Lecture 6:

Proteomics and Metabolomics

From genes to function

Protein dynamics

Metabolic snapshots

Introduction to Biomedical Datascience

Lecture Contents

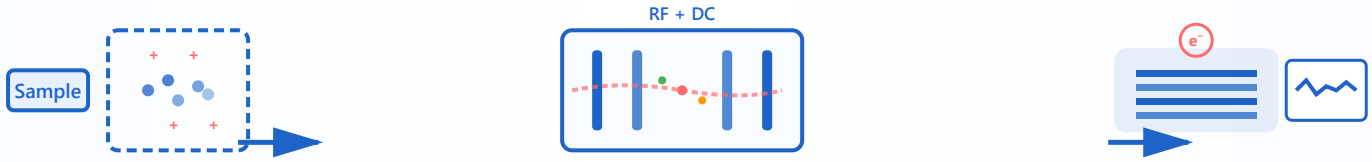
- Mass spectrometry principles
- Proteomics workflows
- Metabolomics approaches

Part 1/3:

Proteomics Technologies

- MS instrumentation
- Sample preparation
- Quantification strategies
- Data analysis

Mass Spectrometry Basics



Ion Generation
ESI / MALDI

Mass Analysis
m/z Separation

Detection
Signal Processing

Resolution & Accuracy

- High resolution: distinguish similar masses
- Mass accuracy: parts per million (ppm)
- Critical for peptide identification

Scan Modes

- Full scan: entire mass range
- Selected ion monitoring (SIM)
- Data-dependent acquisition (DDA)

Sensitivity

- Femtomole to attomole detection
- Dynamic range: 3-5 orders of magnitude
- Low abundance protein detection

Applications

- Protein identification
- Quantification
- PTM analysis

Ionization Methods

ESI (Electrospray)

Soft ionization technique for biomolecules

- Multiple charging states
- Direct coupling with LC
- Ideal for peptides and proteins

MALDI

Matrix-assisted laser desorption

- Crystallized with matrix
- Pulsed laser ionization
- High-throughput screening

Nano-ESI

Enhanced sensitivity ESI variant

- Ultra-low flow rates (nL/min)
- 10-100× more sensitive
- Limited sample volumes

APCI

Atmospheric pressure chemical ionization

- Small molecule focus
- Less polar compounds
- Complementary to ESI

Method Selection

Choose based on: analyte properties, sample complexity, throughput requirements, and sensitivity needs

Mass Analyzers

Quadrupole Filters

- Four parallel rods with RF/DC voltages
- Sequential ion transmission
- Good for targeted analysis

Time-of-Flight (TOF)

- Velocity-based separation
- High mass accuracy
- Unlimited mass range

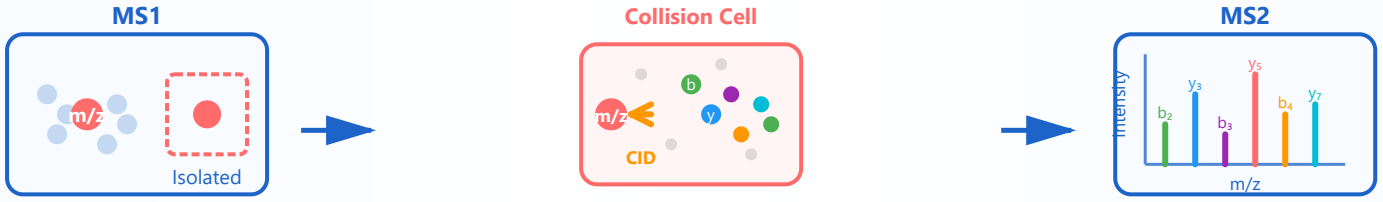
Orbitrap Technology

- Ion orbital trapping
- Ultra-high resolution (>100,000)
- Excellent mass accuracy (<1 ppm)

Ion Trap & Hybrid

- 3D ion confinement
- Multiple MS/MS stages
- Q-TOF, Q-Orbitrap combinations

Tandem MS (MS-MS)



MS1: Select Precursor
Ion Isolation

Fragmentation
CID / HCD / ETD

MS2: Analyze Fragments
Sequence Info

Precursor Selection

- Isolate specific m/z ions
- Top-N data-dependent selection
- Targeted precursor lists

Fragmentation Methods

- CID: collision-induced dissociation
- HCD: higher-energy collisional dissociation
- ETD: electron-transfer dissociation

