February 3, 2009

# 2D electrophoretic and other high-resolution separations and analysis of proteins in biological samples

**Helen Kim**

**934-3880**

helenkim@uab.edu

**Dept of Pharmacology and Toxicology**

**McCallum Building, room 460**

http://www.uab.edu/proteomics

# Learning objectives

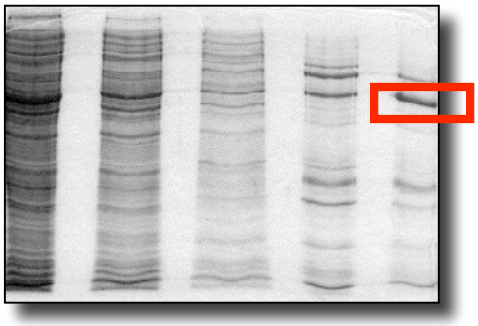
**2-D gel proteomics: What it involves:**

- **2D separation**
- **image, statistical analysis**
- **MS**

**Other types of 2D electrophoretic protein separations;**

- **Free-flow**
- **2D Blue native electrophoresis**

**Initially: The objective of protein separation in proteomics was to get a protein "spot" or "band", for mass spectrometry analysis, to identify the protein/its modifications.**



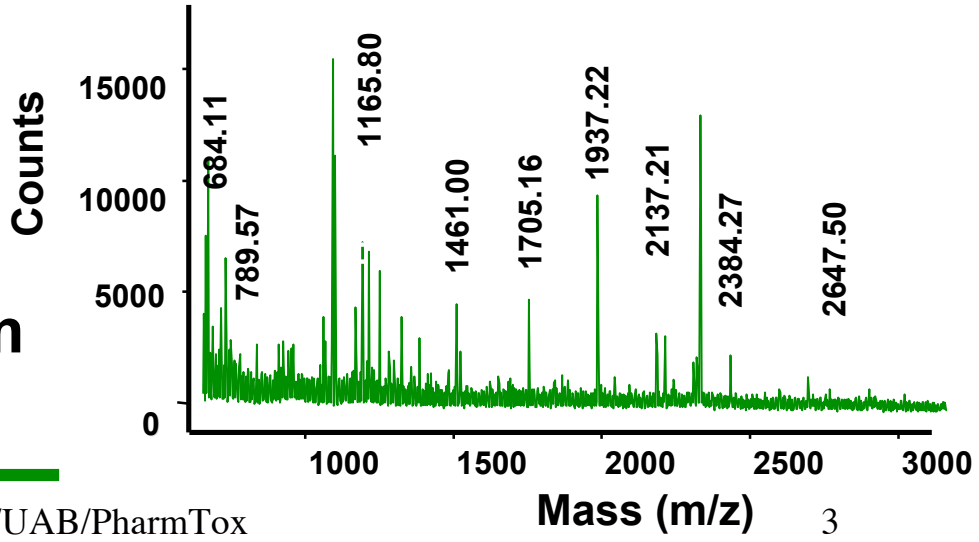
trypsin

MALDI-TOF mass spectrometry

ID of parent polypeptide

MASCOT search engine

http://www.matrixscience.com



# Types of high-throughput separation & analysis technologies

I. 2-dimensional electrophoresis (2-DE)
   A. "regular" IEF/SDS-PAGE
   B. 2D-blue-native electrophoresis
   C. Visualization methods

II. 2-dimensional liquid-based LC/LC

III. Free-flow electrophoresis

IV. "Chip" technology: arrays of ligands for proteins

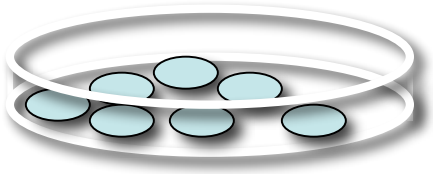
# Parameters that govern the choice of protein separation method

- Purity of protein
- Speed of purification
- Quantity of protein
- What is the question:most important
  - Discovering a new protein/proteome
  - Identifying protein-protein interactions
  - Identifying potential modifications of known proteins

# The Elements of any 2-D separation in proteomics

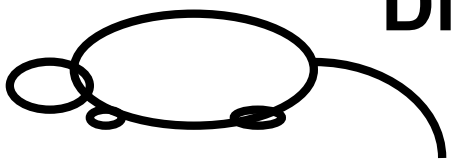
- The experiment!! that generates the "signal"
- Subfractionation to enrich for suspected proteins
- (Trypsin-digestion to generate peptides of the parent proteins)
- Sample work-up
- 2-D separation
- Image or other analysis to identify gel "spot" differences between untreated & treated
- MS of (trypsin-digests) spots/proteins/peptides of interest, to identify and characterize the protein

# Sample preparation for 2DE:



**Harvest, rinse, and pellet the cells;**

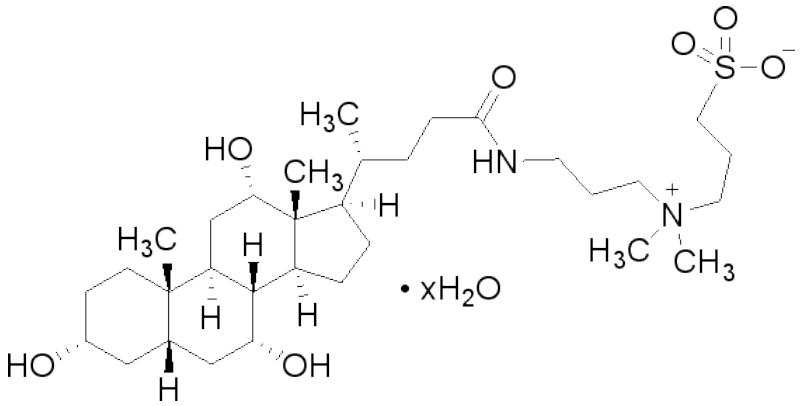
**or**

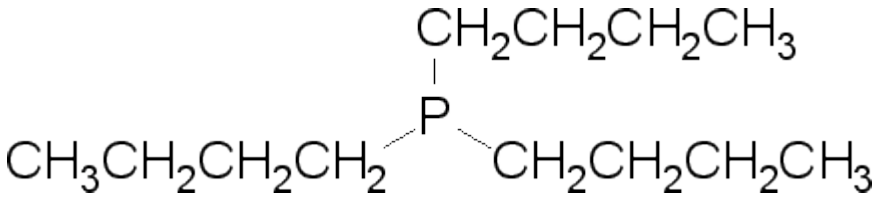


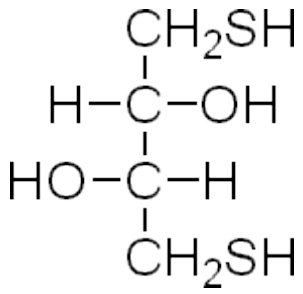
**Dissect out tissue, organ, or fluids;**

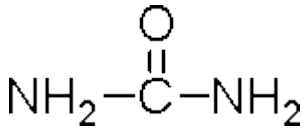
- **Homogenize/lyse in buffer that *unfolds the proteins w/o adding or disturbing the charges*:**
  - **High urea usually 5-8 M---unfolds the protein**
  - **Sometimes 2 M thiourea--unfolds the protein**
  - **1-4% detergent--solubilizes hydrophobic components**
  - **Beta-mercaptoethanol or other reductant, such as TBP, DTT**
  - **Inhibitors: of proteases, kinases, & phosphatases**
- **Clarify by centrifugation to get rid of insoluble matter;**
- **Protein assay to know how much and how concentrated**

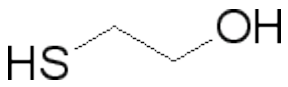
# Structures important in 2D electrophoresis to know and understand

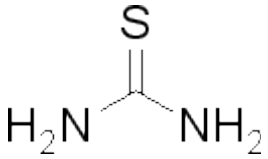


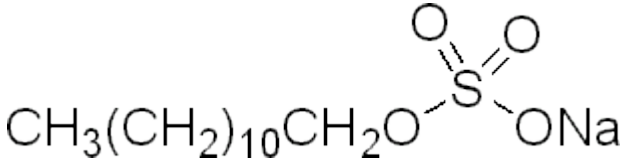



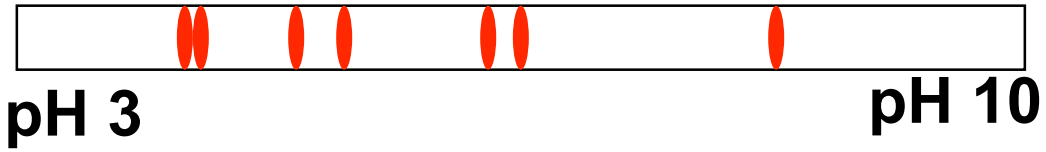



# What 2-D electrophoresis involves:

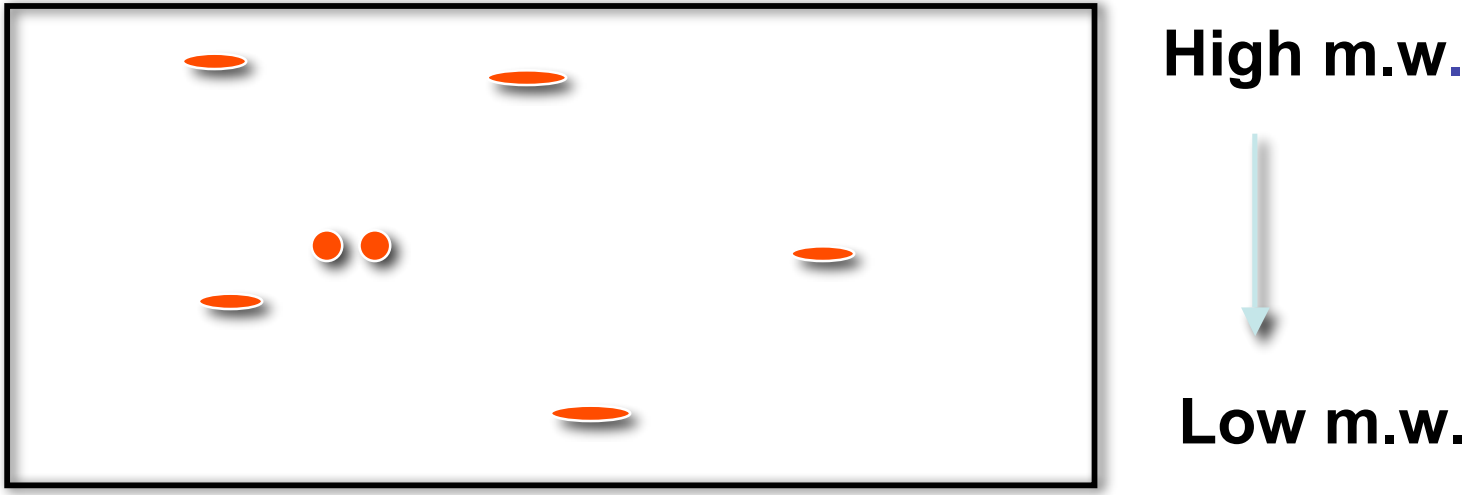
- 1st dimension: **Isoelectric focusing**

