Principles and Applications of Proteomics

Overview

- Why Proteomics?
- 2-DE
  - Sample preparation
  - 1st & 2nd dimension seperation
  - Data Analysis
  - Sample preparation for Mass Spectrometry
- Mass Spectrometry
  - MALDI-TOF, TANDEM MS
  - Identification of MS spectra
- Applications
  - ICAT, Phosphoproteomics, etc.

Roles of Proteins

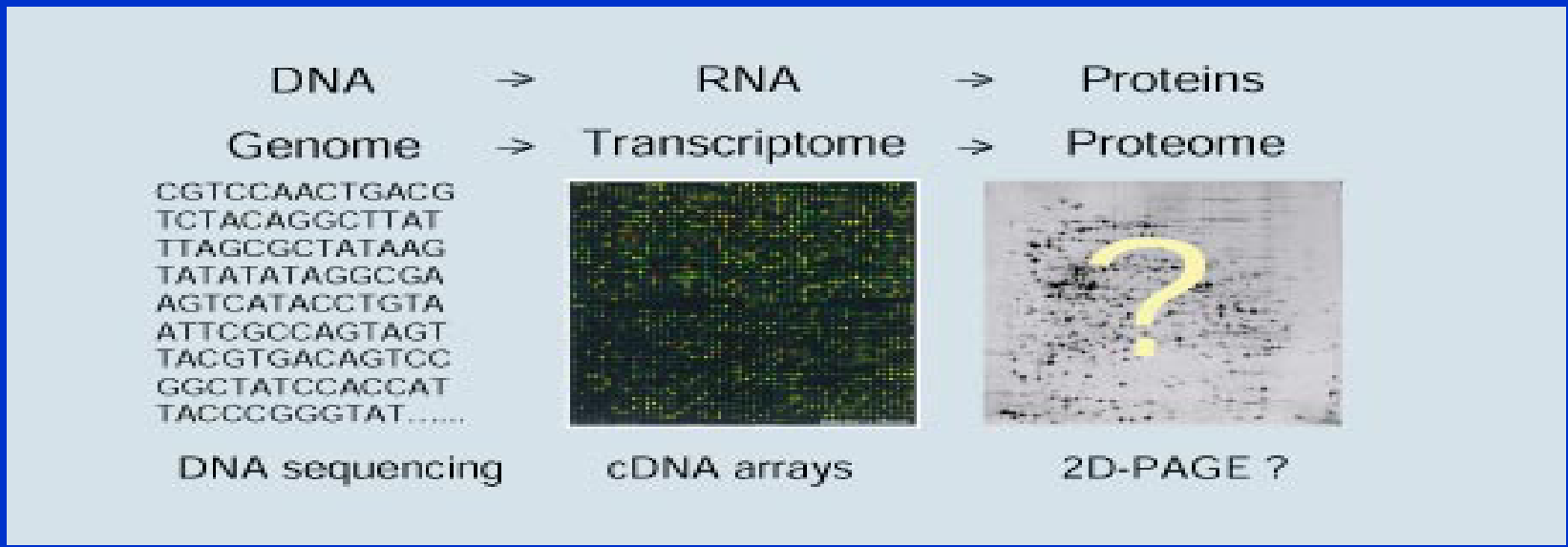
- Proteins are the instruments through which the genetic potential of an organism are expressed = active biological agents in cells

- Proteins are involved in almost all cellular processes and fulfill many functions

- Some functions of Proteins
  - enzyme catalysis, transport, mechanical support, organelle constituents, storage reserves, metabolic control, protection mechanisms, toxins, and osmotic pressure

The Virtue of the Proteome

- Proteome = protein compliment of the genome
- Proteomics = study of the proteome
- Protein world = study of less abundant proteins
- Transcriptomics – often insufficient to study functional aspects of genomics



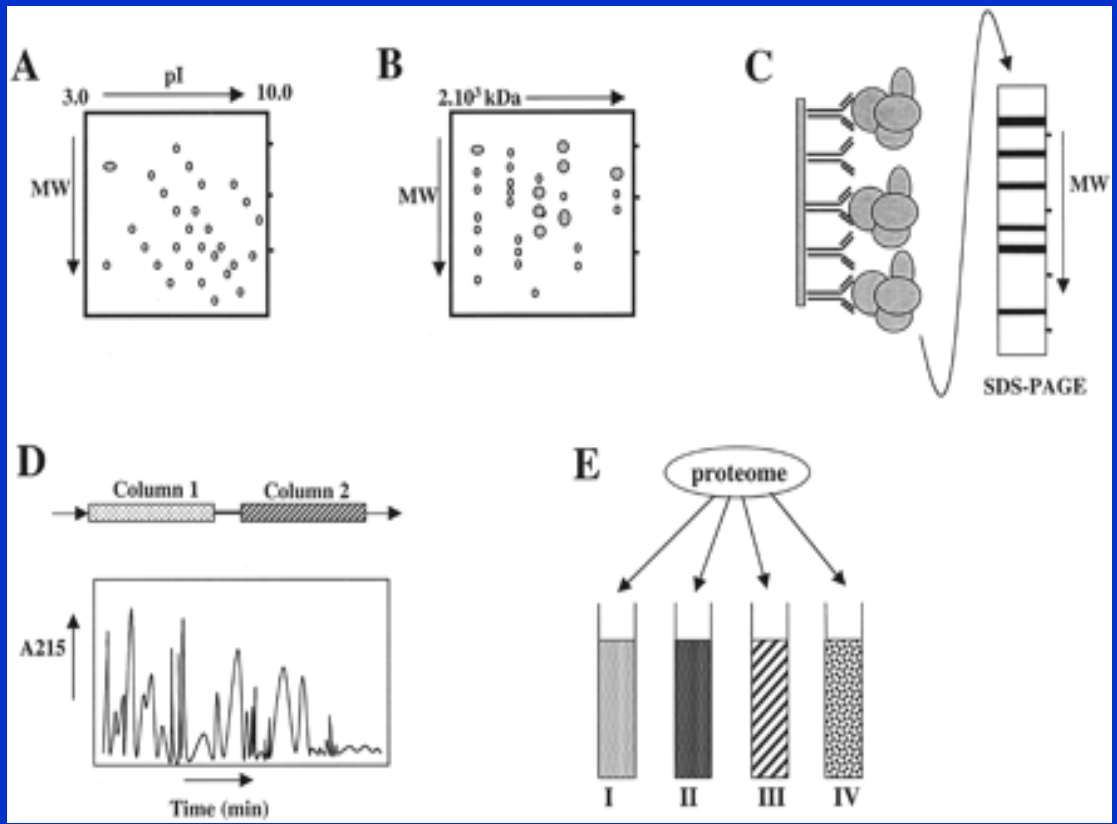
Why Proteomics?

- Whole Genome Sequence – complete, but does not show how proteins function or biological processes occur
- Post-translational modification – proteins sometimes chemically modified or regulated after synthesis
- Proteins fold into specific 3-D structures which determine function
- Gain insight into alternative splicing
- Aids in genome annotation

Some Covalent Post-Translational Modifications

Modification	Residues	Role
Cleavage	Various	Activation of proenzymes and precursors
Glycosylation	Asn,Ser,Thr	Molecular targeting, cell-cell recognition etc
Phosphorylation	Ser,Thr,Tyr	Control metabolic processes & signalling
Hydroxylation	Pro, Lys	Increase H-bonding & glycosylation sites
Acetylation	Lys	Alter charge & weaken interactions with DNA
Methylation	Lys	Alter interactions with other molecules
Carboxylation	Glu	More negative charge, e.g. to bind Ca
Transamidation	Gln, Lys	Formation of crosslinks in fibrin

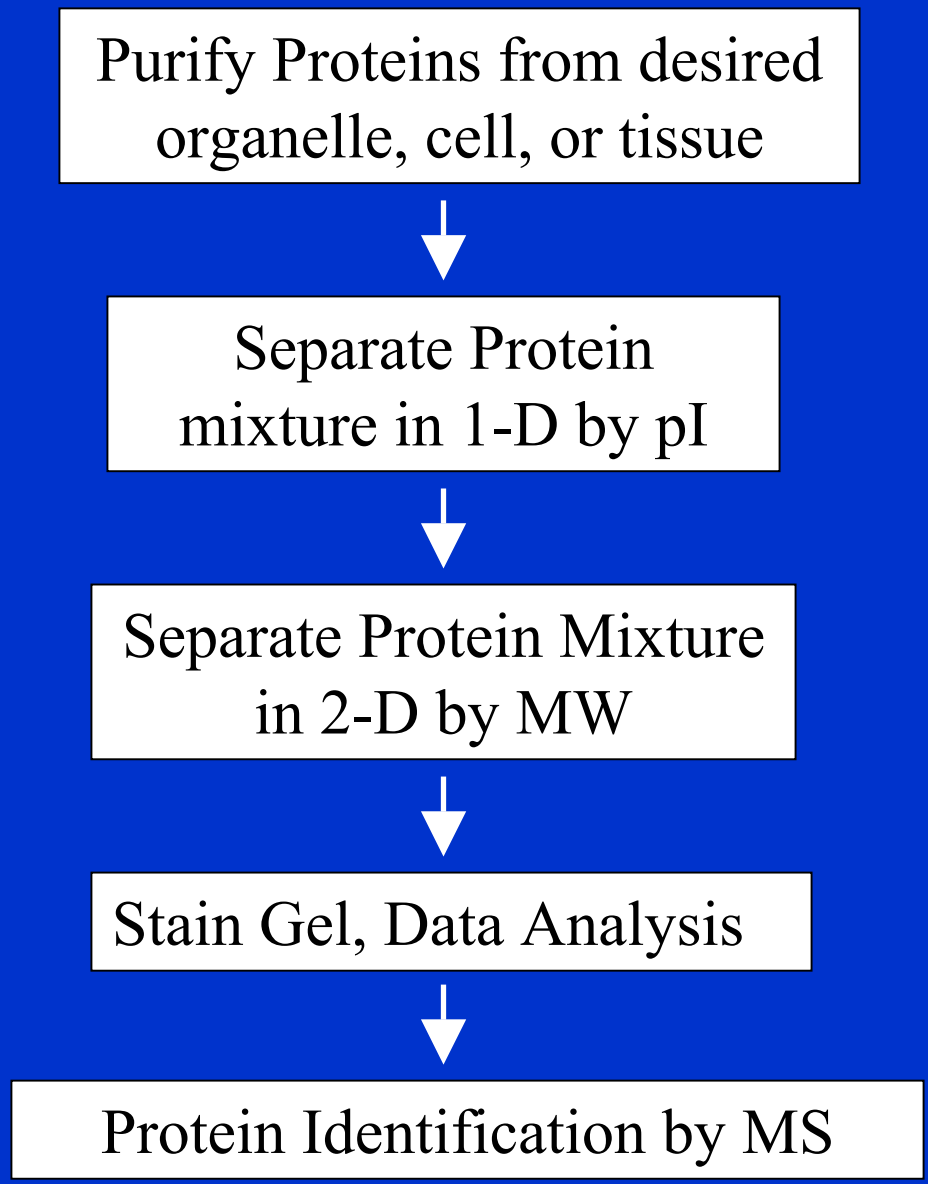
Different Approaches for Proteome Purification and Protein Separation for Identification by MS



