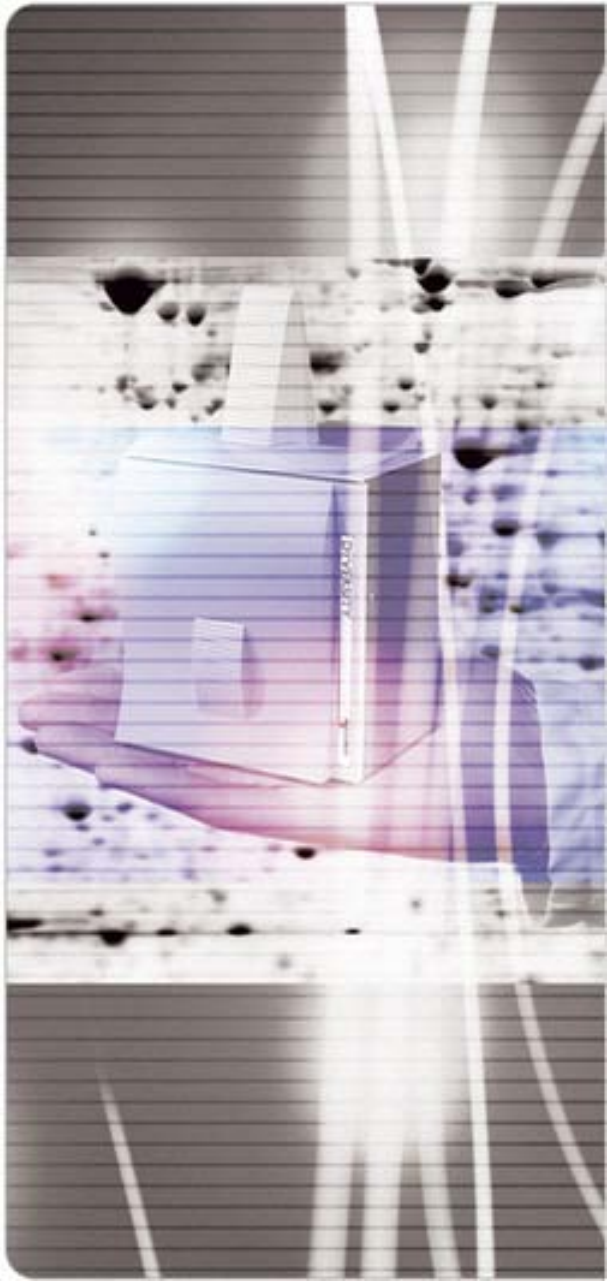
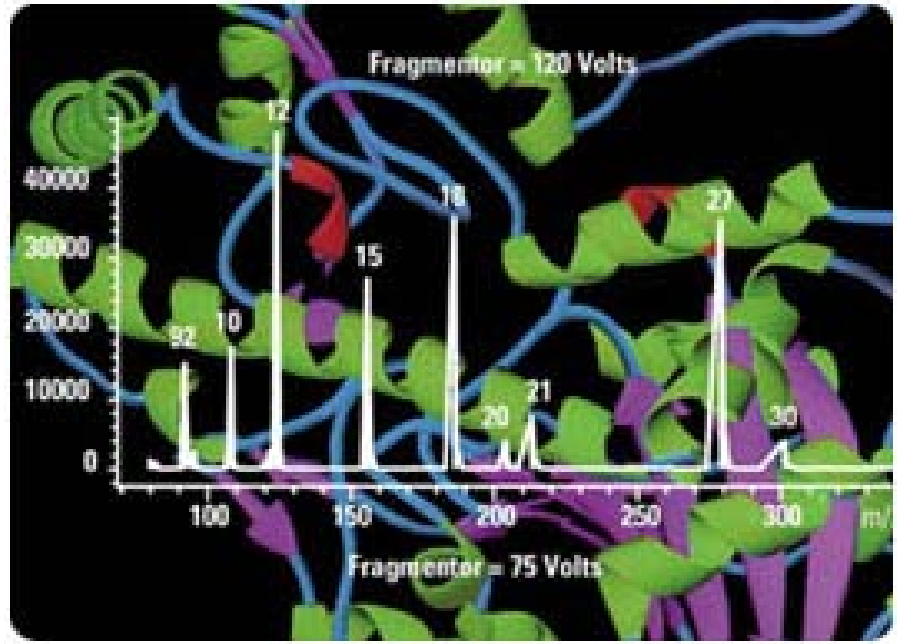


Proteomics: General Background

Michael Hadjiargyrou

Department of Biomedical Engineering



What is proteomics?

- Study of the Proteome: Field that utilizes protein sequences, expression and structure to determine how proteins relate, interact and function in an organism
- includes characterizing and cataloguing proteins and protein libraries, comparing variations in protein expression levels under different conditions, studying protein interactions and functional roles
- techniques are performed in an automated, large scale manner
- may also involve bioinformatic analysis and storage of data

Proteomics

- Proteomics allows the product of genes (proteins) to be studied directly

Why is this important?

1. Not all mRNAs in the cell are translated, so transcriptome may include products that are not present in the proteome (also rates of protein synthesis and protein turnover differ among transcripts, therefore an abundance of transcript may not necessarily correspond to an abundance in the encoded protein

2. Protein activity depends of post-translational modification which are not predictable from the level of the corresponding transcript

The challenges of proteomics

- Splice variants create an enormous diversity of proteins
  - ~25,000 genes in humans give rise to 200,000 to 2,000,000 different proteins
  - Splice variants may have very diverse functions
    - Ex. calcitonin gene
      - Gene variant 1 encodes for calcitonin (increases Ca^{++} uptake in bones)
      - Gene variant 2 encodes for calcitonin gene-related polypeptide (causes blood vessels to dilate)
- Proteins expressed in an organism will vary according to age, health, tissue, and environmental stimuli
- Proteomics requires a broader range of technologies than genomics

