CHAPTER II

THEORETICAL AND LITERATURE REVIEWS

2.1 Leguminosae[35,36]

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).

- 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.

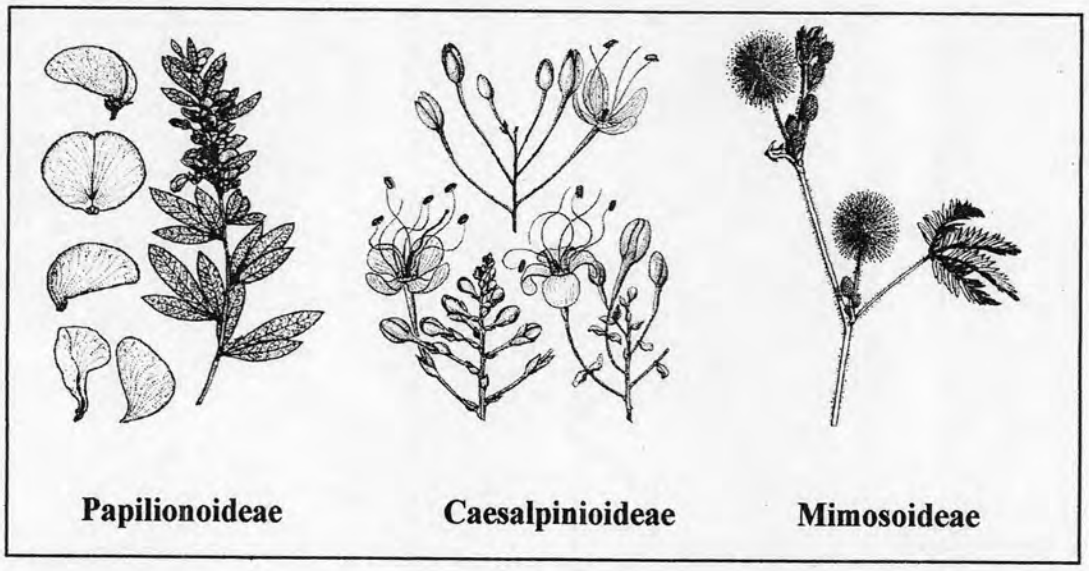

Figure 2.1 Flower morphology of three sub-families belongs to Leguminosae.

Legume is a plant in the family Fabaceae (or Leguminosae), or a fruit of these specific plants. A legume fruit is a simple dry fruit that develops from a simple carpel and usually dehisces (opens along a seam) on two sides. A common name for this type of fruit is a "pod", although pod is also applied to a few other fruit types, such as vanilla. Well-known legumes include alfalfa, clover, peas, beans, lentils, lupins, mesquite, carob, and peanuts.

2.2 *Pithecellobium dulce* [37-40]

2.2.1 General Background

Pithecellobium dulce (Roxb.) Benth. (Ma kam Thet) (Family Leguminosae, subfamily Mimosoideae) is one of 100-200 species in this genus. *Pithecellobium* species are noted for their tolerance of heat, salinity, drought conditions and impoverished soils. The common names of this plant are Manila tamarind, Madras thorn, monkeypod, sweet inga, and sweet tamarind. *Pithecellobium dulce* is a flowering woody legume native to Central, Northern South America and the northwestern regions of Mexico. This plant is the only species that has become widespread outside its origin. It is introduced and widely distributed throughout the Caribbean, Florida, Guam, Hawaii, India and Southeast Asia (Thailand, Malaysia and Indonesia), and can also be found in South Africa and Australia.

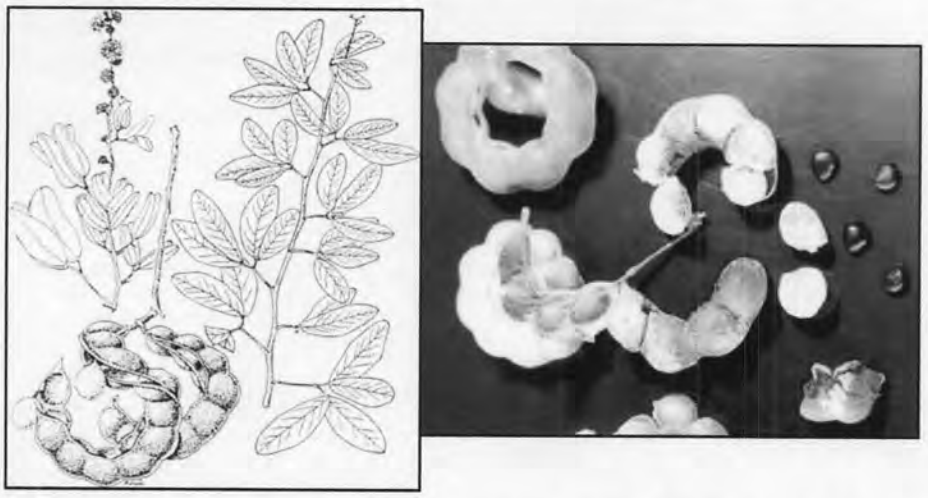

Figure 2.2 *Pithecellobium dulce* flowers, leave, pods, pulps and seeds.

2.2.2 Morphology Description

The height of *P. dulce* is commonly 10-15 meters, but ranges from 5 to 18 m. They are broad-spreading with irregular branches. The bark is grey, becoming rough, furrowed, and then peeling. Leaves are bipinnate, and leaflets oblong to 4 cm in length. Thin spines are in pairs at the base of leaves, and range from 2 to 15 mm in length. Leaves are deciduous. However, new leaf growth coincides with the loss of old leaves, giving the tree an evergreen appearance. The flowers are in small white heads 1 cm in diameter. Each flower has a hairy corolla and calyx surrounding about 50 thin stamens united in a tube at the base. Flowering begins in 3-4 years and is seasonal (October-November in Thailand). The pods with an edible pulp are pinkish, 1-1.5 cm wide, about 12 cm long, and become spiral as they mature (December-March in Thailand). Seeds are about 10 per pod (9,000 to 26,000/kg), black and shiny, hanging on a reddish thread from the pod. The pod splits along both margins.

2.2.3 Chemical Constituents

A pod constitutes 25% of skin, 50% of pulp and 25% of seed. The pulp contains per 100 g, 78.87 kilocalories, 77.8 g water, 3 g protein, 0.5 g fat, 19.6 g carbohydrate, 1.2 g fiber, 0.7 g ash, 13 mg Ca, 42 mg P, 0.5 mg Fe, 19 mg Na, 20.2 mg K, Vitamin A 25 IU (7.5 μg retinol equivalent and 15 μg β-carotene equivalent), 0.24 mg thiamine, 0.1 mg riboflavin, 0.6 mg niacin and 133 mg Vitamin C. The bark contains high tannin. The seed contain13.5% moisture, 17.6% protein, 17.1% crude fat, 7.8% crude fibres, 41.4% starch and 2.6% ash. Seeds have been reported to contain steroids, saponins, lipids, phospholipids, glycosides, glycolipids and polysaccharides. Roots reported to be estrogenic.

2.2.4 Uses and Applications

This tree species has been used for fencing and tanning, but due to its high protein, dietary fiber and unsaturated fatty acids content, seeds of *Pithecellobium* species are traditionally used worldwide for both food and feedstuff. Pods contain a pulp that is variously sweet and acid. The seed and pulp are made into a sweet drink and eaten roasted or fresh. In India, the seeds are used fresh or in curries. Moreover, seeds contain oil that can be used in soap-making or as food, and the residue can be

used as animal feed. The pods are relished by monkeys and livestock. The flowers are attractive to bees as source of pollen. The resulting honey is of high quality. The wood of *P. dulce* is strong and durable vet soft and flexible. It can be used in construction and for posts. The reddish-brown heartwood is dense and difficult to cut. It is commonly used as fuel. Although due to smokiness and low calorific values (5,500 kcal/kg), it is not of high quality. *Pithecellobium dulce* is also very popular as an ornamental and is used in topiarv (plant sculpturing). Trees with variegated leaflets are available as ornamentals in Hawaii. When wounded, the bark exudes a reddish-brown gum similar to gum arabic that dissolves in water to make a mucilage. The bark can also be used for tanning and produces a yellow dye. For medicinal uses as folk medicine, a decoction of its bark can cure frequent bowel movement. Its bark is also used in dysentery, dermatitis and eye inflammation. The leaves can be applied as plasters for pain and venereal sores. Decoctions of leaves are also used for earaches, leprosy, toothaches, larvicide, indigestion and abortifacient.

2.3 Protein Extraction and Precipitation[41-45]

2.3.1 Protein Extraction

The initial step of any purification procedure involves recovery of the protein from its source. The complexity of this step depends largely upon whether the protein of interest is intracellular or extracellular. If the protein required is an intracellular one, collection of the source cells or tissue is followed by their disruption. Most mammalian cells and tissues are relatively easily disrupted. Animal cells, unlike their bacterial or plant counterparts, are devoid of a protective cell wall. Most techniques rely on physical disruption of the cell membrane. A well-known example is homogenization. Upon completion of the homogenization step cellular debris and any remaining intact cells can be removed by centrifugation or by filtration.

2.3.2 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 (p*I*) 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 2.3.

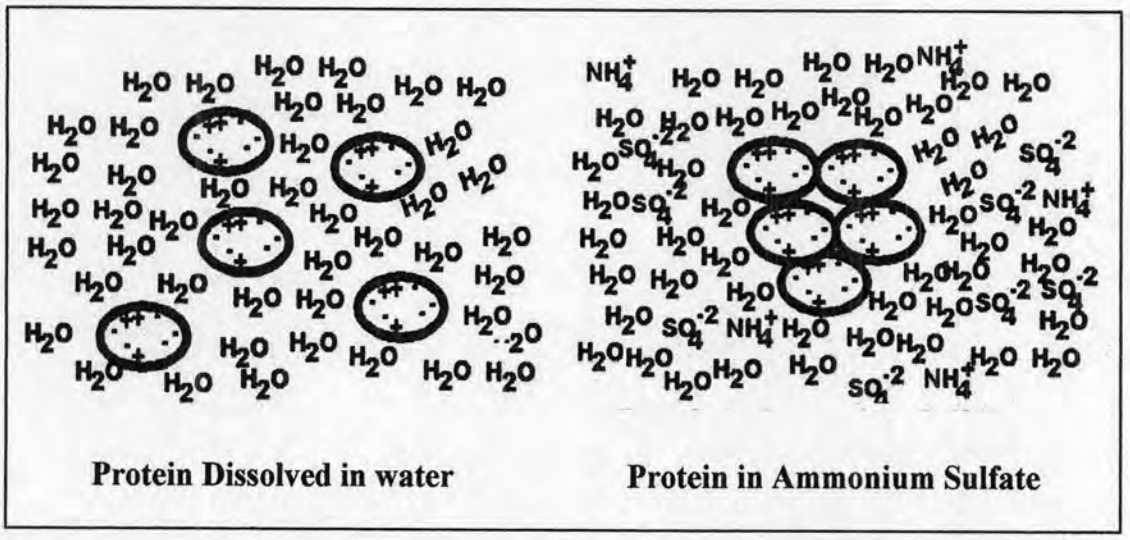

Figure 2.3 Protein dissolving in water when absent salt and increasing salt.