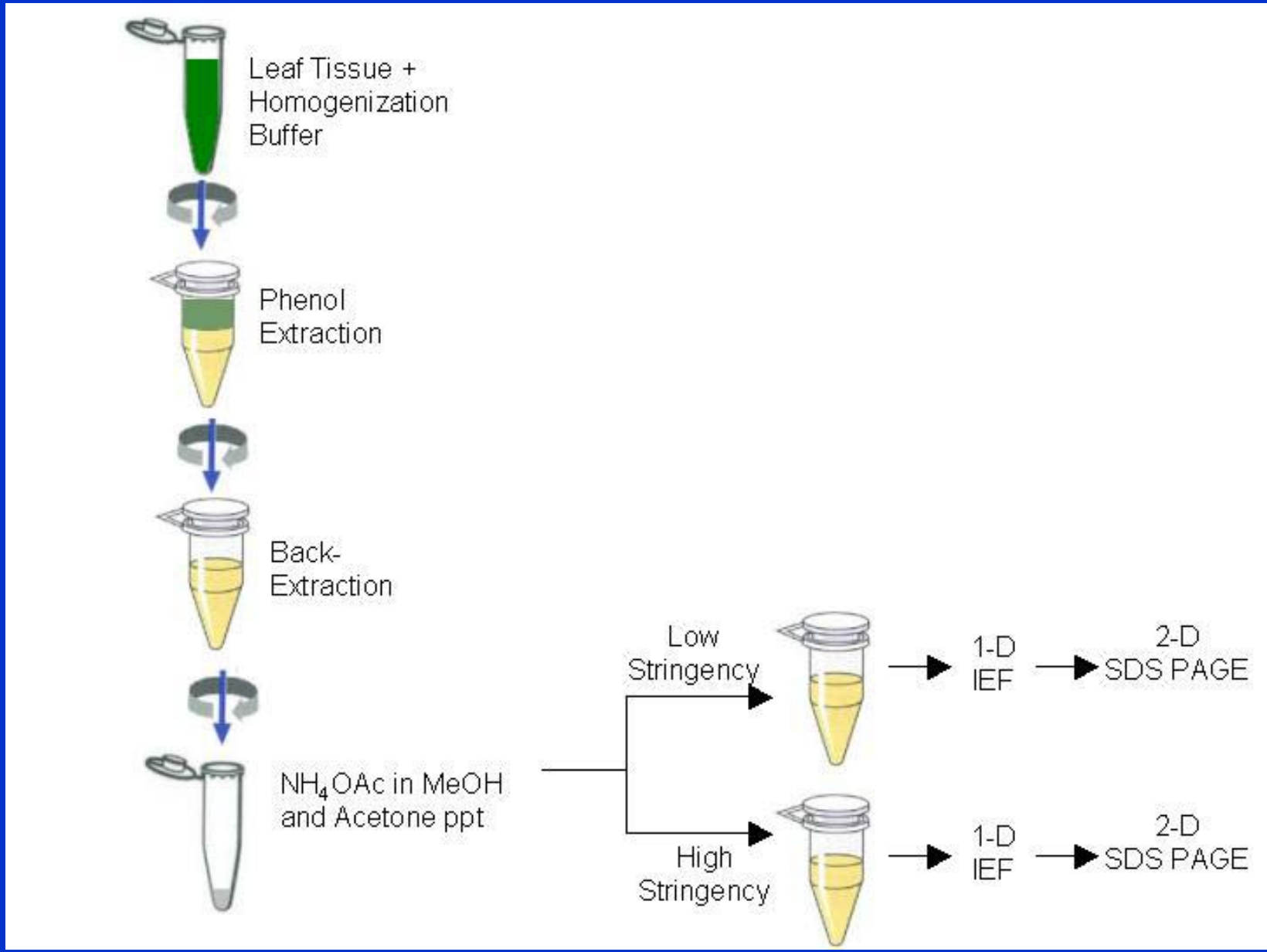
- **A. Separation of individual proteins** by 2-DE
- **B. Separation of protein complexes** by non-denaturing 2-DE
- **C. Purification of protein complexes** by affinity chromatography + SDS-PAGE
- **D. Multidimensional chromatography.**
- **E. Fractionate by Organic Solvent** – separate complex protein mix, hydrophobic membrane proteins

(van Wijk, 2001, Plant Physiology 126, 501-508)

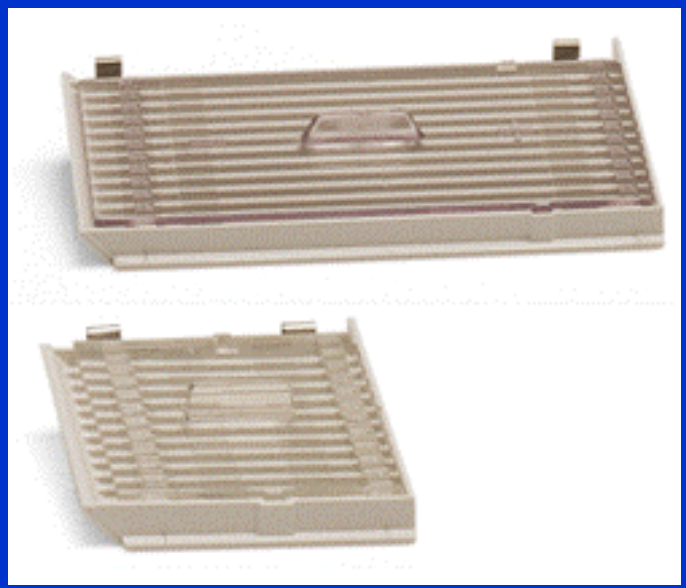
2-Dimensional Protein Electrophoresis (2-DE)



Plant Protein Extraction and Fractionation



First Dimension IEF: Immobilized pH Gradients



IPG principle:

pH gradient is generated by a number (6-8) of well-defined chemicals (immobilines) which are co-polymerized with the acrylamide matrix.

- ✓IPG allows the generation of pH gradients of any desired range between pH 3 and 12.
- ✓sample loading capacity is much higher.

The method of choice for micropreparative separation or spot identification.

Components of IEF Buffer

- Chatotropes
  - 8M Urea
  - OR…7M Urea/2M Thiourea
- Surfactants
  - 4% CHAPS
  - OR….2% CHAPS / 2% SB-14
- Reducing Agents
  - 65mM Dithioerythritol
  - OR….100mM Dithiothretiol
  - OR….2mM tributyl phosphine
- Ampholytes: 2%

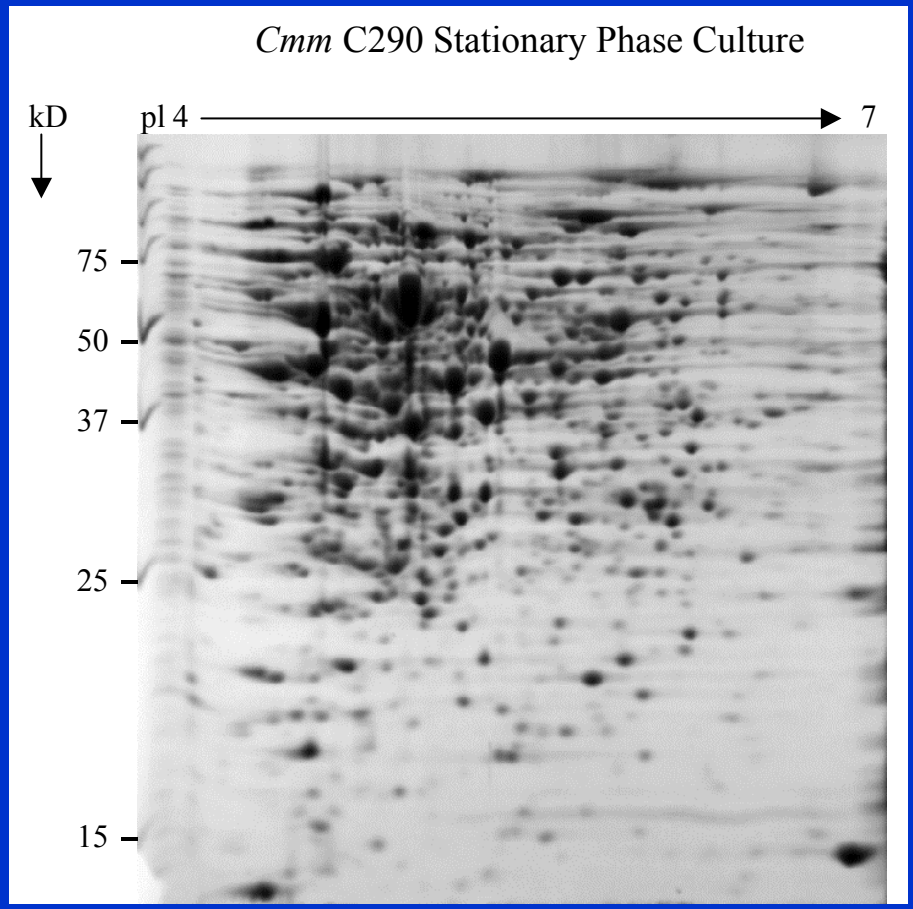
First Dimension IEF: Procedure

- Individual Strips: 24, 18, 11-13, 7cm long; 0.5mm thick

Procedure:

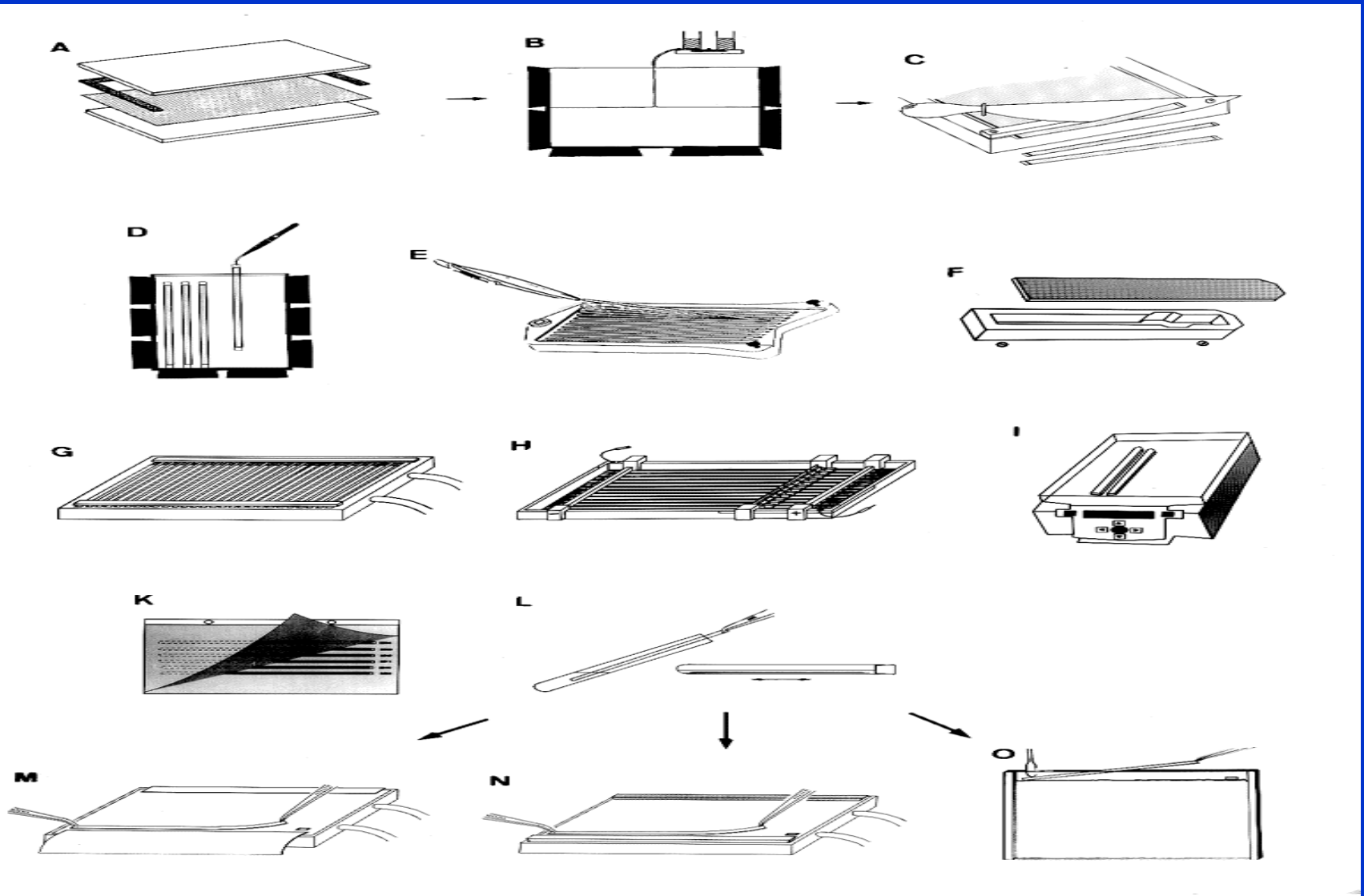
1. Rehydrate dry IPG strips (12h)
2. Apply Sample (during or after rehydration)
3. Run IPG Strips (high V, low current, 20C 4h)

