Lecture 6. Proteomics and Protein Interaction Technologies

Proteome; protein extraction and separation; LC-MS/MS; quantitative proteomics; post-translational modifications; protein identification; interactomics and protein–protein interactions.

Dep. Biotechnology Prof. Kenzhebayeva S.S.

Aim of the lecture. 2. Expected learning outcomes

Aim of the lecture

To introduce the fundamental principles of proteomics, modern methods of protein extraction, separation, identification, and quantification, and approaches for investigating post-translational modifications and protein–protein interactions, emphasizing their applications in functional genomics, biotechnology, and biological research.

2. Expected learning outcomes

- Define the proteome and explain its dynamic nature.
- Distinguish between qualitative, quantitative, and functional proteomics.
- Describe the main principles of protein extraction and separation.
- Explain the operation of LC-MS/MS and its role in protein identification.
- Compare label-free and labeling-based quantitative proteomics.
- Identify major post-translational modifications (PTMs).
- Explain the principles of protein identification using mass-spectrometry data.
- Describe major technologies for studying protein–protein interactions.
- Discuss the role of proteomics and interactomics in modern biotechnology.

3. Lecture plan

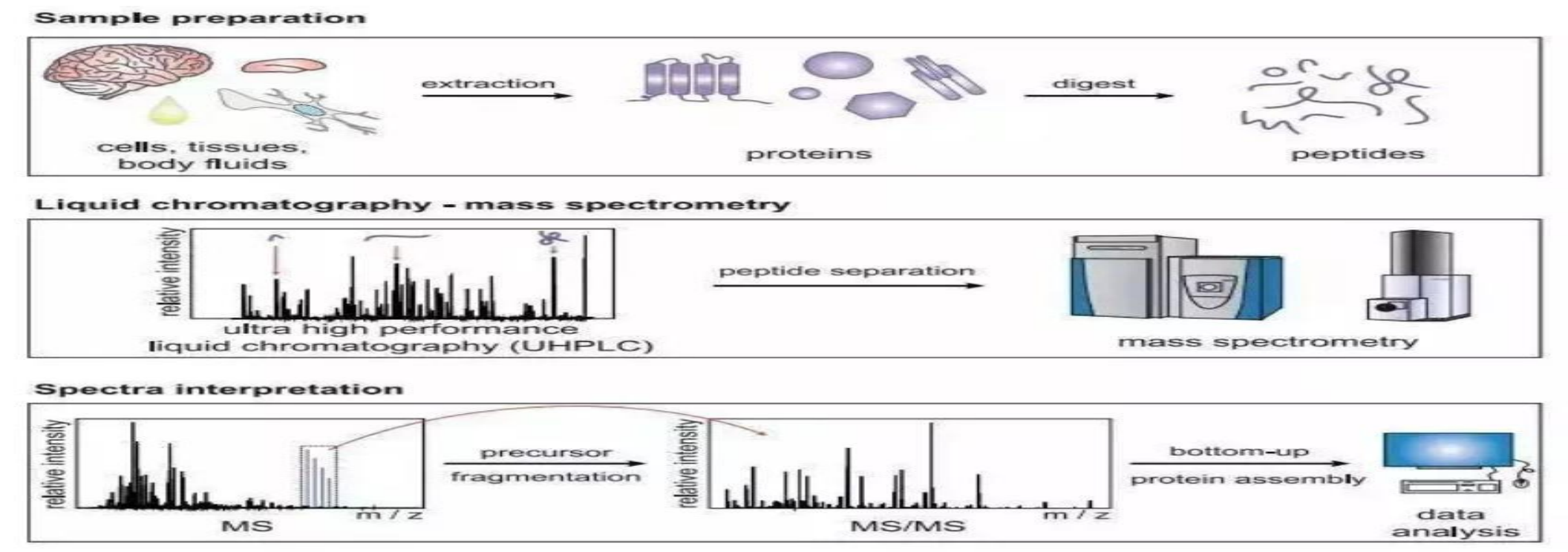
1. Introduction to proteomics: proteome, properties and biological significance

1. Protein extraction and sample preparation
1. Protein separation technologies
1. Mass spectrometry and LC-MS/MS
1. Protein identification and quantitative proteomics
1. Post-translational modifications
1. Interactomics and protein–protein interactions

1. Applications, conclusion, control questions and literature

Section 1. Introduction to proteomics.
Proteomics: concept and significance

- Proteomics is the large-scale study of proteins, including their abundance, structures, functions, modifications, localization, and interactions.
- The proteome is the complete set of proteins and protein forms present in a biological system under defined conditions.



