- Important insights into gene function can be gained by studying gene expression at the mRNA level.
- Disadvantages reflecting the fact that mRNA represents an early stage of gene expression
- Abundance of a given transcript may not reflect the abundance of the corresponding protein due to post-transcriptional gene regulation
- Not all mRNAs are translated at the same rate and some may not be translated at all
- Protein diversity is often generated after transcription, sometimes by the synthesis of two or more types of protein from a given mRNA by post-translational modification (e.g. glycosylation, phosphorylation).

- Function of a protein may depend on its localization, and trafficking between compartments can sometimes be used in a regulatory manner (e.g. to control the activity of transcription factors).
- Abundance of a correctly modified protein cannot be guaranteed to represent the true intracellular activity of the protein.
- All these factors must be taken into account in functional analyses and can only be addressed by studying proteins directly.

- **Protein expression analysis is more challenging than mRNA profiling because proteins cannot be amplified like nucleic acids**

Proteomics

- **Proteome, the entire complement of proteins synthesized by a given cell or organism**
- **Proteomics, the global study of proteins**

MASS SPECTROMETER BASED PROTEOMICS

- **Principles**
  - **Mass spectrometry is based on the fact that ions of differing charge and mass will move differently in a magnetic field. In proteomics the proteins are first separated by some means and then analyzed with a mass spectrometer**

Separating the Proteome

- **The protein genome is separated by several different methods.**
- **Many researchers are first separating portions of the genome, such as isolating organelles, and then analyzing that portion.**
- **This is because often proteins of interest, regulatory proteins are in low abundance.**
- **The most commonly used method is 2-dimensional gel electrophoresis.**
  - **Consists of using isoelectric focusing with SDS polyacrylamide gel electrophoresis**

Isoelectric focusing

- This separates proteins based on isoelectric point
- The isoelectric point is the pH at which the protein has no net charge.
- pH gradients may be large 2-10 or small 6-7
- Typically this is done with an immobilized pH gradient gel strip or with a tube gel containing a low concentration of polyacrylamide.
- Ampholytes are added to create a pH gradient in an electric field and the proteins are loaded.
- The IEF gel is placed in an electrophoresis system for up to 24 hours and the proteins form tight bands at their isoelectric point.
- The IEFgels are now ready for the second method.