**(separation according to charge )**

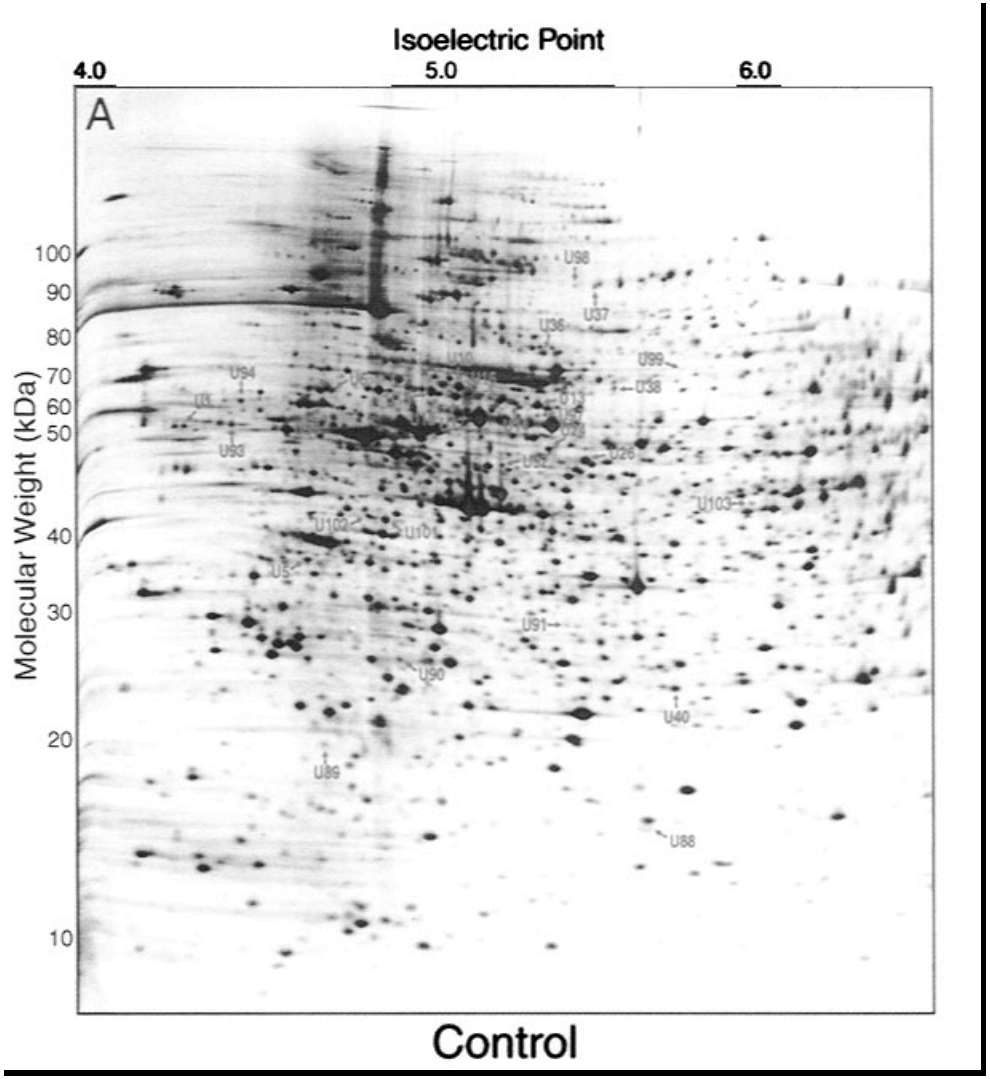



- 2nd dimension: **(SDS)-PAGE**

**(separation according to size )**



# A real 2-D gel



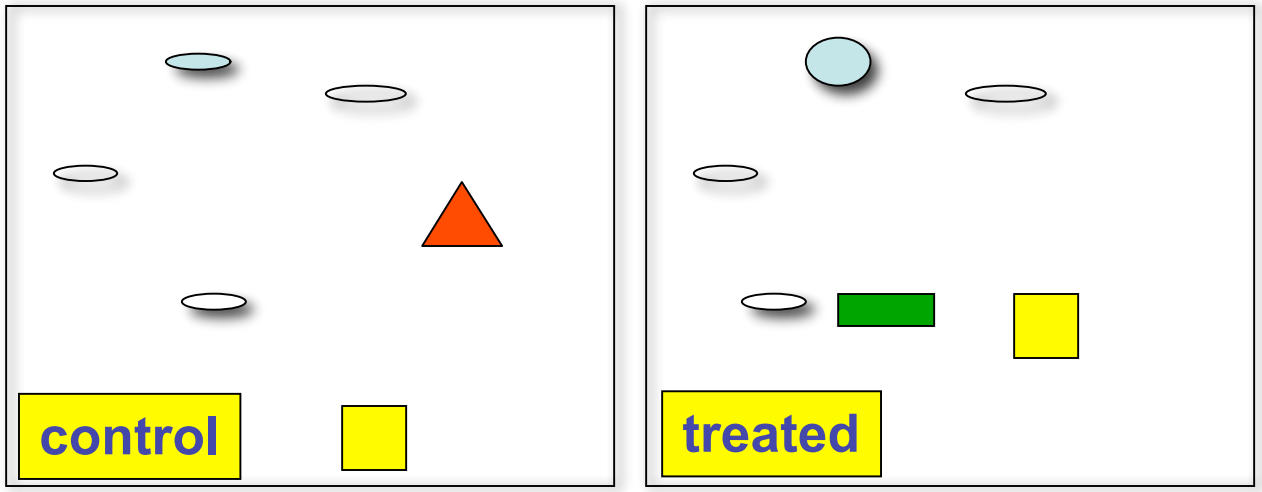

**Find this and other 2-D gels at http:www.expasy.org**

**Lewis et al., [2000] Molec. Cell, 6)**

**(from Natalie Ahn's lab)**

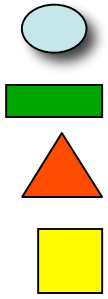
**The pattern itself is information; a change in intensity or position of a spot has biological meaning…..similar to astronomy.**

# Critical part of 2-D gel proteomics: Image analysis




**Either manually or with software: "compare" the images.**

## Types of information:



------ Upregulation of gene

------ Posttranslational modification

------ Downregulation of gene

------ Aberrant processing

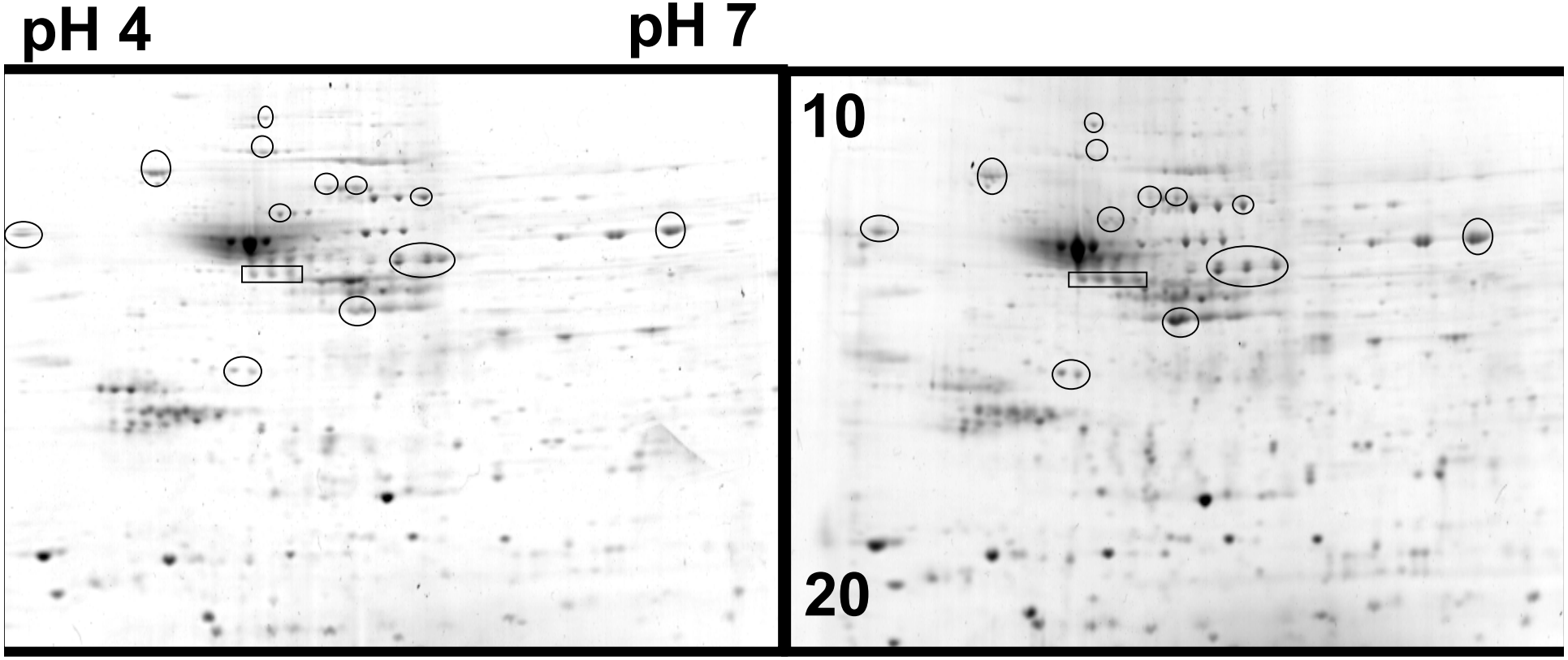
# Elements of image analysis of "regular" 2D gels:

1. Compare the 2D displays of spots
2. Determine total spot number for each display
3. Quantify spot intensities, identify differences
4. Identify spots that may have " moved" horizontally; these are candidates for those that are altered in charge, reflecting posttranslational modifications.

## Ultimate and simple goal of image analysis

to answer the question,
"What is changing, and by how much?"

