Introduction to proteomics: analysis of proteins in complex biological samples

Helen Kim, PhD

Department of Pharmacology & Toxicology

& Director, 2D Proteomics Laboratory

934-3880

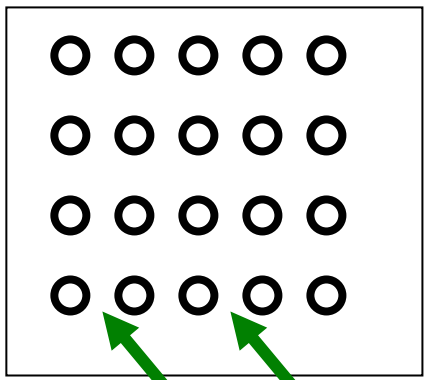
helenkim@uab.edu

McCallum Building, room 460A

Learning objectives

- What is proteomics?
- How proteomics complements genomics approaches in modern research
- Qualitatively different proteomics methodologies
- Importance of reducing the complexity of the protein sample prior to proteomics analysis
  - Rationales
  - Methods
- Intrinsic properties of proteins allow various types of separations and pre-fractionations that should precede proteomics analysis

Genomics: gene microarrays: the good, the bad, and the be-careful-what-you-ask-for.

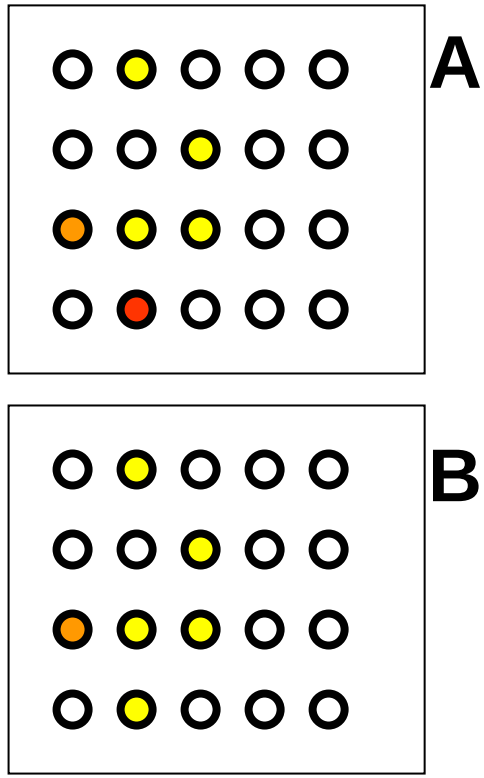


Dots of cDNA's

A gene "chip" or microarray

mRNA's, then fluorescent probe

The good: high-throughput



The bad: list of 528 genes that are "up" or "down"-regulated.

"Potential" proteins Involved in the disease

Genomics methods allow us to determine the changes in expression of tens of thousands of genes at once.

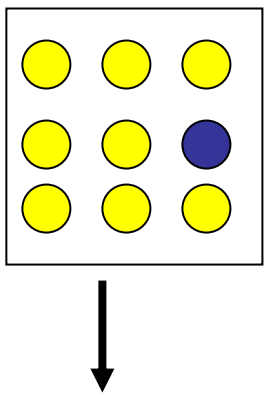
----------So…….why study changes in proteins?

Correct answer #1: The correlation between gene expression and protein levels is poor at best.

- **See graph from Ideker et al., 2001 1-6-04 lecture.**
- **Sometimes changes in the critical proteins are barely detected, because the cell attempts to keep homeostasis with regard to these proteins.**

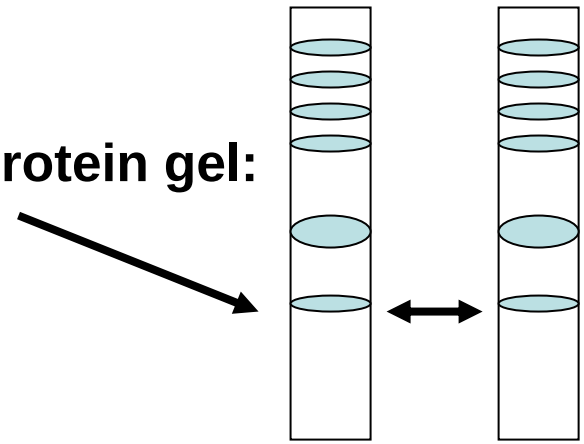
Upregulation of a gene for a crucial protein may have poor correlation at the protein level: example A.

