CHAPTER II

THEORETICAL AND LITERATURE REVIEWS

2.1 Leguminosae

Leguminosae is one of the largest families of flowering plants with 18,000 species classified into around 650 genera. This is usually divided into three sub-families: Papilionoideae, Caesalpinioideae and Mimosoideae, on the basis of flower morphology (specifically, petal shape) [22,23].

- Papilionoideae, previously called Faboideae (Fabaceae): One petal is large and has a crease in it, the two adjacent petals are on the sides, and the two bottom petals are joined together at the bottom, forming a boat-like structure.
- Caesalpinioideae (Caesalpiniaceae): The flowers are irregular (zygomorphic) with five petals which are not differentiated into standard, wings and keel. The stamens are visible externally.
- Mimosoideae (Mimosaceae): The petals are small, regular (actinomorphic) flowers crowded together, and frequently globosely. The stamens are the showiest part of the flower.

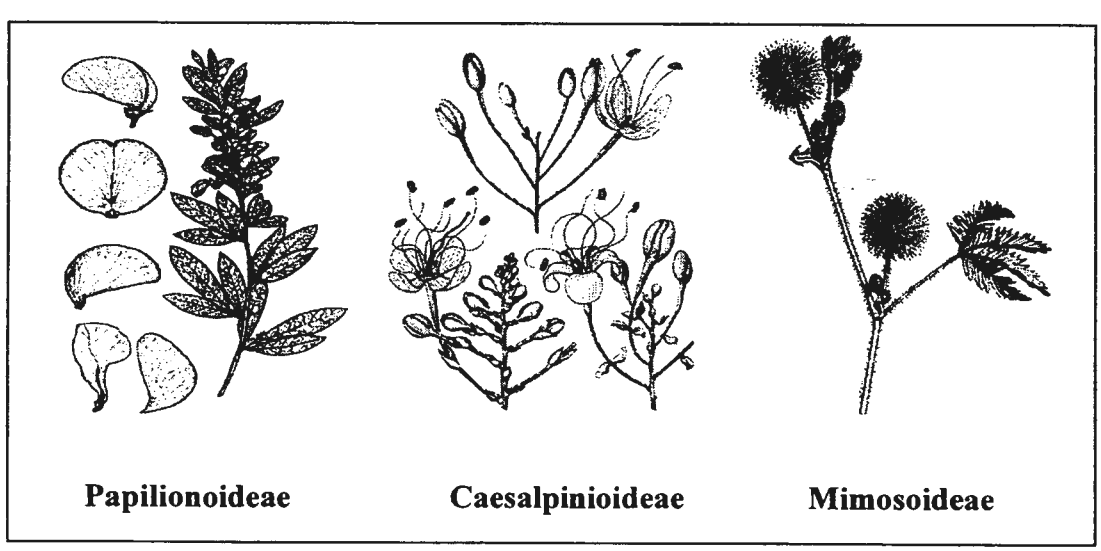


Figure 2.1 Flower morphology of three sub-families belongs to Leguminosae.

Legumes are plants of the pea or bean family, used as crops, forages and green manures. They also synthesize wide varieties of natural products such as flavors, drugs, poisons and dyes [2]. Leguminous plants have formed a popular subject of research owing to the abundance of proteins and polypeptides with important biological activities that they elaborate.

2.2 *Sesbania grandiflora* (L.) Desv.

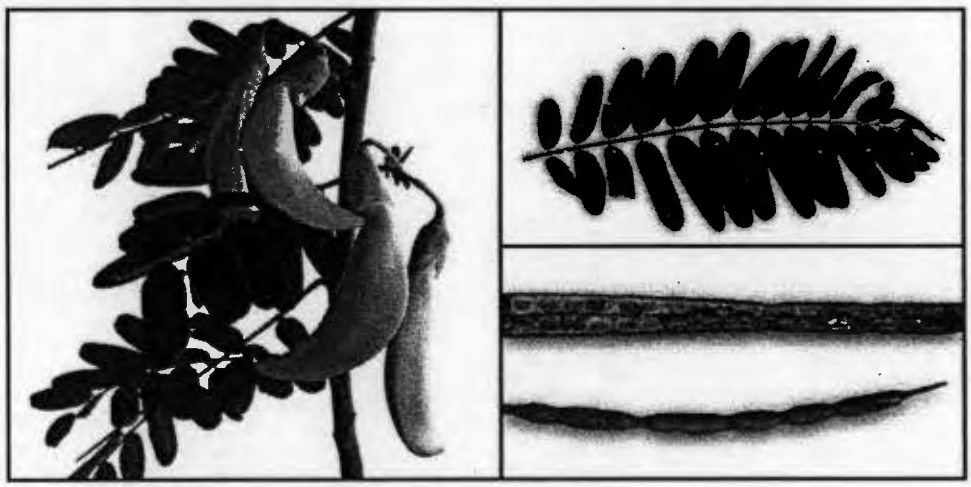


Figure 2.2 *Sesbania grandiflora* flowers, leaves, pods and seeds.

2.2.1 General Background

Sesbania grandiflora (Khae baan) is a small tree in subfamily Papilionoideae belongs to the Leguminosae. It is believed to have originated either in India or Southeast Asia and grows primarily in hot and humid areas of the world. Its native is in Asian countries such as India, Malaysia, Indonesia, Philippines, and Thailand, where it is commonly seen growing on the dikes between rice paddies, along roadsides and in backyard vegetable gardens. This soft-wood tree's leaves are used as fodder and its flowers as food.

2.2.2 Morphology description

Sesbania grandiflora is a small soft-wooded tree to 10 metres in height. Its leaves are pinnate, 15-30 cm long, with 16-30 leaflets in pairs. A dimension of leaves is 12-44 × 5-15 mm, oblong to elliptical in shape. The white or pink-red flowers of this papilionaceous (pea-like flowered) legume are unusually large, with a calyx 15-22 mm long. The standard has dimensions up to 10.5 × 6 cm. Pods are long (20-60 cm) and thin (6-9 mm) with broad sutures containing 15-50 seeds. The seed are tan to red-brown, 6-8 × 3-5 mm, 14-20 weighing 1 g. Bark light gray, corky, deeply furrowed [12,13].

2.2.3 Chemical constituents

Leaves on a zero moisture basis (ZMB) contain per 100 g, 321 calories, 36.3 g protein, 7.5 g fat, 47.1 g carbohydrate, 9.2 g fiber, 9.2 g ash, 1684 mg Ca, 258 mg P, 21 mg Na, 2,005 mg K, 25,679 mg β-carotene equivalent, 1.00 mg thiamine, 1.04 mg riboflavin, 9.17 mg niacin and 242 mg ascorbic acid.

Flowers (ZMB) contain per 100 g, 345 calories, 14.5 g protein, 3.6 g fat, 77.3 g carbohydrate, 10.9 g fiber, 4.5 g ash, 145 mg Ca, 290 mg P, 5.4 mg Fe, 291 mg Na, 1,400 mg K, 636 mg β-carotene equivalent, 0.91 mg thiamine, 0.72 mg riboflavin, 14.54 mg niacin, and 473 mg ascorbic acid. There are several hydrocarbon and alcohol in ethanol extract of this flower such as n-tricosane, n-tetracosane, n-pentacosane, n-hexacosane, n-heptacosane, n-octacosane and montanyl alcohol.

Seeds (ZMB) contain 36.5% CP, 7.4% fat, 51.6% total carbohydrate, and 4.5% ash. The seed oil contains 12.3% palmitic, 5.2% stearic, 26.2% oleic, and 53.4% linoleic acids. The seed testa, which constitutes 20% of the seed, contains 5.2% moisture, 1.3% ash, 0.8% fat, 2.7% CF, 0.1% free reducing sugars, 1.4% sucrose, 2.8% nitrogen, 6.3% pentosans, and 65.4% carbohydrates. Yields of 33% galactomannans are reported for alkali extraction of the testae [24, 25].

2.2.4 Uses and applications

Because of leaves of *Sesbania grandiflora* contain as much as 25-30 percent crude protein so *Sesbania grandiflora* is valued as a fodder, particularly for feeding

cattle and goats, in many regions. Although ruminants readily consume *Sesbania grandiflora* fodder, and its digestibility is high, some feeding studies have indicated that antinutritional factors are present. *Sesbania grandiflora* should be used in feeding to ruminants and should not be fed to chickens or other monogastric animals because its leaf is toxic to them. For Soil improvement, *Sesbania grandiflora* is often maintained in gardens and around crop fields for its contribution of nitrogen. Falling leaflets and flowers recycle nutrients to the ground. Seedlings grow rapidly enough that they have been used similarly to annual green manure crops. The light density wood of *Sesbania grandiflora* makes poor firewood and is not durable as a timber; however it can be used for low quality pulp. The leaves, seed pods and flowers are used as human food in Southeast Asia. The young, tender pods are cooked similarly to other green beans. Unopened white flowers are a common vegetable, steamed or cooked in soups and stews after the stamen and calyx have been removed. The dried leaves of *Sesbania grandiflora* were used in some countries as a tea which is considered to have antibiotic, anthelmintic, anti-tumour and contraceptive properties.

2.2.5 Medicinal application

Sesbania grandiflora has many medicinal uses, e.g. for treating catarrh, headache and epilepsy. In folk medicine, *Sesbania grandiflora* is resorted to be aperient, diuretic, emetic, emmenagogue, febrifuge, laxative, and tonic. Bark, leaves, gums, and flowers are considered medicinal. The astringent bark was used in treating smallpox and other eruptive fevers. The juice from the flowers is used to treat headache, head congestion, or stuffy nose. As a snuff, the juice is supposed to clear the nasal sinuses. Leaves are poulticed onto bruises. There are many regions used *Sesbania grandiflora* as folk medicine, e.g. In Malaysia, the crushed leaves are applied to relieve sprains and bruises, and the bark used as astringestin. It is also used to treat scabies. In Indochina the bark is considered a tonic and used to prevent fever and to treat colic, diarrhoea and dysentery. In the Philippines the bark is used to treat spitting or coughing of blood. Indians apply the roots in rheumatism, the juice of the leaves and flowers for headache and nasal catarrh. Mixed with stramonium and pasted, the root is poulticed onto painful swellings. In Amboina, flower juice is squeezed into the eye to correct dim vision. Cambodians consider the flowers

emollient and laxative, the bark for diarrhea and dysentery. In Java, the bark is used for thrush and infantile disorders of the stomach. An aqueous extract of bark is said to be toxic to cockroaches. Leaves are chewed to disinfect the mouth and throat.

2.3 Literature Reviews

Leguminous plants have formed a popular subject of research owing to the abundance of proteins and polypeptides with important biological activities that they elaborate. Examples of these proteins are lectins, glycoprotein from non-immune origin that is able to agglutinate cells [26]. There are many application of lectins such as biological tools for study of carbohydrates in solution and on cell surface, for identification and separation of cells, using lectins as histochemical and cytochemical probes, application in crop protection or in tumor therapy by immunomodulation, using lectin as mediated drug targeting [27]. In the plant kingdom, seed of legumes have long been known to be a rich source of lectins constituting up to 10% of their total protein [28]. Among the best-characterized seed legume lectins are those isolated from jack bean, soybean, pea, fava bean, lentil, lima bean and red kidney bean. For example of legume lectins, *Cratylia mollis* lectin [10] (Cra) belongs to the mannose/glucose-binding class of lectins has been purified and characterized from the seeds of *Camaratu bean* native to the northeast of Brazil. Antitumor activity of these Cra-loaded liposomes was investigated against mice by Andrade et al. group and the results showed that Cra-loaded liposomes were produced with a high percentage encapsulation ratio and a high tumor inhibition was achieved. In 2001, Ye *et al.* [9] purified a lectin protein with antifungal activity from the seeds of the red kidney bean, *Phaseolus vulgaris,* this lectin manifested inhibitory activity on human immunodeficiency virus-1 reverse transcriptase and α-glucosidase. In same year, Ye and Ng report that a lectin protein from sugar snap [29], *Pisum sativum*, demonstrate high hemagglutinating activity with rabbit erythrocytes.