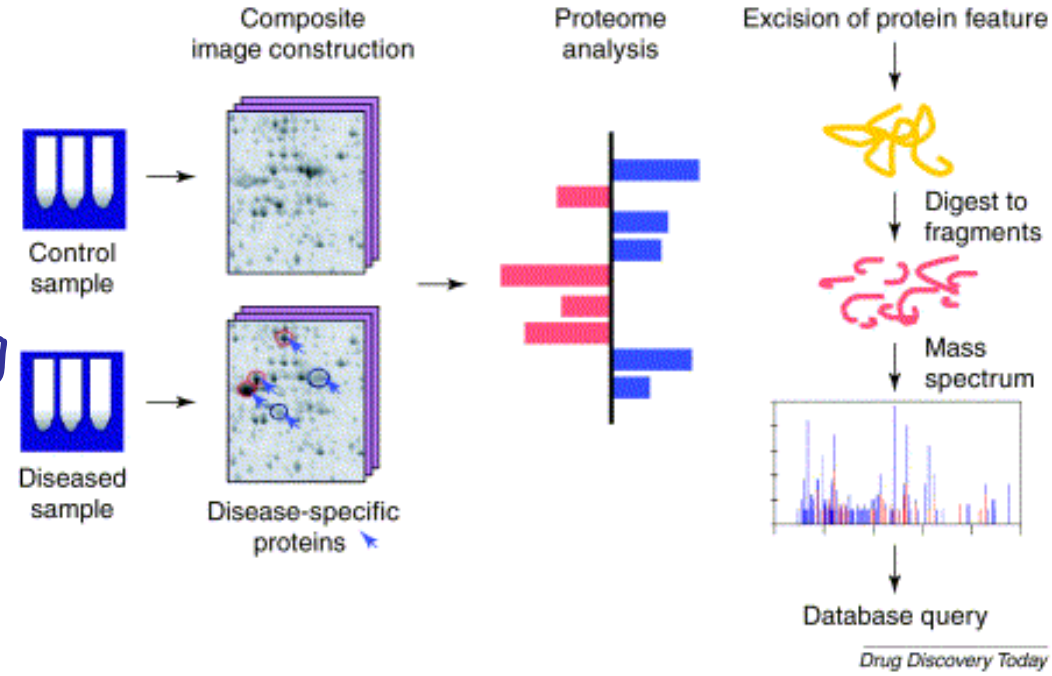
Post-translational modifications

- Proteolytic cleavage - Fragmenting protein

- Addition of chemical groups
  - **Phosphorylation:** activation and inactivation of enzymes
  - **Acetylation:** protein stability, used in histones
  - **Methylation:** regulation of gene expression
  - **Acylation:** membrane tethering, targeting
  - **Glycosylation:** cell–cell recognition, signaling
  - **GPI anchor:** membrane tethering
  - **Hydroxyproline:** protein stability, ligand interactions
  - **Sulfation:** protein–protein and ligand interactions
  - **Disulfide-bond formation:** protein stability
  - **Deamidation:** protein–protein and ligand interactions
  - **Pyroglutamic acid:** protein stability
  - **Ubiquitination:** destruction signal
  - **Nitration of tyrosine:** inflammation

What are some practical applications of proteomics?

- Comparison of protein expression in diseased and normal tissues
  - Likely to reveal new drug targets
    - Today ~500 drug targets
    - Estimates of possible drug targets: 10,000–20,000
- Protein expression signatures associated with drug toxicity
  - To make clinical trials more efficient
  - To make drug treatments more effective

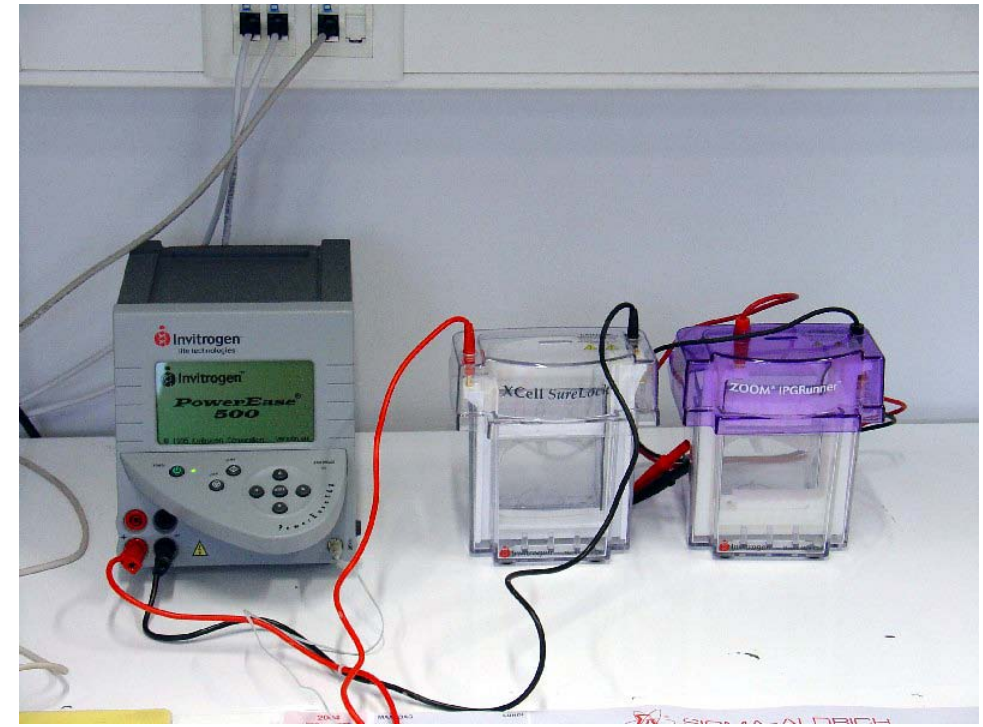


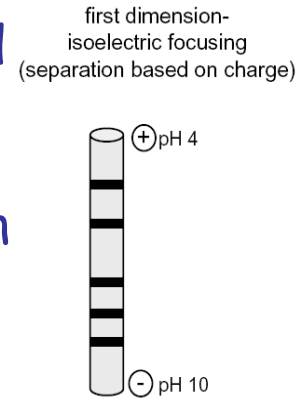
What are some technologies used in proteomics?

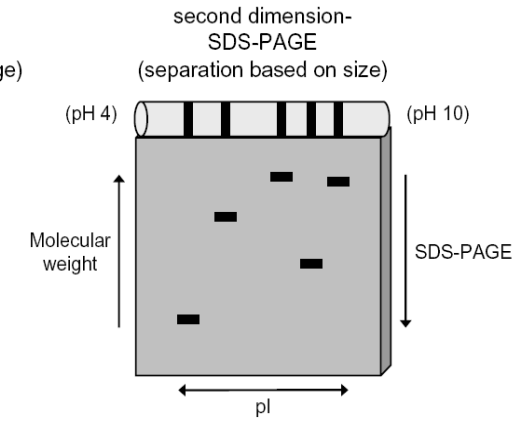
- 2D gel electrophoresis
  - Separates proteins in a mixture on the basis of their molecular weight and charge
- Mass spectrometry
  - Reveals identity of proteins
- Protein chips
  - A wide variety of identification methods
- Yeast two-hybrid method
  - Determines how proteins interact with each other
- X-ray crystallography/NMR spectroscopy
  - Determines protein structure

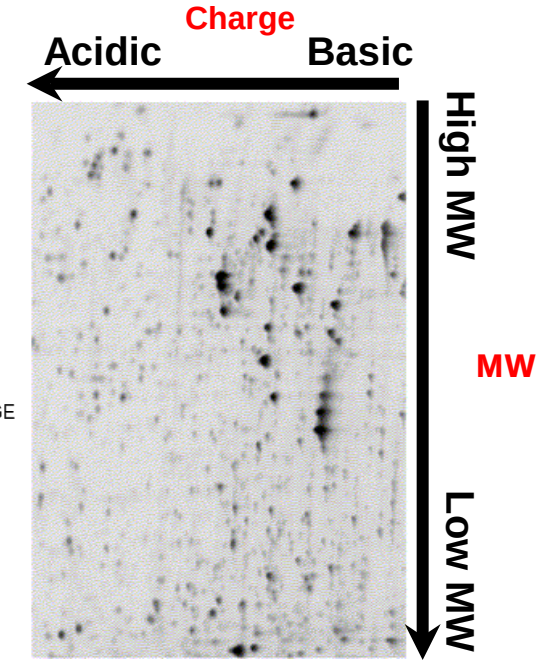
2D Gel Electrophoresis

- Use Polyacrylamide gel
- Voltage across both axes
  - pH gradient along first axis neutralizes charged proteins at different places
  - pH constant on a second axis where proteins are separated by weight
- *x-y* position of proteins on stained gel uniquely identifies the proteins



