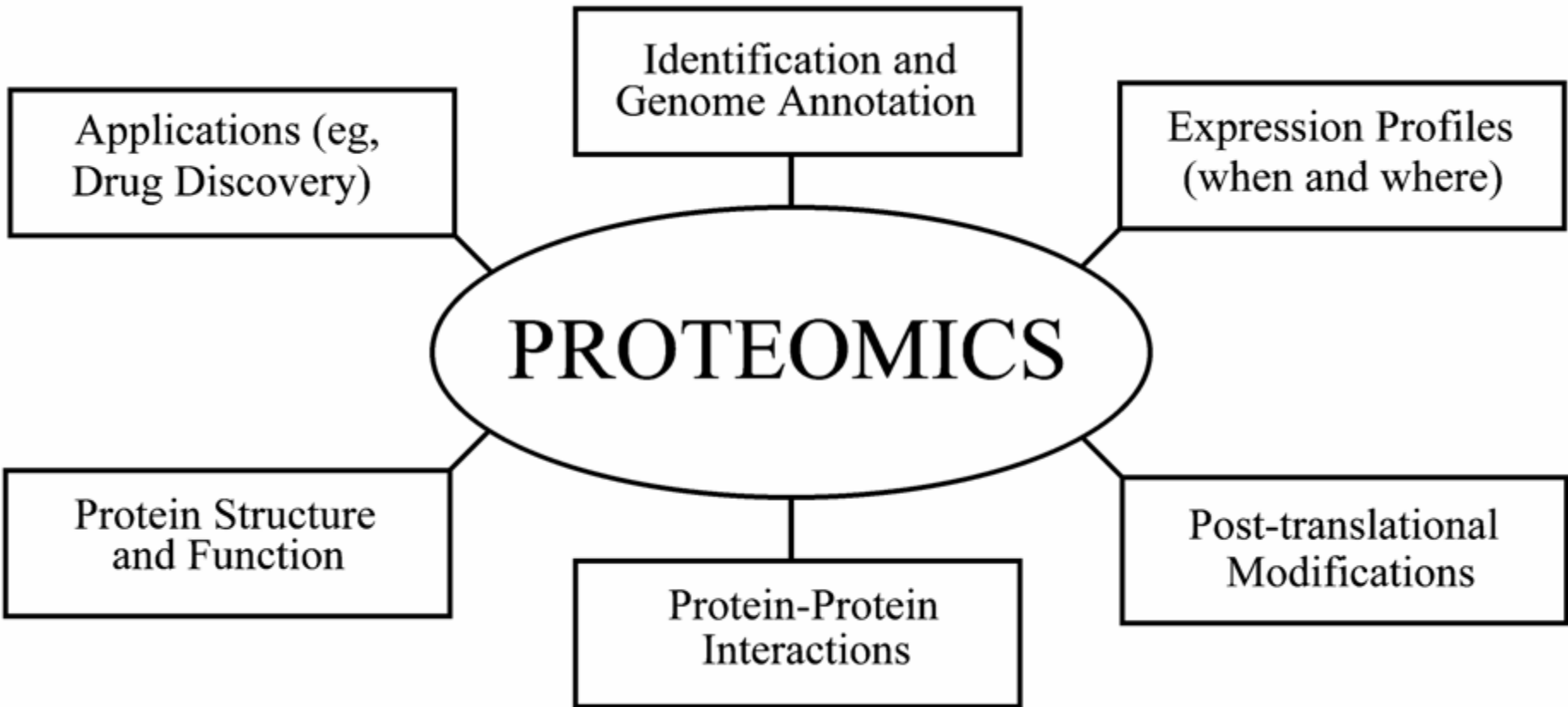


Genomes and Proteomes

- **genome: complete set of genetic information in organism**
 - **gene sequence contains recipe for making proteins (genotype)**
- **proteome: complete set of proteins in cell, tissue, organism, etc.**
 - **much of the information about proteins is not in the genome (phenotype)**



Genomics ↔ Proteomics

- **identify total # of genes and 'functional annotations'**
 - accuracy of exon/intron structure predictions?
 - matching proteins to genes
- **expression profiles (mRNA/protein)**
 - cell/tissue specificity
 - regulation or 'environmental' influences
 - subcellular compartments
- **one gene → more than one protein**
 - alternate splicing
 - post-translational modifications

Initial Identification and Characterization

- **comparison already known (or related) protein and mRNA sequences**
 - prior or currently generated information used to annotate genome databases
 - 30-50% genes unknown
 - among the known sequences most not completely characterized
- **high throughput mRNA methods**
 - expressed sequence tags (ESTs)
 - microarrays (gene chips)