Moreover, lectin proteins, antifungal proteins are known to be elaborated by leguminous species. Recently a protein, Dolichin, a chitinase-like antifungal protein, was isolated from Field Beans (*Dolichos lablab*) by Ye *et al.* [30] This protein also inhibits human immunodeficiency virus (HIV) reverse transcriptase, α- and β-glucosidase which are glycohydrolase implicated in HIV infection. In 2001 Ye and

Ng reported that Unguilin [11], isolated from Black-Eyed pea (*Vigna unguiculata*), was a cyclophilin-like protein with anti-mitogenic, antiviral, and antifungal activities. In 2002, this research group reported that a novel antifungal protein, against a variety of fungal species including *Coprinus comatus*, *Mycosphaerella arachidicola*, *Fusarium oxysporum* and *Botrytis cinerea*, was isolated from French bean legumes [8]. In 2004, Kelemu *et al.* [31] examined seeds from several tropical forage legumes and found that extracts from seeds of *Clitoria ternatea* (L.) exhibited, in vitro, strong antifungal activity on the test fungus *Rhizoctonia solani* Kühn, which this inhibits is an effect from a protein designated Finotin.

Furthermore, lectin and antifungal proteins, pea seeds also contain protease inhibitors. For example, Bowman-Birk protease inhibitor [32] from soybeans has showed anticarcinogenic properties; it has been suggested that a ca. 30 mg per food serving of chymotrypsin inhibitor is expected to protect against several forms of human cancer. In 2004, Vargas *et al.* reported that a Kunitz-type protein inhibitor (trypsin inhibitor) obtained from *Pithecellobium dulce* seeds [33], Leguminosae tree native to Mexico.

In this research, *Sesbania grandiflora* was chosen to study because of this plant is in Leguminosae family which should have high amount of protein same as other plant in this family. For literature of *Sesbania grandiflora* were presented as follow, In 1969, Khanna, S. S. and Perkins, E. G. used gas-liquid chromatography mass spectrometry to identified the chain hydrocarbons in ethanolic extract from the flowers of *Sesbania grandflora* and found that the homologous series of hydrocarbons (C_{23}-C_{31}) were identified as the alcohol n-nonacosanol [34] . In 1982, Kumar *et al.* reported that aqueous extract of the leaves of *Sesbania grandiflora* produced haemolysis of human and sheep erythrocytes even at very low concentrations [35]. Haemolysis was greater when the pH was acidic. The liberation of phospholipids and sterols into the supernatant as a result of haemolysis indicated possible damage to the erythrocyte membrane. The methyl ester of oleanolic acid has been isolated from the flowers of *Sesbania grandiflora* and has shown to have haemolytic effects on sheep and human erythrocytes too [36]. In Thailand, during the outbreak of epizootic ulcerative syndrome (EUS) in 1983 the snake-head fish farmers in Uthaitanie, used the bark of cork wood tree (*Sesbania grandiflora*) for the treatment of haemorrhage

lesions [37]. Most of the fish recovered after treatment. In 1990, Anderson and Weiping studied gum exudates from *Sesbania grandiflora* which found that their gum [38] are strongly dextrorotatory, acidic arabinogalactans and give strongly coloured solution of low viscosity comparable to gum talha *(Acacia seyal).* In 2002, Kasture *et al.* have evaluated the anticonvulsive activity of *Sesbania grandiflora* leaves using a variety of animal models of convulsions. They found that the benzene:ethyl acetate fraction containing triterpene exhibits a wide spectrum of anticonvulsant profile and anxiolytic activity [39]. In 2003, Tewtrakul *et al.* screened for HIV-1 protease (HIV-1 PR) inhibitory activities fifty-two ethanol and water extracts of the plants in Caesalpiniaceae and Papilionaceae families using high performance liquid chromatography (HPLC) technique [40]. Among the tested extracts, *Sesbania grandiflora* flower exhibited the weak activity by 17.55 % inhibition (EtOH extract) and 9.77% inhibition (water extract) at concentration of 100 μg/ml. In addition in this same year, Nakahara *et al.* examined the antimutagenic activity against Trp-P-1 of methanolic extracts of 118 samples of edible Thai plants [41]. The activity was evaluated by the amount of plant extracts which suppressed 90% of the mutagenesis (ED_{90}) exhibited significant activity with antimutagenic. Among to *Sesbania grandiflora* flower and leaves showed higher ED_{90} values than 50 μl/plate. In 2003, Pari and Uma reported that oral administration of an ethanolic extract of *Sesbania grandiflora* leaves produced significant hepatoprotection against erythromycin estolate - induced hepatotoxicity in rats [42]. In 2005, Krasaekoopt and Kongkarnchanatip studied anti-microbial properties of Thai traditional flower vegetables in water extracts, *Sesbania grandiflora, Senna siamea and Telosma minor* and found that the crude extracted solution of *Sesbania grandiflora* and *Senna siamea* flowers had higher microbial inhibition than that of *Telosma* flower due to higher flavonoid contents [43].

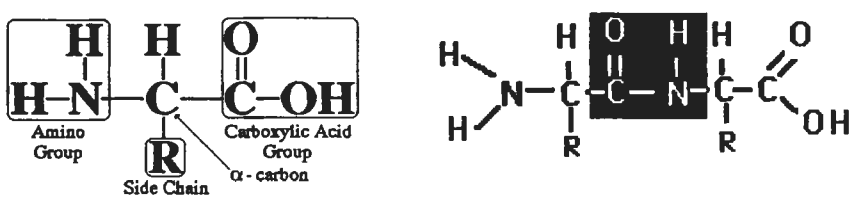
Sesbania grandiflora is potentially useful forage sources [15-19], leaves and pods are reported to be palatable non toxic for cattle. Some other reports suggest that the white flowering variety is non-toxic, while the purple flowering type is toxic. Dried leaves of *Sesbania grandiflora* have been fed (20% of diet) to milking cows and 15% of diet without detrimental effects. *Sesbania grandiflora* contains condensed

tannin precursors (cyanidins) in leaves, whilst no tannin can be detected in of the degradable matter in the rumen.

Almost literatures of *Sesbania grandiflora* reported a few literatures [20,21] about proteomic field in this plant. One of this reports analysis of the plastid gene of maturase K protein in chloroplast of *Sesbania grandiflora* [20]. Led to the idea of this research which concern with proteins of *Sesbania grandiflora* flower which expect to found bioactive proteins same as the proteins isolated from other leguminous plant.

2.4 An introduction to protein

Proteins are macromolecules composed of one or more polypeptides, which are made up of many amino acids linked together as a linear chain. The structure of an amino acid contains an amino group (NH_3^+), a carboxyl group (COO^-), and a R group (side chain) which is usually carbon based and gives the amino acid its specific properties.



Amino acid structure Peptide bond

Figure 2.3 The structure of an amino acid and peptide bond

The amino acids in a protein are held together with peptide bonds. Peptide bonds are bonds between the carboxyl group of one amino acid and the amino group of the next. When three or more amino acids are joined together they are sometimes referred to as a polypeptide chain. Proteins are assembled in four structural levels. The primary structure is the sequence of amino acids, the secondary structure is the formation of α–helices, β–sheets, reverse turns and hairpin loops. The folding of the secondary structure is determined as the tertiary structure and the quaternary structure is the interaction of several subunit proteins, or prosthetic groups, to form a larger protein with a specific function [44].

primary secondary tertiary quaternary

Figure 2.4 The structural levels of protein

2.5 Proteins separations techniques

2.5.1 Protein extraction

The first step in protein purification [44-46] involves a cell disruption step. The method of choice depends on the type of cell. In general, animal cells are easier to disrupt than bacteria, yeast or plant cells. The table below summarizes some of the methods.

Table 2.1 Method for cell disruption

Cell Type	Method	Comment
Bacteria Plant cells	French press	Shearing forces disrupt cell wall as the cells are forced through a small opening under very high pressure. Not practical for large volumes.
Bacteria	Sonication Bead Mill	Disruption of cell walls by shearing and cavitation. Cell wall sheared through abrasion with glass beads.
Animal Cells	Blender	Homogenization of tissue or cells will disrupt cell walls.
Plant Cells Bacteria Spores	Blender	Glass beads are used to disrupt some bacteria, plant cells and bacterial and fungal spores.
Bacteri	Lysis	Solubilization of cell membranes by treatment with lysozyme and EDTA; Grampositive bacteria are more susceptible than Gram-negative bacteria.
Bacteria Yeasts	Lysis	Solubilization of cell membranes with an organic solvent such as toluene.

2.5.2 Protein precipitation

Unlike most other biological organism, the bulk of the solid mass of plant material is not protein but other macromolecules such as polysaccharide and phenolic polymers. Furthermore, the initial homogenization has to be carried out in larger volumes of buffer to allow initial filtration which would be absorption back onto the large mass of insoluble material. As a consequence, there is a resultant initial low protein concentration which necessitates a precipitation step or use of a concentrator.

For the principle of protein precipitation, Proteins are usually soluble in water solutions because they have hydrophilic amino acids on their surfaces that attract water molecules and interact with them. This solubility is a function of the ionic strength and pH of the solution. Proteins have isoelectric points (pI) at which the charges of their amino acid side groups balance each other. If the ionic strength of a solution is either very high or very low proteins will tend to precipitate at their isoelectric point. The most common type of precipitation for proteins is salt induced precipitation. Protein solubility depends on several factors. It is observed that at low concentration of the salt, solubility of the proteins usually increases slightly. This is termed *Salting in*. But at high concentrations of salt, the solubility of the proteins drops sharply. This is termed *Salting out* and the proteins precipitate out, illustrate in Figure 2.5.

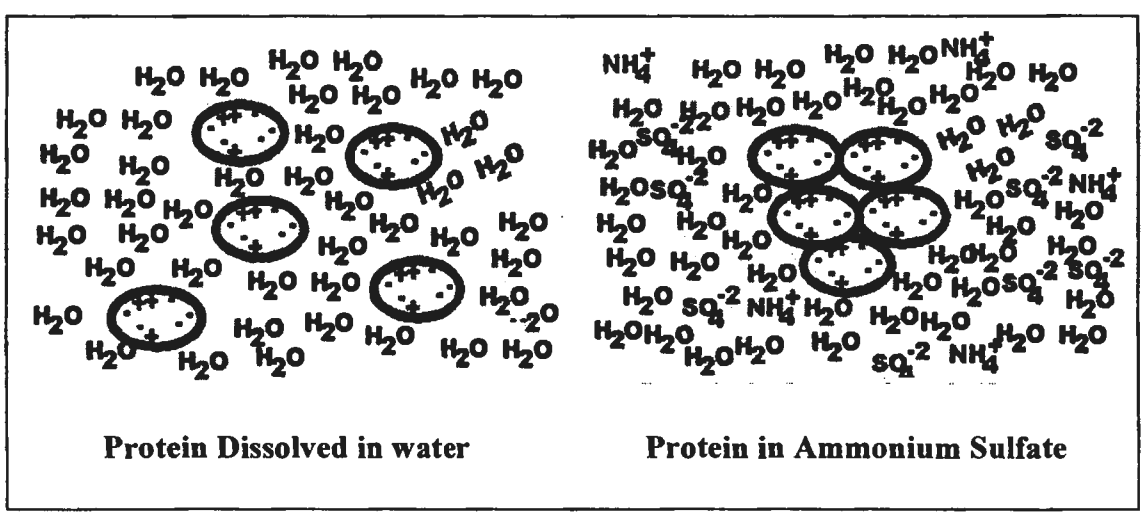


Figure 2.5 Protein dissolving in water when absent salt and increasing salt

Another method is the addition of an organic solvent. If there is a medium decrease in the dielectric constant with the addition of an organic solvent, the the solubility should decrease also. A third method is precipitation by changing the pH of the protein solution. This effect is due to the different functional groups on a protein. There will be some pH, known as the isoeletric point where the net charge on the protein is zero. This is different for different proteins.