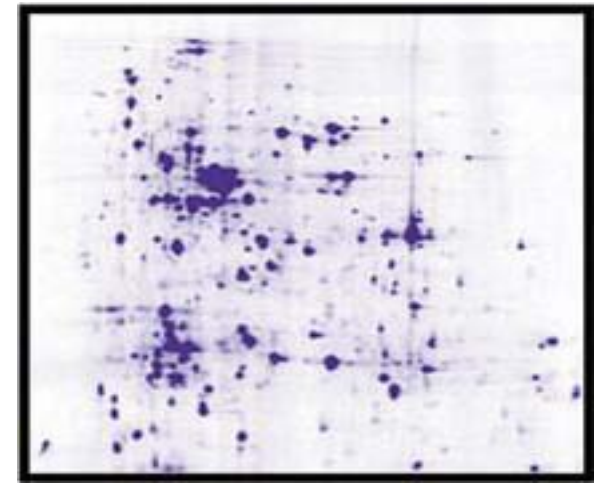




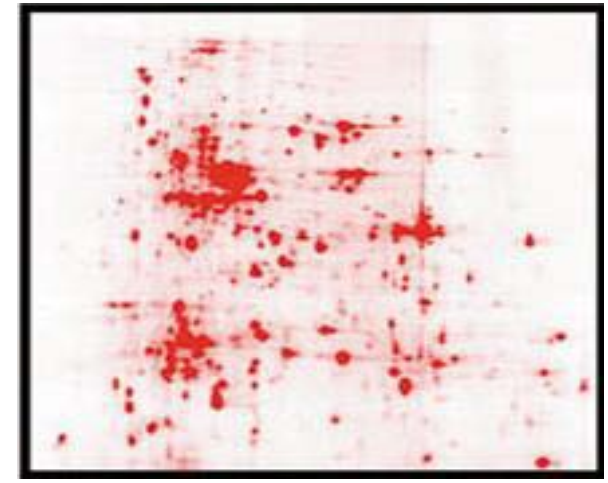


Differential in gel electrophoresis

- Allows comparison of proteomic profiles in two different samples
- Label protein samples from control and experimental tissues
  - Fluorescent dye #1 for control
  - Fluorescent dye #2 for experimental sample
- Mix protein samples together and run on a single gel
- Identify identical proteins from different samples by dye color



Sample 1
Cy3

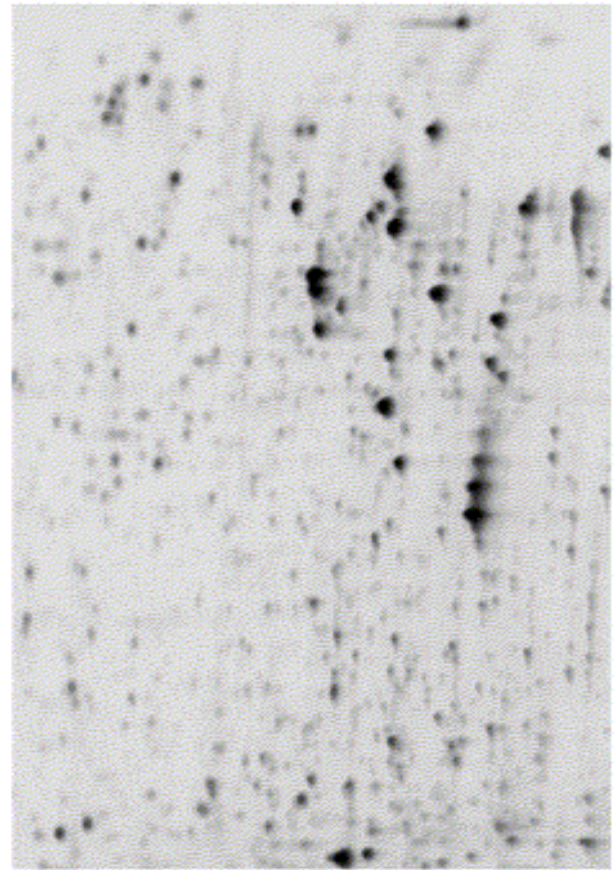


Sample 2
Cy5

Disadvantages encountered with 2D gels

- Laborious technique

- Poor performance of 2D gels for the following:
  - Very large proteins
  - Very small proteins
  - Less abundant proteins
  - Membrane-bound proteins
    - Presumably, the most promising drug targets



Mass spectrometry

- Measures mass-to-charge ratio
- Components of mass spectrometer
  - Ion source
  - Mass analyzer
  - Ion detector
  - Data acquisition unit

Identifying proteins with mass spectrometry

- Preparation of protein sample
  - Extraction from a gel
  - Digestion by proteases — e.g., trypsin
- Mass spectrometer
  - measures mass-charge ratio of peptide fragments
- Identified peptides are compared with database
  - Software used to generate theoretical peptide mass fingerprint (PMF) for all proteins in database
  - Match of experimental readout to database PMF allows researchers to identify the protein

