

Lecture 4 - Fundamental Technologies IV: Proteomics and Metabolomics

BENG168

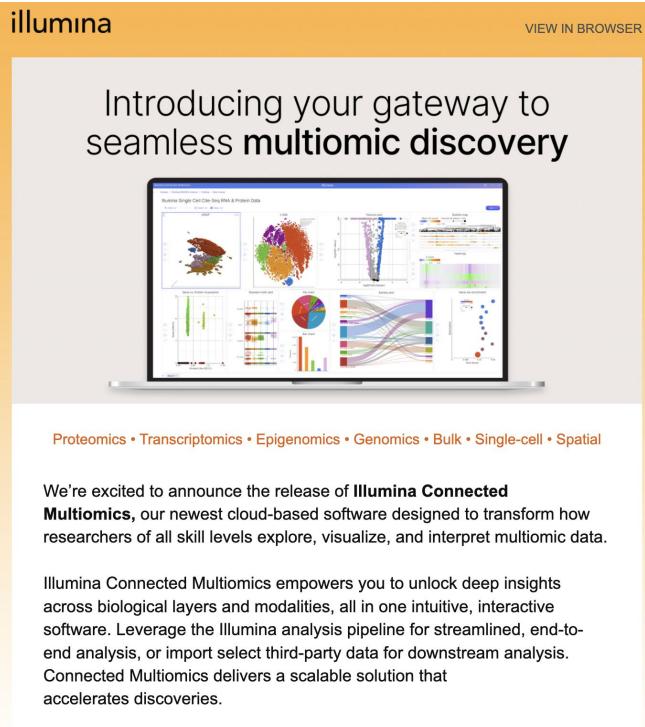
**Instructor: Adam M. Feist, Assistant Professor, Shu Chien -
Gene Lay Department of Bioengineering**

In-Class Announcements & Follow Up

- Update to previous lecture slides
- In-class announcements

Genomics in the News!

- There were some coincidental relevant announcements related to the course material that I wanted to share to show the relevance and current importance of the content.



The image shows the Illumina Connected Multiomics software interface. At the top, it says "Introducing your gateway to seamless multiomic discovery". Below this is a laptop screen displaying various multiomic analysis tools and visualizations. The interface includes sections for "Proteomics • Transcriptomics • Epigenomics • Genomics • Bulk • Single-cell • Spatial". A text block below the interface reads: "We're excited to announce the release of **Illumina Connected Multiomics**, our newest cloud-based software designed to transform how researchers of all skill levels explore, visualize, and interpret multiomic data." Another text block at the bottom states: "Illumina Connected Multiomics empowers you to unlock deep insights across biological layers and modalities, all in one intuitive, interactive software. Leverage the Illumina analysis pipeline for streamlined, end-to-end analysis, or import select third-party data for downstream analysis. Connected Multiomics delivers a scalable solution that accelerates discoveries."



99 Q 10 B

ucsandiegoengineering A new data structure and compression technique developed by engineers at UC San Diego enables the field of pangenomics to handle unprecedented scales of genetic information.

ucsandiegoengineering A new data structure and compression technique developed by engineers at UC San Diego enables the field of pangenomics to handle unprecedented scales of genetic information.

Congrats to @ucsandiegoece Prof. Yatish Turakhia and graduate students Sumit Walia and Yu-Hsiang Tseng, pictured here, for their Nature Genetics paper describing this work!

Pangenomics, a subset of bioinformatics, is the study of many different genomes from one specific species. This can provide a more holistic picture of the natural variation and mutations that occur within a species than using one singular reference genome. This has many practical applications, such as studying how genomic mutations lead to increased transmissibility or drug resistance in pathogens.

The new method provides unmatched compression for pangenomes while also significantly advancing their representative power by encoding additional biologically relevant information, including phylogenies, mutations, and whole-genome alignments.

Learn more at the link in our bio.

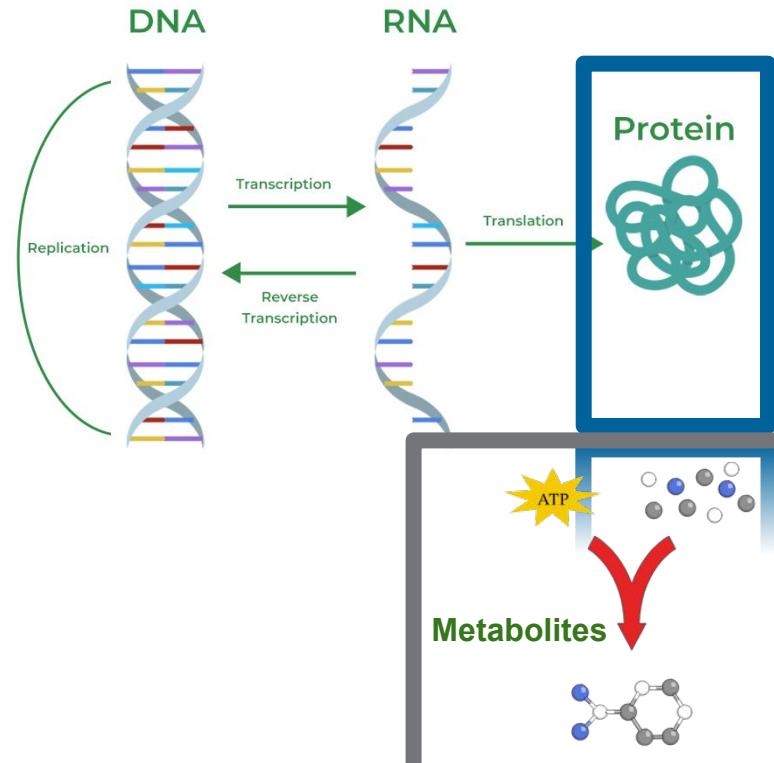
#ucsdengineering #pangenome #bioinformatics
#genomics

21 hours ago

Prerequisite: The Central Dogma

- **Core Concept:** DNA is transcribed into RNA, which is then translated into functional protein machines.
- **Proteomics** identifies the active protein machinery while **Metabolomics** characterizes the resulting physiological metabolic snapshot.
- **Bioengineering Tool:** Mastering these stages is essential for modern bioengineering.
- These technologies drive **Systems Biology** - research to understanding the larger picture by putting pieces together. It's in stark contrast to decades of reductionist biology, which involves taking the pieces apart. (Christopher Wanek, NIH)

Central Dogma of Molecular Biology



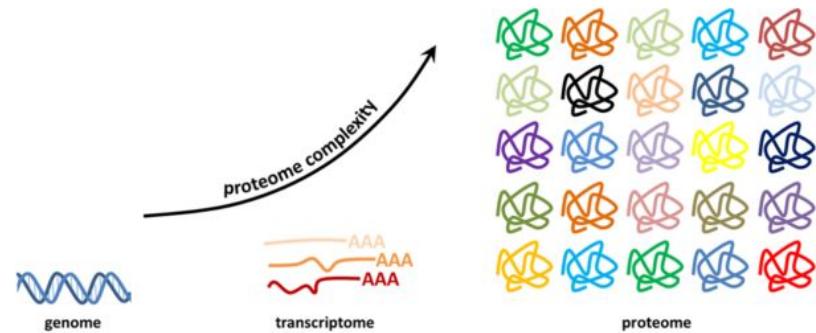
Module 1: Proteomics: Identification and Expression

(Source Pages: Chapter 2: 70–78)

- Proteins are the molecular machines of the cell**
- Proteomics studies structure, function, and expression**
- Disconnect often exists between mRNA and protein levels**

Proteomics Basics

- **Functional Machines:** Proteins serve as the actual molecular machines of the cell, providing structural support and catalyzing essential biochemical reactions.
- **The Proteome:** The proteome represents the complete set of proteins expressed by a genome in a specific cell type or tissue.
- **Analytical Perspective:** Proteomics involves a comprehensive analysis of protein structure, function, and expression profiles.



Proteomicscenter.nl

Why Study Proteomics?

- **The Correlation Gap:** Cellular mRNA levels often fail to correlate accurately with actual protein abundance due to independent regulation of translation.
- **Functional Readout:** Transcriptomics identifies which genes are active but cannot confirm the presence of functional, stable, or active protein machines.
- **Crucial Modifications:** Only proteomics can detect post-translational modifications, such as phosphorylation or glycosylation, which are essential for determining a protein's activity.