Protein Identification

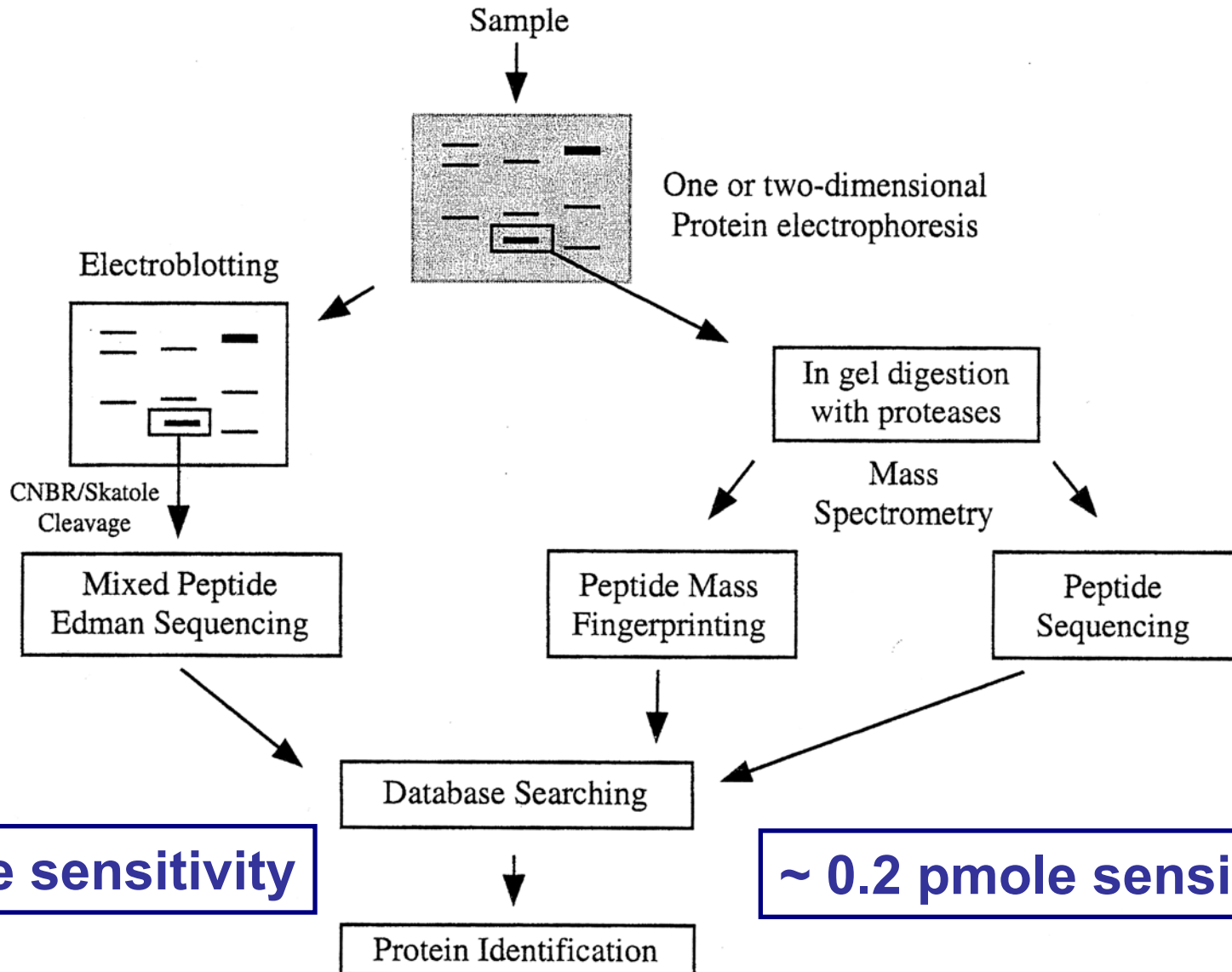
- **proteins more difficult to analyze by 'high throughput methods'**
- **gel electrophoresis (1-D or 2-D) is predominant technique**
- **partial microsequence → protein identification**
 - **N-terminal or internal peptide sequences**
 - **search databases**
- **problems with blocked N-termini or with ability to isolate peptides (abundance)**

Mixed Peptide Sequencing

- **typical microsequencing protocol**
 - gel electrophoresis
 - transfer to membrane
 - excise band
- **treat with CNBr (cleaves at Met), etc.**
 - generates 3-5 fragments
- **subject to 6-12 automated Edman cycles**
- **use computer program to analyze mixed peptide sequences against databases**