Identifying proteins with mass spectrometry

isolate pollen or sperm proteins

2-DE

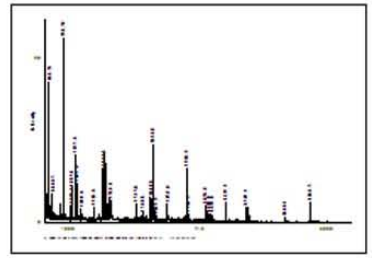
gel spots

Digest Pro

trypsin fragments

Sciex QSTAR

data



spot matching quantification profiling

sequence database

O-MALDI MS: peptide mass mapping
Tandem MS: peptide sequencing

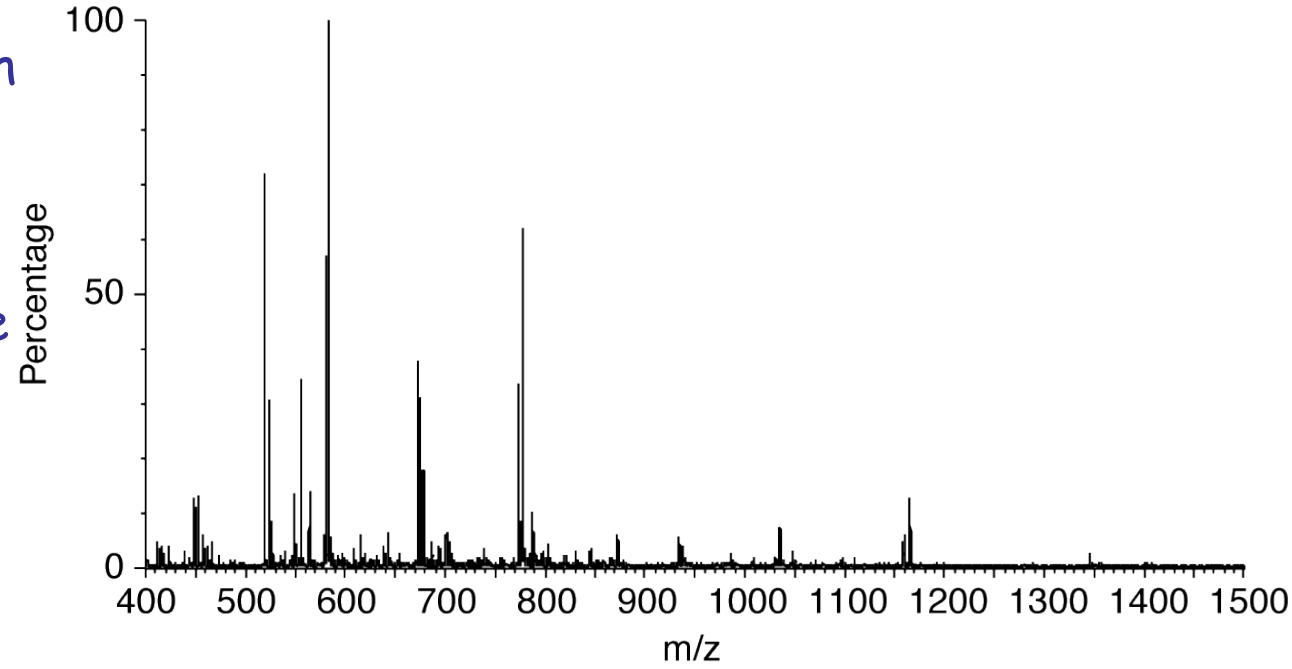
www.pgec.usda.gov

A mass spectrum

Mass spectrum comes from the analysis of a mixture of five proteins that were digested with the protease trypsin.

The mass-to-charge ratio is plotted along the *x*-axis, and the signal intensity for individual ions is plotted along the *y*-axis, which is normalized to the highest peak.

Different peaks represent the presence of individual ionized peptides.

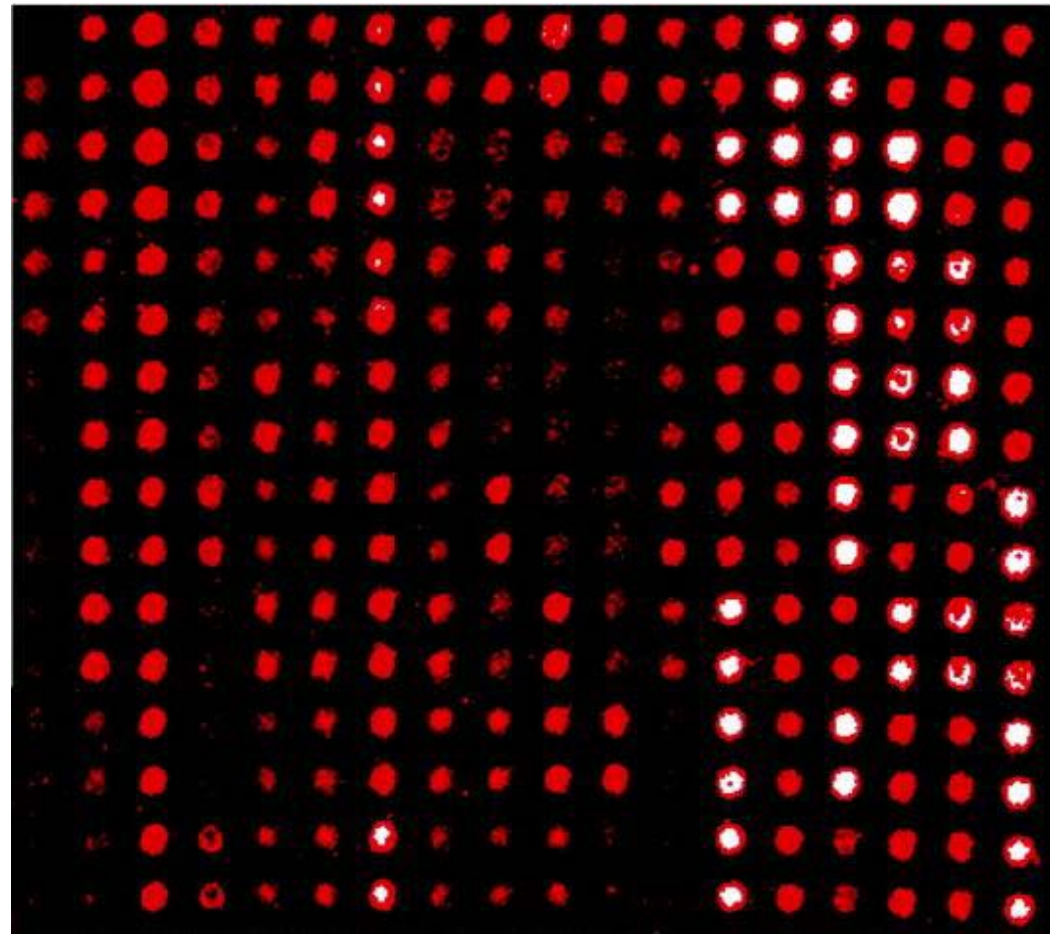


Limitations of mass spectrometry

- Not very good at identifying minute quantities of protein
- Trouble dealing with phosphorylated proteins
- Doesn't provide concentrations of proteins
- Improved software eliminating human analysis is necessary for high-throughput projects

Protein chips/arrays

- Thousands of proteins analyzed simultaneously
- Wide variety of assays:
  - Antibody–antigen
  - Enzyme–substrate
  - Protein–small molecule
  - Protein–nucleic acid
  - Protein–protein
  - Protein–lipid