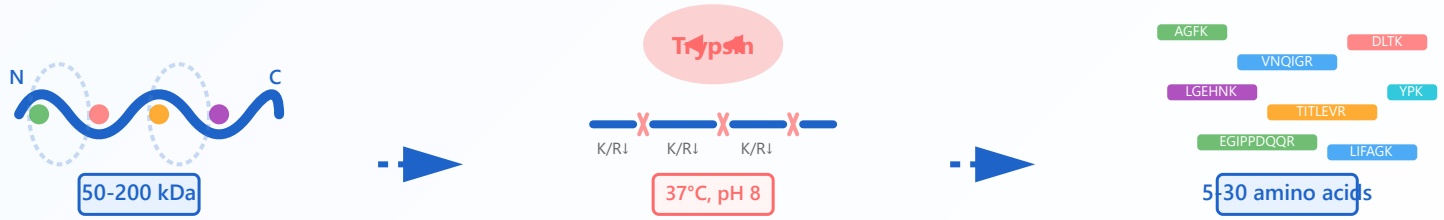
Product Ion Spectra

- b-ions and y-ions from peptides
- Sequence information
- PTM localization

Data Acquisition

- DDA: data-dependent acquisition
- DIA: data-independent acquisition
- Parallel reaction monitoring (PRM)

Bottom-up Proteomics



Intact Protein
Full-length sequence

Enzymatic Digestion
Specific cleavage

Peptide Pool
MS-compatible size

Protein Digestion

- Enzymatic cleavage into peptides
- 5-30 amino acid peptides
- Most common workflow

Trypsin Specificity

- Cleaves after K and R residues
- Predictable peptide generation
- Optimal MS-friendly peptides

Peptide Separation

- Reverse-phase liquid chromatography
- Gradient elution
- Online LC-MS coupling

Data Complexity

- Thousands of peptides
- Multiple charge states
- Requires computational analysis

Top-down Proteomics

Intact Protein

Complete sequence + PTMs

VS

Digested Peptides

Fragmented before analysis

Intact Protein Analysis

- No digestion required
- Analyze whole proteins
- 10-80 kDa typical range

PTM Preservation

- Complete modification pattern
- Combinatorial PTM analysis
- Proteoform characterization

Technical Challenges

- Requires high resolution
- Complex spectra interpretation
- Lower sensitivity than bottom-up

Native MS

- Preserve non-covalent interactions
- Protein complexes
- Quaternary structure information

Quantitative Proteomics

Spectral Counting | Metabolic Labeling | Isobaric Tags | Data Strategies

Label-Free Quantification

- Spectral counting
- Peak intensity measurement
- No labeling required

SILAC Labeling

- Metabolic incorporation
- Heavy amino acids (13C, 15N)
- Cell culture applications

TMT/iTRAQ Tags

- Isobaric mass tags
- Multiplexing 6-16 samples
- Reporter ion quantification

DIA vs DDA

- DIA: all ions fragmented
- DDA: targeted selection
- Trade-offs in coverage and reproducibility

Part 2/3:

Protein Analysis

- Identification algorithms
- PTM characterization
- Interaction studies
- Structural insights

Protein Identification

Database Searching

- Match spectra to sequence databases
- Multiple search engines available
- Statistical scoring algorithms

Peptide-Spectrum Matching

- Compare experimental vs theoretical
- Fragment ion matching
- Mass accuracy requirements

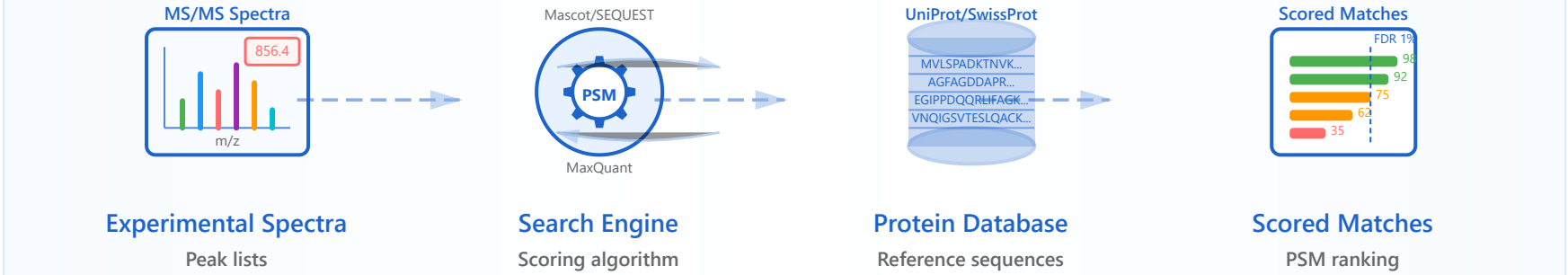
Score Calculations

- Probability-based scoring
- Expectation values (E-values)
- Confidence metrics

FDR Estimation

- False discovery rate control
- Target-decoy approach
- Typically 1-5% FDR threshold