3. From genome to proteome

The transition from genome to proteome represents the flow of genetic information from DNA to functional proteins, which determine cellular structure, metabolism, and physiological responses.

Major stages of gene expression and protein formation:

1. Genome (DNA): Contains genetic information encoding proteins and functional RNA molecules.
2. Transcriptome (RNA): Represents RNA transcripts produced through transcription, including mRNAs that serve as templates for protein synthesis.
3. Translation: Ribosomes decode mRNA sequences to synthesize polypeptide chains.
4. Post-translational modifications (PTMs): Phosphorylation, glycosylation, ubiquitination, and other modifications regulate protein activity, localization, and stability.
5. Proteome: The complete set of proteins and their molecular forms expressed by a cell, tissue, or organism under specific conditions.

Why is the proteome more complex than the genome?

- Alternative splicing can generate multiple protein isoforms from one gene.
- Post-translational modifications increase protein diversity.
- Protein abundance depends on synthesis and degradation rates.
- Protein folding, localization, and interactions influence biological functions.
- The proteome changes dynamically during development and environmental stress.

From genetic information to functional proteins

Genome DNA

Transcriptome RNA

Translation Ribosomes

Protein processing Folding + PTMs

Proteome Functional proteins

Key concept: The genome defines the genetic potential of an organism, whereas the proteome reflects the dynamic protein composition and functional state of its cells.

The proteome as a dynamic molecular system

Unlike the genome, the proteome changes continuously in response to internal and external factors.

1. **Developmental stage and cell differentiation.**
2. **Tissue and organ specificity.**
3. **Environmental conditions and stress.**
4. **Protein synthesis and degradation.**
5. **Post-translational modifications.**

Key concept: One genome can generate many different proteomes depending on cellular conditions.

Main branches of proteomics

Branch	Main objective
Qualitative proteomics	Identify proteins present in a sample
Quantitative proteomics	Measure differences in protein abundance
Structural proteomics	Investigate protein structures
Functional proteomics	Study protein functions and activities
Interaction proteomics	Characterize protein interactions
PTM proteomics	Identify and quantify protein modifications

Major applications of proteomics

Proteomics is used to investigate:

1. **Cellular metabolism and signaling pathways.**
2. **Responses to abiotic and biotic stress.**
3. **Disease biomarkers and molecular diagnostics.**
4. **Protein functions and regulatory networks.**
5. **Crop improvement and industrial biotechnology.**

In plants, proteomics helps identify proteins associated with photosynthesis, drought tolerance, pathogen defense, and nutrient metabolism.

7. Protein extraction methods

Method	Principle	Application
Mechanical homogenization	Physical disruption of cells	Plant and animal tissues
Detergent-based extraction	Solubilization of proteins and membranes	Membrane proteins
Phenol extraction	Partitioning proteins from interfering compounds	Plant proteomics
TCA/acetone precipitation	Protein precipitation and sample cleanup	Plant tissues rich in metabolites
Chaotropic extraction	Protein solubilization using urea/thiourea	Proteomic sample preparation

For plant tissues, special attention is required because phenolic compounds, polysaccharides, pigments, and other metabolites can interfere with protein analysis.

