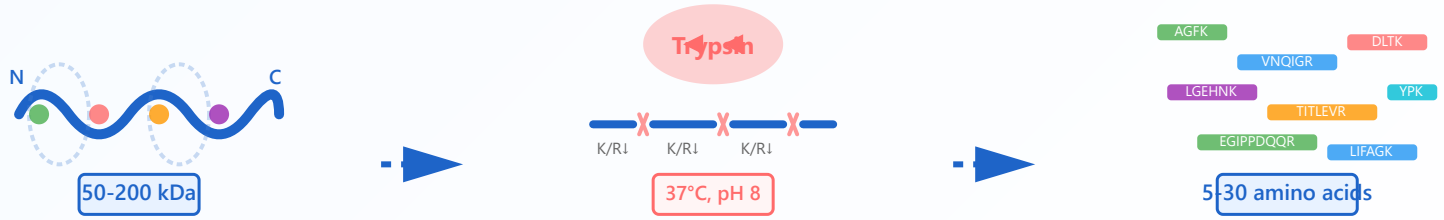
Bottom-up Proteomics



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

1. Protein Digestion: Breaking Down the Target

1.1 Overview and Purpose

Protein digestion is the critical first step in bottom-up proteomics where intact proteins

are enzymatically cleaved into smaller peptide fragments. This process transforms complex, large proteins into manageable peptides that are compatible with mass spectrometry analysis.

The goal is to generate peptides with optimal characteristics for mass spectrometry: typically 5-30 amino acids in length, containing appropriate charge states, and with predictable fragmentation patterns.

Why Digestion is Necessary: Intact proteins are too large and complex for efficient ionization and fragmentation in most mass spectrometers. Peptides provide better sensitivity, more consistent ionization, and interpretable MS/MS spectra.

1.2 Digestion Process

Enzymatic Digestion Workflow

K↓ R↓

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

1000s

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

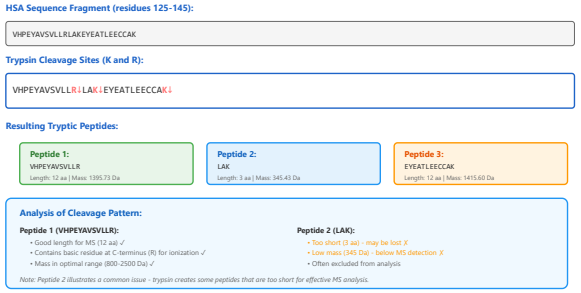
Example: Complete Digestion Protocol

Step-by-Step Trypsin Digestion Protocol:

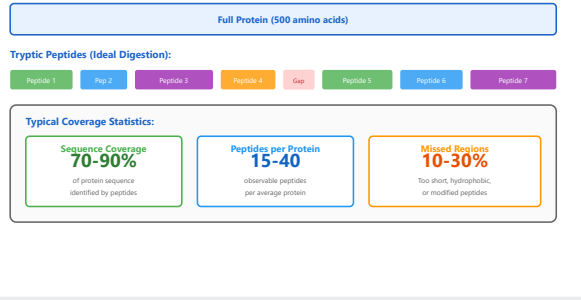
2. Trypsin Specificity: Detailed Analysis

2.1 Real Protein Example

Let's examine how trypsin cleaves a real protein: Human Serum Albumin (HSA), one of the most abundant blood proteins.



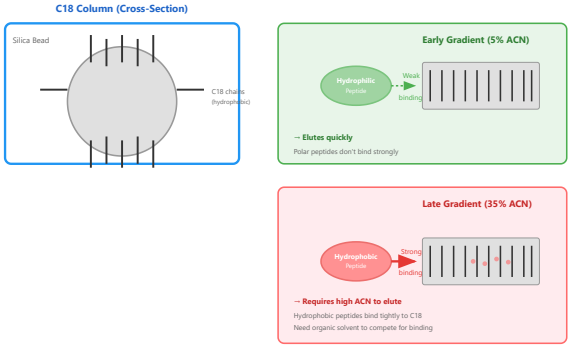
2.2 Peptide Coverage Map



3. Peptide Separation: Chromatography in Detail

3.1 Reverse-Phase Mechanism

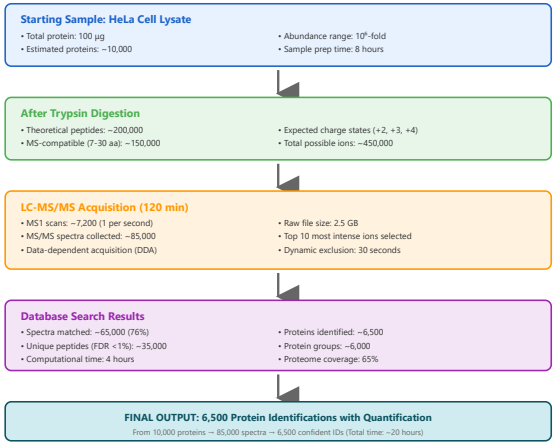
Molecular-Level View of Peptide Separation



4. Data Complexity: Real-World Example

4.1 Case Study: HeLa Cell Proteome Analysis

From Sample to Identification: Numbers



Key Insight: Only about 65% of the proteome is typically identified in a single LC-MS run, even with modern instruments. The missing 35% includes low-abundance proteins, membrane proteins, very large/small proteins, and proteins with unfavorable chemical properties.

Fractionation or enrichment strategies are needed for deeper coverage.