2.5.3 Dialysis of Proteins

After a protein has been ammonium sulfate precipitate and taken back up in buffer at a much greater protein concentration than before precipitation, the solution will contain a lot of residual ammonium sulfate which was bound to the protein. One way to remove this excess salt is to dialysis the protein against a buffer low in salt concentration.

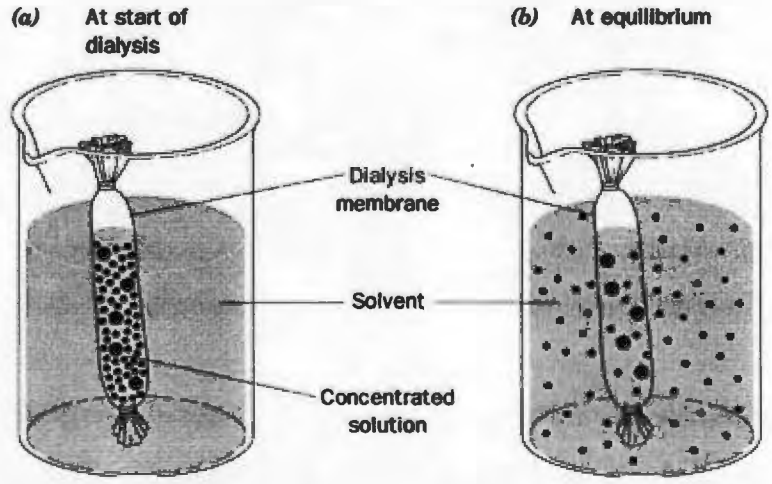


Figure 2.6 The dialysis at start and at equilibrium process.

The dialysis process illustrates in Figure 2.6. First, the concentrated protein solution is placed in dialysis bag with small holes which allow water and salt to pass out of the bag while protein is retained. Next the dialysis bag is placed in a large volume of buffer and stirred for many hours, which allows the solution inside the bag to equilibrate with the solution outside the bag with respect to salt concentration. When this process of equilibration is repeated several times, the protein solution in the bag will reach a low salt concentration.

2.5.4 Electrophoresis

2.5.4.1 Theory of Electrophoresis

Electrophoresis is the movement of an electrically charged substance under the influence of an electric field. This movement is due to the Lorentz force, which may be related to fundamental electrical properties of the body under study and the ambient electrical conditions. The movement of a molecule in an electric field is a function of the net charge q and the frictional coefficient f. The velocity v of a charged molecule in a field is a function of the field strength E by the equation (1) given below.

$$v = \frac{Eq}{f} \quad \text{..... (1)}$$

$$\mu = \frac{v}{E} = \frac{q}{f} \quad \text{..... (2)}$$

In the equation (2), the ratio of velocity to field strength is defined as mobility μ, which equals the charge (q) divided by the frictional coefficient (f). The frictional coefficient is a measure of the hydrodynamic size of molecule. Thus for molecular size and shape are important in defining frictional coefficient. Therefore, the mobility of a molecule is a function of size, sharp, and charge.

2.5.4.2 Gel Electrophoresis

Gel electrophoresis is a technique in which charged molecules, such as protein or DNA, are separated according to physical properties as they are forced through a gel by an electrical current. Smaller molecules can be separated only in Sephadex-type gels. Sephadex-type gels have small pores that exclude larger molecules from access to the stationary phase inside the particle, thereby causing movement outside the pores whereas small molecules are tightly held within the pores. Gels may be prepared from starch, agarose and polyacrylamide.

Polyacrylamide gel are made from acrylamide monomers copolymerized with the cross linker N,N'methylenebisacrylamide in the presence of ammonium persulphate and TEMED (N,N,N',N'-tetramethylethylenediamine) as catalyst. The size of the pores created in the gel is inversely related to the amount of acrylamide

used. Gels with a low percentage of acrylamide are typically used to resolve large proteins and high percentage gels are used to resolve small proteins.

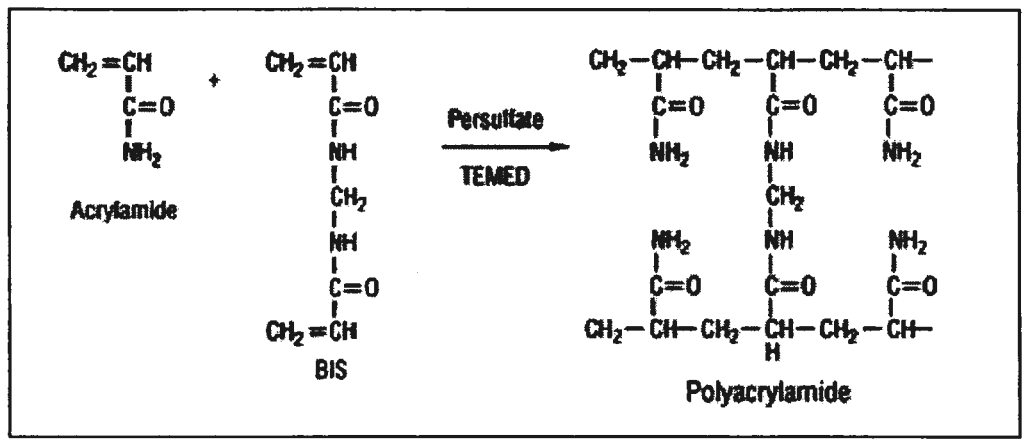


Figure 2.7 Scheme of polymerization and cross-linking of acrylamide.

The size pores in a polyacrylamide gel is determined by two parameters: the total amount of acrylamide present (%T) and the ratio of cross-linker to acrylamide monomer (%C) (Figure 2.8). The %T is the ratio of the sum weights belong to acrylamide monomer and the cross-linker in the solution. As the total amount of acrylamide (%T) increases, the pore size decreases. With cross- linking, 5%C gives the smallest pore size. Any increase or decrease in %C increases the pore size.

$$\%T = \frac{g(\text{acrylamide} + \text{bisacrylamide})}{100\ ml} \times 100$$

$$\%C = \frac{g(\text{bisacrylamide})}{g(\text{acrylamide} + \text{bisacrylamide})} \times 100$$

Figure 2.8 Determination of %T and %C for acrylamide gels.

Proteins are commonly separated using polyacrylamide gel electrophoresis (PAGE) [47,48] to characterize individual proteins in a complex sample, which can be used as a tool to obtain a pure protein sample, or as an analytical tool to provide information on the mass, charge, purity or presence of a protein. Several forms of PAGE can be show below.

SDS-PAGE or Sodium dodecyl sulfate polyacrylamide gel electrophoresis is the method which separates based on the mass of the molecule. Sodium dodecyl sulfate (SDS) is a detergent that breaks up the interactions between proteins. The proteins are dissolved in SDS and reducing agents such as dithiothreitol (DTT), used to remove the remaining tertiary and quaternary structures by reducing disulfide bonds (Figure 2.9 and 9). The SDS-denatured and reduced proteins are flexible rods with uniform negative charge per unit length. Thus, because molecular weight is essentially a linear function of peptide chain length, in sieving gels the proteins separate by molecular weight.The smallest molecules move through the fastest, while larger molecules take longer and result in bands closer to the top of the gel.

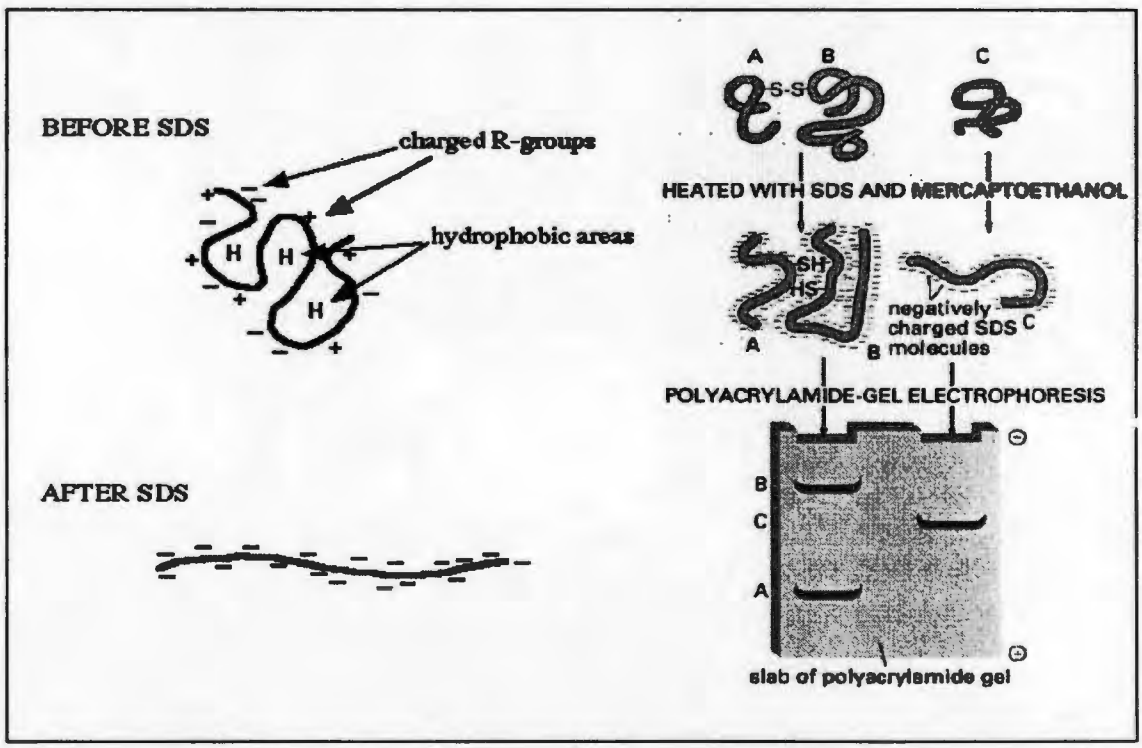


Figure 2.9 Mechanism of SDS to breaks up the interactions between proteins

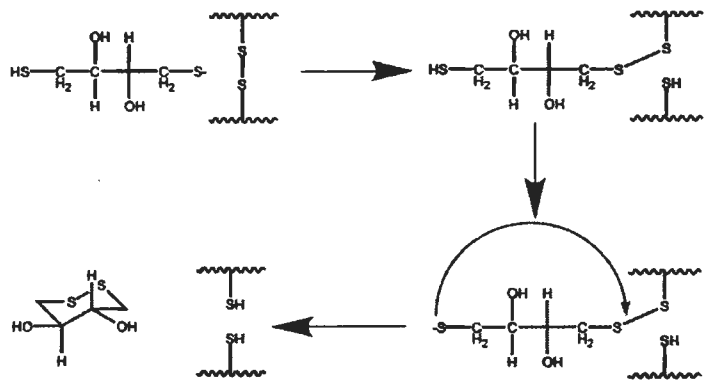


Figure 2.10 Disulfide exchange with DTT

For determination of the apparent molecular weight of proteins, the standards (proteins with known molecular weights) used to generate a standard curve by plotting the log molecular weight (log MW) correlate with relative mobility (Rf), illustrated in Figure 2.11, from which the molecular weight of the unknown sample can be determined. Figure 2.12 illustrates the mobility of standard proteins at several polyacrylamide gel concentrations which the curves are linear only over a limited range of molecular weights. Therefore, the gel concentration should be chosen so that the standards produce a linear curve in the region of the unknown.