Another method is the addition of an organic solvent. The method of protein precipitation by water-miscible organic solvents has been employed since the early

days of protein purification. Addition of a miscible solvent such as ethanol or acetone to an aqueous extract containing proteins has a variety of effects that, combined, lead to protein precipitation. The principal effect is the reduction in water activity. The solvating power of water for a charged, hydrophilic protein molecule is decreased as the concentration of organic solvent increases. This can be described in terms of reduction of the dielectric constant of the solvent, or simply in terms of a bulk displacement of water, plus the partial immobilization of water molecules through hydration of the organic solvent. If there is a medium decrease in the dielectric constant with the addition of an organic solvent, the solubility should decrease also. A two dimensional representation of proteins in water-organic solvent mixture is shown in Figure 2.4 Here aggregation is occurring by interactions between opposite-charged areas on the protein's surfaces.

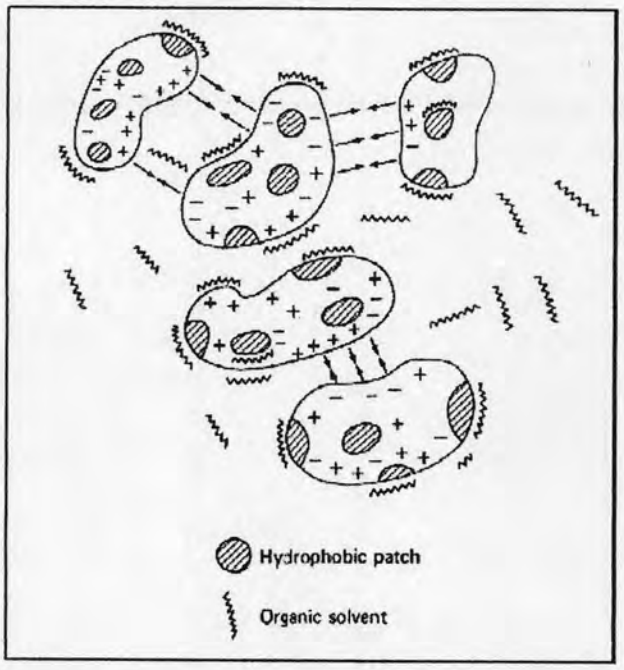

Figure 2.4 Aggregation of proteins by interactions in an aqueous-organic solvent mixture.

2.4 Separation of Proteins [46]

A large number of isolation methods for proteins have been developed. Most of proteins purification involves with chromatography and electrophoresis, for examples, reversed-phase high-performance liquid chromatography (HPLC), ion-

exchange chromatography (IEX), affinity chromatography (AC), hydrophobic interaction chromatography (HIC), gel filtration (GF), capillary electrophoresis (CE), and gel electrophoresis (GE). Each separation method is based on different principles. Proteins are purified using chromatographic techniques which separate according to differences in specific properties, as shown in Table 2.1. In many chromatographic methods, sample components compete with the eluant for sites on the adsorbent and the separation depends on different partitioning between mobile and stationary phase. The components pass with different velocities through the separation column. The high affinity components for the adsorbent are more retarded. Chromatograms are usually monitored by UV absorption at 280 nm. This wavelength is selected because the aromatic residues tyrosine and tryptophan posses strong absorptions in this region.

Table 2.1 Protein properties used during chromatographic purification

Protein Properties	Purification Techniques
Charge	Ion-exchange chromatography (IEX)
Size	Gel filtration (GF)
Hydrophobicity	Reversed-phase high-performance liquid chromatography (HPLC), hydrophobic interaction chromatography (HIC)
Biorecognition (ligand specificity)	Affinity chromatography (AC)

In electrophoretic separation techniques, high separation efficiency can be achieved using a relatively limited amount of equipment. It is mainly applied for analytical rather than for preparative purposes. The electrophoretic separation is carried out under the influence of an electrical field, charged molecules and particles migrate in the direction of the electrode bearing the opposite charge. During this process, the substances are usually in an aqueous solution. Because of their varying

ratios of charges to masses, different molecules and particles of a mixture will migrate at different velocities and will thus be separated into single zone. The electrophoretic mobility, which is a measure of the migration velocity, is a significant and characteristic parameter of a charged molecule or particle. It is dependent on the p*K* values of the charged groups and the size of the molecule of particle. Electrophoretic separations may be carried out in either free solution as in capillary (capillary electrophoresis) or free flow systems, or in stabilizing media such as films or gels (gel electrophoresis).

According to this research, the separations of seed proteins from *Pithecellobium dulce* utilized gel electrophoresis, anion exchange and gel filtration chromatography techniques. The following details of basic principles will be focused on these three separation methods.

2.4.1 Ion Exchange Chromatography[47]

Ion exchange chromatography 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 (p*I*) a protein will bind to an anion exchanger, when below its p*I* 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 2.2) 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 Figure 2.5).

Table 2.2 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	$-O-CH_2COO^-$

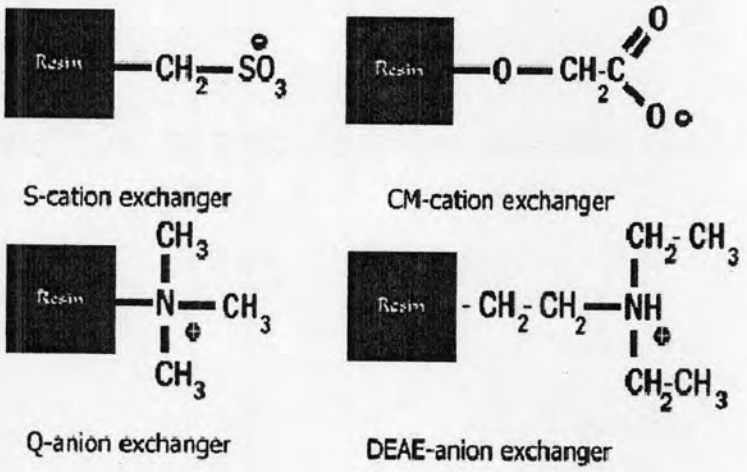

Figure 2.5 Ion exchange resins with functional group structure.

2.4.2 Gel Filtration Chromatography[48]

Gel filtration chromatography was introduced by Porath and Flodin in 1959.[49] Gel filtration chromatography separates proteins with differences in molecular size. Large molecule can not enter the matrix. Intermediate size molecules can enter part of the matrix and small molecules freely enter the matrix. The function of the matrix is to provide continuous decrease in accessibility for the molecule of increasing size. The large molecules are eluted from the column first and the smallest

are eluted last. The solvent in a gel filtration column is both in the space between the particles, *the void volume* V_0, and inside the particles, *the internal volume* V_i. A large molecule that is excluded from the interior has an elution volume $V_e = V_0$. A small molecule which all of V_i is accessible elutes with $V_e = V_0 + V_i$. A molecule of an intermediate size is excluded from the smaller pores, and the internal volume is only partially accessible. A distribution coefficient K_d is defined as the fraction of V_i accessible. The elution volume can be expressed as

$$V_e = V_0 + (K_d \times V_i)$$

and

$$K_d = (V_e - V_0)/ V_i.$$

$K_d = 0$ for molecules that are excluded from the interior of the particles, $K_d = 1$ for molecules to which the solvent in the both the void and the interior volume is accessible, and $K_d > 1$ indicates adsorption to the gel. The elution volume depends on both the shape and the molecular weight of the molecule. Most columns will fractionate within a particular molecular weight range, determined by the pore size of the bead. A good correction exists between molecular weight and elution volume for a set of molecules with similar gross structure, such as globular proteins. The exclusion limit gives the lower limit for molecules eluting in the void.

A great number of gel filtration supports are available, and they can be considered to two general classes: cross-linked natural or synthetic polymers and silica of different pore sized coated with a hydrophilic layer. The cross-linked polymers have a rather limited mechanical stability and restrictive using in conventional chromatography (low flow rate and low pressure column). In contrast to the polymeric supports, matrices suitable for size-exclusion HPLC must consist of small uniform particles that are rigid and can withstand high pressures. The two types of HPLC supports are silica based and organic based polymer. The examples of supports for gel filtration chromatography and typical fractionation ranges are presented in Table 2.3. The proteins of interest should elute between the extremes of $K_d = 0$ and 1. The fractionation range or the most useful working range of some common gels is given in Table 2.4.

Table 2.3 Supports for gel filtration chromatography

	Fractionation ranges (kDa)
Low-pressure supports:	
Sephadex: Dextran, cross-linked with epichlorohydrin	≤0.7-600
Sepharose: Agar, porosity depending on amount of agar in matrix	10-40,000
Bio-Gel A: Agar	<10-150,000
Sepharose CL: Sepharose cross-linked with 2,3-dibromopropanol	10-40,000
Sephacryl: Allyl dextran cross-linked with *N,N'*-methylene bisacrylmide	5-80,000
Bio-Gel P: Polyacrylamide	0.1-400
HPLC supports:	
Silica-based: Porous silica with bonded hydrophilic groups, TSK supports of SW type	5-7000
Organic polymers with TSK supports of PW type: Hydrophilic hydroxylated ether	0.1-8000

Table 2.4 Protein fractionation range of some gels

Sephadex	**kDa**	**HPLC supports**	**kDa**
G-25	1-5	Superose 12	3-100
G-50	1.5-30	Superose 6	5-5000
G-75	3-70	TSK G2000SW	5-100
G-100	4-100	TSK G3000SW	10-500
G-200	5-250	TSK SW	20-7000
Superdex 200	10-600	TSK-30	1-20
Sephacryl S-300	10-1500	TSK-40	2-30

2.4.3 Gel Electrophoresis [50,51]

Gel electrophoresis provides a powerful separation for proteins in complex mixtures. Proteins are separated according to their ability to move under the influence of an electric field. Polyacrylamide gels were first used for electrophoresis by Raymond and Weintraub in 1959.[52] Cross-linked polyacrylmide gels are from the polymerization of acrylamide monomer in the presence of small amounts of *N,N'*-methylene-*bis*-acrylamide (crosslinking reagent) (Figure 2.6). The polymerization of acrylamide is initiated by the addition of ammonium persulfate and the base *N,N,N',N''*-tetramethylenediamine (TEMED). TEMED catalyzes the composition of persulfate ion to give a free radical.

