January 18, 2011

Protein enrichment/purification in proteomics and mass spectrometry

Helen Kim

934-3880

helenkim@uab.edu

Dept of Pharmacology and Toxicology

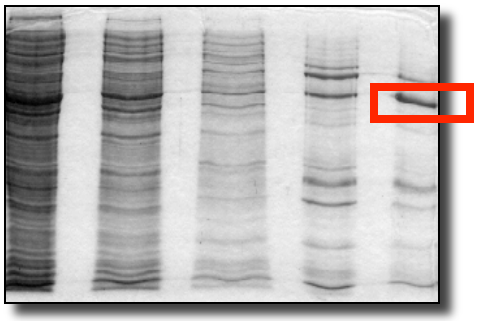
McCallum Building, room 460

http://www.uab.edu/proteomics

Learning objectives

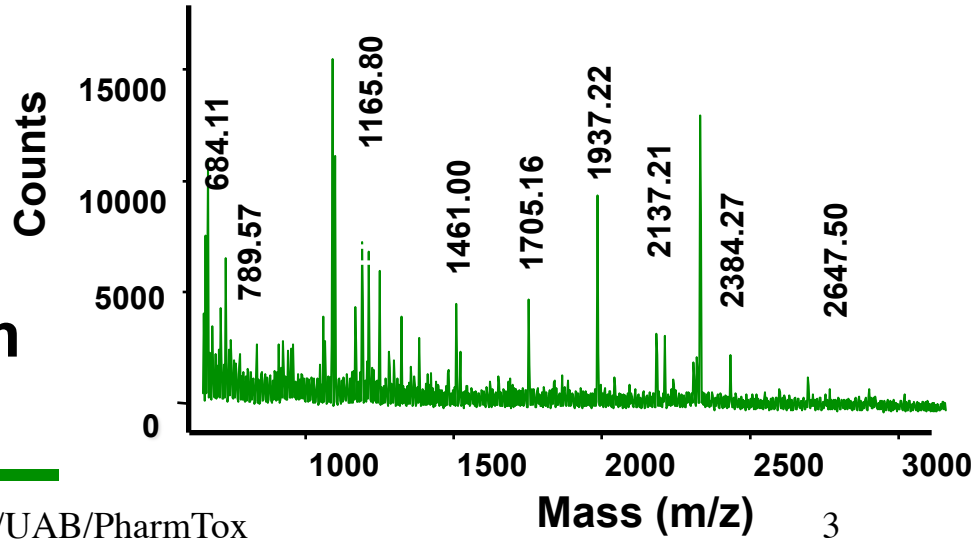
- **Smart proteomics:**
  - **lowest proteome complexity, highest biological specificity**
- **How do we reduce protein complexity**
  - **Purify proteins according to intrinsic properties**
  - **Purify proteins according to biological properties**
- **Proteomics: global assessment of protein differences in biological samples:**
  - **Electrophoresis:**
  - **Western blot**
  - **Importance of controls, normalization and quality control**
  - **Importance of understanding basic chemistries**

The basic elements of intact protein proteomics: (1) separation, (2) analysis, (3) identification and characterization



trypsin

MALDI-TOF mass spectrometry



ID of parent polypeptide

MASCOT search engine

http://www.matrixscience.com

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

Properties of polypeptides that enable separation from each other

- **Intrinsic properties**
  - **Size—number of amino acids**
  - **Net Charge**
  - **Tertiary and quarternary structures**

- **Biological/functional properties**
  - **Intracellular location**
  - **Enzyme activity**
  - **Undergoes oligomerization**
  - **Undergoes modification**

Separating proteins by size

- For purification: Gel filtration chromatography
- Analytic: SDS-PAGE

Separating proteins by charge

- Purification: Ion exchange chromatography
- Analytic: isoelectric focussing

Protein functional properties that can be basis for purification

- Oligomerization:
  - Each cytoskeletal component undergoes reversible oligomerization from its monomers
- Intracellular location:
  - Change in
  - Increase in same location
- Protein-protein interaction

Protein posttranslational modifications and approaches for purification

- Phosphorylation: mass change of 80 daltons
- Glycosylation: mass change (xxx to xxxxx)
- Oxidative modifications:
  - Cysteine oxidations: SH to sulfene:
  - SH to sulfone:
  - SH to SS:

  Reactive aldehyde adduct formation: 4HNE:

  mass changes (xxx, yyy, zzz)

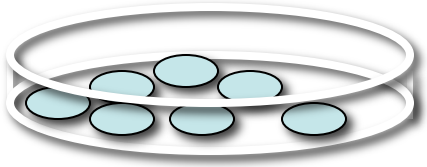
Protein posttranslational modifications and approaches for detection/purification

- Phosphorylation: affinity with antibody, Western blot,
- Glycosylation: affinity with lectin
- Oxidative modifications:
  - Cysteine oxidations: SH to sulfene:
  - SH to sulfone:
  - SH to SS:

  Reactive aldehyde adduct formation: 4HNE:

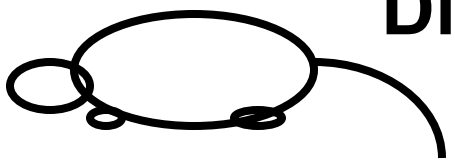
  Western blot, affinity with antibody

Sample preparation for 2D electrophoretic analysis:



Harvest, rinse, and pellet the cells;

or



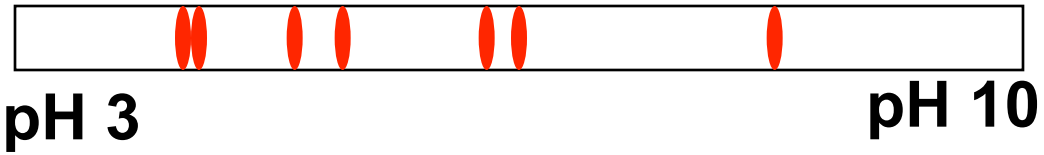
Dissect out tissue, organ, or fluids;

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

What 2-D electrophoresis involves:

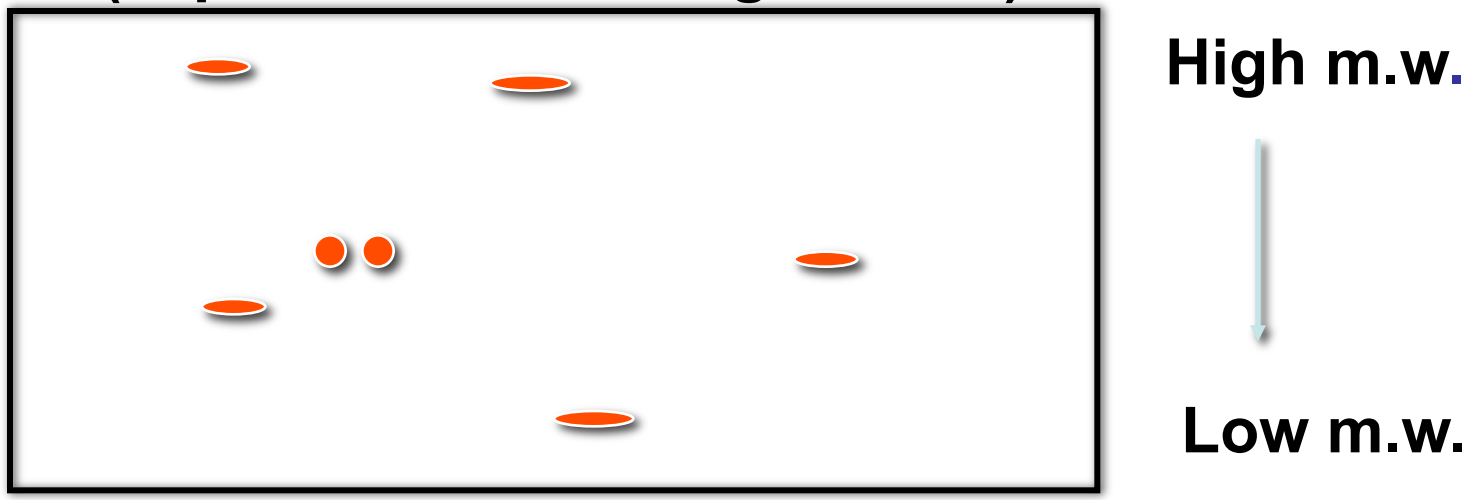
- 1st dimension: **Isoelectric focusing**

(separation according to charge)



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

(separation according to size)

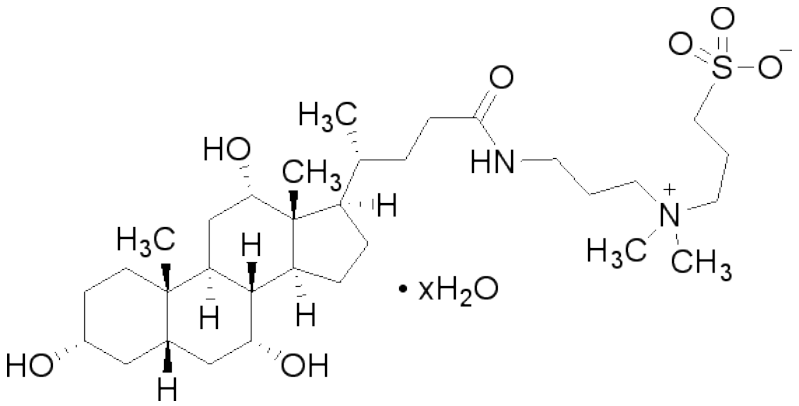


Chemical components in protein sample buffers that distinguish 2D gels from 1D gels.