Database Searching



Experimental Spectra
Peak lists

Search Engine
Scoring algorithm

Protein Database
Reference sequences

Scored Matches
PSM ranking

Search Engines

- Mascot, SEQUEST, X!Tandem
- MaxQuant, Proteome Discoverer
- Each with unique algorithms

Parameter Optimization

- Mass tolerance settings
- Enzyme specificity
- Missed cleavages allowed

Decoy Databases

- Reversed/shuffled sequences
- Estimate false positives
- Quality control

Modifications

- Fixed modifications (e.g., carbamidomethylation)
- Variable modifications (e.g., oxidation)
- Balance between sensitivity and specificity

PTM Analysis

Phosphorylation Sites

- Ser/Thr/Tyr phosphorylation
- Enrichment with TiO2/IMAC
- Site localization algorithms

Glycosylation Patterns

- N-linked and O-linked glycans
- Heterogeneous modifications
- Deglycosylation strategies

Acetylation/Methylation

- Lysine modifications
- Histone PTMs
- Epigenetic regulation

Enrichment Methods

- Immunoprecipitation
- Affinity chromatography
- Chemical derivatization

Protein-Protein Interactions

AP-MS Workflows

- Affinity purification-mass spec
- Pull-down interacting partners
- Identify protein complexes

Proximity Labeling

- BioID, APEX, TurboID
- Spatial proteomics
- In vivo labeling

Cross-linking MS

- Chemical cross-linkers
- Distance constraints
- Protein structure information

Network Construction

- Interaction databases
- Scoring significance
- Pathway analysis

Structural Proteomics

HDX-MS Principles

- Hydrogen-deuterium exchange
- Protein dynamics
- Conformational changes

Cross-linking Constraints

- Distance measurements
- Protein topology
- Complex architecture

Limited Proteolysis

- Protease accessibility
- Structural domains
- Folding states

Ion Mobility

- Gas-phase separation
- Collision cross-section
- Shape information

Clinical Proteomics

Biomarker Discovery

- Disease-specific proteins
- Early detection markers
- Prognostic indicators

Plasma Proteomics

- High dynamic range challenge
- Depletion strategies
- Abundant protein removal

Tissue Proteomics

- FFPE sample analysis
- Spatial proteomics
- Disease pathology

FDA-Approved Tests

- MALDI-TOF bacterial ID
- Targeted protein panels
- Clinical validation requirements

Part 3/3:

Metabolomics

- Small molecules analysis
- Pathway mapping
- Clinical applications

Metabolomics Overview

Targeted vs Untargeted

- Targeted: quantify specific metabolites
- Untargeted: broad metabolite profiling
- Semi-targeted approaches

Primary Metabolites

- Central metabolism (glycolysis, TCA)
- Amino acids, nucleotides
- Energy production molecules

Secondary Metabolites

- Plant natural products
- Signaling molecules
- Defense compounds

Metabolic Flux

- Dynamic metabolite changes
- Isotope tracing (13C, 15N)
- Pathway activity measurement

Sample Preparation

Quenching Metabolism

- Rapid cooling or organic solvents
- Stop enzymatic reactions
- Preserve metabolite levels

Extraction Methods

- Methanol/chloroform extraction
- Solid-phase extraction (SPE)
- Method depends on metabolite class

Matrix Effects

- Ion suppression/enhancement
- Sample cleanup required
- Calibration curve considerations

Internal Standards

- Isotope-labeled compounds
- Normalize for extraction/ionization
- Quality control

LC-MS Methods

Column Chemistry

- Reverse-phase C18 (non-polar)
- HILIC (polar metabolites)
- Mixed-mode columns

Gradient Optimization

- Mobile phase composition
- Flow rate selection
- Peak resolution vs run time

Ion Suppression

- Co-eluting compounds interfere
- Matrix effects
- Mitigated by cleanup and separation

Method Validation

- Linearity, accuracy, precision
- Lower limit of quantification
- Stability testing

GC-MS Methods

Derivatization

- Make metabolites volatile
- Silylation, acetylation
- Improve chromatography

Volatility Requirements

- Low molecular weight compounds
- Thermal stability needed
- Complementary to LC-MS

EI Fragmentation

- Electron ionization
- Reproducible fragmentation
- Library matching possible

Retention Indices

- Normalize retention times
- n-alkane standards
- Cross-lab comparisons

NMR Metabolomics

1H NMR Profiling

- Non-destructive analysis
- All proton-containing metabolites
- Quantitative without standards

2D NMR Experiments

- COSY, TOCSY, HSQC
- Enhanced resolution
- Structure elucidation

Quantification

- Direct concentration measurement
- Internal standard (TSP, DSS)
- No ionization bias

