PROTEIN IDENTIFICATION OF *ANOPHELES MACULATUS* GUTS USING MATRIX-ASSISTED LASER DESORPTION/IONIZATION – TIME OF FLIGHT MASS SPECTROMETRY

by

LIEW HOOI LING

**Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science (Pharmacy)**

OCTOBER 2009

PROTEIN IDENTIFICATION OF *ANOPHELES MACULATUS* GUTS USING MATRIX-ASSISTED LASER DESORPTION/IONIZATION – TIME OF FLIGHT MASS SPECTROMETRY

LIEW HOOI LING

UNIVERSITI SAINS MALAYSIA

OCTOBER 2009

ACKNOWLEDGEMENT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

.

LIST OF PLATES

LIST OF ABBREVIATIONS / SYMBOLS

ACN	Acetonitrile
ACTH	Adrenocorticotropic hormone
APS	Ammonium persulfate/ ammonium peroxodisulphate
ASMS	American Society for Mass Spectrometry
BSA	Bovine serum albumin
BSI	Bioinformatics Solutions Inc
CA	Carbonic anhydrase
CBT	Cow-baited trap
CDC	Center for Disease Control and Prevention
CID	Collision induced dissociation
CHAPS	3-[(3-cholamidopropy)-dimethylammonio]-1-propanesulfonate
CHCA	A-cyano-4-hydroxycinnamic acid
Da	Dalton
DDT	Dichloro-diphenyl-trichloroethane
DTT	Dithiothreitol
EEE	Eastern Equine Encephalitis
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionization
eV	ElectronVolt
IEF	Isoelectric focusing
IPG	Immobilized pH gradient
kDA	KiloDalton
km	Kilometre
kV	KiloVolt
LC	Liquid Chromatography
HPLC	High-Performance Liquid Chromatography
g	Gram
GHz	Gigahertz
GPI-linked	Glypiated
M	Molar
mA	MiliAmphere
MALDI	Matrix-Assisted Laser Desorption/Ionization
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization - Time-of-Flight
MALDI Q-TOF	Matrix-Assisted Laser Desorption/Ionization Quadrupole – Time-of-Flight
Mg	Milligram
Mg/mL	Microgram/microLitre
Ml	MiliLitre
Mm	Milimetre
mM	MiliMolar
mRNA	Messenger Ribonucleic Acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
M_r	Relative molecular mass
MW	Molecular weight
m/z	Mass-to-charge ratio
NH_4HCO_3	Ammonium bicarbonate

ng/µL	Nanogram/microLitre
nm	Nanometre
ns	Nanosecond
p	Probability
pI	Isoelectric point
PM	Pertrophic matrix
PMF	Peptide mass fingerprint
PMFs	Peptide mass fingerprints
Ppm	Part per million
RF	Radio frequency
RP-HPLC	Reverse Phase-High-Performance Liquid Chromatography
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
Tris-HCl	Trizma hydrochloride
TOF	Time-of-flight
µg	Microgram
µL	Microlitre
UV	Ultra-violet
V	Voltage
v/v	Volume/volume
WHO	World Health Organisation
WNV	West Nile Virus
w/v	Weight/volume
1D	One dimensional
2D	Two dimensional
2D-PAGE	Two dimensional- polyacrylamide gel electrophoresis
2D-SDS-PAGE	Two dimensional-sodium dodecyl sulphate-polyacrylamide gel electrophoresis
%	Percentage
%T	Total acrylamide contents
%C	Cross-linker content of the total acrylamide
˙C	Degree Celcius

Note: Three-letter code and one-letter code of the 20 amino acid residues can be referred at page 48 of the Materials and Methods (Chapter 3).

PENGENALPASTIAN PROTEIN BAGI USUS *ANOPHELES MACULATUS* MENGGUNAKAN "MATRIX-ASSISTED LASER DESORPTION/IONIZATION – TIME OF FLIGHT MASS SPECTROMETRY"

ABSTRAK

Merujuk kepada Pertubuhan Kesihatan Sedunia, malaria telah mengancam 300-500 juta manusia sedunia, di mana lebih daripada 90% kes berlaku di kawasan sub-sahara Afrika. *Anopheles maculatus* merupakan vektor utama terhadap penyakit ini di kawasan Selatan Thailand serta Semenanjung Malaysia. Parasit terpaksa melalui halangan yang terkandung di dalam usus nyamuk untuk terus berkembang dan berpindah. Maka, tesis ini menguraikan pendekatan proteomik untuk memberikan gambaran mengenai protein dalam usus *Anopheles maculatus* dari aspek biokimia dengan menggunakan gel elektroforesis dan penjujukan *de novo*. Lima komponen utama dengan berat molekul relative M_r 62,800, M_r 58,600, M_r 28,400, M_r 13,500, dan M_r 12,500 telah diperhatikan pada gel SDS-PAGE, dan kajian ini dijelaskan secara terperinci dengan menggunakan pendekatan mass spektrometri. Selain daripada rantai disulfida diperhatikan antara berat molekul M_r 28,400 dengan M_r 14,000, kehadiran berat monoisotopik dalam kedua-dua protein turut menjelaskan kemungkinan perwujudan hubungan monomer-dimer. Sebanyak 157 peptida telah berjaya dijujuk. Kebanyakan peptida-peptida dengan berat molekul M_r 62,800 dan M_r 58,600 yang dijujuk telah dikenalpasti berpadanan dengan Bovine Serum Albumin. Selain daripada BSA, protein seperti trypsin, 'carbonic anhydrase-2', endopin-1, hemoglobin sub-unit alfa dan beta turut dikenalpasti berpadanan dengan pengkalan data. Lebih kurang 67.1% penjujukan *de vono* peptida tidak berpadanan dengan pengkalan data. Dengan pendekatan mass spektometri bagi protein usus, bersamaan dengan jujukan peptida di dalam data genom *Anopheles maculatus*, tiada kemuskilan

untuk megetahui peranan protein ini terhadap penghadaman darah atau resektor parasit.