Second Dimension Separation: SDS-PAGE



1. Pour linear or gradient standard SDS-PAGE gel (std = 12%)
2. Equilibrate 1-D Gel for SDS-PAGE
3. Load 1-D Gel onto SDS-PAGE gel
4. Apply Protein Ladder with Application Strips
5. Seal 1-D Gel with 0.5% LMP Agarose
6. Run Gel constant mA
7. Stain Gel : Coomassie Blue, Colloidal Coomassie Blue, Silver Stain
8. Visualize Gel & Record Image by Scanning or CCD Camera

2-DE With Immobilized pH Gradients



Gorg, A. 2000, Proteome Research, ch4. Springer

Image Analysis

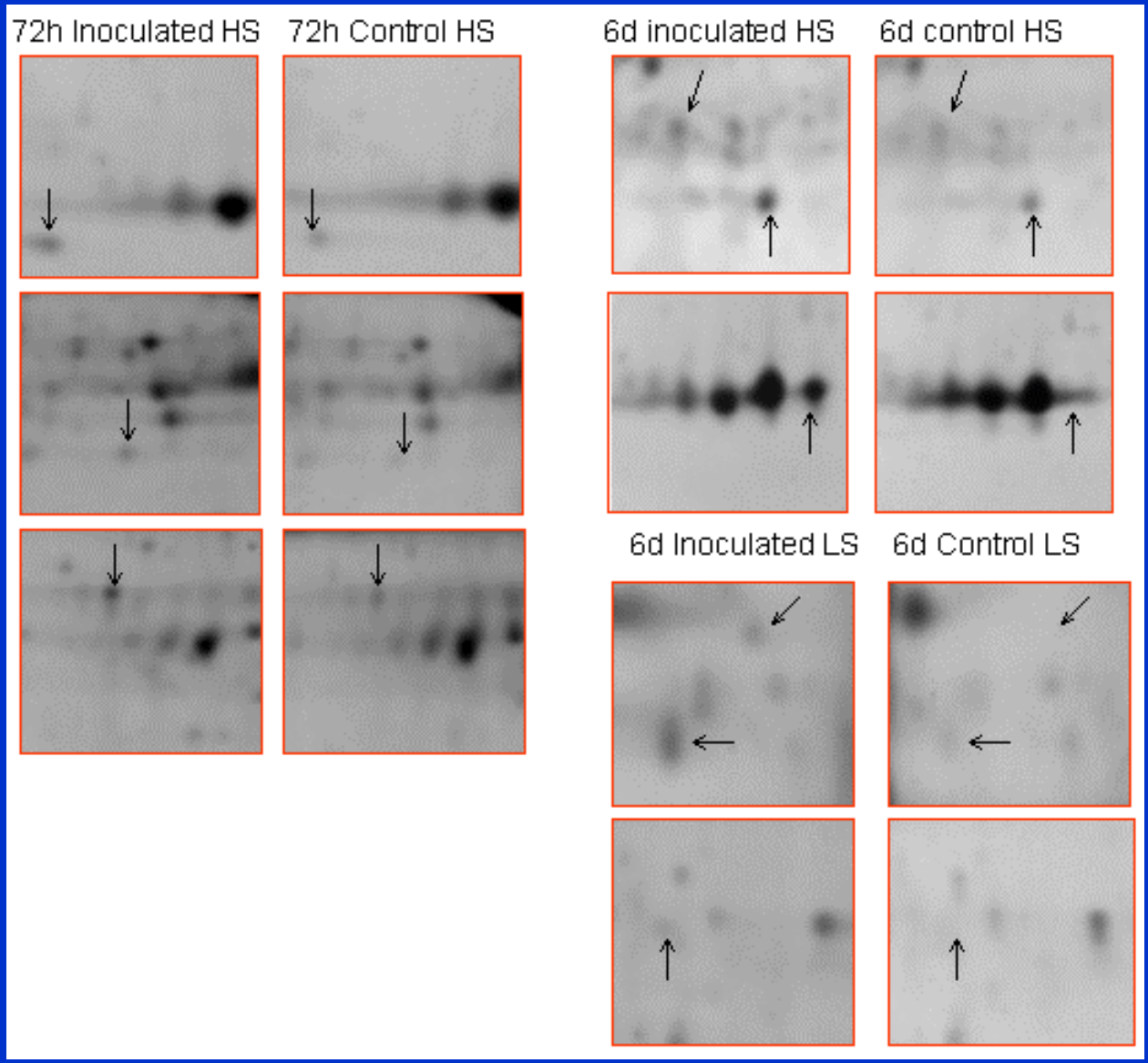
Commonly Used Software:

- ImageMaster™
- Melanie III™
- PDQuest™
- ALL EXPENSIVE- $5-10k

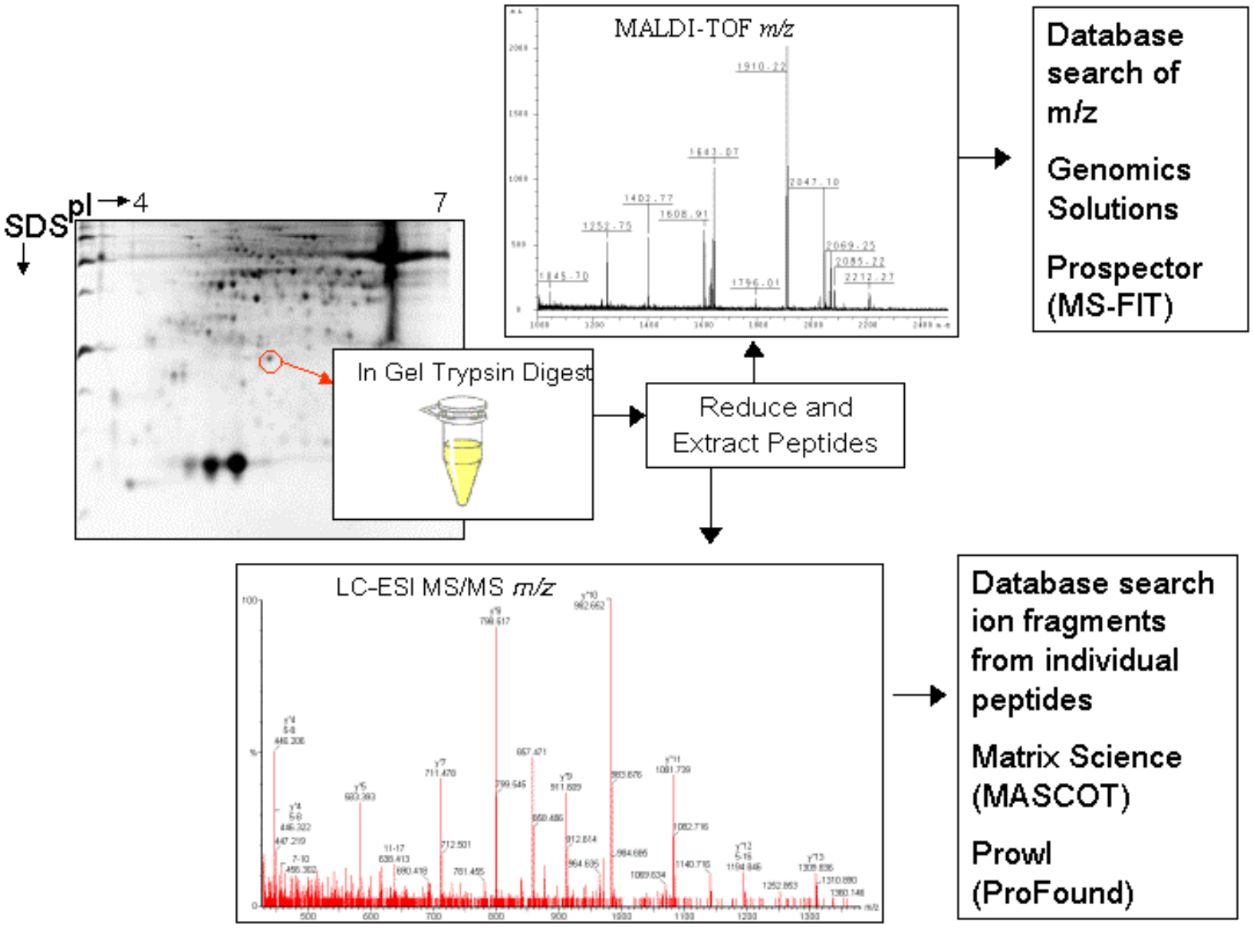
Software Functions:

- Quantification
- Detection
- Alignment
- Comparison
- Matching
- Synthetic Guassian Image from Image of Sample used in all phases

Differential Protein Expression



From Protein To Gene



Spot Picking

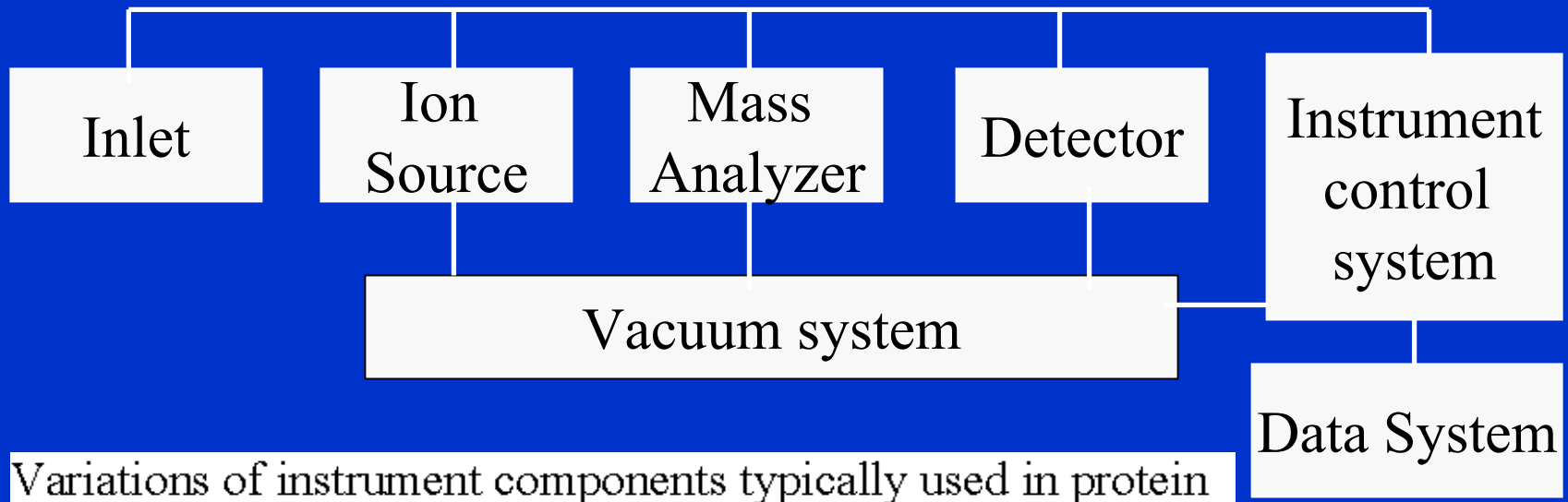
Pick Protein Spot From Gel

- Manual or Automatic

Prepare Sample for MS

- Wash Sample
- Dehydrate Sample
- Dry Sample
- In-gel digestion with trypsin (30ng trypsin, 37C, 16h)
- Extract tryptic peptides from gel
- Desalt and concentrate sample