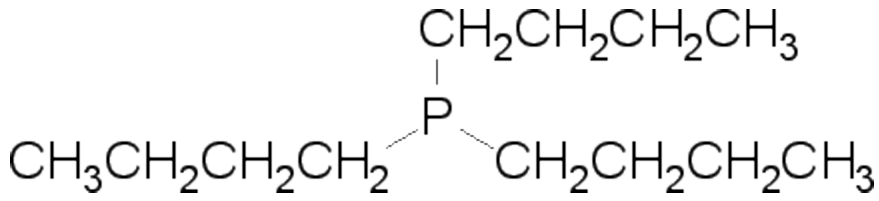
SDS: why do we not want SDS in our initial 2D gel sample buffer
CHAPS:
Which is the ionic detergent? Which is the non-ionic?
Why is it important to know the concentrations of each if switching from SDS to IEF or vice versa.
β-mercaptoethanol, TBP

Remember the ratio: 1.4g/g protein
QUERY: which part of SDS binds to the protein.

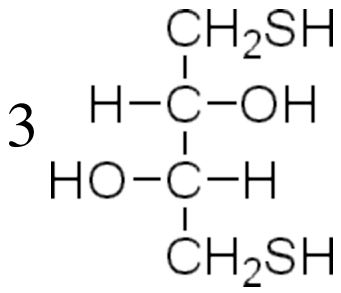
Structures important in protein electrophoresis to know and understand



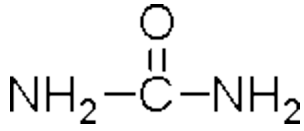
1



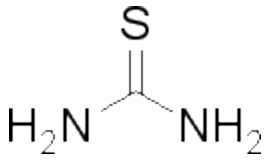
2



3



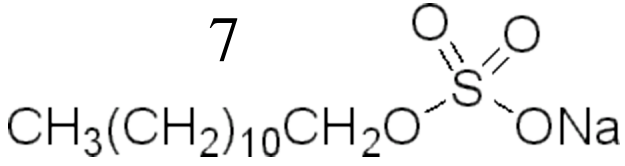
4



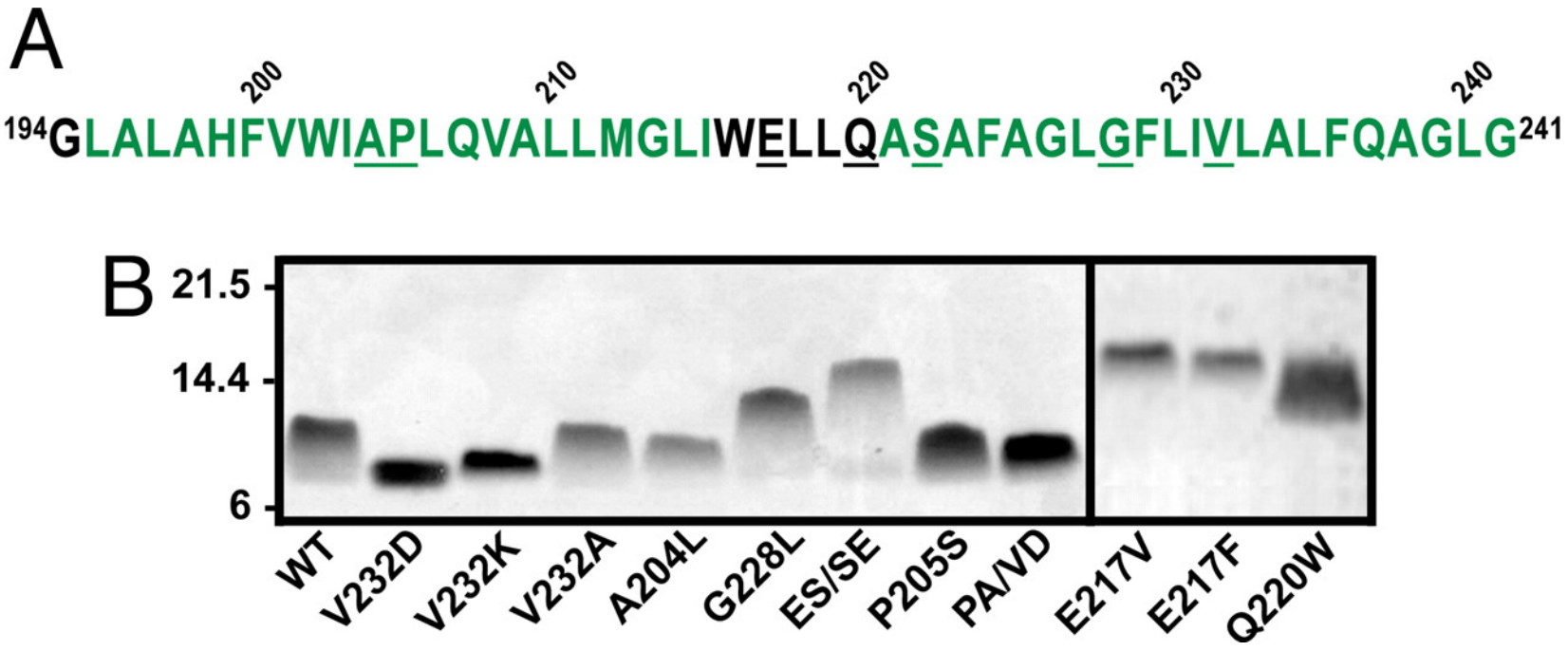
5

6

7

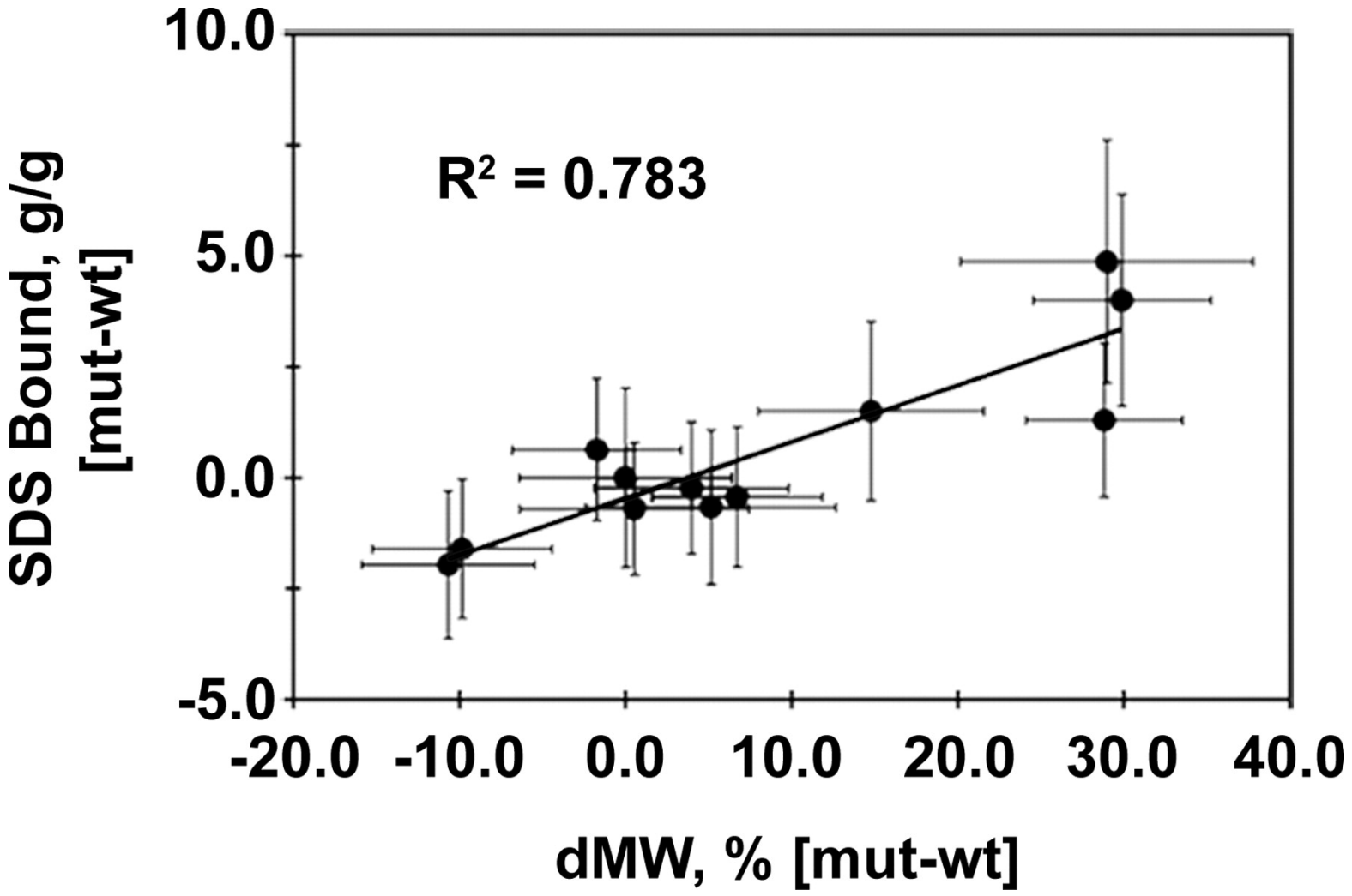


Hairpin sequences and SDS-PAGE analysis.