PROTEIN IDENTIFICATION OF THE *ANOPHELES MACULATUS* GUTS USING MATRIX-ASSISTED LASER DESORPTION/IONIZATION – TIME OF FLIGHT MASS SPECTROMETRY

ABSTRACT

According to the World Health Organization, malaria currently affects 300-500 million people worldwide, of which more than 90% of these cases occur in sub-Saharan Africa and are responsible for over one million deaths each year. *Anopheles maculatus* is the main vector for this disease in Southern Thailand and Peninsular Malaysia. The mosquito gut constitutes a barrier that the parasite must cross if it is to develop and be transmitted. Therefore, this thesis describes the proteomics approach used in the initial investigation work on the biochemical aspects of the gut proteins of *Anopheles maculatus* by gel electrophoresis and *de novo* sequencing. Five main groups of components with relative molecular masses of M_r 62,800, M_r 58,600, M_r 28,400, M_r 13,500, and M_r 12,500, as observed on SDS-PAGE, were subjected to further investigation by mass spectrometry. Apart from the disulphide-link observed for molecular masses of M_r 28,600 to M_r 14,000, the presence of monoisotopic masses in both proteins showed a possible monomer-dimer relationship. There were 157 peptides from the five highly abundant proteins that were successfully sequenced. Most of the peptides in the molecular masses M_r 62,800 & M_r 58,600 were significantly matched to the Bovine Serum Albumin (BSA). Beside BSA, they were five other significant matches to the existing database, which are trypsin, carbonic anhydrase-2, endopin-1, haemoglobin subunit alpha and beta. Approximately 67.1% of *de novo* sequenced peptides were found to be novel components with no match against the database. Mass spectrometry analysis of the proteins found in guts, together with data from the recent sequencing of the

Anopheles maculatus genome, should make it possible to determine the role of these

proteins in blood digestion or parasite receptivity.

CHAPTER 1

INTRODUCTION

Mosquito, a recognized vector, is linked to more human suffering than any other organisms, in which over one million people die from mosquito-borne diseases every year. According to the Center for Disease Control (Atlanta), malaria is a leading cause of premature mortality with around 5.3 million deaths annually, particularly in children under the age of five (WHO, 1996). The disease is most widespread in Africa, whereby a child is killed every 30 seconds (Florens *et al.*, 2002; Knell, 1991). The mosquito of the *Anopheles* genus carries the malaria parasite, *Plasmodium* and the latter infects human and insect hosts alternately. This parasite has to complete a complex life cycle in vector mosquitoes, starting in the midgut lumen, crossing through the midgut epithelial barrier and lastly invading the salivary gland from where they can be transmitted from one host to another during blood feeding (Prévot *et al.*, 1998; Shen *et al.*, 1999; Zieler *et al.*, 2001).

Malaria incidence can be reduced by controlling parasite transmission by the mosquito. The *Plasmodium* develops from gamete to oocyst in the lumen and epithelium of the mosquito midgut (Abraham and Jacobs-Lorena, 2004; Knell, 1991). Trypsin, produced in the mosquito's digestive tract, probably activates parasite chitinase(s), facilitating the route of the parasite through the peritrophic matrix (PM) surrounding the parasite containing blood meal in the mosquito (Warr *et al.,* 2007). Dimopoulos and coworkers (1998) found that early sporogonic stages of *Plasmodium* parasite alter the mosquito midgut immune response and also implied that this immune response molecule can be used to inhibit the development of the

parasites (Tahar *et al.,* 2002; Warr *et al.,* 2007). However, most of the studies were done based on the analyses of mRNA production, moreover limited publication can be found concerning proteome analysis for mosquito midguts.

Anopheles maculatus, a sun-loving insect, is the major vector in Northern Peninsular Malaysia yet there is little information on the population genetics of this mosquito as well as the protein composition of the mosquito midguts. In order for a better understanding towards the research background of this thesis, the next chapter will address the following topics leading to the objectives of this research:

- Mosquito; its definition
- Mosquito as vector of disease
- Definition of malaria
- Malaria transmission
- *Anopheles maculatus*; biology, morphology and distribution
- Research into proteomics-malaria
- Proteomics; its definition
- Overview of proteomics eksperiment workflow
- Sample preparation
- Protein separation
- Protein visualization
- Protein digestion
- Protein identification – mass spectrometry; ionization source; mass analyzer and ion detector
- Tandem mass spectrometry

- *De novo* sequencing

- Database search and protein discovery

1.1 Aim of the Studies

The midgut protein profile of female *Anopheles maculatus* will be analyzed by one-dimensional gel electrophoresis followed by mass spectrometry. Midguts will be isolated more than 12 hours after feeding on cattle blood. Subsequently, this study will lead to the identification of proteins that are specifically produced and regulated in females following blood ingestion. Therefore, the aims of the study are to:

- Characterize the initial protein profiles in the gut of *Anopheles maculatus* species following feeding of the mosquito.

- Identify and characterize proteins present in the gut of *Anopheles maculatus* using matrix-assisted laser desorption/ionization- time-of-flight mass spectrometry techniques.

CHAPTER 2

LITERATURE REVIEW

2.1 Mosquitoes

2.1.1 Definition of Mosquitoes

Mosquitoes also known as *skeeters* are classified in the Phylum of Arthropoda Class of Insecta, Order of Diptera and Family of Culicidae (Rao, 1984; Sandosham and Thomas, 1983). Mosquitoes vary in size and are often smaller than 0.6 inch or 15 mm. They are capable of travelling up to 10 km in a night and fly continuously from 1 hour up to 4 hours. Mosquitoes are nocturnal insects, which are active at night and shelter in shaded humid places during the day. The female adults of most species are recognized by their long proboscis, used for piercing and bloodsucking while male mosquitoes feed on flower nectar and other sweet plant juices (Clements, 1992).

2.1.2 Mosquito as Disease Vector

Among insects, mosquitoes are the most important group in the transmission of several human diseases including malaria, yellow fever, dengue fever, encephalitis, filariasis, elephantiasis and many more. Mosquito vectors transmit a pathogen (disease causing organism or agent) from one host to another either mechanically or biologically. A mechanical vector, also known as carrier will carry the pathogen in its body from one host to another before it develops or multiplies. However, a biological vector is infected by the pathogen that develops or multiplies inside the body of the vector (WHO, 1996). Mosquitoes are not only capable of carrying diseases that afflict humans; they also transmit several diseases and parasites to animals such as dogs and horses. These include dog heartworm, West

Nile Virus (WNV) and Eastern Equine Encephalitis (EEE) (WHO Geneva, 1996). In addition, mosquito bites can cause severe skin irritation through an allergic reaction to the mosquito's saliva and this is what causes the red bump and itching.

2.2 Epidemiology of Malaria

2.2.1 Definition of Malaria

The word malaria comes from the Italian word 'mal aria', which means 'bad air' because it was once thought that the disease came from fetid marshes. In 1880, scientists discovered the real cause of malaria is by protozoan parasites of the genus *Plasmodium* which the parasite is transmitted from one host to another through the bite of a female *Anopheles* mosquito (WHO, 1996). Malaria symptoms will only appear about 9 to 14 days after infection and this varies with different *Plasmodium* species. Typically, malaria causes fever, headache, vomiting, and other flu-like symptoms. It may become life-threatening if no treatments are administrated or sometimes when the parasites become resistant to the drugs which may lead to death. Malaria kills by infecting and destroying the red blood cells and by clogging the capillaries that carry blood to the brain and other organs (WHO Geneva, 1996).