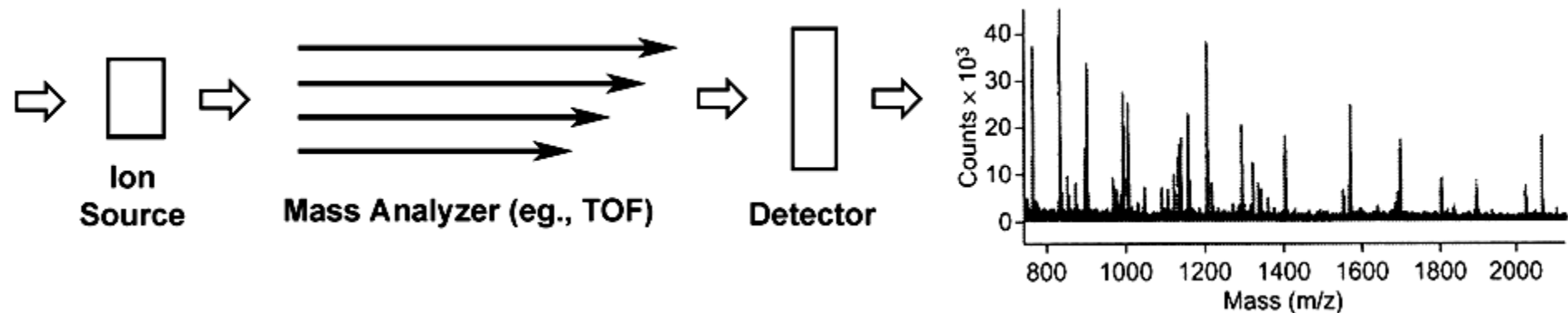
0	M	M	M
1	D	G	E
2	S	Q	V
3	D	T	A
4	A	V	K
5	D	Q	R
6	A	F	D

Strategies for Protein Identification



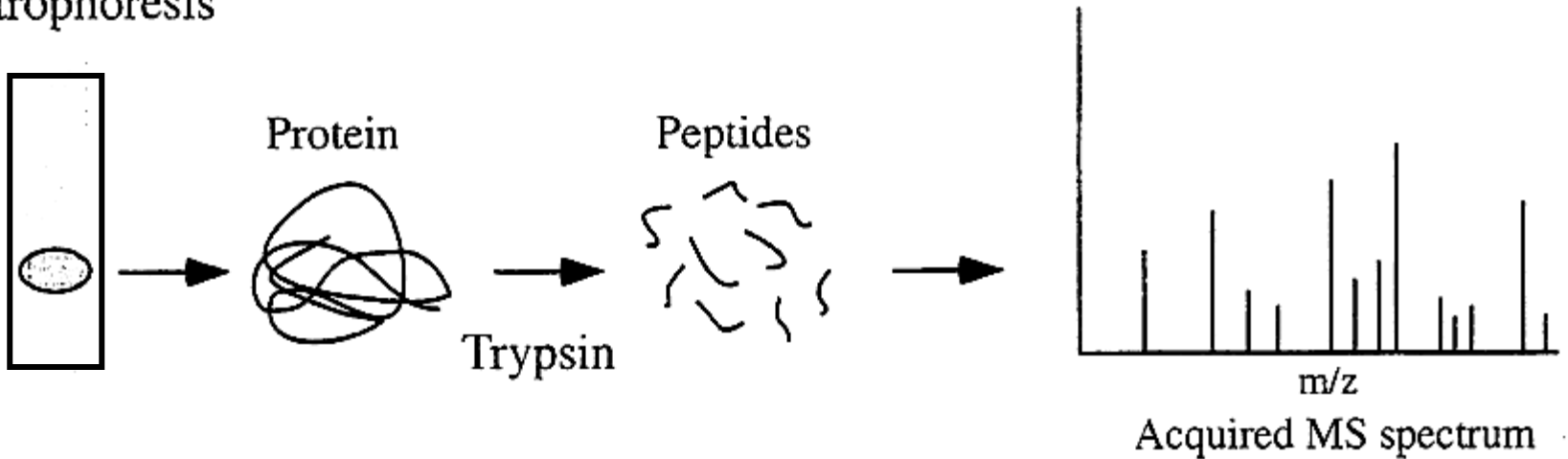
Mass Spectrometer

- **3 principal components**
 - ionization source
 - mass analyzer
 - detector
- **several different types depending on types of ionization source and mass analyzer**
- **accurately measures masses of components in sample and records a mass spectrum**