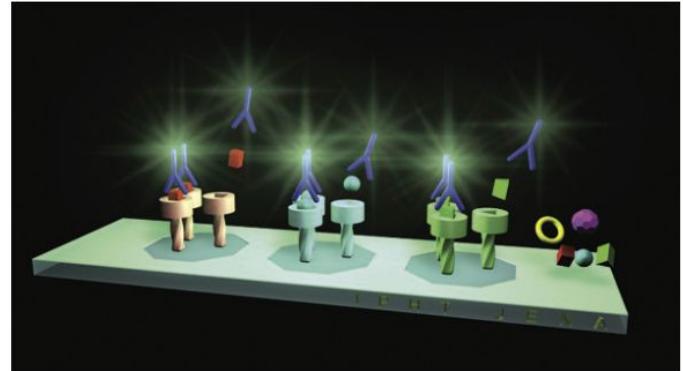


Fig. 10.51 Visionary idea of a protein chip.

Protein Separation: 2D PAGE

- **Multi-Dimensional Sorting:** Two-dimensional polyacrylamide gel electrophoresis (2D PAGE) separates thousands of proteins based on the distinct properties of charge and mass.
- **Matrix Migration:** Extracted proteins move through a gel matrix in an electrical field at rates determined by their unique biochemical characteristics.
- **Resolving Power:** Depending on the gel size and protein abundance, this method can successfully resolve approximately 2,000 different protein spots.

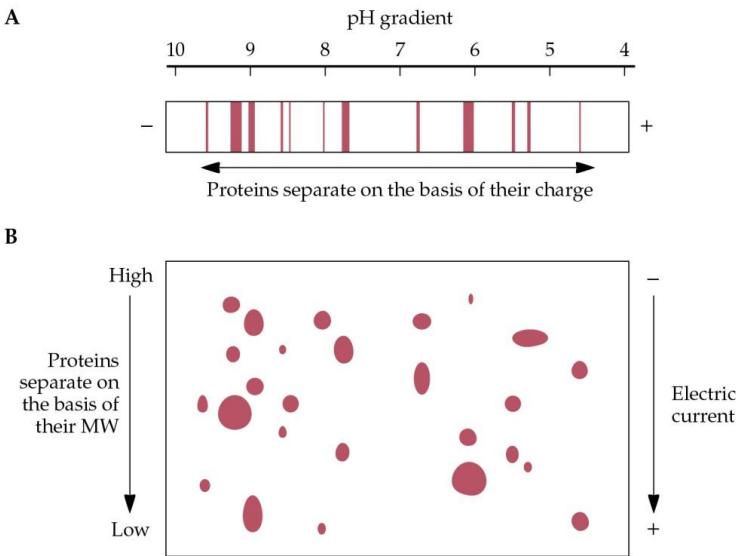


Figure 2.43 2D PAGE for separation of proteins. **(A)** First dimension. Isoelectric focusing is performed to first separate proteins in a mixture on the basis of their net charge. The protein mixture is applied to a pH gradient gel. When an electric current is applied, proteins will migrate toward either the anode (+) or cathode (-) depending on their net charge. As proteins move through the pH gradient, they will gain or lose protons until they reach a point in the gel where their net charge is zero. The pH in this position of the gel is known as the isoelectric point and is characteristic of a given protein. At that point, a protein no longer moves in the electric current. **(B)** Second dimension. Several proteins in a sample may have the same isoelectric point and therefore migrate to the same position in the gel in the first dimension. Therefore, proteins are further separated on the basis of differences in their molecular weights (MW) by electrophoresis, at a right angle to the first dimension, through a sodium dodecyl sulfate-polyacrylamide gel.

Dimension 1: Isoelectric Focusing

- **Charge Separation:** In the first dimension, proteins are separated based on their net charge by migrating through an immobilized pH gradient.
- **Isoelectric Point (pi):** Migration continues until the protein reaches a specific pH where its net charge becomes zero, known as its pi.
- **Immobilized Gradients:** The use of stable pH gradients ensures that each protein stops precisely at its characteristic isoelectric point.

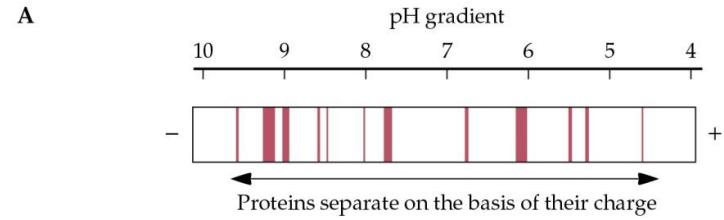
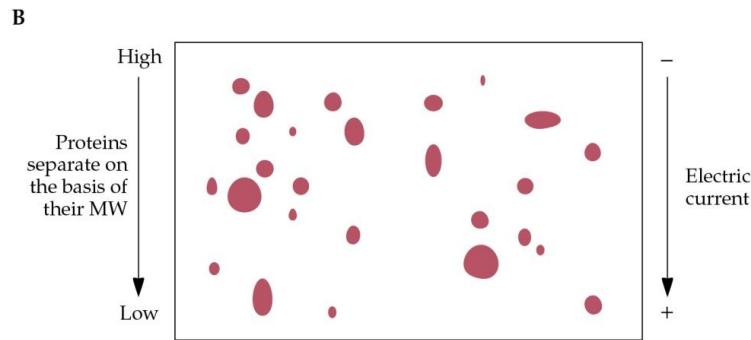


Figure 2.43 2D PAGE for separation of proteins. (A) First dimension. Isoelectric focusing is performed to first separate proteins in a mixture on the basis of their net charge. The protein mixture is applied to a pH gradient gel. When an electric current is applied, proteins will migrate toward either the anode (+) or cathode (-) depending on their net charge. As proteins move through the pH gradient, they will gain or lose protons until they reach a point in the gel where their net charge is zero. The pH in this position of the gel is known as the isoelectric point and is characteristic of a given protein. At that point, a protein no longer moves in the electric current.

Dimension 2: SDS-PAGE

- **Mass Separation:** Proteins are further separated by molecular mass by performing electrophoresis at a right angle to the first-dimension gel.
- **Uniform Charge:** Sodium dodecyl sulfate (SDS) denatures proteins and provides a uniform negative charge-to-mass ratio for consistent separation by size.
- **Sieving Effect:** The polyacrylamide gel acts as a sieve, allowing smaller protein molecules to migrate faster through the matrix than larger ones.



(B) Second dimension. Several proteins in a sample may have the same isoelectric point and therefore migrate to the same position in the gel in the first dimension. Therefore, proteins are further separated on the basis of differences in their molecular weights (MW) by electrophoresis, at a right angle to the first dimension, through a sodium dodecyl sulfate-polyacrylamide gel.

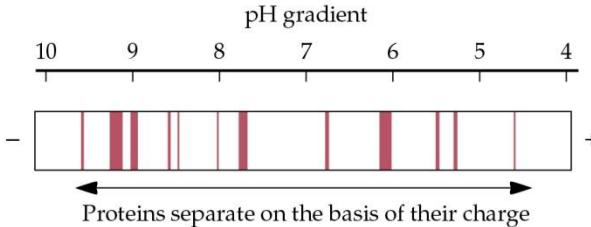
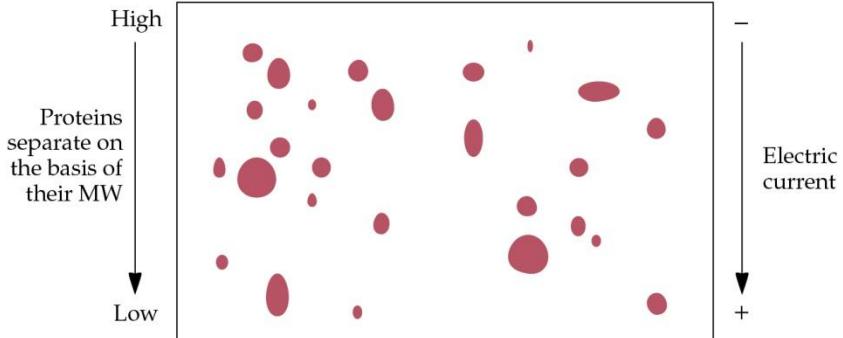
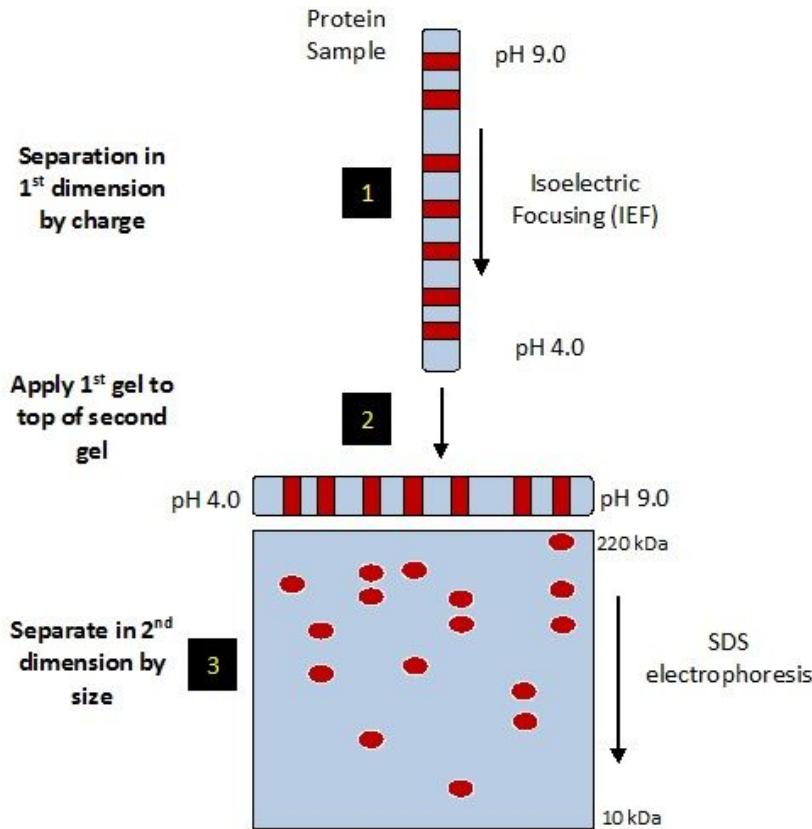
A**B**

