Proteomics

Manickam Sugumaran
Department of Biology
University of Massachusetts
Boston, MA 02125

Genomic studies produced more than 75,000 potential gene sequence targets.

(The number may be even higher than this. Some predict as much as 200,000)

But it was unable to extract anticipated 10,000 therapeutic targets from the crude data.

Hurdles of genomics

- Inability of genomic studies to idenfity drug targets has produced huge gap between genomics and drug discovery process.
- It is mostly because the genomics failed to reveal the protein function and its relevance to disease.

Importance of proteomics

- The true value of genomics will become obvious only when the functional assignments are made to the encoded genes.
- Proteomics seeks to provide this vital information.
- Proteomics is a concept which is a protein based approach.
- It has the capacity to provide functional role to proteins on a genome wide scale.

Use of proteomics

- Nearly 70% of the predicted sequences in multicellular organisms have no know function.
- Proteomics will reveal:
- A) New enzymes
- B) New signal molecules
- C) New pathways
- That may serve as excellent targets for drug discovery process.

Goal of proteomics

- **Provide structural and functional annotation for the entire genome.**
- It is purely an experimental science and not a computational science because, the computational methods are not available to predict a) protein function; b) protein 3-D structure, and c) suitability of a particular protein for drug discovery and development processes.

Genome wide protein purification will provide proteins for

- High throughput screening.
- Solving structure which in turn will give 3-D structure necessary for drug development.
- Thus proteomics will fill in the gap between genomics and drug discovery process.

Large quantities of proteins are in need for

- A) Structural proteomics studies
- B) Functional Proteomics studies
- C) Proteome wide high throughput screening
- D) Examining protein protein interactions

High through put protein purification process is taking place in

- Structure Factory - Germany
- Harvard Institute of Proteomics - USA
- Ontario Center for Structural Proteomics - Canada
- Riken - Genomic Sciences - Japan.
- National Institutes of Health - USA

Proteins and their relevance to disease conditions

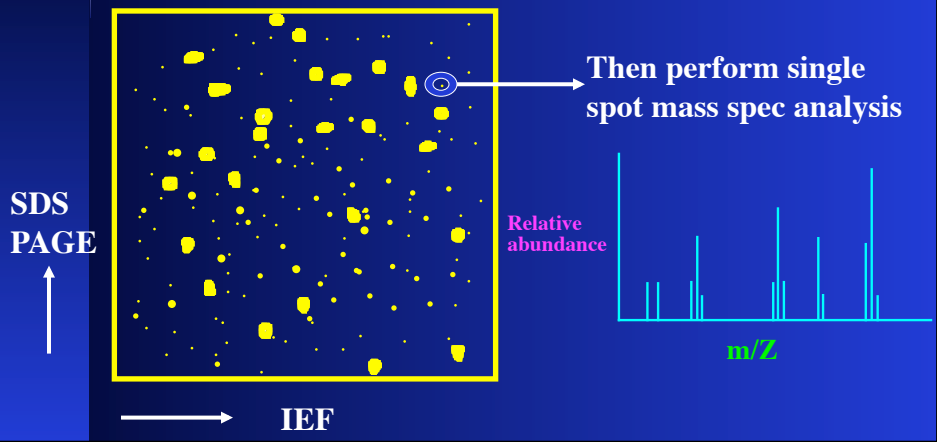
- More than 100,000 proteins exist in humans. Some estimates put this number as high as 300,000.
- But only a fraction of them are expressed in any given cell at any given time.
- To discover and monitor their relevance to disease conditions, it is important to find out A) Which proteins are expressed? B) when they are expressed? C) where they are expressed? And D) to what extent they are expressed in a cell.

DNA microarray technology

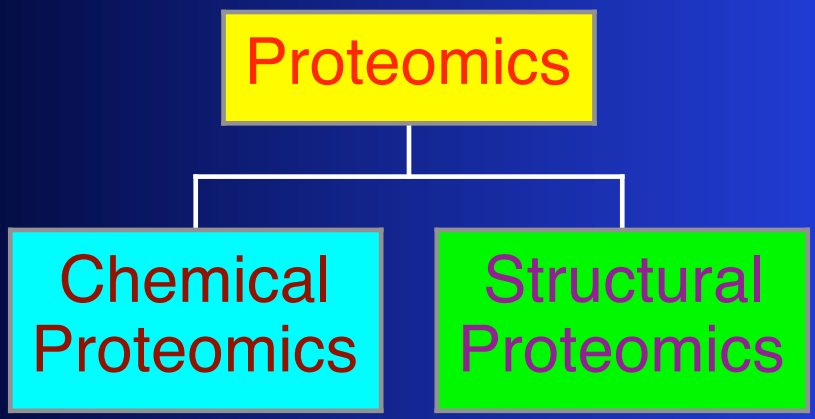
- is useful to monitor the abundance of mRNA in a cell and is even useful to study the poorly expressed genes.
- But it is not accurate for proteins as several of the regulatory processes such as mRNA splicing, covalent modifications etc., occur after transcription.
- Therefore, it is important to know the abundance of proteins.