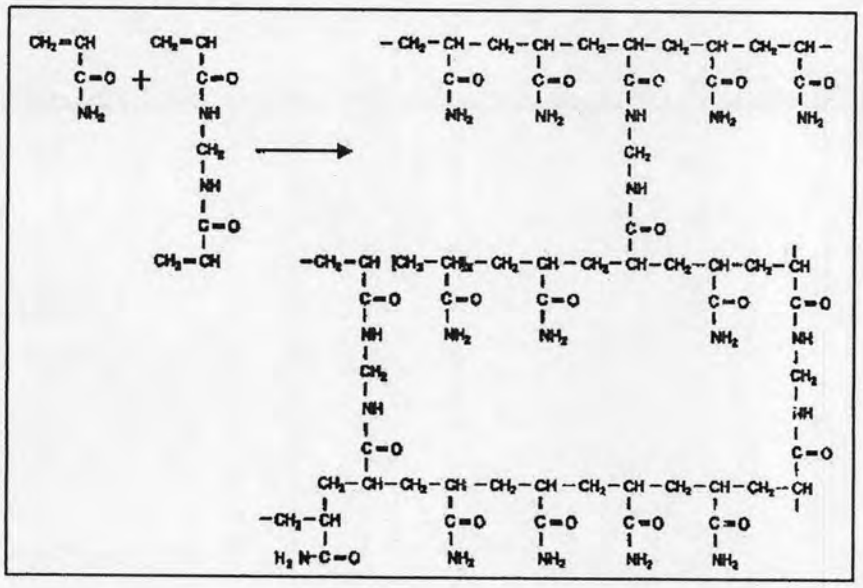

Figure 2.6 The polymerization reaction of acrylamide and methylenebisacrylamide.

The pore size restricts the movement of proteins. The pore size of the gel can be controlled by the total acrylamide concentration *T* and the degree of crosslinking *C*.

$$\%T = [(a + b) \times 100]/V, \quad \%C = (b \times 100)\ (a+b)$$

a is the mass of acrylamide in g.

b is the mass of methylenebisacrylamide is g, and *V* is the volume in ml.

When C remains constant and T increases, the pore size decreases. The range of pore size varies from 4%T to 20 %T. Gel with higher %T has small pore sizes and more restrictive so it favors the movement of smaller proteins with little or no movement of large proteins.

2.4.3.1 Separations Based on Protein Molecular Weight

Sodium-dodecyl sulfate polyacrylamide-gel electrophoresis (SDS-PAGE), reflecting the combination of SDS treatment of proteins with polyacrylamide gel (PAGE), is one-dimensional electrophoresis (1-D electrophoresis). All proteins are given a negative charge and move in the same direction towards the positive electrode. The charge of proteins can be produced by coating with the anionic detergent sodium dodecyl sulfate (SDS). The proteins are denatured and uncoiled by the effect of the SDS saturation, especially when using reducing reagents such as dithiothreitol to cleave disulfide bonds (Figure 2.7). Smaller proteins move more rapidly through the gel than larger proteins such that mixtures of proteins can be separated according to their molecular weight (M_r).

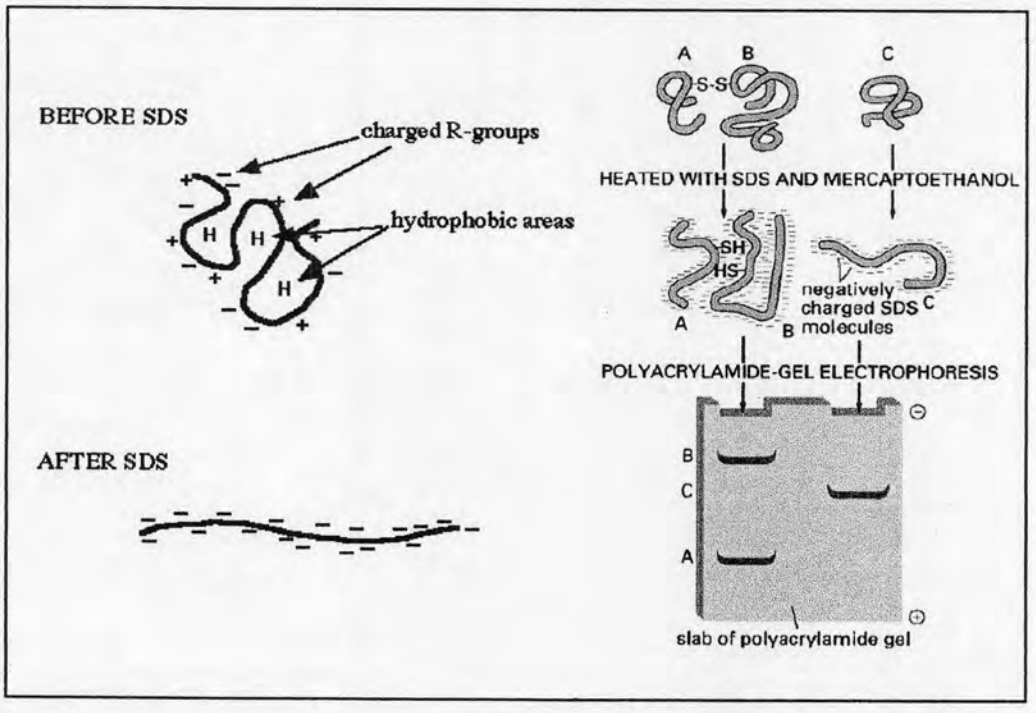

Figure 2.7 Mechanism of SDS to breaks up the interactions between proteins.

2.4.3.2 Separations Based on Protein Isoelectric Point

Isoelectric focusing (IEF) can be described as electrophoresis in a pH gradient set up between a cathode and anode with the cathode at a higher pH than the anode. In electric field, the proteins, which are amphoteric substance, move towards the anode or the cathode until they reach a position in the pH gradient where their net charges are zero. Proteins tend to be positively charged at pH values below their isoelectric pH and negatively charged above. The pH at which a protein has zero net charge is called the "isoelectric point" or "p*I*" of the protein. The protein species migrate and focus (concentrate) at their isoelectric points. The focusing effect of the electrical force is counteracted by diffusion which is directly proportional to the protein concentration gradient in the zone. In an isoelectric focusing gel, the pH conditions are established in polyacrylamide gel by two techniques, carrier ampholytes (low molecular weight amphoteric species) or immobilines (acrylamide derivatives). The most use of isoelectric focusing experiments carried out in immobilized pH gradient (IPG) gel. IPG gel made from mixtures of immobilines with common gradient-mixing systems to establish the pH gradient. The advantages of using IPG gel are allowing a long focusing time to ensure the focusing of the analyte proteins, loading relatively large amount of proteins, available to purchasing IPG in variety of pH ranges, and simplify the physical handling.

2.4.3.3 Two-Dimensional Electrophoresis [53]

Two-dimensional electrophoresis (2-D electrophoresis) is a powerful and high resolution method for the analysis of complex protein mixtures. Several thousand of proteins can be separated in one gel. This technique combines two steps of gel electrophoresis separation, and separates proteins according to two independent properties. The principle of this technique shown in Figure 2.8. The first-dimension step is isoelectric focusing, separates proteins according to their p*I* across the x-axis. Then, all proteins contained in isoelectric focusing gel are saturated with SDS and transferred to SDS-PAGE in the second-dimension step, separates proteins according to their M_r across y-axis. Each spot on the resulting 2-D gel corresponds to a single

protein in the sample. 2-D electrophoresis is also an analytical method for measuring the M_r, p*I*, and relative amount of a protein in the mixture.

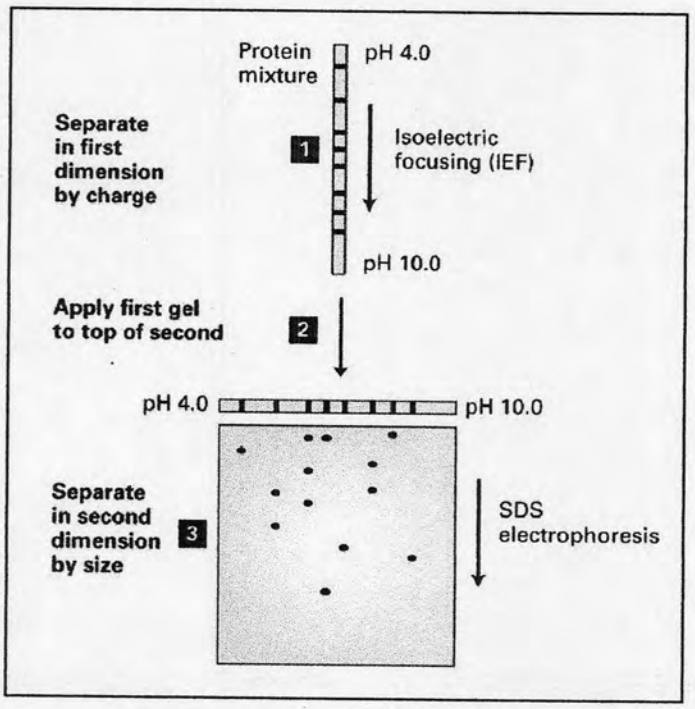

Figure 2.8 Two-dimensional electrophoresis processes.

2.4.3.4 Protein Detection

The mostly applied methods in the detection of separated proteins from gel electrophoresis are Coomassie blue-staining and Silver-staining. Coomassie blue-staining colors the proteins fixed in a polyacrylamide gel with a dark blue dye. Generally, Coomassie blue-staining can detect as little as 0.5 pmol to 1 pmol of protein in a 1-D gel band, and 0.2 pmol to 0.5 pmol for 2-D gel. Silver-staining is an estimated 100-fold more sensitive than Coomassie blue-staining, giving limits of detection for 2-D gel spots in the 5 fmol range.

2.5 Mass Spectrometry [54]

Mass spectrometry (MS) measures the mass weight of molecules in term of the mass-to-charge ratio (m/z). Molecular weight measurements by mass spectrometry are based upon the production, separation, and detection of molecular ions. A typical components of mass spectrometer includes; ion source (Ionizes sample and generates gas phase ions), mass analyzer (Separates ions according to individual mass-to-charge ratios), detector (Detects and amplifies ions). The common two ionization methods for peptides and proteins; electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI), which will be described in detail later. After the gas phase ions have been generated, they are expelled from the ion source and beamed in to mass analyzer, the region of ion separation. Many types of mass analyzer that commonly used for protein analysis applications are time of flight (Tof), quadrupole mass filter, ion trap, combinations of quadrupole Tof and Tof/Tof. Herein, Tof and quadrupole mass filter are mentioned here because these two types of mass analyzer were used in this research.