Basic Components of a Mass Spectrometer



Variations of instrument components typically used in protein sequencing and identification experiments:

Instrument Component	
Sample Inlet	1. Direct probe or stage 2. Capillary column liquid chromatography
Ion Source	1. Electrospray 2. Matrix-assisted laser desorption
Mass Analyzer	1. Quadrupole mass filter 2. Ion trap mass analyzer 3. Time-of-flight mass analyzer

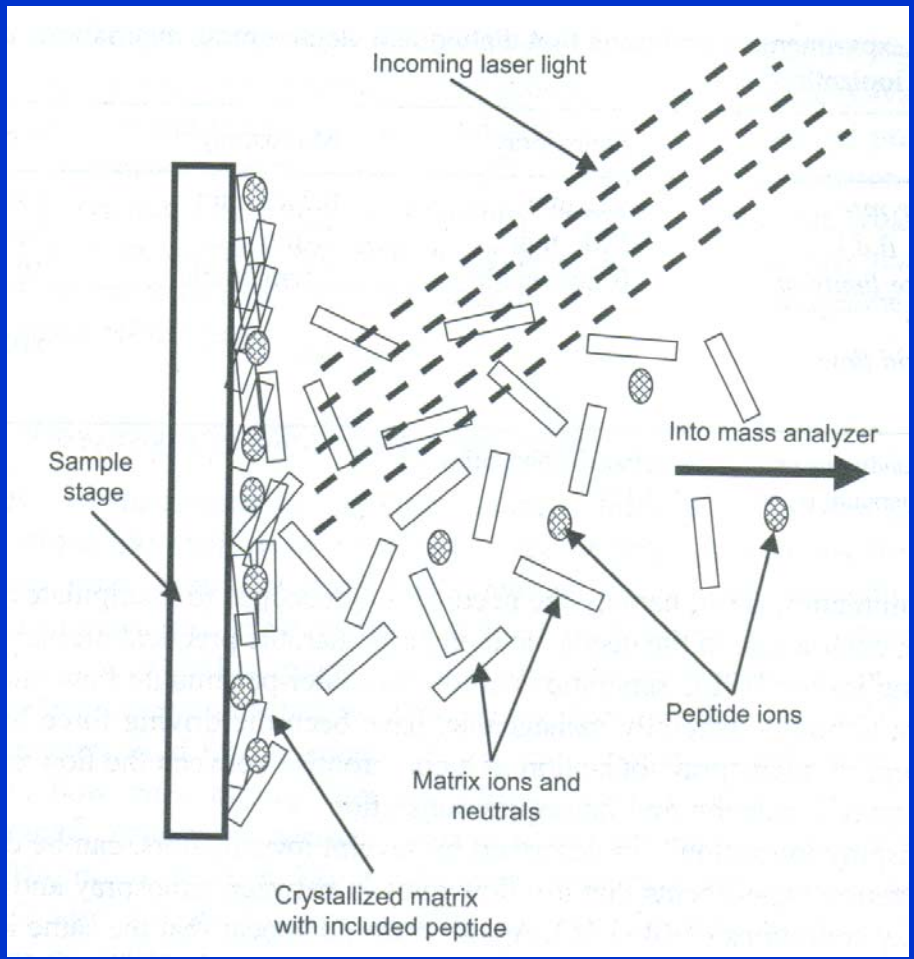
Kinter, M., and Sherman N. *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. Wiley-Interscience: New York, 2000.

Types of Mass Spectrometers

- MALDI-TOF

- ESI TANDEM MASS SPEC INSTRUMENTS
  1. Quadropole Mass Analyzers
  2. Ion Trap Mass Analyzers
  3. TOF Mass Analyzers

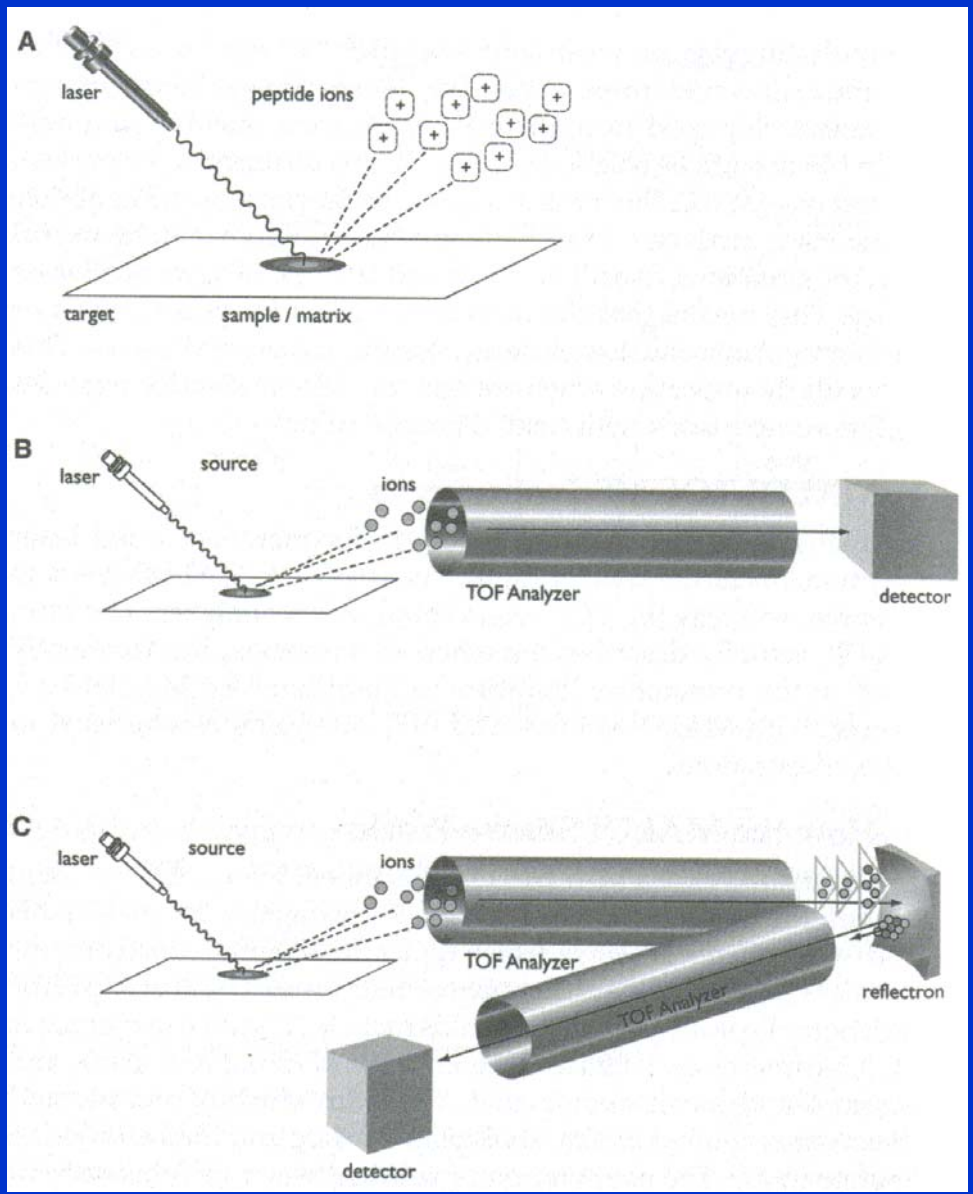
MALDI-TOF: How the MALDI Source Works



- Tryptic peptides co-crystallized with matrix compound on sample stage
- Irradiation with UV-laser
- Matrix compound vaporized and included peptide ions moved to gas phase
- Protonated peptide ions enter MS

Kinter, M., and Sherman N. *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. Wiley-Interscience: New York, 2000.