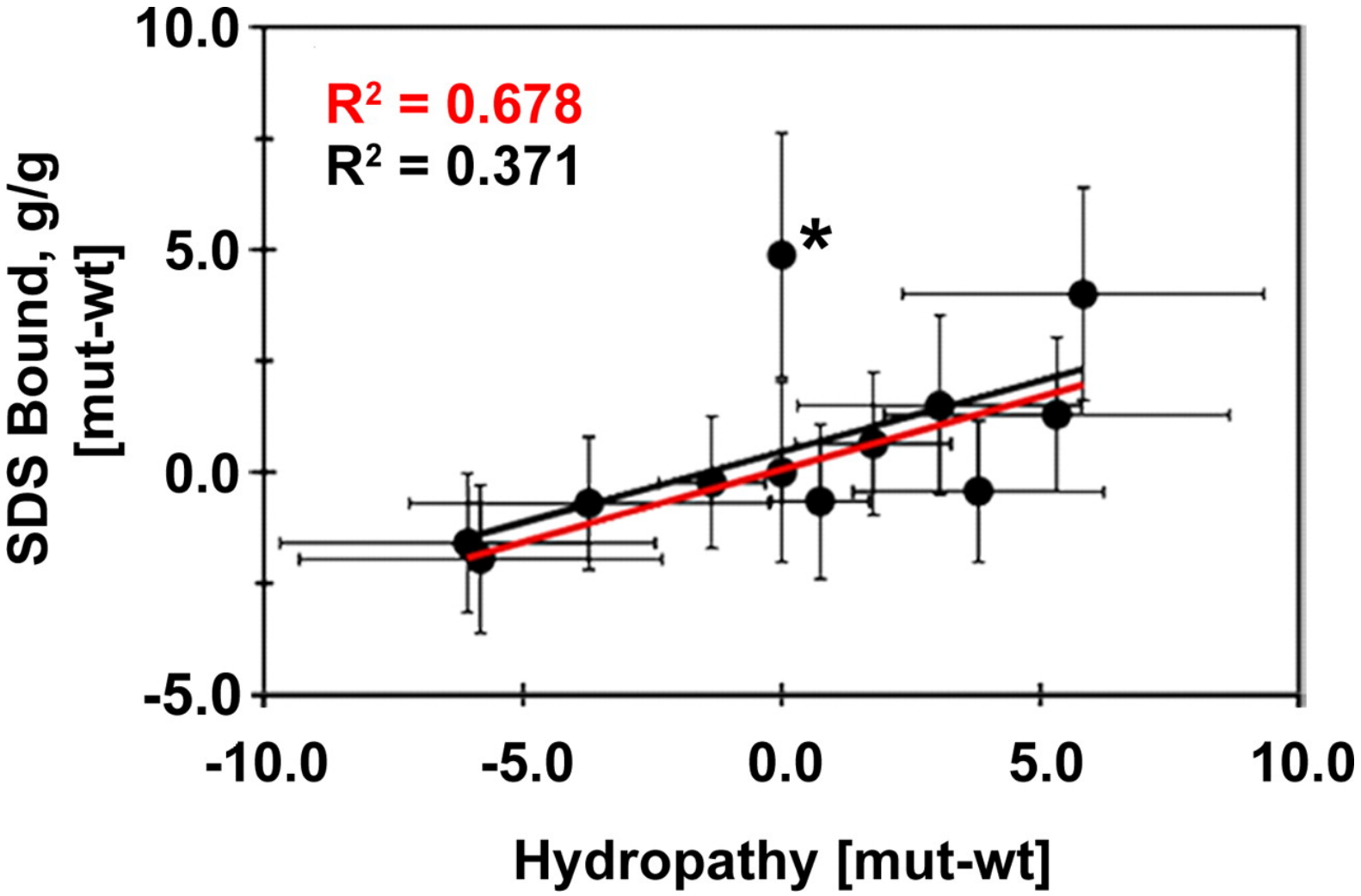
Rath A et al. PNAS 2009;106:1760-1765

Correlation of SDS binding and PAGE mobility.



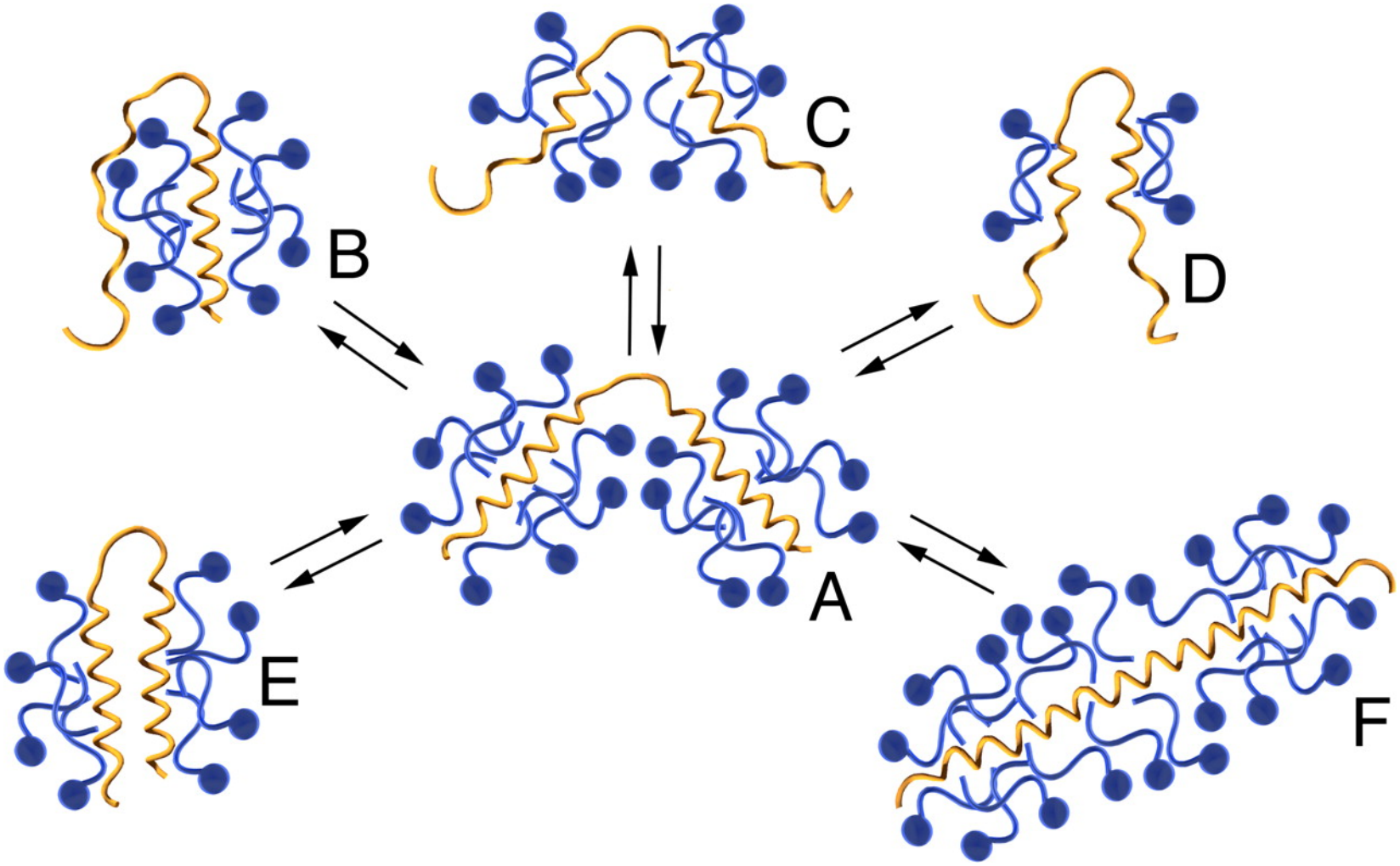
Rath A et al. PNAS 2009;106:1760-1765

Correlation of SDS binding and hydropathy.