Peptide Mass Fingerprinting

A. Electrophoresis



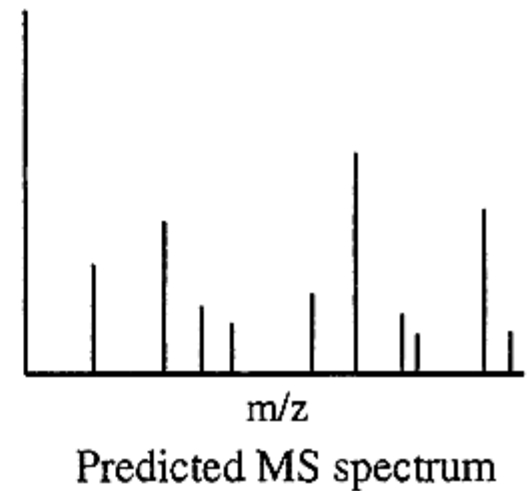
B. MAAVFLTGNWPIHGGC
GICK**GLYSTTVFLAKQ**
HK**MNPTYNQFR**MHSNL
CAHPFTRL**LVSDEGDKC**
GILNFPPS

Protein in
database

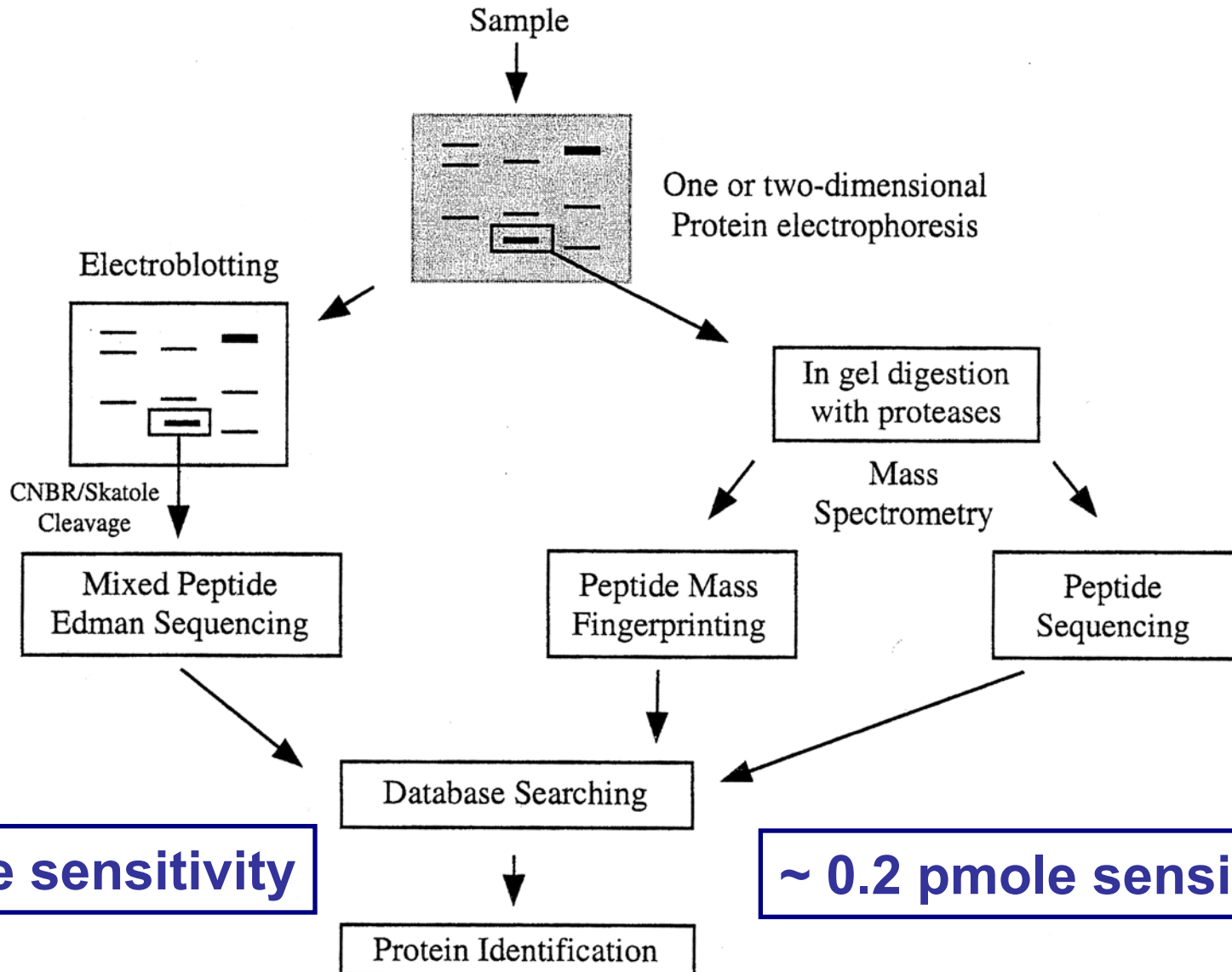


GLYSTTVFLAK
MNPTYNQFR
LVSDEGDK

Predicted peptides
from hypothetical
trypsin treatment

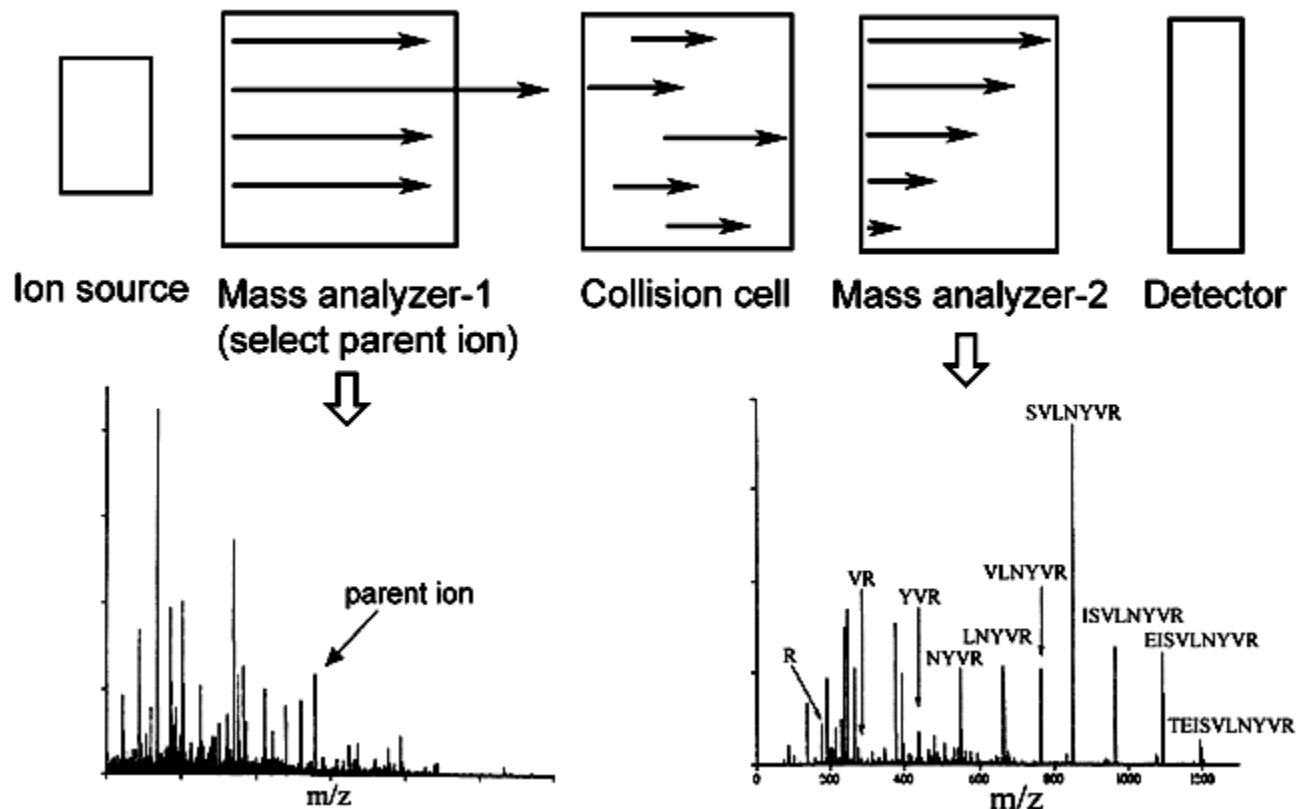


Strategies for Protein Identification



Tandem Mass Spectrometry

- selected peptides fragmented in collision chamber
- resulting spectrum used to deduce the amino acid sequence



Limitations of Gel Electrophoresis for Proteomic Analyses

- **works best on relatively abundant soluble proteins in the low-mid MW range**
- **rather laborious**
- **one protein at a time in complex mixtures**
- **difficult to automate**

High Throughput LC/MS/MS

- **complex protein sample is treated with site-specific protease**
- **subjected to liquid chromatography (HPLC)**
- **peaks automatically subjected to MS/MS**
- **database search with peptide sequences**
 - **identification of genes**
 - **protein expression profiles (~EST data)**
- **restricted proteomes**
 - **subcellular fractions**
 - **protein-protein interactions**
 - **multi-protein complexes**