Proteomics

Proteomics is the study of all proteins within organism.

Challenges

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

   As we have said before we need to know
2. All protein-protein interactions.
3. One protein or peptide may have multiple functions depending on context.
4. Regulation of protein function.
5. Modification
6. Location
   Location will help us to understand the proteins role in the cell, what its function is, and what controls its function.
7. 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 over this large of a concentration range.

What areas are being studied in Proteomics?

The areas being studied under the name of proteomics is changing and often growing as the field develops. I will present a few of the areas in this class.

Some areas in proteomics are:

1. Mass spectrometer based proteomics
    This area is most commonly associated with proteomics and is a method to determine which proteins are expressed and the amounts of those proteins.

2. Array based proteomics
    This method uses various arrays to try and define the function of the proteins, regulation levels and interacting partners within the cell.

3. Informatics
    This is trying to define what information will be needed, stored, accessed and how it can be used to study the proteome of the cell.

4. Clinical
5. Orthagonalomics

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.

Let's begin by discussing various means for separating proteins to prepare them for mass spectrometry and then discuss mass spectrometry and how it is used.
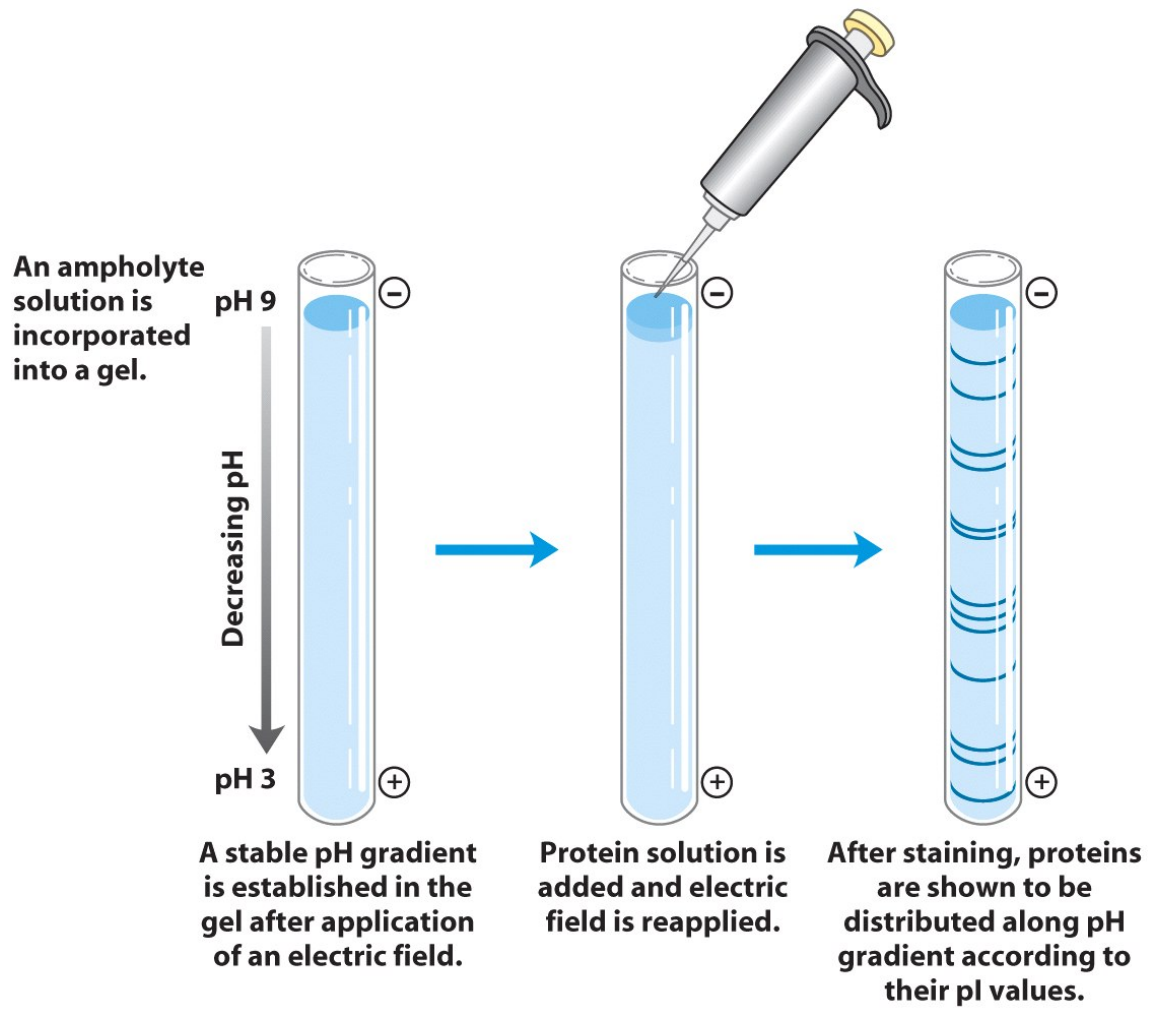
Separating the genome