Sample Requirements

- Larger sample volumes than MS
- Buffer composition matters
- Lower sensitivity

Metabolite Identification

Mass Accuracy

- Sub-5 ppm for confident ID
- High-resolution mass spec
- Elemental formula prediction

Isotope Patterns

- Natural isotope distribution
- Confirm molecular formula
- Chlorine/bromine signatures

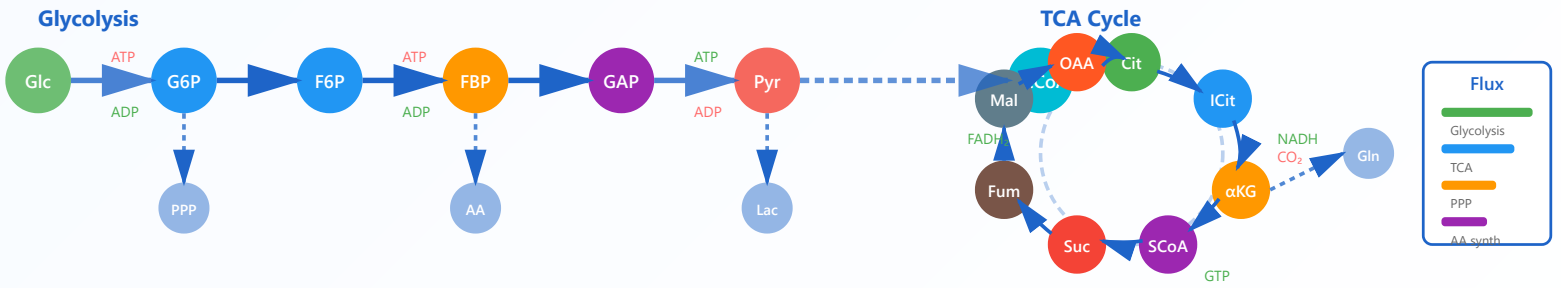
MS/MS Matching

- Fragment ion patterns
- Spectral library search
- In-silico fragmentation

Standards Confirmation

- Authentic chemical standards
- Retention time matching
- Gold standard for identification

Pathway Mapping



KEGG Pathways

- Kyoto Encyclopedia database
- Metabolic pathway maps
- Organism-specific pathways

Metabolic Networks

- Biochemical connections
- Reaction stoichiometry
- Flux balance analysis

Flux Analysis

- 13C glucose/glutamine tracing
- MID (mass isotopomer distribution)
- Pathway activity quantification

Integration Tools

- MetaboAnalyst, XCMS
- Pathway enrichment analysis
- Multi-omics integration

Biomarker Discovery

Study Design

- Case-control studies
- Adequate sample size
- Biological replicates

Statistical Analysis

- Univariate tests (t-test, ANOVA)
- Multivariate (PCA, PLS-DA)
- Multiple testing correction

Validation Cohorts

- Independent sample sets
- Different populations
- Avoid overfitting

ROC Analysis

- Receiver operating characteristic
- AUC (area under curve)
- Diagnostic performance evaluation

Lipidomics

Lipid Classes

- Glycerophospholipids, sphingolipids
- Triacylglycerols, cholesterol esters
- 1000s of lipid species

Extraction Protocols

- Bligh-Dyer, Folch methods
- Biphasic extraction
- Lipid class-specific protocols

Separation Strategies

- Direct infusion (shotgun lipidomics)
- LC-MS with C8/C18 columns
- Supercritical fluid chromatography

Nomenclature

- Lipid MAPS classification
- Fatty acid composition notation
- Standardized reporting

Hands-on MaxQuant Analysis

Raw File Processing

- Load RAW files from mass spec
- Automatic peak detection
- Retention time alignment

Parameter Settings

- Enzyme: Trypsin/P
- Fixed/variable modifications
- FDR thresholds (1% peptide/protein)

Perseus Downstream

- Statistical analysis platform
- Filtering and normalization
- Differential expression analysis

Quality Assessment

- Check identification rates
- Review mass error distributions
- Evaluate quantification reproducibility

Hands-on MetaboAnalyst

Data Upload

- Peak intensity table
- Sample groups defined
- Metabolite IDs (HMDB, KEGG)

Normalization

- Sample-specific normalization
- Log transformation
- Scaling methods (auto, pareto)

Statistical Analysis

- t-tests, ANOVA
- PCA, PLS-DA
- Volcano plots, heatmaps

Pathway Analysis

- Enrichment analysis
- Topology analysis
- Visual pathway maps

Thank You!

- Multi-omics future
- Precision medicine
- Technology advances
- Career paths

Introduction to Biomedical Datascience