2.2.2 Transmission of Malaria

Human malaria is transmitted by 4 *Plasmodium* species, which are *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, and *Plasmodium ovale* (Knell, 1991). These species behave differently in terms of the way they respond to drugs, in the mosquito phase, and once inside the human body. Despite the different species of Plasmodium, the life cycle of each species follows the same basic pathway (Prevot *et al*., 2003; Shahabuddin, 2002; Tahar *et al.*, 2002). The

human malaria parasite has a complex life cycle (Figure 2.1) that requires both a human host (carrier) and an insect host (Knell, 1991; Phillips, 1983). An infected female *Anopheles* bites a human and injects *Plasmodium* parasites in the form of sporozoites into the human blood. In humans, the parasites multiply dramatically, first in the liver cells then enter the human blood. If a female *Anopheles* feeds on this infected human, the parasites will multiply in the mosquito stomach wall then migrate through its body to infest the salivary glands. When the mosquito feeds again, it will inject parasite-containing saliva into another human and this cycle starts again (Florens *et al.*, 2002; Menard, 2000; Menard *et al.*, 1997; Rodriguez and Hernandez, 2004; Shahabuddin, 2002; Simonetti, 1996).
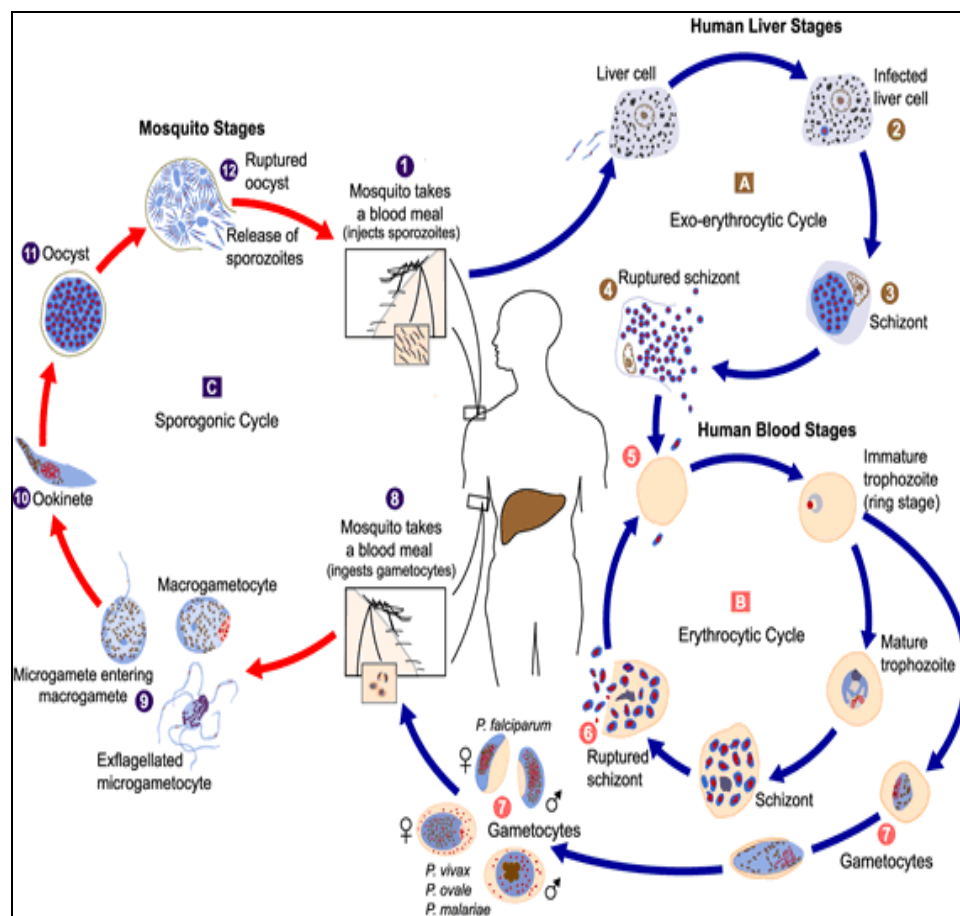
Figure 2.1: The Life Cycle of *Plasmodium.* The malaria parasite exhibits a complex life cycle involving an insect vector (mosquito) and a vertebrate host (human). The major phases of the life cycle are mosquito stages, human liver stages and human blood stages (Source: Centers for Disease Control and Prevention [CDC], 2004).

2.3 *Anopheles*-Malaria mosquitoes

2.3.1 Classification of *Anopheles*

The *Anopheles* mosquitoes belong to the phylum Arthropoda, which constitutes one of the largest divisions of the animal kingdom because their body are segmented and covered with a chitinous exoskeleton. The Anopheles mosquito is classed as an insect or hexapod in which its body is divided into the head, thorax, and abdomen. The mosquitoes are grouped under the order Diptera, known as the true flies but unlike the flies, its wings have scales.

The *Anopheles* mosquitoes are categorized under the suborder Nematocera, members of which have slender bodies with a pair of antennae. In the suborder of Nematocera, the *Anopheles* mosquitoes are placed in the family of Culicidae and subfamily Culicinae (Rao, 1984). The *Anopheles* mosquitoes are then characterized into the genus *Anopheles.* Thus, the classification of the *Anopheles* mosquitoes can be summarized as:

Phylum   : Arthropoda

Class   : Insects

Order   : Diptera

Suborder   : Nematocera

Family   : Culicidae

Subfamily   : Culicinae

Genus   : *Anopheles*

2.3.2 *Anopheles maculatus*

2.3.2.1 Biology and Morphology of *Anopheles maculatus*

Anopheles maculatus (Plate 2.1) is the main malaria vector in Indonesia, Southern Thailand and Peninsular Malaysia (Jaal, 1990; Reid, 1968). Although *Anopheles maculatus* generally feeds on cow and cattle's blood, it still shows much greater preference on human blood compared to the other anophelines. *Anopheles maculatus*, a nocturnal insect which generally feeds during night time ranging between 9 pm to midnight. A study on the biting activity of *Anopheles maculatus* has observed that this mosquito preferred in outdoor biting compared to indoor at a ratio of 4 to 1 (Hassan *et al.*, 2001).