Protein quantification

Before separation or mass-spectrometry analysis, protein concentration is commonly measured using:

- **Bradford assay.**
- **BCA assay.**
- **UV absorbance at 280 nm.**

- **Accurate quantification supports comparable protein loading and reproducible results. Assay selection must account for buffer compatibility.**

9. Preparation of proteins for mass spectrometry

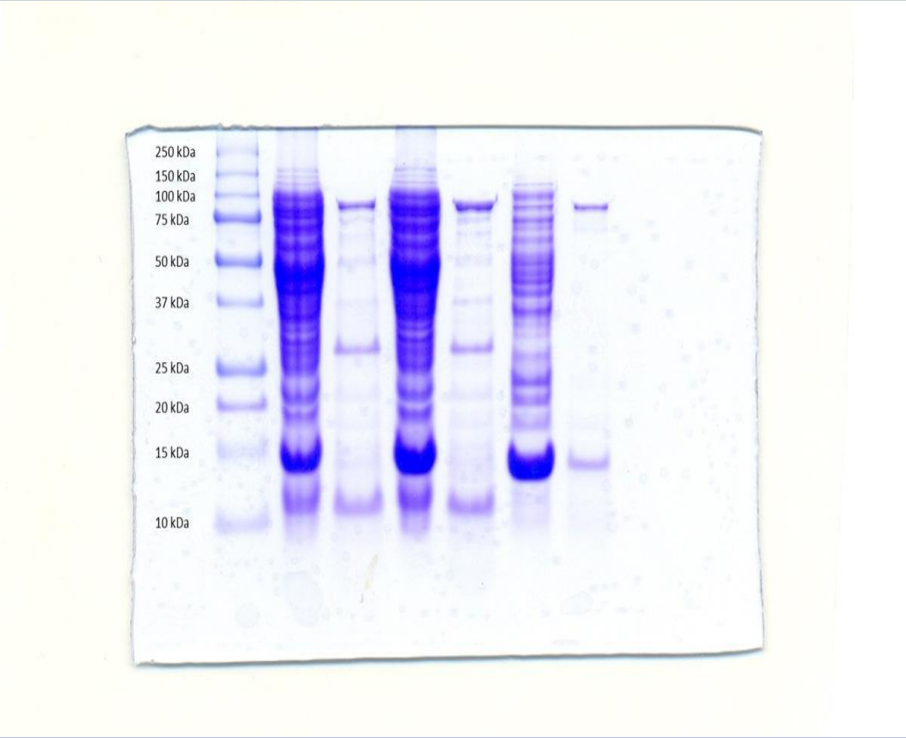
In bottom-up proteomics, proteins are enzymatically digested into peptides, commonly using trypsin.

- Protein extraction
- Protein reduction and alkylation
- Enzymatic digestion
- Peptide purification
- LC-MS/MS analysis

3. Protein separation technologies

SDS-PAGE

- Sodium dodecyl sulfate–polyacrylamide gel electrophoresis separates proteins primarily according to their molecular mass.
- SDS denatures proteins and provides a broadly similar charge-to-mass ratio.
- Smaller proteins generally migrate faster through the gel.
- Protein bands are visualized using appropriate staining methods.



Two-dimensional gel electrophoresis (2-DE)

Two-dimensional electrophoresis separates proteins according to two properties:

First dimension: isoelectric point (pI), using isoelectric focusing.

Second dimension: molecular mass, using SDS-PAGE.

2-DE can reveal differences in protein patterns between samples, although it has limitations for low-abundance, highly hydrophobic, and some extreme-pI proteins.