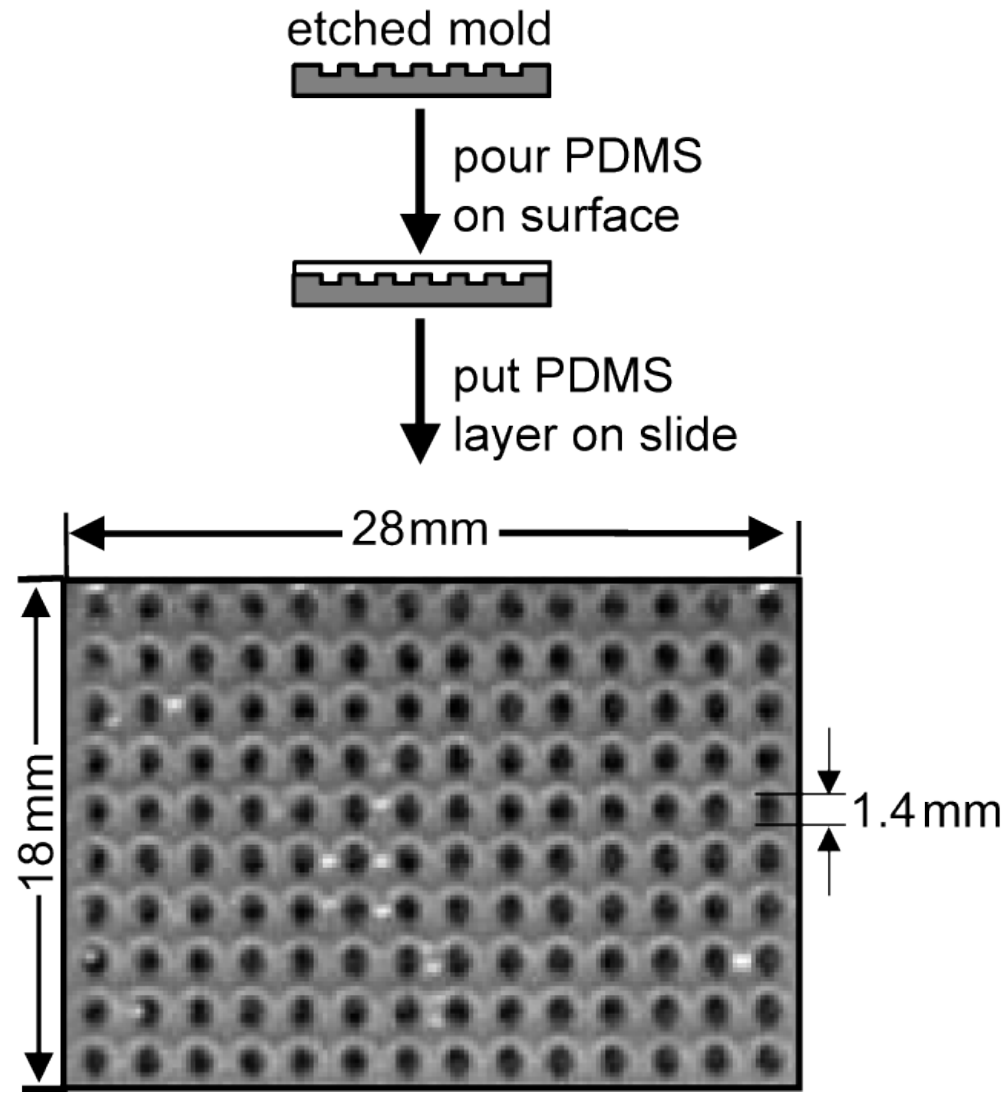


Yeast proteins detected using antibodies

Fabricating protein chips

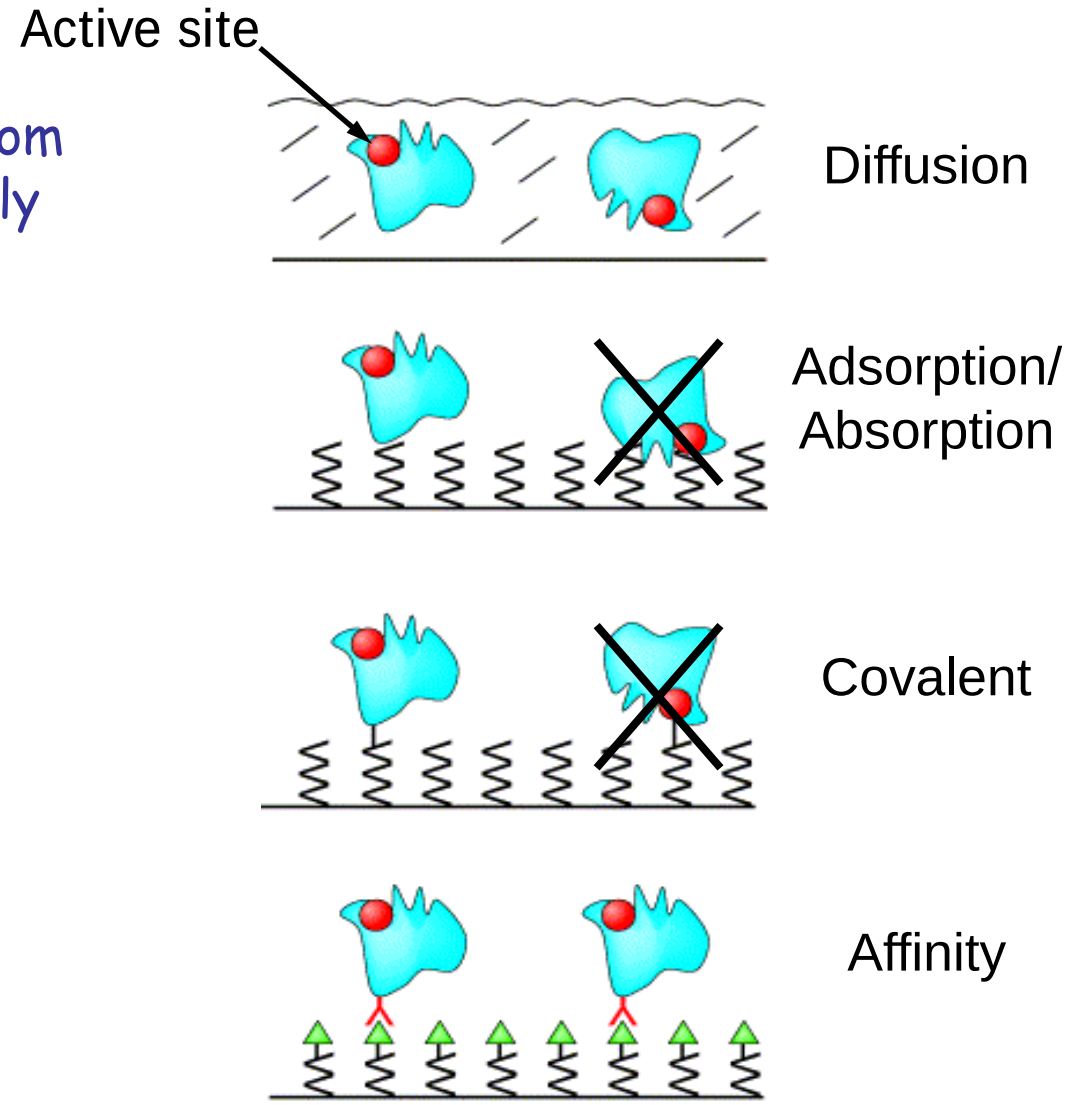
- Protein substrates
  - Polyacrylamide or agarose gels
  - Glass
  - Nanowells
- Proteins deposited on chip surface by robots