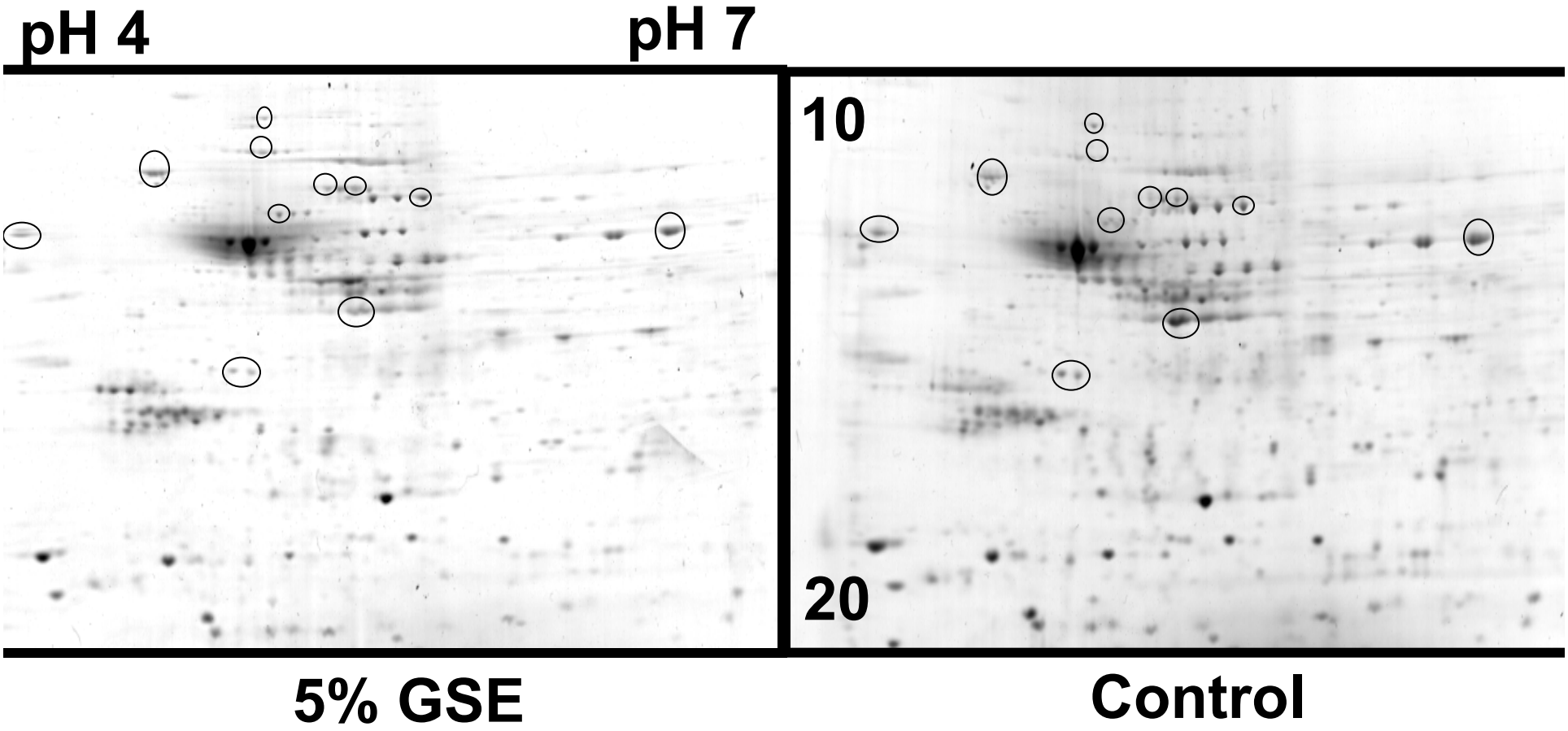
Rath A et al. PNAS 2009;106:1760-1765

Interrelationship between hairpin conformation and detergent binding.



Rath A et al. PNAS 2009;106:1760-1765

A pair of 2D gels from a real experiment: what differences are observed?



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

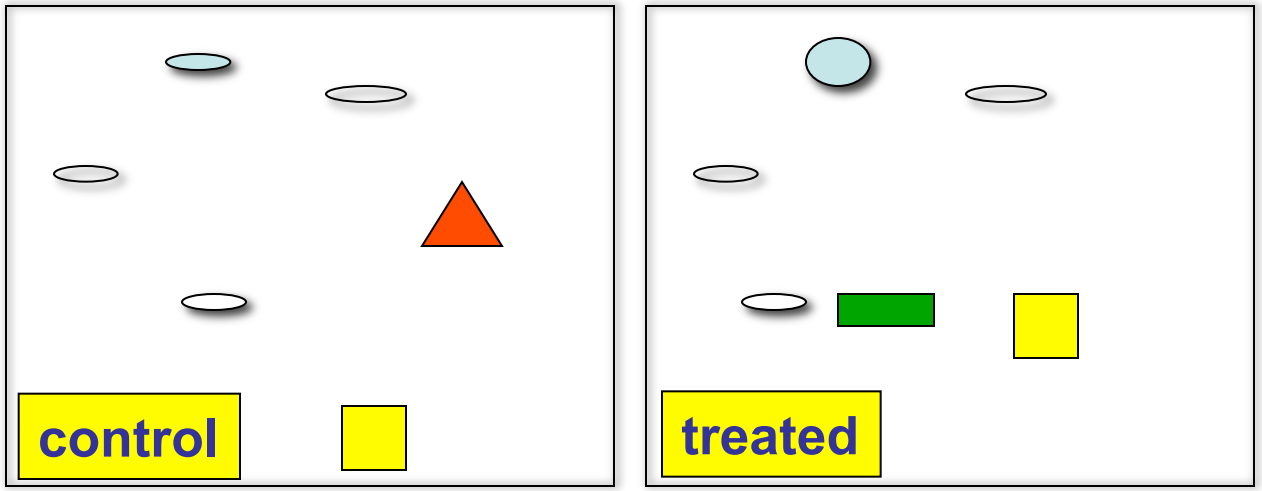
1. Compare the 2D displays of spots
2. Determine total spot number for each display
3. Quantify spot intensities, identify differences
4. Identify spots that may have " moved" horizontally;

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?"

Critical part of 2-D gel proteomics: Image analysis

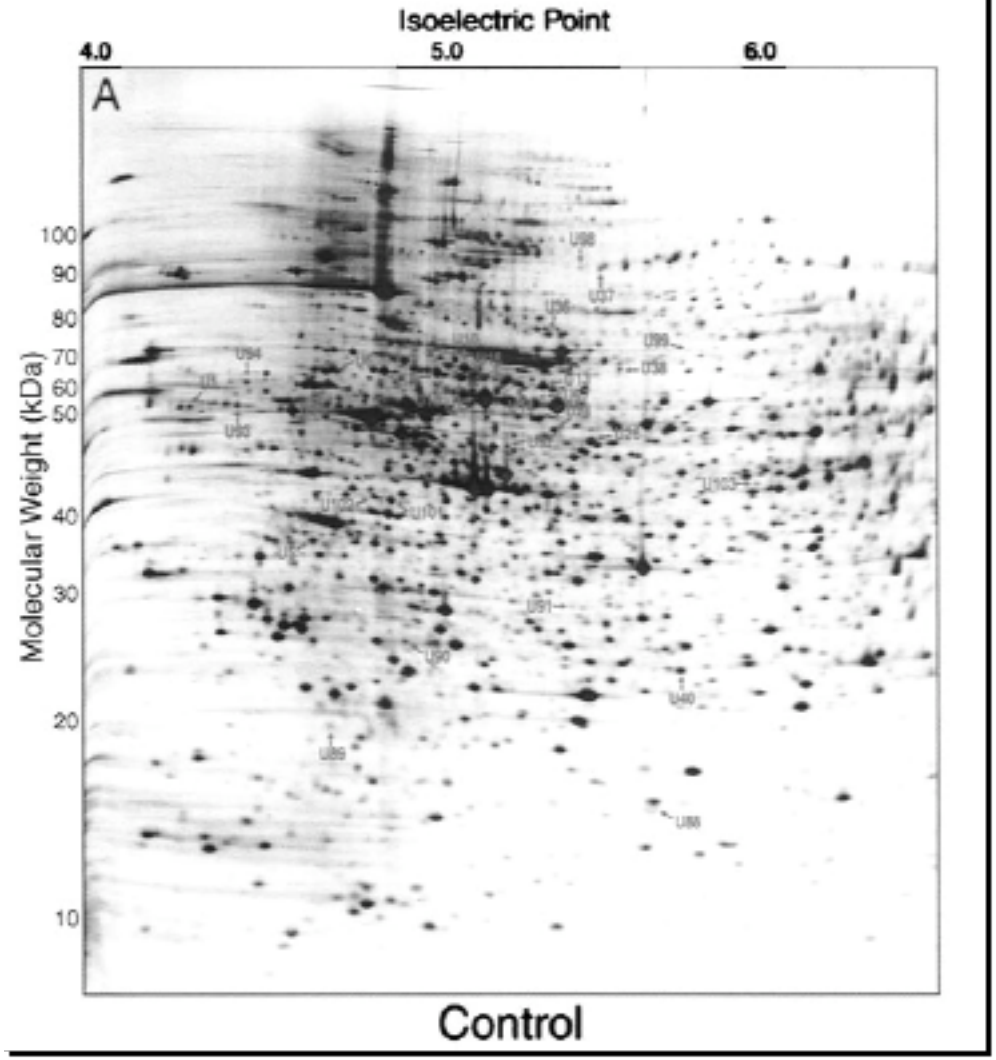


with software: "compare" the images.

Types of information:

- Suggests upregulation of gene
- Suggests new posttranslational modification
- Suggests downregulation of gene
- Suggests "aberrant processing:" the different size and pI indicate part of the protein in control is different from in treated.