Chromatographic separation of proteins and peptides

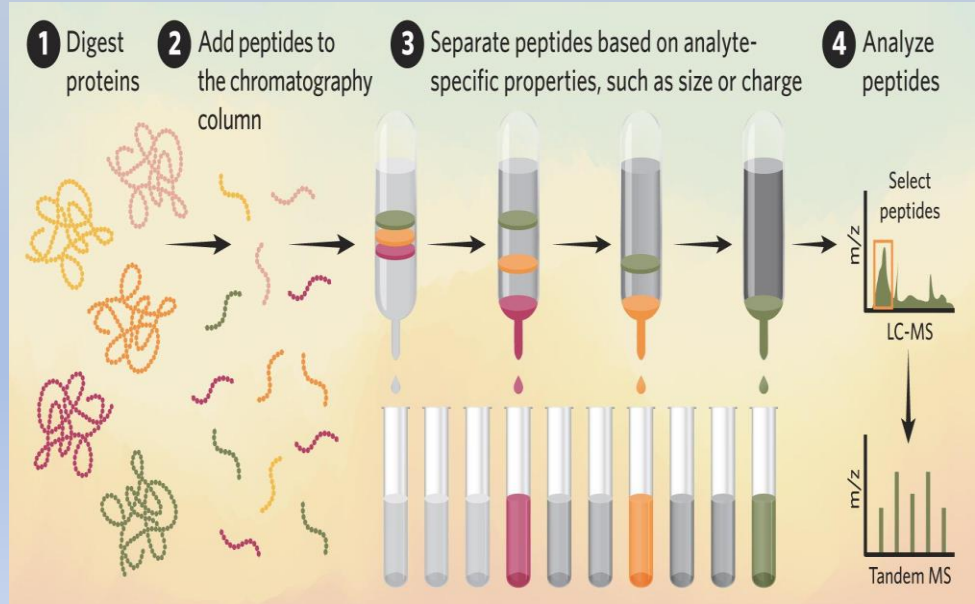
Chromatography is an analytical technique used to separate proteins and peptides according to their physicochemical properties through differential interactions with stationary and mobile phases.

Major chromatographic techniques:

Technique	Separation principle	Applications in proteomics
Size-exclusion chromatography (SEC)	Molecular size and hydrodynamic radius	Protein fractionation and analysis of protein complexes
Ion-exchange chromatography (IEX)	Net electrical charge	Protein purification and fractionation
Affinity chromatography	Specific molecular interactions	Selective purification of target proteins
Reverse-phase liquid chromatography (RP-LC)	Hydrophobic interactions	Peptide separation before mass spectrometry
High-performance liquid chromatography (HPLC)	High-efficiency separation using pressurized mobile phases	Protein and peptide analysis

Role in modern proteomics:

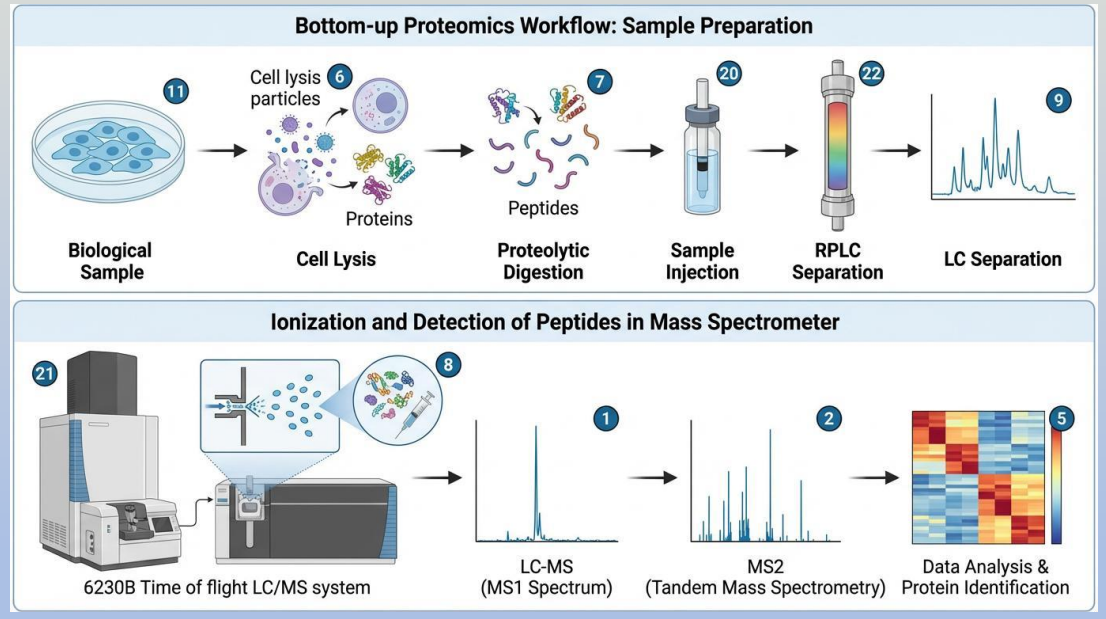
- Reduces sample complexity before mass spectrometry.
- Improves peptide identification and proteome coverage.
- Enables separation of proteins and peptides with different physicochemical characteristics.
- Supports quantitative proteomics and post-translational modification analysis.
- Nano-LC–MS/MS combines high-resolution peptide separation with sensitive mass spectrometric