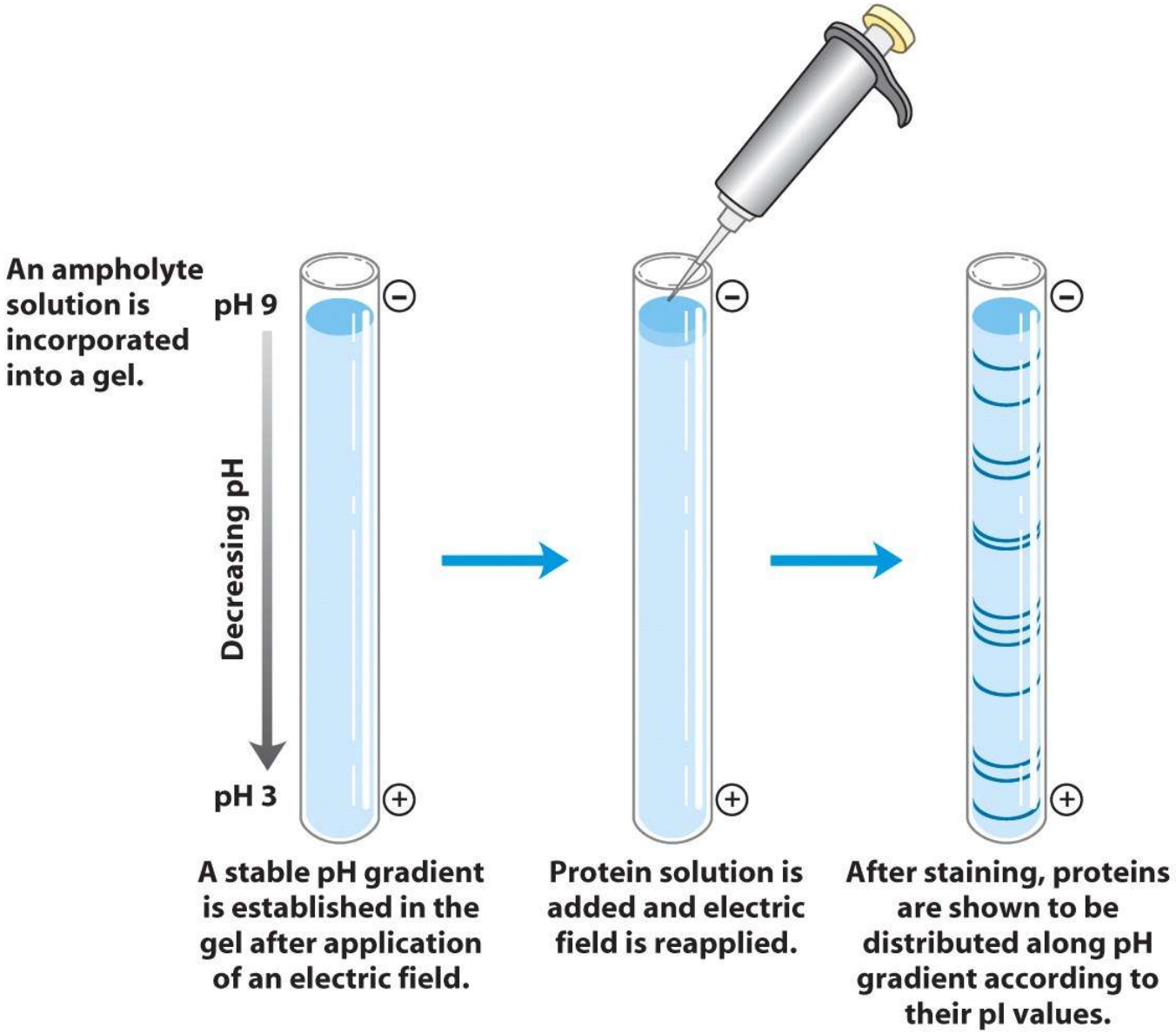


Figure is from "Principles of Biochemistry" Lehninger, Fourth Edition

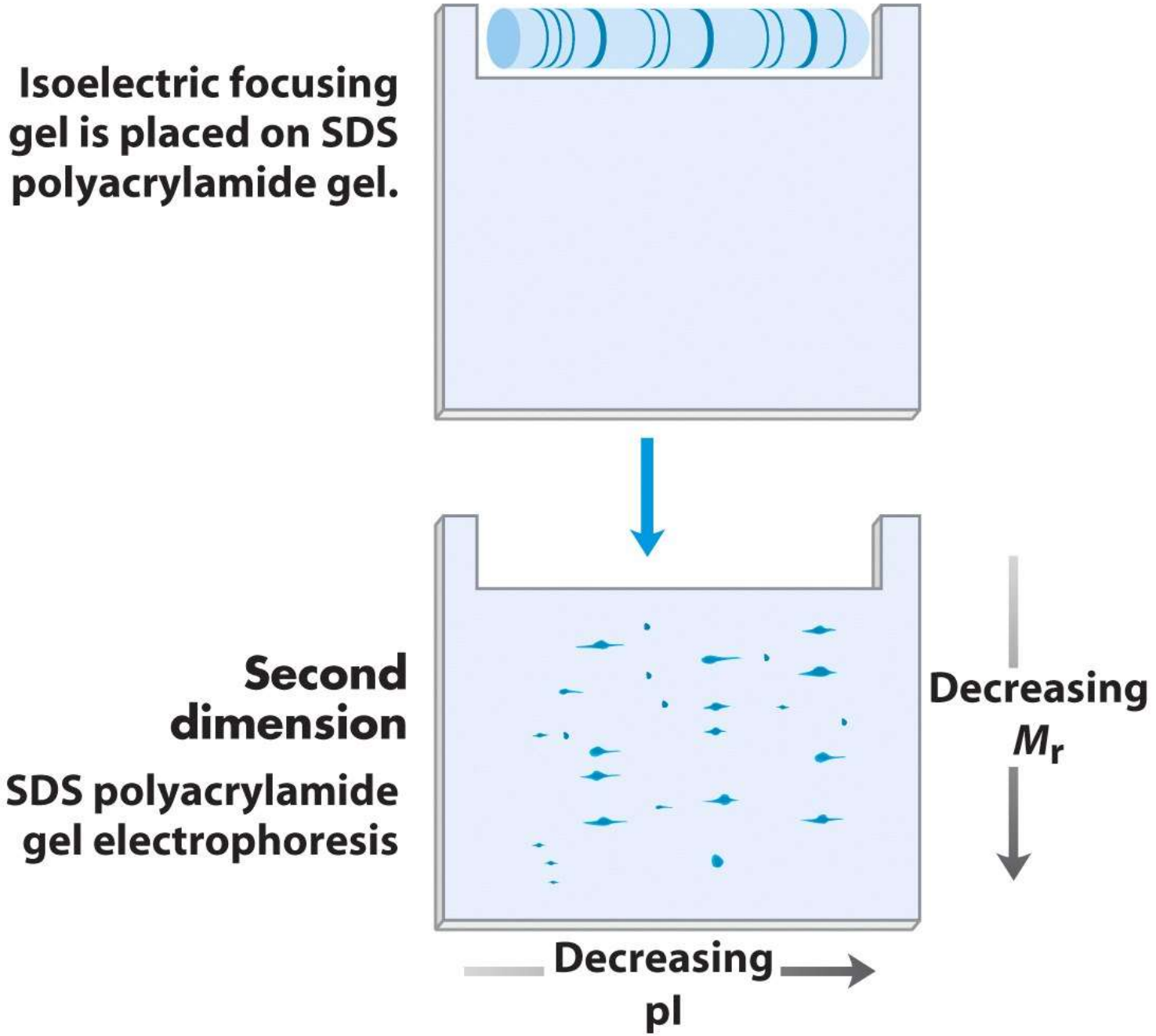
SDS Polyacrylamide Gel Electrophoresis

- **The second dimension separates the proteins based on size.**
- **There are two parts, the stacking gel which concentrates the sample and the running gel (separating gel) that is used to separate the proteins.**
- **The IEF gel is soaked in a solution containing chemical to denature the proteins including sodium dodecyl sulfate a detergent which gives the proteins a net negative charge. This means that all proteins will move in one direction.**
- **The IEF gel is then put in the one long well in the stacking gel, sealed in place with agarose, and the proteins subjected to an electric field to separate.**
- **The larger proteins are found at the top and the smaller ones are found at the bottom of the gel.**