Two dimensional gel electrophoresis and Mass spectral analysis

- **First resolve the entire cellular protein extract by 2-D gel electrophoresis.**



Two major aspects of proteomics



Chemical Proteomics or Chemical Genomics

- 1. Proteome wide screening for known catalytic activities (such as proteases, phosphatases, etc.,)
- Phizicky et. al., fused thousands of protein genes to coding sequence of glutathione-S-transferase and expressed a set of fusion proteins. Expressed proteins were screened for several catalytic activities. They found several unannotated reading frames and ascribed functions to them. (A new cyclic phosphodiesterase and a cytochrome c methyl transferase were discovered by this techniques).

Chemical Proteomics or Chemical Genomics

- 2. Identify small molecules that bind to proteins using combinatorial chemistry libraries and microarray techniques, micro calorimetry, NMR spectroscopy, LC- MS spectrometry etc.,
- Characterize the structure of the ligand.
- This will lead to the identification of the biological function of unknown proteins.

Structural Proteomics or structural genomics

- Proteins of similar function share structural homology even if their primary sequences are not homologous.
- Therefore, unknown proteins may share unrecognized structural homology to known proteins.
- Thus the structural information can lead to functional clues for a number of unannotated proteins.

Structural Proteomics or structural genomics

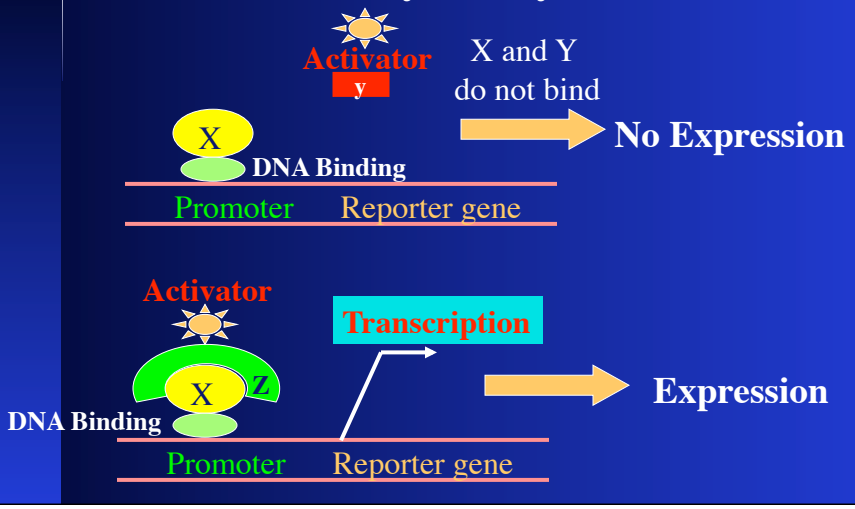
- Goal is to obtain the three dimensional structure of all proteins. This will be useful for
- A) ascribing function to unannotated genes
- B) new drug target selection
- C) validating target based on homology
- D) invalidating targets that do not bind to ligands in terms of their 3-d structure.
- E) developing hits - leads and drugs based on structural based approaches
- F) perfecting structure predicting algorithms

Interaction proteomics

- Protein protein interactions are at the heart of all cellular processes. Knowledge on complex network of protein protein interactions is vital for solving several disease conditions such as autoimmune diseases, cancer, viral infections etc., Faulty macromolecular interactions paly a pivital role in several disease conditions. So elucidating protein protein interactions can lead to discovery of drugs.