Figure 2.43 2D PAGE for separation of proteins. **(A)** First dimension. Isoelectric focusing is performed to first separate proteins in a mixture on the basis of their net charge. The protein mixture is applied to a pH gradient gel. When an electric current is applied, proteins will migrate toward either the anode (+) or cathode (-) depending on their net charge. As proteins move through the pH gradient, they will gain or lose protons until they reach a point in the gel where their net charge is zero. The pH in this position of the gel is known as the isoelectric point and is characteristic of a given protein. At that point, a protein no longer moves in the electric current. **(B)** Second dimension. Several proteins in a sample may have the same isoelectric point and therefore migrate to the same position in the gel in the first dimension. Therefore, proteins are further separated on the basis of differences in their molecular weights (MW) by electrophoresis, at a right angle to the first dimension, through a sodium dodecyl sulfate-polyacrylamide gel.



Identification: Mass Spectrometry (MS)

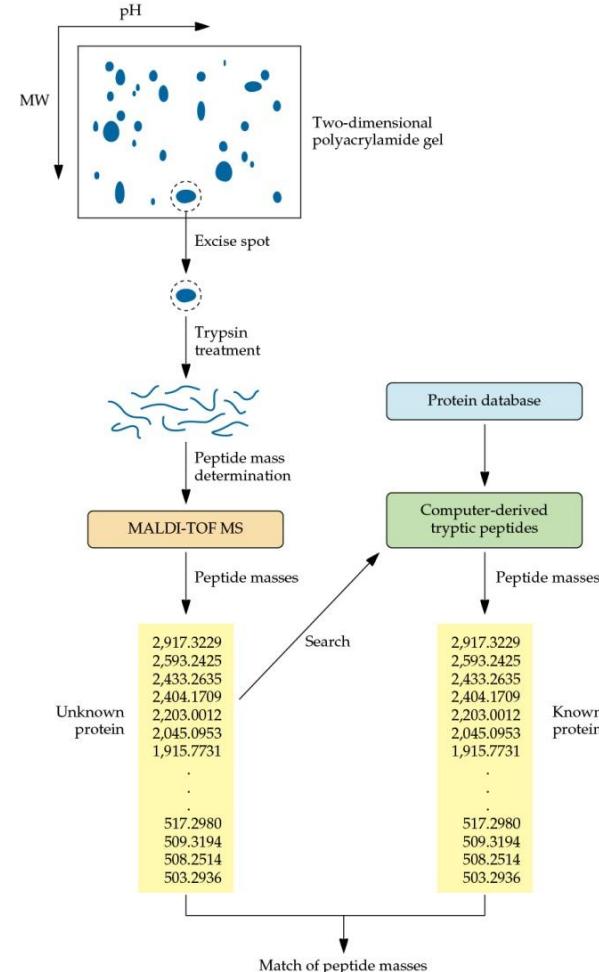
- **Mass Detection:** Mass spectrometry serves as a powerful analytical tool that identifies molecules by precisely detecting the masses of their ionized forms.
- **Analyte Fragmentation:** Before identification, a single protein spot is excised from a gel and digested into smaller peptide fragments using proteases.
- **Ion Sorting:** These ionized peptides are separated within an electromagnetic field based on their specific mass-to-charge (m/z) ratios.



Peptide Mass Fingerprinting

- **Tryptic Digestion:** The enzyme trypsin cleaves proteins at specific lysine and arginine residues to produce a unique set of peptide fragments.
- **MALDI-TOF Accuracy:** MALDI-TOF MS determines the precise masses of these peptides, which are then used as a unique molecular "fingerprint."
- **Database Matching:** Computer algorithms compare these measured masses against theoretical tryptic peptide masses derived from known genomic sequences.

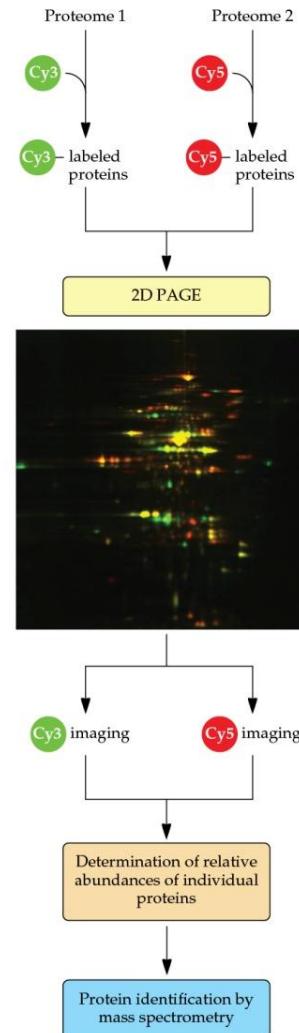
Figure 2.44 Peptide mass fingerprinting. A spot containing an unknown protein that was separated by 2D PAGE is excised from the gel and treated with trypsin. Purified tryptic peptides are separated by MALDI-TOF MS. The set of peptide masses from the unknown protein are used to search a database that contains the masses of tryptic peptides for every known sequenced protein, and the best match is determined. The trypsin cleavage sites of known proteins are determined from the amino acid sequence and, consequently, the masses of the tryptic peptides are easy to calculate. Only some of the tryptic peptide masses for the unknown protein are listed in this example.



Comparing Proteomes: 2D DIGE

- **Differential Labeling:** 2D differential in-gel electrophoresis (2D DIGE) labels two different proteome samples with distinct fluorescent dyes, such as Cy3 and Cy5.
- **Internal Standards:** Both labeled samples are mixed and run on a single gel to eliminate variability between separate electrophoresis runs.
- **Quantitative Comparison:** The ratio of fluorescence intensities in each spot identifies proteins that are either up- or downregulated across conditions.

Figure 2.45 Protein expression profiling using 2D differential in-gel electrophoresis. The proteins of two different proteomes are labeled with fluorescent dyes Cy3 and Cy5, respectively. The labeled proteins from the two samples are combined and separated by 2D PAGE. The gel is scanned for each fluorescent dye, and the relative levels of the two dyes in each protein spot are recorded. Each spot with an unknown protein is excised for identification by MS.



Antibody Microarrays

- **Target Capture:** Immobilized antibodies are arrayed on a surface to capture specific target proteins from a complex cellular lysate.
- **Fluorescence Intensity:** Labeled proteins from two samples are applied to the array, and their relative abundance is measured using a scanner.
- **High-Throughput Diagnostic:** These arrays are commercially used to screen for changes in signaling proteins associated with cancers and autoimmune diseases.