2.5.1 Matrix Assisted Laser Desorption Ionization/Time of flight Mass Spectrometry (MALDI/Tof MS) [55-56]

2.5.1.1 Matrix-Assisted Laser Desorption Ionization (MALDI)

Karas and Hillenkampm[57,58] introduced Matrix Assisted Laser Desorption/Ionization (MALDI) technique that could readily ionize biomolecules in a very sensitive manner. MALDI is a pulsed ionization technique which utilises the energy from a laser to desorp and ionize the analyte molecules in the presence of a light absorpting matrix (Figure 2.9). MALDI ions are created by mixing the analyte with the small organic molecule (the matrix) which absorbs at the wavelength of the laser. The analyte becomes incorporated into the crystal lattice of the matrix and is then irradiated with a laser. The laser causes the desorption and ionization of matrix and analyte, either by protonation (positively charged ions) or desorption (negatively charged ions). The ions are then accelerated into the MS analyzer. As MALDI is a pulsed ionization technique, which is ideally coupled with a Tof analyzer.

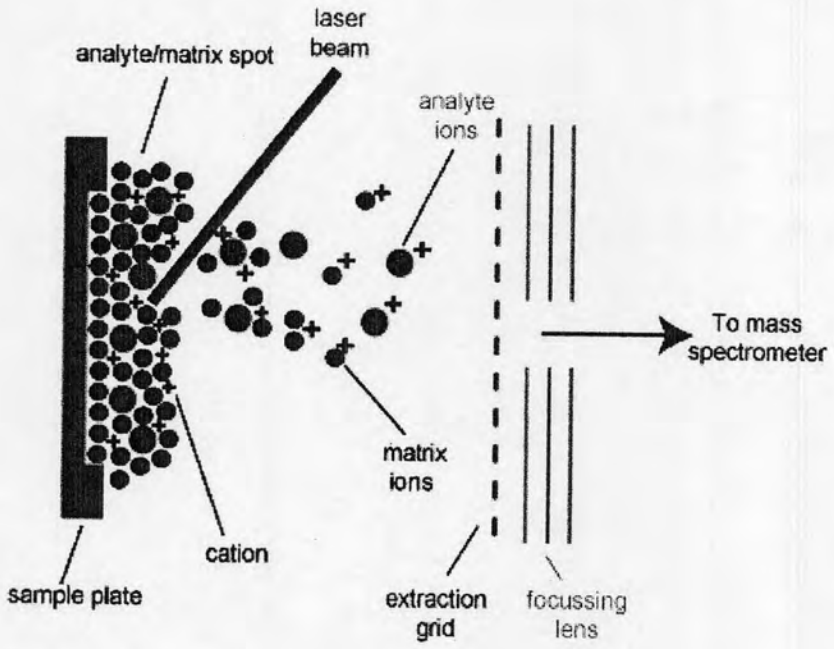

Figure 2.9 Matrix Assisted Laser Desorption/Ionization (MALDI) source.

Among the chemical and physical ionization pathways suggested that MALDI are gas-phase photoionization, excited-state proton transfer, ion-molecule reactions, desorption of preformed ions, etc. The most widely accepted ion formation mechanism involves gas-phase proton transfer with photoionized matrix molecules (Figure 2.10). MALDI produces predominantly singly charged ions. Also, MALDI allows to the desorption and ionization of analytes with very high molecular mass in excess of 100,000 Da.

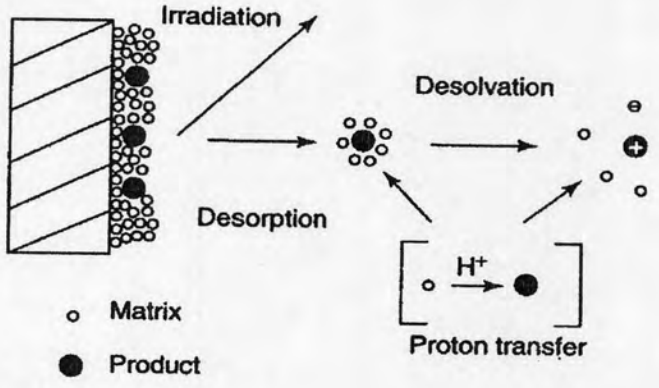

Figure 2.10 MALDI ion formation.

The typical wavelength of the UV lasers utilised is 337 nm. MALDI ions are generated under high vacuum ($5x10^{-6}$mbar). A wide range of matrices for biological mass spectrometry applications have been adopted for use with UV lasers (Table 2.5).

Table 2.5 Common MALDI matrices used in biological applications

Matrix	Application
α-cyano-4-hydroxycinnamic acid	UV laser Peptides analysis & protein digests Analytes < 10 kDa
sinapinic acid (3,5-dimethoxy-4-hydroxycinnamic acid)	Analysis of large polypeptides & proteins Analytes > 10 kDa
2,5-dihydroxybenzoic acid (2,5-DHB)	UV laser Protein digests & Proteins Oligosaccharides released from Glycoproteins
2,4,6-trihydroxyacetophenone (THAP)	UV laser Oligonucleotides < 3 kDa
3-hydroxy picolinic acid	UV laser Oligonucleotides > 3 kDa

The three widely used matrices for peptides and proteins are α-cyano-4-hydroxycinnamic acid (CCA), 2,5-dihydroxybenzoic acid (2,5-DHB), and sinapinic acid.

2.5.1.2 Time of Flight Mass Analyzer

Tof analyzer is one of the simplest mass analyzing devices and commonly used with MALDI ionization. Tof analysis is based on accelerating a set of ions to a detector with the same amount of kinetic energy. Because the ions have the same energy, yet a difference mass, the ions reach the detector at difference times. Mass-to-charge ratios are determined by recording the time that ions use to move in a field free region (a flight tube) from the source to the detector. The arrival time of an ion at the detector is dependent upon the mass, charge, and kinetic energy of the ion. A beam of ions is accelerated through a know potential V_s. When ion leaving the source, an ion with mass m and total charge $q = ze$ has a kinetic energy E_k

$$\frac{mv^2}{2} = qV_s = zeV_s = E_k$$

The time t need to fly the distance d is given by $t = d/v$

Replacing v by its value in the previous equation gives

$$t^2 = \frac{m}{z}\left(\frac{d^2}{2V_s e}\right)$$

The mass-to-charge ratio can be calculated from a measurement of t^2, the term in parentheses being constant. From the equation show that, all others being equal, the lower the mass of ion, the faster it will reach the detector. The linear Tof mass spectrometer is the simplest example of Tof analyzer (Figure 2.11).

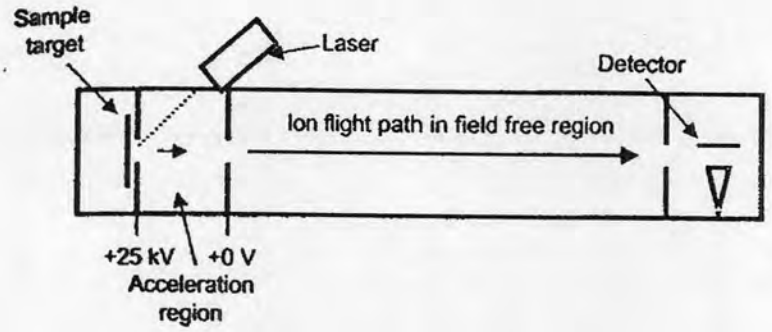

Figure 2.11 Linear time of fight mass spectrometer.

In principle, the upper mass range of a Tof instrument has no limit, which suitable for soft ionization techniques. The advantages of this instrument are high sensitivity and very fast scan speed. The drawback of the linear Tof analyzers is poor mass resolution. Mass resolution is affected by factors that create a distribution in flight times among ions with the same *m/z* ratio. Time of fight reflectron mass analyzer has been developed to improvement of mass resolution. The reflectron is an ion mirror, created by an electric field that reverses the flight path of ion (Figure 2.12). The most significant effect of reflectron is to focus ions with the same *m/z* but different velocities. Focusing is accomplished because these ions penetrate the reflectron to different degrees. The higher-velocity ions penetrate farther, and spend a longer time in the reflectron than the lower velocity ions. Another effect of a reflectron is to increase the effective flight tube. The longer Tof flight tube has better resolution.

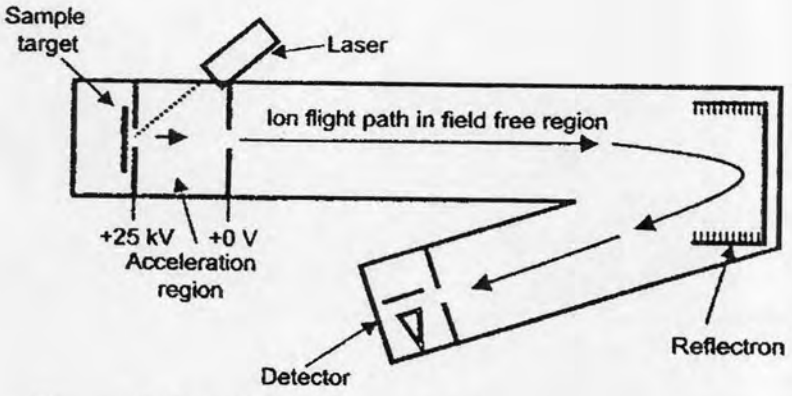

Figure 2.12 Reflectron time of fight mass spectrometer.

As mentioned above, the improving resolution and sensitivity of time of fight reflectron mass analyzer is an important contributor to the mass observed in MALDI/Tof experiment. Also, it is an important consideration in the development of hybrid quadrupole-Tof instruments for tandem mass spectrometry described in detail later.

2.5.2 Electrospray Ionization Mass Spectrometry [54-56, 59]

2.5.2.1 Electrospray Ionization (ESI)

Fenn and co-workers described the use of electrospray for ionizing large biomolecule in 1989.[60,61] In electrospray ionization, analytes that acidic aqueous solution is sprayed through a small-diameter needle. Therefore, ESI ionizes is readily coupled to liquid-based (for example, chromatographic and electrophoretic). A high positive voltage is applied to the needle to produce a Taylor cone from which droplets of the solution are sputtered (Figure 2.13).

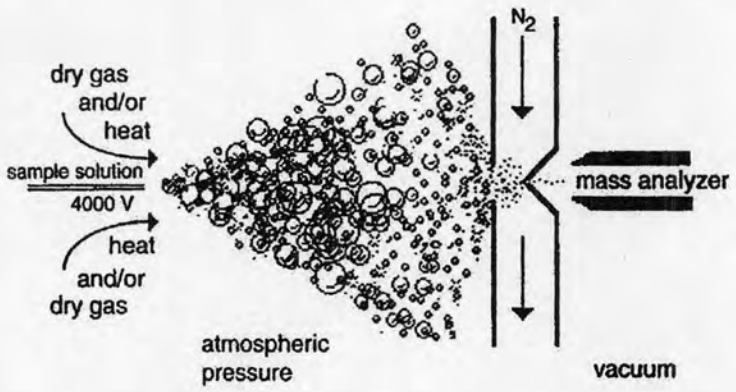

Figure 2.13 The electrospray ionization source.