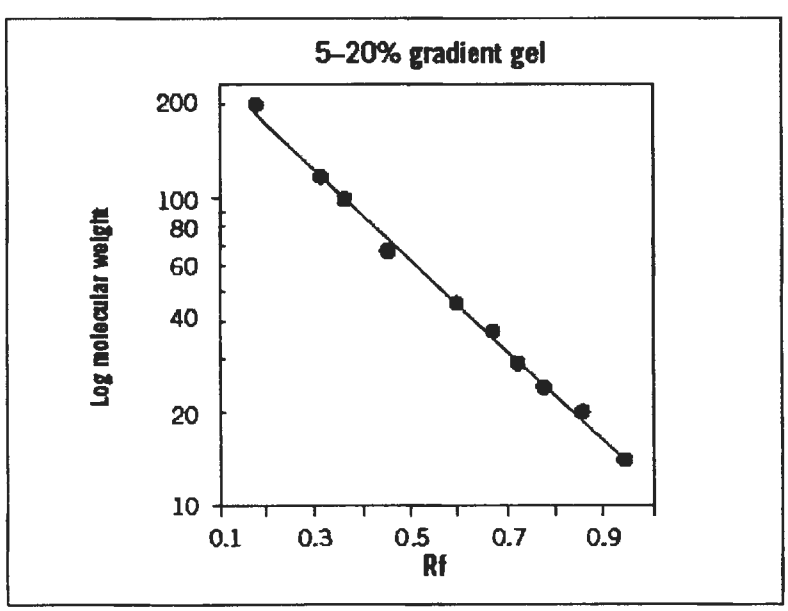


Figure 2.11 Calibration curves of protein standards

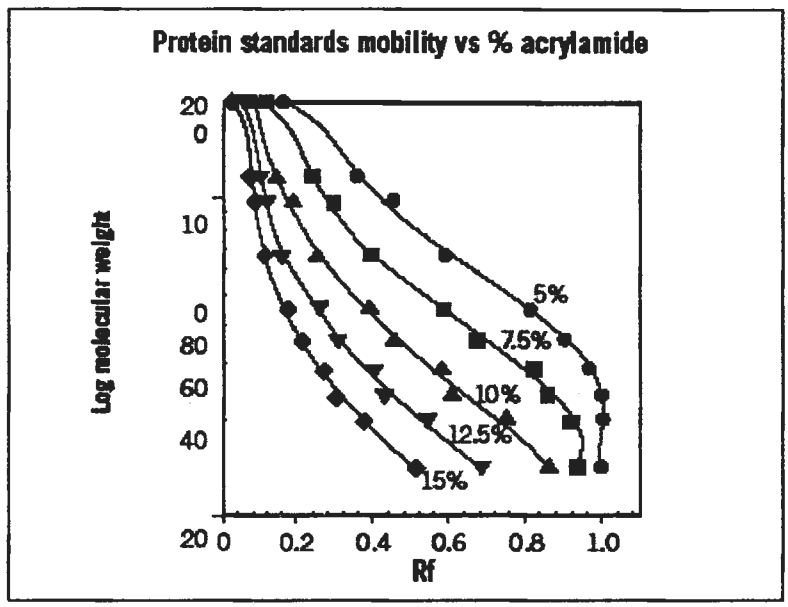


Figure 2.12 Calibration curves of protein standards at several % acrylamide.

Native PAGE or non-denaturing gel electrophoresis is run in the absence of SDS and reducing agent (DTT). In Native PAGE the migration of proteins depends on many factors, including its hydrodynamic size, shape, and native charge. The resolution of nondenaturing electrophoresis is generally not as high as SDS-PAGE, but the technique is useful when preservation of the native structure and conformation of the proteins are desired.

2D-PAGE [49] or Two-dimensional electrophoresis separates proteins by isoelectric point in the first dimension and by mass in the second dimension. It is a very efficient separation technique for proteins. The protein sample is first run on an IPG strip, which after reaching completion is placed in either a horizontal or vertical SDS gel. This technique results in gels that contain spots. Each spot on the gel corresponds to a different protein. 2D Electrophoresis is widely used, and certain methods of it can be coupled with mass spectrometry in order to identify proteins. The principle of this technique shown in Figure 2.13. In the first dimension, Isoelectric focusing (IEF) is an electrophoretic method that separates proteins according to their isoelectric points (pI), the specific pH at which the net charge of the protein is zero.

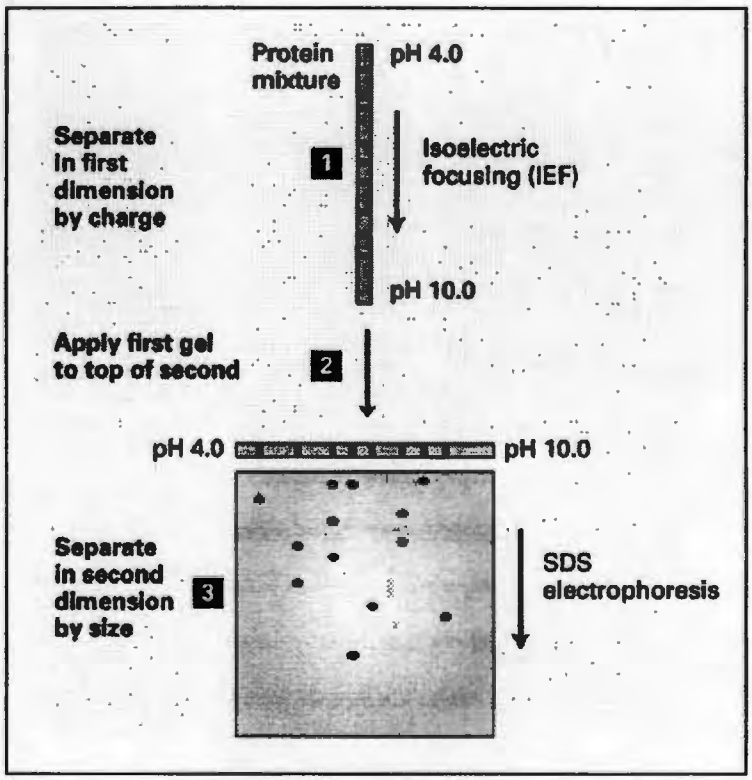


Figure 2.13 Two-dimensional electrophoresis processes.

The net charge of a protein depends on their amino acid composition and the pH of their surroundings. Proteins are positively charged at pH values below their pI and negatively charged at pH values above their pI. In a pH gradient, under an electric field a protein will move to the position in the gradient where its net charge is zero, a positive net charge protein will migrate toward the cathode but a negative net charge protein will migrate toward the anode. This is the focusing effect of IEF, which concentrates proteins at their pIs and separates proteins with very small charge differences. In the second dimension, SDS-PAGE is a method for using to separate proteins according to their molecular weights (MW) because of treating proteins with both SDS and a reducing agent will break up the interactions of proteins and any disulfide linkages present in the proteins. The correlation between the logarithm of the molecular weight and the relative distance of the proteins migration is an approximately linear relationship which can be used to estimate protein molecular weight relatively with proteins standard.

2.5.5 Column Chromatography

Chromatography is the science which is studies the separation of molecules based on differences in their structure and composition [45]. In general, chromatography involves moving a preparation of the materials to be separated over a stationary support. The molecules in the test preparation will have different interactions with the stationary support leading to separation of similar molecules. Test molecules which display tighter interactions with the support will tend to move more slowly through the support than those molecules with weaker interactions. In this way, different types of molecules can be separated from each other as they move over the support material. The basic procedure in chromatography is to flow the solution containing the protein through a column packed with various materials. Different proteins interact differently with the column material, and can thus be separated by the time required to pass the column, or the conditions required to elute the protein from the column. Usually proteins are detected as they are coming of the column by their absorbance at 280 nm. Many different chromatographic methods exist: Proteins are purified using chromatographic purification techniques which separate according to differences in specific properties, as shown in Table 2.2

Table 2.2 Protein properties used during purification.

Protein property	Technique
Charge	Ion exchange (IEX)
Size	Gel filtration (GF)
Hydrophobicity	Hydrophobic interaction (HIC) Reversed phase (RPC)
Biorecognition (ligand specificity)	Affinity (AC)
Charge, ligand specificity or hydrophobicity	Expanded bed adsorption (EBA) follows the principles of AC, IEX or HIC

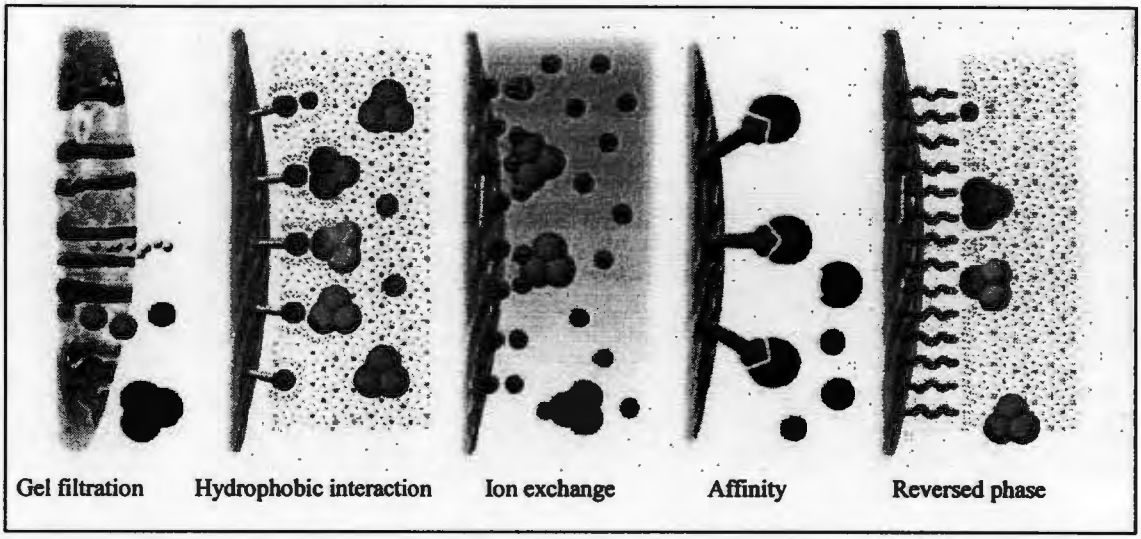


Figure 2.14 Separation principles in chromatographic purification.

2.5.5.1 Ion exchange chromatography

Ion exchange chromatography [50] separates compounds according to the nature and degree of their ionic charge. Proteins are made up of twenty common amino acids. Some of these amino acids possess side groups which are either positively or negatively charged. A comparison of the overall number of positive and negative charges will give a clue as to the nature of the protein. The net surface charge of proteins varies according to the surrounding pH. When above its isoelectric point (pI) a protein will bind to an anion exchanger, when below its pI a protein will behind to a cation exchanger.The column to be used is selected according to its type and strength of charge. Anion exchange resins have a positive charge and are used to retain and separate negatively charged compounds, while cation exchange resins have a negative charge and are used to separate positively charged molecules. The functional groups substituted onto a chromatographic matrix (Table 2.3) determine the charge of an ion exchange medium, a positively-charged anion exchanger or a negatively-charged cation exchanger. The terms strong and weak refer to the extent that the ionization state of the functional groups varies with pH. Commonly used anion exchange resins are Q-resin and DEAE resin while cation exchange resins are S-resin and CM resins (see Figure2.15).