Commonly used techniques to study protein protein interaction

- 1. **Yeast Two Hybrid System:**



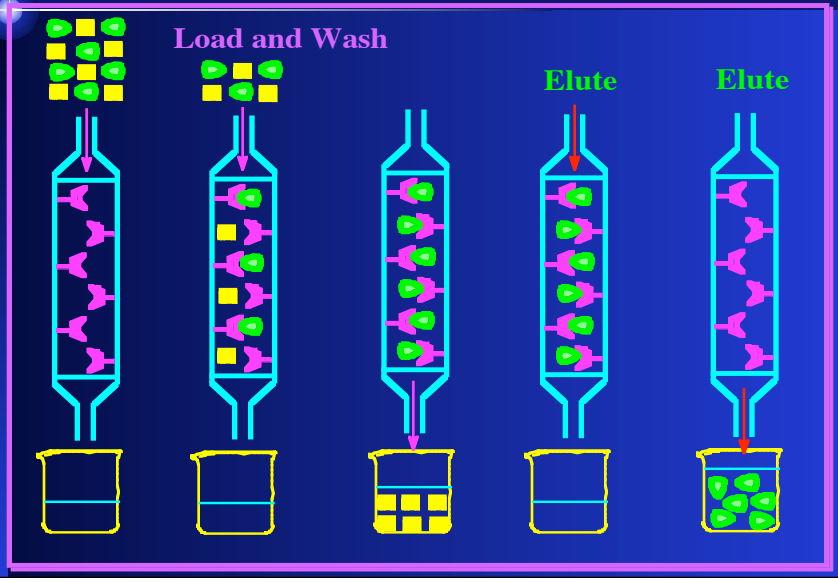
Yeast two hybrid system

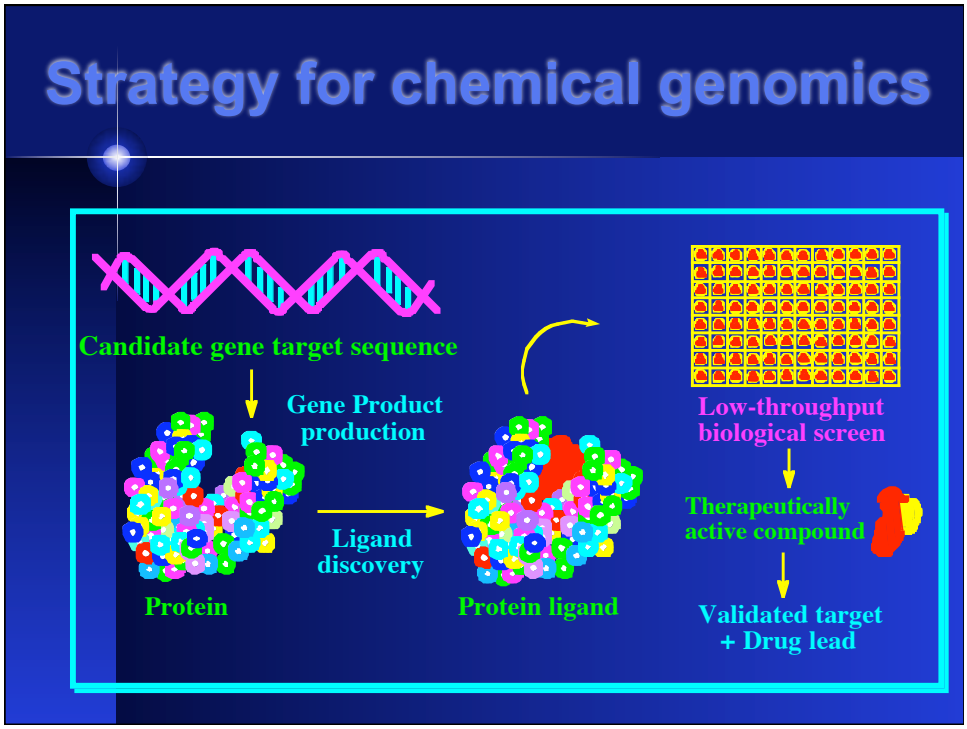
- Powerful to identify many protein protein interactions.
- **Disadvantages:**
- False positive results and time consuming
- Comprehensive genome wide screening has identified only a fraction of known interactions.
- Useful mostly for bimolecular interaction. Larger interactions are not identified.