PDMS: Polydimethylsiloxane



Protein attachment strategies

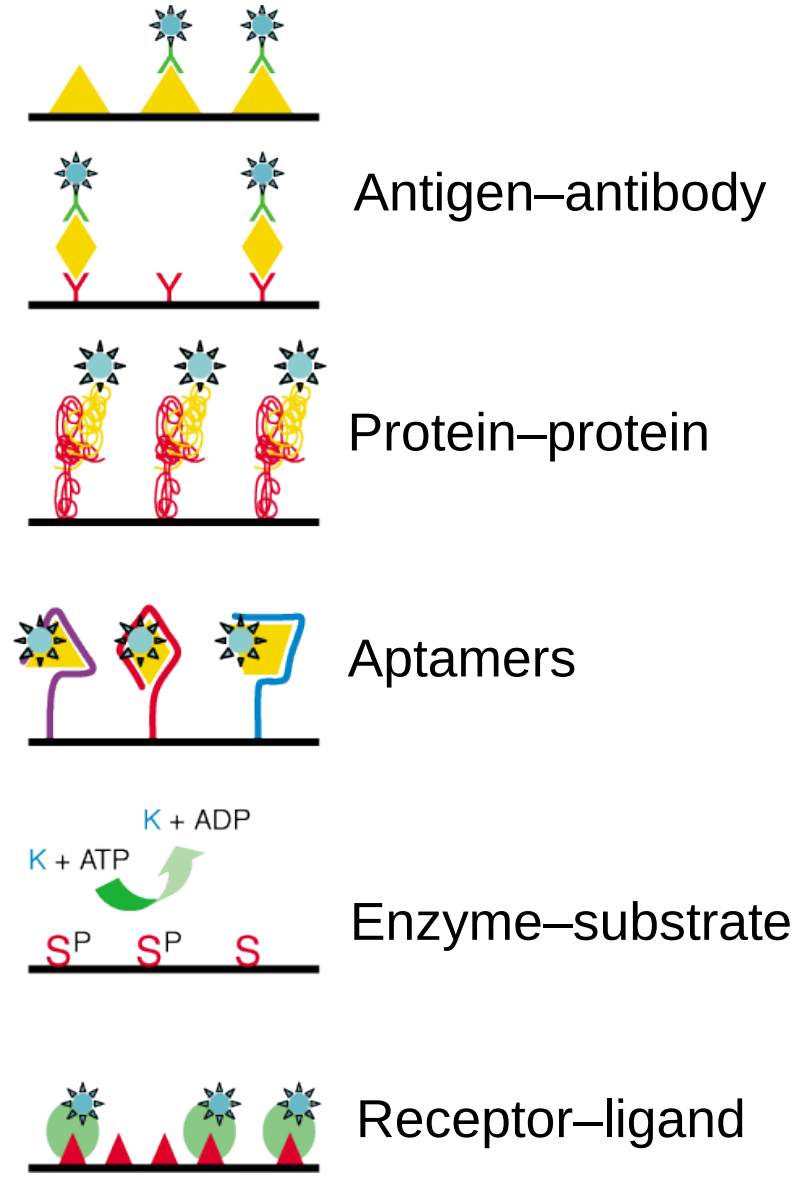
- Diffusion
  - Protein suspended in random orientation, but presumably active
- Adsorption/Absorption
  - Some proteins inactive
- Covalent attachment
  - Some proteins inactive
- Affinity
  - Orientation of protein precisely controlled



Classes of capture molecules

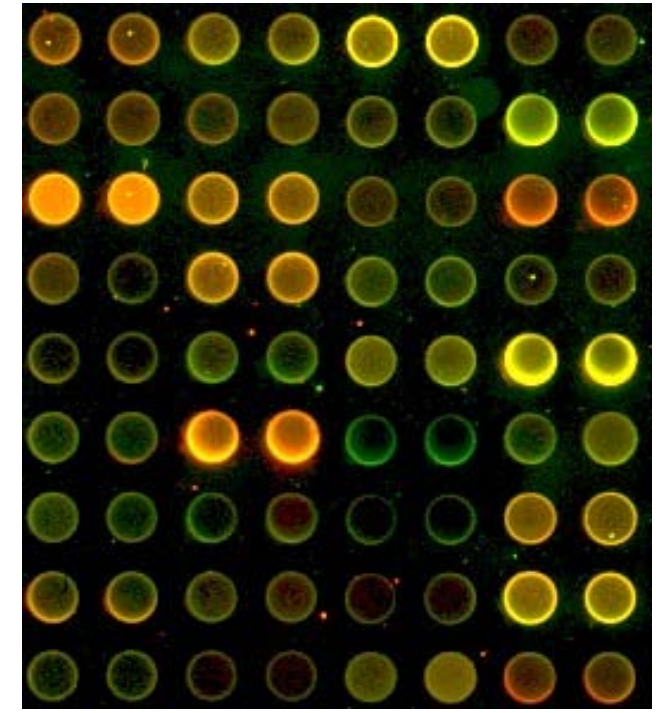
- Different capture molecules must be used to study different interactions

- Examples
  - Antibodies (or antigens) for detection
  - Proteins for protein-protein interaction
  - Enzyme-substrate for biochemical function
  - Receptor-ligand interaction



Protein chips - obtaining results

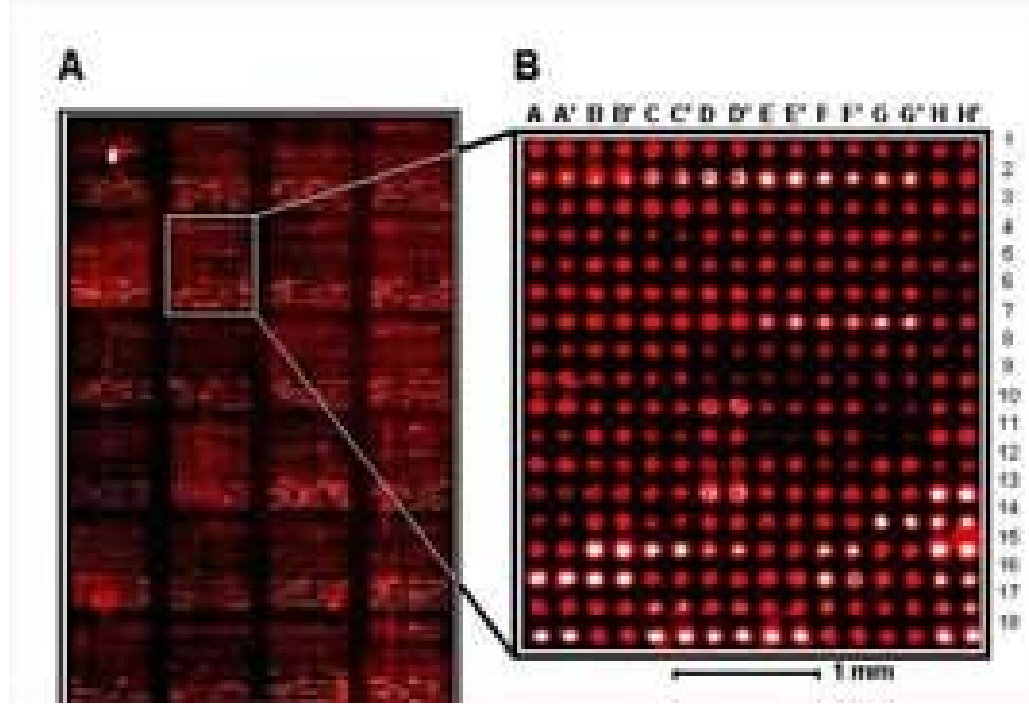
- Fluorescence
  - Most common method
  - Fluorescent probe or tag
  - Can be read out using standard nucleic acid microarray technology
- Surface-enhanced laser desorption/ionization (SELDI)
  - Laser ionizes proteins captured by chip
  - Mass spectrometer analyzes peptide fragments
- Atomic-force microscopy
  - Detects changes in chip surface due to captured proteins