MALDI-TOF MASS SPECTROMETER



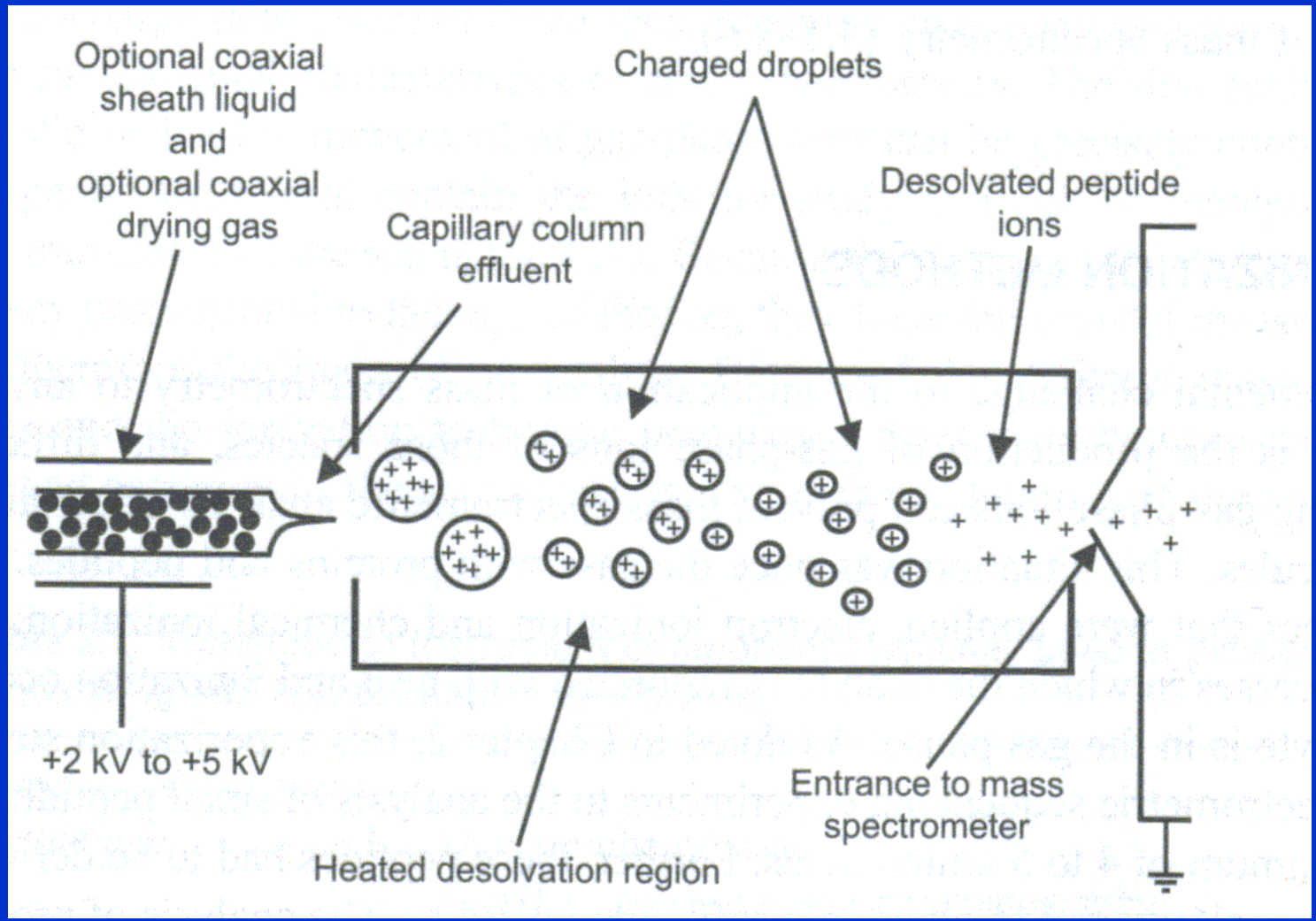
A. MALDI ionization process

B. MALDI-TOF in linear mode

C. MALDI-TOF with reflectron

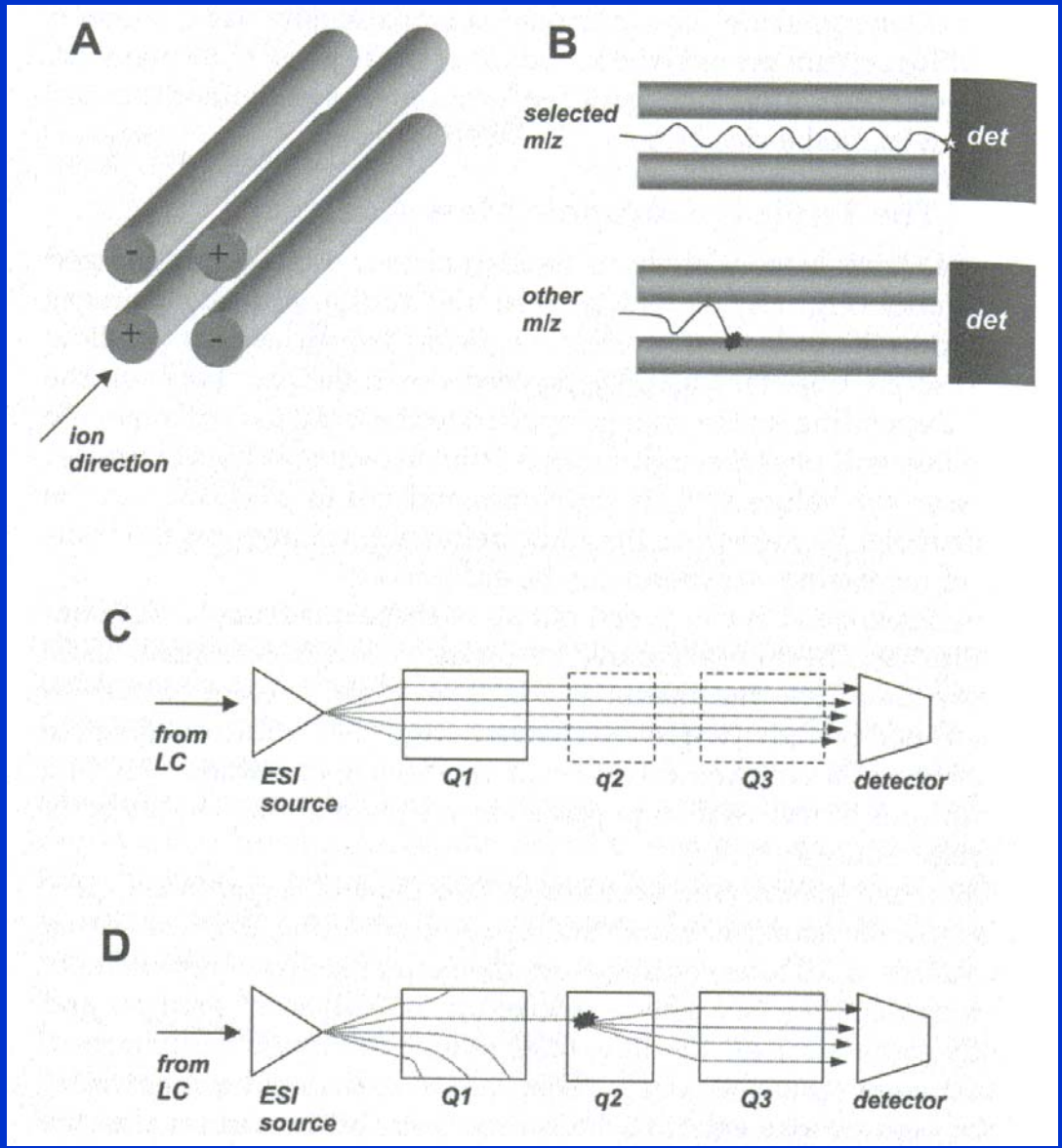
Liebler, D.C. *Introduction to Proteomics: Tools for the new biology*. Humana Press: NJ, 2002.

ELECTROSPRAY IONIZATION (ESI)



Kinter, M., and Sherman N. *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. Wiley-Interscience: New York, 2000.

TANDEM MS- TRIPLE QUADROPOLE MS



A. Quadropole Mass Analyzer

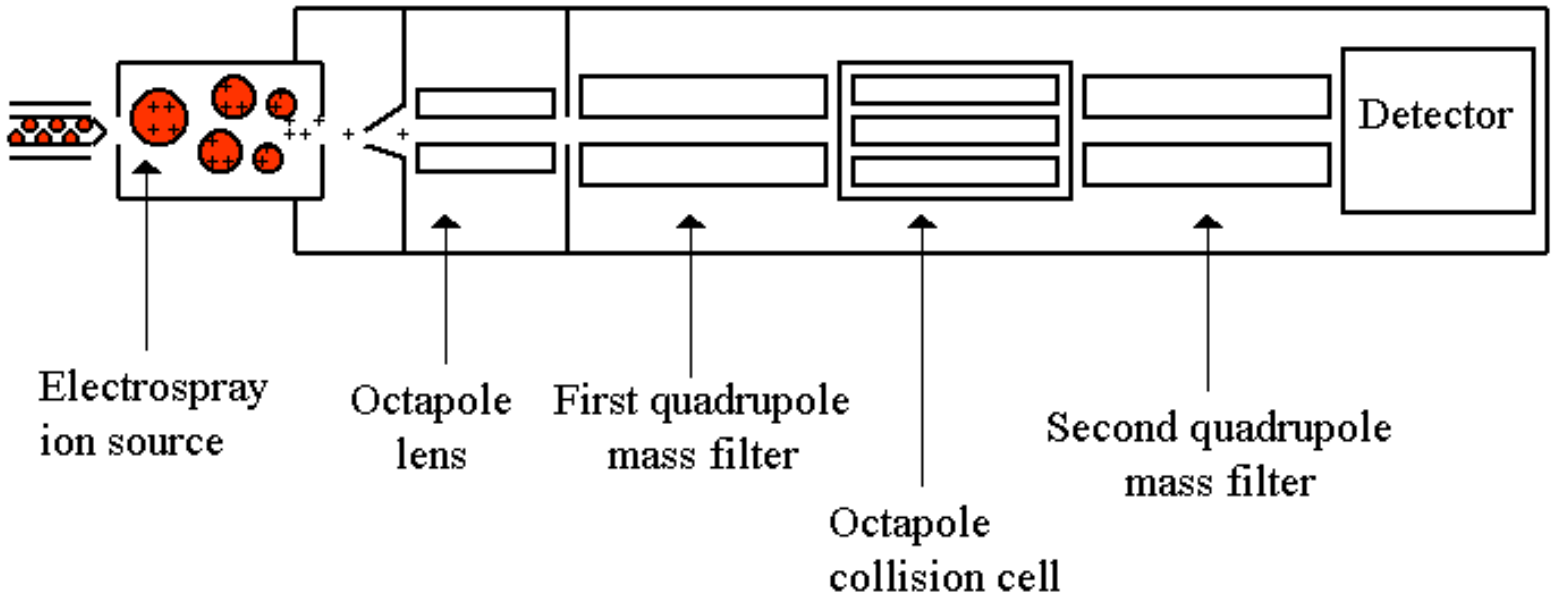
B. Tragetories of ion with selected *m/z* verses ion without selected *m/z*

C. Full-Scan Mode

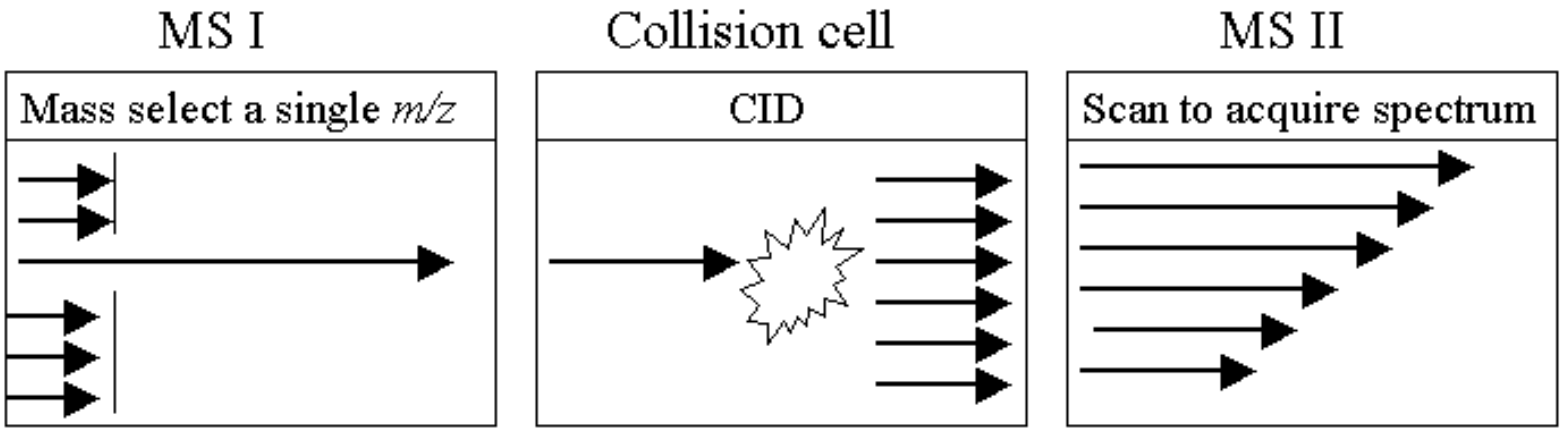
D. Tandem MS-MS Mode

Liebler, D.C. *Introduction to Proteomics: Tools for the new biology*. Humana Press: NJ, 2002.