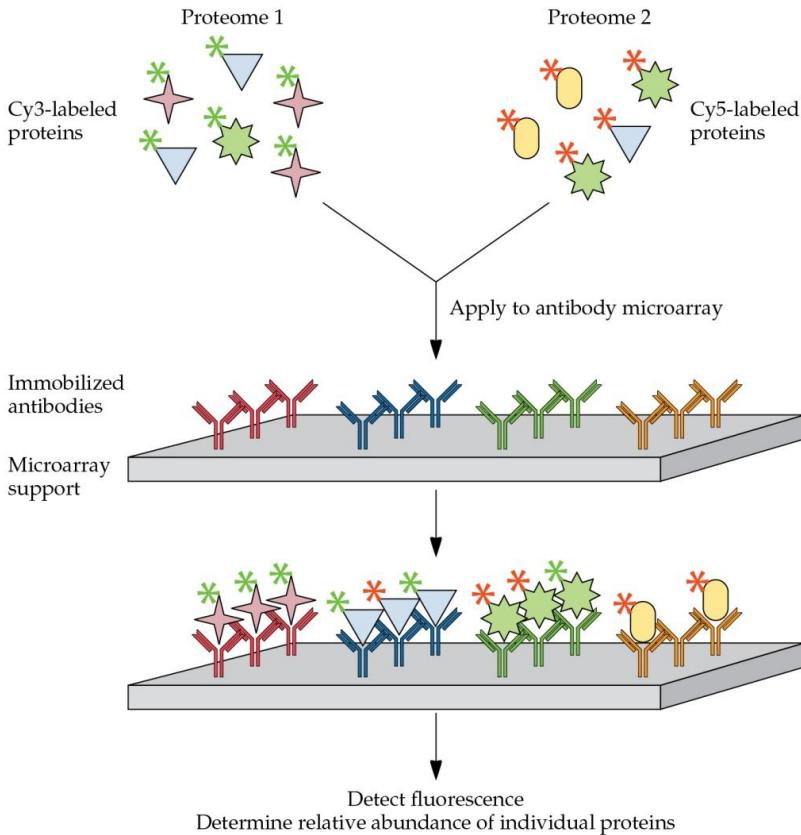


Figure 2.46 Protein expression profiling with an antibody microarray. Proteins extracted from two different samples are labeled with fluorescent dyes Cy3 and Cy5, respectively. The labeled proteins are mixed and incubated with an array of antibodies immobilized on a solid support. Proteins bound to their cognate antibodies are detected by measuring fluorescence, and the relative levels of specific proteins in each sample are determined.

Post-translational Modifications (PTMs)

- **Assay Variety:** "Sandwich" assays use secondary antibodies or lectins to detect modifications like phosphorylation or glycosylation.
- **Functional Influence:** PTMs are critical because they profoundly influence a protein's activity, stability, and its localization within the cell.
- **Glycan Detection:** Lectins are specialized proteins used to bind and identify specific carbohydrate groups on the surface of proteins.

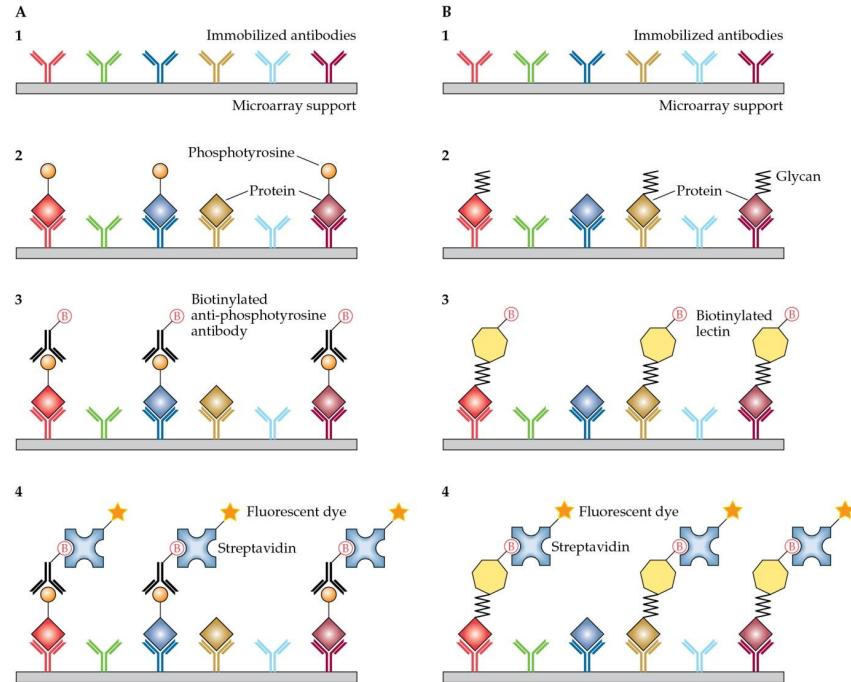


Figure 2.47 Detection of posttranslational modifications with antibody microarrays. **(A)** Detection of tyrosine phosphorylation. An antibody microarray (1) is incubated with a protein sample (2). Biotinylated antiphosphotyrosine antibody is added (3), and for visualization, a streptavidin-fluorescent dye conjugate attaches to the biotin of the antiphosphotyrosine antibody (4). **(B)** Detection of glycan groups. An antibody microarray (1) is incubated with a protein sample (2). A biotinylated molecule (e.g., lectin) that binds to a specific glycan is added (3) and, for visualization, a streptavidin-fluorescent dye conjugate attaches to the biotin of the lectin (4).

Protein Microarrays for Drug Discovery

- **Interaction Mapping:** Purified proteins arrayed on slides are probed with small molecules to identify potential drug-target interactions.
- **Mechanism Action:** This method aids in determining the mechanism of drug action by identifying which proteins interact with therapeutics.
- **Side Effect Prediction:** Arrays help predict undesirable side effects by showing if a drug binds to unintended protein targets.

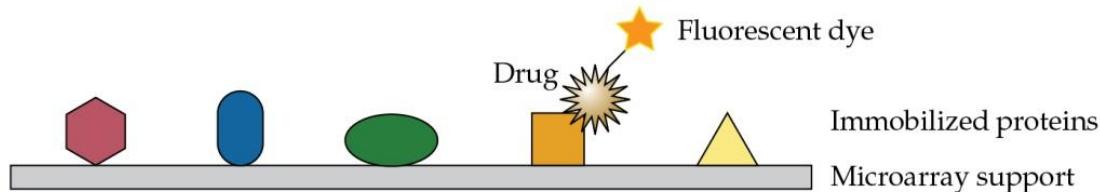


Figure 2.49 Protein microarrays to detect protein-drug interactions. Therapeutic drugs or other small molecules tagged with a fluorescent dye are applied to purified proteins arrayed on a solid support.

Video - Concepts in the module or a demonstration

- [How Does Mass Spectrometry Work](#) (1min47s)
- [Proteomics](#) (2min32s)

More videos

- [David Gonzales - proteomics pro](#)
- [MALDI-TOF MS](#)

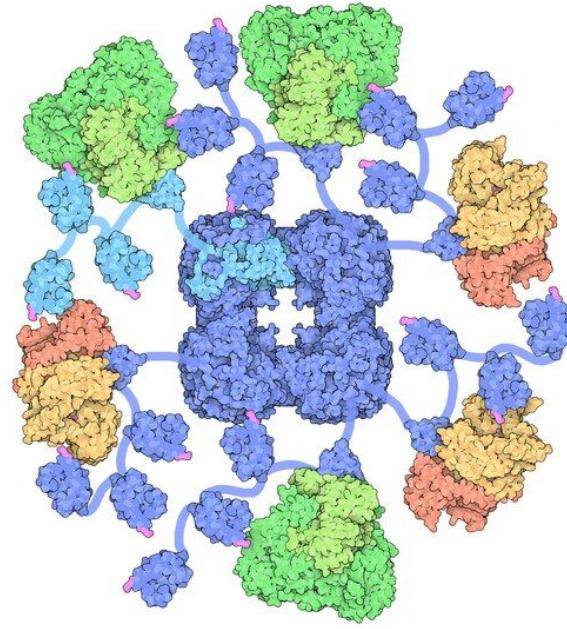
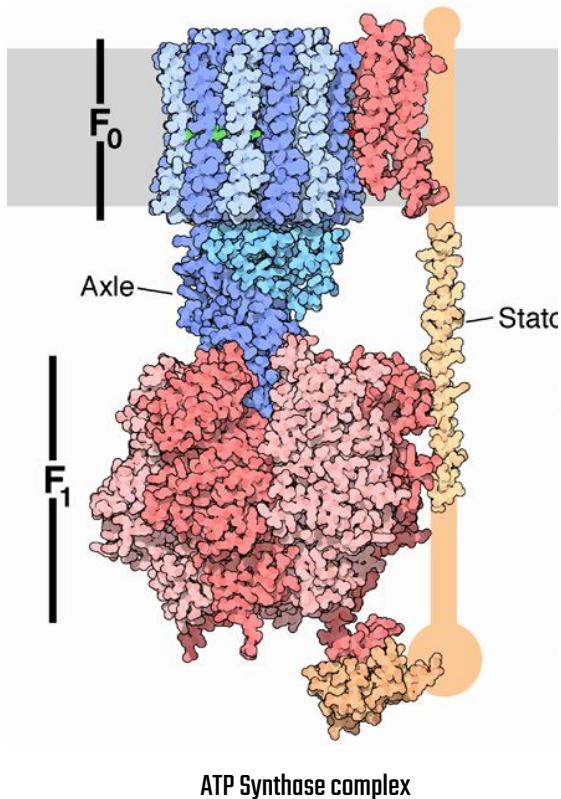
Module 2: Interaction Proteomics and Metabolomics

(Source Pages: Chapter 2: 78–90)

- Mapping functional interconnections in the proteome**
- Proteins function as large, multi-component complexes**
- Metabolomics provides a snapshot of small molecules**

Protein Complexes

- **Protein Complexes:** Multisubunit complexes (as well as single subunit proteins) can interact with each other and other cellular components (DNA, RNA, etc.)