# A pair of 2D gels representing rat brain protein changes induced by ingestion of grape seed



# Categories of data generated by 2D gel image analysis

| SSP | intensity | X coordinate | Y coordinate | Peak quality |
|---|---|---|---|---|
| 1369 | 12,269 | 2.375 | 1.279 | 75 |
| | | | | |

Differential expression?

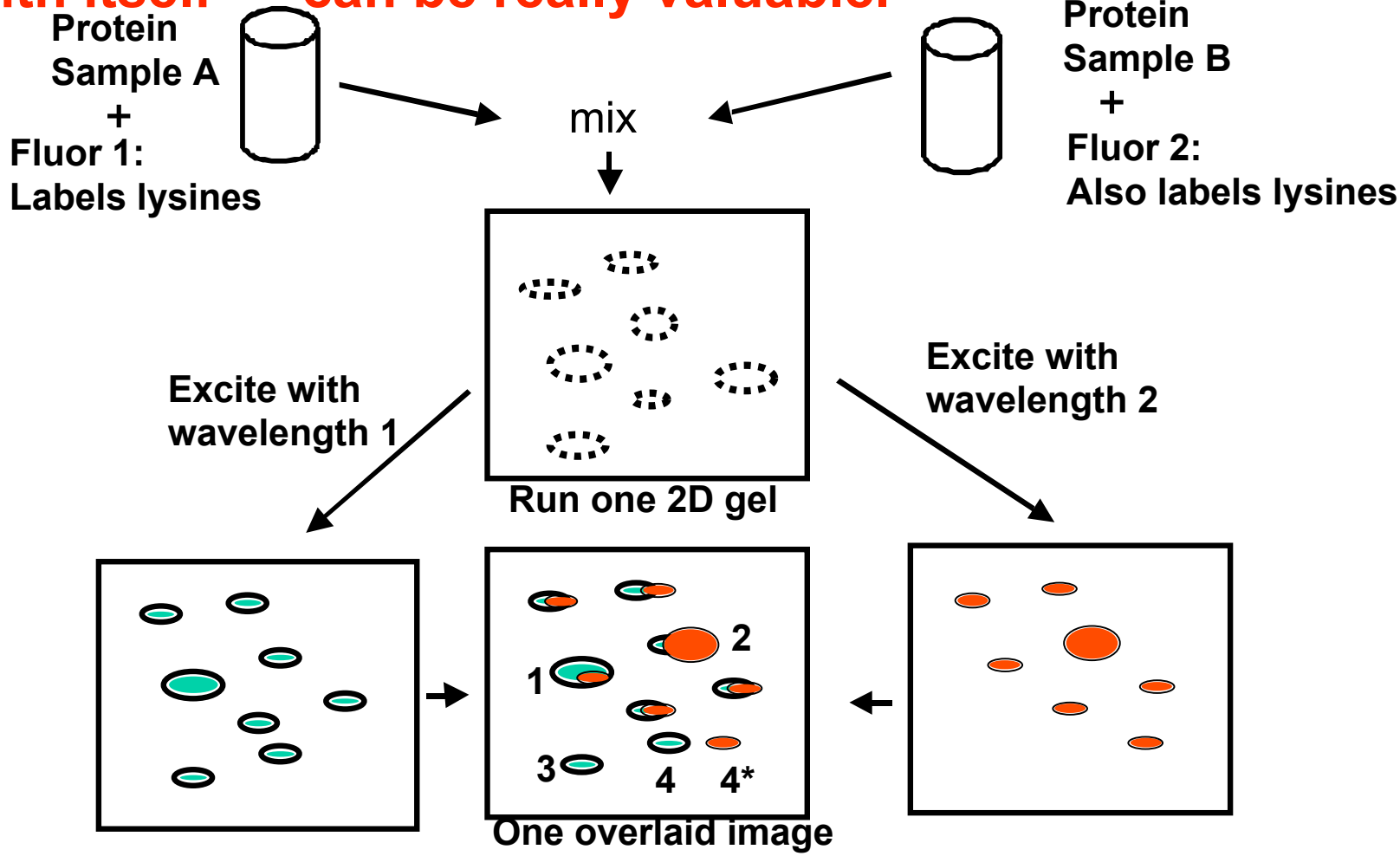
Modifications?

Change in Complexity?

**typical dataset: >100,000 rows of data**

**Statistical analysis**

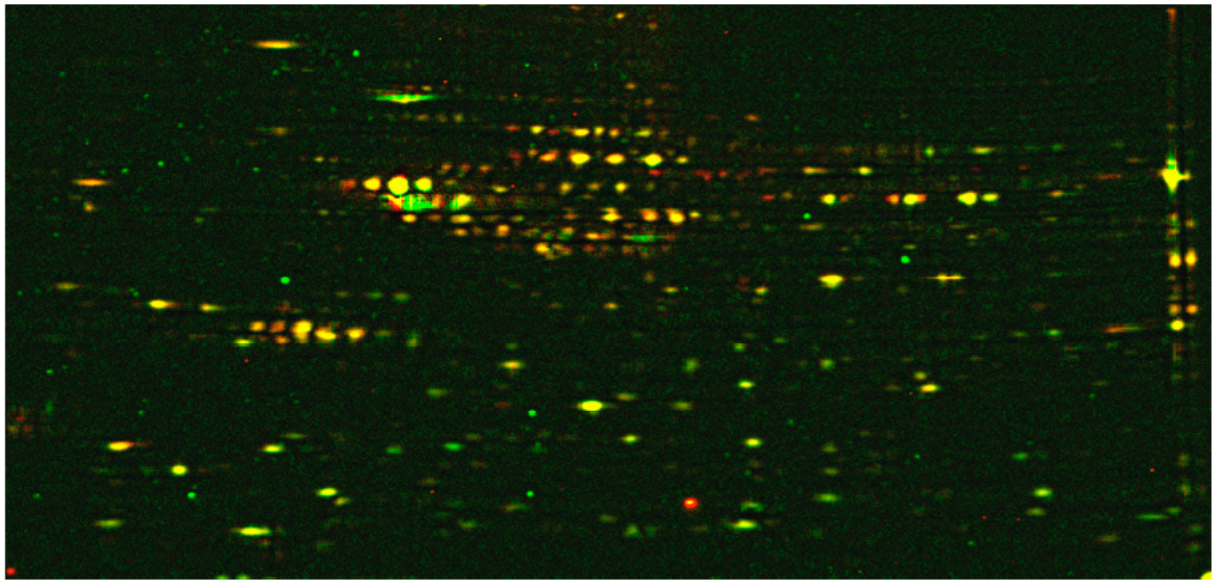
# Difference gel electrophoresis (DIGE); a protein migrates with itself---- can be really valuable.




**Single biggest bottleneck, even with DIGE: even commercially made gels (1st dimension, 2nd dimension) are not perfect. So, the simpler you can keep the system, the better. If you don't HAVE to use a gradient gel, don't.**

Example of DIGE:

**Differential protein labeling with Cy3 and Cy5**
**Superimposed images from the same gel**
**of normal and cancer cell lines from the breast**



Visually: 100% green spots are specific to normal cells, 100% red spots are specific to cancer cells, degrees of yellow-orange indicate differential expression.

# Typical Database obtained from proteomics analysis of breast cancer cell lines

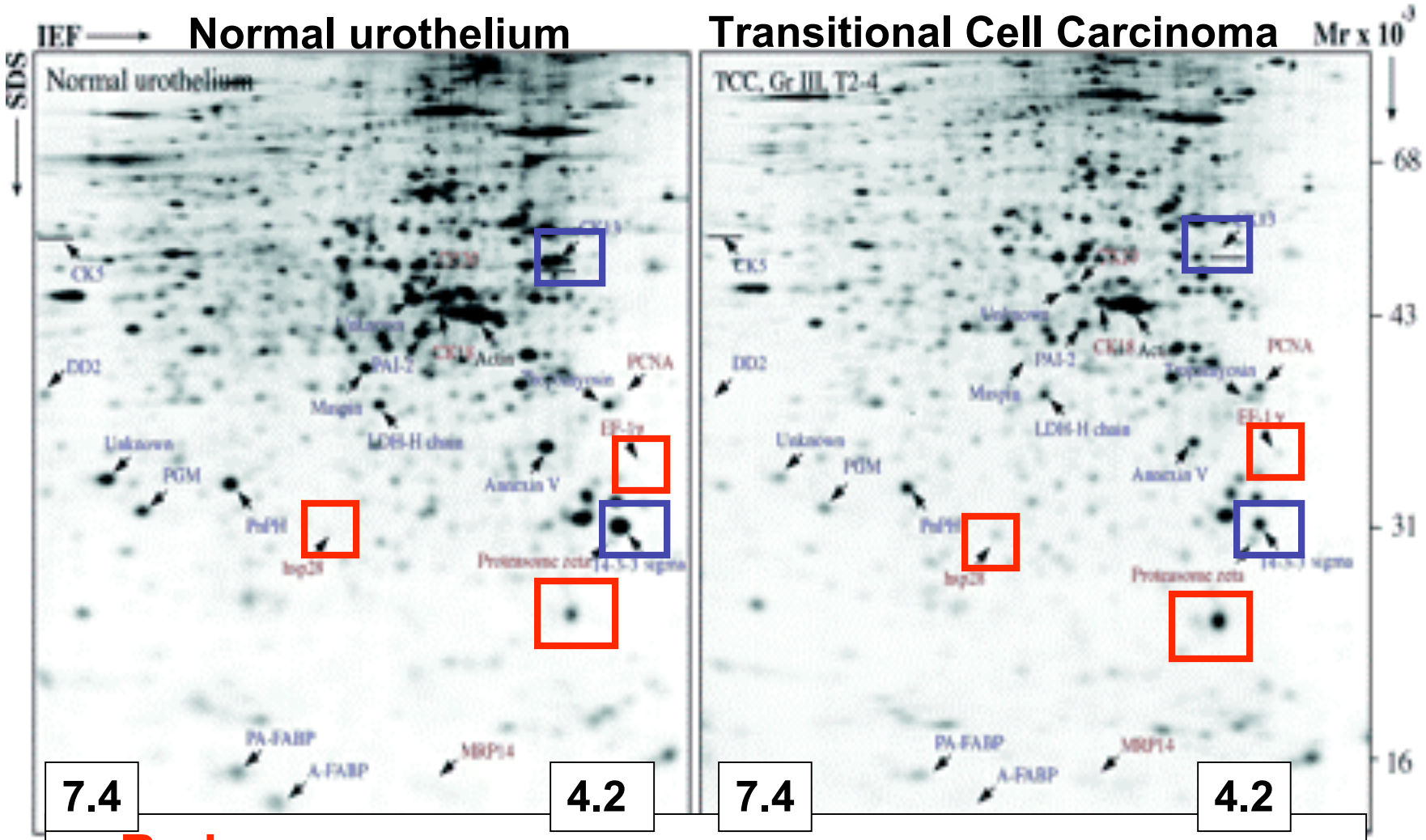
| Sample | Spot# | Protein Name | MOWSE | Accession# | Obs: kDa | Pred: kDa | Obs: pI | Pred: pI |
|---|---|---|---|---|---|---|---|---|
| MCF-7 | BR-8 | Cytoskeletal keratin-7 | 132 | gi12803727 | 20.7 | 51.3 | 8.6 | 5.3 |
| | BR-9 | Alpha-1 anti-trypsin | 90 | gi1942953 | 19.7 | 44.3 | 9.6 | |
| | BR-1 | Human | 94 | gi87303 | 56.2 | 53.5 | 5.9 | 5.6 |
| MCF-10AT | BR-6 | Cytoskeletal keratin-8 Human | 74 | gi18573275 | 23.4 | 24.1 | 6.3 | |
| | BR-5 | hypo:XP109048 | 60 | gi5106591 | 18.1 | 6.4 | 5.7 | |
| | BR-7 | IgG heavy chain variable region | 105 | gi1942953 | 37.0 | 44.3 | 6.9 | |
| | BR-4 | Intact recomb: alpha-1 antitrypsin mutant F-L | 65 | gi5106591 | 10.0 | 6.4 | 4.6 | |
| | | IgG heavy chain var:reg | | | | | | |
| MCF-10A | BR-3 | Human cytokeratin-8 | 65 | gi87303 | 56.0 | 53.5 | 5.8 | 5.6 |
| | BR-2 | DNA replication silencing factor MCM-4 | 58 | gi1705520 | 22.8 | 96.5 | 6 | 6.6 |