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. First I will explain the individual methods that are coupled to create the two dimensions.
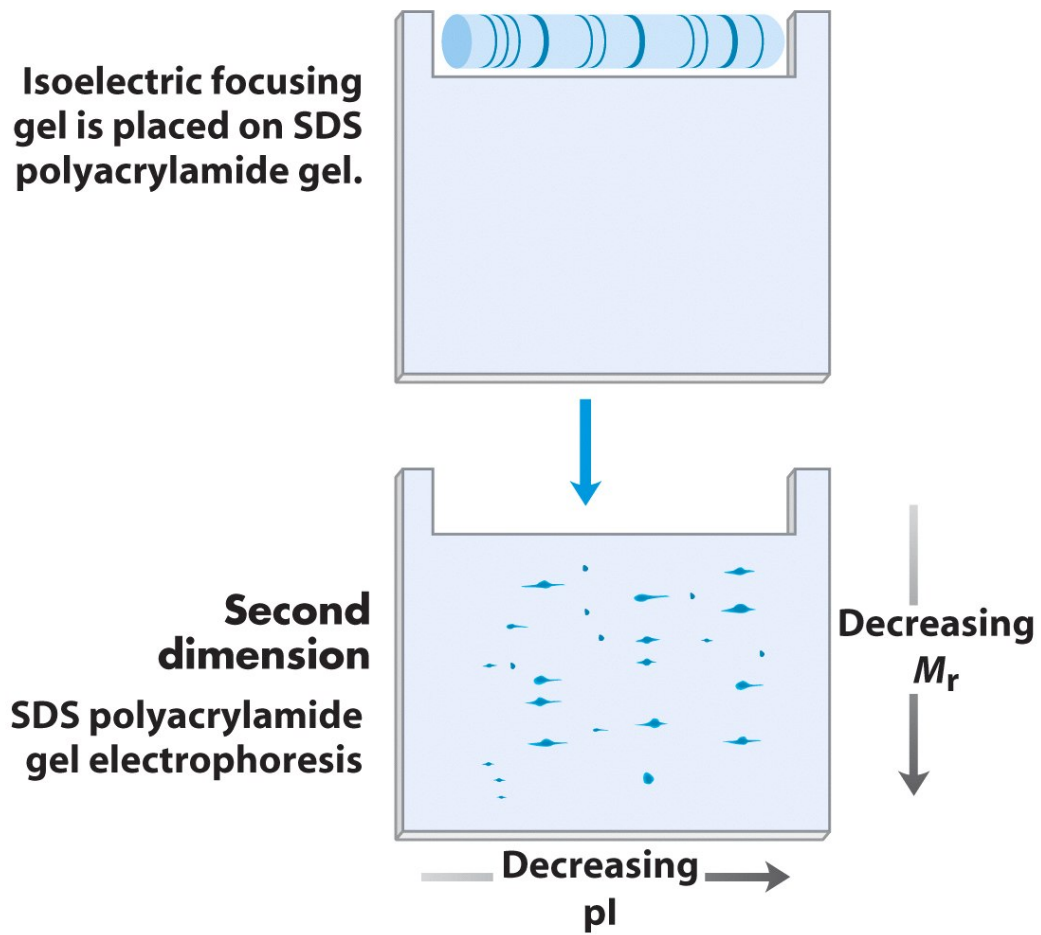
Iso-electric focusing

- This separates proteins based on isoelectric point

- The isoelectric point is the pH at which the protein has no net charge.

- Typically this is done with a tube gel containing a low concentration of poly-acrylamide.

- Ampholytes are added to create a pH gradient in an electric field and the proteins are loaded.

- The tube gel is placed in and electrophoresis system for up to 24 hours and the proteins form tight bands at their iso-electric point.

- The tube gels are now ready for the second method.

An ampholyte solution is incorporated into a gel.

pH 9

Decreasing pH

pH 3

A stable pH gradient is established in the gel after application of an electric field.

Protein solution is added and electric field is reapplied.

After staining, proteins are shown to be distributed along pH gradient according to their pI values.

Figure is from Lehninger fourth edition "Principles of Biochemistry"

- SDS Poly-Acrylamide 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 that is used to separate the proteins.

- The tube 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 tube 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.