A real single-sample/single 2-D gel can be VERRRY complex sometimes



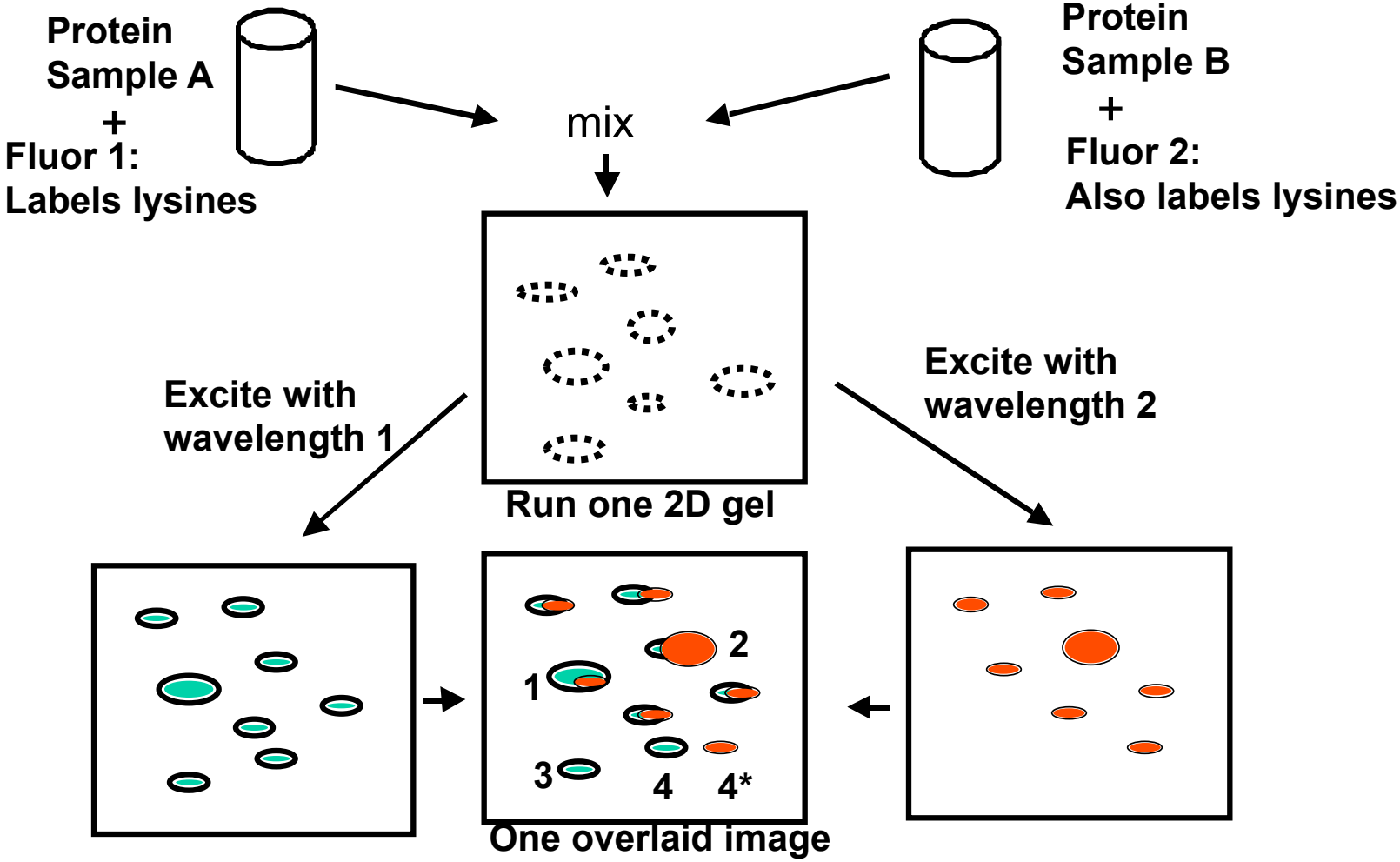
Find this and other 2-D gels at http:www.expasy.org

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

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

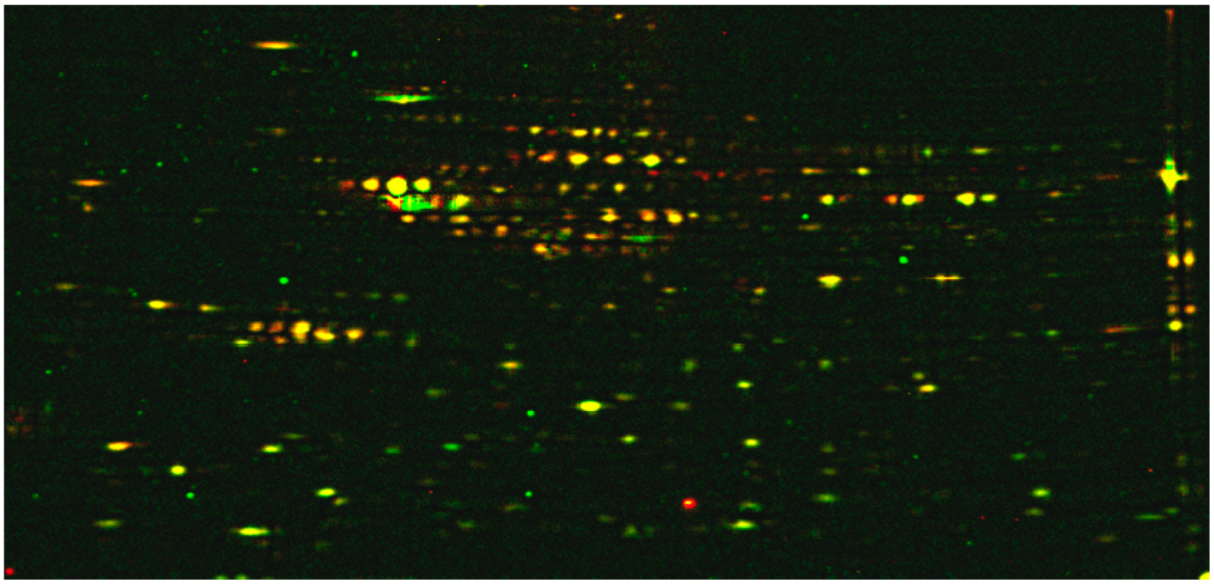
However, you would rather not have to work with 1500 spots!!!

2D difference gel electrophoresis (DIGE)--- more powerful than single sample-single gel



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

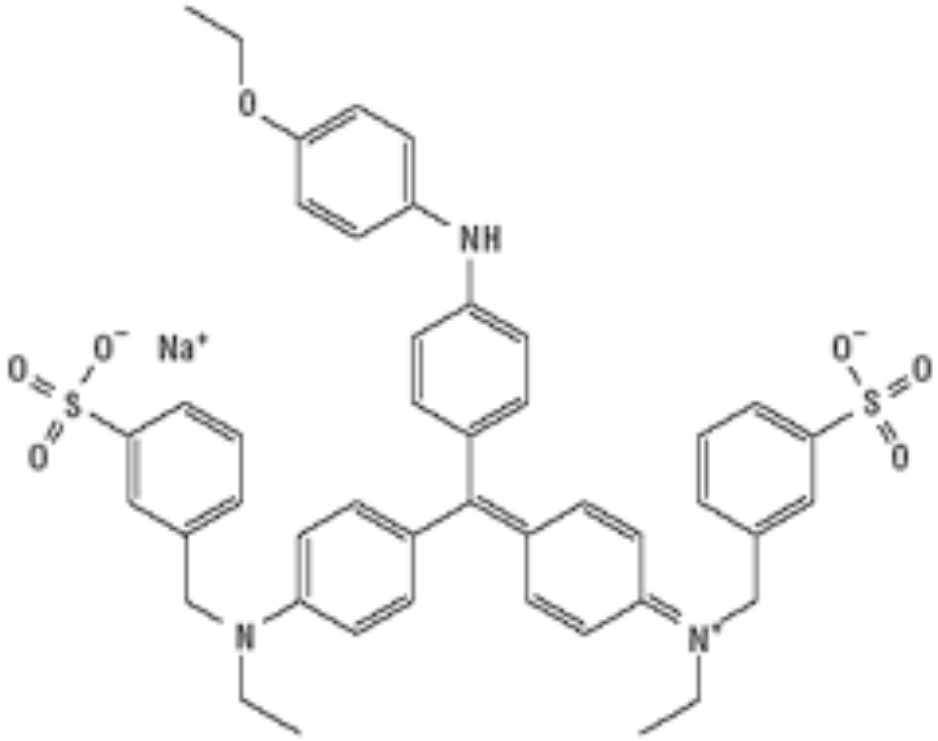


100% green spots are specific to normal cells, 100% red spots are specific to cancer cells, degrees of yellow-orange indicate differential expression.

"Mine" your 2D gel 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;
- 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

Visualizing 2D gels: Coomassie Brilliant Blue



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

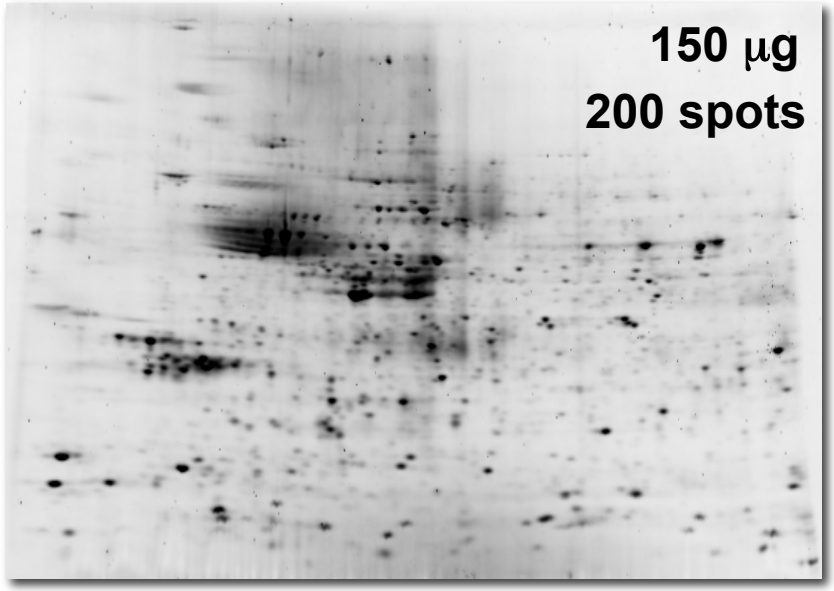
In acidic conditions, the anion of CBBR combines with the protonated amino groups on proteins via electrostatic interactions.

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

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.