Protein–Protein Interactions (PPIs)

- **Interaction Networks:** Proteins rarely act alone and typically function as large complexes comprising different interacting subunits for essential cellular tasks.
- **Functional Interconnections:** Thousands of interactions occur within a cell, ranging from stable structural components to short-lived transient signals.
- **Mapping Complexity:** Determining the functional interconnections among proteome members is essential for understanding cell physiology and signal transduction.

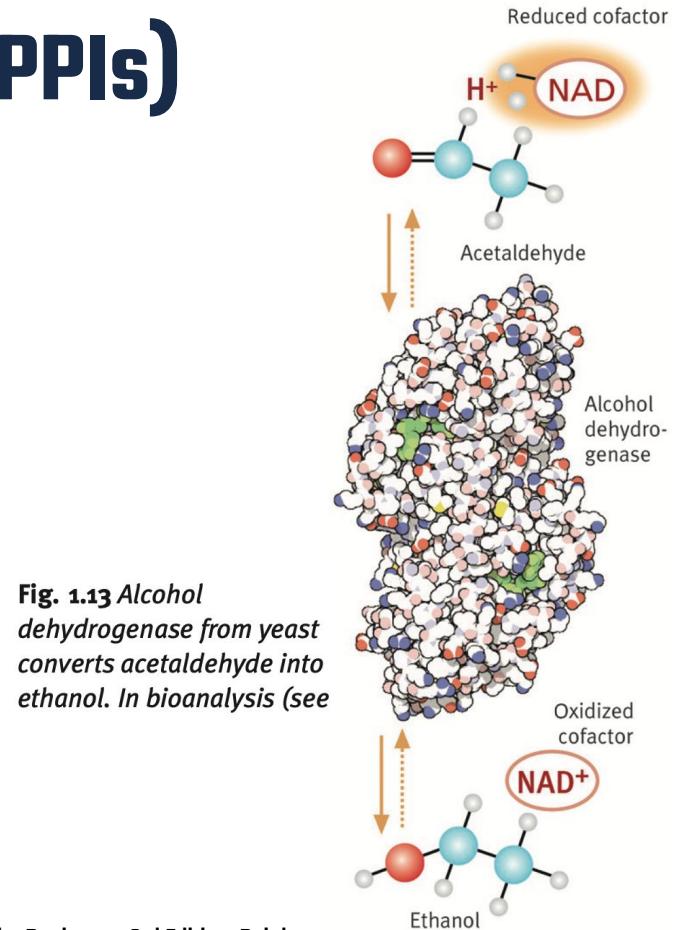
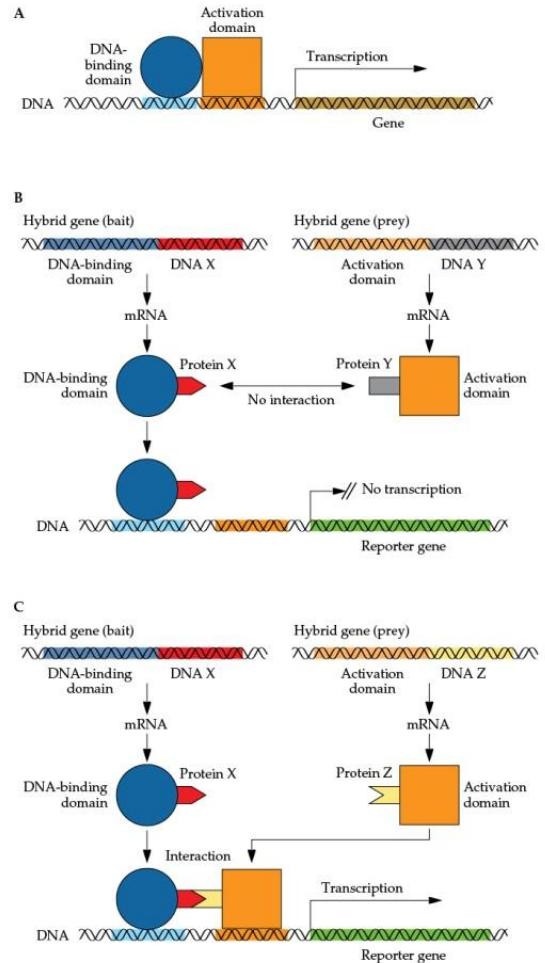


Fig. 1.13 Alcohol dehydrogenase from yeast converts acetaldehyde into ethanol. In bioanalysis (see

Yeast Two-Hybrid (Y2H) Assay

- **Physical Reconstitution:** Pairwise protein interactions are tested by reconstituting an active transcription factor within the nucleus of a living yeast cell.
- **Bait and Prey:** A "bait" protein is fused to a DNA-binding domain, while a potential "prey" protein is fused to an activation domain.
- **Reporter Activation:** If the proteins interact, the two domains connect to initiate the expression of a measurable reporter gene.



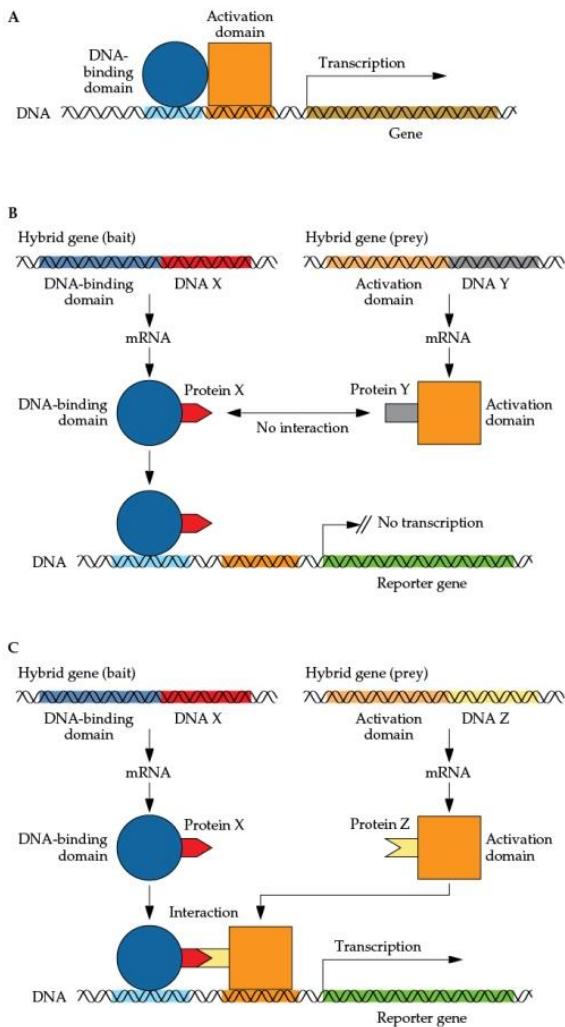
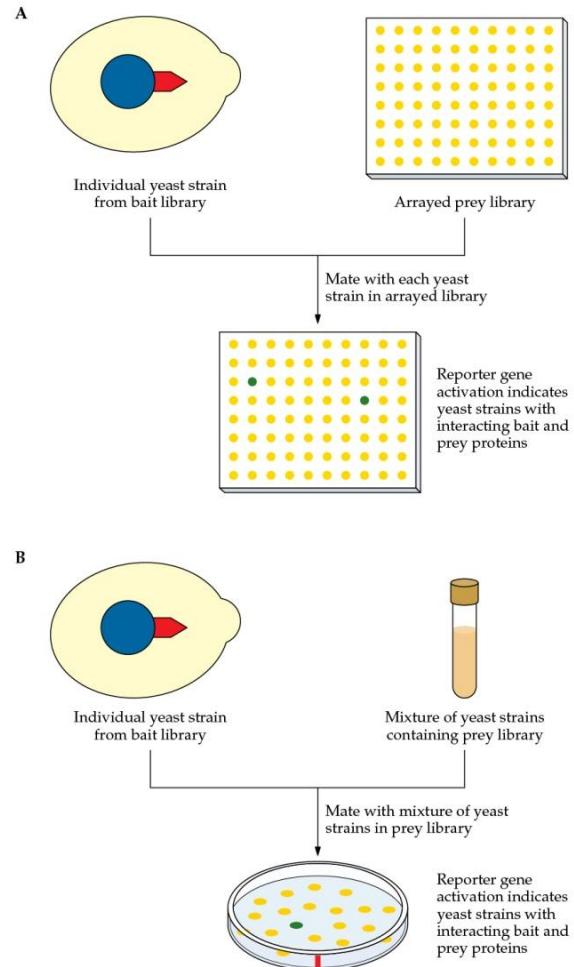