- 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 include trypsin digestion

Isoelectric focusing gel is placed on SDS polyacrylamide gel.

Second dimension

SDS polyacrylamide gel electrophoresis

Decreasing M_r

Decreasing pI

Figure is from Lehninger fourth edition "Principles of Biochemistry"

- Figure is from Lehninger fourth edition "Principles of Biochemistry"

Alternate methods of separation.

Recently there has been a change in the way people are approaching separating the proteins.

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.

The diffused 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.

Mass Spectrometery

1. Separates ions based on mass to charge ratio.
    Charges are placed on the protein or the
    peptide by ionization.

2. Two most common types of ionization are:
    a. 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.
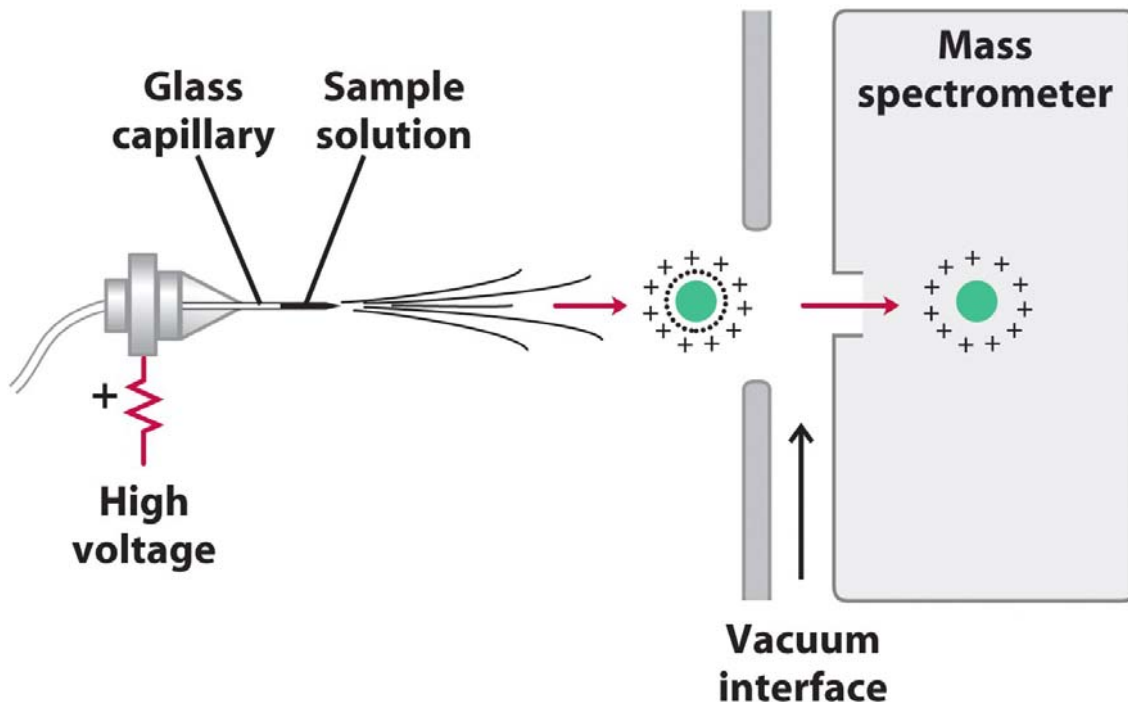
-   Figure is from Lehninger fourth edition "Principles of Biochemistry"

b. 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.