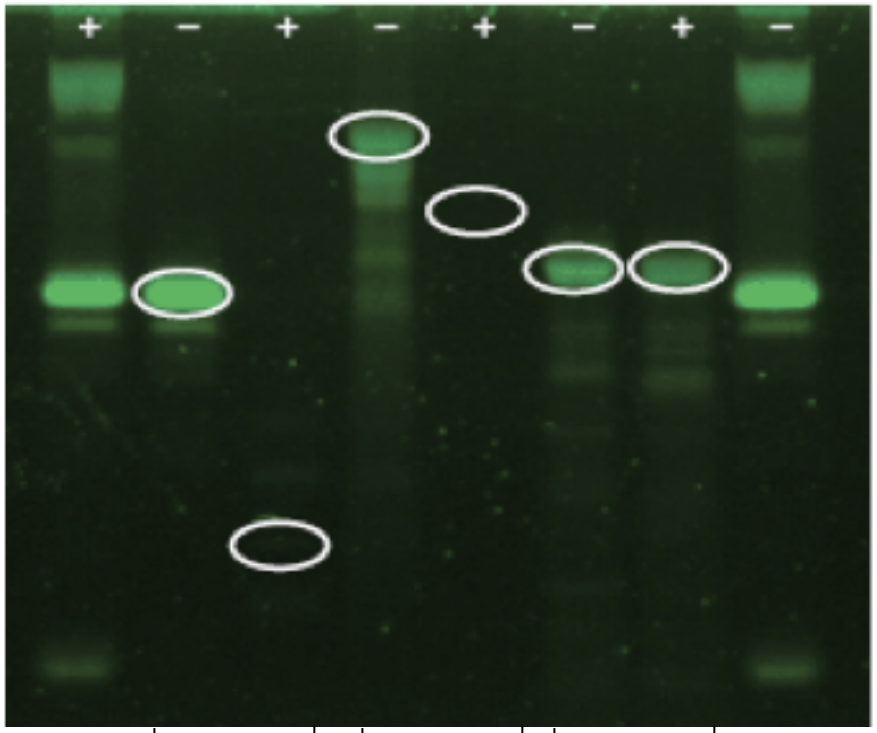
Conclusion: A fluorescently stained 2D gel of an unfractionated sample, only allows detection of the "low hanging fruit."

Wide dynamic range: rationale for protein enrichment

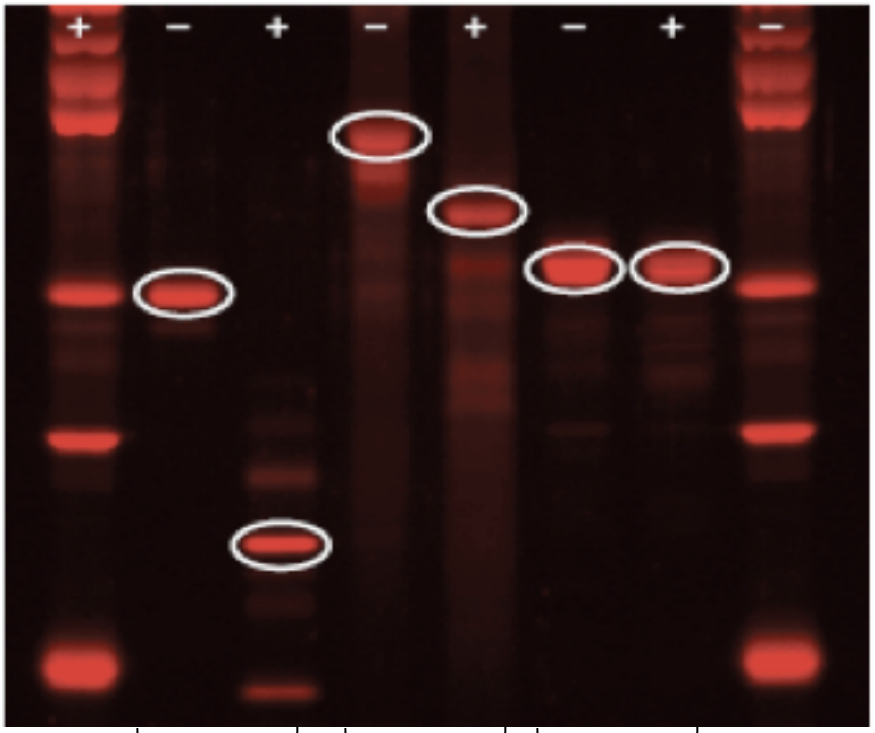
- **A gel that is overloaded with respect to the abundant proteins, may have *barely detectable* amounts of a low abundance protein**
  - **If you can't see it, you dont know it's there;**
  - **Even if software detects and quantitates it, you can't do MS of it, because there isn't enough protein.**
- **No one stain will detect 9 orders of magnitude differences in abundance of proteins.**

Multiplex Proteomics: ProQ Emerald followed by Sypro Ruby identifies multiple glycosylated proteins at once

ProQ Emerald



Sypro Ruby



A- A B- B C- C

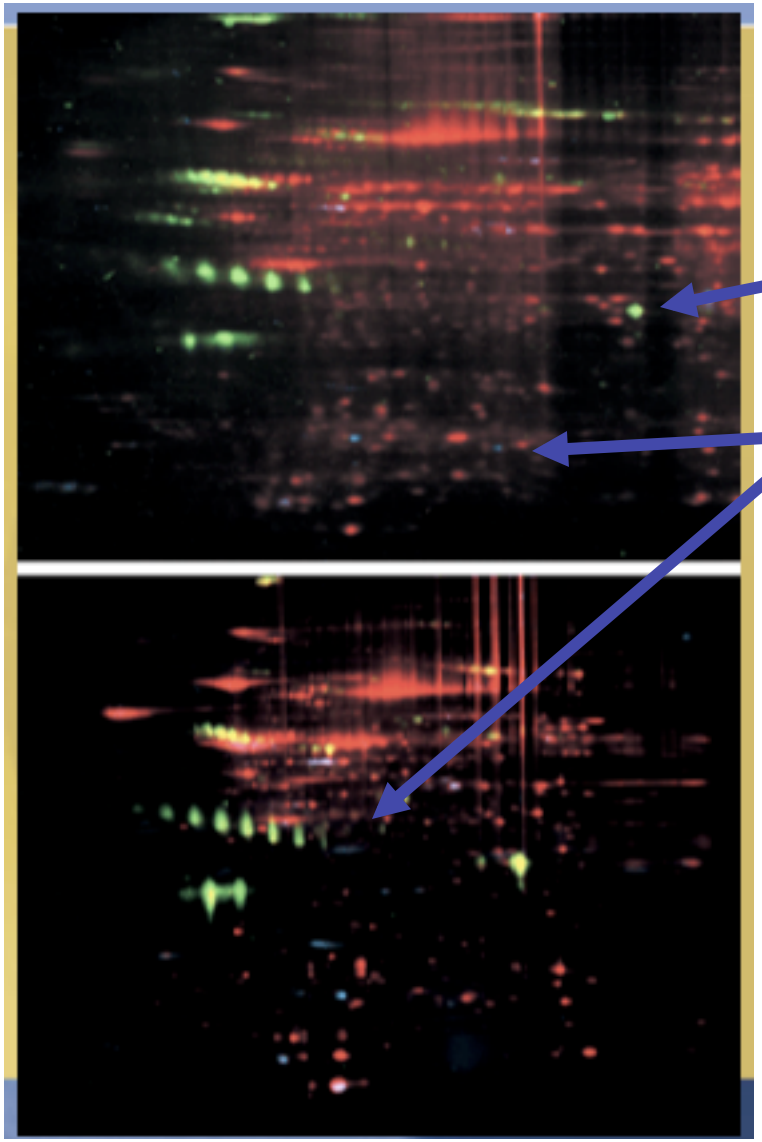
A- A B- B C- C

(Duncan Veal, Fluorotechnics Pty, Ltd.)

Multiplexing: valuable when sample is scarce or difficult to obtain

Normal liver

Diseased liver



Sypro Ruby

Pro Q Emerald

Pro Q Diamond

The value of fluorescent dyes: each is spectrally distinct, i.e. it can be detected separately from others on the gel, due to different excitation and emission spectra.

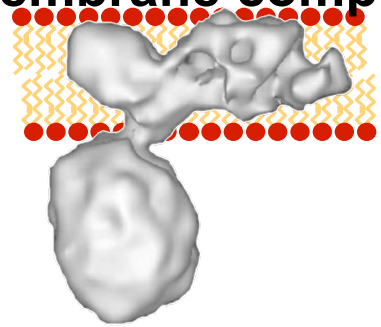
Modified from Duncan Veal, Fluorotechnics Pty, Ltd.

SUMMARY of SDS-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;**
- **Each stain has utility depending on experimental goals, and sample abundance.**

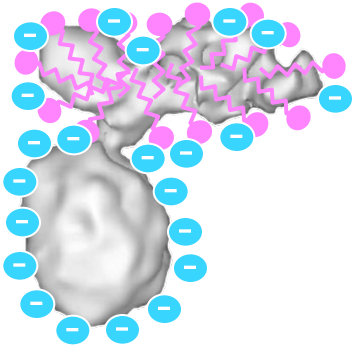
2D-Blue-Native gels: approach for studying membrane-associated protein complexes.

Intrinsic mitochondrial membrane complexes



Detergent

CBBR

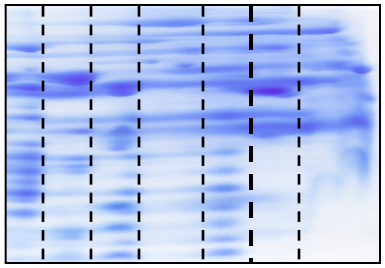


Released complexes, all negatively charged, thanks to the CBBR

1st D: NATIVE electrophoresis

(Modified from Brookes et al., 2002)

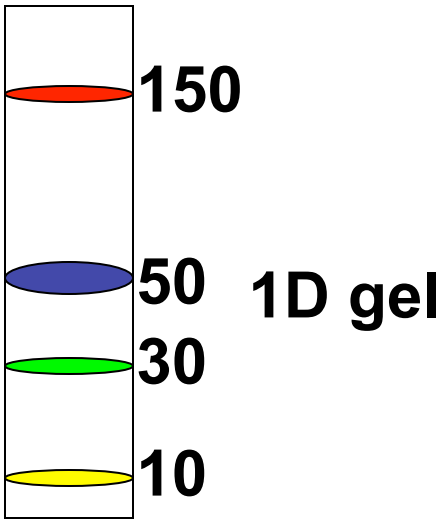
This type of 2D gel has "ladders" of bands.