**(Deshane, Johanning, and Kim, unpublished data)**

# "Mine" your proteomic data

- **Note every difference--eventually it all means something;**
- **But make sure the difference is "real."**
  - **What is the variation in that parameter (mw, pI) for that samespot in that treatment group;**
  - **Quality control issues come into play here;**
    - **Did you design the experiment with a statistician?**
- **Make sure your "basal" mw and pI are consistent with predicted and/or what others have observed;**
- ***Then* you can conclude that a difference in pI, for example, indicates a change in modification**
- **If some/all of a spot is found at a pI different from predicted, it may be constitutively modified in the "unstimulated"/"normal" group**

# Metabolic labelling can enhance 2D gel analysis: i.e. $^{35}$S-methionine-labelling

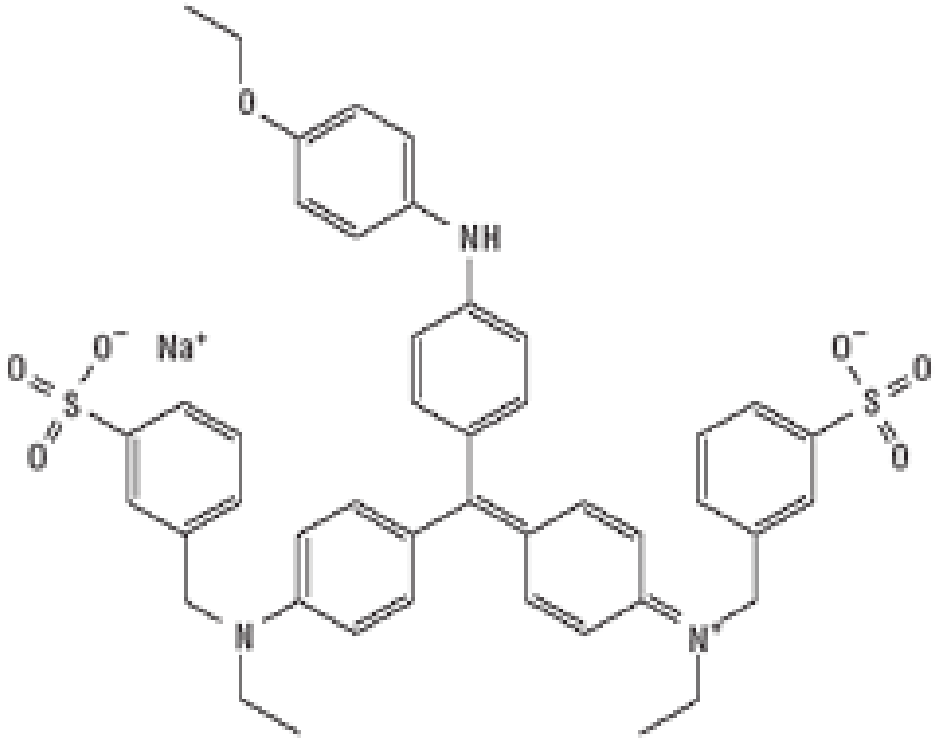



**Red** = upregulated in the tumor; **blue** = down-regulated.



(Celis et al., Mol.Cell Proteomics, 2002, 1:4)

# Visualizing 2D gels: Coomassie Brilliant Blue

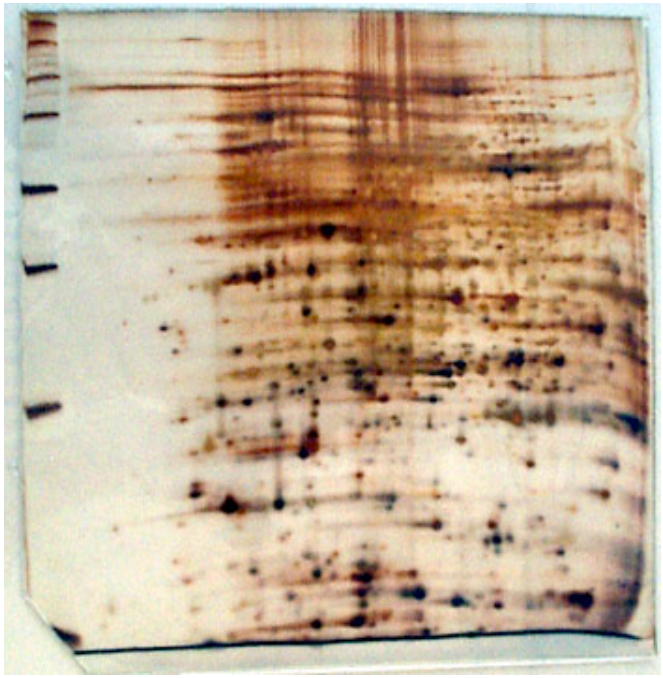



Coomassie Brilliant Blue R-250 Dye
M.W. 825.97

In acidic conditions, the anion of CBBR combines with the protonated animo groups on proteins via electrostatic interactions.

- Inexpensive
- Image readily acquired by scanning at visible wavelengths
- No covalent mass change

# Silver stain



**Silver ions (from silver nitrate) are chemically reduced to metallic silver on lysine residues.**

**This is the most sensitive protein stain, but also the least useful for quantitation because of its low dynamic range.**