- **May be manifested as little or no change in amount, because the cell "damps" out the change.**



The mRNA for protein X is increased 6-fold.

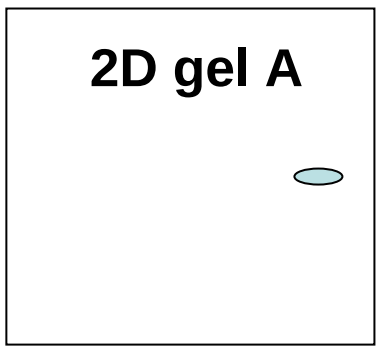
Whereas, on the protein gel:



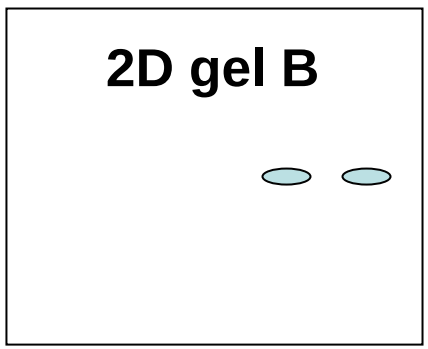
Control Treated

Upregulation of a gene for a crucial protein may have poor correlation at the protein level: example B.

- **May be manifested as a difference in posttranslational modification, as the cell tries to "inactivate" excessive amounts of the protein, by altering it chemically.**



control

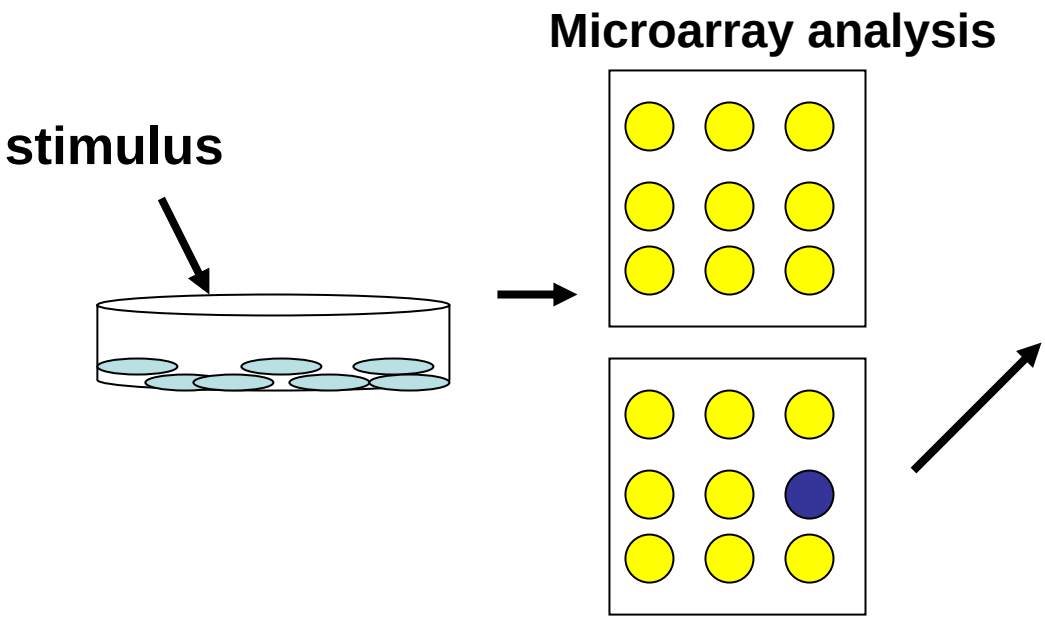


treated

So, the modified protein is not part of the response to the stimulus, but rather, the system telling us that increased amounts of this protein were toxic.

-------How do we know the difference?

Gene expression data may predict the nature of protein modifications, *but not which proteins are modified*.



A kinase is upregulated;

But----which substrates are phosphorylated???

Summary: Why do we need to do broad scale "proteomics"?

- **When you don't have a clue;**
- **When you have only a very small clue;**
  i.e. you've done a microarray experiment, and you have a list of 3,284 genes that are differentially regulated in your system;
- **When you knock out a gene (and hence a protein) that you're convinced is essential for life and health, and the animal pees as usual.**

At best, genomics and proteomics data complement each other.