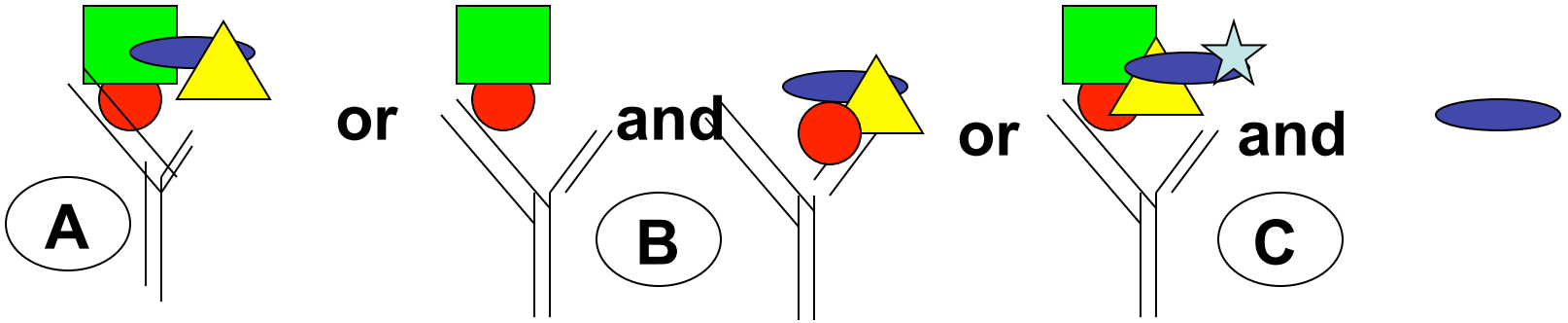
2D native electrophoresis has value beyond identifying components of membrane complexes:

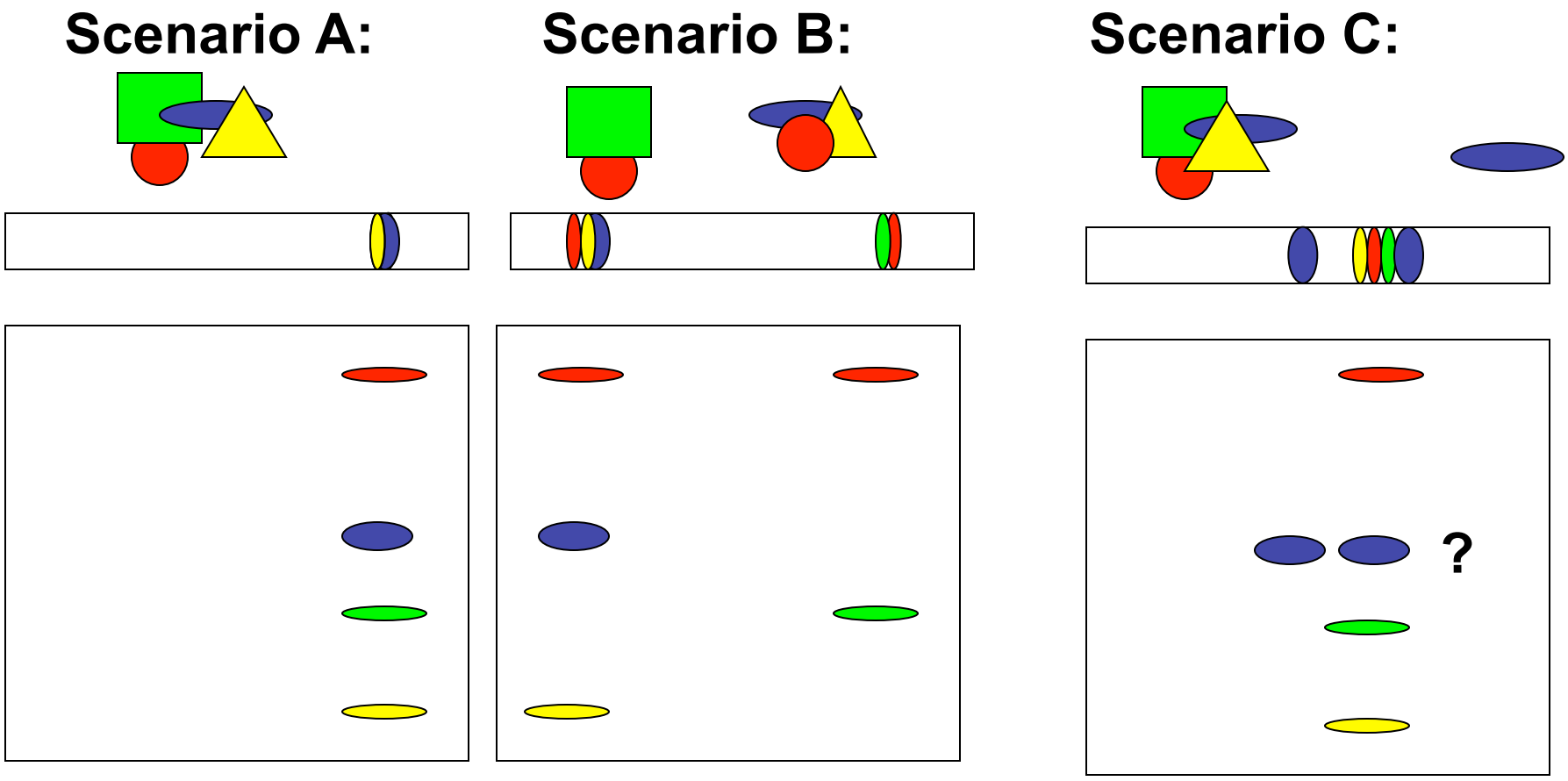
In immunoprecipitates: Which proteins are interacting with which?

Experiment → Lysate + antibody →



It could be:

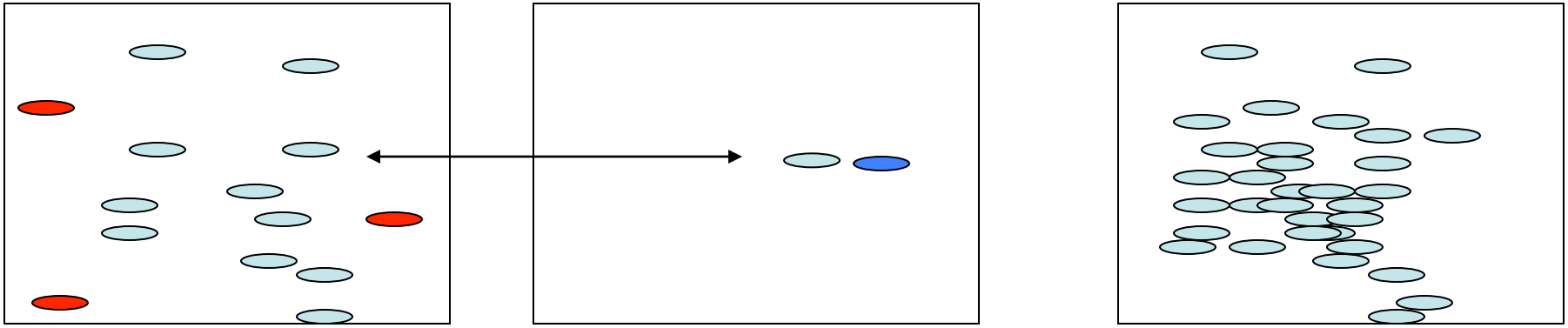




When would you see the same protein in two lanes in the 2nd D?

For greatest sensitivity, and the most biological information:

Combine proteomics with "conventional" approaches i.e. Western blot analysis;



Western blot for phosphoproteins

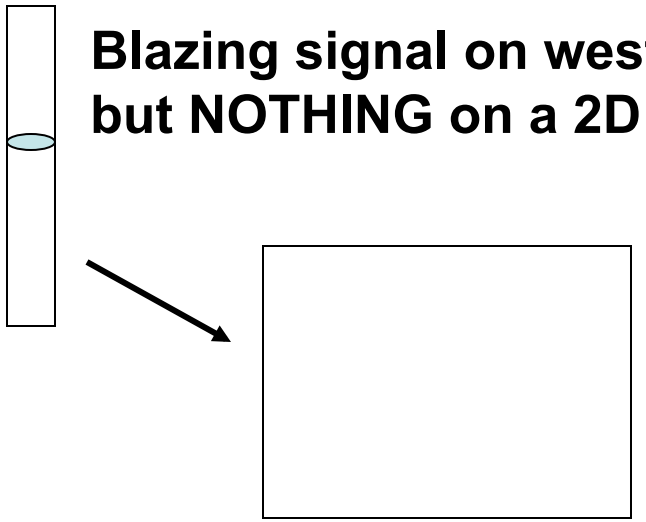
Western blot for a particular protein

Stained gel

Question: what do the red spots in the left blot tell you about those proteins in the stained gel; what does the blue spot in the middle blot tell you?

Issues in going from blots to 2D gels:

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 immunoprecipitate:

Explanations?