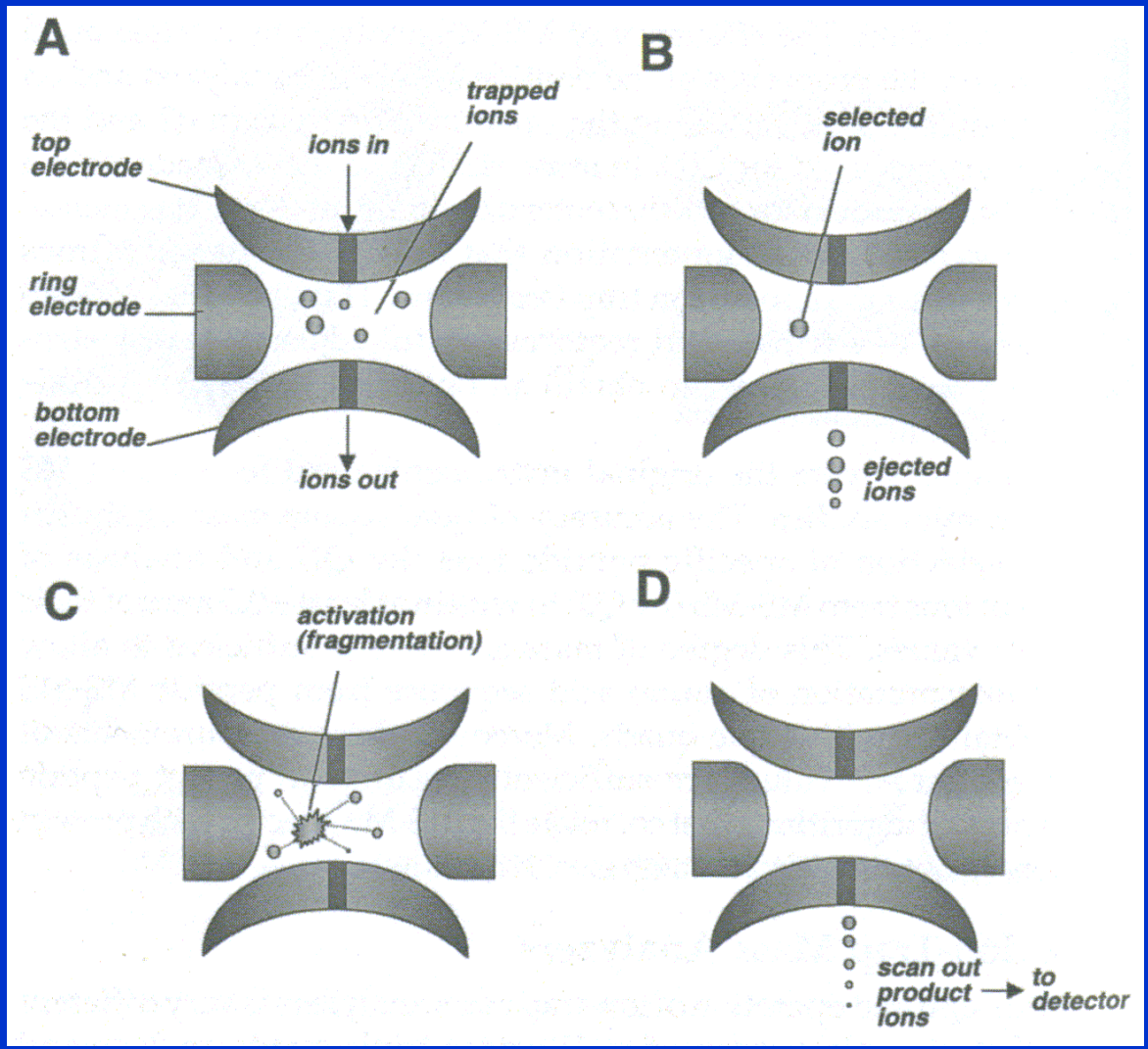
TANDEM MS: TRIPLE QUADRUPOLE MS



Product ion scan



TANDEM MS: ION TRAP MS



A. Ion Trap – Ions collected in trap maintained in orbits by combination of DC and radiofrequency voltages

B. Radiofrequency voltages on selected ions scanned to eject ions based on *m/z* and select particular ion *m/z*

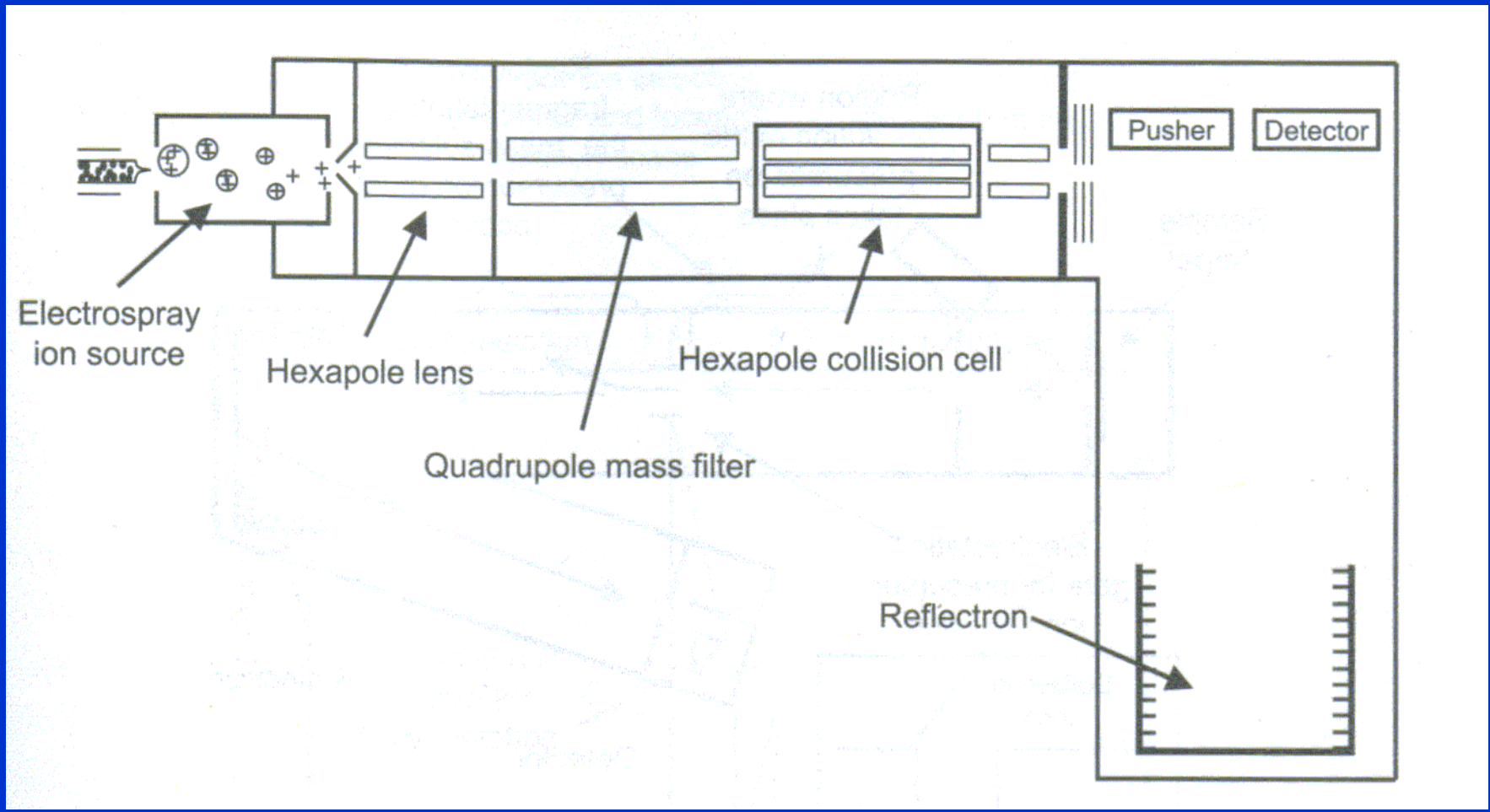
C. Collision-Induced Dissociation

D. Scan out of product ions according to *m/z*

Ion Trap - MS^n

Liebler, D.C. *Introduction to Proteomics: Tools for the new biology.* Humana Press: NJ, 2002.

TANDEM MS: QUADRUPOLE TIME OF FLIGHT MS (Q-TOF)



Liebler, D.C. *Introduction to Proteomics: Tools for the new biology.* Humana Press: NJ, 2002.

Comparison of MALDI-TOF and MS/MS

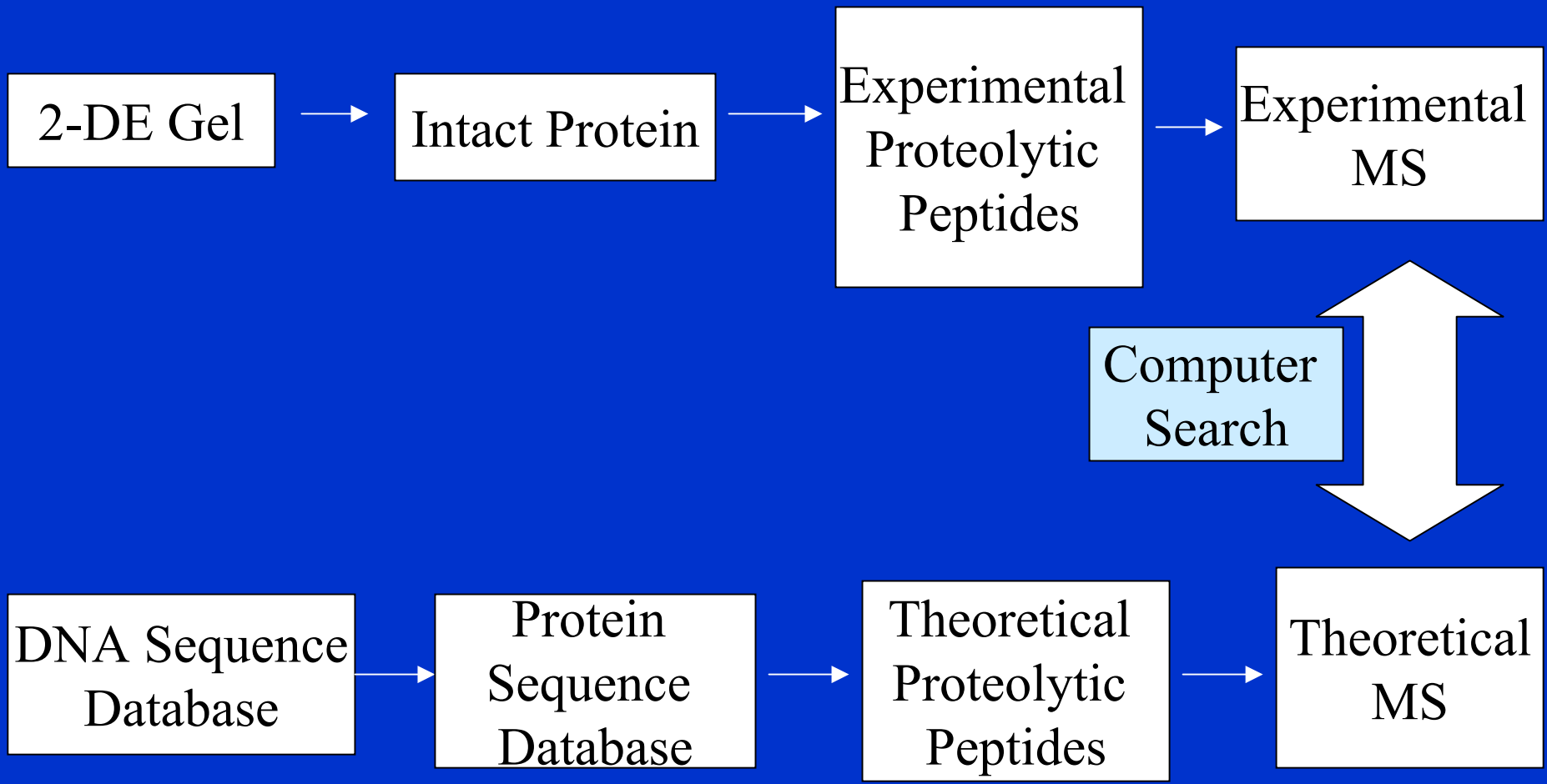
MALDI-TOF

- Sample on a slide
- Spectra generate masses of peptide ions
- Protein Id by peptide mass fingerprinting
- Expensive
- Good for sequenced genomes