Mass spectrometry and LC-MS/MS

- **Principles of mass spectrometry**
- Mass spectrometry identifies and characterizes ions according to their mass-to-charge ratio (m/z).
- A mass spectrometer generally includes:
  - *Ion source.*
  - *Mass analyzer.*
  - *Ion detector.*
  - *Data acquisition and processing system.*
- **Liquid chromatography–tandem mass spectrometry (LC-MS/MS)**
- LC-MS/MS combines peptide separation by liquid chromatography with tandem mass spectrometry.

Main stages of LC-MS/MS



Main stages of LC-MS/MS

1. *Peptide mixture*
2. *Liquid chromatography — peptide separation*
3. *Ionization — usually ESI*
4. *MS1 — precursor-ion measurement*
5. *MS2 — fragment-ion analysis*
6. *Peptide identification and protein inference*

Data-dependent and data-independent acquisition

DDA	DIA
Selects precursor ions for fragmentation based on survey scans	Fragments ions within predefined or variable m/z windows
May miss lower-abundance precursors	Can provide more consistent peptide coverage
Widely used for protein identification	Widely used for large-scale quantitative proteomics

Main stages of LC-MS/MS

- Peptide mixture
- Liquid chromatography — peptide separation
- Ionization — usually ESI
- MS1 — precursor-ion measurement
- MS2 — fragment-ion analysis
- Peptide identification and protein inference

Advantages and limitations of LC-MS/MS

- Advantages:
  - High sensitivity and analytical specificity.
  - Identification of thousands of proteins in suitable samples.
  - Detection of many PTMs.
  - Quantitative comparison between biological conditions.

- Limitations:
  - High instrument and maintenance costs.
  - Complex sample preparation.
  - Incomplete proteome coverage.
  - Challenges with very low-abundance proteins.
  - Need for specialized bioinformatic analysis.

Protein identification and quantitative proteomics

Protein identification

- Protein identification is based on matching experimentally acquired peptide mass spectra to peptide sequences predicted from protein databases or inferred by other computational methods.

Main stages:

- MS/MS spectra → database search → peptide-spectrum matches → statistical validation → protein inference
- Common software environments include MaxQuant, Proteome Discoverer, and FragPipe.

Confidence in protein identification

Protein identification requires statistical quality control.

Important concepts:

1. *Peptide-spectrum match (PSM).*
2. *Peptide identification confidence.*
3. *False discovery rate (FDR).*
4. *Protein inference.*
5. *Unique and shared peptides.*

- A 1% FDR is a commonly used threshold, although the appropriate criteria depend on the experiment and analysis level.

Quantitative proteomics

Label-free versus label-based proteomics

Feature	Label-free	Label-based
Labeling	Not required	Required
Sample processing	Generally simpler	Additional labeling steps
Multiplexing	Samples usually analyzed separately	Multiple samples can be combined
Cost	Often lower reagent cost	Often higher reagent cost
Examples	LFQ, DIA quantification	TMT, SILAC

Quantitative proteomics measures protein abundance and compares it between biological samples.

Two major strategies are:

- Label-free quantification (LFQ) — based on mass-spectrometric signal measurements without isotopic or isobaric labeling.
- Label-based quantification — uses chemical or metabolic labels, such as TMT or SILAC, where applicable.