2-Dimensional Gel Electrophoresis

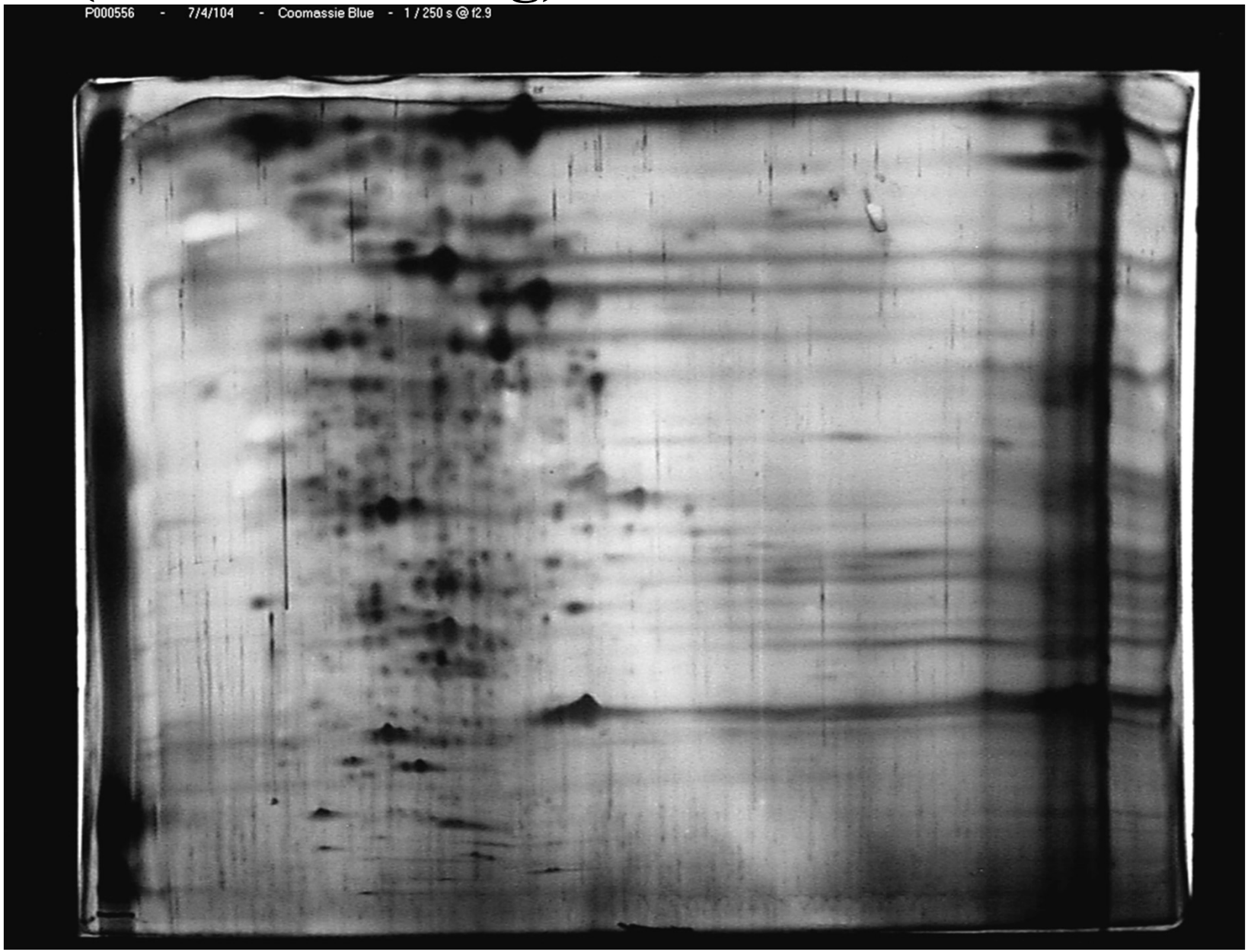
- In a 2D gel the proteins appear as spots on the gel rather than bands. These spots can then be further processed or used for mass spectrometry directly.
- Further processing usually includes spot excision, trypsin digestion, and mass spectromety
- Analysis may also include differential 2D gel electrophoresis
  - In this case a control and sample are separately labeled with a fluorescent molecule.
  - The samples are mixed and electrophoresed in the same gels.
  - A laser scanner is used to identify each spot and a program puts the two images together.

2-D Gel Electrophoresis



2-D Gel (denaturing)

2-D Gel (non-denaturing)



pH 4 **pH 10**

Alternate Separation Methods

- The first dimension is run in larger agarose tube gels with ampholytes.
- This has less resolution than polyacrylamide gels. The tubes are sliced and the proteins are allowed to diffuse out.
- Gel regions are cut, proteins eluted and the proteins are then separated by capillary electrophoresis.
- Capillary electrophoresis has a much greater resolution for the proteins mass.
- Proteins are eluted from the capillary in the process and can be collected. They are readily available for mass spectrometry.

Alternative Separation Methods

- **Whole proteome is analyzed at once.**
- **Proteome is digested with protease (trypsin)**
- **Digested proteome is injected to HPLC with 2 columns in series (mixed bed ion exchange and reverse phase)**
- **Peptides are eluted from ion exchange onto reverse phase and then separated on reverse phase column.**
- **Peptides then enter ESI-MS-MS**