3. Important parameters
    a. sensitivity
    b. resolution
    c. mass accuracy

4. Mass Analyzers
    d. Quadrapole
        i.   Sensitive, acceptable mass accuracy and resolution
        ii. Easily coupled to chromatography.
    e. Time of Flight
        i.   Sensitive, high mass accuracy, high resolution
        ii. Limited to small m/z ratios
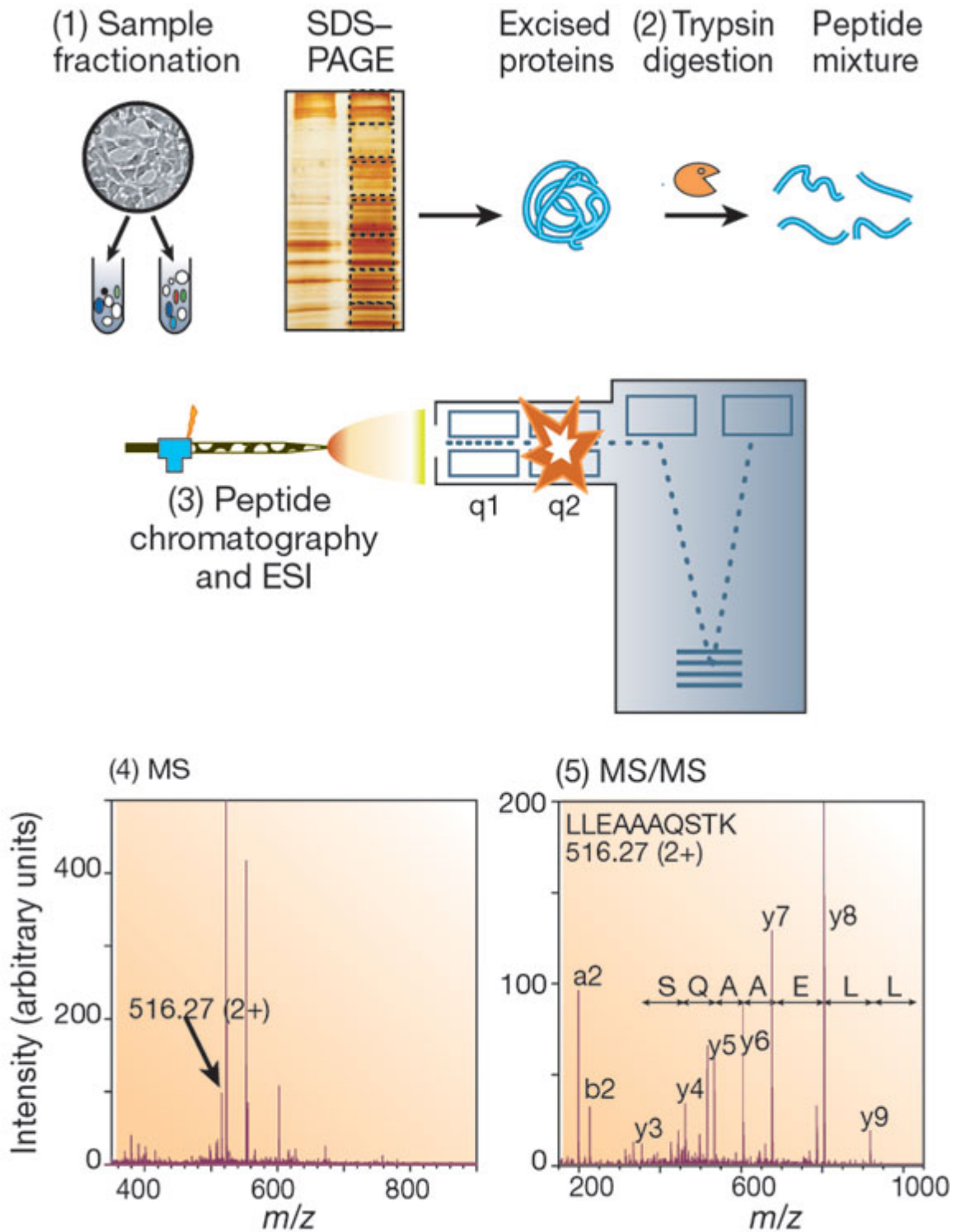   iii. Not easily coupled to chromatography
    f. Ion Trap
        i.   Sensitive
        ii. Low mass accuracy
    g. Fourier Transform ion cyclotron
        i.   High sensitivity, mass accuracy, resolution, dynamic range
        ii. Expensive, difficult to operate, low fragmentation efficiency

4. Instruments are often coupled
   a. MS/MS
   b. Collider - forms collision induced spectra (CID)

(1) Sample fractionation SDS–PAGE Excised proteins (2) Trypsin digestion Peptide mixture

(3) Peptide chromatography and ESI

q1   q2

(4) MS

516.27 (2+)

Intensity (arbitrary units)

400   600   800
m/z

(5) MS/MS

LLEAAAQSTK
516.27 (2+)

y7   y8

a2   S Q A A E L L

y5 y6

b2   y4

y3   y9

200   600   1000
m/z

Protein identification and Quatitation

1. Protein separation by 2 dimensional electrophoresis or similar methods
2. Limited protein purification and automated MS.

Limited purification and MS
  1. multiple chromatography stages
    a. ion exchange, reverse phase, affinity
    b. electrophoresis

  2. To quantitate
    a. add stable isotopes
    b. post separation modification
      i. SH, NH_2, N-linked carbohydrates
    c. incorporation of isotopes in culture

Protein Identification

1. Use collision induced spectra
    i. provides sequence information
    ii. provides unique m/z spectra for each peptide

2. Problem is large number of CID information
    i. need methods for searching
    ii. Filtering has been tried (limited success)
    iii. Most successful with human intervention.
    iv. Improving