Analysis of differentially abundant proteins

Proteomic comparisons may identify differentially abundant proteins (DAPs) between experimental conditions.

- *For example:*
- Control wheat vs drought-stressed wheat
- ↓
- Quantitative LC-MS/MS
- ↓
- Statistical analysis of protein abundance
- ↓
- Identification of DAPs
- ↓
- GO and pathway enrichment
- ↓
- Interpretation of stress-response mechanisms

Post-translational modifications

Post-translational modifications (PTMs) are covalent chemical changes occurring to proteins during or after their synthesis.

- PTMs can regulate:
- Enzymatic activity.
- Protein stability.
- Cellular localization.
- Molecular interactions.
- Signal transduction.

Major types of post-translational modifications (PTMs)

Post-translational modifications (PTMs) are covalent chemical modifications of proteins that occur during or after translation and regulate protein activity, stability, localization, and molecular interactions.

Significance in proteomics:

- Generate multiple molecular forms (proteoforms) from a single protein.
- Regulate enzyme activity and protein–protein interactions.
- Participate in developmental and environmental signaling.
- Can be characterized using enrichment strategies and LC–MS/MS.

Example in plant biotechnology: Protein phosphorylation and SUMOylation participate in plant responses to drought, salinity, and temperature stress, while oxidative modifications can influence antioxidant enzyme function.

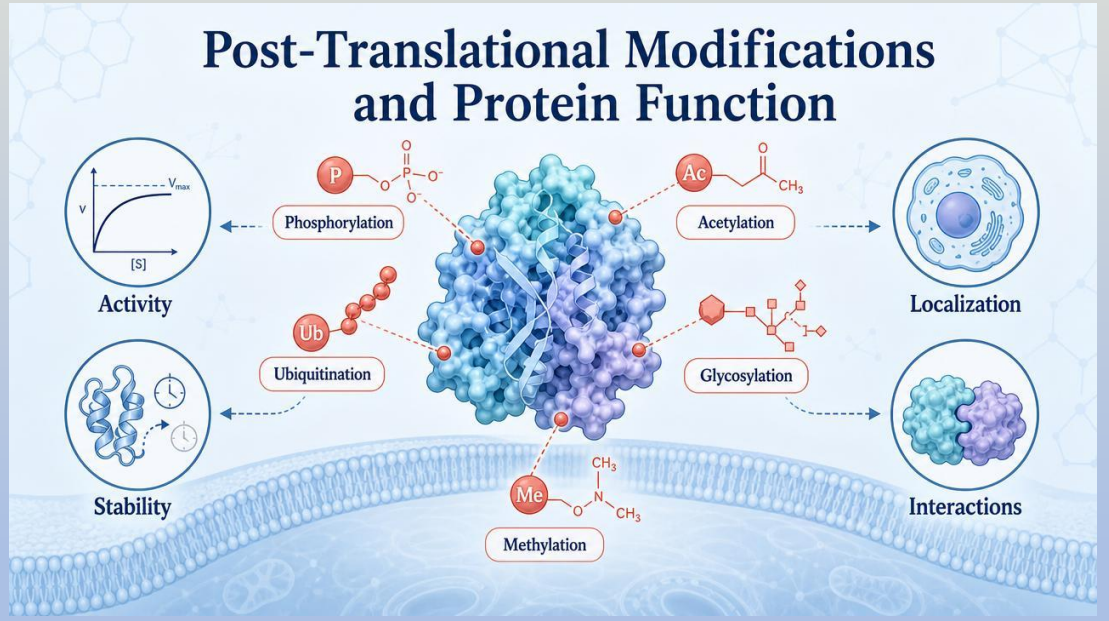
Key concept: PTMs greatly expand proteome diversity and provide an important regulatory layer between protein abundance and biological function.