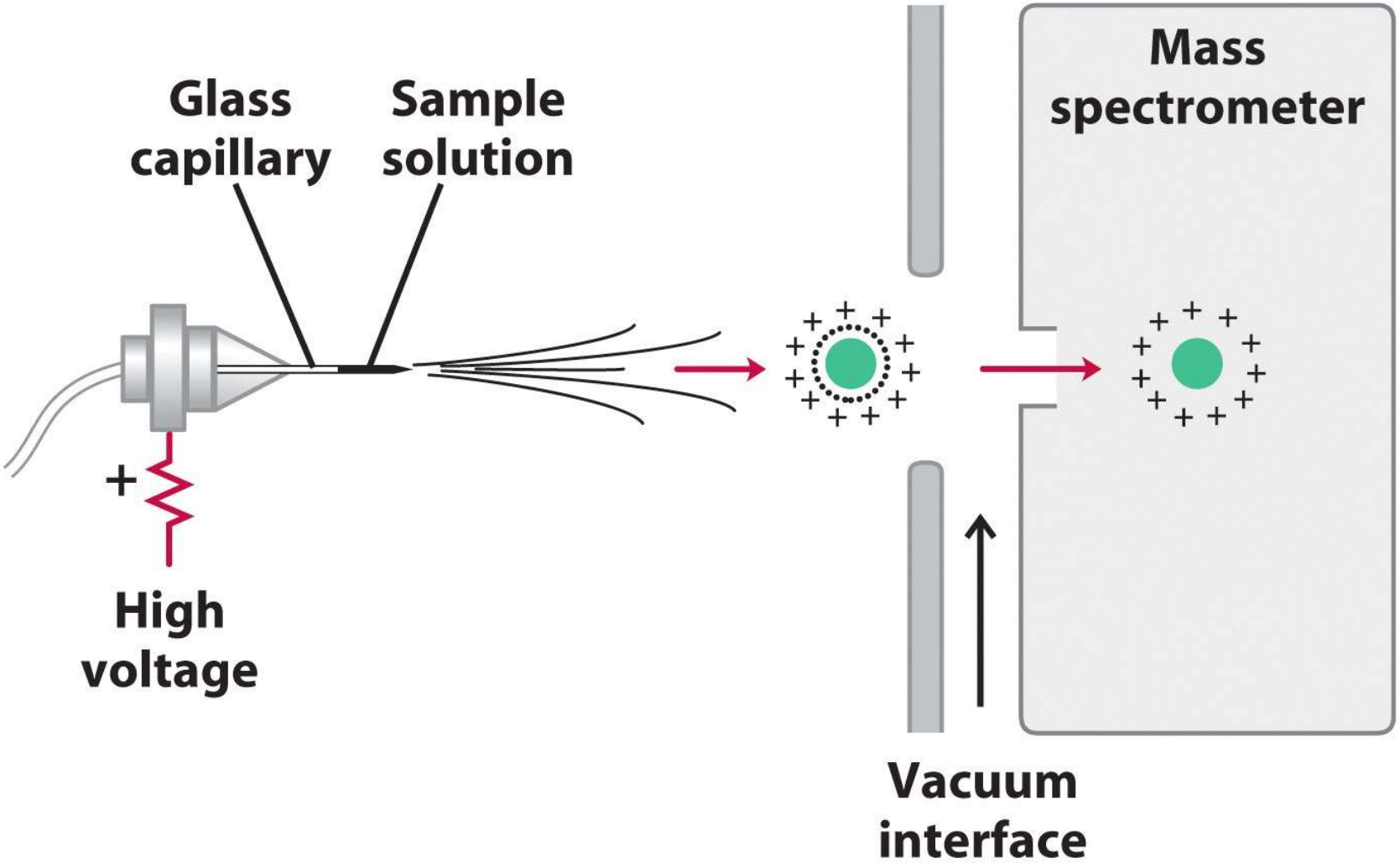
Mass Spectrometery

- **Separates ions based on mass to charge ratio.**
- **Charges are placed on the protein or the peptide by ionization.**
- **Two most common types of ionization are:**
- **Matrix-Assisted Laser Desorption Ionization.**
  - **MALDI causes fragmentation of the protein during ionization. Can be used to get more information about the fragments. Easier to do than ESI.**
- **Electrospray ionization (ESI)**
  - **ESI can give whole protein masses as well as complex masses. If the proteins is first separated by reverse phase HPLC before injection only the subunits masses will be known.**

Matrix-Assisted Laser Desorption Ionization (MALDI)

- **MALDI causes fragmentation of the protein during ionization. Can be used to get more information about the fragments. Easier to do than ESI.**
- **Requires sample to be placed in matrix that absorbs appropriate wavelength light.**
- **Matrix generates heat and forms ions of matrix and what is around it.**

Electrospray Ionization



Mass Analyzers

- **Important parameters**
  - **Sensitivity**
    - **How few ions can be detected.**
  - **Resolution**
    - **How well different masses can be determined.**
  - **mass accuracy**
    - **How reproducible and correct are the masses.**

Mass Analyzers (MS)

- **Quadrapole**
  - **High Sensitivity, acceptable mass accuracy and resolution**
  - **Easily coupled to chromatography**
- **Time of Flight**
  - **High Sensitivity, high mass accuracy, high resolution**
  - **Limited to small m/z ratios**
  - **Not easily coupled to chromatography**
  - **Easily coupled to MALDI**

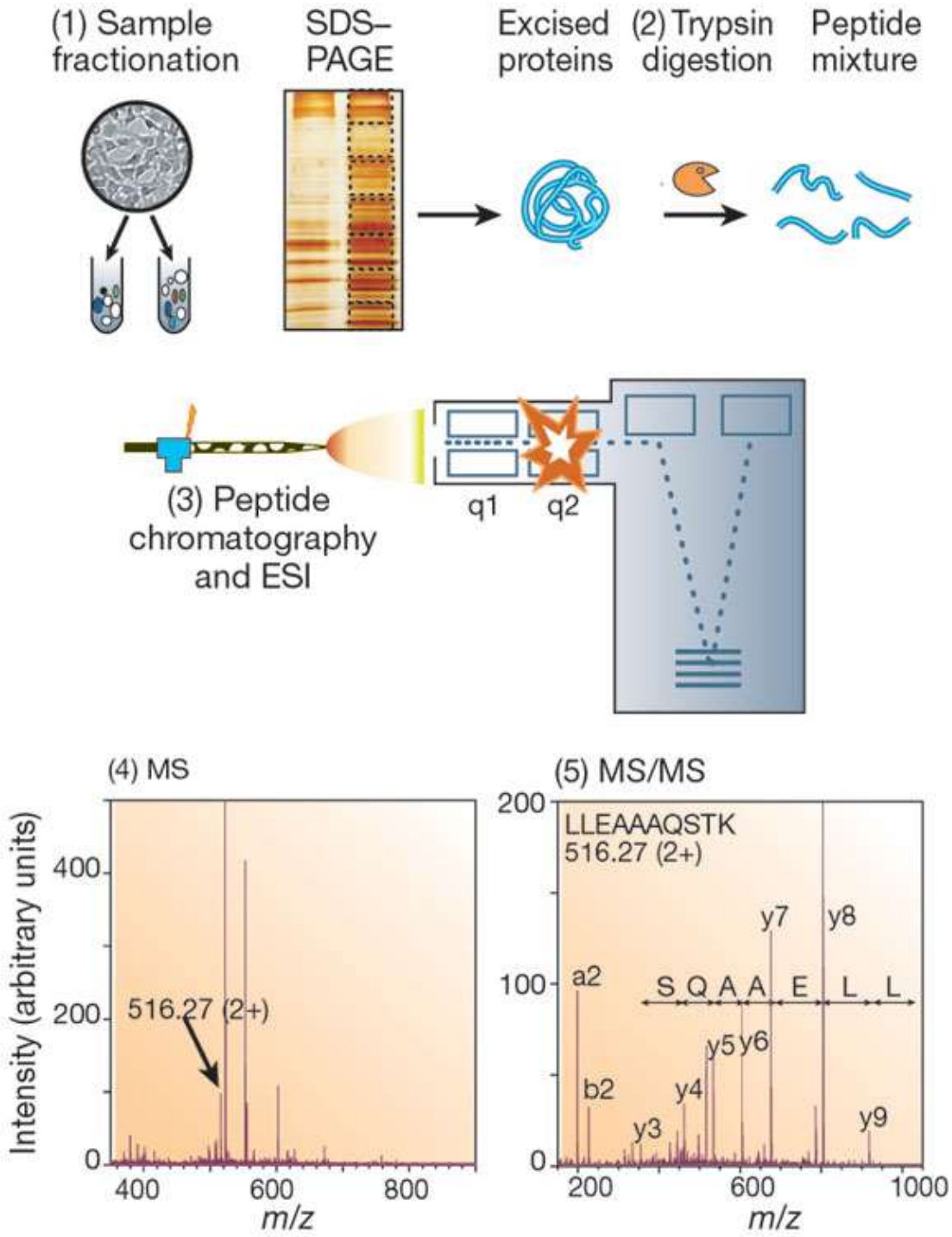
Mass Analyzers (MS)