Figure 2.50 Two-hybrid assay for detecting pairwise protein interactions. **(A)** The DNA-binding domain of a transcription factor binds to a specific sequence in the regulatory region of a gene which orients and localizes the activation domain that is required for the initiation of transcription of the gene by RNA polymerase. **(B)** The coding sequences for the DNA-binding domain and the activation domain are fused to DNA X and DNA Y, respectively, and both constructs (hybrid genes) are introduced into a cell. After translation, the DNA-binding domain–protein X fusion protein binds to the regulatory sequence of a reporter gene. However, protein Y (prey) does not interact with protein X (bait), and the reporter gene is not transcribed because the activation domain does not, on its own, associate with RNA polymerase. **(C)** The coding sequence for the activation domain is fused to the DNA for protein Z (DNA Z) and transformed into a cell containing the DNA-binding domain–DNA X fusion construct. The proteins encoded by the hybrid genes interact, and the activation domain is properly oriented to initiate transcription of the reporter gene demonstrating a specific protein-protein interaction.

Large-Scale Interaction Mapping

- **En Masse Screening:** High-throughput screens mate yeast strains containing whole-proteome libraries to identify thousands of potential bait-prey interactions simultaneously.
- **Mapping the Interactome:** These pairwise maps establish the functional links between multiprotein complexes and essential cellular activities.
- **Validation Challenges:** Large-scale Y2H screens often identify false positives (non-biological interactions) and false negatives (missed interactions), requiring subsequent validation.



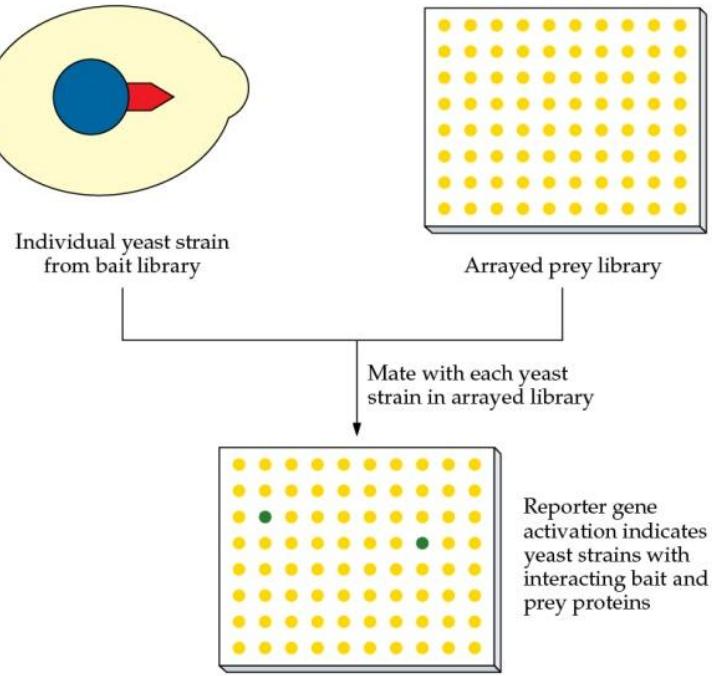
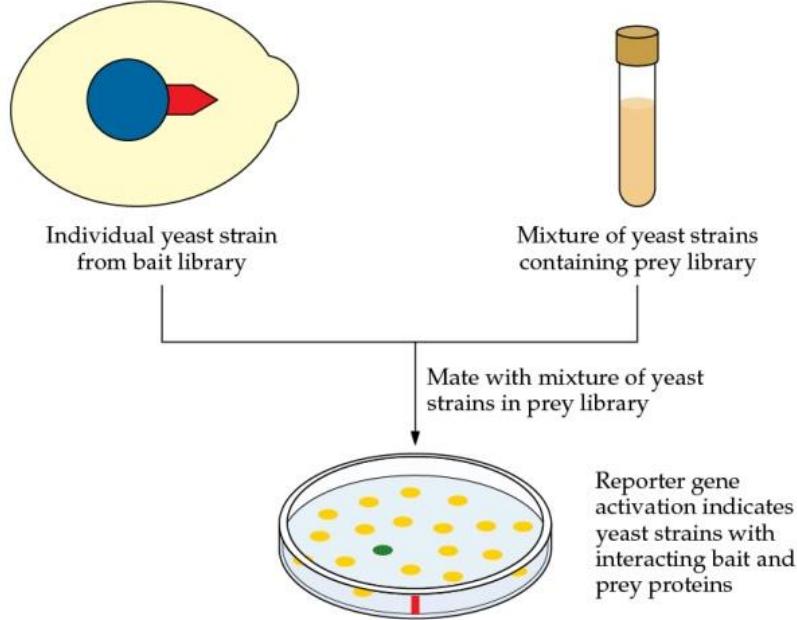
A**B**

Figure 2.51 Large-scale screens for protein interactions using the yeast two-hybrid system. Two libraries are prepared, one containing genomic DNA fragments fused to the coding sequence for the DNA-binding domain of a transcription factor (bait library) and another containing genomic DNA fragments fused to the coding sequence for the activation domain of the transcription factor (prey library). Two methods are commonly used to screen for pairwise protein interactions. **(A)** Individual yeast strains in the

bait library are mated with each yeast strain in an arrayed prey library. Resulting strains in the array that produce bait and prey proteins that interact are detected by assaying for reporter gene activation (cells growing in a multiwell plate that express the reporter gene are indicated in green). **(B)** Yeast strains in the prey library are mated *en masse* with individual strains in the bait library. The mixture of strains is screened for reporter gene activity that identifies strains with interacting bait and prey proteins (green).

Tandem Affinity Purification (TAP)

- **Complex Capture:** The TAP method captures entire multiprotein complexes from cells by using a bait protein expressed with two affinity tags.
- **High Purity:** Successive rounds of affinity binding and protease cleavage ensure that the final purified target complex is free of contaminants.
- **Team Identification:** This technique reveals the entire "team" of proteins that naturally interact with a specific bait protein *in vivo*.