2D gel image courtesy of the U. Va. Mass Spectrometry Shared Facility, 2006

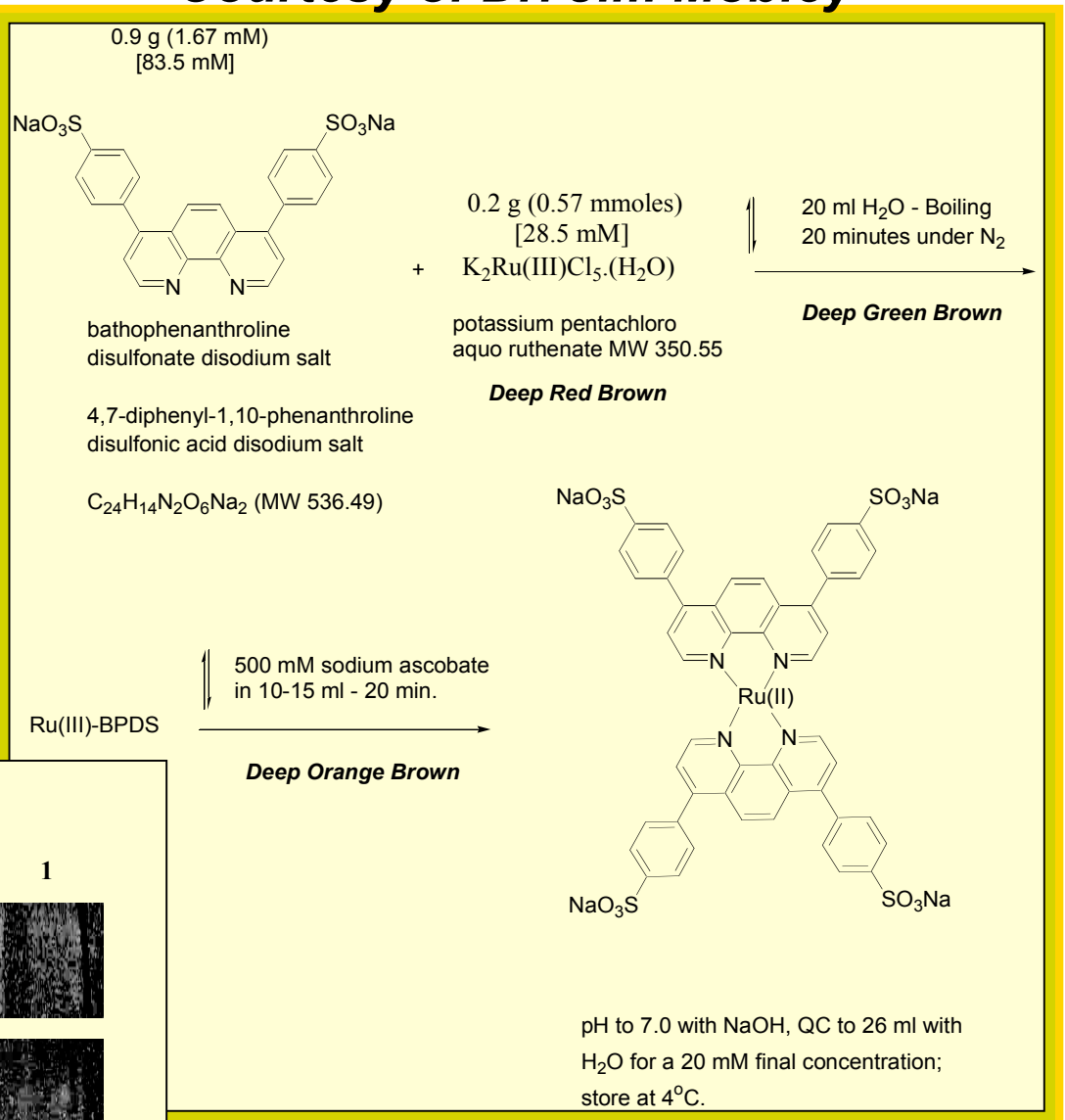
# Sypro Ruby, and Ru(II)-BPDS Dye

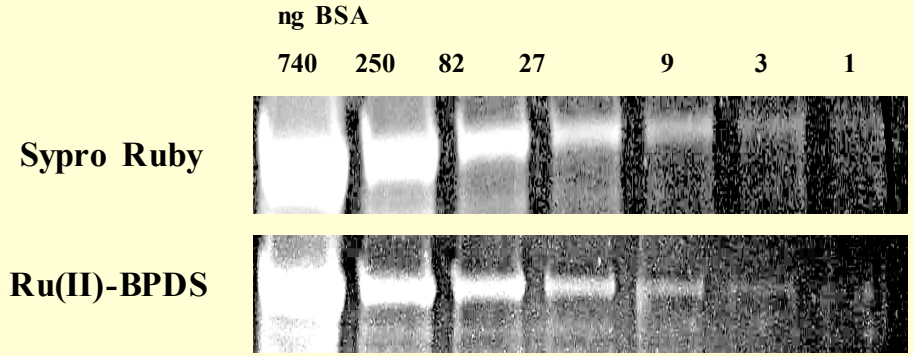
***Courtesy of Dr. Jim Mobley***

**Sypro Ruby (*fluorscent*)**

- **Equal to silver stain in sensitivity.**
- **But ~200-500-fold greater dynamic range.**

**•Compatible for MS analysis.**

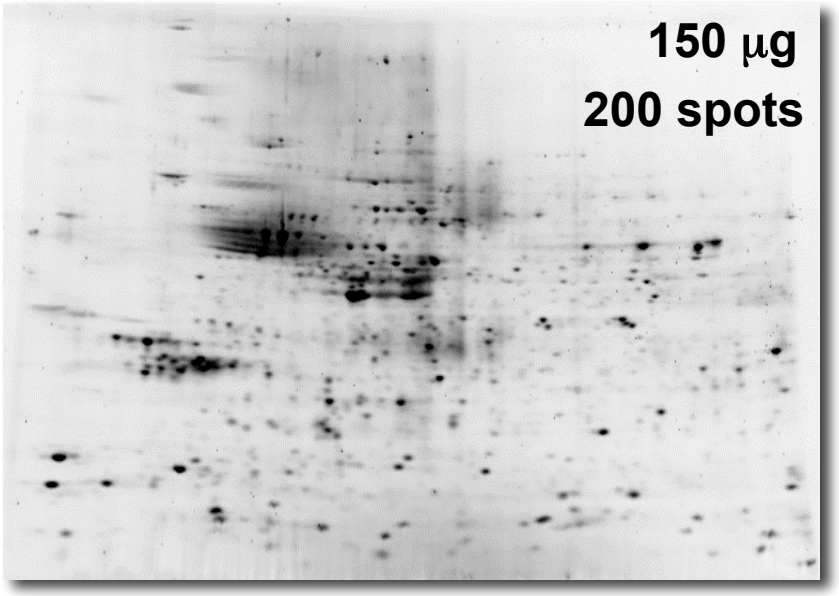


# Advantages and limitations of the types of stains

| | Sensitivity | Dynamic range | MS- compatible |
|---|---|---|---|
| CBBR | 8 ng | 10-30 x | yes |
| Silver | 1 ng | < 10 x | Not without special precautions |
| Fluorescent | 2 ng | 3 orders of magnitude | yes |

# Proteins, proteins everywhere, but where's my receptor?




2D gel of rat brain, stained with Sypro Ruby

**The genome predicts: 20,000-50,000 polypeptides.**

**So, 200 spots is <1% of the total proteome.**

**Conclusion: Even a fluorescently stained 2D gel of an unfractionated sample, only allows detection of the "low hanging fruit."**

# Central issue in proteome analysis: dynamic range

**In cells: protein amounts vary over a wide *dynamic range*:**

**In blood, albumin is 3.5 g/100 ml (35 g/L = 0.5 mM) ($10^{-3}$ M), whereas cytokines are pM ($10^{-12}$ M)**

- **This is a difference of nine (9) orders of magnitude.**

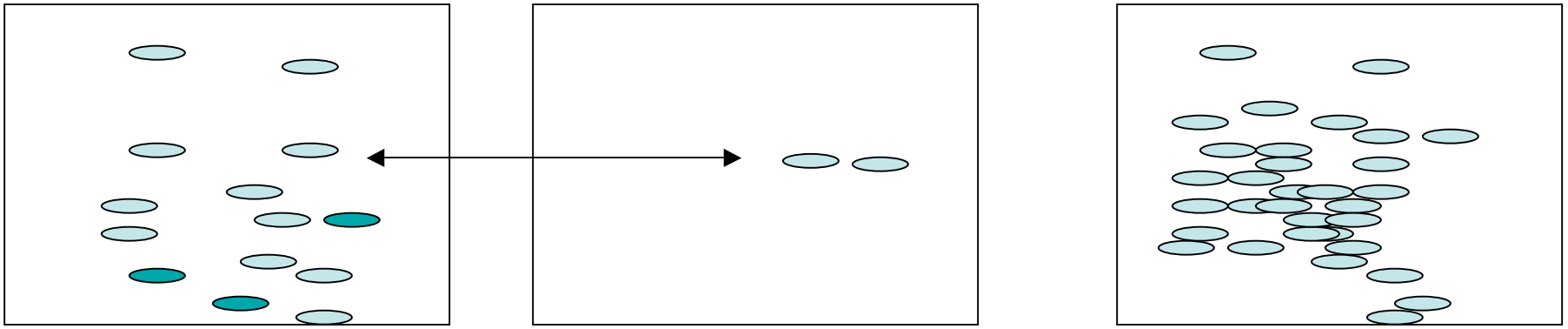
- **A 2D gel that is overloaded with respect to an abundant protein, may have *barely detectable* amounts of a low abundance protein……**
  - **If you can't see it, you don't know a protein is there;**
  - **Even if you know it's there, you can't do MS, because there isn't enough protein.**

- **No one stain will detect 9 orders of magnitude differences in abundance of proteins.**