- **Ion Trap**
  - **High Sensitivity**
  - **Low mass accuracy and resolution**
- **Fourier Transform ion cyclotron**
  - **High sensitivity, mass accuracy, resolution, dynamic range**
  - **Expensive, difficult to operate, low fragmentation efficiency**

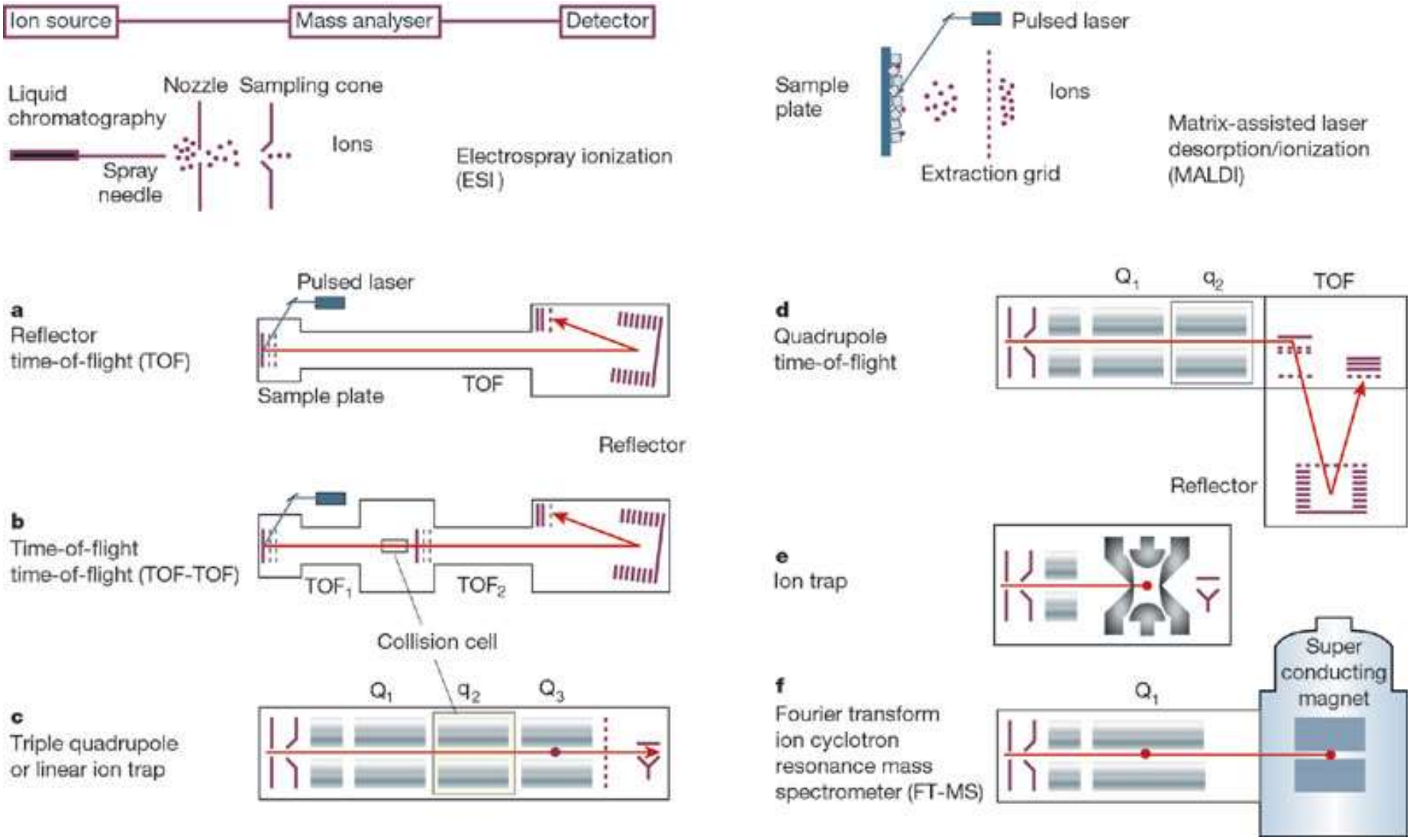
Mass Spectrometers

- **Instruments are often coupled**
  - **MS/MS**
    - **ESI - quadrapole MS -TOF-MS**
  - **Collider**
    - **Is essentially a Quadrapole MS with collision gas included**
    - **Forms collision induced ions**
      - **Gives collision induced spectra (CID)**