Neither by itself gives the whole story.

What *is* proteomics?

- **Genomics: study of genomes of a cell or organism**
- **Proteomics:**
  - **Original definition: study of the proteins encoded by the genome of a biological sample**
  - **Current definition: study of *the whole* protein complement of a biological sample (cell, tissue, animal, biological fluid [urine, serum])**
  - **Usually involves high resolution separation of polypeptides at front-end, followed by mass spectrometry identification and analysis**

Rationales for proteomics approaches in today's research

- **Identify a "marker" protein(s);**
  - **Cancer detection/Monitor response to chemotherapy**
  - **Identify one pathogen from others;**
  - **Distinguish a virulent strain of pathogen from nonvirulent.**

Proteomics & disease analysis, part I:

- **Characterize protein differences between disease and normal tissues--**
  - **For understanding the disease process;**
  - **To develop drug targets;**

- **In cancer, there may be novel proteins due to chromosome instability (ETV6-ABL and BCR-ABL), or inappropriate expression may occur (proteins from embryonic or fetal stages of development)**

Where there is pathology, but genetic basis unknown, proteomics can have critical role in identifying other proteins

Two major diseases:

- **HIV: protease is targeted today;**
  - **are there other proteins, either viral or host, that could be targeted to better deal with the disease?**
- **Alzheimer's disease: 3 known mutations (APP, PS1, PS2) and risk factors (ApoE, estrogen loss);**
  - **50% of AD patients do not have any of the known genetic abnormalities, yet all become demented, all have amyloid plaques and NFT in their brains.**
  - **Remember, every AD patient has AD 100%.**

So, what do we know thus far?

- **Do the experiment to generate enough protein;**
- **If you have less than one million cells (one 10 cm cell culture dish), you'll only be able to detect proteins with copy numbers >100,000**
- **1 g of tissue = 10^8 cells (copy # >1,000)**

Proteome-wide methodologies
(to be covered 1-20-04)

- 2D-isoelectric focusing/SDS-PAGE
- 2D-automated protein liquid chromatography
- MUDPIT; 2D-strong cation exchange-reverse-phase chromatography
  - MUltiDimensional Protein Identification Technology)
- SELDI; Surface-Enhanced Laser Desorption Ionization - protein chips

The whole proteome contains
proteins, proteins, proteins,….

A typical cell proteome can include more than 60,000+ polypeptide isoforms;

- **A really good 2D gel resolves 3,000 proteins;**
- **This is only 5% of the total.**

Avogadro's number comes back to haunt us: a reality check for proteomics

- One gram-mole of anything has 6.02×10^{23} molecules
- For mass spec detection, we need at least 100 femtomole (10^{-13} moles) = 6.02×10^{10} molecules

# expressed per cell	need # cells for 100 fmoles
10	6.02×10^{9}
100	6.02×10^{8}
1000	6.02×10^{7}
10,000	6.02×10^{6}
100,000	6.02×10^{5}

On the other hand: How do we deal with the complexity in proteomics:

as little as possible:

i.e. reduce the numbers of proteins that you have to analyze/separate.