Protons from the acidic conditions give the droplets a positive charge, causing them to move from the needle towards the negatively charged instrument. During this movement, evaporation, which processed by flow of gas (nitrogen) and heat, reduces the size of the droplets until the number and proximity of the positive charges spilt the droplet into a population of smaller, charged droplets. The evaporation and droplet-splitting cycle repeats until nanometer sized droplet are produced and can be directed into the mass spectrometer by appropriate electric field (Figure 2.14).

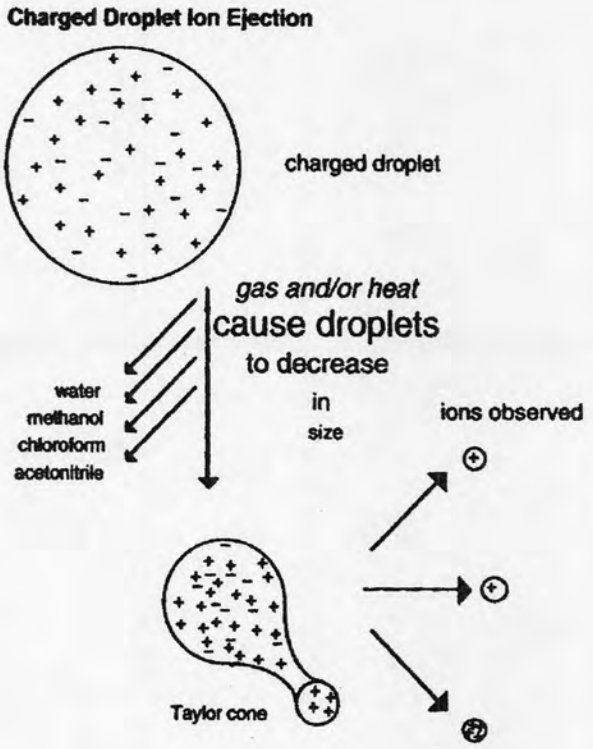

Figure 2.14 The electrospray ionization formation.

The charges are statistically distributed over the analyte's potential charge sites, enabling the formation of multiply charged ions. Each multiply charged ion can be termed a charge state, and a distribution of charge states is characteristic of large macromolecules during ESI analysis. The high electric field in the electrospray source produces multiply-charged ions.

For which

$$\frac{m}{z} = \frac{M + nH}{n}$$

where, M = molecular weight, n = number of charges and H = mass of proton
If a positive ion series is assumed to represent different protonation states, then the mass-to-charge ratio, A_1 and A_2, of adjacent ions in a multiply charged ion series appear at m/z of; $n_1 = n_2 + 1$, where n_1 is the number of charges on A_1 and n_2 is the number of charges on A_2. Then,

$$\frac{M + n_1H}{n_1} = A_1 \quad \text{and} \quad \frac{M + n_2H}{n_2} = A_2$$

Therefore, $$n_2 = \frac{A_1 + H}{A_2 - A_1}$$

Thus, the estimation of mass and charge number can be calculated.

2.5.2.2 Quadrupole Mass Filter

The quadrupole mass filter is composed of four parallel hyperbolic rods, which the gas phase ions have to achieve a stable trajectory. The analyzer is operated by the application of a voltage (DC) and oscillating voltage (Radio Frequency, RF) to one pair of rods and DC voltage of opposite polarity and RF voltage of different phase to the opposite pair of rods (Figure 2.15). The alternating electric field helps to stabilize and destabilize the passing ions. The ions traverse through the space between the rods, according to equation noted in Figure 2.15, and only at specific voltages applied to the rods will certain m/z values be allowed to pass through the rods and reach the detector. The voltages are scanned to allow a wide mass range to be observed.

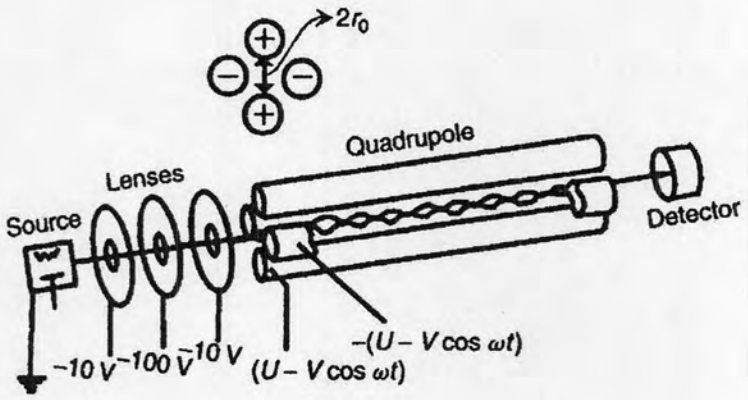

Figure 2.15 The quadrupole instrument.

The ion trajectory can be considered with Mathieu equation, leading to definition of two parameters.

$$a = \frac{8zeU}{m r_o^2 \omega^2} \qquad q = \frac{4zeV}{m r_o^2 \omega^2}$$

Where U = direct current (DC) potential, (V + cos ωt) = radio frequency potential,

ze = charge of ion and r_o = ½ of the distance between opposite rods

The parameter r_o and ω are maintained constants. U and V are the variables. The relation between a and q is

$$\frac{a}{q} = \frac{2U}{V} = \text{constant}$$

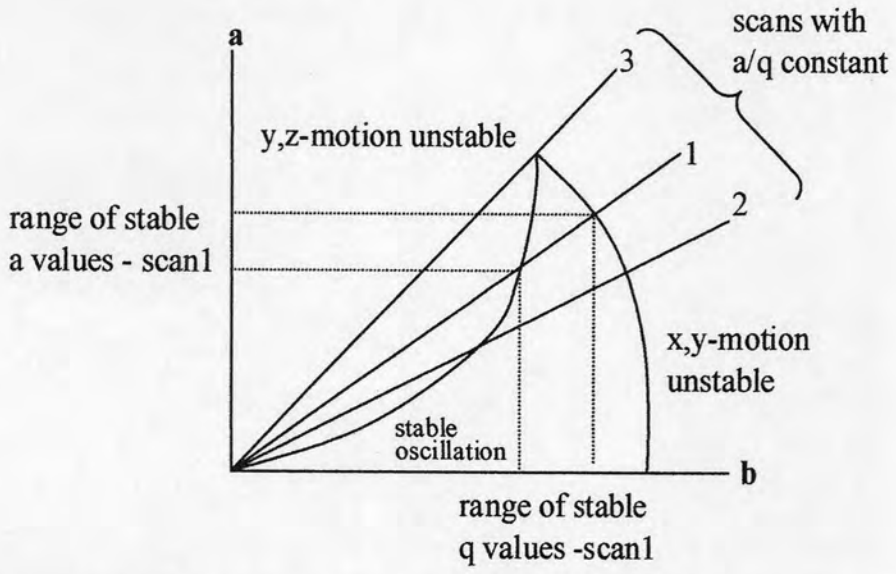

Figure 2.16 The stability region for ion trajectories in the quadrupole mass filter.

For the graph in Figure 2.16, only the stable oscillation area of the graph contains values for both a and q that define stable ion motion along the z-axis, thereby allowing ions to pass along the axis of the quadrupole rods to the detector. For the other values of a and q ions diverge far enough from the z-axis that they collide with, or pass between, the rods. The lines marked 1, 2 and 3 in Figure 2.16 are called scan lines and correspond to variation of U and V such that U/V (a/q) remain constants. The quadrupole are low resolution instrument and thus function as a mass filter. Quadrupole mass analyzers have found commonly utility to interface with ESI and applied for using as instruments in tandem MS (Triple quadrupole and quadrupole Tof).

2.6 Tandem Mass Spectrometry [54, 56, 62]

Tandem mass spectrometry (MS/MS) is general method involving at least two stages of mass analysis. The first analyzer (MS1) is used to isolate a precursor ion, which then undergoes activation a fragmentation to yield product ions. The product ions are subsequently analyzed by a second mass analyzer (MS2). Tandem mass analysis is primarily used to obtain structural information. A tandem mass spectrometer can be classified into two types: tandem-in-space and tandem-in-time. The in-space mass spectrometers have two mass analyzers, and each mass analyzer performs separately to accomplish the difference stages of experiment. The classic example of tandem-in-space instrument is the tandem quadrupole instrument, commonly referred to as a "triple quadrupole" mass spectrometer. The combination of a quadrupole with time-of-flight instrument and time-of-flight with a reflectron are also used as tandem-in-space instruments. The tandem-in-time mass spectrometry performs as appropriate sequence of events in an ion storage device. The in-time instruments have only one mass analyzer. An ion trap mass spectrometer and ion cyclotron resonance instruments, also known as Fourier transform mass spectrometers, are the examples of tandem-in-time systems.

There are three modes of scanning in tandem MS: product ion scan, precursor ion scan, and neutral loss scan (Figure 2.17). These scan modes are distinguished by the relationship between the first and second stage of mass analysis. In standard mode of mass analysis, the first mass analyzer and the collision cell transmit all ions for mass analysis in the second mass analyzer. The product ion scan consists of selecting a precursor ion in the first mass analyzer and determining the product ions resulting from collision-induced fragmentation (CID) in second mass analyzer. The precursor ion scan consists of choosing a product ion and determining the precursor ions. The first mass analyzer is scanned to sequentially transmit the mass-analyzed ions into the collision cell for fragmentation. The second mass analyzer mass-selects the product ion of interest for transmission to the detector. All of precursor ions that produce ions with the selected mass through fragmentation thus are detected. The Neutral loss scan consists of selecting a neutral fragment and detecting all the fragmentations leading to the loss of that neutral fragment. Mass analysis in both mass analyzers is carried out by linking the scanning mass analyzers with a constant mass difference. Fragmentation occurs in the collision cell. The detector signal is the result of precursor ions fragmenting to lose a specific neutral species, forming a product ion with a characteristic mass difference. In tandem instruments, only the tandem quadrupole instrument performs all of the experiments illustrated in Figure 2.17. The precursor ion scan and neutral ion scan mode cannot be preformed with time-based mass spectrometers. Because of the scan requires the focusing of the second spectrometer on a selected ion in the precursor ion scan mode and the scan requires that both mass spectrometers are scanned together in the neutral loss scan mode.