-Quadrapole MS
- Collider
- TOF MS



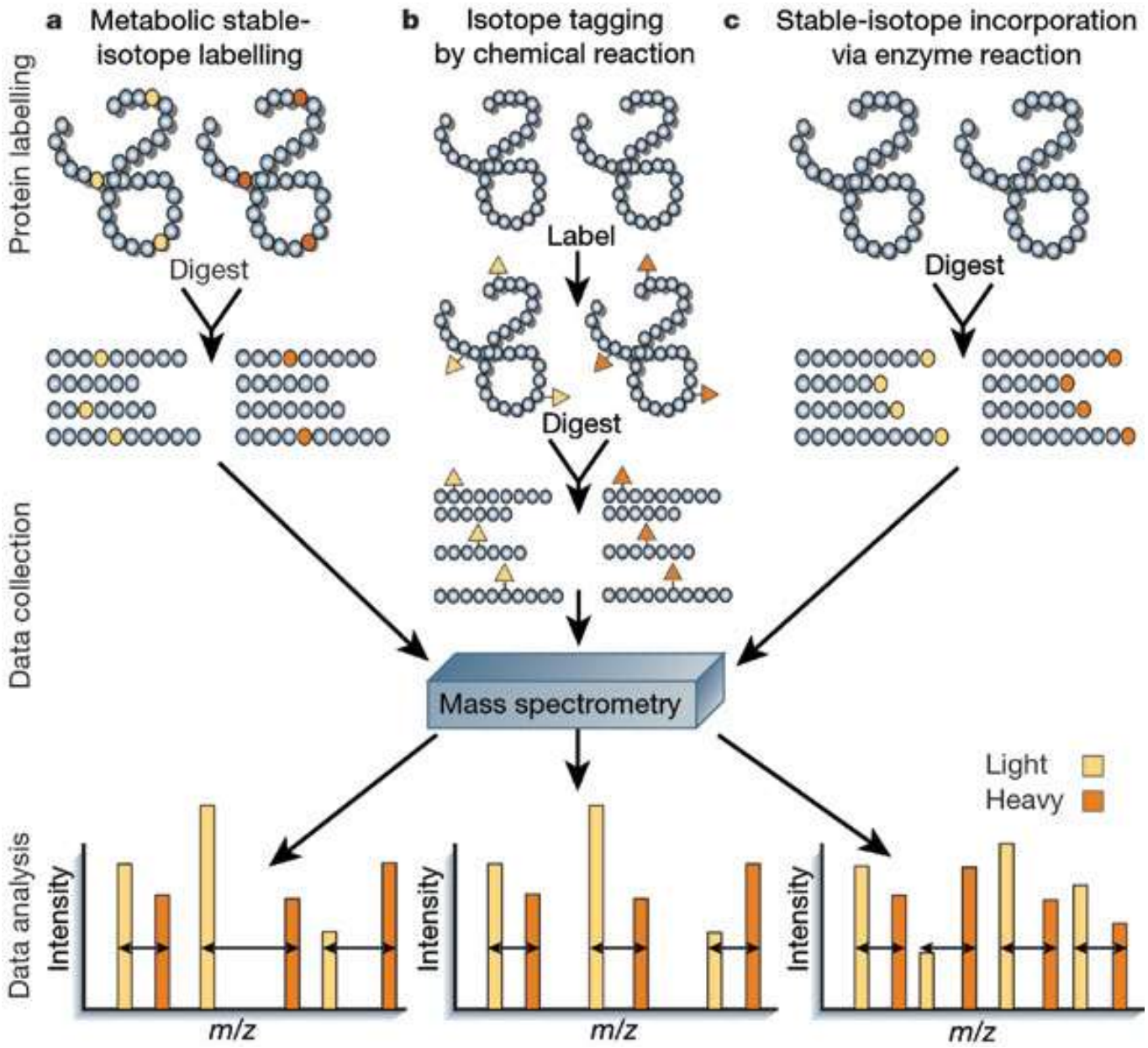
Mass spectrometers



Protein identification and Quatitation

- **To quantitate**
  - **add stable isotopes**
  - **post separation modification**
    - **SH, NH_2, N-linked carbohydrates**
- **Incorporation of isotopes in culture**