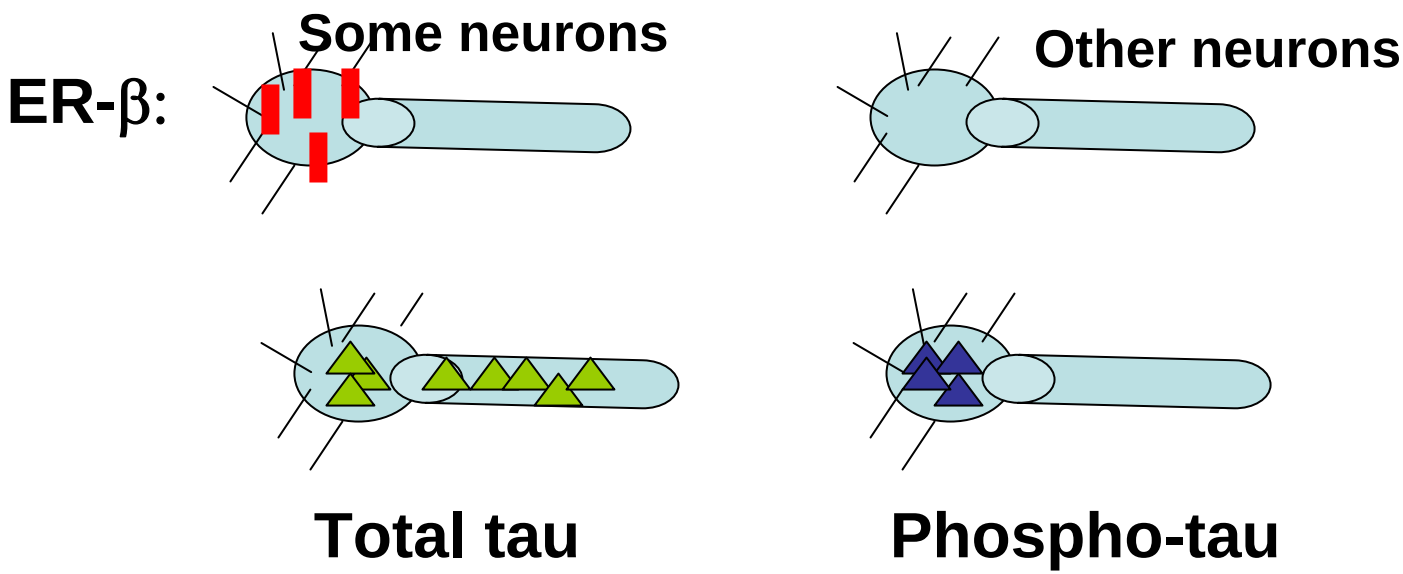
Issues in proteome complexity: I. dynamic range

Cellular proteomes involve a very wide *dynamic range*: proteins can differ in their amounts by nine orders of magnitude;

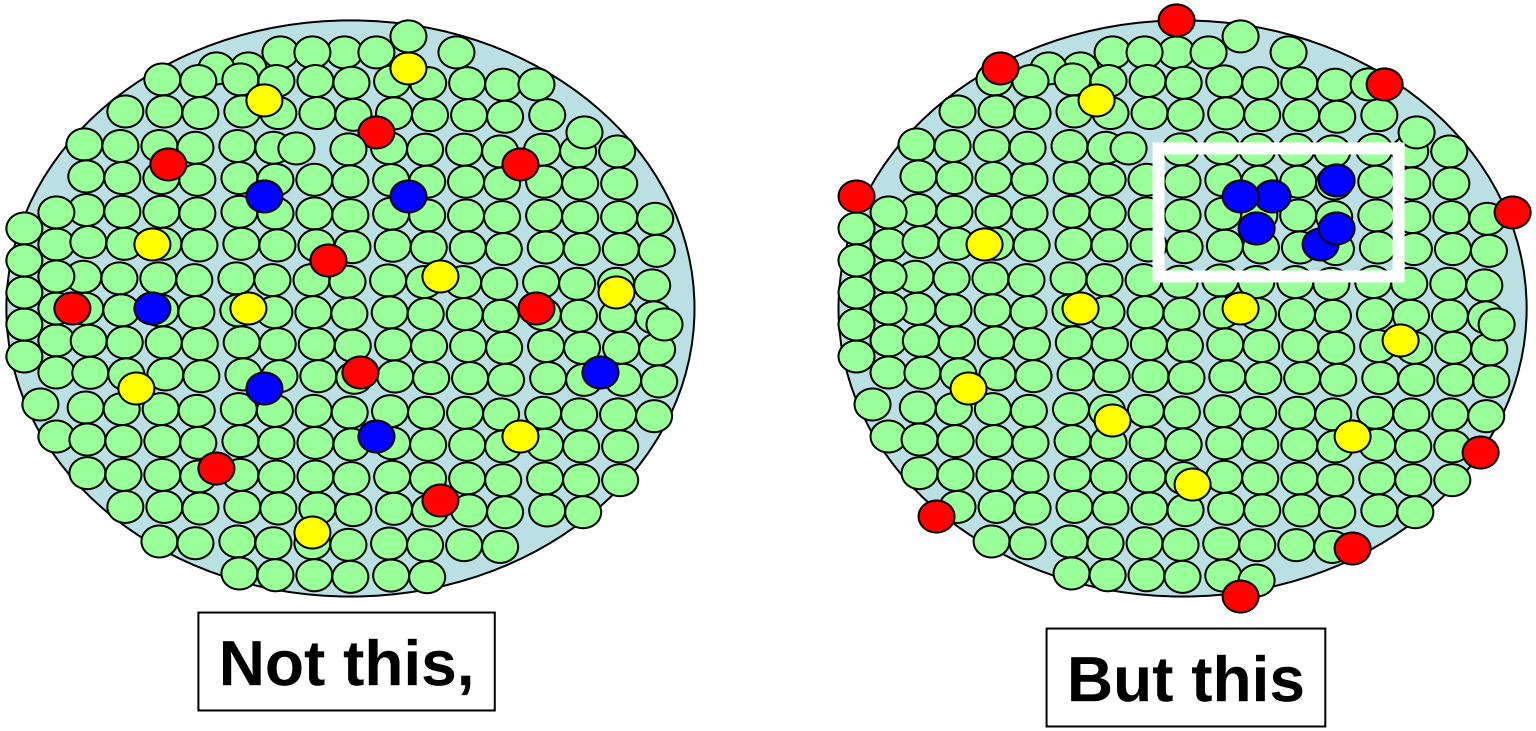
- **In blood, albumin is 3.5 g/100 ml, i.e., 35 g/L = 0.5 mM (10^{-3} M), whereas cytokines are in pM (10^{-12} M)**
- **A 2D gel that is overloaded with respect to an abundant protein, may have *barely detectable* amounts of a low abundance protein…….And if you can't see it, you can't analyze it!!!!!**