TANDEM MS

- Sample in solution
- MS-MS spectra reveal fragmentation patterns – amino acid sequence data possible
- Protein Id by cross-correlation algorithms
- Very Expensive
- Good for unsequenced genomes

Protein Identification Using Peptide Mass Fingerprinting (MALDI-TOF Data)



Databases Available for Id of MS Spectra

- SWISS-PROT – nr database of annotated protein sequences. Contains additional information on protein function, protein domains, known post-translational modifications, etc. (http://us.expasy.org/sprot)
- TrEMBL- computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.
- PIR-International – nr annotated database of protein sequences. (http://www-nbrf.georgetown.edu/)
- NCBInr – translated GenBank DNA sequences, Swiss-Prot, PIR.
- ESTdb – expressed sequence tag database (NIH/NSF)
- UniProt – proposed new database. Will joint Swiss-Prot, TrEMBL, PIR. http://pir.georgetown.edu/uniprot/

Programs Used to Identify Mass Spectra

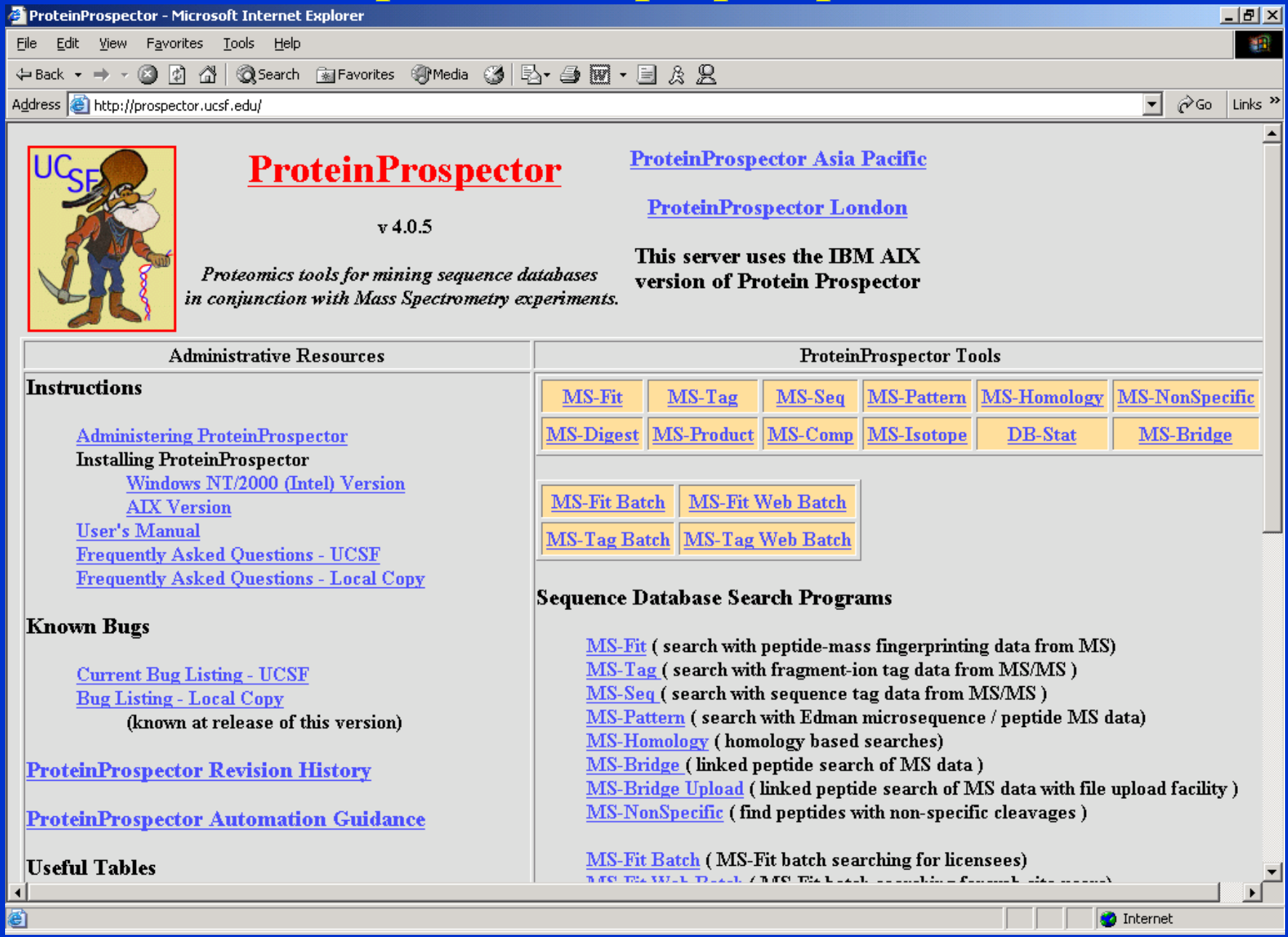
- 3 main types programs available

1. Use proteolytic peptide fingerprint for protein Id (ie MALDI-TOF data).
   - PeptIdent, MultiIdent, ProFound

2. Programs that operate with MALDI-TOF or MS-MS spectra or combination of both
   - PepSea, MASCOT, MS-Fit, MOWSE

3. Programs that operate with MS-MS spectra only
   - SEQUEST, PepFrag, MS-Tag, Sherpa

Protein Prospector - http://prospector.ucsf.edu/



Mass Spec Algorithms for Protein Id (MS-MS only)

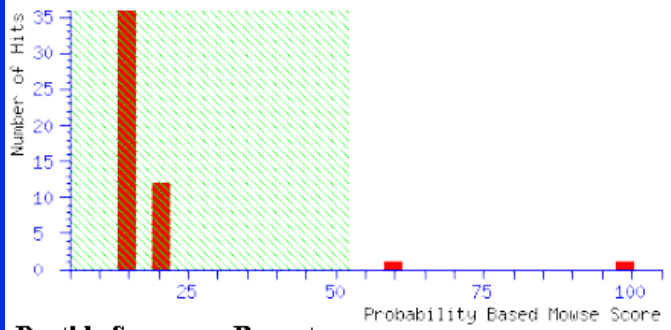
- More perfect algorithms use additional information such as pI, MW, amino acid composition, etc (example: **MOWSE** algorithm).

MATRIX SCIENCE

Mascot Search Results

Probability Based Mowse Score

Score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 52 indicate identity or extensive homology (p<0.05).



Peptide Summary Report

1. gi|6066418 **Mass:** 27447 **Total score:** 99 **Peptides matched:** 2
2. ascorbate peroxidase [Lycopersicon esculentum]

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
10	683.05	2046.13	2045.90	0.23	0	38	1	YAADEDAFFADYAEAHLK
11	690.44	2068.30	2068.04	0.26	0	61	1	QMGLSDQDIVALSGAHTLGR

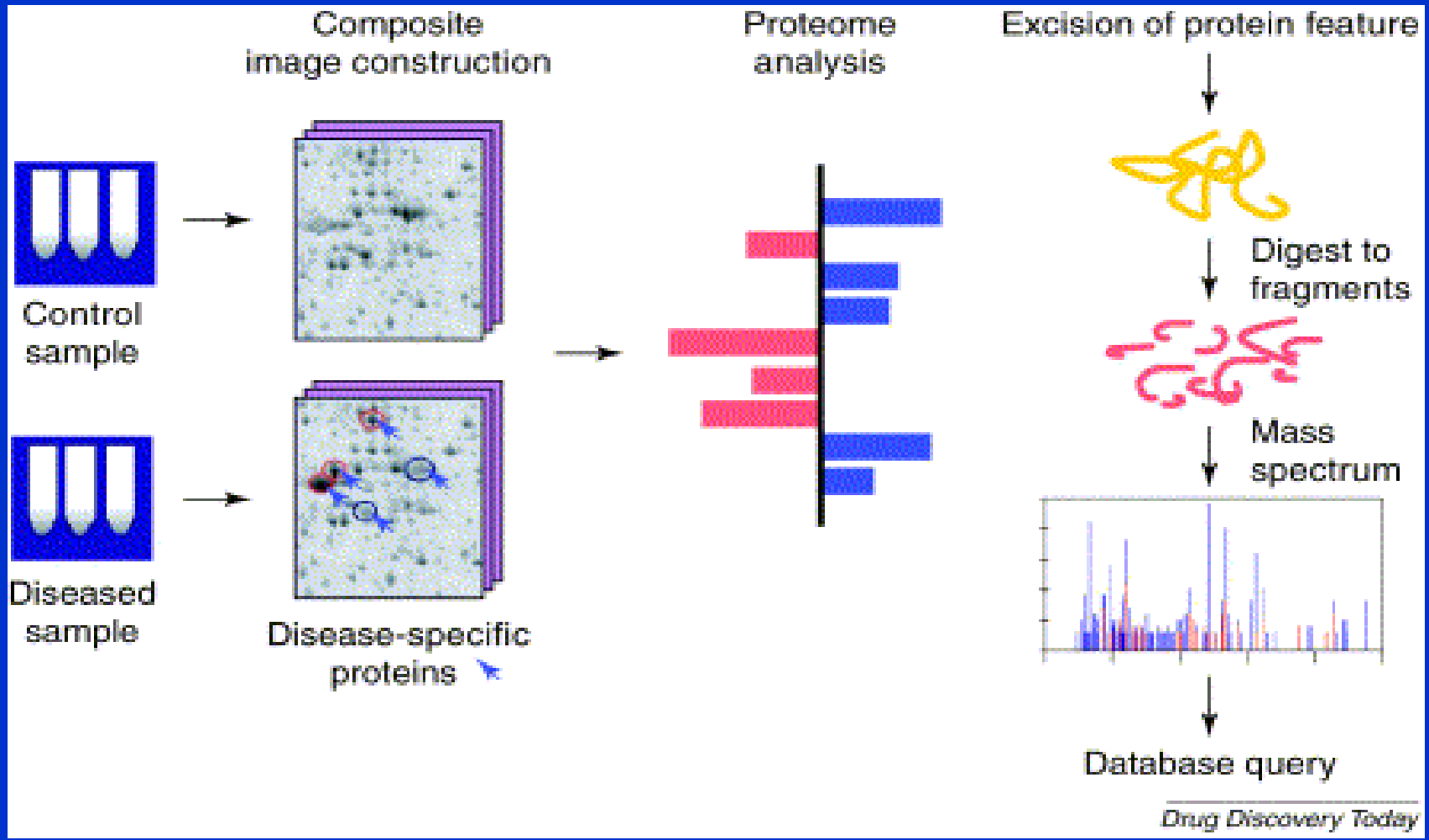
2. gi|99375 **Mass:** 29040 **Total score:** 61 **Peptides matched:** 1
L-ascorbate peroxidase (EC 1.11.1.11) precursor - Arabidopsis thaliana (fragment)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
11	690.44	2068.30	2068.07	0.22	1	61	1	QMGLSDKDIVALSGAHTLGR