Major types of PTMs:

Modification	Molecular mechanism	Biological function
Phosphorylation	Addition of phosphate groups, mainly to Ser, Thr, or Tyr residues	Signal transduction and enzyme regulation
Glycosylation	Attachment of carbohydrate groups	Protein folding, stability, and recognition
Ubiquitination	Covalent attachment of ubiquitin	Protein degradation, trafficking, and signaling
Acetylation	Addition of acetyl groups, often to Lys residues or protein N-termini	Regulation of protein interactions and stability
Methylation	Addition of methyl groups to specific amino acid residues	Protein interactions and regulatory functions
SUMOylation	Attachment of small ubiquitin-like modifier (SUMO) proteins	Stress responses, localization, and transcriptional regulation
Oxidative modifications	Oxidation of susceptible amino acid residues	Redox signaling or oxidative damage

Major types of post-translational modifications (PTMs)

post-translational modifications (PTMs)	Common biological role
Phosphorylation	Signaling and enzyme regulation
Glycosylation	Folding, stability, recognition
Ubiquitination	Protein turnover and signaling
Acetylation	Protein and chromatin regulation
Methylation	Regulation of protein interactions and chromatin
SUMOylation	Stress responses and regulatory functions



Post-translational modifications (PTM)-Focused Proteomics: Enrichment and LC–MS/MS Analysis

PTM-focused proteomics investigates post-translational modifications to determine how they regulate protein activity, molecular interactions, and cellular responses.

Major analytical approaches:

- Selective enrichment: Modified peptides are enriched before LC–MS/MS to improve their detection and identification.
- Phosphoproteomics: Immobilized metal affinity chromatography (IMAC) and titanium dioxide (TiO_2) are commonly used to enrich phosphorylated peptides.
- Other PTMs: Specific enrichment strategies are applied to glycosylated, ubiquitinated, and acetylated peptides.
- LC–MS/MS analysis: Enables identification of modified peptides and characterization of their modification sites.
- Quantitative analysis: Label-free or isotope-labeling approaches measure changes in PTM abundance between experimental conditions.

Main objectives:

1. Identify proteins and peptides carrying specific PTMs.
2. Localize modification sites within protein sequences.
3. Quantify changes in modification abundance.
4. Investigate signaling pathways and regulatory mechanisms.

PTM-focused proteomics workflow

Protein extraction and enzymatic digestion

↓

Selective enrichment of modified peptides

↓

LC–MS/MS analysis

↓

PTM identification and site localization

↓

Quantitative and functional interpretation

Key concept: PTM-focused proteomics reveals regulatory changes that cannot be determined from protein abundance measurements alone.

Interactomics and protein–protein interactions

- Interactomics is the large-scale study of molecular interactions within biological systems.
- A protein–protein interaction (PPI) occurs when two or more proteins associate physically, either transiently or stably.
- PPIs contribute to:
- Protein complex formation.
- Signal transduction.
- Metabolic regulation.
- Transcriptional regulation.
- Cellular defense mechanisms.

Technologies for Investigating Protein–Protein Interactions (PPIs)

- Protein–protein interactions (PPIs) are physical associations between proteins that regulate cellular signaling, metabolic pathways, protein complex formation, and physiological responses.

Applications in biotechnology:

- Identification of protein complexes and signaling networks.
- Investigation of enzyme regulation and metabolic pathways.
- Characterization of plant stress-response mechanisms.
- Functional analysis of candidate genes and proteins.
- *Key concept: Combining complementary PPI technologies improves confidence in identifying and characterizing protein interaction networks.*

Major technologies:

Technology	Principle and application
Yeast two-hybrid (Y2H)	Detects binary protein interactions using reporter gene activation
Co-immunoprecipitation (Co-IP)	Identifies protein associations within cellular complexes
Affinity purification–mass spectrometry (AP–MS)	Identifies proteins associated with a selected bait protein
Proximity labeling (BioID, TurboID)	Labels proteins located near a target protein in living cells
Förster resonance energy transfer (FRET)	Detects close molecular proximity using fluorescence energy transfer

Technologies for investigating protein–protein interactions (PPIs)

Technology	Principle
Yeast two-hybrid (Y2H)	Detects interactions through reporter-gene activation
Co-immunoprecipitation (Co-IP)	Isolates proteins associated with a target protein
Affinity purification–MS (AP-MS)	Identifies proteins associated with an affinity-purified bait
Proximity labeling	Labels proteins near a target protein
FRET/BiFC	Investigates molecular proximity or association in cells

Future Perspectives in Proteomics and Protein Interaction Technologies

Advances in proteomics are transforming the analysis of protein abundance, modifications, structure, and molecular interactions.