Issues in proteome complexity:

Non-equal expression of a protein (ER-β), among like cells; compartmentation of different isoforms (tau) in a single cell



Proteins are non-randomly distributed within cells.



Not this,

But this

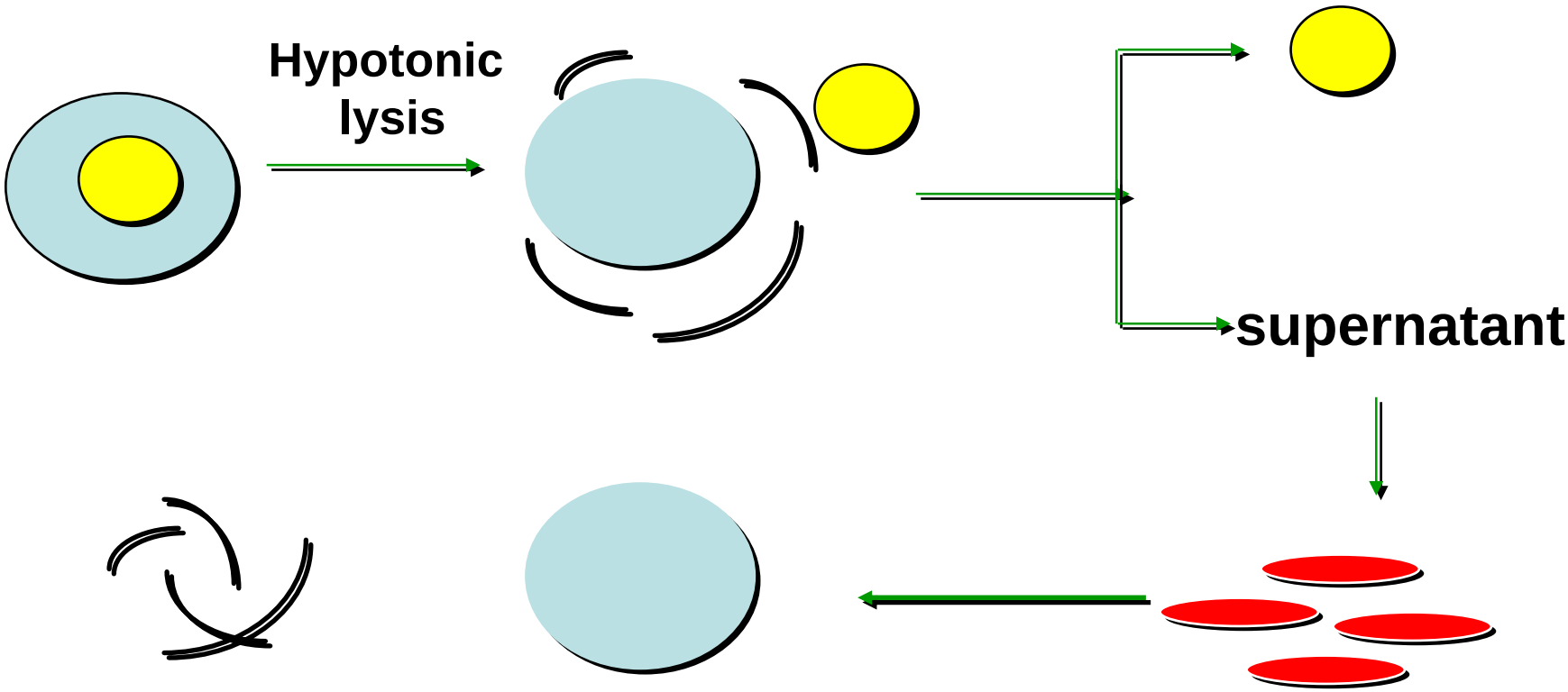
This nonrandom distribution of proteins is the basis for subcellular fractionation.