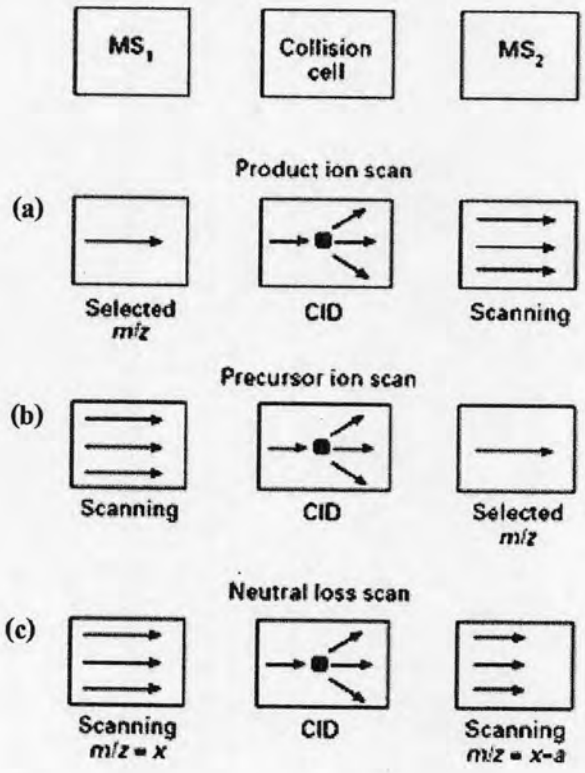

Figure 2.17 A schematic representation of the tandem mass spectrometry scan modes.

2.6.1 Collisionally Activated Dissociation (CAD)

The fragmentation can be achieved by inducing ion-molecule collisions by a process know as collisionally activated dissociation (CAD) or collisionally induced dissociation (CID). The CAD process as a sequence of two steps. The first step is very fast and corresponds to the collision between the ion and the target molecule when a fraction of the ion translational energy is converted into internal energy, bringing the ion into an excited state. The second step is the unimolecular decomposition of activated ion. CAD is accomplished by selecting an ion of interest with a mass analyzer and introducing that ion into the collision cell, where the selected ion collides with the collision gas (Ar or He) atoms, resulting in fragmentation. For collision energy conversion to internal energy, the kinetic energy for internal energy transfers is controlled by the collisions of mobile species (the ion) and a static target (the collision gas). In practice, there are two groups of collision energy: high-energy collision, in the keV range, and low-energy collision, in the range of 1-100 eV. The high-energy collision is common for magnetic instruments

and also uses helium as the common target gas. The low-energy collision occurs in quadrupole or ion trap instruments. The collision gas is more important than it is for the high-energy collisions. Heavier gases such as argon, xenon or krypton are preferred because they allow the transfer of more energy. In comparison of two collision energy, the different fragmentation patterns are observed. The high-energy CAD spectra give simpler, more clear-cut fragmentation, whereas low-energy CAD spectra lead to more diverse fragmentation pathways, often including more rearrangements.

2.6.2 Electrospray Ionization Quadrupole/Time of Flight Mass Spectrometer (ESI-Q/TOF MS)

ESI-Q/TOF mass spectrometry has electrospray ionization for ion source, a quadrupole mass filter for the mass analyzer and a time-of-flight mass analyzer for the second mass analyzer (Figure 2.18). ESI-Q-TOF MS used for molecular weight measurements, reaction monitoring, protein structural studies, peptide sequencing, nucleotide sequencing, macromolecule structure determination due to extended m/z range. Quadrupole time-of-flight or hybrid analyzer is described in 1996 by Morris for oligosaccharide analysis and more recently Lobada *et al.*, 2000; these instrument have rapidly become the instrument standard for MS/MS applications with in the theatre of proteomics. The user is able to acquire MS and most notably MS/MS data with high mass accuracy, resolution and sensitivity. The instrument is generally with interfaced with HPLC.

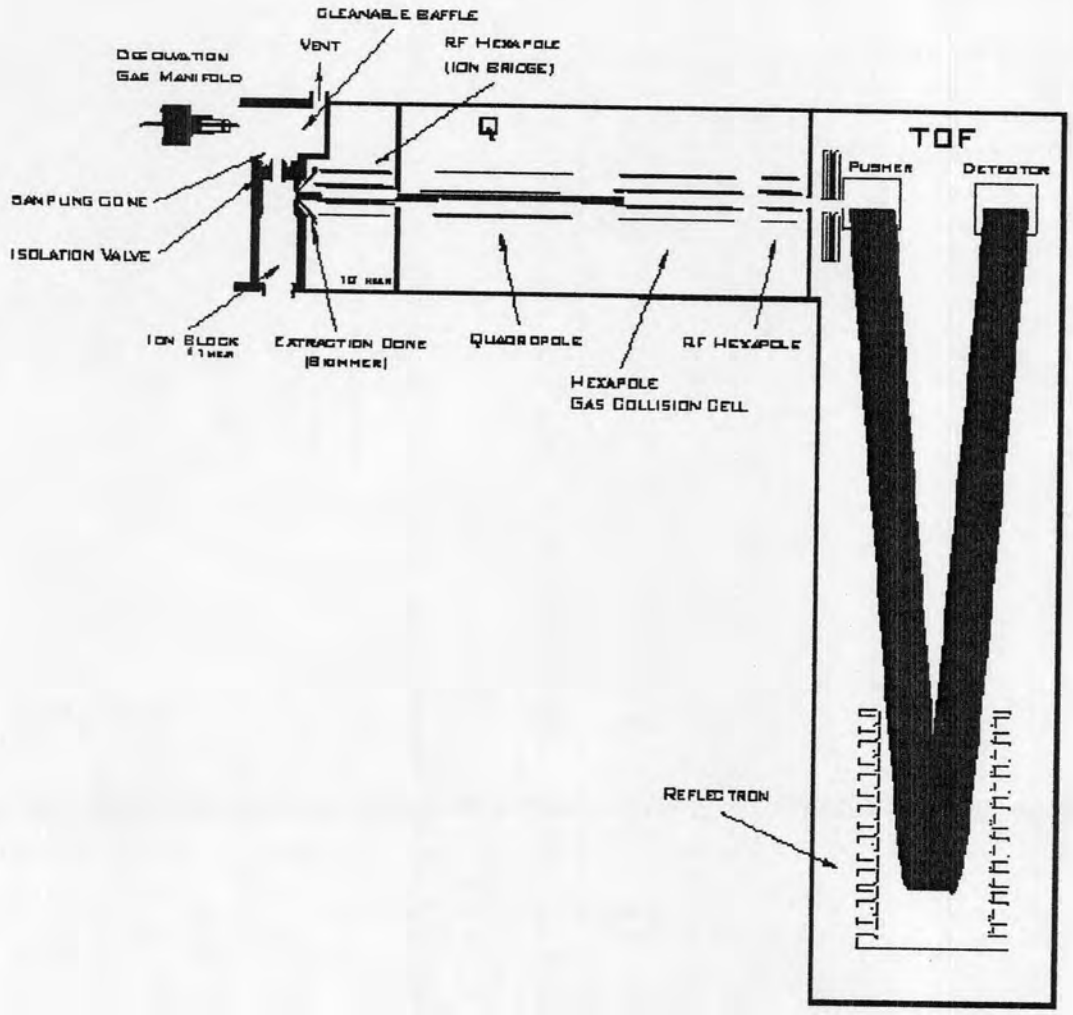

Figure 2.18 Schematic diagram of Electrospray ionization quadrupole/ Time of flight mass spectrometer.

2.6.3 Tandem Mass Spectrometry for Peptide Sequencing

Tandem mass spectrometry obtains structural information on peptide and proteins by CAD, in which precursor ions are subjected to collisions and the structure of resulting ions is characterized. As result of the CID fragmentation the peptide precursor ion fragments predictably at each peptide amide bond along the peptide backbone yielding a distribution of product ions in complimentary ion series forming a ladder which is indicative of the peptide sequence. The nomenclature suggested by Roepstroff and Fohlman and used to describe the different product ions defines two set of ions that are named based on the peptide terminus retained in the ion. The a-, b-, c-ions all contain the N-terminus of the peptide, while the x-, y-, z-

ions all contain the C-terminus (Figure 2.19). There is a marked difference between the fragmentation observed at high and low energy. At high energy described in Figure 2.20 can be generated. The high-energy tandem mass spectra present a greater number of fragment ions but also increasing the complexity and difficulty of interpretation. At low energy, the observed fragments are mostly b- and y-ion. These fragments then lose small molecules such as water or ammonia from functional on amino acid side chains.

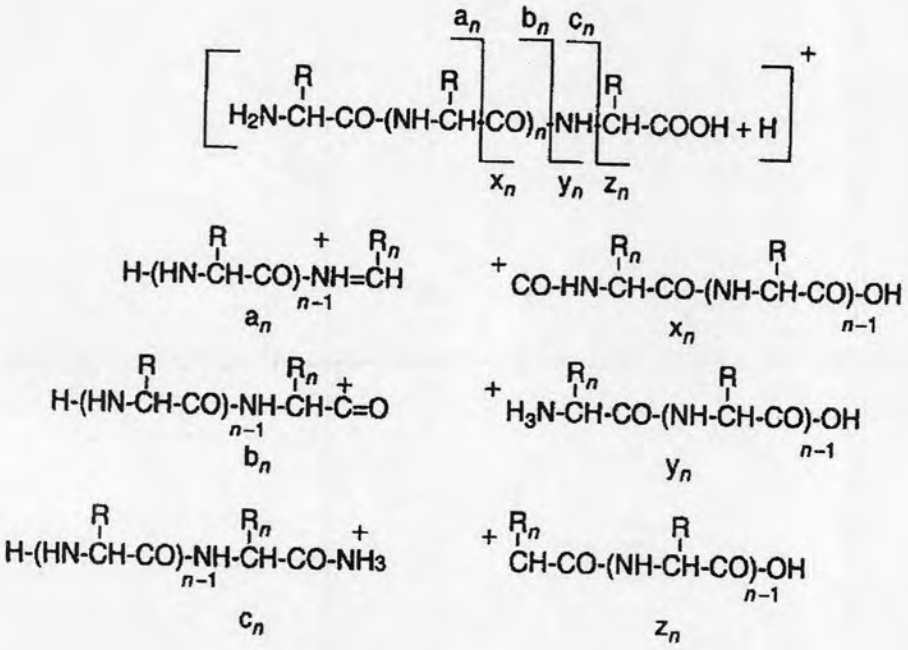

Figure 2.19 The peptide fragmentation.

In addition, there are two other types of fragments which appear among the low masses in the spectrum (Figure 2.20). The first type is called an internal fragment because these fragments have lost the initial N- and C-terminal sides. The second type of fragment is immonium ions of amino acids, labeled by a letter corresponding to the parent amino acid code. Even though these two fragments are rarely observed for all of peptide amino acids, those that appear yield information concerning the amino acid composition of the sample and also confirm the sequence.