## For greatest sensitivity, and the most biological information:

**Stain the gel…..**

**----BUT do a Western blot or two, with replicate gels:**

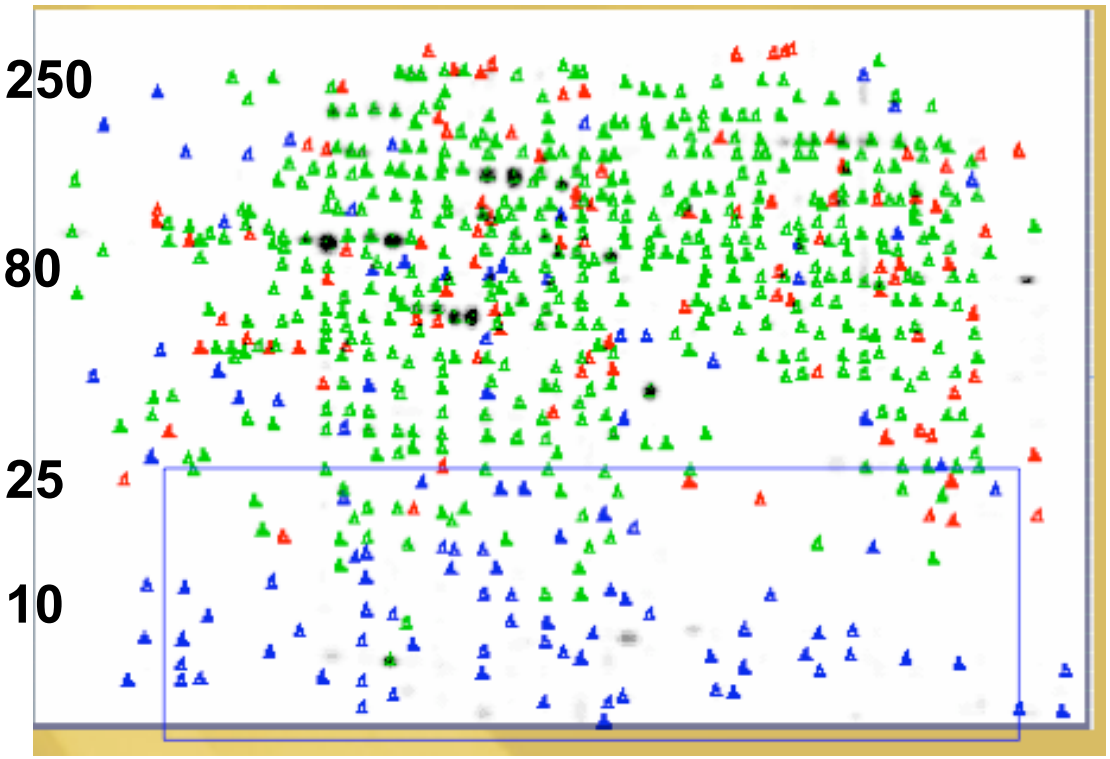


**Western blot for phosphoproteins**

**Western blot for a particular protein**

**Stained gel**

# Deep Purple: the new fluorescence dye from GE




**BLUE**: spots on Deep Purple-stained gels;

**RED**: spots on Sypro Ruby-stained gels;

**GREEN**: spots matched to every gel

**Deep Purple MAY detect more smaller proteins (<25kda).**

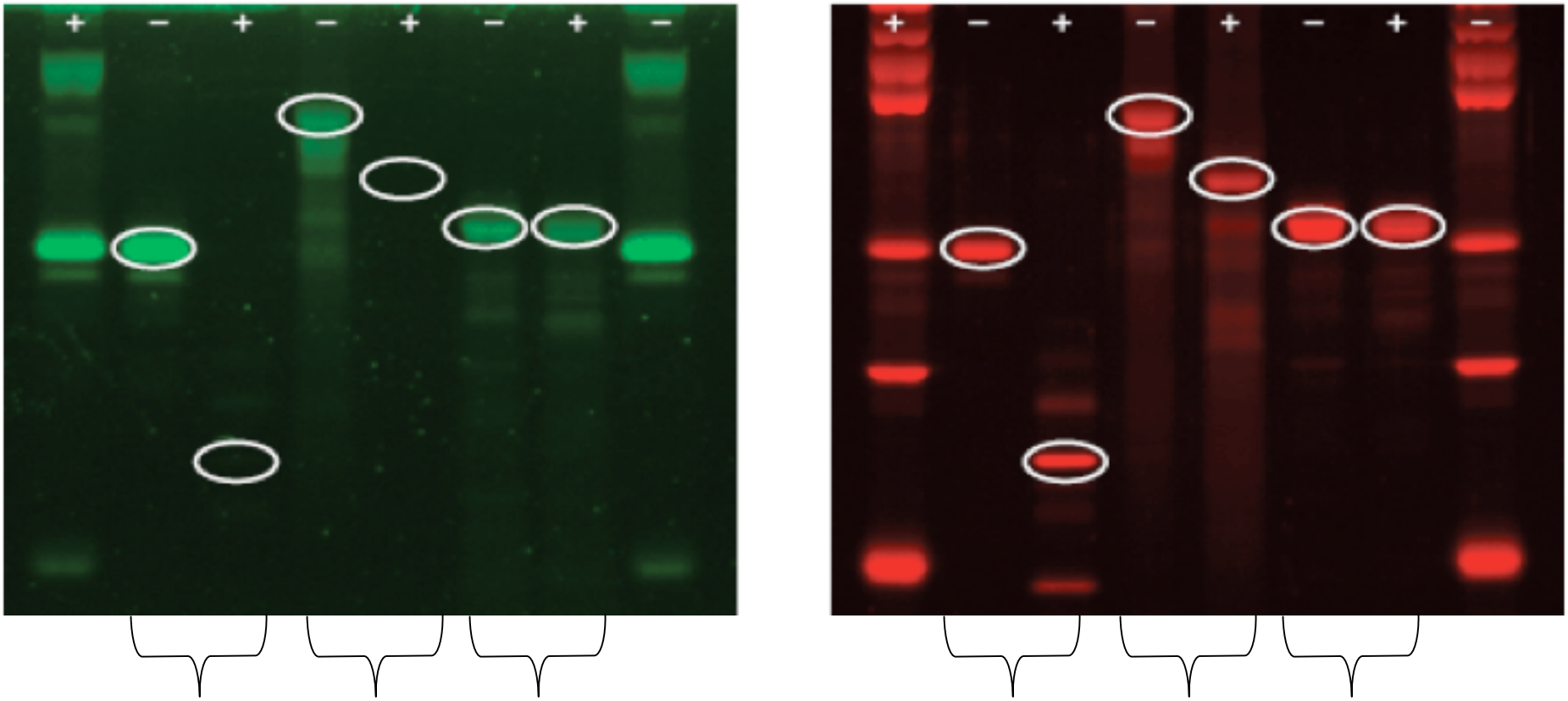


(adapted from Duncan Veal, 2006)

# Multiplex Proteomics: ProQ Emerald followed by Sypro Ruby can identify multiple glycosylated proteins at once

ProQ Emerald

Sypro Ruby

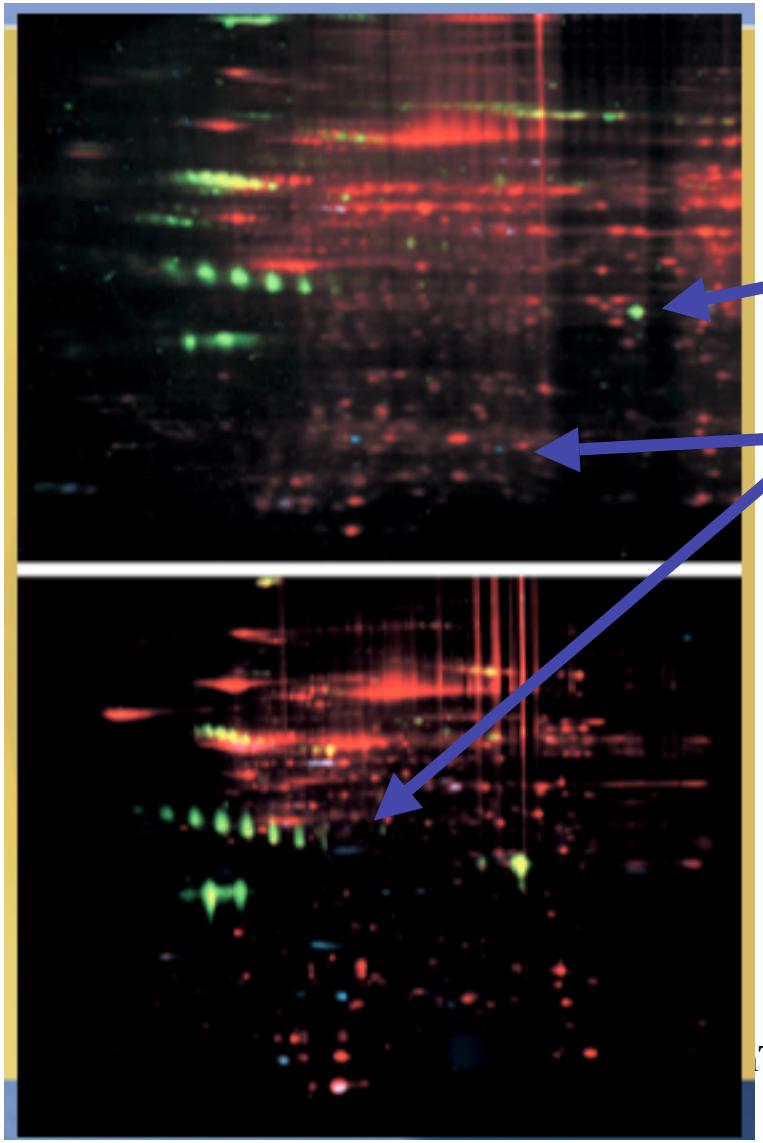



(Duncan Veal, Fluorotechnics Pty, Ltd.)

# Multiplexing: valuable when sample is scarce or difficult to obtain

Normal liver

Liver tumor



Sypro Ruby: total protein

Pro Q Emerald: glycosylated

Pro Q Diamond

**The value of fluorescent dyes: each can be detected separately from others on the gel, due to different excitation and emission spectra.**

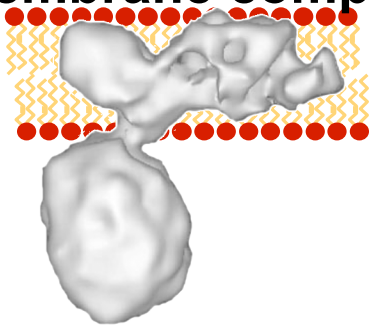
Modified from Duncan Veal, Fluorotechnics Pty, Ltd.

# SUMMARY of 2D gel stains

- **Protein stains differ according to**
  - **Sensitivity/Dynamic range/MS-compatibility/Ease of capture of information**
- **The fluorescent dyes offer unparalleled protein analytical capabilities due to the wide dynamic range, and their MS compatibility;**
- **"Multiplexing" allows analysis of subproteomes in the same gel, maximizing use of scarce samples;**
- **Yet, each stain has utility depending on experimental goals.**

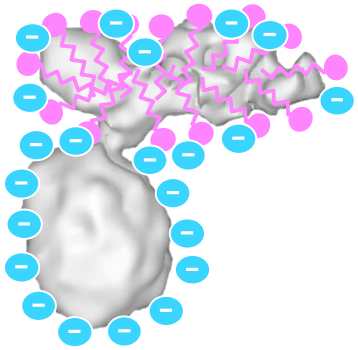
# 2D-Blue-Native gels: for hydrophobic proteins, protein complexes.

**Intrinsic mitochondrial membrane complexes**



**Detergent**

**CBBR**

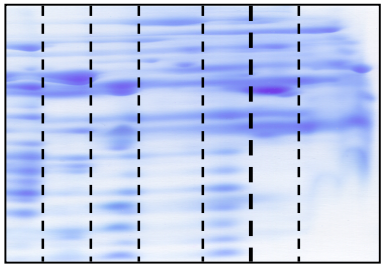


**Released complexes, all negatively charged, thanks to the CBBR**


**1st D: NATIVE electrophoresis**

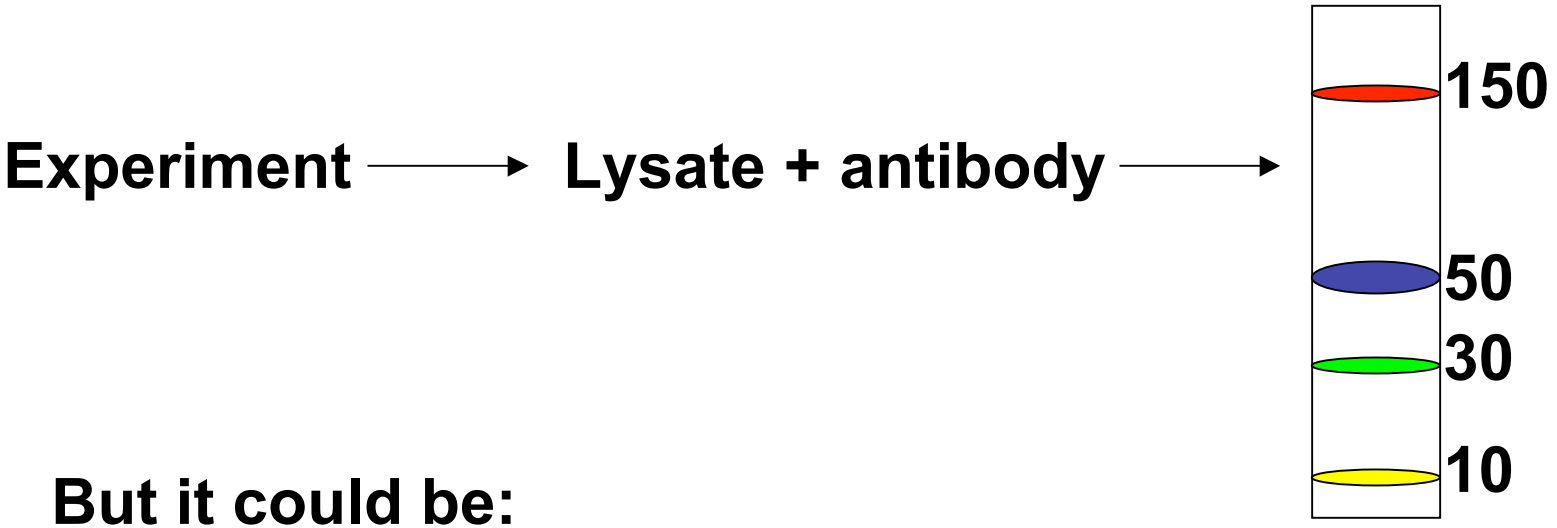
**(Modified from Brookes et al., 2002)**