The wings of *Anopheles maculatus* are dark spotted in which the front margin has at least 4 dark areas involving both the costa and first longitudinal vein of wing (Kay *et al.*, 1991; Oo, 2003). Palps of the female mosquitoes consist of 3 pale white bands, with 2 broad palps located near the tip and one narrow palp. As for the male, their palps are mostly white in colour, with pale stripes along the dorsal surface of the unexpanded portion.

The femora and tibia of the legs are speckled (Oo, 2003; Wattal *et al.*, 1960). Segment 5 of the hind tarsus is either white at the tip end or is completely white in appearance. The last tarsomeres of the hind legs are also white in color. The ventral surface of the abdomen is without tufts of black scales.

Plate 2.1: The Basic Morphology of *Anopheles maculatus* (Source: The Walter Reed Biosystematics Unit, [WRBU]).

2.3.2.2 Distribution of *Anopheles maculatus*

    Anopheles maculatus have different ecological distribution and behaviour pattern that explains why this mosquito is a vector in some areas and not in others. Its distribution extends from China and India to South-East Asia. It is widespread

throughout the East, (Rao, 1984) and is regarded as an important vector of human malarial parasites in some parts of Indonesia and Peninsular Malaysia. In Thailand, *Anopheles maculatus* is recognized as a vector only in the southern areas toward the Malaysia border, although it is distributed throughout the country. The global distribution of *Anopheles* mosquitoes are shown in Figure 2.2.

2.4 Research into Proteomics-Malaria

Malaria still kills up to a million African children each year. Today, malaria can be found throughout the tropical and sub-tropical regions of the world. It is of great interest that malaria mortality remains low in Malaysia. However maintenance of a conservative treatment policy and the strict regulation of private pharma-companies makes it very difficult to obtain alternative treatments (Roll Back Malaria Monitoring and Evaluation, 2005). Therefore, it is necessary to understand the function of the proteins encoded by these genes so that researchers all over the world will be able to elucidate their role in the mechanism of resistance to drugs. Recent studies showed that during the *Plasmodium* development in mosquito midgut, it displays proteins on the surface of the guts (Ghosh *et al*., 2000). Therefore, scientists are in a race to identify and characterize these proteins, which would benefit the coming generations where much effort is needed on developing malaria vaccines.

Many studies on *Anopheles gambiae* have been demonstrated by the researchers throughout the world because it is the principle vector of malaria in Africa. Snake venom, phospholipase A_2 extracted from an eastern diamondback rattlesnake successfully blocks the malaria parasite development in the mosquito midgut by inhibiting the oocyst formation (Zieler *et al*., 2001). Moreira and friends

(2002) also proved that phopholipase of bee venom inhibits malaria parasite development in transgenic mosquitoes. Previously, mouse monoclonal antibodies that recognize the circumsporozoite protein of *Plasmodium gallinaceum* can block sporozoite invasion of *Aedes aegypti* salivary glands (Capurro *et al*., 2000). In 2003, Prévot and coworkers carried out analysis of proteomics on sex-specific and blood-induced proteins of *Anopheles gambiae* midguts using two-dimensional gel electrophoresis and detected that the female midguts contained 23 proteins which are absent in the male mosquito. Therefore, they concluded that with the data obtained from the mass spectrometry analysis together with the data from the sequencing genome, will undoubtedly provide additional information as to whether these proteins are involved in blood digestion or development of the malaria parasites.

Apart from the advanced technologies for vaccine development, such as genome sequence analysis, microarrays, genomics, immunomics, high-throughput cloning and bioinformatics database tools, proteomics technologies also contributed in the malaria vaccine development. Recent studies of the infected human, mosquito vector and *Plasmodium* parasites, both in the areas of genomics and proteomics in parallel with the drug and vaccine development have the potential to reduce the morbidity of malaria disease. However, very little research has been done on *Anopheles maculatus* especially in the areas of proteome analysis. Hence, this research might provide an immensely valuable resource to understand the mosquito, parasite, and their interaction with the human host.

FIGURE 1. Global distribution (Robinson projection) of dominant or potentially important malaria vectors.
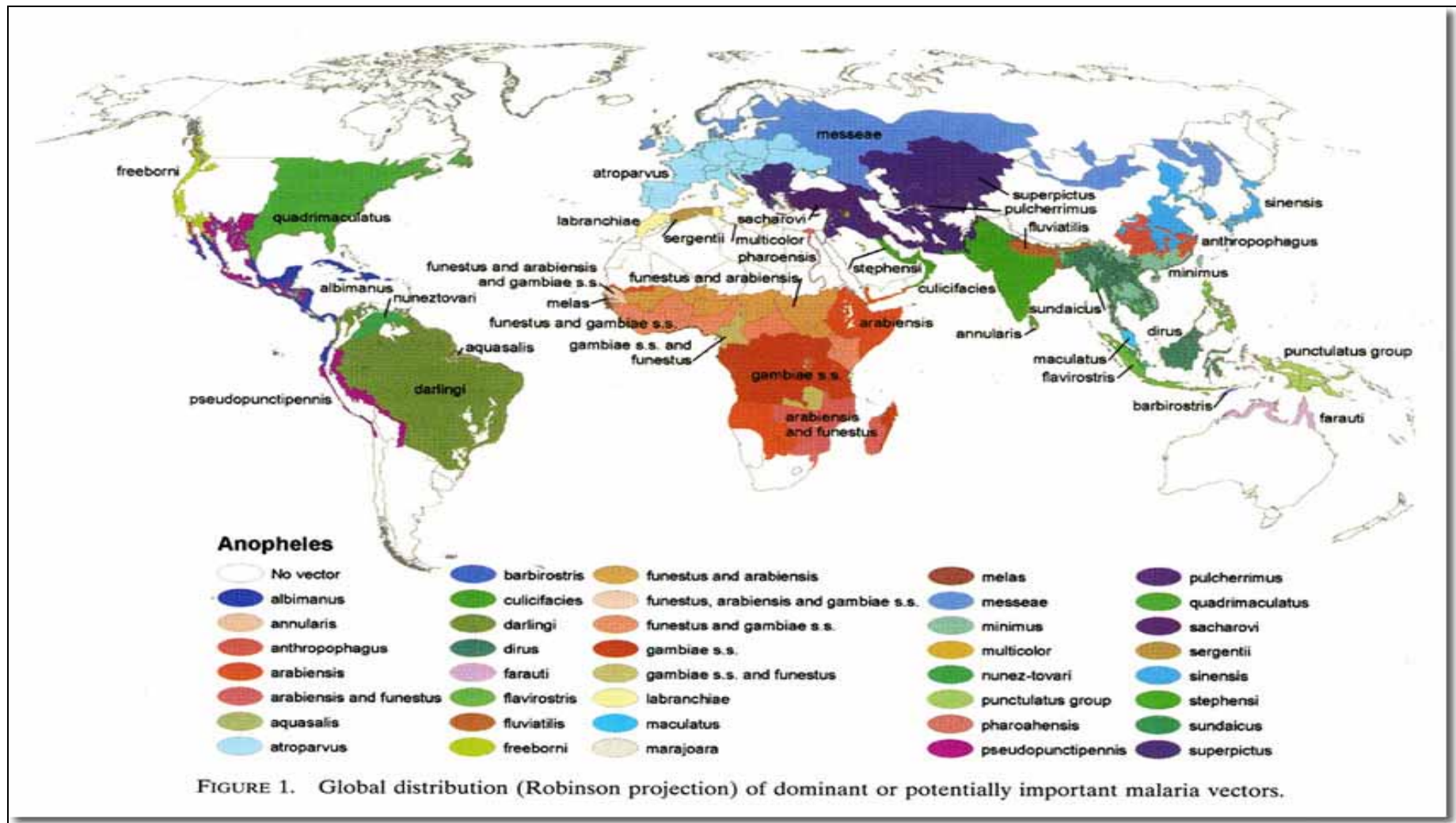
Figure 2.2: The worldwide distribution of Anopheles mosquitoes (Source: Centers for Disease Control and Prevention [CDC], 2004).