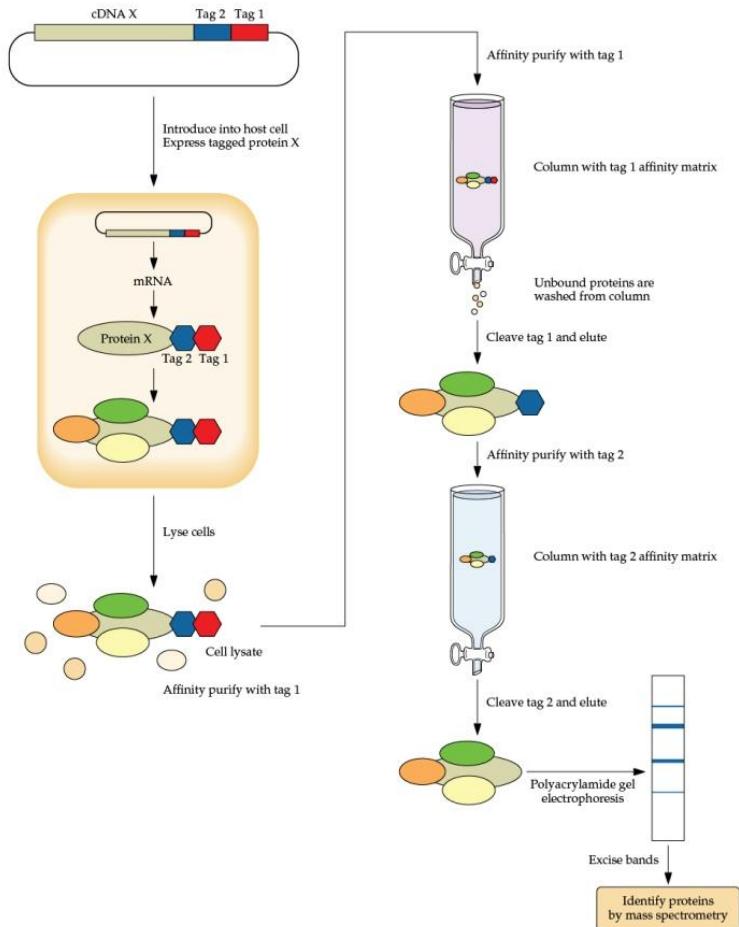
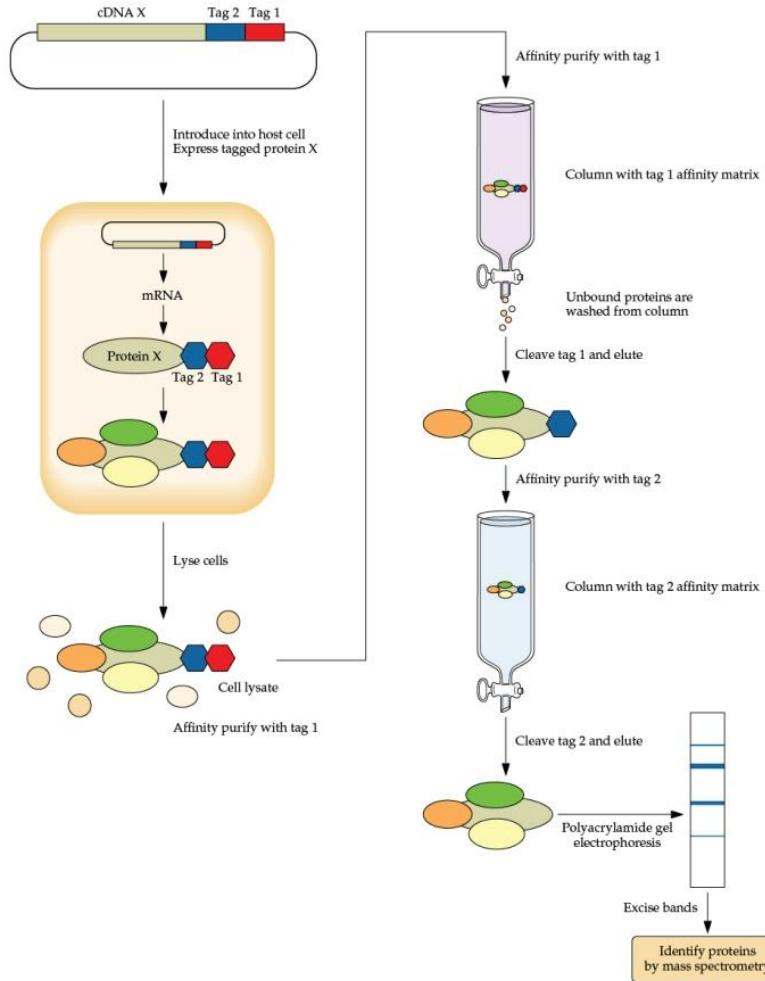


Figure 2.52 Tandem affinity purification to detect multiprotein complexes. The coding region of a cDNA (cDNA X) is cloned into a vector in frame with two DNA sequences (tag 1 and tag 2), each encoding a short peptide that has a high affinity for a specific matrix. The cDNA-tag construct is introduced into a host cell, where it is transcribed and the mRNA is translated. Other cellular proteins bind to the protein encoded by cDNA X (protein X). The complex consisting of protein X and its interacting proteins (colored shapes) is separated from other cellular proteins by the binding of tag 1 to an affinity matrix which is usually fixed to a column. The protein complex is retained on the column, and the noninteracting proteins flow through. The complex is then eluted from the affinity matrix by cleaving off tag 1 with a protease, and a second purification step is carried out with tag 2 and its affinity matrix. The proteins of the complex are separated by one-dimensional PAGE. Single bands are excised from the gel and identified by MS.

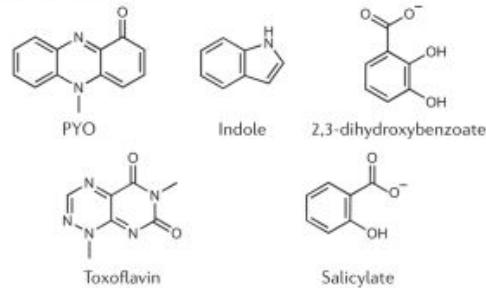


Introduction to Metabolomics

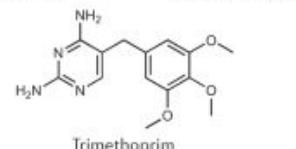
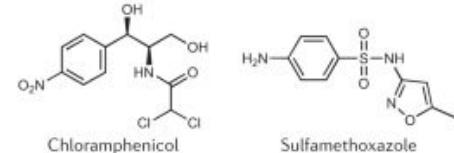
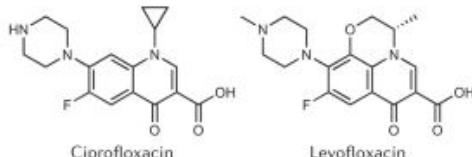
- **Small Molecule Study:** Metabolomics studies the profiles of small molecules, such as sugars (carbohydrates) and lipids, in biological systems.
 - **Diverse Influences:** Metabolite composition is influenced by an organism's genotype, its current health status, and the environment.
 - **Disease Biomarkers:** Comprehensive metabolite profiles serve as biomarkers for diagnosing disease and monitoring the efficacy of treatments.

secondary metabolites

b Metabolite



Antibiotics

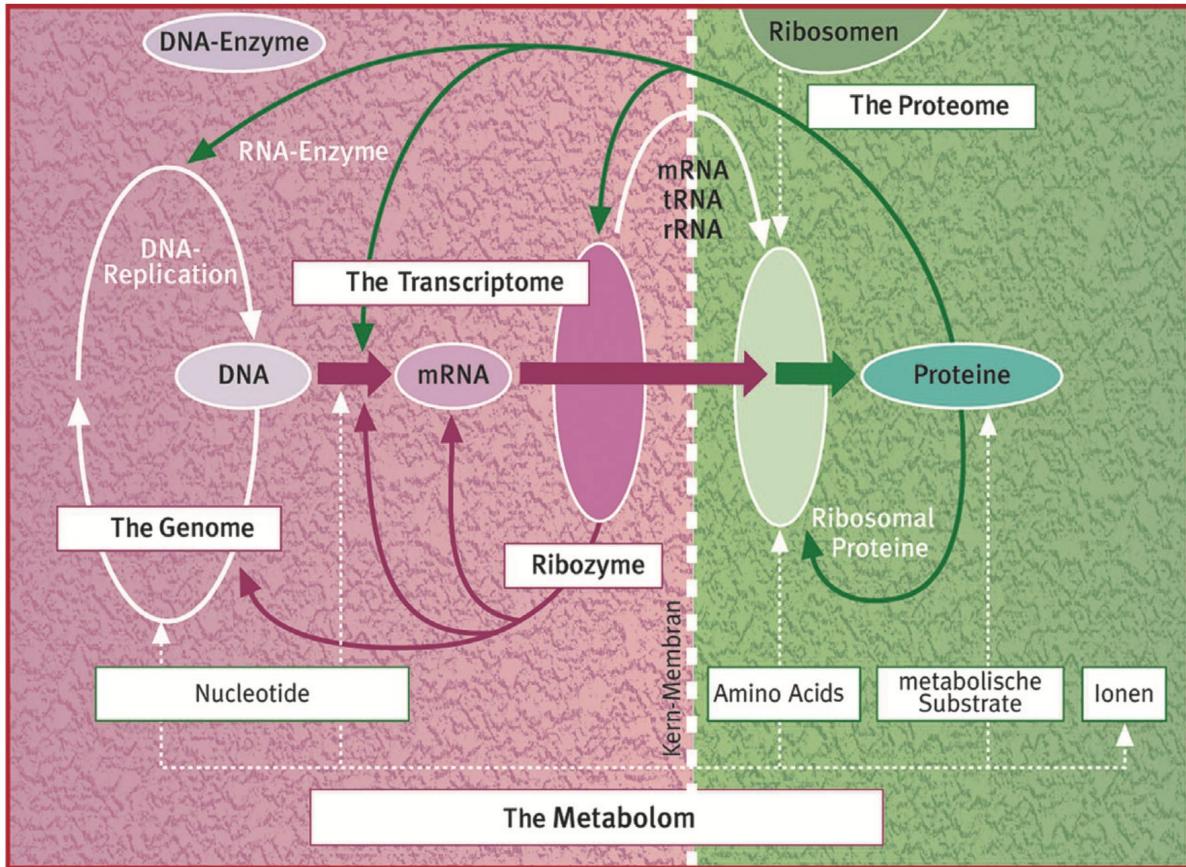


The 4 major Omics and Interactions

Metabolomics
Complete All the
major omics

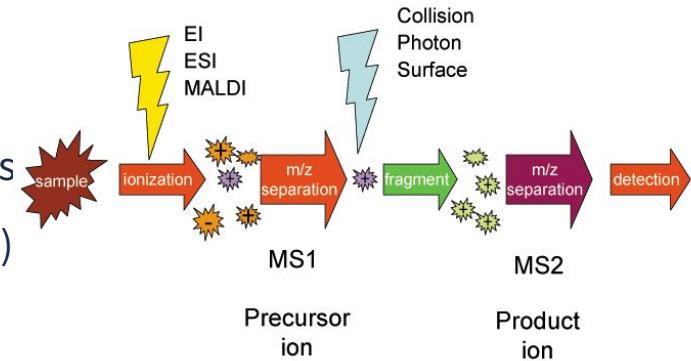
Biotechnology for Beginners, 3rd Edition, Reinhard
Renneberg, eBook ISBN: 9780323855709

Fig. 4.16 Interrelations between
metabolome, genome, transcriptome,
and proteome. The metabolome
comprises all characteristic metabolic
properties of a cell or a tissue.