Table 2.3 Functional groups used on ion exchangers.

Anion exchangers	Type	Functional group
Quaternary ammonium (Q)	strong	$-O-CH_2N^+(CH_3)_3$
Diethylaminoethyl (DEAE)	weak	$-O-CH_2CH_2N^+H(CH_2CH_3)_2$
Diethylaminopropyl(ANX)	weak	$-O-CH_2CHOHCH_2N^+H(CH_2CH_3)_2$
Cation exchangers	**Type**	**Functional group**
Sulfopropyl (SP)	strong	$-O-CH_2CHOHCH_2OCH_2CH_2CH_2SO^-_3$
Methyl sulfonate (S)	strong	$-O-CH_2CHOHCH_2OCH_2CHOHCH_2SO^-_3$
Carboxymethyl (CM)	weak	$-O-CH_2COO^-$

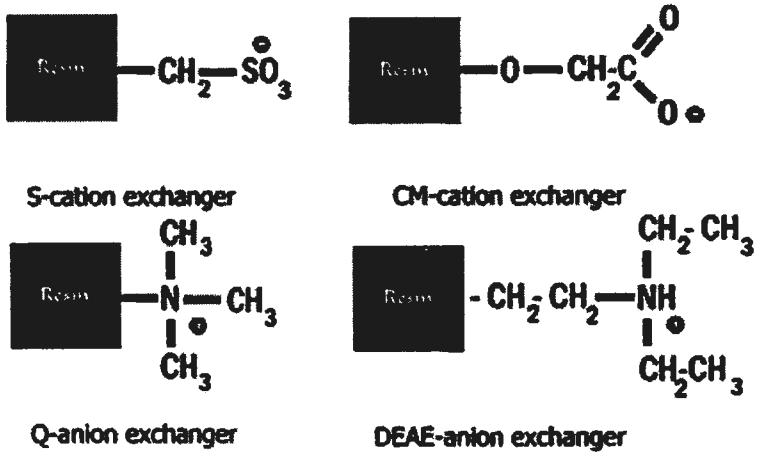


Figure 2.15 Ion exchange resins with functional group structure.

To release the proteins in the order of binding tenacity, conditions are then altered so that bound substances are eluted differentially. This elution is usually performed by changes in pH or increases in salt concentration. At low salt concentration (low ionic strengths), all components will be tightly held on top of the column. When the ionic strength of the mobile phase is increased, the salt ions compete with the adsorbed sample ions for the bonded charges on the column. As a result, some of the sample components will be partially desorbed and start moving through the column. If the salt concentration is higher, the resulting ionic strength causes a larger number of the sample components to be desorbed, and the speed of the movement down the column increases. The proteins come off the column matrix

when the ionic strength of the buffer neutralizes their charge. The least charged molecules come off first and the most highly charged come off last.

2.5.5.2 Gel-Filtration Chromatography

This technique separates proteins based on size and shape. The supports for gel-filtration chromatography are beads which contain holes of given sizes. The size of the pores in the matrix determines the rate at which proteins of various sizes diffuse into the beads and some proteins are completely excluded. Larger molecules, which can't penetrate the pores, move around the beads and migrate through the spaces which separate the beads faster than the smaller molecules, which may penetrate the pores. The column is then eluted with just buffer and the proteins come out in order with the largest first [51].

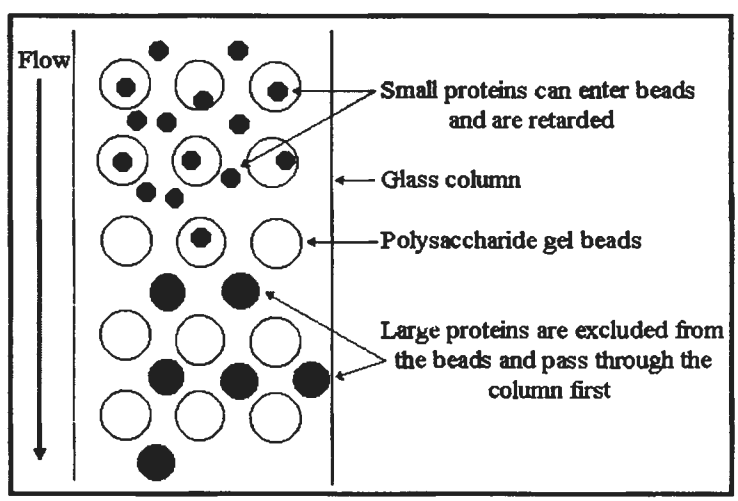


Figure 2.16 Gel-Filtration Chromatography

2.5.5.3 Affinity Chromatography

This is the most powerful technique which separates proteins on the basis of a reversible interaction between a protein (or group of proteins) and a specific ligand attached to a chromatographic matrix. This technique has the potential to be used for the purification of any protein, provided that a specific ligand is available.The target protein is specifically and reversibly bound by a complementary binding substance. The sample is applied under conditions that favor specific binding to the ligand.

Unbound material is washed away, and the bound target protein is recovered by changing conditions to those favoring desorption. Desorption is performed specifically, using a competitive ligand, or none specifically, by changing the pH, ionic strength or polarity [52].

2.5.5.4 Reversed phase chromatography

Reversed phase chromatography (RPC) [53] is based upon interactions between hydrophobic ligands covalently attached to the adsorbent and the hydrophobic patches of molecules that are applied in the aqueous mobile phase. The adsorbents used in this technique are highly substituted with hydrophobic ligands. This leads to that in RPC the hydrophobic interaction is strong enough to adsorb proteins in pure water. However, the very strong interactions that thereby are provided usually require the use of organic solvents and other additives to desorb the protein. This will most often have a denaturing effect on the protein. The majority of reversed phase separation experiments are performed in several fundamental steps as illustrated in Figure 2.17

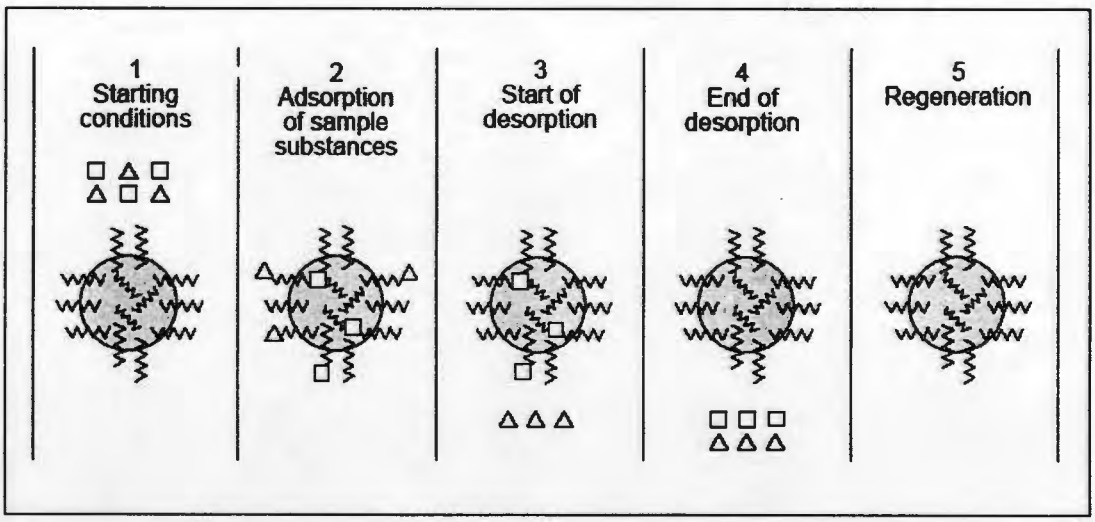


Figure 2.17 Principle of reversed phase chromatography with gradient elution

2.6 Protein Identification techniques

2.6.1 Edman Degradation

Edman degradation, developed by Pehr Edman, is a method of sequencing amino acids in a peptide [54]. This method utilizes phenylisothiocyanate to react with the N-terminal residue under alkaline conditions. The resultant phenylthiocarbamyl derivatized amino acid is hydrolyzed in anhydrous acid. The hydrolysis reaction results in a rearrangement of the released N-terminal residue to a phenylthiohydantoin derivative. The entire sequence of reactions can be repeated over and over to obtain the sequences of the peptide. This process has subsequently been automated to allow rapid and efficient sequencing of even extremely small quantities of peptide. For larger peptides must first be split into smaller peptides by chemical or enzymic digestion before proceeding with the reaction. It is able to accurately sequence up to 30 amino acids with 98% efficiency per amino acid. An advantage of the Edman degradation is that it only uses 10 - 100 picomoles of peptide for the sequencing process.

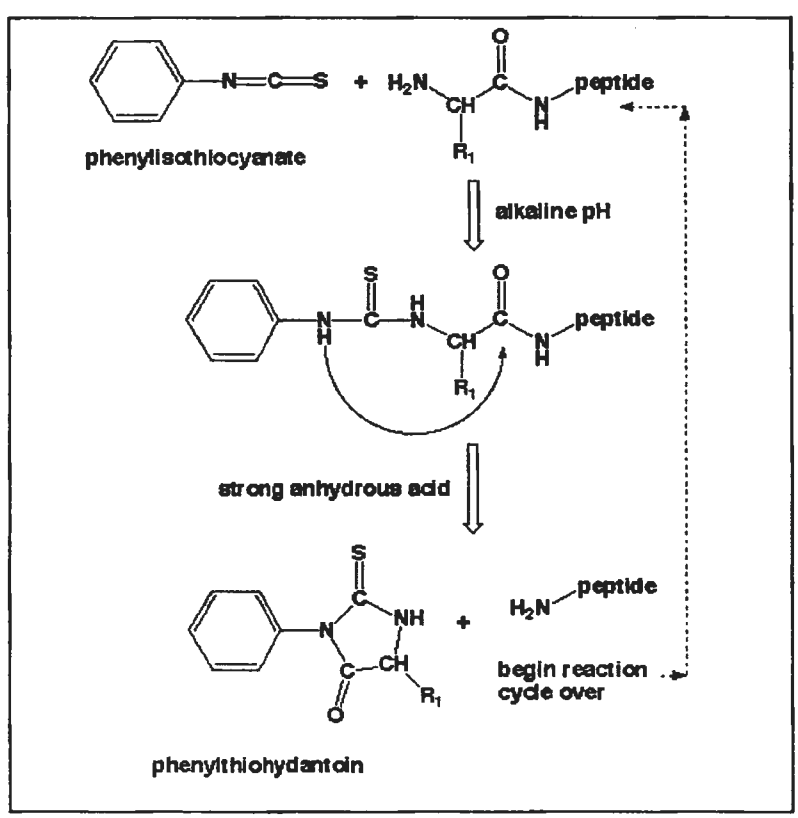


Figure 2.18 The Edman degradation reaction cycle

2.6.2 Mass Spectrometry

2.6.2.1 Basic concept of mass spectrometry

Mass spectrometry [55,56] is an analytical technique for determining the mass of molecules. Mass spectrometers use a variety of techniques to create charged ions or charged ion fragments and separate these ions on the basis of their charge to mass (m/z) ratios. The ionic separation is done by using a combination of magnetic or electric fields. The charged ions are detected by the miniscule current they create when the hit a detector. The resulting mass spectrogram is a series of peaks, with each peak corresponding to a different ion with a different mass-to-charge ratio. The intensity of the peaks is proportional to the number of ions with the corresponding m/z ratio. By analyzing the patterns of peaks seen in a mass spectrum it is possible to determine the mass of the parent ion and to determine something about its chemical structure.