MS Quantitation



Protein Identification

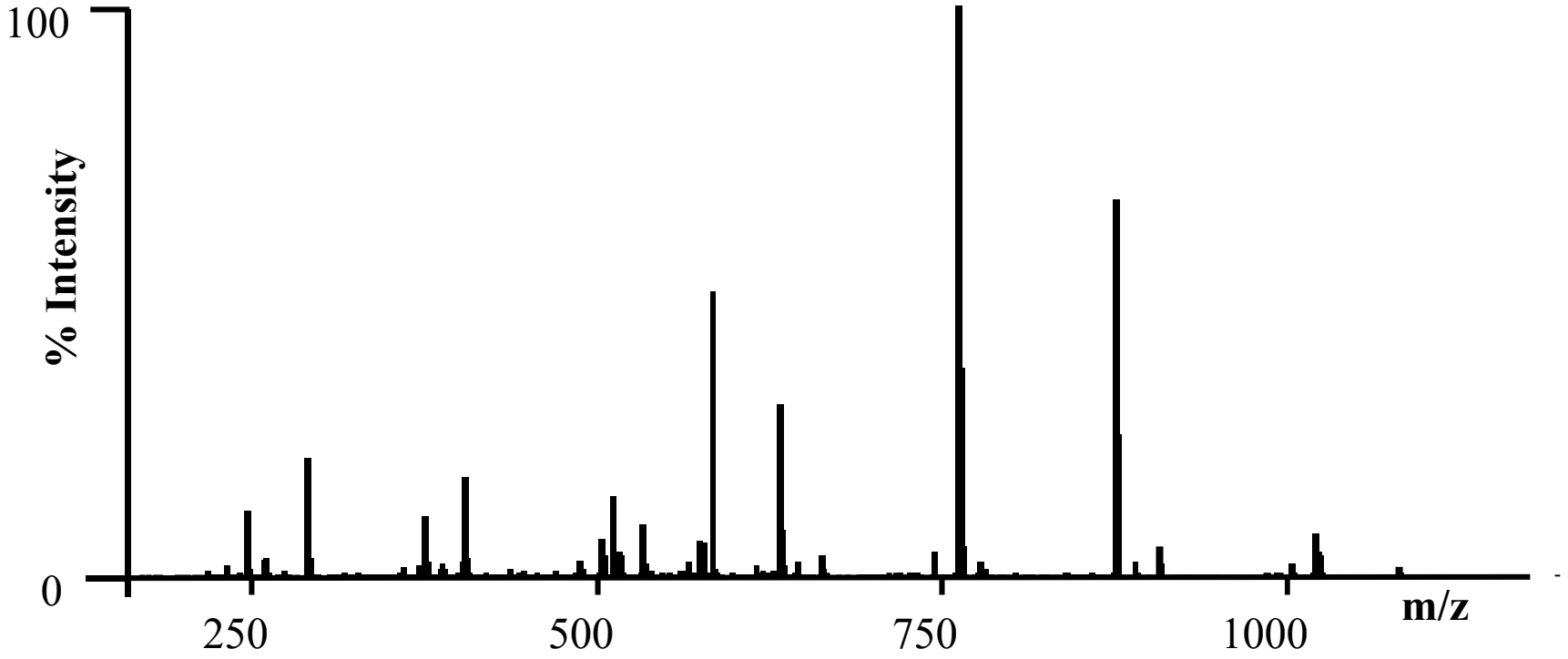
- **Use collision induced spectra**
  - **provides sequence information**
  - **provides unique m/z spectra for each peptide**
- **Problem is large number of CID & large amount of information.**
  - **Need methods for searching**
  - **Filtering has been tried (limited success)**
  - **Most successful with human intervention.**
  - **Improving**
- **Use unique mass of peptides to identify sequence and multiple ions to identify proteins**

Amino Acid Masses

Amino acid	Mass(avg)	Amino acid	Mass(avg)
G	57.0520	D	115.0886
A	71.0788	Q	128.1308
S	87.0782	K	128.1742
P	97.1167	E	129.1155
V	99.1326	M	131.1986
T	101.1051	H	137.1412
C	103.1448	F	147.1766
I	113.1595	R	156.1876
L	113.1595	Y	163.1760
N	114.1039	W	186.2133

Peptide Fragmentation

88	145	292	405	534	663	778	907	1020	1166	b ions
S	G	F	L	E	E	D	E	L	K	
1166	1080	1022	875	762	633	504	389	260	147	y ions



Mass Analyzers (MS)

Quadrapole

- **Sensitive, acceptable mass accuracy and resolution**
- **Easily coupled to chromatography.**

- **Time of Flight**
  - **Sensitive, high mass accuracy, high resolution**
  - **Limited to small m/z ratios**
  - **Not easily coupled to chromatography**
- **Ion Trap**
  - **Sensitive**
  - **Low mass accuracy**
- **Fourier Transform ion cyclotron**
  - **High sensitivity, mass accuracy, resolution, dynamic range**
  - **Expensive, difficult to operate, low fragmentation efficiency**

Proteomics

- **The proteome is larger than the genome due to alternative splicing and protein modification.**

As we have said before we need to know

- **All protein-protein interactions.**
- **Function**
  - **One protein or peptide may have multiple functions depending on context.**
- **Regulation of protein function.**
- **Modification**
- **Location**
  - **Location will help us to understand the proteins role in the cell, what its function is, and what controls its function.**
- **Detection and quantitation**
  - **The concentration of all proteins changes by 10 orders of magnitude within the cell. Currently there are no easy methods for determining the concentrations**

Protein microarrays can also be used for expression analysis

- 2D-electrophoresis is an open system for proteome analysis, rather like direct sequence sampling is an open system for transcriptome analysis.
- Advantage of an open system is that potentially all proteins can be detected
- Disadvantage is that they also have to be characterized, which relies on downstream annotation by mass spectrometry.
- DNA arrays are closed systems in transcriptome analysis, i.e. the data obtained are constrained by the number and nature of sequences immobilized on the array.
- No necessary to characterize any of the features on the array by sequencing because the sequences are already known.
- Similarly, *protein arrays* are emerging as a useful closed system for proteome analysis. These are miniature devices in which proteins, or molecules that recognize proteins, are arrayed on the surface.
- In concept, protein arrays are no different to DNA arrays, but they suffer from several practical limitations.
  - First, the manufacture of DNA arrays is simplified by the availability of methods, such as the polymerase chain reaction, for amplifying any nucleic acid sequence. No amplification procedure exists for proteins.
  - Second, all DNA sequences are made of the same four nucleotides and hence behave

similarly in terms of their chemical properties. The principles of molecular recognition (hybridization between complementary base pairs) apply to all sequences. For this reason, hybridization reactions can be carried out in highly parallel formats using a single complex probe.