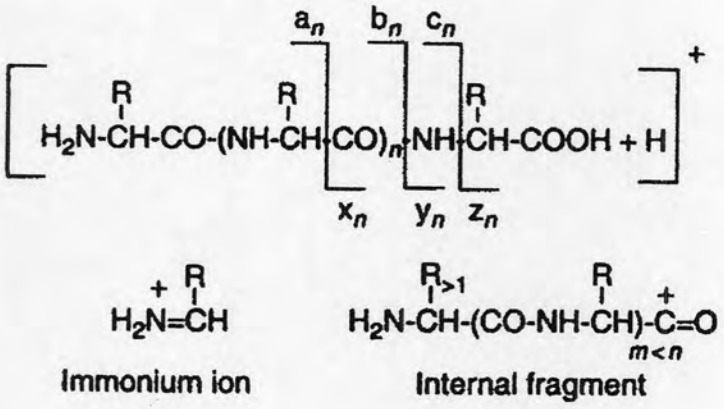

Figure 2.20 Immonium ion and internal fragment.

The mass difference between consecutive ions within a series allows one to determine the identity of the amino acid (Table 2.6) and thus deduce the peptide sequence.

Table 2.6 The Masses of common 20 amino acids

Amino acid	Code (3 letters)	Code (1 letter)	Residue mass (Da)	Immonium ion (m/z)
Glycine	Gly	G	57.02	30
Alanine	Ala	A	71.04	44
Serine	Ser	S	87.03	60
Proline	Pro	P	97.05	70
Valine	Val	V	99.07	72
Threonine	Thr	T	101.05	74
Cysteine	Cys	C	103.01	76
Leucine	Leu	L	113.08	86
Isoleucine	Ile	I	113.08	86
Asparagine	Asn	A	114.04	87
Aspartate	Asp	D	115.03	88
Glutamine	Gln	Q	128.06	101
Lysine	Lys	K	128.09	101
Glutamate	Glu	E	129.04	102
Methionine	Met	M	131.04	104
Histidine	His	H	137.06	110
Phenylalanine	Phe	F	147.07	120
Arginine	Arg	R	156.10	129
Tyrosine	Tyr	Y	163.06	136
Tryptophan	Try	W	186.08	159

2.7 Identification of Proteins

To date, the term proteome analysis or proteomics is well known in biological field. This term was initially named in 1994 at the first Sienna two-dimensional electrophoresis meeting, and defined as the full complement of proteins expressed by a genome.[51,55] Proteins can also be viewed as being composed of smaller multi-amino acid subunit called peptides. It is often either necessary or advantageous to work with peptides rather than proteins. In instances, the protein is chemically or enzymatically digested by hydrolyzing selected amide bonds. Peptides have the same amino acid composition, N- and C-terminus as proteins. The amino acid sequence establishes the identity of protein and defines the primary structure of protein. The two commonly applied methods of generating amino acid sequence information are Edman degradation and mass spectrometry.[51,56] However, Edman degradation is time consuming method. The time required to complete each cycle is about 45 minutes, which limit the analysis of on more two or three samples per day. Furthermore, most native proteins have blocked N-terminally. There is because of post-translational modifications such as the addition of formyl or acetyl groups. Nowadays, mass spectrometry is the most efficient way to identify proteins, especially the short time analysis, the high-sensitivity, and high information content in protein sequencing.[54,56]

MS has become a powerful tool in the proteomic research that has moved dramatically forward with great improvements in mass spectrometers and the subsequent development of improved database searching algorithms. The scale of a proteomic experiment varies according to its aim, such as determining the protein content of a whole organism, analyzing protein in a tissue or targeting proteins in cell sub-compartments. Analysis of complex mixture is significantly more complicated than the analysis of single or pure compounds. In addition to the dynamic range of proteins in their native cellular environment, which cannot be readily accommodated within a single mass spectrometric analysis, different components of a mixture will have differing proton affinities. The ionisation efficiencies of these components will therefore changed and some components may not be detected at all. Independent of whether a protein is being analysed by MS in order to ascertain its primary sequence,

post-translational modifications or structural properties, it usually need to be isolated or purified to enable detection. This can be done either from a native source (e.g. cells, tissues, plasma) or following exogenous expression. A variety of purification techniques may be used for purification from either of these systems, including gel electrophoresis separation (SDS-PAGE) and chromatography by virtue of protein characteristics such as charge, size or hydrophobicity.

Generally, identification of the proteome contains greater difficulties than genome, identification of genes, even mass spectrometry (MS)-based proteomic, taking advantage of high-throughput systems allowing automated proteome analysis. Identification of genomes of several species from the simplest mycoplasm,[63] to the complex, human,[64] has now been elucidated. There are three major differences between genome and proteome analysis. The first difference is that the genome is static, while the proteome of each living cell is dynamic response to the individual metabolic state of cell and reception of intracellular and extracellular signal molecules. Thus while the genome enables a prediction of the proteome simply as the gene products, this cannot be described as the proteome. Also, many proteins which are expressed will be post-translationally by one or more of approximately 200 modifications.[65-68] The second difference concerns the relative amounts of the components within the genome and proteome. For example, the low-expression, rapid-turnover, protein involves in dynamic cellular processes; signal-transduction mechanism. The last difference is functional proteomics study that requires quantification, so have to face an extreme technical and intellectual challenge.

Despite the problems mentioned to above, interest in plant proteomics has increased rapidly in recent years. Areas of specific interest have included the study of protein variations in different plant organs,[69,70] variations in response to physiological events,[71] identification of unknown plant viruses due to their proteome[72] and the identification of microtubule binding proteins in plants.[73] A number of review articles have already been published concerning the study of the plant proteome.[74-78] However, protein analysis of the plant kingdom is still at on early stage. Only three species have yield the complete genomics sequence projects, including rice, poplar and *Arabidopsis thaliana.* Proteomic analyses have been undertaken in a great variety of plants, including potato, tomato, tobacco, grape and spinach. The aim of these

comprehensive analyses of biological network component is to extend our understanding to the system level. At the present, numerous MS-based plant proteomic studies of specific organisms, tissue, and sub-cellular have been performed,[79-84] however, it is far less behind than unicellular prokaryotes, animals and yeast.[85-86] In addition, to the general problem of proteomic as was mentioned above, including physically and chemically of protein much more diverse than nucleic acids, different RNA splicing and post-translational modifications increase the number of protein species several-fold compared to the number of genes, plant proteome approaches face species challenges due to the rigidity of plant cell walls and the accumulation of large quantities of secondary compounds in the vacuole of some plant species. Moreover, the number of fully sequenced plant species is limited that is the cause of another factor limiting plant proteomics studies. The strategy for sequencing a protein using mass spectrometry starts with the precise determination of the molecular weight of that protein using MALDI or ESI MS (Figure 2.21). Accurate determination of molecular weight of a protein is useful for its identification and the determination of its purity.

There are two approaches for protein identification using mass spectrometric data by comparing the mass spectrometric data with the information that has been gathered into the sequence databases. The first approach is using mass spectrometry to detect mass of intact proteins and their digested peptides. These mass spectrometric data are compared with the data of known proteins in database. The second approach is using tandem mass spectrometry and CAD to obtain peptide amino acid sequence. The examples of the protein sequence databases are the GenBank database (www.ncbi.nlm.nih.gov/Genbank/index.html), the SWISS-PROT database (http://www.expasy.ch/sprot/sprot-top.htmL), TrEMBL (http://www.expasy.ch/srs5/), and dbEST (http://www.ncbi.nLm.nih.gov/dbEST/).

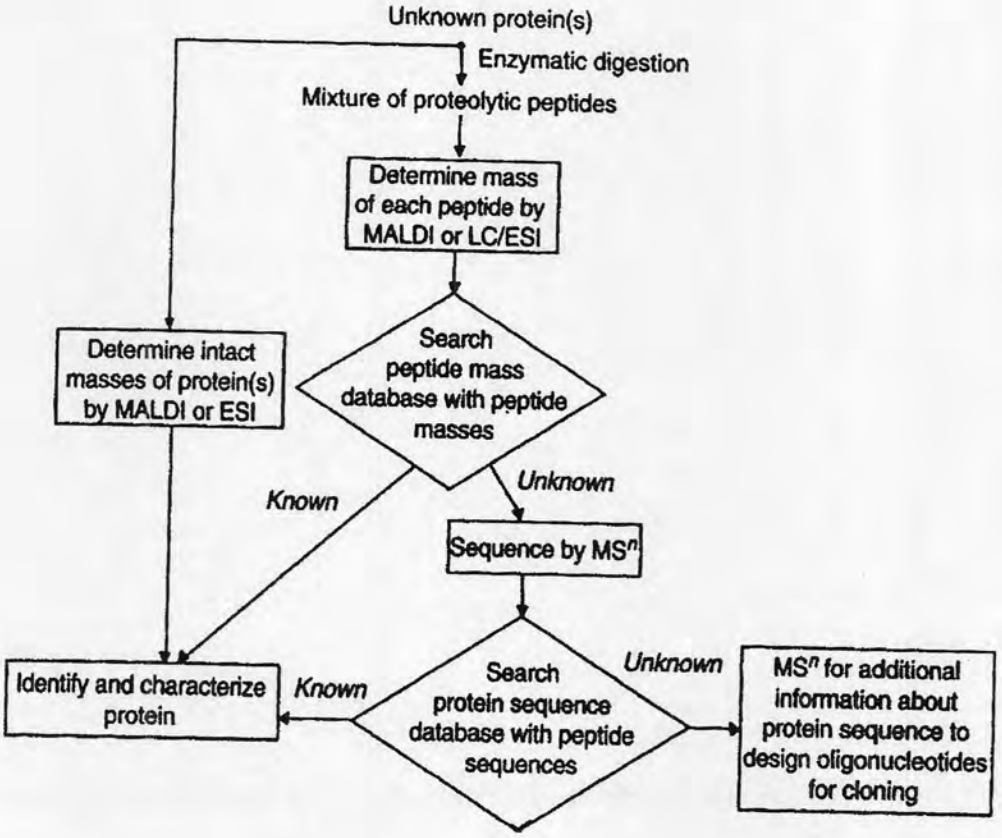

Figure 2.21 Strategy for the identification of proteins based on mass spectrometry.