All mass spectrometers contain at least three major components: an ion source which converts molecules into gas-phase ions, a mass analyzer which separates ions according to their mass-to-charge rations, and an ion collection/detection system which converts the passage of charged ions through a medium into detectable currents or signals. The instrument must also be connected to a computer system to process and record the data and a vacuum pump to control the pressure within the mass spectrometer (see Figure 2.19). Low pressure is necessary to limit the number of ion collisions, which would alter the path of the ions and possibly produce unwanted reaction products or loss of charge.

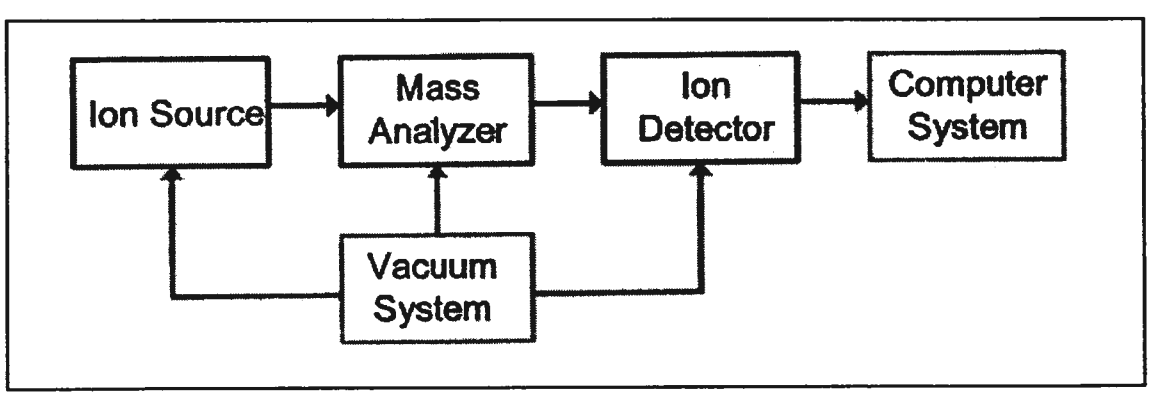


Figure 2.19 Schematic of the basic components of a mass spectrometer.

2.6.2.2 Ion source

Ion source is a piece of equipment used to ionize analyst molecules and, if necessary, free them from the solid or liquid phase. Once the analyst ions are free to move electric fields will direct them into the mass analyzer. There are several types of ion sources:

a) Electron impact (EI) uses an electron beam to ionize gas-phase atoms or molecules. Electrons from the beam knocks an electron off of analyze atoms or molecules to create ions.

b) Chemical ionization (CI) uses a reagent ion to react with the analyze molecules to form ions by either a proton or hydride transfer. The reagent ions are produced by introducing a large excess of methane (relative to the analyze) into an electron impact (EI) ion source.

c) Fast-atom bombardment (FAB) uses a high-energy beam of natural atoms, typically Xe or Ar, strikes a solid sample causing desorption and ionization. It is used for large biological molecules that are difficult to get into the gas phase. FAB causes little fragmentation and usually gives a large molecular ion peak, making it useful for molecular weight determination.

d) Field desorption (FD) uses the formation of ions in the gas phase from a material deposited on a solid surface in the presence of a high electrical field. The sample is deposited onto the emitter and the emitter is biased to a high potential and a current is passed through the emitter to heat up the filament. The analyze molecules are ionized by electron tunneling at the tip of the emitter.

e) Matrix-assisted laser desorption ionization (MALDI) is a method of vaporizing and ionizing large biological molecules such as proteins or DNA fragments. The biological molecules are dispersed in a solid matrix such as nicotinic acid. A UV laser pulse ablates the matrix which carries some of the large molecules into the gas phase in an ionized form so they can be extracted into a mass spectrometer.

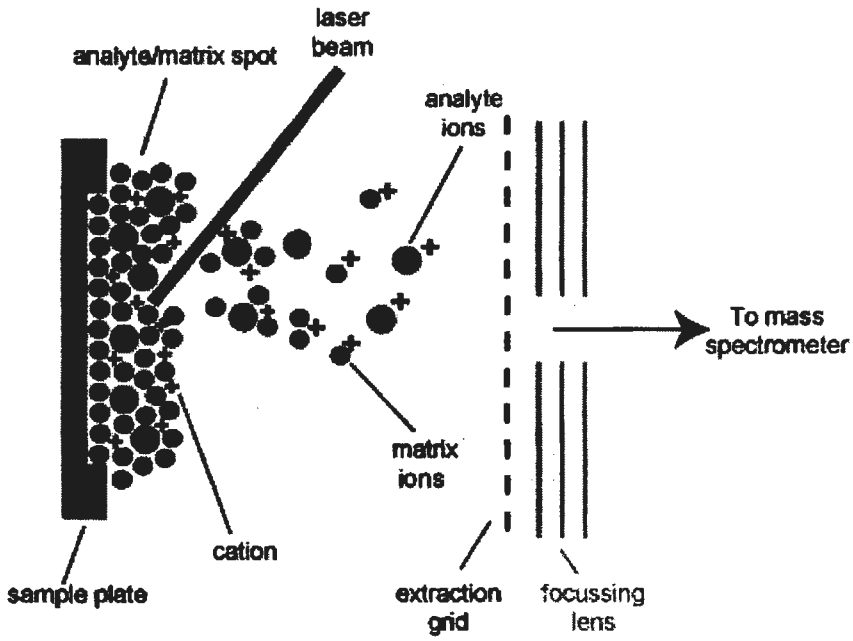


Figure 2.20 The mechanism of MALDI

The mechanism of MALDI is believed to consist of three basic steps:
(i) **Formation of a solid solution**: It is essential for the matrix to be in access thus leading to the analyte molecules being completely isolated from each other. This eases the formation of the homogenous 'solid solution' required to produce a stable desorption of the analyte.
(ii) **Matrix Excitation**: The laser beam is focused onto the surface of the matrix-analyze solid solution. The chromaphore of the matrix couples with the laser frequency causing rapid vibration excitation, bringing about localized disintegration of the solid solution. The clusters ejected from the surface consist of analyze molecules surrounded by matrix and salt ions. The matrix molecules evaporate away from the clusters to leave the free analyte in the gas-phase.
(iii) **Analyte Ionisation**: The photo-excited matrix molecules are stabilised through proton transfer to the analyze. Cation attachment to the analyze is also encouraged during this process. It is in this way that the characteristic $[M+X]^+$ (X= H, Na, K etc.) analyte ions are formed. These ionisation reactions take place in the desorbed matrix-analyte cloud just above the surface. The ions are then extracted into the mass spectrometer for analysis.

f) Electrospray ionization (ESI) consists of a very fine needle and a series of skimmers. A sample solution is sprayed into the source chamber to form droplets. The droplets carry charge when the exit the capillary and, as the solvent evaporates, the droplets disappear leaving highly charged analyze molecules. ESI is particularly useful for large biological molecules that are difficult to vaporize or ionize.

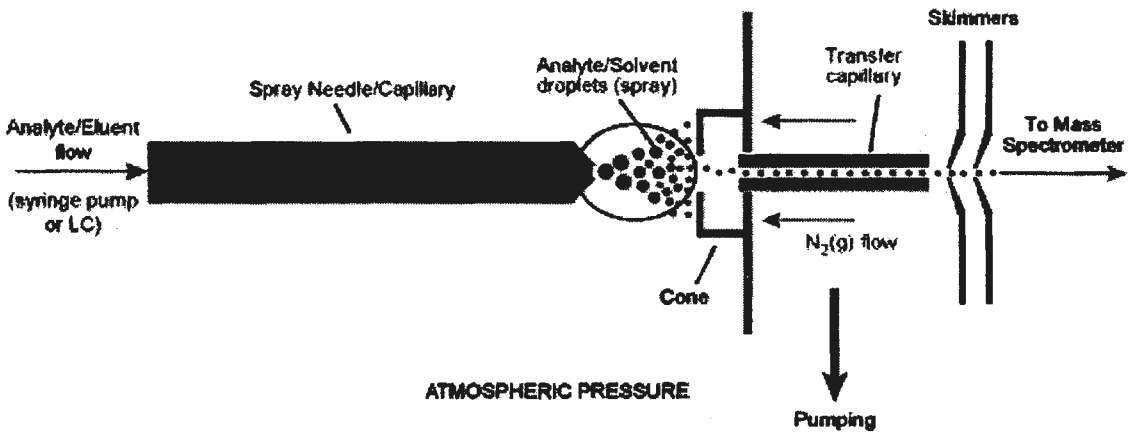


Figure 2.21 A schematic of an ESI source.

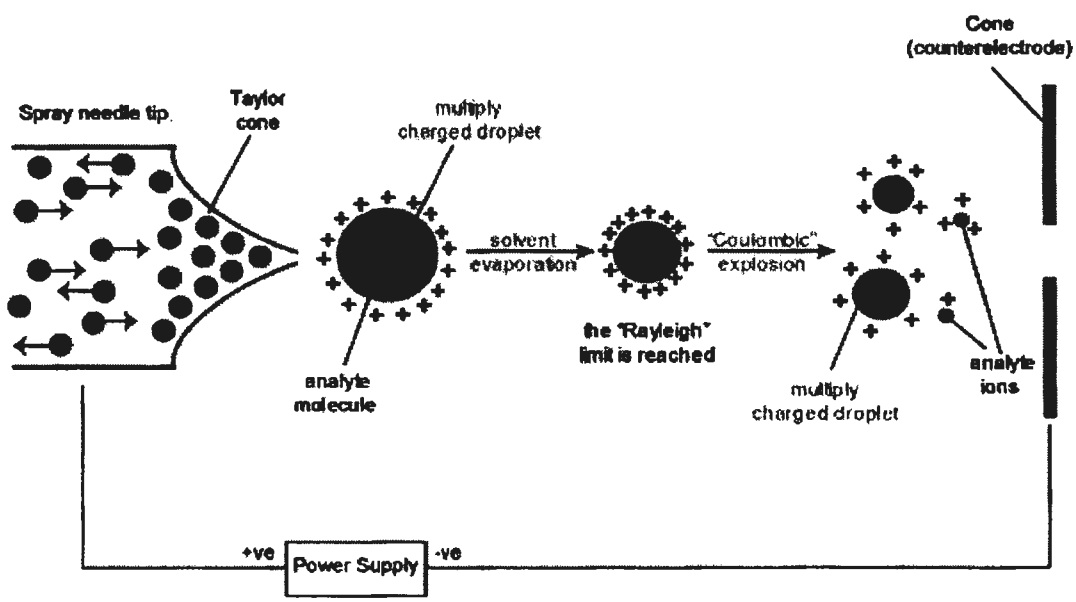


Figure 2.22 A schematic of the mechanism of ion formation in ESI.

2.6.2.3 Mass Analyzer

After the sample has been ionized, the beam of ions is focused and directed into a mass analyzer, which separates the ions based on their mass-to-charge ratio (m/z). All commonly used mass analyzers use electric and magnetic fields to apply a force on charged particles. The relationship between force, mass, and the applied fields can be summarized in Newton's second law and the Lorentz force law:

Newton's second law $F = ma$

Lorentz force law $F = q(E + v \times B)$

Where F is the force applied to the ion, m is the mass of the ion, a is the acceleration, q is the ionic charge, E is the electric field, and $v \times B$ is the vector cross product of the ion velocity and the applied magnetic field. From Newton's second law, it is apparent that the force causes an acceleration that is mass-dependent, and the Lorentz force law show that the applied force is also dependent on the ionic charge. Therefore, the mass spectrometers separate ions according to their mass-to-charge ratio (m/z) rather than by their mass alone.