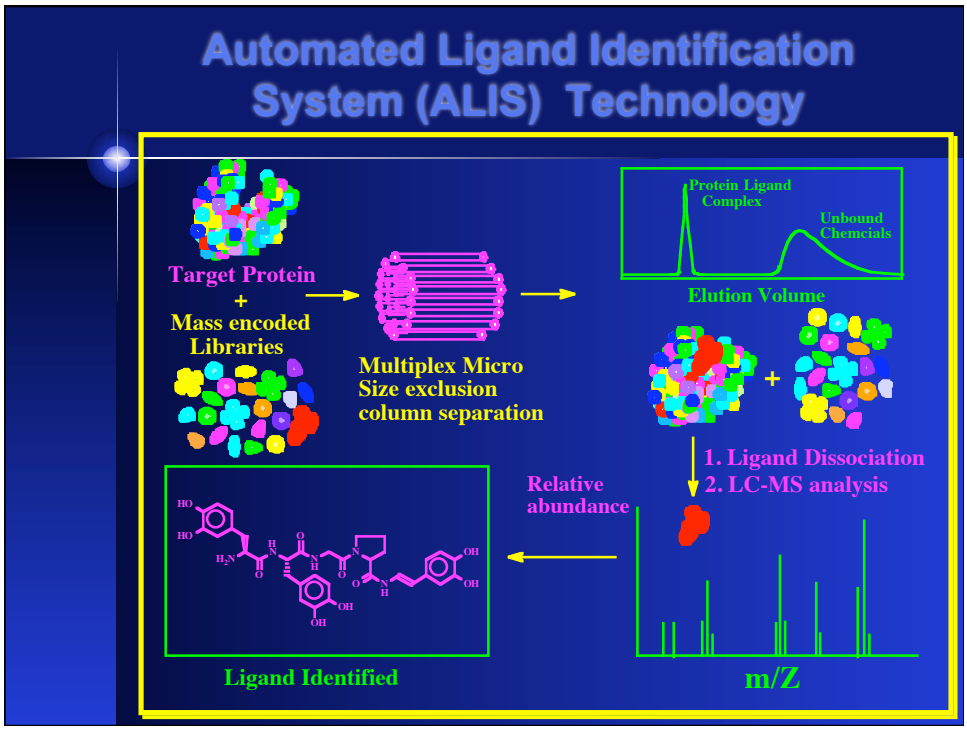
Direct monitoring of protein protein interaction is better.

- **Protein Affinity Chromatography**
- **Advantages:** Better for high throughput screening; more reliable; identifies all interactions; usually no false positive results. Detection can be achieved at as low as 1 in 100,000.
- **Disadvantages**: requires pure protein; weaker interactions (such as 10^{-2} M range) are missed.

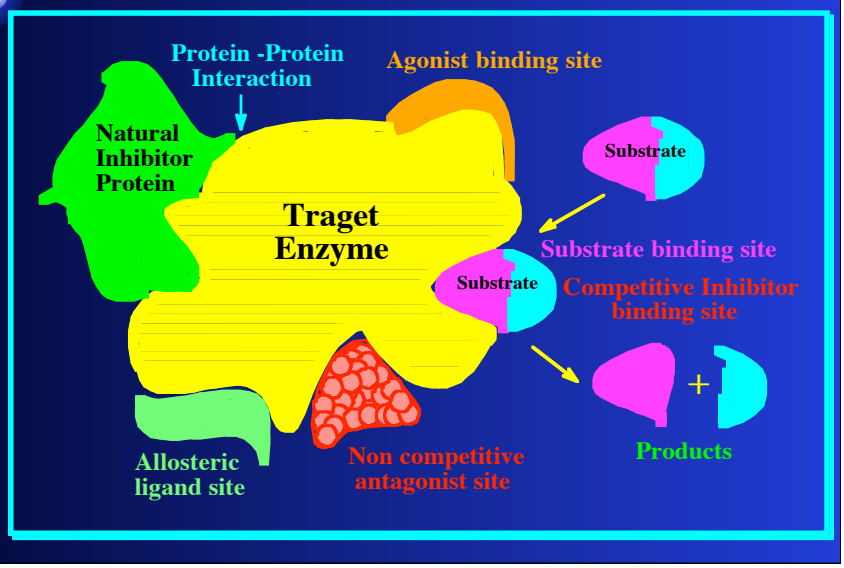
Affinity selection




Strategy for chemical genomics
Candidate gene target sequence
Gene Product production
Protein
Ligand discovery
Protein ligand
Low-throughput biological screen
Therapeutically active compound
Validated target + Drug lead


Automated Ligand Identification System (ALIS) Technology
Target Protein
+
Mass encoded Libraries
Multiplex Micro Size exclusion column separation
Protein Ligand Complex
Unbound Chemicals
Elution Volume
1. Ligand Dissociation
2. LC-MS analysis
Relative abundance
Ligand Identified
m/Z

There are Multiple Target Sites on a Single Protein



ALIS technology

- Can process as much as 300,000 compounds in a day.
- Identify ligands that bind with an affinity of 10^{-6} to 10^{-9} M range easily.
- Less than 5 mg of protein is sufficient to screen several million compounds.
- No need to rely on biochemical knowledge of the target.

ALIS Technology - Advantages

- Reduced assay development time.
- Lower false positive hit rates
- Direct identification of protein ligands
- Well suited for rapid identification.