2.5 Proteomics

2.5.1 Definition of Proteomics

In 1994, Marc Wilkins was the first person to coin the term 'proteome' and 'proteomics', which refers to the study of the total set of proteins expressed in a specific cell at a specific time (Wilkins *et al*., 1996). The aim of 'proteomics' is to identify, characterize, and quantify all proteins involved in a particular pathway, organelle, cell, tissue, organ, or organism and that can be studied simultaneously in order to obtain accurate and comprehensive data (Graves and Haystead, 2002).

2.5.2 Overview of Proteomics Experiment Work Flow

The most common workflow in proteomics experiment and some of the factors affecting the way the experiment is performed are outlined in Figure 2.3.
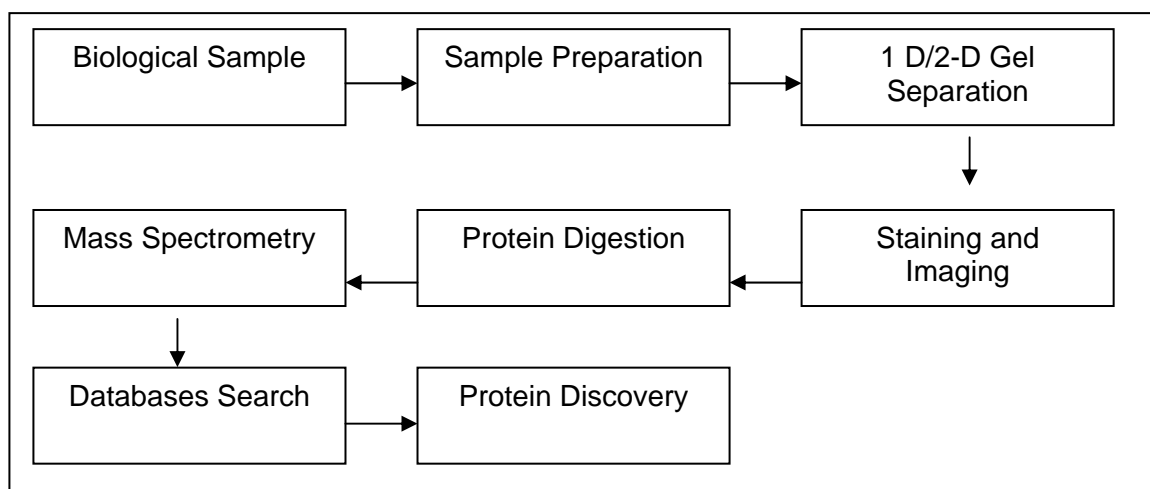
Figure 2.3: Schematic of workflow involved in a typical proteomics experiment.

Due to the great diversity of protein sample types and sources, an appropriate sample preparation method is required in which the samples are being treated prior to its analysis. Proteins are then separated according to their isoelectric focusing point (pI) and molecular weight through one-dimensional (1D) or two-dimentional (2D)

gels separation. For visualization of proteins in gels, a proper staining method is performed. Protein bands or spots of interest can be then excised and digested for mass spectrometry analysis. Apart from acquiring the peptide mass fingerprints of the proteins through mass spectrometry, once the tandem mass spectrum of peptides are generated, it will be used to match to the existing databases for protein identification and discovery.

2.5.3 Sample Preparation

The key to the success of the experiment depends on the method of sample preparation which in turn relies on the aim of the research. Different types of samples require different types of treatments and conditions. Factors such as solubility, size, charge, and isoelectric point (pI) of the proteins of interest will affect sample preparation. The preparation method varies from extraction with simple solubilization solutions to complex mixtures of reducing agents, detergents and chaotropic agents. Sample preparation is also important in reducing the complexity of a protein mixture (Lopez, 2004; McDonald and Yates, 2000).

An efficient sample preparation procedure will prevent protein aggregation and loss of solubility during focusing; will solubilize proteins including hydrophobic proteins; prevent post-extraction chemical modification including enzymatic or chemical degradation of the protein sample and thoroughly digest nucleic acids and other interfering molecules (Molloy *et al.,* 1998). The protein fraction, to be loaded on a one-dimensional (1D) or two-dimensional (2D) gel, must be solubilized in a denaturing buffer with low ionic strength that maintains the native charges of proteins and keeps them soluble.

2.5.4 Protein Separation

Peptide separation, which is achieved primarily by gel electrophoresis, is the first stage in the proteome research strategy. There are a number of electrophoresis methods, for instance one-dimensional (1D), two-dimensional (2D), 2D-polyacrylamide gel (2D-PAGE) and 2D-sodium dodecyl sulphate polyacrylamide gel (2D-SDS PAGE). This separation technique allows the immediate separation of thousands of proteins according to the different chemical properties (Jonsson, 2001; Kinter and Sherman, 2000; Weistermeier, 2001). In one-dimensional electrophoresis, proteins are separated based on the mass or molecular weight, whereby the proteins are dissolved in SDS which binds to the protein in a ratio of 1.4 g SDS per 1.0 g of protein. The negatively charged proteins will migrate across the gel towards the positive pole when an electric current is applied. Depending on their size, smaller proteins will migrate faster through the gel pores compared to the larger proteins (Weistermeier and Naven, 2002).