There are many types of mass analyzer; the most common mass analyzers are the magnetic sector, the quadrupole, the ion trap, the TOF, and the Fourier transform-ion cyclotron resonance analyzer.

a) Magnetic sector In a magnetic deflection mass spectrometer, ions leaving the ion source are accelerated.to a high velocity. The ions then pass through a magnetic sector in which the magnetic field is applied in a direction perpendicular to the direction of ion motion.

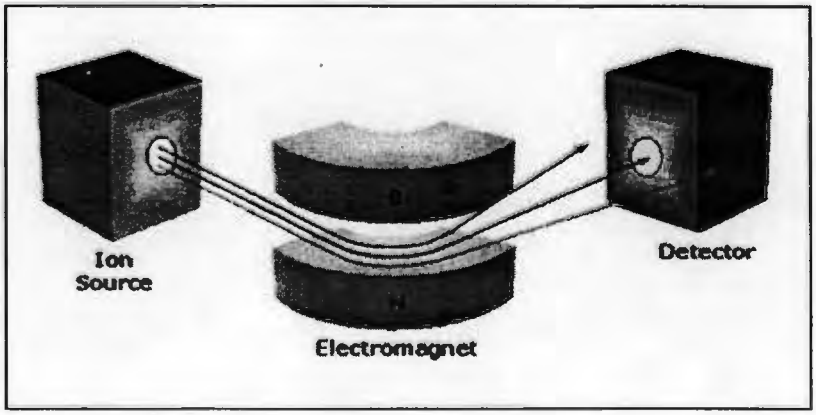


Figure 2.23 Magnetic sector mass analyzer

b) Quadrupole mass analyzer consists of four parallel rods that have fixed DC and alternating RF potentials applied to them. Ions produced in the source of the instrument are then focused and passed along the middle of the quadrupoles. Their motion will depend on the electric fields so that only ions of a particular m/z will be in resonance and thus pass through to the detector. The RF is varied to bring ions of different m/z into focus on the detector and thus build up a mass spectrum.

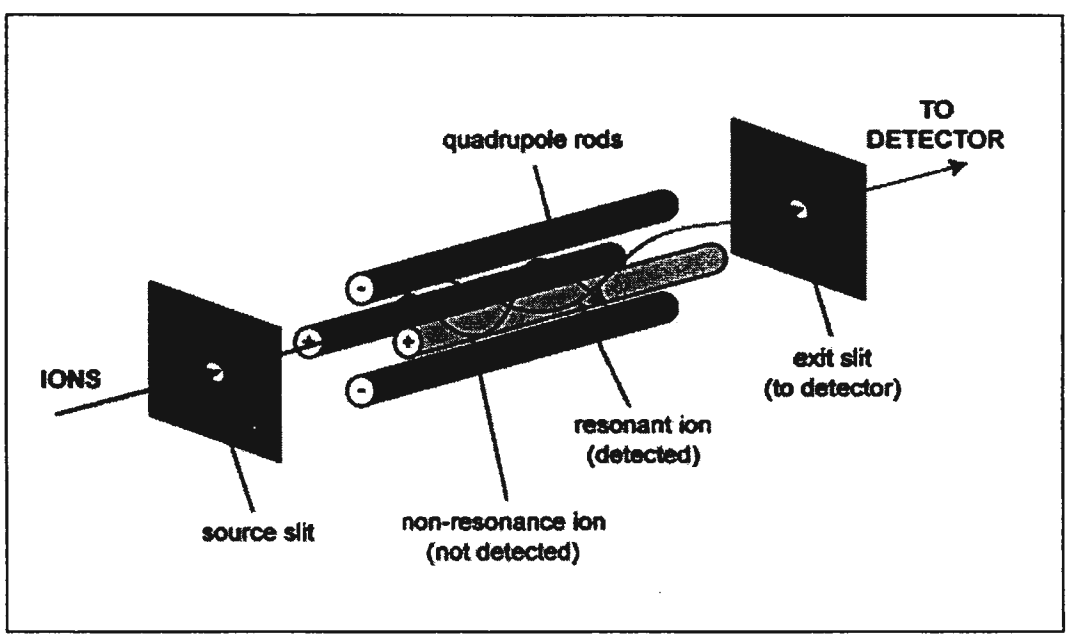


Figure 2.24 Quadrupole mass analyzer

c) Time-of-flight (TOF) mass analyzer is based on a simple mass separation principle. Consider ionized species starting from the same position at the same time, being accelerated by means of a constant homogeneous electrostatic field. Their velocities are unambiguously related to their mass-to-charge ratio and times of arrival at a detector directly indicate their masses, If the particles all have the same charge, then their kinetic energies will be identical, and their velocities will depend only on their masses. Lighter ions will reach the detector first.

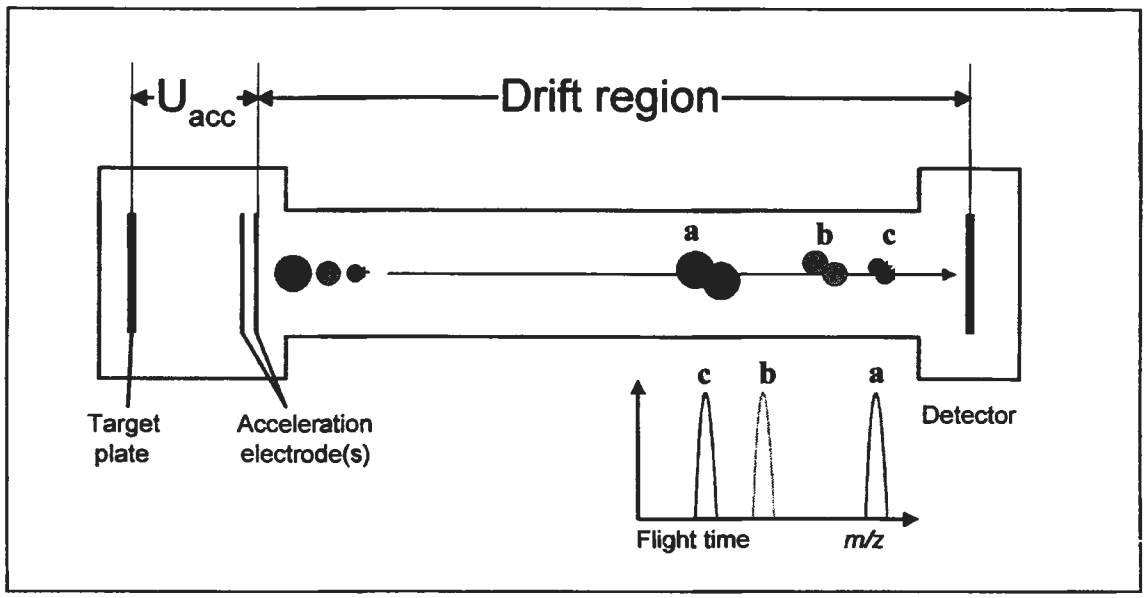


Figure 2.25 Time-of-flight (TOF) mass analyzer

d) Fourier transforms ion cyclotron resonance analyzer is a type of mass analyzer for determining the mass to charge ratio (m/z) of ions based on the cyclotron frequency of the ions in a fixed magnetic field. The ions are trapped in a Penning trap (a magnetic field with electric trapping plates) where they are excited to a larger cyclotron radius by an oscillating electric field perpendicular to the magnetic field. The excitation also results in the ions moving in phase (in a packet). The signal is detected as an image current on a pair of plates which the packet of ions passes close to as they cyclotron. The resulting signal is called a free induction decay (fid), transient or interferogram that consists of a superposition of sine waves. The useful signal is extracted from this data by performing a Fourier transform to give a mass spectrum. Specifically a fast Fourier transform (FFT) is usually used to transform the discrete fid data.

2.6.2.4 Detector

The final element of the mass spectrometer is the detector. The detector monitors the ion current, amplifies it and the signal is then transmitted to the data system where it is recorded in the form of mass spectra .The m/z values of the ions are plotted against their intensities to show the number of components in the sample, the molecular mass of each component, and the relative abundance of the various components in the sample. The type of detector is supplied to suit the type of analyzer; the more common ones are the photomultiplier, the electron multiplier and the micro-channel plate detectors.

2.6.3 Tandem mass spectrometry (MS-MS)

Tandem mass spectrometers [57] are instruments that have more than one analyzer and so can be used for structural and sequencing studies. Two, three and four analyzers have all been incorporated into commercially available tandem instruments, and the analyzers do not necessarily have to be of the same type. In an MS-MS experiment a precursor ion is mass-selected by mass analyzer 1 (MS_1) and focused into a collision region preceding a second mass analyzer (MS_2) followed in Figure 2.26a Inert gas is generally introduced into the collision region and collisions occur between the precursor ion and inert gas atoms. Excited precursor ions decompose to product ions in a process termed "collision-induced dissociation" (CID). MS-MS analysis can be carried out in either of three scan mode illustrate in Figure 2.26. From this Figure, product ions scan is the specific scan function or process that records product ion spectrum. Precursor-ion scan is the specific scan function or process that will record a precursor ion spectrum. And Neutral loss scan is a scan that determines, in a single instrument, all the parent ion mass/charge ratios which react to the loss or gain of a selected neutral mass.

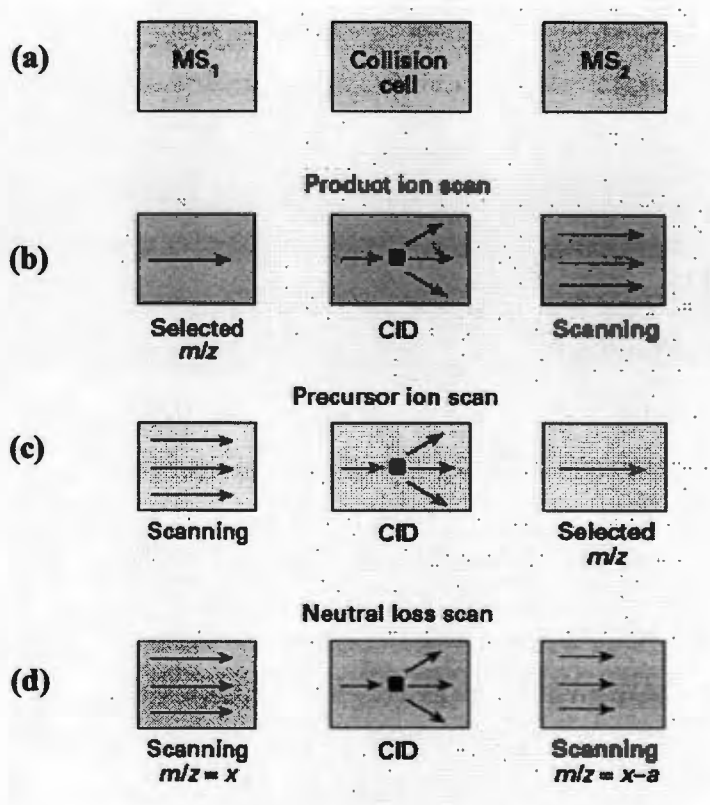


Figure 2.26 MS-MS experiments on spatially separated mass analyzers

In order to obtain protein and peptide sequence information by tandem mass spectrometry, fragments of an ion must be produced by introducing the selected ion in to a collision cell. These fragment ions can be separated into two classes. One class retains the charge on the N-terminal while cleavage is observed from the C-terminal which can occur at three different positions, A_n, B_n, C_n. The second class of fragment ions generated from the N-terminal retains the charge on the C-terminal, while cleavage is observed from the N-terminal which can occur at three different positions, types X_n, Y_n, and Z_n. The mechanism of peptide-bond cleavage is now fairly well understood. Considering, for example, a doubly protonated tryptic peptide, one proton will be localized on the C-terminal Arg or Lys side chain and the second may be localized at one of the amide bonds (or the N-terminus). In the collision process this second proton can be mobilized, giving a heterogeneous population of ions in which the second proton may reside at any one of the peptide bonds. Protonation of amide nitrogen will weaken the amide bond and lead to cleavage.