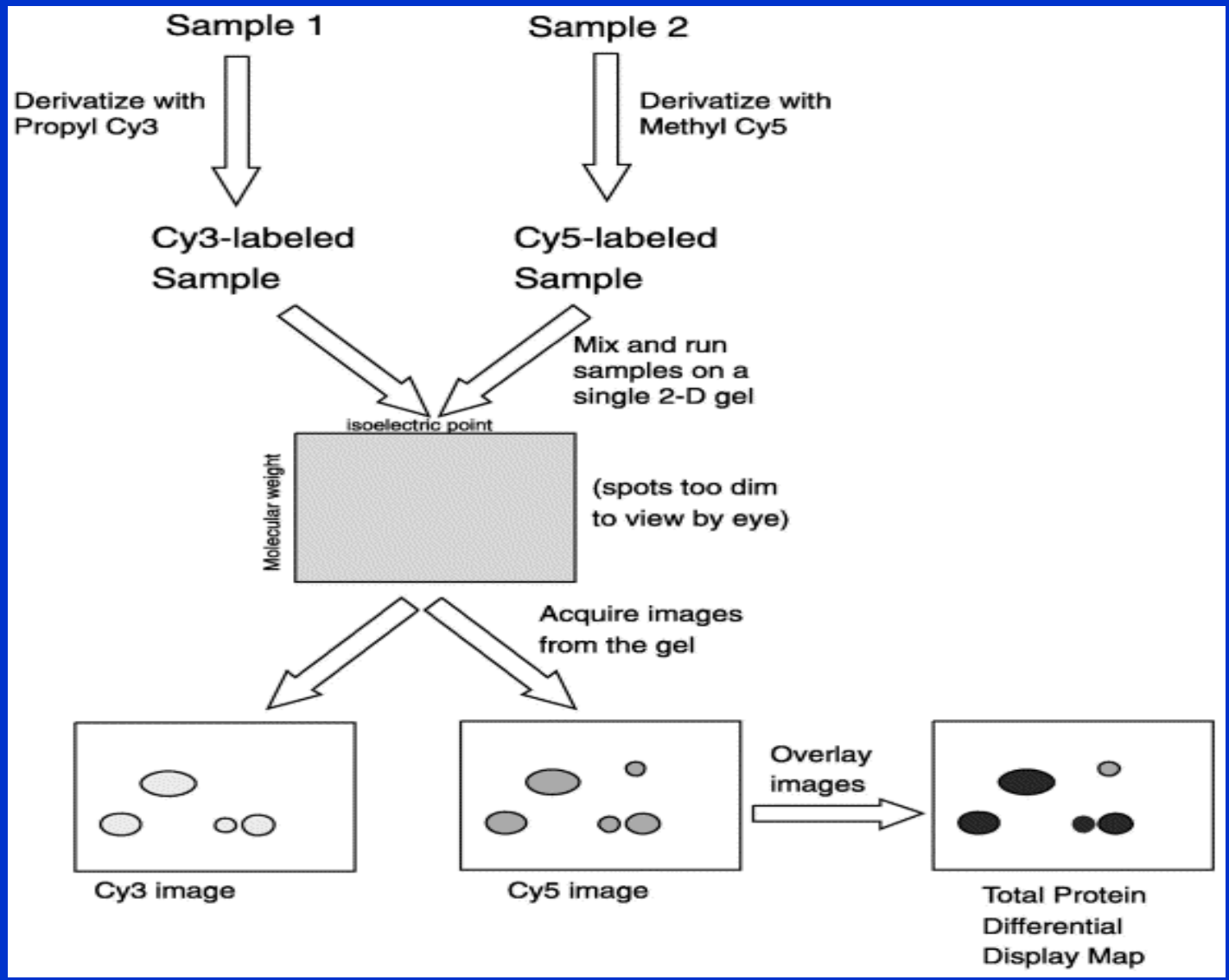
Proteomics Applications

- Differential Display Proteomics
  - DIGE – Difference gel electrophoresis
  - MP – multiplexed proteomics
  - ICAT – isotope coded affinity tagging

Protein Expression Profile Analysis

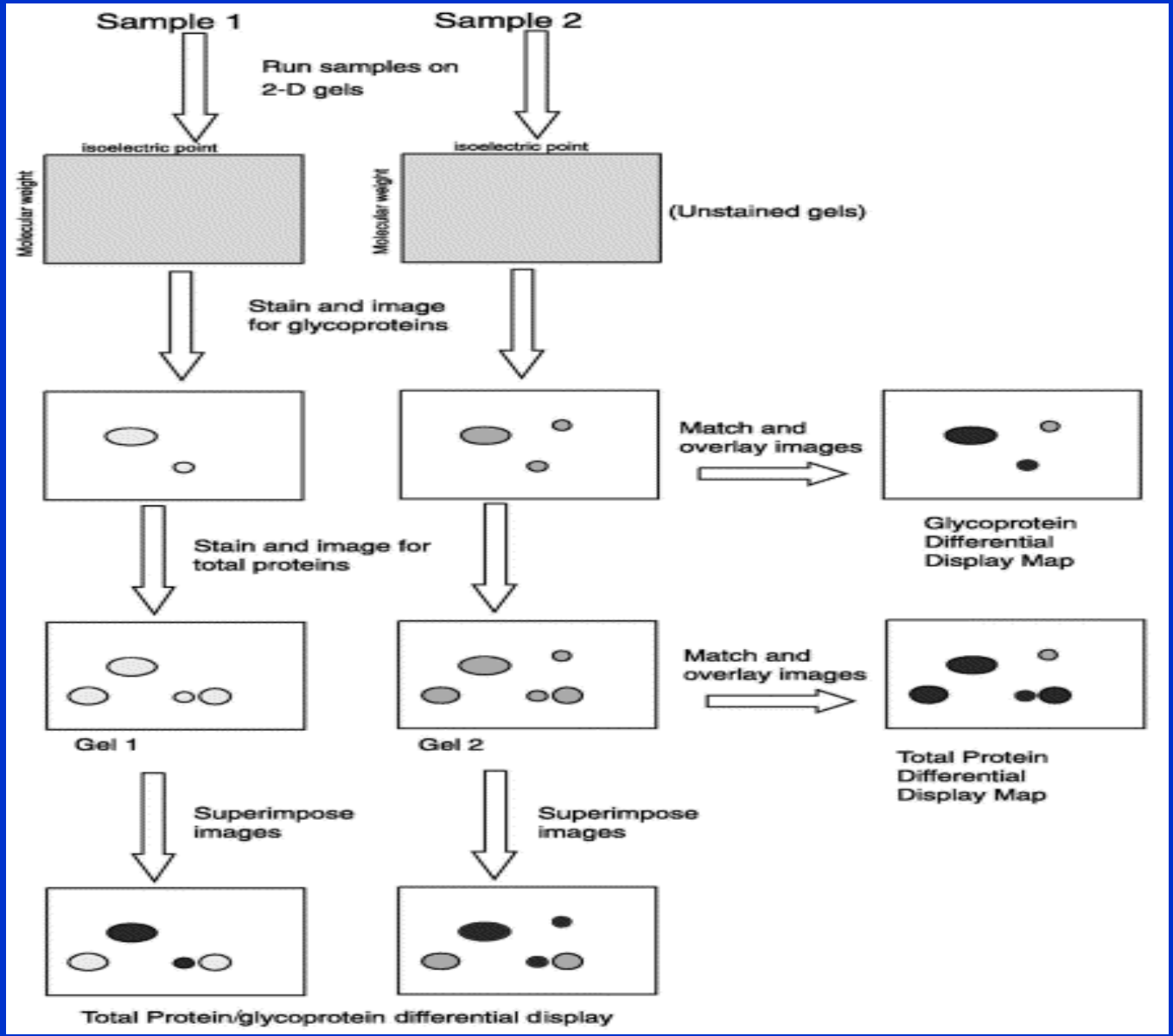


Difference Gel Electrophoresis (2D-DIGE)

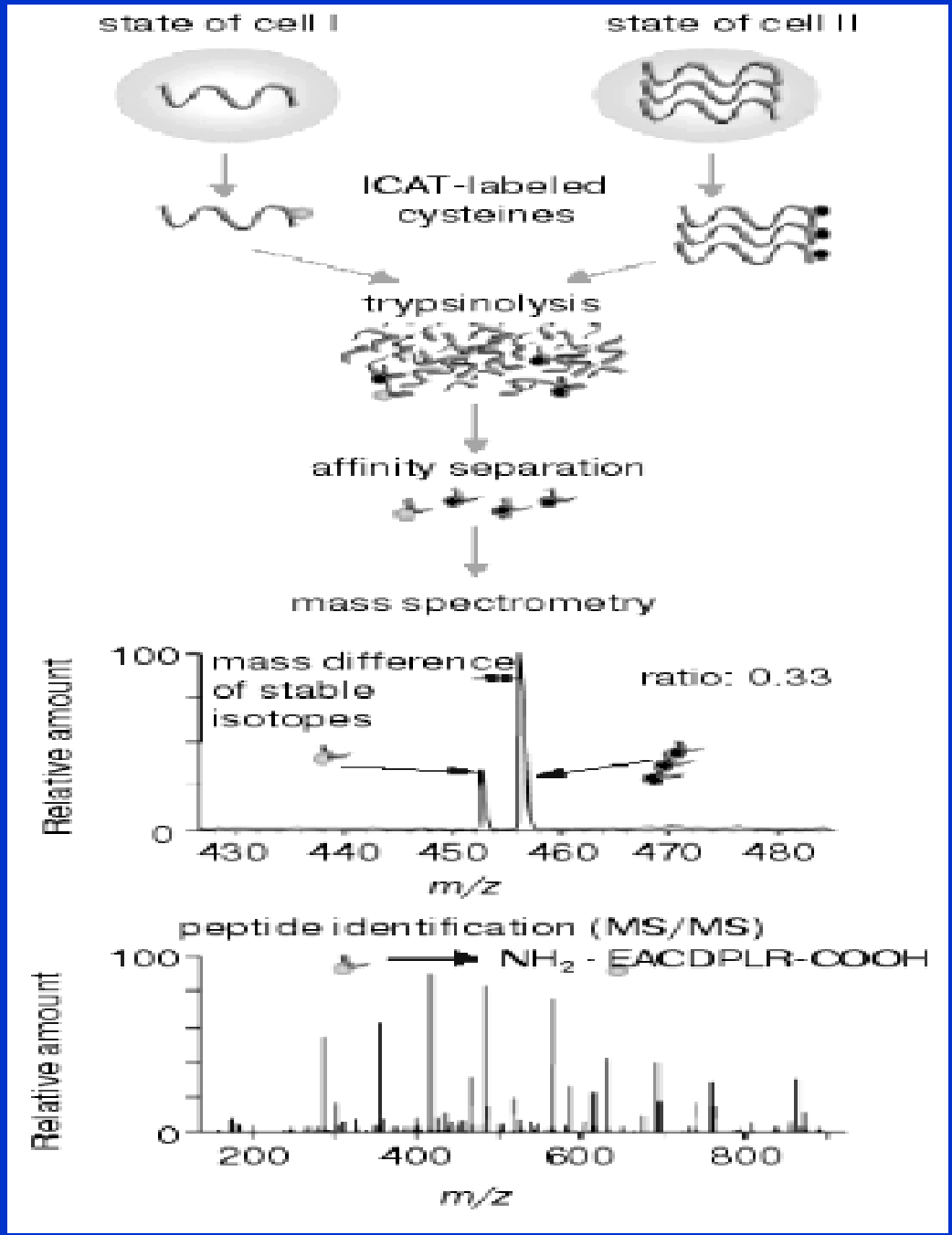


(Unlu, 1997, electrophoresis 18, 2071)

Multiplexed Proteomics (MP)



(Steinberg, 2001, Proteomics 1,841, 2071)



Isotope-Coded Affinity Tagging (ICAT)

(Smolka, 2002, Mol Cell Proteomics 1, 19-29)

Conclusions

- 2-DE is a powerful technique to separate of complex protein mixtures and analyze proteomes.
- Mass Spectrometry microsequencing can identify proteins from 2-DE gels and other samples.
- There are multiple databases and computer programs available to analyze MS data for protein Identification
- Proteomics approach can be used to identify all proteins in particular sample, elucidate additional components of biochemical pathway(s), or analyze post-translational modifications at a small or large scale.