Two-dimensional electrophoresis is a powerful technique that can be used to resolve more than 10,000 proteins simultaneously in a highly reproducible. In this reported study, most post-translationally modified isoforms of proteins are readily differentiated using this technique (James, 2001). The weaknesses of 2D electrophoresis is its inability to deal with certain classes of proteins, such as highly hydrophobic proteins, very small proteins, and those with isoelectric points at either extreme of the pH scale (Kinter and Sherman, 2000; Weistermeier, 2001). Proteins are first separated by their isoelectric focusing (pI) in which the technique is known as isoelectric focusing (IEF). Isoelectric focusing point is defined as the pH at which a protein will not migrate in an electric field and is determined by the number and

types of charged groups in a protein. Thereby, positively charge proteins will be pulled towards the more negative end while the negatively charge proteins will be pulled to the more positive end (Garfin, 2000). The second dimension is then separated by protein molecular weight perpendicularly to the first dimension. Molecular weight can be determined in SDS-PAGE by comparing the migration of protein spots to the migration of standards (Weistermeier and Naven, 2002).

For protein separation, liquid chromatography (LC) is also commonly used. In liquid chromatography, proteins are dissolved in a liquid phase, and subsequently passed through one or more columns that separate the proteins on a number of dimensions. Separation on the basis of hydrophobicity using Reverse Phase - High-Performance Liquid Chromatography (RP-HPLC) technique is commonly employed in proteome research (Bakthiar and Tse, 2000). Liquid chromatography can be used for direct analysis of the samples, but increasingly, it is being used in tandem with 2D electrophoresis to further concentrate proteins, and to increase proteome coverage (Thomas, 2001).

2.5.5 Protein Visualization

In order to visualize proteins in gels, they must be stained in some manner. The choice of staining method is determined by several factors, including preferred sensitivity, linear range, ease of use, cost, and the type of imaging equipment available. Sometimes proteins are detected by immunoblotting, glycoprotein analysis or total protein stain after transferring to a membrane supported by Western blotting. Each type of protein stain has its own characteristics and limitations with regard to

the sensitivity of detection and the types of proteins that stain best. The sensitivity of staining can be best achieved by the following factors:

- Amount of stain that binds to the proteins;

- Intensity of the coloration;

- Difference in coloration between stained proteins and the residual background in the body of the gel (Weistermeier, 2001).

Coomassie brilliant blue staining and silver staining are the common staining methods used for in-gel protein detection.

2.5.6 Protein Digestion

Once proteins of interest are selected by differential analysis or other criteria, the proteins can be excised from gels and identified. Protein digestion with proteolytic enzymes, such as trypsin, is the most common method used for producing proteolytic peptides for mass spectrometry analysis. In-gel digestion and in-solution digestion are the two general methods of protein digestion carried out prior to mass spectrometry analysis. Throughout these digestions steps, it is important to avoid any contamination of the sample as well as the need to minimize the handling of the sample in order to obtain the optimum results (Jonsson, 2001; Weistermeier and Naven, 2002).

2.5.7 Protein Identification – Mass Spectrometry

The ability to determine the accurate molecular weight (MW) by mass spectrometry (MS) and to search databases for peptide mass matches has made protein identification possible. Mass spectrometry is a powerful analytical technique that is used to identify unknown compounds, to quantify known compounds, and to

elucidate the structure and chemical properties of molecules (ASMS, 2001; Gross, 2004). Such technologies has undergone significant improvements allowing for its application to amino acids, proteins, peptides, lipids, carbohydrates, DNA, drugs and many other biological relevant molecules (Siudzak, 2006). Developments of certain mass spectrometry technologies such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) have made it possible to ionize and analyze large biomolecules (Kinter and Sherman, 2000). Generally, a mass spectrometer measures the mass of a molecule by converting them into ions and sorting them via a stream of electrical fields according to the mass-to-charge ratio (*m/z*). Similarly, there are 3 basic components in all mass spectrometer (Figure 2.4):

- An ionization source that convert particles into ions;

- A mass analyzer that sorts ions according to their mass-to-charge ratio (*m/z*) by applying electromagnetic fields;

- An ion detector that measures the mass-to-charge ratio (*m/z*) and thus calculating the abundances of each ion present (Hoffmann and Stroobant, 2005; Kinter and Sherman, 2000; Siudzak, 2006).
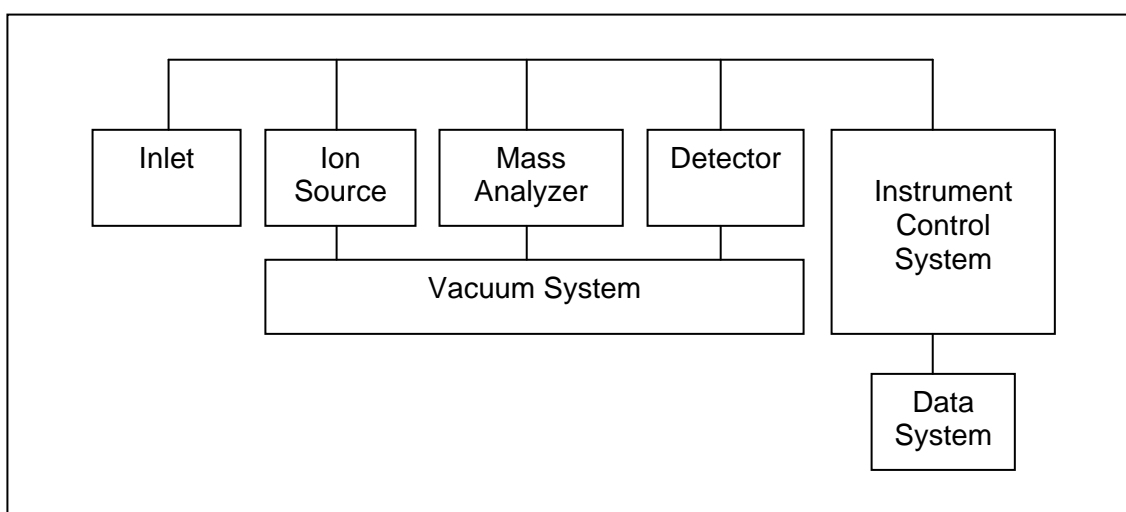
Figure 2.4: The basic components of a mass spectrometer. The mass analyzer, detector and parts of the ion source are maintained under vacuum.

2.5.7.1 Ionization Sources

The two most common ionization sources used for the majority of biochemical analyses are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) and these are described in more details in the following sections respectively.