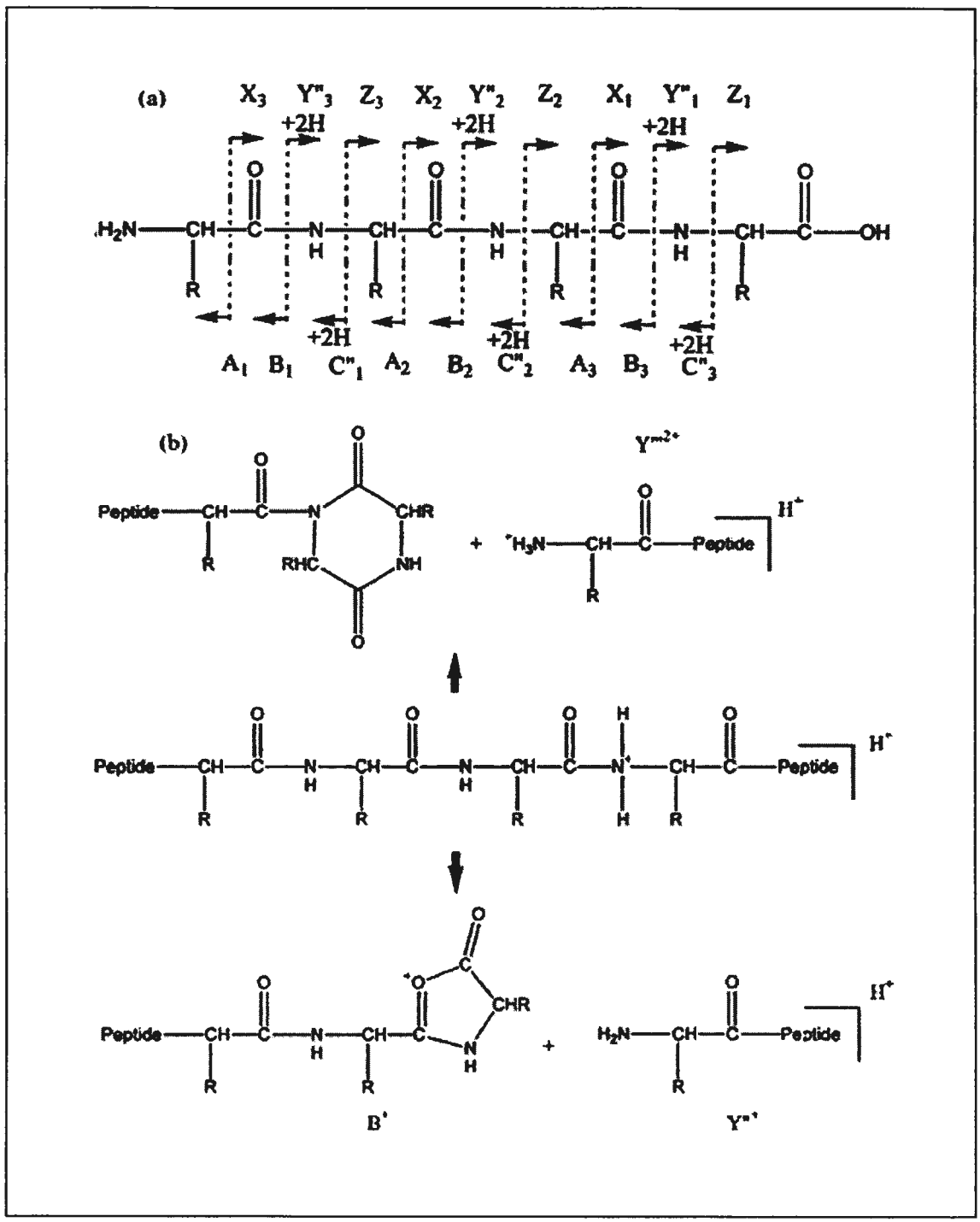


Figure 2.27 Schematic the cleavage of peptide bonds and fragment nomenclature

2.6.4 Using Mass Spectrometry for Protein Identification

Mass spectrometry has been used for the analysis of proteins and peptides since 1989, when two new "soft" techniques for gas phase ionization of large, polar, and highly charged molecules were introduced [58,59]. In one of these, electrospray ionization (ESI), ions are formed from a liquid solution at atmospheric pressure, while in the second, matrix assisted laser desorption ionization (MALDI), a laser pulse induces the sample to sublimate out of a dry crystalline matrix. The devices that bring the analytes into gas phase and ionize them are known as "sources". Whole protein mass analysis is primarily conducted using either time-of-flight (TOF) MS, or Fourier transform ion cyclotron resonance (FT-ICR). These two types of instrument are preferable because of their wide mass range, and in the case of FT-ICR, its high mass accuracy. The most widely used instrument for peptide mass analysis is the quadrupole ion trap. Multiple stages Q-TOF and MALDI –TOF instruments also find use in this application.

2.6.5 Protein sequence databases

Protein sequence databases [60] are a central resource for storing the data generated by these and more conventional efforts, and making them available to the scientific community. There are a number of different protein sequence databases available with different aims. It is important to distinguish between universal databases covering proteins from all species and specialized data collections storing information about specific families or groups of proteins, or about proteins of a specific organism. Furthermore, two categories of universal protein sequence databases can be discerned: expertly curated databases and sequence repositories.

2.6.5.1 Sequence repositories

These databases add little or no additional information to the sequence records they contain and generally make no effort to provide a nonredundant collection of sequences to users. Example of these databases such as GenPept, NCBI's Entrez Protein, RefSeq.

a) **GenPept**: GenBank Gene Products Data Bank (GenPept) database was produced by the National Center of Biotechnology Information (NCBI). The entries in the

database are derived from translations of the sequences contained in the nucleotide database. This entries lack additional annotation and does not contain proteins derived from amino acid sequencing.

b) **NCBI's Entrez Protein**: The database contains sequence data translated from the nucleotide sequences of the DDBJ/EMBL/GenBank database. The database differs from GenPept in that many of the entries contain additional information that has been extracted from curated databases such as Swiss-Prot and PIR. As with GenPept, the sequence collection is redundant.

c) **RefSeq**: Reference Sequence was produced by the NCBI. The aim of the project is to provide a non-redundant collection of reference protein sequences. The main features of the RefSeq collection include nonredundancy, explicitly linked nucleotide and protein sequences, updates to reflect current knowledge of sequence data and biology, data validation and format consistency, and distinct accession series.

2.6.5.2 Expertly curated databases

The curated databases enrich the sequence data by adding additional information, which gets validated by expert biologists before being added to the databases to ensure that the data in these collections can be considered to be highly reliable. There is also a large effort invested in maintaining non-redundant datasets by compiling all reports for a given protein sequence into a single record. Example of these databases present in below.

a) **PIR-PSD** (http://pir.georgetown.edu/) is the oldest universal curated protein sequence database established in 1984 as a successor to the original National Biomedical Research Foundation Protein Sequence Database.

b) **Swiss-Prot** (http://www.ebi.ac.uk/swissprot/index.html) is the leading universal curated protein sequence database. It strives to provide a high level of annotation, a minimal level of redundancy, a high level of integration with other biomolecular databases, and extensive external documentation.

c) **TrEMBL** (http://www.ebi.ac.uk/trembl/) was introduced to make new sequences available as quickly as possible. TrEMBL consists of computer-annotated entries

derived from the translation of all coding sequences in the DDBJ/EMBL/ GenBank nucleotide sequence database that are not yet included in Swiss-Prot.

d) **UniProt** (http://www.uniprot.org) is the next generation of protein sequence databases. UniProt comprises three components: first, the UniProt Knowledgebase which will continue the work of Swiss-Prot, TrEMBL and PIR by providing an expertly curated database; second, the UniProt Archive (UniParc) into which new and updated sequences are loaded on a daily basis; third, the UniProt non-redundant reference databases (UniProt NREF), which provide non-redundant views on top of the UniProt Knowledgebase and UniParc.

2.6.6 Protein Identification Using Mass Spectrometry Data

2.6.6.1 Protein Identification from Peptide Mass Fingerprinting

Peptide mass fingerprinting (PMF) developed in the early 1990's by John Yates and colleagues, is an analytical technique for protein identification. Meaning of Peptide mass fingerprinting is a set of peptide molecular masses from an enzyme digest of a protein. The principle of protein identification using peptide mass fingerprinting is based on the comparison of the list of experimental masses of known proteins (Figure 2.28). The goal is to find the protein or proteins whose peptide masses provide the best match with the experimental fingerprint. Search engines for peptide mass fingerprint could be accessed on the web, with early sites including Protein Prospector (UCSF), ProFound (Rockefeller University), Mascot (Matrix Science), and others (Table 2.4).

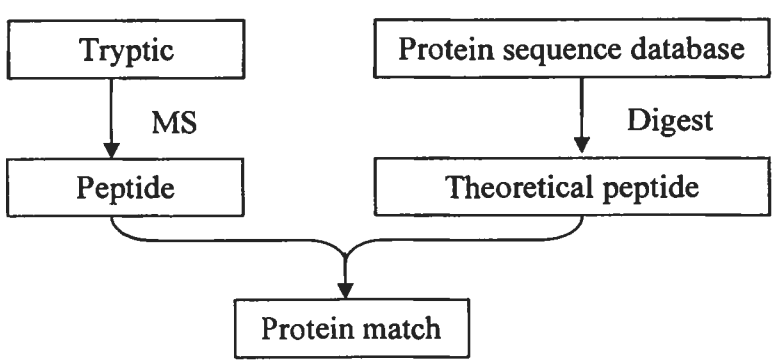


Figure 2.28 A flowchart illustrating the concept of peptide mass fingerprinting.

Table 2.4 PMF Servers on the web

PMF Servers	Web side
Aldente (Phenyx)	http://www.expasy.org/tools/aldente/
Mascot	http://www.matrixscience.com
MassSearch	http://cbrg.inf.ethz.ch/Server/MassSearch.html
Mowse	http://srs.hgmp.mrc.ac.uk/cgi-bin/mowse
MS-Fit	http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm
PeptideSearch	http://www.mann.emblheidelberg.de
Profound (Prowl)	http://bioinformatics.genomicsolutions.com
XProteo	http://xproteo.com:2698/
PepMAPPER	http://wolf.bms.umist.ac.uk/mapper/

2.6.6.2 Proteins Identification from tandem (MS/MS) mass spectra.

Because peptide mass fingerprinting does not always work for unambiguous protein identification there has been increasing emphasis on using tandem mass spectrometers equipped with collision induced dissociation (CID) cells to provide more precise and interpretable peptide data. SEQUEST and Mascot are two software packages that can be used to analyze tandem mass data of peptide fragments. Both programs take uninterpreted tandem mass spectral data, perform sequence database searches, and identify probable peptides or protein matches.