**This type of 2D gel has "ladders" of bands.**

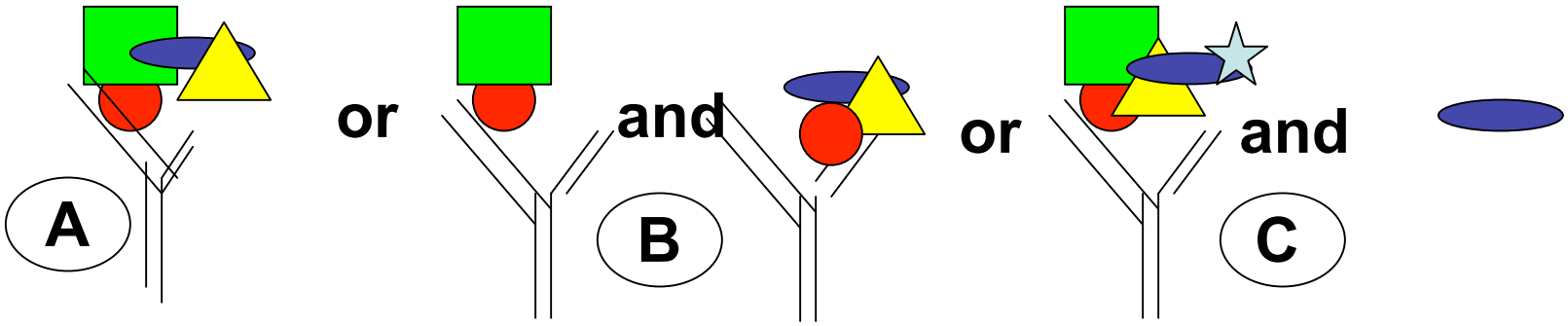


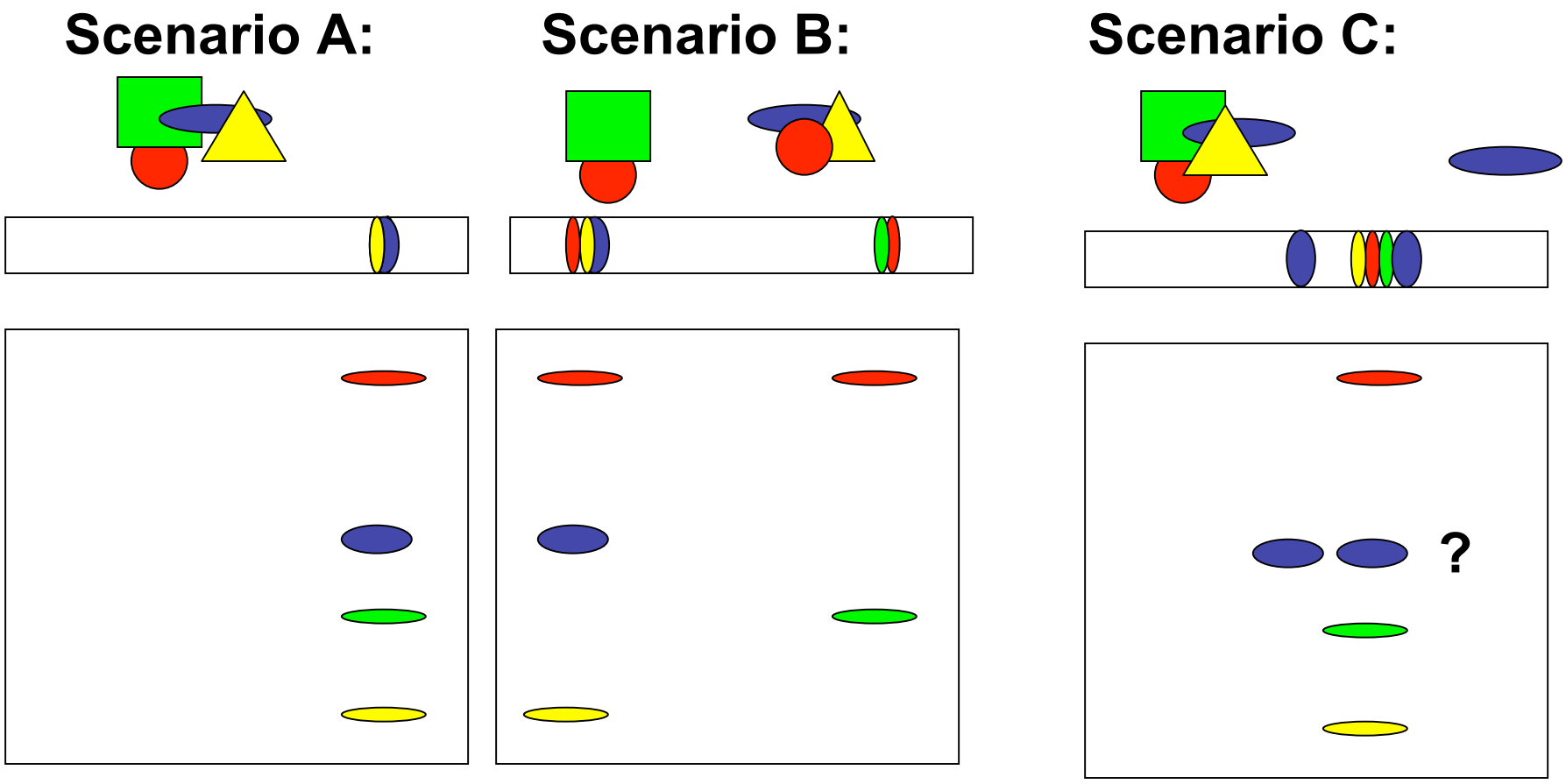
# What questions does BN electrophoresis address:

## (a) Which proteins are actually interacting with which?




**But it could be:**

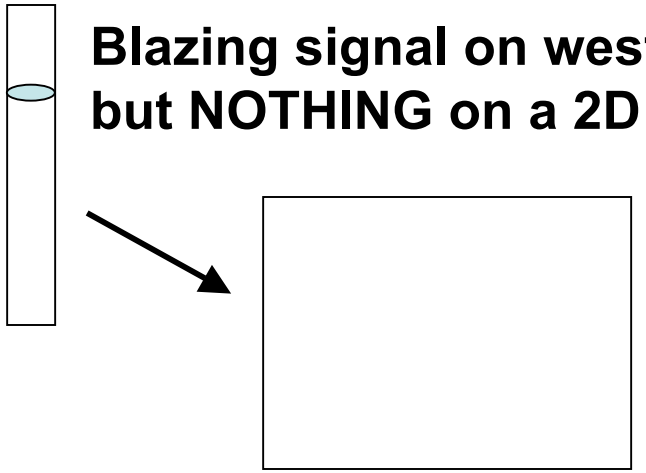





**When would you see the same protein in two lanes on the 2nd D?**

# Issues in 2D gel analysis:

I. **Now you see it, now you don't:**



**Blazing signal on western blot of 1D SDS gel; but NOTHING on a 2D gel of the same lp:**

Explanations?

1. Solubilization
2. Quantity
3. Modification

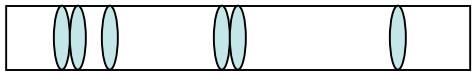
II. **I see the band on my 8 cm long 1D gel that runs with my western blot band;**

**Why do I need to run a 2D gel anyway?**

**1. Information about what proteins that associate with the antigen;**

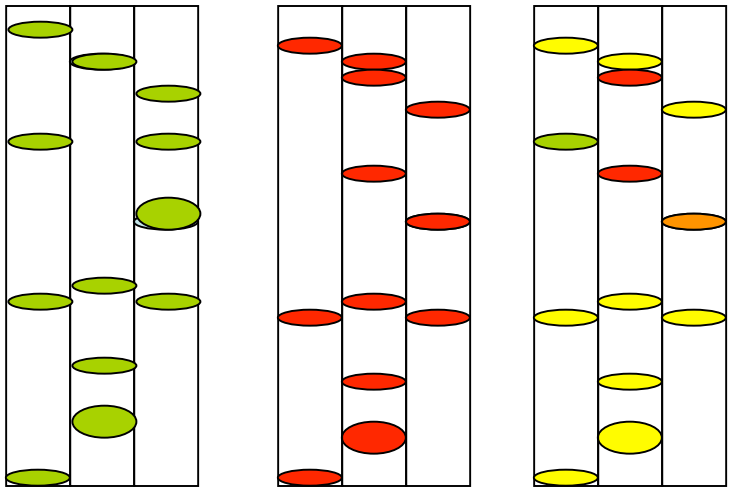
**2. Separate multiple proteins in that one band**

# 2D-LC-LC



1st dimension: chromatofocusing (like IEF, but *in solution*, so can take higher protein loads)