www.willson.cm.utexas.edu

Disadvantages encountered with designing protein chips

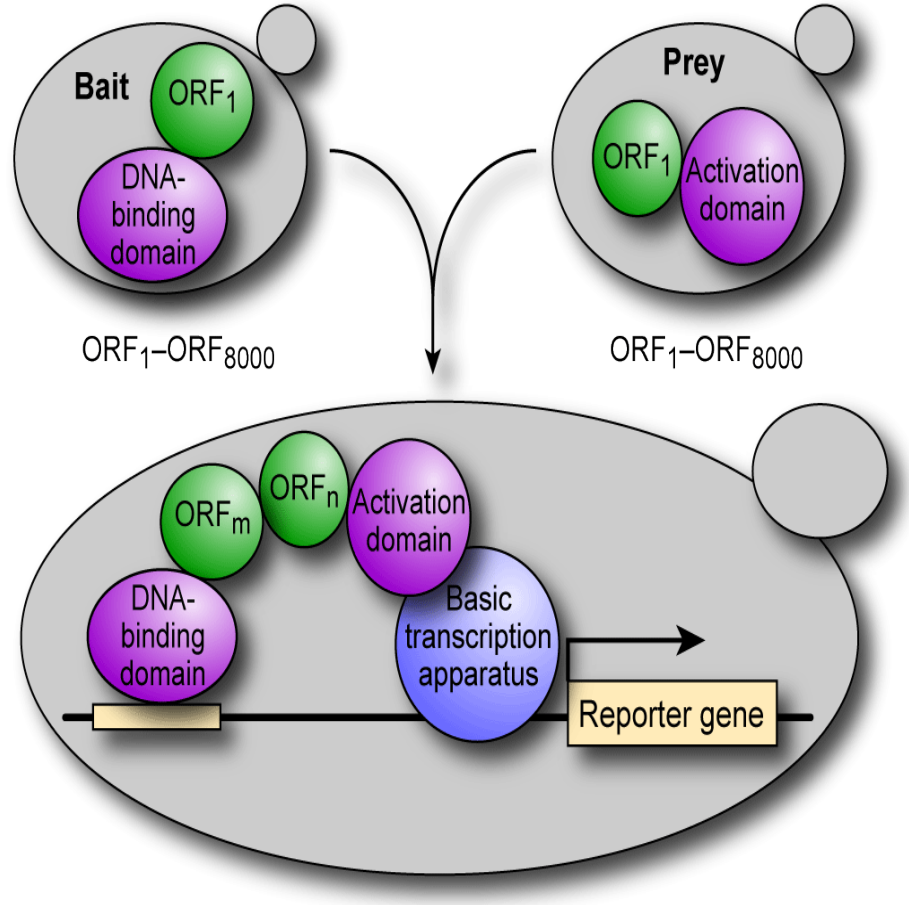
- Unique process is necessary for constructing each probe element
- Challenging to produce and purify each protein on chip
- Proteins can be hydrophobic or hydrophilic
  - Difficult to design a chip that can detect both



www.smalltimes.com

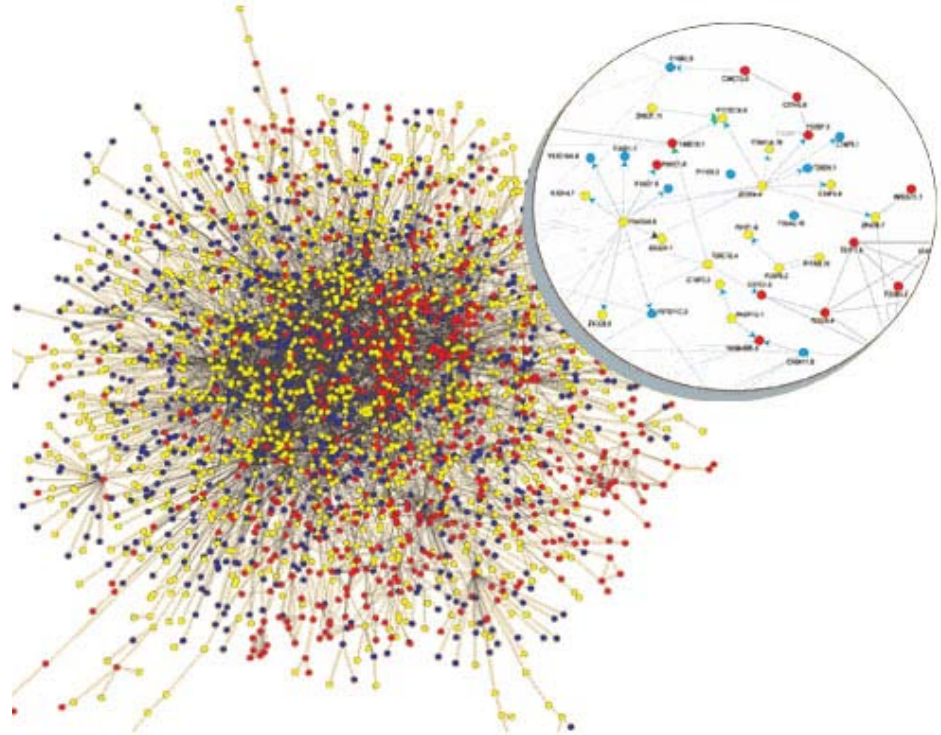
Yeast two-hybrid method

- Goal: Determine how proteins interact with each other
- Method
  - Use yeast transcription factors
  - Gene expression requires the following:
    - A DNA-binding domain
    - An activation domain
    - A basic transcription apparatus
  - Attach protein$_1$ to DNA-binding domain (bait)
  - Attach protein$_2$ to activation domain (prey)
  - Reporter gene expressed only if protein$_1$ and protein$_2$ interact with each other



Results from a yeast two-hybrid experiment – "Interactome"