2.7.1 Protein Identification Using Peptide Mass Mapping

The peptide mass mapping usually performed using MALDI-MS. The method comprises protein digestion, MALDI/Tof analysis and sequence database search algorithms. The protein of interest can be enzymatically or chemically cleaved into its constituent peptides. There are many enzymes and chemical reagents available for protein cleavage (Table 2.7). The enzymatic digestion offers several advantages, including high specificity, minimal side reactions and good cleavage. The most common used enzyme is trypsin, which specifically cleaves protein on the C-terminal side of Lys and Arg, except those bonds are to proline (P) residues (K-P or R-P bonds). There are two types of digestion, in-gel digestion and in-solution digestion. The in-gel digestion methodology has become routine for proteins separated by 2D electrophoresis. Briefly, the in-gel digestion steps are protein spot picking, destaining, reduction and alkylation, tryptic digestion, and extraction tryptic peptides for analysis. For protein separated by other techniques (not gel

electrophoresis), the digestion method is carried out by digest proteins in solution, so called "in-solution digestion".

Table 2.7 Specific chemical and enzymatic cleavage of protein

(The cleavage is carboxyl side of X residue and amino acid side of Y residue)

Reagent	Cleavage site
Chemical cleavage	
Cyanogen bromide	Met-Y
Formic acid	Asp-Pro
Hydroxylamine	Asn-Gly
2-Nitro-5-thiocyanobenzoate	X-Cys
Phenyl isothiocyanate (Edman)	Terminal amino group of peptide
Enzymatic cleavage	
Trypsin	Arg-Y, Lys-Y
Chymotrypsin	Tyr-Y, Phe-Y, Trp-Y
Endoprotease V8	Glu-Y (Asp-Y)
Endoprotease Asp-N	X-Asp (X-Cys)

The resulting peptides from protein digestion are mass analyzed. The experimental peptides masses are compared with expected values of theoretical peptide masses from sequence database. The score is calculated and assigned. The score reflects the match between the theoretically and experimentally determined masses. Then, the protein can be identified as the most probability match. Several programs are available to perform the peptide mass mapping search. There are MASCOT at www.matrixscience.com, profound at www.prowl.com, and MS-FIT at www.prospector.ucsf.edu/. The four important parameters for a peptide mass mapping search are peptide mass list, the cleavage agent, error tolerance (mass accuracy), and knowledge of peptide modifications. The peptide mass mapping is the easiest and fastest technique to identify proteins. However, not all the proteins using this technique can be identified. This can be caused by several reasons, for instance: insufficient of peptide information, post-translational modifications of the proteins, and query protein not exist in the protein database. Then, more specific information than molecular weight of peptides is required to identify protein, such as amino acid sequence using tandem MS and N-terminal-sequencing.

2.7.2 Protein Identification Using Tandem Mass Spectrometry

Tandem mass spectrometry can produce information specific to the amino acid sequence of peptide. The 'peptide sequence tag' is the approach using the peptide sequence information obtained by MS/MS. This approach allows a protein to be identified from a partial sequence and from mass differences between this sequence and the N-terminal and the C-terminal of the peptide resulting from cleavage of the protein. The other method that based on MS/MS but does not require interpretation in terms of the sequence of observed fragments is the method using the SEQUEST algorithm. An algorithm SEQUEST has been developed for the interpretation of fragmentation spectra from tandem MS. A correlation is searched between the fragmentation spectrum and peptide sequences contained in a database. First, the algorithm looks for all the peptides in the database that are the same mass as the precursor ions. Then, a measure of the similarity between the predicted fragments from the sequence obtained from the database and the fragments in the sample spectrum allows the most probable sequence to propose. The sequence or even part of the sequence of a peptide is more specific than its molecular weight. There are many database search programs for use with mass spectrometric protein sequencing data (Table 2.8). The programs have been divided into three categories: programs that use amino acid sequences that must be produced by the interpreting the spectra, programs that use peptide molecular weight information, and programs that use the data from uninterpreted product ion spectra.

Table 2.8 Examples of the databases search programs and their internet addresses that can be used for protein identification

Program Name	Internet Address
Programs that use amino acid sequences for the search query	
FASTA	fasta.bioch.virginia.edu/fasta
BLAST	www.ncbi.nlm.nih.gov/blast/blast.cgi
MS-Edman	prospector.ucsf.edu/mshome3.2.html
Programs that use peptide molecular weights for the search query	
MS-Fit	prospector.ucsf.edu/mshome3.2.htm
MOWSE	srs.hgmp.mrc.ac.uk/cgi-bin/mowse
PeptideSearch	www.mann.embl-heidelberg.de/Services/PeptideSearch/PeptideSearchIntro.html
Programs that use the data from uninterpreted product ion spectra for the search query	
SEQUEST	thompson.mbt.washington.edu/sequest/
MS-Tag	prospector.ucsf.edu/mshome3.2.htm
PeptideSearch	www.mann.embl-heidelberg.de/Services/PeptideSearch/PeptideSearchIntro.html

2.8 Literature Reviews

There are a number of reports on the studies of the *Pithecellobium dulce* in the literature. Most of these studies involve the compounds isolated from different parts of *P.dulce* and biological activities of their extracts or the chemical constituents. In 1968 and 1970, Nigam *et al.* isolated the antitumor compound of sterol glucosidal mixture (sitosterol, campesterol, stigmasterol and spinasterol) from the heart wood and the fruit pulp of *P.dulce* [87] and reported that the alcoholic extract of the leaves of this plant on solvent segregation and chromatography of the different fractions yield octacosanol, beta-D-glucoside of alpha-spinasterol, alpha-spinasterol and kaempferol-3-rhamnoside.[88] In 1980, Zapesochnaya *et al.* found a flavonoid mixture of kaempferol and isoquercetin effective against the bacteria (*Escherichia coli, Bacillus anthracis* and *B. subtilis*) and fungi (*Aspergillus niger* and *A. flavus*) from the leaves of *P.dulce*.[89] In 1985, Adinarayana *et al.* reported the cyclitol, dulcitol from acetone leaf extract of this plant.[90] Saxena *et al.* found a novel isoflavonoidal glycoside, Geinstein 4'- O-alpha -L rhamnopyranoside from the ethanolic root extract of this plant in 1998 and a new flavonoid, 3'-prenyl apigenine-7-O-rutinoside, from *P. dulce* stem in 1999.[91,92] The bark contains tannin of a catechol type, which varies according to the age of the plant. For example, an acetone extract of the bark is to consist mostly of 3,4,7,3',4'- pentahydroxy flavan a compound which combines the properties of both leucoanthocyanidin and phlobotannin.[87] Khatri *et al.*, Hosamani K.and Banerjee B. studied the seed fatty acid profile and alkylated resin.[93-95] The isoflavonoid isolated from root extract, tested on female rats, showed dose dependent estrogenic activity by increasing uteri weight from 15.5±0.25 mg in control to 34.2±068 mg in orally treated rats (1.25 mg/ kg/day for 4 days).[95] In 1992, Kulkarni *et al.* studied the fatty acids distribution by Gas liquid chromatography in total lipids, phospholipids and glycolipids of this plant.[96] Total lipids from *P.dulce* seeds had a significant content of arachidic acid (16%) and the major fatty acids present are palmitic, stearic, oleic, linoleic , myristic , linolenic and arachidic acids. In 1994, Sahu *et al.* reported that two bisdesmodic triterpenoid saponins isolated from the seeds of *P. dulce* have anti-inflammatory activity.[97] In 1997, Nigam *et al.* reported seven saponins, termed pithedulosides A-G[98] and Yoshikawa *et al.* reported four oleanane-type triterpene glycosides, pithedulosides H-K[99] from *P. dulce* seeds. In 1999,

Niranjan *et al.* found a novel acylated triterpenoid saponin, designated pithecelloside, isolated from the seeds.[100] There are several studies of the antimicrobial activity of the seed extracts. For example, the powder, methanolic and aqueous extracts of *P. dulce* seeds have proved fungistatic and possess fungicidal effects against plant pathogens, *Fusarium oxysporum*, *Botrytis cinerea*, *Penicillium digitatum* and *Rhizopus stolonifer.*[101] The ^{1}H NMR spectrum of active fractions in these extracts indicated the presence of two triacyl glycerols, glycerol 1,3-dilinoleoyl-2-decanoic, glycerol 1-linoleoyl-2- docosanoic-3- olein. In 2003, Barrera *et al.* reported that several triterpene saponins, pitheduloside A, B, E, F and I inhibited in vitro mycelial growth of *Rhizopus stolonifer* and *colletotrichum gloeosporioides*, respectively.[102] Furthermore, the less-polar hexane extract and polar methanolic extract of this plant screened against various bacteria and fungi were also found to be active.[103] In 2005, Pithayanukul *et al.* tested polyphenols from the aqueous extract of *P. dulce* for their inhibitory activities against *Naja kaouthia* (NK) venom by in vitro neutralization method.[104] The extract could completely inhibit the lethality of the venom at 4 LD_{50} concentration and the venom necrotizing activity at the minimum necrotizing dose while also inhibited up to 90% of the acetylcholinesterase activity of NK venom at much lower tannin concentration. There has been suggested anti-venom activity of this plant polyphenols by selectively blocking the nicotinic acetylcholine receptor and non-selectively by precipitation of the venom proteins.

On the other hand, there have been a few protein studies from *P. dulce*. In 2004, Vargas *et al.* reported for the first time the isolation and characterization of a protease inhibitor from the seeds of *P. dulce*. The purification of the *P. dulce* trypsin inhibitor (PDTI) was a direct process. After its extraction (pH 8.0) and precipitation (80% $(NH_4)_2 SO_4$), the pH was adjusted to 4.0, the supernatant was loaded onto a CM-Sepharose column, and a single peak of trypsin inhibitory activity was eluted (CM-TIA). The main component of CM- TIA was PDTI, a protein composed of two polypeptide chains joined by disulfide bridge(s), with a p*I* of 4.95 and a molecular weight determined by electrospray mass spectrometry of 19614 Da. The N-terminal sequence of PDTI has the highest similarity with the seed inhibitor of *Acacia confusa*. Protein inhibitory activity has been registered for the seeds of several *Pithecellobium* spp.[106,107] Trypsin inhibitory activity in *P. keyense* was higher than that described for

soybean. The inhibitory activity identified in extracts of *P. dulce* seeds was against trypsin, chymotrypsin, and papain. From database searching, there has been no report of sequence protein from *P. dulce* found in SWISS-PROT and TrEMBL database (www.expasy.org).

Although, the existing chemical, pharmacological and clinical literature on the plant is impressive, the majority of protein components in *P. dulce* seeds and their biological activities are still unknown. There is a considerable lack of information of the protein from *P. dulce* seeds. The characterization and studies of protein from *P. dulce* seeds remain open to future investigation. Therefore, it is interesting to study the protein profile of *P. dulce* seeds and purification of the bioactive proteins from *P. dulce* seeds. This may serve as a starting point to better understand the property of *P. dulce* seed proteins.