The basis for protein separation:

- **Biological properties:**
  - **Intracellular location**
  - **Protein-protein interactions**
  - **Posttranslational modifications**

- **Intrinsic properties:**
  - **Net charge**
  - **Size**
  - **Hydrophobicity**

Subcellular compartments differ in size and density.

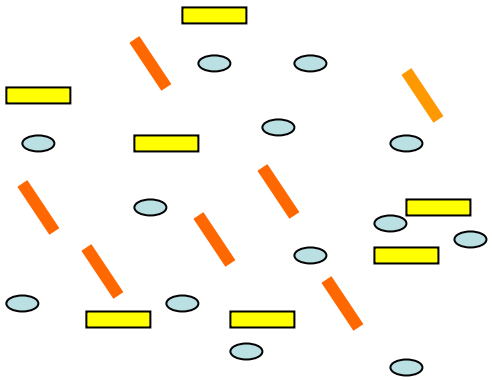


The good news: subcelluar proteomes are readily “catalogued.”

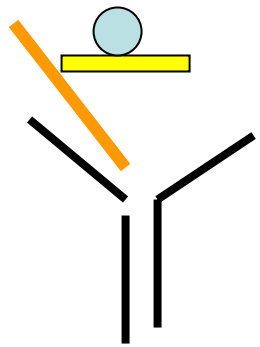
Compartment	# polypeptides in SWISSPROT as of 2000
Mitochondria (1000/cell)	269------5% of total
Lysosome (400/cell)	50------1% of total
Peroxisome	35------0.6%
ER and Golgi apparatus	157------3%
Nuclei (5% cell volume)	964-----17%
Others (cytosol, membrane)	4228----75%
	total: 5703

(Jung et al. [2000] Electrophoresis)

Biological specificity of antibodies is invaluable in reducing the complexity of the proteome



A cell lysate: 6,739 polypeptides



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

What if an antibody to your protein doesn't exist?

Make a protein that has a "tag" for which there *is* an antibody (make molecular biology work for you).

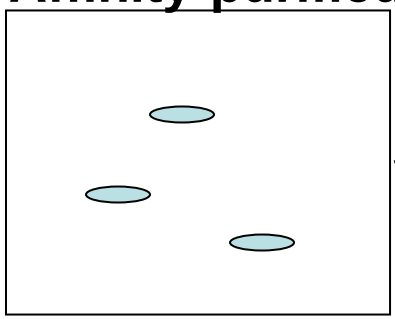
Express one of the following:

- HA-tagged protein
- FLAG-tagged protein
- Histidine-tagged protein

→ **homogenate**

↓ **+ Antibody to the "tag"**

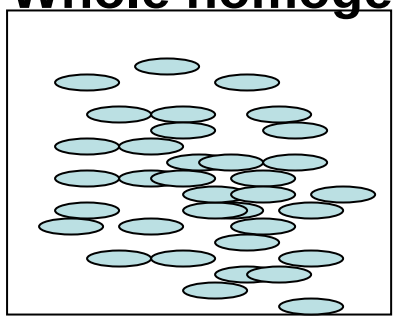
Affinity-purified



2D gel

vs

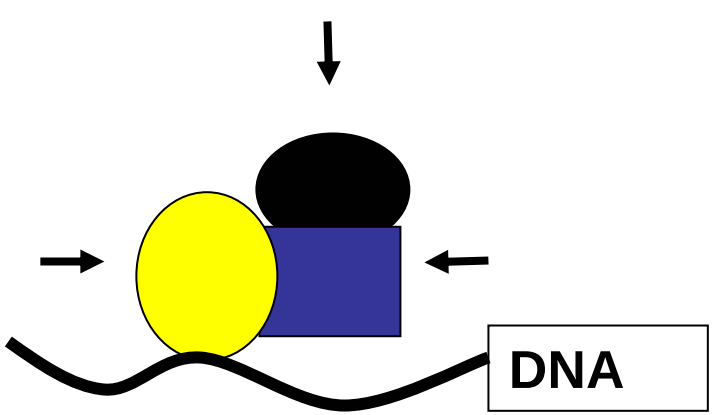
Whole homogenate



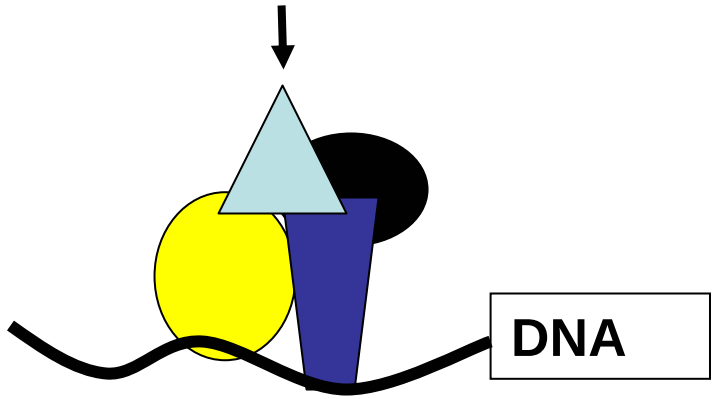
2D gel

← **Enrichment of your protein maybe 10,000-fold**

A valuable role of the proteomics approach: Discovery & analysis of components of protein-complexes



Inactive transcription complex



Active transcription complex

(The arrows indicate the proteins that could be antigens for immunoaffinity purification of the complex.)

Intrinsic properties of proteins provide the basis for different separation methods, used both pre-proteomics, and in proteomics.

I. Proteins are charged.

These will likely bind differently to the same ION-EXCHANGE column, although they are the same mass.

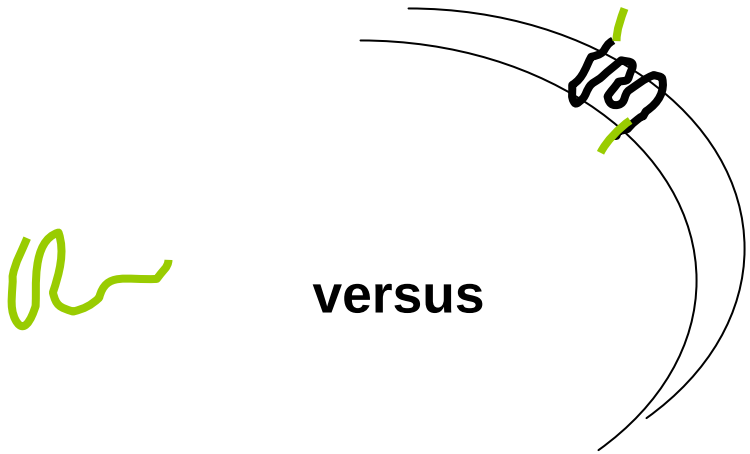
II. Proteins of the same molecular weight can have different shapes.

Protein A: 41,293 daltons

Protein B: 41,293 daltons

These two proteins will behave very differently on “sieving” columns.

III. Proteins can be mostly hydrophilic, or mostly hydrophobic.



The same intrinsic properties also are the basis of proteomics separations

- 2D electrophoresis:
  - Charge: IEF
  - Mass: SDS-PAGE
- 2D LC-LC
  - Charge: Ion-exchange
  - Hydrophobility
- Mass: Mass spectrometry

Take home lessons in studying proteins with proteomics methods

- **The fewer proteins in the proteome you analyze, the better the chances of detecting the ones that "matter."**
- **Genomics data can complement proteomics data.**
- **Understanding the biological properties of the proteins of interest can enhance proteomics analysis.**
- **Intrinsic properties of proteins form the basis of invaluable prefractionation prior to proteomics analysis.**