2.5.7.1.1 Electrospray Ionization

The overall electrospray ionization (ESI) process can be divided into 2 steps; the initial production of charged droplets and the subsequent ion formation processes which occur largely as a result of droplet evaporation (Matsuo *et al*., 1994). During standard electrospray ionization, the sample is dissolved with a polar and volatile solvent and introduced into a capillary that ends in a very fine tip at a certain flow rate. A high voltage is applied to the tip of the capillary, creating a fine spray of highly charged droplets in the presence of an electric field. With the assistance of a flow of nitrogen gas, heat or both are applied to the droplets at atmospheric pressure thus causing the solvent to evaporate gradually, forcing the charges in the molecules within these droplets to be pushed closer together. Eventually, a 'Coulombic explosion' occurs and this cycle is repeated until the droplets are evaporated completely and decreased into smaller droplets up to the point where analyte ions are produced into the gas phase (Desiderio and Nibbering, 2001; Hoffmann and Stroobant, 2005; Kinter and Sherman, 2000; Siudzak, 2006). The electrospray ionization process is illustrated in Figure 2.5 and the advantages and disadvantages of ESI are shown in Table 2.1.
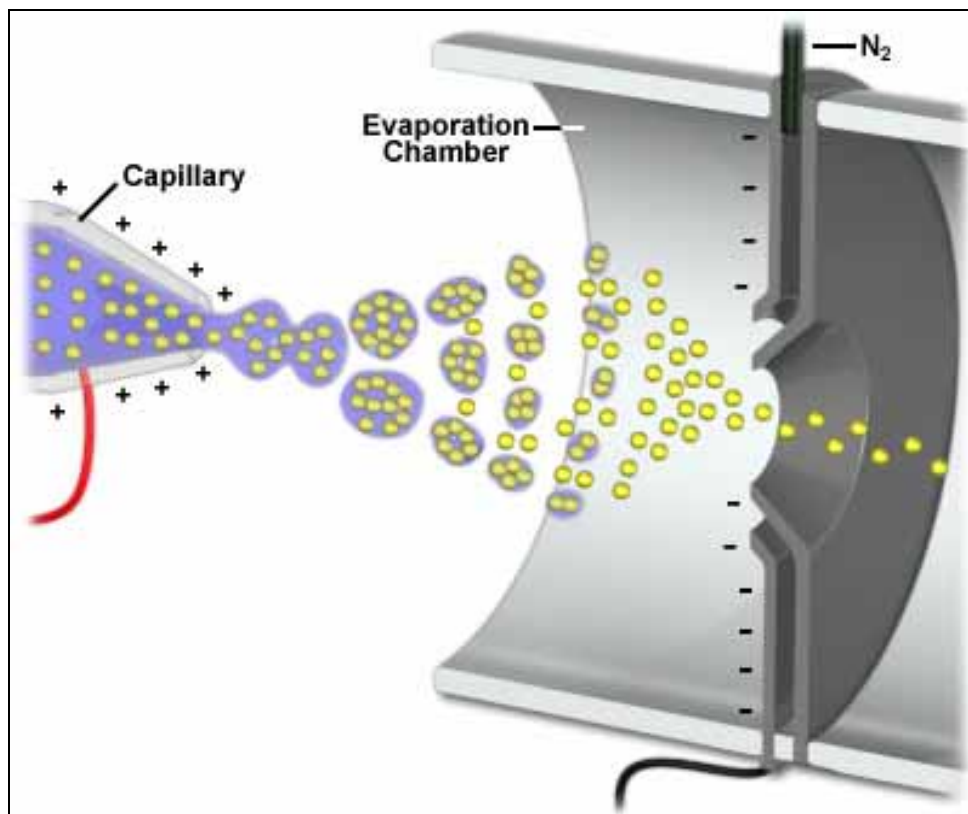
Figure 2.5: The generalized processes associated with electrospray ionization. Charged droplets are sputtered from a Taylor cone are reduced in size through a desolvation process that ultimately produces ions that enter the mass spectrometer (Source: National High Magnetic Field Laboratory, 1995).

Table 2.1: Advantages and disadvantages of electrospray ionization [ESI] (Source: Siudzak, 2006).

Advantages	Disadvantages
Practical mass range of up to 70,000 Da	Simultaneous mixture analysis can be poor
Good sensitivity with femtomole to low picomole	Complex mixtures can reduce sensitivity
Softest ionization method, capable of generating noncovalent complexes in the gas phase	The presence of salts and ion-pairing agents like TFA which reduce sensitivity
Easily adaptable to liquid chromatography	Carryover from sample to sample
Multiple charging allows for analysis of high mass ions with a relatively low m/z range instrument	Multiple charging can be confusing especially in mixture analysis
No matrix interference	Sample purity is important

2.5.7.1.2 Matrix-Assisted Laser Desorption/Ionization

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) was first introduced by Tanaka, Karas and Hillenkamp in 1988 (Hoffmann and Stroobant, 2005; Kinter and Sherman, 2000; Siudzak, 2006). MALDI is achieved in 2 steps; in the first step, the compound to be analyzed is mixed in solvent containing small organic molecules in solution (known as matrix) and the matrix has a strong absorption at the laser wavelength. The second step involves ablation of bulk portions of this dried mixture by intense pulses of UV laser, which then vaporize the matrix compound and produce a plume that carries the protonated peptide or protein into the gas phase (Splenger, 1997; Matsuo *et al.*, 1994). Once ions are formed in the gas phase, the desorbed charged molecules can be directed electrostatically from the MALDI ionization source to the mass analyzer. As illustrated in Figure 2.6, ionization on MALDI occurs by protonation in the acidic environments produced by the acidity of matrix compounds and by the addition of appropriate volume of dilute acid usually with a trace of trifluoroacetic acid to the samples. Because the laser desorption generates ions in discrete packets, MALDI is usually associated with time-of-flight (TOF) mass analysis (Desiderio and Nibbering, 2001; Kinter and Sherman, 2000; Siudzak, 2006; Spengler, 1997).