Metabolomic Methods: MS and NMR

- **Mass Spectrometry (MS):** Measures **mass-to-charge ratios of ionized metabolites**, typically following chromatographic separation.
- **LC-MS vs. GC-MS:** Liquid chromatography separates dissolved molecules via retention times, while gas chromatography requires volatile or derivatized samples
- **NMR Spectroscopy:** Nuclear Magnetic Resonance (NMR) provides structural identification by measuring how atomic nuclei absorb electromagnetic energy in a magnetic field; this **non-destructive method** can visualize living cells.
- **Tandem MS (MS-MS):** Utilizes two stages of mass analysis, ion selection and subsequent fragmentation, to provide high-specificity structural identification.



https://en.wikipedia.org/wiki/Tandem_mass_spectrometry

Metabolomics Case Study: PAH

- **Disease Profiles:** Researchers used metabolomics to compare 376 metabolites in lung tissue from patients with pulmonary arterial hypertension (PAH) versus healthy controls.
- **Metabolic Shifts:** Diseased tissue showed a significant decrease in glycolytic intermediates and a large increase in tricarboxylic acid (TCA) cycle intermediates.
- **Diagnostic Targets:** The results suggest a shift from carbohydrate to fatty acid energy supply, providing new targets for metabolic modulation therapy.

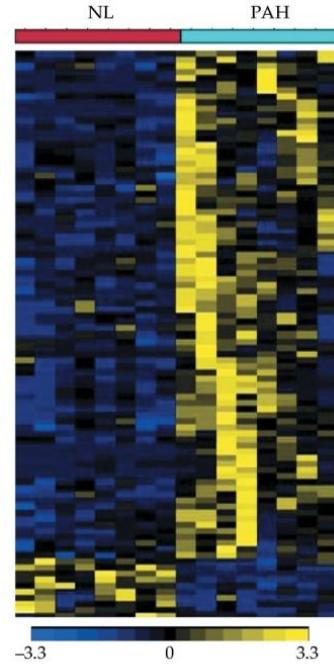
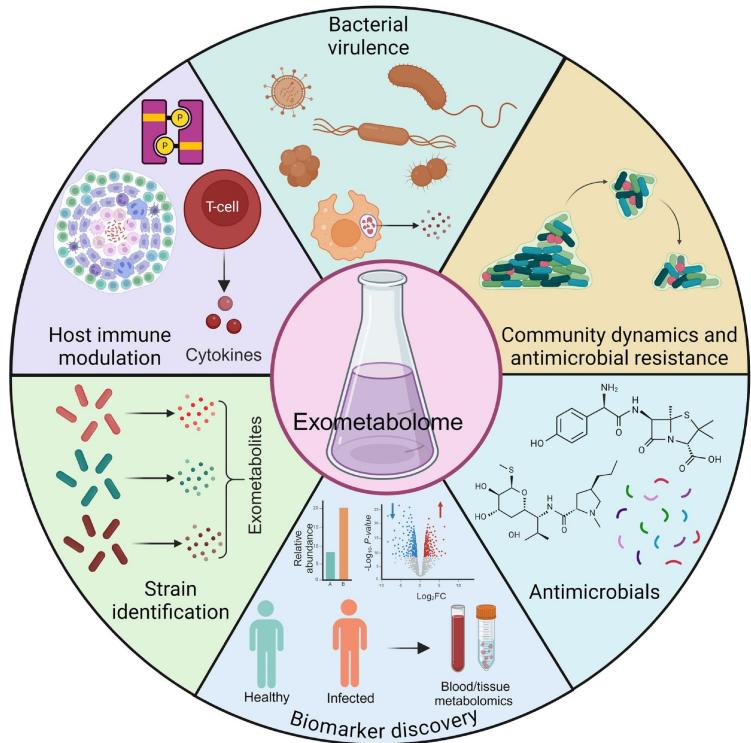


Figure 2.54 Metabolite profiles associated with severe pulmonary arterial hypertension (PAH). The relative levels of 93 metabolites in lung tissue samples from eight patients with PAH and eight individuals without the disease (NL) are represented as a heat map. Levels of a metabolite (rows) in each tissue sample (columns) were compared to the median metabolite level (black); shades of yellow represent increased levels, and shades of blue indicate decreased levels. In tissue samples from patients with PAH, 83 metabolites, including glucose, tricarboxylic acid cycle intermediates, and dicarboxylic fatty acids, were more abundant, and 10 metabolites, including intermediates in glycolysis, were less abundant than in normal lung tissue. Reprinted from Zhao et al., *PLoS ONE* 9:e88727, 2014, © 2014 Zhao et al.

Intra - and Exo - metabolomics

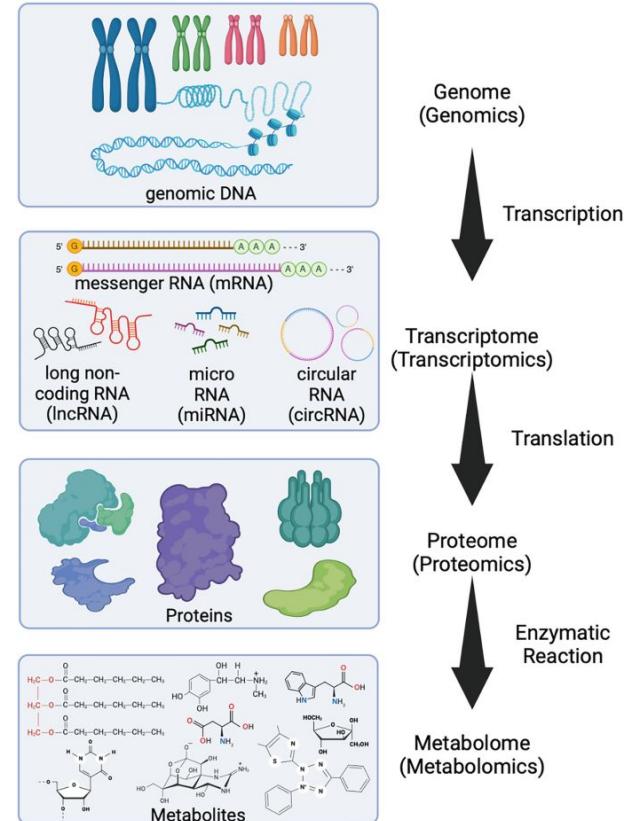
- **Biological Location:** Intra-metabolomics measures the "endometabolome" within cell or tissue lysates, whereas exo-metabolomics analyzes small molecules in cell secretions or culture media.
- **Concentration Ranges:** Intracellular metabolites are typically found in higher millimolar (mM) to micromolar (μM) ranges, while extracellular metabolites are can be high mM for major components down to the nanomolar (nM) ranges.
- **Typical Counts:** High-throughput analysis typically identifies hundreds to over a thousand metabolites within cells, whereas secreted profiles often consist of dozens to several hundred measurable molecules.



Chugh, Saurabh et al. Trends in Microbiology, Volume 33, Issue 5, 546 - 557

Systems Biology Integration

- **Integrated Omics:** Genomics defines potential, Proteomics reveals activity, and Metabolomics provides the final physiological snapshot of the entire biological system.
- **Predictive Modeling:** Combined data from all "omics" levels are used to build computational models that predict how a cell responds to perturbations.
- **Engineering Focus:** These technologies allow bioengineers to understand the larger picture by putting molecular pieces together through data integration.



Video - Concepts in the module or a demonstration

- Difference Between Genome, Transcriptome, Proteome, & Metabolome (5min44s)
- What is Metabolomics (1min22s)

More videos

- Yeast Two Hybrid System (4min39s)
- Spatial Transcriptomics (5min33s)

The End