringent the search can be

Protein Identification Using MS Data (Pt.I)



- need good Dbase
- need good algorithms to search the Dbase
- what if your sequence is not in the Dbase ?
    - Need sequence data !!!

Protein Identification Using MS Data (Pt.II)

Without errors the comparison is rather trivial:

Real m/z vs Expected m/z

But methods have to tolerate errors:

a) Recording error < 1%

b) Searched sequence not verbatim in database (due to mutations)

c) Mutations may cause different digestions

d) Impurities in the sample produce spurious data

e) Partial or incorrect digestion

f) Systematic error of apparatus

Peptide	Theoretical MW	Experimental MW	% Mass Accuracy
Bradykinin	1059.56	1059.62	0.006
Angiotensin I	1295.68	1295.63	0.004
Calciseptine	7031.20	7030.80	0.006
Interleukin-8 (rat)	7840.09	7840.38	0.004
Cytochrome C peptides:			
TGQAPGFTYTDANK	1469.68	1469.66	0.001
KTEREDLIAYLK	1477.82	1477.86	0.003
GITWKEETLMEYLENPKK	2208.11	2208.19	0.004

System is Automatically Tuned and Calibrated Once With External Standard

Average Mass Accuracy = 0.004% (~40 ppm) [Also MS/MS]

• A mass accuracy of 0.004% (40 ppm) is good, but 1 ppm would be even better (more on this later)

• More than one peptide can have similar MW !!!! Especially as errors get compounded