Sinapinic acid (SA) is a common matrix used for protein analysis; alpha-cyano-4 hydroxycinnamic acid (CHCA) is often used for peptide analysis while 2,5-dihydroxy benzoic acid (DHB) is often used for small proteins analysis. Apart from the tolerance of sample contamination, the matrix also serves to minimize sample damage from the laser pulse by absorbing most of the energy and increases the

efficiency of energy transfer from laser to the analyte. Table 2.2 shows the advantages and disadvantages of MALDI.
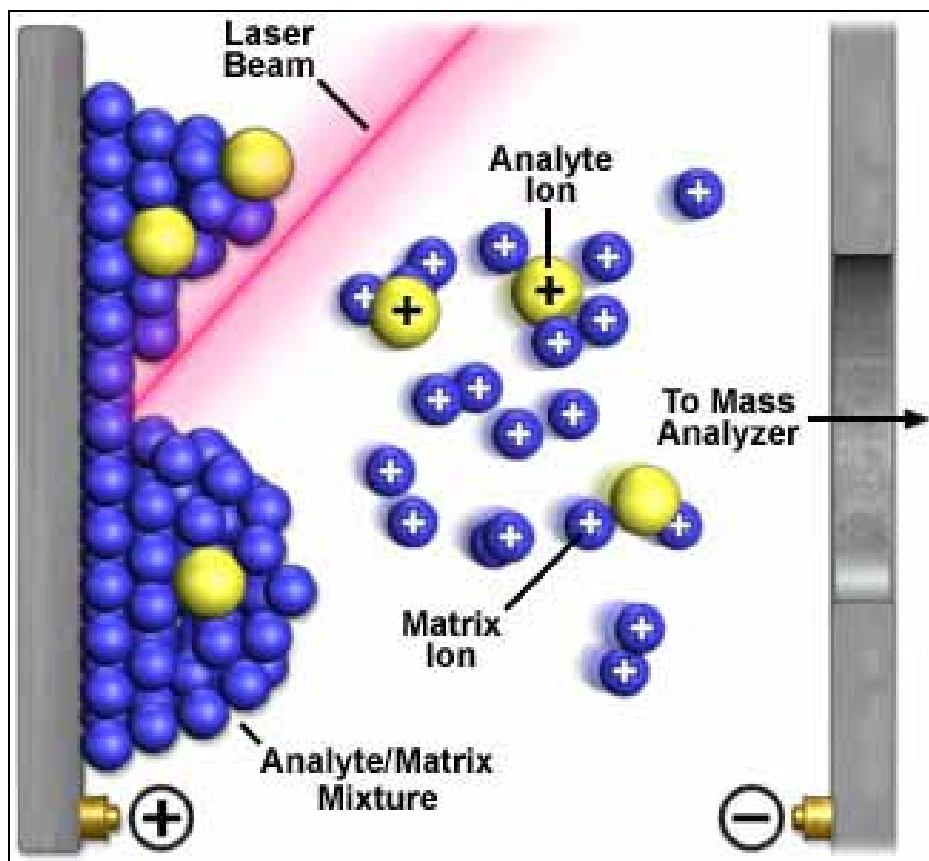
Figure 2.6: The generalized processes associated with matrix-assisted laser desorption/ionization. The analyte are co-crystallized with the matrix compound and are irradiated with UV-laser pulses. The laser pulses vaporize the matrix mixture and produce a plume that carries protonated peptide or protein into the gas phase before directed into the mass analyzer (Source: National High Magnetic Field Laboratory, 1995).

Table 2.2: Advantages and disadvantages of matrix-assisted laser desorption/ionization [MALDI] (Source: Siudzak, 2006).

Advantages	Disadvantages
Practical mass range of up to 300,000 Da	Matrix background can be a problem for compounds below a mass of 700 Da
Typical sensitivity with low femtomole to low picomole. Attomole sensitivity is possible	Possibility of photo-degradation by laser desorption/ionization
Suitable for the analysis of complex mixtures	Acidic matrix used may cause degradation on some compounds
Tolerance of salts in milimolar concentrations	

2.5.7.2 Mass Analyzers

The main function of the mass analyzer is to separate or sort the ions formed in the ionization source of the mass spectrometry according to their mass-to-charge ratio. Quadrupole, quadrupole ion trap, time-of-flight (TOF), fourier transform ion cyclotron resonance (FTICR) and others are the mass analyzers available. Quadropole mass analyzer consists of 4 parallel poles or rods, forming 2 sets of opposing rods connected electrically in which a radio frequency (RF) and DC potential are applied to each pair of rods to produce a complex oscillating movement of the ions as they move from the beginning of the mass filter to the end. This mass analyzer acts as a mass selective filter that only allow ions of a specific mass range to pass through the quadrupole at a specific RF field (Desiderio and Nibbering, 2001; Siudzak, 2006). As for quadrupole ion trap mass analyzer, the principles are similar to a quadrupole analyzer in which the ions of all masses are initially trapped or sequentially ejected and oscillated in a radio frequency quadrupole field, rather than passing through the analyzer (Gross, 2004; Hoffmann and Stroobant, 2005).

The operating principles of time-of-flight mass analyzers are simple in which all ions are accelerated by an electric field resulting in the ions entering a field-free region where they travel at a velocity that is inversely proportional to their mass-to-charge ratio (m/z). Hence, the time taken for the ions to reach the detector is measured and converted to mass-to-charge ratio. Because all ions are given the same amount of kinetic energy, heavier ions will need a longer time to reach the detector compared to the lighter ions due to their lower velocity and higher masses (Bakthiar and Tse, 2000; Siudzak, 2006). Other techniques like the fourier transform ion cyclotron resonance mass analyzers which determined the mass of ions based on the

cyclotron frequency of the ions in a certain magnetic field. The ions trapped in the analyzer cell are excited into a cyclotron orbit using a RF pulse (Marshall *et al.*, 1998).

These mass analyzers have various features, including the mass range, mass accuracy, scan speed, resolution as well as tandem analysis capabilities (Siudzak, 2006). Different analyzers can be combined with different ionization methods. For example, TOF analyzer is usually coupled with MALDI ion sources. In contrast, ion trap and quadrupole mass analyzers can be integrated with ESI ion sources (Kinter and Sherman, 2000).

2.5.7.3 Ion Detector

Once the ions are separated by the mass analyzer, the ions will reach the ion detector whereby the mass-to-charge ratio will be measured. A mass spectrum will be generated when a signal is produced in a detector through a mass spectrometry scanning. Beside the most commonly used detector which is the electron multiplier, detectors such as Faraday cup, ion-to-photon, array and many more are also used (Hoffmann and Stroobant, 2005; Matsuo *et al.*, 1994; Siudzak, 2006).

2.5.8 Tandem Mass Spectrometry

Generally, mass spectrometry can be used in 2 different phases for identification purposes, in which the mass-to-charge (*m/z*) ratios of each peptide is measured in the first phase. When peptides undergo mass spectrometry, they obtain one or more positive charge (z) based on a range of parameters used during the operation of the mass spectrometer and the presence of various chemical groups that