1. Solubilization by SDS but maybe not by CHAPS; ways to check
2. Quantity---Western blots more sensitive than people realize; need to scale up at least 10-50 fold for detection by stain

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

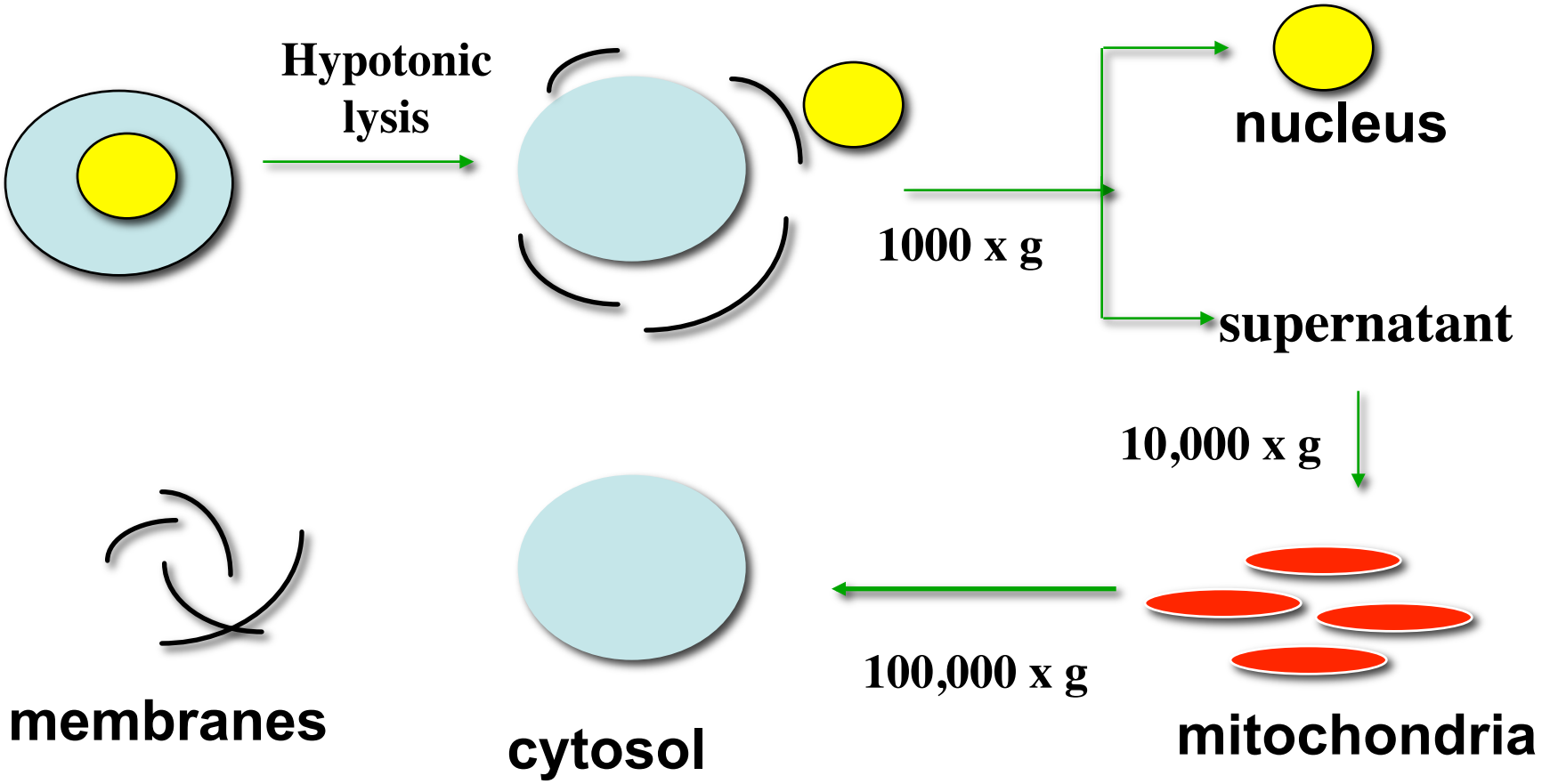
1. maximim resolution of proteins associated with the antigen;
2. Separate multiple proteins in one band
3. Ensure IgG bands are not hiding proteins of interest

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**

Functional properties of proteins: intracellular location: Basis for reduction of proteome complexity:

Subcellular fractionation by differential centrifugation



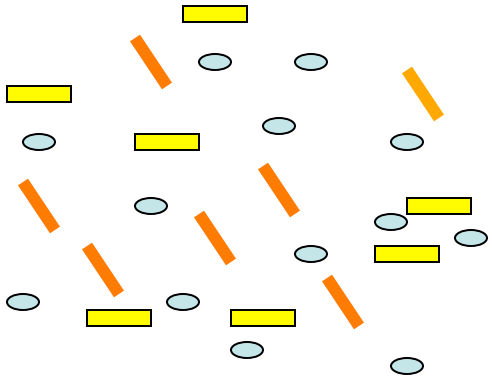
The good news: subcellular proteomes are readily "catalogued."

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

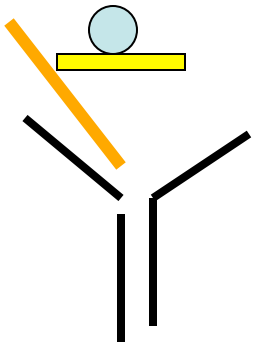
(Jung et al. [2000] Electrophoresis)

Note date: this is old!!! But the principle is the same

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



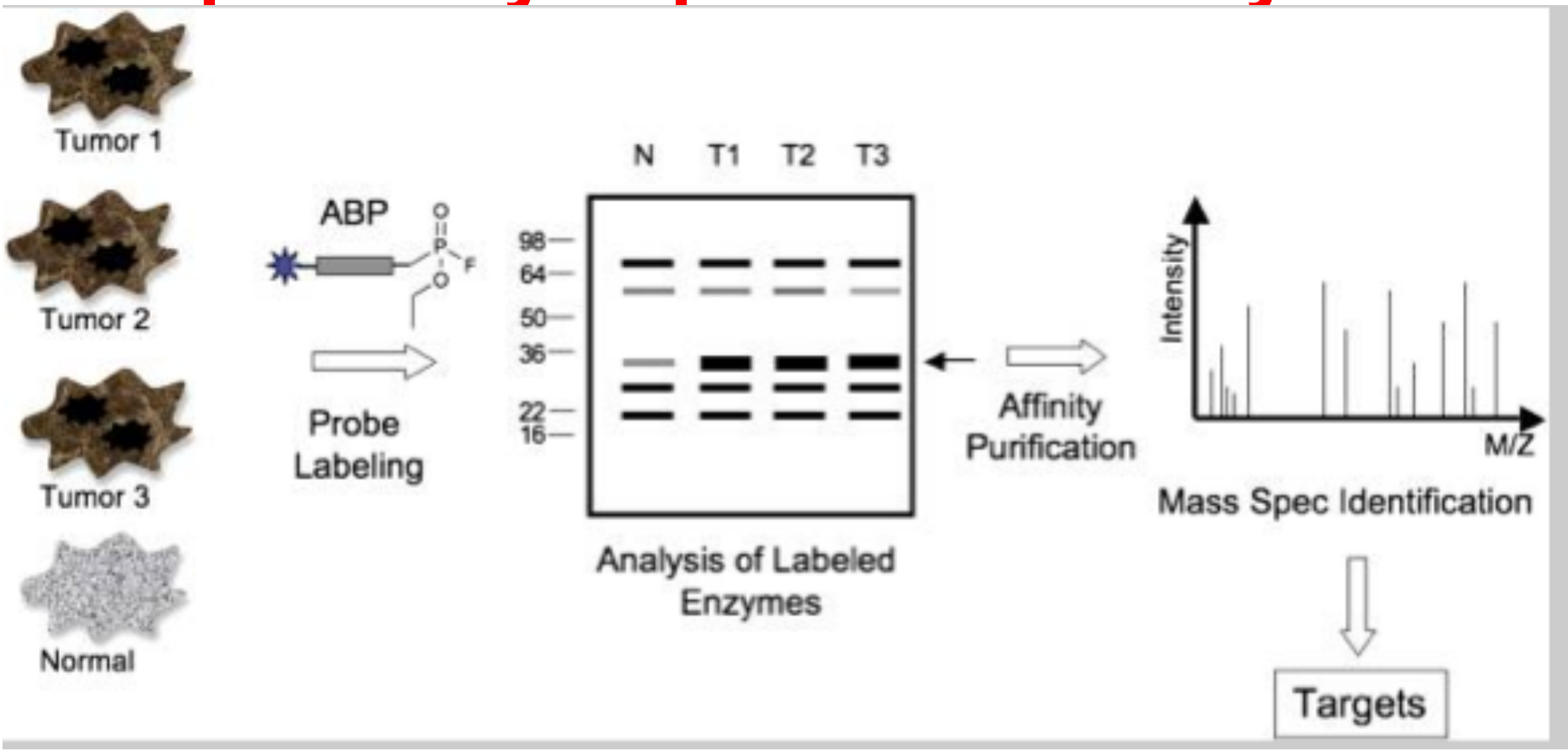
A cell lysate: 6,739 polypeptides



An immune complex of 1-10 polypeptides (why might there be more than one polypeptide?)

Which sample would you rather deal with on a 2D gel?

Activity-based protein profiling (ABPP) to increase biological specificity of proteomic anslysis



Bogyo, PNAS, Feb, 2010

Affinity purification for studying certain PTMs

- Dan Liebler's work studying 4-hydroxynonenal-modified proteins using "click chemistry"

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:

J. Biol. Chem.

Proteomics

Molecular & Cellular Proteomics

J. Proteome Research

J. Agric. Food Chem.

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

- **2D electrophoretic patterns yield mass and charge information, thus expression or posttranslational modification differences;**
- **This biological information is not easily obtained in analysis of digests.**
- **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 technologies you can access readily**
  - **What you can afford**

Future directions in intact protein analysis

I. **Subcellular fractionation will regain importance in proteome analysis;**
II. **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;**
III. **2D gel positional information, *without protein identities*, is information itself.**
IV. **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: the quality is 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:
  I. What changes are causal to the disease/phenotype
  II. Some are "real" but not causal;
  III. Some could be response of the cells/tissues TO the disease, not causing the disease.