- Conversely, proteins are made of 20 amino acids specified by the genetic code plus many others generated by post-translational modification, so they have diverse chemical properties.
  - For example, some proteins are soluble in water while others are lipophilic. Recognition parameters vary widely so the same reaction conditions could never be used for all proteins.
- Third, the homogeneity of DNA molecules means that labels are incorporated evenly and labeling does not interfere with hybridization. Binding to solid substrates such as nylon and glass does not interfere with hybridization either.
- However, the labeling of proteins is much more variable and both labelling and attachment to a substrate could interfere with protein binding, either by affecting the way the protein folds or by blocking the binding site.
- Despite these differences, protein arrays have been manufactured in many of the ways discussed above for DNA arrays, including variations on standard contact printing, inkjetting, and photolithography. The concept of the protein array, protein microarray, or protein chip covers a wide range of different applications. Some of these are……

Antibody arrays contain immobilized antibodies or antibody derivatives for the capture of specific proteins

- Antibodies are attached to the array surface, so the protein array or chip can be thought of as a miniaturized solid-state immunoassay.
- Antibodies interact with specific proteins and are highly discriminatory, so they are suited to the detailed analysis of protein profiles and expression levels.

  eg, using a recombinant Staphylococcal protein A covalently attached to a gold surface. The recombinant protein A has five immunoglobulin G binding domains, allowing antibodies to be attached by the Fc region, therefore exposing the antigen-binding domain.

  Antibody arrays have also been generated by using banks of bacterial strains expressing recombinant antibody molecules.

- Three different formats for this type of assay.
  - A standard immunoassay in which the antibodies are immobilized and are used to capture labelled proteins from solution. Protein expression levels are quantified by measuring the signal (usually fluorescent) which has been incorporated into the proteins. A recent report (Haab *et al.* 2001) describes the use of such an array comprising 115 different antibodies. Another ground breaking aspect of this report was that two protein samples, each labeled with a different fluorophore, were exposed to the array simultaneously, and differential protein expression could be monitored.
  - The second format is a miniature sandwich assay, in which unlabelled proteins are captured from solution and detected with a second, labelled antibody. Although this format requires two antibodies recognizing distinct epitopes for each protein, it is not necessary to label the target population of proteins, a process that is inefficient and variable.

- **The third format involves a tertiary detection system and offers even greater sensitivity.**
  - **One example is the *immunoRCA technique*, which involves rolling-circle amplification. The principle of this technique is that a protein, captured by an immobilized antibody, is recognized by a second antibody in a sandwich assay, but the second antibody has an oligonucleotide covalently attached to it. In the presence of a circular DNA template, DNA polymerase, and the four dNTPs, rolling-circle amplification of the template occurs resulting in a long concatamer comprising hundreds of copies of the circle, which can be detected using a fluorescent-labeled oligonucleotide probe.**

Antigen arrays are used to measure antibodies in solution

- On these devices, protein antigens are attached to the array surface and used for reverse immunoassays, i.e. detecting antibodies in solution. The antigens may be proteins or other molecules such as peptides or carbohydrates. Several reports have been published in which arrays of allergens have been used to screen serum samples for IgE reactivity in allergic responses and autoimmune diseases. In some cases it has proven possible not only to confirm the presence of such antibodies but also to carry out quantitative analysis. More recently, antigen arrays have been used to serodiagnose patients with viral infections.

General protein arrays can be used for expression profiling and functional analysis

- Any type of protein is used to assay protein–protein interactions and protein interactions with other molecules. A range of detection strategies may be used, including labelling of the interacting molecules, or label-independent methods such as surface plasmon resonance. For example, Ge *et al.* (2000) developed a system for studying molecular interactions using a universal protein array (UPA) system where protein samples are transferred from 96-well microtiter plates to nylon membranes.
- The technology has also been applied to the arraying of cDNA expression libraries, such that screening can be carried out not only with nucleic acid probes but also with antibodies or other ligands directed at the recombinant proteins. One of the most impressive demonstrations of the power of protein array technology was provided by MacBeath & Schreiber (2000). They used an array of proteins on a glass slide to screen for ligands, enzyme substrates, and protein–protein interactions. Functional assays can also be carried out on such arrays. For example, Zhu *et al.* (2000) produced an array containing nearly all the protein kinases of the yeast proteome and carried out kinase assays on 17 different substrates. More notably, the same group has also produced a glass microarray containing nearly all the proteins in the yeast proteome (5800 spots) and used this to screen for various functions such as phospholipid binding and interactions with calmodulin (Zhu *et al.* 2001).

Other molecules may be arrayed instead of proteins

- These are not protein arrays in the strict sense because they do not consist of arrayed proteins. They are considered here because they are analogous to antibody arrays, i.e. they contain specific capture agents that interact with proteins.
- DNA arrays fall within this class if they are used to analyze DNA–protein interactions. Aptamers, single-stranded nucleotides that interact specifically with proteins, could also be used in this manner.
- Another area of active current research is the development of chips containing artificial recognition sites for proteins. The concept of *molecularly imprinted polymers* (MIPs) depends on the ability to imprint (block) a polymeric substrate with recognizable molecular imprints that mimic the actual recognition molecules.

Some biochips bind to particular classes of protein

- Instead of specific molecular interactions, these devices use broad-specificity capture agents. As above, they can be termed protein chips but not protein arrays. For example, Ciphergen Biosystems Inc. market a range of ProteinChips with various surface chemistries to bind different classes of proteins.
- Although relatively non-specific compared to antibodies, complex mixtures of proteins can be simplified and then analyzed by mass spectrometry. An
- advantageous feature of this system is the ease with which it is integrated with downstream MS analysis, since the ProteinChip itself doubles as a modified MALDI plate (Fung *et al.* 2001, Weinberger *et al.* 2001). After the chip has been washed to remove unbound proteins, it is coated with a matrix solution and analyzed by time-of-flight MS. This allows surface-enhanced laser desorption and ionization (SELDI), which provides more uniform mass spectra than MALDI and allows protein quantification.
- Other protein chip platforms use surface plasmon resonance to detect and quantify protein binding. This involves measuring changes in the refractive index of the chip surface caused by increases in mass (Malmqvist & Karlsson 1997). Protein chips produced by the US company BIAcore are based on this concept, and other chips combine surface plasmon resonance measurements with mass spectrometry