Emerging research directions:

- Single-cell proteomics: Characterization of protein expression at cellular resolution.
- Spatial proteomics: Investigation of protein localization within tissues and cellular compartments.
- Advanced mass spectrometry: Improved sensitivity, throughput, and quantitative reproducibility.
- Structural proteomics: Integration of experimental structural methods and AI-based protein structure prediction.
- Multi-omics integration: Combining proteomics with genomics, transcriptomics, metabolomics, and phenomics.
- Interactomics: Large-scale mapping of protein interaction networks.
- Precision biotechnology: Identification of proteins and regulatory networks associated with desirable biological traits.

Challenges:

- Wide dynamic range of protein abundance.
- Detection of low-abundance proteins and transient interactions.
- Complexity of post-translational modifications.
- Reproducibility, data standardization, and functional validation.

Key concept: Integrating quantitative, structural, and interaction proteomics will advance systems biology and biotechnology.

Conclusion

- Proteomics investigates the composition, abundance, modifications, localization, and functions of proteins.
- Protein diversity arises from alternative RNA processing, translation, and post-translational modifications.
- Chromatographic separation and mass spectrometry are fundamental technologies for protein identification and quantification.
- PTM-focused proteomics provides insight into protein regulation and cellular signaling.
- Protein interaction technologies reveal molecular associations and functional protein networks.
- Bioinformatics and multi-omics integration support the interpretation of complex proteomic datasets.
- Emerging technologies, including single-cell proteomics and AI-assisted structural analysis, expand the possibilities of functional proteomics.

Overall conclusion: Proteomics and protein interaction technologies connect molecular information with cellular function and provide essential tools for understanding biological processes and developing modern biotechnological applications.

Control Questions

1. What is proteomics, and how does it differ from genomics and transcriptomics?
2. Why is the proteome more complex and dynamic than the genome?
3. What are the main stages of a typical proteomic analysis workflow?
4. What chromatographic methods are used for separating proteins and peptides?
5. What are the principles of mass spectrometry and tandem mass spectrometry (MS/MS)?
6. What are the major types and biological functions of post-translational modifications?
7. Why is selective peptide enrichment important in PTM-focused proteomics?
8. What technologies are used to investigate protein–protein interactions?
9. How do Y2H, Co-IP, AP–MS, and proximity labeling differ?
10. What are the major applications and future perspectives of proteomics in plant biotechnology?

Recommended Literature

- 1. Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. Nature, 537, 347–355. https://doi.org/10.1038/nature19949
- 2. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., & Yates, J. R. III. (2013). Protein analysis by shotgun/bottom-up proteomics. Chemical Reviews, 113, 2343–2394. https://doi.org/10.1021/cr3003533
- 3. O'Connell, J. D., Paulo, J. A., O'Brien, J. J., & Gygi, S. P. (2018). Proteome-wide evaluation of two common protein quantification methods. Journal of Proteome Research, 17, 1934–1942. https://doi.org/10.1021/acs.jproteome.8b00016
- 4. Gingras, A.-C., Gstaiger, M., Raught, B., & Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. Nature Reviews Molecular Cell Biology, 8, 645–654. https://doi.org/10.1038/nrm2208
- 5. Branon, T. C., et al. (2018). Efficient proximity labeling in living cells and organisms with TurboID. Nature Biotechnology, 36, 880–887. https://doi.org/10.1038/nbt.4201
- 6. Jumper, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2
- 7. Meier, F., et al. (2020). diaPASEF: Parallel accumulation–serial fragmentation combined with data-independent acquisition. Nature Methods, 17, 1229–1236. https://doi.org/10.1038/s41592-020-00998-