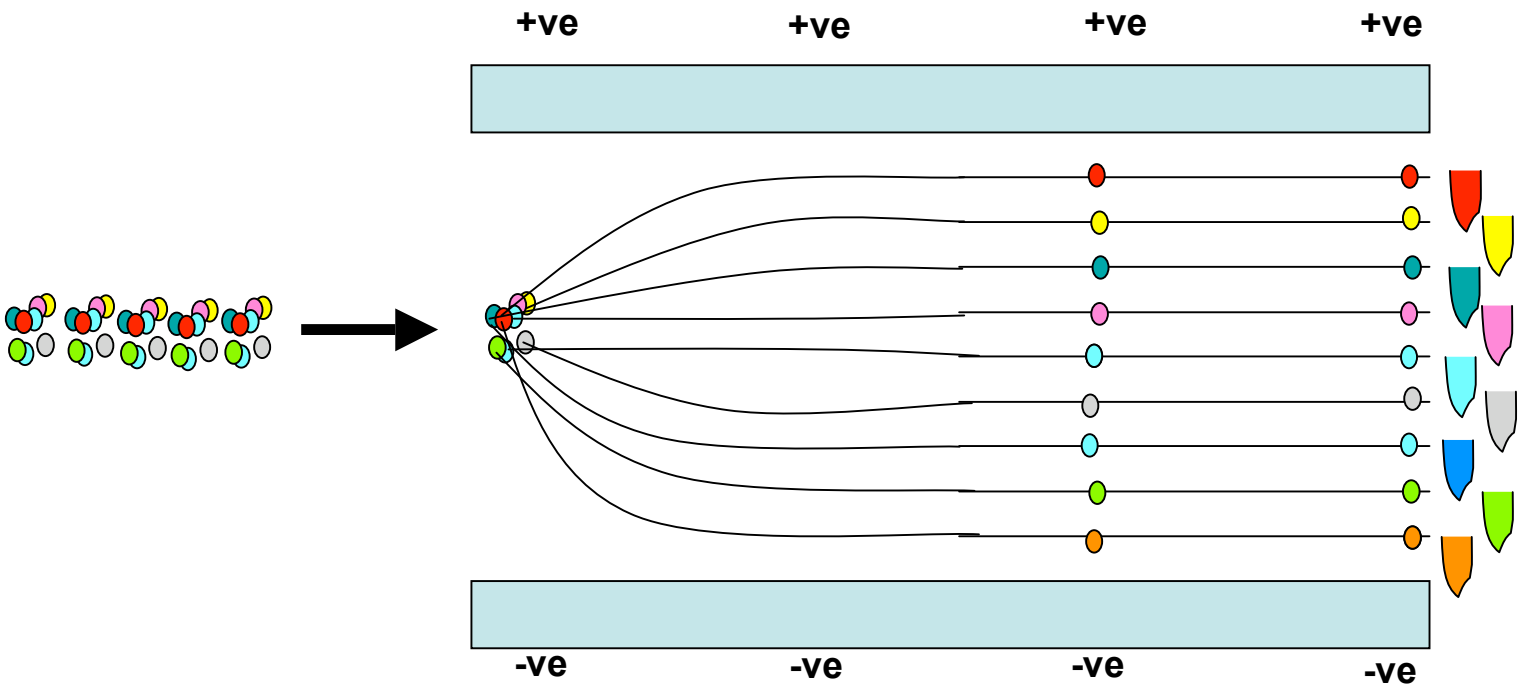
**2nd D: RPLC**



**normal disease**

**Digitized super-imposed images; Image analysis fundamentally different from 2D gels; Quantitation MUCH Faster;**
**CAVEATS?**
**cost;**
**relatively new;**
**reproducibility**

# Free flow electrophoresis




**There are subpopulations of mitochondria in most cells, each with a different net charge. These can be passed through a "chamber" which has a potential difference across the sides of the chamber. Each particle will find a position where its surface charge corresponds to that of the gradient, and move along that position parallel with others moving at their own positions. This allows separation of multiple subgroups of mitochondria and physical collection into 96-well plates at the end of the free-flow chamber.**

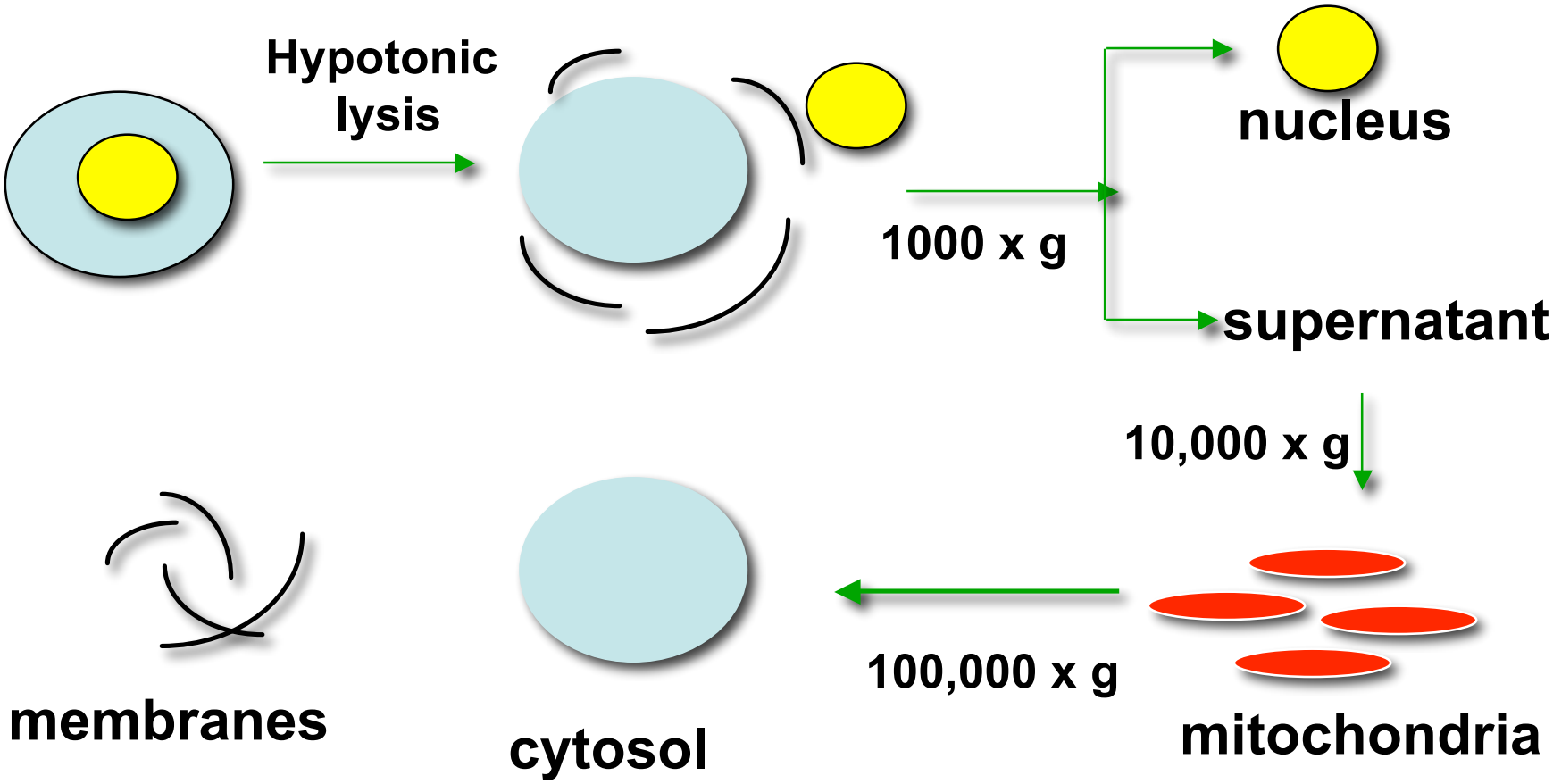
**http://www.bd.com/proteomics/technology/**

# Bottom line for effective 2D gel separation and analysis

- **Reduce proteome complexity by incorporating biological information or properties :**
  - **Intracellular location--subcellular fractions**
  - **Protein-protein interactions--immunoprecipitations, BN gels**
  - **Different states of oligomerization in vivo: microtubules**

## Various ways to reduce proteome complexity:

Subcellular fractionation by differential centrifugation

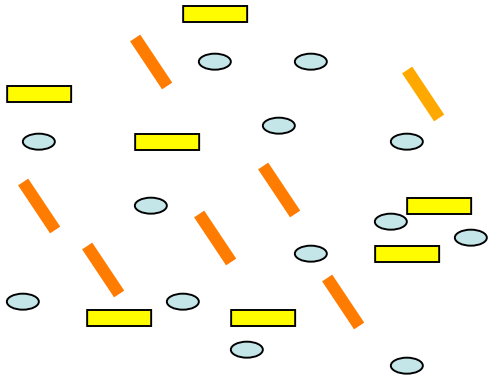


# The good news: subcellular proteomes are readily "catalogued."

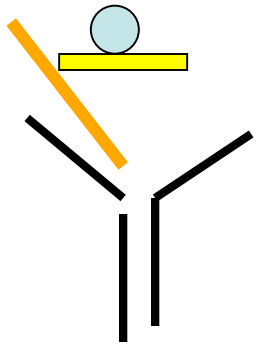
| Compartment | # polypeptides in SWISSPROT as of 2000 |
|---|---|
| Mitochondria (1000/cell) | 269------5% of total |
| Lysosome (400/cell) | 50------1% of total |
| Peroxisome | 35------0.6% |
| ER and Golgi apparatus | 157------3% |
| Nuclei (5% cell volume) | 964-----17% |
| Others (cytosol, membrane) | 4228----75% |
| | total:5703 |

(Jung et al. [2000] Electrophoresis)
Note date of article: this is old data

# Biological specificity of antibodies is invaluable in reducing the complexity of the proteome to be analyzed



**A cell lysate: 6,739 polypeptides**



**An immune complex of 1 - 3 polypeptides (why might there be more than one polypeptide?)**

**Which sample would you rather resolve on a 2D gel?**

# Make use of databases and the internet:

**I. Check existing databases and web-links:**

**www.expasy.org**

***many are annotated***

**helpful links:**

**proteomics tools**

**II. Keep up with the literature/ competition:**

***Electrophoresis***

***Proteomics***

***Molecular & Cellular Proteomics***

***J. Proteome Research***

**III. Use genomics information when available:**

**The polypeptide sequence (from the cDNA) can predict electrophoretic parameters-- m.w. & pI;**

***helpful in setting up 2D gel conditions***

# Take home message

- **Use of proteomics technologies enables *global* analysis of protein changes ;**
- **2D electrophoresis can indicate both differential expression or posttranslational modifications;**
- **Other newer protein separation approaches can be valuable in reducing complexity, or in biological information;**
- **Choice of separation governed by**
  - **Abundance of sample**
  - **Question being asked**
  - **What you can access readily**
  - **Cost**

# Future directions in intact protein analysis approaches

I. DIGE and Cy-dye labelling will enhance 2D gel analysis of complex proteomes;
II. Subcellular fractionation will regain importance in proteome analysis;
III. While automated 2D LC/LC-MS/MS may appear more highthroughput for "discovery," every resolved spot on a 2D gel is a purification, *and* a discovery;
IV. 2D gel positional information, *without protein identities*, is information itself.
V. Where 2D gels may play larger roles is in validation of results generated by other approaches;

# Final thoughts

- What proteomics technology gives back is like any other analytical approach: it's as good as what you put in;
- Be mindful of distinguishing between low abundance proteins vs low level contaminants;
- Keep in mind "conventional" approaches like Western blotting to validate proteomic results;
- Purify, purify, purify before running any proteomic experiment.
- Formulating a hypothesis forces you to incorporate biological information that can enhance proteomic analysis.
- Identifying differentially expressed proteins is a beginning:
  - What changes are causal to the disease/phenotype
    - Some could be "real" but not causal
  - Some could be response of the cells/tissues TO the disease, not causing the disease.
  - How to distinguish between the above?