- Goal: To characterize protein-protein interactions among 6,144 yeast ORFs
  - 5,345 were successfully cloned into yeast as both bait and prey
  - Identity of ORFs determined by DNA sequencing in hybrid yeast
  - 692 protein-protein interaction pairs
  - Interactions involved 817 ORFs



www.the-scientist.com

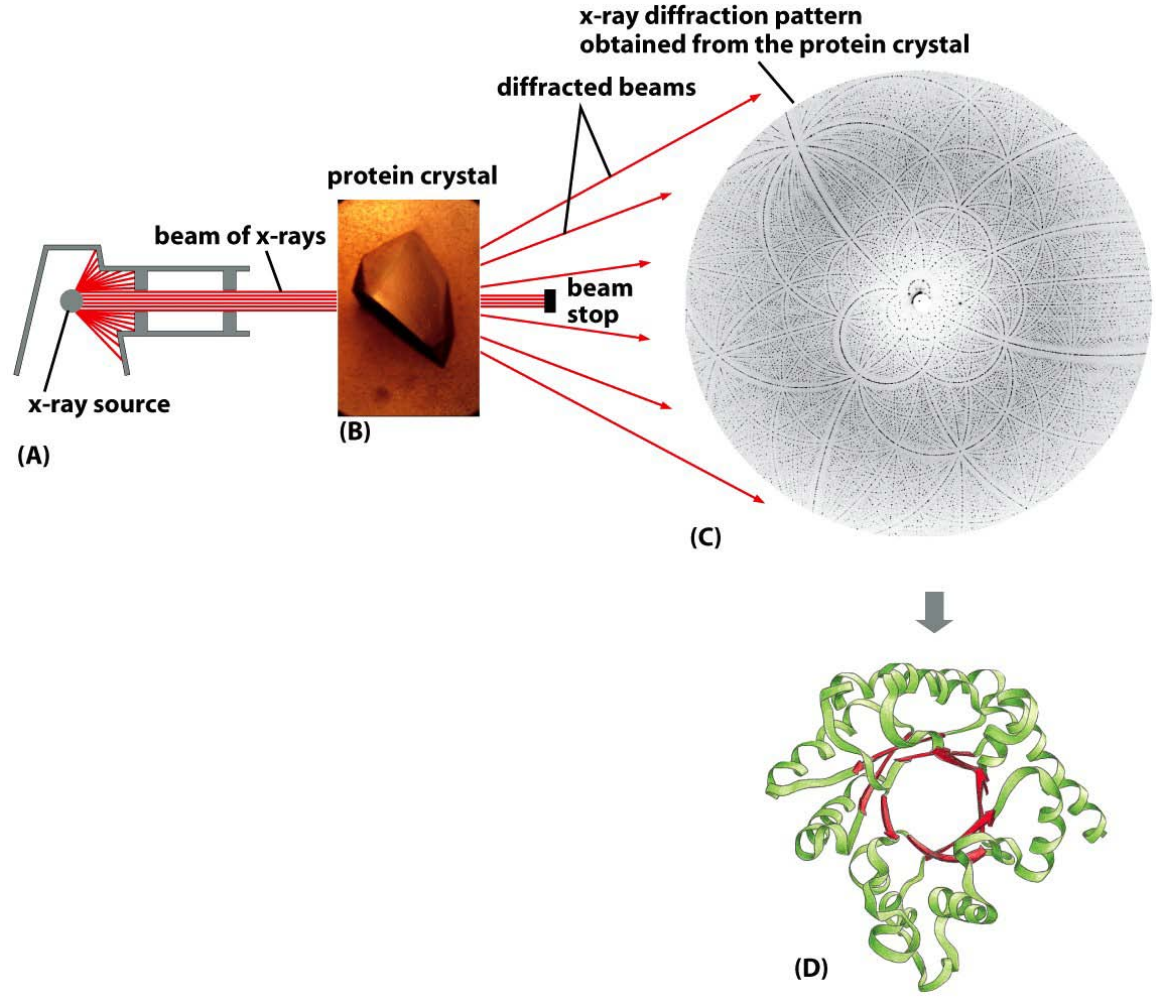
Disadvantages associated with the yeast two-hybrid method

- There is evidence that other methods may be more sensitive
- Some inaccuracy reported when compared against known protein–protein interactions
  - False positives
  - False negatives

X-ray Crystallography

- A technique used to determine the detailed, three-dimensional structure of molecules

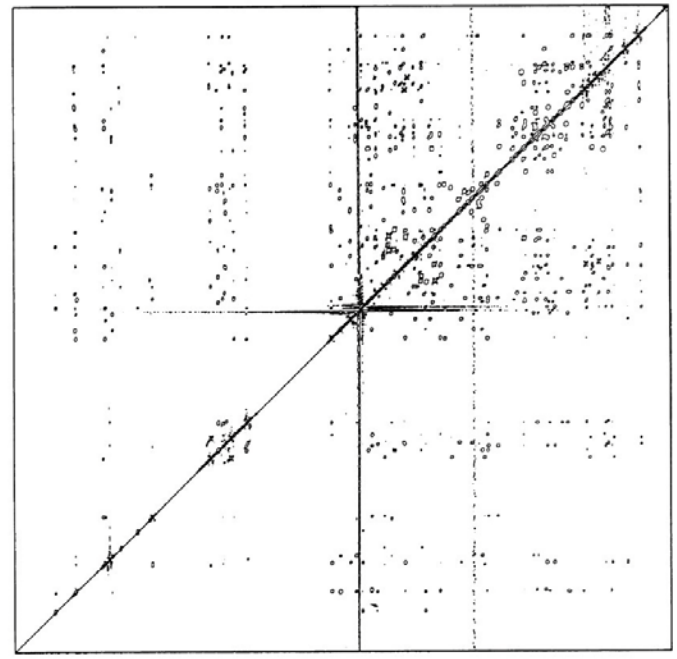
- It is based on the scattering of X-rays through a crystal of the molecule under study.



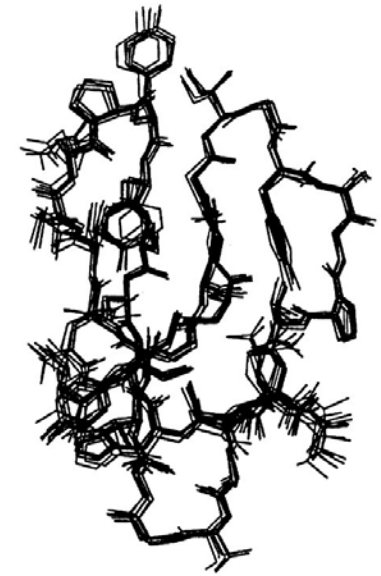
Nuclear magnetic resonance (NMR) spectroscopy

- A technique used to determine the detailed, three-dimensional structure of molecules and, more broadly, to study the physical, chemical, and biological properties of matter.

- It uses a strong magnet that interacts with the natural magnetic properties in atomic nuclei.



(A)



(B)

Future prospects of Proteomics

- The next decade may see the complete deciphering of the proteome of yeast
- More initiatives, like the Human Liver Proteome Project, are underway
- Better understanding of disease
- Expansion of:
  - Structural proteomics - in-depth analysis of protein structure
  - Expression proteomics - analysis of expression and differential expression of proteins
  - Interaction proteomics - analysis of interactions between proteins to characterize complexes and determine function.